NEW ECONOMIC WINDOWS

F. Abergel • B.K. Chakrabarti
A. Chakraborti • M. Mitra (Eds.)

Econophysics of Order-driven Markets

Springer

Econophysics of Order-driven Markets

New Economic Windows

Series Editors
MARISA FAGGINI, MAURO GALLEGATI, ALAN KIRMAN

Series Editorial Board

Jaime Gil Aluja
Departament d'Economia i Organització d'Empreses, Universitat de Barcelona, Spain
Fortunato Arecchi
Dipartimento di Fisica, Università degli Studi di Firenze and INOA, Italy
David Colander
Department of Economics, Middlebury College, Middlebury, VT, USA
Richard H. Day
Department of Economics, University of Southern California, Los Angeles, USA
Steve Keen
School of Economics and Finance, University of Western Sydney, Australia
Marji Lines
Dipartimento di Scienze Statistiche, Università degli Studi di Udine, Italy
Thomas Lux
Department of Economics, University of Kiel, Germany
Alfredo Medio
Dipartimento di Scienze Statistiche,Università degli Studi di Udine, Italy
Paul Ormerod
Directors of Environment Business-Volterra Consulting, London, UK
Peter Richmond
School of Physics, Trinity College, Dublin 2, Ireland
J. Barkley Rosser
Department of Economics, James Madison University, Harrisonburg, VA, USA
Sorin Solomon Racah
Institute of Physics, The Hebrew University of Jerusalem, Israel
Pietro Terna
Dipartimento di Scienze Economiche e Finanziarie, Università degli Studi di Torino, Italy
Kumaraswamy (Vela) Velupillai
Department of Economics, National University of Ireland, Ireland
Nicolas Vriend
Department of Economics, Queen Mary University of London, UK
Lotfi Zadeh
Computer Science Division, University of California Berkeley, USA

Editorial Assistant:

Anna Parziale
Dipartimento di Studi Economici, Università degli Studi di Napoli "Parthenope", Italy

Frédéric Abergel · Bikas K. Chakrabarti
Anirban Chakraborti · Manipushpak Mitra
Editors

Econophysics of Order-driven Markets

Proceedings of Econophys-Kolkata V

Springer

Editors
Frédéric Abergel
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris
Châtenay-Malabry, France

Anirban Chakraborti
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris
Châtenay-Malabry, France

Bikas K. Chakrabarti
Centre for Applied Mathematics and Computational Science
Saha Institute of Nuclear Physics
Kolkata, India,
Economic Research Unit
Indian Statistical Institute
Kolkata, India

Manipushpak Mitra
Economic Research Unit
Indian Statistical Institute
Kolkata, India

New Economic Windows
ISSN print edition: 2039-411X
ISSN electronic edition: 2039-4128

ISBN 978-88-470-1765-8
e-ISBN 978-88-470-1766-5
DOI 10.1007/978-88-470-1766-5

Springer Milan Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010941697

© Springer-Verlag Italia 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher and the authors accept no legal responsibility for any damage caused by improper use of the instructions and programs contained in this book and the CD. Although the software has been tested with extreme care, errors in the software cannot be excluded.

Cover design: Simona Colombo, Milano
Typesetting: le-tex publishing services GmbH, Leipzig, Germany
Printing and binding: Grafiche Porpora, Segrate (MI)

Printed on acid-free paper

Springer-Verlag Italia S.r.l, Via Decembrio 28, I-20137 Milano
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The challenge met by the 5th Econophys-Kolkata conference held at the Saha Institute for Nuclear Physics in March 2010 was an interesting one. It was actually more a question than a challenge: to what extent can one depart from a zoomed-out point of view providing the usual stylized facts of financial markets and shift towards a small scale, microscopic type of approach, and still retain a conceptual view of the way financial markets work?

This shift of perspective was motivated by the necessity of addressing the case of general order-driven markets. These markets, which have set the standards for major financial markets such as the Equity, Forex, or Futures contracts markets, have led to the production of a huge amount of data, and an almost equally huge amount of scientific and econometric literature, owing to the seemingly perfect information that they provide to the interested researcher.

As a matter of fact, the paradigmatic case of european equity markets between (roughly speaking) the years 2000 and 2005 is almost a scientist's dream come true: any transaction on a given equity would take place in a single venue of a single exchange, the liquidity up to the five (or ten, sometimes twenty ...) first limits was public information accessible through the order book data, every event affecting the order book was time-stamped to the millisecond, and sometime even the identity of the sender of any given order was accessible.

This clearly transparent situation is still prevailing in some sense, although the existence of competing exchanges since the early 90's in the US and 2007 in Europe has made the situation a bit more difficult to apprehend. More to the point, the existence of dark pools, or hidden liquidity reservoirs operated by brokers for their "large" clients, is making this situation even more complex. However, one can still think that a data-based, in-depth study of the order-driven markets will lead to a good understanding of the mechanisms of price formation in financial markets.

And indeed, it is the case, at least up to a point: as we shall see from many contributions in the present volume, the mechanisms ruling short time scales microstructural effects, price impact, volatility clustering, relaxation of the order book following large trades ... are available for direct observation, and can be well described using either agent-based or purely statistical models. Of particular interest is the

question of whether the so-called "zero-intelligence" models may provide a faithful and useful representation of the dynamics of order-driven markets. One can find in this volume some pieces of empirical evidence against this, in particular regarding the "cat and mouse" game collectively played by liquidity takers and providers, and several other important effects must therefore be accounted for when designing more realistic models. Interestingly enough, there still exists, as a by-product of pure zero-intelligence, some sort of stochasticity in the order flow and therefore, in the price and spread volatility, due to memory effects in the order book. But these effects may sometimes hinder the researcher's path, as they play the role of an artificial source of volatility at the smallest time scales, rather than highlight some more fundamental behaviours of actual market participants.

In selecting the contributions for the present volume, the editors have tried and followed the intellectual approach of an experimental physicist: data should come first, in that they, and only they, can be considered a reliable ground for building up any kind of theories. Then modelling of course, follows. Models address one particular phenomenon that is observed and highlighted, they provide some information as to how one can explain and reproduce this phenomenon. Then equations are studied, both from a theoretical point of view and from that of numerical simulation. As a matter of fact, markets tend to behave in such a complicated manner that analytically tractable models, no matter how appealing they seem to the scientist, hardly provide a reasonable representation of reality. And the use of numerical simulations, particularly when one considers "intelligent" agents, is often the best – and sometimes, the only – way to understand and measure the microscopic and macroscopic impacts of such specific behaviours as market making, order splitting, order flow correlation between two tradable assets ...

These proceedings are organized as follows: a first section is concerned with the study of limit order books – data first, and then models. A second section is devoted to results on high frequency data and modelling. We have also included a "miscellaneous" section so as to incorporate relevant contributions from other areas of Econophysics. Finally, we have summarized in a brief "panel discussion" section some of the remarks made by the participants during the various interesting and animated exchanges that took place during the conference.

The editors would like to thank the Centre for Applied Mathematics and Computational Science of the Saha Institute for Nuclear Physics, the Indian Statistical Institute and the École Centrale Paris for their support in organizing this conference. They also address their warm thanks to Ioane Muni Toke, who not only contributed in a significant way to the scientific content of these proceedings, but also provided an invaluable help and support during the preparation of the manuscript.

We are grateful to the Editorial Board of the New Economic Windows series for agreeing again to the publish the Proceedings of Econophys-Kolkata V, in their esteemed series. The earlier Econophys-Kolkata proceedings volumes were: (i) Econophysics & Economics of Games, Social Choices and Quantitative Techniques, Eds. B. Basu, B.K. Chakrabarti, S.R. Chakravarty, K. Gangopadhyay, Springer-Verlag, Italia, Milan, 2010; (ii) Econophysics of Markets and Business Networks, Eds. A. Chatterjee, B.K. Chakrabarti, New Economic Windows, Springer-

Verlag Italia, Milan 2007; (iii) Econophysics of Stock and other Markets, Eds. A. Chatterjee, B.K. Chakrabarti, New Economic Windows, Springer-Verlag Italia, Milan 2006; (iv) Econphysics of Wealth Distributions, Eds. A. Chatterjee, S. Yarlagadda, B.K. Chakrabarti, New Economic Windows, Springer-Verlag Italia, Milan, 2005.

September 2010

Frédéric Abergel
Bikas K. Chakrabarti
Anirban Chakraborti
Manipushpak Mitra

Contents

List of Contributors xi

Part I Order Book Data and Modelling

Trade-throughs: Empirical Facts and Application to Lead-lag Measures 3
F. Pomponio, F. Abergel

Are the Trading Volume and the Number of Trades Distributions Universal? 17
V.S. Vijayaraghavan, S. Sinha

Subpenny Trading in US Equity Markets 31
R. Delassus, S. Tyč

"Market Making" in an Order Book Model and Its Impact on the Spread 49
I. Muni Toke

Price-Time Priority and Pro Rata Matching in an Order Book Model of Financial Markets 65
T. Preis

High-Frequency Simulations of an Order Book: a Two-scale Approach . . . 73
C.-A. Lehalle, O. Guéant, J. Razafinimanana

A Mathematical Approach to Order Book Modelling 93
F. Abergel, A. Jedidi

Reconstructing Agents' Strategies from Price Behavior 109
V. Alfi, M. Cristelli, L. Pietronero, A. Zaccaria

Market Influence and Order Book Strategies 125
F. Ghoulmié

Multi-Agent Order Book Simulation: Mono- and Multi-Asset High-Frequency Market Making Strategies 139
L. Foata, M. Vidhamali, F. Abergel

Part II High-Frequency Data and Modelling

The Nature of Price Returns During Periods of High Market Activity 155
K. Al Dayri, E. Bacry, J.F. Muzy

Tick Size and Price Diffusion 173
G. La Spada, J. Doyne Farmer, F. Lillo

High Frequency Correlation Modelling 189
N. Huth, F. Abergel

The Model with Uncertainty Zones for Ultra High Frequency Prices and Durations: Applications to Statistical Estimation and Mathematical Finance 203
C.Y. Robert, M. Rosenbaum

Exponential Resilience and Decay of Market Impact 225
J. Gatheral, A. Schied, A. Slynko

Part III Miscellaneous

Modeling the Non-Markovian, Non-stationary Scaling Dynamics of Financial Markets 239
F. Baldovin, D. Bovina, F. Camana, A.L. Stella

The von Neumann–Morgenstern Utility Functions with Constant Risk Aversions 253
S.R. Chakravarty, D. Chakrabarti

Income and Expenditure Distribution. A Comparative Analysis 259
K. Gangopadhyay, B. Basu

Two Agent Allocation Problems and the First Best 271
M. Mitra

Opinion Formation in a Heterogenous Society 277
M.-T. Wolfram

Opinion Formation in the Kinetic Exchange Models 289
A. Chakraborti, B.K. Chakrabarti

Panel Discussion 305

List of Contributors

Frédéric Abergel
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
frederic.abergel@ecp.fr

Khalil Al Dayri
CMAP, École Polytechnique
khalil.aldayri@polytechnique.fr

Valentina Alfi
Centro Studi e Ricerche "E. Fermi", Rome
Department of Physics
University of Rome La Sapienza, Rome
valentina.alfi@roma1.infn.it

Emmanuel Bacry
CMAP, École Polytechnique
emmanuel.bacry@polytechnique.fr

Fulvio Baldovin
Department of Physics, INFN e CNISM
University of Padua, Padua, Italy
baldovin@pd.infn.it

Banasri Basu
Physics and Applied Mathematics Unit
Indian Statistical Institute, Kolkata, India
banasri@isical.ac.in

Dario Bovina
Department of Physics, INFN e CNISM
University of Padua, Padua, Italy
bovina@pd.infn.it

Francesco Camana
Department of Physics, INFN e CNISM
University of Padua, Padua, Italy
camana@pd.infn.it

Bikas K. Chakrabarti
Centre for Applied Mathematics and Computational Science
Saha Institute of Nuclear Physics, Kolkata, India
Economic Research Unit, Indian Statistical Institute, Kolkata, India
bikask.chakrabarti@saha.ac.in

Debkumar Chakrabarti
Ramakrishna Mission Vidyamandira, Howrah, India
debkchak@yahoo.com

Anirban Chakraborti
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
anirban.chakraborti@ecp.fr

Satya R. Chakravarty
Indian Statistical Institute, Kolkata, India
satya@isical.ac.in

Matthieu Cristelli
Istituto dei Sistemi Complessi, CNR, Rome
Department of Physics
University of Rome La Sapienza, Rome
matthieu.cristelli@roma1.infn.it

Romain Delassus
École Polytechnique – BNP Paribas, Paris
romain.delassus@polytechnique.org

J. Doyne Farmer
Santa Fe Institute, Santa Fe, USA
jdf@santafe.edu

Laurent Foata
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
laurent.foata@student.ecp.fr

Kausik Gangopadhyay
Indian Institute of Management Kozhikode, Kerala, India
kausik@iimk.ac.in

Jim Gatheral
Department of Mathematics
Baruch College, CUNY, New York
jim.gatheral@baruch.cuny.edu

François Ghoulmié
Académie de Versailles, Paris
francois.Ghoulmie@gmail.com

Olivier Guéant
MFG R&D, Paris, France
olivier.gueant@post.harvard.edu

Nicolas Huth
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
Natixis CIB, Paris
nicolas.huth@ecp.fr

Aymen Jedidi
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
aymen.jedidi@ecp.fr

Gabriele La Spada
Department of Economics, Princeton University, Princeton, USA
Department of Economic Sciences, LUISS, Rome
gla@princeton.edu

Charles-Albert Lehalle
Head of Quantitative Research
Crédit Agricole Cheuvreux, Courbevoie, France
clehalle@cheuvreux.com

Fabrizio Lillo
Department of Physics and Technologies
University of Palermo, Palermo, Italy
Santa Fe Institute, Santa Fe, USA
lillo@unipa.it

Manipushpak Mitra
ERU – Indian Statistical Institute, Kolkata, India
mmitra@isical.ac.in

Jean-François Muzy
CNRS-UMR 6134, Université de Corse, Corte, France
muzy@univ-corse.fr

Fabrizio Pomponio
BNP-Paribas, Equity Derivatives Quantitative Analytics
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
fabrizio.pomponio@bnpparibas.com

Tobias Preis
Center for Polymer Studies, Department of Physics, Boston, USA
Artemis Capital Asset Management GmbH, Holzheim, Germany
Institute of Physics, Johannes Gutenberg University Mainz, Germany
mail@tobiaspreis.de

Luciano Pietronero
Department of Physics, University of Rome La Sapienza, Rome
Istituto dei Sistemi Complessi, CNR, Rome
luciano.pietronero@roma1.infn.it

Julien Razafinimanana
Crédit Agricole Quantitative Research, Courbevoie, France
jrazafinimanana@cheuvreux.com

Christian Y. Robert
CREST – ENSAE Paris Tech, Paris
chrobert@ensae.fr

Mathieu Rosenbaum
CMAP, École Polytechnique
mathieu.rosenbaum@polytechnique.edu

Alexander Schied
Department of Mathematics
Mannheim University, Mannheim, Germany
schied@uni-mannheim.de

Sitabhra Sinha
The Institute of Mathematical Sciences
C.I.T. Campus, Chennai, India
sitabhra@imsc.res.in

Alla Slynko
Department of Mathematics
Mannheim University, Mannheim, Germany
alla.a.slynko@gmail.com

Attilio L. Stella
Department of Physics, INFN e CNISM
University of Padua, Padua, Italy
stella@pd.infn.it

Ioane Muni Toke
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
ioane.muni-toke@ecp.fr

Stéphane Tyč
BNP-Paribas, Head of Equity Derivatives Quantitative Analytics
stephane.tyc@bnpparibas.com

Michael Vidhamali
Chair of Quantitative Finance
Laboratory of Mathematics Applied to Systems
École Centrale Paris, Châtenay-Malabry, France
michael.vidhamali@student.ecp.fr

Vikram S. Vijayaraghavan
Department of Physics
Indian Institute of Technology Madras, Chennai, India
vikramsv@smail.iitm.ac.in

Marie-Therese Wolfram
Faculty of Mathematics
University of Vienna, Vienna, Austria
marie-therese.wolfram@univie.ac.at

Andrea Zaccaria
Department of Physics, University of Rome La Sapienza, Rome
Istituto dei Sistemi Complessi, CNR, Rome
andrea.zaccaria@roma1.infn.it

Part I
Order Book Data and Modelling

Trade-throughs: Empirical Facts and Application to Lead-lag Measures

Fabrizio Pomponio and Frédéric Abergel

Abstract. Order splitting is a well-known behavior in trading: traders constantly scan the limit order book and choose to limit the size of their orders to the quantity available at the best limit. Order splitting allows traders not to reveal their intention to the market so as not to move too much the price against them. In this note, we focus on the other trades, called trade-throughs, which are trades that go through the best available price in the order book. We provide various statistics on trade-throughs: their liquidity, their intraday distribution and the spread relaxation that follows them. We also present a new method to get empirical distributions of lead-lag parameters between assets, sectors or even markets. This empirical study is based on tick-by-tick data of major EU and US equity futures from TRTH (Thomson Reuters Tick History) database.

1 Introduction

We know from Bouchaud [5] that the trading flow time series is a long-memory process.[1] For example, if we look at the following figures, we see that the trade sign autocorrelation for the French stock BNP-Paribas is approximately fitted by a power-law of exponent 0.5 (which is a typical kind of long-memory process).

Fabrizio Pomponio
BNP-Paribas, Equity Derivatives Quantitative Analytics, Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
e-mail: fabrizio.pomponio@bnpparibas.com

Frédéric Abergel
Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
e-mail: frederic.abergel@ecp.fr

[1] A centered process X is said to exhibit long-memory behavior when its series of auto-covariances is not summable, i.e. $\sum_{h\in\mathbb{N}} |\gamma(h)| = +\infty$ where $\gamma(h) = E[X_t X_{t-h}]$.

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

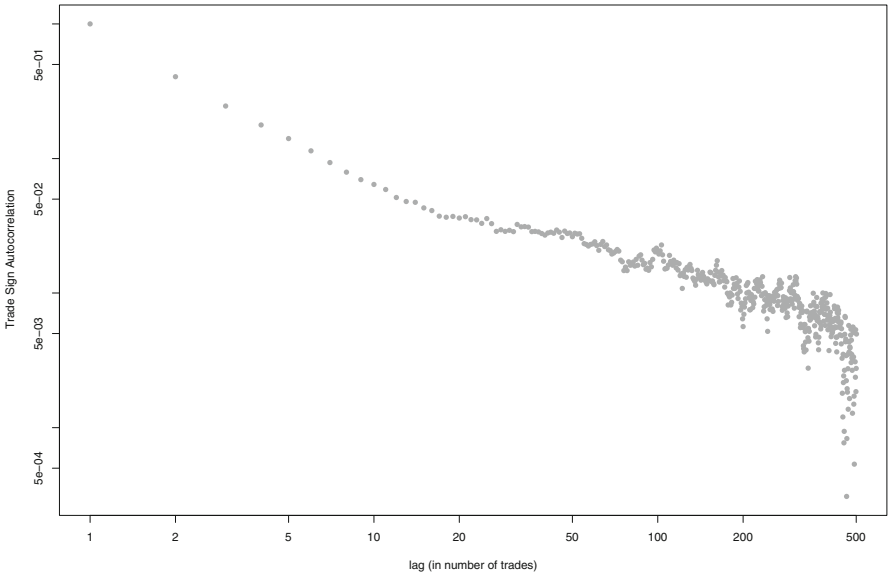


Fig. 1 Trade sign autocorrelation for BNP-Paribas stock on March–April 2010

Lillo [4] argues this is mainly explained by the splitting of large orders. Traders don't want to reveal their intentions to the markets so the price will not move too much against them. Assume a trader wants to trade a large order, he will then split it in several orders to minimize his price impact. In practice, traders constantly scan the limit order book and mainly restrict the size of their orders to the quantity available at the best limit. Conversely, sometimes speed of execution is more important than minimizing market-impact. It may be interesting to study trade-throughs which are the trades that stand outside this usual pattern of limiting trade size to the best limit liquidity.

In Sect. 2, we give a general insight about TRTH's data we used and how they are processed at BNP-Paribas. We also present the data we focused on (major US and EU equity futures and major French stocks).

In Sect. 3, we first look at the fraction of the trading volume that is taken from each limit of the order book. We precisely define what trade-throughs are and then provide various statistics[2] on them: their occurrence and volume proportions (we also test the stability of the definition of trade-thoughs with respect to the mean trade volume). We also provide the intraday timestamp distribution of trade-throughs, on which we highlight an important peak for equity futures at the same moment of the day as the release-time of major macro-economic news. This is a first confirmation of their higher informational content.

In Sect. 4, we focus on the spread behavior after a trade-through. We use the framework of complex systems to model the limit order book and consider trade-

[2] All the statistical computations are made by using the free statistical software R, available at http://cran.r-project.org.

throughs as excitations of this complex system. We show that there is a typical power-law relaxation of the excess spread after trade-throughs in physical time.

In Sect. 5, we use trade-throughs to obtain empirical distributions of lead-lag parameter. We apply this technique to the global lead-lag between US and EU equity markets and show that US equity markets are globally leading EU equity markets.

Sect. 6 summarizes the note and presents further research.

The advantages of this general framework based on trade-throughs are:

- We are restricting our set of data to extract very informative events. This simplifies some usual statistics but, more importantly, this allows one to do new statistics, like lead-lag empirical distributions, that are not possible using all tick-by-tick data.
- The framework of trade-throughs is flexible enough to adapt the rare events we are focusing on to the data we are looking at and to the goals we are achieving (restriction on the limit number touched by the trades, on the size of the trades, ...).
- We can study individual asset (stocks, futures, options, ...) but also global sectors, markets or economies.

2 Data Presentation

2.1 General TRTH Data Presentation and BNP-Paribas's Processing

The data used for this study come from TRTH (Thomson Reuters Tick History). The informations stored in the database include quotes (grouped in the 'Quotes' file) and trades (grouped in the 'Time And Sales' file). Both quotes and trades are timestamped in milliseconds by Thomson-Reuters's timestamping machines located in London. Quotes entries are composed of Bid/Ask/BidSize/AskSize. Trades entries contain Price/Volume of each transaction.

An important point to be mentioned in the data presentation is that TRTH data are flagged. Each entry of both quotes and trades files has a flag indicating information to be taken into account in the data analysis (e.g. this line from the 'Time And Sales' file is a trade from the auction phase, ...). Those flags are markets and exchanges dependent in the sense that specific knowledge from each market and exchange is necessary to correctly interpret each TRTH/exchange flag. After this flag processing[3], we end up with trades tagged within a limited number of flags' categories. The most important of trades flags (normal, auction, OTC, offbook, block trade, rck, market closed, cancelled, late0day & lateNdays, late report) are detailed in the next figure.

Last, some data sent by exchanges are corrections of previous entries. A basic example of those correction may be the cancellation of a previous trade that is to be re-

[3] The flag processing is done by BNP-Paribas Equities & Derivatives Quantitative R&D Histo team. It is an ongoing process that seems to be quite reliable on the most important exchanges.

Flag	Information and signification
normal	Trades occurring during the continuous trading session.
auction	Trades occurring at the end of the auction phase.
OTC	Over-the-counter trades occurring directly between two parties. They are opposed to trades occurring on centralized exchanges.
off book	Trades occurring outside the usual trading system of the considered exchange. They may be trades reported from a broker, for example. In Euronext, trades outside of the NSC (Nouveau Système de Quotation) fall inside this category.
block trade	Reporting of block trades (very large trades).
rck	TRTH/exchange's threshold break alerting that price and/or volume of this trade (as reported by the exchange) seem too different from an usual behavior and should not be considered as relevant.
market closed	Trades occuring before or after the regular trading session.
cancelled	Cancelled trades.
late 0 day & late N days	Trades that were reported the same day or N days after.
late report	Trades reported later in market data feeds. They include late0day and lateNdays trades.
unknown	All the other trades.

Fig. 2 TRTH's most important trades flags

placed with another one. Corrections are the only case when TRTH's data are modified at BNP-Paribas before being accessible to users. So, the tick-by-tick markets data we used in this study are the data after all corrections have been taken into account.

2.2 Data Used in This Study

Throughout the article, we only considered trades flagged as 'normal' trades. In particular, we did not consider any block-trade or off-book trade in the following. We explain here how we selected the perimeter of the data used for the statistics and lead-lag empirical determination between US and EU equity markets. First, we wanted to select a few instruments among the most representative of the US and EU equity markets. To this end, we ranked all equity financial instruments available in TRTH according to their ADV (Average Daily Volume). Then, we picked the most liquid ones (3 from US equity markets and 3 from EU equity markets). By doing so, we ended up with a minimal number of financial instruments that represent the most liquid instruments of the markets we want to study. This choice is based on the intuition that market moves are first expressed in the most liquid instruments.

The final financial instruments perimeter is composed of E-mini S&P500, Nasdaq E-mini and Dow Jones E-mini futures (for US equity markets) and of Eurostoxx,

DAX and Footsie futures (for European equity markets). Moreover, in order to consider the most liquid instruments, we chose to focus on the futures with the nearest maturity.

For the basic statistics on occurrences and volumes of trade-throughs (Sect. 2 of this article), we only considered data of March 2010 (from 16 to 21.30, Paris Time reference). For the lead-lag study (Sect. 5 of this article), we used data from the beginning of December 2009 to mid-March 2010 and restricted our data time-frame to the period of the day when both EU and US equity markets are opened and are widely trading (from 15.30 to 17.30, Paris Time reference) because we think that this lead-lag phenomenon is particularly relevant at that moment.

3 Some Statistics on Trade-throughs

Usually, traders scan the limit order book and restrict the size of their orders to the available liquidity. They split a large order in several orders in order to restrict the size of their trades to the dynamic quantity available at the best limit. As mentioned in the introduction, we are interested in the trades that deviates from this usual behavior. We will focus on the trades that consume the liquidity available in the order book in an aggressive way: the trade-throughs.

3.1 Definition of Trade-throughs

We call a x-th limit trade-through any trade that consumes at least one share at the x-th limit available in the order book. This definition is inclusive in the sense that, if p is greater than q, any p-th limit trade-through is also part of the q-th limit trade-throughs. For example, a 2nd-limit trade-through completely consumes the first limit available and begins to consume the second limit of the order book. In the following figures, we show an example of such a trade.

3.2 Basic Statistics on Occurrences and Volumes of Trade-throughs

We present in this section basic statistics on occurrences and volumes of trade-throughs in order to better measure the significance of this phenomenon on trading.[4]

[4] We recall that those statistics are done excluding the first and last half-hours of the usual trading session not to be impacted by auction phases. As we'll see in the section on intraday timestamp distribution of trade-throughs, those are parts of the day where trade-throughs also occur. So, on a global daily basis, following statistics are in fact underestimations. However, we do believe that they help to better understand the links between trade-throughs and both relative tick value of the asset and liquidity in the order book.

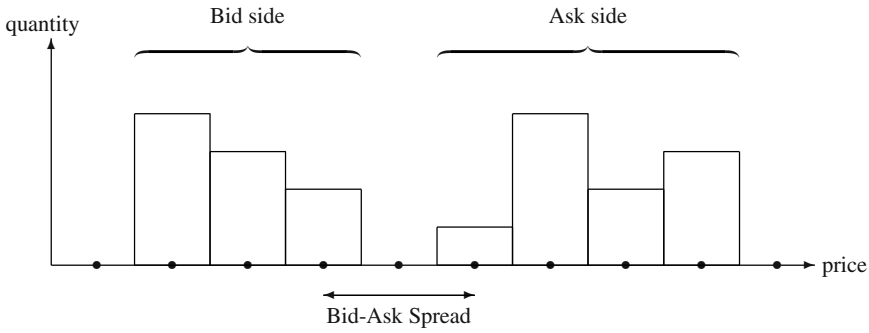


Fig. 3 Initial limit order book configuration

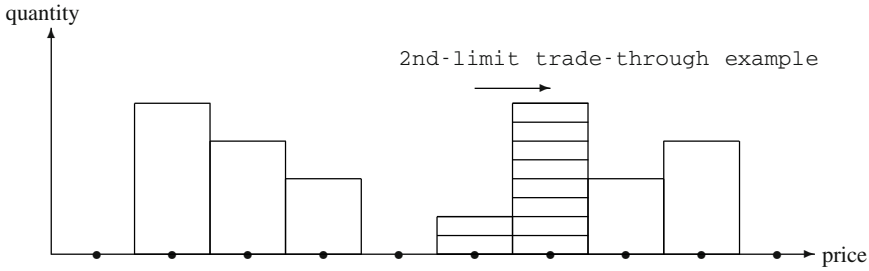


Fig. 4 2nd-limit trade-through example

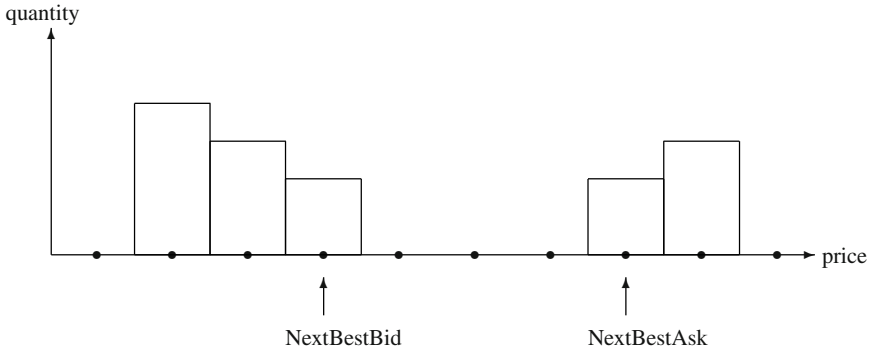


Fig. 5 After-trade-through limit order book configuration

We can notice that even if trade-throughs may rarely occur (with a probability occurrence of less than 5%), they include a non-negligible part of the daily-volume (up to 20% for the DAX index future).

An important remark should be made at this stage: the smaller the relative tick value, the more the trade-throughs are present, both in occurrence and volume. This result seems intuitive as we know that the smaller the relative tick value is on a particular asset, the more aggressively this asset is traded.

Financial asset considered	**2nd-limit TT** Occurrence (in %)	**2nd-limit TT** Volume (in %)	**Relative tick value** (indicative, in bp)
E-mini S&P500 – ES@	1,43	2,66	2,2
Nasdaq E-mini – NQ@	1,91	5,73	1,3
Dow Jones E-mini – YM@	3,17	9,65	1,0
Eurostoxx – STXE@	1,88	7,37	3,5
Footsee – FFI@	3,30	8,95	0,9
DAX – FDX@	7,75	19,10	0,8

Fig. 6 Basic statistics on 2nd-limit trade-throughs (estimations based on March 2010 data)

Financial asset considered	**3rd-limit TT** Occurrence (in %)	**3rd-limit TT** Volume (in %)	**Relative tick value** (indicative, in bp)
E-mini S&P500 – ES@	0,0012	0,0067	2,2
Nasdaq E-mini – NQ@	0,014	0,26	1,3
Dow Jones E-mini – YM@	0,062	0,79	1,0
Eurostoxx – STXE@	0,012	0,31	3,5
Footsie – FFI@	0,17	1,52	0,9
DAX – FDX@	0,50	3,28	0,8

Fig. 7 Basic statistics on 3rd-limit trade-throughs (estimations based on March 2010 data)

3.3 Stability of Trade-through's Definition with the Size of the Trade

In this section, we want to check that our definition of trade-throughs is stable with the volume of the trade. In other words, we want to verify that the trades we are focusing on are not due to little quantity available at the best limit of the order book (in which case the trader is almost forced to consume several limits in order to obtain the quantity he's usually looking for). To do so, we first compute the Mean Trade Volume of the usual trades (the trades that only consume less than the quantity available at the first limit of the order book). Then, we will check whether the statistical set of data we defined as trade-throughs is relatively stable when we add a second restriction to their definition, which is they have to consume more than the Mean Trade Volume.

For the US and EU futures, if we look at the 3rd-limit trade-through, there is no impact (both on occurrence and volume) on the set of trade-throughs if we add the previous restriction on trade volume. For the 2nd-limit trade-through, global volumes of trade-throughs is not really changed when restricting trade-throughs definition with the additional constraint on volume. But there exists a difference in occurrences, which reflects that a fraction of 2nd-limt trade-throughs for EU and US futures corresponds to trades split over the first two limits of the order book because of an unusual lack of liquidity on the first limit.

Financial asset considered	**2nd-limit TT** Occurrence (in %)	**2nd-limit TT** Volume (in %)	**Relative tick value** (indicative, in bp)
E-mini S&P500 – ES@	0,42	2,3	2,2
Nasdaq E-mini – NQ@	1,19	5,73	1,3
Dow Jones E-mini – YM@	2,19	8,95	1,0
Eurostoxx – STXE@	0,80	7,04	3,5
Footsee – FFI@	1,97	7,94	0,9
DAX – FDX@	5,53	17,7	0,8

Fig. 8 Basic statistics on 2nd-limit trade-throughs with volumes higher than the Mean Trade Volume (estimations based on March 2010 data)

Financial asset considered	**3rd-limit TT** Occurrence (in %)	**3rd-limit TT** Volume (in %)	**Relative tick value** (indicative, in bp)
E-mini S&P500 – ES@	0,0010	0,0065	2,2
Nasdaq E-mini – NQ@	0,014	0,26	1,3
Dow Jones E-mini – YM@	0,057	0,78	1,0
Eurostoxx – STXE@	0,0106	0,305	3,5
Footsie – FFI@	0,16	1,50	0,9
DAX – FDX@	0,49	3,27	0,8

Fig. 9 Basic statistics on 3rd-limit trade-throughs with volumes higher than the Mean Trade Volume (estimations based on March 2010 data)

As a conclusion, our definition of a trade-through as a trade consuming several limits of the order book is pretty stable with the size of the trades. A trade-through is generally a trade with an higher volume than usual trades.

3.4 Intraday Timestamp Distribution of Trade-throughs

We present in this paragraph the intraday distributions of timestamps[5] for 2nd-limit and 3rd-limit trade-throughs for the different equity futures of interest. When looking at this distribution for the 2nd-limit trade-throughs, we recognize a U-shape part, similar to the one of intraday volume distribution. But what strikes here is the presence of peaks at very precise hours, observed for both US and European equity futures. If we restrict the study only to the 3rd-limit trade-throughs, the U-shape part is almost completely removed from the distribution which is now composed only of those specific peaks.

The precise hours when this important statistical pattern of intraday distribution of trade-throughs timestamps happens are:

- 07.50: Eurex trading phase beginning (FESX, FDAX).
- 09.00: Euronext trading phase beginning (FTSE).

[5] All timestamps presented in this article are referenced in the time-reference of Europe/Paris = CET = UTC/GMT + 1 h.

- 14.30: CME open-outcry trading phase beginning (major equity index futures).
- 15.30: NYSE regular trading phase beginning.
- 16.00: US major macro news releases (ISM Manufacturing Index, . . .).[6]
- 17.30: End of the calculation of the DAX index using Xetra electronic trading system.
- 22.00: Eurex trading phase end (FESX, FDAX). Euronext trading phase end (FTSE).

4 Spread Relaxation

In this section, we want to focus on the behavior of spread after trade-throughs. But the approach is slightly different from what we previously did. We already know that trade-throughs are rare and informative events, whereas trades that only consume less than the quantity available at the best limit are what happens usually in the limit order book. So, if we consider the limit order book as a complex system, it is natural to see a trade-through as an excitation of this system.

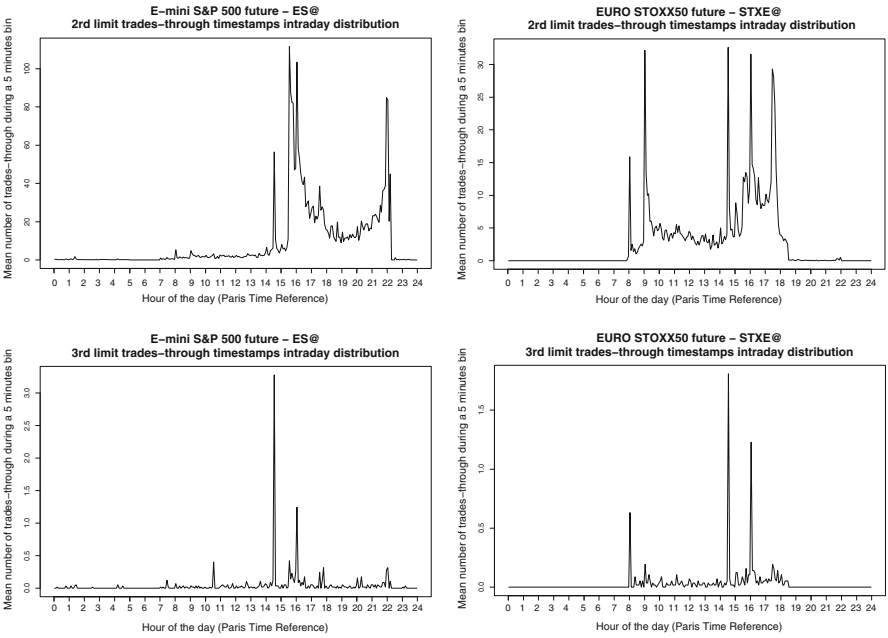


Fig. 10 Trade-throughs timestamps intraday distributions for US equity indexes futures (E-mini S&P500, Nasdaq E-mini and Dow Jones E-mini) and European equity indexes futures futures (Eurostoxx, DAX and Footsie)

[6] The possible links between news and major events in stock prices have already been studied, for example by Joulin [2].

From an initial state of the limit order book, an incoming trade-through excites the limit order book and brings it to an excited state. After this excitation, the complex system evolves in time. It may relax to a state close to the initial state, it may also stay at a state close to the excited one, ... After that, there are some usual trades that arrive in the limit order book. And then, a new trade-through occurs and excites the system again, and so on.

We think that this approach of considering the limit order book as a complex system and trade-throughs as excitations is particularly relevant to study the spread after a trade-through. So, we are interested in measuring the spread of the limit order book in order to know:

- what level of spread is reached at the excited state (after the trade-through)?
- is there a spread relaxation after the trade-through?
- if so, what kind of relaxation is it?

In the following figures, we detailed the behavior of the excess spread of the French stock BNP-Paribas after trade-throughs. Excess spread is the difference between the value of the spread after a trade-through and the value of the spread just before it (here-above called the initial state). The excited value of excess-spread is approximately one tick which seems reasonable as we know that 2nd-limit trade-throughs (which increase the spread one tick from the best limit) are the main part of trade-throughs. Then, we can see that, after a trade-through, there is a power-law relaxation of the excess spread in physical time (with an exponent close to 0.25).

Remember from the Sect. 2.7.2 where we studied the intraday distribution of trade-throughs, that there is on average 5 trade-throughs per bin of 5 minutes for the stock BNP-Paribas during the period of the day we are studying (from 9.30 to 17.00, Paris Time reference). So, on average, there is approximately one trade-through per minute for this stock which means that the limit order book is excited every minute and that the excess-spread relaxation we are looking at is statistically reliable on the first minute after each trade-through (before another trade-through arrives on average). For delays bigger than the minute, we have significantly less data to compute our statistics for this stock and we should be very carefull with any conclusion regarding this part of the graph.

We considered physical time in this section because it seemed to us this was the good measure to look at the relaxation of our complex system. There seems to be no such result if we look at the relaxation of the spread in trade-time (sampling the spread time-series before each trade). Our explanation of this difference is that when you look at the spread in trade-time, you are not considering any of the spread changes from the arrival time of the trade-through until the next trade. And this is precisely where the relaxation of the spread may occur. Basically, you are throwing away of your statistics the data where the phenomenon you're studying is the most relevant.

Lillo [1] studied a similar problem in the relaxation of the spread. They conditioned on a move of the spread and tried to measure a relaxation. Empirically, they get a power-law behavior of the excess spread in trade time (exponent 0.4 or 0.5). Power-law behavior seems to indicate a very slow relaxation but no specific

timescale for the spread decay. They have no explanation for this empirical observation.

Another example is given by Kertesz [3] where they studied relaxations after large price moves. They filtered large price moves both in 'absolute' values (intra-day price changes larger than 2% of the current price in time windows not longer than 120 minutes) and in 'relative' values (intra-day price changes in time windows not longer than 120 minutes, exceeding 6 times the normal volatility during that period of the day). They obtained a relaxation of the excess bid-ask spread decreasing in physical time as a power-law of exponent 0.38.

In both cases, the methodologies used are similar but different with respect to ours as the conditioning are done in spread moves or large price events, and not on the order-book limit reached by a trade. We all obtain very slow relaxation of the excess-spread whatever the timescale used, with power-law fits with exponents of the same order of magnitude.

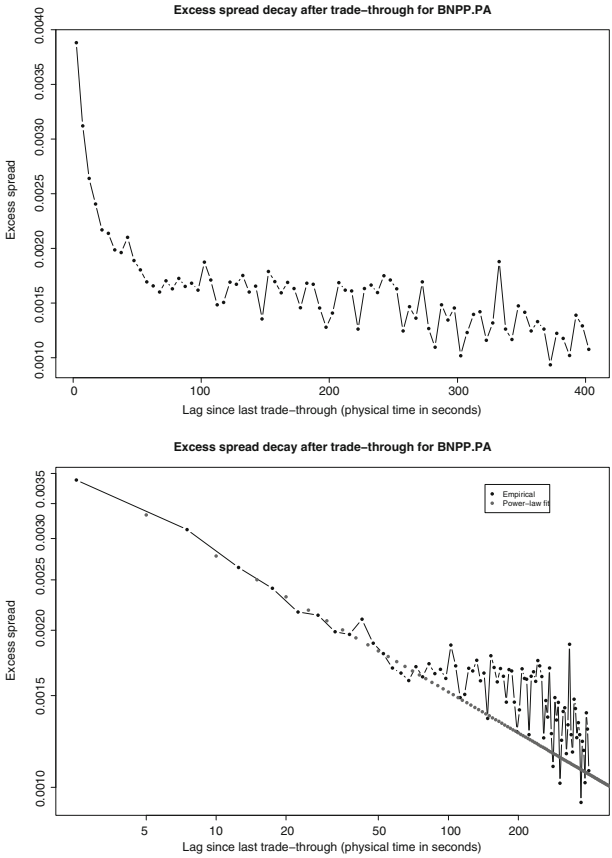


Fig. 11 Excess-spread decay after trade-throughs for BNP-Paribas on March-April 2010

5 Lead-lag Parameter Estimation from Trade-throughs Time-Series

5.1 Using Trade-throughs to Measure Empirical Distribution of Lead-lag Parameter

We think that our approach of using trade-throughs to obtain empirical distribution of lead-lag parameter if well-founded because:

- We are not using all trades to obtain all possible lead-lag between any of those (which would have no sense) but less data.
- We are using trades that are more informational (cf response function).
- If any lead-lag relation exists, this should reflect on major events occurring in limit order book.
- The method is empirical and its result is a full distribution of lead-lag parameter, and not only one value of lead-lag parameter that maximizes some contrast correlation criterium.

5.2 Empirical Measure of Lead-lag Parameter

5.2.1 Lead-lag Parameter: the Estimation Technique

For the sake of simplicity, assume we have two grids representing the timestamps of trade-throughs for two different assets and we want to get an empirical distribution of the lead-lag parameter between the two assets.

First, we put in relation every timestamp in one grid with the closest one in the other grid. So, every trade-through of the first grid has a correspondent trade-through in the other grid. Sometimes, a few trades in the other grid will not be the correspondent of any trade in the first grid. At the end of this stage, every trade-through of asset 1 is linked with the closest trade-through of asset 2. We show in the

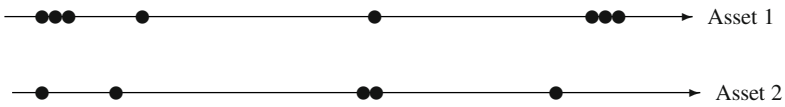


Fig. 12 Initial trade-throughs timestamp grids for assets 1 and 2

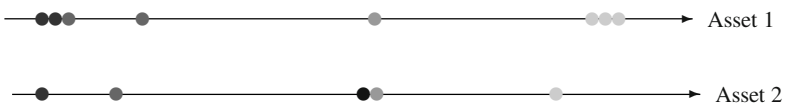


Fig. 13 Trade-throughs of asset 1 are put in relationship with those of asset 2

following figures how such two different timestamps grids of trade-throughs are put in relationship one with another.

Then, to obtain the empirical distribution of the lead-lag parameter between those two assets, we compute the difference of timestamps between any trade-through of the first grid and the correspondent trade-through of the other grid. We end up with as many measures of the lead-lag parameter as the number of trade-throughs in the first grid. Finally, we compute the distribution of those measures of the lead-lag parameter.

This method may be generalized to any groups of assets by merging in a previous step the timestamps of trade-throughs in the two groups. For example, in the study of the lead-lag between EU and US equity markets, we began by obtaining the two general grids (one for US equity market and one for EU equity market) by merging together the timestamps of the trade-throughs of E-mini S&P500, Nasdaq E-mini and Dow Jones E-mini futures (for the US general grid) and of Eurostoxx, DAX and Footsie futures (for the EU general grid).

5.3 Empirical Results for the Lead-lag Between US and EU Equity Markets

In the following graphs, we plot two useful statistics: the positive and negative cumulate of the lead-lag parameter between any trade-through occurring in the US

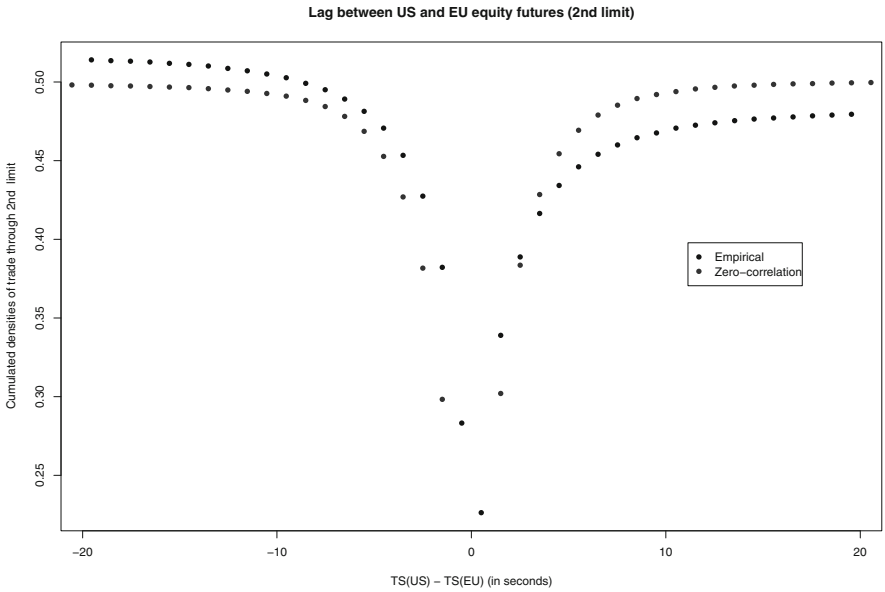


Fig. 14 Positive and negative lag cumulate between US and EU equity markets (zoom on the 20 seconds time-window - 200 seconds timescale)

group and the closest one in the EU group. On the left side of each graph, we have the part where US leads Europe. On the right side of the graph, it is the contrary: Europe leads US.

In order to compare the empirical distribution with a zero-correlation framework, we added the distribution obtained with the same method but where we randomized the arrival time of every trade-through for every future on any day. In the zero-correlation framework, the positive and negative cumulate should start from zero at the zero lag (like any cumulate) and both symmetrically grow and reach the value of 0.5.

We can notice that the empirical distribution of the lead-lag parameter is higher than the zero-correlation framework on the side where US lead Europe. And it is the contrary on the other side of the graph where Europe leads the US. So, if we believe that our financial products are representative of the global US ad European equity markets (which is the case because they are the most liquid instruments), then on average US equity market leads the European equity market.

6 Conclusions

In this paper, we defined particular trades that consume the liquidity in an aggressive way: the trade-throughs. We provide various statistics on trade-throughs: their liquidity, their intraday distribution and the spread relaxation that follows them. We studied their arrival timestamps and highlighted an important peak in this distribution at 4pm (Paris time reference), time of the day when major macro-economic news are released (ISM Manufacturing Index, . . .). We also used trade-throughs to obtain empirical distributions of lead-lag parameter between financial markets.

Acknowledgements The authors would like to thank the members of BNP-Paribas Equities & Derivatives Quantitative R&D team for fruitful discussions, and especially Boris LEBLANC and Stéphane TYC.

References

1. A. Ponzi, F. Lillo, R.N. Mantegna: Market reaction to a bid-ask spread change: A power-law relaxation dynamics. Physical Review E 80, 016112 (2009)
2. A. Joulin, A. Lefevre, D. Grunberg, J.-P. Bouchaud: Stock price jumps: news and volume play a minor role. URL: arXiv:0803.1769v1 (2008)
3. B. Tóth, J. Kertész, J.D. Farmer: Studies of the limit order book around large price changes. Eur. Phys. J. B 71, 499–510 (2009)
4. F. Lillo, J. Doyne Farmer: The Long Memory of the Efficient Market. Studies in Nonlinear Dynamics & Econometrics, vol. 8, issue 3 (2004)
5. J.-P. Bouchaud, J. Doyne Farmer, F. Lillo: How markets slowly digest changes in supply and demand. Handbook of financial markets: dynamics and evolution

Are the Trading Volume and the Number of Trades Distributions Universal?

Vikram S. Vijayaraghavan and Sitabhra Sinha

Abstract. Analysis of dynamical phenomena in financial markets have revealed the existence of several features that appear to be invariant with respect to details of the specific markets being considered. While some of these "stylized facts", such as the inverse cubic law distribution of price returns indeed seem to be universal, there is less consensus about other phenomena. In particular, there has been a long-running debate in the literature about whether the distributions of trading volume $V_{\Delta t}$ and the number of trades $N_{\Delta t}$ occurring in a given time interval Δt, are universal, and whether the volume distribution is Levy-stable. In this article, we analyse data from the National Stock Exchange of India, both daily and high frequency tick-by-tick, to answer the above questions. We observe that it is difficult to fit the $V_{\Delta t}$ and $N_{\Delta t}$ distributions for all stocks using the same theoretical curve, e.g., one having a power-law form. Instead, we use the concept of the stability of a distribution under temporal aggregation of data to show that both these distributions converge towards a Gaussian when considered at a time-scale of $\Delta t = 10$ days. This appears to rule out the possibility that either of these distributions could be Levy-stable and at least for the Indian market, the claim for universality of the volume distribution does not hold.

1 Introduction

A financial market comprising a large number of interacting components, viz., agents involved in trading assets whose prices fluctuate with time as a result of

Vikram S. Vijayaraghavan
Department of Physics, Indian Institute of Technology Madras, Chennai, India,
e-mail: vikramsv@smail.iitm.ac.in

Sitabhra Sinha
The Institute of Mathematical Sciences, C.I.T. Campus, Chennai, India,
e-mail: sitabhra@imsc.res.in

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

the constant stream of external news and other information affecting the actions of the traders, is a paradigmatic example of a complex system. However, despite such inherent complexity, markets appear to exhibit several statistically regular features which make them amenable to a rigorous analysis by techniques based on the statistical mechanics of physical systems, a discipline that is often referred to as econophysics [1–3]. Indeed, many of the empirical relations obtained by such analysis appear to be statistically invariant or *universal* with respect to different markets, periods of observations and the type of assets being considered. These *stylized facts* of the market (as they are often referred to in the economic literature) include the celebrated *inverse cubic law* for the distribution of price (or index) fluctuations as measured by their logarithmic returns [4]. First observed in the developed markets of advanced economies [5], it has later been reported also in emerging markets at all stages of their development [6, 7]. Another robust feature characterizing financial markets is *volatility clustering*, i.e., the occurrence of long-range temporal correlations in the magnitude of price fluctuations [8]. Thus, periods marked by large fluctuations (i.e., high volatility) often tend to be persistent, as is seen across many different markets.

There have also been claims that other quantifiers of market activity, such as the distributions for order size q (i.e., the number of shares traded in a particular transaction), trading volume V_t (i.e., the total number of shares traded in a given period) and the number of trades N_t over a specific time interval, possess universal forms [9, 10]. However, the evidence for the invariance of these distributions seems less unequivocal. Note that the three distributions are not completely independent of each other, as the volume $V_{t,\Delta t}$ over a particular time interval $[t, t + \Delta t]$ is related to the number of trades $N_{t,\Delta t}$ and the sizes of each trade q_i that takes place in the interval as

$$V_{t,\Delta t} = \sum_{i=1}^{N_{t,\Delta t}} q_i . \tag{1}$$

For US markets, the N_t cumulative distribution appears to follow an approximately "inverse cubic" form, i.e., $P(N_t > x) \sim x^{-\beta}$ with $\beta \simeq 3.4$ [9]. Both the trade size and volume cumulative distributions have been claimed to be Levy-stable with exponents $\zeta_q \simeq 1.53$ (the so-called "inverse half-cubic law") and $\zeta_V \simeq 1.7$, respectively [11]. However, not only has the universality of these exponents been challenged, even the power-law form of the distributions appear to be dependent on the type of stock and the market being considered. For example, an early study of the volume distribution of several stocks in the London Stock Exchange (LSE) did not show any evidence of power-law scaling [12], but it was pointed out later that this depended on whether one was considering the downstairs or upstairs market in LSE. As splitting a large order into several smaller parts is regularly practised in the downstairs market (but rare in the upstairs market) it is probably not surprising that long tails can only be seen when the trades in the upstairs market are included in the volume data [13]. A re-analysis of the US stock data complicated the issue further by showing that the cumulative distribution of trading volume over 15-minute intervals has a tail exponent of around 2.2, i.e., outside the Levy-stable regime [14].

More recent work on emerging markets such as the Korean [15] and Chinese [16] exchanges have also revealed significant deviations from the Levy-stable power-law tails of volume distribution reported for the developed markets of US, London and Paris [17] (see also [18, 19]). There have also been related studies that try to fit the entire distribution of trading volume rather than focusing only on the tail, e.g., by using the q-Gamma distribution [20, 21].

The reason that the universality (or otherwise) of the distributions for q, V_t and N_t is of interest to the econophysics community is because this may provide insights towards understanding the statistical relationship between price returns and market activity. It is frequently said that it takes volume to move prices, implying that the dynamics of price fluctuations (measured by the log-returns) can be understood in terms of the distributions of trade size, number of trades and trading volume. Indeed, the price impact function, that measures how the volume of shares traded affects the price movement, tries to quantify such a relation. By assuming a square-root functional form for the impact (based on empirical analysis of US markets), Gabaix *et al.* [22] have developed a theory of market movements where the long-tailed return distribution arises as a consequence of the long-tailed volume distribution. The square-root relation between price and volume leads to the result that the price return distribution exponent ($\simeq 3$) is twice the volume distribution exponent (~ 1.5), thereby connecting the inverse cubic and half-cubic laws. However, we have recently shown that the occurrence of power-law tailed distributions for price and volume with their characteristic exponents do not critically depend on the assumption of a square-root price impact function [23], nor does the existence of the inverse cubic law for returns necessarily imply an exponent of around $3/2$, or even a power-law nature, for the distribution of trading volume [24][1].

It is in this context that we report our analysis of the data for market activity in the National Stock Exchange (NSE) of India in this article. As this market has already been shown to exhibit the inverse cubic law of returns [6, 7], the absence of a Levy-stable nature for the volume distribution would appear to argue against the theoretical work relating the return and volume distribution exponents on the basis of a square-root form for the price impact function. While our earlier work on the trade and volume distributions in this market had also shown the absence of a clear power-law functional form for either [25], here we use an alternative procedure to show that the two distributions do not have the same behavior as that reported for the developed markets. In particular, we use the concept of stability of a distribution under temporal aggregation of data to show that both the quantities converge to a Gaussian distribution at a time-scale of $\Delta t = 10$ days. This evidence against the Levy-stable nature of the volume distribution (even though the return distribution follows the inverse cubic law) suggests that the theoretical framework of [23, 24] can better explain the market dynamics than arguments based on square-root price impact function whose predictions about the relations between return and volume is not matched by the empirical data. It is of course possible that the deviation from the

[1] In fact, our numerical results show that even a log-normal distribution of trading volume can result in a power-law tailed return distribution.

Levy-stable nature for volume and trade size distributions is a result of the emerging nature of the market which is yet to evolve into a completely developed form. Just as the network representing the relations between price movements of different stocks (measured by the cross-correlation between returns) has been suggested to change over time from being homogeneous to one having a clustered organization as the market matures [26], the volume distribution could, in principle, become more and more heavy tailed as market activity increases, eventually becoming Levy-stable at a certain stage of market development.

2 The Indian Financial Market

There are 23 different stock markets in India. The largest of these is the National Stock Exchange (NSE) which accounted for more than half of the entire combined turnover for all Indian financial markets in 2003–04 [27], although its market capitalization was comparable to that of the second largest market, the Bombay Stock Exchange. The NSE is considerably younger than most other Indian markets, having commenced operations in the capital (equities) market from Nov 1994. However, by as early as 2004 it had become the world's third largest stock exchange (after NASDAQ and NYSE) in terms of transactions [27]. It is thus an excellent source of data for studying the trading frequency and volume statistics in an emerging market.

Description of the data set. The low-frequency data that we analyze consists of the daily volume and number of trades for the entire NSE market, as well as, for individual stocks, available from the exchange web-site [28]. The period we have considered begins at March 1994 (for the entire market) or the date from which data for a particular stock has been recorded in the NSE database (for individual stocks) and ends at May 2010. For the market data, this corresponds to 3910 working days. We also consider high-frequency tick-by-tick data containing information of all transactions carried out in the NSE between Jan 1, 2003 and Mar 31, 2004. This information includes the date and time of trade, the price of the stock during transaction and the number of shares traded. This database is available in the form of CDs published by NSE.

3 Results

To investigate the nature of the volume and number of trades distribution in detail, we first consider the high-frequency tick-by-tick data. To calculate these quantities we use a time-interval $\Delta t = 5$ minutes and normalize the resulting variables by subtracting the mean and dividing by their standard deviation. The resulting distributions of normalized trading volume $v = \frac{V_{t,\Delta t} - \langle V \rangle}{\sqrt{\langle V^2 \rangle - \langle V \rangle^2}}$ and number of trades $n = \frac{N_{t,\Delta t} - \langle N \rangle}{\sqrt{\langle N^2 \rangle - \langle N \rangle^2}}$, where $\langle \ldots \rangle$ represents time average, for all stocks that are traded

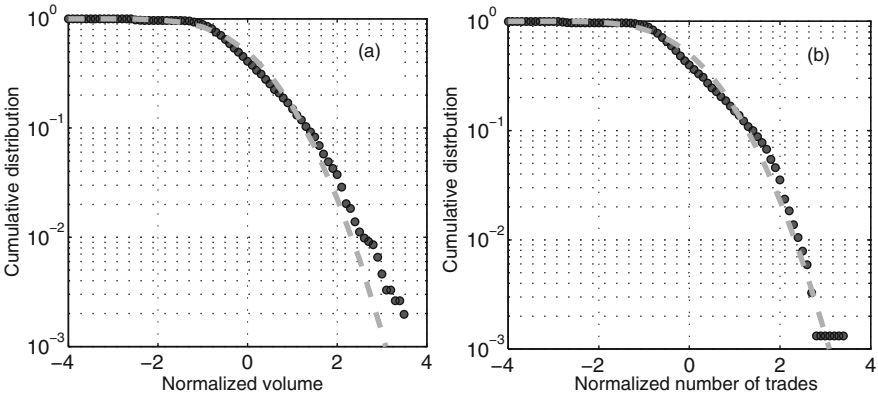


Fig. 1 Cumulative distribution of (a) normalized trading volume and (b) normalized number of trades in $\Delta t = 5$-minute time intervals for all stocks traded in NSE in December 2003. The cumulative standard normal distribution, i.e., $\mathcal{N}(0, 1)$, is shown for comparison (*broken line*)

in NSE are shown in Fig. 1. Direct comparison with the standard normal distribution $\mathcal{N}(0, 1)$ shows that both of these quantities are distributed differently from a Gaussian.

As the exact nature of the distributions for the entire market is difficult to characterize, we now consider the volume and number of trades data for *individual stocks*. Fig. 2 shows the corresponding distributions for a particular stock which appear to possess tails described by a power-law decay. Using the Clauset-Shalizi-Newman (CSN) estimator based on maximum likelihood and Kolmogorov-Smirnov statistic [29], we obtain exponents of -2.87 and -3.11 for the volume and number of trades respectively, both of which lie outside the Levy-stable regime. However, the values of these exponents differ from stock to stock. More importantly, the power-law nature of the decay itself is not entirely representative of the ensemble of stocks. The deviation of the distributions from a power-law is quite apparent visually for several frequently traded stocks (e.g., the volume distribution of SBI).

As the best-fit distributions for the high-frequency volume and number of trades statistics of the NSE do not appear to have a form that is common to all stocks, we cannot readily use this data to decide whether these distributions are Levy-stable or not. Instead, we shall use an indirect approach based on the idea of the stability of a distribution under time-aggregation of the corresponding random variables[2]. A distribution is said to be *stable*, when a linear combination of random variables independently chosen from the distribution has the same distribution, up to a trans-

[2] It should be noted here that the convergence to a stable form, which follows from the Central Limit Theorem, is strictly valid only when the variables being aggregated are statistically independent. However, if correlations do exist between the variables, then provided that these correlations decay sufficiently fast, the theorem still holds and the convergence result can be applied. We have explicitly verified that the auto-correlation function for trading volume shows an exponential decay with time-lag.

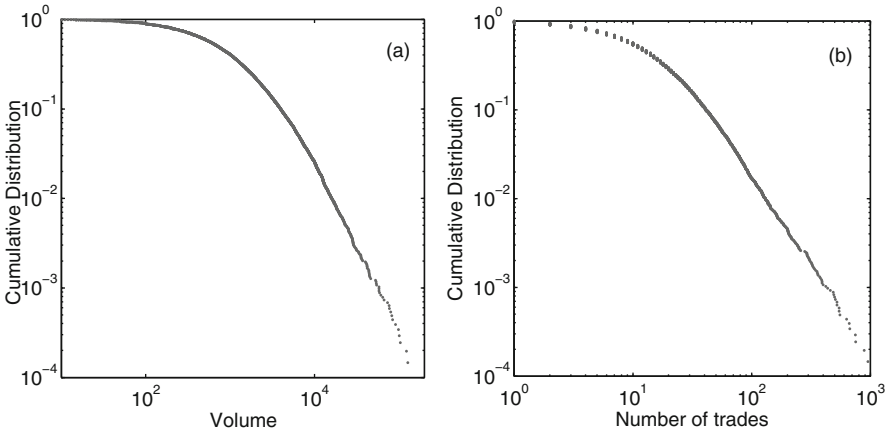


Fig. 2 Cumulative distribution of (a) volume and (b) number of trades in $\Delta t = 5$-minute time intervals for a particular stock (Colgate) traded at NSE between Jan 2003 and March 2004

lation in the mean and a scale factor in the variance. Thus, a sum of independent, identically distributed random variables always converge to a stable distribution. In terms of symbols, if x_1 and x_2 are random variables chosen from a stable distribution $P_{\text{stable}}(x)$, then for any pair of positive constants a, b, the composite variable $a\ x_1 + bx_2$ has the same distribution, but possibly with a different mean and variance. If the mean is identical to the original distribution, then it is said to be *strictly stable* (or stable in the narrow sense) [30]. This is a generalization of the classical *Central Limit Theorem*, according to which, a variable generated by adding a large number of random numbers from arbitrary distributions having finite variance will eventually be seen to follow a Gaussian distribution. Removing the restriction of finite variance results in other possible stable distributions, including the Cauchy and Levy distributions. In particular, a cumulative probability distribution having a power-law tail exponent $\alpha > -2$ has an unbounded second moment. It is, thus, Levy-stable and will not converge to a Gaussian even if we consider an aggregate quantity generated by summing together many random variables generated using such a distribution. Here, we shall use the fact that if the volume or the number of trades, when aggregated over long time periods, converges to a Gaussian distribution, then the original distribution of $V_{t,\Delta t}$ or $N_{t,\Delta t}$ (respectively) could not have been Levy-stable.

Fig. 3 shows the time-series of daily trading volume and number of trades for all stocks traded at NSE. To address the non-stationary nature of the variation in both the quantities, we calculate the mean (μ_t) and standard deviation (σ_t) over a moving window. The data is then de-trended by subtracting the mean and normalized by dividing by the standard deviation, i.e., $x_{t,\text{daily}} = (X_{t,\text{daily}} - \mu_t)/\sigma_t$, where $X_{t,\text{daily}}$ can represent either the daily volume or number of trades. The window used in Fig. 3 has a width of 10 days but small variations in the window size do not critically affect the results. One can also check whether the fluctuations from the mean values

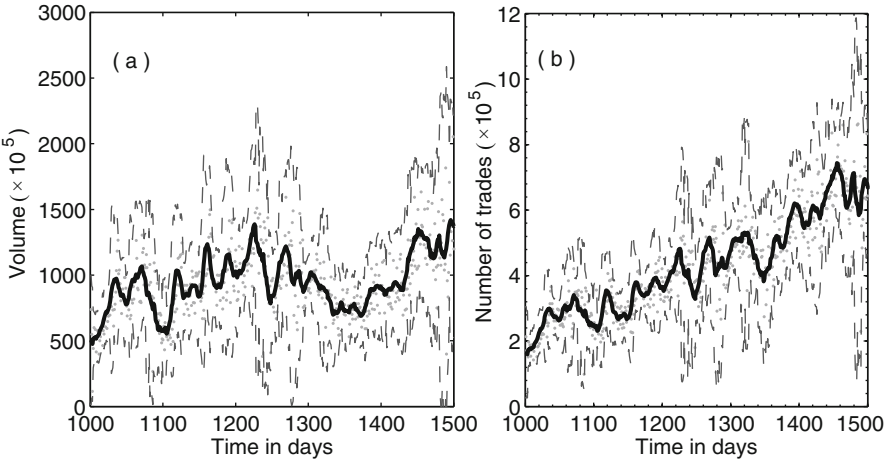


Fig. 3 The time-series of (a) the total volume of all stocks traded $V_{t,\text{daily}}$ and (b) the total number of trades in the market $N_{t,\text{daily}}$ (*dots*), shown for the interval $T = 1000$–1500 days in the daily NSE data. The continuous curves represent the moving average(μ_t) of the corresponding quantities calculated over a moving window (having a width of 10 days and shifted in steps of 1 day). The standard deviation (σ_t) calculated over the window is used to show the range of fluctuations (*dotted lines*) in the quantities $V_{t,\text{daily}}$ and $N_{t,\text{daily}}$ expected from a Gaussian distribution (i.e., $\mu \pm 3\sigma$)

observed in these quantities agree with those expected from a Gaussian distribution by verifying if most data points lie within the bounds representing three standard deviations above and below the mean (which account for about 99.7% of all data points if they are normally distributed). As seen from Fig. 3, this indeed appears to be the case.

To obtain a more reliable comparison between the empirical and normal distributions, we next use a graphical method, specifically the Quantile-Quantile or Q-Q plots [31], for comparing the de-trended, normalized volume and number of trades data with the standard normal distribution. The abscissa and ordinate of any point in such a Q-Q plot correspond to the quantiles (i.e., points taken at regular intervals from the cumulative distribution function) of the theoretical and empirical distributions being compared, respectively. Linearity of the resulting curve implies that the empirical distribution is indeed similar to the theoretical distribution, in this case, the standard normal distribution. While the daily data (Fig. 4a,d) shows deviation from linearity at the ends, the agreement between the two distributions become better when the data is aggregated over several days. Indeed, when we consider the volume and number of trades over a 10-day period, the corresponding distributions appear to match a normal distribution fairly well as indicated by the linearity of the Q-Q plots (Fig. 4c,f). This is also shown by direct graphical comparison of the distributions of these quantities aggregated over 10 days with the normal distribution shown in Fig. 5.

For a more rigorous determination of the nature of the distributions for the temporally aggregated volume and number of trades data, we turn to statistical tests for

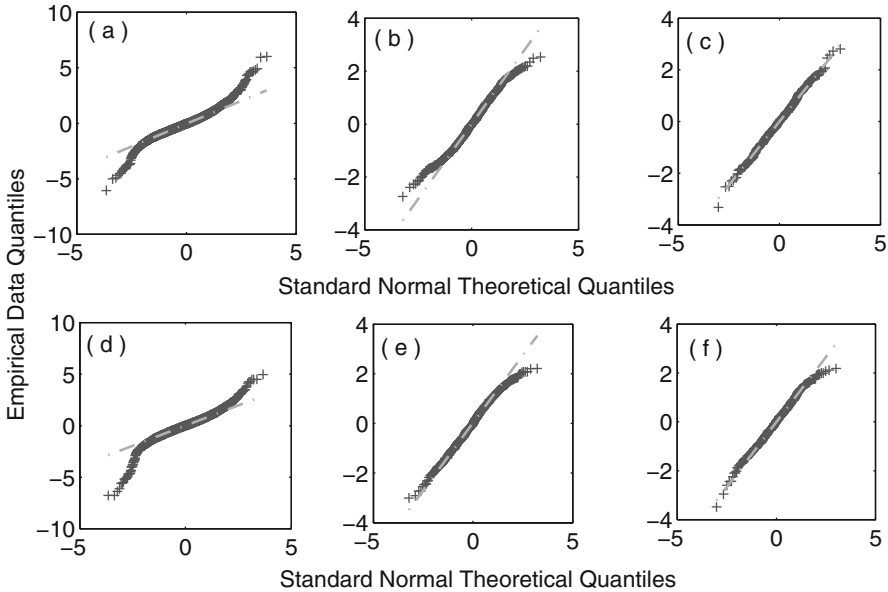


Fig. 4 Q-Q plots comparing the distributions of normalized and de-trended volume (a-c) and number of trades (d-f) for the entire market to a standard normal distribution at different scales of temporal aggregation. The aggregation is over 1 day for (a,d), 5 days for (b,e) and 10 days for (c,f). The *broken line* used for evaluating linearity of the plots is obtained by extrapolating the line joining the the first and third quartiles of each distribution. The linear nature of the plots for both volume and number of trades aggregated over 10 days suggest that the quantities converge to a Gaussian distribution at this level of temporal aggregation

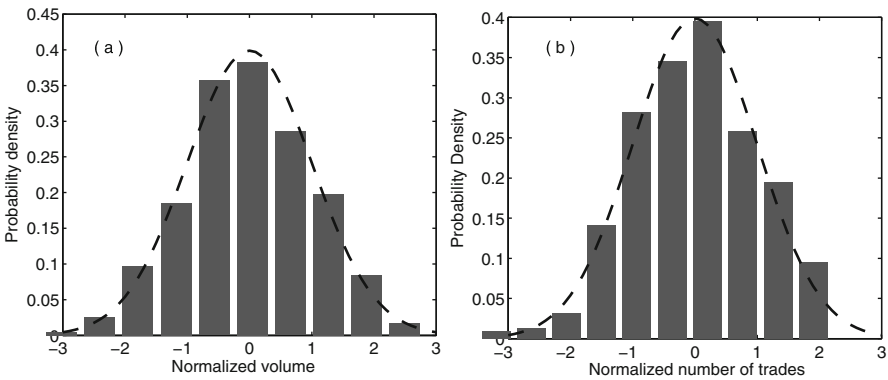


Fig. 5 Probability distribution of the de-trended and normalized trading volume (a) and number of trades (b) for all stocks in NSE aggregated over 10 days. For comparison, the standard normal distribution is shown (*broken curve*)

normality. Such tests go beyond simple regression-based best-fit of an empirical distribution by a theoretical curve and provides measures for the goodness of fit of the theoretical distribution to the data. Here, we use the Lilliefors test and the Anderson–

Darling test for testing whether the distribution of $v_{t,\text{daily}}$ and $n_{t,\text{daily}}$ approaches the Gaussian form as the data aggregation is done over longer and longer time-scales. For both these tests the null hypothesis (H_0) considered is that the empirical data is described by a Gaussian distribution. The Lilliefors test begins by estimating the mean and standard deviation of the underlying distribution from the data. It then calculates the test statistic, which is the maximum deviation of the empirical distribution from a normal distribution with the estimated mean and standard deviation. The null hypothesis is rejected when the maximum deviation becomes statistically significant. For all results reported here, we have fixed the level of significance at 5%. The p-value for the test indicates the probability of obtaining the observed maximum deviation assuming H_0 to be true, and a small value indicates that it is very unlikely that the empirical data follows a Gaussian distribution. The Anderson–Darling test is a non-parametric method for determining whether the empirical data is generated by a specific probability distribution and is considered to be one of the most powerful statistical tests for identifying deviations from normality [32]. It estimates the value of a test statistic, A^2, which is then compared with standard critical values of the theoretical distribution against which the empirical data is being tested. For example, the null hypothesis that a Gaussian distribution explains the empirical data can be rejected if the estimated test statistic A^2 exceeds 0.751.

The results of both statistical tests for the volume and number of trades data for the entire market is shown in Table 1. While the daily data clearly does not fit a Gaussian distribution, when aggregated over 10 days the trading volume does appear to be normally distributed, as the null hypothesis cannot be rejected for either of the tests we have used. Similarly, as the temporal aggregation is increased to 10 days for the number of trades data, the resulting distribution does appear to converge to a Gaussian form according to both the tests. As the time-period over which the daily data has been collected is relatively large ($\sim$ 16 years) we also checked whether the convergence to a Gaussian with increasing temporal aggregation also holds for subsets of the entire data-set. We have verified that even when the data is split into three approximately equal parts, with each sub-set corresponding to a period of about 5 years, the time-aggregated volume and number of trades distributions approach a Gaussian distribution according to the statistical tests.

Table 1 Normality tests for trading volume and number of trades for the entire NSE market at different scales of temporal aggregation

	Temporal Aggregation	Anderson–Darling test Reject H_0?	Anderson–Darling test Statistic (A^2)	Lilliefors test Reject H_0?	Lilliefors test p-value
Volume	1 day	Y	26.631	Y	0
	5 day	Y	2.1738	Y	0.0097
	10 day	N	0.2110	N	0.7100
Trades	1 day	Y	28.694	Y	0
	5 day	Y	1.4519	Y	0.0050
	10 day	N	0.3764	N	0.7950

Thus far we have been considering together all stocks that are traded in the NSE. In order to verify if the convergence to Gaussian distribution is also seen when the trading volume data for *individual stocks* is aggregated over longer periods, we shall now look at a few representative stocks from different market sectors. The cumulative distributions of the volume traded over the course of a single day for two stocks (Colgate and SBIN) are shown in Fig. 6. Both appear approximately linear in a semi-logarithmic graph, suggesting that the distribution may be fit by

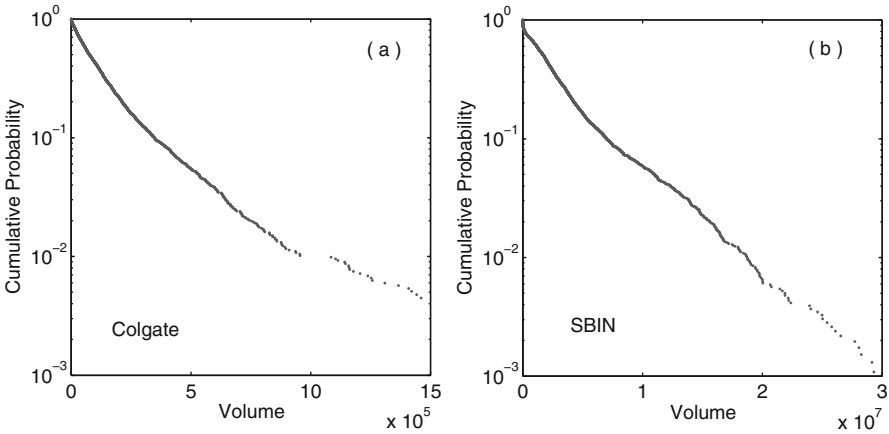


Fig. 6 Cumulative distribution of the total daily volume for two representative stocks: (a) Colgate and (b) SBIN, during the period March 1994 to May 2010. Note that the ordinate has a logarithmic scale. Thus, the linear nature of the distributions suggest that they are approximately exponentially decaying

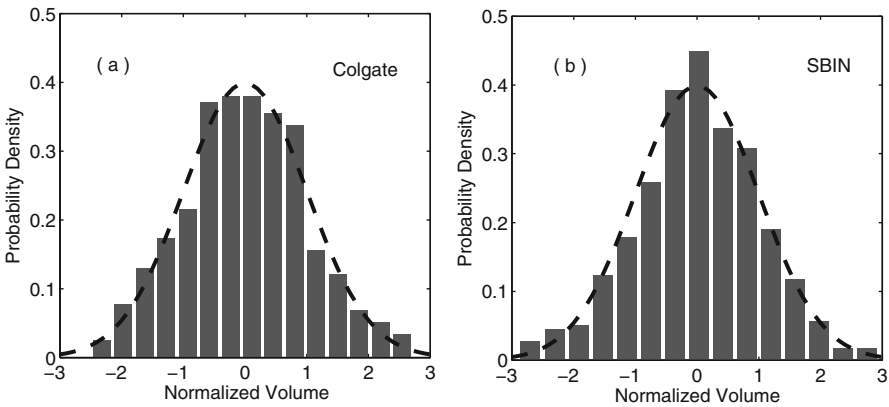


Fig. 7 Probability distribution of the de-trended and normalized trading volume aggregated over 10 days for two stocks: (a) Colgate and (b) SBIN. For comparison, the standard normal distribution is shown (*broken curve*). The time-series was de-trended by subtracting the mean calculated over a moving window (having a width 10 days and shifted in steps of 1 day) and normalized by dividing with the standard deviation calculated over the same window

an exponential form. However, when we look at the volume traded over 10 days, the corresponding de-trended and normalized distributions appear to be reasonably well-fit by the standard normal distribution (Fig. 7).

Table 2 Lilliefors test for normality of trading volume distribution for representative individual stocks in NSE

Stock	5-day aggregate		10-day aggregate	
	Reject H_0?	p-value	Reject H_0?	p-value
ABANLLOYD	Y	0.0	N	0.127
ACC	Y	0.0	N	0.727
COLGATE	Y	0.0	N	0.508
DABUR	Y	0.0	Y	0.002
DRREDDY	Y	0.001	N	0.002
GAIL	Y	0.0	Y	0.017
GLAXO	Y	0.0	Y	0.010
GODREJIND	Y	0.0	N	0.636
HCLTECH	Y	0.0	Y	0.014
HDFCBANK	Y	0.0	N	0.899
ICICIBANK	Y	0.0	N	0.558
INFOSYSTCH	Y	0.0	Y	0.0
IOC	Y	0.0	Y	0.0
RELCAPITAL	Y	0.0	Y	0.029
RELIANCE	Y	0.0	Y	0.0
SATYAMCOMP	Y	0.003	N	0.104
SBIN	Y	0.0	N	0.502
TCS	Y	0.036	N	0.787

Table 3 Anderson–Darling test for normality of trading volume distribution for representative individual stocks in NSE

Stock	5-day aggregate		10-day aggregate	
	Reject H_0?	Statistic (A^2)	Reject H_0?	Statistic (A^2)
ABANLLOYD	Y	6.696	N	0.546
ACC	Y	2.199	N	0.326
COLGATE	Y	6.691	N	0.229
DABUR	Y	7.013	Y	0.876
DRREDDY	Y	2.9960	N	0.389
GAIL	Y	4.084	Y	1.212
GLAXO	Y	6.833	Y	1.281
GODREJIND	Y	4.757	N	0.270
HCLTECH	Y	1.891	Y	1.099
HDFCBANK	Y	3.954	N	0.298
ICICIBANK	Y	3.504	N	0.611
IOC	Y	3.585	N	0.491
RELCAPITAL	Y	4.135	Y	1.113
RELIANCE	Y	7.806	Y	14.07
SATYAMCOMP	Y	2.001	N	0.521
SBIN	Y	2.537	N	0.339
TCS	Y	1.300	N	0.439

As in the case of the data for the entire market, we have carried out the Lilliefors test (Table 2) and the Anderson–Darling test (Table 3) for the volume data at different levels of aggregation. As is seen from the test results, while the volume traded over 5 days cannot be described by a Gaussian distribution for any of the stocks, when we consider the volume traded over 10 days, the Gaussian distribution appears to be a reasonable fit for many of the stocks considered.

Thus, our results indicate that, at least for the Indian market, the proposed invariant forms for the volume and number of trade distributions that have been observed in the developed markets of USA, London and Paris [17] do not hold true. In particular, the trading volume distribution does not follow a Levy-stable form. It has been suggested that, in the developed markets, the Levy-stability of the V_t distribution is a consequence of the Levy-stable trade size (q) distribution. Thus, a reason for the deviation of the volume distribution from Levy-stability could be inferred by looking at Eq. (1). If the distribution of q_i is Levy-stable but not that of N_t, the heavier tail of the former distribution would appear to dominate the nature of the tail of the V_t distribution. Presumably, this is what is happening in developed markets where we note that ζ_q and ζ_V are almost same (within error bars) [11]. However, in the Indian market, the distribution of q_i, even though it appears to fit a power-law, is clearly outside the Levy-stable regime. For instance, the exponent obtained by the CSN estimator for all trades carried out in December 2003 at NSE is $\zeta_q \simeq 2.63$ (Fig. 8). Thus, in the Indian financial market, the nature of the distribution for V_t may be dominated by that of N_t instead of the q distribution. Indeed, our earlier analysis had shown that there is a strong (almost linear) correlation between N_t and V_t [25], which would appear to support this hypothesis. It suggests that, for emerging markets where the trade size distribution has not yet become Levy-stable, the volume distribution would closely follow the distribution of the number of trades which is outside the Levy-stable region (as seen also for developed markets).

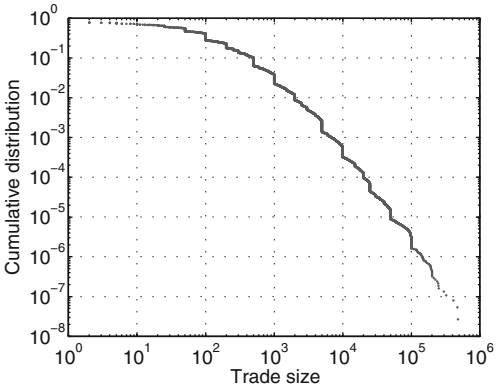


Fig. 8 Cumulative distribution of trade sizes for all transactions carried out at NSE during December 2003

4 Conclusions

In this article, we have examined the statistical properties of the distributions of the trading volume and the number of trades in the National Stock Exchange, the largest Indian financial market. Using both low-frequency (daily) and high-frequency (tick-by-tick) data, we have tried to characterize the nature of these distributions. In particular, we have sought to establish whether or not the distributions are Levy-stable by examining their stability under temporal aggregation. Our results show that although from the tick-by-tick or daily data it is difficult to exactly characterize the nature of the distribution of volume and number of trades, when we consider these quantities aggregated over a period of several days (e.g., 10 days), the resulting distribution approaches a Gaussian form. This has been verified both graphically using Q-Q plots and plots of the probability distribution functions, as well as, with statistical tests of normality, such as the Lilliefors test and the Anderson–Darling test. This suggests that the distributions of volume and number of trades are not Levy-stable, as otherwise they could not have converged to a Gaussian distribution when aggregated over a long period. Our results are significant in the context of the ongoing debate about the universality of the nature of the volume and number of trades distributions. Unlike the Levy-stable nature of the volume and trade size distributions seen in developed markets, the emerging financial market of India appears to show a very different form for these distributions, thereby undermining the claim for their universality.

Acknowledgements We are grateful to M. Krishna for invaluable assistance in obtaining and analyzing the high-frequency NSE data. We gratefully acknowledge the assistance of R. K. Pan in processing both the daily and high-frequency data.

References

1. Mantegna RN, Stanley HE (2000) An Introduction to Econophysics. Cambridge University Press, Cambridge
2. Farmer JD, Smith DE, Shubik M (2005) Is economics the next physical science? Physics Today 58(9):37–42
3. Sinha S, Chakrabarti BK (2009) Towards a physics of economics, Physics News 39 (2):33–46
4. Gopikrishnan P, Meyer M, Amaral LAN, Stanley HE (1998) Inverse cubic law for the distribution of stock price variations, Eur Phys J B 3:139–140
5. Lux T (1996) The stable paretian hypothesis and the frequency of large returns: An examination of major German stocks, Applied Financial Economics 6:463–475
6. Pan RK, Sinha S (2007) Self-organization of price fluctuation distribution in evolving markets, Europhys Lett 77:58004
7. Pan RK, Sinha S (2008) Inverse-cubic law of index fluctuation distribution in Indian markets, Physica A 387:2055–2065
8. Bouchaud JP (2005) The subtle nature of financial random walks, Chaos 15:026104
9. Plerou V, Gopikrishnan P, Amaral LAN, Gabaix X, Stanley HE (2000) Economic fluctuations and anomalous diffusion, Phys Rev E 62: R3023–R3026

10. Gopikrishnan P, Plerou V, Gabaix X, Stanley HE (2000) Statistical properties of share volume traded in financial markets, Phys Rev E 62:R4493–R4496
11. Plerou V, Gopikrishnan P, Gabaix X, Amaral LAN, Stanley HE (2001) Price fluctuations, market activity and trading volume, Quant Finance 1: 262–269
12. Farmer JD, Lillo F (2004) On the origin of power-law tails in price fluctuations, Quant Finance 4:C7–C11
13. Plerou V, Gopikrishnan P, Gabaix X, Stanley HE (2004) On the origin of power-law fluctuations in stock prices, Quant Finance 4:C11–C15
14. Eisler Z, Kertesz J (2006) Size matters: Some stylized facts of the stock market revisited, Eur Phys J B 51:145–154
15. Hong BH, Lee KE, Hwang JK, Lee JW (2009) Fluctuations of trading volume in a stock market, Physica A 388:863–868
16. Qiu T, Zhong LX, Chen G, Wu XR (2009) Statistical properties of trading volume of Chinese stocks, Physica A 388:2427–2434
17. Plerou V, Stanley HE (2007) Tests of scaling and universality of the distribution of trade size and share volume: Evidence from three distinct markets, Phys Rev E 73:046109
18. Racz E, Eisler Z, Kertesz J (2009) Comment on "Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets", Phys Rev E 79:068101
19. Plerou V, Stanley HE (2007) Reply to "Comment on 'Tests of scaling and universality of the distribution of trade size and share volume: Evidence from three distinct markets' ", Phys Rev E 79:068102
20. Queiros SMD (2005) On the emergence of a generalised Gamma distribution: Application to traded volume in financial markets, Europhys Lett 71:339–345
21. Mu G-H, Chen W, Kertesz J, Zhou W-X (2009) Preferred numbers and the distribution of trade sizes and trading volumes in the Chinese stock market, Eur Phys J B 68:145–152
22. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE (2003) A theory of power-law distributions in financial market fluctuations, Nature 423:267–270
23. Vikram SV, Sinha S (2010) A mean-field model of financial markets: Reproducing long-tailed distributions and volatility correlations. In: Basu B *et al* (eds) Econophysics and Economics of Games, Social Choices and Quantitative Techniques. Springer, Milan
24. Vikram SV, Sinha S (2010) Emergence of universal scaling in financial markets from mean-field dynamics, arXiv:1006.0628
25. Sinha S, Pan R K (2006) The power (law) of Indian markets: Analysing NSE and BSE trading statistics. In: Chatterjee A, Chakrabarti B K (eds) Econophysics of Stock and Other Markets. Springer, Milan
26. Pan RK, Sinha S (2007) Collective behavior of stock price movements in an emerging market, Phys Rev E 76:046116
27. National Stock Exchange (2004) Indian securities market: A review. (http://www.nseindia.com/content/us/ismr2005.zip)
28. http://www.nseindia.com/
29. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data, SIAM Review 51:661-703
30. Nolan JP (2010) Stable Distributions: Models for heavy tailed data. Birkhauser, Boston
31. Thode H (2002) Testing for Normality. Marcel Dekker, New York
32. Kvam PH, Vidakovic B (2007) Nonparametric Statistics with Applications to Science and Engineering. Wiley, Hoboken

Subpenny Trading in US Equity Markets

Romain Delassus and Stéphane Tyč

Abstract. We study sub-penny trading in the US equity markets.

1 Equity Market Structure – Introduction

1.1 SEC Concept Release

In January 2010, the Securities and Exchange Commission (SEC) issued a concept release[1] seeking public comments on various issues concerning the U.S. equity markets structure. This concept release is mainly based on the observation that the U.S equity markets have undergone significant changes in recent years, most of which were due to technological innovations. Markets now rely on advanced computer technology, trading speed has accelerated to microseconds, and trading centers now offer a large variety of order types and other trading services.

The concept release seeks public comment on such issues as high frequency trading, co-location trading terminals, markets that do not publicly display price quotation, and more generally on any matter related to market structure. To date, some 200 comments have been received by the Commission [4]. Submissions come from individual equity investors, investment firms, banks, broker-dealers … and comments therefore reflect a complete range of opinions on the subject. Reading those comments, one issue kept coming back, brought up by individual equities investors

Romain Delassus
École Polytechnique – BNP Paribas, Paris,
e-mail: romain.delassus@polytechnique.org

Stéphane Tyč
BNP-Paribas, Head of Equity Derivatives Quantitative Analytics
e-mail: stephane.tyc@bnpparibas.com

[1] All information can be found at http://www.sec.gov/news/press/2010/2010-8.htm.

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

as well as investment firms. This issue is well summarized in one of Bright Trading comments [7]:

> Underlying Problem: Undisplayed Trading Center Compromising the NBBO through Sub-Penny Trading

1.2 Minimum Price Variation

A crucial feature of the market structure is the minimum price variation (MPV), also called tick size. For stocks priced over \$ 1, the one penny minimum price variation was chosen in 2002 as a conclusion of a series of studies and comments. The main idea was to balance the positive effects of a reduction of the tick size, which was then defined as a fraction (usually $\frac{1}{16}$) of a dollar, (reduction of spreads, greater opportunities for price improvement ...) against the negative effects of a too small MPV (flickering quotations, market depth reduction ...). The rule however, didn't apply to all trading venues, and in August 2003, NASDAQ submitted a proposed rule change to adopt an MPV of \$ 0.001 for Nasdaq-listed securities in order to remain competitive with Electronic Communication Networks (ECNs).

In February 2004, the Commission proposed new Rule 612 that would create a consistent monetary increment for all trading venues. The sub-penny rule (Rule 612 of regulation NMS) indeed:

> prohibits market participants from displaying, ranking, or accepting quotations, orders, or indications of interest in any NMS stock priced in an increment smaller than \$ 0.01 if the quotation, order, or indication of interest is priced equal to or greater than \$ 1.00 per share.[2]

However, this rule concerned only order and quote submission and did not prohibit sub-penny trading.

The problem raised in today's concept release comments concerning sub-penny trading is closely linked to SEC's questions on undisplayed liquidity. Indeed, each of these comments (we counted over 50 of them) points out the fact that this rule applies neither for broker-dealers internalization, nor for dark-pools. Some say this leads to a beneficial price-improvement for the client (which seemed to be SEC's purpose at the time this exemption was decided), but others declare that this practice has become much too important, and now compromises the NBBO by discouraging liquidity providers.

The comments submitted can be divided in two groups which, in turn, can be schematically mapped on the position of the submitting firms in the equity trading landscape. It will help guide the understanding to draw an oversimplified caricature of these actors.

[2] Extracted from SEC's Responses to Frequently Asked Questions Concerning Rule 612 of Regulation NMS [6].

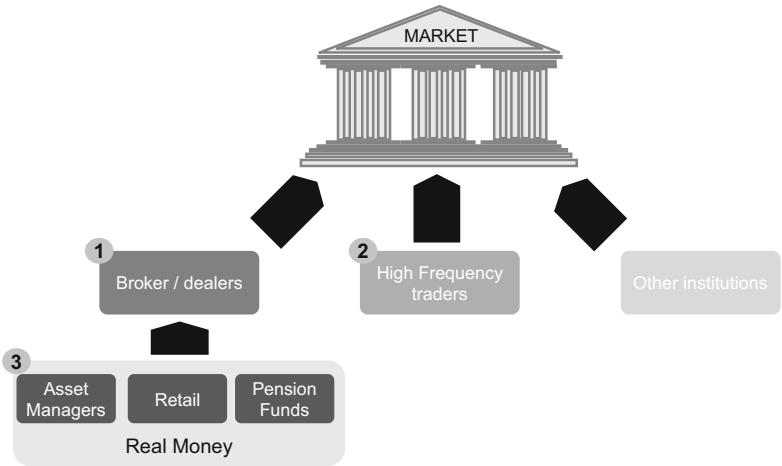


Fig. 1 Schematic diagram showing two different actors in the U.S. equity market

1.3 Two Different Points of View

In a very simplified way, one could describe two different actors in the U.S. equity market by the following graph:

The broker-dealers execute their clients orders on and off the market, and the high frequency traders interact directly with the market.

Of course, frontiers are not always that clear, and we do not pretend to categorize anyone in such or such category. Indeed, many firms interact with equity markets in both capacities. However, we do believe this schematic representation helps understanding the different points of view reflected by the comments. The main idea is that the different actors are distinguished according to their role in the equity trading ecosystems. In this oversimplified view, we separate actors trying to provide best execution to their clients (the broker-dealers) from actors trying to maximize the gains from their strategies (High Frequency traders) and who portray themselves as liquidity providers.

1.3.1 Broker-dealers

We pool in this category the different actors serving as intermediaries for investors. In general, this type of actor receives client order flow, can interact with it at risk, tries to pre match the flow and provides best execution in the public venues.

Most actors in this category believe that a well-diversified offer serves the interests of long term investors. For example, they are often in favor of keeping some level of opacity in the market, in order to control market impact. This argument especially holds for large trades. On the subject of sub-penny trading, they often reject

the proposal to reduce the minimum quoting increment, arguing that it would only exacerbate the problem of trading ahead of their orders by high-frequency traders.

Those actors also reject the concept of a trade-at rule, which would introduce some kind of time priority on the NBBO. The key aspect is that today, at a given price point, all venues are considered equal. However, not all executions at the same price are equal, since the structure of fees varies, depending on the market. They argue that a trade-at rule would reduce competition between exchanges, ATSs and OTC market makers and would most likely lead to an increase in long-term investors' costs.

1.3.2 High Frequency Traders

This term is simply to be understood as a category of actors whose holding period is short. They interact directly with the market, often acting as market makers. Their time horizon can vary from a few seconds (or even less) for the fastest ones, to a few days. This group is much more heterogenous than the previous one, but a certain number of characteristics stand out, the main one being that they use open competition and technological innovation to narrow spreads.

Almost all comments from this category criticize the use of sub-penny trading. The problem lies in the fact that only a minority of actors is allowed to use such a strategy, which automatically leads to the formation of a " two tiers market". Their main argument is that in today's quote-driven markets, there is an intensive competition between market-makers that has conducted to a substantial reduction of costs for long-term investors. Their view point is supported by the comment of one major equity investor. The Vanguard Group [2] indeed estimates that "transaction costs have declined 50 bps, or 100 bps round trip".

As a result of this reduction,

> if an average actively managed equity mutual fund with a 100% turnover ratio would currently provide an annual return of 9%, the same fund would have returned 8% per year without the reduction in transaction costs over the past decade.

In view of this estimation, high frequency traders conclude that transparency and competition are always beneficial, and urge the SEC to take action against sub-penny trading. To support this argument, they argue that today:

> [...] almost every single market order placed in these retail brokerage accounts, is checked by the broker-dealer's OTC market maker to decide if they can make money by trading against their customer.[3]

The conclusion is therefore that the only NBBO orders that are filled are those that are most likely "toxic" in the short-term. This can only lead to a widening of spreads, and a reduction of depth in the market.

[3] Extracted from Bright Trading's comment.

This issue of sub-penny trading raises a certain number of regulatory questions. The idea is to determine whether the argument of price improvement holds against the negative impact that this strategy has on liquidity providers. Should sub-penny trading simply be banned? Would a regulatory change in the minimum quoting increment be beneficial in some specific cases?

2 Empirical Study

2.1 Methodology and Data

In order to study the usage of sub-penny trading, we should first evaluate its importance. In a comment submitted to the SEC, Knight Capital Group wrote that, in 2009, it "provided over $ 63 million in price improvement on 26.3 billion shares". This figure translates into a price improvement of 0.24 cent per share. On the contrary, the many Bright Trading comments show the usage of sub-penny trading in order to step in front of the NBBO with a price improvement so little as 0.01 cent per share. In order to confront those 2 points of vue, we conducted a simple statistical review of sub-penny trading in U.S. equity stocks.

The data used to obtain the following results all come from Thomson Reuters tick-by-tick historical data (see [3] for more information). We used Reuters Indice Codes (RIC) that provided us directly with the consolidated trade tape, mainly on stocks listed on NASDAQ Capital Market. Our main problem was then to calculate the price improvement. As we had very few information on each trade (essentially its price and quantity, and eventually the NBBO at the time of the trade), we simplified this problem by considering that for each trade, the price improvement was equal to the difference between the real price and the price approximated to the closest cent value. For example, when we saw a trade price of **$ 24.9899**, we implicitely supposed that the best national ask at the time was of **$ 24.99** and therefore that the price improvement per share of the trade was **$ 0.0001**.

Another important aspect is that we sometimes find trade prices with a precision superior to 1/100th of a cent. It seems unlikely that those were the true prices of the trades, and in order to get around this problem, we chose to round all prices to a precision of 1/100th of a cent (corresponding to a price improvement of 1 cent per lot). A price of **493.149994** is therefore not considered as a sub-penny trade in our study.

Table 1 Extract from Tomson Reuters data for WPO.II on the 12th Apr. 2010

Timestamp	Last	LastQ	Ask	Bid
34786.946	**493.149994**	100	493.3	468.52
34786.965	**493.299988**	100	493.3	468.52
35950.657	492.5	100	493	490.51

2.2 Results

2.2.1 Price Improvement Distribution

Our first result concerns the distribution of the price improvement for sub-penny trades on NASDAQ100 stocks. We calculated this distribution by analyzing a one week long data set of all trades on stocks listed in the NASDAQ100 index. The graph should be read as follows: for example, the first bar states that **0.9%** of the total traded volume (counted in stocks) that took place on those stocks during the third week of March 2009 was traded with a sub-penny increment comprised between **$ 0.0001** and **$ 0.0004**. The volume traded in sub-penny represents 7.38% of the total traded volume for the sample week in 2009, and 10.62% for the week chosen in 2010.

