New Economic Windows

Frédéric Abergel
Bikas K. Chakrabarti
Anirban Chakraborti
Nivedita Deo
Kiran Sharma *Editors*

New Perspectives and Challenges in Econophysics and Sociophysics

Springer

New Perspectives and Challenges in Econophysics and Sociophysics

New Economic Windows

More information about this series at http://www.springer.com/series/6901

Frédéric Abergel · Bikas K. Chakrabarti ·
Anirban Chakraborti · Nivedita Deo ·
Kiran Sharma
Editors

New Perspectives and Challenges in Econophysics and Sociophysics

Springer

Editors
Frédéric Abergel
Centrale Supélec
Gif-sur-Yvette, France

Bikas K. Chakrabarti
Saha Institute of Nuclear Physics
Kolkata, India

Anirban Chakraborti
Jawaharlal Nehru University
New Delhi, India

Nivedita Deo
University of Delhi
New Delhi, India

Kiran Sharma
Jawaharlal Nehru University
New Delhi, India

ISSN 2039-411X ISSN 2039-4128 (electronic)
New Economic Windows
ISBN 978-3-030-11363-6 ISBN 978-3-030-11364-3 (eBook)
https://doi.org/10.1007/978-3-030-11364-3

Library of Congress Control Number: 2018967429

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

The search for truth should be the goal of our activities; it is the sole end worthy of them.

—Henri Poincaré

Preface

This volume contains essays that were mostly presented in the joint international workshop entitled "Econophys-2017" and "APEC-2017," held at the Jawaharlal Nehru University and University of Delhi, New Delhi, during November 15–18, 2017. For the first time, the Econophys series and the Asia Pacific Econophysics Conference (APEC) series merged together to have a great workshop, which was organized jointly by the Jawaharlal Nehru University, University of Delhi, Saha Institute of Nuclear Physics, and CentraleSupélec. We received great support and encouragement from the steering committee of the APEC.

Economic and financial markets appear to be in a permanent state of flux. Billions of agents interact with each other, giving rise to complex dynamics of economic quantities at the micro- and macro-levels. With the availability of huge data sets, researchers are able to address questions at a more granular level than were possible earlier. Fundamental questions of aggregation of action and information, coordination, complexity, and evolution of economic and financial networks have received significant importance in the current research agenda of the Econophysics literature. In parallel, the Sociophysics literature has focused on large-scale social data and their inter-relations. Empirical approach has become a front-runner in finding short-lived patterns within the data. The essays appearing in this volume include the contributions of distinguished experts and researchers and their co-authors from varied communities—economists, sociologists, financial analysts, mathematicians, physicists, statisticians, and others. A positive trend for this interdisciplinary track is that more and more sociologists, economists, and statisticians have started interacting with the physicists, mathematicians, and computer scientists! Evidently, most have reported their recent works and reviews on the analyses of economic and social behaviors. A few papers have been included that were accepted for presentation but were not presented at the meeting since the contributors could not attend due to unavoidable reasons. The contributions are organized into three parts. The first part comprises papers on "Econophysics". The papers appearing in the second part include studies in "Sociophysics". Finally, the third part is Miscellaneous, containing a proposal for an Interdisciplinary research center, and an "Epilogue", which discusses the advent of "Big data" research.

We are grateful to all the local organizers and volunteers for their invaluable roles in organizing the meeting, and all the participants for making the workshop a success. We acknowledge all the experts for their contributions to this volume. We also thank Vishwas Kukreti, Arun Singh Patel, Hirdesh Kumar Pharasi, and Amrita Singh for their help in the LaTeX compilation of the articles. The editors are also grateful to Mauro Gallegati and the Editorial Board of the New Economic Windows series of the Springer-Verlag (Italy) for their continuing support in publishing the Proceedings in their esteemed series.[1] The conveners (editors) also acknowledge the financial support from Jawaharlal Nehru University, University of Delhi, and CentraleSupélec.

Gif-Sur-Yvette, France — Frédéric Abergel
Kolkata, India — Bikas K. Chakrabarti
New Delhi, India — Anirban Chakraborti
New Delhi, India — Nivedita Deo
New Delhi, India — Kiran Sharma
September 2018

[1] Past volumes:

1. Econophysics and Sociophysics: Recent Progress and Future Directions, Eds. F. Abergel, H. Aoyama, B. K. Chakrabarti, A. Chakraborti, N. Deo, D. Raina and I. Vodenska, New Economic Windows, Springer-Verlag, Milan, 2017.
2. Econophysics and Data Driven Modelling of Market Dynamics, Eds. F. Abergel, H. Aoyama, B. K. Chakrabarti, A. Chakraborti, A. Ghosh, New Economic Windows, Springer-Verlag, Milan, 2015.
3. Econophysics of Agent-based models, Eds. F. Abergel, H. Aoyama, B. K. Chakrabarti, A. Chakraborti, A. Ghosh, New Economic Windows, Springer-Verlag, Milan, 2014.
4. Econophysics of Systemic Risk and Network Dynamics, Eds. F. Abergel, B. K. Chakrabarti, A. Chakraborti and A. Ghosh, New Economic Windows, Springer-Verlag, Milan, 2013.
5. Econophysics of Order-driven Markets, Eds. F. Abergel, B. K. Chakrabarti, A. Chakraborti, M. Mitra, New Economic Windows, Springer-Verlag, Milan, 2011.
6. Econophysics & Economics of Games, Social Choices and Quantitative Techniques, Eds. B. Basu, B. K. Chakrabarti, S. R. Chakravarty, K. Gangopadhyay, New Economic Windows, Springer-Verlag, Milan, 2010.
7. Econophysics of Markets and Business Networks, Eds. A. Chatterjee, B. K. Chakrabarti, New Economic Windows, Springer-Verlag, Milan 2007.
8. Econophysics of Stock and other Markets, Eds. A. Chatterjee, B. K. Chakrabarti, New Economic Windows, Springer-Verlag, Milan 2006.
9. Econophysics of Wealth Distributions, Eds. A. Chatterjee, S. Yarlagadda, B. K. Chakrabarti, New Economic Windows, Springer-Verlag, Milan, 2005.