Two main features can be seen in these distributions:

- We first observe a significantly higher frequency for trades with half a cent in price improvement. This can be explained by the fact that some trades are priced at the NBBO midprice, a method that sometimes leads to a price with a half cent. This kind of trades actually supports the view put forward by "broker dealers" in our classification.
- Another interesting point is that the percentages are distributed symmetrically around the half a cent price improvement. This comforts us in the method we used to calculate the price improvement. We also observe the decreasing relationship between the volume traded and the price improvement (from 0.0001 to 0.0049), with a maximum clearly located at the first bar. This observation seems to confirm Bright Trading's opinion on sub-penny trading, a simple strategy used to step ahead of protected limit orders.

2.2.2 Correlation with the Stock's Price

Our next result about the importance of sub-penny trading concerns the understanding of the characteristics of the stocks on which this strategy concentrates. As we have seen in the previous paragraph, sub-penny trading is a reality. However, the previous graph showed average figures on all NASDAQ100 index stocks. We here want to show that the previous distribution is actually strongly correlated with the stock's value.

In order to do so, we present on the next page a graph showing a similar distribution as the one shown before, but calculated on a list of more than 440 stocks quoted on the NASDAQ and priced between 1 $ and 5 $. We can observe that the first and last bars of the distribution have a relative weight much more important for those stocks than for the NASDAQ100 stocks.

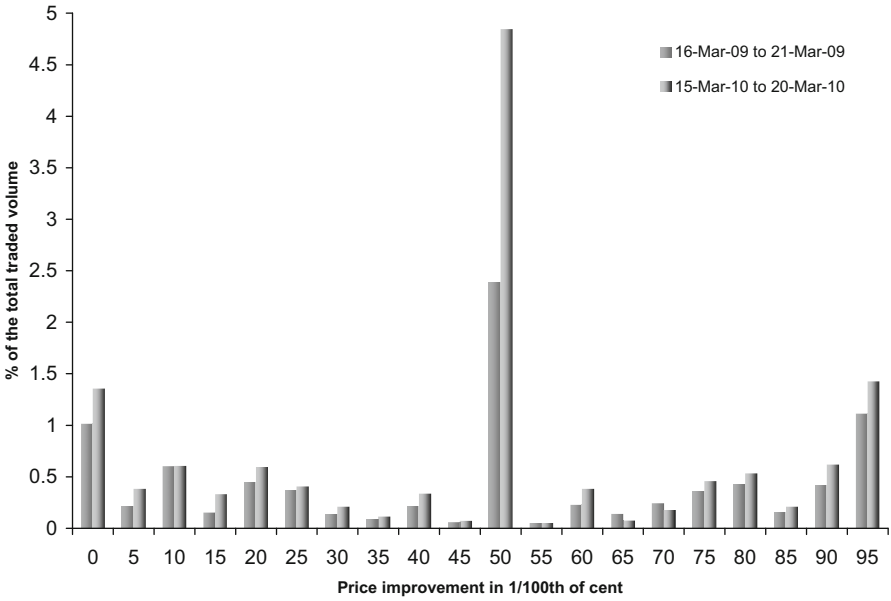


Fig. 2 NASDAQ distribution

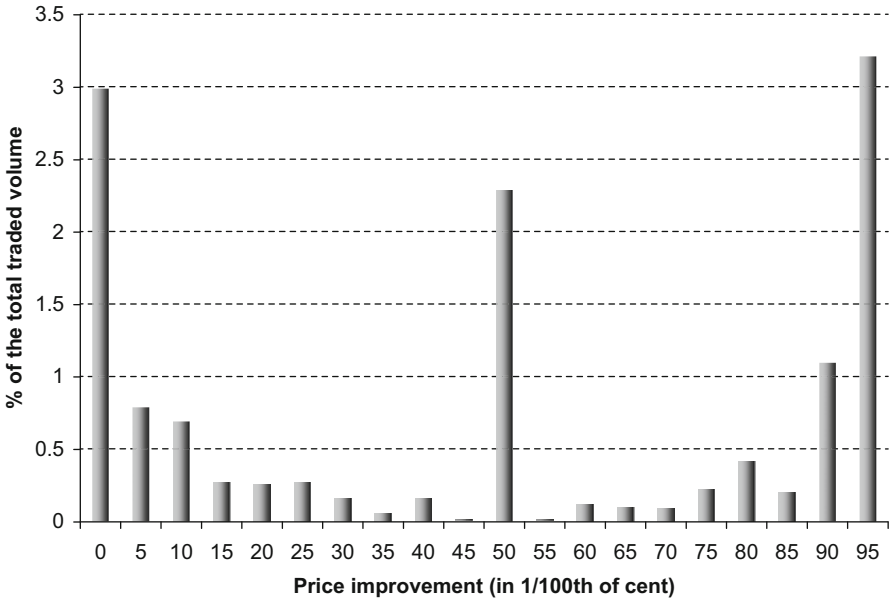


Fig. 3 Price improvement distribution on 440 NASDAQ stocks between 1 $ and 5 $

2.2.3 On the Importance of Queue Jumping

In order to assert the validity of the price-improvement argument, we wanted to quantify this improvement in a more detailed way than the aggregate number given by Knight Capital Group. The previous results on the price improvement distribution seem to support both views. There are however two kinds of sub-penny trading. The first one results from a crossing at the midprice, but the second one is a simple strategy to step ahead. Indeed, in stocks priced under 5 \$, where spreads often are narrowed to one tick (the minimum for high frequency traders), we observed that more that 6% of the traded volume is traded with a price-improvement of less than 1/20th of a cent (first and last bars). In the following of this paper, we call "queue jumping" the usage of those prices (such as 9.9999 \$ or 10.0001 \$). In such cases, the rebate given to the client is of 1 to 4 cents per lot, which most certainly doesn't justify the loss of opportunity for the liquidity provider.

The previous distributions seemed to show a negative relation between the queue jumping percentage and the value of the stock. In order to illustrate this fact, we calculated for more than 1800 NASDAQ stocks the average effective spread[4] and the percentage of the volume traded in queue jumping for the 11th week of 2010. The result is given in the following graph, with the effective spread given in a percentage of the stock's value. The red line represents a 1 tick effective spread, and the yellow squares the mean of the queue jumping percentage.

The main result is given by the queue jumping percentage mean. For stock values over 5 \$, the yellow square of abscisse x gives the mean for all the stocks priced between x and $x + 1$ dollars, whereas for stock values under 5 \$, the mean is calculated for stocks priced between x and $x + 0.5$ dollars. We observe that, as the stock value decreases and the relative spread increases, the percentage of queue jumping becomes more and more important. An intuitive interpretation of this result is that the the lower the stock value is, the higher the relative spread and therefore the more profitable queue jumping becomes.

2.2.4 Sub-penny Trading Provides Price Improvement

The previous paragraph highlighted the correlation between the queue jumping percentage and the stock value, showing that sub-penny trading strategies that only provide a negligible price improvement (the so called "queue jumping") are actually very important for low priced stocks. In a more general way however, the justification of sub-penny trading by the price-improvement argument holds, as we show on the following graph.

On this graph, we calculated for each stock the total price improvement in dollars (i.e. for each trade the price improvement, as defined in the methodology, multiplied by the volume) as well as the effective spread multiplied by the number of stocks

[4] The effective spread is here defined as twice the absolute difference between the trade price and the NBBO midprice. The average is done on all trades of the week.

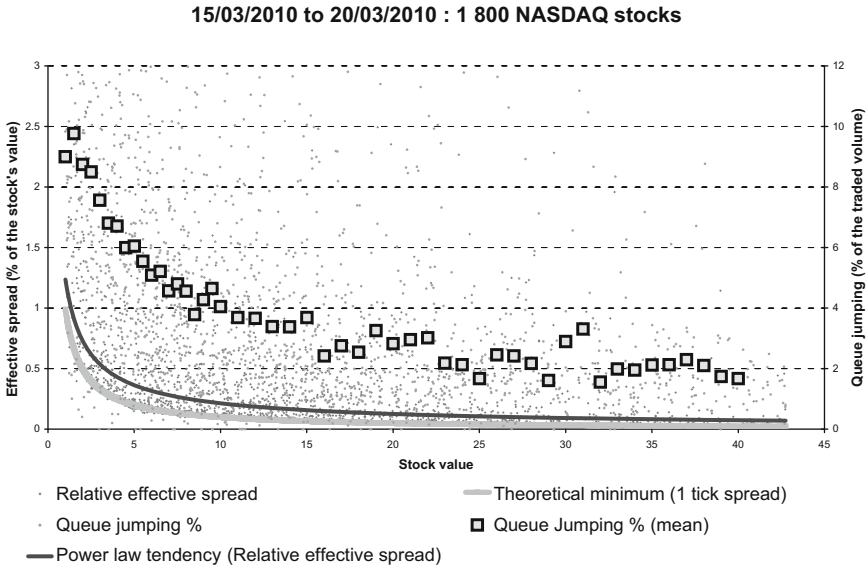


Fig. 4 Effective spread and queue jumping versus stock value for 1800 NASDAQ stocks

traded with a sub-penny price improvement. The main idea is that market makers essentially make money by "earning the spread". On average, high frequency traders gains can be estimated by $x \times \text{Spread} \times \text{Volume traded}$, where x is a certain discount factor to take the various costs as well as adverse selection into account (from our definition of the effective spread, we have $x < 0.5$). The effective spread multiplied by the sub-penny volume (and multiplied by x) therefore represents an approximation of the high frequency traders' loss to broker-dealers due to sub-penny trading. On the other hand, the total price improvement represents the money paid back by the broker-dealers to their customers.

The graph on Fig. 4 shows that on average, broker-dealers pay each final investor some 20% of the spread in sub-penny increments. This corresponds to a reduction of the effective spread of 40% for the trades executed in sub-pennies, and increments add up to \$5 million on 2 billion shares over the week. We find the same ratio that was announced by Knight Capital Group. However, this figure is not equally distributed between midprice and queue jumping.

The graph on Fig. 5 is equivalent to the one on Fig. 4, but we only took into account non-midprice trades. On this type of trades, broker-dealers only pay 10% of the spread to final investors. Once again, these graphs point out the fact that sub-penny trades are separated into 2 different categories:

- Midprice crossings: those trades are not intermediated at risk, since they generally are due to the crossing of the order flow before sending it to the exchange. In this case, the gains (compared to a normal trade) are entirely paid to the clients (half a spread each).

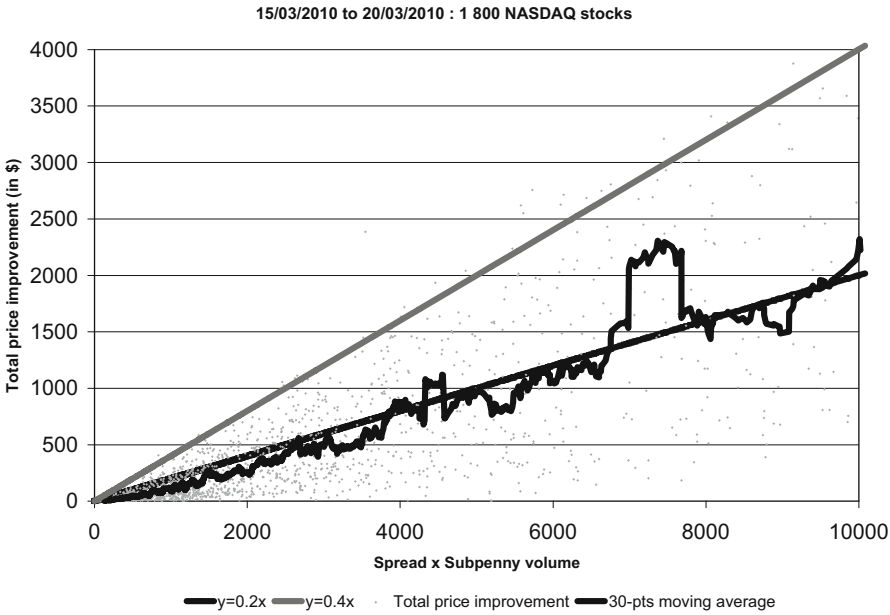


Fig. 5 Price improvement for 1800 NASDAQ stocks

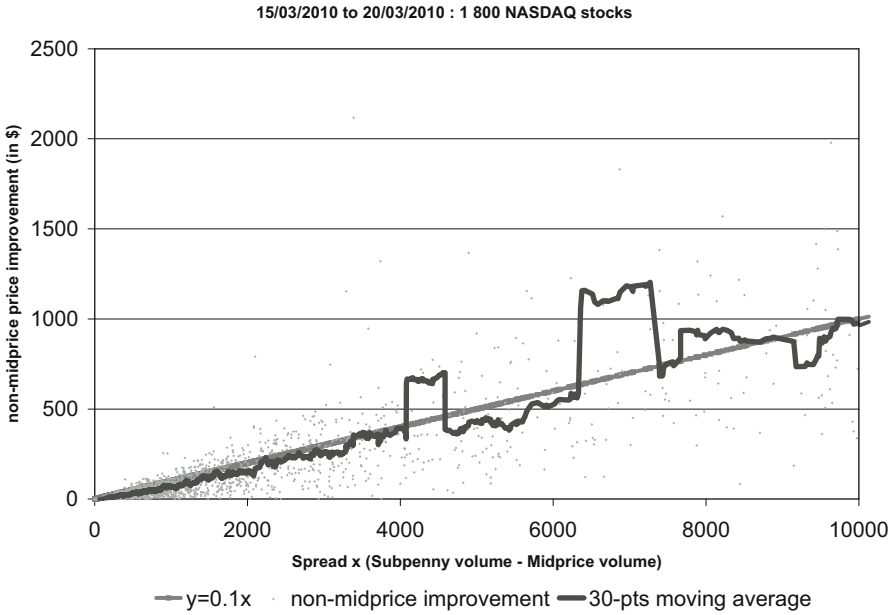


Fig. 6 Price improvement for 1800 NASDAQ stocks

- Non-midprice crossings: generally, a market maker needs to support the risk and trade on his own account. In this case, 10% of the spread is economically transferred to each final investors.

The fact that the gains for final investors are inferior in the second case is entirely justifiable. The market maker supports a risk, and is paid in order to do so. It is however, the increasing proportion of this kind of trades that is much criticized in the SEC concept release comments. Since some market makers are able to reduce the spread by some 20%, why not encourage competition and authorize this practice to a larger proportion of actors? The Sub-penny rule in 2007 imposed a minimum tick of 1 cent for stocks over 1 $. The reason at the time was that a further reduction of the tick would have caused too many problems compared to the beneficial reduction of the spread. The rule however was kept flexible enough to allow that a small propotion of trades could still take place with a sub-penny increment. The next section will show that this proportion has actually tripled since (increasing from 4% to nearly 12% of the total traded volume) and cannot be concidered now as marginal.

2.3 Historical Evolution

2.3.1 Evolution of the Importance of Sub-penny Trading

We now concentrate on studying the historical evolution of the different results presented earlier. Our first graphic shows the evolution, since 2003, of the height of different bars of the price improvement distribution. *Total subpenny* represents the sum of all bars, *0.005 $ improvement* represents the volume traded with a half a cent

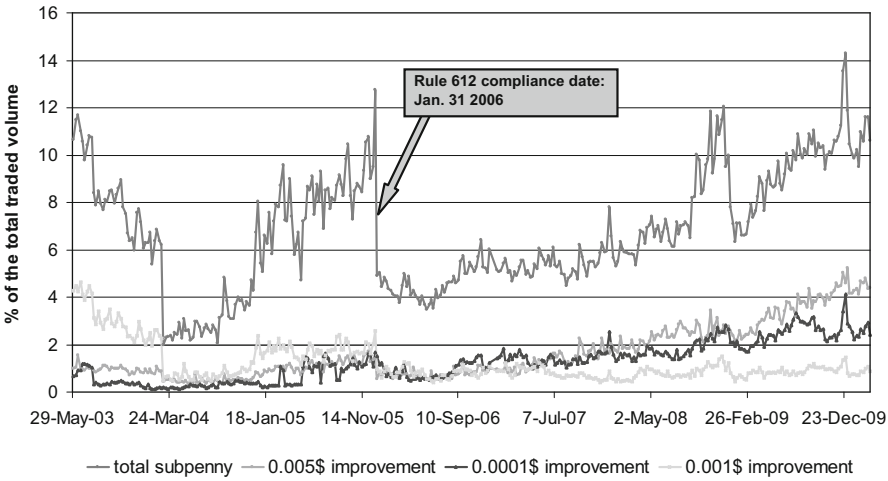


Fig. 7 Evolution of sub-penny trading for NASDAQ100 stocks

increment, and *0.0001 $ improvement* refers to the queue jumping trades. We also added a curve named *0.001 $ improvement* which adds up the third and the next to last bars.

The percentage drop from nearly 13% to less than 5% at the end of 2005 can be explained by the compliance date for the rule 612 (Minimum pricing increment). Since 2006 and the adoption of Rule 612, we can observe a general increasing tendency for the total, half a cent, and queue jumping curves. An interesting point to notice is that the curve named *0.001 $ improvement* seems stable or even slightly decreasing. This would tend to confirm the high frequency traders' viewpoint on the abusive useage of the queue jumping strategy. In order to confirm this fact, we study in the next paragraph the historical evolution of the price improvement distribution.

2.3.2 Evolution of the Price Improvement Distribution

This second graphic shows the evolution of the relative height of the first, third and mid bars. The figures are simply obtained by dividing the volume of each bars with the total volume traded in sub-penny.

What can be concluded from the linear regressions is that the relative growth of the queue jumping volume has been superior to the one of the total sub-penny volume. The global tendency of the evolution of the price improvement distribution can therefore be resumed by two stylized facts:

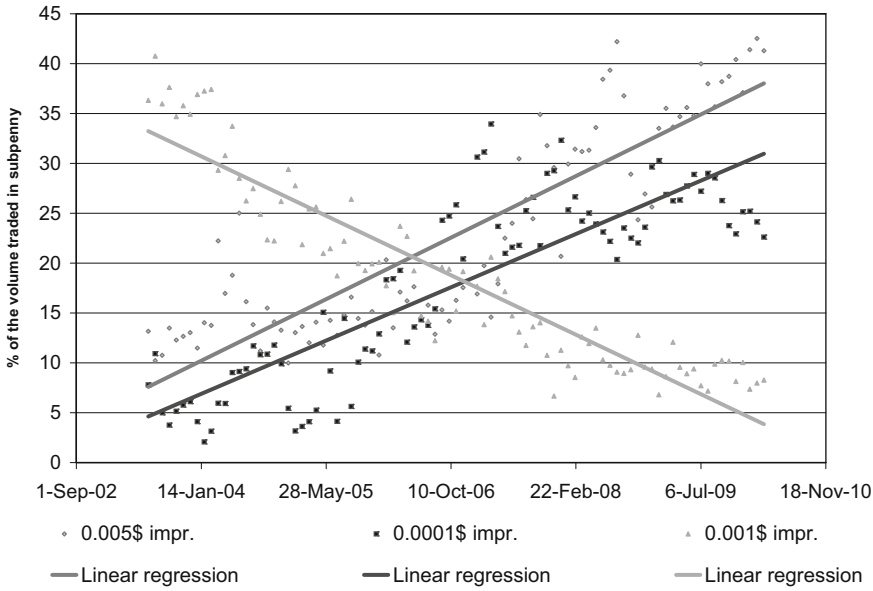


Fig. 8 Distribution evolution for NASDAQ100 stocks

- An increase of the relative weight of the central bar (midprice crossing).
- A concentration of the rest of the weight on the first and last bars (queue jumping).

This tendency is illustrated in the next figure, where we compare two price improvement distributions, one from 2004 and the other one from 2010. Whereas in 2004, queue jumping was made with a 0.1 cent of price improvement, it is now done with a 0.01 cent improvement.

3 Discussion

3.1 Minimum Price Variation

As empirical results show, sub-penny trading today is essentially used for midprice crossing and queue jumping. Queue jumping appears to be a strategy used to buy price priority at a negligible cost (1% of the tick). We have observed that this strategy decreases with the stock's price, and becomes very important (6% to 10 % of the total volume) for stocks between 1 $ and 5 $. The problem seems to come essentially from the fact that for these stocks, the MPV is an important barrier that artificially increases the relative effective spread.

Intuitively, there is a strong incentive to use queue jumping strategies when the relative effective spread becomes large enough. For stocks valued under 5 $, market making strategies are very profitable: there is an important competition between market makers, but since the spread is already reduced to the minimum price variation, this competition focuses on order priority and not on price improvement. The fact that not all actors are able to provide price improvement discourages some market makers (liquidity providers) and therefore diminishes the quality of the open order book. This is clearly explained in many comments received by the SEC, where the fact that sub-penny trading creates a "two-tier market" and discourages liquidity providers is pointed out and criticized.

The tick size is therefore at the center of the problem. It is a strong impediment for the "high frequency traders" of our classification, and an important advantage for the "broker-dealers" that are allowed to provide sub-penny price improvement. Two radical solutions have been put forward in the comments in order to improve the price formation process: either ban sub-penny for everyone (that could be achieved by implementing the "trade-at" rule proposed by the SEC, or by only allowing crossing at midpoint) or allow it for all actors (by, for example, reducing the MPV for stocks under $ 10).

3.2 Sub-penny for Everyone

The idea to reduce the minimum price variation for low priced stocks sounds especially appealing considering the graph shown in Fig. 3. As described in the previous paragraph, the competition between market makers is, in a way, saturated for low priced stocks. A reduction of the tick size could therefore foster competition on price improvement rather than on price priority.

This intuition is highlighted by the graph on Fig. 9. This graph was constructed in the same way as the one shown on Fig. 3, using the same data. The yellow squares are the exact same points as the ones on Fig. 3. They correspond to the mean of the queue jumping percentage. The triangles named "Mid crossing %" correspond to the percentage of the traded volume that was traded with a half a cent increment (which means prices from *.**495 to *.**505 due to our rounding procedure), and the squares named "Subpenny %" correspond to the total percentage of the volume traded in sub-penny.

This graph shows that for stocks over $ 10, sub-penny trading is mostly used for midpoint crossing. This type of trades have a true economical interest since both parties gains, the gains being equally distributed between the buyer and the seller. Only some 2 or 3 percent of the volume is traded with a price improvement under 1/20th of a cent, and if we consider this percentage to be acceptable, subpenny trading isn't an important issue for these stocks. The figures however are very different

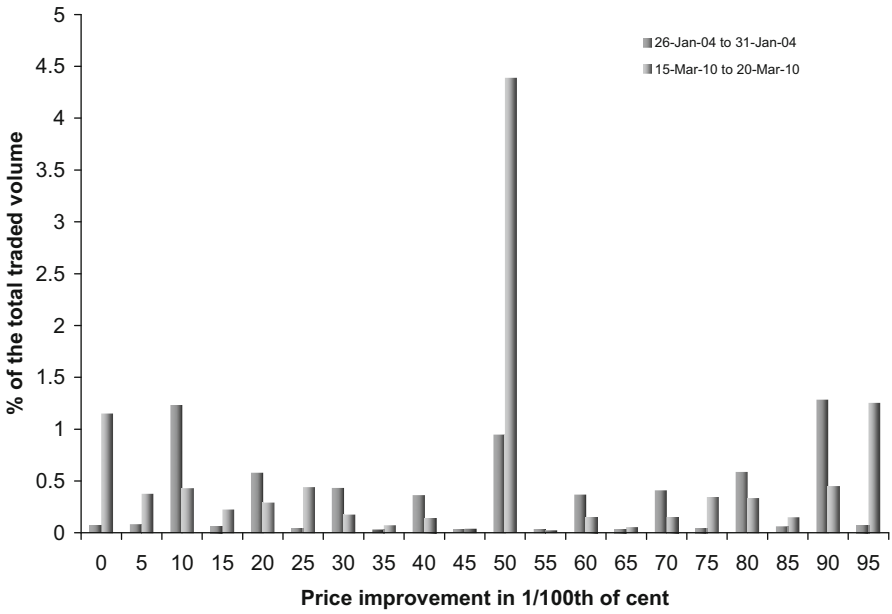


Fig. 9 2004 price improvement distribution compared to today's

for low priced stocks. Queue jumping becomes an important phenomena, whereas midprice crossing becomes negligible. We therefore observe a change of strategies:

- The relative weight of midprice crossing strategies decreases whereas the one of queue jumping strategies increases, accentuating the asymmetry of the distribution of the gains from the trade.
- Subpenny strategies become much more important.

The observation that the percentage of the volume traded in sub-penny increment increases when the stock value decreases highlights the fact that a one cent MPV is a constraining barrier for low priced stocks. This fact was experimentally tested in 2002 by Bidisha Chakrabarty and Kee H. Chung [1]. They compared spreads between six Electronic Communications Network (ECNs), three that allowed subpenny quotes (group S) and three that did not (group P).Their conclusion was that

> [...] for a sample of stocks that trade on all six of these ECNs [...] group S ECNs have narrower spreads than group P ECNs, especially for low-price stocks. Even after correcting for left-truncation and price discreteness [...] spreads for the same stocks are tighter on group S ECNs, suggesting that a smaller tick size fosters greater price competition [...] The one penny tick is frequently a binding constraint on the inside spread and the relaxation of the binding constraint would result in a 0.7 cent (16%) reduction in the inside spread.

One solution could therefore be to reduce this MPV for stocks priced under $ 10. The idea is simply to recreate the same conditions as for higher priced stocks, where open competition is enough to control sub-penny trading. For example, a 1/10th of a cent MPV for stocks under $ 10 would intuitively put those stocks in the same trading conditions as the ones priced between $ 10 and $ 100. Midprice crossing should benefit from this regulation change and queue jumping strategies become less important.

3.3 Banning Queue Jumping

The diminution of the MPV for low priced stocks does not come without problems, and could exacerbate arbitrage between exchanges with different fee models. For example, decreasing the MPV lower than twice the liquidity rebate given to liquidity providers by some exchanges would compromise the NBBO. An ask price inferior to the bid could indeed be observed without any possibility for those orders to be rerouted. The problem comes from the fact that today, exchanges fees are calculated proportionally to the traded volume. This could however be resolved simply by switching to rebates proportional to the traded volume price, or by making exchange fees proportional to the MPV.

Another solution could also be to only allow crossing to be done at midprice. The problem is that this would limit the possibilities of price improvement. For our

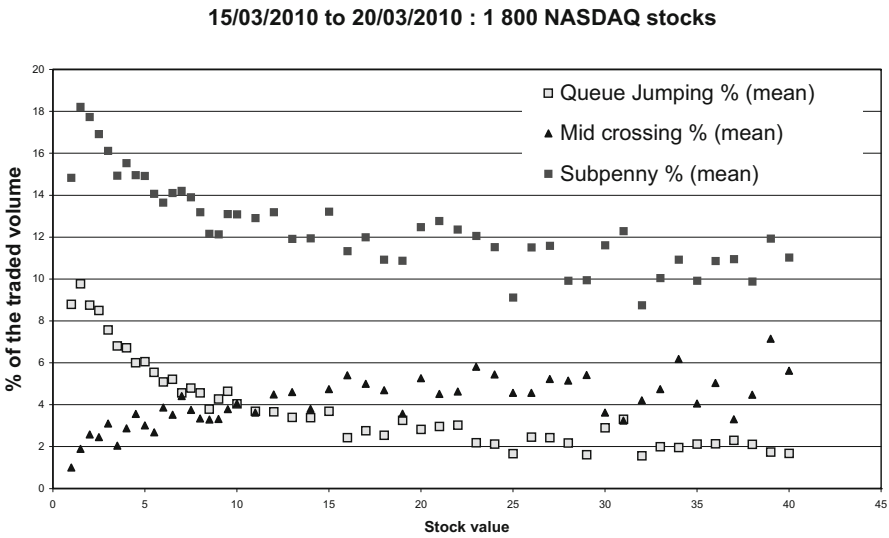


Fig. 10 Queue jumping and midprice crossing versus stock value

1 800 NASDAQ stock list, the price improvement due to non-midprice increment weights $ 2.8 million (whereas midprice crossing permitted a gain of $ 2.1 million). It is however difficult to quantify the loss due to the practice of queue jumping. The many actors that responded to the SEC's concept release evoked that queue jumping had a negative impact on the spread. It makes liquidity providing strategies less profitable, discouraging liquidity providers which automatically widens the spread. Banning this kind of practice would therefore have a positive impact of the spread, that could compensate the loss in price improvement.

4 Conclusions

Our historical results have shown a growing importance of the practice of sub-penny trading since January 2005 and the MPV rule compliance date. An even more disturbing fact is highlighted by the study of the evolution of the price improvement distribution: the growth of the relative weight of midprice crossing came along with a similar growth for queue jumping. Even if on average, the total price improvement has been increasing, the correlation between sub-penny volumes and stock value underlines the fact that the legal minimum price variation is a critical issue in this problem.

Our recommendation is therefore to lower the MPV for low priced stocks in order to let open competition reduce the problem of queue jumping. We suggest an MPV of 0.1 cent for stocks priced between $ 1 and $ 10.

We also are making available the complete data set[5] of the various statistics we used in this study. Weekly statistics on the traded volume and the price improvement, as well as the effective spread and the average trade price, are given for every stock appearing in the composition of the NASDAQ100 index since 2003. All formulas and explanations are given in the explanation sheet[6].

Acknowledgements We would like to thank F. Pomponio and the historical data team of BNP Paribas for guidance and support.

References

1. Bidisha Chakrabarty and Kee H. Chung. Can sub-penny pricing reduce trading costs? (2004) http://www.sec.gov/rules/proposed/s71004/bchakrabarty2.pdf
2. The Vanguard Group Inc. response to sec's concept release (2010) http://www.sec.gov/comments/s7-02-10/s70210-122.pdf
3. Fabrizio Pomponio and Frédéric Abergel. Trade-throughs: empirical facts and application to lead-lag measures, this volume
4. Securities and exchange commission. Comments on concept release on equity market structure (2010) http://www.sec.gov/comments/s7-02-10/s70210.shtml
5. Securities and exchange commission. Concept release on equity market structure (2010) http://www.sec.gov/rules/concept/2010/34-61358.pdf
6. Securities and exchange commission. Division of market regulation: Responses to frequently asked questions concerning rule 612 (minimum pricing increment) of regulation nms http://www.sec.gov/divisions/marketreg/subpenny612faq.htm
7. Bright Trading. Undisplayed trading centers compromising the nbbo through sub-penny trading (2010) http://www.sec.gov/comments/s7-02-10/s70210-63.pdf

[5] Available on Ecole Centrale's Chair of Quantitative Finance website: http://fiquant.mas.ecp.fr/Exchange/Subpenny/.

[6] http://fiquant.mas.ecp.fr/Exchange/Subpenny/data_explanation_sheet.pdf.

"Market Making" in an Order Book Model and Its Impact on the Spread

Ioane Muni Toke

Abstract. It has been suggested that marked point processes might be good candidates for the modelling of financial high-frequency data. A special class of point processes, Hawkes processes, has been the subject of various investigations in the financial community. In this paper, we propose to enhance a basic zero-intelligence order book simulator with arrival times of limit and market orders following mutually (asymmetrically) exciting Hawkes processes. Modelling is based on empirical observations on time intervals between orders that we verify on several markets (equity, bond futures, index futures). We show that this simple feature enables a much more realistic treatment of the bid-ask spread of the simulated order book.

1 Introduction

Arrival Times of Orders: Event Time Is Not Enough

As of today, the study of arrival times of orders in an order book has not been a primary focus in order book modelling. Many toy models leave this dimension aside when trying to understand the complex dynamics of an order book. In most order driven market models such as [1, 8, 18], and in some order book models as well (e.g. [21]), a time step in the model is an arbitrary unit of time during which many events may happen. We may call that clock *aggregated time*. In most order book models such as [7, 9, 19], one order is simulated per time step with given probabilities, i.e. these models use the clock known as *event time*. In the simple case where these probabilities are constant and independent of the state of the model, such a time treatment is equivalent to the assumption that order flows are homogeneous Poisson processes. A probable reason for the use of non-physical time in order

Ioane Muni Toke
Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
e-mail: ioane.muni-toke@ecp.fr

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

book modelling – leaving aside the fact that models can be sufficiently complicated without adding another dimension – is that many puzzling empirical observations (see e.g. [6] for a review of some of the well-known "stylized facts") can (also) be done in event time (e.g. autocorrelation of the signs of limit and market orders) or in aggregated time (e.g. volatility clustering).

However, it is clear that physical (calendar) time has to be taken into account for the modelling of a realistic order book model. For example, market activity varies widely, and intraday seasonality is often observed as a well known U-shaped pattern. Even for a short time scale model – a few minutes, a few hours – durations of orders (i.e. time intervals between orders) are very broadly distributed. Hence, the Poisson assumption and its exponential distribution of arrival times have to be discarded, and models must take into account the way these irregular flows of orders affect the empirical properties studied on order books.

Let us give one illustration. On Fig. 1, we plot examples of the empirical density function of the observed spread in event time (i.e. spread is measured each time an event happens in the order book), and in physical (calendar) time (i.e. measures are weighted by the time interval during which the order book is idle). It appears that density of the most probable values of the time-weighted distribution is higher than in the event time case. Symmetrically, the density of the least probable event is even smaller when physical time is taken into account. This tells us a few things about the dynamics of the order book, which could be summarized as follows: the wider the spread, the faster its tightening. We can get another insight of this empirical property by measuring on our data the average waiting time before the next event, conditionally on the spread size. When computed on the lower one-third-quantile (small spread), the average waiting time is 320 milliseconds. When computed on the upper one-third-quantile (large spread), this average waiting time is 200 mil-

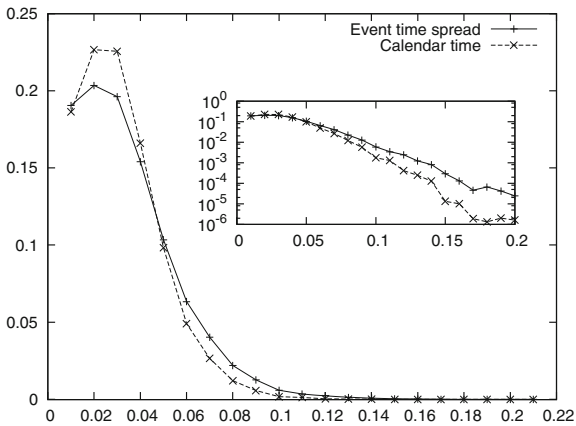


Fig. 1 Empirical density function of the distribution of the bid-ask spread in event time and in physical (calendar) time. In *inset*, same data using a semi-log scale. This graph has been computed with 15 four-hour samples of tick data on the BNPP.PA stock (see Sect. 2.1 for details)

liseconds. These observations complement some of the ones that can be found in the early paper [4].

Counting Processes with Dynamic Intensity

There is a trend in the econometrics literature advocating for the use of (marked) point processes for the modelling of financial time series. One may find a possible source of this interest in [10, 11], which introduce autoregressive conditional duration (ACD) and autoregressive conditional intensity (ACI) models. [12] fit that type of models on the arrival times of limit, market and cancellation orders in an Australian stock market order book.

A particular class of point processes, known as the Hawkes processes, is of special interest for us, because of its simplicity of parametrization. A univariate linear self-exciting Hawkes process $(N_t)_{t>0}$, as introduced by [14, 15], is a point process with intensity:

$$\lambda(t) = \lambda_0 + \int_0^t \nu(t-s)\mathrm{d}N_s, \tag{1}$$

where the kernel ν is usually parametrized as $\nu(t) = \alpha \mathrm{e}^{-\beta t}$ or in a more general way $\nu(t) = \sum_{k=1}^{p} \alpha_k t^k \mathrm{e}^{-\beta t}$. Statistics of this process are fairly well-known and results for a maximum likelihood estimation can be found in [20]. In a multivariate setting, mutual excitation is introduced. A bivariate model can thus be written:

$$\begin{cases} \lambda^1(t) = \lambda_0^1 + \int_0^t \nu_{11}(t-s)\mathrm{d}N_s^1 + \int_0^t \nu_{12}(t-s)\mathrm{d}N_s^2 \\ \\ \lambda^2(t) = \lambda_0^2 + \int_0^t \nu_{21}(t-s)\mathrm{d}N_s^1 + \int_0^t \nu_{22}(t-s)\mathrm{d}N_s^2. \end{cases} \tag{2}$$

The use of these processes in financial modelling is growing. We refer the reader to [3] for a review and [13] for a textbook treatment. In [5], a bivariate (generalized) Hawkes process is fitted to the time series of trades and mid-quotes events, using trading data of the General Motors stock traded on the New York stock Exchange. In [17] a ten-variate Hawkes process is fitted to the Barclay's order book on the London Stock Exchange, sorting orders according to their type and aggressiveness. It is found that the largest measured effect is the exciting effect of market orders on markets orders. [16] fits a bivariate Hawkes model to the time series of buy and sell trades on the EUR/PLN (Euro/Polish Zlotych) FX market. Using the simplest parametrization of Hawkes processes and some (very) constraining assumptions, some analytical results of trade impact may be derived. [2] fits a bivariate Hawkes process to the trade time series of two different but highly correlated markets, the "Bund" and the "Bobl" (Eurex futures on mid- and long-term interest rates).

Organization of This Paper

In this paper, we propose to enhance a basic order book simulator with arrival times of limit and market orders following mutually (asymmetrically) exciting Hawkes processes. Modelling is based on empirical observations, verified on several markets (equities, futures on index, futures on bonds), and detailed in Sect. 2. More specifically, we observe evidence of some sort of "market making" in the studied order books: after a market order, a limit order is likely to be submitted more quickly than it would have been without the market order. In other words, there is a clear reaction that seems to happen: once liquidity has been removed from the order book, a limit order is triggered to replace it. We also show that the reciprocal effect is not observed on the studied markets. These features lead to the use of unsymmetrical Hawkes processes for the design of an agent-based order book simulator described in Sect. 3. We show in Sect. 4 that this simple feature enables a much more realistic treatment of the bid-ask spread of the simulated order book.

2 Empirical Evidence of "Market Making"

2.1 Data and Observation Setup

We use order book data for several types of financial assets:

- BNP Paribas (RIC[1]: BNPP.PA): 7th component of the CAC40 during the studied period.
- Peugeot (RIC: PEUP.PA): 38th component of the CAC40 during the studied period.
- Lagardère SCA (RIC: LAGA.PA): 33th component of the CAC40 during the studied period.
- Dec.2009 futures on the 3-month Euribor (RIC: FEIZ9).
- Dec.2009 futures on the Footsie index (RIC: FFIZ9).

We use Reuters RDTH tick-by-tick data from September 10th, 2009 to September 30th, 2009 (i.e. 15 days of trading). For each trading day, we use only 4 hours of data, precisely from 9:30 am to 1:30 pm. As we are studying European markets, this time frame is convenient because it avoids the opening of American markets and the consequent increase of activity.

Our data is composed of snapshots of the first five limits of the order books (ten for the BNPP.PA stock). These snapshots are timestamped to the millisecond and taken at each change of any of the limits or at each transaction. The data analysis is performed as follows for a given snapshot:

1. if the transaction fields are not empty, then we record a market order, with given price and volume;

[1] Reuters Identification Code.

2. if the quantity offered at a given price has increased, then we record a limit order at that price, with a volume equal to the difference of the quantities observed;
3. if the quantity offered at a given price has decreased without any transaction being recorded, then we record a cancellation order at that price, with a volume equal to the difference of the quantities observed;
4. finally, if two orders of the same type are recorded at the same time stamp, we record only one order with a volume equal to the sum of the two measured volumes.

Therefore, market orders are well observed since transactions are explicitly recorded, but it is important to note that our measure of the limit orders and cancellation orders is not direct. In Table 1, we give for each studied order book the number of market and limit orders detected on our 15 4-hour samples. On the studied period, market activity ranges from 2.7 trades per minute on the least liquid stock (LAGA.PA) to 14.2 trades per minute on the most traded asset (Footsie futures).

Table 1 Number of limit and markets orders recorded on 15 samples of four hours (Sep 10th to Sep 30th, 2009; 9:30am to 1:30pm) for 5 different assets (stocks, index futures, bond futures)

Code	Number of limit orders	Number of market orders
BNPP.PA	321,412	48,171
PEUP.PA	228,422	23,888
LAGA.PA	196,539	9,834
FEIZ9	110,300	10,401
FFIZ9	799,858	51,020

2.2 Empirical Evidence of "Market Making"

Our idea for an enhanced model of order streams is based on the following observation: once a market order has been placed, the next limit order is likely to take place faster than usual. To illustrate this, we compute for all our studied assets:

- The empirical probability density function (pdf) of the time intervals of the counting process of all orders (limit orders and market orders mixed), i.e. the time step between any order book event (other than cancellation).
- The empirical density function of the time intervals between a market order and the immediately following limit order.

If an independent Poisson assumption held, then these empirical distributions should be identical. However, we observe a very high peak for short time intervals in the second case. The first moment of these empirical distributions is significant: one the studied assets, we find that the average time between a market order and the following limit order is 1.3 (BNPP.PA) to 2.6 (LAGA.PA) times shorter than the average time between two random consecutive events.

On the graphs shown in Fig. 2, we plot the full empirical distributions for four of the five studied assets[2]. We observe their broad distribution and the sharp peak for the shorter times: on the Footsie future market for example, 40% of the measured time steps between consecutive events are less that 50 milliseconds; this figure jumps to nearly 70% when considering only market orders and their following limit orders. This observation is an evidence for some sort of market-making behaviour of some participants on those markets. It appears that the submission of market orders is monitored and triggers automatic limit orders that add volumes in the order book (and not far from the best quotes, since we only monitor the five best limits).

In order to confirm this finding, we perform non-parametric statistical test on the measured data. For all four studied markets, omnibus Kolmogorov–Smirnov and Cramer–von Mises tests performed on random samples establish that the considered distributions are statistically different. If assuming a common shape, a Wilcoxon–Mann–Withney U test clearly states that the distribution of time intervals between

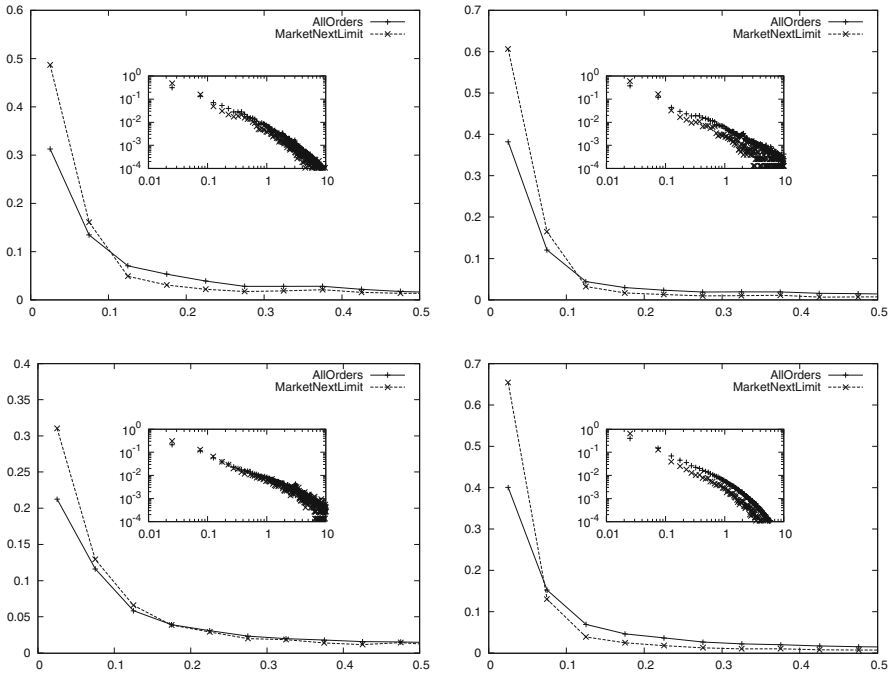


Fig. 2 Empirical density function of the distribution of the time intervals between two consecutive orders (any type, market or limit) and empirical density function of the distribution of the time intervals between a market order and the immediately following limit order. x-axis is scaled in seconds. In *insets*, same data using a log-log scale. Studied assets: BNPP.PA (*top left*), LAGA.PA (*top right*), FEIZ9 (*bottom left*), FFIZ9 (*bottom right*)

[2] Observations are identical on all the studied assets.

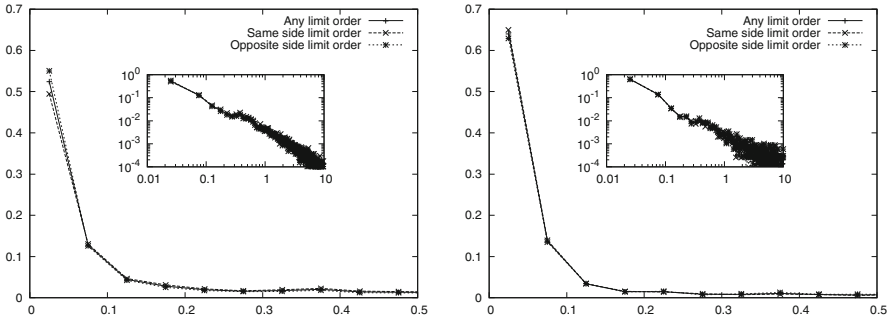


Fig. 3 Empirical density function of the distribution of the time intervals between a market order and the immediately following limit order, whether orders have been submitted on the same side and on opposite sides. x-axis is scaled in seconds. In *insets*, same data using a log-log scale. Studied assets: BNPP.PA (*left*), LAGA.PA (*right*)

a market orders and the following limit order is clearly shifted to the left compared to the distributions of time intervals between any orders, i.e. the average "limit following market" reaction time is shorter than the average time interval between random consecutive orders.

Note that there is no link with the sign of the market order and the sign of the following limit order. For example for the BNP Paribas (resp. Peugeot and Lagardere) stock, they have the same sign in 48.8% (resp. 51.9% and 50.7%) of the observations. And more interestingly, the "limit following market" property holds regardless of the side on which the following limit order is submitted. On Fig. 3, we plot the empirical distributions of time intervals between a market order and the following limit order, conditionally on the side of the limit order: the same side as the market order or the opposite one. It appears for all studied assets that both distributions are roughly identical. In other words, we cannot distinguish on the data if liquidity is added where the market order has been submitted or on the opposite side. Therefore, we do not infer any empirical property of placement: when a market order is submitted, the intensity of the limit order process increases *on both sides* of the order book.

This effect we have thus identified is a phenomenon of liquidity replenishment of an order book after the execution of a trade. The fact that it is a bilateral effect makes its consequences similar to "market making", event though there is obviously no market maker involved on the studied markets.

2.3 A Reciprocal "Market Following Limit" Effect?

We now check if a similar or opposite distortion is to be found on market orders when they follow limit orders. To investigate this, we compute for all our studied assets the "reciprocal" measures:

- The empirical probability density function (pdf) of the time intervals of the counting process of all orders (limit orders and market orders mixed), i.e. the time step between any order book event (other than cancellation).
- The empirical density function of the time step between a market order and the previous limit order.

As previously, if an independent Poisson assumption held, then these empirical distribution should be identical. Results for four assets are shown on Fig. 4. Contrary to previous case, no effect is very easily interpreted. For the three stocks (BNPP.PA, LAGA.PA and PEUP.PA (not shown)), it seems that the empirical distribution is less peaked for small time intervals, but difference is much less important than in the previous case. As for the FEI and FFI markets, the two distributions are even much closer. Non-parametric tests confirms these observations.

Performed on data from the three equity markets, Kolmogorov tests indicate different distributions and Wilcoxon tests enforce the observation that time intervals between a limit order and a following market order are stochastically larger than time intervals between unidentified orders. As for the future markets on Footsie (FFI) and 3-month Euribor (FEI), Kolmogorov tests does not indicate differences in

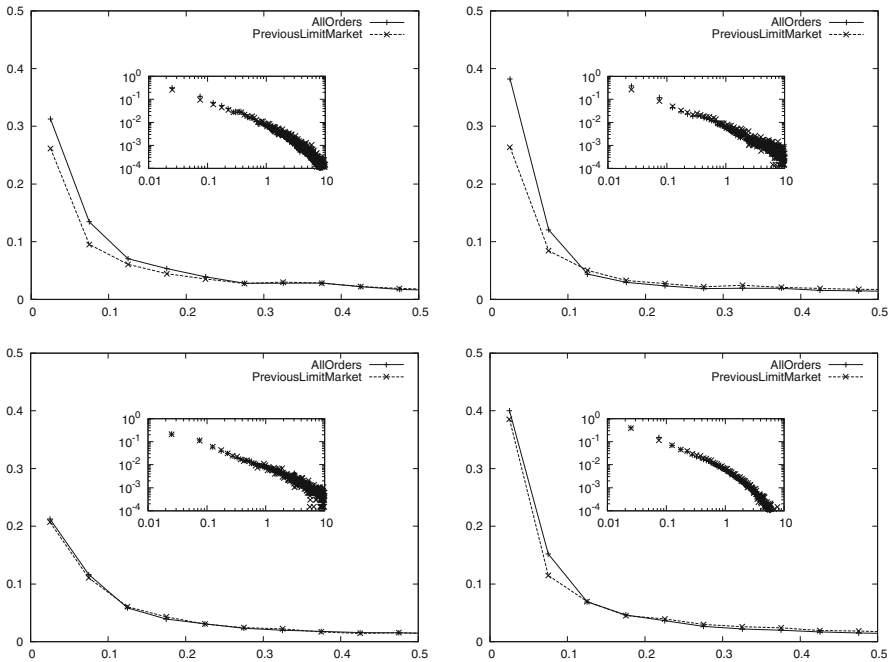


Fig. 4 Empirical density function of the distribution of the time intervals between two consecutive orders (any type, market or limit) and empirical density function of the distribution of the time intervals between a limit order and an immediately following market order. In *insets*, same data using a log-log scale. Studied assets: BNPP.PA (*top left*), LAGA.PA (*top right*), FEIZ9 (*bottom left*), FFIZ9 (*bottom right*)

the two observed distributions, and the result is confirmed by a Wilcoxon test that concludes at the equality of the means.

3 Order Book Models with Mutually Exciting Order Flows

Following these previous observations, we enhance a basic agent-based order book simulator with dependence between the flows of limit and market orders.

3.1 The Basic Poisson Model

We use as base model a standard zero-intelligence agent-based market simulator built as follows. One agent is a liquidity provider. This agent submits limit orders in the order books, orders which he can cancel at any time. This agent is simply characterized by a few random distributions:

1. submission times of new limit orders are distributed according to a homogeneous Poisson process N^L with intensity λ^L;
2. submission times of cancellation of orders are distributed according to homogeneous Poisson process N^C with intensity λ^C;
3. placement of new limit orders is centred around the same side best quote and follows a Student's distribution with degrees of freedom ν_1^P, shift parameter m_1^P and scale parameter s_1^P;
4. new limit orders' volume is randomly distributed according to an exponential law with mean m_1^V;
5. in case of a cancellation, the agent deletes his own orders with probability δ.

The second agent in the basic model is a noise trader. This agent only submits market order (it is sometimes referred to as the liquidity taker). Characterization of this agent is even simpler:

6. submission times of new market orders are distributed according to a homogeneous Poisson process N^M with intensity μ;
7. market orders' volume is randomly distributed according to an exponential law with mean m_2^V.

For all the experiments, agents submit orders on the bid or the ask side with probability 0.5. This basic model will be henceforth referred to as "HP" (Homogeneous Poisson).

Assumptions 1, 2 and 6 (Poisson) will be replaced in our enhanced model. Assumption 3 (Student) is in line with empirical observations in [19]. Assumptions 4 and 7 are in line with empirical observations in [6] as far as the main body of the distribution is concerned, but fail to represent the broad distribution observed in empirical studies. All the parameters except δ, which we kept exogenous, can be more

or less roughly estimated on our data. In fact δ is the parameter of the less realistic feature of this simple model, and is thus difficult to calibrate. It can be used as a free parameter to fit the realized volatility.

3.2 Adding Dependence Between Order Flows

We have found in Sect. 2.2 that market data shows that the flow of limit orders strongly depends on the flow of market order. We thus propose that in our experiment, the flow of limit and market orders are modelled by Hawkes processes N^L and N^M, with stochastic intensities respectively λ^L and μ defined as:

$$\begin{cases} \mu(t) = \mu_0 + \int_0^t \alpha_{MM} \mathrm{e}^{-\beta_{MM}(t-s)} \mathrm{d}N_s^M \\ \lambda^L(t) = \lambda_0^L + \int_0^t \alpha_{LM} \mathrm{e}^{-\beta_{LM}(t-s)} \mathrm{d}N_s^M + \int_0^t \alpha_{LL} \mathrm{e}^{-\beta_{LL}(t-s)} \mathrm{d}N_s^L. \end{cases} \tag{3}$$

Three mechanisms can be used here. The first two are self-exciting ones, "MM" and "LL". They are a way to translate into the model the observed clustering of arrival of market and limit orders and the broad distributions of their durations. In the empirical study [17], it is found that the measured excitation MM is important. In our simulated model, we will show (see 4.2) that this allows a simulator to provide realistic distributions of the durations of trades.

The third mechanism, "LM", is the direct translation of the empirical property we have presented in Sect. 2.2. When a market order is submitted, the intensity of the limit order process N^L increases, enforcing the probability that a "market making" behaviour will be the next event. We do no implement the reciprocal mutual excitation "ML", since we do not observe that kind of influence on our data, as explained in Sect. 2.3.

Rest of the model is unchanged. Turning these features successively on and off gives us several models to test – namely HP (Homogeneous Poisson processes), LM, MM, MM+LM, MM+LL, MM+LL+LM – to try to understand the influence of each effect.

4 Numerical Results on the Order Book

4.1 Fitting and Simulation

We fit these processes by computing the maximum likelihood estimators of the parameters of the different models on our data. As expected, estimated values varies

with the market activity on the day of the sample. However, it appears that estimation of the parameters of stochastic intensity for the MM and LM effect are quite robust. We find an average relaxation parameter $\hat{\beta}_{MM} = 6$, i.e. roughly 170 milliseconds as a characteristic time for the MM effect, and $\hat{\beta}_{LM} = 1.8$, i.e. roughly 550 milliseconds characteristic time for the LM effect. Estimation of models including the LL effect are much more troublesome on our data. In the simulations that follows, we assume that the self-exciting parameters are similar ($\alpha_{MM} = \alpha_{LL}$, $\beta_{MM} = \beta_{LL}$) and ensure that the number of market orders and limit orders in the different simulations is roughly equivalent (i.e. approximately 145000 limit orders and 19000 market orders for 24 hours of continuous trading). Table 2 summarizes the numerical values used for simulation. Fitted parameters are in agreement with an assumption of asymptotic stationarity.

Table 2 Estimated values of parameters used for simulations

Model	μ_0	α_{MM}	β_{MM}	λ_0^L	α_{LM}	β_{LM}	α_{LL}	β_{LL}
HP	0.22	–	–	1.69	–	–	–	–
LM	0.22	–	–	0.79	5.8	1.8	–	–
MM	0.09	1.7	6.0	1.69	–	–	–	–
MM LL	0.09	1.7	6.0	0.60	–	–	1.7	6.0
MM LM	0.12	1.7	6.0	0.82	5.8	1.8	–	–
MM LL LM	0.12	1.7	5.8	0.02	5.8	1.8	1.7	6.0

Common parameters: $m_1^P = 2.7, \nu_1^P = 2.0, s_1^P = 0.9$
$_1^V = 275, m_2^V = 380$
$\lambda^C = 1.35, \delta = 0.015$

We compute long runs of simulations with our enhanced model, simulating each time 24 hours of continuous trading. Note that using the chosen parameters, we never face the case of an empty order book. We observe several statistics on the results, which we discuss in the following sections.

4.2 Impact on Arrival Times

We can easily check that introducing self- and mutually exciting processes into the order book simulator helps producing more realistic arrival times. Fig. 5 shows the distributions of the durations of market orders (left) and limit orders (right). As expected, we check that the Poisson assumption has to be discarded, while the Hawkes processes help getting more weight for very short time intervals.

We also verify that models with only self-exciting processes MM and LL are not able to reproduce the "liquidity replenishment" feature described in Sect. 2.2. Distribution of time intervals between a market order and the next limit order are

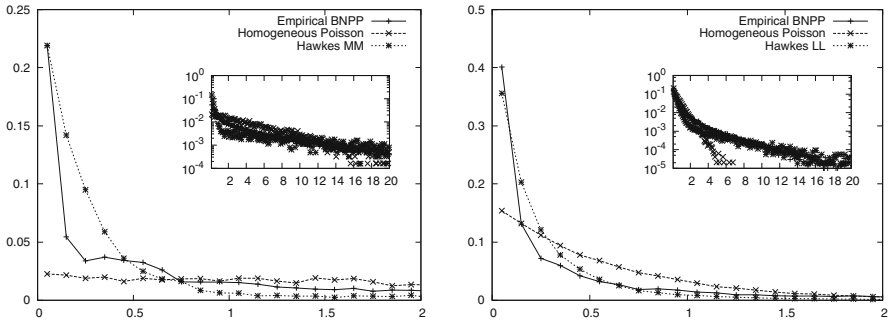


Fig. 5 Empirical density function of the distribution of the durations of market orders (*left*) and limit orders (*right*) for three simulations, namely HP, MM, LL, compared to empirical measures. In *inset*, same data using a semi-log scale

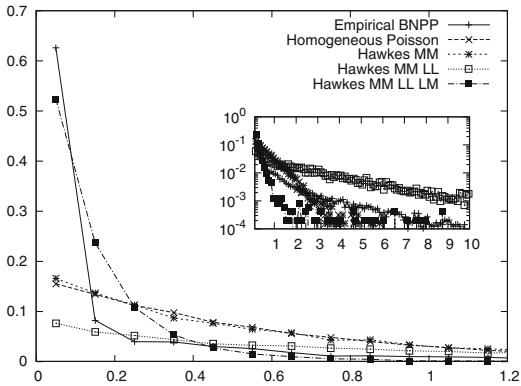


Fig. 6 Empirical density function of the distribution of the time intervals between a market order and the following limit order for three simulations, namely HP, MM+LL, MM+LL+LM, compared to empirical measures. In *inset*, same data using a semi-log scale

plotted on Fig. 6. As expected, no peak for short times is observed if the LM effect is not in the model. But when the LM effect is included, the simulated distribution of time intervals between a market order and the following limit order is very close to the empirical one.

4.3 Impact on the Bid-ask Spread

Besides a better simulation of the arrival times of orders, we argue that the LM effect also helps simulating a more realistic behaviour of the bid-ask spread of the order book. On Fig. 7, we compare the distributions of the spread for three models – HP, MM, MM+LM – in regard to the empirical measures. We first observe that the model with homogeneous Poisson processes produces a fairly good shape for

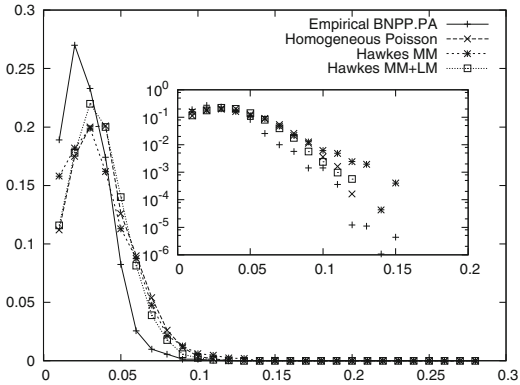


Fig. 7 Empirical density function of the distribution of the bid-ask spread for three simulations, namely HP, MM, MM+LM, compared to empirical measures. In *inset*, same data using a semi-log scale. x-axis is scaled in euro (1 tick is 0.01 euro)

the spread distribution, but slightly shifted to the right. Small spread values are largely underestimated. When adding the MM effect in order to get a better grasp at market orders' arrival times, it appears that we flatten the spread distribution. One interpretation could be that when the process N^M is excited, markets orders tend to arrive in cluster and to hit the first limits of the order book, widening the spread and thus giving more weight to large spread values. But since the number of orders is roughly constant in our simulations, there has to be periods of lesser market activity where limit orders reduce the spread. Hence a flatter distribution.

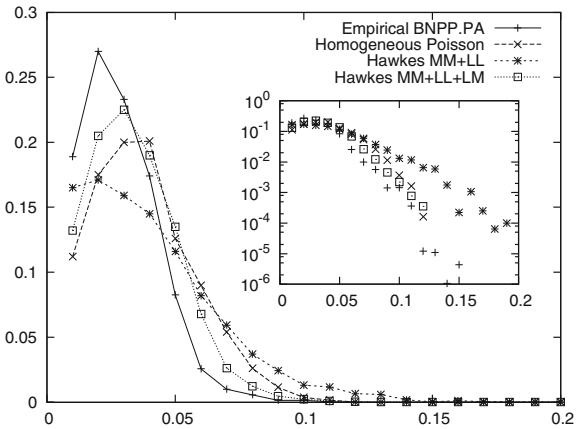


Fig. 8 Empirical density function of the distribution of the bid-ask spread three simulations, namely HP, MM, MM+LM, compared to empirical measures. In *inset*, same data using a semi-log scale. x-axis is scaled in euro (1 tick is 0.01 euro)

Here, the MM+LM model produces a spread distribution much closer to the empirical shape. It appears from Fig. 7 that the LM effect reduces the spread: the "market making" behaviour, i.e. limit orders triggered by market orders, helps giving less weight to larger spread values (see the tail of the distribution) and to sharpen the peak of the distribution for small spread values. Thus, it seems that simulations confirm the empirical properties of a "market making" behaviour on electronic order books.

We show on Fig. 8 that the same effect is observed in an even clearer way with the MM+LL and MM+LL+LM models. Actually, the spread distribution produced by the MM+LL model is the flattest one. This is in line with our previous argument. When using two independent self exciting Hawkes processes for arrival of orders, periods of high market orders' intensity gives more weight to large spread values, while periods of high limit orders' intensity gives more weight to small spread values. Adding the cross-term LM to the processes implements a coupling effect that helps reproducing the empirical shape of the spread distribution. The MM+LL+LM simulated spread is the closest to the empirical one.

4.4 A Remark on Price Returns in the Model

It is somewhat remarkable to observe that these variations of the spread distributions are obtained with little or no change in the distributions of the variations of the mid-price. As shown on Fig. 9, the distributions of the variations of the mid-price sampled every 30 seconds are nearly identical for all the simulated models (and much tighter than the empirical one). This is due to the fact that the simulated order books are much more furnished than the empirical one, hence the smaller standard deviation of the mid price variations. One solution to get thinner order books and

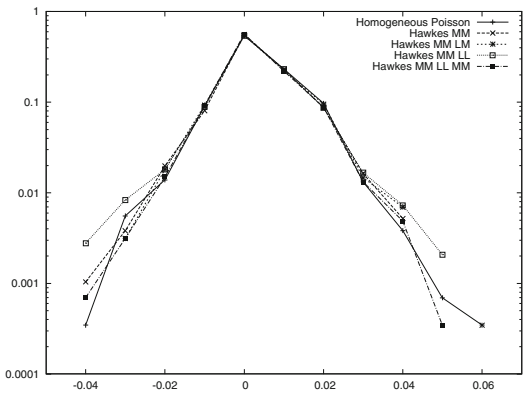


Fig. 9 Empirical density function of the distribution of the 30-second variations of the mid-price for five simulations, namely HP, MM, MM+LM, MM+LL, MM+LL+LM, using a semi-log scale. x-axis is scaled in euro (1 tick is 0.01 euro)

hence more realistic values of the variations of the mid-price would be to increase our exogenous parameter δ. But in that case, mechanisms for the replenishment of an empty order book should be carefully studied, which is still to be done.

5 Conclusions

We have shown the the use of Hawkes processes may help producing a realistic shape of the spread distribution in an agent-based order book simulator. We emphasize on the role of the excitation of the limit order process by the market order process. This coupling of the processes, similar to a "market making" behaviour, is empirically observed on several markets, and simulations confirms it is a key component for realistic order book models.

Future work should investigate if other processes or other kernels (ν_{LM} in our notation) might better fit the observed orders flows. In particular, we observe very short characteristic times, which should lead us to question the use of the exponential decay. Furthermore, as pointed out in the paper, many other mechanisms are to be investigated: excitation of markets orders, link with volumes, replenishment of an empty order book, etc.

References

1. Alfi, V., Cristelli, M., Pietronero, L., Zaccaria, A.: Minimal agent based model for financial markets i – origin and self-organization of stylized facts. The European Physical Journal B **67**(3), 13 pages (2009). DOI 10.1140/epjb/e2009-00028-4
2. Bacry, E.: Modeling microstructure noise using point processes. In: Finance and Statistics (Fiesta) Seminar, Ecole Polytechnique (2010)
3. Bauwens, L., Hautsch, N.: Modelling financial high frequency data using point processes. In: Handbook of Financial Time Series, pp. 953–979 (2009)
4. Biais, B., Hillion, P., Spatt, C.: An empirical analysis of the limit order book and the order flow in the paris bourse. The Journal of Finance **50**(5), 1655–1689 (1995). URL http://www.jstor.org/stable/2329330
5. Bowsher, C.G.: Modelling security market events in continuous time: Intensity based, multivariate point process models. Journal of Econometrics **141**(2), 876–912 (2007). DOI 10.1016/j.jeconom.2006.11.007. URL http://www.sciencedirect.com/science/article/B6VC0-4MV1P46-1/2/2b2c30c69257dc531c7f90517586b771
6. Chakraborti, A., Muni Toke, I., Patriarca, M., Abergel, F.: Econophysics: Empirical facts and agent-based models. 0909.1974 (2009). URL http://arxiv.org/abs/0909.1974
7. Challet, D., Stinchcombe, R.: Analyzing and modeling 1+ 1d markets. Physica A: Statistical Mechanics and its Applications **300**(1-2), 285–299 (2001)
8. Cont, R., Bouchaud, J.-P.: Herd behavior and aggregate fluctuations in financial markets. Macroeconomic Dynamics **4**(02), 170–196 (2000)
9. Cont, R., Stoikov, S., Talreja, R.: A Stochastic Model for Order Book Dynamics. SSRN eLibrary (2008)
10. Engle, R.F.: The econometrics of Ultra-High-Frequency data. Econometrica **68**(1), 1–22 (2000). URL http://www.jstor.org/stable/2999473

11. Engle, R.F., Russell, J.R.: Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. Journal of Empirical Finance **4**(2-3), 187–212 (1997). DOI 10.1016/S0927-5398(97)00006-6. URL http://www.sciencedirect.com/science/article/B6VFG-3SX0D5B-5/2/709b3c532191afd84b0dbffd14b040d8
12. Hall, A.D., Hautsch, N.: Modelling the buy and sell intensity in a limit order book market. Journal of Financial Markets **10**(3), 249–286 (2007). DOI 10.1016/j.finmar.2006.12.002. URL http://www.sciencedirect.com/science/article/B6VHN-4MT5K6F-1/2/1a71db7f8aab72c017e407033edc5734
13. Hautsch, N.: Modelling irregularly spaced financial data. Springer (2004)
14. Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. Biometrika **58**(1), 83–90 (1971). DOI 10.1093/biomet/58.1.83. URL http://biomet.oxfordjournals.org/cgi/content/abstract/58/1/83
15. Hawkes, A.G., Oakes, D.: A cluster process representation of a Self-Exciting process. Journal of Applied Probability **11**(3), 493–503 (1974). URL http://www.jstor.org/stable/3212693
16. Hewlett, P.: Clustering of order arrivals, price impact and trade path optimisation. In: Workshop on Financial Modeling with Jump processes, Ecole Polytechnique (2006)
17. Large, J.: Measuring the resiliency of an electronic limit order book. Journal of Financial Markets **10**(1), 1–25 (2007). DOI 10.1016/j.finmar.2006.09.001. URL http://www.sciencedirect.com/science/article/B6VHN-4M936YC-1/2/45b513a31fc6df3b6921987eaafded84
18. Lux, T., Marchesi, M., Italy, G.: Volatility clustering in financial markets. Int. J. Theo. Appl. Finance **3**, 675–702 (2000)
19. Mike, S., Farmer, J.D.: An empirical behavioral model of liquidity and volatility. Journal of Economic Dynamics and Control **32**(1), 200–234 (2008)
20. Ogata, Y.: On Lewis' simulation method for point processes. IEEE Transactions on Information Theory **27**(1), 23–31 (1981)
21. Preis, T., Golke, S., Paul, W., Schneider, J.J.: Multi-agent-based order book model of financial markets. EPL (Europhysics Letters) **75**(3), 510–516 (2006). URL http://www.iop.org/EJ/abstract/0295-5075/75/3/510/

Price-Time Priority and Pro Rata Matching in an Order Book Model of Financial Markets

Tobias Preis

Abstract. Using our recently introduced order book model of financial markets we analyzed two different matching principles for order allocation – price-time priority and pro rata matching. Price-time priority uses the submission timestamp which prioritizes orders in the book with the same price. The order which was entered earliest at a given price limit gets executed first. Pro rata matching is used for products with low intraday volatility of best bid and best ask price. Pro rata matching ensures constant access for orders of all sizes. We demonstrate how a multiagent-based model of financial market can be used to study microscopic aspects of order books.

1 Introduction

In recent years, econophysicists started to investigate and understand the price formation process in detail on a microscopic level. In this context, a statistical model of the continuous double auction [1, 2] was developed. Based on this model, we proposed an multiagent-based order book model recently. These Monte Carlo based simulations of financial markets' order books were introduced in [3] and studied in detail in [4]. Here we will provide simulation based evidence for two different order matching principles which can be found in order books of real exchanges[1].

The definition of the order book model and its main results are provided in Sect. 2. Sect. 3 will focus on Monte Carlo based simulations of the order book model using both price-time priority and pro rata matching. Sect. 4 summarizes our findings.

Tobias Preis
Center for Polymer Studies, Department of Physics, Boston, USA,
Artemis Capital Asset Management GmbH, Holzheim, Germany,
Institute of Physics, Johannes Gutenberg University Mainz, Germany,
e-mail: mail@tobiaspreis.de

[1] Supplementary information can be found on http://www.tobiaspreis.de.

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

2 Continuous Double Auction and the Definition of the Order Book Model

The order book model [3, 4] is based on microscopic structures which one can find at electronic financial markets. The function of an exchange based order book is to store buy orders and sell orders of all market participants. In our simulations, only one order book is used, in which one individual asset – e.g. a stock – is traded. There are various types of orders in financial markets' exchange systems. In our order book model, only the two most important types are implemented, namely limit orders and market orders. Limit orders are executed only at a specific price or a price, which is better for the trader, who placed this limit order in the book, whereas market orders are executed immediately against the best available limit order stored in the order book. Limit orders are added chronologically, which realizes a price time priority based matching algorithm. Thus, at a given price level with more than one limit order, the limit order, which was inserted first, has execution priority.

Limit and market orders are performed by agents in the model, in which we also distinguish between two types of market participants – liquidity providers and liquidity takers. These two groups of agents differ in the types of orders they are permitted to submit. On the one hand, N_A liquidity providers transmit only limit orders. In the case of a limit sell order, a liquidity provider offers an asset for sale at a set limit price or a higher price. Analogously, a limit buy order indicates a demand for buying an asset and the order is executed at a set limit price or any better price for the liquidity provider.

Let p_a be the best ask price, which is the lowest price level of all limit sell prices in the order book, and analogously p_b the best bid price, being the highest price level for which at least one limit buy order is stored in the order book. In the order book model, limit orders are placed around the midpoint $p_m = \frac{p_a+p_b}{2}$ with a rate α, i.e., $\alpha \cdot N_A$ new limit orders are submitted per time step. Let q_{provider} be the probability with which a limit order is a limit buy order. Thus, with probability $1 - q_{\text{provider}}$, the limit order to be placed is a limit sell order. The liquidity provider, which can be identified as market maker, supplies liquidity to the market in order to exploit the spread $s = p_a - p_b$: such market participants intend, e.g., to sell an asset at price p_a or higher and then to buy it back at price p_b or lower. Thus, they have earned at least the spread s. As seen in this example, short sales are allowed, i.e., it is allowed for agents to sell assets even if they do not possess them.

On the other hand, there are N_A liquidity takers, who transmit only market orders. These market orders are submitted with rate μ, i.e., $\mu \cdot N_A$ market orders are inserted per time step into the order book. A market order will be immediately executed after arrival. A market sell order is executed at price level p_b, a market buy order at price level p_a. A market order of a liquidity taker is a market buy order with probability q_{taker} and a market sell order with probability $1 - q_{\text{taker}}$. In this basic version of the order book model, the simple case $q_{\text{provider}} = q_{\text{taker}} = \frac{1}{2}$ is applied. Thus, all orders will be produced symmetrically around the midpoint. In practice, it is possible, that the limit price of a limit sell order is lower than the current best

bid price and the limit price of a limit buy order is higher than the current best ask price. Such "crossing" limit orders degenerate to market orders and thus, they are executed immediately. In our order book model, only pure limit orders will be used.

Limit orders, which are stored in the order book, can also expire or can be deleted. In the model, this canceling process is realized in the way that each stored order is deleted with probability δ per time unit. As there are overall $2N_A$ agents in the multiagent system, each Monte Carlo step (MCS) consists of $2N_A$ moves. In each move, one agent is randomly selected and can perform one action according to the probability rates. If the chosen agent is a liquidity provider, then a limit order with probability α is submitted by the chosen agent. On the other hand, if the selected agent is a liquidity taker, then a market order with probability μ is placed in the order book which will be immediately executed. Orders in our model have the constant order volume 1. Thus, it is possible only to buy or sell one asset unit with an individual order.

Based on this simple rules, first an unrealistic independent identically distributed order placement depth can be applied. This is realized in the way that limit buy orders are entered on each price level in the interval of $[p_a - 1 - p_{\text{int}}; p_a - 1]$ with the same probability, and accordingly, limit sell orders are transmitted uniformly distributed in the interval of $[p_b + 1; p_b + 1 + p_{\text{int}}]$. Already with this definition of the order book model, profits and losses of the agents can be analyzed. Using this setup, the averaged wealth value of liquidity takers and liquidity providers drifts apart linearly in time [3]. Comparing these results with real financial markets, it has to be stated that liquidity takers are systematically disadvantaged in comparison to liquidity providers. The distinction in our model between the two groups of market participants reflects the two types of orders. In general, market participants are not restricted to one order type in reality.

In the next step, a more realistic order placement depth will be integrated in the order book model. The order book depth of real financial markets can be described by a log-normal distribution [5]. And, to take this into account the independent identically distributed limit order placement is replaced by an exponentially distributed order placement depth. Thus, for placing a limit order, the limit price p_l is determined for a limit buy order through

$$p_l = p_a - 1 - \eta \tag{1}$$

and for a limit sell order according to

$$p_l = p_b + 1 + \eta \tag{2}$$

whereby η is an exponentially distributed integer random number created by $\eta = \lfloor -\lambda_0 \cdot \ln(x) \rfloor$ with x being a uniformly distributed random number in the interval $[0; 1)$ and $\lfloor z \rfloor$ denoting the integer part of z. With this construction, the submission process of limit orders has the tendency to reduce the gap between best bid price and best ask price. Also, crossing limit orders are avoided, as the price of a limit buy order cannot be equal or larger than the best ask price and the price of a limit sell order cannot be equal or lower than the best bid price.

As result of applying the exponential order placement rule, a log-normally distributed order book depth profile is obtained [3, 4]. This basic version is used in order to study both order matching algorithms. This simple variant is also able to reproduce the results of [1, 2]. The price time series of this basic version possesses an antipersistent price behavior on short time scales which is due to the order book structure. On medium and long time scales the Hurst exponent converges towards a diffusive regime. The price change distributions exhibit an almost Gaussian shape. The model, which is characterized by a symmetry created by identical buy and sell probabilities, describes a stationary market. However, when one additionally introduces a symmetry disturbance, the order book model is displaced from its stationary state. This extension is implemented by a temporal modulation of the buy probability q_{taker} of the liquidity takers or the buy probability q_{provider} of the liquidity providers [3]. Qualitatively identical results are achieved, if both probabilities are modulated independently of each other. Employing a feedback random walk to introduce micro market trends into the market, one additionally obtains a persistent price behavior on medium time scales. However, no fat tails can be reproduced with such a symmetry-breaking extension of the order book model. When one furthermore couples the characteristic order placement depth to the prevailing market trend, widened price change distributions are achieved, with so-called fat tails. Thus, with these extensions of our order book model, we could demonstrate that the generation of a nontrivial Hurst exponent is independent of the generation of fat tails [4]. This disproves the implication which can be often found in the literature that a persistent price behavior corresponds to non-Gaussian price changes. Furthermore, we are able to support the statement in [6, 7] that $H > 1/2$ implies not necessarily long time correlations.

3 Matching Principles

When orders are entered into the electronic order book, they are sorted by type, price, and submission time. Market orders are always given the highest priority for matching purposes. Orders at a given price level are aggregated, although the number of orders remains unknown. Market participants only see the specific details of their own limit orders[2]. Most exchange traded products – e.g. equity index derivatives at the European Exchange (EUREX) in Germany – follow the matching principle which is known as price-time priority. This is not the case for money market products which show typically smaller intraday fluctuations. These products follow pro rata matching.

Price-time priority can be described as follows. When an order is entered into the order book, it is assigned a timestamp with a resolution of milliseconds. This timestamp is used to prioritize orders in the book with the same price. The order which was entered earliest at a given price limit gets executed first.

[2] More information can be found, e.g., on www.eurexchange.com.

Pro rata matching is used for products with low intraday volatility of best bid or best ask price. Pro rata matching ensures constant access for orders of all sizes. Otherwise, using price-time priority a large order may prevent smaller orders from participating in the order matching process. When matching existing orders in the electronic order book against an incoming order, the pro rata matching algorithm takes into account every book order at best bid or best ask price according to its percentage of the overall volume bid or offered at the price. Its timestamp is neglected. Thus, pro rata principles avoid conflicts in priority between orders with small and large volumes [8].

In this section, we study the consequences of both matching algorithms using our order book model. The price-time priority based matching algorithm was already implemented in the basic version of the order book model [3, 4]. Without loss of generality, we study in both versions the simple case that the order volume is set to 1 – for the price-time priority and for the pro rata matching. However, in this case it is not possible to observe the situation that a large order prevent smaller orders from participating in the matching process. Thus, one can not distinguish orders based on volume. Additionally, pro rata matching ignores timestamps. In our framework with constant order volume the matching process can be realized by executing a randomly chosen limit order at the best available price (best bid or best ask) if there are more than one limit orders at this price level.

Before we compare both order matching algorithms, we have to think about an appropriate macroscopic variable for that purpose. As the change of matching algorithms does not effect the times and sales records, it is not useful to analyze the price time series. Only variables referring to execution times of individual orders are useful to study. Thus, we will analyze the time-to-fill distributions of limit orders. The time-to-fill quantity T_i of an individual limit order i is given by the time interval which starts with the submission of the limit order to the central order book at time t_i^l and which ends with its execution at time t_i^e. The time-to-fill T_i is given by

$$T_i = t_i^e - t_i^l. \tag{3}$$

As the order volume is set to 1 one has not to handle the special cases of partial executions. Deleted limit orders do not contribute to the time-to-fill distributions. In order to measure the time-to-fill distributions, we have to store for each limit order in the order book the submission timestamp t_i^l. Thus, we can determine the time difference T_i. Please note that also market orders are neglected for the calculation of this distribution as they have an execution time interval of 0 MCS by definition – of course, this is not true if the order book is empty.

We choose the same parameters for the simulation of the order book model as used in [3, 4]: $\alpha = 0,15$, $\mu = 0,025$, $\delta = 0,025$, $\lambda = 100$, and $N_A = 250$. In Fig. 1a, the time-to-fill distribution of limit orders is shown for this parameter set for both order allocation methods. The calculation of the time-to-fill distributions is based on simulations lasting 10^5 MCS. Results are averaged over 10 runs. In Fig. 1a, one can clearly see that the price-time priority matching algorithm has a larger probability in comparison to the pro rata based allocation only for $T \in [10; 50]$. This

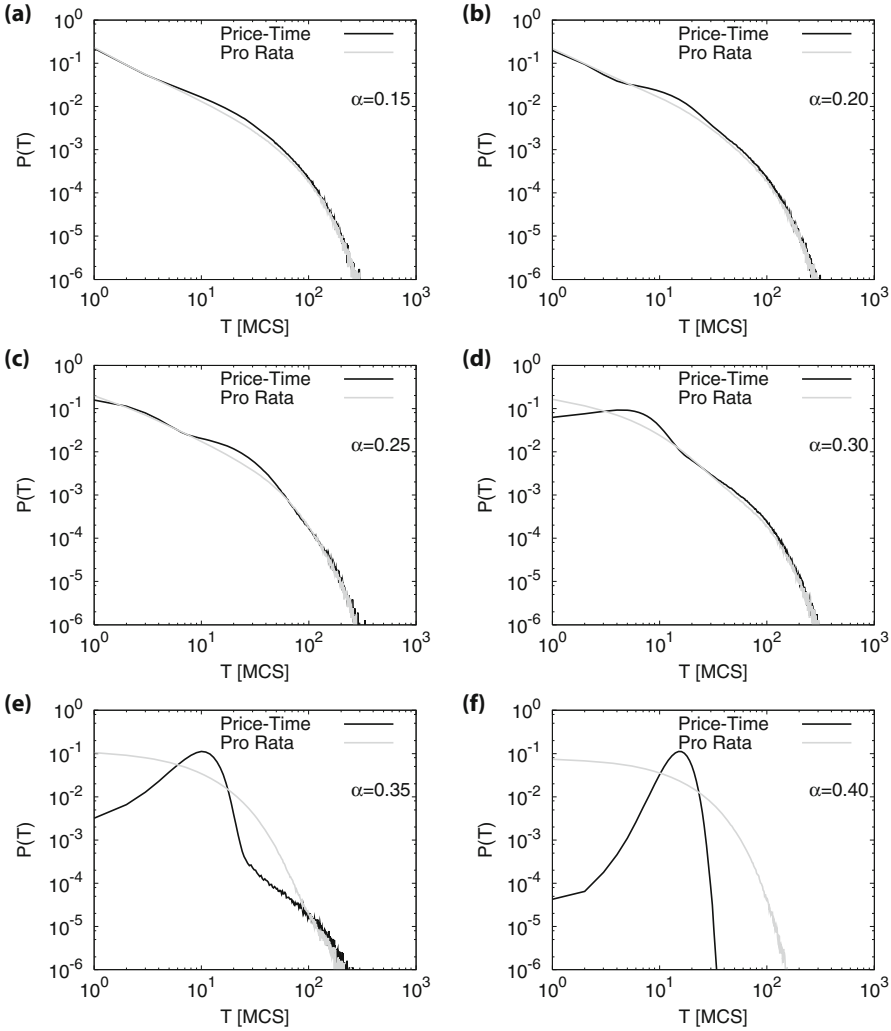


Fig. 1 Time-to-fill distributions of a price-time priority and pro rata based order matching for (a) $\alpha = 0,15$, (b) $\alpha = 0,20$, (c) $\alpha = 0,25$, (d) $\alpha = 0,30$, (e) $\alpha = 0,35$, (f) $\alpha = 0,40$

effect will be analyzed in more detail when we start to "freeze" the order book stepwise, i.e., when we reduce the volatility of best bid and best ask price.

Based on the results which were obtained by the parameter space analysis [4], freezing can be realized increasing the limit order rate α. If the market order rate μ is constant and we increase the limit order rate, then more and more limit orders are placed around around the midpoint. In the end of this process, α is so large that the number of market orders per time unit is too less in order to change the best bid price and the best ask price. Thus, the last traded price is jumping from the constant best

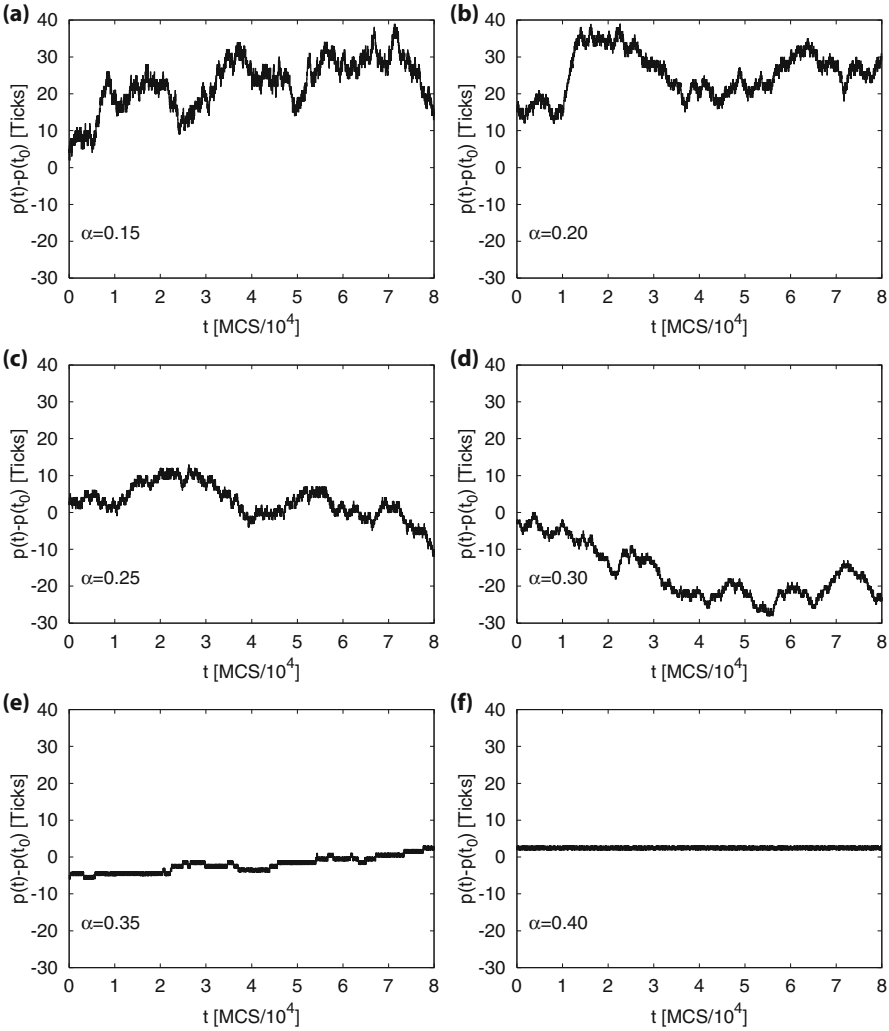


Fig. 2 Freezing of the price time series: Subsets comprising 8×10^4 MCS for (a) $\alpha = 0,15$, (b) $\alpha = 0,20$, (c) $\alpha = 0,25$, (d) $\alpha = 0,30$, (e) $\alpha = 0,35$, (f) $\alpha = 0,40$

bid price to the constant best ask price and vice versa. Fig. 2 shows such a stepwise freezing of the price time series for various values of α. The corresponding time-to-fill distributions are shown in Fig. 1. The larger α the larger is the qualitative change of the distribution shape. Using a pro rata allocation method, we find in all cases a strictly monotonic decreasing distribution. However, if we apply a price-time priority allocation method in our order book, we end up with a distribution which has a distinct maximum located at $T > 1$. This supports in an impressive way why exchange operators are using pro rata matching algorithms for low volatility products

as it guarantees more fairness. On the other hand, the question arises whether price-time priority should be replaced by pro rata matching in general as the time-to-fill profile is not changing when we use the pro rata approach. This result supports the tendency in the USA in recent years. There, a larger number of products obeys pro rata matching.

In case of a completely frozen order book and price-time priority (see Fig. 1f and Fig. 2f), a limit buy or limit sell order has to pass the whole queue at best bid or best ask before it can be matched with an arriving market order. If we use pro rata matching the limit orders are randomly selected. Thus, strictly monotonic decreasing shape of the distribution persists also for a frozen order book.

4 Conclusions

Based on our recently introduced order book model of financial markets [3, 4] we analyzed two different matching principles for order allocation – price-time priority and pro rata matching. Price-time priority uses the submission timestamp which prioritizes orders in the book with the same price. The order which was entered earliest at a given price limit gets executed first. Pro rata matching is used for products with low intraday volatility of best bid and best ask price. Pro rata matching ensures constant access for orders of all sizes. The results obtained from simulations of the order book model show that the larger the limit order rate the larger is the qualitative change of the time-to-fill distribution shape. Using a pro rata allocation method, we find in all cases a strictly monotonic decreasing distribution. However, if we apply a price-time priority allocation method in our order book, we end up with a distribution which has a distinct maximum located at $T > 1$.

References

1. Daniels MG, Farmer JD, Gillemot L, Iori G, Smith E (2003) Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. Physical Review Letters 90:108102
2. Smith E, Farmer JD, Gillemot L, Krishnamurthy S (2003) Statistical theory of the continuous double auction. Quantitative Finance 3:481–514
3. Preis T, Golke S, Paul W, Schneider JJ (2006) Multi-agent-based order book model of financial markets. Europhysics Letters 75:510–516
4. Preis T, Golke S, Paul W, Schneider JJ (2007) Statistical analysis of financial returns for a multiagent order book model of asset trading. Physical Review E 76:016108
5. Maslov S, Mills M (2001) Price fluctuations from the order book perspective – empirical facts and a simple model. Physica A 299:234–246
6. Bassler KE, Gunaratne GH, McCauley JL (2006) Markov processes, Hurst exponents, and nonlinear diffusion equations: With application to finance. Physica A 369:343–353
7. McCauley JL, Gunaratne GH, Bassler KE (2007) Hurst exponents, Markov processes, and fractal Brownian motion. Physica A 379:1–9
8. Preis T (2011) Ökonophysik – Die Physik des Finanzmarktes. Gabler Research, Springer Fachmedien Wiesbaden

High-Frequency Simulations of an Order Book: a Two-scale Approach

Charles-Albert Lehalle, Olivier Guéant and Julien Razafinimanana

1 Introduction

Models of market microstructure at the order book scale can be split into two families:

- First, the agent-based models [5] aiming at simulating a large number of agents, each of them having its utility function or feedback rule. The philosophy of this kind of modelling is similar to Minsky's paradigm in artificial intelligence in the eighties: build each agent so that if you stealthily replace, one by one, each real person interacting in the market with such a virtual ersatz, you will finally obtain a full synthetic replica of a real market. The actual limits faced by this research programme are: first, the difficulty to rationalise and quantify the utility function of real persons, and then the computational capabilities of today's computers. Last but not least, the lack of analytical results of this fully non-parametric approach is also a problem for a lot of applications. It is, for instance, usually difficult to know how to choose the parameters of such models to reach a given intra-day volatility, given sizes of jumps, average bid-ask spread, etc.
- Second, the "zero intelligence" models [9] aiming at reproducing stylised facts (Epps effect on correlations, signature plot of volatility, order book shapes, etc.) using random number generators for time between orders, order sizes, prices, etc.

Charles-Albert Lehalle
Head of Quantitative Research, Crédit Agricole Cheuvreux, Courbevoie, France,
e-mail: clehalle@cheuvreux.com

Olivier Guéant
MFG R&D, Paris, France,
e-mail: olivier.gueant@post.harvard.edu

Julien Razafinimanana
Crédit Agricole Quantitative Research, Courbevoie, France
e-mail: jrazafinimanana@cheuvreux.com

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

This approach is more oriented to "knowledge extraction" from existing recordings than the agent-based one. Its focus on stylized facts and our capability to emulate them using as simple as possible generators is close to the usual definition of "knowledge" (following for instance Kolmogorov or Shannon in terms of complexity reduction). It succeeds in identifying features like short-term memory, Epps effect on correlations, signature plots for high-frequency volatility estimates, dominance of power laws [25], and the general profile of market impact [11], among others, that are now part of the usual benchmarks to validate any microscopic market model. The limits of this approach are: first, the usual stationarity assumptions that are made, and the difficulty of linking the microscopic characteristics with macroscopic ones, for instance linking characteristics of the underlying probability distributions to market volatility (even if recent advances have been made in this direction using Hawkes models [2] or usual distributions [7]). The search for such links is motivated by the fact that as they are probability-based, their diffusive limits (or equivalent) should behave similarly to usual quantitative models on a large scale (for instance Levy processes [24]).

In this paper we present an approach that endeavours to take the best aspects of the two previous ones proposing a "zero-intelligence" microscopic model "pegged" at an agent-based one. We use new approaches to answer to the classical drawbacks described earlier:

- The computational need and the lack of closed form formula is solved via the use of the recent advances in the Mean Field Games (MFG) mathematical theory [18]. This enables a continuum of agents to be considered without the need to simulate each of them, but "simply" solving (stochastic) Partial Differential Equations (PDEs): one forward PDE (for the transport of the mass of the agents and of their views) and backward Hamilton-Jacobi-Bellman PDEs (to determine optimal strategies according to the agents' utility functions).
- The usual stationarity assumptions of the zero-intelligence models is here replaced with a conditioning of the probability distributions used (for order sizes, prices, arrival rates, etc.) by characteristics of the distance between agents' views and the realised microscopic order books.

Before few comments on the layout of this paper, let's just underline two aspects of the work presented here: the motivations behind the development of such a new model and a real life illustration of it.

Motivations of Such a Two-layered Model

As this research is mainly funded by Crédit Agricole Cheuvreux (the broker arm of Crédit Agricole's investment bank), it has been guided by a practical need that other approaches do not answer. The main practical use of the models obtained inside this new framework is to **test trading algorithms**. As electronic brokerage execution and proprietary high-frequency trading are used more extensively month after month [16], the usual drawbacks of zero intelligence and agent-based models

prevent their wide use in testing such trading strategies. To expose the needs for such testing, here is a brief summary of the essence of trading algorithms. Such an algorithm lives in a state space made up of three kinds of variables:

- The **instantaneous market data** X_t: this is usually the state of the Limit Order Book (LOB) on several trading venues, trades when they occur, news, etc.
- With the increase in high-frequency trading activity (around 250 updates per second of the LOB on liquid European stocks, an increase in the usual size of iceberg orders of 30% in the last three years, mirrored orders, etc.) a trading algorithm has to filter the instantaneous market data X_t to build **its view on the state of the market**: Y_t. It will include some real-time analytics as estimates of the instantaneous volatility, usual betas (or covariance of the stocks against synthetic market axes), estimates of liquidity imbalances (between buyers and sellers), etc. The internal view on the market Y_t takes the market data into account via an updating rule H_θ:

$$Y_t = H_\theta(Y_{t-\delta t}, X_t)$$

  where θ is a set of static parameters of the algorithm (like its risk aversion, window sizes, etc.). The functional H_θ is clearly part of the strategy of the algorithm. The behaviour of such views as stochastic processes obtained by an updating rule can be studied theoretically thanks to the theory of stochastic algorithms (see [21] for a detailed application on "Dark Liquidity seeking").
- And **the internal state of the algorithm** Z_t containing for instance its inventory, its pending orders in the books, its risk constraints, etc.

The trading algorithm has to make decisions based on available information (i.e. its view and its state):

$$D_t = F_\theta(Y_t, Z_t).$$

This decision-making process can embed computations like the one of the reservation price for market making algorithms [1], or optimal liquidation trajectories [3].

Trading between $t = 0$ and $t = T$, an algorithm tries to maximise the expectation of a functional of its trajectory and terminal state:

$$V_T(H, F, \theta) = \mathbb{E}\left(\int_{t=0}^{T} g(Z_t)\,\mathrm{d}t + G(Z_T)\right)$$

choosing at least the best value for θ, but also choosing the proper class for functionals H (view on market) and F (decision process).

This type of maximisation is common in quantitative finance, especially in derivative pricing [22], and we know that the "greeks" of $V_T(H, F, \theta)$ with respect to some market characteristics are needed to understand the obtained optimal trajectories. They measure the sensitivity of V_T to changes in state variables of the

market. For derivatives pricing, they are usually the Delta (sensitivity to the underlying price), Vega (sensitivity to volatility), Theta (sensitivity to time to maturity) and Gamma (sensitivity of the Delta to the underlying price). For algorithmic trading, we propose to use at least the following:

- Vega – sensitivity to intra-day volatility:

$$\mathscr{V} = \frac{\partial V_T(H, F, \theta)}{\partial \sigma}.$$

- Psi – sensitivity to bid-ask spread:

$$\Psi = \frac{\partial V_T(H, F, \theta)}{\partial(\text{bid-ask spread})}.$$

- Phi – sensitivity to trading frequency:

$$\Phi = \frac{\partial V_T(H, F, \theta)}{\partial(\text{trading frequency})}.$$

- Iota – sensitivity to imbalance between sell and buy orders:

$$\iota = \frac{\partial V_T(H, F, \theta)}{\partial(\text{sell vs buy imbalance})}.$$

To obtain such greeks, one has to be able to calibrate the market model to obtain a given volatility, spread, or trading frequency, which is not that easy for classic zero-intelligence or agent-based models.

The model proposed here can consequently be used at order one (i.e. to obtain back-tests: estimates of V_T) but also for higher order estimates that are needed for stress tests or sensitivity analyses.

A Real-life Equivalent of This Model

As this is a project on-going research, the model presented in Sect. 1 and 2 below is an application of our two-layer framework (a zero-intelligence model pegged at a MFG agent-based one) to a stylised real-life situation: **a market in which access to order books is provided by only one pure agency broker**. In such a market, the portfolio managers, the proprietary traders and all other investors make decisions using information on the value of quoted assets they all share. The state space of their decisions will be named the **market of the views** (or *viewed market*) and the consolidation of all their interests will be called the **order book of the views** (or *viewed order book*). Each investor will consequently adjust continuously his view according to the feedback he has from the consolidation of the whole market via the order book of the views. In the proposed framework, the order book of the views is continuous in prices, quantities and time. The **real market** and **real market order book** are discrete in prices (tick size), quantities and time (event based). The dynamics of the views will be modelled by a mean field model, as presented in Sect. 1.

To implement their views on the real market, investors will continuously mandate a pure agency broker that will buy or sell stocks on the **real market order book** in order to comply with its clients' views. The behaviour of the pure agency broker is here modelled by a zero intelligence model whose probability distributions are conditioned by the distance between the real market order book and the order book of the views. From a qualitative real-life perspective, the broker will act optimally from a best execution point of view, as described by the actual European or US regulators [20]. Since there is only one broker in the example described in Sect. 2, the latter will act with respect to the distance between the real order book and the order book of all the views.