Contents

Part I
Econophysics

Chapter 1
Strategic Behaviour and Indicative Price Diffusion in Paris Stock Exchange Auctions

Damien Challet

Abstract We report statistical regularities of the opening and closing auctions of French equities, focusing on the diffusive properties of the indicative auction price. Two mechanisms are at play as the auction end time nears: the typical price change magnitude decreases, favoring underdiffusion, while the rate of these events increases, potentially leading to overdiffusion. A third mechanism, caused by the strategic behavior of traders, is needed to produce nearly diffusive prices: waiting to submit buy orders until sell orders have decreased the indicative price and vice-versa.

Introduction

Research in market micro-structure has focused on the dynamical properties of open markets [5, 9]. However, main stock exchanges have been using auction phases when they open and close for a long time.[1] Auctions are known to have many advantages, provided that there are enough participants: for example, auction prices are well-defined, correspond to larger liquidity, and decrease price volatility (and bid-ask spreads) shortly after the opening time and before closing time (see e.g. [8, 10, 11]).

Only a handful of papers are devoted to the dynamics of auction phases, i.e., periods during which market participants may send limit or market orders specifically for the auction. Reference [6] investigates when fast and slow traders send their orders during the opening auction phase of the Paris Stock Exchange and find markedly different behaviors: the slow brokers are active first, while high-frequency traders are mostly active near the end of auctions. In the same vein, [3] shows how and when

[1] London Stock Exchange and XETRA (Germany) recently added a mid-day short auction phase.

D. Challet (✉)
Laboratoire de Mathématiques et Informatique pour la Complexité et les Systèmes, Centralesupélec, Université Paris Saclay, Saint-Aubin, France
e-mail: damien.challet@centralesupelec.fr

F. Abergel et al. (eds.), *New Perspectives and Challenges in Econophysics and Sociophysics*, New Economic Windows,
https://doi.org/10.1007/978-3-030-11364-3_1

low-latency traders (identified as high-frequency traders) add or remove liquidity in the pre-opening auction of the Tokyo Stock Exchange. Accordingly, [12] finds typical patterns of high-frequency algorithmic trading in the auctions of XTRA. Finally, [7] analyzes anonymous data from US equities and compute the response functions of the final auction price to the addition or cancellation of auction orders as a function of the time remaining until the auction, which have strikingly different behaviors in the opening and closing auction phases.

Auctions, Data and Notations

The opening auction phase of Paris Stock Exchange starts at 7:15 and ends at 9:00 while the closing auction phase is limited to the period 17:30 to 17:35. The auction price maximises the matched volume.

From the Thomson Reuters Tick History, we extract auction phase data for the 2013-04-16 components of the CAC40 index. This database contains all the updates to either the indicative match price or the indicative matched volume in the 2010-08-02 to 2017-04-12 period, which amounts to 8,095,524 data points for the opening auctions and 15,007,048 for the closing auctions. Note that the closing auction phase has about twice as many updates despite being considerably shorter.

For each asset α, we denote the indicative price of auction $x \in \{\text{open}, \text{close}\}$ of day d at time t by $\pi^x_{\alpha,d}(t)$, the time of auction x by t^x and the auction price by $p^x_{\alpha,d}$. Dropping the index α since this paper focuses on a single asset at a time, the i-th indicative price change occurs at physical time $t^x_{i,d}$ and its log-return equals $\delta p^x_{i,d} = \log \pi^x_d(t^x_{i,d}) - \log \pi^x_d(t^x_{i-1,d})$. It is useful to work in the time-to-auction (TTA henceforth) time arrow: setting $\tau = t^x - t$, the log-return between the final auction price and the current indicative is then $\Delta p^x_d(\tau) = \log p^x_d - \log \pi^x_d(t)$.

Similarly, the indicative matched volume is written as $W^x_d(t)$, while the final volume is V^x_d. Finally, when computing averages over days, since updates occur at random times, we will use time coarsening by $\delta\tau$ seconds, i.e. compute quantity averages over days within time slices of $\delta\tau$ seconds.

Figure 1.1 illustrates why auctions deserve attention: the relative importance of the closing auction volume has more than doubled in the last 10 years. Note that the relative opening auction volume of French equities is quite small (typically around 1%) and has stayed remarkably constant.

From Collisions in Event Time to Diffusion in Physical Time

It is useful to consider the price as the position of a uni-dimensional random walker and assume that each price change is caused by a collision: if collision i shifts the price p by δp_i, after N collisions the mean square displacement equals

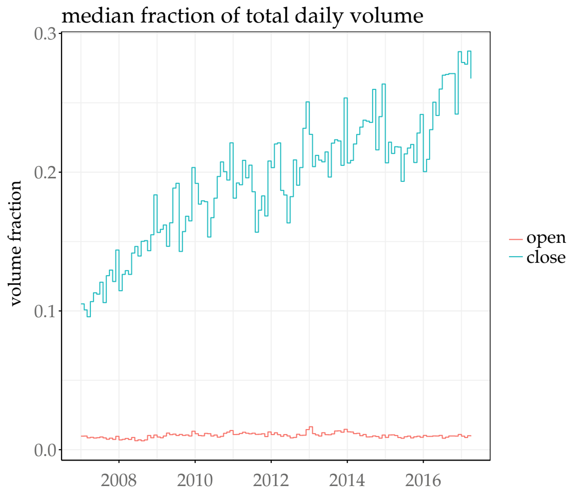

Fig. 1.1 Opening and closing fraction of the total daily volume (median computed over all the tickers) since 2007, showing the global increase of the relative importance of the closing auction, but not of the opening auction. Medians over assets of monthly medians for single assets

$$E\left(\left[\sum_{i=1}^{N}\delta p_i\right]^2\right) \propto N \tag{1.1}$$

if the increments δp_i are i.i.d, a straightforward consequence of the central limit theorem. This corresponds to standard diffusion. In addition, if the collisions occur at a constant rate ρ, then time is homogeneous and $E(t_i) = i\rho$. As we shall see, none of these assumptions is true during auctions, which makes them quite interesting dynamical systems.