This paper follows the description of the two layers: Sect. 1 presents the macroscopic agent-based MFG layer. It outlines how the MFG framework can deduce (stochastic) PDEs from a setting with a continuum of agents: one forward PDE describing the transport of the mass of the agents in their state space (the market of the views) and backward equations implementing the real-time optimisation of the utility functions of the agents. **In the specific example presented in this paper, the agents are of three kinds: mean reverters, trend followers and noise traders**. Since in the specific case presented here each agent is anticipating of the future value of the stock with respect to his style of investing (i.e. a continuation of the price move for the trend followers, an alternation around its moving average for the mean reverters, and an unexpected one for the noise traders), there will be no backward induction to determine the views and hence no Hamilton-Jacobi-Bellman PDEs. The remaining forward PDE therefore fully describes the behaviour of agents within their state space. Once the dynamics of the views of the investors are defined, Sect. 2 describes how the zero-intelligence layer is conditioned by the distance between the order book of the views and the real order book. This section is illustrated by analyses of real datasets to support some of the assumptions made. As the development of this two-layer framework is on-going research, the conclusion gives some guidelines on our future work and main directions.

2 The Macroscopic Scale: a Mean Field Game Model of Agents' Views

The macroscopic scale of the model is based on a price formation model developed by Jean-Michel Lasry and Pierre-Louis Lions in [18]. This price formation model was built at the margin of a new framework in economic modelling associated with mean field games. Mean field games enable modelling of economic situations in which a large number of agents interact with one another and have strategic behaviours. Here, the interactions occur in the order book and the number of agents is arguably large enough for the mean field game framework to be applied. However, as in the model presented in [18] and in the papers following this seminal

one, we are not going to use, in this specific paper[1], the optimisation dimension of mean field games, remaining therefore in line with the classic mean field approach of econophysicists.

In what follows, we are first going to present the mean field model as it appears in [18] and as it has been studied in [12, 13, 19]. Then, we will propose a dynamic and random version of the model that allows us to model limit order book dynamics.

2.1 Presentation of the Theoretical Framework

Limit order books are made up of buy and sell orders, and we adopt a continuous version of each side of the limit order book. Namely, we introduce two functions that we assume to be smooth, $(t, p) \mapsto m_B(t, p)$ and $(t, p) \mapsto m_A(t, p)$, representing, at time t, the density (number) of buy (resp. sell) orders at price p.

These densities represent the willingness of buyers and sellers to exchange the stock under consideration, and we will denote $p^*(t)$ the agreement price[2]. We suppose that agents modelled by the two densities are subject to signals that make them move their orders in the limit order book. These signals are not going to be modelled *per se*. Rather, we suppose, in relation to what may be inferred from real data[3], that orders evolve according to a seemingly random process.

Hence, execution will take place when buy orders meet sell orders (both of them being moved by the random process), that is when m_B and m_A intersect.

In the initial model (see [18]), the authors introduced a trading cost a and assume that each buyer (resp. seller) reintroduces a sell (resp. buy) order at price $p^*(t) + a$ (resp. $p^*(t) - a$).

In this context, limit order books are modelled by a coupled system of free-boundary evolution PDEs:

$$\partial_t m_B\,(t, p) - \frac{\varepsilon^2}{2}\partial^2_{pp} m_B(t, p) = \lambda(t)\delta_{p=p^*(t)-a}$$

$$\partial_t m_A\,(t, p) - \frac{\varepsilon^2}{2}\partial^2_{pp} m_A(t, p) = \lambda(t)\delta_{p=p^*(t)+a}$$

with

$$m_B(t, p) > 0, \forall p < p^*(t) \quad m_B(t, p) = 0, \forall p \geq p^*(t)$$

$$m_A(t, p) > 0, \forall p > p^*(t) \quad m_A(t, p) = 0, \forall p \leq p^*(t)$$

[1] Although our generic framework is aimed at using the whole MFG framework.

[2] $p^*(t)$ will be defined more precisely hereafter.

[3] We know from our dataset that a very large proportion of limit order book modifications are not due to trades but are rather due to cancel orders, insert orders or order updates. For instance, there were 934,763 changes in the first five limits of the order books for Total SA on May 5th 2010 and only 36,310 of these were actually due to a trade. Similarly, for France Telecom, the respective figures are 386,896 changes in limit order books for 21546 trades.

where ε stands for the standard deviation of the random process that moves the orders and where $\lambda(t)$ is the flow of sell orders that meet buy orders. This flow is

$$\lambda(t) = -\frac{\varepsilon^2}{2}\partial_p m_B\left(t, p^*(t)\right) = \frac{\varepsilon^2}{2}\partial_p m_A\left(t, p^*(t)\right).$$

The last equation represents the supply/demand equilibrium condition in written form. Since the slopes of m_B and m_A are the same (up to the sign), we introduce a regular function m defined as:

$$m(t, p) = \begin{cases} m_B(t, p)\ , & \text{if } p \leq p^*(t) \\ -m_A(t, p)\ , & \text{if } p > p^*(t) \end{cases}$$

m satisfies a unique parabolic equation:

$$\partial_t m\,(t, p) - \frac{\varepsilon^2}{2}\partial^2_{pp} m(t, p) = -\frac{\varepsilon^2}{2}\partial_p m\left(t, p^*(t)\right)\left(\delta_{p=p^*(t)-a} - \delta_{p=p^*(t)+a}\right) \quad (1)$$

and the limit conditions are $m(0, \cdot)$ given on the domain $[p_{\min}, p_{\max}]$ and, for instance, Neumann conditions at $p_{\min}$ and $p_{\max}$.

This equation has been studied in [12,13,18,19] and the main results (under very light and natural assumptions or slight modifications of the problem) are that the problem is well formulated, that the solutions globally exists and there is a single solution.

In what follows, we will modify this price formation equation to model the dynamics of limit order books.

2.2 Modelling Limit Order Books and Price Dynamics

Although the preceding model was continuous and there was no bid/ask spread, we see it as a good macroscopic model for limit order books. The reason for that lies in the dynamics of the free-boundary.

Since the equilibrium price $p^*(t)$ is implicitly defined by $m(t, p^*(t)) = 0$, the dynamics of $p^*(t)$ are characterised by:

$$\frac{\mathrm{d}p^*(t)}{\mathrm{d}t} = -\frac{\partial_t m\,(t, p^*(t))}{\partial_p m\,(t, p^*(t))} = -\frac{\varepsilon^2}{2}\frac{\partial^2_{pp} m\,(t, p^*(t))}{\partial_p m\,(t, p^*(t))}.$$

This equation says that **the dynamics of the price are given by the shape of the limit order book** modelled by m in a way that is in accordance with practitioners' experience. In spite of this feature, the model is far too smooth to be able to model properly the fast dynamics of limit order books and to replicate price volatility. In order to replicate real limit order books, we will modify the model in several directions.

First, we will no longer consider that orders are reinserted in the limit order book according to a transaction cost rule. Instead, we will consider that each executed order corresponds to an agent that can either be, as explained above, a noise trader, a trend follower or a mean reverter, depending on the past dynamics of the price. In other words we have:

$$\partial_t m(t,p) - \frac{\varepsilon^2}{2}\partial^2_{pp} m(t,p) = -\frac{\varepsilon^2}{2}\partial_p m\left(t, p^*(t)\right)[\text{source}(t,p)] \tag{2}$$

where the source term is complex and depends on $\{p^*(s)/s < t\}$.

As mentioned in the introduction, we have three types of agents:

- **Trend-followers** who buy when the price has increased and sell if the price has decreased. These agents thus compare the price at time t and the price T seconds ago (that is $p^*(t-T)$). These agents reinsert orders symmetrically: for instance, if the price has increased, we believe that a trend follower bought a stock and reinserted it in the ask part of the limit order book at price $p^*(t) + [p^*(t) - p^*(t-T)]$.
- **Mean-reverters** who buy when, on average, the price has moved down and, similarly, sell when, on average, the price has moved up. These agents thus compare the price at time t and the average price during the last T seconds[4] (that is $\bar{p}_T(t) = \frac{1}{T}\int_{t-T}^{t} p^*(s)ds$). As above, reinsertion of orders is made symmetrically to the reference price $\bar{p}_T(t)$.
- **Noise traders** that reinsert orders according to the current distribution of orders.

In this context, the partial differential equation has a stationary solution[5]:

$$m(p) = -\frac{\pi}{p_{\max} - p_{\min}} \sin\left(\frac{\pi}{p_{\max} - p_{\min}}(p - p^*)\right)$$

with $p^* = \frac{p_{\min}+p_{\max}}{2}$.

This stationary solution will be used as the initial solution of the problem after our second modification which consists of inserting noise on the limit order book. To model the insertion of new orders and the cancellation of existing ones, we indeed **add noise and we end up with the following (stochastic) equation**:

$$\mathrm{d}m = \frac{\varepsilon^2}{2}\partial^2_{pp} m(t,p)\mathrm{d}t - \frac{\varepsilon^2}{2}\partial_p m\left(t, p^*(t)\right)[\text{source}(t,p)]\,\mathrm{d}t + \nu g(p, p^*(t))\mathrm{d}W_t^p$$

with $m(0,p) = -\frac{\pi}{p_{\max}-p_{\min}} \sin\left(\frac{\pi}{p_{\max}-p_{\min}}(p - \frac{p_{\min}+p_{\max}}{2})\right)$ and Neumann conditions at the frontier.

[4] Time horizons for both the trend-followers and the mean-reverters are distributed according to a Gamma distribution to emulate agents with different time horizons.

[5] We choose the stationary solution with mass 1 on each side of the order book.

ν is here the intensity of the noise and $g(p, p^*)$ is a smooth function[6] with $g(p^*, p^*) = g(p_{\min}, p^*) = g(p_{\max}, p^*) = 0$ and chosen so that m does not vanish or change sign without any economic reason (in practice the noise under consideration is bounded in addition to the above hypothesis and the bound takes the form of an absolute cap at $\varphi \times |m(t, p)|^{0.4} m_0(p)^{0.6}$ where m_0 is the uniform distribution on $[p_{\min}, p_{\max}]$).

2.3 Calibration and Confrontation to Reality

To calibrate the model, that is seen as a model to describe the views on the order book, we need to choose the domain $[p_{\min}, p_{\max}]$ and the values of the parameters ε, ν and φ. In this text, we are going to focus on the two parameters ε and φ that are the most important two in the calibration.

Our model provides a dynamic description for both the views and the price. Hence, our goal will be to reproduce the actual volatility of prices σ_p and the dynamics of the limit order book around the equilibrium price $p^*(t)$. For this second purpose, we recall the (still valid) formula $\frac{\mathrm{d}p^*(t)}{\mathrm{d}t} = -\frac{\varepsilon^2}{2}\partial_p \ln\left(-\partial_p m\left(t, p^*(t)\right)\right)$ that invites us to introduce the "slope" $\ell(t) = -\partial_p m\left(t, p^*(t)\right)$. The "volatility" of $\ell(t)$, denoted σ_ℓ, will serve as a proxy for the dynamics of the limit order books.

The calibration process is therefore aimed at targeting real couples (σ_p, σ_ℓ) while varying the values of ε and φ. Intuitively, this may be possible[7] since the roles played by ε and φ are the same as far as the price is concerned although an increase in ε smooths[8] m whereas an increase in φ does the contrary.

These variations can be summed up in the following table:

	ε	φ
$\sigma_p(\varepsilon, \varphi)$	$\uparrow$	$\uparrow$
$\sigma_\ell(\varepsilon, \varphi)$	$\downarrow$	$\uparrow$

To calibrate the model we need to "invert"[9] the function $(\varepsilon, \varphi) \mapsto (\sigma_p, \sigma_\ell)$ in the domain of couples (σ_p, σ_ℓ) characterising the real markets.

[6] In practice we took

$$g(p, p^*) = \underbrace{\frac{1}{p^{*2.5}}}_{\text{normalization}} \times \underbrace{\sqrt{|p - p^*|}}_{\text{no noise at } p^*} \times \underbrace{(p - p_{\min})(p_{\max} - p)}_{\text{no noise at the boundaries}} .$$

[7] It's very important to be able to drive the model with a chosen volatility to be able to carry out a sensitivity analysis as described in the introduction.

[8] This is a classic regularisation result.

[9] This is not a proper inversion since the dynamic is stochastic. We could have carried out Monte-Carlo simulations but we decided to focus on couples (ε, φ) inducing low variance in the resulting volatility couples.

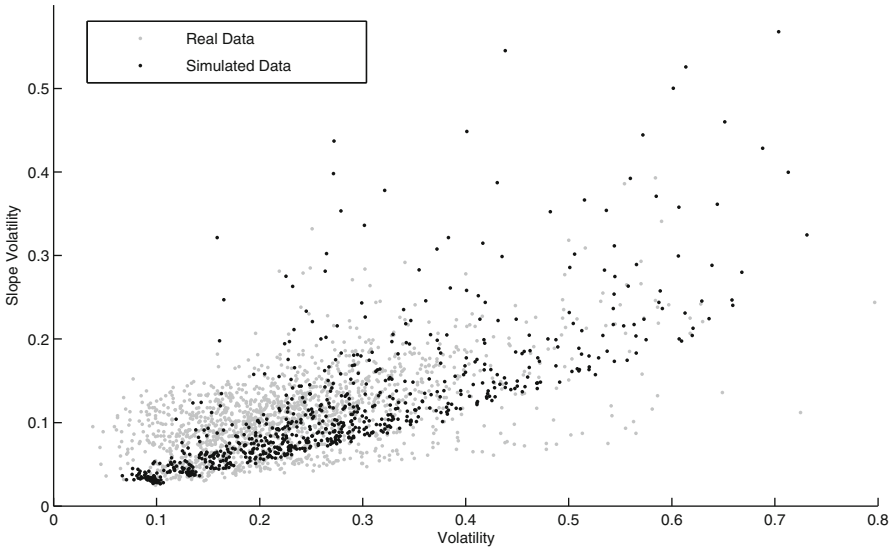


Fig. 1 Calibration

We calculated the values of (σ_p, σ_ℓ) for the 40 stocks in the CAC 40 index during May and June 2010 using the first five limits of order books and we simulated around 700 "stable" simulations[10]. The result is displayed below:

Although it is difficult to replicate some points[11], a wide range of couples[12] can be obtained using the right choices of ε and φ.

Hence, if one wants to (re-)play a day characterised by a given couple (σ_p, σ_ℓ) then one has to choose[13] a couple (ε, φ) whose stability will guarantee a result close to the targeted value (σ_p, σ_ℓ). The corresponding simulation will provide series of prices, sampled at any frequency, along with series of continuous limit order books.

These time series, and more exactly the times series $(p^*(t), \ell(t))$, will serve as a target time series to generate orders at the microscopic scale (see below).

To illustrate the price obtained at the macroscopic scale, we provide the following plot where the average five-minute Garman-Klass volatility is around 25% (in annualised terms) and where the excess kurtosis (for 1-minute return) is just below 5 (Fig. 2).

[10] We dropped the less stable couples (ε, φ).

[11] At least by a sufficiently "stable" couple (ε, φ).

[12] The most dense part of the "real" data set is well replicated.

[13] The inversion procedure is not detailed here and can be a tabulation or a linearised inversion.

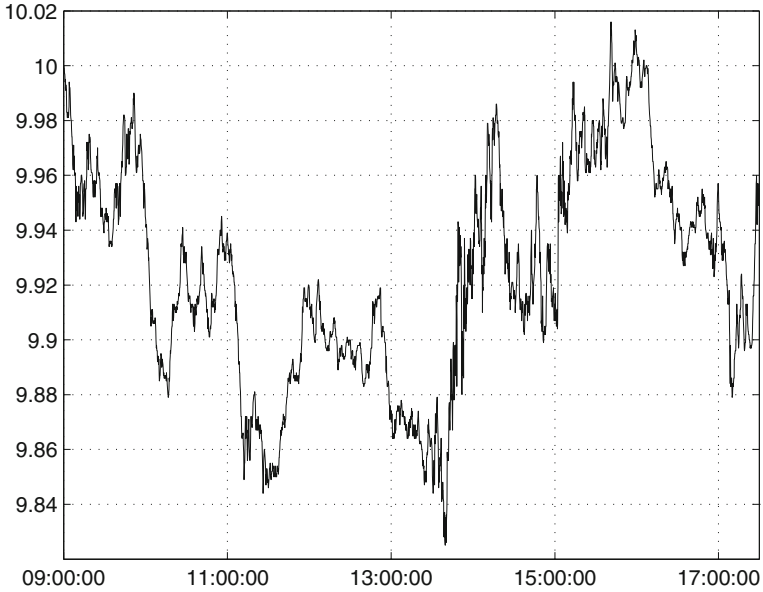


Fig. 2 Simulated price series

3 The Microscopic Scale: a Conditioned Zero-intelligence Model

3.1 Main Principles

Three elements are needed to implement a model of the characteristics of orders sent to a limit order book (LOB) based on conditional distributions:

- a conditioning variable C;
- a marginal law $\mathscr{L}_\theta$, where θ are the parameters of the law;
- a relationship between the parameters and the conditioning variable:

$$\theta = g(C). \tag{3}$$

The considered characteristics of an order are:

- the time between two consecutive orders;
- the type of each order, i.e. either *market* (liquidity consuming order), *cancel* or *limit* (liquidity providing order, or *resting order*) and buy (i.e. *ask*) or sell (i.e. *bid*);
- and the size and price of each sent order.

Conditioning Variables

In the example of one broker taking the routing decisions for all investors' orders with respect to its clients' views, it is natural to condition the characteristics of the orders to distance measurements between the execution conditions (i.e. the investors' views) and the real state of the market. In our model it will be read as the distance between the *order book of the views* and the *real order book*.

Thanks to the macroscopic layer, the continuous order book of the views is known via the density $m(t, p)$ of aggregated views at price p (i.e. the signed relative mass of investors having interests at price p). The state of the real order book is discrete in price (because of the tick size δp) and in volume (i.e. a number of shares is an integer). The notation $\tilde{m}(t, p)$ will be used for the interests of investors as disclosed on the real order book. If the two order books (the one of the views and the real one) are perfectly aligned, the following stationary master equation is verified ($\lfloor \cdot \rfloor$ is the rounding function):

$$\tilde{m}(t, p) = \left\lfloor \int_{p-\delta p/2}^{p+\delta p/2} m(t, p)\, \mathrm{d}p \right\rfloor \tag{4}$$

$\tilde{p}_t^*$ will be used for the instantaneous equilibrium price of the real order book, while p_t^* is the instantaneous equilibrium price of the views as defined by the MFG layer.

In the microscopic model presented here, the characteristics of orders sent to the real market by the broker will mainly be conditioned by the values of $\tilde{p}_t^*$ and p_t^* (first order condition) and of the slopes of $\tilde{m}_t$ and m_t around $\tilde{p}_t^*$ and p_t^* respectively (second order condition). To justify these choices, let's note $Q(p; \tilde{p}^*, p^*, \tilde{m}, m)$ the probability for a resting order at price p to be executed given $\tilde{p}^*$, p^*, $\tilde{m}$, and m.

Assumption 1 (First order assumption) *When the price p of a resting order is higher (respectively lower) than the current market price $\tilde{p}^*$ (i.e. it is a sell order, resp. a buy order), then its probability to be executed $Q(p)$ decreases (resp. increases) with the signed distance $\tilde{p}^* - p^*$ (when p^* moves):*

$$(p - \tilde{p}^*) \cdot \frac{\partial Q(p; \tilde{p}^*, p^*, \tilde{m}, m)}{\partial(\tilde{p}^* - p^*)} \leq 0. \tag{5}$$

Assumption 2 (Second order assumption) *When the price p of a resting order is close to the current market price $\tilde{p}^*$ and when the price of the viwes p^* is close to $\tilde{p}^*$, then its probability to be executed $Q(p)$ increases with the signed distance $\tilde{\ell}(\tilde{p}^*) - \ell(p^*)$, where $\ell(p)$ is the slope of the order book of the views at price p and $\tilde{\ell}(p)$ its counterpart in the real order book:*

$$\frac{\partial Q(p; \tilde{p}^*, p^*, \tilde{m}, m)}{\partial(\tilde{\ell}(\tilde{p}^*) - \ell(p^*))} \geq 0. \tag{6}$$

This last assumption comes from a more obvious one:

$$(p - \tilde{p}^*) \cdot \frac{\partial Q(p; \tilde{p}^*, p^*, \tilde{m}, m)}{\partial(\tilde{m}(p) - m(p))} \leq 0$$

and the Taylor expansion of $\tilde{m}(p) - m(p)$ around $\tilde{p}^*$ and p^* with the assumption that $\tilde{p}^*$ is close to p^*.

Marginal Laws

As stated before, parametrised probability distributions will be needed for the time between two consecutive orders, the type of orders and the size and price of the sent order. Taking our assumptions into account, the parameters of these laws will be functionals of the instantaneous distance between the equilibrium prices $\Delta p_t^* = \tilde{p}_t^* - p_t^*$ and between the slopes at these prices $\Delta \ell_t = \tilde{\ell}(\tilde{p}^*) - \ell(p^*)$.

The choice of specific distributions for these characteristics will be illustrated later in this paper using some real data and exploratory simulations.

Time between consecutive orders – It is natural to take a *Poisson law* $\mathrm{d}\nu_t(\delta t)$ with intensity $\lambda(\Delta p_t^*, \Delta \ell_t)$ to model δt, the time between two consecutive orders.

Type of order – For the type of the next order, the choice between events $e_{T,s}$ where T is in $\{M, L, C\}$ (i.e. Market, Limit, Cancel) and s in $\{B, A\}$ (for Bid and Ask) can be made through a *multinomial law* of parameters $(r_{T,s}(\Delta p_t^*, \Delta \ell_t))_{T,s}$ (i.e. $r_{T,s} = \mathbb{P}(e_{T,s})$) such that:

$$\sum_{T \in \{M,L,C\}, s \in \{B,A\}} r_{T,s}(\Delta p_t^*, \Delta \ell_t) = 1. \tag{7}$$

Size and price of an order – Each type of order (i.e. each event $e_{T,s}$) has different properties: market orders have only a quantity $q_{T,s}$ while limit and cancel orders have a quantity $q_{T,s}$ and a price $p_{T,s}$. The mass of the interests present in the real order book has to be conserved, a master equation (that can be considered as the microscopic counterpart of Eq. (2): the macroscopic MFG EDP) needs to be satisfied in expectation: the mass of the inserted order on one side (Limit Bid or Limit Ask) must be equal to the removed one (Cancel Bid plus Market Ask or Cancel Ask plus Market Bid):

$$\mathbb{E}\left((\mathbf{1}_{T=L,s=B} - (\mathbf{1}_{T=C,s=B} + \mathbf{1}_{T=M,s=A}))\, q_{T,s}\right) = 0$$
$$\mathbb{E}\left((\mathbf{1}_{T=L,s=A} - (\mathbf{1}_{T=C,s=A} + \mathbf{1}_{T=M,s=B}))\, q_{T,s}\right) = 0.$$

From point process to diffusions

Prior to a precise description of the application of our framework to a one-agency-broker model presented here, let's just look at a simplification of the microscopic layer of the generic framework. At a given time t, the state of the first limits of the limit order book (prices p^{ask}, p^{bid} and quantities q^{ask}, q^{bid}) confronted to the characteristics of the next order gives the probability for the mid-price p^{mid} to change from t to the next event in $t + \delta t$:

$$\begin{aligned}
\mathbb{P}(p^{\text{mid}}_{t+\delta t} > p^{\text{mid}}_t|\delta t) &= \mathbb{P}(e_{C,A}(t+\delta t), p_{C,A}(t+\delta t) = p^{\text{ask}}_t, q_{C,A}(t+\delta t) = q^{\text{ask}}_t) \\
&\quad + \mathbb{P}(e_{M,B}(t+\delta t), q_{M,B}(t+\delta t) \geq q^{\text{ask}}_t) \\
&\quad + \mathbb{P}(e_{L,B}(t+\delta t), p_{L,B}(t+\delta t) > p^{\text{bid}}_t) \\
\mathbb{P}(p^{\text{mid}}_{t+\delta t} < p^{\text{mid}}_t|\delta t) &= \mathbb{P}(e_{C,B}(t+\delta t), p_{C,B}(t+\delta t) = p^{\text{bid}}_t, q_{C,B}(t+\delta t) = q^{\text{bid}}_t) \\
&\quad + \mathbb{P}(e_{M,A}(t+\delta t), q_{M,A}(t+\delta t) \geq q^{\text{bid}}_t) \\
&\quad + \mathbb{P}(e_{L,A}(t+\delta t), p_{L,A}(t+\delta t) < p^{\text{ask}}_t) \\
\mathbb{P}(p^{\text{mid}}_{t+\delta t} = p^{\text{mid}}_t|\delta t) &= 1 - (\mathbb{P}(p^{\text{mid}}_{t+\delta t} > p^{\text{mid}}_t|\delta t) + \mathbb{P}(p^{\text{mid}}_{t+\delta t} < p^{\text{mid}}_t|\delta t)).
\end{aligned}$$

Since all the events and random variables of upper equalities are conditioned by ΔM_t (the instantaneous distance between the real market and the viewed market), the probability that the mid price moves up, down or stays as it is can be modelled by a point process of stochastic intensity Λ_t that is conditioned by the $\Delta M_{t-\tau}$ for all $\tau < t$.

It is worth noting that any model of the mid price as a point process with path-dependent stochastic intensity as Hawkes processes (see for instance [17] or [15]) can be seen as an aggregation of this macroscopic one.

Moreover, as the limit of a rescaled compensated Poisson process with intensity Λ is a Brownian motion with volatility $\sqrt{2\Lambda}$ (see [23]), and as a Brownian motion is the solution of the heat equation $\partial_t m(t,p) = \sqrt{\Lambda/2} \cdot \partial_{pp} m(t,p)$, close to our macro Eq. (1), we have a qualitative explanation of how a "zoom out" (i.e. rescaling) of our microscopic layer could be close to the PDEs of our macroscopic layer.

3.2 Choice of Distributions for the Characteristics of Orders

Returning to a specific application of our model to the case of a market with many different investors and only one agency broker, here are the details of our conditioned models.

Dataset for Explorations, Illustrations and Fitting

Crédit Agricole Cheuvreux maintains an order-by-order database going back to 2007 on a large scope of European stocks (around 500 stocks). This order-by-order database is updated at every change on the first five limits of the order book on the primary market or on any multilateral trading facility with more than 1% market share on the stock.

In the context of this study, this database has been *de-stripped* to translate changes within the first five limits and trades into *events*: a side (bid or ask), a type (market, limit or cancel), a quantity and a price.

All the explorations, illustrations and figures used here come from this de-stripped flow. For that, we made the assumption that *the future of the real market can be considered as the viewed market*. This comes from the fact that the real market can be seen as continuously tracking its own future and sending orders to make this future happen.

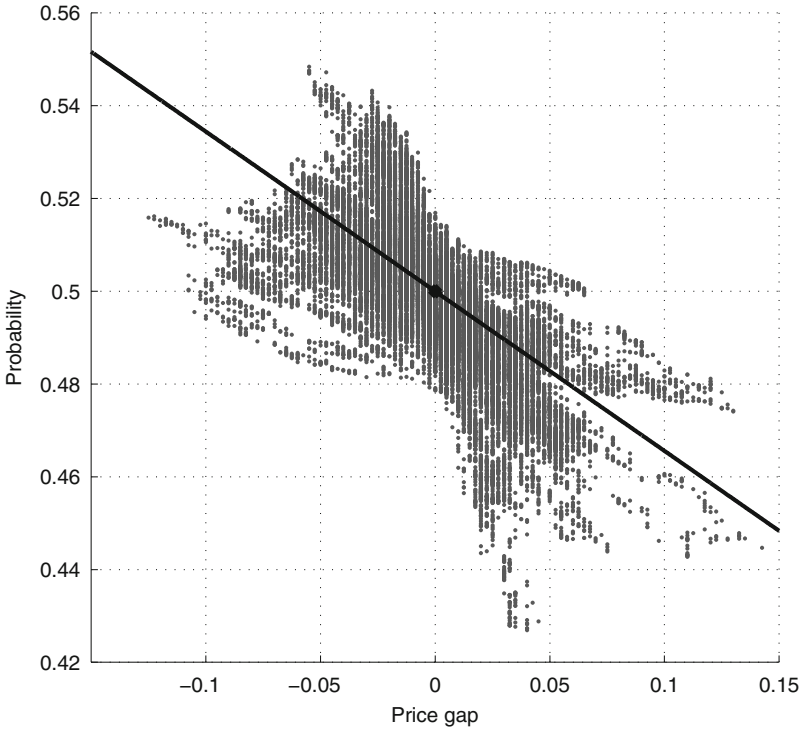


Fig. 3 Probability for the next event $e_{T,s}$ to have (T, s) in $\mathscr{S}^-$ rather than in $\mathscr{S}^+$ with respect to Δp^*

Arrival Rate of Orders

As stated in previous sections, we will use a Poisson arrival rate of orders to obtain δt, the time to the next order. In this we will follow Cohen *et al.* [6], Foucault, Kadan and Kandel [10] and Daniels *et al.* [8]. Our empirical data enabled us to identify a dependence between the time to the next order and $\Delta p_t^* = \tilde{p}_t^* - p_t^*$, so we use for the intensity $\lambda(\Delta p_t^*)$:

$$\lambda(\Delta p_t^*) = \lambda_0 + \lambda_1 \cdot \left(\Delta p_t^*\right)^2 . \tag{8}$$

The intensity of the Poisson distribution increases with the square of the distance between the real price and the viewed price. This uses our **first order assumption**.

Qualitatively, this means that the further away the price of the real market is from the price of the aggregated views, the faster orders are sent to the real market.

Type of Order
We use a model in two steps:

- Determination of the type of the order between: Cancel Bid, Insert Ask and Market Bid on the one hand and Cancel Ask, Insert Bid and Market Ask on the other hand. The orders in the first set increase the mass of the Ask side relatively to the mass of the Bid side, and orders of the latter subset have the opposite effect. Fig. 3 plots the empirical probability for the next event $e_{T,s}$ to have (T, s) in $\mathscr{S}^- = \{(C, B), (L, A), (M, A)\}$ rather than in $\mathscr{S}^+ = \{(C, A), (L, B), (M, B)\}$ with respect to Δp_t^*. The empirical (not conditioned) proportion to be in $\mathscr{S}^+$ rather than to be in $\mathscr{S}^-$ is close to 50%, and the linear regression corresponding to the line of the figure has a R^2 of around 40% (real data for one typical day of trading of France Telecom, a French stock on the local main index). The first step to determine the type of order is therefore modelled, using our **first order assumption** and a binomial distribution between $\mathscr{S}^+$ and $\mathscr{S}^-$ with:

$$\mathbb{P}(e_{T,s}, (T, x) \in \mathscr{S}^-) = \frac{1}{2} + r_0 \cdot \Delta p_t^*. \tag{9}$$

- Then the discrimination has to be made between the three elements of each set. Here we use our **second order assumption** and model this using $\Delta \ell_t$.

Qualitatively, this means that the higher the price of the real market is compared to the price of the aggregated views, the higher the ratio between the size of the bid side of the real order book and the ask side. Moreover, the difference between the slopes of the order books (real vs. viewed) influences the type of orders (Limit, Cancel or Market).

Size of the Next Order
Next we define the distribution of size of the order to be sent, given its type (Market, Limit or Cancel). After explorations of our empirical data, we decided to **use Gamma distributions**. In [4] Bouchaud, Mezard and Potters already observed that the available volumes at the bid and ask of the order book look like Gammas on the french stocks, as in [14], other authors observe that this property of the order book for Chinese stocks is best modelled using log-normal laws.

In Fig. 4, we see the comparison of a fitted Gamma and the empirical repartition of order sizes for Limit orders for France Telecom during one trading day. The data had to be regularised because *the size of the orders is highly discrete*, probably because some agents uses round numbers or multiples of estimates of the average trade size computed on previous days. The regularisation process used here is simply a blurring noise. The chosen Gamma distributions are independent of Δp_t^* and $\Delta \ell_t$.

Qualitatively, this means that the impact of the distance between the real market and the viewed one changes the shape of the order books through the number and the type of sent orders rather than via their size.

Price of the Next Order (When Needed)
The price of an order to be sent is only needed for Limit or Cancel orders, as Market

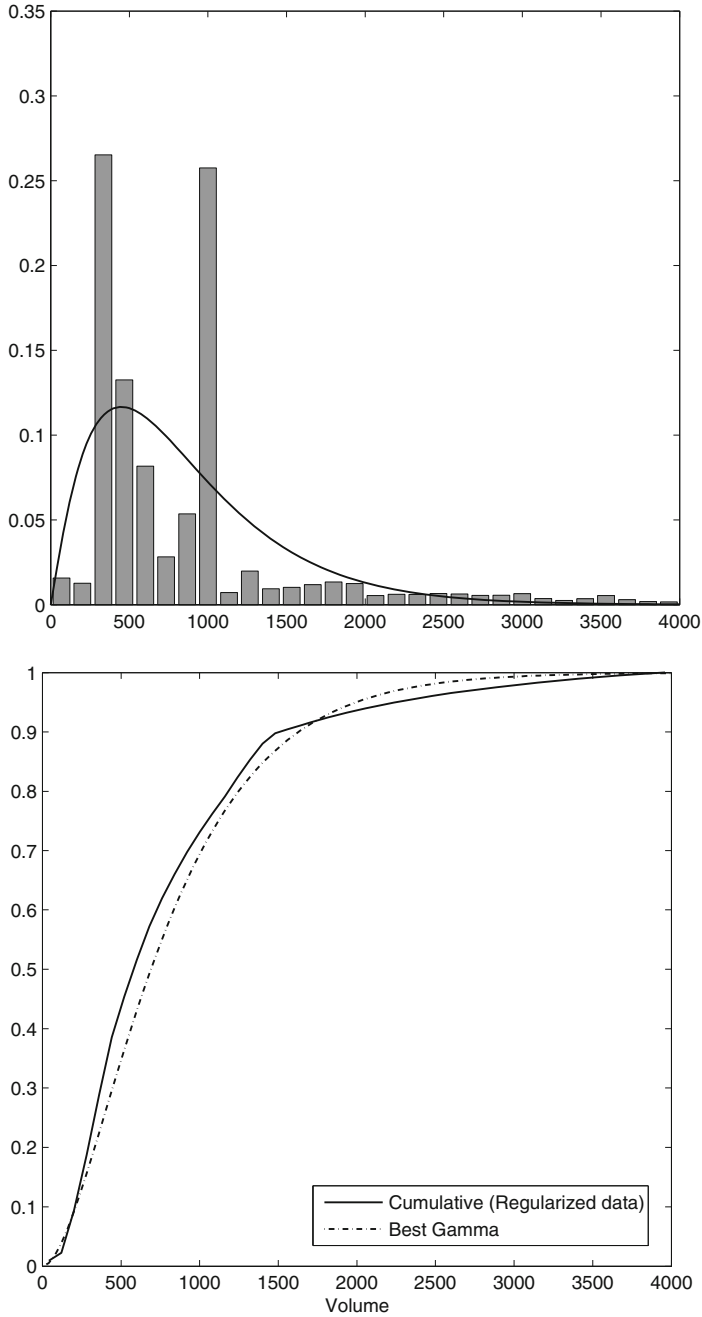


Fig. 4 Empirical density (*top*) and its regularised cumulative repartition function (*bottom*) of the size of Limit orders on France Telecom share on the 19 May 2010 (large cap on the French index) vs a fitted Gamma density

orders consume the liquidity available in the books at any price. The distribution of the price of the order expressed in multiples of the current bid-ask spread $2\psi_t$ and relative to the "best opposite" (p_t^{bid} for a Sell order and p_t^{ask} for a Buy one) seems to be more stable than other measures of the price according to our empirical data. In this illustration of our two-layer framework, a Gamma distribution conditioned by Δp_t^* is taken, for $u = (p - p_t^{ask/bid})/\psi$:

$$\mathrm{d}\mu\left(u \,\middle|\, e_{T,B/A}, T \in \{L, C\}\right) = \Gamma(u; \alpha(\Delta p_t^*), \beta(\Delta p_t^*))\, du$$

with α and β, the parameters of the Gamma law, are functionals of Δp_t^* so that the variance is constant and the mean depends linearly on Δp_t^*, using our **first order assumption**.

Qualitatively, this means that the price of inserted or cancelled orders is relative to the best opposite, can be expressed in bid-ask spreads, and increases with the distance between the real market and the viewed one.

3.3 Overall Simulations

Finally, we obtain a full simulation of the order book dynamics using the conjunction of our two layers: the macroscopic one, based on a Mean Field Game (MFG) framework, modelling the aggregated views of the investing agents (living in a continuous state space) and the microscopic one, sending orders according to zero-intelligence-like models of distributions of orders to send to the real (discretised) market, given the distance between the viewed order book and the real one.

Fig. 5 shows an example of 20 minutes of simulations: the smooth grey curve is the trajectory of p_t^* following the macroscopic MFG layer, the black stepped curve is the trajectory of $\tilde{p}_t^*$ following the microscopic layer pegged to the MFG one.

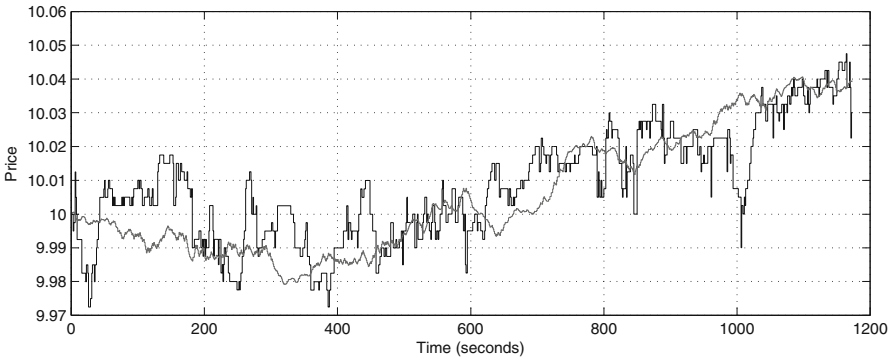


Fig. 5 An example for 20 minutes of simulations: the *smooth grey curve* is the trajectory of p_t^* following the macroscopic MFG layer, the *black stepped curve* is the trajectory of $\tilde{p}_t^*$ following the microscopic layer pegged to the MFG one

4 Conclusions

The model of the market microstructure at the order book level presented here endeavours to answer the practical need for a tool to conduct back-tests, stress tests and sensitivity analysis of the payoff of a trading algorithm.

It is structured similarly to real markets with investors making decisions according to their views on the future prices in the market (i.e. macroscopic layer) and using brokers to access the order books (microscopic layer). **The views of the investors are modelled via a Mean Field Game (MFG) and orders are sent to the market according to a zero-intelligence-type model pegged to the instantaneous distance between the views and the state of the order books**.

This paper describes an application of this framework to a market with three kinds of investors (mean reverters, trend followers and noise traders) and only one pure agency broker, and provides illustrations of the outcomes of this model as well as some analyses of real data at the order book scale.

As this is an on-going research project for Crédit Agricole Cheuvreux's research team, a more generic study of this framework is currently being conducted where any configuration of market access (i.e. not only through one pure agency broker) and any type of investors (implementing stochastic control to solve their explicit utility functions) is available. Plans for future research include studying the effect of investors taking into account not only their views on the future of the market, but also the current market price (i.e. using intra-day marked-to-market valuation of their positions in their utility functions). A natural extension of this work is also to consider the PDEs obtained thanks to the MFG layer as the diffusive limit of the Markovian zero-intelligence one.

Acknowledgements The authors wish to thank Adrian Iuga for producing some useful real order book data analyses during his internship at Crédit Agricole Cheuvreux.

References

1. M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. Quantitative Finance, 8(3):217–224, 2008
2. E. Bacry, S. Delattre, M. Hoffman and J. F. Muzy. Deux modèles de bruit de microstructure et leur inférence statistique. Présentation, 2009
3. B. Bouchard, N.-M. Dang and C.-A. Lehalle. Optimal control of trading algorithms: a general impulse control approach. Technical report, 2009
4. J. P. Bouchaud, M. Mezard and M. Potters. Statistical properties of stock order books: empirical results and models. 2002
5. A. Chakraborti, I. Muni Toke, M. Patriarca and F. Abergel. Econophysics: Empirical facts and agent-based models. 2009
6. K.J. Cohen, S.F. Maier, R.A. Schwartz and D.K. Whitcomb. Transaction costs, order placement strategy, and existence of the bid-ask spread. The Journal of Political Economy, 89(2):287–305, 1981

7. R. Cont and A. de Larrard. Stochastic modeling of order books: high-frequency dynamics and diffusion limits. In 28th European Meeting of Statisticians, 2010
8. M.G. Daniels, D.J. Farmer, L. Gillemot, G. Iori and E. Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. Physical Review Letters, 90(10):108102+, 2003
9. Z. Eisler, J. Kertesz, F. Lillo and R.N. Mantegna. Diffusive behavior and the modeling of characteristic times in limit order executions. Social Science Research Network Working Paper Series, 2007
10. T. Foucault, O. Kadan and E. Kandel. Limit order book as a market for liquidity. Review of Financial Studies, 18(4), 2005
11. J. Gatheral. No-dynamic-arbitrage and market impact. Social Science Research Network Working Paper Series, 2008
12. M. González and M. Gualdani. Asymptotics for a symmetric equation in price formation. Applied Mathematics and Optimization, 59(2):233–246, 2009
13. M. González and M. Gualdani. Asymptotics for a free-boundary problem in price formation. To appear
14. G. Gu, W. Chen and W. Zhou. Empirical shape function of limit-order books in the chinese stock market. Physica A: Statistical Mechanics and its Applications, 387(21):5182–5188, 2008
15. A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. Biometrika, 58(1):83–90, 1971
16. B. Jovanovic and A.J. Menkveld. Middlemen in Limit-Order Markets. SSRN eLibrary, 2010
17. J. Large. Measuring the resiliency of an electronic limit order book. Journal of Financial Markets, 10(1):1–25, 2007
18. J.-M. Lasry and P.-L. Lions. Mean field games. Japanese Journal of Mathematics, 2(1):229–260, 2007
19. P.A. Markowich, N. Matevosyan, J.F. Pietschmann and M.T. Wolfram. On a parabolic free boundary equation modeling price formation. Mathematical Models and Methods in Applied Sciences, 11(19):1929–1957, 2209
20. Official Journal of the European Union. Directive 2004/39/ec of the european parliament. Directive 2004/39/EC of the European Parliament, 2004
21. G. Pagès, S. Laruelle and C.-A. Lehalle. Optimal split of orders across liquidity pools: a stochatic algorithm approach. Technical report, 2009
22. H. Pham. Continuous-time Stochastic Control and Optimization with Financial Applications (Stochastic Modelling and Applied Probability). Springer, 1 edition, 2009
23. F.P. Schoenberg. On rescaled poisson processes and the brownian bridge. Annals of the Institute of Statistical Mathematics, 54(2):445–457, 2002
24. A.N. Shiryaev. Essentials of Stochastic Finance: Facts, Models, Theory. World Scientific Publishing Company, 1st edition, 1999
25. M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters and M. Vettorazzo. Relation between bid-ask spread, impact and volatility in double auction markets. Technical report, 2006

A Mathematical Approach to Order Book Modelling

Frédéric Abergel and Aymen Jedidi

Abstract. We present a mathematical study of the order book as a multidimensional continuous-time Markov chain where the order flow is modelled by independent Poisson processes. Our aim is to bridge the gap between the microscopic description of price formation (agent-based modelling), and the Stochastic Differential Equations approach used classically to describe price evolution in macroscopic time scales. To do this we rely on the theory of *infinitesimal generators*. We motivate our approach using an elementary example where the spread is kept constant ("perfect market making"). Then we compute the infinitesimal generator associated with the order book in a general setting, and link the price dynamics to the instantaneous state of the order book. In the last section, we prove the stationarity of the order book and give some hints about the behaviour of the price process in long time scales.

1 Introduction and Background

The emergence of electronic trading as a major means of trading financial assets makes the study of the order book central to the understanding of the mechanism of price formation. In order driven markets, buy and sell orders are matched continuously subject to price and time priority. The *order book* is the list of all buy and sell limit orders, with their corresponding price and size, at a given instant of time. Essentially, three types of orders can be submitted:

- *Limit order*: Specify a price (also called "quote") at which one is willing to buy or sell a certain number of shares.
- *Market order*: Immediately buy or sell a certain number of shares at the best available opposite quote.
- *Cancellation order*: Cancel an existing limit order.

Frédéric Abergel, Aymen Jedidi
Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
e-mail: frederic.abergel@ecp.fr; aymen.jedidi@ecp.fr

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

In the econophysics literature, "agents" who submit exclusively limit orders are referred to as *liquidity providers*. Those who submit market orders are referred to as *liquidity takers*.

Limit orders are stored in the order book till they are either executed against an incoming market order or cancelled. The *ask* price P^{A} is the price of the best (i.e. lowest) limit sell order. The *bid* price P^{B} is the price of the best (i.e. highest) limit buy order. The gap between the bid and the ask

$$S := P^{\mathrm{A}} - P^{\mathrm{B}},$$

is always positive and is called the *spread*. Prices are not continuous, but rather have a discrete resolution ΔP, the *tick*, which represent the smallest quantity by which they can change. We define the *mid-price* as the average between the bid and the ask

$$P := \frac{P^{\mathrm{A}} + P^{\mathrm{B}}}{2}.$$

The price dynamics is the result of the interplay between the incoming order flow and the order book. Fig. 1 is a schematic illustration of this process. Note that we chose to represent quantities in the bid side of the book by negative numbers.

Although in reality orders can have any size, we shall assume throughout this note that all orders are of unit size τ. This assumption is convenient to carry out our analysis and is, for now, of secondary importance to the problem we are interested in.

2 An Elementary Approximation: Perfect Market Making

We start with the simplest agent-based market model:

- The order book starts in a full state: All limits above P_0^{A} and below P_0^{B} are filled with one limit order of unit size. The spread starts equal to 1 tick.
- The flow of market orders is modelled by two independent Poisson processes M_t^+ (buy orders) and M_t^- (sell orders) with constant arrival rates (or intensities) λ^+ and λ^-.
- There is one liquidity provider, who reacts immediately after a market order arrives so as to maintain the spread constantly equal to 1. He places a limit order in the same side of the market order (i.e. a limit buy order after a market buy order and vice versa) with probability q and in the opposite side with probability $1-q$.

The mid-price dynamics can be written under the following form

$$\begin{aligned} \mathrm{d}P_t &= \frac{\Delta P}{2}(\mathrm{d}M_t^+ - \mathrm{d}M_t^-)Z \\ &:= \Delta(\mathrm{d}M_t^+ - \mathrm{d}M_t^-)Z \end{aligned} \tag{1}$$

where $\Delta := \frac{\Delta P}{2}$ (1/2 of a tick) and Z is a Bernoulli-type random variable

$$\begin{cases} Z = 0 \text{ with probability } (1-q), \\ Z = 1 \text{ with probability } q. \end{cases} \tag{2}$$

The infinitesimal generator associated to this dynamics is

$$Lf(P) = q\left(\lambda^+(f(P+\Delta) - f) + \lambda^-(f(P-\Delta) - f)\right). \tag{3}$$

It is well known that a continuous limit obtains with suitable assumptions on the intensity and tick size. Noting that (3) can be rewritten as

$$\begin{aligned} Lf(P) = {} & \frac{1}{2}q(\lambda^+ + \lambda^-)\Delta^2 \frac{f(P+\Delta) - 2f + f(P-\Delta)}{\Delta^2} \\ & + q(\lambda^+ - \lambda^-)\Delta P \frac{f(P+\Delta) - f(P-\Delta)}{2\Delta}, \end{aligned} \tag{4}$$

and under the following assumptions

$$\begin{cases} q(\lambda^+ + \lambda^-)\Delta^2 \longrightarrow \sigma^2 \text{ as } \Delta \to 0, \\ q(\lambda^+ - \lambda^-)\Delta \longrightarrow \mu \text{ as } \Delta \to 0, \end{cases} \tag{5}$$

the generator converges to the classical diffusion operator

$$\frac{\sigma^2}{2}\frac{\partial^2 f}{\partial P^2} + \mu \frac{\partial f}{\partial P}, \tag{6}$$

corresponding to a Brownian motion with drift. This simple case is worked out as an example of the type of limit theorems that we will be interested in in the sequel. One should also note that a more classical approach using the Functional Central limit Theorem (FCLT) as in [1] yields similar results: For given fixed values of λ^+, λ^- and Δ, the rescaled, centred price process

$$\sqrt{\delta}\frac{P_{\frac{t}{\delta}} - \frac{\mu}{\delta}t}{\sigma} \tag{7}$$

converges as $\delta \to 0$, to a standard Brownian motion B_t where

$$\begin{cases} \sigma = \Delta\sqrt{(\lambda^+ + \lambda^-)q}, \\ \mu = \Delta(\lambda^+ - \lambda^-)q. \end{cases} \tag{8}$$

Let us mention that one can easily achieve more complex diffusive limits such as a local volatility model by imposing that the limit is a function of P and t

$$\begin{cases} q(\lambda^+ + \lambda^-)\Delta^2 \to \sigma^2(P,t), \\ q(\lambda^+ - \lambda^-)\Delta \to \mu(P,t), \end{cases} \tag{9}$$

always a possibility if the original intensities are functions of P and t themselves. The case of stochastic volatilities is also encompassed by diffusive limits of these models. For instance consider the following pure-jump model for P and the intensities λ^+, λ^-

$$\begin{cases} \mathrm{d}P_t = \Delta(\mathrm{d}M_t^+ - \mathrm{d}M_t^-)Z \\ \dfrac{\mathrm{d}\lambda_t^+}{\lambda_t^+} = (J^{++} - 1)\mathrm{d}N_t^{\mathrm{vol}+} + (J^{+-} - 1)\mathrm{d}N_t^{\mathrm{vol}-}, \\ \dfrac{\mathrm{d}\lambda_t^-}{\lambda_t^-} = (J^{-+} - 1)\mathrm{d}N_t^{\mathrm{vol}+} + (J^{--} - 1)\mathrm{d}N_t^{\mathrm{vol}-}. \end{cases} \tag{10}$$

Its infinitesimal generator is given by the following operator

$$\begin{aligned} Lf(P, \lambda^+, \lambda^-) = q\left(\lambda^+(f(P + \Delta P) - f) + \lambda^-(f(P - \Delta P) - f)\right) \\ + \nu^+\left(f(J^{++}\lambda^+, J^{-+}\lambda^-) - f\right) \\ + \nu^-\left(f(J^{+-}\lambda^+, J^{-+}\lambda^-) - f\right), \end{aligned} \tag{11}$$

where ν^+, ν^- are the intensities of λ_t^+, and λ_t^- respectively.

In a more general fashion, we will use the expression *generalized Bachelier market* to designate a market model where the best bid and ask prices, or equivalently the mid-price and the spread, obey the following type of dynamics

$$\mathrm{d}P_t = \Delta \sum_{i=1}^{Q} Z_t^i \mathrm{d}M_t^i, \tag{12}$$

where the M_t^i are independent Poisson processes and the marks Z_i are random variables with finite means and variances. The same standard results on martingale convergence as the one used above show that the rescaled, centred price process

$$\sqrt{\delta}\left(P_{\frac{t}{\delta}} - \mathbb{E}\left[P_{\frac{t}{\delta}}\right]\right) \tag{13}$$

converges to a Gaussian process with a diffusion coefficient determined by the variance of the marks and the intensities of the Poisson processes. This extension of the Bachelier market will come in handy in Sect. 6, where we study the stationary order book in a more general setting.

3 Order Book Dynamics

3.1 Model Setup

We now consider the dynamics of a general order book under a Poisson type assumption for the arrival of new market orders, limit orders and cancellations. We shall assume that the whole order book is fully described by a fixed number of limits N, ranging at each time from 1 to N ticks away from the best available opposite

quote. By doing so, we adopt the representation described e.g. in [2] or [3], but depart slightly from it by adopting *a finite moving frame*, as we think it more realistic and also, more convenient when scaling in tick size will be addressed. Let us now recall the type of events that may happen:

- arrival of a new market order;
- arrival of a new limit order;
- cancellation of an already existing limit order.

These events are described by independent Poisson processes:

- $M_t^{\pm}$: arrival of new market order; with intensity $\lambda_M^{\pm}$;
- $L_t^{\pm i}$: arrival of a limit order i ticks away from the best opposite quote; with intensity $\lambda_L^{i\pm}$;
- $C_t^{\pm i}$: cancellation of a limit order i ticks away from the best opposite quote; with intensity $\lambda_C^{i+}\dfrac{a_i}{\tau}$, $\lambda_C^{i-}\dfrac{b_i}{\tau}$;

where, as usual, ΔP is the tick size, τ the size of any new incoming order, and the superscript "+" (respectively "−") refers to the ask (respectively bid) side of the book. The intensity of the cancellation process at level i is proportional to the available quantity at that level.

We impose constant boundary conditions outside the moving frame of size $2N$: Every time the moving frame leaves a price level, the number of shares at that level is set to a^{∞} (or b^{∞} depending on the side of the book). Note that this makes the model Markovian as we do not keep track of the price levels that have been visited (then left) by the moving frame at some prior time.

3.2 Evolution of the Order Book

Denoting by $\boldsymbol{a}_t = \left(a_t^1, \ldots, a_t^N\right)$ the vector of available quantities of sell limit orders (the ask side), and similarly $\boldsymbol{b}_t = \left(b_t^1, \ldots, b_t^N\right)$ the bid side of the order book, we can write the following coupled SDEs:

$$\begin{cases} da_t^i = -\left(\tau - \displaystyle\sum_{k=1}^{i-1} a^k\right)_+ \mathrm{d}M_t^+ + \tau \mathrm{d}L_t^{i+} - \tau \mathrm{d}C_t^{i+} \\ \qquad + (J_M^b(\boldsymbol{a}) - \boldsymbol{a})_i \mathrm{d}M_t^- + \displaystyle\sum_{i=1}^{N} (J_{L_i}^-(\boldsymbol{a}) - \boldsymbol{a})_i \mathrm{d}L_t^{i-} + \sum_{i=1}^{N} (J_{C_i}^-(\boldsymbol{a}) - \boldsymbol{a})_i \mathrm{d}C_t^{i-}, \\ db_t^i = -\left(\tau - \displaystyle\sum_{k=1}^{i-1} b^k\right)_+ \mathrm{d}M_t^- + \tau \mathrm{d}L_t^{i-} - \tau \mathrm{d}C_t^{i-} \\ \qquad + (J_M^a(\boldsymbol{b}) - \boldsymbol{b})_i \mathrm{d}M_t^+ + \displaystyle\sum_{i=1}^{N} (J_{L_i}^+(\boldsymbol{b}) - \boldsymbol{b})_i \mathrm{d}L_t^{i+} + \sum_{i=1}^{N} (J_{C_i}^+(\boldsymbol{b}) - \boldsymbol{b})_i \mathrm{d}C_t^{i+}, \end{cases} \tag{14}$$

where the J's are *shift operators* corresponding to the number of ticks by which the best bid (respectively ask) moves following an event on the ask side (respectively bid side) of the book. For instance the shift operator corresponding to the arrival of a sell market order of size τ is

$$J_M^b(\boldsymbol{a}) = \left(\underbrace{0, 0, \ldots, 0}_{k\text{times}}, a^1, a^2, \ldots, a^{N-k}\right), \tag{15}$$

with

$$k = \inf\{p : \sum_{j=1}^{p} b^j > \tau\} - \inf\{p : b^p > 0\}. \tag{16}$$

Similar expressions can be derived for the other set of events affecting the order book. Also one should note that the last two summations in (14) are in fact taken only on the range of indexes smaller than or equal to the spread (in ticks).

In the next sections, we will study some general properties of such models, starting with the generator associated to this $2N-$dimensional continuous-time Markov chain.

4 Infinitesimal Generator

Let us now work out the infinitesimal generator associated to the jump process described above. One can derive the following result:

$$\begin{aligned}
Lf\,(\boldsymbol{a};\boldsymbol{b}) = {} & \lambda_M^+\left(f\left((a^i - (\tau - A^{i-1})_+)_+; J_M^a(\boldsymbol{b})\right) - f\right) \\
& + \sum_{i=1}^{N} \lambda_L^{i+}(f\left(a^i + \tau; J_L^{i+}(\boldsymbol{b})\right) - f) \\
& + \sum_{i=1}^{N} \lambda_C^{i+}\frac{a_i}{\tau}(f\left(a^i - \tau; J_C^{i+}(\boldsymbol{b})\right) - f) \\
& + \lambda_M^-\left(f\left(J_M^b(\boldsymbol{a}); (b^i - (\tau - B^{i-1})_+)_+\right) - f\right) \\
& + \sum_{i=1}^{N} \lambda_L^{i-}(f\left(J_L^{i-}(\boldsymbol{a}); b^i + \tau\right) - f) \\
& + \sum_{i=1}^{N} \lambda_C^{i-}\frac{b_i}{\tau}(f\left(J_C^{i-}(\boldsymbol{a}); b^i - \tau\right) - f),
\end{aligned} \tag{17}$$

where, to ease the notations, we note $f(a_i;\boldsymbol{b})$ instead of $f(a_1, \ldots, a_i, \ldots, a_N;\boldsymbol{b})$ etc. The operator above, although cumbersome to put in writing, is simple to decipher: a series of standard difference operators corresponding to the "deposition-evaporation" process of orders at each limit, combined with the shift operators expressing the moves in the best limits and therefore, in the origins of the frames for

the two sides of the order book. Clearly, the interesting part lies in the coupling of the two sides of the book: the shifts on the a's depend on the b's, and vice versa. More precisely the shifts depend on the profile of the order book on the other side, namely the cumulative number of orders

$$\begin{cases} A_i = \sum_{k=1}^{i} a_k, \\ B_i = \sum_{k=1}^{i} b_k \end{cases} \tag{18}$$

and the generalized inverse functions thereof

$$\begin{cases} A^{-1}(\tau) = \inf\{p : \sum_{j=1}^{p} a_j > \tau\}, \\ B^{-1}(\tau) = \inf\{p : \sum_{j=1}^{p} b_j > \tau\}. \end{cases} \tag{19}$$

Note that the index corresponding to the best opposite quote equals the spread S in ticks, that is

$$\begin{cases} i_A = A^{-1}(0) = \inf\{p : \sum_{j=1}^{p} a_j > 0\} = \dfrac{S}{\Delta P} = i_S, \\ i_B = B^{-1}(0) = \inf\{p : \sum_{j=1}^{p} b_j > 0\} = \dfrac{S}{\Delta P} = i_S. \end{cases} \tag{20}$$

5 Price Dynamics

Let us focus on the dynamics of the best ask and bid price, denoted by P_t^{A} and P_t^{B}. One can easily see that they satisfy the following SDE:

$$\begin{cases} \mathrm{d}P_t^{\mathrm{A}} = \Delta P \Big[(A^{-1}(\tau) - A^{-1}(0))\mathrm{d}M_t^{+} - \sum_{i=1}^{N} (A^{-1}(0) - i)_{+} \mathrm{d}L_t^{i+} \\ \qquad\qquad + (A^{-1}(\tau) - A^{-1}(0))\mathrm{d}C_t^{i_A +} \Big] \\ \mathrm{d}P_t^{\mathrm{B}} = -\Delta P \Big[(B^{-1}(\tau) - B^{-1}(0))\mathrm{d}M_t^{-} - \sum_{i=1}^{N} (B^{-1}(0) - i)_{+} \mathrm{d}L_t^{i-} \\ \qquad\qquad + (B^{-1}(\tau) - B^{-1}(0))\mathrm{d}C_t^{i_B -} \Big] \end{cases} \tag{21}$$

which describes the various events that affect them: change due to market order, change due to new limit orders inside the spread, and change due to the cancellation of limit orders at the best price. One can summarize these two equations in order to highlight, in a more traditional fashion, the respective dynamics of the mid-price and the spread:

$$\begin{aligned}\mathrm{d}P_t = \frac{\Delta P}{2}\Big[&(A^{-1}(\tau) - A^{-1}(0))\mathrm{d}M_t^+ - (B^{-1}(\tau) - B^{-1}(0))\mathrm{d}M_t^- \\ &- \sum_{i=1}^{N}(A^{-1}(0) - i)_+\mathrm{d}L_t^{i+} + \sum_{i=1}^{N}(B^{-1}(0) - i)_+\mathrm{d}L_t^{i-} \\ &+ (A^{-1}(\tau) - A^{-1}(0))\mathrm{d}C_t^{i_A+} - (B^{-1}(\tau) - B^{-1}(0))\mathrm{d}C_t^{i_B-}\Big]. \end{aligned} \tag{22}$$

$$\begin{aligned}\mathrm{d}S_t = \Delta P\Big[&(A^{-1}(\tau) - A^{-1}(0))\mathrm{d}M_t^+ + (B^{-1}(\tau) - B^{-1}(0))\mathrm{d}M_t^- \\ &- \sum_{i=1}^{N}(A^{-1}(0) - i)_+\mathrm{d}L_t^{i-} - \sum_{i=1}^{N}(B^{-1}(0) - i)_+\mathrm{d}L_t^{i+} \\ &+ (A^{-1}(\tau) - A^{-1}(0))\mathrm{d}C_t^{i_A+} + (B^{-1}(\tau) - B^{-1}(0))\mathrm{d}C_t^{i_B-}\Big]. \end{aligned} \tag{23}$$

The set of equations above are interesting in that they relate in an explicit way the profile of the order book with the size of an elementary jump of the mid-price or the spread, therefore linking the volatility dynamics with order arrival. For instance the "infinitesimal" drifts of the mid-price and of the spread, conditional on the shape of the order book at time t, are given by

$$\begin{aligned}\mathbb{E}\left[\mathrm{d}P_t|(\boldsymbol{a}_t;\boldsymbol{b}_t)\right] = \frac{\Delta P}{2}\Big[&(A^{-1}(\tau) - A^{-1}(0))\lambda_M^+ - (B^{-1}(\tau) - B^{-1}(0))\lambda_M^- \\ &- \sum_{i=1}^{N}(A^{-1}(0) - i)_+\lambda_L^{i+} + \sum_{i=1}^{N}(B^{-1}(0) - i)_+\lambda_L^{i-} \\ &+ (A^{-1}(\tau) - A^{-1}(0))\lambda_C^{i_A+} - (B^{-1}(\tau) - B^{-1}(0))\lambda_C^{i_B-}\Big]\mathrm{d}t, \end{aligned} \tag{24}$$

and

$$\begin{aligned}\mathbb{E}\left[\mathrm{d}S_t|(\boldsymbol{a}_t;\boldsymbol{b}_t)\right] = \Delta P\Big[&(A^{-1}(\tau) - A^{-1}(0))\lambda_M^+ + (B^{-1}(\tau) - B^{-1}(0))\lambda_M^- \\ &- \sum_{i=1}^{N}(A^{-1}(0) - i)_+\lambda_L^{i+} - \sum_{i=1}^{N}(A^{-1}(0) - i)_+\lambda_L^{i-} \\ &+ (A^{-1}(\tau) - A^{-1}(0))\lambda_C^{i_A+} + (B^{-1}(\tau) - B^{-1}(0))\lambda_C^{i_B-}\Big]\mathrm{d}t. \end{aligned} \tag{25}$$

6 Stationary State and Diffusive Limit

6.1 Ergodicity of the Order Book

Of the utmost interest is the behaviour of the order book in a stationary state. We have the following result[1]:

Proposition 1 *If* $\underline{\lambda_C} = \min_{1\le i\le N}\{\lambda_C^{\pm i}\} > 0$, *then* $\boldsymbol{X}_t = (\boldsymbol{a}_t; \boldsymbol{b}_t)$ *is an ergodic Markov process. In particular* $\boldsymbol{X}_t$ *has a* stationary distribution.

Proof. This is an application of the stochastic Lyapunov ergodicity criterion (Appendix A). Let $\varphi(\boldsymbol{a};\boldsymbol{b}) = \sum_{i=1}^N a^i + \sum_{i=1}^N |b^i|$ be the total number of shares in the book. Using the expression of the infinitesimal generator (17) we have

$$\begin{aligned} L\varphi(\boldsymbol{a};\boldsymbol{b}) &\le \sum_{i=1}^N(\lambda_L^{i+} + \lambda_L^{i-})\tau - (\lambda_M^+ + \lambda_M^-)\tau - \sum_{i=1}^N(\lambda_C^{i+}a^i + \lambda_C^{i-}|b^i|) \\ &\quad + \sum_{i=1}^N \lambda_L^{i-}(i_S - i)_+ a^\infty + \sum_{i=1}^N \lambda_L^{i+}(i_S - i)_+|b^\infty| \qquad (26) \\ &\le N(\overline{\lambda_L^+} + \overline{\lambda_L^-})\tau - (\lambda_M^+ + \lambda_M^-)\tau - \underline{\lambda_C} f(\boldsymbol{a};\boldsymbol{b}) \\ &\quad + N(N+1)(\overline{\lambda_L^-}a^\infty + \overline{\lambda_L^+}|b^\infty|) \qquad (27) \end{aligned}$$

where $\overline{\lambda_L^\pm} = \max_{1\le i\le N}\{\lambda_L^{i\pm}\}$ and $\underline{\lambda_C} = \min_{1\le i\le N}\{\lambda_C^{i\pm}\} > 0$.

The first three terms in the r.h.s. of inequality (26) correspond respectively to the arrival of a limit, market, or cancellation order—ignoring the effect of the shift operators. The last two terms are due to shifts occurring after the arrival of a limit order within the spread. The terms due to shifts occurring after market or cancellation orders (which we do not put in the r.h.s) are negative, hence the inequality. To obtain inequality (27), we used the fact that the spread i_S (and hence $(i_S - i)_+$) is bounded by $N + 1$—a consequence of the boundary conditions we impose.

Let γ be a positive constant and K a constant such that

$$K > \max\{\tau, \frac{\gamma + N(\overline{\lambda_L^+} + \overline{\lambda_L^-})\tau + N(N+1)(\overline{\lambda_L^-}a^\infty + \overline{\lambda_L^+}|b^\infty|) - (\lambda_M^+ + \lambda_M^-)\tau}{\underline{\lambda_C}}\}.$$

We have:

- if $\varphi(\boldsymbol{a};\boldsymbol{b}) > K$ then
$$L\varphi(\boldsymbol{a};\boldsymbol{b}) \le -\gamma;$$

[1] A similar result, albeit in a slightly different model, was proved in [3]. We still sketch our proof because it follows a different route.

- the random variables $\sup\{\varphi(\boldsymbol{X}_t) : t \leq 1\}$ and $\int_0^1 |L\varphi(\boldsymbol{X}_t)|\mathrm{d}t$ are integrables;
- the set $F = \{(\boldsymbol{a};\boldsymbol{b}) : \varphi(\boldsymbol{a};\boldsymbol{b}) \leq K\}$ is finite.

By Proposition 8.14 in [4] (Appendix A), we conclude that the Markov process $\boldsymbol{X}_t = (\boldsymbol{a}_t;\boldsymbol{b}_t)$ is ergodic. □

Remark 1 *As a corollary of Proposition 1, the spread $S = A^{-1}(0)$ has a well-defined stationary distribution (this is expected as by construction the spread is bounded by $N+1$).*

6.2 Large Scale Limit of the Price

The stationarity of the order book is essential for the study of the long term behaviour of the price. We recall Eq. (22) giving the mid-price increments

$$\begin{aligned}\mathrm{d}P_t = \frac{\Delta P}{2}\Big[&(A_t^{-1}(\tau) - A_t^{-1}(0))\mathrm{d}M_t^+ - (B_t^{-1}(\tau) - B_t^{-1}(0))\mathrm{d}M_t^- \\ &- \sum_{i=1}^{N}(A_t^{-1}(0) - i)_+\mathrm{d}L_t^{i+} + \sum_{i=1}^{N}(B_t^{-1}(0) - i)_+\mathrm{d}L_t^{i-} \\ &+ (A_t^{-1}(\tau) - A_t^{-1}(0))\mathrm{d}C_t^{i_A+} - (B_t^{-1}(\tau) - B_t^{-1}(0))\mathrm{d}C_t^{i_B-}\Big]. \end{aligned} \tag{28}$$

In order to clarify the dependence of the price process on the order book dynamics, let us introduce the following *deterministic* functions:

- $\Phi : \mathbb{N}^N \to \{0,\ldots,N\}$; $\Phi(\boldsymbol{a}) = A^{-1}(\tau) - A^{-1}(0)$. Φ is the *virtual price impact* in ticks of a market order of size τ (or of a cancellation order at the best price);
- $\Psi : \mathbb{N}^N \to \{1,\ldots,N+1\}$; $\Psi(\boldsymbol{a}) = A^{-1}(0)$. Ψ gives the value of the *spread* in ticks.

Eq. (28) can now be rewritten as

$$\begin{aligned}\mathrm{d}P_t = \frac{\Delta P}{2}\Big[&\Phi(\boldsymbol{a}_t)(\mathrm{d}M_t^+ + \mathrm{d}C_t^{\Psi(\boldsymbol{a}_t)+}) - \Phi(|\boldsymbol{b}_t|)(\mathrm{d}M_t^- + \mathrm{d}C_t^{\Psi(|\boldsymbol{b}_t|)-}) \\ &- \sum_{i=1}^{N}(\Psi(\boldsymbol{a}_t) - i)_+\mathrm{d}L_t^{i+} + \sum_{i=1}^{N}(\Psi(|\boldsymbol{b}_t|) - i)_+\mathrm{d}L_t^{i-}\Big]. \end{aligned} \tag{29}$$

Should the order book be in its stationary state, then the price increments are stationary, and the model is recast under the form described in Sect. 2

$$\mathrm{d}P_t = \Delta\sum_{i=1}^{Q} Z_t^i \mathrm{d}N_t^i.$$

We are interested in the existence of a stochastic process limit $\widetilde{P}_t$ after proper scaling[2]

$$\widetilde{P}_t = \lim_{n\to\infty} \frac{P_{nt} - \mathbb{E}(P_{nt})}{\sqrt{n}}. \tag{30}$$

In order to apply a generalized version of the Functional Centrale Limit Theorem, one needs to measure the dependence of price increments $\mathrm{d}P_t$. The analytical study of the dependence structure (mainly the autocovariance of price increments) is not obvious and we shall investigate it in a future work [5], but numerical simulation (see Figs. 5–6) suggests that the large scale limit process $\widetilde{P}_t$ is a Brownian motion with a volatility determined by the variance of the marks Z^i and the intensities of the Poisson processes.

7 Conclusions

This note provides a simple Markovian framework for order book modelling, in which elementary changes in the price and spread processes are explicitly linked to the instantaneous shape of the order book and the order flow parameters. Two basic properties were investigated: the stationarity of the order book and the large scale limit of the price process. The first property, to which we answered positively, is desirable in that it assures the stability of the order book in the long run. The scaling limit of the price process is more subtle to establish and one can ask if, within our framework, stochastic volatility effects can arise. Of course, more realistic stylized facts of the price process (in particular fat tails and long memory) can be added if we allow more complex assumptions on the order flow (e.g. feedback effects [6]). Further properties of this model and its extensions will be discussed in detail in future work.

References

1. W. Whitt. Stochastic – Process Limits. Springer, 2002
2. E. Smith, D. Farmer, L. Guillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. Quantitative Finance, 2003
3. R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. Operations Research, 2008
4. P. Robert. Stochastic Networks and Queues. Springer, 2003
5. F. Abergel and A. Jedidi. In preparation
6. T. Preis, S. Golke, W. Paul, and J.J. Schneider. Multi-agent-based order book model of financial markets. Europhysics Letters, 2006
7. A. Ferraris. Equity market impact models (Presentation). 2008
8. S. Asmussen. Applied Probability and Queues. Springer, 2003

[2] Technically the convergence happens in $D([0,\infty[)$, the space of $\mathbb{R}$-valued càdlàg functions, equipped with the Skorohod topology.

9. M. Bossy and N. Champagnat. Markov processes and parabolic partial differential equations. Encyclopedia of Quantitative Finance. R. Cont (Ed). *Wiley*, 2010
10. J.P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order books: empirical results and models. Quantitative Finance, 2002
11. S. Ethier and T. Kurtz. Markov Processes: Characterization and Convergence. Wiley, 1986
12. J. Gatheral and R. Oomen. Zero-intelligence realized variance estimation. Finance and Stochastics, 2007
13. F. Klebaner. Introduction to Stochastic Calculus with Applications. Imperial College Press, 2005
14. L.C.G. Rogers and D. Williams. Diffusions, Markov Processes, and Martingales. Cambridge University Press, 2000

Appendix A. Lyapunov Stochastic Stability Criterion

Let (X_t) be an irreducible Markov jump process on the countable state space $\mathscr{S}$, and L its infinitesimal generator. If there exists a function $f : \mathscr{S} \to \mathbb{R}_+$ and constants $K, \gamma > 0$ such that:

- *if $f(x) > K$ then $Lf(x) \leq -\gamma$;*
- *the random variables $\sup\{f(X_t) : t \leq 1\}$ and $\int_0^1 |Lf(X_t)|\mathrm{d}t$ are integrables;*
- *the set $F = \{x : f(x) \leq K\}$ is finite,*

then (X_t) is ergodic.

Appendix B. Figures

We provide below some representative figures obtained by numerical simulation of the order book (Figs. 2–6).

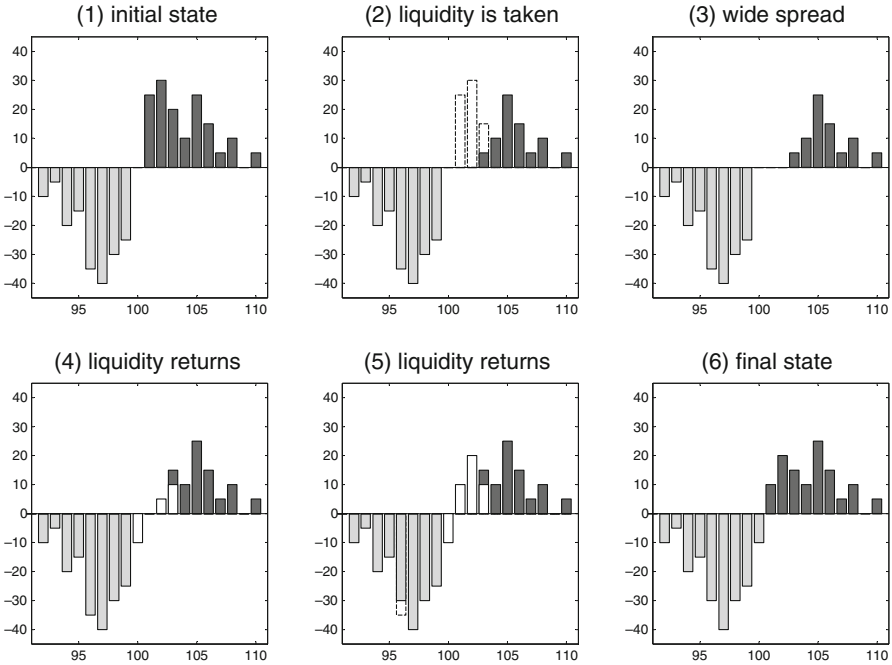


Fig. 1 Order book schematic illustration: A market buy order arrives and removes liquidity from the ask side, then limit sell orders are submitted and liquidity is restored [7]

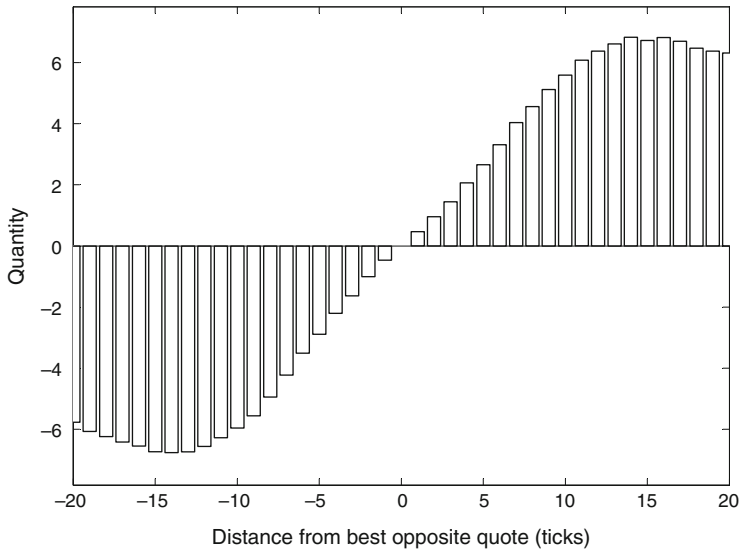


Fig. 2 Average depth profile

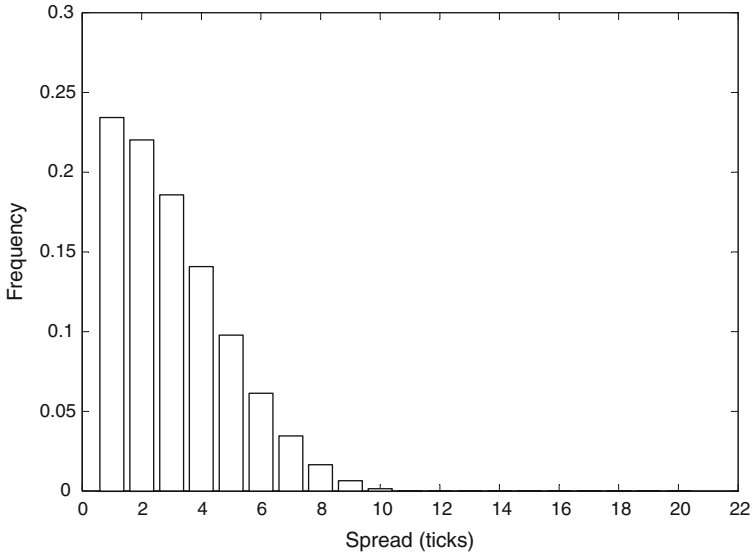


Fig. 3 Histogram of the spread

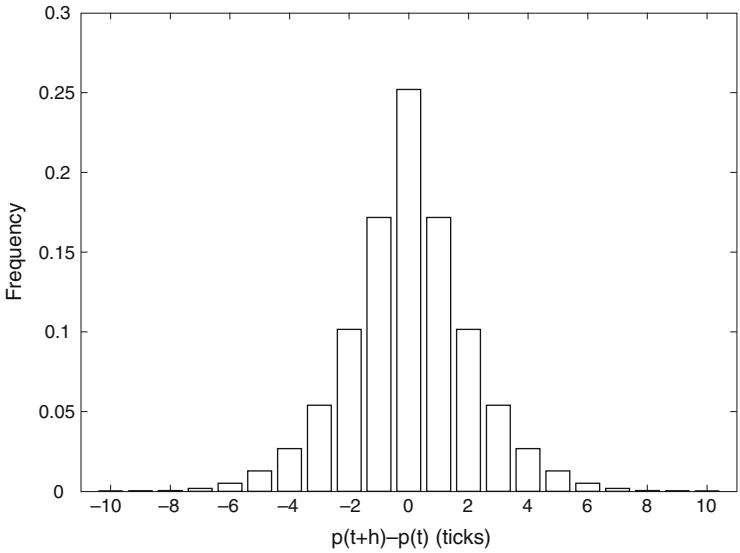


Fig. 4 Histogram of price increments

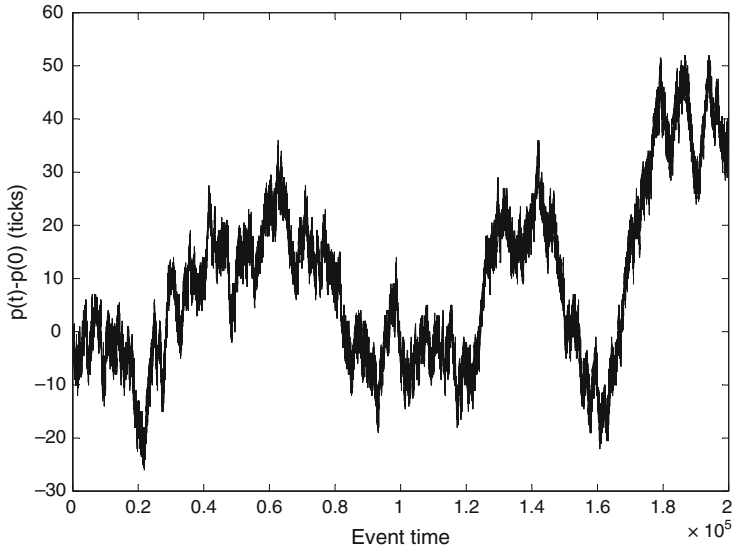


Fig. 5 Price sample path

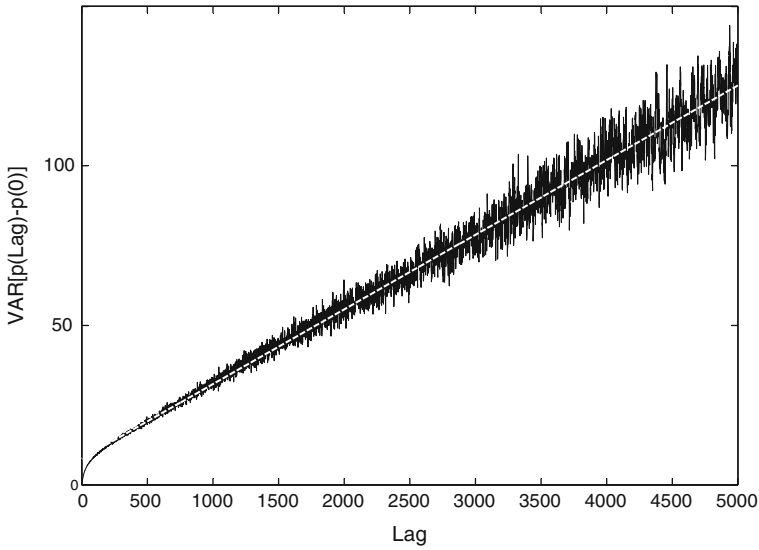


Fig. 6 Variance in event time. The dashed line is a linear fit

Reconstructing Agents' Strategies from Price Behavior

Valentina Alfi, Matthieu Cristelli, Luciano Pietronero and Andrea Zaccaria

1 Introduction

In the past years several Agents Based Models (ABMs) have been introduced to reproduce and interpret the main features of financial markets [7, 14]. The ABMs go beyond simple differential equations with the aim of being able to address the complex phenomenology of a dynamics. This phenomenology is usually interpreted in terms of the Stylized Facts (SF) which correspond to complex correlations beyond the simple Random Walk (RW). The ABMs give the possibility to describe the intrinsic heterogeneity of the market which seems to be responsible for many of these SF [6, 12]. The main SF are the fat tails for the fluctuations of price-returns, the arbitrage condition, which implies no correlations in the price returns, and the volatility clustering which implies long memory correlations for volatility.

Many of the existing ABMs are able to reproduce these SF, however, often these models contain a large number of parameters and variables and also introduce specific *ad hoc* assumptions. In this situation an analysis of the quality of these models is rather difficult. In addition, given the mentioned problems, it is difficult to point out which features of the model lead to some specific properties.

An important common element of most models is the competition between different agents' strategies. For example in the Lux-Marchesi (LM) [10, 11] model fundamentalist agents have a fixed reference price and bet that the real price will converge towards this value. These agents therefore induce a stabilizing tendency. Chartist agents instead tend to follow a price trend and to amplify it, so they have a destabilizing effect. Other important elements of many models are the herding

Valentina Alfi, Matthieu Cristelli, Luciano Pietronero, Andrea Zaccaria
Department of Physics, University of Rome La Sapienza, Rome,
e-mail: matthieu.cristelli@roma1.infn.it

Valentina Alfi
Centro Studi e Ricerche "E. Fermi", Rome

Matthieu Cristelli, Luciano Pietronero, Andrea Zaccaria
Istituto dei Sistemi Complessi, CNR, Rome

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

phenomenon and the role of price behavior with respect to the agents strategies. A long standing mystery of the LM model has been the fact that the interesting behavior of the SF is recovered only for a very particular, intermediate number of agents. In both limits of many and few agents the SF disappear. In addition this model is so complicated that it has not been possible to develop a systematic study of its properties.

Starting from these observations and considering the basic ingredients of the LM model, we have tried to build a minimal model which is still able to reproduce the main SF [2–5]. This "workable" model has permitted to clarify all the puzzling properties mentioned before and to consider a number of new questions in a systematic way. The main result is that indeed the SF correspond to a finite number of effective agents and this number is fixed by the market dynamics with a mechanism of self-organization. In this perspective the SF correspond to finite size effects and should not lead to genuine critical exponents. This is a basic result with important implications both conceptual and practical. Also the dynamics and fluctuations of the effective number of agents represents a new important element.

The simplicity of our model permits a systematic development of the field in various directions. In the present paper we show an example by considering the inverse problem, namely the reconstruction of the agents' strategies from the properties of the price time-series. This question is very relevant from both a conceptual and applied point of view and we use the model as a suitable playground to develop the appropriate methods. The reduced number of parameters permits in fact a reconstruction of these parameters from the price behavior. We study in detail the signal to noise ratio and show that, with a suitable data set, the reconstruction is indeed possible. Key elements in this respect are the size of the sliding window and the memory parameter of the chartists' strategies. There is a trade-off between the accuracy of the results and the speed to adjust to a change of strategy. We also show a first example of analysis of real data.

In Sect. 2 we summarize briefly the main elements of our model. In Sect. 3 we identify the criteria for the reconstruction method. In Sect. 4.1 we show the signal to noise ratio with respect to the various parameters of the analysis and discuss some specific examples. In Sect. 5 we present an analysis of an experimental price time series. In Sect. 6 we discuss the possible developments and applications of this method.

2 The Minimal Model

We consider two classes of agents: fundamentalists and chartists. Fundamentalist agents believe that a fundamental price p_f exists and they try to drive the actual price $p(t)$ towards the fundamental price which we can consider as a constant in the following without any loss of generality. The action of fundamentalists is then to buy if the actual price is under the p_f and to sell otherwise so they try to stabilize the market. We introduce also a positive parameter γ to quantify the strength of

fundamentalists' actions. In formula, at a certain time step t, the action of a single fundamentalist agent on the next price increment is:

$$A_f = \gamma(p_f - p(t)). \tag{1}$$

The chartists instead are noise traders who analyze the price time series in order to detect trends and gain profit from them. The trend detection is based on the evaluation of the distance between the actual price and a smoothed profile of it. In our model the smoothed profile of the price is simply the moving average $p_M(t)$ of the price $p(t)$ performed on the previous M time steps:

$$p_M(t) = \frac{1}{M}\sum_{k=0}^{M-1} p(t-k). \tag{2}$$

Chartist agents buy if the price $p(t)$ is above the moving average $p(t)$ and sell otherwise, so they can create bubbles and crashes which destabilize the market. Similarly to Eq. 1 the action of the chartists' can be written as:

$$A_c = \frac{b}{M-1}(p(t) - p_M(t)) \tag{3}$$

where the parameter b ($0 \leq b < 2$) tunes the strength of the signal and the factor $M - 1$ is needed to make the chartists' action independent on the choice of the parameter M [13, 15].

Considering a population of N agents divided in N_f fundamentalists and N_c chartists, the collective signal is given by the sum of all agents' actions plus an endogenous source of noise:

$$p(t+1) - p(t) = \frac{1}{N}\left[\sum_{i=1}^{N_c}\frac{b_i}{M_i-1}(p(t)-p_{M_i}(t)) + \sum_{i=N_c+1}^{N_f}\gamma_i(p_f-p(t))\right] + \sigma\xi. \tag{4}$$

In general we can consider an heterogeneous group of agents where any agent has a different characterizing parameter. We make a simplification by endowing any agent with the same parameter, but in the following we will comment also on the more general situation with heterogeneous agents. With this simplification Eq. 4 can be rewritten as:

$$p(t+1) - p(t) = \frac{N_c}{N}\frac{b}{M-1}(p(t) - p_M(t)) + \frac{N_f}{N}\gamma(p_f - p(t)) + \sigma\xi. \tag{5}$$

This mechanism for the price formation makes the model very simple and workable. For example it is interesting to stress that Eq. 5 can be solved analytically in the two limits cases where agents are all fundamentalists or all chartists. It also interesting to notice that here we have only two classes of agents (fundamentalists and chartists) while other models, for example the LM one, need 3 classes of agents to obtain a realistic dynamics.

Now we give the agents the possibility to decide to switch their strategy from fundamentalist to chartist and vice-versa. So we define two transition rates, which contain two terms. The former one is an herding term which encourages agents to switch strategy if the majority of the other agents have a different behavior. The latter term considers the possibility that one agent changes strategy following her own considerations about the market, simply by looking to the market signal she perceives.

Here we consider for the sake of simplicity the case where this second term is constant leading to the transition rates:

$$P_{fc} = B(1-\delta)(K + \frac{N_c}{N}) \tag{6}$$

$$P_{cf} = B(1+\delta)(K + \frac{N_f}{N}) \tag{7}$$

where P_{fc} is the probability to switch from fundamentalists to chartists and P_{cf} the opposite one. The parameter K rules the probability to change strategy independently on the other agents. The parameter δ introduces an asymmetry between the strategies. We can notice that in the limit $\delta = 0$ we recover the Kirman's Ants Model [8] for the dynamics of agents' strategies. More details about the model and its capability to reproduce the SF can be found in [3–5].

3 The Reconstruction Method

In this section we introduce a method to analyze the output of the model described in Sect. 2 in order to reconstruct the agents' strategies. We consider our model as a black-box generating a price series and we try to detect a posteriori the different regimes of the dynamics, i.e. when the system is dominated by fundamentalists or chartists. This could be very useful in view of an analysis of experimental data from real markets [1, 13, 15].

Before describing the method we notice that, while the moving average p_M can be always estimated from data, this is not the case for the fundamental price p_f. So one has to consider two different situations, one in which the fundamental price p_f is known (for example from economic considerations) and the other in which p_f is unknown. In this last case one can approximate it with the moving average $p_M(t)$. It is worth noticing that this approximation is a priori valid only in the limit of very large M but, as we will show later, we can obtain good results also for intermediate values of M. In the following we indicate the strength of the fundamentalist agents with $s_f = \gamma$, the one of chartists with $s_c = b/(M-1)$ and with x the fraction of chartists N_c/N. With this notation Eq. 5 becomes:

$$p(t+1) - p(t) = s_c x(p(t) - p_M(t)) - s_f(1-x)(p(t) - p_f) + \sigma\xi. \tag{8}$$

The price return $p(t+1) - p(t)$ is a function of the two variables $p(t) - p_M(t)$ and $p(t) - p_f$ and so it can be seen as a plane whose slopes depend on the parameters

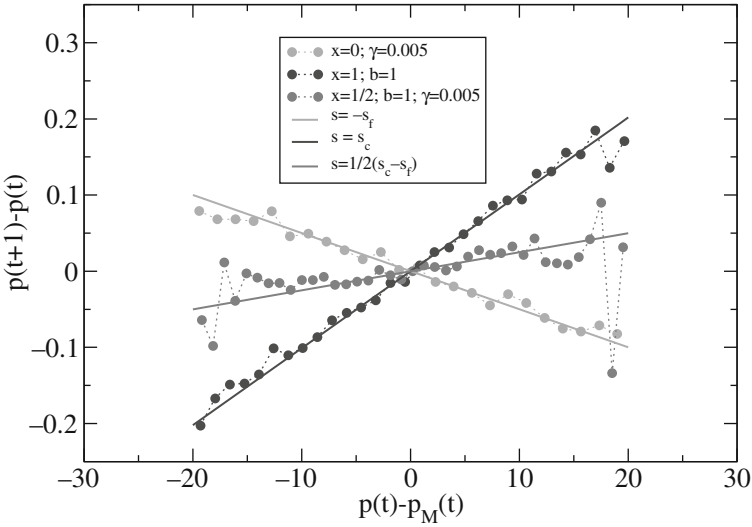


Fig. 1 The method we are describing consists in reconstructing the agents' strategies by considering the linear relation between the 1-step price increment and the distance between the price and its moving average. By fitting this linear relation one can recover the sign and strength of the model's parameters. In this picture we show three different situations corresponding to three different slopes. A positive slope indicates a destabilizing tendency and vice-versa. The results of the reconstruction are given by the *dots* while the *straight lines* represent the expected values known from the model's parameters

of the model (b, M, γ) and on the chartist fraction x. In particular the gradient of this plane is positive in the $p(t) - p_M(t)$ direction and negative in the other one. One can calculate the quantity $p(t+1) - p(t)$ as a function of $p(t) - p_M(t)$ and $p(t) - p_f$ from a given data-set, that in this case is the output of the model, and then obtain an estimation of the gradients $s_c x$ and $-s_f(1-x)$ by simply plotting these quantities and performing two linear regressions.

Approximating p_f with p_M we obtain a simplification of the procedure because $p(t+1) - p(t)$ becomes a function of one variable. So, approximating p_f with a suitable $p_M(t)$, Eq. 5 becomes:

$$p(t+1) - p(t) \approx \Big[s_c x - s_f(1-x)\Big](p(t) - p_M(t)) + \sigma\xi. \tag{9}$$

In this way the plane described by Eq. 8 becomes a straight line with slope given by:

$$s = s_c x - s_f(1-x) = (s_c + s_f)x - s_f \tag{10}$$

and Eq. 9 can be compactly written as

$$p(t+1) - p(t) \approx s(p(t) - p_M(t)) + \sigma\xi. \tag{11}$$

We will plot $p(t+1) - p(t)$ as a function of $p(t) - p_M(t)$ and with a linear fitting procedure we can estimate the slope s. Due to the noise term, we need to consider reasonably long intervals to extract a precise value of s. Given the slope s and the values of the parameters b, M and γ, we can recover the fraction of chartists x

$$x = \frac{s - s_f}{s_c + s_f}. \tag{12}$$

Even if the model's parameters are not known, being the slope s proportional to the fraction of chartists x, the estimation of s gives an idea of the "sentiment" of the market, i.e. if it is in a stable regime $s < 0$ or in an unstable one $s > 0$. From Eq. 10 we can observe that, if the fundamentalists' parameter $s_f = \gamma$ is strong enough, one can obtain a stable market also with a larger number of chartists and, on the contrary, if the parameter s_c is larger than s_f one can obtain an unstable market even if fundamentalists are more than chartists.

In Fig. 1 we show an example of three straight lines obtained plotting $p(t+1) - p(t)$ as a function of $p(t) - p_M(t)$. We have considered three different situations where we have run our model keeping fixed the fraction of chartist agents x. The parameters of the simulation are $b = 1$, $M = 100$ and $\gamma = 0.005$ and we plot the results for the cases $x = 0$, $x = 1$ and $x = 1/2$. We can observe that the resulting slope s is indeed proportional to the fraction of chartists x.

4 Reconstruction for the Agent Model

4.1 Window Size Dependence

The noise term in Eq. 11 introduces some fluctuations in the estimate of the slope s. If the parameters to be estimated remain constant, a larger dataset will give more accurate results. On the other hand the shorter is the interval on which the analysis is performed the more reactive will be the response to a change in market sentiment, so we need a trade-off between these two competing effects.

In order to look for the minimal interval for the analysis, we consider the limit case in which there are only chartist agents. In this case $s = s_c = b/(M-1)$ and so, generating time series of different length, we can check the dependence of the estimation of the slope s_c on the size of the dataset. We have considered a long dataset with parameters $b = 1$ and $M = 50$ and we have divided it into windows of size T, then we have performed the estimation of the parameter b inside each window. Varying T we can observe that for small windows, as expected, the fluctuations are very large and the results are inconclusive. In Fig. 2 we show the results for five different window sizes and we can observe that for small values of T the fluctuations are too large and the estimation of the parameter could be sometimes misleading. In Fig. 3 we show the dependence of the accuracy of the estimations of b on the window size. We can observe that for window sizes smaller than $T = 1000$

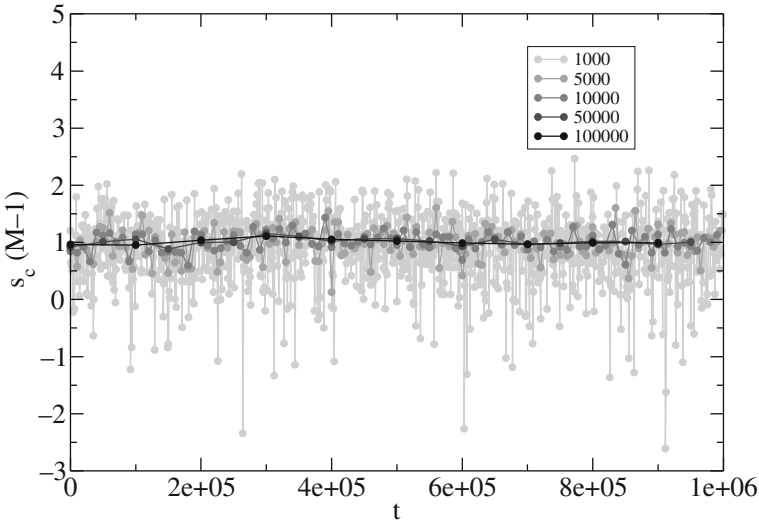


Fig. 2 Signal to noise ratio. Analysis of a simulation of the model with only chartist agents with parameter $b = 1$ and $M = 50$. We plot the estimation of the parameter $b = s_c(M - 1)$ for different sizes of the sliding window. The fluctuations in the estimation of b become larger for small values of T

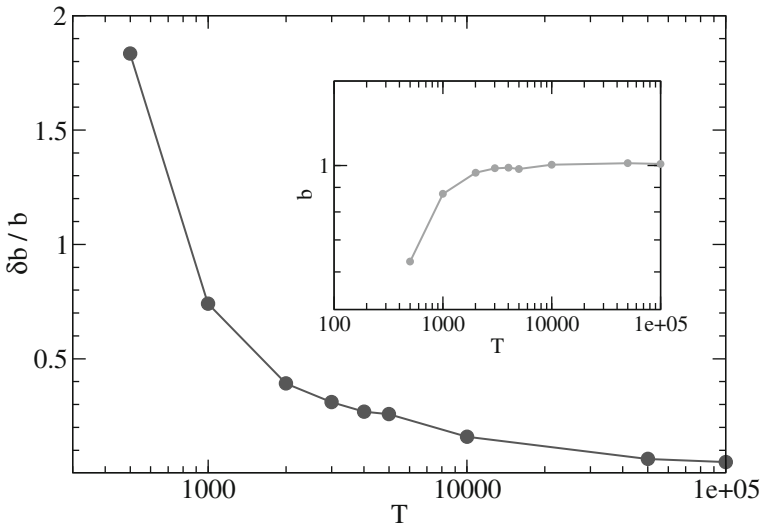


Fig. 3 We consider a population of only chartists with parameters $b = 1$ and $M = 50$. We show the reconstructed values of the parameter b and its relative fluctuations for different values of T. We observe that a window of at least 1000 points is needed to perform an efficient analysis. In the *inset* we report the estimation of b as a function of the window size

the relative fluctuations are very large ($\delta b/b > 1$) and also the average value of the parameter b is precise, as one can see in the inset of Fig. 3. For the next analysis we will use windows of size $T \geq 1000$ in order to have a good compromise between accuracy and sensitivity in parameters' changes.