Event Rates

In the case of indicative auction prices, the event rate is not constant: the activity usually increases just before the auction time. This finding is a generic feature of auctions with fixed end time [4], and more generally of human procrastinating nature when faced with a deadline, be it conference registration [1] or paying its fee [2].

Let us denote by $N_d^x(t) = \sum_{i,\ 0<t_{i,d}^x\leq t} 1$ the number of price events (changes) having occurred up to time t on day d for auction x. The activity pattern of day d can be measured by the ratio between the number of events up to time t on day d and the total number of events which happened that day, defined as $\nu_d(t) = N_d^x(t)/N_d^x(t^x)$. The average and median $\nu(t) = M(\nu_d)(t)$, where M stands for either average or median over days, can be seen in Fig. 1.2. One similarly defines the fraction between the indicative matched volume at time t and the auction volume $\omega_{d(t)} = M(W_d^x(t)/V_d^x)$, reported in the same figure.

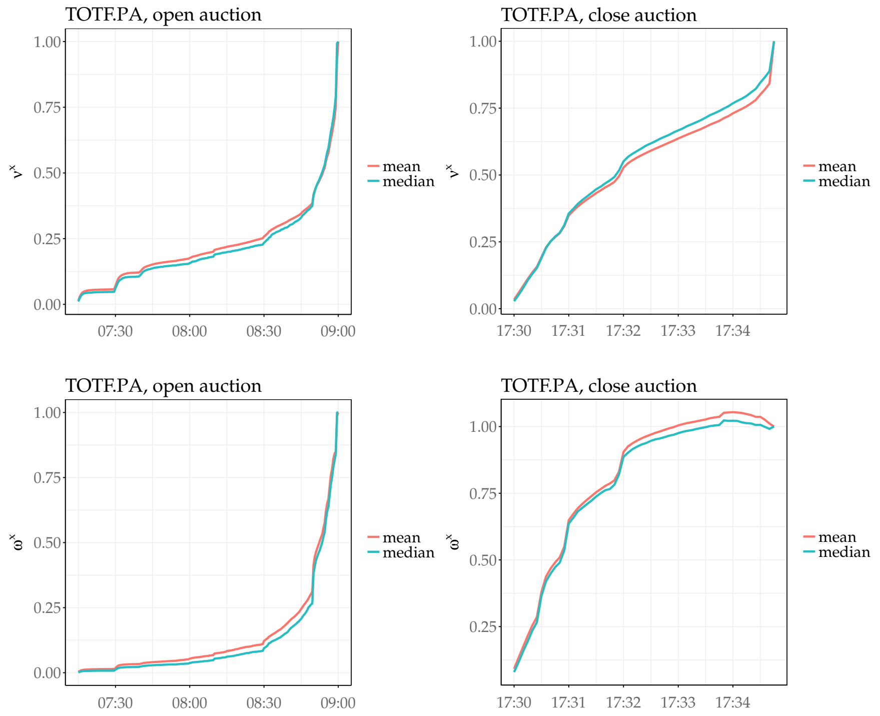

Fig. 1.2 Average activity patterns for opening and closing auctions (left and right plots, respectively) for the most active asset (Total). Upper plots: scaled price change events ν at a function of physical time t. Lower plots: scaled indicative volume fraction ω as a function of t. $\delta\tau = 30$ s for opening auctions and 5 s for closing auctions

There are clear peaks of changes for both ν and ω at unimaginative physical times such as 7:30, 8:30, etc., and at round minutes and multiples of 30 s during the closing auction. This, of course, denotes a regular behavior of some investors. If each peak is systematically caused by a single trader, there are reasons to think that this regularity does inject information and that it will be exploited by more flexible traders. However, sending orders at the same time as other traders is a rational behavior as it allows one to hide in the crowd, unless one's orders are systematically of the same imbalance sign as the aggregate volume at that time. Thus, from a game theoretical point of view, the emergence of activity peaks is self-organized and stable. Nothing constrains the number of peaks and their locations, which are hence instances of emerging, self-organized conventions. The closing auction being much shorter than the opening one, it is natural that the peaks should appear at round minutes, as this somehow provides more obvious peak locations than the opening auction. When the closing auction lasts for a much longer time, e.g. for US equities, there are much fewer price activity peaks [7].

The global pattern of price changes and total volume matched clearly differs between both types of auctions. During opening auctions, the price change rate increases much, starting from a low baseline. During closing auctions, the opposite happens: price change activity is first large, slows down during the first 2–3 minutes and then picks up again just before the cut-off time (17:34:45). The average relative matched volume $\omega(t)$ behaves similarly as $\nu(t)$ during the opening auctions, probably because prices changes are mostly caused by the arrival of new matchable volume, not cancellations. Indeed, half of the open auction events typically happen in the last 10 minutes for most assets, and half of the volume is matched in the last minutes. Closing auctions display a different behavior: more than half of the volume is matched during the first minute, and 80% during the first two minutes. For a few assets (TOTF.PA, UNBP.PA, for ones), there is a peak of indicative matched volume up to 10% larger than the auction volume about one minute before the end of the auction; the same behavior is found in US equity markets [7].

Activity Acceleration

The acceleration pattern of price change rate follows some regularity. To characterize it in a simpler way, it is useful to work in Time-To-Auction τ frame. Since the latter reverts the time arrow, the activity decelerates as a function of τ. Let us denote the average event rate $\rho^x(\tau)$ so that the expected number of event in the period τ to $\tau + \delta\tau$ is $n^x(\tau) = E[N_d^x(\tau + \delta\tau) - N_d^x(\tau)] = \rho^x(\tau)\delta\tau$. Figure 1.3 shows $n^{\text{open}}(\tau)$ of several assets, together with the smoothed version of n^x, denoted by $n^x_{\text{smoothed}} = N^x(\tau)/\tau$: if $n^x \propto \tau^{-\beta}$, so does n^x_{smoothed} but with much less noise, which helps assessing the presence of a power-law visually. We shall drop the x superscript when no confusion is possible.

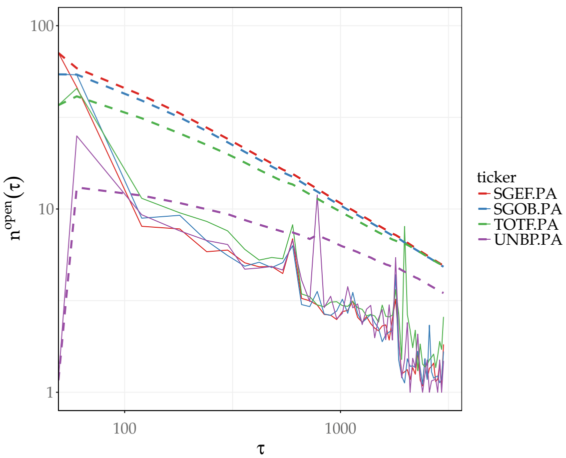

Fig. 1.3 Average number of price changes as a function of the time to auction τ, in seconds, for the opening auction. Dashed lines refer to n_{smoothed}. Time coarsing factor $\delta\tau = 60$s

Assuming that $n(\tau) \propto \tau^{-\beta}$, we perform a robust linear fit of $\log n(\tau) = cst - \beta\tau$ for $\tau \in [100, 300]$ seconds and only keep the fits whose t-statistics associated with β is larger than 5. This particular choice of interval for τ corresponds to a typical period during which the autocorrelation of δp_i at one lag is roughly constant (see section "Diffusive Properties of Indicative Prices"). In addition, for each asset, we only keep days during which there were at least 50 price changes.

If the typical absolute value of price change σ does not depend on τ and is still i.i.d., Eq. (1.1) becomes

$$E\left(\left[\sum_{i=1}^{N} \delta p_i\right]^2\right) = \sum_{i:\, t_i <= \tau} E(\delta p_i)^2 \propto \sigma \tau^{1-\beta} \tag{1.2}$$

hence the Hurst exponent in τ time, denoted by h, equals $(1 - \beta)/2$: the price change rate influences the diffusive pattern in a simple way, given the above approximations. It is worth noting at this juncture that in the normal time frame the price is overdiffusive if σ does not depend on τ and if $\beta > 0$, i.e., if the rate of price changes increases near the auction end time and the Hurst exponent in the normal time arrow is $H = (1 + \beta)/2$.

Typical Price Change

When the indicative price changes, it jumps to the next non-empty tick of the auction order book. Thus, the typical indicative price change reflects the density of the latter, which increases as the auction time nears. As a consequence, the typical price change magnitude σ is not constant but decreases near the auction end time, or equivalently increases as a function of τ. Once again, for opening auctions, we find an approximate power-law relationship $\sigma(\tau) \propto \tau^{\alpha}$ (see Fig. 1.4). We apply the same method as for $n(\tau)$ to estimate α: we only keep days during which there were at least 50 price changes for a given asset; robust fits of $\log \delta p(\tau) = cst + \alpha\tau$ for $\tau \in [100, 300]$ are carried out. Only fits whose t-statistics associated with α are larger than 5 are kept.

Diffusive Properties of Indicative Prices

It is easy to see why the increase of activity and decrease of the typical magnitude of price changes have antagonistic and purely mechanistic effects on the diffusive properties of the indicative auction price in the simplest case: neglecting the auto-correlations and cross-correlations of both $n(\tau)$ and δp_i, Eq. (1.2) becomes indeed

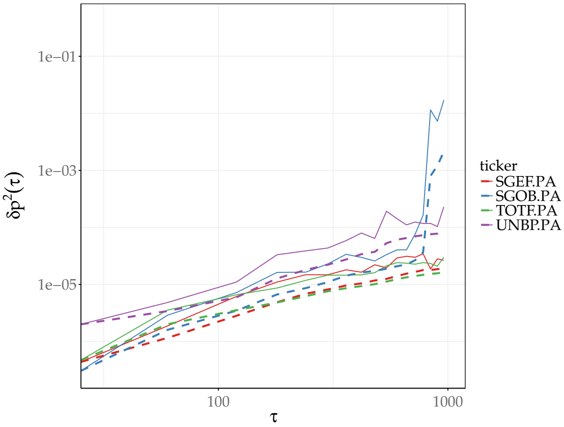

Fig. 1.4 Average scale of the log price increment as a function of the time to auction τ, in seconds, for the opening auction. Dashed lines refer to the smoothed quantity

$$E\left(\Delta p^2\right)(\tau) \simeq \sum_{\tau'<=\tau} E(n(\tau')\delta p^2(\tau')) \propto \tau^{h_0} \tag{1.3}$$

$$\simeq \sum_{\tau'<=\tau} E(n(\tau'))E(\delta p^2(\tau')) \propto \tau^{1+\alpha-\beta} = \tau^{h_0^{(\alpha\beta)}}, \tag{1.4}$$

The first approximation assumes that all δp_i within a time slice are i.i.d, while the second one assumes no correlation between n and δp^2. The relative merits of both approximations can be assessed in Fig. 1.5. The first approximation corresponds to the continuous black line and the second one to the black dots. Both curves are close together; however neglecting the dependence between n and $|\delta p|$ underestimates the typical magnitude of Δp. The same figure makes it clear that something is wrong even in the first approximation, as $\sum_{\tau'<=\tau} E(n(\tau')\delta p^2(\tau'))$ is about 10 times larger than $E(\Delta p^2)$. This discrepancy is mainly due to the bouncing behavior of π for large τ: a large δp_i is typically followed by large δp_{i+1} of opposite sign, which inflates $E(\delta p^2)$ and does not correspond to significant price change as the latter reverts immediately to a value close to that before event i. This is why trimmed means, which removes a given fraction of the largest δp_i for each time slice and each day, decrease much this discrepancy. The latter is also due in part to a simple strategic behavior: during the auction phase, negative indicative price change triggers the sending of buy orders and vice-versa, causing an intrinsically smaller than expected $\Delta p(\tau)^2$ (see below for a more detailed discussion) (Fig. 1.6).