4.2 Dependence on the Length of the Moving Average

Then we check the method for the case of all fundamentalists. In this case p_f is approximated by the moving average of the price p_M. In order to estimate the parameter γ one should plot the increments $p(t+1)-p(t)$ as a function of $p(t)-p_f$. In Fig. 4 we have plotted the estimation of the parameter $s_f = \gamma$ for different values of the number of steps used to perform the moving average p_M. We can see that even for small values of M ($M = 10$) the absolute value of the parameter γ is well detected and one can obtain a good fit for values of M starting from 50.

Before analyzing the output of the model we test the dependence of the analysis on the choice of the parameter M which is the number of time steps used for the computation of the moving average $p_M(t)$.

We have analyzed the two limit cases of populations consisting in only fundamentalist agents or only chartist agents ($x = 0$ and $x = 1$ respectively). In the

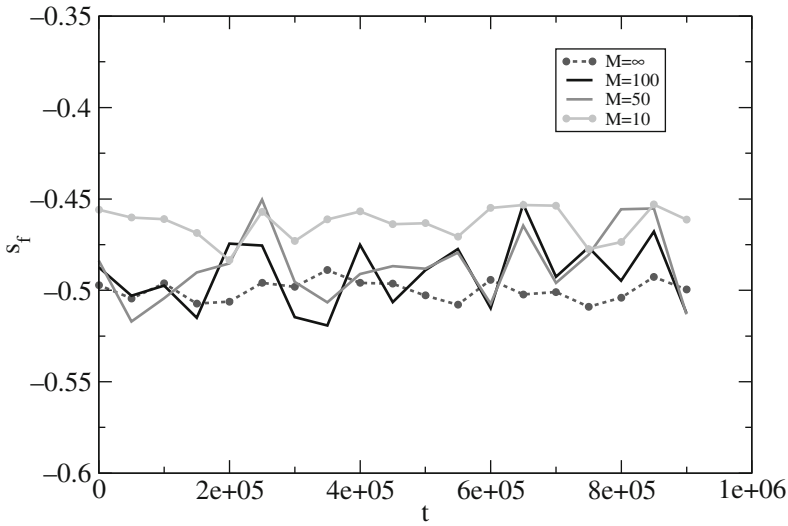


Fig. 4 Here we show a test of the approximation $p_f \approx p_M(t)$. We have run a simulation with only fundamentalists with $s_f = \gamma = 0.5$ and, in performing the backward analysis, we have approximated the fundamental price with a moving average of different lengths. For comparison we also show the analysis performed with the true p_f (dashed line). We can see that the estimation of γ improves for increasing values of M but the result is reasonable already for $M = 10$

case $x = 0$ we estimate the parameter $s_f = \gamma$ by considering the approximation of $p_f \approx p_M(t)$. In this case the price $p(t)$ describes a diffusive process only on very short time scales but on longer time scales the process is no more diffusive and the variance tends to a constant value [4]. So we expect the fluctuations δs_f of the parameter $s_f = \gamma$ to be proportional to the fluctuations of $p(t) - p_M(t)$ and hence to the variance of the underlying process. In Fig. 5 we plot the quantity $\delta s_f / s_f$ as a function of the moving average length M and, as expected, the fluctuations increase for short time scales ($M < 1000$) and become almost constant for longer time scales. In the inset we have plotted an average value of the parameter s_f as a function of M which, as already noted in Fig. 4, is converges to the correct value as M increases.

The situation is different in the opposite limit $x = 1$ in which the process is superdiffusive (with respect to a simple RW) for short time scales and recovers a normal diffusion for longer time scales. Considering the scaling properties of the process, the fluctuations δb increase with M as M^α with $1/2 \leq \alpha < 1$ and so, being $s_c = b/(M-1)$ and $\delta s_c = \delta b/(M-1)$, one obtain that δs_c decreases for increasing values of M while $\delta s_c / s_c = \delta b / b$ has the opposite behavior.

In the case $x = 1$ the value of M appears both in the basic model dynamics and in the reconstruction method. The value of M of the model is in principle not known in the reconstruction analysis. Here we start by using the same value of M for both

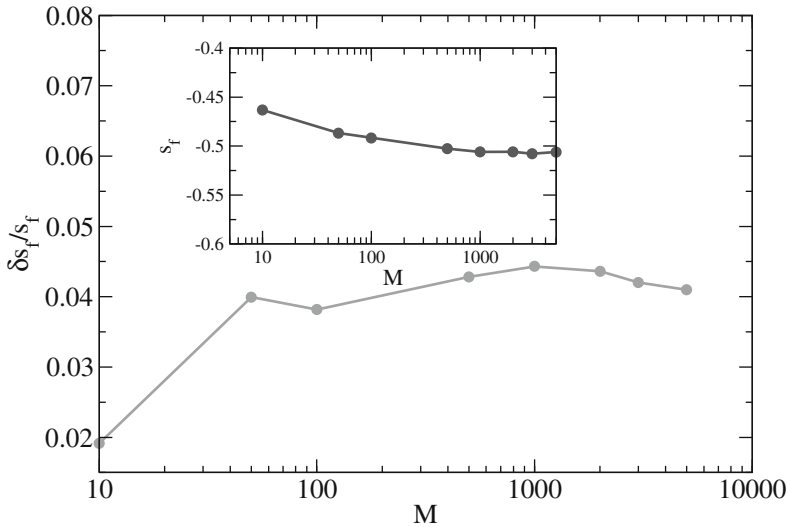


Fig. 5 Fluctuations of the parameters' estimation as a function of the value of M used in the backward analysis. The case shown corresponds to $x = 0$ (only fundamentalists). In this case the dynamics is diffusive only on very small time scales and becomes non diffusive for longer times. The relative fluctuations $\delta s_f / s_f$ increase for small values of M and then converge to a constant value (*main figure*) while the estimation of the parameter s_f is more and more precise for increasing values of M

cases while in the following we will consider the more realistic case in which the real value of M used by the chartists is unknown.

In Fig. 6 we have plotted the quantities $\delta s_c/s_c$ and δs_c (in the inset) as functions of M. We can observe that the first quantity is an increasing function of M while the second one decreases. If the parameter M (that we indicate with M_{true}) of the chartist strategy is unknown one must perform the backward analysis by fixing a parameter M which will not be in general the correct one. In Fig. 7 we show a simulation with only chartists ($x = 1$) and with parameters $b = 1$ and $M_{\text{true}} = 50$ and then we perform the backward analysis with different values of the parameter M ranging from 10 to 500. We also plot the values of the parameters estimation with the right value of M. The result is that, apart from values of M smaller than M_{true}, the backward analysis is able to produce a good estimation of the parameter b for any value of $M > M_{\text{true}}$.

We have seen in Fig. 6 that the fluctuations in the estimation of the parameter b increase with M. When the value of M used for the backward analysis is different from the simulation of the dynamics (except for the case of $M = 50$), the situation is slightly different. In fact, while the fluctuations $\delta b/b$ increase with large M (like in Fig. 6) when $M > M_{\text{true}}$, they exhibit a minimum for $M = M_{\text{true}}$, so the backward analysis is able to identify the value of M used to simulate the dynamics.

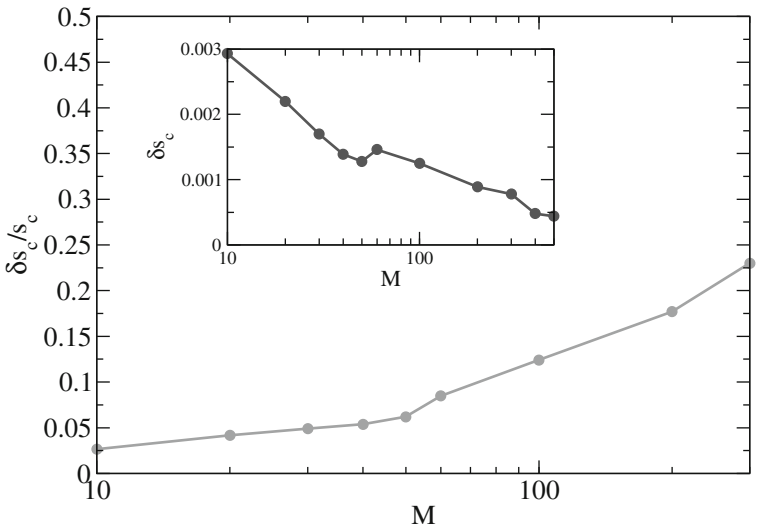


Fig. 6 Fluctuations of the parameters' estimation as a function of the value of M used in the backward analysis for the case $x = 1$ (only chartists). We consider various simulations with different values of M and then performed the reconstruction analysis with the corresponding M. If there are only chartists the dynamics is diffusive and so the relative fluctuations $\delta s_c/s_c$ increase for increasing values of M (*main figure*) while the standard deviation $\delta s_c = \delta b/(M-1)$ decreases with M due to the $M - 1$ factor (*inset*)

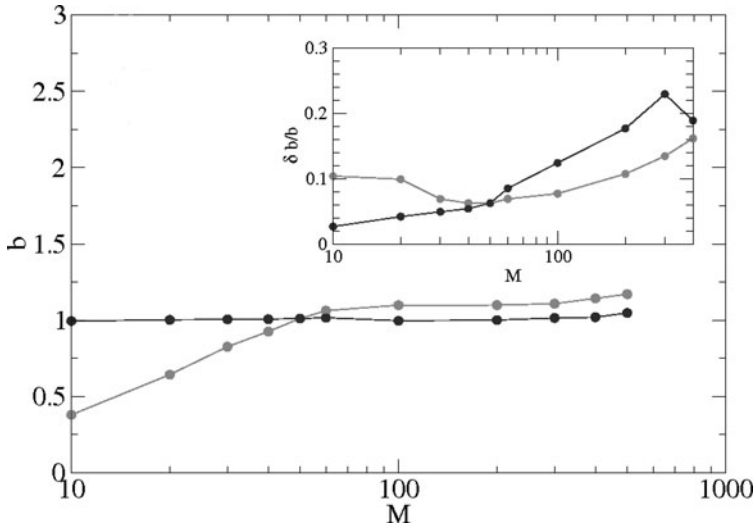


Fig. 7 Here we test the situation in which the parameter M used for the backward analysis is different from the one of the model's simulation. We consider a population of chartists ($x = 1$) with parameters $b = 1$ and $M_{\text{true}} = 50$. Then we perform the reconstruction method for different values of M (*light gray line*). For comparison we also show a reconstruction made using the correct value $M_{\text{true}} = 50$ (*dark gray line*). As we can see from the *main picture* the method underestimates the parameter b when $M < M_{\text{true}}$ but the reconstruction is good for values $M > M_{\text{true}}$. In Fig. 6 we have seen that $\delta b/b$ is an increasing function of M but in this case we have an anomalous behavior when $M < M_{\text{true}}$. In fact $\delta b/b$ decreases up to a minimum when $M = M_{\text{true}}$ disclosing the "ability" of the model to recognize the right value of $M = M_{\text{true}}$ used for the dynamics

4.3 Backward Analysis on the Agent Model

Now we are going to check the sensitivity of the method developed in the previous sections to detect the changes in agents' strategies of the ABM described in Sect. 2. We start from the simple 1-agent case in which the agent switches from chartist to fundamentalist and vice versa with a constant probability BK. We have run a 1-agent simulation with parameters $b = 1$, $\gamma = 0.1$ and $M = 50$. Therefore the fraction of chartists x can assume only two values, i.e. "0" and "1", and consequently the total slope s is alternatively equal to $-s_f$ or s_c.

In Fig. 8 we plot the time evolution of the agent's strategy and the corresponding backward analysis of the total slope s defined in Eq. 10. We have chosen a window size of 5000 points in order to have a rather reactive analysis. In fact we can observe that the estimation of the total slope s is precise and sensitive.

Before analyzing the general multi-agents model, let us consider that the total slope s depends on the fraction of chartists x in the way described by Eq. 10. To check this relation with the backward analysis we have performed several simulations keeping x fixed and then we have measured the corresponding slope.

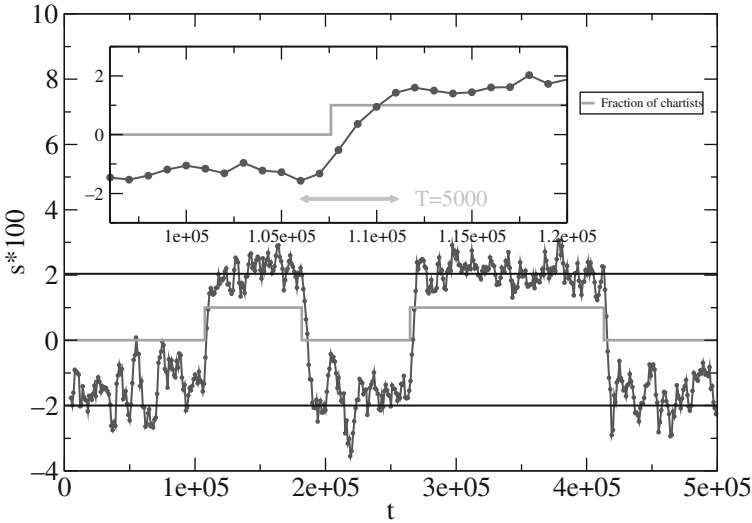


Fig. 8 Analysis of the 1-agent case with parameters $b = 1$, $M = 50$ and $\gamma = 0.02$ and sliding window size $T = 5000$. In the *main picture* we show the model dynamics (*light gray line*) which corresponds to $x = 1$ if the agent is chartist and $x = 0$ if fundamentalist. The *dark gray line* is the a posteriori reconstruction which corresponds ideally to s_c or $-s_f$ respectively. In the *inset* we report a magnification of a period in which the agent switches from fundamentalist to chartist. We can see that the method is reasonably reactive with respect to the change of strategy

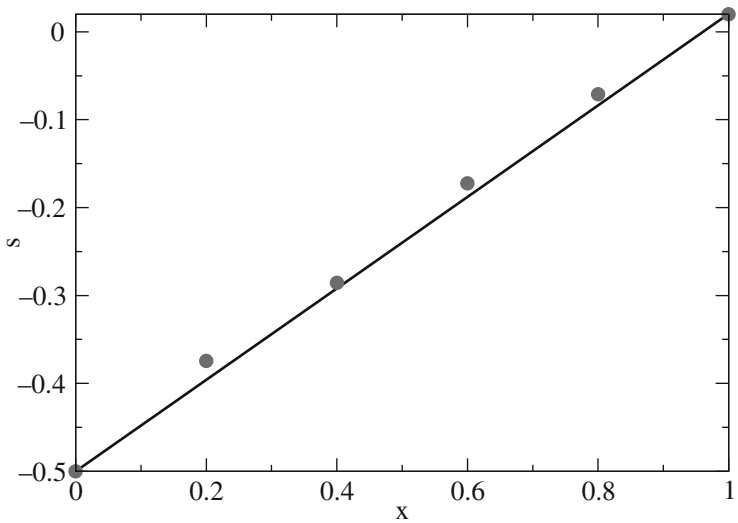


Fig. 9 Test of the linear relation between the total slope s and the fraction of chartists x. We consider a population of $N = 100$ agents with parameters $b = 1$, $M = 50$ and $\gamma = 0.5$. We observe a good agreement between the simulations and the prediction given by Eq. 10

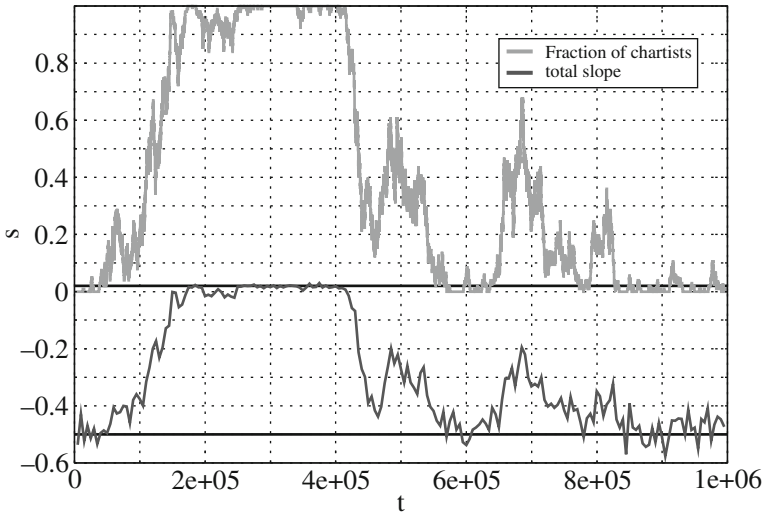


Fig. 10 Strong chartist situation. Reconstruction analysis for $N = 100$ agents with parameters $b = 1$, $M = 50$ and $\gamma = 0.5$ and asymmetry parameter $\delta = 0$. We plot the fraction of chartists (*light gray line*) as a function of time and the total slope s (*dark gray line*) obtained by the reconstruction method

An analysis of the slope as a function of the number of chartists for a 100-agents simulation with parameters $b = 1$, $\gamma = 0.5$ and $M = 50$ is shown if Fig. 9. In this case the fundamentalists' parameter s_f is larger than the chartists' one s_c of about a factor 25. Consequently the total slope s can be negative even if the chartists are the majority. In the case of Fig. 9 we have an example of this situation because the slope s become positive only when the fraction of chartists x is greater than the quantity $s_f/(s_c + s_f)$ that in this case is about 0.96. In the symmetric case with $s_f = s_c$ we recover the situation where the sign of the slope s indicate the predominant strategy.

We now consider the model with 100 agents with a symmetric dynamics ($\delta = 0$). The parameters are the same used in the previous analysis. We perform a sliding linear regression and we reconstruct the time-dependent values of the total slope s.

In Fig. 11 we show a comparison between the time estimation of the parameter s and the dynamics of the fraction of chartists x. We can observe that we obtain an accurate reconstruction of the model's dynamics. In this symmetric case the states $x \simeq 0$ and $x \simeq 1$ are equally probable, so one can observe a complete inversion of population from one strategy to the other.

In the asymmetric case $\delta \neq 0$ the situation is slightly different because there is a preferred state (the fundamental one in our case) and, depending on the asymmetry's strength, it may be not possible to observe a complete inversion of population. In this asymmetric situation the system is almost locked in the fundamentalists state with small and short (with respect to the time window size) bursts of chartists, so

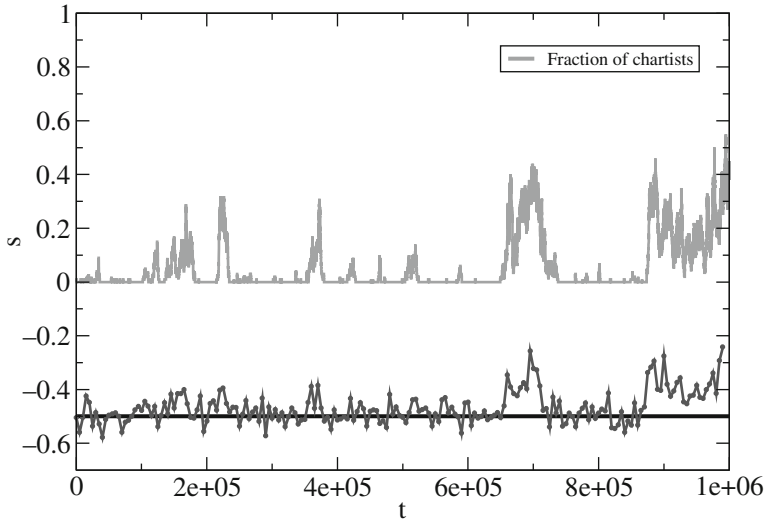


Fig. 11 Weak chartists situation. Reconstruction analysis of a $N = 100$ agents with parameters $b = 1$, $M = 50$ and $\gamma = 0.5$ and asymmetry parameter $\delta > 0$. We plot the fraction of chartists (*light gray line*) as a function of time and the total slope s (*dark gray line*) obtained by the reconstruction method. In this asymmetric case where the system does not undergo to complete inversion of population and the signal is rather weak, the reconstruction analysis is less efficient

our backward analysis is much more difficult to be carried out. In Fig. 11 we show the asymmetric dynamics $\delta \neq 0$ of the fraction of chartists x and the slopes s reconstructed from the analysis. If the bursts of chartists are too small in amplitude ($x < 0.3$) the signal can not be distinguished from the noise. The duration of the bursts is an important aspect too, because we have to consider very small windows to detect a short signal and the shorter is the window, the noisier is the result.

5 Analysis on Experimental Data

Now we are going to consider a first example of an analysis of real data. In this case we do not know the fundamental price p_f so our analysis is similar to the one of [1, 13, 15, 16]. However, in our scheme there are several combinations of parameters which can lead to the same observed slope. The knowledge of p_f (from fundamental analysis or other sources) would permit to remove at least partially this degeneracy and to use all the information the method can give.

In Fig. 12 we have plotted the tick-by-tick time series of the [rice of the stock *General Electrics* for a period of three months. Then we have performed the backward analysis with sliding windows of size $T = 5000$ to estimate the total slope s. We can observe that the values of the slope are almost always negative indicating a stable market. In view of the results shown in Fig. 9 a negative slope can be due both to a larger number of fundamentalists or to a stronger strength of the

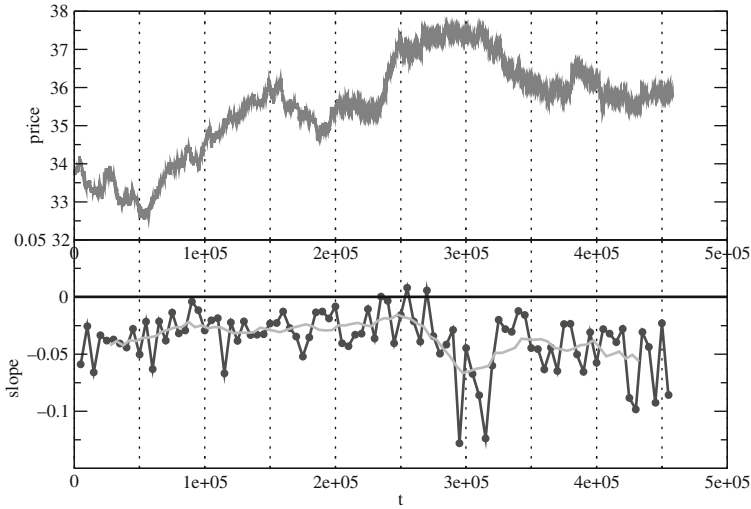


Fig. 12 Tick-by-tick price time series of the stock *General Electrics* for a period of about three months (*upper panel*). In the *lower panel* we plot the results of the reconstruction method performed with windows of size $T = 5000$. We can see that the total slope s is almost always negative and then the resulting dynamics can be interpreted as dominated by a stabilizing tendency. This can be due both to a dominance of fundamentalists or to the presence of a small number of fundamentalists but with a very strong weight on the market

fundamentalists independently on their number. In this paper we show this preliminary result as a simple test and we will consider a more detailed analysis on larger data-sets in future works. It should be mentioned that recently the reconstruction of agents' strategies have been performed from inventory variations [9]. It would be interesting to compare the various methods on the same time series.

6 Conclusions and Perspectives

In this paper we have developed a method for reconstructing the agents' strategies from the price time series. The method is based on a new agent based model we recently introduced [3–5].

In this work we used it as a black-box to develop and test a method to reconstruct the agents' strategies from price behavior. Having very few parameters and variables we can easily reconstruct the strategies because of their direct relation with the parameters. In principle an optimal reconstruction would imply the knowledge or the estimation of the value of p_f. This would lead to a more complex situation with respect to the one discussed in the present paper. In future we are going to consider also this possibility. For the moment the value of p_f is approximated by a suitable moving average and in the present scheme its knowledge would add some information on the optimal combination of parameters corresponding to price dynamics. The reconstruction method is essentially based on a backward analysis of

the model's dynamics. A basic ingredient is the size of the sliding window considered. We study in detail the signal to noise ratio corresponding to the size of the window. We also consider the response to a change in agents' strategies. We have found that there is a necessity of a trade-off between large windows, which lead to accurate results, and small windows which give a quick response to a change. In future works we intend to explore the optimization of this trade-off in real data. We also intend to generalize the presented method in order to achieve the maximum information in the case in which the value of p_f is known from fundamental analysis.

References

1. Alfi, V., Coccetti, F., Marotta, M., Pietronero, L., Takayasu, M.: Hidden forces and fluctuations from moving averages: A test study. Physica A **370**, 30–37 (2006)
2. Alfi, V., Cristelli, M., Pietronero, L., Zaccaria, A.: Mechanisms of self-organization and finite size effects in a minimal agent based model. J. Stat. Mech. p. P03016 (2009)
3. Alfi, V., Cristelli, M., Pietronero, L., Zaccaria, A.: Minimal agent based model for financial markets I: origin and self-organization of stylized facts. Eur. Phys. J. B **67**, 385–397 (2009)
4. Alfi, V., Cristelli, M., Pietronero, L., Zaccaria, A.: Minimal agent based model for financial markets II: statistical properties of the linear and multiplicative dynamics. Eur. Phys. J. B **67**, 399–417 (2009)
5. Alfi, V., Pietronero, L., Zaccaria, A.: Self-organization for the stylized facts and finite-size effects in a financial-market model. EPL **86**(5), 58,003 (2009). DOI 10.1209/0295-5075/86/58003. URL http://dx.doi.org/10.1209/0295-5075/86/58003
6. Bouchaud, J.P., Potters, M.: Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management. Cambridge University Press (2003)
7. Challet, D., Marsili, M., Zhang, Y.C.: Minority Game: interacting agents in financial markets. Oxford University Press (2005)
8. Kirman, A.: Ants, rationality and recruitment. Quarterly Journal of Economics **180**, 137–156 (1993)
9. Lillo, F., Moro, E., Vaglica, G., Mantegna, R.N.: Specialization and herding behavior of trading firms in a financial market. New Journal of Physics **10**(4), 043,019 (2008). URL http://stacks.iop.org/1367-2630/10/i=4/a=043019
10. Lux, T.: Turbulence in financial markets: the surprising explanatory power of simple cascade models. Quantitative Finance **1**, 632–640 (2001)
11. Lux, T., Marchesi, M.: Scaling and criticality in a stochastic multi-agent model of a financial market. Nature **397**, 498–500 (1999)
12. Mantegna, R.N., Stanley, H.: An Introduction to Econophysics: Correlation and Complexity in Finance. Cambridge University Press, New York, NY, USA (2000)
13. Mizuno, T., Takayasu, H., Takayasu, M.: Analysis of price diffusion in financial markets using PUCK model. Physica A Statistical Mechanics and its Applications **382**, 187–192 (2007). DOI 10.1016/j.physa.2007.02.049
14. Samanidou, E., Zschischang, E., Stauffer, D., Lux, T.: Microscopic models of financial markets. Tech. rep. (2006)
15. Takayasu, M., Mizuno, T., Takayasu, H.: Potential force observed in market dynamics. Physica A Statistical Mechanics and its Applications **370**, 91–97 (2006). DOI 10.1016/j.physa.2006.04.041
16. Alfi V., De Martino, A.T., Pietronero, L.: Detecting the traders' strategies in minority-majority games and real stock-prices. Physica A **382**, 1 (2007)

Market Influence and Order Book Strategies

François Ghoulmié

Abstract. I review in this paper[1] my findings on order driven market modeling. Following my previous works on robust agents based modeling in finance [1–3, 5], I study specific characteristics of order book markets. By controlling the descriptive time scale of the dynamics involved, I show how market impact, linear by definition, and trading strategies lead to precise pictures for clarifying order book dynamics, consistent with what is observed empirically. I then discuss more specifically the role of market impact in the created dynamics and structure of the book and the economic implications of my studies.

The article is organized as follows. In Sect. 1, I describe financial market dynamics in an agent-based market model that clarifies the role of volatility in characteristics observed on a wide range of descriptive time scales. I define in Sect. 2 the limit order book model, agents' strategies, and link liquidity provision to volatility estimates. I focus the analysis on the dynamics and structures of the book in Sect. 3. I discuss the economic implications of the results and draw conclusions in the last sections.

1 Properties of the Single Asset Market Model

The model describes a market where a single asset is traded by n agents and I recall its main properties for pedagogical purposes [1–3]. Trading takes place at discrete time steps t. Provided the parameters of the model are chosen in a certain range, these periods may be interpreted as "trading days". At each time period, every agent receives public news about the asset's performance. If the news is judged to be significant the agent places for a unit of asset a buy or sell order, depending on whether

François Ghoulmié
Académie de Versailles, Paris,
e-mail: Francois.Ghoulmie@gmail.com

[1] Paper prepared for the Kolkata V Econophysics Conference at the Saha Institute of Nuclear Physics, India 2010. The discoveries presented in this paper are consistent with the logics and structures of my research works at École Normale Supérieure, Paris, at the Santa Fe Institute and at the Australian National University.

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

the news received is pessimistic or optimistic. Prices then move up or down according to excess demand. The model produces stochastic heterogeneity and sustains it through the updating of agents' strategies. Let us recall in a mathematical way the ingredients of the single asset model. At each time period:

- All the agents receive a common signal ϵ_t generated by a Gaussian distribution with 0 mean and standard deviation D, namely $N(0, D^2)$.
- Each agent i compares the signal to its threshold $\theta_i(t)$.
- If $|\epsilon_t| > \theta_i(t)$ the agent considers the signal as significant and generates an order $\phi_i(t)$ according to
$$\phi_i(t) = 1_{\epsilon_t > \theta_i} - 1_{\epsilon_t < -\theta_i}, \tag{1}$$
where $\phi_i(t) > 0$ is a buy order, $\phi_i(t) < 0$ is a sell order and $\phi_i(t) = 0$ is an order to remain inactive.
- The market price p_t is affected by the excess demand and moves according to
$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right) = g\left(\frac{\sum_i \phi_i(t)}{n}\right), \tag{2}$$
where r_t are the returns at time t and g is the price impact function.
- Each agent updates, with probability s, her threshold to $|r_t|$.

The evolution of the thresholds distribution is described with the following master equation:
$$f_{t+1}(\theta) = (1-s)\ f_t(\theta) + s\ \delta_{|r_t|,\theta}, \tag{3}$$
with $|r_t| = |g(\mathrm{sign}(\mathrm{ffl_t})\mathrm{F_t}(|\mathrm{ffl_t}|))|$, F_t being the cumulative distribution of the thresholds.

The solution of Eq. 3 can be derived analytically and reads as
$$f_t(\theta) = (1-s)^t\ f_0 + s\sum_{j=1}^{t}(1-s)^{j-1}\delta_{|r_{t-j}|,\theta}. \tag{4}$$

Moreover, numerical tests confirm the validity of the former solution. Stationary solutions are the limiting cases: without feedback $s = 0$, and without heterogeneity $s = 1$. We specify now the range of parameters that leads to realistic price behaviors. First of all, we want a large number of agents in order to guarantee heterogeneity in the market. Indeed, when the number of agents is lowered, the distribution of returns becomes multi-modal with 3 local maxima, one at zero, one positive maximum and a negative one. This can be interpreted as a disequilibrium regime: the market moves either one way or the other. The updating frequency s should be chosen small, $s << 1$, in order to guarantee heterogeneity. When the amplitude of the noise is small, $D << g(1/n)$, the absolute value of the returns evolves through a series of periods characterized by "jumps" whose amplitudes decay exponentially in time. The sensitivity of the thresholds increases when the noise level increases and the behavior of the returns is closer to the Gaussian signal. On the other hand, when the amplitude of the news is too high, $D >> g(1)$, the returns distribution has two

peaks: the maximum at $g(1)$ and the minimum at $g(-1)$. We thus want the following condition in order to get realistic returns dynamics:

$$g(1/n) << D << g(1). \tag{5}$$

If we consider now the linear price impact function, $g(x) = x/\lambda$, with λ the market depth characterizing how much the market moves when filling one unit of asset, the above condition leads to the parameter reduction $D_{\text{eff}} = D\lambda$ and $1/n << D_{\text{eff}} << 1$. We then get clusters of volatility of length $1/s$ consistent with the correlation structure suggested by the stationary solution. This slow feedback mechanism generates endogenous heterogeneity, excess volatility, volatility clustering and transforms Gaussian news into semi heavy-tailed price returns. When a majority of agents have a low value for their threshold a large price fluctuation becomes very probable. Because only a small fraction of agents increases its threshold response when a large fluctuation occurs the probability of also getting a large fluctuation at the next time step remains high. In other words, the slow feedback mechanism causes persistence in the fluctuations.

2 Definition of the Agent-Based Limit Order Book Structure

I now extend the agent-based study to limit order book dynamics [5] and show how by linking the previous volatility based model to liquidity I obtain the order book high fluctuations. The model describes a limit order book where a single asset is traded by various types of agents. These traders populate the book with limit and market orders. Trading takes place at discrete time steps t. These periods may be interpreted on a wide range of time scales: from tick level to trading days or weeks. The trades in the model result from matching market orders against limit orders. The first group of agents is composed of n market neutral traders operating as liquidity providers: at each time period, each agent i of this group places limit orders (bid and ask) of one unit size at distance $\theta_i(t)$ ($\sim$ estimation for the volatility during [t,t+1]) from the price (mid-price which is the average of the best bid and best ask). Given a signal on expected price change ("new information") IID Gaussian noise $\epsilon_t \sim N(0, D^2)$ with D= noise level, informed agents then send market orders to buy or sell limit orders that are at a distance below the expected market price. Prices then move up or down according to the direction of the trades. The model produces stochastic heterogeneity and sustains it through the updating of agents' strategies. Let us define in a more mathematical way the ingredients of the limit order dynamic model.

2.1 Strategic Forces in the Trading Arena

I describe first the adaptive decision-making rules at each time period. Liquidity providers' strategies are distances or thresholds $\theta_i(t)$ that determines their bid and

ask orders. These positions results from their expectations of price movements and their risk aversion related for example to the uncertainty of execution and the waiting time where informed traders can take advantage of the liquidity they are providing. Each agent i places a bid and ask according to:

$$\begin{aligned} \text{bid}(i) &= p_t - \theta_i(t) \\ \text{ask}(i) &= p_t + \theta_i(t). \end{aligned} \tag{6}$$

Given an information ϵ_t on the price change, which can have many potential origins (networks of friends, public news, private information) informed agents then send market orders to buy or sell limit orders that are at a distance below the expected market price. The market order flow is thus determined by the direction of the price change and the number of available limit orders in the book. Each liquidity provider's limit order will be likely to be executed or not according to:

- buy order likely to be executed if $\epsilon_t > \theta_i(t)$: $\phi_i(t) = +1$ (buy);
- sell order likely to be executed if $\epsilon_t < \theta_i(t)$: $\phi_i(t) = -1$ (sell);
- limit orders are canceled if not executed: $\phi_i(t) = 0$ (canceled).

So:

$$\phi_i(t) = 1_{\epsilon_t > \theta_i(t)} - 1_{\epsilon_t < -\theta_i(t)}. \tag{7}$$

In the model, orders placements and execution are very sensitive to the limit price, but not to the volume of the order. By construction, the time period is the typical cancellation time of limit orders' strategies responsible for the market activity at the studied time scale. One should be aware that the frequencies of execution and cancellation are comparable. The capabilities of the model demonstrated throughout the paper show that the typical cancellation time can be a relevant choice for determining the actors of the returns dynamics at a characteristic timescale. This quest and focus on the actors of the dynamics at a given time scale is indeed the strength and utility of the statistical approach when attacking problems with high number of interacting entities.

2.2 Price Response to Aggregate Demand

I now describe the dynamics of the model ruled by the evolution of price and updating of strategies. The price is adjusted by the excess demand $Z_t = \sum \phi_i(t)$, which corresponds to the orders likely to be executed, through a price impact function g which depends on the total number of traded shares n:

$$r_t = \ln\left(\frac{p_{t+1}}{p_t}\right) = g(Z_t). \tag{8}$$

The empirical behavior [6] of this function indicates increased linearity and decreased slope while increasing n. I focus on the linear case defined as: $r_t = \frac{Z_t}{\lambda n}$ where $ng'(0) = \lambda^{-1}$. λ represents market depth at the studied time scale, the typical order imbalance needed to move the price by one point, normalized by total

number of traded shares n and in this framework characterizes the order book depth. A market with increased market depth is more liquid. By changing this function, one can explore the impact of mechanical execution on market learning dynamics.

2.3 Learning Strategies for the Fitness Thresholds

Initially, each liquidity provider agent has a trading rule given by the choice of a fitness threshold $\theta_i(0)$. At each time step, if the change in price is greater than the parameter δ, each agent i consider the price movement as significant and update with probability s according to the following defined rule: $\theta_i(t+1) = |r_t|$.

As a consequence of this rule, a fraction $s \in [0, 1]$ of these agents updates their strategies/thresholds using recent information. Because the limit order positions are directly related to the price changes, this updating rule implies that the limit orders are placed outside a spread 2δ. I introduced here the parameter δ which controls the spread, the difference between the best bid and best ask, and often fix it to 2δ. The reason of this choice for modeling the creation of a spread in the book is that agents' threshold are indeed an estimate of risk and it is thus wise to bound it at a non-zero minimal value.

Introducing IID random variables $u_i(t), i = 1\ldots n, t \geq 0$ uniformly distributed on [0, 1], which indicate whether agent i updates his threshold or not, we can write the learning rule as

$$\theta_i(t) = 1_{|r_{t-1}|>\delta}[1_{u_i(t)<s}|r_{t-1}| + 1_{u_i(t)\geq s}\theta_i(t-1)] + 1_{|r_{t-1}|\leq\delta}\theta_i(t-1). \quad (9)$$

Here ϵ_t represents randomness due to public news arrivals whereas the random variables $u_i(t)$ represent idiosyncratic sources of randomness. This way of updating can be seen as a stylized version of various estimators of volatility based on moving averages or squared returns. As a consequence of this mapping, a feedback loop is created between the volatility and the orders placement in the book.

The updating rule allows to differentiate between indistinguishable rational players. Indeed, given this probabilistic updating model, even if we start from an initially homogeneous population $\theta_i(0) = \theta_0$, heterogeneity develops into the population through the learning process which corresponds to changes in the strategies motivated by trading costs reduction. In this sense, the heterogeneity of agents strategies is endogenous in this model and, as we will see below, evolves through high fluctuations.

2.4 Features of the Model

Let us recall the main ingredients of the model defined above. At each time period:

- Informed agents receive a common signal $\epsilon_t \sim N(0, D^2)$.
- Each liquidity provider agent i uses a threshold $\theta_i(t)$ to set his bid and ask according to (1).

- If $|\epsilon_t| > \theta_i(t)$ informed agent considers the signal as significant and are likely to generate a market order $\phi_i(t)$ according to (2).
- The market price is affected by the excess demand and moves according to (3).
- Each liquidity provider agent adjusts, with probability s, her threshold according to (4).

With regard to some of the agent-based models considered in the literature, some important aspects are the following:

- Prices move through market fluctuations of supply and demand. Players can be "fundamentalist" and "chartist" traders. Price formation results from direct interaction of agents in the book.
- Information asymmetry: the model is based on the dynamics between informed traders and liquidity providers. Liquidity providers use market neutral strategies and differ in the way they process the information.
- Liquidity providers are localized in the book.
- Endogenous heterogeneity: heterogeneity of agents behavioral rules appears endogenously due to their learning strategies. There is also a "structural" heterogeneity in the model between informed traders placing market orders and liquidity providers placing limit orders. This structural diversity is also compatible with a distinction between fundamentalists and chartists, between speculators with directional strategies and hedgers with market-neutral strategies [4]. The dynamics of this heterogeneity is responsible for the shape of the book.
- The spread is the consequence of the decision making under uncertainty. The order placement strategies are indeed based on an estimate of risk which is wise to set strictly greater than zero.

The model has tractable parameters: s describes the average updating frequency, D is the standard deviation of the news arrival process and λ is the market depth. Furthermore, as we will observe in the next section, if we require to interpret specifically the trading periods this will put a further restriction on the parameters, reducing the effective number of parameters. Nevertheless, the clear structure of the model generates time series of returns with interesting complex dynamics and with properties similar to those observed empirically.

3 Dynamics and Structure of the Order Book

The model is straightforward to simulate and the results can be reproduced in a robust way [5].

3.1 Occupation Number Dynamics

As defined in Sect. 4, the order book is composed of limit orders placed around the price at distances defined as thresholds which belong to the set $\{\frac{k}{n\lambda}, k = 0, \ldots, n\}$ because of the updating mechanism that maps the strategies to estimates of the

volatility. The order book is symmetric by construction, however additional directional traders can be added to the model to take into account the temporary imbalances observed in the book. I focus the study now on the dynamics of the number of limit orders occupying the places at distance $\frac{k}{n\lambda}$ of the price, for each $k = 0, \ldots, n$.

Because only when the absolute returns are greater than δ, after an initialization period, orders are placed at distances greater than δ, and 2δ represents the spread. Fig. 1 shows the dynamics of the number of orders occupying a distance in the book. Less orders are placed deeper in the book. However the dynamics is similar for all the occupied distances. The occupation numbers decays exponentially in time and increases through upward "jumps": this behavior is actually similar to that of a class of stochastic volatility models, used to describe various econometric properties of returns. The upward jumps are understood as a fraction of agents updating their thresholds to this distance, and the exponential decay is understood as the fact that

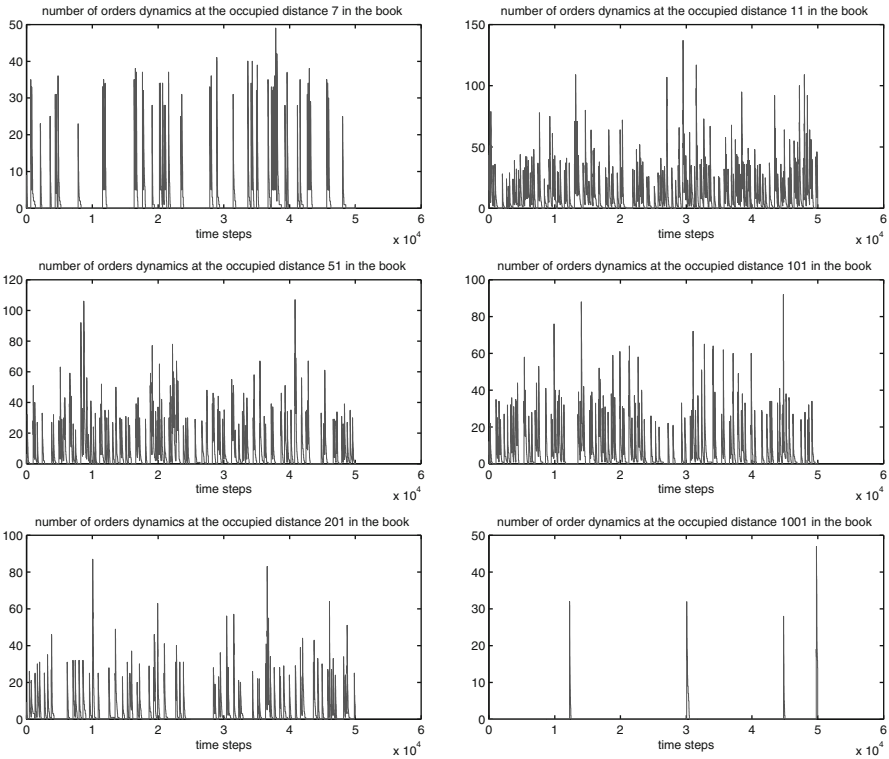


Fig. 1 Occupation number dynamics generated numerically by the agent-based market model for $n = 2000$, $D = 0.001$, $\lambda = 10$, $s = 0.015$ and $\delta = \frac{5}{n\lambda}$. The x-axis represents the time periods, and the y-axis the number of orders located at the same distance on one side of the book from the set $\{\frac{k}{n\lambda}, k = 0, \ldots, n\}$. The pictures shows how the various levels of depth in the book are occupied during a simulated run. Less orders are placed deeper in the book. The occupation numbers evolve through upward jumps and exponential decays

a fraction s of these agents adjust their thresholds to other values in the following time periods.

3.2 Profile of the Book

One note in the previous paragraph, the high fluctuations in the dynamics of occupation numbers. The question now is if there is a stable shape that characterizes the order book. To answer this question, I compute the average shape of the book over periods of increasing lengths and show the results in Fig. 2 for the following parameter values $n = 2000$, $s = 0.015$, $\lambda = 10$, $D = 0.001$. The threshold distribution reaches a stationary state and one get a stable form for the average shape of the book, however this is obtained only after averaging over a large period. The stable shape

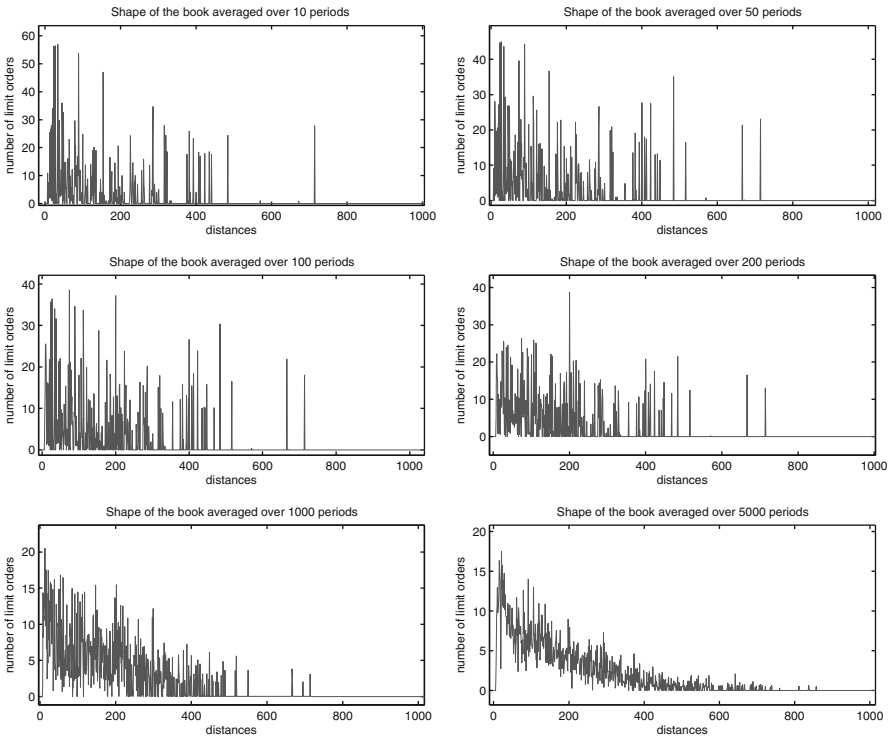


Fig. 2 Shapes of the book generated numerically by the agent-based market model for $n = 2000$, $D = 0.001$, $\lambda = 10$, $s = 0.015$ and $\delta = \frac{5}{n\lambda}$. The x-axis represents the set $\{k = 0, \ldots, n\}$ of possible occupied distances $\frac{k}{n\lambda}$ on each side of the symmetric book. The y-axis represents the number of orders located at the same distance in the book. The pictures show the results for various periods of averaging. One note the high fluctuations of the state of the book before reaching the stable shape

of the book that emerges from the simulations is a slow, continuous, symmetric decay of limit order volume as a function of the distance from the spread, as shown in Plot 1 of Fig. 3 with a logarithmic scale. This result demonstrates that the current agent-based framework is capable of replicating reasonably well the hump-shape of the book as reported in previous empirical studies [7] and offers a robust alternative explanation compared to recent works on this subject [8]. However, this pattern obtained on an average over a large period is not typical of the state of the book at any given time and this is also true with the observed dynamics in order book data. The shape of the book varies strongly from a time period to the other one, and is always very different from the average pattern. One can distinguish three type of orders: (i) a first block of orders around the spread consisting of a substantial fraction of the volume in the book, (ii) a second long area moderately occupied in orders, (iii) and finally some largely occupied distances deep in the book.

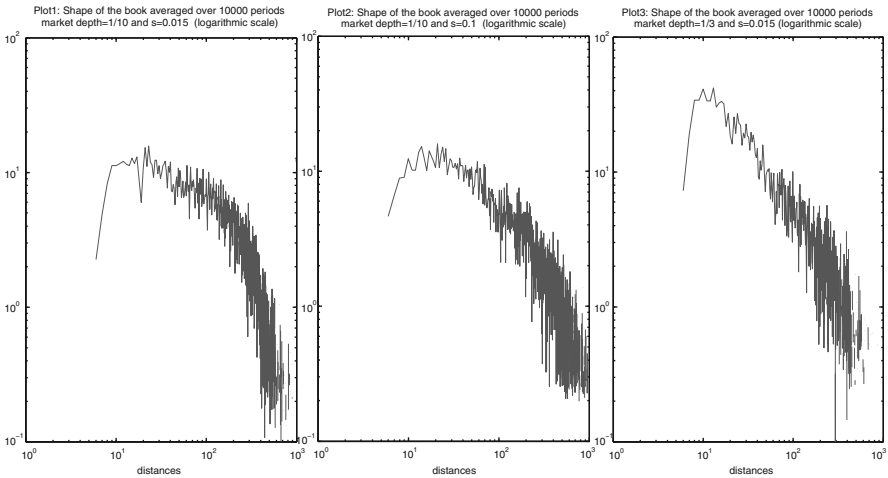


Fig. 3 *Plot 1*: Shape of the book, with logarithmic scale, generated numerically by the agent-based market model for $n = 2000$, $D = 0.001$, $\lambda = 10$, $s = 0.015$ and $\delta = \frac{5}{n\lambda}$. The hump-shape looking of the book is consistent with the empirical profile. *Plot 2*: Shape of the book, with logarithmic scale, generated numerically by the agent-based market model for $n = 2000$, $D = 0.001$, $\lambda = 10$, $s = 0.1$ and $\delta = \frac{5}{n\lambda}$. Robustness check for the profile of the book towards the updating frequency. The orders tend to be more concentrated near the spread. *Plot 3*: Shape of the book, with logarithmic scale, generated numerically by the agent-based market model for $n = 2000$, $D = 0.001$, $\lambda = 3$, $s = 0.015$ and $\delta = \frac{5}{n\lambda}$. Robustness check for the profile of the book towards the descriptive time scale. This plot also shows that when decreasing liquidity, orders tend to be more concentrated near the spread and the maximum of the shape of the book is sharper and closer to the spread

3.3 Master Equation

I now derive analytical results by studying the evolution of the state of the book. The state of the book is defined by the threshold distributions or limit orders' location distribution. The evolution of the thresholds distribution is described with the following master equation:

$$f_{t+1}(\theta) = (1 - s1_{|r_t|>\delta})\ f_t(\theta) + s1_{|r_t|>\delta}\ \delta_{|r_t|,\theta}, \tag{10}$$

with $|r_t| = |g(sign(\epsilon_t)F_t(|\epsilon_t|))|$, F_t being the cumulative distribution of the thresholds.

The solution of Eq. 10 can be derived analytically and reads as

$$f_t(\theta) = f_0 \prod_{j=0}^{t-1}(1 - s1_{|r_j|>\delta}) + s\sum_{j=1}^{t}(1-s)^{j-1}1_{|r_{t-j}|>\delta}\delta_{|r_{t-j}|,\theta}. \tag{11}$$

After an initial phase, the state of the book is independent of the initial distribution and is described as a sum integrating weighted elements of a function of the price path, expressed with the following equation:

$$f_t(\theta) = s\sum_{j=1}^{t}(1-s)^{j-1}1_{|r_{t-j}|>\delta}\delta_{|r_{t-j}|,\theta}. \tag{12}$$

The state of the book can thus be understood as a filter of the series $(1_{|r_t|>\delta}\delta_{|r_t|,\theta})_t$ obtained with the convolution with the series $(s(1-s)^{t-1})_t$.

The present multi-agent study acts as a double transform because it provides a double, spatial and temporal, multi-scale resolution. Indeed, Gaussian news are transformed into a price series at a specifically tuned time resolution and the state of the book is the result of a finely tractable transformation of past returns.

3.4 Robustness of the Results

The results obtained by varying the parameters of the limit order book model confirm the robustness of the results discussed in the previous sections.

Plot 2 in Fig. 3 is a picture of the shape of the book varying $1/s$ with $n = 2000$, $s = 0.1$, $\lambda = 10$, $D = 0.001$ and demonstrates the fact that the stable average shape of the book is independent of the updating time scale. This is consistent with the fact that the stationary state reached by the thresholds distribution must be independent of the updating time scale. One note that varying $1/s$ does affect the occupation number dynamics which is consistent with the high fluctuations before reaching the stationary state.

Plot 3 in Fig. 3 is another picture of the average shape of the book obtained with $n = 2000$, $s = 0.015$, $\lambda = 3$, $D = 0.001$, parameters that are consistent with intraday dynamics. The stable shape of the book that emerges from the simulations

is again a slow, continuous, symmetric decay of limit order volume as a function of the distance from the spread. This illustrates the robustness of this pattern towards the descriptive time scale of the simulated financial dynamics.

One finally note that the maximum of the shape of the book tends to get sharper and closer to the spread when liquidity or heterogeneity is decreased. Orders tend to be nearer to the spread and more concentrated in these cases.

4 Economic Implications

4.1 Robustness Claim

Due to the comprehensive structure of the model, the results obtained through computer controlled experiments can be explained and traced back to agents' adaptive behavior and analytical results are also produced. The properties are the consequence of agents' learning strategies rather than the consequence of structural effects detected for example with zero-intelligence models [9]. The robustness claim holds when talking about the parameters which can take a wide range of values, however the dynamics can change dramatically by increasing the complexity of the model and by adding more strategies. It is indeed an interesting question to explore the effects of more heterogeneity in the model, however this has to be consistent with the adopted statistical mean fields approaches to model in a sophisticated way sensible and rational players. The dynamics change radically as well when taking other distributions for the information flow. In particular, with a power law distribution, the market fluctuations are more determined by the excessive fluctuations of the information flow. Large fluctuations are indeed more the results of large fluctuations in the information flow rather than due to the values of the strategies and the impact of learning is reduced expressed by the absence of volatility clustering. One therefore questions if it is relevant to consider other than Gaussian distributions for mean field approaches to modeling the information flow that is not yet "processed" in the markets. Regarding the impact function, one can also observe the reduction of learning impact and even trimodal distribution of asset returns in some limit cases when considering concave impact functions as suggested by empirical studies. However, the descriptive time scale in the model is longer than the trade duration of market orders filling the excess demand and this corresponds to large aggregation times for computing empirically the market impact function. Empirically, market impact tends indeed to be linear for long aggregation times. The model is thus compatible with the empirically computed concave price impact function. It seems finer to consider the linear case at the descriptive time scale, for market influence control.

4.2 Market Impact and Excessive Volatility

I now discuss specifically the role of market impact, its plausible origins and how it is related to orders splitting. One can see in the discrete time model that at each

time step, if the amplitude of the market order flow was directly executed with a single market order, the dynamics of the price would match the information dynamics, which means that the distribution of returns would be Gaussian. We deduce from this that additional orders are necessary to obtain the observed dynamics. Moreover, one can expect indeed that this market order flow comes in a fragmented way, simply because it is generated by many market players whose orders are executed sequentially. One can see here as well the importance of latency for these market participants and understand the technological race to gain a competitive edge. Now, by construction, the time interval corresponds in the model to the cancellation time of the limit orders placed by the agents who are the main contributors for the excess demand at the descriptive time scale, and market impact is linearly modeled as resulting from the mesoscopic smoothing of market frictions effects. What could be the origins of market impact is the placement during this time interval of market orders hitting the other side or because the market order flow amplitude is fixed at the beginning of the interval, the execution process involve other limit orders that are placed during the time interval. Orders splitting, a strategy to minimize market impact, seems to be directly linked in generating market impact. Indeed, these orders can trigger more market activities. In a sense, this view reconciles the idea that orders splitting [10] is responsible for market order flows correlations. However, in my opinion, its role is more linked to market impact and the longer term correlations are more the consequence of the updating of order book strategies and the effects of assets allocation dynamics [11]. Market impact on the other hand is the major factor in determining the excessive fluctuations. This is a strong result from an economic point of view as it confirms that microstructure effects of price formation have dramatic consequences for market dynamics. It thus opens interesting perspectives for policymaking in order to regulate excess volatility.

4.3 Market Impact and Market Efficiency

I summarize now some arguments to respond to the critique [12] on the linear modeling of price impact. No, market impact is not simple. However, it appeared that it is a good strategy to consider the linear case in order to asses its role and therefore boost the explanatory power of the approach in order to show its consequences on markets dynamics, including the nonlinear shape of the book. There are also less contradictions than it seems. It is necessary to model market frictions, otherwise the price dynamics would be the same as the information dynamics. The reported nonlinear relation and the temporary decay are compatible with the framework because we are considering the descriptive time scale large enough compared to the execution process to be able to linearize market impact. Correlation in orders signs could very likely be due to the fragmentation of orders in the execution process in order to fill the excess demand. The persistence observed when considering the tick by tick process could be compatible with exponential decays of the various execution processes described at the various descriptive time scales. Longer memo-

ries effects, especially in the amplitude of order flows could be due to other causes such as the updating of order book strategies. In the time interval corresponding to trade duration, market frictions can be due to new limit orders in the spread and also market orders hitting the other side until excess demand is filled, and that can both lead to market efficiency, absence of autocorrelations in the returns, and market impact. Orders splitting can lead to the observed nonlinear relations for market impact. A general theory on market impact should be cautious when describing the phenomenon at different time scales, because the more orders splitting is important, as it is perceived for small aggregation times, the orders placement provoking market frictions is more diffusive and strategically non directional. The volume being proportional to trade duration, that can explain for example the exponents in the nonlinear relations. However, if the fragmentation of orders splitting is less important as perceived for large aggregation times, then market impact can be mainly due to strategic contrarian trading directly proportional to the volume.

4.4 What Are the Effects of Algorithmic Trading?

Algorithmic trading has witnessed an important growth the last decade and the momentum is still strong. How the present study is compatible with algorithmic trading and what can it tell us on its effects? The implementation of these algorithms can aim for example orders splitting and in that case it affects market impact. Observations in the market have witnessed a tightened spread, therefore according to the agent-based model an increased market depth. Orders splitting seems to play its role of minimizing market impact. According to the model, this should provoke a decreased level of volatility and what is observed is that volatility behaviors remained similar and trading volume has increased. That means for a given descriptive time scale, that D_{eff} remained the same and D decreased, in other words that the information flow and thus market order flow is getting faster as one can expect from the technological development of the last decade in finance.

5 Conclusions

In the present work I have focused on the combined role of market impact and order book strategies in limit order book dynamics and structures with the introduced robust agent-based market model capable of reproducing the major empirical stylized facts observed in the returns of financial assets for all timescales. The high instabilities in the state of the book emerge endogenously as a result of trading strategies and not from the input signal that can be interpreted as the fundamentals of the asset. Because of market impact, these high fluctuations provoke excessive volatility. The framework puts back the central microstructure role of price formation in generating excessive volatility. The present studies bring thus many elements of answers to

the question of how markets process supply and demand, and on how to understand markets dynamics for fundamental purposes, reduce volatility for regulatory goals and manage it in practice. My agent-based works lead to a better appreciation of the relationships between market forces and the various time scales and open interesting perspectives to explore other markets features with a solid tested base that constitutes the present agent-based market model of learning.

Acknowledgements The author would like to thank Fabrizio Lillo, Anirban Chakraborti, Bikas K. Chakrabarti, Tiziana Di Matteo, Tomaso Aste and Frédéric Abergel for fruitful discussions and advices.

References

1. F. Ghoulmié. Statistical mechanics of adaptive agents in socio-economic systems. PhD thesis. Ecole Normale Supérieure. 2005
2. F. Ghoulmié, R. Cont, J.-P. Nadal, Heterogeneity and feedback in an agent-based market model, Journal of Physics C, **17**, p. S1259–S1268, 2005
3. F. Ghoulmié et al. Effects of diversification among assets in an agent-based market model, Proceeding to the 2007 SPIE conference on Complex Systems in Canberra
4. J.C. Hull, Options, futures, and other derivatives, 6th edition, Prentice Hall
5. F. Ghoulmié. Behind Order Book Dynamics: Advances Using The Global Approach, Wilmott Journal (Quantitative Finance) 1:5–6, 227–244, 2009
6. J.-P. Bouchaud, J.D. Farmer, F. Lillo, How markets slowly digest changes in supply and demand, 2008. arXiv:0809.0822v1
7. J.-P. Bouchaud, M. Mézard, M. Potters, Statistical properties of stock order books: empirical results and models, Quantitative Finance **2**, p. 251–256, 2002
8. C. Chiarellaa, G. Iori, J. Perell, The impact of heterogeneous trading rules on the limit order book and order flows, Journal of Economic Dynamics and Control Volume **33**, Issue 3, p. 525–537, March 2009
9. E. Smith, J.D. Farmer, L. Gillemot, S. Krishnamurty, Statistical theory of the continuous double auction, Quantitative Finance **3**, p. 481–514, 2003
10. Lillo et al., Statistical identification with hidden Markov models of large order splitting strategies in an equity market, 2010
11. F. Ghoulmié. Order Book Market Complexity, 2010
12. J.-P. Bouchaud, Price impact, 2009

Multi-Agent Order Book Simulation: Mono- and Multi-Asset High-Frequency Market Making Strategies

Laurent Foata, Michael Vidhamali and Frédéric Abergel

Abstract. We present some simulation results on various mono- and multi-asset market making strategies. Starting with a zero-intelligence market, we gradually enhance the model by taking into account such properties as the autocorrelation of trade signs, or the existence of informed traders. We then use Monte Carlo simulations to study the effects of those properties on some elementary market making strategies. Finally, we present some possible improvements of the strategies.

1 Introduction

In the context of electronic markets, with trade and price transparency, market making has evolved from being a highly privileged, volume-based, broker-like activity to one of the many strategies commonly used by most market members. High frequency traders can more or less be considered as market makers, since they provide liquidity on markets via limit orders. It is therefore natural, and quite important for the applications, to understand the type of strategies that market makers can use, and to assess their profitability. The aim of this paper is to provide some insight into this question, using simulation. In particular, we will address the following three questions: can a market maker make money in a zero-intelligence market? In a market with order flow autocorrelation? In a market with informed traders? The last two questions arise naturally when one want to reproduce some known features of real-life markets, whereas the answer to the first one provides an indirect way of validating or invalidating the use of zero-intelligence order book models in the context of trading strategies. As we will show in this paper, the use of more realistic as-

Laurent Foata, Michael Vidhamali, Frédéric Abergel
Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
e-mail: laurent.foata@student.ecp.fr; michael.vidhamali@student.ecp.fr; frederic.abergel@ecp.fr

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

sumptions will make the simplest market making strategies lose money. It will then become necessary to use more sophisticated ones involving both market making and directional strategies in order to recover some of the profitability.

2 Market Models

2.1 Zero-intelligence Market

Our aim is to accurately reproduce the structure and the mechanisms of a typical order book in financial markets. In the zero-intelligence setting as introduced by J. Doyne Farmer [4], a market simply comprises liquidity providers and takers which send their orders independently from one another. The structure of an order book is recreated through interactions between agents. For numerical purposes, we have used the basic algorithm defined by Preis [8].

The liquidity provider only places limit orders which fill the order book. It submits orders at a random time, and with a random volume, all following exponential distributions. The prices are given by

$$p_{\text{buy}} = p_a - 1 - \eta \tag{1}$$

$$p_{\text{sell}} = p_b + 1 + \eta \tag{2}$$

where p_a is the best ask price, p_b the best bid price, η is an exponentially distributed random variable.

The order type – sell limit, buy limit, cancel buy or cancel sell – is chosen according to a 45/45/5/5 uniform probability. At a given price, orders are queued and are processed according to a first in, first out system.

The liquidity taker places market orders with timing and volumes also following exponential laws; though it is constrained to buy or sell at the best available ask or bid price. The order type – market buy or market sell – is chosen according to a 50/50 uniform probability, by default. We will use the notation $\lambda_{\text{buy}} = 1 - \lambda_{\text{sell}}$ as parameter of this Bernoulli process.

This kind of market is called a "zero-intelligence market" [4], for agents do not take into account interactions with other agents and do not follow any strategy of their own. In other words, they do not know why they are trading, nor what the others are doing. Note that with such settings, a log-normally distributed order book is achieved, as found in real financial markets.

2.2 Market with Autocorrelated Order Flow

In the zero-intelligence case, the price process is unrealistic, for it fails to exhibit many properties found in real financial markets according to the stylized facts [1] as,

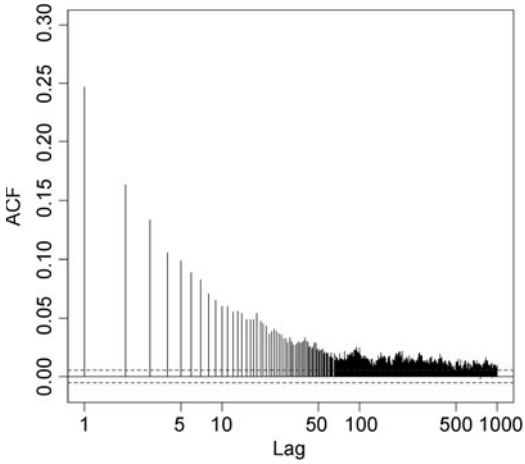


Fig. 1 ACF from Voss algorithm with $H = 0.7$

for instance, *volatility clustering*. We will then add in our model one of those properties: the autocorrelation of trade signs in $n^{-\alpha}$ [6, 7]. In reality, this phenomenon can be explained by the splitting of large orders by brokers seeking optimal execution.

The fractional Brownian motion exhibits long-range dependence and the signs of its increments are autocorrelated. We use this fact to correlate trade signs of the liquidity taker, i.e. λ_{buy}. We have chosen a quick algorithm called *Voss random addition algorithm* [11] to simulate a fBm, which gives satisfactory results (1). It consists in adding a white noise in the middle of a segment, and on the other, already computed points. The correlation of trade signs varies with the Hurst exponent.

2.3 Market with Informed Traders

In order to simulate the behavior of, for instance, fundamentalist traders who believe an asset should have a particular value, we implement an enhanced liquidity taker with a reference price [9]. The liquidity taker will seek to match this price, updating his buy frequency λ_{buy} according to the formula

$$x = \lambda_{\text{buy}}^{t} + \text{sgn}(\text{RP} - S_t)(e^{\epsilon|\text{RP}-S_t|} - 1) \tag{3}$$

$$\lambda_{\text{buy}}^{t+1} = x\mathbf{1}_{[\lambda_{\text{buy}}^{-},\lambda_{\text{buy}}^{+}]}(x) \tag{4}$$

where RP *is the reference price,* S_t *the current mid-price,* ϵ *an elasticity coefficient,* $\lambda_{\text{buy}}^{\pm}$ *the boundaries of* λ_{buy}.

ϵ and $\lambda_{\text{buy}}^{\pm}$ parameters determine the convergence speed toward the reference price. We will use $\epsilon = 0.01$ tick^{-1}. Also, we take values such as $\lambda_{\text{buy}}^{-} + \lambda_{\text{buy}}^{+} = 1$.

3 High Frequency Market Making Strategies

3.1 Default Strategy

Theoretically, a market maker places sell and buy limit orders in a symmetric fashion, hoping to earn (to make) the bid-ask spread by simultaneously selling and buying to two liquidity takers with opposite interests. In this model, the agent identified as market maker bears some similarities with the standard liquidity provider introduced in Sect. 2.1: random action times and orders volumes. But its strategy is more refined: it will submit two limits orders, and improve the current best limits by one tick whenever possible, in order to increase the likelihood that its orders be executed. Moreover, it will also follow some specific rules related to overnight positions and inventory risk. These rules are detailed below.

Position Closing
At the end of each trading day, the market maker closes its position in order to avoid overnight risk. Thus, it sends market orders at the end of the day like a liquidity taker. This effectively lowers its P&L. We analyze this in the simulations. Note that, in our simulations, the market is very liquid, so that the market maker has no trouble closing its positions.

Inventory Limits
At any given time during a trading day, the market maker may not hold more than a limit quantity of stock - this particular type of market risk is known as inventory risk. Since there is no overnight position, this restriction is imposed by fixing the maximum number of orders originated by the market maker at 5% of the average number of limit orders. The limits can be modified manually via the program's options.

Orders Cancellation
In the simplest case, the market maker only places orders at the best bid and ask. Market maker's orders which are not at the best limits will be cancelled when it is the acting agent. This rule will be adapted for other strategies.

3.2 Full Priority

We make the purely-theoretical, unrealistic assumption that the market maker can place limit orders which have priority amongst all the other orders at the same price. At a given price, the market makers orders will always be at the top of the queue and will be executed before the others. Actually, this hypothesis is common in the literature [12] and is equivalent to supposing that the market maker provides liquidity for a given percentage of any new incoming market order. This assumption is also implicit when replaying marginal strategies on data.

3.3 Spread-based Priority

We have coded a more realistic priority based on the spread value. When the spread is greater than or equal to 2 ticks, the agent improves the best limits and its orders are executed before those of any other liquidity provider. This priority mimics a technological advantage and can be replicated in practice.

3.4 Orders Placement Strategy

Instead of always improving the best limits, we study the effects of symmetrically submitting orders at various distances from the best limits. For instance, the market maker can chose to submit its orders only at the third limit in the order book. These orders will therefore be in a good position if and when the price reaches these limits.

3.5 Advanced Strategies

With a default liquidity taker, the price processes exhibit no significant trend. However, when there is a reference price different than the actual asset price, a trend emerges. In general, this implies that the market maker will lose money, due to *mean aversion*. Trying to hit the bid-ask spread the usual way will not work even with theoretical full-priority orders. As a result, the market maker has to design more sophisticated strategies.

3.5.1 Trend Follower

In order to detect a trend, the agent computes a coefficient $\theta \in [-1; 1]$ equal to the average sign of market orders along a sliding window of depth d. If $|\theta|$ is higher than a certain threshold t, then, in agreement with the positive (negative) trend, it places a sell (buy) limit order at a fixed, further limit f than usual, and a buy (sell) market order at the best limit. Otherwise, the usual market making strategy is applied. We will use $d = 10$, $t = 0.3$ and $f = 3$.

3.5.2 Mean Reversion

To develop more complex strategies, we add another asset in the modelled market which will be correlated with the asset #1. This is a realistic assumption, for we often find, in real markets, pairs of assets which are strongly correlated, either positively or negatively.

For each asset, the liquidity taker on one asset keeps track of the last orders of the liquidity taker on the other asset up to a certain depth p, and updates its probability parameter λ_{buy} with an exponential regression

$$x = 0.5 + 0.5 \sum_{i=0}^{p-1} c \exp(-\frac{i}{p})(\mathbf{1}_{\{\text{order}^{t-i}=\text{buy}\}} - \mathbf{1}_{\{\text{order}^{t-i}=\text{sell}\}}) \tag{5}$$

$$\lambda_{\text{buy}}^{t} = x\mathbf{1}_{[\lambda_{\text{buy}}^{-},\lambda_{\text{buy}}^{+}]}(x) \tag{6}$$

where p is the number of trades taken into account and $c = \frac{1-e^{1/p}}{1-e^{-1}}$ a normalization coefficient

Here, we will use $p = 50$. The interval $[\lambda_{\text{buy}}^{-}, \lambda_{\text{buy}}^{+}]$ can be chosen in order to stick to a precise correlation factor between assets prices. In our model, when the reference price option is enabled, asset #1 will have a reference price, and asset #2 will follow the trend thanks to the cross-correlation between assets.

With this simple algorithm, we have managed to reproduce a particular effect present in all high frequencies markets, called the *Epps effect*: time series analysis exhibit a dependence of stock returns correlations upon the sampling time scale Δt of data [5, 23] (2).

Note that this algorithm can be seen as a discrete time version of the model presented by Bacry and coauthors [13]. Given those two correlated assets, we then develop a simple strategy exploiting mean reversion. Using a Dickey-Fuller test on $y_t = \log(S_t^2) - \log(S_t^1)$, we find a p-value below 0.01, which means that the model produces mean reversion between the two assets [2].

We build as indicators Bollinger bands of the price ratio $\frac{S_1}{S_2} \approx \log(\frac{S_1}{S_2}) - 1$: the moving average MA, and the upper and lower bands ($MA \pm 2\sigma$). Whenever the ratio

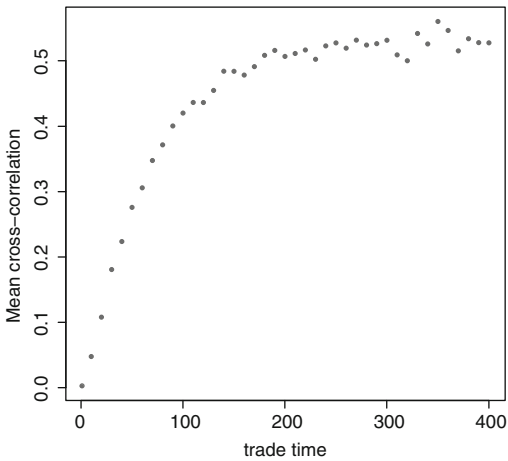


Fig. 2 Mean cross-correlation for different sampling time scales Δt (trade time) with $\lambda_{\text{buy}}^{+} = 0.7$

goes higher (resp. lower) than the upper (resp. lower) band, then the market maker opens a position exploiting mean reversion. Otherwise, the market maker follows its default strategy.

Several approaches were possible to take a position exploiting mean reversion. At first, we tried only with limit orders. For example, let us assume that the price ratio is over $(MA + 2\sigma)$. Then, we place a limit buy on order book #1 at a further limit than usual and also and a limit sell at the best ask. We do the opposite on order book #2: limit sell at a further limit, and limit buy at the best bid.

We can also use mixed strategies with market orders in order to control inventory risk. With the same assumptions as before, we can place a limit buy and a market sell on order book #1 and a limit sell and a market buy on order book #2. Note that this last strategy is not a pure market-making strategy anymore, but rather a market-making *with trend inclusion* strategy.

4 Analysis

Once our strategies are implemented, we test them using Monte Carlo simulations. Each simulation corresponds to a trading day of 6 hours 30 minutes (NYSE). The most interesting outputs are the agents' P&L and P&L per trade at the end of the trading day, the traded volumes and the price volatility.

Agents' laws of probability are fixed for simulations as shown in Table 1.

4.1 Zero-intelligence Market

We start by analyzing simple market making strategies on zero-intelligence market. Results of simulations are given in Table 2, and on Figs. 3 and 4. Boxplots indicate

Table 1 Agents' probability distributions

	Liquidity Provider	**Liquidity Taker**	**Market Maker**
Buy frequency (uniform)	$\lambda_{\text{buy}} = 0.5$	$\lambda_{\text{buy}} = 0.5$	*determinist*
Cancellation (uniform)	$\delta = 0.1$	*none*	*determinist*
Action time (exponential)	$\alpha = 12$	$\alpha = 10$	$\alpha = 10$
Orders price (exponential)	$\beta = 7$	*best limits*	*depends on strategy*
Orders volume (exponential)	$\mu = 120$	$\mu = 80$	$\mu = 10$

minimum, 1st quartile (0.25), median, 3rd quartile (0.75) and maximum of distributions. P&L is always given in hundreds of ticks.

Adopting the default market making strategy on a zero-intelligence market leads to positive gains on average: the P&L distribution is Gaussian, centered on 116.7. Spread-based priority increases P&L because the volume traded per day increases, even though P&L per trade is somewhat lower. With full priority, P&L increases because of a rise in P&L/trade, the traded volume being constant.

Table 2 Basic market making strategies on zero-intelligence market

Strategy	Output	**Min.**	1st **Qu.**	**Median**	**Mean**	3rd **Qu.**	**Max.**	**SD**
Default	P&L	−216.3	72.1	120.0	116.7	163.4	468.1	71.42
	P&L/trade	−0.006873	0.002186	0.003644	0.003567	0.00494	0.01486	0.002196
	Volume				31550			6074
	Volatility				0.3914			
Spread priority	P&L	−145.5	113.5	148.5	147.6	184.8	371.5	57.52
	P&L/trade	−0.00353	0.002694	0.003575	0.003554	0.004426	0.008791	0.001394
	Volume				41270			3696
	Volatility				0.3830			
Full priority	P&L	−138.9	143.1	185	180.9	218.7	445.2	61.04
	P&L/trade	−0.004012	0.004073	0.005288	0.005163	0.006239	0.01299	0.001756
	Volume				34120			5023
	Volatility				0.3679			

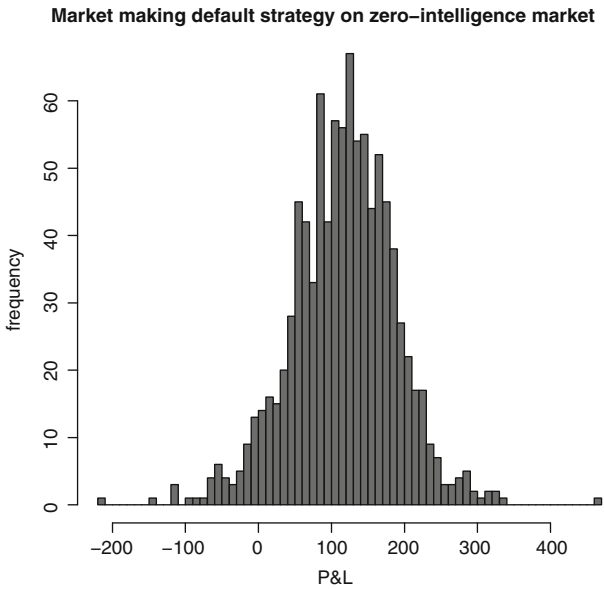


Fig. 3 Market maker with default strategy P&L distribution at end of the trading day

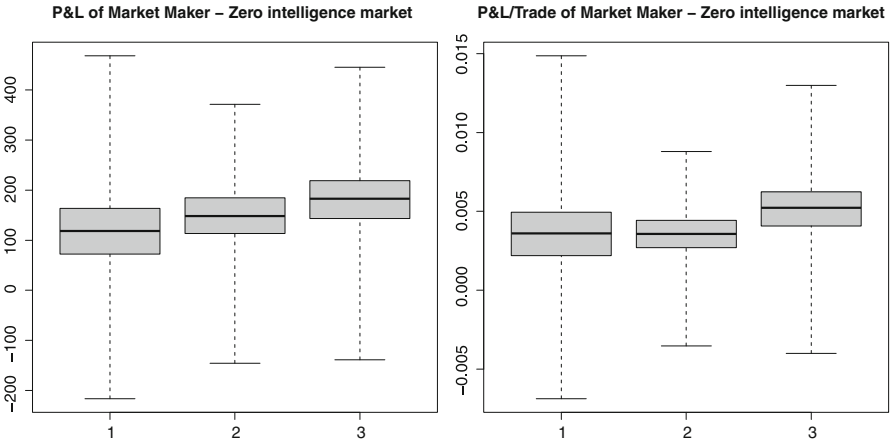


Fig. 4 Different market making strategies – default (1), spread-based priority (2) and full priority (3)

We then vary the placement of orders for the default market maker. P&L/trade increases as the limit orders are placed further and total P&L too, albeit not as dramatically (see Table 3, Fig. 5), because the market maker trades less.

This strategy slightly increases the P&L, compared with the default strategy and the spread-based priority. This rise may be explained by an implicit priority of market maker's orders which would be higher on the queue than with other strategies, for they are submitted long before the market price is actually close to the order price.

The P&L seems to reach a peak then decreases when the placement lag increases, because of the competition between the increasing P&L/trade and the lowering vol-

Table 3 Orders placement strategy on zero-intelligence market

Strategy	Output	**Min.**	1st **Qu.**	**Median**	**Mean**	3rd **Qu.**	**Max.**	**SD**
1st limit	P&L	−386.4	86.83	134.8	139.5	191.5	1306	93.5545
	P&L/trade	−0.009765	0.003176	0.004857	0.00498	0.006909	0.02021	0.003038
	Volume				27110			5023
2nd limit	P&L	−540.2	73.99	128.7	130.3	192.3	732.3	107.92
	P&L/trade	−0.02375	0.003505	0.006237	0.006173	0.009195	0.024020	0.004945
	Volume				21140			8370
3rd limit	P&L	−126.9	98.13	164.7	165.6	230.8	706	104.06
	P&L/trade	−0.008498	0.005720	0.009735	0.009559	0.01342	0.03391	0.0055784
	Volume				16790			7251
4th limit	P&L	−340	119.3	180	182.4	245.4	705.2	109.17
	P&L/trade	−0.017840	0.008364	0.01275	0.0127	0.01721	0.04933	0.0072725
	Volume				14130			6192

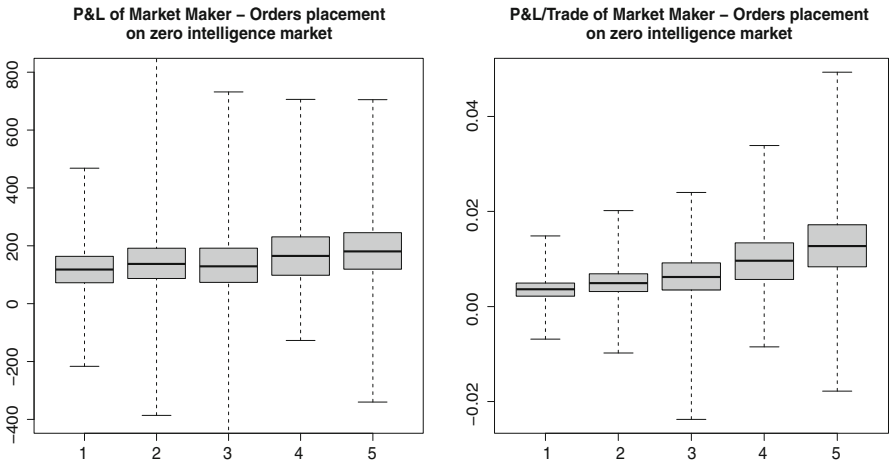


Fig. 5 Orders placement at 1st limit (2), at 2nd limit (3), at 3rd limit (4), at 4th limit (5), compared with the default strategy (1)

ume traded. When the P&L reaches its peak, we are close to that obtained with full priority.

4.2 Autocorrelated Market

The market making P&L decreases on autocorrelated markets ($H > 0.5$) and increases on anti-correlated markets ($H < 0.5$), as shown on Fig. 6. This was expected because, on positively correlated markets, the next market order type has a high probability to be the same as the previous one, thus, the market maker makes the spread less often.

We now apply previous strategies on an autocorrelated market with a Hurst exponent of $H = 0.7$. Regarding priority-based strategies, the spread-based one is less efficient than in a zero-intelligence market, because a high frequency strategy based on high volumes and low returns per trade cannot be successful due to the local correlation. The full priority strategy remains efficient because it leads to a higher P&L per trade.

We can also compare our results with the model developed by Bouchaud *et al.* [12], which is quite similar to the case of an autocorrelated market with full priority as said in Sect. 3.2. By taking this model and our own parameters, we find the following interval for the market maker P&L per trade: $[0.00429, 0.00621]$, with the lower (resp. upper) boundary corresponding to slow (resp. fast) market making. Therefore, the value obtained with our simulations (a mean P&L/trade of 0.004391 – see Fig. 6) seems to correspond to a rather slow market making in their model.

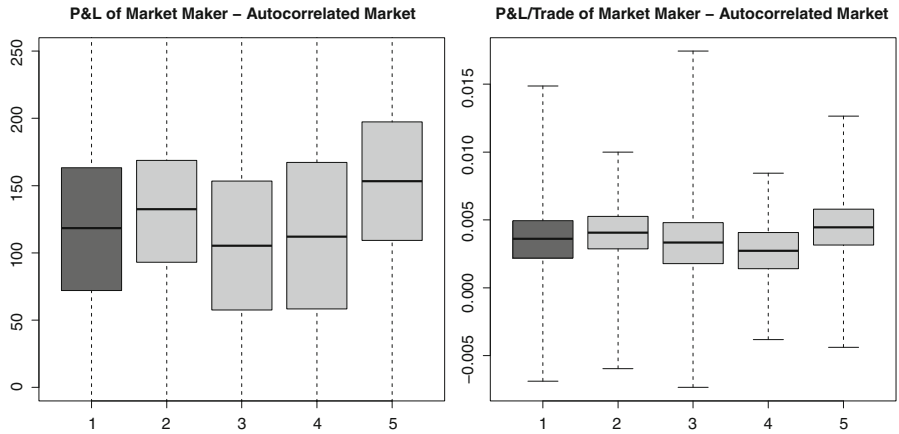


Fig. 6 Impact of autocorrelation – normal (1), $H = 0.3$ (2) and $H = 0.7$ (3). Priority strategies results with $H = 0.7$: spread-based (4) and full (5) priorities

Now, we study the impact of order placement. Results on Fig. 7 shows a better efficiency of this strategy compared with priority-based ones, on autocorrelated markets. Both P&L and P&L per trade are quite higher than with priority-based strategies.

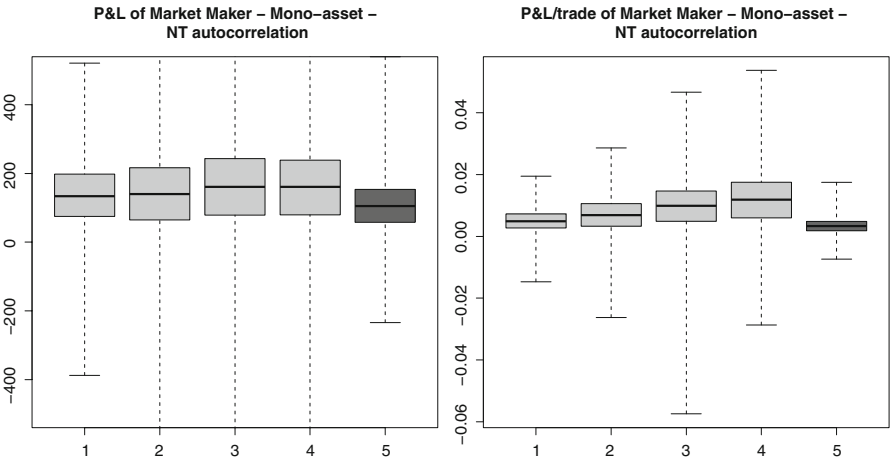


Fig. 7 Orders placement at 1st limit (1), at 2nd limit (2), at 3rd limit (3), at 4th limit (4), compared with the default strategy (5)

4.3 Market with Informed Traders

We now consider a market with informed traders who have a reference price for one asset. The market maker does not have any information on this price. We study the consequences of this information asymmetry on the previous strategies.

As shown on Fig. 8, the P&L/trade of the market maker is now negative. This fact is well known by traders on real financial markets who estimate that a basic market making strategy leads to losing roughly 1 tick per trade. Therefore, when competing against a market which has a clear trend, the market maker's performance drops as expected if it plays his default strategies.

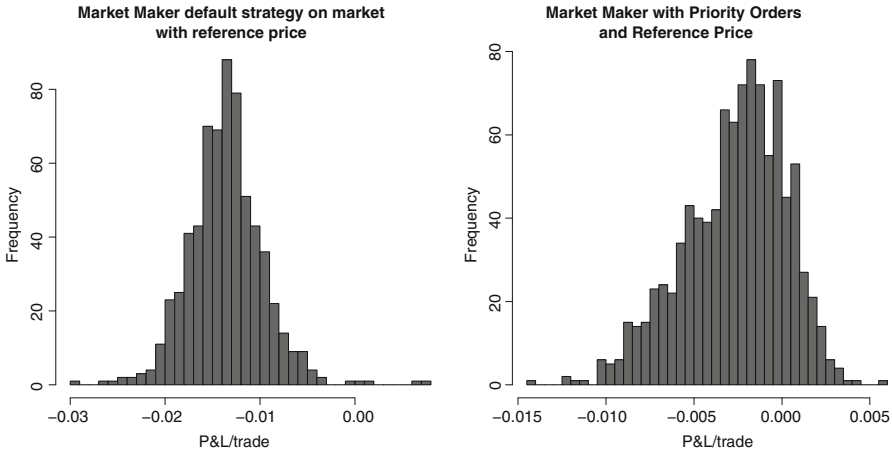


Fig. 8 P&L/trade of the market maker on a market with informed traders – default (1) and full priority (2) strategies

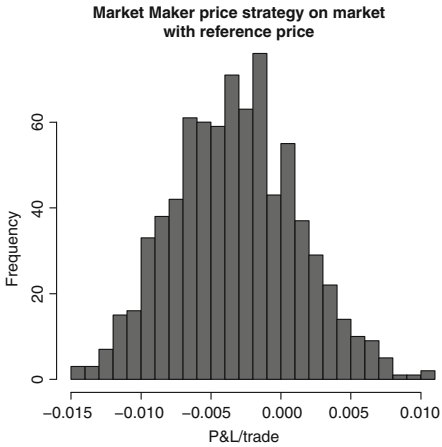


Fig. 9 Market maker with price guessing strategy

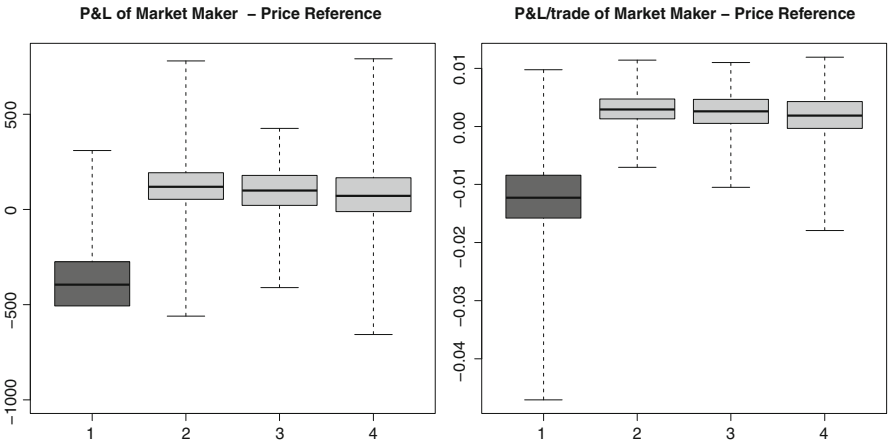


Fig. 10 Reference price 0.5% of the initial price – default strategy (1), mean reversion with limit orders (2) and with mixed orders (3). Reference price 1% of the initial price: mean reversion with limit orders (4)

We now study performances of our strategies developed against a reference price. The strategy with a rolling mean translates the P&L/trade distribution so that it is now centered around 0 ($mean = -0.004$), but the agent is still losing money on each trade as shown on Fig. 9.

The mean reversion strategy helps to recover positive gains (Fig. 10). The results are even better than with full priority orders, so it appears that one can exploit the assets correlation in order to build a good indicator. We see that P&L/trade is lower with the mixed strategy than with the strategy with only limit orders. Actually, with the former, the market maker reduces its inventory risks, but loses part of the bid-ask spread because of the market orders.

5 Conclusions and Current Research

Although basic market making strategies are efficient on zero-intelligence markets, when the market becomes complex, with autocorrelation of trade signs and reference price, those strategies fail and the market maker starts losing money. We have developed other strategies in order to cope with such a market. Guessing the trend with a rolling mean computation is not enough to recover positive gains. More interestingly, in the case of a market with two correlated assets, we have implemented a strategy based on mean reversion which gives satisfactory results even on markets with clear trends.

However, our enhanced zero-intelligence market is still too restrictive to reproduce efficiently real markets. New empirical observations verified on several markets (equities, futures on index, futures on bonds, ...), show evidence of an excitation of the limit order process by the market order process [10]: after a market order,

a limit order is likely to be submitted more quickly than it would have been without the market order. This has a direct impact on the bid-ask spread shape, and thus on the market maker profit.

In further study, this phenomenon could be added in the model and the modified competition between the market maker and the liquidity provider analyzed.

References

1. Chakraborti, A., Toke, I.M., Patriarca, M., Abergel, F.: Econophysics: Empirical facts and agent-based models (2009. http://arxiv.org/abs/0909.1974.). URL http://arxiv.org/abs/0909.1974
2. Chan, E.P.: Quantitative Trading: How to Build Your Own Algorithmic Trading Business (Wiley Trading) (2008)
3. Epps, T.W.: Comovements in stock prices in the very short-run. Journal of the American Statistical Association **74** (1979). URL http://www.jstor.org/pss/2286325
4. Farmer, J.D., Patelli, P., Zovko, I.I.: The predictive power of zero intelligence in financial markets. Quantitative Finance Papers. cond-mat/0309233, arXiv.org (2003). URL http://ideas.repec.org/p/arx/papers/cond-mat-0309233.html
5. Huth, N., Abergel, F.: The Times Change: Multivariate Subordination, Empirical Facts. SSRN eLibrary (2009. [to be published in Quantitative Finance Papers])
6. Lillo, F., Farmer, J.D.: The long memory of the efficient market. Studies in Nonlinear Dynamics & Econometrics **8**(3) (2004). URL http://ideas.repec.org/a/bpj/sndecm/v8y2004i3n1.html
7. Mike, S., Farmer, J.D.: An empirical behavioral model of liquidity and volatility. Quantitative finance papers, arXiv.org (2007). URL http://econpapers.repec.org/RePEc:arx:papers:0709.0159
8. Preis, T., Golke, S., Paul, W., Schneider, J.J.: Statistical analysis of financial returns for a multiagent order book model of asset trading. Phys Rev E Stat Nonlin Soft Matter Phys **76**(1 Pt 2), 016,108 (2007)
9. Qi, Z.: Simulation of multi-agents in financial market (2009). Done under the supervision of Abergel, Frederic
10. Toke, I.M.: "market making" behaviour in an order book model and its impact on the bid-ask spread. Quantitative finance papers, arXiv.org (2010). URL http://econpapers.repec.org/RePEc:arx:papers:1003.3796
11. Voss, R.F.: Fractals in nature: from characterization to simulation pp. 21–70 (1988)
12. Wyart, M., Bouchaud, J.P., Kockelkoren, J., Potters, M., Vettorazzo, M.: Relation between bid-ask spread, impact and volatility in order-driven markets. Quantitative Finance **8**(1), 41–57 (2008). URL http://ideas.repec.org/a/taf/quantf/v8y2008i1p41-57.html
13. Bacry, E.: Modeling microstructure noise using point processes, Econophys-Kolkata V Conference, submitted to Quantitative finance papers

Part II
High-Frequency Data and Modelling

The Nature of Price Returns During Periods of High Market Activity

Khalil Al Dayri, Emmanuel Bacry and Jean-François Muzy

Abstract. By studying all the trades and best bids/asks of ultra high frequency snapshots recorded from the order books of a basket of 10 futures assets, we bring qualitative empirical evidence that the impact of a single trade depends on the intertrade time lags. We find that when the trading rate becomes faster, the return variance per trade or the impact, as measured by the price variation in the direction of the trade, strongly increases. We provide evidence that these properties persist at coarser time scales. We also show that the spread value is an increasing function of the activity. This suggests that order books are more likely empty when the trading rate is high.

1 Introduction

During the past decade, the explosion of the amount of available data associated with electronic markets has permitted important progress in the description of price fluctuations at the microstructure level. In particular the pioneering works of Farmer's group [8, 9, 11, 12] and Bouchaud *et al.* [4, 5, 7] relying on the analysis of order book data, has provided new insights in the understanding of the complex mechanism of price formation (see e.g [3] for a recent review). A central quantity in these works and in most approaches that aim at modeling prices at their microscopic level, is the market impact function that quantifies the average response of prices to "trades". Indeed, the price value of some asset is obtained from its cumulated variations caused by the (random) action of sell/buy market orders. In that respect, the price dynamics is formulated as a discrete "trading time" model like:

$$p_n = \sum_{i<n} G(n-i, V_i)\varepsilon_i + \textit{diffusion} \tag{1}$$

Khalil Al Dayri, Emmanuel Bacry
CMAP, École Polytechnique,
e-mail: khalil.aldayri@polytechnique.fr; emmanuel.bacry@polytechnique.fr

Jean-François Muzy
CNRS-UMR 6134, Université de Corse, Corte, France,
e-mail: muzy@univ-corse.fr

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

where n and i are transaction "times", i.e., integer indices of market orders. V_i is the quantity traded at index i, ε_i is the sign of the ith market order ($\varepsilon_i = -1$ if selling and $\varepsilon_i = +1$ if buying). The function $G(k, V)$ is the bare impact corresponding to the average impact after k trades of a single trade of volume V. Among all significant results obtained within such a description, one can cite the weak dependence of impact on the volume of market orders, i.e., $G(n, V) \sim G(n) \ln V$, the long-range correlated nature of the sign of the consecutive trades ε_i and the resulting non-permanent power-law decay of impact function $G(n)$ [3]. Beyond their ability to reproduce most high frequency stylized facts, models like (1) or their continuous counterparts [1] have proven to be extremely interesting because of their ability to control the market impact of a given high frequency strategy and to optimize its execution cost [10].

Another well known stylized fact that characterizes price fluctuations is the high intermittent nature of volatility. This feature manifests at all time scales, from intradaily scales where periods of intense variations are observed, for instance, around publications of important news to monthly scales [2]. Since early works of Mandelbrot and Taylor [13], the concept of subordination by a trading or transaction clock that maps the physical time to the number of trades (or the cumulated volume) has been widely used in empirical finance as a way to account for the volatility intermittency. The volatility fluctuations simply reflects the huge variations of the activity. The observed intradaily seasonal patterns [6] can be explained along the same line. Let us remark that according to the model (1), the physical time does not play any role in the way the market prices vary from trade to trade. This implies notably that the variance per trade (or per unit of volume traded) is constant and therefore that the volatility over a fixed physical time scale, is only dependent on the number of trades.

The goal of this paper is to critically examine this underlying assumption associated with the previously quoted approaches, namely the fact that the impact of a trade does not depend in any way on the physical time elapsed since previous transaction. Even if one knows that volatility is, to a good approximation, proportional to the number of trades within a given time period (see Sect. 3), we aim at checking to what extent this is true. For that purpose we use a database which includes all the trades and *level 1* (i.e., best ask and best bid) ultra high-frequency snapshots recorded from the order books of a basket of 10 futures assets. We study the statistics of return variations associated to one trade conditioned by the last inter-trade time. We find that the variance per trade (and the impact per trade) increases as the speed of trading increases and we provide plausible interpretations to that. We check that these features are also observed on the conditional spread and impact. Knowing that the spread is a proxy to the fullness of the book and the available liquidity [15], we suspect that in high activity periods the order books tend to deplete. These "liquidity crisis" states would be at the origin of considerable amounts of variance not accounted for by transaction time models.