Let us now compare the TTA Hurst exponents of the above quantities, plotted in Fig. 1.7 for the 6 stocks whose fits of both α and β are deemed significant. Two features stand out. First, h_0 overestimates h, even when accounting for the fairly large error bars. This implies that the dynamics caused by the interplay between typical price change shrinking and the acceleration of the activity is more subtle than the

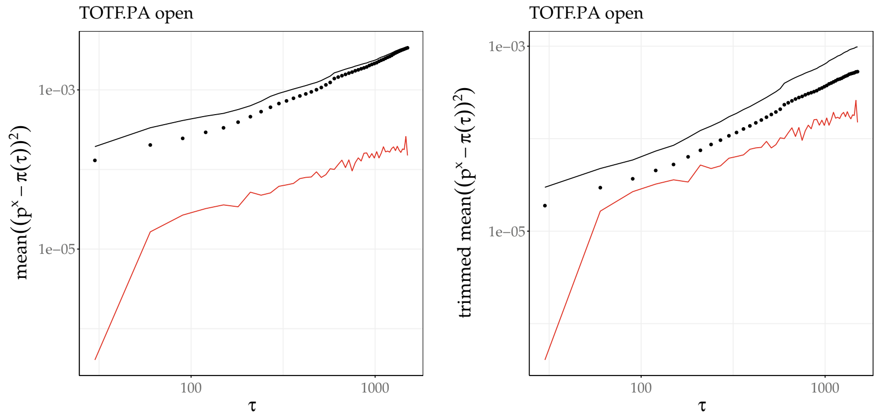

Fig. 1.5 Average square difference between the auction price and the indicative price τ seconds before the opening auction. Continuous red lines (bottom of the figure) refer to $E(\Delta p^2)(\tau)$, The upper black continuous line is $\sum_{\tau'<=\tau}[n(\tau')\delta p^2(\tau')]$, and the black dots are $\sum_{\tau'<=\tau} E[n(\tau')]E[\delta p^2(\tau')]$. Left plot: plain averages over all values of δp_i; right plot: trimmed means where the 20% largest (in absolute value) δp_i for each day and each time slice of $\delta\tau = 30$ s have been removed in the computation of the averages of quantities based on δp_i

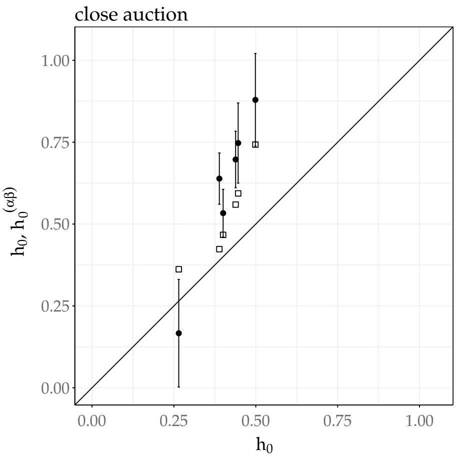

Fig. 1.6 Autocorrelation between two consecutive price changes within time slices of $\delta\tau = 60$ s, averaged over all days for TOTF.PA

simple approximation above. In fact, interestingly, $h_0^{\alpha\beta}$ also overestimates the Hurst h_0 exponent: this emphasizes the fact that the δp_i are not i.i.d.

Indeed, in practice, even linear autocorrelation of both $\rho(\tau)$ and δp_i and the cross-correlation between them are not negligible. Let us focus on the autocorrelation of δp_i, denoted by $C_{\delta p}(\delta i)$. For each time slice $[\tau, \tau + \delta\tau[$, we average $C_{\delta p,d}(1)$ over all the days for a given asset. Figure 1.6 plots this quantity versus τ for TOTF.PA, the most active asset in our dataset. Generally, $C_{\delta p}(1) < 0$; even more, it becomes more and more negative near the auction time, i.e., for small τ. Since the price changes become

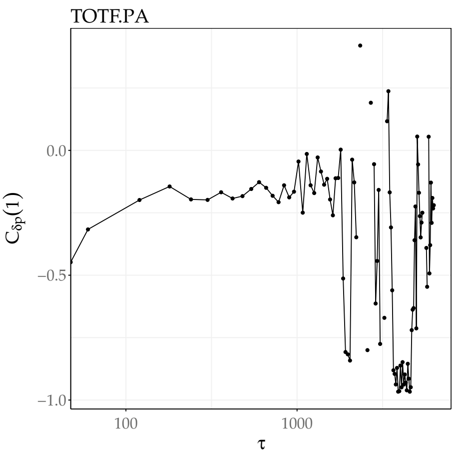

Fig. 1.7 Hurst exponents h_0 and $h_0^{(\alpha\beta)}$ versus the actual Hurst exponent for the 6 assets of the CAC40 that yield good power-law fits of both $\sigma^2 \propto \tau^\alpha$ and $n \propto \tau^{-\beta}$; open auction, $\delta\tau = 60[s]$; Time-To-Auction arrow. Error bars correspond to one standard deviation. When no error bar is visible, the error is at most as large as the symbol

relatively smaller in that limit, this reflects a purposeful bounce of the indicative auction price between two close price ticks; the large negative autocorrelation points to strategic behavior, by which traders try to decrease the immediate impact of their auction orders by submitting their orders after other orders of the opposite sign (hence to hide their actions); in fact, the autocorrelation of the sign of δp, $C_{\text{sign}\,\delta p}(1)$ is even smaller than $C_{\delta p}(1)$ for small τ. For large τ, this auto-correlation also tends to have very small values, which is reinforced by the fact that an outstandingly large δp_i is often followed by a similarly large δp_{i+1} of opposite sign. Thus strategic behavior is more common for small τ.

When $C_{\delta p}(1)$ does not depend on τ, it only modifies the prefactor of τ in Eq. (1.3) by a factor of the order $\frac{1+C_{\delta p}(1)}{1-C_{\delta p}(1)}$, not the Hurst exponent, and thus explains in part the discrepancy between $E(\Delta p^2)(\tau)$ and $\sum_{\tau'=1}^{N} E[n(\tau')\delta p(\tau')^2]$. The dependence of $C_{\delta p}(1) < 0$ on τ modifies the apparent Hurst exponent in a nontrivial way. This is why we measured h for $\tau \in [100, 300]$, i.e., in a region where $C_{\delta p}(1) < 0$ is the most constant.