The paper is structured as follows: in Sect. 2 we describe the futures data we used and introduce some useful notations. In Sect. 3, we study the variance of price increments and show that if it closely follows the trading activity, the variance per

trade over some fixed time interval is not constant and increases for strong activity periods. Single trade variance of midpoint prices conditioned to the last intertrade duration are studied in Sect. 4. We confirm previous observations made over a fixed time interval and show that, as market orders come faster, their impact is greater. We also show that, for large tick size assets, the variations of volatility for small intertrade times translates essentially on an increase of the probability for a trade to absorb only the first level of the book (best bid or best ask). There is hardly no perforation of the book on the deeper levels. In Sect. 5.1 we show that the single trade observations can be reproduced at coarser scales by studying the conditional variance and impact over 100 trades. We end the section by looking at the spread conditioned to the intertrade durations. This allows us to confirm that in period of high activity, the order book tends to empty itself and therefore the increase in the trading rate corresponds to a local liquidity crisis. Conclusions and prospects are provided in Sect. 6.

2 Data Description

In this paper, we study highly liquid futures data, over two years during the period ranging from 2008/08 till 2010/03. We use data of ten futures on different asset classes that trade on different exchanges. On the EUREX exchange (localized in Germany) we use the futures on the DAX index (DAX) and on the EURO STOXX 50 index (ESX), and three interest rates futures: 10-years Euro-Bund (Bund), 5-years Euro-Bobl (Bobl) and the 2-years Euro-Schatz (Schatz). On the CBOT exchange (localized in Chicago), we use the futures on the Dow Jones index (DJ) and the 5-Year U.S. Treasury Note Futures (BUS5). On the CME (also in Chicago), we use the forex EUR/USD futures (EURO) and the the futures on the SP500 index (SP). Finally we also use the Light Sweet Crude Oil Futures (CL) that trades on the NYMEX (localized in New-York). As for their asset classes, the DAX, ESX, DJ, and SP are equity futures, the Bobl, Schatz, Bund, and BUS5 are fixed income futures, the EURO is a foreign exchange futures and finally the CL is an energy futures.

The date range of the DAX, Bund and ESX spans the whole period from 2008/08 till 2010/03, whereas, for all the rest, only the period ranging from 2009/05 till 2010/03 was available to us. For each asset, every day, we only keep the most liquid maturity (i.e., the maturity which has the maximum number of trades) if it has more than 5000 trades, if it has less, we just do not consider that day for that asset. Moreover, for each asset, we restrict the intraday session to the most liquid hours, thus for instance, most of the time, we close the session at settlement time and open at the outcry hour (or what used to be the outcry when it no longer exists). We refer the reader to Table 1 for the total number of days considered for each asset (column D), the corresponding intraday session and the average number of trades per day. It is interesting to note that we have a dataset with a variable number of trading days (around 350 for the DAX, Bund and ESX, and 120 for the rest) and a variable

average number of orders per day, varying from 10 000 trades per day (Schatz) to 95 000 (SP). Our data consist of *level 1* data: every single market order is reported along with any change in the price or in the quantity at the best bid or the best ask price. All the associated timestamps are the timestamps published by the exchange (reported to the millisecond).

It is important to point out that, since when one market order hits several limit orders it results in several trades being reported, we chose to aggregate together all such transactions and consider them as one market order. We use the sum of the volumes as the volume of the aggregated transaction and as for the price we use the last traded price. In our writing we freely use the terms transaction or trade for any transaction (aggregated or not). We are going to use these transactions as our "events", meaning that all relevant values are calculated at the time of, or just before such a transaction. As such, we set the following notations:

Notations 1 *For every asset, let D be the total number of days of the considered period. We define:*

1. $N_k, k \in \{1 \dots D\}$ *the total number of trades on the* k^{th} *day*
2. t_i *is the time of the* i^{th} *trade* ($i \in [1, \sum_k N_k]$)
3. b_{t_i} *and* a_{t_i} *are respectively the best bid and ask prices right before the* i^{th} *trade*
4. $p_{t_i} = \frac{b_{t_i} + a_{t_i}}{2}$ *is midpoint price right before the* i^{th} *trade*
5. $s_{t_i} = a_{t_i} - b_{t_i}$ *is spread right before the* i^{th} *trade*
6. $r_{t_i} = p_{t_{i+1}} - p_{t_i}$ *is the return caused by the* i^{th} *trade, measured in ticks*
7. $NT[s,t] = \#\{t_i, s \leq t_i < t\}$ *corresponds to the number of trades in the time interval* $[s,t]$
8. $\mathrm{e}_t[...]$ *or* $\mathrm{e}_i[...]$ *indifferently refers to the historical average of the quantity in between backets, averaging on all the available days and on all the trading times* $t = t_i$. *The quantity is first summed up separately on each day (avoiding returns overlapping on 2 consecutive days), then the so-obtained results are summed up and finally divided by the total number of terms in the sum.*

Let us note that in the whole paper, we will consider that the averaged returns are always 0, thus we do not include any mean component in the computation of the variance of the returns.

"Perceived" Tick Size and Tick Value

The *tick value* is a standard characteristic of any asset and is measured in its currency. It is the smallest increment by which the price can move. In all the following, all the price variations will be normalized by the tick value to get them expressed in ticks (i.e., in integers for price variations and half-integers for midpoint-price

Table 1 Data Statistics. The assets are listed *from top to bottom* following the increasing order of the $P_=$ column (see (2)), i.e., from the smaller (*top*) to the greater (*bottom*) "perceived" tick size. D: number of days that are considered. The *Tick Value* is the smallest variation (expressed in the local currency) by which a trading price can move. The *Session* column indicates the considered trading hours (local time). The *# Trades/Day* is the average of the daily number of trades (i.e., $\sum_{k=1}^{D} N_k/D$ using Notations 1). P_0 and $P_=$ are defined in Eqs. (4) and (2) and reported here in percent

Futures	Exchange	Tick Value	D	Session	# Trades/Day	1/2-η	P_0	$P_=$
DAX	EUREX	12.5€	349	8:00-17:30	56065	0.082	49	67.9
CL	NYMEX	10$	127	8:00-13:30	76173	0.188	72.8	79.8
DJ	CBOT	5$	110	8:30-15:15	36981	0.227	72.6	92.2
BUS5	CBOT	7.8125$	126	7:20-14:00	22245	0.288	81.6	95.1
EURO	CME	12.5$	129	7:20-14:00	42271	0.252	79.5	95.2
Bund	EUREX	10€	330	8:00-17:15	30727	0.335	80.9	97.6
Bobl	EUREX	10€	175	8:00-17:15	14054	0.352	86.5	99.1
ESX	EUREX	10€	350	8:00-17:30	55083	0.392	88.3	99.2
Schatz	EUREX	5€	175	8:00-17:15	10521	0.385	89.3	99.4
SP	CME	12.5$	112	8:30-15:15	97727	0.464	96.6	99.8

variations). As one can see in Table 1, column *Tick Value*, our assets have very different tick values. It is important to note a counter-intuitive though very well known fact: the tick value *is not* a good measure of the *perceived size* (by pratitionners) of the tick. A trader considers that an asset has a small tick when he "feels" it to be negligible, consequently, he is not averse at all to price variations of the order of a single tick. For instance, every trader "considers" that the the ESX index futures has a much greater tick than the DAX index futures though the tick values are of the same orders! There have been several attempts to quantify the perceived tick size. Kockelkoren, Eisler and Bouchaud in [7], write that "large tick stocks are such that the bid-ask spread is almost always equal to one tick, while small tick stocks have spreads that are typically a few ticks". Following these lines, we calculate the number of times (observed at times t_i) the spread is equal to 1 tick:

$$P_= = \frac{\#\{i, s_{t_i} = 1\}}{N} \tag{2}$$

and show the results in Table 1. We classify our assets according to this criterion and find SP to have the largest tick, with the spread equal to 1 99.8% of the time, and the DAX to have the smallest tick.

In a more quantitative approach, in order to quantify the aversion to price changes, Rosenbaum and Robert in [14] give a proxy for the perceived tick size using last traded non null returns time-series. If N_t^a (resp. N_t^c) is the number of times a trading price makes two jumps in a row in the same (resp. different) directions, then the perceived tick size is given by $1/2 - \eta$ where η is defined by

$$\eta = \frac{N_t^c}{2N_t^a} \tag{3}$$

For each asset, we computed η for every single day, and average over all the days in our dataset and put the result in the $1/2 - \eta$ column in Table 1. We find that the rankings of the assets using this criterion almost matches the ranking using [7]'s $P_=$ criterion (two slight exceptions being the ESX/Schatz and BUS5/EURO ranking). One interpretation of the η based proxy is that if the tick size is large, market participant are more averse to changes in the midpoint price and market makers are happy to keep the spread collapsed to the minimum and the midpoint would only move when it becomes clear that the current price level is unsustainable. To check that, we calculate the number of times (observed at times t_i) the return (as defined in notation 1) is null:

$$P_0 = \frac{\#\{i, r_{t_i} = 0\}}{N} \tag{4}$$

and show the result in Table 1. Again, it approximately leads to the same ranking which has nothing to do with the ranking using Tick Values.

3 Realized Variance versus Number of Trades

It is widely known that, in a good approximation, the variance over some period of time is proportional to the number of trades during that time period (see e.g. [6]).

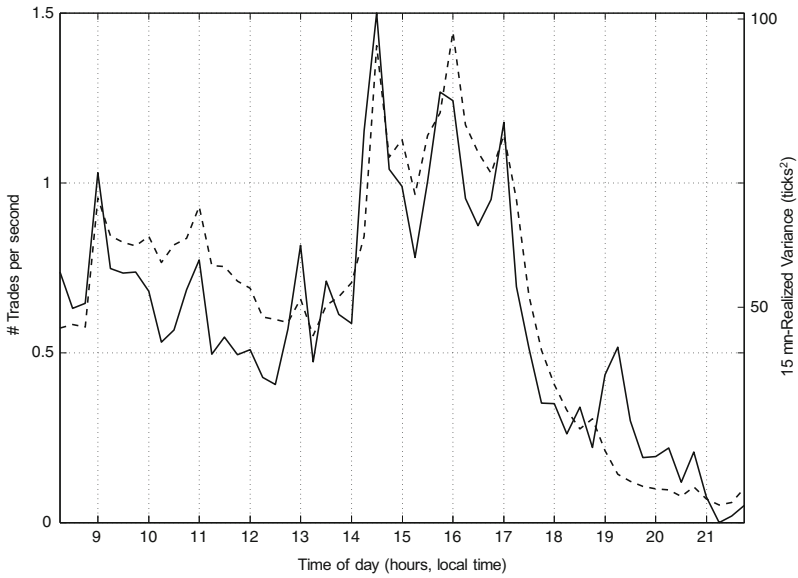


Fig. 1 Bund intraday seasonality of both trading rate and volatility (abscissa are expressed in hours, local time). Averages are taken on all available days. *Dashed line*: average intraday rate of trading (average number of trades per second) using 15mn bins. *Solid line*: average 15-minutes-realized variance (estimated summing on 15 1mn-squared returns)

Fig. 1 illustrates this property on Bund data. On 15 minutes intraday intervals, averaging on every single day available, we look at (dashed curve) the average intraday rate of trading (i.e., the average number of trades per second) and (solid curve) the average (15 minutes) realized variance (estimated summing on 1mn-squared returns $(p_{t+1mn} - p_t)^2$). We see that the so-called intraday seasonality of the variance is highly correlated with the intraday seasonality of the trading rate [6]. In order to have more insights, we look at some daily statistics: Fig. 2 shows a scatter plot in which each point corresponds to a given day k whose abscissa is the number of trades within this day, i.e., N_k, and the ordinate is the daily variance (estimated summing over 5-mn quadratic returns) of the same day k. It shows that, despite some dispersion, the points are mainly distributed around a mean linear trend confirming again the idea shown in Fig. 1 that, to a good approximation, the variance is proportional to the number of trades. In that respect, trading time models (Eq. (1)) should capture most of the return variance fluctuations through the dynamics of the transaction rate. However, in Fig. 2, the points with high abscissa values (i.e., days with a lot of activity) tend to be located above the linear line, whereas the ones with low abscissa (low activity) cluster below the linear line, suggesting that the variance per trade is dependant on the daily intensity of trading.

Before moving on, we need to define a few quantities. Let Δt be an intraday time scale and let N be a number of trades. We define $V(\Delta t, N)$ as the estimated price variance over the scale Δt conditioned by the fact that N trades occurred. Using notations, 1 (*7*) and (*8*), from a computational point of view, when $\Delta t = \Delta t_0$ is fixed and N is varying, $V(\Delta t = \Delta t_0, N)$ is estimated as:

$$V(\Delta t = \Delta t_0, N) = \mathrm{e}_t \left[(p_{t+\Delta t_0} - p_t)^2 | NT[t, t+\Delta t_0] \in [N - \delta_N, N + \delta_N] \right] \quad (5)$$

where δ_N is some bin size. And, along the same line, when we study $V(\Delta t, N)$ for a fixed $N = N_0$ value over a range of different values of Δt, one defines a temporal bin size $\delta_{\Delta t}$ and computes $V(\Delta t, N = N_0)$ as[1]

$$V(\Delta t, N = N_0) = \mathrm{e}_i \left[(p_{t_{i+N_0}} - p_{t_i})^2 | t_{i-1+N_0} - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \quad (6)$$

Let us note that, in both cases, the bins are chosen such that each bin involves approximately the same number of terms. We also define the corresponding conditional variance per trade as:

$$v(\Delta t, N) = \frac{V(\Delta t, N)}{N}. \quad (7)$$

In order to test the presence of an eventual non-linear behavior in the last scatter plots (Fig. 2), we show in Fig. 3 the 5-minutes variance per trade $v(\Delta t = 5mn, N)$ as a function of the average intertrade duration $\frac{5mn}{N}$ as N is varying. We clearly see that the estimated curve (solid line) is below the simple average variance (dashed line) for large intertrade durations and above the average variance when the trades

[1] Let us point out that we used the index $i - 1$ in the condition of (6) and not the index i since, for the particular case $N_0 = 1$ (extensively used in Sect. 4), we want to use a *causal* conditionning of the variance. For N_0 large enough, using one or the other does not really matter.

are less than 600 milliseconds apart. Note that we observed a similar behavior for most of the futures suggesting a universal behavior.

To say that the realized variance is proportional to the number of trades is clearly a very good approximation as long as the trading activity is not too high as shown both on a daily scale in Fig. 2 and on a 5mn-scale in Fig. 3. However, as soon as the trading activity is high (e.g., average intertrade duration larger than $600ms$ on a 5mn-scale), the linear relationship seems to be lost. In the next section we will focus on the impact associated with a single trade.

4 Single Trade Impact on the Midpoint Price

In this section, we will mainly focus on the impact of a trade i, and more specifically on the influence of its arrival time t_i on the return $r_{t_i} = p_{t_{i+1}} - p_{t_i}$. In order to do so, it is natural to consider the return r_{t_i} conditioned by $t_i - t_{i-1}$, the time elapsed since previous transaction. We want to be able to answer questions such as: how do

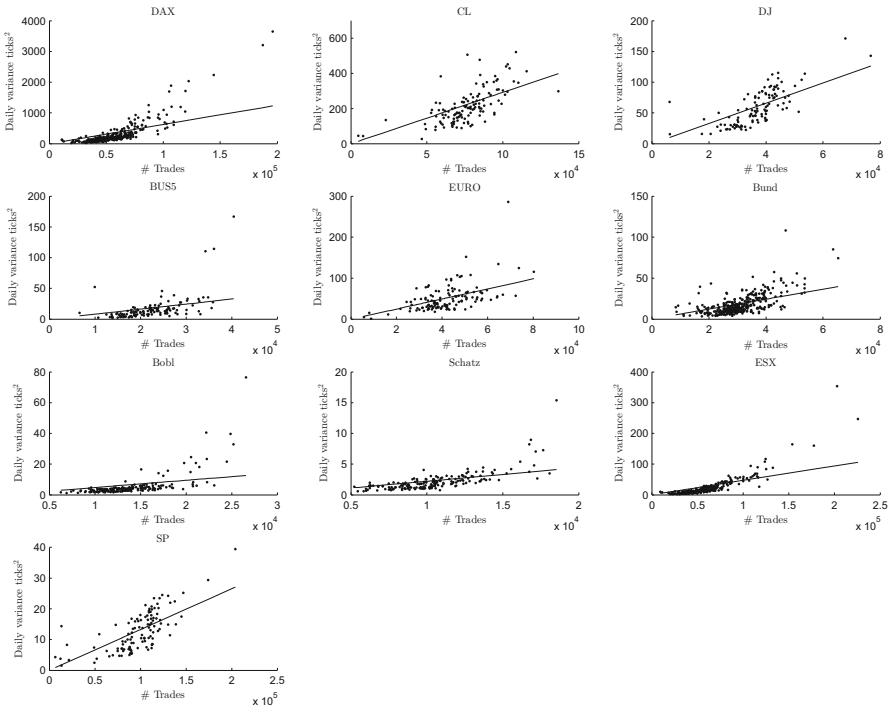


Fig. 2 For each asset (in increasing perceived tick size $P_=$): Daily variance (estimated summing over 5-mn quadratic returns) against daily number of trades. Each dot represents a single day. The *solid line* is the linear regression line with zero intercept. We see strong linearity between the variance and the number of trades but there seem to be clustering of *dots above* (resp. *below*) the *solid line* for days with high (resp. low) activity

compare the impacts of the ith trade depending on the fact that it arrived right after or long after the previous trade? Of course, in the framework of trading time models this question has a very simple answer: the impacts are the same! Let us first study the conditional variance of the returns.

4.1 Impact on the Return Variance

In order to test the last assertion, we are naturally lead to use Eqs (6) and (7) for $N_0 = 1$, i.e,

$$v(\Delta t, N = 1) = \mathrm{e}_i \left[r_{t_i}^2 \mid t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{8}$$

Let us illustrate our purpose on the DAX and the SP futures. They trade on two different exchanges, (EUREX and CME) and have very different daily statistics (e.g., DAX has the smallest perceived tick and SP the largest as one can see in

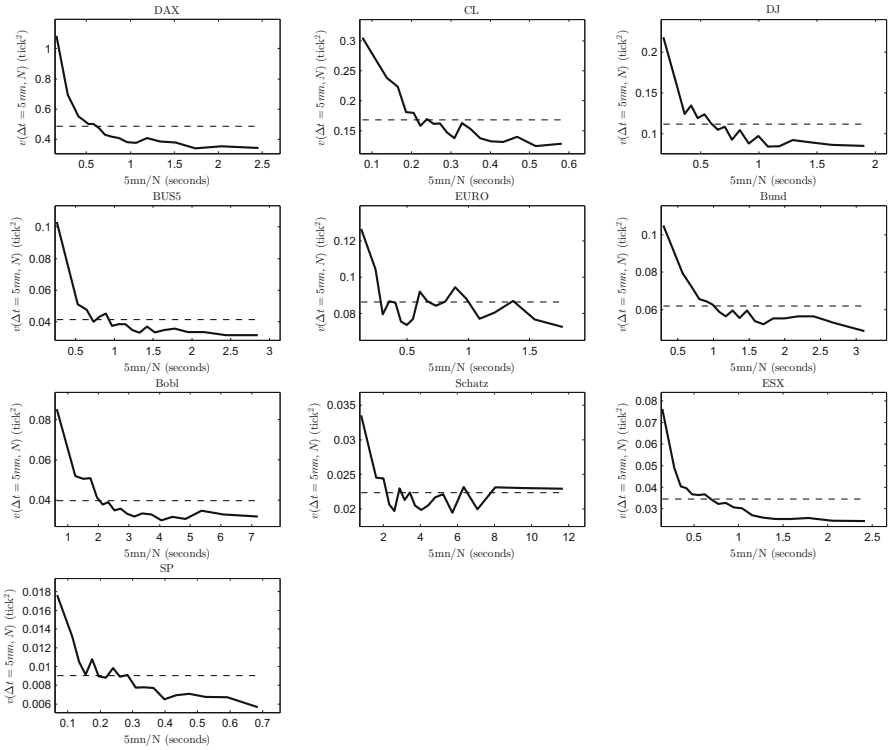


Fig. 3 For each asset (in increasing perceived tick size $P_=$), *Solid line*: conditional $v(\Delta t = 5mn, N)$ variance per trade (see (5) as a function of the average intertrade duration $\frac{5mn}{N}$ when varying N. *Dashed line*: unconditional 5mn-variance per trade. The *solid line* is almost constant for average times above 0.6 seconds, and it increases when the trading becomes faster

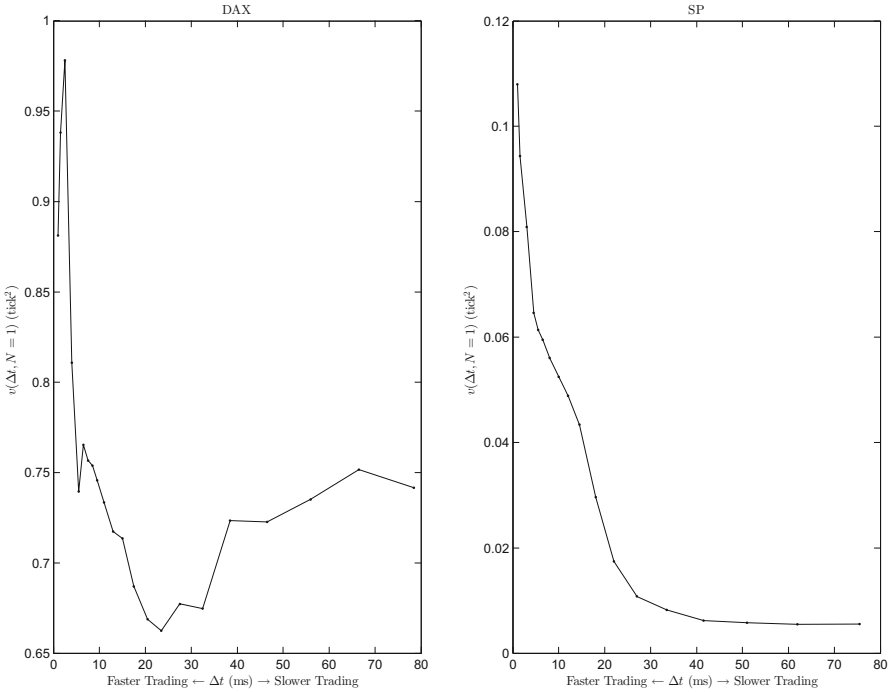


Fig. 4 $v(\Delta t, N = 1)$ as a function of Δt over very short Δt's for DAX and SP. The variance per trade increases dramatically below a certain Δt

Table 1). Fig. 4 shows for both assets $v(\Delta t, N = 1)$ (expressed in squared $tick$) as a function of Δt (in milliseconds). We notice that both curves present a peak for very small Δt and stabilize around an asymptotic constant value for larger Δt. This value is close to 0.7 ticks^2 for the DAX and to 0.005 ticks^2 for the SP. The peak reaches 0.95 (35% above the asymptote) for the DAX, and 0.1 (2000% above the asymptote).

Fig. 5 switches (for all assets) to a log scale in order to be able to look at a larger time range. A quick look at all the assets show that they present a very similar behavior. One sees in particular for the ESX curve that the variance increases almost linearly with the rate of trading, and then suffers an explosion as Δt becomes smaller than 20 ms. The "same" explosion can be qualitatively observed over all assets albeit detailed behavior and in particular the minimal threshold Δt may vary for different assets.

Let us note that the variance $v(\Delta t, N = 1)$ as defined by (8) can be written in the following way:

$$v(\Delta t, N = 1) = P(\Delta t)A(\Delta t), \tag{9}$$

where $P(\Delta t)$ is the probability for the return to be non zero conditioned by the intertrade duration $t_i - t_{i-1} = \Delta t$, i.e.,

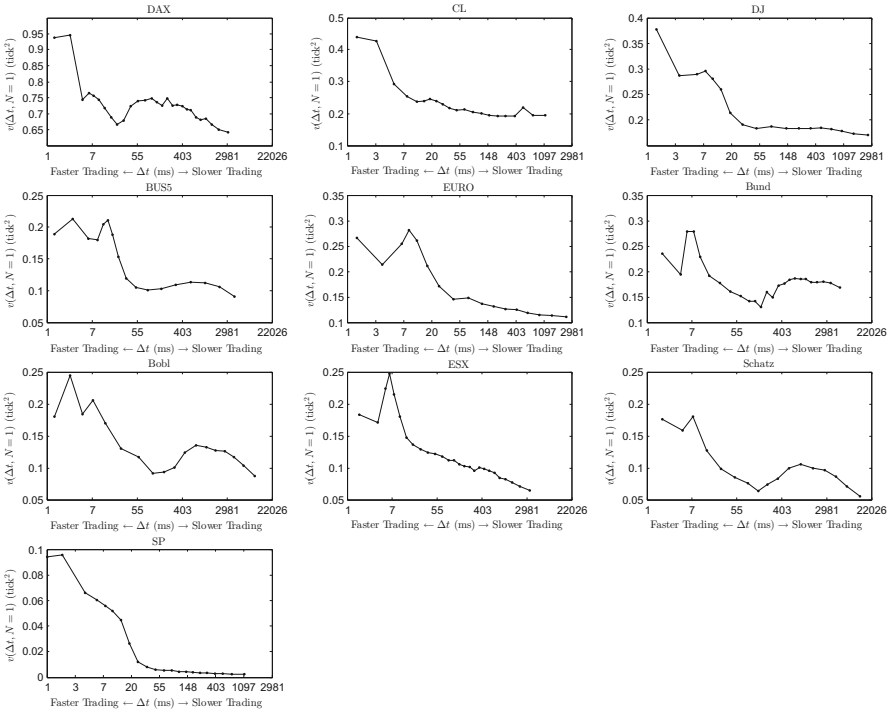


Fig. 5 For each asset (in increasing perceived tick size $P_=$), $v(\Delta t, N = 1)$ as a function of Δt (logarithm scale). We see an "explosion" of the variance when the trading is getting faster

$$P(\Delta t) = Prob\{r_{t_i} \neq 0 \mid t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}]\} \tag{10}$$

and where $A(\Delta t)$ is the expectation of the squared return conditioned by the fact that it is not zero and by the intertrade duration $t_i - t_{i-1} = \Delta t$, i.e.,

$$A(\Delta t) = \mathrm{e}_i \left[r_{t_i}^2 \mid r_{t_i} \neq 0 \text{ and } t_i - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{11}$$

In short $P(\Delta t)$ is the probability that the midpoint price moves while $A(\Delta t)$ is the squared amplitude of the move when non-zero. In Fig. 6, we have plotted, for all assets, the function $P(\Delta t)$ for different Δt. One clearly sees that, as the trading rate becomes greater ($\Delta t \to 0$), the probability to observe a move of the midpoint price increases. One mainly recovers the behavior we observed for the analog variance plots. Let us notice that (except for the DAX), the values of the moving probabilities globally decrease as the perceived ticks $P_=$ increases (for large ticks, e.g. SP, at very low activity this probability is very close to zero). The corresponding estimated moving squared amplitudes $A(\Delta t)$ are displayed in Fig. 7. It appears clearly that, except for the smallest perceived ticks assets (DAX and CL basically), the amplitude can be considered as constant. This can be easily explained: large tick assets never make moves larger than one tick while small tick assets are often "perforated" by

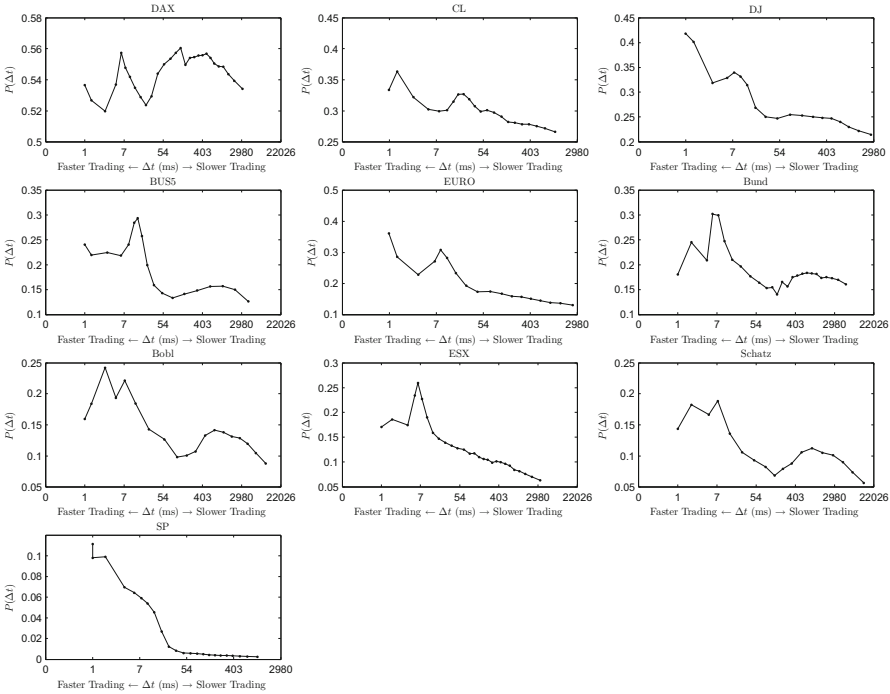


Fig. 6 For each asset (in increasing perceived tick size $P_=$), Probability $P(\Delta t)$ as defined by (10) as a function of Δt. We see that the probability of getting a price move increases with market order rate for most assets

a market order. One can thus say that, except for very small ticks assets, the variance increase in high trading rate period is mostly caused by the increase of the probability that a market order absorb only the first level of the book (best bid or best ask). There is hardly no perforation of the book on the deeper levels.

4.2 Impact on the Return

Before moving to the next section, let us just look at the direct impact on the return itself, as defined for instance by [3], conditioned by the intertrade time:

$$I(\Delta t, N = 1) = \mathrm{e}_i \left[\varepsilon_i r_{t_i} | t_i - t_{i-1} = \Delta t \right]. \tag{12}$$

According to [15], we expect the impact to be correlated with the variance per trade and therefore for $I(\Delta t)$ to follow a very similar shape to that of $v(\Delta t, N = 1)$ shown in Fig. 5. This is confirmed in Fig. 8 where one sees that, for all assets, the impact goes from small values for large intertrade intervals to significantly higher values for small intertrade durations.

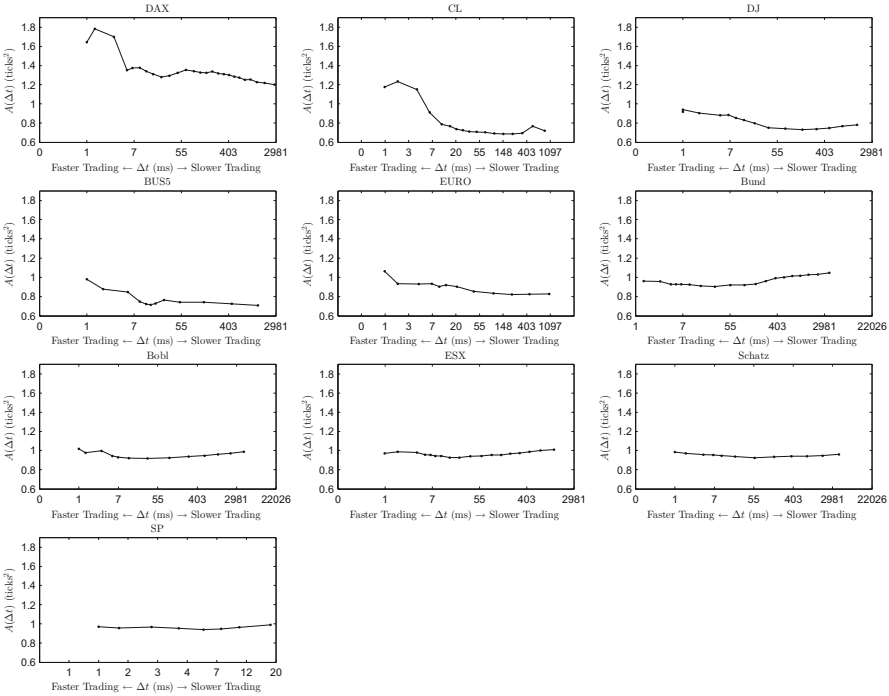


Fig. 7 For each asset (in increasing perceived tick size $P_{=}$), Size of Absolute squared returns $A(\Delta t)$ as a function pf Δt. For very small tick assets, like DAX and CL we see that the absolute size of a return increases with the rate of market orders. This propriety quickly stops being true as the tick increases. The orderbook of a large tick asset is generally much thicker than that of a small tick asset and therefore it is extremely hard to to empty more than one level

5 From Fine to Coarse

5.1 Large Scale Conditional Variance and Impact

One of the key issue associated to our single trade study is the understanding of the consequences of our findings to large scale return behavior. This question implies the study of (conditional) correlations between successive trades, which is out of the scope of this paper and will be addressed in a forthcoming work. However one can check whether the impact or the variance averaged locally over a large number of trades still display a dependence as respect to the trading rate. Indeed, in Fig. 3 we have already seen that this feature seems to persist when one studies returns over a fixed time (e.g., 5 min) period conditioned by the mean intertrade duration over this period. Along the same line, one can fix a large $N = N_0$ value and compute $v(\Delta t, N = N_0)$ and $I(\Delta t, N = N_0)$ as functions of Δt. Note that $v(\Delta t, N = N_0)$ is defined in Eq. (7) while the aggregated impact can be defined similarly as:

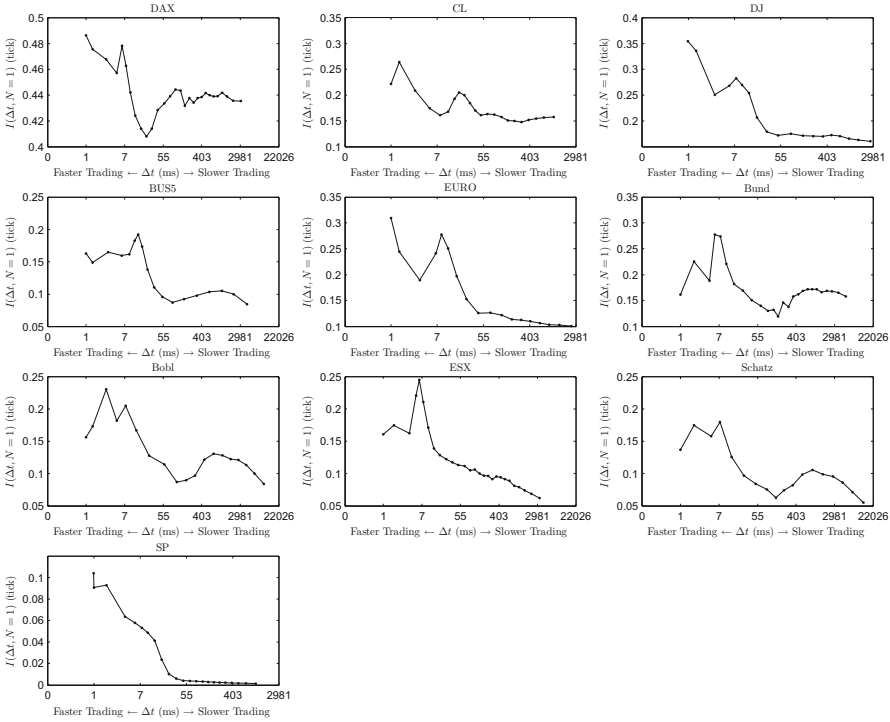


Fig. 8 For each asset (in increasing perceived tick size $P_=$), $I(N = 1|\Delta t)$ as defined by (12) as a function of Δt. The shape of the curves confirms the idea that the impact is high correlated with the variance per trade

$$I(\Delta t, N = N_0) = \mathrm{e}_i \left[\epsilon_i (p_{t_{i+N_0}} - p_{t_i}) \mid t_{i-1+N_0} - t_{i-1} \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{13}$$

In Fig. 9 and 10 are plotted respectively the variance $v(\Delta t, N = 100)$ and the return impact $I(\Delta t, N = 200)$ as functions of Δt. One sees that at these coarse scales, the increasing of these two quantities as the activity increases is clear (except maybe for the variance of the EURO). As compared to single trade curves, the threshold-like behavior are smoothed out and both variance and return impacts go continuously from small to large values as the trading rate increases.

5.2 Liquidity Decreases when Trading Rate Increases

One possible interpretation of these results would be that when the trading rate gets greater and greater, the liquidity tends to decrease, i.e., the order book tends to empty.

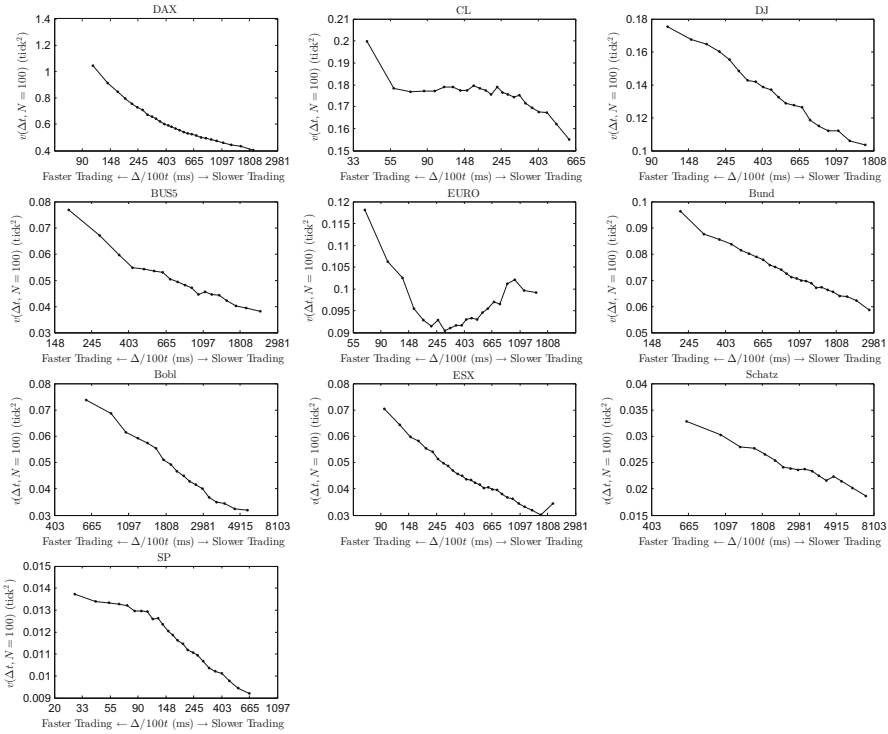


Fig. 9 For each asset (in increasing perceived tick size $P_=$), $v(\Delta t, N = 100)$, as defined by (7), as a function of Δt. Clearly the variance of a speedy 100 trades is higher than the variance of 100 slow trades

In [15], the authors mention that the spread is an indicator of the thinness of the book and that the distance from the best bid or ask to the next level of the order book is in fact equivalent to the spread. Moreover, they bring empirical evidence and theoretical no-arbitrage arguments suggesting that the spread and the variance per trade are strongly correlated. Accordingly, we define the mean spread over N trades as

$$s_{t_i,N} = \frac{1}{N} \sum_{k=0}^{N-1} s_{t_{i+k}}, \tag{14}$$

and the conditional spread at the fixed scale $N = N_0$ as

$$S(\Delta t, N = N_0) = \mathrm{e}_i \left[s_{t_i,N} \mid t_{i+N} - t_i \in [\Delta t - \delta_{\Delta t}, \Delta t + \delta_{\Delta t}] \right]. \tag{15}$$

Fig. 11 displays, for each asset, $S(\Delta t, N = 100)$ as a function of $\Delta t/100$ (using log scale). One observes extremely clearly an overall increase of the spread value with the rate of trading for all assets, This clearly suggests that the order book is thinner during periods of intense trading. This seems to be a universal behavior not depending at all on the perceived tick size.

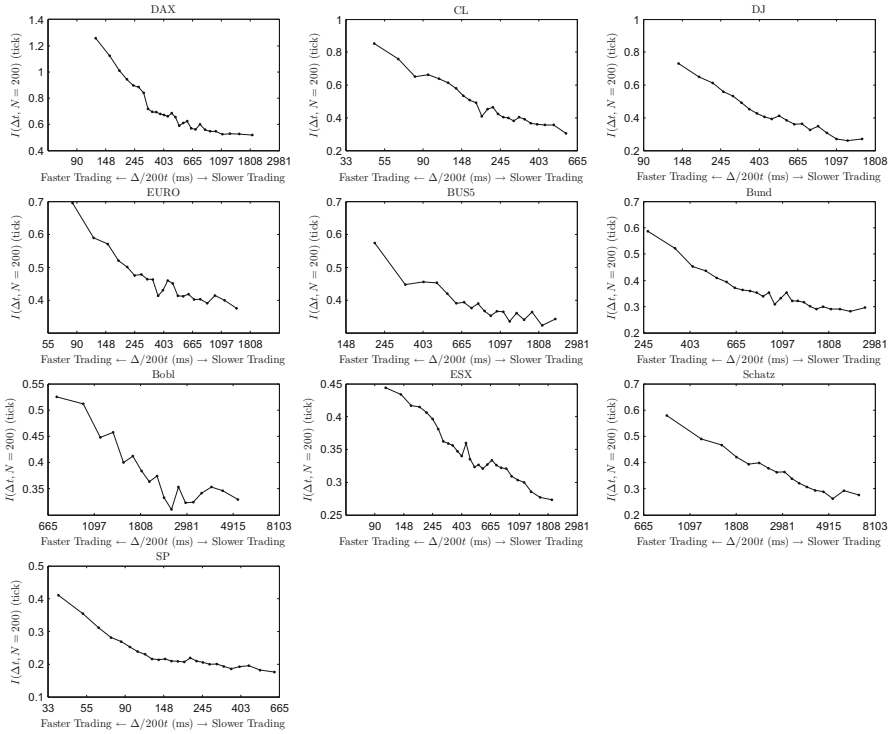


Fig. 10 For each asset (in increasing perceived tick size $P_=$), $I(\Delta t, N = 200)$, as defined by (13), as a function of Δt. The impact of speedy trades propagates well into the future. Even 200 trades away, one speedy trade has caused more impact than a slower one

6 Conclusions

In this short paper we provided empirical evidence gathered from high frequency futures data corresponding to various liquid futures assets that the impact (as measured from the return variance or using the standard definition) of a trading order on the midpoint price depends on intertrade duration. We have also shown that this property can also be observed at coarser scale. A similar study of the spread value confirmed the idea that order books are less filled when trading frequency is very high. Notice that we did not interpret in any causal manner our findings, i.e., we do not assert that a high transaction rates should be responsible for the fact that books are empty. It just appears that both phenomena are highly correlated and this observation has to be studied in more details. In a future work, we also plan to study the consequences of these observations on models such those described in the introductory section (Eq. (1)). A better understanding of the aggregation properties (i.e., large values of N) and therefore of correlations between successive trades will also be addressed in a forthcoming study.

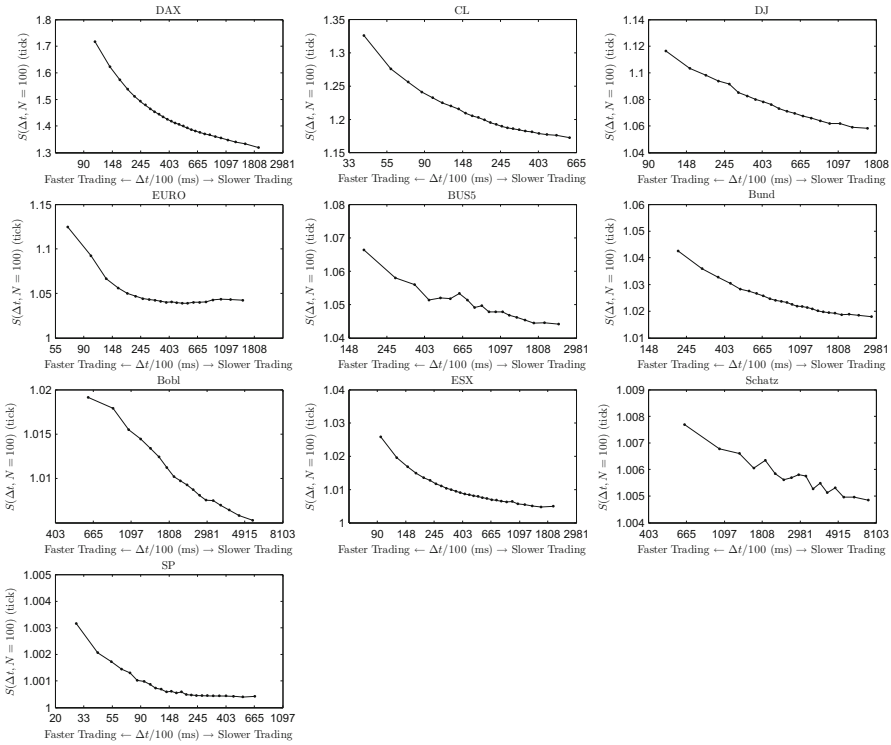


Fig. 11 For each asset (in increasing perceived tick size $P_=$), $S(\Delta t, N = 100)$ as defined in (15) as a function of $\Delta t/100$. The form of the curves confirms that there is a strong liquidity decrease when the trading rate is increasing

Acknowledgements The authors would like to thank Robert Almgren (Quantitative Brokers and NYU) for many useful comments and support. They would also like to thank Mathieu Rosenbaum, Arthur Breitman and Haider Ali for inspiring discussions. This research is part of the Chair *Financial Risks* of the *Risk Foundation.* The financial data used in this paper have been provided by the company *QuantHouse EUROPE/ASIA*, http://www.quanthouse.com.

References

1. Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. Risk, 2005
2. J.P. Bouchaud and M. Potters. Theory of Financial Risk and Derivative Pricing. Cambridge University Press, Cambridge, 2003
3. Jean-Philippe Bouchaud, J.D. Farmer, and Fabrizio Lillo. How markets slowly digest changes in supply and demand. In Thorsten Hens and Klaus Schenk-Hoppe, editors, Handbook of Financial Markets: Dynamics and Evolution, pages 57–156. Elsevier: Academic Press, 2008

4. Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. Quantitative Finance, 10:176–190, 2004
5. Jean-Philippe Bouchaud, Julien Kockelkoren, and Marc Potters. Random walks, liquidity molasses and critical response in financial markets. Quantitative Finance, 6(2):115–123, 2006
6. M.M. Dacorogna, R. Gençay, U.A. Müller, R.B. Olsen, and O.V. Pictet. An introduction to high frequency finance. Academic Press, 2001
7. Zoltan Eisler, Bouchaud Jean-Philippe, and Julien Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. Quantitative finance papers, arXiv.org, 2010
8. J. Doyne Farmer, Austin Gerig, Fabrizio Lillo, and Szabolcs Mike. Market efficiency and the long-memory of supply and demand: is price impact variable and permanent or fixed and temporary? Quantitative Finance, 6(2):107–112, 2006
9. J. Doyne Farmer, Laszlo Gillemot, Fabrizio Lillo, Szabolcs Mike, and Sen Anindya. What really causes large price changes? Quantitative Finance, 4:383–397(15), 2004
10. Jim Gatheral. No-dynamic-arbitrage and market impact. Quantitative Finance, 10(7):749–759, 2010
11. László Gillemot, J. Doyne Farmer, and Fabrizio Lillo. There's more to volatility than volume. Quantitative Finance, 6(5):371–384, 2006
12. Fabrizio Lillo and J. Doyne Farmer. The long memory of the efficient market. Studies in Nonlinear Dynamics & Econometrics, 8(3), 2004
13. B.B. Mandelbrot and H.W. Taylor. On the distribution of stock price differences. Operation Research, 15:1057–1062, 1967
14. Christian Y. Robert and Mathieu Rosenbaum. A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones. Forthcoming, Journal of Financial Econometrics, 2010
15. Matthieu Wyart, Jean-Philippe Bouchaud, Julien Kockelkoren, Marc Potters, and Michele Vettorazzo. Relation between bid-ask spread, impact and volatility in order-driven markets. Quantitative Finance, 8(1):41–57, 2008

Tick Size and Price Diffusion

Gabriele La Spada, J. Doyne Farmer and Fabrizio Lillo

1 Introduction

A tick size is the smallest increment of a security price. Tick size is typically regulated by the exchange where the security is traded and it may be modified either because the exchange enforces an overall tick size change or because the price of the security is too low or too high. There is an extensive literature, partially reviewed in Sect. 2 of the present paper, on the role of tick size in the price formation process. However, the role and the importance of tick size has not been yet fully understood, as testified, for example, by a recent document of the Committee of European Securities Regulators (CESR) [1].

Tick size can affect security price in direct and indirect ways. It is clear that at the shortest time scale on which individual orders are placed the tick size has a major role which affects where limit orders can be placed, the bid ask spread, etc. This is the realm of market microstructure and in fact there is a vast literature on the role of tick size on market microstructure. However, tick size can also affect price properties at longer time scales, and relatively less is known about the effect of tick size on the statistical properties of prices. The rationale is that since market microstructure affects price diffusion, if tick size affects microstructure it is likely

Gabriele La Spada
Department of Economics, Princeton University, Princeton, USA,
and
Department of Economic Sciences, LUISS, Rome,
e-mail: gla@princeton.edu

J. Doyne Farmer
Santa Fe Institute, Santa Fe, USA,
e-mail: jdf@santafe.edu

Fabrizio Lillo
Department of Physics and Technologies, University of Palermo, Palermo, Italy,
and
Santa Fe Institute, Santa Fe, USA,
e-mail: lillo@unipa.it

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

to also affect price diffusion. This point of view is strengthened by the observation that there is growing evidence that microstructural events are important factors in explaining longer term price fluctuations [2]. For example, large fluctuations in price returns are observed also at the individual transaction scale [3] and their effects last for long periods [4]. Nonetheless the relation between market microstructure and price diffusion on a longer time scale is still not fully understood. It is therefore worth investigating the effect of tick size on price diffusion properties.

The present paper is divided in two parts. In the first (Sect. 2) we review the effect of tick size change on the market microstructure and the diffusion properties of prices. The second part (Sect. 3) presents original results obtained by investigating the tick size changes occurring at the New York Stock Exchange (NYSE). We show that tick size change has three effects on price diffusion. First, as already shown in the literature, tick size affects price return distribution at an aggregate time scale. Second, reducing the tick size typically leads to an increase of volatility clustering. We give a possible mechanistic explanation for this effect, but clearly more investigation is needed to understand the origin of this relation. Third, we explicitly show that the ability of the subordination hypothesis in explaining fat tails of returns and volatility clustering is strongly dependent on tick size. While for large tick sizes the subordination hypothesis has significant explanatory power, for small tick sizes we show that subordination is not the main driver of these two important stylized facts of financial market. Finally Sect. 4 concludes.

2 Literature Review

In this section we review some literature on the effect of tick size on market microstructure and on price diffusion. Most of the studies we consider are case studies of tick size changes in different markets. This section is divided in two parts. In the first we review the effect of tick size on market microstructure, while in the second we discuss recent papers on the effect of tick size on price diffusion.

2.1 Tick Size and Market Microstructure

Crack [5] and Ahn *et al*. [6] studied the impact of the September 3, 1992 AMEX reduction in the minimum tick size from 1/8 to 1/16 for stocks priced under five dollars. They found approximately a 10% decline in quoted spreads and depths in addition to an increase in average daily trading volume of 45-55 %. Niemeyer and Sandås [7] studied the Stockholm Stock Exchange and found that tick size is positively correlated to spread and market depth, and negatively correlated to volume. Angel [8] found that small tick size narrows the bid-ask spread, but diminishes liquidity by making limit order traders and market makers more reticent to supply shares.

A series of studies [9–12] investigated the April 15, 1996 Toronto Stock Exchange (TSE) reduction in the minimum tick size to five cents. They found a significant decline in the quoted bid-ask spreads of 17-27% and in the quoted depth of 27-52%, while average trading volume displayed no statistically significant increase.

Bessembinder [13] studied the Nasdaq stocks whose price level passes through $10, and thus changed tick size from 1/8 to 1/32 and found that the effective spread fell by 11%. Ronen and Weaver [14] studied the May 7, 1997 switch to 1/16 at the AMEX and found reduced quote spreads and depths. They concluded that the implemented reduction to the minimum tick size has decreased transaction costs and increased liquidity. Bollen and Whaley [15] and Ricker [16] studied the 1997 tick size reduction from 1/8 to 1/16 at the NYSE and found that volume weighted bid-ask spread declined by approximately 13-26% while quoted depth decreased between 38% and 45%. They concluded that the NYSE tick size reduction has improved the liquidity of the market especially for low-priced shares. Jones and Lipson [17] used institutional data to study the effect of tick size changes at the Nasdaq and NYSE. They found that trading costs decreased for smaller trades, but increased for larger trades.

Goldstein and Kavajecz [18] studied the tick size change from 1/8 to 1/16 at the NYSE. They found that the quoted bid-ask spread narrowed by 14.3% (note that this is the spread quoted by the specialist). For the most infrequently traded stocks the spread increased. The quoted depth declined by 48% while the limit order book spread (i.e. the spread between the highest buy order and the lowest sell order) increased by 9.1% (note that this result is in disagreement with previous studies). The tick size reduction had also an effect on transaction cost. Transaction costs for small orders decreased even if this benefit is reduced for infrequently traded and low-price stocks. Transaction costs for large orders either did not change (for frequently traded stocks), or increased (for infrequently traded stocks). The authors were also able to track the behavior of different market participants. They found that after tick size reduction floor members were less frequently providing additional depth at existing limit order book prices while they were more frequently improving best prices. Contribution to displayed depth from floor members decreased by 35% on average. Finally, limit order traders increased the ratio of cancelled limit orders to total limit orders by 6.2%.

Huang and Stoll [19] compared two different market structures, namely the NYSE, which is an auction market with a tick size rule, and the LSE, which in the investigated period was a dealer market with no minimum tick size. They found that dealer market spreads are higher than auction market spreads, because in auction markets limit orders narrow spreads. Similarly depth is lower in auction markets because limit orders narrow the spreads, and these spread narrowing limit orders are small. Finally, in both markets they found evidence of clustering, which is the tendency for prices to fall on a subset of available prices. Quote clustering is highly correlated with spreads, while trade clustering is smaller in a auction market because limit orders break up quote clustering as they seek to gain priority.

Finally, Bacidore *et al.* [20] studied the tick size change from 1/16 to 1/100 at the NYSE ("decimalization") by considering separately the NYSE and the Consolidated Quotation. They found that the NYSE's quoted spread fell by 30% while non-NYSE spreads also generally decreased, albeit by relatively small amounts. Overall, the NBBO[1] quoted spread fell by 30.7%. Overall, NYSE quoted size fell by 70.5% while non-NYSE average size declined by 26.9% (in both cases the greatest percentage decrease occurred in the high-volume, low-price group). NBBO size fell by 61.5%, which suggests that the non-NYSE markets became more likely to add to the NBBOs depth. The authors also investigated the overall shape of the limit order book. They found that displayed liquidity fell dramatically with decimal pricing, with a drop in displayed size of the order of 50%. This effect appears to be greatest for low-priced stocks. As for the order properties, decimalization had only a minimal effect on the relative fraction of market and limit orders. However the average limit order size decreased by 33.4% and the average market order size decreased by 15.7% after decimalization. Limit order traders became more aggressive after decimalization and the cancellation rate increased from 43% to 53%. The average time between quote updates declined and the total number of quotes across all markets increased; this effect was stronger in non-NYSE markets. NYSE's share of quotes declined on average from 40% to 34%. The average number of NBBO per stock per day doubled, while the number of quotes increased only by 27% and the fraction of the average trading day that the NYSE is at both sides of the NBBO declined from 93% to 82%. All these results are consistent with a more competitive quoting environment. The average transaction cost greatly increased after decimalization[2]. However, the effect is different depending on the volume. For all stocks, the cost of small orders decreased while the cost of large orders increased after decimalization.

Taken together these studies indicate that a reduction in tick size (i) narrows the spread; (ii) decreases the quoted depth and the overall depth in the limit order book, i.e. the displayed liquidity decreases; (iii) modifies the order flow in a way that the rate of orders (and cancellations) per unit time increases, but their size becomes smaller; (iv) the transaction cost decreases for small orders, but increases for large orders (at least when the cost is computed by assuming that a large order climbs up the book); (v) creates a more competitive environment for liquidity provision inside a market (for example limit orders become more aggressive) and across different market segments. These two last aspects are probably important factors which contribute significantly to the practice of order splitting and algorithmic trading.

---

[1] The National Best Bid and Offer (NBBO) includes prices from all competing exchanges and refers to the price at the time of entry into the market.

[2] As in other studies, this is the average cost of trading a given number of shares if the only liquidity in the market is the liquidity displayed in the book. Assuming that the midpoint of the spread is a proxy for the value of a security, the cost of displayed liquidity for an order is the product of the additional shares available in the limit order book at each price point times the distance of the price point from the spread midpoint summed through the number of shares in the order of interest. Dividing that sum by the total number of shares in the order provides the per share cost of obtaining the displayed liquidity. This might be different from the actual cost of a large order obtained for example with order splitting.

It is worth noting that the literature has given different sign to the relation between tick size and liquidity. On one side a small tick size reduces the spread, i.e. increases liquidity. On the other hand, it reduces depth and makes liquidity providers more reluctant to display their orders, with an effect of reducing liquidity. These contradictory results can be better reconcilied by considering that a complex concept such as liquidity is hardly captured by one metric.

2.2 Tick Size and Price Properties on Longer Time Scales

As said above, the literature on the relation between tick size and the statistical properties of prices is smaller than the huge literature on tick size and market microstructure. Here we review two recent papers from the Econophysics literature.

Onnela *et. al.* [21] studied how tick size affects price return distribution. They assumed a continuous price process that is discretized by the tick size. By using numerical simulations they found that the effect of discretization on return distribution due to tick size is negligible when the tick-to-price ratio is small, while it is significant when this ratio is large. Moreover, the proportion of zero returns appears to be much higher for stocks with a high tick-to-price ratio. They performed an empirical study by considering stocks traded at the same time in two exchanges with different tick sizes (NYSE and TSE). They observed that on average the proportion of zero returns increases as the tick size increases, but this is better accounted for by the tick-to-price ratio, whose variation explains roughly 69% of variation in zero returns. Moreover 57% of cases exhibit price clustering, such that the effective tick size deviates from the nominal tick size. In particular at the NYSE there is a strong preference for even-eights.

In a recent paper Munnix *et al.* [22] considered two effects of tick size on price diffusion properties. The first is, as in the previous paper, on the return distribution and the conclusions are similar. The second is the effect of tick size on the estimation of cross correlation between price returns of two stocks. It is known [23] that cross correlation between returns of two stocks declines when one reduces the length of the time interval used to compute returns. There has been several explanations for this, which is called the Epps effect, ranging from those based on learning to those on non-syncronous trading [24, 25]. By using numerical simulations and analytical calculations authors of [22] showed that even for synchronous time series, the discretization due to the tick size induces a distortion of the correlation coefficient toward smaller intervals. They then test their model on real financial data and they find that the discretization effect is responsible for up to 40% of the Epps effect. Moreover, the contribution of the discretization effect is particularly large for stocks that are traded at low prices. This highlights the importance of the tick-to-price ratio as compared to the absolute tick size. We will see below that more generally tick size affects the correlation properties of the second moment including its temporal autocorrelation.

3 Tick Size and Price Diffusion

In this section we investigate the role of tick size in the statistical properties of price fluctuations, namely the return distribution, volatility clustering, and the subordination hypothesis.

3.1 Return Distribution

The first question is how tick size change affects the distributional properties of returns. We have mentioned above two recent papers [21,22] that discuss theoretically and empirically how the return distribution changes when tick size changes.

We use here a set of 5 high cap stocks (KO, MRK, PEP, T, WMT) traded at the New York Stock Exchange (NYSE). There have been two tick size changes in the NYSE. On June 24, 1997 the tick size changed from $1/8$ to $1/16$ of a dollar, and on January 29, 2001 the tick size changed from $1/16$ to $1/100$ of a dollar. We consider short timescale log returns, namely, fifteen-minute returns. For each tick size change we consider two time intervals of length 100 trading days, one before and one after the tick size change. We consider absolute returns $|r|$ and we compute the complementary cumulative distribution function defined as $F_c(x) = \mathrm{P}(|r| \geq x)$. In Fig. 1 we show the complementary cumulative distribution function of 15 minute absolute returns for the stock MRK before and after the first tick size change (left panel), and the second tick size change (right panel). When the tick size changed from 1/8 to 1/16, both smaller non-zero returns and larger returns became more likely, i.e. the distribution function of the absolute returns became more fat-tailed. On the other hand, when the tick size changed from 1/16 to 1/100, only smaller non-zero returns became more likely. In this second case the distribution function of absolute returns did not become more fat-tailed, but the number of small non-zero returns increased significantly. Similar results hold for all the other stocks.

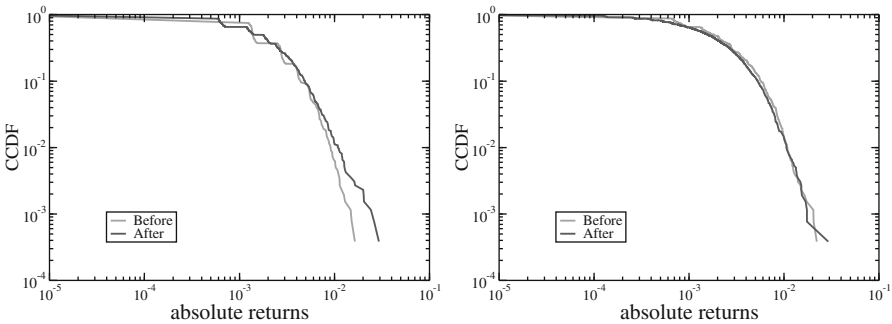


Fig. 1 Complementary cumulative distribution function of the stock MRK before and after the first tick size change (*left panel*), and the second tick size change (*right panel*)

We use the probability of zero price change and the tail exponent to quantify the difference in the return distribution before and after a tick size change. We denote the frequency of zero returns by p_0. We define the difference in the frequency of zero returns just before (p_0^-) and just after (p_0^+) the tick size change, as $\Delta p_0 = p_0^+ - p_0^-$. For each tick size change we perform a t-test of the null hypothesis that the mean of Δp_0 is statistically greater than or equal to zero. We obtain a p value equal to 0.0031 and 0.0014 for the first and the second tick size change, respectively. We can therefore reject the null that Δp_0 did not change or increase at a level of 1% for both tick size changes. This means that in both cases the frequency of zero returns diminished. On average it reduced by 50% on the first tick size change and by 70% on the second tick size change. In other words, the price becomes less "sticky" when the tick size decreases. In order to study the tail properties of the return distribution, we make use of the Hill estimator α_H of the tail exponent. Fig. 2 shows the value of the Hill estimator of the 5 stocks before and after the two tick size changes. As before, we define the difference in the Hill estimator before and after a tick size change as $\Delta\alpha_H = \alpha_H^+ - \alpha_H^-$ and we perform a t-test of the null hypothesis that $\Delta\alpha_H$ is greater or equal than zero. For the first tick size change we can reject the null hypothesis at a level of 10% ($p = 0.06$), while for the second tick size change we cannot reject the null hypothesis ($p = 0.62$). By using a shorter time period (50 trading days) on a larger set of 9 stocks we obtain a 5% significant rejection of the null for the first tick size change, while again we cannot reject the null for the second tick size change. This confirms the intuition obtained from Fig. 1 that the tail of the return distribution changed after the first tick size change, but remained statistically the same in the second tick size change. The explanation of this difference remains an open issue.

In conclusion, our empirical results confirm that the return distribution can be different before and after a tick size change. In both cases the frequency of zero returns is higher for large tick sizes, while the behavior of the tails seems to be different in the two tick size changes.

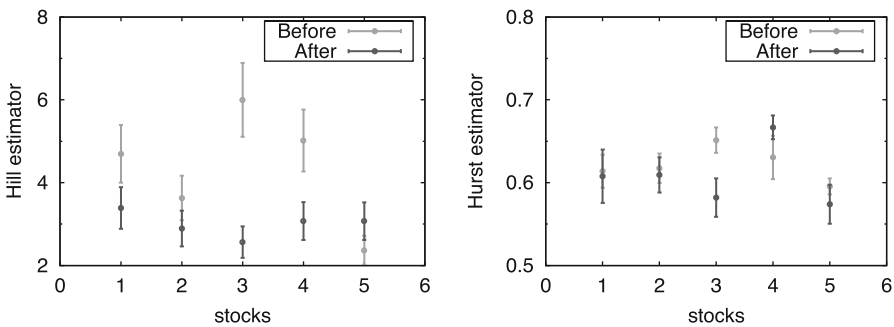


Fig. 2 Hill estimator for the tail exponent of the absolute return distribution before and after a tick size change from 1/8 to 1/16 (*left panel*) and from 1/16 to 1/100 (*right panel*). The stocks are alphabetically ordered and the error bars are 95% confidence intervals

3.2 Volatility Clustering

The other important stylized fact of financial time series is volatility clustering. Roughly speaking, volatility clustering means that periods of high (low) volatility are more likely followed by periods of high (low) volatility. A method to quantify the degree of volatility clustering is through its autocorrelation function (ACF). It is commonly observed that the ACF of volatility decreases slowly to zero and is statistically different from zero for long time periods of the orders of months. Moreover, there is a consensus that volatility is a long memory process [26], which means that for large lags the ACF $\rho(k)$ decays as a power law, $\rho(k) \sim k^{-\gamma}$ with $0 < \gamma < 1$. A long memory process lacks a typical time scale and can be characterized by the Hurst exponent H, which for long memory processes is given by $H = 1 - \gamma/2$. Here we investigate how the ACF and the (estimated) Hurst exponent changes when the tick size changes. We consider the same stocks and time periods as in the previous section using the absolute price return for 15 minute intervals as a proxy of volatility

Let us denote the ACF of the absolute returns by $\rho(k)$, where k is thenumber of 15 minute intervals. Fig. 3 shows the ACF of volatility for the stock MRK for 100 days before and after the tick size change. In both cases the volatility ACF before the tick size change is smaller than the one after the tick size change. This suggests that volatility is less clustered (autocorrelated) when the tick size is large. In order to test this conclusion more quantitatively we define the difference in the autocorrelation function before and after the tick size change as $\Delta\rho(k) = \rho^+(k) - \rho^-(k)$, where $\rho^-(k)$ is the ACF just before the change, and $\rho^+(k)$ is the ACF just after the change. For each lag $k = 1, \ldots, 4$ we perform a t-test of the null hypothesis that the mean of $\Delta\rho(k)$ is smaller than zero. Fig. 3 shows the statistics and the p value. In all but one case we reject the null at the 1% confidence level. In one case ($k = 1$ in the second tick size change) the p value is 0.019. This shows that the volatility becomes more correlated, i.e. more clustered, when the tick size is reduced. In the next subsection we present a simple model explaining this effect.

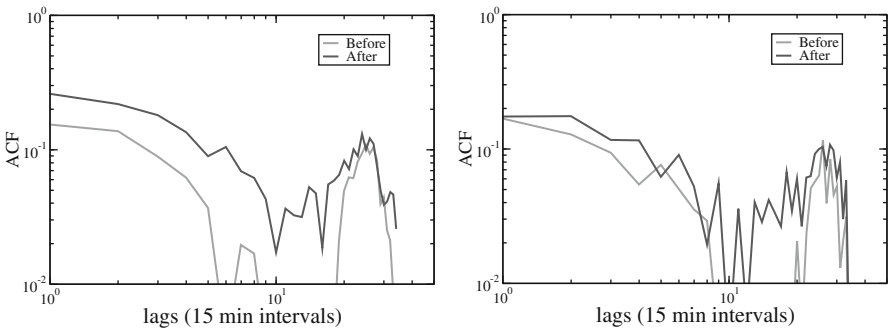


Fig. 3 Autocorrelation function of 15 min absolute returns for the stock MRK in a 100 trading day period before and after the tick size change from 1/8 to 1/16 (*left panel*) and from 1/16 to 1/100 (*right panel*)

Table 1 Result of a t-test of the null hypothesis that the difference $\Delta\rho(k)$ of the autocorrelation function of 15 min absolute returns before and after a tick size change is smaller than zero

	$k=1$		$k=2$		$k=3$		$k=4$	
	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value
tick size change #1	3.7982	0.0096	5.8893	0.0021	5.6963	0.0023	7.0142	0.0011
tick size change #2	3.0323	0.0193	7.4501	0.0009	3.9608	0.0083	8.7527	0.0005

The previous analysis focused mainly on the autocorrelation for small lags. In order to investigate the change in the ACF for large lags before and after a tick size change we compute the Hurst exponent of absolute returns. It was suggested in [27] that the Hurst exponent of volatility becomes larger when the tick size becomes smaller. This is based on empirical evidence obtained by measuring the Hurst exponent in periods of about three years. Here we want to see if this effect persists on much shorter time intervals (5 months). The main problem is that we have shorter time series and therefore the estimation of the Hurst exponent is noisier. We estimate the Hurst exponent by using the detrended fluctuation analysis (DFA) [28]. Fig. 4 shows the estimated Hurst exponent of volatility for the two tick size changes. The reduction of tick size from 1/8 to a 1/16 is associated with an increase of the estimated Hurst exponent, while in the second tick size change this phenomenon is not evident. A t-test of the null hypothesis that the variation $\Delta H = H^+ - H^-$ in the estimated Hurst exponent is smaller than zero confirms this intuition.

In conclusion, we showed evidence that a reduction in tick size leads to an increase of volatility clustering, measured as the autocorrelation function of 15 min absolute returns. In the next section we give some indication that this phenomenon can be explained with a mechanistic model of price dynamics. A more detailed explanation of this effect is presented elsewhere [29]. We found less evidence that the Hurst exponent changes when the tick size is modified. The increase of the Hurst exponent is clear after the first tick size reduction, while it is not observed during

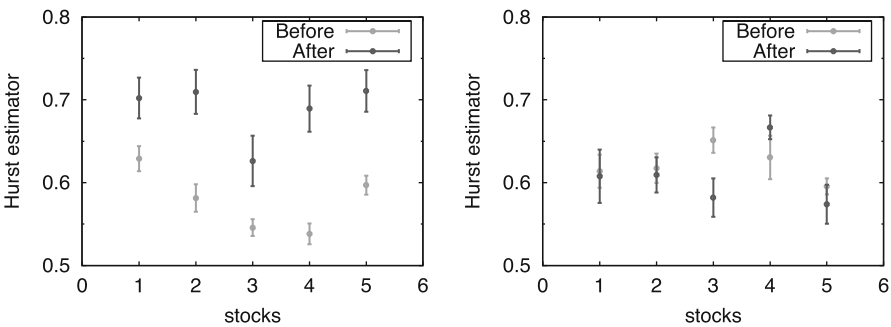


Fig. 4 DFA estimator of the Hurst exponent of the absolute returns before and after a tick size change from 1/8 to 1/16 (*left panel*) and from 1/16 to 1/100 (*right panel*). The stocks are alphabetically ordered and error bars are standard deviations

the second tick size change. This might be due to a poor estimation of the Hurst exponent. Mechanistic models [29] predict that the Hurst exponent should remain the same, but they also show that larger tick sizes have a lower *estimated* Hurst exponent.

3.3 A Simple Model

We present here a simple mechanistic model that reproduces the increase in volatility clustering when the tick size is reduced. A more detailed explanation is given in [29]. Consider a price process described by a simple ARCH(1) process [30]

$$r_t = \sigma_t z_t \qquad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 \tag{1}$$

where r_t is the price returns on a given time scale and z_t is a Gaussian noise term with zero mean and unit variance. We generated a return time series of length 2^{16} with parameters $\alpha_0 = 0.1$ and $\alpha_1 = 0.9$. We then construct the price time series p by integrating the return time series[3]. We assumed that this (unobservable) price is coarse grained by the tick size grid. Specifically, if the tick size is δ the observed price at the given time scale is $p_{obs} = [p/\delta]\delta$, where $[x]$ indicates the integer part of x. Finally, we construct observed returns from the observed price and we compute the autocorrelation function of absolute observed returns.

In Fig. 5 we show the autocorrelation function of absolute returns of the original process and of the discretized process. Note that the squared returns of an ARCH(1) price process are exponentially autocorrelated with a time scale $1/|\log \alpha_1|$. It is evident that coarse graining due to tick size reduces the ACF and that the larger the tick size, the smaller the autocorrelation function, similarly to what we observed in

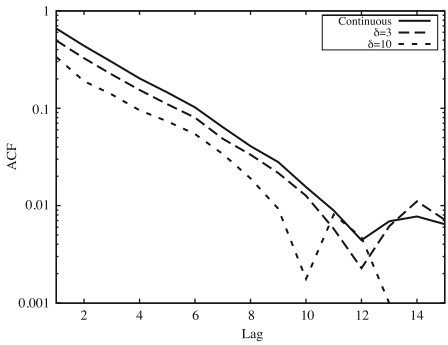


Fig. 5 Autocorrelation function of absolute returns of an ARCH(1) model and of a coarse grained version of it with different tick sizes δ

[3] Note that we are considering additive rather than multiplicative returns. For the small time scales we are considering here the two give nearly equal results.

real data. This is a fully mechanistic model of the effect of tick size on volatility clustering. Other reasons can be at the origin of the empirically observed increase of volatility clustering after a tick size reduction. These may include microstructural effects, change in the strategic behavior of traders, etc. Our simple model and its extensions [29] show that part of the effect could be due to purely mechanical reasons.

3.4 Tick Size and the Subordination Hypothesis

The origin of fat tails and clustered volatility in price fluctuations is an important problem in financial economics Although the cause is still debated, the view has become increasingly widespread that in an immediate sense both of these features of prices can be explained by fluctuations in volume, particularly as reflected by the number of transactions. The original idea dates back to a paper by Mandelbrot and Taylor [31] that was developed by Clark [32]. Mandelbrot and Taylor proposed that prices could be modeled as a subordinated random process $Y(t) = X(\tau(t))$, where Y is the random process generating returns, X is Brownian motion and $\tau(t)$ is a stochastic time clock whose increments are IID and uncorrelated with X. Clark hypothesized that the time clock $\tau(t)$ is the cumulative trading volume in time t, but more recent works indicated that the number of transactions is more important than their size [33]. Gillemot, Lillo, and Farmer [27] performed a series of shuffling experiments and showed that neither number of transactions nor volume are the principal cause of heavy tails in price returns and clustered volatility. Specifically, they compared returns (or volatilities) computed under different measures of time. They found that volatility is still very strong even if price movements are recorded at intervals containing an equal number of transactions (or volume), and that the volatility observed in this way is highly correlated with volatility measured in real time. In contrast, when they shuffle the order of events, but control for the number of transactions so that it matches the number of transactions in real time, they observe a much smaller correlation to real time volatility.

For the purpose of this paper we would like to stress the importance of the tick size in assessing the relative role of the subordination hypothesis in explaining fat tails and clustered volatility (in part discussed also in [27]). We consider here the 626 trading day period from Jan 1, 1995 to Jun 23, 1997, when the tick size at NYSE was 1/8 of a dollar, and the 734 trading day period from Jan 29, 2001 to December 31, 2003, when the tick size was a penny. We consider returns in real time, transaction time, and shuffled transaction time. Real time returns are simply 15 minute returns. Transaction time returns are obtained by considering time intervals containing an equal number of transactions. To make series with different time measures comparable we compute returns in transaction time by considering a number of transactions equal to the average number of transactions in 15 min. Finally, shuffled transaction time is obtained in the following way: We first measure the return of each transaction. We then shuffle the time series of individual trans-

action returns and we aggregate individual transaction returns so that we match the number of transactions in each real time interval. In transaction time returns we are destroying any fluctuation of number of transactions while we preserve the temporal sequence of real trades. In contrast, in shuffled transaction time we preserve the fluctuations of trading activity (as measured by the number of transactions), but we destroy any temporal correlation between the amplitude of consecutive individual transaction price movements. If the subordination hypothesis is correct, then real time returns should be closer to shuffled transaction time returns than to transaction time returns.

In Fig. 6 we show the complementary cumulative distribution function of absolute returns of the stock PG in the two tick size regimes. When the 1/8 tick size was shuffled, transaction time returns are closer than transaction time returns to real time returns. This is qualitatively consistent with the subordination hypothesis. The period when the tick size was a penny shows a completely different pattern. The right panel of Fig. 6 shows that in the small tick size regime the transaction time returns are closer than the shuffled transaction time returns to the real time returns. Clearly in this case the subordination hypothesis is not the main driver of fat tails. In other words, in order to reproduce fat tails of returns it is more important to preserve the temporal order of individual transaction returns than the fluctuations of their arrival rate. Since individual price returns are largely determined by liquidity fluctuations, these results indicate that when the tick size is small liquidity fluctuations are more important than volume (i.e. number of trades) fluctuations.

A similar conclusion can be drawn by investigating volatility clustering. Fig. 7 shows the autocorrelation of absolute returns for PG under different tick size and time measures. When tick size is large (left panel), shuffled transaction time returns show volatility clustering that is close to the real one, while transaction time absolute returns are relatively less correlated. This again is in agreement with the subordination hypothesis. On the contrary, in the small tick size regime (right panel)

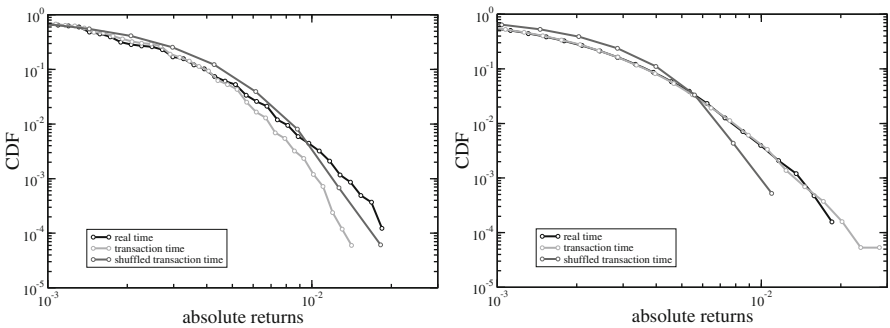


Fig. 6 Complementary cumulative distribution function of absolute returns of the stock PG in the period when the tick size was 1/8 (*left*) and when it was a penny (*right*). We show the distribution of returns computed in real time (*black*), transaction time (*light gray*), and shuffled transaction time (*dark gray*)

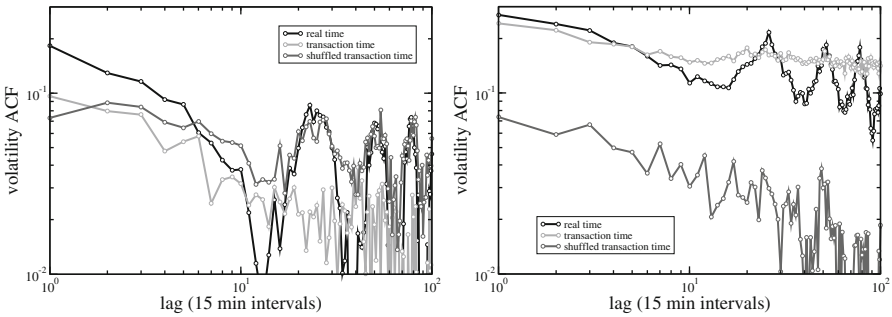


Fig. 7 Autocorrelation of absolute returns for the stock PG in the period when the tick size was 1/8 (*left*) and when it was a penny (*right*). We show the autocorrelation of absolute returns computed in real time (*black*), transaction time (*light gray*), and shuffled transaction time (*dark gray*)

the transaction time returns show volatility clustering very close to the real case[4]. Shuffled transaction time returns show a volatility clustering that is almost an order of magnitude smaller than the real case. Again, when the tick size is small, the subordination hypothesis plays a minor role in explaining volatility clustering. We conjecture that the strong liquidity autocorrelation is an important driver of volatility clustering. In fact, most measures of liquidity, such as the spread [34] and distance between occupied levels in the order book (gaps) [35], display strong autocorrelation, often consistent with long memory. Finally, note that the level of absolute return ACF in real time is much larger in the small tick size regime (right panel) than in the large tick size regime (left panel). This is in agreement with the analysis and the mechanistic model of the previous section.

In conclusion we have shown that the role of the subordination hypothesis in fat tails of returns and volatility clustering is strongly dependent on tick size. This is important also because many empirical studies of the subordination hypothesis have been performed on time periods of large tick size. Our study raises the question of the validity of these studies if they are applied to recent periods of small tick size.

4 Conclusions

In conclusion we have shown that tick size has multiple roles in influencing the statistical properties of price diffusion. As expected, and shown in the literature, tick size affects return distribution. When tick size is large (compared to the price) the price is more "sticky", i.e. there is an higher fraction of zero returns. It is less clear

[4] Note that real time and shuffled transaction time absolute return ACF show peaks due to the daily periodicity of trading activity. This is not observed for the transaction time absolute return ACF because of the way in which we have constructed the time series. Our discussion here refers to the global level of the ACF and not to the daily periodicities.

whether the tails of return distributions are affected by tick size. We have shown that tick size affects volatility clustering. A decrease in tick size leads to more clustered volatility. There might be many reasons for this effect that should be investigated. Here we have shown that a simple mechanistic model is able to qualitatively capture this effect. Finally, tick size influences the relative role of volume (i.e. number of transactions) vs. liquidity fluctuations in explaining return distribution and volatility clustering. This relates to the importance of the subordination hypothesis in price diffusion. The original results presented here are preliminary in the sense that one should consider different time windows, use control windows to check for global trends, and consider different more tick size changes.

The approach we have followed in this research is to directly study the effect of tick size on price diffusion properties. We know that this effect is mediated by the price formation process, i.e. by the market microstructure. The long term goal of this research is to understand how tick size affects market microstructure and how in turn microstructure affects price diffusion.

References

1. http://www.cesr-eu.org/index.php?page=consultation_details&id=158
2. Bouchaud JP, Farmer JD, Lillo F (2009) How markets slowly digest changes in supply and demand. In Hens T, Schenk-Hoppé KR (eds) Handbook of Financial Markets: Dynamics and Evolution, North-Holland
3. Farmer JD, Gillemot L, Lillo F, Mike S, Sen A (2004) What really causes large price changes? Quantitative Finance 4: 383–397
4. Ponzi A, Lillo F, Mantegna RN (2009) Market reaction to a bid-ask spread change: A power-law relaxation dynamics. Physical Review E 80:016112
5. Crack TF (1994), Tinkering with ticks: choosing minimum price variation for U.S. equity markets. Working Paper, Massachusetts Institute of Technology, unpublished
6. Ahn H, Cao CQ, Choe H (1996) Tick size, spread and volume. Journal of Financial Intermediation 5:2–22
7. Niemeyer J, and Sandås P (1994) Tick size, market liquidity and trading volume: evidence from the Stockholm Stock Exchange. Working Paper, Stockholm School of Economics, unpublished
8. Angel JJ (1997) Tick size, share price, and stock splits. Journal of Finance 52:655–681
9. Bacidore J (1997) The impact of decimialization on market quality: an empirical investigation of the Toronto Stock Exchange. Journal of Financial Intermediation 6:92–120
10. Porter D and Weaver D (1997) Decimalization and market quality. Financial Management 26:5–26
11. Huson M, Kim Y, Mehrotra V (1997) Decimal quotes, market quality, and competition for order flow: evidence from the Toronto Stock Exchange. Working Paper, University of Alberta, unpublished
12. Ahn H, Cao CQ, Choe H (1998) Decimalization and competition among stock markets: evidence from the Toronto Stock Exchange cross-listed securities. Journal of Financial Markets 1:51–87
13. Bessembinder H (1997) Endogenous changes in the minimum tick: an analysis of Nasdaq securities trading near ten dollars. Working Paper, Arizona State University, unpublished
14. Ronen T, and Weaver DG (1998) Teenies' anyone: the case of the American Stock Exchange. Working Paper, Rutgers University, unpublished

15. Bollen NPB, Whaley RE (1998) Are "teenies" better? Journal of Portfolio Management 25:10–24
16. Ricker JP (1998) Breaking the eighth: sixteenths on the New York Stock Exchange. Working Paper, unpublished
17. Jones CM, Lipson ML (2001) Sixteenths: Direct Evidence on Institutional Trading Costs. Journal of Financial Economics 59:253–278
18. Goldstein MA, Kavajecz KA (2000) Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. Journal of Financial Economics 56:125–149
19. Huang RD, Stoll HR (2001) Tick Size, Bid-Ask Spreads, and Market Structure. The Journal of Financial and Quantitative Analysis 36:503–522
20. Bacidore J, Battalio R, Jennings R (2001) Changes in Order Characteristics, Displayed Liquidity, and Execution Quality on the New York Stock Exchange around the Switch to Decimal Pricing. New York Stock Exchange Working Paper
21. Onnela JP, Töyli J, Kaski K (2009) Tick size and stock returns. Physica A: Statistical Mechanics and its Applications 388:441–454
22. Münnix MC, Schfer R, Guhr T (2010) Impact of the tick-size on financial returns and correlations. preprint at `arXiv:1001.5124v4`
23. Epps TW (1979) Comovements in stock prices in the very short run. Journal of the American Statistical Association 74:291–298
24. Renó R (2003) A closer look at the Epps effect. International Journal of Theoretical and Applied Finance 6:87–102
25. Tóth B, Kertész J (2009) The epps effect revisited. Quantitative Finance 9:793–802
26. Ding Z, Granger CWJ, Engle RF (1993) A long-memory property of stock market returns and a new model. Journal of Empirical Finance 1:83–106
27. Gillemot L, Farmer JD, Lillo F (2006) There's more to volatility than volume. Quantitative Finance 6:371–384
28. Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Physical Review E 49: 1685–1689
29. La Spada G *et al.* (2010) Effect of coarse-graining on the diffusion properties of a long memory process. in preparation
30. Engle RF (1982) Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. Econometrica 50:987–1008
31. Mandelbrot BB, Taylor HM (1967) On distribution of stock price differences. Operations Research 15:1057–1062
32. Clark PK (1973) Subordinated stochastic process model with finite variance for speculative prices. Econometrica 41:135–155
33. Ane T, Geman H (2000) Order flow, transaction clock, and normality of asset returns. The Journal of Finance 55 2259–2284
34. Plerou V, Gopikrishnan P, Stanley HE (2005) Quantifying fluctuations in market liquidity: analysis of the bid-ask spread. Physical Review E 71:046131
35. Lillo F, Farmer JD (2005) The key role of liquidity fluctuations in determining large price changes. Fluctuation and Noise Letters 5:L209–L216

High Frequency Correlation Modelling

Nicolas Huth and Frédéric Abergel

1 Introduction

Many statistical arbitrage strategies, such as pair trading or basket trading, are based on several assets. Optimal execution routines should also take into account correlation between stocks when proceeding clients orders. However, not so much effort has been devoted to correlation modelling and only few empirical results are known about high frequency correlation. Depending on the time scale under consideration, a plausible candidate for modelling correlation should:

- at high frequency: reproduce the Epps effect [1], take into account lead-lag relationships between assets [2];
- at the daily scale: avoid purely Gaussian correlations [3].

We develop a theoretical framework based on correlated point processes in order to capture the Epps effect in Sect. 1. We show in Sect. 2 that this model converges to correlated Brownian motions when moving to large time scales. A way of introducing non-Gaussian correlations is also discussed in Sect. 2. We conclude by addressing the limits of this model and further research on high frequency correlation.

2 A Model for High Frequency Correlation

In this section, we start by reviewing the most famous empirical fact about high frequency correlation, namely the Epps effect. Then we suggest a theoretical framework that captures this salient feature of high frequency data.

Nicolas Huth, Frédéric Abergel
Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
Natixis CIB, Paris
e-mail: nicolas.huth@ecp.fr; frederic.abergel@ecp.fr

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

2.1 Empirical Fact: the Epps Effect

In 1979, T.W. Epps [1] observed that, in his own words:

> Correlations among price changes [...] are found to decrease with the length of the interval for which the price changes are measured.

This result was then recovered on more recent data and on several markets [4–7]. Two reasons for the Epps effect were advocated in the literature:

- Market orders on two assets are asynchronous: as $\Delta t \to 0$

$$P\left(\left(\left\{X_{i\Delta t} - X_{(i-1)\Delta t}\right) \neq 0\right\} \cap \left\{\left(Y_{i\Delta t} - Y_{(i-1)\Delta t}\right) \neq 0\right\}\right) \to 0 \qquad (1)$$

  so that $\sum_i (X_{i\Delta t} - X_{(i-1)\Delta t})(Y_{i\Delta t} - Y_{(i-1)\Delta t}) \to 0$, where $X_{i\Delta t}$ and $Y_{i\Delta t}$ are the prices of two assets recorded at time $i\Delta t$ (Δt being the sampling period) with any scheme of interpolation (previous-tick, linear, etc...). Indeed, when sampling prices at very high frequency, it is highly unlikely that both assets will experience a price jump.

- Information needs a human time scale to be processed [2]. On financial markets, some assets, called the leaders, which are often the most liquid, incorporate information onto their prices faster than others, called the laggers. As a result, when dealing with two assets that exhibit a so-called lead-lag relationship, there will be only partial correlation at timescales shorter than the characteristic lead-lag time.

Fig. 1 illustrates the Epps effect on French high frequency data. The stocks studied are BNPP.PA, SOGN.PA, RENA.FP, VLOF.PA, LVMH.PA, LYOE.PA during the time period from 2008-02-04 to 2008-03-20. Clearly, the correlation is almost nil when sampling at few seconds of trading. The correlation converges to an asymptotic level after about half an hour of trading.

As a testimony that the Epps effect is not only due to asynchrony, Fig. 2 plots the empirical probability of having both assets jumping in a time window of a given length for the same three pairs of stocks. Obviously, this probability is quite small for few seconds. It is noteworthy that the characteristic time of this probability is not only smaller than the Epps effect one but is also the same for all the three pairs of stocks. If asynchrony was to be the sole reason for the Epps effect, then all the Epps curves should converge after about ten minutes.

2.2 Correlated Point Processes as a Model for Correlation

Because of the existence of a minimal price change on markets, called the tick, prices are closer to point processes rather than diffusions, which is the standard assumption for daily data. When looking at several assets, a natural question that

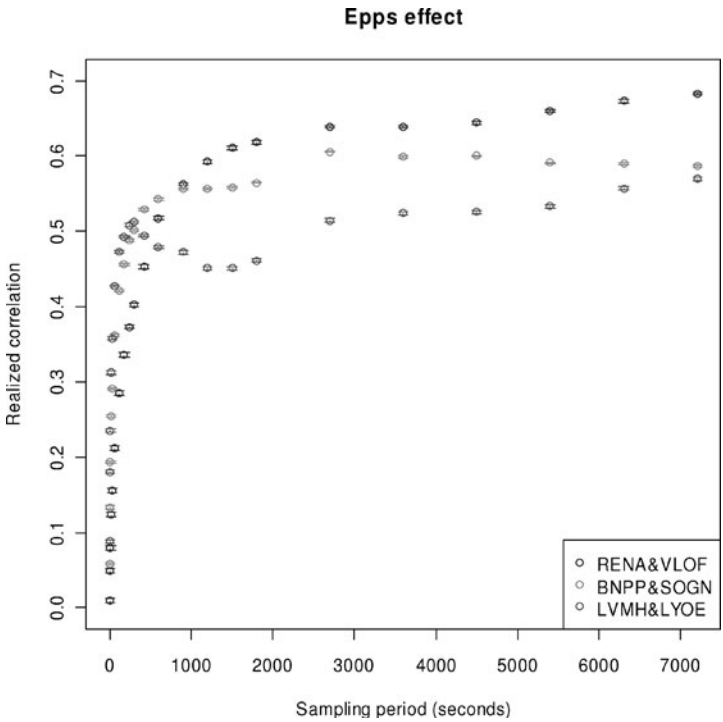


Fig. 1 Correlation coefficient for three pairs of French stocks as a function of the sampling period of price changes

arises is therefore how to correlate point processes. As a point process is fully described by either the associated counting process or the intensity process, there are basically three ways of addressing the issue of correlation, that we explain in Table 1.

Table 1 Four ways of correlating two point processes

	N_2	λ_2
N_1	sync. jumps (ex: $N_2 = N_1 + N_0$)	Hawkes processes $\lambda_2 = f(N_1, \ldots)$
λ_1		async. jumps λ's must be random

From the microstructure of markets point of view, the upper left solution is unrealistic because it is highly unlikely that two market orders on two different assets are executed at the very same time. Therefore, the most appropriate solution is to use correlated stochastic intensities. The upper right solution, i.e. Hawkes processes,

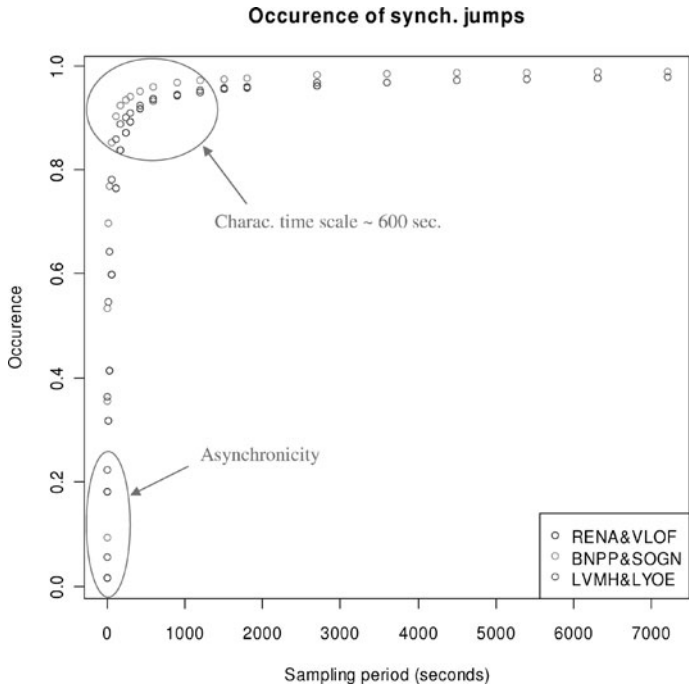


Fig. 2 Empirical probability of having both prices jumping in a time window of a given length as a function of this length for three pairs of French stocks

can be interpreted as particular choice for the correlation structure: past events on one asset tend to trigger (or inhibit) events on the other asset, so that correlation is endogenous.

Let us consider the following model:

$$\mathrm{d}P_t^i = \Delta P^i \left(\mathrm{d}N_t^{i,+} - \mathrm{d}N_t^{i,-}\right) \tag{2}$$

where

- P_t^i is the price of asset i at time t. It can be all kind of prices: last traded price, midprice, best limit, etc…
- ΔP^i is the jump size for asset i. It is assumed to be constant for the sake of simplicity.
- $N_t^{i,+}$ (resp. $N_t^{i,-}$) is the number of market orders that triggered upward (resp. downward) price changes for asset i up to time t.

This model is a pure price model in the sense that it does not explicitly take into account the dynamics of the order book, which impacts ΔP^i. In this model, we can explicitly compute the Epps curve[1]:

[1] See Appendix A for the proof.

$$\rho_{\Delta t} = Corr\left(\mathrm{d}P^1_{\Delta t}, \mathrm{d}P^2_{\Delta t}\right) = \frac{a}{\sqrt{b_0 + \frac{b_1}{\Delta t} + \frac{b_2}{\Delta t^2}}} \tag{3}$$

where

$$\begin{aligned}
a &= Cov\left(\lambda^{1,+} - \lambda^{1,-}, \lambda^{2,+} - \lambda^{2,-}\right)\\
b_0 &= Var\left(\lambda^{1,+} - \lambda^{1,-}\right) Var\left(\lambda^{2,+} - \lambda^{2,-}\right)\\
b_1 &= E\left(\lambda^{1,+} + \lambda^{1,-}\right) Var\left(\lambda^{2,+} - \lambda^{2,-}\right) + E\left(\lambda^{2,+} + \lambda^{2,-}\right) Var\left(\lambda^{1,+} - \lambda^{1,-}\right)\\
b_2 &= E\left(\lambda^{1,+} + \lambda^{1,-}\right) E\left(\lambda^{2,+} + \lambda^{2,-}\right).
\end{aligned}$$

The features of $\rho_{\Delta t}$ are:

- $\rho_{\Delta t} \to 0$ as $\Delta t \to 0$,
- $\rho_{\Delta t} \to \frac{a}{\sqrt{b_0}} = Corr\left(\lambda^{1,+} - \lambda^{1,-}, \lambda^{2,+} - \lambda^{2,-}\right)$ as $\Delta t \to +\infty$,
- $\rho_{\Delta t}$ is increasing and concave,

which reproduce most of the curves plotted on Fig. 1. In the case of time-dependent intensities $\left(\lambda^{1,\pm}_t, \lambda^{2,\pm}_t, t \geq 0\right)$, we get

$$\rho_{\Delta t} = Corr\left(\mathrm{d}P^1_{\Delta t}, \mathrm{d}P^2_{\Delta t}\right) = \frac{a_{\Delta t}}{\sqrt{b_{0,\Delta t} + \frac{b_{1,\Delta t}}{\Delta t} + \frac{b_{2,\Delta t}}{\Delta t^2}}} \tag{4}$$

$$\begin{aligned}
a_{\Delta t} &= Cov\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{1,+}_s - \lambda^{1,-}_s\right)\mathrm{d}s, \frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{2,+}_s - \lambda^{2,-}_s\right)\mathrm{d}s\right)\\
b_{0,\Delta t} &= Var\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{1,+}_s - \lambda^{1,-}_s\right)\mathrm{d}s\right) Var\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{2,+}_s - \lambda^{2,-}_s\right)\mathrm{d}s\right)\\
b_{1,\Delta t} &= E\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{1,+}_s + \lambda^{1,-}_s\right)\mathrm{d}s\right) Var\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{2,+}_s - \lambda^{2,-}_s\right)\mathrm{d}s\right)\\
&\quad + E\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{2,+}_s + \lambda^{2,-}_s\right)\mathrm{d}s\right) Var\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{1,+}_s - \lambda^{1,-}_s\right)\mathrm{d}s\right)\\
b_{2,\Delta t} &= E\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{1,+}_s + \lambda^{1,-}_s\right)\mathrm{d}s\right) E\left(\frac{1}{\Delta t}\int_0^{\Delta t}\left(\lambda^{2,+}_s + \lambda^{2,-}_s\right)\mathrm{d}s\right).
\end{aligned}$$

In particular, the use of time-dependent intensities allows for non-monotonous shapes of $\rho_{\Delta t}$.

Fig. 3 shows least squares fits of the empirical curves with Eq. 3. The last fit is not as accurate as the two firsts because the Epps curve is not monotonous. This is the typical case where time-dependent intensities are needed to achieve a better fit.

The least squares fit is not really natural and can not allow us to decide whether our model is sophisticated enough to reproduce market mechanisms or not. We should rather use maximum likelihood or moments estimation. If our model is right, then the asymptotic value of the correlation should be equal to $Corr\left(\lambda^{1,+}-\lambda^{1,-}, \lambda^{2,+}-\lambda^{2,-}\right)$, that we approximate by the empirical correlation of the imbalances of upwards/downwards moves $Corr\left(N^{1,+}-N^{1,-}, N^{2,+}-N^{2,-}\right)$. We also compute the correlation of the imbalances of buy/sell market orders and the correlation of the total number of trades $Corr\left(N^{1,+}+N^{1,-}, N^{2,+}+N^{2,-}\right)$ for comparison. Fig. 4 shows how these quantities behave.

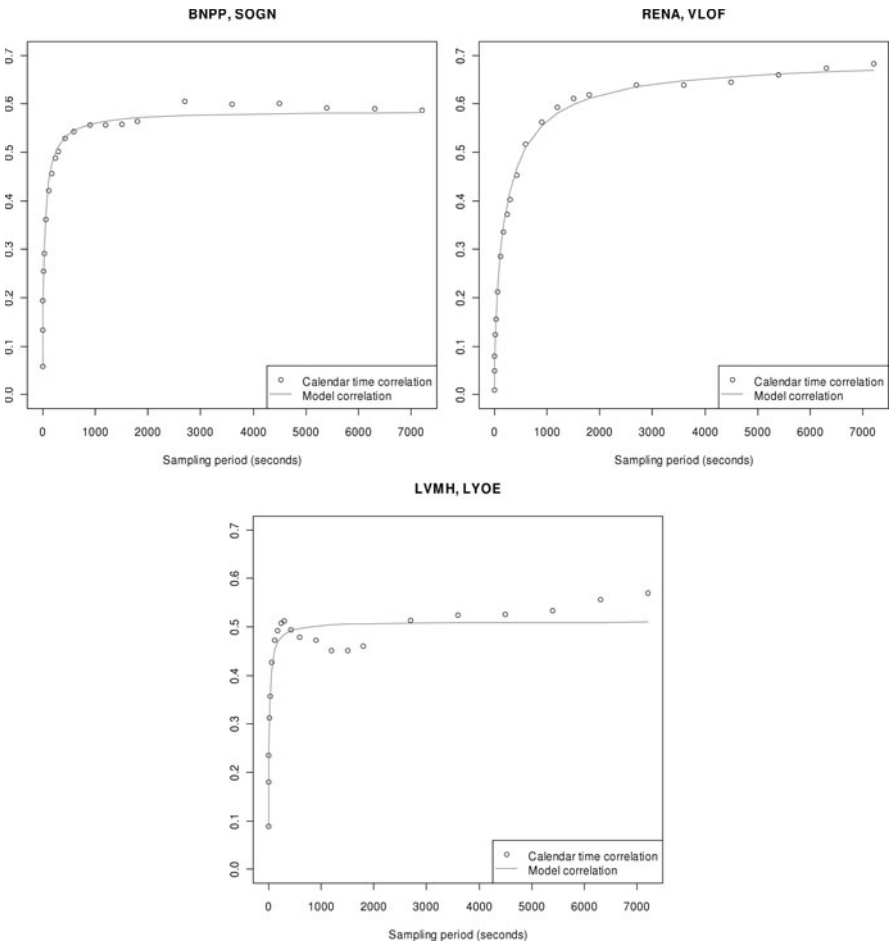


Fig. 3 Least squares fit of Eq. 3 for three pairs of French stocks

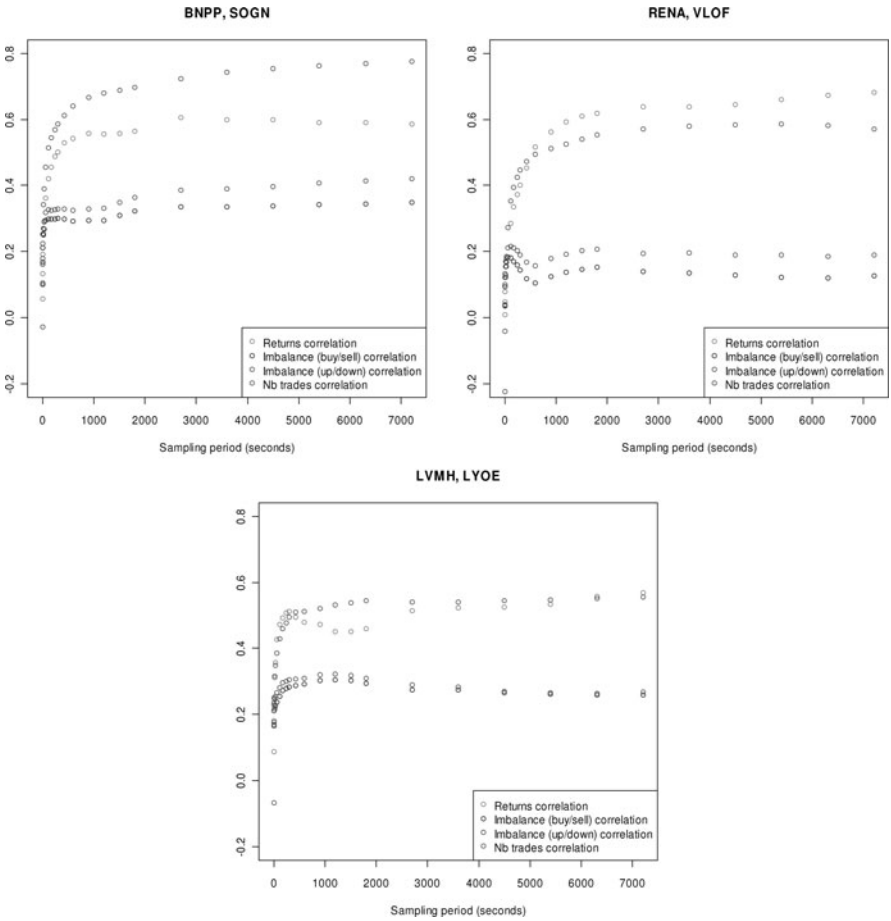


Fig. 4 Comparison between various correlations for three pairs of French stocks

The asymptotic level of the empirical correlation is rather far from what the model predicts, i.e. the upwards/downwards imbalance correlation. So there might be other market mechanisms that are important to take into account. In particular, it seems that the correlation of returns is close to the correlation of the total number of trades $Corr\big(N^{1,+} + N^{1,-}, N^{2,+} + N^{2,-}\big)$, which is a good proxy for the correlation of the variances of returns $Corr\big(\big(\mathrm{d}P^1\big)^2, \big(\mathrm{d}P^2\big)^2\big)$.

3 Large Scale Limit

So far we have been concerned with high frequency correlation modelling. We might be interested in how a model behaves after a long time, a day, say. More stylized

facts are known about daily time series [3]. In particular, we would like our model on a daily scale to

- exhibit a diffusive behavior: $Var\,(\mathrm{d}P_{\Delta t}) \propto \Delta t$;
- display non-Gaussian tails for the distribution of returns;
- reproduce volatility clustering;
- allow for a non-Gaussian correlation structure, such as extreme correlations.

In order to study the large scale limit of the point process model, let us consider the special case of Hawkes processes, which was introduced in [8]

$$\mathrm{d}P_t^{i,(\delta)} = \Delta P^i \sqrt{\delta}\left(\mathrm{d}N_{t/\delta}^{i,+} - \mathrm{d}N_{t/\delta}^{i,-}\right)$$
$$\lambda_t^{i,\pm} = \mu + \int_0^t \phi(t-s)\mathrm{d}N_s^{i,\mp} + \int_0^t \psi(t-s)\mathrm{d}N_s^{j,\pm}$$

with the stationarity constraint that the spectral radius of the matrix of the L^1-norm of regression kernels is strictly less than one. The large scale limit is reached as $\delta \to 0$. It can be shown[2] that this model converges to correlated Brownian motions

$$\mathrm{d}P_t^{i,0} = \sigma.\Delta P_i.C\left(\mathrm{d}B_t^i + D.\mathrm{d}B_t^j\right)$$
$$\sigma := \sqrt{\frac{2\mu}{1-(\|\phi\|+\|\psi\|)}}$$
$$C := \frac{1+\|\phi\|}{(1+\|\phi\|)^2 - \|\psi\|^2}$$
$$D := \frac{\|\psi\|}{1+\|\phi\|}$$

where (B^1, B^2) is a standard bi-dimensional Brownian motion. So this model achieves a diffusive behaviour but fails to reproduce non-Gaussian tails and non-Gaussian correlation and volatility clustering.

We suggest a way of making the correlation structure more complex by introducing a common exogenous noise in the dynamics of the intensities

$$\mathrm{d}P_t^{i,\delta} = \Delta P_i \sqrt{\delta}\left(\mathrm{d}N_{t/\delta}^{i,+} - \mathrm{d}N_{t/\delta}^{i,-}\right)$$
$$\lambda_t^{i,\pm} = \mu + \int_0^t \phi(t-s)\mathrm{d}N_s^{i,\mp} + \int_0^t \psi(t-s)\mathrm{d}N_s^{j,\pm} + M_t^{\pm}.$$

[2] See Appendix B for the proof.

Then the diffusive limit reads

$$
\begin{aligned}
\mathrm{d}P_t^{i,0} &= \Delta P_i.C\left(\sigma\left(\mathrm{d}B_t^i + D.\mathrm{d}B_t^j\right) + E.\mathrm{d}M_t^0\right)\\
\sigma &:= \sqrt{\frac{2\mu}{1-(\|\phi\|+\|\psi\|)}}\\
C &:= \frac{1+\|\phi\|}{(1+\|\phi\|)^2-\|\psi\|^2}\\
D &:= \frac{\|\psi\|}{1+\|\phi\|}\\
E &:= \frac{1+\|\phi\|+\|\psi\|}{1+\|\phi\|}\\
M_t^0 &:= M_t^{0,+} - M_t^{0,-} := \lim_{\delta\to 0}\sqrt{\delta}\int_0^{t/\delta}\left(M_s^+ - M_s^-\right)\mathrm{d}s.
\end{aligned}
$$

As a result, the correlation coefficient reads

$$
\rho = \frac{2\sigma^2 D + E^2\frac{\mathrm{d}\langle M^0\rangle_t}{\mathrm{d}t}}{\sigma^2\left(1+D^2\right) + E^2\frac{\mathrm{d}\langle M^0\rangle_t}{\mathrm{d}t}}
$$

which is stochastic as long as $\frac{\mathrm{d}\langle M^0\rangle_t}{\mathrm{d}t}$ is[3]. For instance, M^0 can be interpreted as a market driver which is common to all stocks. Therefore the correlation depends on the market volatility. Indeed, many studies on daily correlation matrices tend to show that the largest part of the spectra of the correlation is explained by a market factor, rather than by direct interactions between stocks.

4 Conclusions and Further Research

We studied a framework for high frequency correlation modelling based on point processes. This model exhibits a correlation structure that depends on the time scale, in agreement with the Epps effect. However, it seems that there are still market mechanisms to be included to make it more in agreement with real data.

There are still open questions regarding high frequency correlation such as lead-lag estimation [9, 10] and modelling. Furthermore, the role played by the order book shape in the dynamics of correlation has never been studied. We are also investigating ways of achieving better fits of empirical data by adding new parameters in the dynamics of the intensities of market orders, such as an exogenous noise.

[3] The brackets stand for the quadratic variation of a stochastic process.

Appendix A

Standard computations show that

$$\begin{aligned} Cov\left(\mathrm{d}P^1_{\Delta t}, \mathrm{d}P^2_{\Delta t}\right) &= \Delta P^1 \Delta P^2 \left(Cov\left(N^{1,+}_{\Delta t}, N^{2,+}_{\Delta t}\right) + Cov\left(N^{1,-}_{\Delta t}, N^{2,-}_{\Delta t}\right)\right) \\ &\quad - Cov\left(N^{1,+}_{\Delta t}, N^{2,-}_{\Delta t}\right) - Cov\left(N^{1,-}_{\Delta t}, N^{2,+}_{\Delta t}\right)\Big) \\ Var\left(\mathrm{d}P^i_{\Delta t}\right) &= \left(\Delta P^i\right)^2 \left(Var\left(N^{i,+}_{\Delta t}\right) + Var\left(N^{i,-}_{\Delta t}\right) - 2Cov\left(N^{i,+}_{\Delta t}, N^{i,-}_{\Delta t}\right)\right) \end{aligned}$$

and

$$\begin{aligned} Cov\left(N^{1,a}_{\Delta t}, N^{2,b}_{\Delta t}\right) &= E\left(Cov\left(N^{1,a}_{\Delta t}, N^{2,b}_{\Delta t}\right)|\lambda^{1,a}, \lambda^{2,b}\right) \\ &\quad + Cov\left(E\left(N^{1,a}_{\Delta t}|\lambda^{1,a}\right), E\left(N^{2,b}_{\Delta t}|\lambda^{2,b}\right)\right) \\ &= 0 + \Delta t^2.Cov\left(\lambda^{1,a}, \lambda^{2,b}\right) \end{aligned}$$

for $a, b \in \{+, -\}$ and

$$Var\left(N^{i,\pm}_{\Delta t}\right) = \Delta t.E\left(\lambda^{i,\pm}\right) + \Delta t^2.Var\left(\lambda^{i,\pm}\right).$$

Finally, we get

$$Corr\left(\mathrm{d}P^1_{\Delta t}, \mathrm{d}P^2_{\Delta t}\right) = \frac{a}{\sqrt{b_0 + \frac{b_1}{\Delta t} + \frac{b_2}{\Delta t^2}}}$$

where

$$\begin{aligned} a &= Cov\left(\lambda^{1,+} - \lambda^{1,-}, \lambda^{2,+} - \lambda^{2,-}\right) \\ b_0 &= Var\left(\lambda^{1,+} - \lambda^{1,-}\right) Var\left(\lambda^{2,+} - \lambda^{2,-}\right) \\ b_1 &= E\left(\lambda^{1,+} + \lambda^{1,-}\right) Var\left(\lambda^{2,+} - \lambda^{2,-}\right) + E\left(\lambda^{2,+} + \lambda^{2,-}\right) Var\left(\lambda^{1,+} - \lambda^{1,-}\right) \\ b_2 &= E\left(\lambda^{1,+} + \lambda^{1,-}\right) E\left(\lambda^{2,+} + \lambda^{2,-}\right) \end{aligned}$$

so that

$$\begin{aligned} \lim_{\Delta t \to 0} Corr\left(\mathrm{d}P^1_{\Delta t}, \mathrm{d}P^2_{\Delta t}\right) &= 0 \\ \lim_{\Delta t \to +\infty} Corr\left(\mathrm{d}P^1_{\Delta t}, \mathrm{d}P^2_{\Delta t}\right) &= \frac{a}{\sqrt{b_0}} = Corr\left(\lambda^{1,+} - \lambda^{1,-}, \lambda^{2,+} - \lambda^{2,-}\right). \end{aligned}$$

In the case of time-dependent intensities, the same line of computing goes except that λ's have to be replaced by their time average $\frac{1}{\Delta t}\int_t^{t+\Delta t} \lambda(s)\mathrm{d}s$.

Appendix B

Let us consider the following Hawkes model

$$\mathrm{d}P_t = \Delta P \left(\mathrm{d}N_t^+ - \mathrm{d}N_t^-\right)$$
$$\lambda_t^{\pm} = \mu + \int_0^t \phi(t-s)\mathrm{d}N_s^{\mp}$$

where $\Delta P \in R^+$ and $\|\phi\| < 1$. We are looking for the diffusive limit of this model. Therefore, we introduce the rescaled model

$$\mathrm{d}P_t^\delta = \Delta P \sqrt{\delta}\left(\mathrm{d}N_{t/\delta}^+ - \mathrm{d}N_{t/\delta}^-\right).$$

We are interested in the limit $\delta \to 0$. Let us split the price into two parts: the martingale and the compensator

$$\mathrm{d}P_t^\delta = \mathrm{d}M_t^\delta + \mathrm{d}C_t^\delta$$
$$M_t^\delta = \Delta P \sqrt{\delta}\left(\left(N_{t/\delta}^+ - N_{t/\delta}^-\right) - \int_0^{t/\delta}\left(\lambda_s^+ - \lambda_s^-\right)\mathrm{d}s\right)$$
$$C_t^\delta = \Delta P \sqrt{\delta}\int_0^{t/\delta}\left(\lambda_s^+ - \lambda_s^-\right)\mathrm{d}s.$$

Regarding the compensator, we have

$$C_t^\delta = \Delta P \sqrt{\delta}\int_0^{t/\delta}\left(\lambda_s^+ - \lambda_s^-\right)\mathrm{d}s$$
$$= -\|\phi\|\int_0^t \mathrm{d}P_u^\delta \int_0^{\frac{t-u}{\delta}} \frac{\phi(x)}{\|\phi\|}\mathrm{d}x \to -\|\phi\|\,\mathrm{d}P_t^0.$$

Let us compute the quadratic variation of the martingale part

$$\frac{\left\langle M^\delta\right\rangle_t}{\Delta P^2} = \delta\int_0^{t/\delta}\left(\lambda_s^+ + \lambda_s^-\right)\mathrm{d}s$$

$$
\begin{aligned}
&= 2\mu t + \delta \int_0^{t/\delta} \left(\mathrm{d}N_u^+ + \mathrm{d}N_u^-\right) \int_0^{t/\delta - u} \phi(x)\mathrm{d}x \\
&= 2\mu t + \delta \int_0^{t/\delta} \mathrm{d}N_u^+ - \lambda_u^+ \mathrm{d}u + \mathrm{d}N_u^- - \lambda_u^- \mathrm{d}u \int_0^{t/\delta - u} \phi(x)\mathrm{d}x \\
&\quad + \delta \int_0^{t/\delta} \left(\lambda_u^+ + \lambda_u^-\right) \mathrm{d}u \int_0^{t/\delta - u} \phi(x)\mathrm{d}x \\
&\to 2\mu t \sum_{n \geq 0} \|\phi\|^n = \left(\frac{2\mu}{1 - \|\phi\|}\right) t
\end{aligned}
$$

since

$$
\begin{aligned}
&\lim_{\delta \to 0} \delta \int_0^{t/\delta} \left(\left(\mathrm{d}N_u^+ - \lambda_u^+ \mathrm{d}u\right) + \left(\mathrm{d}N_u^- - \lambda_u^- \mathrm{d}u\right)\right) \int_0^{t/\delta - u} \phi(x)\mathrm{d}x \\
&= \lim_{\delta \to 0} \sqrt{\delta}\, \|\phi\| \left(B_t^+ + B_t^-\right) = 0.
\end{aligned}
$$

Therefore, we get $M_t^\delta \to \sqrt{\frac{2\mu}{1-\|\phi\|}}\,(\Delta P)\, B_t$ by using the following lemma.

Lemma 1 *Let M be a local martingale and $< M >_\infty := \lim_{t \to +\infty} < M >_t$. Then*

$$
E\left(\sup_{t \geq 0} M_t^2\right) \leq 4.E(< M >_\infty).
$$

If these two quantities are finite, then M is a martingale which converges a.s. and in L^2 towards a random variable M_∞ as $t \to +\infty$.

Finally, the diffusive limit reads

$$
\mathrm{d}P_t^0 = \sqrt{\frac{2\mu}{1 - \|\phi\|}} \frac{\Delta P}{1 + \|\phi\|} \mathrm{d}B_t.
$$

In the case of a bivariate price model

$$
\begin{aligned}
\mathrm{d}P_t^{i,\delta} &= \Delta P_i \sqrt{\delta} \left(\mathrm{d}N_{t/\delta}^{i,+} - \mathrm{d}N_{t/\delta}^{i,-}\right) \\
\lambda_t^{i,\pm} &= \mu + \int_0^t \phi(t-s)\mathrm{d}N_s^{i,\mp} + \int_0^t \psi(t-s)\mathrm{d}N_s^{j,\pm}.
\end{aligned}
$$

The very same line of proof yields

$$\begin{aligned}
\mathrm{d}P_t^{i,0} &= \sigma.\Delta P_i.C\left(\mathrm{d}B_t^i + D.\mathrm{d}B_t^j\right)\\
\sigma &:= \sqrt{\frac{2\mu}{1-(\|\phi\|+\|\psi\|)}}\\
C &:= \frac{1+\|\phi\|}{(1+\|\phi\|)^2-\|\psi\|^2}\\
D &:= \frac{\|\psi\|}{1+\|\phi\|}
\end{aligned}$$

so that the correlation coefficient of the two assets is constant

$$\rho = \frac{\left\langle P_t^{1,0}, P^{2,0}\right\rangle_t}{\sqrt{\langle P^{1,0}\rangle_t \langle P^{2,0}\rangle_t}} = \frac{2D}{1+D^2}.$$

Finally, we introduce an exogenous noise in the dynamics of the intensities

$$\begin{aligned}
\mathrm{d}P_t^{i,\delta} &= \Delta P_i\sqrt{\delta}\left(\mathrm{d}N_{t/\delta}^{i,+} - \mathrm{d}N_{t/\delta}^{i,-}\right)\\
\lambda_t^{i,\pm} &= \mu + \int_0^t \phi(t-s)\mathrm{d}N_s^{i,\mp} + \int_0^t \psi(t-s)\mathrm{d}N_s^{j,\pm} + M_t^{\pm}.
\end{aligned}$$

We assume that there exists two random processes $\left(M_t^{0,\pm}, t \geq 0\right)$ such that

$$\lim_{\delta\to 0}\sqrt{\delta}\int_0^{t/\delta} M_s^{\pm}\mathrm{d}s = M_t^{0,\pm}$$

and $\left\langle M^{0,\pm}\right\rangle_t < \infty$. Then the diffusive limit reads

$$\begin{aligned}
\mathrm{d}P_t^{i,0} &= \Delta P_i.C\left(\sigma\left(\mathrm{d}B_t^i + D.\mathrm{d}B_t^j\right) + E.\mathrm{d}M_t^0\right)\\
\sigma &:= \sqrt{\frac{2\mu}{1-(\|\phi\|+\|\psi\|)}}\\
C &:= \frac{1+\|\phi\|}{(1+\|\phi\|)^2-\|\psi\|^2}\\
D &:= \frac{\|\psi\|}{1+\|\phi\|}\\
E &:= \frac{1+\|\phi\|+\|\psi\|}{1+\|\phi\|}\\
M_t^0 &:= M_t^{0,+} - M_t^{0,-}.
\end{aligned}$$

References

1. Epps T. W. (1979) Comovements in Stock Prices in the Very Short-Run. Journal of the American Statistical Association 74, 291–298
2. Toth B., Kertesz J. (2009) The Epps Effect Revisited, Quantitative Finance 9(7), 793–802
3. Bouchaud J.-P., Potters M. (2004) Theory of Financial Risk and Derivative Pricing, From Statistical Physics to Risk Management. Cambridge University Press
4. Reno R. (2003) A Closer Look at the Epps Effect. International Journal of Theoretical and Applied Finance 6: 87–102
5. Toth B., Kertesz J. (2006) Increasing Market Efficiency: Evolution of Cross-Correlations of Stock Returns, Physica A 360, 505–515
6. Iori G., Precup O.V. (2006) Cross-correlation Measures in the High Frequency Domain, Working Paper
7. Toth B., Kertesz J. (2007) On the origin of the Epps Effect, Physica A 383(1), 54–58
8. Bacry E. (2010) Modeling microstructure noise using point processes, Econophys- Kolkata V Conference, submitted to Quantitative finance papers
9. Robert C. Y., Rosenbaum M. (2009) On the limiting spectral distribution of the covariance matrices of time-lagged processes, to appear in Journal of Multivariate Analysis
10. Robert C. Y., Rosenbaum M., Hoffman M., Yoshida N. (2010) Estimation of the lead-lag parameter from non-synchronous data, Working Paper

The Model with Uncertainty Zones for Ultra High Frequency Prices and Durations: Applications to Statistical Estimation and Mathematical Finance

Christian Y. Robert and Mathieu Rosenbaum

1 Introduction

The goal of this note is to describe a model for ultra high frequency prices and durations, the model with uncertainty zones developed in [27]. We also give some results from [28] and [29] which show how it can be used in practice for statistical estimation or in order to hedge derivatives. Before introducing this model, we briefly recall the classical approaches of price modelling in the so-called microstructure noise literature.

1.1 Microstructure Noise

Since the publication of the celebrated Black-Scholes article [6], continuous time processes have become usual modelling tools in mathematical finance. Among these processes, semi-martingales[1] have a major role. Indeed, it is well known that the "no free lunch" assumption is essentially only compatible with semi-martingale type price dynamics, see [9]. However, since the middle of the nineties and the massive availability of high frequency financial data (data recorded every second or even millisecond), numerous empirical studies have shown that, over a short time period, it is not reasonable to assume that prices are observations from a semi-martingale, see for example [1, 16]. Of course, one obvious reason for this is prices discreteness (see Sects. 1.2 and 1.3). Thus, series of high frequency prices are not of the same

Christian Y. Robert
CREST – ENSAE Paris Tech, Paris,
e-mail: chrobert@ensae.fr

Mathieu Rosenbaum
CMAP, École Polytechnique,
e-mail: mathieu.rosenbaum@polytechnique.edu

[1] To fix ideas, we have in mind in this note the case where these semi-martingales are continuous.

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

nature as series of low frequency prices. Said differently, the scale invariance property of Brownian type dynamics is only satisfied below some sampling frequency. This phenomenon is called "microstructure noise".

1.2 Additive Microstructure Noise

Assume here our time horizon is [0, 1]. A way of modelling prices enabling to take into account this microstructure noise and to stay consistent with the mathematical finance theory is the "additive microstructure noise approach". This means that one simply considers that the price P_t observed at time $t = i/n,\ i = 0, \ldots, n$ is of the form

$$\log(P_{i/n}) = \log(X_{i/n}) + \varepsilon_i^n, \tag{1}$$

where X is a continuous-time semi-martingale which represents an efficient price and ε_i^n is a microstructure noise term. Of course, one can always write Decomposition (1). The term "additive" means that in this approach, the statistician essentially focuses on the properties of the noise process $\varepsilon_t = P_t - X_t$. This kind of model is probably the most simple way to obtain:

- data close to data from a continuous-time semi-martingale in the low frequencies: $P_{i/n} - P_{j/n} \approx X_{i/n} - X_{j/n}$ as soon as $(i - j)/n$ is large enough;
- data very different from those of a continuous-time semi-martingale in the high frequencies, the noise term becoming important at these scales.

In the framework of model (1), simple considerations enable to deduce elementary desirable properties for the observed price and the microstructure noise. First, recall that

(i) Observed prices are discrete.

Indeed, market prices are on a tick grid. This is essentially incompatible with observations of a semi-martingale at exogenous sampling times. Remark that the important impact of this discretization effect on various statistical procedures has been emphasized since the middle of the eighties, see for example [2, 8, 14]. Let us also recall a well known stylized fact:

(ii)We observe quick oscillations of the transaction price between two values ("bid-ask bounce").

Consider now the quadratic variation of $(\log X_t)_{t\geq 0}$ and the covariation between $(\varepsilon_t)_{t\geq 0}$ et $(\log X_t)_{t\geq 0}$. These quantities are respectively defined by the following formulae (provided they make sense):

$$[\log X]_t = \mathbb{P}\text{-}\lim_{n\to\infty} \sum_{j=1}^{n} \big(\log(X_{s_{j,n}}) - \log(X_{s_{j-1,n}})\big)^2,$$

$$[\varepsilon, \log X]_t = \mathbb{P}\text{-}\lim_{n\to\infty} \sum_{j=1}^{n} (\varepsilon_{s_{j,n}} - \varepsilon_{s_{j-1,n}})\big(\log(X_{s_{j,n}}) - \log(X_{s_{j-1,n}})\big),$$

for any deterministic sequence of partitions $(0 = s_{0,n} < s_{1,n} < \ldots < s_{n,n} = t)$ such that $\sup_j \{s_{j+1,n} - s_{j,n}\}$ tends to zero as n goes to infinity. Assuming that $(P_t)_{t \geq 0}$ is a finite activity jump process, we derive

$$[\log P, \log X]_t = [\log X]_t + [\varepsilon, \log X]_t = 0$$

and

$$[\varepsilon]_t = [\log P]_t - [\log X]_t - 2[\varepsilon, \log X]_t = [\log P]_t + [\log X]_t.$$

It follows that:

(iii) $[\varepsilon]_t$ is almost surely finite.

The standard approach of microstructure modelling considers an essentially iid microstructure noise, independent of the efficient price, with a law which does not depend on the sampling frequency n. These kind of noise is, by extension, called additive microstructure noise. The models with additive microstructure noise have been largely used, notably in order to build estimating procedures for the integrated volatility of the efficient price, see in particular [3, 4, 13, 33, 35, 36]. Note that these models have been extended in several directions: heteroskedastic noise, correlated with the efficient price, see [24] or contamination of the efficient price through a complex Markov kernel, see [20, 26].

Standard additive microstructure noise models are convenient for carrying out calculations and are not so unrealistic when sampling returns at some period larger than about five minutes. Nevertheless, they do not satisfy any of the properties (i), (ii), (iii). Indeed, in the ultra high frequencies, errors due to price discreteness and diurnal patterns lead to a non-linear dependence between the microstructure noise and the efficient price and to an intricate heteroskedastic time dependence for the noise. Thus, it is natural to consider models with rounding errors.

1.3 Model with Rounding Errors

In this section, we mention a very simple model allowing to satisfy properties (i), (ii), (iii), together with the assumption of a semi-martingale efficient price. The idea is to focus on properties of the observed price itself, not of the microstructure noise. Thus, it is very natural to consider the model with rounding errors introduced in [10]. In this model, the efficient price is a Brownian diffusion X_t and we observe the sample

$$(X_{i/n}^{(\alpha_n)}, i = 0, \ldots, n),$$

where $X_{i/n}^{(\alpha_n)} = \alpha_n \lfloor X_{i/n}/\alpha_n \rfloor$. Therefore, $X_{i/n}^{(\alpha_n)}$ is the observation of $X_{i/n}$ with rounding error α_n (the tick size), with $\alpha_n \to 0$. In this framework, prices are discrete and behave like a diffusion process in the low frequencies, the rounding effect becoming negligible. Moreover, property (ii) and, for some specifications of α_n, (iii), are satisfied.

It is shown in [30] that one can also build estimator of the integrated volatility of the efficient price process in this model. Interestingly enough, estimators built in the additive microstructure noise case are not robust in the rounding framework when α_n goes to zero too slowly, see [26]. Remark also that beyond the will to reproduce property (iii), the assumption $\alpha_n \to 0$ is necessary from a statistical point of view. Indeed, one can not estimate the integrated volatility if α_n is constant, see [20].

1.4 Towards a Joint Dynamic for Prices and Durations

Assuming the price is a semi-martingale observed with rounding error is not entirely satisfying. First, such model is not realistic in the ultra high frequencies since it produces too many jumps (because the sample path crosses an infinite number of times each level it reaches). Mostly, durations are not modelled, only an exogenous sampling is considered. Therefore, if the practitioner wants to use such model, the following questions arise:

- What is the right sampling frequency to choose: 1 second? 1 minute? 5 minutes?
- What is the right price to choose? Bid price? Mid price? VWAP Price? Last traded price?

In Sect. 2, we explain a model which is relevant and allows to avoid the preceding questions. Statistical applications of this model are given in Sect. 3 and its use for hedging derivatives is explained in Sect. 4.

2 Model with Uncertainty Zones

We consider in [27] a model for both prices and durations, the model with uncertainty zones. This model enables to satisfy (i), (ii), (iii) and the main stylized facts of high frequency data (see [12, 15]), in particular:

- a negative autocorrelation of the returns, which vanishes when the frequency decreases (in calendar time and transaction time);
- intraday seasonalities and specific relations between variables. For example, the U shape of the intraday volatility and an inverse relation between durations and volatility.

Beyond reproducing the empirical features, we insist on the fact that a financial model is really of interest only if the practitioner can use it. We show in Sect. 3 how the model can be used in order to estimate relevant quantities for trading and in Sect. 4 how it enables to compute hedging errors in a microstructure context.

2.1 Description of the Model

We describe now the model with uncertainty zones for the last traded price. In an idealistic framework, where the efficient price would be observed, market partici-

pants would trade when the efficient price crosses the tick grid. In practice, there is some uncertainty about the efficient price value so that market participants are reluctant to price changes. Hence, there is a modification of the transaction price only if some buyers and sellers are truly convinced that the efficient price is sufficiently far from the last traded price. We introduce a parameter η that quantifies the aversion to price changes (with respect to the tick size) of the market participants and propose a model that takes into account this aversion.

Let $(X_t)_{t\geq 0}$ denote the theoretical, efficient price of the asset. On a rich enough filtered probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, we assume that the logarithm of the efficient price is a $\mathscr{F}_t$ -adapted continuous semi-martingale of the form

$$\mathrm{d}\log X_t = a_t \mathrm{d}t + \sigma_{t-}\mathrm{d}W_t,$$

where $(W_t)_{t\geq 0}$ is a standard $\mathscr{F}$-Brownian motion, $(a_t)_{t\geq 0}$ is a progressively measurable process with locally bounded sample paths and $(\sigma_t)_{t\geq 0}$ is a positive $\mathscr{F}_t$-adapted process with càdlàg sample paths.

The tick grid where transaction prices are bound to lie on is defined as $\{k\alpha; k\in\mathbb{N}\}$, with α the tick size. For $k \in \mathbb{N}$ and $0 < \eta < 1$, we define the zone U_k by $U_k = [0, \infty) \times (d_k, u_k)$ with

$$d_k = (k + 1/2 - \eta)\alpha \text{ and } u_k = (k + 1/2 + \eta)\alpha.$$

Thus, U_k is a band around the mid-tick grid value $(k + 1/2)\alpha$, see Fig. 1. Note that when η is smaller than $1/2$, there is no overlap between the zones.

We assume that the transaction price may jump from price $k'\alpha$ to price $k\alpha$ with $k' \neq k$ only once the efficient price exited down the zone U_k or exited up the zone U_{k-1} and provided that market conditions are favorable for a transaction to occur. In a way, the transaction price only changes when the efficient price is close from a new multiple value of α and market participants want to trade. The zones $(U_k)_{k\in\mathbb{N}}$ represent bands inside of which the efficient price can not trigger a change of the transaction price. Consequently, they will be referred to as the uncertainty zones.

More specifically, let us precise the construction of the sequence $(\tau_i)_{i\geq 0}$ of the exit times from the uncertainty zones which will lead to a change in the transaction price. Let $\tau_0 = 0$ and assume without loss of generality that τ_1 is the exit time of $(X_t)_{t\geq 0}$ from the set (d_{k_0-1}, u_{k_0}) where $k_0 = X_0^{(\alpha)}/\alpha$, with $X_0^{(\alpha)}$ the value of X_0 rounded to the nearest multiple of α. We introduce a sequence $(L_i)_{i\geq 1}$ of $\mathscr{F}_{\tau_i}$-measurable discrete random variables which represent the absolute value in number of ticks of the price jump between the i-th and the $(i + 1)$-th transaction leading to a price change. As explained later, the distribution of this variable will depend on the value of some market quantities at time τ_i. Then define recursively τ_{i+1} as the exit time of $(X_t)_{t>\tau_i}$ from the set $(d_{k_i-L_i}, u_{k_i+L_i-1})$, where $k_i = X_{\tau_i}^{(\alpha)}/\alpha$, that is

$$\tau_{i+1} = \inf\left\{t : t > \tau_i, X_t = X_{\tau_i}^{(\alpha)} - \alpha(L_i - \frac{1}{2} + \eta) \text{ or } X_t = X_{\tau_i}^{(\alpha)} + \alpha(L_i - \frac{1}{2} + \eta)\right\}.$$

In particular, if $X_{\tau_i} = d_j$ for some $j \in \mathbb{N}$, τ_{i+1} is the exit time of $(X_t)_{t>\tau_i}$ from the set (d_{j-L_i}, u_{j+L_i-1}), and if $X_{\tau_i} = u_j$ for some $j \in \mathbb{N}$, τ_{i+1} is the exit time of $(X_t)_{t>\tau_i}$ from the set (d_{j-L_i+1}, u_{j+L_i}).

Finally, let $t_0 = 0$ and $P_0 = X_0^{(\alpha)}$. We assume that the couples (t_i, P_{t_i}) satisfy for $i \geq 1$

$$\tau_i \leq t_i < \tau_{i+1} \text{ and } P_{t_i} = X_{\tau_i}^{(\alpha)}.$$

It means that, between τ_i and τ_{i+1}, at least one transaction has occurred at price P_{t_i} and t_i is the time of the first of these transactions. The difference $t_i - \tau_i$ can be viewed as the delay caused by the reaction times of the market participants and/or by the trading process.

We also assume that the jump sizes are bounded (what is empirically not restrictive) and denote by m their maximal value. For $k = 1, ..., m$ and $t > 0$, let

$$N_{\alpha,t,k}^{(a)} = \sum_{t_i \leq t} \mathbb{I}_{\{(P_{t_i} - P_{t_{i-1}})(P_{t_{i-1}} - P_{t_{i-2}}) < 0 \text{ and } |P_{t_i} - P_{t_{i-1}}| = k\alpha\}}$$

and

$$N_{\alpha,t,k}^{(c)} = \sum_{t_i \leq t} \mathbb{I}_{\{(P_{t_i} - P_{t_{i-1}})(P_{t_{i-1}} - P_{t_{i-2}}) > 0 \text{ and } |P_{t_i} - P_{t_{i-1}}| = k\alpha\}}$$

be respectively the number of alternations and continuations of k ticks. An alternation (continuation) of k ticks is a jump of k ticks whose direction is opposite to (the same as) that of the preceding jump, see Fig. 1. We now precise the conditional distribution of the jumps in ticks between consecutive transaction prices. Let $(\chi_t)_{t \geq 0}$ be a $\mathscr{F}_t$-adapted multidimensional continuous Ito semi-martingale with progressively measurable with locally bounded sample paths and positive $\mathscr{F}_t$-adapted volatility matrix whose elements have càdlàg sample paths. We define the filtration $\mathscr{E}$ as the complete right-continuous filtration generated by $(X_t, \chi_t, N_{\alpha,t,k}^{(a)}, N_{\alpha,t,k}^{(c)},\ k = 1, \ldots, m)$. We assume that conditional on $\mathscr{E}_{\tau_i}$, L_i is a discrete random variable on $[\![1, m]\!]$ satisfying

$$\mathbb{P}_{\mathscr{E}_{\tau_i}}[L_i = k] = p_k(\chi_{\tau_i}),\ \ 1 \leq k \leq m, \tag{2}$$

for some unknown positive differentiable with bounded derivative functions p_k. In practice, χ_t may represent quantities related for example to the traded volume, the bid-ask spread, or the bid and ask depths. For the applications, specific form for the p_k are given in [27].

2.2 Discussion

- The model with uncertainty zones accommodates the inherent properties of prices, durations and microstructure noise together with a semi-martingale efficient price. In particular, this model allows for discrete prices, a bid-ask bounce and an inverse relation between durations and volatility. Moreover the usual behaviors of the autocorrelograms and cross correlograms of returns and microstructure noise, both in calendar and tick time, are reproduced. Eventually, it

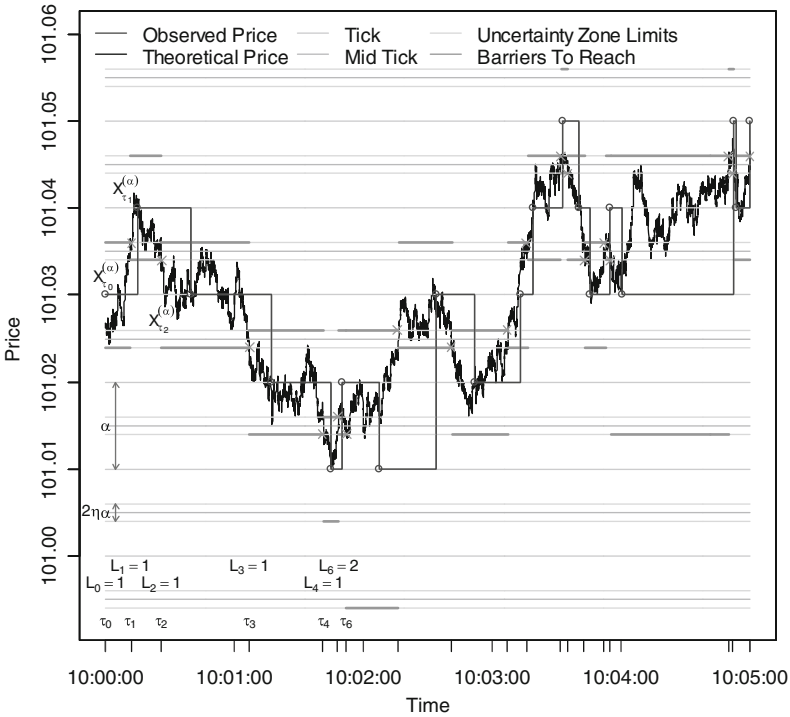


Fig. 1 Example of trajectories of the latent price and of the observed price. The *gray crosses* denote the exit points associated to the τ_i

leads to jumps in the price of several ticks, the size of the jumps being determined by explanatory variables involving for example the order book. Mostly, the model with uncertainty zones is clearly validated on real data. These results are studied in details in [27].

- As explained in the previous section, η quantifies the aversion to price changes (with respect to the tick size) of the market participants. Indeed, η controls the width of the uncertainty zones. In tick unit, the larger η, the farther from the last traded price the efficient price has to be so that a price change occurs. In some sense, a small η ($< 1/2$) means that the tick size appears too large to the market participants and a large η means that the tick size appears too small.
- There are several other ways to interpret the parameter η, notably from a practitioner's perspective. For example, one can think that in the very high frequencies, the order book can not "follow" the efficient price and is reluctant to price changes. This reluctancy could be characterized by η. Another possibility is to view η as a measure of the usual prices depth explored by the transaction volumes.

2.3 Properties of the Last Traded Price and Its Durations

Fig. 2 represents the volatility function used in our simulation together with the durations between price changes. The chosen U-shaped form for the volatility is classical and the durations have a behavior which is in inverse relation to those of the volatility. This reproduces a usual empirical characteristic of high frequency financial data, see for example [12, 15].

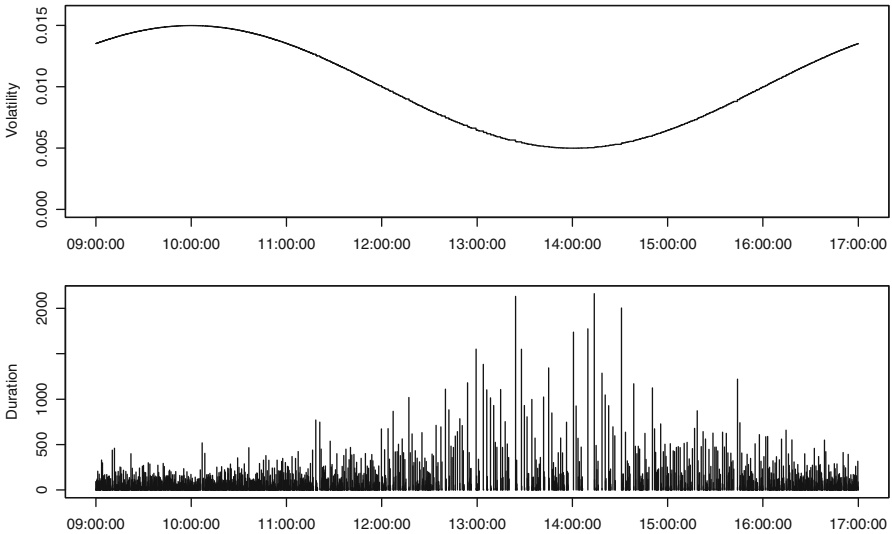


Fig. 2 Volatility trajectory and durations

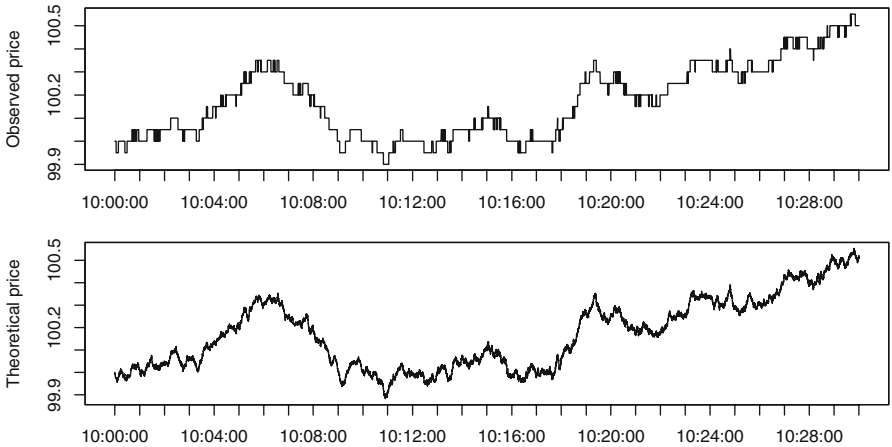


Fig. 3 Sample paths of the last traded and efficient prices

On Fig. 3, we show the sample paths of the last traded and efficient prices over half an hour. The commonly observed numerous and quick oscillations of one tick of the last traded price (bid-ask bounce) are reproduced thanks to the behavior of the semi-martingale efficient price around the uncertainty zones ($\eta < 1/2$).

We finally draw some autocorrelation functions for the logarithmic returns on Fig. 4. Note that the aim of such graphs is just to show that our model reproduces the stylized facts of real data. Indeed, it is shown in [27] that very similar patterns are observed on real data. However, one has to be cautious with their interpretation because of stationary issues. Thanks to the uncertainty zones, we observe in our model the stylized fact of a significative negative first order correlation between the returns for sampling frequencies between one and thirty seconds. Moreover, in tick time, many of the higher order autocorrelations are significant and systematically alternate sign.

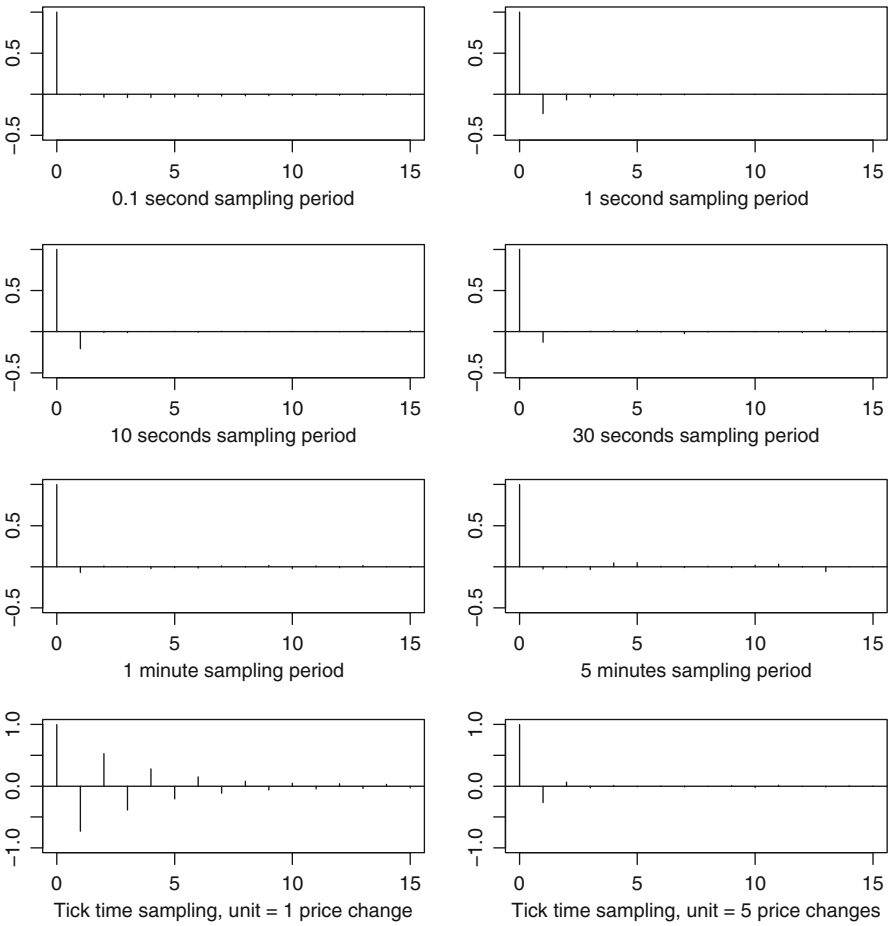


Fig. 4 Autocorrelation functions of the log returns

3 Statistical Procedures

In this section, we show that the model with uncertainty zones is very convenient in order to design statistical procedures.

3.1 Estimation of the Efficient Price

In the model with uncertainty zones, it is extremely easy to build estimators of relevant quantities. Indeed, we have the following nice property:

$$X_{\tau_i} = P_{t_i} - \alpha\left(\frac{1}{2} - \eta\right) sign(P_{t_i} - P_{t_{i-1}}).$$

Thus, we can very simply retrieve the value of the efficient price at time τ_i from P_{t_i}, $P_{t_{i-1}}$ and η. Remark here that the P_{t_i} and $P_{t_{i-1}}$ are the transaction prices at times t_i and t_{i-1}, hence observable. The parameter η can be estimated without difficulty (see next section). Therefore, we can estimate the efficient price at time τ_i the following way:

$$\hat{X}_{\tau_i} = P_{t_i} - \alpha(\frac{1}{2} - \hat{\eta}) sign(P_{t_i} - P_{t_{i-1}}). \tag{3}$$

This relation is very important in practice: we are able to retrieve the efficient price. Thus, any statistical procedure designed for a semi-martingale can be applied in the microstructure context, provided one uses the $\hat{X}_{\tau_i}$ and not the P_{t_i}. Nevertheless, one should be careful. First we do not have the exact values of the prices but only estimations. Mostly, the price is estimated at times τ_i, endogenous stopping times. This is a major difficulty in order to prove properties of estimators. Indeed, the usual theorems essentially assume the observation times are exogenous. Thus they cannot be used in this context.

3.2 Estimation of η

We define the estimator of η by

$$\hat{\eta}_{\alpha,t} = \left(0 \vee \sum_{k=1}^{m} \lambda_{\alpha,t,k} u_{\alpha,t,k}\right) \wedge 1,$$

with

$$\lambda_{\alpha,t,k} = \frac{N^{(a)}_{\alpha,t,k} + N^{(c)}_{\alpha,t,k}}{\sum_{j=1}^{m}\left[N^{(a)}_{\alpha,t,j} + N^{(c)}_{\alpha,t,j}\right]} \text{ and } u_{\alpha,t,k} = \frac{1}{2}\left(k(\frac{N^{(c)}_{\alpha,t,k}}{N^{(a)}_{\alpha,t,k}} - 1) + 1\right).$$

The idea behind this estimator is that the $u_{\alpha,t,k}$ are consistent estimators of η for each k. The $\lambda_{\alpha,t,k}$ are then natural weighting factors. Note in particular that $N^{(c)}_{\alpha,t,1}/N^{(a)}_{\alpha,t,1}$ is an estimator of 2η. Consequently, if η is smaller than $1/2$, we may expect more alternations than continuations in the last traded price and conversely. We assume α tends to zero (which is a natural asymptotic, as in the rounding model of Sect. 1.3). We have the following result, where $\overset{u.c.p.}{\rightarrow}$ denotes uniform convergence in probability over compact sets included in $[0, T]$:

Theorem 3.1 *As α tends to* 0,

$$\hat{\eta}_{\alpha,t} \overset{u.c.p.}{\rightarrow} \eta.$$

Table 1 shows some daily estimations of η on assets of the CAC 40 index, the week of 2007, January 15 (we use the usual abbreviations of the names of the companies). These estimated values are remarkably stable within the week.

Table 1 Measures of η on CAC 40 assets, week of 2007, January 15

TICKER	$\hat{\eta}$ Day 1	$\hat{\eta}$ Day 2	$\hat{\eta}$ Day 3	$\hat{\eta}$ Day 4	$\hat{\eta}$ Day 5
AIRF	0.277	0.273	0.250	0.236	0.250
ALSO	0.189	0.215	0.199	0.199	0.194
BNPP	0.100	0.111	0.146	0.157	0.110
CAGR	0.193	0.242	0.238	0.215	0.209
DANO	0.076	0.110	0.086	0.081	0.110
EAD	0.179	0.243	0.256	0.234	0.227
FTE	0.192	0.221	0.246	0.274	0.192
RENA	0.167	0.274	0.190	0.294	0.279
SGOB	0.088	0.136	0.131	0.129	0.118
TOTF	0.048	0.058	0.083	0.065	0.069

3.3 Estimation of the Integrated Volatility

The integrated volatility of (X_t) on $[0, t]$, $t \leq T$, is defined by

$$IV_t = \int_0^t \sigma_s^2 \mathrm{d}s.$$

In our framework, a natural idea for estimating this quantity is to consider the approximate realized volatility given by

$$\widehat{RV}_{\alpha,t} = \sum_{t_i \leq t} \left(\log(\hat{X}^t_{\tau_i}) - \log(\hat{X}^t_{\tau_{i-1}})\right)^2,$$

where for $t_i < t$,

$$\hat{X}^t_{\tau_i} = P_{t_i} - \alpha\left(\frac{1}{2} - \hat{\eta}_{\alpha,t}\right) sign(P_{t_i} - P_{t_{i-1}}).$$

Let us now give the definition of stable convergence in law. Let Z^α be a family of random variables (taking their values in the space of càdlàg functions endowed with the Skorokhod topology J_1). Let α_n be a deterministic sequence tending to zero as n tends to infinity and $\mathscr{I}$ be a sub-σ-field of $\mathscr{F}$. We say that Z^{α_n} converges $\mathscr{I}$-stably to Z as α_n tends to zero ($Z^{\alpha_n} \overset{\mathscr{I}-\mathscr{L}s}{\to} Z$) if for every $\mathscr{I}$-measurable bounded real random variable V, (V, Z^{α_n}) converges in law to (V, Z) as n tends to infinity. This is a slightly stronger mode of convergence than the weak convergence, see [19] for details and equivalent definitions. Finally, we say that Z^α converges $\mathscr{I}$-stably to Z as α tends to zero if for any sequence α_n tending to zero, $Z^{\alpha_n} \overset{\mathscr{I}-\mathscr{L}s}{\to} Z$.

Next we introduce the following notation: ∇_1 and ∇_2 are two $(2m+1)$ valued vectors defined by $\nabla_{1,1} = 1$, $\nabla_{1,i} = 0$ for $i = 2, \ldots, 2m+1$, $\nabla_{2,1} = 0$ and for $i = 1, \ldots, m$,

$$\nabla_{2,2i} = i + \eta - 1/2,\ \nabla_{2,2i+1} = -i^{-1}(i + \eta - 1/2)(i + 2\eta - 1).$$

The processes $(f_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$ are defined by

$$f_t = \int_0^t \varphi(\chi_u)\sigma_u^2 X_u^2 \mathrm{d}u,\ \ \mu_t = \int_0^t \sum_{k=1}^m \frac{2k(k-1+2\eta)}{2k-1+2\eta} p_k(\chi_u)\varphi(\chi_u)\sigma_u^2 \mathrm{d}u,$$

with

$$\varphi(\chi_u) = \left(\sum_{j=1}^m p_j(\chi_u) j(j-1+2\eta)\right)^{-1}.$$

We are now able to state our limit theorem. Note that we are in the unusual situation where the observation times are random and endogenous. The key idea for the proof is to work in a modified time in which the observation times are equidistant and to use stability properties of the convergence in law in the Skorohod space. Let $\mathbb{D}[0, T]$ denote the space of càdlàg functions on $[0, T]$ and $^\top$ the transpose operator. We have the following result:

Theorem 3.2 *Let $\mathscr{I}$ be the filtration generated by the processes X and χ. As α tends to* 0*, we have*

$$\alpha^{-1}(\widehat{RV}_{\alpha,t} - \int_0^t \sigma_s^2 ds) \overset{\mathscr{I}-\mathscr{L}s}{\to} (\nabla_1^\top + \frac{\mu_t}{f_t}\nabla_2^\top) \int_0^t b_{f_s} d\mathscr{W}_{f_s},$$

in $\mathbb{D}[0, T]$, where $\mathscr{W}$ is a $(2m+1)$ Brownian motion which is defined on an extension of the filtered probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and is independent of all the preceding quantities and b_s is a $(2m+1)\times(2m+1)$ matrix defined in [28].

Note that the assumptions on the efficient price are very weak. In particular, σ_u is not necessarily an Ito semi-martingale as in [4] or [20]. Moreover, assuming that the process χ is observed and that we are able to estimate the functions p_k (which is not restrictive in practice, as shown in [27]), all the quantities in the asymptotic variance can naturally be estimated. Thus, for practical use, confidence intervals can be built.

3.4 Estimation of the Integrated Covariation

We now turn to the problem of estimating the integrated covariation when two assets are observed. In our context, one can not use the classical realized covariation estimator for two reasons: the asynchronicity of the data and the presence of microstructure noise. The problem of the asynchronicity of the data has to be taken with great care since intuitive ideas such as the previous tick interpolation may lead to a systematic bias called Epps effect, see [18] and [34] for details. This issue has been treated by Hayashi and Yoshida in [18]. Nevertheless, the proposed estimator is in general not robust to microstructure noise, see [32]. We show in [28] that if the Hayashi-Yoshida estimator is used on the estimated values of the efficient price given by Eq. (3), it is consistent.

More precisely we consider a $\mathscr{F}_t$-adapted bidimensional continuous Ito semi-martingale $(X_t^{(1)}, X_t^{(2)})$ such that for $j = 1, 2$

$$Y_t^{(j)} = \log X_t^{(j)} = \log X_0^{(j)} + \int_0^t a_u^{(j)} \mathrm{d}u + \int_0^t \sigma_{u-}^{(j)} \mathrm{d}W_u^{(j)}$$

and

$$\langle W^{(1)}, W^{(2)} \rangle_t = \int_0^t \rho_s \mathrm{d}s,$$

where ρ_s is an adapted process with càdlàg sample paths such that for all s, $-1 < \rho_s < 1$. We impose the same assumptions on $a^{(j)}$ and $\sigma^{(j)}$ as previously. The quantities $\alpha^{(j)}$, $\eta^{(j)}$, $L_i^{(j)}$, $\tau_i^{(j)}$, $t_i^{(j)}$ and $P_{t_i^{(j)}}^{(j)}$ are also defined in the same way as previously. The usual Hayashi-Yoshida covariation estimator is given by

$$HY_t = \sum_{t_{i_1}^{(1)} \le t} \sum_{t_{i_2}^{(2)} \le t} \left(\log(P_{t_{i_1}^{(1)}}^{(1)}) - \log(P_{t_{i_1-1}^{(1)}}^{(1)})\right)\left(\log(P_{t_{i_2}^{(2)}}^{(2)}) - \log(P_{t_{i_2-1}^{(2)}}^{(2)})\right) v_{i_1,i_2}$$

with

$$v_{i_1,i_2} = \mathbb{I}_{[t_{i_1-1}^{(1)}, t_{i_1}^{(1)}] \cap [t_{i_2-1}^{(2)}, t_{i_2}^{(2)}] \neq \emptyset}.$$

Naturally, we define our estimator by

$$\widehat{RCV}_t = \sum_{t_{i_1}^{(1)} \le t} \sum_{t_{i_2}^{(2)} \le t} \left(\log(\hat{X}_{\tau_{i_1}^{(1)}}^{(1)}) - \log(\hat{X}_{\tau_{i_1-1}^{(1)}}^{(1)})\right)\left(\log(\hat{X}_{\tau_{i_2}^{(2)}}^{(2)}) - \log(\hat{X}_{\tau_{i_2-1}^{(2)}}^{(2)})\right) v_{i_1,i_2}.$$

We have the following theorem:

Theorem 3.3 *Assume that for all k and m, the rank of $t_k^{(m)}$ among the set of all the $t_i^{(1)}$ and $t_j^{(2)}$ is the same as the rank of $\tau_k^{(m)}$ among the set of all the $\tau_i^{(1)}$ and $\tau_j^{(2)}$. Suppose also that $\alpha^{(2)} = c\alpha^{(1)}$ with $c > 0$. As $\alpha^{(1)}$ tends to zero, we have*

$$\widehat{RCV}_t \overset{u.c.p.}{\rightarrow} \int_0^t \rho_s \sigma_s^{(1)} \sigma_s^{(2)} ds.$$

Thus, the problem of estimating the integrated covariation of two assets is another example which shows that our method consisting in estimating the values of the efficient price is very convenient to adapt classical statistical procedures to the microstructure noise context.

4 Hedging Error and Microstructure Noise

In [29], we consider a very different problem. Indeed, we are interested in the issue of hedging a European derivative in the context of the model with uncertainty zones. In mathematical finance, the questions of pricing and hedging a derivative were initially treated under the assumption of a "frictionless" market. The fundamental conditions for such a market can be summarized as follows:

- It is possible to borrow and lend cash at a risk-free interest rate.
- The transaction price is equal to the efficient price, irrespectively of the volume of the transaction and of its sign (buy or sell).
- One can buy or sell instantaneously and continuously.
- There are no transaction costs.
- The asset is perfectly divisible (it is possible to buy or sell any fraction of a share). Moreover, short selling is authorized.

The failure of one of the preceding conditions makes the problem of hedging a derivative security more complex. For example, the case of restrictions on short selling is treated in [21] and the presence of liquidity costs is studied in [7]. The consequences of transaction costs in conjunction with discrete-time hedging operations has also been extensively studied, see among others [22, 23, 25].

In [29], we consider the model with uncertainty zones since it is quite a reasonable model for ultra high frequency prices and durations. In particular, prices stay on the tick grid. This has two important consequences. The first one is the impossibility to buy or sell a share at the efficient price: the microstructure noise leads to a cost (possibly negative) that can not be avoided. The second one comes from the fact that the transaction price changes a finite number of times on a given time period. Therefore, it is reasonable to assume that one waits for a price change before rebalancing the hedging portfolio. Thus, we consider a framework where the second of the preceding assumptions is no more in force and where the third one becomes irrelevant.

4.1 Hedging Strategies

We consider a European derivative security with expiration date T and payoff $F(X_T)$, where F is a regular enough payoff function[2]. We present in this section our benchmark frictionless hedging strategy and two strategies adapted to our uncertainty zone market. We assume in the remainder of the paper that all assets are perfectly divisible and, without loss of generality, that the riskless borrowing and lending rate is zero.

4.1.1 Benchmark Frictionless Hedging Strategy

The benchmark frictionless hedging strategy is those of an agent deciding (possibly wrongly) that the volatility of the efficient price at time t is equal to $\sigma(t, X_t)$, for a regular enough function $\sigma(t, x)$. If true, such an assumption on the volatility enables to build a self-financing replicating portfolio of stocks and riskless bonds whose marked to model price at time t is of the form $C\,(t, X_t)$. The function C satisfies

$$\dot{C}_t\,(t,x) + \frac{1}{2}\sigma^2\,(t,x)\,x^2\ddot{C}_{xx}\,(t,x) = 0, \;\; C\,(T,x) = F\,(x)$$

with $\dot{C}_t\,(t,x) = \partial C\,(t,x)\,/\partial t$, $\ddot{C}_{xx}\,(t,x) = \partial^2 C\,(t,x)\,/\partial x^2$.

Suppose the agent implements this strategy in a frictionless market. It leads to a benchmark frictionless hedging portfolio whose value Π_t satisfies

$$\Pi_t = C\,(0, X_0) + \int_0^t \dot{C}_x\,(u, X_u)\,\mathrm{d}X_u.$$

Note that, if the model is misspecified, Π_t is different from $C\,(t, X_t)$, see [11].

Finally, we assume that for some $M > 0$, there exists a sequence of closed sets $\mathscr{S}_n \subset [0, M]$, such that that the function $C\,(t,x)$ satisfies

$$\left|\frac{\partial^{\gamma+\beta} C\,(t,x)}{\partial t^{\gamma}\,\partial x^{\beta}}\right| < +\infty,$$

for all $(t,x) \in \{[0, T-1/n) \times [1/n, +\infty)\} \cup \left\{[T-1/n, T] \times \{[1/n, +\infty)/\overset{\circ}{\mathscr{S}}_n\}\right\}$, and

$$\rho_n = \inf\{T - 2/n \le t \le T, X_t \in \mathscr{S}_n\}$$

is such that $\mathbb{P}[\rho_n > T] \to 1$, with the convention $\inf\{\emptyset\} = +\infty$. Remark that for example, if the benchmark frictionless hedging portfolio is built thanks to the Black-Scholes model, the preceding assumption holds for a European call with strike K taking $\mathscr{S}_n = [K - 1/n, K + 1/n]$.

[2] This is a simplifying, slightly incorrect, framework since we should consider a payoff $F(P_T)$. However, the order of magnitude of the difference is clearly negligible in our context.

4.1.2 Hedging Strategies in an Uncertainty Zone Market

We assume the real market is a model with uncertainty zones, with, for simplicity, $t_i = \tau_i$ for all i and where one can buy and sell at the last traded price, irrespectively of the volume of the transaction and of its sign. So, we naturally impose that the times when the hedging portfolio may be rebalanced are the times where the transaction price moves[3]. Thus, we assume that the hedging portfolio can only be rebalanced at the transaction times τ_i. Therefore, the trading strategies are here piecewise constant. In this setting, we consider strategies such that, if τ_i is a rebalancing time, the number of shares in the risky asset at time τ_i is $\dot{C}_x\left(\tau_i, X_{\tau_i}\right)$. In the next section, we consider two hedging strategies: (i) the hedging portfolio is rebalanced every time that the transaction price moves, (ii) the hedging portfolio is rebalanced only once the transaction price has varied by more than a selected value.

4.2 Asymptotic Results for the Hedging Error

In order to study the hedging error, we develop an asymptotic approach, in the spirit of [5], [17] et [31]. In our setting, the microstructural hedging error is due to:

- discrete trading: the hedging portfolio is rebalanced a finite number of times;
- microstructure noise on the price: between two rebalancing times, the variation of the market price (multiple of the tick size) differs from the variation of the efficient price.

We analyze this microstructure hedging error in two steps. First we assume that there is no microstructure noise on the price although the trading times are endogenous (for all i, $P_{\tau_i} = X_{\tau_i}$): we expect more or less similar results as in [5], [17] and [31] where exogenous trading times are considered. Second we assume the presence of the endogenous microstructure noise and discussed the two hedging strategies.

4.3 Hedging Error Without Microstructure Noise on the Price

We first study the effect of discrete trading. Thus we assume here there is no microstructure on the prices, that is $P_{\tau_i} = X_{\tau_i}$. Let $\phi(t) = \sup\{\tau_i : \tau_i < t\}$. The hedging error can be written

$$L^{(1)}_{\alpha,t} = \int_0^t [\dot{C}_x(u, X_u) - \dot{C}_x(\phi(u), X_{\phi(u)})]\mathrm{d}X_u,$$

and we show the following result in [29]:

[3] The agent is supposed to be price follower, which means that the transaction price never moves because of its own trading.

Theorem 4.1 *As α tends to* 0,

$$N_{\alpha,t}^{1/2} L_{\alpha,t}^{(1)} \overset{\mathscr{I}-\mathscr{L}s}{\rightarrow} L_t^{(1)} := f_t^{1/2} \int_0^t c_{f_s}^{(1)} \mathrm{d}\mathscr{W}_{f_s}^{(1)},$$

where $\mathscr{W}^{(1)}$ is a Brownian motion defined on an extension of the filtered probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and independent of all the preceding quantities, and $c_s^{(1)}$ is such that

$$(c_s^{(1)})^2 = \frac{1}{6} \ddot{C}_{xx}^2 \left(\theta_s, X_{\theta_s}\right) \mu_4 \left(\chi_{\theta_s}\right).$$

We see that the variance of the hedging error is proportional to reciprocal of the number of rebalancing transactions and depends on the local volatility gamma of the derivative security.

4.4 Hedging Error with Microstructure Noise

In the presence of microstructure noise on the price, the transaction prices differ from the efficient prices. The hedging error is now given by

$$L_{\alpha,t}^{(2)} = \int_0^t \dot{C}_x \left(u, X_u\right) \mathrm{d}X_u - \int_0^t \dot{C}_x(\phi\left(u\right), X_{\phi(u)}) \mathrm{d}P_u.$$

Let

$$\pi_a \left(\chi_t\right) = \lim_{\alpha\to 0} \mathbb{E}_{\mathscr{E}_{\tau_{i\alpha}}} \left[\mathbb{I}_{\{\Delta X_{\tau_{i\alpha}} \Delta X_{\tau_{i\alpha}-1} < 0\}}\right] = \sum_{k=1}^m \frac{k}{2k-1+2\eta} p_k \left(\chi_t\right)$$

be the asymptotic conditional probability that the next price change at time t is due to an alternation and let

$$\begin{aligned}\mu_{1,a}^* \left(\chi_t\right) &= \lim_{\alpha\to 0} \mathbb{E}_{\mathscr{E}_{\tau_{i\alpha}}} \left[\alpha^{-1} |\Delta X_{\tau_{i\alpha}}| \mathbb{I}_{\{\Delta X_{\tau_{i\alpha}} \Delta X_{\tau_{i\alpha}-1} < 0\}}\right] \\ &= \sum_{k=1}^m \frac{k(k-1+2\eta)}{2k-1+2\eta} p_k \left(\chi_t\right)\end{aligned}$$

be the asymptotic conditional expectation of the absolute value of the normalized price change at time t when the price change is due to an alternation. We have the following result:

Theorem 4.2 *As α tends to* 0,

$$L_{\alpha,t}^{(2)} \overset{\mathscr{I}-\mathscr{L}s}{\rightarrow} L_t^{(2)} := \int_0^t a_{f_s}^{(2)} \mathrm{d}s + \int_0^t b_{f_s}^{(2)} \mathrm{d}X_s + \int_0^t c_{f_s}^{(2)} \mathrm{d}\mathscr{W}_{f_s}^{(2)},$$

in $\mathbb{D}[0, T]$, where $\mathscr{W}^{(2)}$ is a Brownian motion defined on an extension of the filtered probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and independent of all the preceding quantities,

and $a_s^{(2)}$, $b_s^{(2)}$ *and* $c_s^{(2)}$ *are such that*

$$\begin{aligned} a_s^{(2)} &= -(1-2\eta)\ddot{C}_{xx}\left(\theta_s, X_{\theta_s}\right)\mu_{1,a}^*\left(\chi_{\theta_s}\right)\varphi\left(\chi_{\theta_s}\right) \\ b_s^{(2)} &= (1-2\eta)\dot{C}_x\left(\theta_s, X_{\theta_s}\right)\mu_{1,a}^*\left(\chi_{\theta_s}\right)\varphi\left(\chi_{\theta_s}\right) \\ (c_s^{(2)})^2 &= (1-2\eta)^2\dot{C}_x^2\left(\theta_s, X_{\theta_s}\right)\varphi\left(\chi_{\theta_s}\right)\left(\pi_a\left(\chi_{\theta_s}\right)\varphi^{-1}\left(\chi_{\theta_s}\right) - (\mu_{1,a}^*\left(\chi_{\theta_s}\right))^2\right). \end{aligned}$$

It is worth noticing that the microstructural hedging error process is not renormalized as in the previous case. It means that the hedging error does not vanish even if the number of rebalancing transactions goes to infinity. However, if $\eta = 1/2$, the changes of the transaction prices coincide with the changes of the efficient prices at the exit times of the uncertainty zones ($\Delta P_{\tau_i} = \Delta X_{\tau_i}$) and the error due to the microstructure noise on the price vanishes.

The first two components of the asymptotic hedging error are quite unusual in this kind of asymptotic distribution theories. The first term can be interpreted as a bias due to implicit transaction costs linked to bid-ask bounce. The second one is due to the asymmetry between alternations and continuations. Indeed, when an alternation occurs, $\Delta P_{\tau_i} - \Delta X_{\tau_i} = (1-2\eta)sign(\Delta X_{\tau_i})$ while, when a continuation occurs, $\Delta P_{\tau_i} - \Delta X_{\tau_i} = 0$. Moreover, remark that the quadratic variation of the asymptotic hedging error is

$$(1-2\eta)^2 \int_0^t \dot{C}_x^2\left(s, X_s\right)\pi_a\left(\chi_s\right)\mathrm{d}f_s.$$

It now depends on the local volatility delta of the derivative security and on the proportion of alternations. Consequently, the variance of the microstructural hedging error increases with the position delta and the proportion of alternation in the price. Indeed, compared with the previous case, one now faces an additional microstructural hedging error of the first order.

4.5 Optimal Rebalancing Level in the Presence of Microstructure Noise

We now build strategies where the portfolio is rebalanced when the price moves "significantly". It is probably more natural than a timely based rebalancing strategy and should reduce the impact of the microstructure noise. So we now assume that the hedging portfolio is rebalanced only once the price changes of l_α ticks. For simplicity we assume here that $m = 1$. We choose l_α such that $l_\alpha \to \infty$ and $\alpha l_\alpha \to 0$, as $\alpha \to 0$. In this way hedging errors due respectively to microstructure noise on the price and discrete-time rebalancing will disappear as the tick goes to zero. The hedging error is now given by

$$L_{\alpha,t}^{(3)} = \int_0^t \dot{C}_x(u, X_u)\mathrm{d}X_u - \int_0^t \dot{C}_x(\phi^{(l)}(u), X_{\phi^{(l)}(u)})\mathrm{d}P_u.$$

with $\phi^{(l)}(t) = \sup\{\tau_i^{(l)} : \tau_i^{(l)} < t\}$ and the $\tau_i^{(l)}$ are the stopping times associated to moves of l_α ticks. We have the following result:

Theorem 4.3 *Let $l_\alpha = \lfloor \alpha^{-1/2} \rfloor$. As α tends to 0,*

$$(N_{\alpha,t})^{1/4} L_{\alpha,t}^{(3)} \overset{\mathscr{I}-\mathscr{L}s}{\to} L_t^{(3)} := (f_t^{(l)})^{1/4} \times \left(\int_0^t a_{f_s^{(l)}}^{(3)} \mathrm{d}s + \int_0^t b_{f_s^{(l)}}^{(3)} \mathrm{d}X_s + \int_0^t c_{f_s^{(l)}}^{(3)} \mathrm{d}\mathscr{W}_{f_s^{(l)}}^{(3)} \right),$$

where $\mathscr{W}^{(3)}$ is Brownian motion defined on an extension of the filtered probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and independent of all the preceding quantities, and $b_s^{(3)}$ and $c_s^{(3)}$ are such that

$$\begin{aligned} a_s^{(3)} &= -(1-2\eta)\ddot{C}_{xx}(\theta_s, X_{\theta_s}) \\ b_s^{(3)} &= \frac{(1-2\eta)}{2}\dot{C}_x(\theta_s, X_{\theta_s}) \\ (c_s^{(3)})^2 &= \frac{(1-2\eta)^2}{4}\dot{C}_x^2(\theta_s, X_{\theta_s}) + \frac{1}{6}\ddot{C}_{xx}^2(\theta_s, X_{\theta_s}). \end{aligned}$$

This optimal strategy (in term of choice of l_α) allows to reduce significantly the hedging error in the presence of microstructure noise. Interestingly, the quadratic variation of the asymptotic hedging error now depends both on the delta and on the gamma of the derivative security.

4.6 One Numerical Study

We give in this section numerical results about the hedging errors of a European call with strike $K = 100$ and maturity $T = 1$. We consider the following model for the underlying asset for the trading period. The efficient price is given by the Black-Scholes dynamics

$$\mathrm{d}X_t = \sigma X_t \mathrm{d}W_t, \; x_0 = 100, \; t \in [0, T],$$

where $\sigma = 0.01$ and we take $\alpha = 0.05$, $\eta = 0.05$ and $m = 1$. These parameters are chosen to be in agreement with real data, see [27]. Finally, we assume the benchmark strategy of the agent is the Black-Scholes strategy with $\sigma = 0.01$. We give here statistics for the number of rebalancings (NR) and the histogram of the hedging

Table 2

	Average of NR	Standard Deviation of NR
Every price move rebalancing	3487	108
5 ticks rebalancing	20.22	3.55

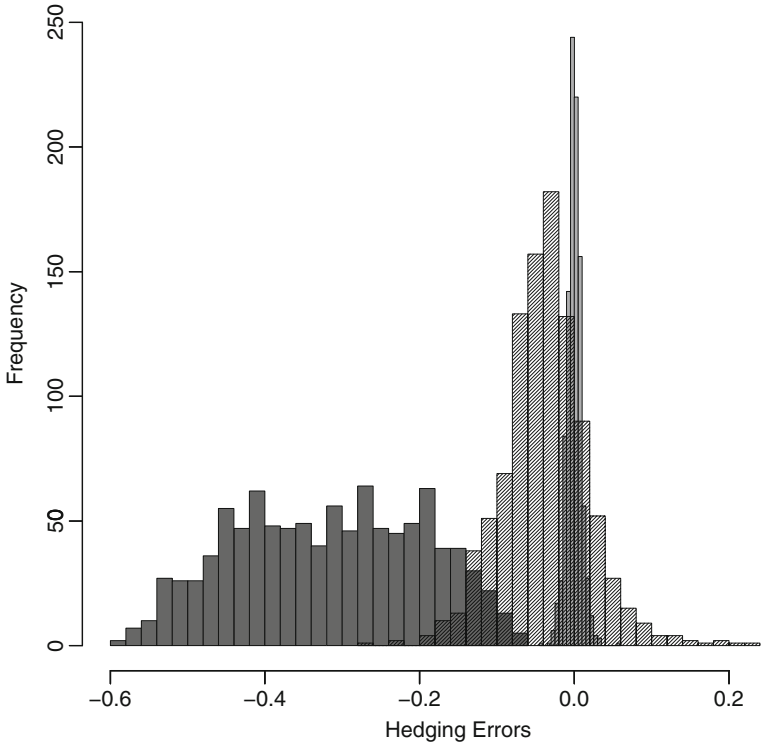


Fig. 5 Histograms of $L_{\alpha,T}^{(i)}$, for $i = 1$ (in green, error due to discrete trading times), 2 (in red, error due to discrete trading times and microstructure noise), 3 (in grey, error in the case of "optimal" rebalancing)

error for the different strategies over 1000 Monte-Carlo simulations (Fig. 5). Note that $L_{\alpha,T}^{(3)}$ is computed for price moves of 5 ticks.

We clearly see that rebalancing the portfolio each time the price changes induces a strong negative bias in the hedging error. Rebalancing less frequently enables to significantly improve the hedging error, which is in agreement with the theoretical results.

References

1. Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2000) Great Realizations. Risk, 13, 105–108
2. Ball, C. (1988) Estimation bias induced by discrete security prices. Journal of Finance, 43, 841–865
3. Bandi, F.M., Russel, J.R. (2005) Microstructure noise, realized variance and optimal sampling. To appear in Review of Economic Studies

4. Barndorff-Nielsen O.E., Hansen P., Lunde A., Shephard, N. (2008) Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica, 76, 1481–1536
5. Bertsimas, D., Kogan, L., Lo, A.W. (2000) When is time continuous? Journal of Financial Economics, 55, 173–204
6. Black, F., Scholes, M. (1973) The pricing of options and corporate liabilities. Journal of Political Economy, 81, 637–654
7. Cetin, U., Soner, M., Touzi, N. (2009) Option hedging under liquidity costs. To appear in Finance and Stochastics
8. Cho, D., Frees, E. (1988) Estimating the volatility of discrete stock prices. Journal of Finance, 43, 451–466
9. Delbaen, F., Schachermayer, W. (1994) A general version of the fundamental theorem of asset pricing. Mathematische Annalen, 300, 463–520
10. Delattre, S., Jacod, J. (1997) A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round-off errors. Bernoulli, 3, 1–28
11. El Karoui, N., Jeanblanc-Picqué, M., Shreve, S. (1998) Robustness of the Black and Scholes formula. Mathematical Finance, 8, 93–126
12. Engle, R.F. (2000) The econometrics of ultra high frequency data. Econometrica, 68, 1–22
13. Gloter, A., Jacod, J. (1997) Diffusions with measurement errors, I. Local Asymptotic Normality, II. Optimal estimators. ESAIM PS, 5, 225–260
14. Gottlieb, G., Kalay, A. (1985) Implications of the Discreteness of Observed Stock Prices. Journal of Finance, 40, 135–154
15. Gourieroux, C., Jasiak, J. (2001) Financial Econometrics: Problems, Models, and Methods. Princeton University Press
16. Hansen, P.R., Lunde, A. (2006) Realized variance and market microstructure noise. Journal of Business and Economics Statistics, 24, 127–161
17. Hayashi, T., Mykland, P.A. (2005) Hedging errors: an asymptotic approach. Mathematical Finance, 15, 309–343
18. Hayashi, T., Yoshida, N. (2005) On covariance estimation of non-synchronously observed diffusion processes. Bernoulli, 11, 359–379
19. Jacod, J., Shiryaev, A.N. (2002) Limit Theorems for Stochastic Processes. Springer-Verlag, Second Edition
20. Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M. (2009) Microstructure noise in the continuous case: the pre-averaging Approach. Stochastic Processes and their Applications, 119, 2249–2276
21. Jouini, E., Kallal, H. (1995) Arbitrage in securities markets with short sales constraints. Mathematical Finance, 5, 197–232
22. Kabanov, Y.M., Last, G. (2002) Hedging under transaction costs in currency markets: a continuous- time model. Mathematical Finance, 12, 63–70
23. Kabanov, Y.M., Safarian, M.M. (1997) On Leland's strategy of option pricing with transactions costs. Finance and Stochastics, 1, 239–250
24. Kalnina, I., Linton, O.B. (2008) Estimating quadratic variation consistently in the presence of correlated measurement error. Journal of Econometrics, 147, 47–59
25. Koehl, P.F., Pham, H., Touzi, N. (1999) Hedging in discrete time under transaction costs and continuous-time limit. Journal of Applied Probability, 36, 163–178
26. Li, Y., Mykland, P.A. (2007) Are volatility estimators robust with respect to modelling assumptions? Bernoulli, 13, 601–622
27. Robert, C.Y., Rosenbaum, M. (2009) A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones. To appear in Journal of Financial Econometrics
28. Robert, C.Y., Rosenbaum, M. (2009) Volatility and covariation estimation when microstructure noise and trading times are endogenous. To appear in Mathematical Finance
29. Robert, C.Y., Rosenbaum, M. (2010) On the microstructural hedging error. SIAM Journal on Financial Mathematics, 1, 427–453
30. Rosenbaum, M. (2009) Integrated volatility and round-off error. Bernoulli, 15, 687–720

31. Tankov, P., Voltchkova, E. (2008) Asymptotic analysis of hedging errors in models with jumps. Stochastic Processes and their Applications, 119, 2004–2027
32. Voev, V., Lunde, A. (2007) Integrated covariance estimation using high-frequency data in the presence of noise. Journal of Financial Econometrics, 5, 68–104
33. Zhang, L. (2006a) Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli, 12, 1019–1043
34. Zhang, L. (2006b) Estimating covariation: Epps effect and microstructure noise. A paraître dans Journal of Econometrics
35. Zhang, L., Mykland, P.A., Aït-Sahalia, Y. (2005) A Tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association, 100, 1394–1411
36. Zhou, B. (1996) High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics, 14, 45–52

Exponential Resilience and Decay of Market Impact

Jim Gatheral, Alexander Schied and Alla Slynko

Abstract. Assuming a particular price process, it was shown by Gatheral in [6], that a model that combines nonlinear price impact with exponential decay of market impact admits price manipulation, an undesirable feature that should lead to rejection of the model. Subsequently, Alfonsi and Schied proved in [2] that their model of the order book which has nonlinear market impact and exponential resilience, is free of price manipulation. In this paper, we show how these at-first-sight incompatible results are in reality perfectly compatible.

1 Two Different Generalizations of the Obizhaeva and Wang Model

We begin by reviewing the models presented in [6] and [1] respectively, both of which turn out to be (different) generalizations of the model of [8]. In what follows, we work in continuous time and consider only absolutely continuous strategies. Although this restriction might at first seem very limiting, one can think of approximating any trading strategy (which may even include block trades) by a sequence of interval VWAP trades.

1.1 The Gatheral (JG) Model

In [6], the stock price S_t at time t is given by

$$S_t = S_0 + \int_0^t f(v_s)\, G(t-s)\, ds + \int_0^t \sigma_s\, dZ_s \tag{1}$$

Jim Gatheral
Department of Mathematics, Baruch College, CUNY, New York,
e-mail: jim.gatheral@baruch.cuny.edu

Alexander Schied, Alla Slynko
Department of Mathematics, Mannheim University, Mannheim, Germany,
e-mail: schied@uni-mannheim.de; alla.a.slynko@gmail.com

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

where v_s is our rate of trading in dollars at time $s < t$, the *instantaneous market impact function* $f(v_s)$ represents the impact of trading at time s and the *decay kernel* $G(\tau)$ is a convex monotonic decreasing function of its argument. Z is a Brownian motion. S_t thus follows an arithmetic random walk with a drift that depends on the accumulated impacts of previous trades.

Expression (1) may be regarded as a generalization of processes previously considered by [4], [5] and [8].

Assuming the price process (1), the cost $\mathscr{C}_{JG}(X)$ associated with an absolutely continuous trading strategy X is given by

$$\mathscr{C}_{JG}(X) = \int_0^T v_t \, dt \int_0^t ds \, f(v_s) \, G(t-s) \tag{2}$$

with $v_t = \dot{X}_t$.

1.2 A Version of the Alfonsi–Fruth–Schied (AFS) Model in Continuous Time

[1] and [2] assume a continuous ask price distribution for available shares in the limit order book: the number of shares offered at a price x greater than the current stock price is given by $\varphi(x)\,dx$ where $\varphi(\cdot)$ (assumed continuous) is the *density* of the limit order book. All trade executions are assumed to be with market orders. Each time a market (buy) order arrives, it eats into the order book; limit order arrivals replenish the order book, up to the original target density $\varphi(x)$. The *volume impact* E_t quantifies the size of the resulting hole in the order book at time t. The increase in E_t resulting from the arrival of market orders is offset by the decrease in E_t resulting from the arrival of limit orders into the order book.

In the discrete time model of [1] and [2], as illustrated in Fig. 1, when a trade of size ξ_t is placed at time t, E_t changes to $E_{t+} = E_t + \xi_t$. Following the assumption of exponential resilience of the order book, the size of the hole at time $u > t$ is modeled as

$$E_u = e^{-\varrho\,(u-t)}\, E_{t+} = e^{-\varrho\,(u-t)}\, (E_t + \xi_t)\,.$$

By induction, we see that

$$E_u = \sum_{t<u} e^{-\varrho\,(u-t)}\, \xi_t\,.$$

In an obvious generalization of this discrete time model to continuous time and from exponential to arbitrary resilience, we describe the evolution of the volume impact process by

$$E_t = \int_{[0,t)} \psi(t-s)\, dX_s$$

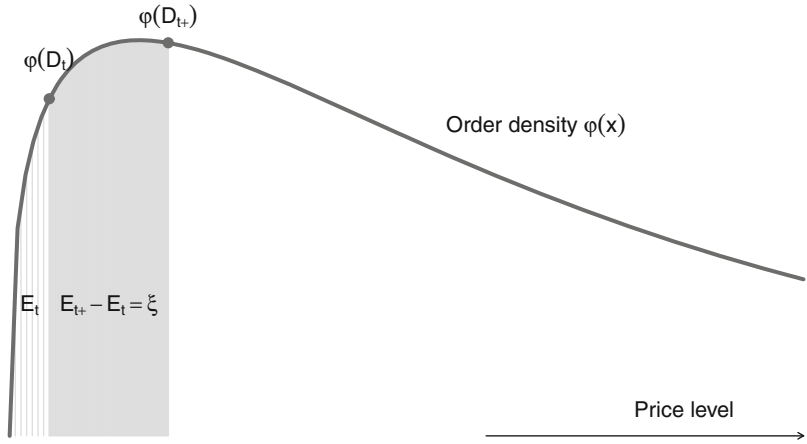


Fig. 1 Schematic of the order book in the AFS model. When a trade of size ξ is placed at time t, E_t changes to $E_{t+} = E_t + \xi$ and $D_t = F^{-1}(E_t)$ to $D_{t+} = F^{-1}(E_{t+}) = F^{-1}(E_t + \xi)$

where the strategy X may consist of both block-trades and periods of continuous trading, and $\psi(\tau)$ is the *resilience function*.
The *spread* D_t at time t is given by

$$D_t = F^{-1}(E_t)$$

where $F(x)$ is the cumulative density of orders and given by the antiderivative of $\varphi(\cdot)$:

$$F(x) = \int_0^x \varphi(y)\, dy.$$

The price process is then given by

$$S_t = S_t^0 + D_t = S_t^0 + F^{-1}(E_t), \tag{3}$$

where S^0 is the unaffected price process. It can be taken as $S_t^0 = S_0 + \int_0^t \sigma_s\, dZ_s$ as in (1).

To compute the expected cost of a trading strategy X, consider the purchase of a block of ξ shares. Once again, as illustrated in Fig. 1, when a trade of size ξ is placed at time t, E_t changes to $E_{t+} = E_t + \xi$ and $D_t = F^{-1}(E_t)$ to

$$D_{t+} = F^{-1}(E_{t+}) = F^{-1}(E_t + \xi).$$

The corresponding expected cost increment is

$$\int_{D_t}^{D_{t+}} x\, \varphi(x)\, dx = \int_{F^{-1}(E_t)}^{F^{-1}(E_{t+})} x\, \varphi(x)\, dx = \int_{E_t}^{E_{t+}} F^{-1}(x)\, dx = H(E_{t+}) - H(E_t)$$

where

$$H(x) = \int_0^x F^{-1}(x)\, dx.$$

For a sequence of block trades $\xi = (\xi_1, ..., \xi_N)$ we get as total expected execution costs:

$$\mathscr{C}(\xi) = \sum_{n=1}^{N} \left[H(E_{n+}) - H(E_n) \right].$$

When passing to the limit, in the case that execution strategy X is absolutely continuous with $\dot{X}_t = v_t$, this quantity converges to

$$\int_0^T H'(E_t)\, dX_t = \int_0^T H'(E_t)\, v_t\, dt.$$

Hence, in the Alfonsi–Fruth–Schied model, the expected execution costs of the strategy X are:

$$\mathscr{C}_{AFS}(X) = \int_0^T v_t\, dt\, F^{-1}\left(\int_0^t \psi(t-s)\, dX_s \right). \tag{4}$$

By comparing (2) and (4), we see that in some sense, decay of price impact and order book resilience are dual to each other.

2 Price Manipulation in the Gatheral and Alfonsi–Fruth–Schied Models

We begin with a proposition generalizing the result proved in [6] that exponential decay of market impact is compatible only with linear market impact. It turns out that any model with a nonlinear market impact function $f(\cdot)$ and a decay kernel $G(\cdot)$ that is nonsingular at time zero admits price manipulation.

Proposition 1 *Assuming the price process (1), consider a model with a general nonlinear instantaneous market impact function $f(\cdot)$ and a nonincreasing decay kernel $G(t)$ with $G(0) := \lim_{t\downarrow 0} G(t) < \infty$. Then, such a model admits price manipulation.*

Proof. We can assume without loss of generality that $G(0) = 1$. Continuity of this decay kernel at $t = 0$ means that

$$\forall \varepsilon > 0,\ \ \exists \nu > 0 \ \text{ such that } \ \forall t \ \text{ with } \ |t| < \nu,\ \ |G(t) - G(0)| < \varepsilon. \tag{5}$$

To construct the desired round trip we follow Sect. 3.2. of [6]. Thus we first choose $v_1, v_2 > 0$ such that

$$2\varepsilon := \frac{f(v_1)v_2 - f(v_2)v_1}{2v_2 f(v_1)} > 0. \tag{6}$$

With ε given, we then take the corresponding $\nu > 0$ from (5). Then we consider the strategy where shares are purchased at the (positive) constant rate v_1 and then liquidated at the (positive) constant rate v_2 over the time horizon $[0, \nu]$. Therefore, $\dot{X}_t = v_1$ for $t \in [0, \theta\nu]$ and $\dot{X}_t = -v_2$ for $t \in (\theta\nu, \nu]$.

To compute the expected cost of this trading strategy, we use formula (3) in [6] and derive

$$\begin{aligned}
\mathscr{C}(X) &= v_1 f(v_1) \int_0^{\theta\nu} dt \int_0^t G(t-s)ds + v_2 f(v_2) \int_{\theta\nu}^{\nu} dt \int_{\theta\nu}^t G(t-s)ds \\
&\qquad - v_2 f(v_1) \int_{\theta\nu}^{\nu} dt \int_0^{\theta\nu} G(t-s)ds \\
&\le v_1 f(v_1) \int_0^{\theta\nu} dt \int_0^t ds + v_2 f(v_2) \int_{\theta\nu}^{\nu} dt \int_{\theta\nu}^t ds \\
&\qquad + v_2 f(v_1)(\varepsilon - 1) \int_{\theta\nu}^{\nu} dt \int_0^{\theta\nu} ds \\
&= v_1 f(v_1) \frac{\theta^2\nu^2}{2} + v_2 f(v_2)\Big(\frac{\nu^2 - \theta^2\nu^2}{2} - (\theta\nu^2 - \theta^2\nu^2)\Big) \\
&\qquad + v_2 f(v_1)(\varepsilon - 1)\nu^2\theta(1-\theta),
\end{aligned}$$

where we get the inequality by using the fact that $G(\cdot) \le 1$ in order to estimate the first two terms and the fact that $-G(t) \le \varepsilon - 1$, in order to estimate the last one.

Then, from the definition of a round trip, it follows that

$$\theta = \frac{v_2}{v_1 + v_2}$$

and we obtain

$$\begin{aligned}
\mathscr{C}(X) &= v_1 f(v_1) \frac{\nu^2}{2} \frac{v_2^2}{(v_1+v_2)^2} + v_2 f(v_2) \frac{\nu^2}{2} + v_2 f(v_2) \frac{\nu^2}{2} \frac{v_2^2}{(v_1+v_2)^2} \\
&\quad - v_2 f(v_2)\nu^2 \frac{v_2}{v_1+v_2} + \varepsilon v_2 f(v_1)\nu^2 \frac{v_2 v_1}{(v_1+v_2)^2} - v_2 f(v_1)\nu^2 \frac{v_2 v_1}{(v_1+v_2)^2} \\
&= f(v_1)\nu^2 \Big\{ \frac{v_1 v_2^2}{2(v_1+v_2)^2} + \varepsilon \frac{v_1 v_2^2}{(v_1+v_2)^2} - \frac{v_1 v_2^2}{(v_1+v_2)^2} \Big\} \\
&\quad + f(v_2)\nu^2 \Big\{ \frac{v_2}{2} + \frac{v_2^3}{2(v_1+v_2)^2} - \frac{v_2^2}{v_1+v_2} \Big\}
\end{aligned}$$

$$= f(v_1)v^2\Big\{\frac{-v_1v_2^2 + 2\varepsilon v_1 v_2^2}{2(v_1+v_2)^2}\Big\} + f(v_2)\tilde{\varepsilon}^2\Big\{\frac{v_1^2 v_2}{2(v_1+v_2)^2}\Big\}$$

$$= \frac{\tilde{\varepsilon}^2}{2(v_1+v_2)^2}\Big\{2\varepsilon f(v_1)v_1v_2^2 - f(v_1)v_1v_2^2 + f(v_2)v_2v_1^2\Big\}. \tag{7}$$

From (6) it follows that

$$2\varepsilon f(v_1)v_1v_2^2 - f(v_1)v_1v_2^2 + f(v_2)v_2v_1^2 < 0. \tag{8}$$

It follows from (7) that $\mathscr{C}(X) < 0$, and so X is a price manipulation strategy. □

The following special case of the preceding proposition should be compared with the result of [6] mentioned at the beginning of this section.

Corollary 1 *A model with price process (1), $f(\cdot)$ nonlinear, and $G(\tau) = e^{-\varrho\tau}$ for some $\varrho > 0$, admits price manipulation.*

Thus, a JG model with nonlinear market impact and exponential decay always admits price manipulation. On the other hand, it has been shown in [2] that an AFS model with resilience function $\psi(\tau) = e^{-\varrho\tau}$ for some $\varrho > 0$, *does not* admit price manipulation.

Remark 1 It is shown in [7] that in the special case of linear instantaneous market impact, convexity of a nonincreasing decay kernel $G(\cdot)$ is sufficient to ensure that a JG model does not admit price manipulation. Moreover, in this case, the optimal execution strategy consists of trades of only one sign: an optimal buy execution consists of only buy orders.

3 VWAP-equivalent Models

At first sight, the expression (4) for the cost of trading in the Alfonsi–Fruth–Schied model seems to be incompatible with expression (2) for the cost of trading in the Gatheral model: The JG and AFS models are different. Nevertheless, in this section, we find conditions under which the JG and AFS models generate identical estimates for the cost of VWAP executions characterized by $\dot{X}_t = v$, with a given constant $v \in \mathbb{R}$. We call such models *VWAP-equivalent*.