Discussion

Indicative auction prices display non-trivial properties due in part to the antagonistic effects of both the acceleration of activity and the reduction of the typical price change magnitude. However, the indicative price is much less over-diffusive than what these two effects alone imply. In other words, the deviation from purely mechanistic effect points to a more subtle dynamics. This makes sense, as the traders have a clear incentive to minimize their easily detectable impact. Their strategic behavior results in often alternatively positive and negative indicative price changes, i.e., in a clearly anti-correlated price changes. Quite tellingly, this negative auto-correlation is more and more pronounced as the auction end nears.

So far, we have used a basic data type, which nevertheless has a rich behavior. More detailed data, such as data from the auction book, will allow us to characterize order strategic placement, the evolution of the average auction book density and the price impact of new orders and order cancellations much before the auction time, in the spirit of the response function of [7], but accounting for both the volume of new auction orders and their immediate impact on the auction order book.

References

1. Alfi, V., Parisi, G., Pietronero, L.: Conference registration: how people react to a deadline. Nat. Phys. **3**(11), 746 (2007)
2. Alfi, V., Gabrielli, A., Pietronero, L.: How people react to a deadline: time distribution of conference registrations and fee payments. Open Phys. **7**(3), 483–489 (2009)
3. Bellia, M., Pelizzon, L., Subrahmanyam, M.G., Uno, J., Yuferova, D.: Low-latency trading and price discovery: Evidence from the Tokyo stock exchange in the pre-opening and opening periods (2016)
4. Borle, S., Boatwright, P., Kadane, J.B.: The timing of bid placement and extent of multiple bidding: an empirical investigation using eBay online auctions. Stat. Sci. 194–205 (2006)
5. Bouchaud, J.-P., Farmer, J.D., Lillo, F.: How markets slowly digest changes in supply and demand. In: Hens, T., Schenk-Hoppè, K.R. (eds.) *Handbook of Financial Markets: Dynamics and Evolution*, pp. 57–160. Elsevier (2009)
6. Boussetta, S., Lescourret, L., Moinas, S.: The role of pre-opening mechanisms in fragmented markets (2016)
7. Challet, D., Gourianov, N.: Dynamical regularities of US equities opening and closing auctions. Submitted to Market Microstructure and Liquidity (2018). arXiv:1802.01921
8. Chelley-Steeley, P.L.: Market quality changes in the London stock market. J. Bank. Financ. **32**(10), 2248–2253 (2008)
9. O'Hara, M.: Market Microstructure Theory. Wiley, New York (1997)
10. Pagano, M.S., Peng, L., Schwartz, R.A.: A call auction's impact on price formation and erder routing: Evidence from the NASDAQ stock market. J. Financ. Mark. **16**(2), 331–361 (2013)
11. Pagano, M.S., Schwartz, R.A.: A closing call's impact on market quality at Euronext Paris. J. Financ. Econ. **68**(3), 439–484 (2003)
12. Yergeau, G.: Machine learning and high-frequency algorithms during batch auctions (2018)

Chapter 2
Complex Market Dynamics in the Light of Random Matrix Theory

Hirdesh K. Pharasi, Kiran Sharma, Anirban Chakraborti and Thomas H. Seligman

Abstract We present a brief overview of random matrix theory (RMT) with the objectives of highlighting the computational results and applications in financial markets as complex systems. An oft-encountered problem in computational finance is the choice of an appropriate epoch over which the empirical cross-correlation return matrix is computed. A long epoch would smoothen the fluctuations in the return time series and suffers from non-stationarity, whereas a short epoch results in noisy fluctuations in the return time series and the correlation matrices turn out to be highly singular. An effective method to tackle this issue is the use of the power mapping, where a non-linear distortion is applied to a short epoch correlation matrix. The value of distortion parameter controls the noise-suppression. The distortion also removes the degeneracy of zero eigenvalues. Depending on the correlation structures, interesting properties of the eigenvalue spectra are found. We simulate different correlated Wishart matrices to compare the results with empirical return matrices computed using the S&P 500 (USA) market data for the period 1985–2016. We also briefly review two recent applications of RMT in financial stock markets: (i) Identification of "market states" and long-term precursor to a critical state; (ii) Characterization of catastrophic instabilities (market crashes).

H. K. Pharasi
Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México,
62210 Cuernavaca, México
e-mail: hirdeshpharasi@gmail.com

K. Sharma · A. Chakraborti (✉)
School of Computational and Integrative Sciences, Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: anirban@jnu.ac.in

K. Sharma
e-mail: kiransharma1187@gmail.com

T. H. Seligman
Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México,
62210 Cuernavaca, México
e-mail: seligman@icf.unam.mx

Centro Internacional de Ciencias, 62210 Cuernavaca, México

F. Abergel et al. (eds.), *New Perspectives and Challenges in Econophysics and Sociophysics*, New Economic Windows,
https://doi.org/10.1007/978-3-030-11364-3_2

Introduction

With the advent of the "Big Data" era [9, 13], large data sets have become ubiquitous in numerous fields—image analysis, genomics, epidemiology, engineering, social media, finance, etc., for which we need new statistical and analytical methods [3, 5, 6, 15, 29]. Empirical correlation matrices are of primal importance in big data analyses, since various statistical methods strongly rely on the validity of such matrices in order to isolate meaningful information contained in the "observational" signals or time series [2]. Often the time series are of finite lengths, which can lead to spurious correlations and make it difficult to extract the signal from noise [11, 26]. Hence, it is very important to understand quantitative effects of finite-size time series in determination of empirical correlations [8, 11, 26, 33].

Random matrix theory (RMT) tries to describe statistics of eigenvalues of random matrices, often in the limit of large dimensions. The subject came up first in a celebrated paper of Wishart [39] in 1929 where he proposed that the correlation matrix of white noise time series was an adequate prior for correlation matrices. E. Cartan proposed the classical random matrix ensembles in an important but little known paper [4]. After that there was increasing interest in the subject among which it is important to mention work by L.G. Hua, who published the first monographs on the subject in 1952; an English translation is available [12].