Proposition 2 *JG and AFS models are VWAP-equivalent if and only if $f(v) \propto v^\delta$ for some $\delta > 0$,*

$$G(t) = A\,\partial_t\left(\int_0^t \psi(t-s)\,ds\right)^\delta = A\,\delta\,\psi(t)\Big(\int_0^t \psi(s)\,ds\Big)^{\delta-1}$$

for some positive constant A and

$$F^{-1}(x) = A\,x^\delta.$$

Proof. We require $\mathscr{C}_{AFS}(X)$ and $\mathscr{C}_{JG}(X)$ to be equal for all completion times $T > 0$, so by (2) and (4), we must have

$$F^{-1}\left(\int_0^t \psi(t-s)\, v\, ds\right) = \int_0^t ds\, f(v)\, G(t-s) \tag{9}$$

for all t and all v. Differentiating with respect to t and v respectively gives

$$F^{-1\prime}\left(\int_0^t \psi(t-s)\, v\, ds\right) v\, \psi(t) = f(v)\, G(t)$$

and

$$F^{-1\prime}\left(\int_0^t \psi(t-s)\, v\, ds\right) \left\{\int_0^t ds\, \psi(t-s)\right\} = f'(v) \int_0^t G(t-s)\, ds.$$

Eliminating the factor $F^{-1\prime}(\cdot)$ between these last two equations gives

$$\frac{v\, f'(v)}{f(v)} = \frac{G(t)}{\psi(t)} \frac{\int_0^t ds\, \psi(t-s)}{\int_0^t G(t-s)\, ds}$$

from which we conclude that the models can be VWAP-equivalent only if $f(v) \propto v^{\delta}$ for some $\delta > 0$. In this case we also get

$$\delta\, \frac{\psi(t)}{\int_0^t ds\, \psi(t-s)} = \frac{G(t)}{\int_0^t G(t-s)\, ds}$$

which has the solution

$$\delta \log \int_0^t ds\, \psi(t-s) = \log \int_0^t G(t-s)\, ds + \text{const.}$$

This last equality can be rearranged to obtain

$$G(t) = A\, \partial_t \left(\int_0^t ds\, \psi(t-s)\right)^{\delta}$$

for some positive constant A. Finally, from (9)

$$F^{-1}(x) = A\, x^{\delta}$$

as required. □

Example 1 (Exponential resilience) Put $A = 1$ and take

$$\psi(\tau) = \varrho\, e^{-\varrho\,\tau} \quad \text{for } \varrho > 0.$$

Then the JG and AFS models are VWAP-equivalent only if

$$G(t) = \partial_t \left(\int_0^t ds\, \varrho\, e^{-\varrho\, s} \right)^{\delta} = \frac{\delta\, \varrho\, e^{-\varrho\, t}}{(1 - e^{-\varrho\, t})^{1-\delta}}$$

and also

$$F^{-1}(x) = x^{\delta};\ f(v) = v^{\delta}.$$

Thus in the JG model VWAP-equivalent to the AFS model of [1], decay of market impact is exponential for large times $\varrho\, t \gg 1$ but power-law for small times $\varrho\, t \ll 1$. In particular, exponential resilience of the order book seems not to imply exponential decay of market impact.

Example 2 (No resilience) When $A = 1$ and

$$\psi(\tau) = 1,$$

we must have

$$G(t) = \partial_t t^{\delta} = \frac{\delta}{t^{1-\delta}},$$

for the JG and the AFS models to be VWAP-equivalent. That is, power-law decay of market impact is equivalent (in some sense) to no order book resilience.

Example 3 (Power-law resilience) When $A = 1$ and

$$\psi(\tau) = \frac{1}{\tau^{\gamma}}$$

for some $0 < \gamma < 1$, we must have

$$G(t) = \partial_t \left(\frac{t^{1-\gamma}}{1-\gamma} \right)^{\delta} = \frac{\delta\,(1-\gamma)}{(1-\gamma)^{\delta}}\, \frac{1}{t^{1-\delta\,(1-\gamma)}}$$

for the JG and the AFS models to be VWAP-equivalent. We see that power-law book resilience is equivalent (in some sense) to power-law decay of market impact.

Remark 2 According to [6], price manipulation is possible if the sum of the exponent δ of the power-law of instantaneous market impact and the exponent of the power-law of decay of market impact sum to less than one. We see from Example 3 that in the JG model VWAP-equivalent to an AFS model with power-law resilience, we have

$$\delta + 1 - \delta\,(1-\gamma) = 1 + \delta\,\gamma > 1$$

and the lower bound derived in [6] is not violated.

3.1 When Are VWAP-equivalent Models Identical?

For a JG model and an AFS model to give identical costs for any given absolutely continuous strategy, we would need

$$F^{-1}\left(\int_0^t \psi(t-s)\, v_s\, ds\right) = \int_0^t ds\, f(v_s)\, G(t-s)$$

for arbitrary v_s. Imposing conditions we derived in the previous paragraph for the models to be VWAP-equivalent, we obtain

$$\left(\int_0^t \psi(t-s)\, v_s\, ds\right)^{\delta} = \int_0^t ds\, v_s^{\delta}\, \partial_s \left(\int_0^s du\, \psi(s-u)\right)^{\delta}.$$

By (for example) considering the special case $\psi(\tau) = 1$, $v_s = s$, we find that we must have $\delta = 1$ in which case, equivalence of the two formulations holds trivially. Moreover, in the linear case, $G(\tau) = \psi(\tau)$.

Thus, both the order book model of [1] and the price process of [6] are generalizations of the model of [8] with linear market impact and exponential decay. However, they are quite different generalizations. The two formulations give identical predictions for the cost associated with any given absolutely continuous trading strategy only when market impact is linear.

Example 4 (Square-root market impact with square-root resilience) Consider the two VWAP-equivalent models with $F^{-1}(x) = \sqrt{x}$, $\psi(\tau) = 1/\sqrt{\tau}$, $f(v) = \sqrt{v}$ and

$$G(\tau) = \partial_\tau \left(2\sqrt{\tau}\right)^{1/2} = \frac{1}{2\sqrt{2}} \frac{1}{\tau^{3/4}}.$$

By substitution into (2) and (4), it is easy to verify that for a VWAP execution with constant trading rate v and time to completion T,

$$\mathscr{C}_{JG} = \mathscr{C}_{AFS} = \frac{4}{5}\sqrt{2}\, v^{3/2}\, T^{5/4}.$$

and in both models, for $t < T$, the expected stock price is given by

$$S_t - S_0 = \sqrt{2\, v}\, t^{1/4}.$$

However, after completion of the execution, the expected stock price differs in the two models. In the JG model, the expected stock price is given for $t > T$ by

$$S_t^{JG} - S_0 = \int_0^T \sqrt{v}\, G(t-s)\, ds = \sqrt{2\, v}\, \left\{t^{1/4} - (t-T)^{1/4}\right\}. \tag{10}$$

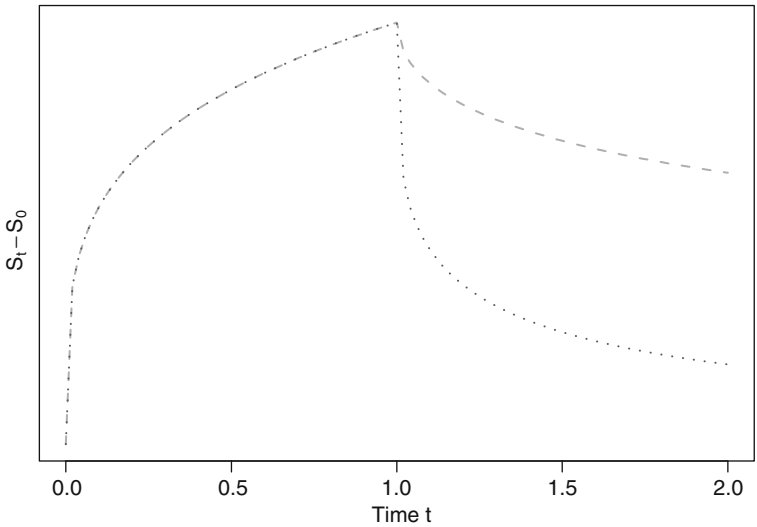


Fig. 2 With $T = 1$ and $v = 1$, graphs of $S_t^{AFS} - S_0$ (the *dashed line*) and $S_t^{JG} - S_0$ (the *dotted line*). VWAP-equivalent JG and AFS models generate identical expected price paths before completion of a VWAP execution but very different price paths after completion

In the VWAP-equivalent AFS model, the expected stock price is given for $t > T$ by

$$S_t^{AFS} - S_0 = \sqrt{\int_0^T v\,\psi(t-s)\,ds} = \sqrt{2\,v}\,\left(\sqrt{t} - \sqrt{t-T}\right)^{1/2}. \tag{11}$$

For $t > T$ then, S_t^{JG} and S_t^{AFS} have very different behaviors as illustrated in Fig. 2.

4 Compatibility with the Square-root Formula

The *square-root formula* has the following form:

$$\text{Cost} = \text{Spread term} + c\,\sigma\,\sqrt{\frac{n}{V}}$$

where n is the number of shares to be traded, σ is the volatility of the stock (in daily units), and V is the average daily volume of the stock. One could think of the spread term as representing VWAP slippage and the second square-root term as representing price impact.

As previously discussed in [6], the square-root formula has been widely used for many years to generate a pre-trade estimate of transactions cost; there is ample empirical evidence that this formula does indeed generate a reasonable rough estimate of transactions costs for simple execution strategies such as VWAP.

Interestingly, the square-root formula implies that the cost of liquidating a stock is independent of the time taken: the formula refers neither to the duration of the trade nor to the trading strategy adopted. Fixing market volume and volatility, price impact depends only on trade-size.

In the JG model, with

$$f(v) = \frac{3}{4}\sigma\sqrt{\frac{v}{V}}$$

and $G(\tau) = \tau^{-1/2}$, the total cost in dollars associated with a VWAP execution with duration T is given by

$$\mathscr{C}_{JG} = \frac{3}{4}\sigma\, v\sqrt{\frac{v}{V}}\int_0^T dt\int_0^t \frac{ds}{\sqrt{t-s}} = \sigma\, v\sqrt{\frac{v}{V}}\,T^{3/2}$$

so the cost per share is given by

$$\hat{\mathscr{C}} = \frac{\mathscr{C}_{JG}}{v\,T} = \sigma\sqrt{\frac{v\,T}{V}}.$$

Noting that $v = n/T$, we obtain

$$\hat{\mathscr{C}} = \sigma\sqrt{\frac{n}{V}}$$

which coincides with the square-root formula!

We conclude that the square-root formula is consistent with a JG model that has both a square-root market impact function and square-root decay of market impact. From Example 2, we see that the VWAP-equivalent AFS model is also compatible with the square-root formula. Specifically, this AFS model has

$$\psi(\tau) = 1;\ F^{-1}(x) \propto \sqrt{x}$$

which corresponds to a linear order book profile with no resilience; the volume impact E_t is non-decreasing, increasing each time a market buy order arrives. In particular, there is no price reversion after completion of a VWAP execution in this VWAP-equivalent AFS model that is compatible with the square-root formula.

5 Conclusions

First in Proposition 1 we showed that, assuming the price process (1), a model that has a nonlinear market impact function $f(\cdot)$ and a decay kernel $G(\cdot)$ that is non-singular at time zero admits price manipulation. This generalizes the result of [6] on the incompatibility of nonlinear market impact with exponential decay of market impact.

We then showed by explicit computation that exponential order book resilience does not necessarily imply exponential decay of market impact. The JG model of [6] and the AFS model of [1] are different models; even in the special case where both models make identical predictions for the cost of a VWAP execution, predictions of the two models for the decay of market impact post completion of the execution are very different.

A practical corollary of this is that a large database of VWAP (or VWAP-like) executions is not sufficient to fix a reasonable process for the impacted asset price. In Example 4, we presented two quite different dynamical models with identical predictions for the cost of a VWAP execution. In practice, algorithmic executions tend to be all VWAP-like in the sense that they would not permit differentiation between models in this way. The only way to ultimately distinguish between models would be to keep track of the evolution of the asset price *after* completion of each algorithmic execution.

Finally, we noted that there exist both JG and AFS models compatible with the popular square-root formula for pre-trade estimation of market impact although the AFS model predicts no price reversion after completion of the execution. On the other hand, while the JG price process is posited as an obvious extension of previously suggested price processes, AFS models are directly motivated by a simple model of the order book.

References

1. Alfonsi A., Fruth A., Schied A. (2010) Optimal execution strategies in limit order books with general shape functions. Quantitative Finance 10:143–157
2. Alfonsi A., Schied A. (2010) Optimal execution and absence of price manipulations in limit order book models. SIAM J. Financial Math. 1:490–522
3. Alfonsi A., Schied A., Slynko A. (2009) Order book resilience, price manipulations, and the positive portfolio problem. Available at SSRN: papers.ssrn.com/sol3/papers.cfm?abstract_id=1498514
4. Almgren R., Chee T., Hauptmann E., Li H. (2005) Equity market impact Risk Magazine July 2005:57–62
5. Bouchaud J.-P., Gefen Y., Potters M., Wyart M. (2004) Fluctuations and response in financial markets: the subtle nature of 'random' price changes. Quantitative Finance 4:176–190
6. Gatheral J. (2010) No-Dynamic-Arbitrage and Market Impact. Quantitative Finance 10: Forthcoming
7. Gatheral J., Schied A., Slynko A. (2010) Transient linear price impact and Fredholm integral equations. Available at SSRN: papers.ssrn.com/sol3/papers.cfm?abstract_id=1531466
8. Obizhaeva A., Wang J. (2005) Optimal trading strategy and supply/demand dynamics. Available at SSRN: papers.ssrn.com/sol3/papers.cfm?abstract_id=686168

Part III
Miscellaneous

Modeling the Non-Markovian, Non-stationary Scaling Dynamics of Financial Markets

Fulvio Baldovin, Dario Bovina, Francesco Camana and Attilio L. Stella

Abstract. A central problem of Quantitative Finance is that of formulating a probabilistic model of the time evolution of asset prices allowing reliable predictions on their future volatility. As in several natural phenomena, the predictions of such a model must be compared with the data of a single process realization in our records. In order to give statistical significance to such a comparison, assumptions of stationarity for some quantities extracted from the single historical time series, like the distribution of the returns over a given time interval, cannot be avoided. Such assumptions entail the risk of masking or misrepresenting non-stationarities of the underlying process, and of giving an incorrect account of its correlations. Here we overcome this difficulty by showing that five years of daily Euro/US-Dollar trading records in the about three hours following the New York market opening, provide a rich enough ensemble of histories. The statistics of this ensemble allows to propose and test an adequate model of the stochastic process driving the exchange rate. This turns out to be a non-Markovian, self-similar process with non-stationary returns. The empirical ensemble correlators are in agreement with the predictions of this model, which is constructed on the basis of the time-inhomogeneous, anomalous scaling obeyed by the return distribution.

1 Introduction

The analysis of many natural and social phenomena is hindered by the fact that one cannot replicate the dynamical evolution of the system under study. This may happen, for instance, for earthquakes [1], solar flares [2], large eco-systems [3], and

Fulvio Baldovin, Dario Bovina, Francesco Camana, Attilio L. Stella
Department of Physics, INFN e CNISM, University of Padua, Padua, Italy
e-mail: baldovin@pd.infn.it; bovina@pd.infn.it; camana@pd.infn.it; stella@pd.infn.it

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

financial markets [4]. If with a single time series available we try to accommodate the historical data within a stochastic process description, we must assume a priori the existence of some statistical quantities which remain stable over time [4]. This entitles us to sample their values at different stages of the historical evolution, rather than at different instances of the process. For example, in the analysis of historical series in Finance it is usual to assume the stationarity of the distribution of return fluctuations and hence to detect their statistical features through sliding time interval empirical sampling. However, the plausible [5–9] nonstationarity of these fluctuations at intervals ranging from minutes to months would drastically alter the relation between some of the stylized empirical facts detected in this way, and the underlying stochastic process. In order to identify the correct model, one has to overcome this difficulty. The breaking of time-translation invariance possibly signalled by increments non-stationarity would represent a challenge in itself, being a genuine manifestation of dynamics out of equilibrium, like the aging properties observed in glassy systems [10].

In order to detect the possible presence of nonstationarity at certain time-scales for the distribution of the increments, one would need to have access to many independent realizations of the same process, repeated under similar conditions. Quite remarkably, high-frequency financial time-series offer an opportunity of this kind, in which it is possible to directly sample an ensemble of histories. In [7] it has been proposed that when considering high-frequency EUR/USD exchange rate data as recorded during the first three hours of the New York market activity, independent process realizations can tentatively be identified in the daily repetitions of the trading. This gives the interesting possibility of estimating quantities related to ensemble, rather than time-averages. Here we profit of this opportunity by showing that a proper analysis of the statistical properties of this ensemble of histories naturally leads to the identification and validation of an original stochatic model of market evolution. The main idea at the basis of this model is that the scaling properties of the return distribution are sufficient to fully characterize the process in the time range within which they hold. The same type of model has been recently proposed by some of the present authors to underlie more generally the evolution of financial indices also in cases when only single realizations are available [5]. In those cases the application of the model is less direct, and rests on suitable assumptions about the relation between the stationarized empirical information obtainable from the historical series and the underlying driving process.

An interesting feature of the model discussed here and in [5,6], is that the anomalous scaling of the return PDF enters in its construction on the basis of a property of correlated stability which generalizes the stability of Gaussian PDF's under independent random variables summation. This correlated stability was shown recently to allow the derivation of novel, constructive limit theorems for the PDF of sums of many strongly dependent random variables obeying anomalous scaling [11]. In this perspective, the model we present offers a valid alternative to more standard models of Finance based on Gaussianity and independence. At the same time, the probabilistic framework provided by our modelization presents clear formal analogies and parallels with those standard models.

2 An Ensemble of Histories Based on the Returns of the EUR/USD Exchange Rate

To address the above points, given the EUR/USD exchange rate at time t (t measured in tens of minutes) after 9.00 am New York time, $S(t)$, let us define the return in the interval $[t-T, t]$ as $R(t,T) \equiv \ln S(t) - \ln S(t-T)$, where $t = 1, 2, \ldots, t \geq T$. By storing the daily repetitions of the returns from March 2000 to March 2005, we obtain an ensemble of $M = 1,282$ realizations $\{r^l(t,T)\}_{l=1,2,\ldots,M}$ of the discrete-time stochastic process $R(t,T)$, with t ranging in almost three hours after 9.00 am NY time, i.e., $1 \leq t \leq 17$. Below, the superscript "e" labels quantities empirically determined on the basis of this ensemble. The first key observation is that the empirical second moment $m_2^e(t,1) \equiv \sum_{l=1}^{M} [r^l(t,1)]^2/M$ systematically decreases as a function of t in the interval considered (see Fig. 1a). This is a clear indication of return non-stationarity of the underlying process at this time scale. In addition, an analysis of the nonlinear moments m_α^e of the total return $R(t,t) = \ln S(t) - \ln S(0)$ for $t \geq 1$,

$$m_\alpha^e(t,t) \equiv \frac{1}{M} \sum_{l=1}^{M} \left| r^l(t,t) \right|^\alpha, \quad \alpha \in \mathbb{R}_+, \tag{1}$$

shows that such a nonstationarity is accompanied by an anomalous scaling symmetry. Indeed, to a good approximation one finds $m_\alpha(t,t) \sim t^{\alpha D}$ in this range of t, where $D \simeq 0.364\ldots$ is essentially independent of α (Fig. 1b). Accordingly, the ensemble histograms for the PDF's of aggregated returns in the intervals $[0,t]$, $p_{R(t,t)}$, are consistent with the scaling collapse

$$t^D \; p_{R(t,t)} \left(t^D \; r \right) = g(r) \tag{2}$$

reported in Fig. 2. The scaling function g identified by such collapse plot is manifestly non-Gaussian. It may also be assumed to be even to a good approximation[1].

To further simplify our formulas below, wherever appropriate we will switch to the notations: $R_i \equiv R(i,1)$ and $r_i \equiv r(i,1)$. Similarly $r_i^l \equiv r^l(i,1)$ will indicate the i-th return on a 10 min-scale in the l-th history realization of our ensemble.

An important empirical fact (Fig. 1c) is that the linear correlation between returns for non-overlapping intervals

$$c_{lin}^e(1,n) \equiv \frac{\frac{1}{M} \sum_{l=1}^{M} \left[r_1^l \; r_n^l \right]}{\sqrt{m_2(1,1) \; m_2(n,1)}}, \tag{3}$$

with $n = 2, \ldots$, is negligible in comparison with the correlation of the absolute values of the same returns. At this time scale also correlators of odd powers of a return with odd or even powers of another return are negligible. Only even powers of the returns are strongly correlated.

[1] We have detrended the data by subtracting from $r^l(t,T)$ the average value $\sum_{l=1}^{M} r^l(t,T)/M$. Data skewness can be shown to introduce deviations much smaller than the statistical error-bars in the analysis of the correlators.

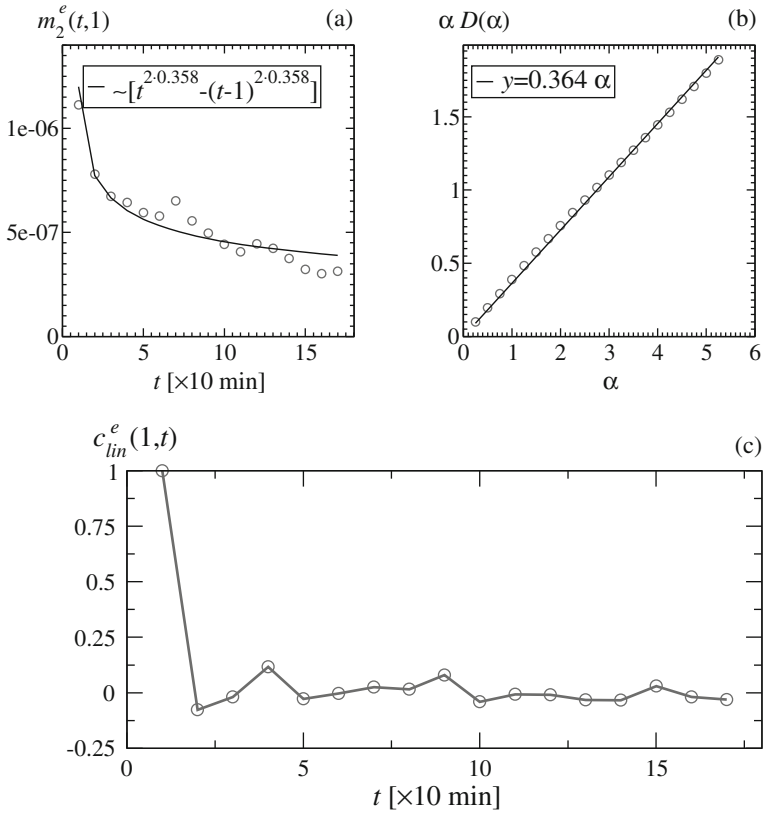


Fig. 1 Empirical ensemble analysis of the returns. (a) The line is given by $\langle\sigma^2\rangle_\rho\ \left[t^{2D}-(t-1)^{2D}\right]$, with $\langle\sigma^2\rangle_\rho = \langle r_1^2\rangle_p = 2.3\cdot10^{-7}$ and the best-fitted $D = 0.358$. (b) Analysis according to the ansatz in Eq. (2). The *straight line* characterizes a simple-scaling behavior with a best-fitted $D = 0.364$. (c) The linear correlation vanishes for non-overlapping returns

3 Self-similar Model Process

The empirical facts listed above already enable us to suggest a very plausible model for the stochastic process expected to generate the data. Both in physics and in Finance, a well established trend in modeling anomalous scaling is that of expressing the scaling functions, like our g, as convex combinations of Gaussian PDF's with varying widths. This has clear mathematical advantages, since it is possible to express very general scaling functions with such convex combinations. In physics the representation in terms of mixtures of Gaussians often reflects the presence of some heterogeneity or polydispersity in the problem [12]. In Finance, the use of convex combinations of Gaussians to represent return PDF's is naturally suggested by the fact that return time series show a variety of more or less long intervals characterized

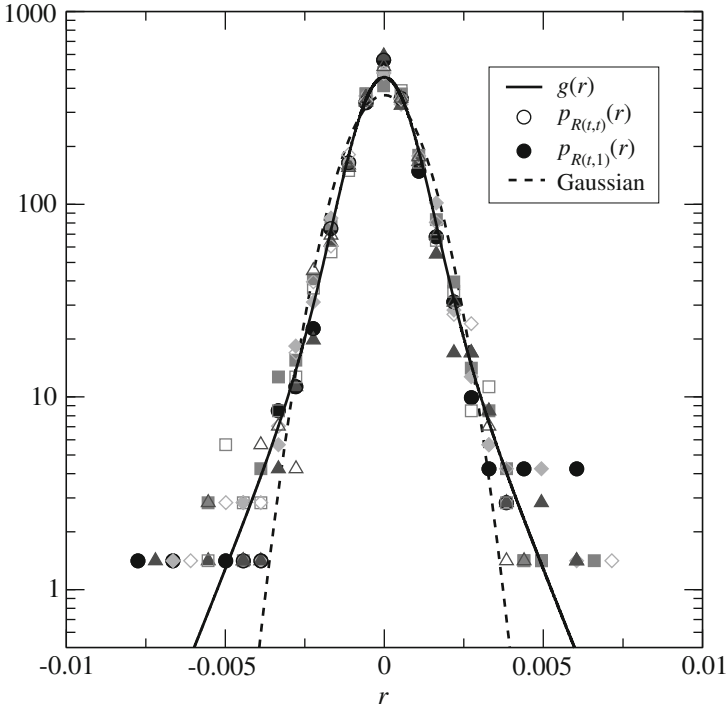


Fig. 2 Non-Gaussian scaling function g. *Empty* [*full*] *symbols* are obtained by rescaling $p_{R(t,t)}$ [$p_{R(t,1)}$] according to Eq. (2) [Eq. (8)] for $t = 1, 5, 10, 17$

by peculiar values of the volatility (volatility clustering). The idea that $p_{R(t,t)}$ can be represented as a mixture of Gaussians of varying widths is suggested by the same basic motivations which lead to the introduction of stochastic volatility models in Finance [13–16]. In the light of the empirical facts, such a representation of the scaling function in the PDF of the aggregated return naturally suggests an adequate full modelization of the process generating the successive partial returns. Let us indicate by $\rho(\sigma)$ a normalized, positive measure in $]0, +\infty[$ such that we can represent g as:

$$g(x) = \int_0^{+\infty} d\sigma \rho(\sigma) \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}. \tag{4}$$

A suitable form of ρ can be easily identified, e.g. by matching its moments with those of g, and by relating the large σ behavior of $\rho(\sigma)$ with the large $|r|$ behavior of $g(r)$. For instance, ρ may decay as a power law at large σ's if the moments of g are expected to be infinite above a given order. These conditions enable us to fix a number of parameters in ρ such that the scaling function in Eq. (4) fits the data in the empirical collapse in Fig. 2. As discussed below, in our case the set of

data on which we can count to construct histograms of g is relatively poor. So, our determinations of ρ will be rather qualitative.

Once identified ρ, more ambitiously we may try to use it for a weighted representation of the joint PDF's of the successive elementary returns R_i, $i = 1, 2, \ldots$ generated in the process. Indeed, we may tentatively write the joint PDF of these returns in the following form:

$$p^n(r_1, r_2, \ldots, r_n) = \int_0^{+\infty} d\sigma \rho(\sigma) \prod_{i=1}^{n} \frac{\exp\left(-\frac{r_i^2}{2a_i^2\sigma^2}\right)}{\sqrt{2\pi\, a_i^2\, \sigma^2}}, \tag{5}$$

with $n = 1, 2, \ldots, 17$. The coefficients a_i in the last equation have to be chosen consistent with the non-stationarity of the elementary returns reported in Fig. 1a and with the other statistical properties of the elementary and aggregated returns discussed in the previous section. It is straightforward to realize that $\langle r_i^2 \rangle_p = \langle \sigma^2 \rangle_\rho\, a_i^2$, while $\langle r_i \rangle_p = 0$ and $\langle r_i r_j \rangle_p = 0$ for $i \neq j$, where $\langle \cdot \rangle_p$ denotes averages with respect to the joint PDF in Eq. (5), whereas $\langle \cdot \rangle_\rho$ those with respect to the PDF ρ. Likewise, we immediately realize that odd-odd or odd-even correlators of the R_i's are strictly zero. Assuming validity of Eq. (5) means in first place that the i-dependence of a_i must be chosen such to fit the values reported in Fig. 1a. The choice of the i dependence of a_i must be also consistent with the simple scaling of the PDF of aggregated returns. Indeed, taking into account that $R(t, t) = R_1 + R_2 + \cdots + R_t$, for $t = 1, 2, \ldots, 17$, Eq. (5) implies that for the same t values

$$p_{R(t,t)}(r) = \frac{g\left(r/\sqrt{a_1^2 + a_2^2 + \cdots + a_t^2}\right)}{\sqrt{a_1^2 + a_2^2 + \cdots + a_t^2}}. \tag{6}$$

Comparing this result with Eq. (2), we see that it is necessary to choose the a_i's such that $a_1^2 + a_2^2 + \cdots + a_t^2 = t^{2D}$ in order to be consistent with the empirical scaling in Eq. (2). This last requirement is satisfied if we put

$$a_i = \sqrt{i^{2D} - (i-1)^{2D}}, \quad i = 1, 2, \ldots\,. \tag{7}$$

A first problem is then to see whether this form of the a_i coefficients is compatible with the i-dependence already implied by the non-stationarity. Eq. (7) appears to be reasonably well compatible with the trend of the empirical mean square elementary returns $m_2(i, 1)$. Indeed, given $\langle \sigma^2 \rangle_\rho = \langle r_1^2 \rangle_p = 2.3 \cdot 10^{-7}$, the best fit in Fig. 1a is obtained with $D = 0.358\ldots$ in the expression for $\langle r_i^2 \rangle_p$. The expectation value of σ^2 is with respect to the ρ entering the integral representation (3) already chosen for g. Remarkably, the value of D is very close to the estimate of D obtained above through the analysis of the moments of $p_{R(t,t)}$.

Summarizing, Eq. (5) and the above conditions on the a_i's define a non-Markovian stochastic process with linearly uncorrelated increments and a PDF of returns satifying a time inhomogeneous scaling of the form:

$$p_{R(t,T)}(r) = \frac{1}{\sqrt{t^{2D} - (t-T)^{2D}}}\ g\left(\frac{r}{\sqrt{t^{2D} - (t-T)^{2D}}}\right), \tag{8}$$

where both t and T are understood to be integer multiples of the 10 minutes unit. In Fig. 2 it is shown that the data collapse of both $p_{R(t,t)}$ and $p_{R(t,1)}$ are indeed compatible with the same non-gaussian PDF g.

From the point of view of probability theory, the structure of our process in Eq. (5) rests on a stability property for PDF's of sums of dependent random variables [11]. Indeed, if we indicate by $\tilde{p}^n(k_1, k_2, ...k_n)$ the Fourier transform (characteristic function) of the joint PDF of the first n returns ($1 \leq n \leq 17$), a direct calculation yields

$$\tilde{p}^n(k, k, ..., k) = \tilde{p}^1\left(n^D K\right) \tag{9}$$

and

$$\tilde{p}^n(0, .., k_i, .., 0) = \tilde{p}^1(a_i k_i), \quad i = 1, \ldots, n. \tag{10}$$

For $D = 1/2$ these relations have the the same form as those holding in the case of independent variables, when $\tilde{p}^n(k_1, \ldots, k_n) = \tilde{p}^1(k_1)\,\tilde{p}^1(k_2)\,\ldots\,\tilde{p}^1(k_n)$, and $\tilde{p}^1$ is a Gaussian characteristic function. However, even for $D = 1/2$ a general $\rho(\sigma)$ implies dependence of the R_i's. To recover the independent case one needs further to choose $\rho(\sigma) = \delta(\sigma - \sigma_0)$. Thus, the superposition of independent Gaussian processes with different σ's in Eq. (5) implies an extension of the basic stability properties of the independent Gaussian variables case to the dependent case. This extension also allows to derive limit theorems for the anomalous scaling of sums of many dependent random variables [11].

4 Correlations Structure

As discussed above, the identification of ρ may be used to reconstruct the joint PDF of the returns R_i's as in Eq. (5). In this section we elaborate further on this point, by performing a detailed comparison between model predictions (based on an explicit expression for ρ) and empirical determinations of various two-point correlators.

Considering the data collapse of both $p_{R(t,t)}$ and $p_{R(t,1)}$ in Fig. 2, we propose the following functional form for ρ (see also [11]):

$$\rho(\sigma) = A\,\frac{\sigma^\gamma}{d + \sigma^\delta}, \quad \sigma \in [\sigma_{min}, +\infty[, \quad 0 < \gamma < \delta, \tag{11}$$

where A is a normalization factor, and $d > 0$ is a parameter influencing the width of the distribution g. Notice that $\rho(\sigma) \sim \sigma^{-(\delta-\gamma)}$ for $\sigma \gg 1$. The rational behind this choice for ρ is that one can use the exponents γ, δ to reproduce the large $|x|$ behavior of $g(x)$, and then play with the other parameters to obtain a suitable fit of the scaling function, for instance the one reported in Fig. 2.

The first two-point correlator we consider in our analysis is

$$\kappa_{\alpha,\beta}(1,n) \equiv \frac{\langle |R(1,1)|^{\alpha} \ |R(n,1)|^{\beta} \rangle}{\langle |R(1,1)|^{\alpha} \rangle_p \ \langle |R(n,1)|^{\beta} \rangle_p} = \frac{\langle |r_1|^{\alpha} \ |r_n|^{\beta} \rangle_p}{\langle |r_1|^{\alpha} \rangle_p \ \langle |r_n|^{\beta} \rangle_p}, \tag{12}$$

with $n > 1$, and $\alpha, \beta \in \mathbb{R}_+$. A value $\kappa_{\alpha,\beta} \neq 1$ means that returns on non-overlapping intervals are dependent. Using Eq. (5) it is possible to express a general many-return correlator in terms of the moments of ρ. For example, from Eq. (5) we have

$$\langle |r_1|^{\alpha} \ |r_n|^{\beta} \rangle_p = B_{\alpha} \, B_{\beta} \ \ a_1^{\alpha} \, a_n^{\beta} \ \langle \sigma^{\alpha+\beta} \rangle_{\rho}, \tag{13}$$

with

$$B_{\alpha} \equiv \int_{-\infty}^{+\infty} dr \ |r|^{\alpha} \ \frac{e^{-r^2/2}}{\sqrt{2\pi}}. \tag{14}$$

We thus obtain

$$\kappa_{\alpha,\beta}(1,n) = \frac{\langle \sigma^{\alpha+\beta} \rangle_{\rho}}{\langle \sigma^{\alpha} \rangle_{\rho} \ \langle \sigma^{\beta} \rangle_{\rho}} = \frac{B_{\alpha} B_{\beta}}{B_{\alpha+\beta}} \frac{\langle |r_1|^{\alpha+\beta} \rangle_p}{\langle |r_1|^{\alpha} \rangle_p \ \langle |r_1|^{\beta} \rangle_p}. \tag{15}$$

Two model-predictions in Eq. (15) are: (i) Despite the non-stationarity of the increments R_i's, $\kappa_{\alpha,\beta}(1,n)$ is independent of n; (ii) The correlators are symmetric, i.e., $\kappa_{\alpha,\beta} - \kappa_{\beta,\alpha} = 0$.

We can now compare the theoretical prediction of the model for $\kappa_{\alpha,\beta}(1,n)$, Eq. (15), with the empirical counterpart

$$\kappa^e_{\alpha,\beta}(1,n) \equiv \frac{\sum_{l=1}^{M} \left[\left| r_1^l \right|^{\alpha} \ \left| r_n^l \right|^{\beta} \right]}{\frac{1}{M} \sum_{l=1}^{M} \left| r_1^l \right|^{\alpha} \ \sum_{l=1}^{M} \left| r_n^l \right|^{\beta}}, \tag{16}$$

which we can calculate from the EUR/USD dataset. Notice that once ρ is fixed to fit the one-time statistics in Fig. 2, in this comparison we do not have any free parameter to adjust. Also, since our ensemble is restricted to $M = 1,282$ realizations only, large fluctuations, especially in two-time statistics, are to be expected.

Fig. 3 shows that indeed non-overlapping returns are strongly correlated in the about three hours following the opening of the trading session, since $\kappa^e_{\alpha,\beta} \neq 1$. In addition, the constancy of $\kappa^e_{\alpha,\beta}$ is clearly suggested by the empirical data. In view of this constancy, we can assume as error-bars for $\kappa^e_{\alpha,\beta}$ the standard deviations of the sets $\{\kappa^e_{\alpha,\beta}(1,n)\}_{n=2,3,\ldots,17}$. The empirical values for $\kappa^e_{\alpha,\beta}$ are also in agreement with the theoretical predictions for $\kappa_{\alpha,\beta}$ based on our choice for ρ. In this and in the following comparisons it should be kept in mind that, although not explicitly reported in the plots, the uncertainty in the identification of ρ of course introduces an uncertainty in the model-predictions for the correlators.

In Fig. 4 we report that also the symmetry $\kappa_{\alpha,\beta} = \kappa_{\beta,\alpha}$ is emipirically verified for the EUR/USD dataset. The validity of this symmetry for a process with non-stationary increments like the present one is quite remarkable.

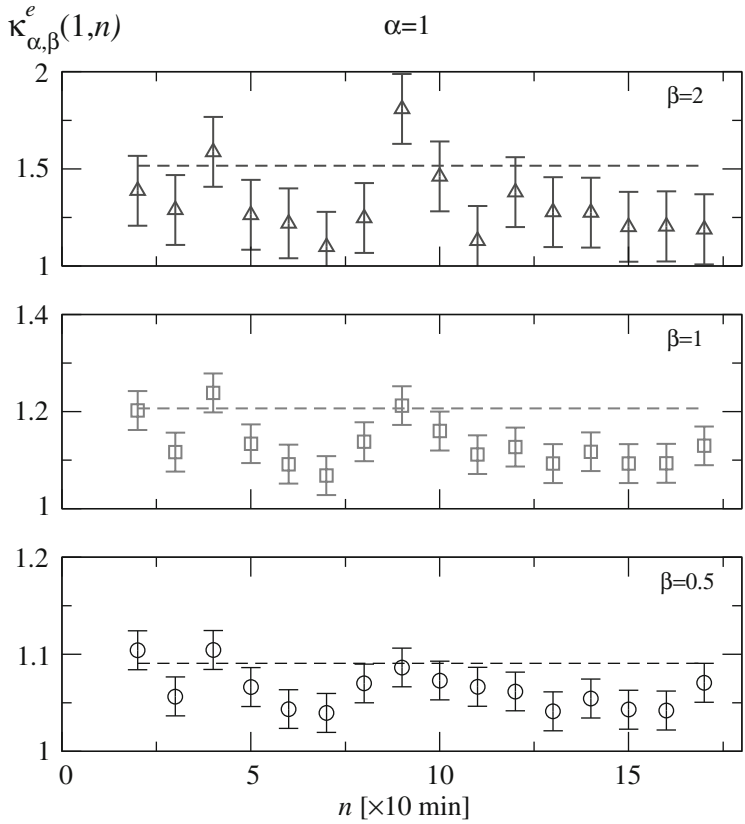


Fig. 3 Constancy of $\kappa^e_{\alpha,\beta}$. *Dashed lines* are model-predictions

A classical indicator of strong correlations in financial data is the volatility autocorrelation, defined as

$$c(1,n) \equiv \frac{\langle |r_1|\ |r_n|\rangle_p - \langle |r_1|\rangle_p\ \langle |r_n|\rangle_p}{\langle |r_1|^2\rangle_p\ - \langle |r_1|\rangle_p^2}. \tag{17}$$

In terms of the moments of ρ, through Eq. (13) we have the following expression for c :

$$c(1,n) = \frac{B_1^2\ a_1\ a_n\ \left[\langle\sigma^2\rangle_\rho - \langle\sigma\rangle_\rho^2\right]}{a_1^2\ \left[B_2\ \langle\sigma^2\rangle_\rho - B_1^2\langle\sigma\rangle_\rho^2\right]}. \tag{18}$$

Unlike $\kappa_{\alpha,\beta}$, c is not constant in n. The comparison with the empirical volatility autocorrelation,

$$c^e(1,n) \equiv \frac{\sum_{l=1}^M \left[\left|r_1^l\right|\ \left|r_n^l\right|\right] - \frac{1}{M}\sum_{l=1}^M \left|r_1^l\right|\ \sum_{l'=1}^M \left|r_n^{l'}\right|}{\sum_{l=1}^M \left|r_1^l\right|^2 - \frac{1}{M}\sum_{l=1}^M \left|r_1^l\right|\ \sum_{l'=1}^M \left|r_1^{l'}\right|}, \tag{19}$$

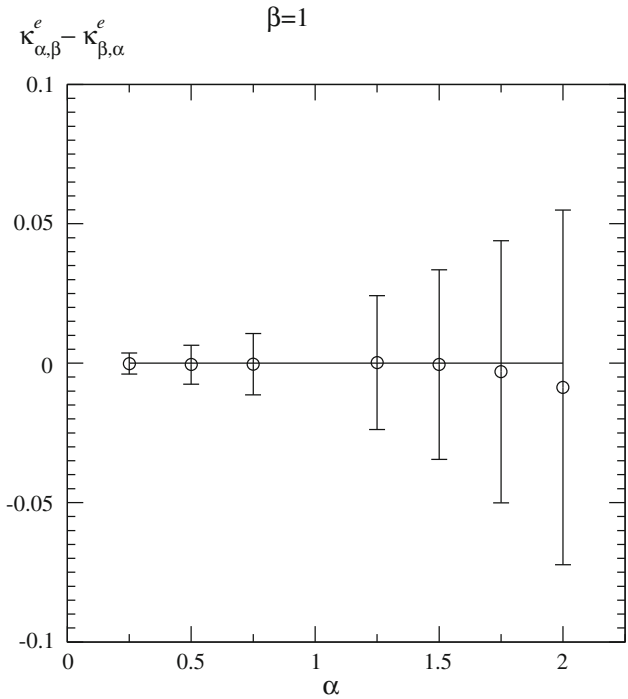


Fig. 4 Symmetry of $\kappa^e_{\alpha,\beta}$. Error-bars are determined as in Fig. 3

yields a substantial agreement (see Fig. 5). The error-bars in Fig. 5 are obtained by dynamically generating many ensembles of $M = 1,282$ realizations each, according to Eq. (5) with our choice for ρ, and taking the standard deviations of the results. Again, the uncertainty associated to the theoretical prediction for c is not reported in the plots Problems concerning the numerical simulation of processes like the one in Eq. (5) are discussed in [11].

A further test of our model can be made by analyzing, in place of those of the increments, the non-linear correlators of $R(t,t)$, with varying t. To this purpose, let us define

$$K_{\alpha,\beta}(t_1,t_2) \equiv \frac{\langle |R(t_1,t_1)|^\alpha \ |R(t_2,t_2)|^\beta \rangle}{\langle |R(t_1,t_1)|^\alpha \rangle \ \langle |R(t_2,t_2)|^\beta \rangle}, \tag{20}$$

with $t_2 \geq t_1$. Model calculations similar to the previous ones give, from Eq. (5),

$$K_{\alpha,\beta}(t_1,t_2) = \frac{B^{(2)}_{\alpha,\beta}(t_1,t_2)}{t_1^{\alpha D} t_2^{\beta D} B_{\alpha+\beta}} \frac{\langle |r_1|^{\alpha+\beta} \rangle_p}{\langle |r_1|^\alpha \rangle_p \ \langle |r_1|^\beta \rangle_p}, \tag{21}$$

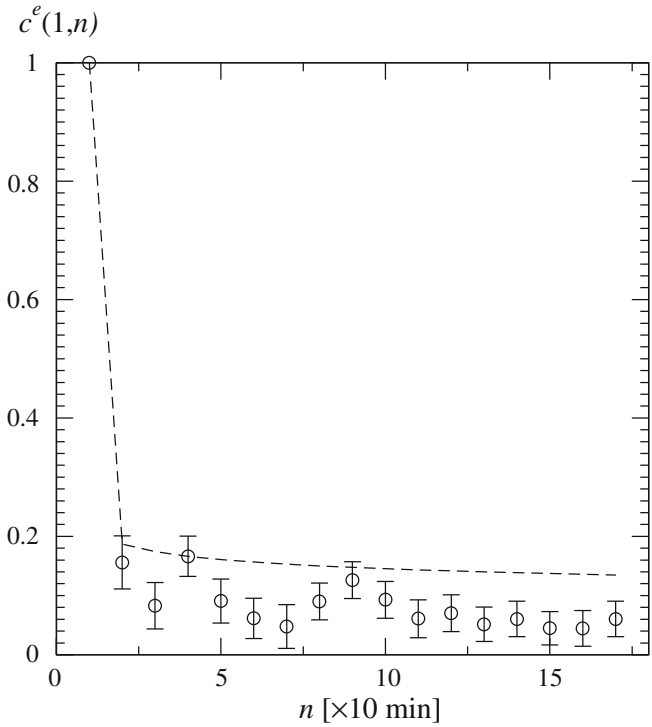


Fig. 5 Volatility autocorrelation. *Dashed line* is the model-prediction

where

$$B^{(2)}_{\alpha,\beta}(t_1,t_2) \equiv \int_{-\infty}^{+\infty} dr_1\ |r_1|^{\alpha}\ \frac{\exp\left(-r_1^2/\left(2t_1^{2D}\right)\right)}{\sqrt{2\pi t_1^{2D}}}$$
$$\int_{-\infty}^{+\infty} dr_2\ |r_2|^{\beta}\ \frac{\exp\left[-(r_1-r_2)^2/\left(2t_2^{2D}-2t_1^{2D}\right)\right]}{\sqrt{2\pi\left(t_2^{2D}-t_1^{2D}\right)}}. \tag{22}$$

According to Eq. (21), $K_{\alpha,\beta}$ is now identified by both ρ and D. Moreover, it explicitly depends on t_1 and t_2. The comparison between Eq. (21) and the empirical quantity

$$K^e_{\alpha,\beta}(t_1,t_2) \equiv \frac{\sum_{l=1}^{M}\left[\left|r^l(t_1,t_1)\right|^{\alpha}\ \left|r^l(t_2,t_2)\right|^{\beta}\right]}{\frac{1}{M}\sum_{l=1}^{M}\left[\left|r^l(t_1,t_1)\right|^{\alpha}\right]\ \sum_{l=1}^{M}\left[\left|r^l(t_2,t_2)\right|^{\beta}\right]}, \tag{23}$$

reported in Fig. 6 (the error-bars are determined as in Fig. 5) supplies thus an additional validation of our model.

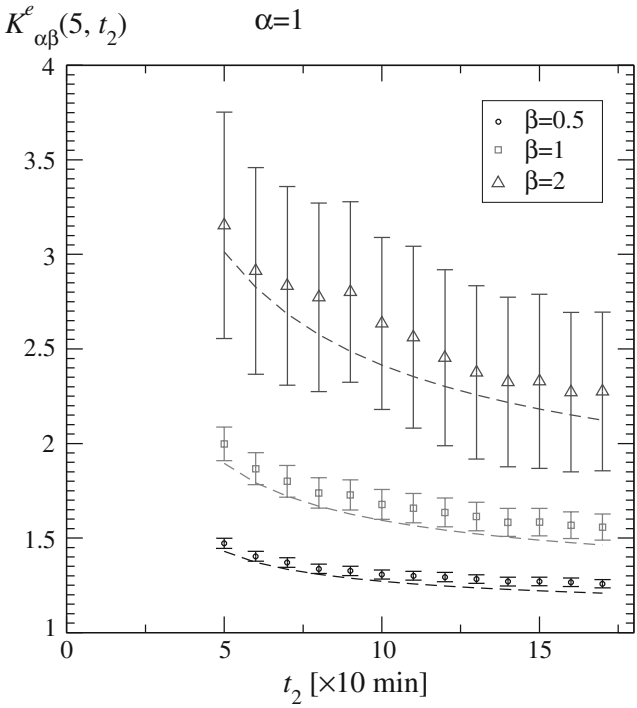


Fig. 6 Correlators $K^e_{\alpha,\beta}$. *Dashed lines* are model-predictions

5 Conclusions

In the present work we addressed the problem of describing the time evolution of financial assets in a case in which one can try to compare the predictions of the proposed model with a relatively rich ensemble of history realizations. Besides the fact that considering the histories at disposal for the EUR/USD exchange rate as a proper ensemble amounts to a main working assumption, a clear limitation of such an approach is the relative poorness of the ensemble itself. Indeed, the simulations of our model suggest that in order to reduce substantially the statistical fluctuations one should dispose of ensembles larger by at least one order of magnitude.

In spite of these limitations, we believe that the non-Markovian model we propose [5, 6, 11] is validated to a reasonable extent by the analysis of the data, especially those pertaining to the various correlators we considered. In this respect it is important to recall that the first proposal of the time inhomogeneous evolution model discussed here has been made in a study of a single, long time series of the DJI index in [5]. In that context, the returns time inhomogeneity, Eq. (8), was supposed to underlie the stationarized information provided by the empirical PDF of the returns. This assumption allowed there to give a justification of several stylized facts, like the scaling and multiscaling of the empirical return PDF and the power

law behavior in time of the return autocorrelation function. We believe that the results obtained in the present report, even if pertaining to a different time-scale (tens of minutes in place of days), constitute an interesting further argument in favor of a general validity of the model.

The peculiar feature of this model is that of focussing on scaling and correlations as basic, closely connected properties of assets evolution. This was strongly inspired by what has been learnt in the physics of complex systems in the last decades [17–19], where methods like the renormalization group allowed for the first time systematic treatments of these properties [6]. At the same time, through the original probabilistic parallel mentioned in Sect. 3, our model maintains an interesting direct contact with the mathematics of standard formulations based on Brownian motion, of wide use in Finance. This last feature is very interesting in the perspective of applying our model to problems of derivative pricing [13–16, 20].

Acknowledgements We thank M. Caporin for useful discussions. This work is supported by "Fondazione Cassa di Risparmio di Padova e Rovigo" within the 2008-2009 "Progetti di Eccellenza" program.

References

1. Scholz CH (2002) The Mechanics of Earthquakes and Faulting. Cambridge University Press, New York
2. Lu ET, Hamilton RJ, McTiernan JM, Bromond KR (1993) Solar flares and avalanches in driven dissipative systems. Astrophys. J. 412: 841–852
3. Bjornstad ON, Grenfell BT (2001) Noisy clockwork: time series analysis of population fluctuations in animals. Science 293: 638–643
4. Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quant. Finance 1: 223–236
5. Baldovin F, Stella AL (2007) Scaling and efficiency determine the irreversible evolution of a market. Proc. Natl. Acad. Sci. 104: 19741–19744
6. Stella AL, Baldovin F (2008) Role of scaling in the statistical modeling of Finance. Pramana – Journal of Physics 71, 341–352
7. Bassler KE, McCauley JL, Gunaratne GH (2007) Nonstationary increments, scaling distributions, and variable diffusion processes in financial markets. Proc. Natl. Acad. Sci. USA 104: 17287–17290
8. D. Challet and P.P. Peirano (2008) The ups and downs of the renormalization group applied to financial time series. arXiv:0807.4163
9. A. Andreoli, F. Caravenna, P. Dai Pra, G. Posta (2010) Scaling and multiscaling in financial indexes: a simple model. arXiv:1006.0155
10. De Benedetti PG, Stillinger FH (2001) Supercooled liquids and the glass transition. Nature 410: 259–267
11. Stella AL, Baldovin F (2010) Anomalous scaling due to correlations: Limit theorems and self-similar processes. J. Stat. Mech. P02018
12. Gheorghiu S, Coppens MO (2004) Heterogeneity explains features of "anomalous" thermodynamics and statistics. Proc. Natl. Acad. Sci. USA 101: 15852–15856
13. Bouchaud JP, Potters M (2000) Theory of Financial Risks Cambridge University Press, Cambridge, UK

14. Heston L (1993) A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. The Review of Financial Studies 6: 327–343
15. Gatheral J (2006) The volatility surface: a practitioner's guide. J. Wiley and Sons, New Jersey
16. Fouque JP, Papanicolaou G, Sircar KR (2000) Derivatives in financial markets with stochastic volatility. Cambridge University Press, Cambridge, UK
17. Goldenfeld N, Kadanoff LP (1999) Simple Lessons from Complexity. Science 284: 87–90
18. Kadanoff LP (2005) Statistical Physics, Statics, Dynamics and Renormalization. World Scientific Singapore
19. Jona-Lasinio G (2001) Renormalization group and probability theory. Phys. Rep. 352: 439–458
20. Black F, Scholes M. (1973) The Pricing of Options and Corporate Liabilities. J. Polit. Econ. 81: 637-654

The von Neumann–Morgenstern Utility Functions with Constant Risk Aversions

Satya R. Chakravarty and Debkumar Chakrabarti

Abstract. Two Arrow–Pratt measures of risk aversion indicate attitudes of the individuals towards risk. In the theory of finance often these measures are assumed to be constant. Using certain intuitively reasonable conditions, this paper develops axiomatic characterizations of the utility functions for which the Arrow–Pratt measures are constant.

1 Introduction

Consumers often have to make choices under conditions of uncertainty. In the context of uncertainty a consumer's preferences can be represented by a utility function that satisfies the expected utility hypothesis, which says that the expected value of the utility function should be maximized. A utility function satisfying the expected utility property is known as the von Neumann–Morgenstern utility function. This utility function has been extensively used to study an individual's attitudes towards risk.

Arrow (1971) and Pratt (1964) suggested two measures of risk aversion, absolute and relative measures, which rely on the von Neumann -Morgenstern utility functions. These measures have been used extensively for analyzing problems on uncertainty. Often it is assumed that the Arrow–Pratt measures of risk aversion are constant. Constancy of these measures has interesting implications in financial economics. For instance, if the measure of absolute risk aversion is constant, then the consumer's wealth, on which the utility function is defined, is not an inferior

Satya R. Chakravarty
Indian Statistical Institute, Kolkata, India,
e-mail: satya@isical.ac.in

Debkumar Chakrabarti
Ramakrishna Mission Vidyamandira, Howrah, India,
e-mail: debkchak@yahoo.com

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

good (Arrow, 1970 and Demange and Laroque,2006). However, there does not exist any characterization of the von Neumann–Morgenstern utility functions that display constant absolute and relative risk aversions.

In this paper we develop characterizations of the von Neumann–Morgenstern utility functions for which the two Arrow–Pratt measures of risk aversion are constant. The characterizations are based on the certainty equivalent The certainty equivalent of an uncertain prospect with state-contingent returns, where each state occurs with certain probability, is defined as that level of return, which if assigned to each state of the prospect will produce the same expected utility as the prospect itself. The certainty equivalent has a negative monotonic relationship with the cost of risk in an uncertain prospect. The cost of risk becomes positive if the situation is characterized by uncertainty. In the absence of uncertainty there is no cost of risk.

After presenting the preliminaries and definitions in the next section, we present the characterization theorems in Sect. 3. Finally, Sect. 4 concludes.

2 Preliminaries and Definitions

For any uncertain prospect P there is a probability distribution on different levels of returns on the prospect. We denote these prospective returns by $r_1, r_2, \ldots, r_n$. Suppose that the consumer assigns a probability p_i to realization of the level of return r_i, $1 \leq i \leq n$. The vector of probabilities $(p_1, p_2, \ldots, p_n)$ is denoted by p, while $r = (r_1, r_2, \ldots, r_n)$ is the vector of probable returns on the prospect P. Clearly, $0 \leq i \leq n$ for all $1 \leq i \leq n$ and $\sum_{i=1}^{n} p_i = 1$. Each r_i is assumed to be drawn from the non-degenerate interval $[a, b]$ in the real line R. We use the compact notation $P = (p, r)$ for the prospect P. The expected return $\sum_{i=1}^{n} p_i r_i$ on the prospect is denoted by $\mu(p, r)$.

Let $U : [a, b] \to R$ be a von Neumann–Morgenstern utility function. Assume that U is twice differentiable and denote the first and second derivatives of U by U' and U'' respectively. Assume further that U is increasing so that $U' > 0$. Existence of U' ensures that U is continuous. Then the expected utility from the prospect P is given by $\sum_{i=1}^{n} p_i U(r_i)$. Given any two prospects $P = (p, r)$ and $P' = (p', r')$, the inequality $\sum_{i=1}^{n} p_i U(r_i) \geq \sum_{i=1}^{n} p'_i U(r'_i)$ remains preserved under positive linear transformations of U That is, if $V = a + bU$, where $b > 0$ and a are constants, then $\sum_{i=1}^{n} p_i U(r_i) \geq \sum_{i=1}^{n} p'_i U(r'_i)$ implies that $\sum_{i=1}^{n} p_i V(r_i) \geq \sum_{i=1}^{n} p'_i V(r'_i)$. That is, the von Neumann–Morgern utility function satisfies the 'cardinally measurable, fully comparable' measurability and comparability assumption (Sen, 1977).

Two questions that generally arise in the context of revelation attitude towards risk of a decision maker are: would he prefer to receive the expected payoff from a prospect with certainty rather than facing the prospect? How much money he would like to give up to avoid the risk involved in the prospect? We can look at the issue in terms of the certainty equivalent of a prospect.

The certainty equivalent of a prospect is that level of return, which if associated with each state of the prospect will give the same level of expected utility as that

generated by the actual prospect. That is, for any person with the von Neumann–Morgern utility function U and for any prospect $P = (p, r)$, we define the certainty equivalent r_e corresponding to $P = (p, r)$ as follows:

$$\sum_{i=1}^{n} p_i U(r_e) = \sum_{i=1}^{n} p_i U(r_i). \tag{1}$$

Since U is increasing, r_e exists and is given by $r_e = U^{-1}\left(\sum_{i=1}^{n} p_i U(r_i)\right)$. Clearly, r_e is an average of the state-contingent returns. It is easy to verify that r_e remains invariant under positive linear transformations of U.

The Arrow–Pratt measure of absolute risk aversion $R_A(W)$ for an individual with utility function U and level of wealth W is defined as

$$R_A(W) = -\frac{U''(W)}{U'(W)}. \tag{2}$$

This measure is positive, zero or negative depending on whether U is strictly concave, linear or strictly convex. That is, for a risk averse individual R_A is positive, whereas it is zero or negative according as the individual is risk neutral or risk lover. A higher value of R_A indicates that the person is more risk averse in absolute sense.

The Arrow–Pratt relative risk aversion measure is defined as

$$R_R(W) = -\frac{WU''(W)}{U'(W)}. \tag{3}$$

The measure R_R takes on positive, zero or negative values according as the individual is risk averse, risk neutral or risk lover. Increasingness of R_R with W means that the individual becomes more relative risk averse as his level of wealth increases. Both R_A and R_R remain invariant under positive linear transformations of U.

3 The Characterization Theorems

Before we characterize the utility functions associated with constant values of R_A and R_R, we demonstrate a property of the certainty equivalent.

Theorem 1 Given the probability vector p, the certainty equivalent r_e is a continuous function of state-contingent returns.

Proof. As we have noted r_e can be written explicitly as $r_e = U^{-1}\left(\sum_{i=1}^{n} p_i U(r_i)\right)$. Since each r_i is drawn from the compact set $[a, b]$, the domain of the function $U(r_i)$ is $[a, b]$. Now, since U is increasing and the continuous image of a compact set is compact (Rudin, 1976, p.89), $U(r_i)$ takes values in the compact set $[U(a), U(b)]$. For a given p, continuity and increasingness of the function U implies that the average function $\sum_{i=1}^{k} p_i U(r_i)$ is continuous and takes values in $[U(a), U(b)]$. Continuity and increasingness of U^{-1} on $[U(a), U(b)]$ now follows from Theorem 4.53 of Apostol (1974, p.95). This in turn demonstrates continuity of r_e. □

For a strictly concave utility function U, by Jensen's inequality (Marshall and Olkin, 1979, p.454) $U\left(\sum_{i=1}^{n} p_i r_i\right) > \left(\sum_{i=1}^{n} p_i U(r_i)\right)$ Hence $\mu(p,r) - r_e > 0$. Given strict concavity of U, the difference

$$C_A(p,r) = \mu(p,r) - r_e \tag{4}$$

can be interpreted as a measure of cost of risk because uncertainty reduces the expected utility from the prospect exactly by this amount. If there is no uncertainty in the prospect, that is, if the prospect is risk-free, then $C_A = 0$. This difference gives us the risk premium that the individual would be willing to pay to avoid investment in the risky prospect. Since this cost measure depends on nominal value of the returns we call it an absolute measure.

Suppose that all state-contingent returns on an uncertain prospect increase or decrease by the same absolute amount. Formally, the state -contingent return vector $(r_1, r_2, \ldots, r_n)$ becomes $(r_1 + c, r_2 + c, \ldots, r_n + c)$, where c is a scalar and the probabilities $(p_1, p_2, \ldots, p_n)$ remain unchanged. Then the expected return on the prospect changes by c. A cost indicator C satisfies translation invariance if its value does not alter under a change of this type, that is, $C(p,r) = C(p, r_1 + c, r_2 + c, \ldots, r_n + c)$ for all allowable values of c. The following theorem shows that the only utility function for which the absolute cost measure C_A satisfies translation invariance is the one with constant absolute Arrow–Pratt measure of risk aversion. Formally,

Theorem 2 The cost measure C_A satisfies translation invariance if and only if the underlying utility function has a constant value of the Arrow–Pratt measure of absolute risk aversion.

Proof. If all the r_i's change by the same amount c, expected return also changes by c. Therefore, translation invariance of C_A demands that it should also change by c Using the expression for r_e we can write this condition as

$$U^{-1}\left(\sum_{i=1}^{n} p_i U(r_1 + c)\right) = U^{-1}\left(\sum_{i=1}^{n} p_i U(r_i)\right) + c. \tag{5}$$

This is a quasuilinear functional equation with only continuous solution given by $U(r_i) = -ae^{-br_i}$, where a and b are constants (Aczel, 1966, p.153). Since U is increasing and strictly concave we must have $a > 0$ and $b > 0$. It is easy to check that the absolute Arrow–Pratt measure of risk aversion for this utility function is the constant b.

Conversely, C_A for the utility function $U(r_i) = -ae^{-br_i}$ is given by $(\log \sum_{i=1}^{n} p_i e^{b(\mu(p,r)-r_i)})/b$. This cost measure is translation invariant. □

While C_A specifies cost of risk in difference form, we can also measure the cost in ratio form. Thus, we can consider the following as a measure of the cost of risk:

$$C_R(p,r) = 1 - \frac{r_e}{\mu(p,r)}. \tag{6}$$

Since C_R is stated in a ratio form it is referred to as a relative measure. If the utility function is strictly concave, C_R is bounded between zero and one, where the lower bound is attained whenever state contingent returns are the same.

The returns are stated as nominal values. Now, suppose that the currency unit of the returns changes from euros to cents. Then since the returns remain unchanged and their probabilities also do not change, we can argue that the cost of risk should not alter. A cost indicator satisfying this postulate is called a scale invariant measure. Formally, we say that a cost measure C is scale invariant if $C(p, r) = C(p, cr)$, where $c > 0$ is a constant. In the following theorem we show that the only utility function for which the cost measure C_R satisfies scale invariance is the one with constant relative Arrow–Pratt measure of risk aversion. Formally,

Theorem 3 The cost measure C_R satisfies scale invariance if and only if the underlying utility function has a constant value of the Arrow–Pratt measure of relative risk aversion.

Proof. Observe that $\mu(p, cr) = c\mu(p, r)$, where $c > 0$. Therefore C_R will be a relative indicator if $r_e(p, cr) = cr_e(p, r)$, where $c > 0$. Using the expression for r_e we can write this condition as

$$cU^{-1}\left(\sum_{i=1}^{n} p_i U(r_i)\right) = U^{-1}\left(\sum_{i=1}^{n} p_i U(cr_i)\right), \tag{7}$$

where $c > 0$ is a scalar. The quasilinear functional equation in (7) has only one continuous solution and it is given by

$$U(r_i) = \begin{cases} a + b\dfrac{r_i^{1-\beta}}{1-\beta}, & \beta \neq 1, \\ a + b\log r_i, & \beta = 0, \end{cases} \tag{8}$$

where a and b are constants (Aczel, 1966, p.153). Since U is increasing and strictly concave we must have that $b > 0$ and $\beta > 0$ in the first expression for $U(r_i)$ in (8). The risk aversion measure R_R for this utility function is β, a positive constant.

Conversely, for the utility function given by (8) the C_R measure is given by

$$C_R(p, r) = \begin{cases} 1 - \dfrac{\left(\sum_{i=1}^{n} p_i r_i^{1-\beta}\right)^{\frac{1}{(1-\beta)}}}{\mu(p,r)}, & \beta > 0, \beta \neq 1, \\ 1 - \dfrac{\prod_{i=1}^{n} (r_i)^{p_i}}{\mu(p, r)}, & \beta = 1. \end{cases} \tag{9}$$

Clearly, the C_R function given (9) by satisfies scale invariance. □

Thus, Theorems 2 and 3 uniquely identify the utility functions for which the values of the absolute and relative measures of risk aversion are constant. In other words, these theorems provide necessary and sufficient conditions for the utility functions that display respectively constant absolute and relative Arrow–Pratt measures of risk aversion.

4 Conclusions

Constancy of the Arrow–Pratt measures of risk aversion is a standard assumption in the theory of finance. While many results in the theory of finance are based on this assumption, there have been no characterizations of the underlying utility functions. In this paper we made an attempt to fill in this gap by developing characterizations of these utility functions using the cost of risk involved in a risky prospect.

Acknowledgements The authors thank Dr. Manipushpak Mitra for heplpful comments.

References

1. Aczel, J. (1966) Lectures on functional equations and their applications. Academic Press, London
2. Apostol, M. (1974) Mathematical analysis. Addison-Wesley, London
3. Arrow, K. (1971) Essays on the theory of risk bearing. Markham, Chicago
4. Demange, G. and Laroque, G. (2006) Finance and the economics of uncertainty, Blackwell, London
5. Marshall, A.W. and Olkin, I. (1979) Inequalities: theory of majorization and its applications, Academic Press, New York
6. Pratt, J. (1964) Risk aversion in the small and the large. Econometrica, 32; 122–136
7. Rudin, W. (1976) Principles of mathematical analysis. McGraw-Hill, London
8. Sen, A (1977) On Weights and measures: information constraints in social welfare analysis. Econometrica 45; 1539–1572

Income and Expenditure Distribution. A Comparative Analysis

Kausik Gangopadhyay and Banasri Basu

Abstract. There are empirical evidences regrading the Pareto tail of the income distribution and the expenditure distribution. We formulate a simple economic framework to study the relation between them. We explain the Pareto tails in both the distributions with a Cobb–Douglas felicity function to describe the preferences of agents. Moreover, the Indian data suggest a thicker Pareto tail for the expenditure distribution in comparison to the income distribution. With a uniform distribution of taste parameters for various goods, we identify a process that can give rise to this empirical phenomenon. We also verify our observation with appropriate simulation results.

1 Introduction

Economics is the science of economic decisions at an individual level as well as at a collective level. Manifold ways of individual decision making give rise to various distributions in the aggregate population as far as economic variables are concerned. Lack of appropriate and detailed micro-data is a constraint in analyzing the economic decisions as made by the decision makers. Often the answer lies in the study of distributions. As income and expenditure are integral elements for individual economic decision making, the distribution of income is a prominent topic for study among the economists and econophysicists. Income and wealth measure, to some extent, potential opportunities available to a particular individual; whereas expenditure is an indicator of realized opportunity of a family or individual. For a normative

Kausik Gangopadhyay
Indian Institute of Management Kozhikode, Kerala, India,
e-mail: kausik@iimk.ac.in

Banasri Basu
Physics and Applied Mathematics Unit, Indian Statistical Institute, Kolkata, India,
e-mail: banasri@isical.ac.in

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

analysis of social conditions, study of income and wealth distribution is vital. For a positive analysis of economic theory, the joint study of income and expenditure distributions is key to exploring the savings decisions of economic decision makers. Moreover, economic decisions are taken over dynamic considerations. An individual's economic decisions are bound to reflect not only her present level of income, but also expected future income levels of that individual. It is well-accepted that expenditure of an individual is more robust to a present shock compared to the income. Therefore, expenditure distribution is potentially more robust over time compared to its income counterpart.

The present literature documents well the study of income distribution over different societies – Australia [1, 2], Brazil [3, 4], China [5], India [6], Japan [7, 8], Poland [9], France [10], Germany [10], United Kingdom [11, 12] and USA [13, 14]. Almost all the works with some exceptions [15] document the evidence of a Pareto tail, named after Italian economist and sociologist Vilfredo Pareto [16], at the right end. The consensus is rather low on the lower tail in these works. Typically, income distribution is portrayed by a log-normal distribution at the lower tail. Nevertheless, the empirical documentation resorts to gamma, generalized beta of the second kind, Weibul or Gompertz as some of the possibility distributaries to model the income data. A synopsis [17] of the entire study points to various competing theories to understand this empirical fact as a physical phenomenon [1, 11, 18]. The Pareto law or the power law is illustrative of an extremely high magnitude of inequality for the fact that the complementary cumulative distribution[1] of the personal income (m) diminishes exponentially. Mathematically speaking, $F^C(m) \sim m^{-(1+\alpha)}$ with the exponent α lying between 1 and 2.

Even though study of expenditure distribution is somewhat mandated by the theoretical demands, the data are far scarce for expenditure compared to income and even wealth. The expenditure distribution is found [19] to follow Pareto law with an exponent of -2 by analyzing a huge point-of-scale (POS) dataset obtained from a convenience store chain in Japan. The analysis [20] of Indian expenditure data collected by the National Sample Survey Organization (NSSO) for the years of 1983–2007 also indicates the existence of a Pareto tail. In this particular case, the expenditure distribution is fitted to be a mixture of log-normal and Pareto distribution at the right tail. The weight of the Pareto tail is in the range of 10 to 20% depending on the particular year and urban or rural nature of the data. The Pareto exponent is in the range of 1.4 to 2.5. It is in contrast to a study [15] using the U.K. and the U.S. data, which finds no such Pareto tail, but expenditure (and also income) distribution as a log-normal distribution. However, it is very much in resonance with a study [21] on the asset distribution of Indian households showing a Pareto law with the exponent ranging from 1.8 to 2.4.

[1] The cumulative density function (CDF) of a random variable X is often denoted by $F(x)$. It is the probability of X attaining all the values lesser than or equal to x and is given by $P(X \leq x)$. The complementary cumulative density function (CCDF) of the random variable X at a point x is the probability of that random variable to attain a value larger than x and is denoted by $F^C(x)$, which is $P(X \geq x)$.

The present article builds a foundation between income and expenditure distributions starting from an individual's preferences. The expenditure occurs when an individual chooses to spend over a good. Each good finds its place in the felicity function of the consumer. In the behavioural set-up, money acts like a good in itself. The savings propensity is modeled along this behavioural formulation [22, 23], in which, an individual holds money for the sake of deriving utility. We derive some properties of the income and expenditure distribution under this baseline formulation. The theoretical properties are analysed and contrasted along with the empirical findings in the literature. Moreover, the emphasis is placed to meet the perceptions from both ends.

When the parameter values are kept constant across agents, it induces constant savings in this standard economic environment. Therefore, any Pareto tail in the income distribution would naturally be transformed to a similarly Pareto tailed expenditure distribution under this framework. From a different perspective, we can not have a thicker tailed expenditure distribution compared to the Pareto tail of the income distribution. The data may indicate to this anomaly. To sort out this anomaly, we model the taste parameters of agents following certain simple distributional assumption. The resulting expenditure distribution is theoretically shown to possess a thicker Pareto tail compared to the corresponding income distribution. We simulate our model with a fixed income for all agents to find that the resulting expenditure distribution indeed possesses a Pareto tail.

In the next section, the theoretical framework is discussed along with pertinent assumptions on agents' tastes and preferences. In Sect. 3, we discuss various cases arising out of our theoretical framework. The simulation results are thereby elaborated. Finally, we conclude our discussion in Sect. 4.

2 Theoretical Framework

Let an agent consume τ goods. All the goods are available in the market with prices $p_1, p_2, \ldots, p_\tau$, respectively. The consumed quantities of these goods are denoted by $g_1, g_2, \ldots, g_\tau$. Moreover, the agent has felicity for saving money. The agent's savings, noted by the random variable s, is another input argument in the utility function of the agent. We make the following assumptions on the agent's preferences.

Assumption *The agent's preferences are continuous.*

The definition of continuity means the following. Given a sequence of pairs of bundles, (x_n, y_n), if first bundle is always (weakly) preferred over the later throughout the sequence, i.e. $x_n \succeq y_n$ for all n, then this property is preserved in the limit as well. In other words, there is no sudden jumps in the agent's preferences, which is considered somewhat intuitive as far as the preferences are considered. If the preferences are continuous, the utility function must be continuous. Moreover, any continuous function could be approximated by a suitable polynomial. A utility function is

a cardinal concept and therefore could be positive without violating any fundamental premises of the utility theory. We conclude that the utility function is a positive polynomial. The contribution of different inputs to this function is related to substitutability of various inputs. A benchmark case is illustrated with Cobb–Douglas utility function. In this particular case, the substitutability of the inputs is proportional to the relative prices of the inputs. We write the mathematical form of the utility function as,

$$u\left(g_1, g_2, \ldots, g_\tau, s\right) = g_1^{\delta_1} \cdot g_2^{\delta_2} \cdots g_\tau^{\delta_\tau} \cdot s^{\delta_s}, \tag{1}$$

where $\delta_1, \delta_2, \ldots, \delta_\tau$ are parameters indicating the significance of each of the goods in the felicity function of the agent. The savings propensity parameter is denoted by δ_s. We assume that $1 > \delta_i > 0$ for all $i = 1, 2, \ldots, \tau$ and for $i = s$. This assumption is totally innocuous. If $\delta_j < 0$ for some j, it means that j^{th} good is not preferred and less of the good is better. Therefore that particular good would not be consumed and we can exclude it from our consideration. Without loss of generality, we can assume that $\delta_s + \sum_{i=1}^{\tau} \delta_i = 1$.[2]

In a particular case, with with two goods and no savings (*i.e.* $\tau = 2$ and $\delta_s = 0$) [24], where all agents are endowed with different types of goods and agents have a preference for consumption of all the goods, the mechanism of trade occurs like a kinetic exchange model for gases. The income distribution of agents with the proposed utility function is found to follow a Pareto upper tail under suitable restrictions on the values of parameters. Our effort goes for a more general scenario where the occurrence of trade can be modeled with the same utility function including expenditure and savings.

The economic environment is static in nature. The holding of money addresses the concern for the dynamic necessities to a certain extent. An agent holds money [23] not only for static considerations, but also as an asset against future uncertainties. An agent receives an income of size m and maximizes her utility described by Eq. (1) subject to the following budget constraint,

$$p_1 \cdot g_1 + p_2 \cdot g_2 + \cdots p_\tau \cdot g_\tau + s \leq m \tag{2}$$

where s is the agent's savings as specified. The agent chooses $g_1, g_2, \ldots, g_\tau$ as well as s and maximizes her utility. The utility function is increasing in all these arguments. Therefore, an agent shall choose more and more of all the goods and savings if she has a free hand. And, that is where the budget constraint described in Eq. (2) becomes pertinent. Due to the constraint on resources imposed by the total income, the agent is compelled to consume limited amount of resources of each kind. She has to consume different goods in such a manner that consumption of each good, quantity-wise, deserves the due importance of the good placed in the utility function. In other words, the agent solves an optimization exercise, which is based on the principle that on the margin, equal amount of money spent on each

[2] If $\delta_s + \sum_{i=1}^{\tau} \delta_i$ is not equal to 1, we can normalize the δ_is through division of all the δ_is by this sum. It will not change any of the results to be obtained from the model.

good should bring about same change in utility. The marginal utility of the ith good is the change in utility per additional unit of good i consumed, when an infinitesimal amount of ith good is consumed to the current consumption bundle. The agent pays an amount of p_i per unit of ith good for that matter. Therefore the change in utility per unit of money is the marginal utility of the ith good divided by p_i for the agent. To maximize utility, the agent should choose such a consumption bundle that the marginal utility of the ith good divided by p_i is same for all the goods.

The *equilibrium* of this model is defined as a bundle of commodities and a savings decision for the agent that *maximizes* her utility subject to the budget constraint noted in Eq. (2). This decision is contingent on the values of the taste parameters, for example, $\delta_1, \delta_2, \ldots, \delta_\tau$, and δ_s of the agent as well as the income of the agent, m. In other words, the equilibrium of the model is best described as a function, which is mapped from the domain of the space of parameters, $\delta_1, \delta_2, \ldots, \delta_\tau, \delta_s$, and m to the range of space of consumption bundles for all the commodities.

Mathematically speaking, let mu_i be the marginal utility of the ith good in equilibrium. This can be calculated by considering the derivative of the utility function of the agent with respect to the quantity of ith good consumed,

$$mu_i = \frac{\partial u}{\partial g_i} = g_1^{\delta_1} \cdot g_2^{\delta_2} \cdots \delta_i g_i^{\delta_i - 1} \cdots g_\tau^{\delta_\tau} = \frac{\delta_i \cdot u\,(g_1, g_2, \ldots, g_\tau)}{g_i}. \tag{3}$$

If $\frac{mu_i}{mu_j} > \frac{p_i}{p_j}$ then the consumption pattern is such that marginal utility from the ith good is more than its price when compared to the jth good. Economic efficiency demands that the consumer find it suitable to consume more of the ith good relative to the jth good and consequently the marginal utility of ith good falls to the extent that $\frac{mu_i}{mu_j}$ becomes equal to $\frac{p_i}{p_j}$. Similarly if $\frac{mu_i}{mu_j} < \frac{p_i}{p_j}$, the consumer increases the consumption of the jth good and eventually the equality is restored. In equilibrium, we observe the equality when the consumer maximizes her utility. If we use this equality along with the expression of mu_i from Eq. (3), we obtain that $\frac{p_i\, g_i}{\delta_i} = \frac{p_j\, g_j}{\delta_j}$. Moreover, this equality holds valid for any arbitrary i and j. Moreover, the way savings has been introduced in the utility function, the savings of an agent, s, acts also like a commodity with unit price. Therefore, the following equation is satisfied in equilibrium:

$$\frac{p_1\, g_1}{\delta_1} = \frac{p_2\, g_2}{\delta_2} = \cdots = \frac{p_\tau\, g_\tau}{\delta_\tau} = \frac{s}{\delta_s}. \tag{4}$$

As marginal utility of each of the goods in positive amount is positive, the budget constraint (2) holds with equality in equilibrium. We additionally use (4) to obtain the consumption of each of the goods in equilibrium. The main result could be summarized as follows:

Proposition *The agent uses $\delta_1, \delta_2, \ldots,$ and δ_τ fractions of the total income m for purchasing goods 1, 2, …,and τ. Moreover, the agent saves remaining δ_s fraction of the total income.*

This proposition comes directly from the first order condition in (4). It demands an observation that the proportions in (4) are all equal to another fraction, which is formed by adding all the numerators of the proportions divided by all the denominators. The numerator, thus formed, is equal to the total income m by appealing to the budget constraint noted in (2). The denominator of the fraction is $\delta_s + \sum_{i=1}^{\tau} \delta_i$, which is equal to unity by dint of our model.

3 Comparison of Income and Expenditure Distribution

In this section, we compare the income and expenditure distribution, in particular the Pareto tails. We analyse how likely it is the case of Pareto tail from a theoretical viewpoint. If there is a bunch of possible alternative theories which predict the Pareto tail, the valid ones may be sorted out from the empirical analysis of the distributions.

3.1 The Case of Same Parametric Value for Agents

As a simple case, we consider the situation, in which the parameters of the economic environment is constant across different agents. For example, savings decision is a constant δ_s in the economic environment presented above. Additionally, we assume the following with strong empirical support.

Assumption *The distribution of income has a Pareto tail.*

In our economic environment, income is exogenous. This assumption, if added to that economic environment, awards a structure to the income distribution. It is quite consistent with our framework as the current framework models from a consumer's viewpoint irrespective of the income to be earned.

Proposition *The expenditure distribution is exactly similar to the income distribution under these assumptions.*

In other words, if $x\%$ of the population follow income distribution at the upper tail of the income distribution, it is the same $x\%$ who will be in the Pareto tail of the expenditure distribution. How consistent is this with the data? Study with expenditure distribution is rather limited. Analysis of India income data yields a Pareto tail of 5–10% [6] in contrast to the expenditure distribution with the Indian data depicting [20] a tail of 10–20%. Therefore there is a considerable diversion between the two distributions.[3]

[3] It should be mentioned that these two estimates of Pareto tails come from two different sources. With the exception of a study [15] there is no such empirical comparison. The relevant study finds both the distribution to be lognormal without any significant Pareto tail, in dealing with the data from USA and UK.

It is natural to ask the question that under what conditions it is possible to find such an empirical result. Obviously, the assumption of *same* parameter values for all agents has to be dropped and should be replaced with agents be distributed. We illustrate a case where expenditure distribution could have a Pareto tail without the income distribution having Pareto tail. A benchmark model is studied where the income is constant for all the individuals.

3.2 Asymptotic Case with No Savings and One Dominant Good

Without loss of generality, we can assume that good 1 represents the basic necessities of life. The importance of each good is denoted by the corresponding parameter in the utility function. If good 1 is the pre-dominant good, compared to all the other goods put together, the sum of values of parameters $\delta_2, \delta_3, \ldots, \delta_\tau$ is small compared to δ_1. Further, we *assume* that the savings propensity, δ_s, is equal to zero.

$$\begin{aligned} c &= p_1\, g_1 + p_2\, g_2 + \cdots p_\tau\, g_\tau \\ &= p_1\, g_1 \left(1 + \frac{\delta_2}{\delta_1} + \cdots + \frac{\delta_\tau}{\delta_1}\right). \end{aligned} \tag{5}$$

Since good 1 represents the basic necessities of life, the variance in consumption of this good is rather small across individuals and we can replace it with a constant, κ. We incorporate this in 5 to gather,

$$c = \kappa \left(1 + \frac{\delta_2}{\delta_1} + \cdots + \frac{\delta_\tau}{\delta_1}\right).$$

Taking logarithm of both the sides and using the rule of approximation that $\log(1 + \epsilon) \approxeq \epsilon$, where $|\epsilon|$ is sufficiently small, we arrive at,

$$\log c \approxeq \log \kappa + \frac{\delta_2 + \delta_3 + \cdots + \delta_\tau}{\delta_1}, \tag{6}$$

where $\delta_2 + \delta_3 + \cdots + \delta_\tau$ is sufficiently small compared to the value of δ_1.

According to the tastes and priorities of individuals, the values of the parameters $\delta_2, \delta_3, \ldots, \delta_\tau$ differ. In general, we can treat $\frac{\delta_2}{\delta_1}, \frac{\delta_3}{\delta_1}, \ldots, \frac{\delta_\tau}{\delta_1}$ as random variables and assume that they are identically and independently drawn from a distribution with a finite mean and finite variance. If τ is sufficiently large, we appeal to the Central Limit Theorem to conclude that $\left(\frac{\delta_2}{\delta_1} + \cdots + \frac{\delta_\tau}{\delta_1}\right)$ follows a normal distribution. From (6), it is noted that c follows a lognormal distribution.

In a more general scenario, the number of goods itself is a random variable. With the assumption that τ is geometrically distributed, c follows a double Pareto distribution as proved in [25]. The double Pareto distribution has both its upper and lower tails following a Pareto distribution with different parameters (say, α and β).

The standard form of a double Pareto density function is given by:

$$f_{dp}(x) = \begin{cases} \dfrac{\alpha\beta}{\alpha+\beta} x^{\beta-1} & \text{for} \quad 0 < x \leq 1 \\ \\ \dfrac{\alpha\beta}{\alpha+\beta} x^{-\alpha-1} & \text{for} \quad x > 1\,. \end{cases} \tag{7}$$

A related possibility occurs when the population is divided into two strata, comprising π and $1-\pi$ fractions. The second fraction is the poorer section consuming only the necessary items whereas the affluent class, the first section, consumes a relatively higher number of goods – both necessary and luxury items. It is quite reasonable to assume that the total number of necessary items consumed is fixed and as explained above, the expenditure distribution for the poorer section should follow a log-normal distribution. However, the number of luxury items consumed can be treated as a random variable, so that the expenditure distribution of the affluent class can be modeled as a double Pareto confined to the upper tail, which is nothing but a Pareto distribution. This is consistent with the fact that higher end of the expenditure distribution should follow a Pareto law, similar to the income distribution. The overall expenditure distribution is then given by a mixture of lognormal and Pareto distribution. The probability density function of such a distribution is expressed as,

$$f_m(x) = \pi \cdot f_p(x) + (1-\pi) \cdot f_{ln}(x), \tag{8}$$

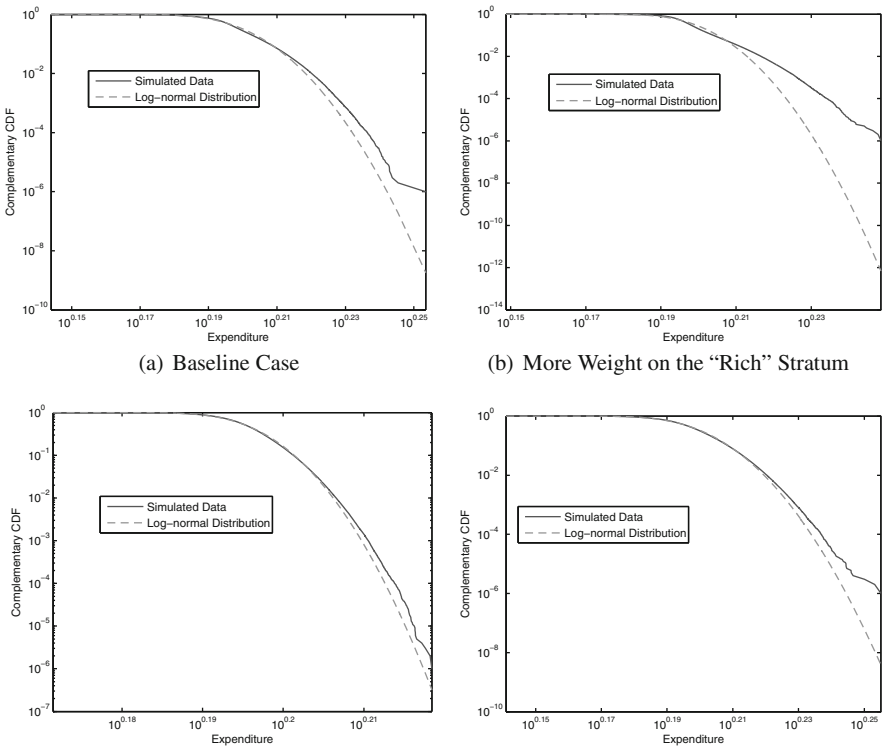
where $f_{ln}(\cdot)$ and $f_p(\cdot)$ are the probability density functions for the log-normal and Pareto distribution with π as the relative weight. More explicitly,

$$\begin{aligned} f_{ln}(x) &= \frac{1}{\sqrt{2\pi}\sigma\, x} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \\ f_p(x) &= \nu \frac{x_0^{\nu}}{x^{\nu+1}} \cdot 1_{x>x_0} \end{aligned} \tag{9}$$

where μ and σ^2 are the parameters associated with the log-normal distribution. It may be noted that expectation of the Pareto distribution exists if and only if $\nu > 1$. The value of this Pareto exponent, ν, is an important parameter along with x_0, the cut-off of the Pareto tail.

3.3 A Simulation Study

We elaborate our intuition with a simulation study. The simulation study assumes that there are τ goods. A suggested number for τ is 10000 as considered in the conducted study. The consumption of each good i is determined by the taste (δ_i) of an agent. There are two strata from which agents come from. The first stratum is the so-called "poor" stratum. The persons belonging to this stratum consume only



(a) Baseline Case (b) More Weight on the "Rich" Stratum

(c) More Number of Goods for "Poor" (d) Less Number of Expected Goods for "Rich"

Fig. 1 Simulation Study: This study uses one million agents. The agents' expenditure is plotted against the Complementary Cumulative Distribution Function (CCDF) in the log–log scale. The *solid line* indicates the simulated data; whereas the *dotted line* shows the CCDF of the log-normal distribution with the mean and variance same as that of the simulated data. The baseline case is when $\pi = 0.3$, $\tau_0 = 100$ and Expected value of $\tau = 1100$. The second graph depicts a case with $\pi = 0.70$. The third case is achieved when τ_0 equal to 500. The expected value of τ is 600 for the fourth case. All the agents possess the same income. However the savings rate or the expenditure is different as the taste of agents differ. All the panels reveal the existence of a Pareto tail in the expenditure distribution. The magnitude of the Pareto tail depends on the values of parameters

first τ_0 goods, which is equal to 100 for the study. The parameters of $\delta_1, \delta_2, \ldots, \delta_{\tau_0}$ are chosen from a uniform distribution between 0 and ζ. ζ denotes the maximum possible share of a good in the income. We pick ζ such that the total money spent for all τ_0 goods together do no exceed 90% of an agent's income. The other stratum is known as the "rich" stratum. In this particular stratum, agents can consume a random number of goods upto τ. The minimum number of τ_0 goods are consumed always. The additional number of goods consumed is drawn from a geometric distribution. The weights of the "poor" and the "rich" strata is given by $1 - \pi$ and π, respectively.

We would like to point out that even if we are not explicitly modeling the savings decision, it is quite imbibed in the model. The proportion of savings is chosen out of the expenditure decision. The value of δ_s actually varies between $1 - \zeta$ and 1. The expenditure is a random variable and savings is the part of income which is not going to the expenditure. Therefore, the savings parameter is also a random variable. The income is constant. And, therefore whatever Pareto tail we could find in the expenditure distribution, it is definitely coming out of the preferences of the consumers.

The illustrations show that we can obtain a power tail law in the distribution of expenditures. The length of the Pareto tails depends on the values of the various parameters. The baseline case assumes that the weight of the "rich" strata is 30% represented by the parameter π. The other parameter is average number of goods consumed for this strata. It is the expected value of τ for the rich stratum. In the baseline case, it is taken to be 1100 as opposed to the 100 goods consumed by the members of the "poor" strata.

The other cases are somewhat similar to the baseline case with the exception of the value of one parameter. The comparative statics is intuitive and could be figured out from the illustrations. The case of more consumers in the "rich" strata could be obtained with a high value of π, say, 70%. Obviously, this helps the Pareto tail to become fatter. On the other hand, if we increase the total number of goods consumed by the "poor" strata, the Pareto tail moves to be slimmer as illustrated here with τ_0 equaling 500. The last experiment is about expected number of goods consumed by the rich strata. It shows that with an expected number of goods of 600, the Pareto tail is still prominent with small reductions.

4 Discussion

We have experimented with a simple model about agents' tastes. In particular, an agent's felicity function is modeled as Cobb–Douglas. The question is what kind of restrictions do we need on the parameters to obtain a Pareto tail for the expenditure distribution? And secondly, if we do get a Pareto tail out of the expenditure distribution, when agents behave in accordance with this felicity function, what is the magnitude of that Pareto tail with respect to the magnitude of the Pareto tail in the income data? Could it be bigger than the income data?

For any model with agents having Cobb–Douglas felicity functions, we have found that the expenditure distribution possesses the Pareto tail as much as the income distribution. We look for more assumptions in the model to develop the case for magnitude wise a greater Pareto tail of the expenditure distribution compared to the income distribution. We assume that agents have heterogenous preferences for goods. In particular, if agents' preferences vary uniformly within a given range, it may lead to a log-normal distribution. We add to this model in two ways. First, there are two strata in the society – "poor" and "rich". The "poor" stratum consumes only a fixed number of goods. The members of the "rich" stratum consume more number

of goods compared to the "poor" stratum. Secondly, the number of goods consumed by the members of the "rich" stratum follows a geometric distribution.

We predict that we can obtain a lognormal expenditure distribution with a Pareto tail with these assumptions, even if we start with constant income across agents. The simulation results confirm our theoretical prediction. The role of different parameters are also analysed. Thereby we resolve the anomaly of the Indian data portaying a greater Pareto tail in expenditures compared to the income distribution. In this process, we implicitly assume that not all members of the society have access to all the resources. The two strata are identified in the model by available opportunities rather than income or tastes. There is a significant difference between the opportunities available to rural and urban segments of Indian population which justifies the assumption of discrepancy in opportunities for the individuals. This also explains the difference in the Pareto cut off in the expenditure distribution of the urban and rural population [20]. Indeed our attempt is towards a general theory of preferences which may be taken care of while studying the distribution analysis for the related study of income and expenditure.

References

1. A. Banerjee, V.M. Yakovenko and T.Di. Matteo, Physica A **370** (2006) 54–59
2. T.Di. Matteo, T. Aste, S.T. Hyde, Exchanges in compex networks: income and wealth distributions in The Physics of Complex Systems, eds. F. Mallamace and H. E. Stanley, IOS Press Amsterdam (2004), 43
3. F.A. Cowell, F.H.G. Ferreira, J.A. Litchfeld, J. Income Distr. **8** (1998) 63–76
4. N.J. Moura Jr and M.B. Rebeiro, Eur. Phys. J. B 67 (2009) 101–120
5. D. Chotikapanich, D.S. Prasada Rao, K.K. Tang, Rev. Incom. and Wealth **53** (2007) 127–147
6. S. Sinha, Physica A 359 (2006) 555–562
7. H. Aoyama et al., Fractals **8** (2000) 293–300
8. Y. Fujiwara et al. Physica A **321** (2003) 598–604
9. P. Lukasiewicz, A. Orlowski, Physica A **344** (2004) 146–151
10. C. Quintano and A. D'Agostino, Rev. of Income and Wealth **52** (2006) 525–546
11. A. Dragulescu and V.M. Yakovenko, Physica A 299 (2001) 213–221
12. A. Harrison, Rev. of Econ. Studies, **48** (1981) 621–631
13. A. Christian Silva, V.M. Yakovenko, Europhys. Lett. **69** (2005) 304–310
14. A. Dragulescu and V.M. Yakovenko, Eur. Phys. J. B. **20** (2001) 585–589
15. E. Battistin, R. Blundell, and A. Lewbel, Journal of Political Economy Vol. 117, No. 6: pp. 1140–1154, 2009
16. V. Pareto, Cours d'économie politique, (Lausanne, 1897)
17. V.M. Yakovenko and J.B. Rosser, Jr., Rev. Mod. Phys. 81, (2009) 1703
18. A. Chatterjee, B.K. Chakrabarti and S.S. Manna, Physica A 335 (2004) 155–163
19. T. Mizuno et al. Physica A **387** (2008) 3931–3935
20. A. Ghosh, K. Gangopadhyay, and B. Basu, Physica A (article in press), doi:1016/j.physa.2010.06.018
21. A. Jayadev Physica A 387 (2008) 270–276
22. Don Patinkin, Money, Interest and Prices, Harper & Row (1956)
23. Robert C. Feenstra, Journal of Monetary Economics 17 (1986) 271–291
24. A.S. Chakrabarti and B.K. Chakrabarti, Physica A **388**, (2009), 4151–4158
25. M. Mitzenmacher: Internet Math, 1, (2004) 305–333

Two Agent Allocation Problems and the First Best

Manipushpak Mitra

Abstract. We consider a general class of two agent allocation problems and identify the complete class of first best rules. By first best rules we mean allocation rules for which we can find efficient, strategyproof and budget balanced mechanisms. We show that the only first best rules are the fixed share allocation rules.

1 Introduction

We consider a general class of allocation problems in a two agent setting and in the presence of money. Allocation of indivisible goods in the presence of money has been analyzed extensively (see Fujinaka and Sakai [3], Miyagawa [6], Ohseto [7], [9], [10], Sakai [11], Svensson and Larsson [12], Tadenuma and Thomson [13]). Allocation of divisible goods in the presence of money has also been analyzed (see Back and Zender [1], Wilson [15], Yang and Hajek [16]).

In this paper we consider a situation where a good or a set of goods (divisible or indivisible) needs to be allocated among two agents and that agents' preferences are quasilinear in the allocated good and money. Assuming unknown types of the agent we ask the following question: can we find allocation rules for which we can find efficient, strategyproof and budget balanced mechanisms? Such allocation rules are called first best rules. We show that the only first best rules are the fixed share allocation rules.

2 The Framework

Let $N = \{1, 2\}$ be the set of agents to whom a good or a set of goods needs to be allocated. Let $U_i(d_i, t_i; \theta_i) = d_i\theta_i + t_i$ where d_i is the amount allocated to agent i,

Manipushpak Mitra
ERU – Indian Statistical Institute, Kolkata, India,
e-mail: mmitra@isical.ac.in

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

$\theta_i \in \Re_+$ is the type (or valuation) of agent i and t_i is the monetary transfer to agent i. In this paper we assume that the type of the agent is unknown. Since the total amount to be allocated is fixed, $d_1 + d_2 = m$. An allocation of the good can be done in many ways. We assume that **D** represents the set of possible allocations. Note that a generic element of the set **D** is $d \equiv (d_1, d_2)$ with $d_1 + d_2 = m$. An allocation rule is a mapping $d : \Re^2_+ \to \mathbf{D}$ that specifies for each profile $\theta = (\theta_1, \theta_2) \in \Re^2_+$ an allocation vector $d(\theta) = (d_1(\theta), d_2(\theta)) \in \mathbf{D}$.

A transfer rule is a mapping $t : \Re^2_+ \to \Re^2$ that specifies for each profile $\theta \in \Re^2_+$ a transfer vector $t(\theta) = (t_1(\theta), t_2(\theta)) \in \Re^2$. While an allocation rule is concerned with the distribution of good(s), the transfer rule is concerned with the distribution of money. We assume that the type of any agent is private information and introduce the notion of a direct revelation mechanism that deals with the allocation of good(s) as well as the allocation of money in order to elicit this private information.

A direct revelation mechanism (or simply a mechanism) (d, t) constitutes of an allocation rule d and a transfer rule t. Under a direct revelation mechanism agents announce their types simultaneously and given the announcement of the agents the mechanism specifies the allocations and transfers. Fix an allocation rule $d : \Re^2_+ \to \mathbf{D}$. Given this allocation rule consider a mechanism (d, t). The mechanism specifies that given an announcement $\theta' = (\theta'_1, \theta'_2) \in \Re^2_+$ the utility of agent i with true type θ_i is given by $U_i(d_i(\theta'), t_i(\theta'), \theta_i) = d_i(\theta')\theta_i + t_i(\theta')$.

Definition 1 A mechanism (d, t) is *efficient* if for each $\theta \in \Re^2_+$, $d(\theta) = (d_1(\theta), d_2(\theta)) \in \arg\max_{d \in \mathbf{D}}(d_1\theta_1 + d_2\theta_2)$.

Definition 2 A mechanism (d, t) is *strategyproof* if for all $\theta_1, \theta'_1 \in \Re_+$ and all $\theta_2 \in \Re_+$, $U_1(d_1(\theta_1, \theta_2), t_1(\theta_1, \theta_2), \theta_1) \geq U_1(d_1(\theta'_1, \theta_2), t_1(\theta'_1, \theta_2), \theta_1)$ and $U_2(d_2(\theta_1, \theta_2), t_2(\theta_1, \theta_2), \theta_2) \geq U_2(d_2(\theta_1, \theta'_2), t_2(\theta_1, \theta'_2), \theta_2)$ for all $\theta_2, \theta'_2 \in \Re_+$ and all $\theta_1 \in \Re_+$.