Wigner introduced RMT to physics, based on the assumption that the interactions between the nuclear constituents were so complex that they could be modeled as random fluctuations in the framework of his R-matrix scattering theory [36]. This culminated in the presentation of the Hamiltonian $\hat{H}$ as a large random matrix, such that the energy levels of the nuclear system could be approximated by the eigenvalues of this matrix, and indeed the spacings between the energy levels of nuclei could be modeled by the spacing of eigenvalues of the matrix [37, 38]. The use of RMT has spread over many fields from molecular physics [14] to quantum chromodynamics [28]. Lately, RMT has become a popular tool for investigating the dynamics of financial markets using cross-correlations of empirical return time series [25, 30].

In this chapter, we present recent techniques of random matrix theory (RMT) mainly focused on computational results and applications of correlations in financial markets viewed as complex systems [1, 10, 30, 31]. A central problem that often arises in computational finance is the choice of the epoch size over which the empirical cross-correlation return matrix needs to be computed. A very long epoch would smoothen the fluctuations in return time series and also the time series suffers from the problem of non-stationarity [19], whereas a short-time epoch would result in noisy fluctuations in return time series and the correlation matrix turns out to be highly singular (with many zero eigenvalues) [8]. Among others, an effective method to tackle this issue has been the use of the power mapping [8, 11, 26, 33], where a non-linear distortion is applied to a short epoch correlation matrix. Here, we demonstrate how the value of distortion parameter controls the noise-suppression. It also removes the degeneracy of the zero eigenvalues (which for very small values of the distortion parameter leads to a well separated "emerging spectum" near zero). Depending on the correlation structures, interesting properties of the eigenvalue spectra are found. Correlation matrices constructed from white noise were introduced by Wishart and their eigenvalue spectrum gets a shape of Marčenko-Pastur distribution [16]; there are significant deviations when a correlation structure is introduced [7]. We simulate different correlated Wishart matrices [18, 39] to compare the results with empirical return matrices computed using S&P 500 (USA) market data for the period 1985–2016 [8]. We also briefly review two recent applications of RMT in financial stock markets: (i) Identification of "market states" and long-term precursor to a critical state [23]; (ii) Characterization of catastrophic instabilities (market crashes) [8].

This chapter is described as follows. Section "Data Description, Methodology and Results" discusses the data description, methodology and results in details. Section "Recent Applications of RMT in Financial Markets" contains applications of RMT in financial markets. Finally, section "Concluding Remarks" contains concluding remarks.

Data Description, Methodology and Results

Data Description

We have used the database of Yahoo finance [40], for the time series of adjusted closure prices for S&P 500 (USA) market, for the period 02/01/1985–30/12/2016 ($T = 8068$ days); number of stocks $N = 194$, where we have included the stocks that are present in the index for the entire duration. The sectoral abbreviations are given in Table 2.1.

Methodology and Results

Correlations between different financial assets play fundamental roles in the analyses of portfolio management, risk management, investment strategies, etc. However, one only has finite time series of the assets prices; hence, one cannot calculate the exact

Table 2.1 Abbreviations of ten different sectors for S&P 500 index

Labels	Sectors	Labels	Sectors
CD	Consumer discretionary	ID	Industrials
CS	Consumer staples	IT	Information technology
HC	Health care	MT	Materials
EG	Energy	TC	Technology
FN	Financials	UT	Utilities

correlation among assets, but only an approximation. The quality of the estimation of the true cross-correlation matrix strongly depends on the ratio between the length of the financial price time series T and the number of assets N. The larger the ratio $Q = T/N$, the better the estimation is; though for practical limitations, the ratio can be even smaller than unity. However, such correlation matrices are often too noisy, and thus need to be filtered from noise. To build the correlation matrices, we first calculate the return r_i from the daily price P_i of stocks $i = 1, \ldots, N$, at time t (trading day):

$$r_i(t) = \ln P_i(t) - \ln P_i(t-1), \tag{2.1}$$

where $P_i(t)$ denotes the price of stock i at time t. Since different stocks have varying levels of volatility, we define the equal-time Pearson cross-correlation coefficient as

$$C_{ij}(\tau) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sigma_i \sigma_j}, \tag{2.2}$$

where $\langle \ldots \rangle$ denotes the time average and σ_k denotes the standard deviation of the return time series r_k, $k = 1, \ldots, N$, computed over an epoch of M trading days ending on day τ. The elements C_{ij} are restricted to the domain $-1 \le C_{ij} \le 1$, where $C_{ij} = 1$ corresponds to perfect correlations, $C_{ij} = -1$ to perfect anti-correlations, and $C_{ij} = 0$ to uncorrelated pairs of stocks. The difficulties in analyzing the significance and meaning of the empirical cross-correlation coefficients C_{ij} are due to several reasons, which include the following:

1. Market conditions change with time and the cross-correlations that exist between any pair of stocks may not be stationary if an epoch chosen is too long.
2. Too short epoch, for estimation of cross-correlations, introduces "noise", i.e., fluctuations.

For these reasons, the empirical cross-correlation matrix $\boldsymbol{C}(\tau)$ often contains "random" contributions plus a part that is not a result of randomness [22, 24]. Hence, the eigenvalue statistics of $\boldsymbol{C}(\tau)$ are often compared against those of a large random correlation matrix—a correlation matrix constructed from mutually uncorrelated time series (white noise) known as Wishart matrix.

We first reproduce the basic results of RMT, e.g., the Marčenko-Pastur distribution, or Marčenko-Pastur law, which describes the asymptotic behavior of eigenvalues of square random matrices [16]. Then, we present a study of time evolution of the empirical cross-correlation structures of return matrices for N stocks and the eigenvalues spectra over different time epochs, and try to extract some new properties or information about the financial market [8, 23].

Wishart and Correlated Wishart Ensembles

Let us construct a large random matrix $\boldsymbol{B}$ arising from N random time series each of length T, where the entries of a time series are real independent random variables drawn from a standard Gaussian distribution with zero mean and variance σ^2, such that the resulting matrix $\boldsymbol{B}$ is $N \times T$. Then the Wishart matrix can be constructed as

$$\boldsymbol{W} = \frac{1}{T}\boldsymbol{B}\boldsymbol{B}'. \tag{2.3}$$

In RMT, the ensemble of Wishart matrices is known as the *Wishart orthogonal ensemble*. In the context of a time series, $\boldsymbol{W}$ may be interpreted as the *covariance* matrix, calculated over N stochastic time series, each with T statistically independent variables. This implies that on average, $\boldsymbol{W}$ does not have cross-correlations.

A correlated Wishart matrix can be constructed as

$$\boldsymbol{W} = \frac{1}{T}\boldsymbol{G}\boldsymbol{G}', \tag{2.4}$$

where $\boldsymbol{G} = \boldsymbol{\zeta}^{1/2}\boldsymbol{B}$, is a $N \times T$ matrix; $\boldsymbol{G}'$ is the $T \times N$ transpose matrix of $\boldsymbol{G}$, and the $N \times N$ positive definite symmetric matrix ζ controls the actual correlations. If $\boldsymbol{\zeta}$ is a diagonal matrix with the diagonal entries as unity and off-diagonal entries as zero (i.e., $\zeta = \mathbb{1}$, the identity matrix), then the resulting matrix $\boldsymbol{W}$ reduces to one of the former *Wishart orthogonal ensemble*. If the diagonal entries of $\boldsymbol{\zeta}$ are unity and off-diagonal elements are non-zero and real, then the resulting matrices form the *correlated Wishart orthogonal ensemble*. For simplicity, in this chapter, we have generated and used $\boldsymbol{\zeta}$ for which all the off-diagonal elements are same (equal to a constant U, which lies between zero and unity).

The spectrum of eigenvalues for the Wishart orthogonal ensemble can be calculated analytically. For the limit $N \to \infty$ and $T \to \infty$, with $Q = T/N$ fixed (and bigger than unity), the probability density function of the eigenvalues is given by the Marčenko-Pastur distribution:

$$\bar{\rho}(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \tag{2.5}$$

where σ^2 is the variance of the elements of $\boldsymbol{G}$, while λ_{min} and λ_{max} satisfy the relation:

$$\lambda_{min}^{max} = \sigma^2\left(1 \pm \frac{1}{\sqrt{Q}}\right)^2. \tag{2.6}$$

For $Q \leq 1$, positive semi-definite matrices $\boldsymbol{W}$, the density $\bar{\rho}(\lambda)$ in the above Eq. 2.5 is normalized to Q and not to unity. Therefore, taking into account the $(N - T)$ zeros, we have

$$\bar{\rho}(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} + (1 - Q)\delta(\lambda). \tag{2.7}$$

First, we have generated a Wishart matrix $\boldsymbol{W}$ (with $\zeta = \mathbb{1}$) of size $N \times N$ constructed from N time series of real independent Gaussian variables, each of finite length T, zero mean and unit variance ($\sigma^2 = 1$). Figure 2.1 shows the effect of finite sizes of the sets of parameters N and T on the probability distributions of the elements W_{ij} of the Wishart ensemble and the corresponding eigenvalue spectra. Figure 2.1a shows the probability distribution of the elements of the Wishart matrix of dimensions, where $N = 1024$ and $T = 10240$. Figure 2.1d shows the corresponding density of eigenvalues $\bar{\rho}(\lambda)$, which takes the shape of the theoretical Marčenko-Pastur distribution (red dashed line) [16]. Similarly, Fig. 2.1b, c show the respective probability distributions of the elements of Wishart matrices generated using the sets of parameters $N = 10240$ and $T = 102400$, and $N = 30720$ and $T = 307200$. We can see that with increase in system size (both N and T) the shape of the distribution becomes narrower, implying that the amount of spurious cross-correlations decreases. Ideally, the distribution should be a Dirac-delta at zero, since true cross-correlations do not exist. The eigenvalue spectra are less sensitive to the parameters N and T, as can be seen in Fig. 2.1e, f, which show the corresponding eigenvalue spectra. For all of the above simulations, we find the simulated data agree closely with the theoretical Marčenko-Pastur distributions (red dashed lines) with $\lambda_{max} = 1.732$ and $\lambda_{min} = 0.468$ (theoretically calculated using Eq. 2.6, and $Q = 10$).

As we have mentioned earlier, the assumption of stationarity fails for a very long return time series, so it is often useful to break one long time series of length T into n shorter epochs, each of size M (such that $T/M = n$). The assumption of stationarity then improves for each of the shorter epochs. However, if there are N return time series, such that $N >> M$, then the corresponding cross-correlation matrices are highly singular with $N - M + 1$ zero eigenvalues, which lead to poor eigenvalue statistics. We use the power map technique [11, 34] to break the degeneracy of eigenvalues at zero. In this method, a non-linear distortion is given to each element (W_{ij}) of the Wishart matrix $\boldsymbol{W}$ (or later in each correlation coefficient C_{ij} of the empirical cross-correlation matrix $\boldsymbol{C}$) of short epoch by:

$$W_{ij} \rightarrow (\text{sign } \ W_{ij})|W_{ij}|^{1+\varepsilon}, \tag{2.8}$$

where ε is a noise-suppression parameter. For very small distortions, e.g., $\varepsilon = 0.001$ (as used here), we get an "emerging spectrum" of eigenvalues, arising from the degenerated eigenvalues at zero which is well separated from the original spectrum. The power mapping method suppresses noise present in the correlation structure of short-time series (see e.g., Refs. [8, 17, 21, 23, 32] for recent studies and applications). Later in this chapter, we study different aspects of the power mapping method by varying the value of distortion ε from 0 to 0.8.

In Fig. 2.2, we have studied the effect of non-linear distortion on the behavior of Wishart ensemble ($U = 0$), where $N >> M$. The top row of Fig. 2.2 shows semi-log plots of the ensembles with parameters: (a) $N = 1024$ and $M = 512$, and (b) $N = 1024$ and $M = 64$. Then small non-linear distortions with $\varepsilon = 0.001$ are given to the ensembles to display the emerging spectra, shown in Fig. 2.2c, d. Interestingly, the shape of the emerging spectrum changes from a semi-circle to a strongly distorted one, as M becomes shorter. Also, note that emerging spectrum shifts towards the left