Proposition 1 *A mechanism (d, t) is strategyproof only if*

(ND1) $d_1(\theta_1, \theta_2)$ *is non-decreasing in* θ_1 *and*
(ND2) $d_2(\theta_1, \theta_2)$ *is non-decreasing in* θ_2.

Since Proposition 1 is a very standard result we omit its formal proof.

Definition 3 A mechanism (d, t) is *budget balanced* if for all $\theta \in \Re^2_+$, $t_1(\theta) + t_2(\theta) = 0$.

Definition 4 An allocation rule $d : \Re^2_+ \to \mathbf{D}$ is *implementable* if there exists a mechanism (d, t) that satisfies efficiency of decision and strategyproofness.

Definition 5 A mechanism (d, t) is a VCG mechanism, if for all $\theta = (\theta_1, \theta_2) \in \Re^2_+$,

$$t_i(\theta) = d_j(\theta)\theta_j - h_i(\theta_j), \ h_i : \Re_+ \to \Re, \ \forall \ i, j \in N, \ i \neq j. \tag{1}$$

Proposition 2 *Any allocation rule $d : \Re^2_+ \to \mathbf{D}$ is implementable uniquely by the class of VCG mechanisms.*

Proposition 2 states that all allocation rules $d : \Re^2_+ \to \mathbf{D}$ are implementable. Moreover, the associated mechanism must be a VCG mechanism (see Vickrey [14], Clarke [2] and Groves [4]). Proposition 2 follows directly from Holmström [5] since domain of preferences specified in this paper satisfy Holmström's definition of "convex" domains.

3 The Result

Our main result deals with first best implementable allocation rules.

Definition 6 An allocation rule $d : \Re^2_+ \to \mathbf{D}$ is *first best implementable* if there exists a mechanism (d, t) that satisfies efficiency of decision, strategyproofness and budget balancedness.

We show that the only first best rules are the fixed share allocation rules.

Definition 7 An allocation rule $d : \Re^2_+ \to \mathbf{D}$ is a *fixed share allocation rule* if the set $\mathbf{D}$ is a singleton, that is, if there exists a real number r such that $d(\theta) = (d_1(\theta) = r, d_2(\theta) = m - r)$ for all $\theta = (\theta_1, \theta_2) \in \Re^2_+$.

Theorem 1 *An allocation rule $d : \Re^2_+ \to \mathbf{D}$ is first best implementable only if it is a fixed share allocation rule.*

To prove Theorem 1 we will use the following lemma which shows that an allocation rule is first best implementable only if it satisfies an 'appropriate' symmetry condition.

Lemma 1 *An allocation rule $d : \Re^2_+ \to \mathbf{D}$ is first best implementable only if for all $\theta = (\theta_1, \theta_2) \in \Re^2_+$ with $\theta_1 \neq \theta_2$, $d(\theta_1, \theta_2) = d(\theta_2, \theta_1)$.*

Proof. Using condition (1) (given by the VCG mechanism) and using budget balancedness it follows that an allocation rule $d : \Re^2_+ \to \mathbf{D}$ is first best implementable only if we can find function $h_i : \Re_+ \to \Re$ for $i = 1, 2$ such that for all $\theta = (\theta_1, \theta_2) \in \Re^2_+$,

$$d_1(\theta_1, \theta_2)\theta_1 + d_2(\theta_1, \theta_2)\theta_2 = h_1(\theta_2) + h_2(\theta_1). \tag{2}$$

Since from (2) it follows that $d_1(1, 1) + d_2(1, 1) = h_1(1) + h_2(1) = m$, by setting $\theta_1 = \theta_2 = x \in \Re_+$ in (2) and then using $d_1(x, x) + d_2(x, x) = m = h_1(1) + h_2(1)$ we get

$$h_1(x) + h_2(x) = h_1(1)x + h_2(1)x, \ \forall \ x \in \Re_+. \tag{3}$$

An immediate consequence of (3) is that

$$h_1(x) = h_1(1)x + K(x) \text{ and } h_2(x) = h_2(1)x - K(x), \ \forall \ x \in \Re_+ \tag{4}$$

where $K(x)$ is any real valued function. Using (2) and (4) for states $\theta = (\theta_1, \theta_2)$ and $\theta' = (\theta'_1 = \theta_2, \theta'_2 = \theta_1)$ we obtain the following.

(A) $d_1(\theta_1,\theta_2)\theta_1 + d_2(\theta_1,\theta_2)\theta_2 - h_1(1)\theta_2 - h_2(1)\theta_1 = K(\theta_2) - K(\theta_1)$ and
(B) $d_1(\theta_2,\theta_1)\theta_2 + d_2(\theta_2,\theta_1)\theta_1 - h_1(1)\theta_1 - h_2(1)\theta_2 = K(\theta_1) - K(\theta_2)$.

Summing **(A)** and **(B)** and then using $h_1(1) + h_2(1) = m$ we get

$$[d_1(\theta_1,\theta_2) + d_2(\theta_2,\theta_1) - m]\theta_1 + [d_2(\theta_1,\theta_2) + d_1(\theta_2,\theta_1) - m]\theta_2 = 0. \quad (5)$$

Substituting $m = d_1(\theta_1,\theta_2) + d_2(\theta_1,\theta_2)$ in the first term of (5) and substituting $m = d_1(\theta_2,\theta_1) + d_2(\theta_2,\theta_1)$ in the second term of (5) we get

$$[d_2(\theta_2,\theta_1) - d_2(\theta_1,\theta_2)](\theta_1 - \theta_2) = 0. \quad (6)$$

From (6) it follows that $d_2(\theta_1,\theta_2) = d_2(\theta_2,\theta_1)$ for all $\theta_1 \neq \theta_2$. Moreover, substituting $d_2(\theta_2,\theta_1) = m - d_1(\theta_2,\theta_1)$ and $d_2(\theta_1,\theta_2) = m - d_1(\theta_1,\theta_2)$ in (6) we also get

$$[d_1(\theta_1,\theta_2) - d_1(\theta_2,\theta_1)](\theta_1 - \theta_2) = 0. \quad (7)$$

From (7) it also follows that $d_1(\theta_1,\theta_2) = d_1(\theta_2,\theta_1)$ for all $\theta_1 \neq \theta_2$. □

Proof of Theorem 1. Consider any state $\theta = (\theta_1 = x, \theta_2 = x)$ with $x \in \Re_+$. We first show that **(i)** $d_i(x,x) = d_i(y,x) = d_i(x,y)$ for all $y \in \Re_+$ and for $i = 1,2$. Consider $y > x$. From (ND1) it follows that $d_1(y,x) \geq d_1(x,x) \Rightarrow m - d_2(y,x) \geq m - d_2(x,x) \Rightarrow d_2(x,x) \geq d_2(y,x) = d_2(x,y)$ where the last equality follows from Lemma 1. Since (ND2) gives $d_2(x,y) \geq d_2(x,x)$, it follows that $d_2(x,x) \geq d_2(y,x) = d_2(x,y) \geq d_2(x,x)$ implying $d_2(x,x) = d_2(x,y) = d_2(y,x)$. Since $d_1(\theta) + d_2(\theta) = m$ for all $\theta \in \Re_+^2$, we also get $d_1(x,x) = d_1(x,y) = d_1(y,x)$. For $y < x$ a very similar kind of argument with inequalities in the opposite direction gives $d_i(x,x) = d_i(y,x) = d_i(x,y)$ for $i = 1,2$.

Consider any state $\theta' = (\theta_1' = z, \theta_2' = z)$ with $z \in \Re_+$ and $z \neq x$. Applying steps similar to those used for proving **(i)** above, it follows that **(ii)** $d_i(z,z) = d_i(y',z) = d_i(z,y')$ for all $y' \in \Re_+$ and for $i = 1,2$. Finally by substituting $y = z$ in **(i)** and by substituting $y' = x$ in **(ii)** we get **(iii)** $d_i(x,x) = d_i(x,z) = d_i(z,x) = d_i(z,z)$ for $i = 1,2$. Applying **(i)**, **(ii)** and **(iii)** it follows that if in any state $\hat{\theta} = (\hat{\theta}_1 = x, \hat{\theta}_2 = x)$, $d(\hat{\theta}) = (d_1(\hat{\theta}) = r, d_2(\hat{\theta}) = m - r)$ where $x \in \Re_+$ and r is any arbitrary real number then for all $\theta \in \Re_+^2$, $d(\theta) = (d_1(\theta) = r, d_2(\theta) = m - r)$ and the result follows. □

4 Conclusions

We have restricted our search of mechanisms to direct revelation mechanisms only. This is without loss of generality due to The Revelation Principal (see Myerson [8]). The class of allocation problems for which Theorem 1 is valid is quite broad since the set of possible allocations **D** is very general as it allows for the allocation of one or more than one divisible or indivisible good(s). Our result is a negative one since the only first best rules are the fixed share allocation rules for which the set **D**

is a singleton. Hence for such allocation rules efficiency and strategyproofness are trivially satisfied and there is no need for agent specific transfers. The type space is common to both agents is a crucial assumption. In the absence of this assumption Lemma 1 is not valid.

Acknowledgements The author is thankful to Satya Ranjan Chakravarty, Conan Mukherjee and Arunava Sen for their comments and suggestions.

References

1. Back, K. and Zender, J. F., 1993. Auctions of Divisible Goods: On the Rationale for the Treasury Experiment. The Review of Financial Studies, 6, 733–764
2. Clarke, E. H., 1971. Multi-part pricing of public goods. Public Choice 11, 17–33
3. Fujinaka, Y., Sakai, T., in press. The manipulability of fair solutions in allocation of an indivisible object with money. Journal of Public Economic Theory
4. Groves, T., 1973. Incentives in teams. Econometrica 41, 617–631
5. Holmström, B., 1979. Groves' schemes on restricted domains. Econometrica 47, 1137–1144
6. Miyagawa, E., 2001. House allocation with transfers. Journal of Economic Theory 100, 329355
7. Ohseto, S., 1999. Strategy-proof allocation mechanisms for economies with an indivisible good. Social Choice and Welfare 16, 121136
8. Myerson, R., 1979. Incentive-compatibility and the bargaining problem. Econometrica 47, 6173
9. Ohseto, S., 2000. Strategy-proof and efficient allocation of an indivisible good on finitely restricted preference domains. International Journal of Game Theory 29, 365374
10. Ohseto, S., 2006. Characterizations of strategy-proof and fair mechanisms for allocating indivisible goods. Economic Theory 29, 111121
11. Sakai, T., 2007. Fairness and implementability in allocation of indivisible objects with monetary compensations. Journal of Mathematical Economics 43, 549563
12. Svensson, L.-G., Larsson, B., 2002. Strategy-proof and nonbossy allocation of indivisible goods and money. Economic Theory 20, 483502
13. Tadenuma, K., Thomson, W., 1993. The fair allocation of an indivisible good when monetary compensations are possible. Mathematical Social Sciences 25, 117132
14. Vickrey, W., 1961. Counterspeculation, auctions and competitive sealed tenders. Journal of Finance 16, 8–37
15. Wilson, R. 1979. Auctions of Shares. Quarterly Journal of Economics, 93, 675–698
16. Yang, S. and Hajek, B., 2004. An efficient mechansim for allocation of a divisible good with its application to network resource allocation, IMA Workshop 6: Control and Pricing in Communication and Power Networks (University of Minnesota)

Opinion Formation in a Heterogenous Society

Marie-Therese Wolfram

1 Introduction

Opinion formation and opinion leadership has attracted a lot of research among sociologists and physicists in the last decades. The first concept of opinion leadership goes back to Lazarsfeld *et al*. [8] in 1944. Larzarsfeld *et al*. found out that during the presidential elections in 1940 interpersonal communication showed greater influence than direct media effects. In their theory of two-step flow communication opinion leaders, who are active media users, select, modify and transmit information from the media to the less active part of the community. In later models sociologists gained a different view of opinion leadership by introducing the notion of public individuation. Public individuation describes how people want to differentiate and act differently from other people, see [9]. This attitude is a necessary prerequisite for an opinion leader, since she or he has to stand out against the masses. Characteristic features of opinion leaders are their high self esteem and confidence as well as their ability to withstand criticism. Although new technologies like the internet, blogs or instant messaging changed the way of communication and information dissemination globally, opinion leadership still plays a critical role in opinion formation processes.

Social and political networks are another important factor. Traditional social networks include sport clubs or religious groups, but nowadays online networks like Facebook or Twitter connect more people than ever before. The importance of these social and political networks in the process of opinion formation can not be underestimated. Within these networks information is disseminated much faster than outside the network.

In the last years physicists got more interested in the mathematical modeling of opinion formation. They established a new research field, known as sociophysics. This notion was motivated by the pioneering work of Galam *et al*. [7]. For further information we refer to the work by Cominiciolo *et al*. [3] and Galam [10].

Marie-Therese Wolfram
Faculty of Mathematics, University of Vienna, Vienna, Austria,
e-mail: marie-therese.wolfram@univie.ac.at

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

Various approaches can be found in the literature like cellular automata, mean-field type models (leading to partial differential equations) or kinetic models. Kinetic models were introduced by Toscani [13] in 2006. The general idea is that the behavior of a sufficiently large number of interacting individuals in a society can be described by methods of statistical physics just as the collision of gas molecules in a container. In kinetic models the exchange of opinion between individuals is defined by pairwise, microscopic interactions. Then the whole society evolves by a certain macroscopic opinion distribution, which depends on the specific microscopic interactions.
In this work we present a kinetic model for opinion formation. It includes several aspects of opinion formation: self-thinking and compromise processes [4, 14]. Self thinking describes the diffusive way, in which individuals change their opinion depending on information sources like the media. Furthermore people tend to reach a compromise after exchanging opinions. Toscani introduced a opinion formation model in [13] which includes both processes. Here opinion is exchanged between individuals through pairwise interactions. These interaction laws lead to a partial differential equation of Fokker–Planck type in a suitable scaling limit. Similar diffusion equations were obtained in [11] as a mean field limit the Sznajd model [12]. A similar kinetic approach has been proposed by Boudin *et al*. [1, 2] using different microscopic interaction laws.

Our work is based on Toscani's model [13] and the later generalizations of opinion formation processes in the presence of strong leaders by Cominciolo *et al*. [3] and Düring *et al*. [5]. We present a generalization of this model for a society, in which individuals are connected via social or political networks. Our approach is a simplification of the complex reality, therefore it has its limitations. The statistical description is only valid for a large number of individuals. Furthermore we assume that an individual can be a member of one network only. Nevertheless the proposed model is a first step to a better understanding of opinion leadership and social networks.

This paper is organized as follows. In Sect. 2 we review Toscani's model for opinion formation and present our generalized approach. Sect. 3 is devoted to the limiting Fokker Planck system. Finally we present numerical examples in Sect. 4.

2 Kinetic Approaches for Opinion Formation

We start with a review of Toscani's model [13], discuss the motivations and underlying assumptions and continue with our extend approach.

Toscani's Model

Toscani's model is based on binary interactions between two individuals. Here the opinion is represented by the continuous variable $w \in \mathscr{I}$ with $\mathscr{I} = [-1, 1]$. The endpoints of the interval $\mathscr{I}$, i.e. $w = \pm 1$, correspond to extreme opinions. Let v

and w denote the pre-interaction opinion for two individuals and v^* and w^* their post-interaction opinion. Then the post-interaction opinions are given by

$$v^* = v - \gamma P(|v - w|)(v - w) + \eta_1 D(v), \tag{1a}$$
$$w^* = w - \gamma P(|w - v|)(w - v) + \eta_2 D(w). \tag{1b}$$

Here $\gamma \in (0, \frac{1}{2})$ is the constant compromise parameter, the quantities η_1 and η_2 denote random variables with zero mean and variance σ^2. The self thinking process of each individual due to the global access to information, e.g. through the press, television or internet, is modeled via random diffusion by the parameters η_i, $i = 1, 2$. The functions $P(\cdot)$ and $D(\cdot)$ model the local relevance of compromise and self-thinking for a given opinion. Both functions have to satisfy additional constraints to guarantee that the post-interaction opinions stay in the interval $\mathscr{I}$.

2.1 A Kinetic Model with Several Groups

Based on Toscani's approach we would like to present a model that describes opinion formation among several groups or networks. This approach is a further extension of the work of Düring *et al.* [5], which studied opinion formation between two groups of individuals, namely 'strong opinion leaders' and 'ordinary people'. We extend this approach for several groups, i.e. 'strong opinion leaders' and groups or networks of 'ordinary people'. Here we make the hypothesis that several groups in the human society are strongly connected via networks, societies or simply their age. Furthermore we assume that an individual can be a member of a single network only and that individuals among one group share a common compromise parameter.

Let N denote the number of groups or networks in a society, the N-th group always denotes the strong opinion leaders. The interaction rules between different groups are given by:

- If two individuals from groups i and j with either $i, j < N$ or $i, j = N$ meet, then their interaction rule is given by

$$v^* = v - \gamma_{ij} P_{ij}(|v - w|)(v - w) + \eta_{ij} D(v), \tag{2a}$$
$$w^* = w - \gamma_{ij} P_{ij}(|v - w|)(v - w) + \eta_{ij} D(v). \tag{2b}$$

  This means that if two individuals from the groups of ordinary people or two strong leaders meet, their post interaction opinion is determined by Toscani's model (1).
- If an individual from the i-th group of ordinary people with opinion v (with $i < N$) meets a strong opinion leader with opinion w their post-interaction opinions are given by

$$v^* = v - \gamma_{iN} P_{iN}(|v - w|)(v - w) + \eta_{iN} D(v), \tag{2c}$$
$$w^* = w. \tag{2d}$$

Here the post-interaction opinion of the strong leader does not change, while the post-interaction opinion of the ordinary person is influenced by the opinion leader. This reflects the assumption that opinion leaders promote opinions, show high self confidence and withstand criticism.

We reiterate that the parameter γ_{ij} denotes the constant compromise parameter, which describes the willingness to make a compromise between two opinions. The quantity η_{ij} are random variables with zero mean and variance σ^2 modeling self-thinking via diffusion.

Let $f_i = f_i(w,t)$, $i = 1,\dots,N$ denote the distribution function for each group, which depends on the opinion $w \in \mathscr{I}$ and time $t \in \mathbb{R}^+$. Then the time-evolution of opinion distribution satisfies a system of Boltzmann-like equations, given by

$$\frac{\partial}{\partial t} f_i(w,t) = \sum_{j=1}^{N} \frac{1}{\tau_{ij}} \mathscr{Q}_{ij}(f_i, f_j)(w), \qquad \text{for } i = 1,\dots,N-1 \tag{3a}$$

$$\frac{\partial}{\partial t} f_N(w,t) = \frac{1}{\tau_{NN}} \mathscr{Q}_{NN}(f_N, f_N)(w). \tag{3b}$$

The parameters τ_{ij} are suitable relaxation times which allow to control the interaction frequencies between the different groups. The Boltzmann-like collision operators can be derived by standard methods of kinetic theory, considering that the change in time of $f_i(w,t)$ due to binary interaction.

Let $\langle\cdot\rangle$ denote the mean operator with respect to the random quantities η_{ij}. Then the collision operator in weak form is given by

$$\begin{aligned} &\int_{\mathscr{I}} \mathscr{Q}_{ij}(f_i, f_j)(w)\phi(w)\,dw \\ &\quad = \frac{1}{2}\left\langle \int_{\mathscr{I}^2} \big(\phi(w^*) + \phi(v^*) - \phi(w) - \phi(v)\big) f_i(v) f_j(w)\,dv\,dw \right\rangle, \end{aligned} \tag{4}$$

for all smooth test function $\phi(w)$.

3 The Limiting Fokker–Planck System

To understand the behavior of solutions for large times, i.e. close to the steady state, we study the quasi-invariant opinion limit ($\gamma_{ij}, \sigma_{ij} \to 0$ and $\sigma_{ij}^2/\gamma_{ij} = \lambda_{ij}$) by formal asymptotics. This approach is well known in kinetic theory and has been introduced by Toscani for opinion formation processes, see [13]. We introduce for $\gamma_{ij} \ll 1$ the transformation

$$\tau = \gamma_{ij} t, \quad g_i(w,\tau) = f_i(w,t), \quad i = 1,\dots,N,$$

which implies $f_i(w,0) = g_i(w,0)$. The mass of each group is given by $M_i = \int g_i \, dv$ for $i = 1, \ldots, N$. Then the limiting Fokker–Planck system is given by

$$\frac{\partial}{\partial \tau} g_i(w,\tau) = \sum_{j=1}^{N} \frac{\partial}{\partial w} \left(\left(\frac{1}{\tau_{ij}} \mathscr{K}_{ij}(w,\tau) \right) g_i(w,\tau) \right) \tag{5a}$$

$$+ \sum_{j=1}^{N} \left(\frac{\lambda_{ij} M_j}{c_{ij} \tau_{ij}} \right) \frac{\partial^2}{\partial w^2} \left(D_{ij}^2(w) g_i(w,\tau) \right) \quad \text{for } i = 1, \ldots, N-1,$$

$$\begin{aligned} \frac{\partial}{\partial \tau} g_N(w,\tau) = & \frac{\partial}{\partial w} \left(\frac{1}{\tau_{NN}} \mathscr{K}_{NN}(w,\tau) g_N(w,\tau) \right) \\ & + \frac{\lambda_{NN} M_N}{c_{NN} \tau_{NN}} \frac{\partial^2}{\partial w^2} \left(D_2^2(w) g_N(w,\tau) \right), \end{aligned} \tag{5b}$$

where $c_{ii} = 2$ for all $i, j = 1, \ldots N$ and $c_{ij} = 4$ if $i \neq j$. The convolution operators $\mathscr{K}_{ij}$ are given by

$$\mathscr{K}_{ij}(w,\tau) = \int_{\mathscr{I}} P_{ij}(|w-v|)(w-v) g_i(v,\tau) \, dv, \quad \text{for } i = 1, \ldots, N. \tag{6}$$

System (5) is supplemented with the following no flux boundary conditions at $w \pm 1$ (which result from the integration by parts)

$$\sum_{j=1}^{N} \left[\frac{1}{\tau_{ij}} \mathscr{K}_{ij}(w,\tau) g_i(w,\tau) + \frac{\lambda_{1j} M_j}{c_{ij} \tau_{1j}} \frac{\partial}{\partial w} \left(D_{ij}^2(w) g_i(w,\tau) \right) \right] = 0, \tag{7a}$$

$$\frac{1}{\tau_{NN}} \mathscr{K}_{NN}(w,\tau) g_N(w,\tau) + \frac{\lambda_{NN} M_N}{c_{NN} \tau_{NN}} \frac{\partial}{\partial w} \left(D_N^2(w) g_N(w,\tau) \right) = 0, \tag{7b}$$

and

$$D_i^2(w) g_i(w) = 0 \qquad \text{on } w = \pm 1 \text{ for all } i = 1, \ldots N-1. \tag{7c}$$

To ensure that the post interaction opinions v^* and w^* in (2) stay in the interval $\mathscr{I}$ we assume that $D(\pm 1) = 0$. Therefore condition (7c) is automatically satisfied, if the solution g_i, $i = 1, \ldots, N$ are sufficiently regular.

4 Numerical Simulations

We consider a heterogenous society with different age groups. Here we assume that people within an age group are better connected than outside their group. For example, senior citizens are more likely to discuss current issues with other seniors and have a different opinion formation behavior than young people. We would like to understand the behavior of our mathematical model for different parameter sets.

As a very simple example we divide the 'ordinary' people into three different age groups – young people (15–25 years), workers (25–60 years) and senior citizens (> 60 years).

We model the diffusion of opinion by the function

$$D_i(w) = D(w) := (1 - w^2)^\alpha. \tag{8}$$

for $i = 1, \ldots, N$ with $\alpha > 1/2$. The compromise propensity $P_i(\cdot)$ for $i = 1, \ldots N$ by

$$P_i(|v - w|) = \mathbf{1}_{|v-w| \le r_i}, \tag{9}$$

where $\mathbf{1}_A$ denotes the indicator function on the set A and r_i denotes the interaction radius. The following parameters are fixed throughout this section, if not mentioned otherwise:

- relaxation times: $\tau_{ij} = 1$; for $i, j = 1, \ldots, N$;
- ratio of normal people to opinion leaders: $\sum_{i=1}^{N-1} M_i = 0.95$ and $M_N = 0.05$;
- exponent of the diffusion function in (8): $\alpha = 2$.

The initial datum for every group of normal people is given by a Gaussian

$$g_i(w, 0) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(w-\mu_1)^2}{2\sigma_i^2}} \qquad \text{for } i = 1, \ldots, N-1, \tag{10}$$

with $\sigma_i = 0.1$ and weights $\sum_{k=1}^{m} p_k = 1$. The opinion leaders are initially distributed as

$$g_N(w, 0) = \sum_{k=1}^{n} \frac{q_k}{\sqrt{2\pi}\sigma_N} e^{-\frac{(w-\mu_k)^2}{2\sigma_N^2}} \tag{11}$$

with weights $\sum\limits_{k=1}^{n} q_k = 1$.

4.1 Monte Carlo Simulations

We performed a series of kinetic Monte Carlo simulations for the presented Boltzmann model (2) to understand the influence of the model parameters and compare the results to the behavior of the limiting Fokker–Planck system (5). This kind of simulations are known as direct simulation Monte Carlo (DSMC). Pairs of individuals are chosen randomly and non-exclusively for binary collisions, where they exchange opinion according to the rule under consideration. We denote by N_i, $i = 1, \ldots, N$ the number of individuals in every single group. Furthermore we average over $M = 20$ time steps to obtain an approximate steady state opinion distribution. Each simulation is carried out for 10^6 time steps, where one time step

corresponds to $\sum_{i=1}^{N} N_i$ interactions. We choose $N_1 = 1000$ young individuals, $N_2 = 2000$, working people, $N_3 = 1750$ senior citizens and $N_4 = 250$ strong leaders. The parameter γ is set to 0.02. The random variables η_{ij} are chosen such that they assume only values $\pm\nu = \pm 0.01$ with equal probability. The initial distributions are given by the discrete analogues of (10) and (11).

4.2 Numerical Solution of the Fokker–Planck System

To illustrate the long-time behavior of the proposed model we discretize the nonlinear Fokker–Planck system (5) using a hybrid discontinuous Galerkin (DG) method introduced by Egger and Schöberl in [6]. This hybrid DG method was initially developed for convection diffusion equations and yields stable discretizations for convection dominated problems. We consider the following semi-implicit (in time) discretization of the Fokker–Planck system (5), which fits into the framework of [6]. Note that the method is conservative, which is consistent with the assumption that the initial mass of the Fokker–Planck system is preserved in time.

We choose a partition of the time interval $[0, T]$, with $0 = t_0 < t_1 < \ldots < t_k < \ldots < t_m = T$, and define $\Delta t_k = t_{k+1} - t_k$. Then the semi-discrete scheme is given by

$$\begin{aligned}\frac{g_i^{k+1} - g_i^k}{\Delta t_k} &= \sum_{j=1}^{N} \frac{\partial}{\partial w}\left(\frac{1}{\tau_{ij}} \mathscr{K}_1(g_j^k; w, t) g_1^{k+1}(w,t)\right) \\ &\quad + \sum_{j=1}^{N} \frac{\lambda_{ij} M_j}{c_{ij}\tau_{ij}} \frac{\partial^2}{\partial w^2}\left(D^2(w) g_i^{k+1}(w,t)\right),\end{aligned} \tag{12a}$$

$$\begin{aligned}\frac{g_N^{k+1} - g_N^k}{\Delta t_k} &= \frac{\partial}{\partial w}\left(\frac{1}{\tau_{NN}} \mathscr{K}_{NN}(g_N^k; w, t) g_N^{k+1}(w,t)\right) + \\ &\quad + \frac{\lambda_{NN} M_N}{c_{NN}\tau_{NN}} \frac{\partial^2}{\partial w^2}\left(D^2(w) g_N^{k+1}(w,t)\right).\end{aligned} \tag{12b}$$

Here $g_i^k, i = 1, \ldots, N$ denotes the solution at time $t = t_k$. We choose an equidistant mesh of mesh size $h = \frac{1}{400}$ to discretize the interval $\mathscr{I} = [-1, 1]$. The time steps Δt_k are set to 0.01.

4.3 Numerical Examples

In our first example we would like to simulate a simple opinion formation process. Two groups of opinion leaders are located at $w \pm 0.5$. Every age group is given by a Gaussian (10) with $\mu = 0$. In this example we assume that the interaction radius between senior citizens and other age groups are much smaller than the radii

of workers. Young people have larger interaction radii. The interaction radii in our first example are given in Table 1. The evolution and the stationary solutions of the Boltzmann and Fokker–Planck system are depicted in Fig. 1.

Table 1 Interaction radii – the value left of the backslash corresponds to the interaction radius of the individual in the left colum, the value right to the individual in the top row

	young	workers	senior	leaders
young	0.75	0.5	0.4	0.5\0
workers	0.5	0.5	0.5	0.5\0
senior	0.4	0.4	0.4	0.4\0
leaders	0\0.5	0\0.5	0\0.4	0.4

In our second example we decrease the interaction radii of the senior citizens even further(see Table 2) and increase the groups of opinion leaders (located at $w = \pm 0.25$ and $w = \pm 0.75$ with weight $q_k = 0.25$ for $k = 1, \ldots, 4$). All age groups are initially distributed around $\mu_i = \pm 0.5$ with weight $p_k = 0.5$ for $k = 1, 2$ in (10). Furthermore we set the parameter α in (8) to 0.75.

Table 2 Interaction radii

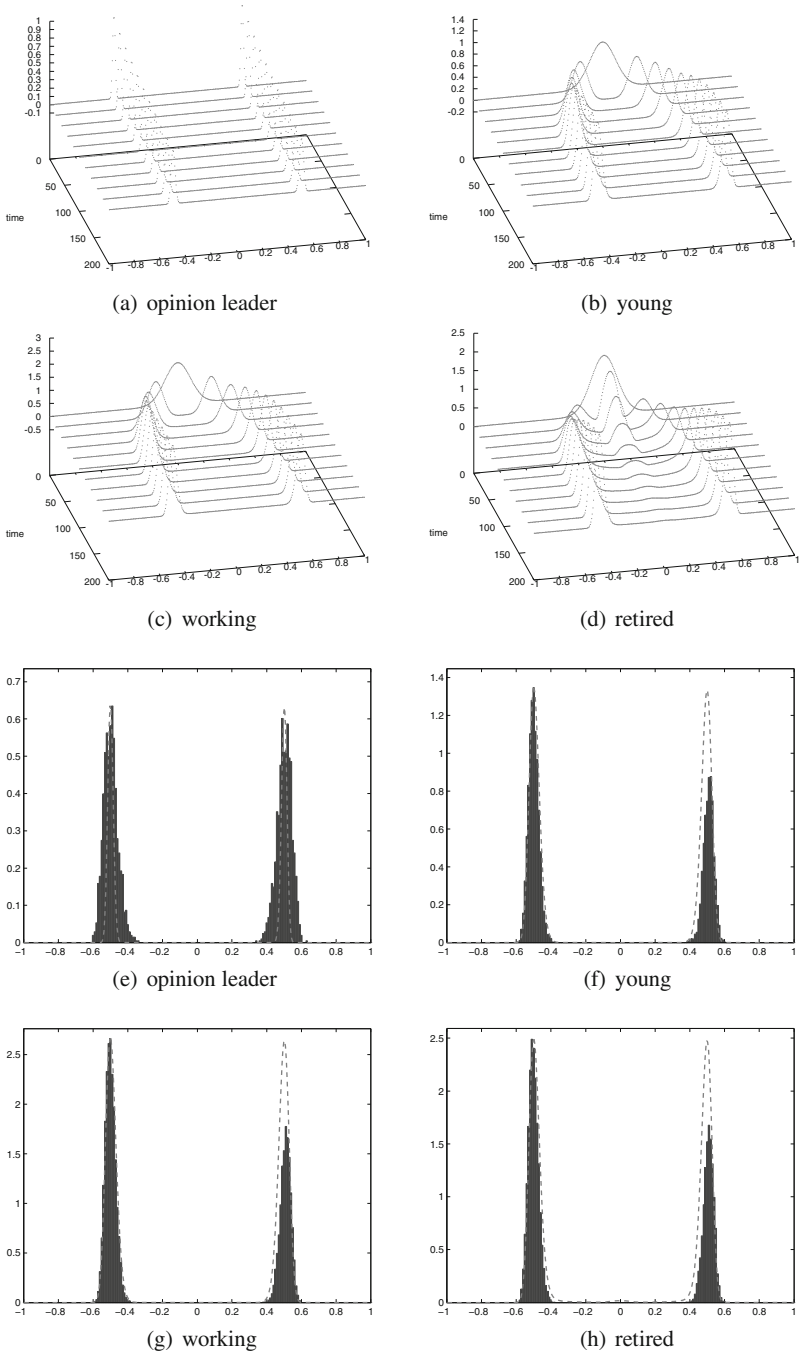
	young	workers	senior	leaders
young	0.75	0.5	0.2	0.5\0
workers	0.5	0.75	0.5	0.5\0
senior	0.2	0.2	0.2	0.2\0
leaders	0\0.5	0\0.5	0\0.2	0.2

The evolution and the stationary solutions of the Boltzmann and Fokker–Planck system are depicted in Fig. 2. We observe that the smaller diffusivity coefficient α causes a greater spread in the different age groups as well as the opinion leaders. The groups of opinion leaders at $w \pm 0.25$ diffuse and merge during the time evolution. Therefore we can only observe two peaks in the stationary profiles of all age groups.

If we increase the interaction frequencies within the networks $\tau_{ii} = 0.1$ for $i = 1, \ldots, 4$, the interaction with the strong opinion leaders loose their influence in the opinion formation process. The evolution of the Fokker Planck system is shown in Fig. 3. Young people and workers merge at the center $w = 0$, while retired individuals remain centered around the initial datum (due to the small interaction radii of senior citizens).

5 Conclusions

We introduced and discussed a nonlinear kinetic model for a heterogenous society. We assume that this society includes several different groups or social networks as



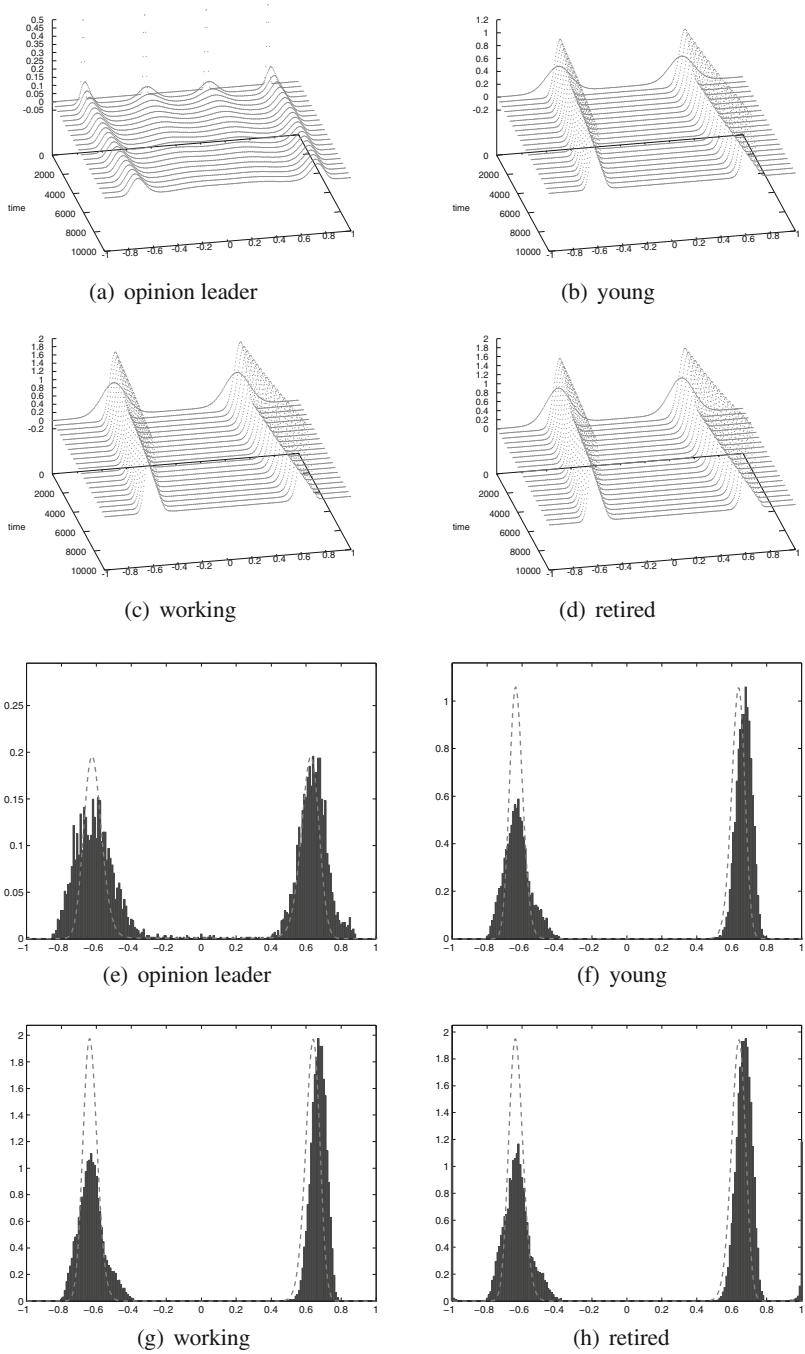
(a) opinion leader (b) young

(c) working (d) retired

(e) opinion leader (f) young

(g) working (h) retired

Fig. 1 (a)–(d) Time evolution of the Fokker–Planck solution, (e)–(h) Stationary solution of the MC simulation and the Fokker–Planck system



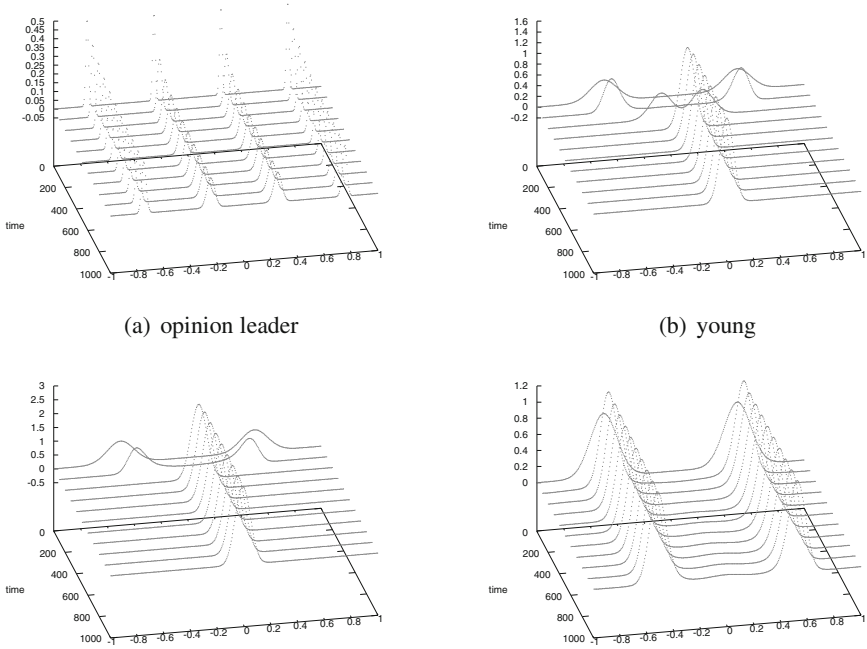
(a) opinion leader (b) young

(c) working (d) retired

(e) opinion leader (f) young

(g) working (h) retired

Fig. 2 (a)–(d) Time evolution of the Fokker–Planck solution, (e)–(h) Stationary solution of the MC simulation and the Fokker–Planck system



(a) opinion leader (b) young

(c) working (d) retired

Fig. 3 (a)–(d) Time evolution of the Fokker–Planck solution

well as a group of strong opinion leaders. The evolution of opinion is described by a system of Boltzmann-like equations, where collisions describe the binary exchange of opinion and self-thinking. For suitably large times the system of Boltzmann equations is well approximated by the nonlinear Fokker–Planck system. We discuss the behavior of the Boltzmann type system as well as the Fokker–Planck system for various model parameters in different examples.

Acknowledgements M.T. Wolfram has been supported by the FWF research grant T 456-N23.

References

1. Boudin, L., Monaco, R., Salvarani, F.: A kinetic model for multidimensional opinion formation. Phys. Rev. E **81**, 036, 109 (2010)
2. Boudin, L., Salvarani, F.: A kinetic approach to the study of opinion formation. M2AN Math. Model. Numer. Anal. **43**(3), 507–522 (2009). DOI 10.1051/m2an/2009004. URL http://dx.doi.org/10.1051/m2an/2009004
3. Comincioli, V., Della Croce, L., Toscani, G.: A Boltzmann-like equation for choice formation. Kinet. Relat. Models **2**(1), 135–149 (2009). DOI 10.3934/krm.2009.2.135. URL http://dx.doi.org/10.3934/krm.2009.2.135

4. Deffuant, G., Amblard, F., Weisbuch, G., Faure, T.: How can extremism prevail? a study based on the relative agreement interaction model. JASSS J. Art. Soc. Soc. Sim **5**(4) (2002)
5. Düring, B., Markowich, P., Pietschmann, J.F., Wolfram, M.T.: Boltzmann and Fokker-Planck equations modelling opinion formation in the presence of strong leaders. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **465**(2112), 3687–3708 (2009). DOI 10.1098/rspa.2009.0239. URL http://dx.doi.org/10.1098/rspa.2009.0239
6. Egger, H., Schoberl, J.: A hybrid mixed discontinuous Galerkin finite-element method for convection-diffusion problems. IMA J Numer Anal p. drn083 (2009). DOI 10.1093/imanum/drn083. URL http://imajna.oxfordjournals.org/cgi/content/abstract/drn083v1
7. Galam, S., Gefen, Y., Shapir, Y.: Sociophysics: a new approach of sociological collective behavior. J. Math. Sociology **9**, 1–13 (1982)
8. Lazarsfeld, P., Berelson, B., Gaudet, H.: The people's choice: How the voter makes up his mind in a presidential campaign. Duell, Sloan, Pearce (1944)
9. Maslach, C., Stapp, J., Santee, R.: Individuation: Conceptual analysis and assessment. J. Personality Social Psychology **49**(3), 729–738 (1985)
10. S., G.: Heterogeneous beliefs, segregation, and extremism in the making of public opinions. Phys. Rev. E **71**(4), 046, 123 (2005)
11. Slanina, F., Lavicka, H.: Analytical results for the sznajd model of opinion formation. Eur. Phys. J. B **35**(2), 279–288 (2003)
12. Sznajd-Weron, K., Sznajd, J.: Opinion formation in closed community. Int. J. Mod. Phys. C **11**(6), 1157–1165 (2000)
13. Toscani, G.: Kinetic models of opinion formation. Commun. Math. Sci. **4**(3), 481–496 (2006). URL http://projecteuclid.org/getRecord?id=euclid.cms/1175797553
14. Weidlich, W.: A systematic approach to Mathematical modelling in the social sciences. Harwood Academic Publishers (2000)

Opinion Formation in the Kinetic Exchange Models

Anirban Chakraborti and Bikas K. Chakrabarti

Abstract. We review the minimal multi-agent model (LCCC) for the collective dynamics of opinion formation in the society, which was based on the kinetic exchange dynamics studied in the context of income, money or wealth distributions in a society. This model has an intriguing spontaneous symmetry breaking transition to polarized opinion state starting from non-polarized opinion state. We also briefly review the simple variants and extensions of this model that have been proposed recently.

1 Introduction

It has only been a few decades that physicists have started studying social phenomena and dynamics, leading to the growth of the interdisciplinary field of "Sociophysics" [1]. One of the problems is of "opinion formation", which is a collective dynamical phenomenon, and as such is closely related to the problems of competing cultures or languages [2, 3]. It deals with a "measurable" response of the society to e.g., political issues, acceptances of innovations, etc. Numerous models of competing options have been introduced to study this phenomenon, e.g., the "voter" model (which has a binary opinion variable with the opinion alignment proceeding by a random choice of neighbors) [4], or the Sznajd-Weron discrete opinion formation model (where more than just a pair of spins is associated with the decision

Anirban Chakraborti
Chair of Quantitative Finance, Laboratory of Mathematics Applied to Systems, École Centrale Paris, Châtenay-Malabry, France,
e-mail: anirban.chakraborti@ecp.fr

Bikas K. Chakrabarti
Centre for Applied Mathematics and Computational Science, Saha Institute of Nuclear Physics, Kolkata, India
and
Economic Research Unit, Indian Statistical Institute, Kolkata, India,
e-mail: bikask.chakrabarti@saha.ac.in

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

making procedure) [5]. There have been other studies of systems with more than just two possible opinions [6], or where the opinion of individuals is represented by a "continuous" variable [7–9] using real numbers. Also, since opinion formation in a human society is mediated by social interactions between individuals, such social dynamics has been considered to take place on a network of relationships (see [2] for recent review on such models).

A two body exchange dynamics has already been developed in the context of modelling income, money or wealth distributions in a society [10–14]. The general aim was to study a many-agent statistical model of closed economy (analogous to the kinetic theory model of ideal gases) [15], where N agents exchange a quantity x, that may be defined as wealth. The states of agents are characterized by the wealth $\{x_i\}$, $i = 1, 2, \ldots, N$, such that $x_i > 0, \quad \forall i$ and the total wealth $W = \sum_i x_i$ is conserved. The question of interest is: "What is the equilibrium distribution of wealth $f(x)$, such that $f(x)dx$ is the probability that in the steady state of the system, a randomly chosen agent will be found to have wealth between x and $x + dx$?" The evolution of the system is carried out according to a prescription, which defines the trading rule between agents. The agents interact with each other through a pair-wise interaction characterized by a "saving" parameter λ, with $0 \leq \lambda \leq 1$. The dynamics of the model (CC) is as follows [15]:

$$\begin{aligned} x_i' &= \lambda x_i + \epsilon(1-\lambda)(x_i + x_j)\,, \\ x_j' &= \lambda x_j + (1-\epsilon)(1-\lambda)(x_i + x_j)\,, \end{aligned} \tag{1}$$

where ϵ $(0 \leq \epsilon \leq 1)$ is a stochastic variable, changing with time. It can be noticed that in this way, the quantity x is conserved during the single transactions: $x_i' + x_j' = x_i + x_j$, where x_i' and x_j' are the agent wealths after the transaction has taken place. In general, the functional form for steady state distribution $f(x)$ is seen to be close to the Γ-distribution [16, 17]. As a further generalization, the agents could be assigned different saving propensities and the steady state distribution $f(x)$ show Pareto-like power-law behavior asymptotically [18, 20]. This model (CCM) is described by the trading rule

$$\begin{aligned} x_i' &= \lambda_i x_i + \epsilon[(1-\lambda_i)x_i + (1-\lambda_j)x_j]\,, \\ x_j' &= \lambda_j x_j + (1-\epsilon)[(1-\lambda_i)x_i + (1-\lambda_j)x_j]\,. \end{aligned} \tag{2}$$

One of the main features of this model, which is supported by theoretical considerations [19–21], is that the wealth distribution exhibits a robust power-law in the asymptotic limit of x. Detailed analytical structure of the collective dynamics in these models are now considerably well-developed [21, 22].

Earlier, Toscani [23] had introduced and discussed kinetic models of (continuous) opinion formation involving both exchange of opinion between individual agents and diffusion of information. Based on this model, During et al [24] proposed another mathematical model for opinion formation in a society that is built of two groups, one group of "ordinary" people and one group of "strong opinion leaders". Starting from microscopic interactions among individuals, they arrived at

a macroscopic description of the opinion formation process. They discussed the steady states of the system, and extended it to incorporate emergence and decline of opinion leaders.

Here, we review the studies of the minimal model (LCCC) for the collective dynamics of opinion formation in the society, based on kinetic exchanges, and its variants or extensions.

2 Model for Opinion Formation and Results

2.1 Homogeneous Multi-agent Model

Following the CC model described in the earlier section, we present a minimal model [25, 26] for the collective dynamics of opinion $O_i(t)$ of the i-th person in the society, consisting of N ($N \longrightarrow \infty$) persons. We assume that any particular person can discuss (interact) only with one other person each time (time increases discretely by unity after each such discussion). A two-person "discussion" is viewed here as a simple two-body *scattering process* in physics (depicted schematically in Fig. 1. Persons in the society may bump onto each other randomly and exchange opinions through such random two-person discussions. In general, a person i could have any opinion O_i between two extreme polarities denoted by $+1$ and -1. In any discussion at time $t+1$, a person *retains* a fraction of his/her older opinion $O_i(t)$, determined by his/her "conviction", parameterized by λ_i. This parameter value is characteristic of a person and does not change with time t. Additionally, the person i is "influenced" stochastically by the other person j during the discussion having the "influence" parameter equal to his/her conviction parameter λ_j. We further assume for simplicity that all agents are *homogeneous* – have the *same* conviction parameter λ. Mathematically the dynamics may be represented by

$$\begin{aligned} O_i(t+1) &= \lambda(O_i(t) + \epsilon_t O_j(t)) \,, \\ O_j(t+1) &= \lambda(O_j(t) + \epsilon'_t O_i(t)) \,, \end{aligned} \tag{3}$$

where the opinion $-1 \leq O_i(t) \leq 1$ for all agents i and time t, the conviction parameter $0 \leq \lambda \leq 1$ is *quenched* (does not change with time), and the stochastic parameters ϵ_t and ϵ'_t are *annealed* variables (change with time) – uncorrelated random numbers uniformly distributed between zero and unity. Note that the equations are linear, but non-linearity is introduced in this model by imposing that $-1 \leq O_i(t) \leq 1$ for all agents i and times t.

The question we are interested is that if such social dynamics continually take place, can any consensus be reached or polarity evolve after a long time? Mathematically, we are interested in the steady state distribution of O and other statistical properties. It is noteworthy that unlike in the market models, here we have no conservation of opinion. Rather, the steady state of value of $\bar{O}(t) = (1/N)|\Sigma_i O_i(t)|$ represents the order of the average opinion in the society after a long time t. We

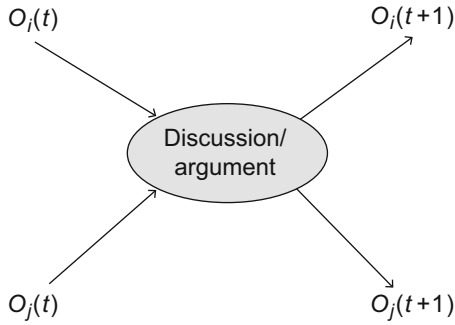


Fig. 1 Random discussions

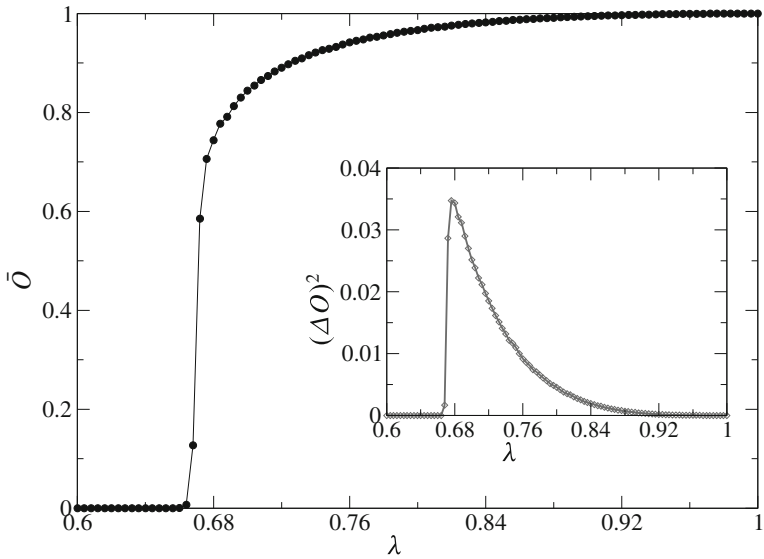


Fig. 2 Numerical results for the variation of the average opinion $\bar{O}(t)$ for large t (steady state value of $\bar{O}$) against λ, following dynamics of Eq. (3). (*Inset*) Numerical results for the variation of the variance $(\Delta O)^2 \equiv \overline{(O - \bar{O})^2}$ against λ, following dynamics of Eq. (3)

study the relaxation dynamics in the society: the relaxation and fluctuation of $\bar{O}$, the steady state value of $\bar{O}(t)$ for $t > \tau$, the relaxation time.

Remarkably, we find there is appearance of "polarity" or consensus, starting from initial random disorder (where O_i's are uniformly distributed with positive and negative values). In the language of physics, there is a "spontaneous symmetry breaking" transition in the system: starting from $\bar{O}(0) = 0$ the system evolves either to the "para" state with $\bar{O} \equiv \bar{O}(t > \tau) = 0$ (where all individual agents have the opinion 0) for $\lambda \leq 2/3$, or (*continuously*) to the "symmetry broken" state

$\bar{O} \equiv \bar{O}(t > \tau) \neq 0$ (where *all* individuals have either positive or negative opinions) for $\lambda \geq 2/3$ (see Fig. 2) for times $t > \tau$. We note, however, that the fluctuation in $\bar{O}$ does not diverge, and shows a cusp near λ_c (see inset of Fig. 2). We also study the relaxation behavior of $\bar{O}(t)$ and the critical divergence of the relaxation time τ near $\lambda = \lambda_c = 2/3$ (see Sect. 2.3, Fig. 5).

2.2 Random Multiplier Map

The basic nature of transition produced by Eq. (3), can perhaps be reproduced by the following simple iterative map

$$O(t+1) = \lambda(1+\epsilon_t)O(t) \tag{4}$$

with the restriction that $O(t) \leq 1$, which is ensured by assuming that if $O(t) \geq 1$, $O(t)$ is set equal to 1. As usual, ϵ_t is a stochastic variable ranging between 0 and 1 (assumed to be uniformly distributed in our case). In a mean-field approach, the above equation reduces effectively to a multiplier map like $O(t+1) = \lambda(1+\langle\epsilon\rangle)O(t)$, where $\langle\epsilon\rangle = 1/2$. Clearly for $\lambda \leq 2/3$, $O(t)$ converges to zero. The initial value $O(0)$, is assigned either a positive or negative value. If it starts from a positive (negative) value, $O(t)$ remains positive (negative). We note that there are subtle differences in the dynamics of Eq. (3) and Eq. (4). Apart from the absence of

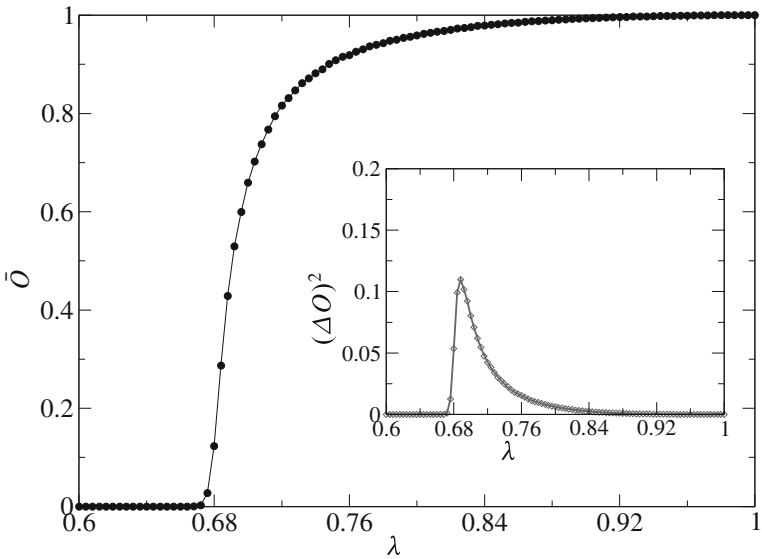


Fig. 3 Numerical results for the variation of the average opinion $\bar{O}(t)$ for large t (steady state value of $\bar{O}$) against λ, following dynamics of Eq. (4). (*Inset*) Numerical results for the variation of the variance $(\Delta O)^2 \equiv \overline{(O-\bar{O})^2}$ against λ, following dynamics of Eq. (4)

"spontaneous symmetry breaking" of the multi-agent model (from $\pm O_i(0)$ values to all positive or all negative transition beyond λ_c), the nature of the phase transition (singularity) in the iterative map is also slightly different. The critical value $\lambda_c = \exp\{-(2\ln 2 - 1)\} \approx 0.6796$ has an analytical derivation [27], but for most numerical studies done here, we take $\lambda_c = 0.68$. The time variation of the average opinion $\bar{O}(t) = (1/N)\Sigma_i|O_i(t)|$, where i refers to different initial realizations and N refers to the total of all such realizations, and its fluctuations are studied numerically (see Fig. 3). We study the relaxation behavior of $\bar{O}(t)$ and the critical divergence of the relaxation time τ near $\lambda = \lambda_c = 0.68$ (see Sect. 2.3, Fig. 6). We note again that the fluctuation in $\bar{O}$ does not diverge, and shows a cusp near $\lambda_c = 0.68$ (see inset of Fig. 3). We also note that the steady state fluctuation ΔO near λ_c, is generally much higher in magnitude for the map case.

2.3 Results and Analyses

For both the multi-agent model and the iterative map, we study the variation (with λ) of the fraction p of the agents having $O_i = \pm 1$ at any time t in the steady state ($t > \tau$). This parameter p gives the average "condensation" fraction (of people in the society having extreme opinions $|O_i| = 1$) in the steady state. The growth of p, as shown in Fig. 4, is seen to be similar to that of $\bar{O}$. The inset shows that the growth

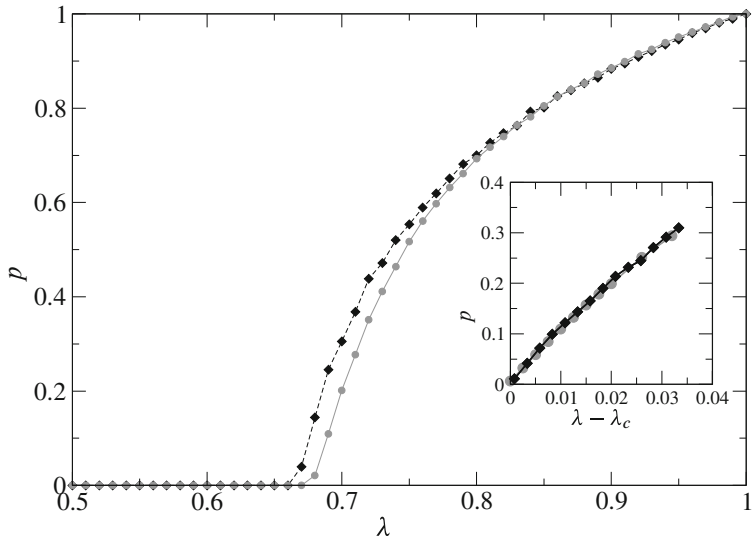


Fig. 4 Numerical results for the variation of the average condensate fraction $p(t)$ for large t (steady state value of p) against λ, following dynamics of Eq. (3) in black diamonds, and dynamics of Eq. (4) in *gray circles*. (*Inset*) Numerical results for the growth of p, following dynamics of Eq. (3) in *black diamonds*, and dynamics of Eq. (4) in *gray circles*, close to λ_c

behavior for p above (respective) λ_c, for both the multi-agent model and map, are identical.

We studied the relaxation behavior of $\bar{O}$ and p, for both the multi-agent model and map. In each case, the relaxation time is estimated numerically from the time value at which $\bar{O}$ or p first touches the steady state value within a pre-assigned error limit. We find diverging growth of relaxation time τ (for both $\bar{O}$ and p) near $\lambda = \lambda_c$ (see Fig. 5 and Fig. 6). The values of exponent z for the divergence in $\tau \sim |\lambda - \lambda_c|^{-z}$ have been estimated numerically for both the multi-agent model and the map (for

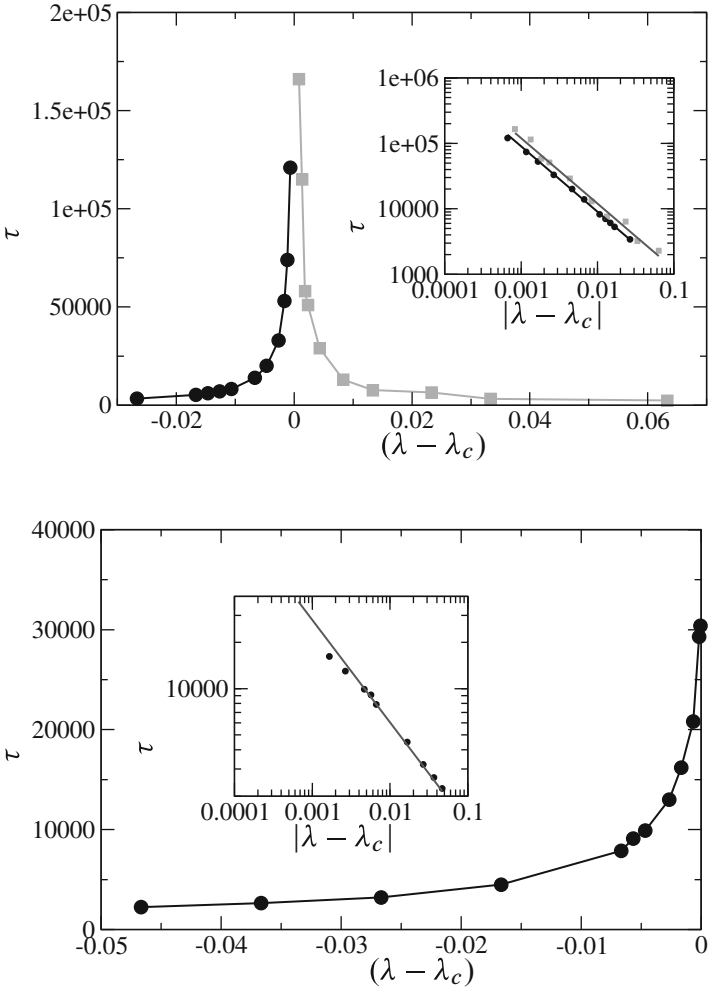


Fig. 5 Numerical results for relaxation time behaviors τ versus $\lambda - \lambda_c$, for (*top*) multi-agent model with $\bar{O}$ and (*bottom*) multi-agent model with p. (*Insets*) Determination of exponent z from numerical fits of $\tau \sim |\lambda - \lambda_c|^{-z}$

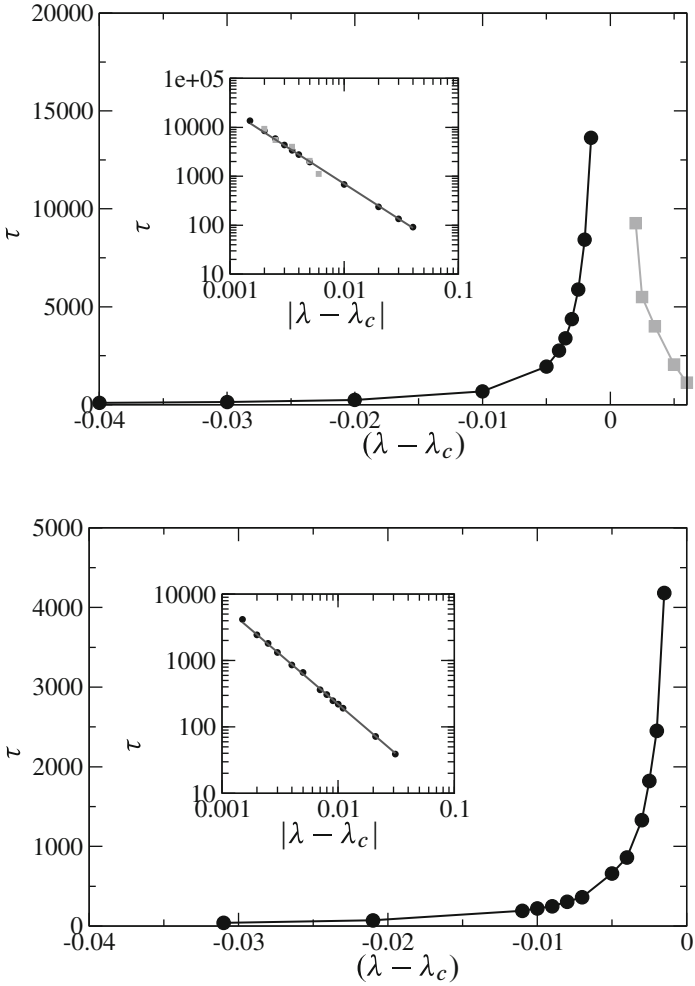


Fig. 6 Numerical results for relaxation time behaviors τ versus $\lambda - \lambda_c$, for (*top*) map with $\bar{O}$ and (*bottom*) map with p. (*Insets*) Determination of exponent z from numerical fits of $\tau \sim |\lambda - \lambda_c|^{-z}$

both $\lambda > \lambda_c$ and $\lambda < \lambda_c$, wherever accurate data were obtained). For the multi agent model, the fitting values for exponent z corresponding to $\bar{O}$ and p, respectively, are $z \approx 1.0 \pm 0.1$ and $z \approx 0.7 \pm 0.1$. For the map case, the fitting values for exponent z corresponding to both $\bar{O}$ and p, turn out to be the same: $z \approx 1.5 \pm 0.1$.

For the iterative map Eq. (4), we study carefully the time evolution of the condensation fraction p of $|O| = 1$ in different realizations at different values of λ. The variation of the steady state value p against λ is shown in Fig. 3. It may be noted that while the steady state value of $\bar{O}$ starts to grow from $\lambda \approx 2/3$ (see Fig. 3), the steady state value of p starts growing at $\lambda \approx 0.68$ (see Fig. 4). Numerical results for

the growth of the relaxation time τ for both $\bar{O}$ and p, against λ are shown in Fig. 6. Both diverge at $\lambda \approx 0.68$. This clearly indicates that p, rather than $\bar{O}$, is the order parameter for the transition.

An approximate analysis of the above transition for λ closer to unity can be done for the iterative map Eq. (4) as follows. In Fig. 7, we give the numerical results for the steady state distribution opinion, $P(|O|)$ for three different values of λ; we observe roughly a bi-modal nature of the distribution as $\lambda \to 1$: one mode is the uniform distribution within the range $|O_{min}| < |O| < 1$ (and $|O_{min}| \approx \lambda$) and another a delta function at $|O| = 1$. We therefore approximate the steady state distribution of opinion by assuming that opinion $O(t)$ is distributed uniformly starting from a minimum O_{min} upto unity with (integrated) probability $(1 - p)$, and a δ-function at exactly unity with probability p. Then

$$\bar{O} = (1 - p)O_{av} + p.1, \tag{5}$$

where $O_{av} = (O_{min} + 1)/2$. We have assumed that the value $O(t)$ stays in those two regions (from λ to 1 and exactly at 1) with probability $(1 - p)$ and p. Hence, the corresponding equations are

$$O(t + 1) = \lambda(1 + \epsilon)O(t) \qquad \text{with probability } 1 - p,$$

and

$$O(t + 1) = 1 \qquad \text{with probability } p.$$

Note that the first equation is realized only if $\lambda(1 + \epsilon)O(t) < 1$ or $\epsilon < \epsilon_{max} = \frac{1}{\lambda O_{av}} - 1$. This cut-off implies that $(1 - p) = \int_0^{\epsilon_{max}} d\epsilon = \frac{1}{\lambda O_{av}} - 1$, since

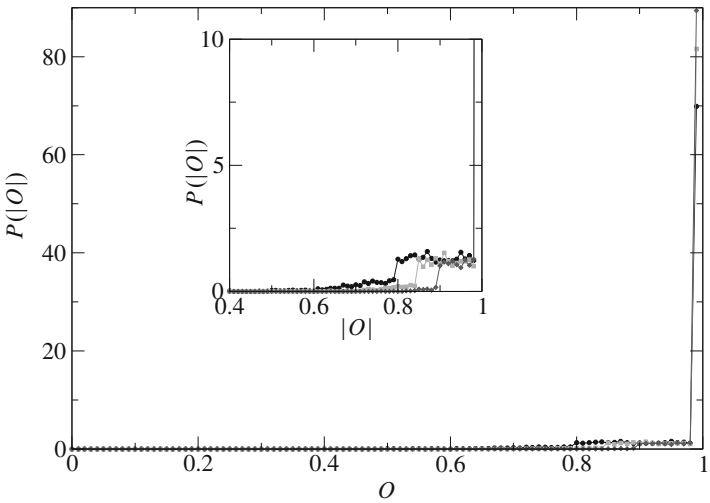


Fig. 7 Numerical results for for the steady state distribution opinion, $P(|O|)$, for three values $\lambda = 0.8, 0.85, 0.9$ showing bi-modal distributions in each case. (*Inset*). The same steady state distribution $P(|O|)$, for three values $\lambda = 0.8, 0.85, 0.9$, but close to $|O| = 1$

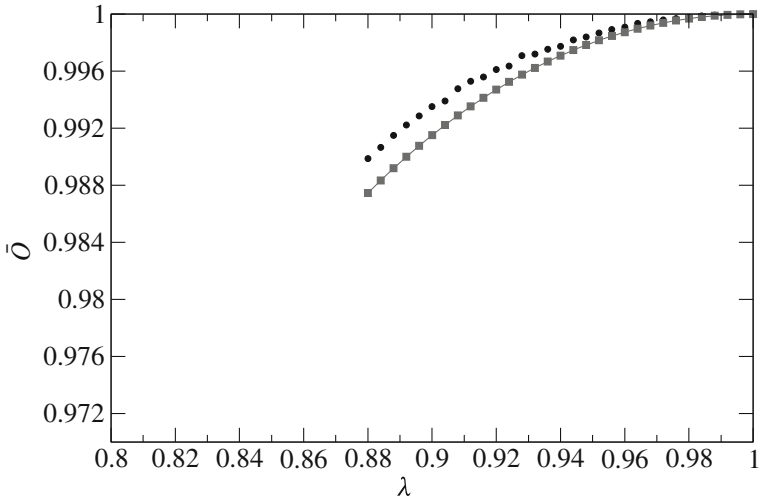


Fig. 8 Fit of the approximate theoretical calculation Eq. (6) (in *gray squares*) with the numerical simulations for $\lambda \to 1$, following dynamics of Eq. (4) (in *black circles*)

$\epsilon \sim$uni[0,1]. By substituting O_{av} and p in Eq. (5), we derive the result that

$$\bar{O} = \frac{5\lambda + 2\lambda^2 - \lambda^3 - 2}{2\lambda(1+\lambda)} \tag{6}$$

which is compared with the numerical simulations for $\lambda \to 1$ in Fig. 8. It is evident that the approximation holds well, only for $\lambda \to 1$.

3 Discussion and Extension

Above, we proposed a minimal model for the collective dynamics of opinion formation in the society, by modifying kinetic exchange dynamics studied in the context of markets. The multi-agent model (dynamics given by Eq. (3)) and its map version (dynamics given by Eq. (4)) have kinetic exchange like linear contributions from random two-person discussions or scattering processes, though the saturation of $|O_i| \leq 1$ induces non-linearity in the dynamics. This model has an intriguing spontaneous symmetry breaking transition to polarized opinion state starting from non-polarized opinion state. Specifically, in the multi-agent model, we see that for $\lambda > \lambda_c = 2/3$, starting from random positive and negative O_i values (or for that matter any arbitrary state), at $t = 0$, the system eventually evolves to a state at $t > \tau$ where *all* O_i are either positive or negative, with $|\bar{O}|$ determined by the λ value! This is similar to the growth of spontaneous magnetization in Ising magnets (where starting from arbitrary up and down spin states, a preferred direction is chosen by fluctuation), with magnetization determined by the temperature below

its transition value. The appearance of spontaneous symmetry breaking in this simple kinetic opinion exchange model is truly remarkable. It appears to be one of the simplest collective dynamical model of many-body dynamics showing non-trivial phase transition behaviour. Indeed, it may be noted that for $\lambda \leq \lambda_c$, at $t > \tau$, *all* O_i's become identically zero (without any fluctuation), while for $\lambda > \lambda_c$, O_i's have fluctuations but the average has a steady state value depending on the value of λ. We have only one absorbing state in our model and, as such, the nature of the phase transition in this model is quite different and does not fit to the commonly studied two absorbing state models (see e.g., [28]). In order to understand the nature of the transition, we also studied a simple iterative map and derived approximate result for the order parameter variation under certain limits, which compares quite well with the numerical simulations. Specifically, we find that the fraction p of people with extreme opinion $|O_i| = 1$, and its fluctuations determine the nature of the phase transition in our model and locate the critical point accurately (from numerical studies). With the two mode distribution (uniform and delta) of O, valid close to $\lambda \to 1$ (see Fig. 7), we could develop an approximate analysis of the variation of the steady state mean opinion $|\bar{O}|$ against λ as in Eq. (6). In any case, further investigations are necessary for understanding this phase transition. Also, the study of this phase transition behavior for an extended model with separate"conviction" and "influence" parameters in Eq. (3), has recently been reported [29], which we briefly discuss below. We also briefly summarize below, the lattice version of the above model and non-equilibrium relaxation in these systems, as studied by Biswas et al [30]. Additional studies for the *heterogeneous* conviction factors λ_i's, in influence of "field terms" that represent the external influence of media, etc. are indicated below.

3.1 Two Parameter Model: Sen

In [29], Sen generalised the model to incorporate two parameters, λ, to represent conviction and μ, to represent the influencing ability of individuals.

In the model, let $O_i(t)$ be the opinion of the ith individual at time t; then after an interaction of the ith and jth agents, the opinions of the two individuals are changed according to

$$O_i(t+1) = \lambda_i O_i(t) + \epsilon_1 \mu_j O_j(t) \tag{7}$$

$$O_j(t+1) = \lambda_j O_j(t) + \epsilon_2 \mu_i O_i(t) \tag{8}$$

where ϵ_1 and ϵ_2 are independent random variables ranging from zero to one. Making $\lambda_i = \mu_i$, one gets back the LCCC model [25, 26]. The opinions of both the individuals are changed at the same time, and the interacting individuals are chosen randomly. In the simplest case they keep the two parameters λ and μ independent of the agents, i.e, assume a homogeneous population having identical λ and μ.

In the steady state, the condition for nonzero solutions of $\langle x_i \rangle$ is obtained as,

$$(1-\lambda)^2 = \langle \epsilon_1 \epsilon_2 \rangle \, \mu^2. \tag{9}$$

Since ϵ_1 and ϵ_2 are independent random variables with mean value equal to 0.5, and as λ, μ cannot exceed 1, the above condition reduces to

$$\lambda = 1 - \mu/2. \tag{10}$$

In the simulations with N individuals whose opinions are randomly distributed initially, they first investigated the steady state behaviour and found that indeed, there is a threshold phenomena as the average opinions shows spontaneous symmetry breaking above a phase boundary occurring in the $\lambda - \mu$ plane (the phase boundary obtained numerically matches exactly with (10). For $\lambda = 1$, they found that the final state is not only ordered, but it is completely polarised in the sense that the opinions of all the individuals are equal and exactly 1 (or -1), irrespective of the value of μ. Therefore, there is a line in the ordered phase where the fluctuations vanish completely. The effect of μ for other values of λ is not irrelevant; for any $\lambda \neq 1$, the nature of the phase is dictated by both μ and λ. The phase diagram is shown in Fig. 9. They investigated the nature of the transition at different points on the phase boundary, by varying the parameters close to the transition points (λ_c, μ_c) on the phase boundary, which could be done in several ways in a two dimensional plane. They chose two trajectories: path A, where μ is kept fixed at μ_c and λ is varied, and path B, where λ is fixed at λ_c and μ is varied. In some special cases, all possible trajectories could not be explored, e.g., for $\lambda_c = 1, \mu_c = 0$, the path A does not exist.

They also studied the behaviour of some quantities close to the phase transition point. The equilibrium value of the order parameter m showed a power law behaviour with $(\lambda - \lambda_c)$,

$$m \propto (\lambda - \lambda_c)^{\beta} \tag{11}$$

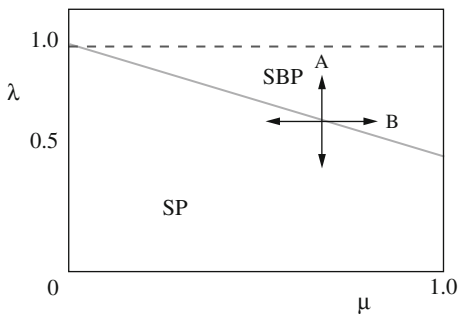


Fig. 9 The phase boundary obtained by numerical simulation coincides exactly with that given in Eq. (10). SP denotes the symmetric phase and SBP the symmetry broken phase. The paths A and B are possible trajectories along which the different studies can be made. Along the *dashed line* $\lambda = 1$, the opinions of all the agents are equal and take extreme values in two possible ways, either $x_i = 1$ or $x_i = -1$ for all i. Courtesy P. Sen [29]

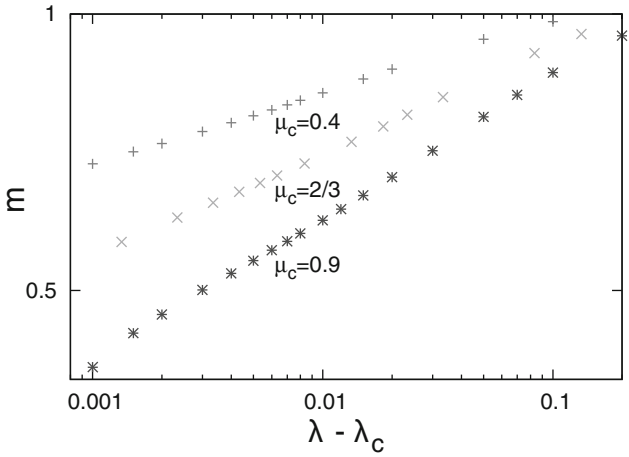


Fig. 10 The equilibrium value of order parameter as a function of $\lambda - \lambda_c$ is shown for different values of μ_c along path A for $N = 256$. The exponent is seen to be appreciably dependent on μ_c. Courtesy P. Sen [29]

where β is quite strongly dependent on the point on the phase boundary, e.g., $\beta = 0.079 \pm 0.001$ at $\mu_c = 0.4$ and $\beta = 0.155 \pm 0.001$ at $\mu_c = 0.9$ (shown in Fig. 10). This result shows that the phase transition is non-universal.

In summary, in this model a phase boundary given by $\lambda = 1 - \mu/2$ is obtained separating the symmetric and symmetry broken phases. The phase transition along the boundary is shown to be non-universal as the exponents have different at different points on the boundary. The time scale diverges near the phase boundary in a power law manner and the order parameter also shows a power law behaviour.

3.2 Lattice Version and Non-equilibrium Relaxation Studies: Biswas et al.

In [30], the authors study the lattice version of the model (LCCC) [26], where the agents are arranged on a 1D lattice, and a randomly chosen nearest neighbor pair update their opinions according to Eq. (3), such that an agent only exchange opinion with one of its two nearest neighbors. Then they study the non-equilibrium relaxation, using the non-equilibrium relaxation technique proposed in [31]. In non-equlibrium relaxation the simulation is started from a fully ordered state. Away from the critical point, the order parameter shows an exponential relaxation, while at the critical point the relaxation is a power law.

In the original version of the LCCC model, the order parameter p decays as $p \sim t^{-\delta}$, where $\delta = 1.2 \pm 0.1$. The critical point was estimated to be $\lambda_c \approx 0.66659 \pm 0.00002$. For the lattice version, however, the critical point changes slightly to $\lambda_c = 0.66679 \pm 0.00001$ ($N = 1200$). The critical exponent $\delta = 1.15 \pm 0.01$.

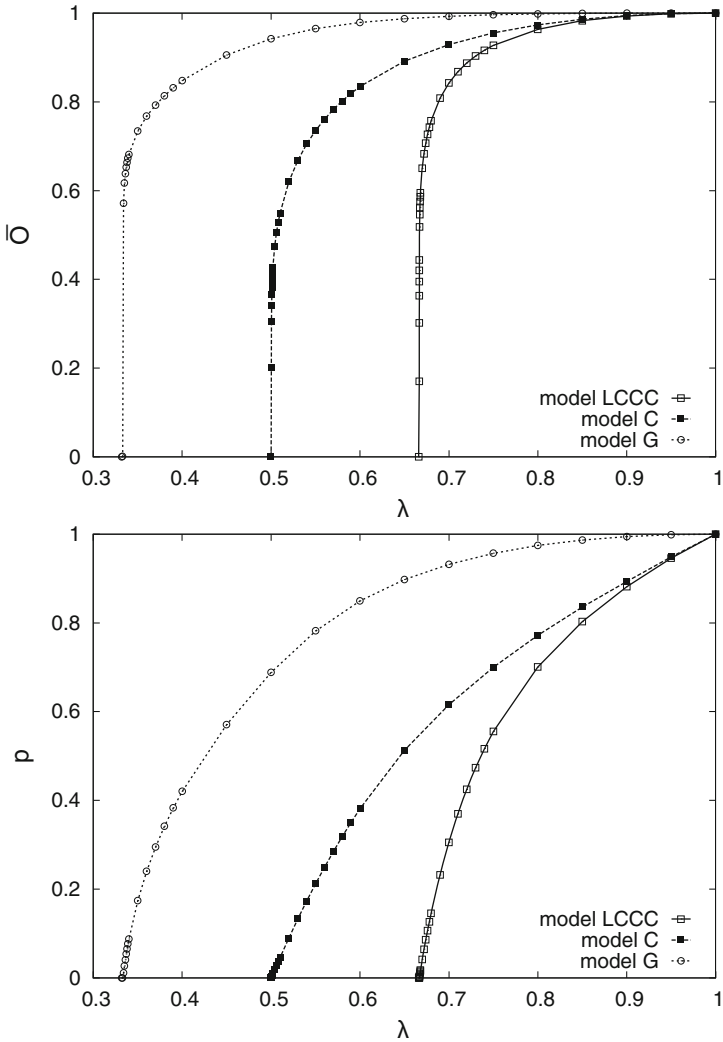


Fig. 11 Variation of order parameters O and p with λ for the 3 models. Courtesy A. Chatterjee [30]

A simpler model (C) was also proposed. There, the dynamics was

$$\begin{aligned} O_i(t+1) &= \lambda O_i(t) + \epsilon O_j(t) \\ O_j(t+1) &= \lambda O_j(t) + \epsilon' O_i(t). \end{aligned} \tag{12}$$

The critical point, using a mean field approach, was found to be $\lambda_c = 1 - \langle \epsilon \rangle$. In the case of a uniform distribution of ϵ, the critical point is $\lambda_c = 1/2$.

They also studied the effect of adding a feedback of the global opinion in the LCCC model (G). The dynamics follows

$$\begin{aligned} O_i(t+1) &= \lambda[O_i(t) + \epsilon O_j(t)] + \epsilon' \bar{O}(t) \\ O_j(t+1) &= \lambda[O_j(t) + \eta O_i(t)] + \eta' \bar{O}(t) \end{aligned} \tag{13}$$

where ϵ, ϵ', η, η' are drawn randomly from an uniform distribution.

From a mean-field approach, it is easy to estimate the critical point at $\lambda_c = 1/3$. It was also verified numerically.

The phase diagram is shown in Fig. 11 for the three models using the order parameter p, as well as $\bar{O}$.

3.3 Heterogeneous Multi-agent Model

Following the CCM model, described in the introduction, we could extend the minimal model [25] for the collective dynamics of opinion $O_i(t)$ of the i-th person in the society of N ($N \longrightarrow \infty$) persons, as:

$$\begin{aligned} O_i(t+1) &= \lambda_i O_i(t) + \epsilon_t \lambda_j O_j(t)\,, \\ O_j(t+1) &= \lambda_j O_j(t) + \epsilon'_t \lambda_i O_i(t)\,, \end{aligned} \tag{14}$$

where $-1 \leq O_i(t) \leq 1$ for all i and t, and $0 \leq \lambda_i \leq 1$'s are *quenched* variables (do not change with time, but vary from person to person), and ϵ_t and ϵ'_t are *annealed* variables (changing with time), that are random numbers uniformly distributed between zero and unity. As before, nonlinearity is kept in this model by assuming that $-1 \leq O_i(t) \leq 1$ for all agents i and times t, by ensuring the bound on $O_i(t)$: if $O_i(t) \leq -1$ or $O_i(t) \geq 1$, then $O_i(t) = -1$ or 1 respectively. Here, we assume λ_i's to be uniformly spread in the interval [0,1) (equivalent to the CCM model for market dynamics). We study similarly, starting from "symmetric" states (with random positive and negative values of $O_i(0)$, the evolution of the system. The dynamics here leads collectively to the "Polarized" or "Symmetry broken" state ($O_i(t)$ are either all positive or all negative, for all i, and times $t > \tau$) only. The "indifferent" states (with $O_i(t) = 0$ for all i, for times $t > \tau$) disappear in the large system size limit, although this is clearly a *fixed point* of the dynamics given by Eq. 14.

It may be noted that the above dynamics can be considerably modified by the presence of "polarizing field" terms h_i (fixed over time t but dependent on person i), added linearly to the dynamical equations Eq. 14 of $O_i(t)$. Such "fields" can be provided by the "influences" of the media in the society. Detailed analyses of the field terms, etc. will be reported elsewhere [32].

Acknowledgements The authors acknowledge F. Abergel, A. Chatterjee, A. Jedidi, K. Kaski, S.N. Majumdar, M. Marsili, N. Millot, M.A. Muñoz, M. Patriarca and S. Pradhan for useful discussions, comments and correspondences.

References

1. B.K. Chakrabarti, A. Chakraborti and A. Chatterjee (Eds.), Econophysics and Sociophysics: Trends and Perspectives (Wiley-VCH, Berlin, 2006)
2. C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009)
3. C. Schulze and D. Stauffer, in [1], p 311; G. Weisbuch, in [1], p 339; S. Galam, in [1], p 367
4. R. A. Holley and T.M. Liggett, Ann. Probab. **3**, 643 (1975)
5. K. Sznajd-Weron and J. Sznajd, Int. J. Mod. Phys. **C 11**, 1157 (2000)
6. F. Vazquez, P.L. Krapivsky, and S. Redner, J. Phys. **A 36**, L61 (2003)
7. R. Hegselman, U. Krause, J. Artif. Soc. Soc. Simul. **5**, 2 (2002)
8. G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Adv. Complex Syst. **3**, 87 (2000)
9. S. Fortunato, Int. J. Mod. Phys. **C 16**, 17 (2005)
10. V. Yakovenko, J. Rosser, Rev. Mod. Phys. **81**, 1703 (2009)
11. A. Chatterjee, B.K. Chakrabarti, Eur. Phys. J. **B 60**, 135 (2007)
12. M. Patriarca, E. Heinsalu, A. Chakraborti, Eur. Phys. J. **B 73**, 145 (2010)
13. A. Chakraborti, I. Muni Toke, M. Patriarca, F. Abergel, arXiv:0909.1974v2 (2010)
14. P. L. Krapivsky, S. Redner, Science and Culture (Kolkata) **76**, 424 (2010); arXiv:1006.4595 [physics.soc-ph]
15. A. Chakraborti, B.K. Chakrabarti, Eur. Phys. J. **B 17**, 167 (2000)
16. A. Chakraborti, M. Patriarca, Pramana **71**, 233 (2008)
17. M. Patriarca, A. Chakraborti, K. Kaski, Phys. Rev. **E 70**, 016104 (2004)
18. A. Chatterjee, B.K. Chakrabarti, S.S. Manna, Physica **A 335**, 155 (2004)
19. P.K. Mohanty, Phys. Rev. E 74, 011117 (2006)
20. A. Chakraborti, M. Patriarca, Phys. Rev. Lett. **103**, 228701 (2009)
21. P. Repetowicz, S. Hutzler, P. Richmond, Physica A 356, 641 (2005)
22. D. Matthes, G. Toscani, J. Stat. Phys. 130, 1087 (2008); F. Bassetti, G. Toscani, arxiv:1002.3689 (2010)
23. G. Toscani, Commun. Math. Sci. **4**, 481 (2006)
24. B. During, P. Markowich, J.-F. Pietschmann and M.-T. Wolfram, Proc. R. Soc. **A 465**, 3687 (2009)
25. M. Lallouache, A. Chakraborti and B.K. Chakrabarti, Science and Culture (Kolkata) **76**, 485 (2010); arXiv:1006.5921 [physics.soc-ph]
26. M. Lallouache, A.S. Chakrabarti, A. Chakraborti and B.K. Chakrabarti, Phys. Rev. E **82**, 056112 (2010)
27. M. Marsili, private communication
28. O. Al Hammal, H. Chaté, I. Dornic, and M.A. Muñoz, Phys. Rev. Lett. **94**, 230601 (2005)
29. P. Sen, arXiv:1009.4315 (2010)
30. S. Biswas, A.K. Chandra, A. Chatterjee, arxiv:1010:3190 (2010)
31. Y. Ozeki, N. Ito, J. Phys. A: Math Theor. **40**, R149 (2007)
32. M. Lallouache, A. Chakraborti, B.K. Chakrabarti and K. Kaski, in preparation

Panel Discussion

In this brief section, we compile some of the remarks and comments made during the panel discussion session that took place during the conference.

A general theme that was addressed in a recurrent fashion was that of the quality of financial data: as a matter of fact, the study of financial markets as a physical phenomenon is made possible – and only up to a point - thanks to the possibility offered by the gigantic data sets one can use to analyze, model and simulate the properties of markets. However, it is clear that data made available to researchers do not always contain the right amount of information one would love to have. For instance, the identity of the agent (a market member) sending a given order is not always available – and some very interesting results – due to Fabrizio Lillo and co-workers – show that an *a priori* knowledge of the type of agents one considers plays a definite role in interpreting some seemingly purely statistical results such as the order flow autocorrelation. Going further down this alley, one would obviously want some information on the final investor when there is one, thereby allowing the researcher to account separately for effects that come from broker-style best execution routines and for those related to actual investment strategies, both at high and low frequencies. To rephrase it a bit more generally, there was a general impression that statistical models of market, no matter how large the calibration set seems to be, still need to incorporate some behavioural knowledge in order to offer a satisfactory representation of real markets.

Another related point that was discussed is the question of whether models should – and could – stay simple, thereby explaining in a satisfactory fashion a particular phenomenon but at the risk of reproducing only very partially the whole of the market, or whether one should accept more complexity in the models, and hence, rely more on numerical simulations. From an application-minded point of view, it is clear that more robust, more general models are in order, and the numerical tools available to the scientific and industrial world are obviously powerful enough to allow realistic simulations of a financial market. But the extent to which such an approach may help one understand the various competing effects influencing the behaviour of markets is debatable. Models should help one have a clear representa-

F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra (Eds.)
Econophysics of Order-driven Markets. © Springer-Verlag Italia 2011

tion of the financial markets, but they should also aim at mimicking and predicting their behaviour in a realistic fashion. Sometimes, these two targets require very different approaches

Below are two contributions to this panel discussion, kindly sent by their respective authors.

The first one is by Tobias Preis:

- Econophysics research has been addressing a key question of interest: quantifying and understanding large (and small) stock market fluctuations. Previous work has focused on the challenge of quantifying the behavior of the probability distributions of large fluctuations of relevant variables such as returns, volumes, and the number of transactions. Sampling the far tails of such distributions require a large amount of data. A very large amount of precise historical financial market data has already been collected, many orders of magnitude more than for other complex systems.
- In the meantime, we get very positive feedbacks from physics and econometrics (especially econometrics) as well as from FINANCIAL INDUSTRY (trading companies, hedge funds, banking institutions). Not only IDEAS stemming from "Econophysics" are (more and more) accepted ... In addition, TECHNOLOGY aspects used in statistical physics are becoming mainstream in financial market trading systems (algorithmic trading). That refers to GPU-Computing which was successfully used, e.g., for Monte Carlo Simulations of the Ising model in statistical physics (see T. Preis et al., J. Comp. Phys. 228, 4468 (2009)) -> An example for the efficient usage of GPUs to analyze financial market time series can be found in T. Preis et al., New J. Phys. 11, 093024 (2009).

The second one is by Sitabhra Sinha:

The fifth of the series of Econophys-Kolkata meetings is a rare opportunity to take stock of what is by all measures a remarkable series of meetings. From the first workshop in 2005 which focused on income and wealth distributions to the present one with its emphasis on order-driven markets, these meetings have provided an opportunity for physicists and some economists to discuss several important themes in the rapidly evolving field of econophysics. In view of the economic and financial turmoil that the world has undergone in the past few years, the advent of econophysics has taken on an especially pertinent role in providing possible alternative foundations to the study of economic phenomena. Indeed, just as the last decade of the previous century saw the collapse of one of the dominant economic ideologies, i.e., centralized planning characteristic of many erstwhile socialist countries, we may hope that the tribulations at the end of the first decade of the first century will strike a similar death-blow to another much-vaunted economic ideology, that of neo-liberal free-market capitalism.

Much of the problem with mainstream economics appears to stem from its obsession with axiomatic principles in the tradition of mathematics. Indeed, economics as it is practised today should more aptly be titled "econo-mathematics" rather than the study of empirical phenomena [1]. Econophysics provides a refreshing alternative to this view by considering economics as a natural science (along the lines of physics)

where theory is driven by experimental observations. Thus, we need to seek areas in economics where there exist a set of established empirical "laws" (based on rigorous observations and analysis) that we can seek to explain. It is this approach that primarily differentiates econophysics from mainstream economics, the latter in general having little connection to reality (save the few attempts by game theorists to verify certain results in behavioral economics through controlled experiments which have met with only mixed success).

The two main themes of econophysics, that of financial markets and affluence distribution, illustrates this need for a well-established body of facts upon which theory can be based. The availability of large quantities of empirical data about markets at both the gross (trade data) and finer (order book) levels, and the ability to process them using high-performance computers, has initiated a series of discoveries of universal features in such systems. Similarly, the use of income tax and other related data for various countries to measure the distribution of affluence (as measured by wealth or income) of individuals and organizations, has rejuvenated work in the puzzle of why inequality seems to be universal across societies - a question that had been first raised by Vilfredo Pareto more than a century earlier. Indeed, newer discoveries seem to show the near-ubiquity of the power-law distribution that characterizes inequality in society, so much so, that we may ask whether one can also observe it in primate societies.

However, despite the advances made in these two areas, the reason they have not yet made significant impact among mainstream economists is because neither are central themes in the traditional discipline of economic science. What are then the emerging new areas of econophysics that can shake up conventional economists? As a result of the ongoing financial crisis, the study of economic networks have recently emerged as a major challenge in the general scientific consciousness [2]. The empirical studies of inter-bank money transfer networks in the US and European settings have shown how liable these systems are to cascading failures [3–5]. The connection between network topology and dynamical stability that had earlier been carried out in the context of eco-systems [6–9] suggests that we are in the threshold of achieving a profound understanding of why large systemic collapses occur in the economy. Just as there are several empirical "laws" establishing scaling behavior between ecological variables (e.g., the celebrated relation between area and number of species), one can look for similar *stylized facts* in the context of economic networks.

How about the two central themes of contemporary economics, viz., (a) developmental economics or the theory of economic growth and (b) sustainable economies? The studies of Hidalgo *et al.* on how trajectories in the abstract product space by different countries can be analyzed to understand growth provides a glimpse of the new viewpoint that physics-based approaches provide to study this problem [10]. The other area of how to foster a sustainable economy has taken on a fresh urgency in view of the catastrophic state that our planet has been pushed to by the relentless push for economic development [11]. Economists now are coming round to the viewpoint that the traditional input-output (or "throughput") model of the economy may need to be replaced by a more efficient closed-cycle model more along the lines

of how nature recycles nutrients. The zero-growth economy of Japan has been proposed as a possible model of how we can have a desirable living standard without sacrificing the well-being of future generations in the process. The future challenge to econophysicists is to establish solid theoretical foundations of such a economic model, possibly based on analogies in ecological or environmental sciences. One hopes that future editions of the Econophys-Kolkata meetings will focus on one or more of these themes.

References

1. Samuelson P A (1947) Foundations of Economic Analysis. Harvard University Press, Cambridge, Mass
2. Schweitzer F et al (2009) Economic networks: The new challenges, Science 325:422
3. Soramäki K et al. (2007) The topology of interbank payment flows, Physica A 379:317
4. Iori G et al (2008) A network analysis of the Italian overnight money market, J Econ Dyn Control 32:259
5. May R M, Levin S A, Sugihara G (2008) Ecology for bankers, Nature 451:893
6. Sinha S (2005) Complexity vs stability in small-world networks, Physica A 346:147
7. Sinha S, Sinha S (2005) Evidence of universality for the May-Wigner stability theorem for random networks with local dynamics, Phys Rev E 71:020902(R)
8. Sinha S, Sinha S (2006) Robust emergent activity in dynamical networks, Phys Rev E 74:066117
9. Sinha S (2010) Are large complex economic systems unstable?, Science and Culture (in press, also at arxiv:1009.0972)
10. Hidalgo C A et al (2007) The product space conditions the development of nations, Science 317:482
11. Meadows D H, Randers J, Meadows D L (2004) Limits to Growth: The 30-Year Update. Earthscan, London

New Economic Windows Series

Chatterjee A., Yarlagadda S., Chakrabarti B.K. (Eds.)
Econophysics of Wealth Distributions
2005, ISBN 978-88-470-0329-3

Salzano M., Kirman A.P. (Eds.)
Economics: Complex Windows
2005, ISBN 978-88-470-0279-1

Chatterjee A., Chakrabarti, B.K. (Eds.)
Econophysics of Stock and other Markets
2006, ISBN 978-88-470-0501-3

Salzano M., Colander D. (Eds.)
Complexity Hints for Economic Policy
2007, ISBN 978-88-470-0533-4

Chatterjee A., Chakrabarti, B.K. (Eds.)
Econophysics of Markets and Business Networks
2007, ISBN 978-88-470-0664-5

Gatti D., Gaffeo E., Gallegati M., Giulioni G., Palestrini A.
Emergent Macroeconomics. An Agent-Based Approach to Business Fluctuations
2008, ISBN 978-88-470-1561-6

Faggini M., Lux T. (Eds.)
Coping with the Complexity of Economics
2009, ISBN 978-88-470-1082-6

Basu B., Chakrabarti B.K., Chakravarty S.R., Gangopadhyay K. (Eds.)
Econophysics & Economics of Games, Social Choices and Quantitative Techniques
2010, ISBN 978-88-470-1500-5

Faggini M., Vinci C.P. (Eds.)
Decision Theory and Choices: a Complexity Approach
2010, ISBN 978-88-470-1777-1

Abergel F., Chakrabarti B.K., Chakraborti A., Mitra M. (Eds.)
Econophysics of Order-driven Markets
2011, ISBN 978-88-470-1765-8