INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2022IPPAG574

Thèse de doctorat

ÉCOLE POLYTECHNIQUE

IP PARIS

Microfounded theories of price formation

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à École Polytechnique

École doctorale n° 574
École Doctoral de l'Institut polytechnique de Paris (EDIPP)
Spécialité de doctorat : Physique

Thèse présentée et soutenue à Palaiseau, le 18 Novembre 2022, par

MICHELE VODRET

Composition du Jury :

Damien Challet CentraleSupélec	Rapporteur
Johannes Muhle-Karbe Imperial College London	Rapporteur
Giulio Bottazzi Scuola Superiore Sant'Anna	Examinateur
Albert Kyle University of Chicago	Examinateur
Sophie Laruelle Université Paris-Est Créteil	Examinatrice
Michael Benzaquen CNRS & Ecole polytechnique	Directeur de thèse
Iacopo Mastromatteo Capital Fund management	Co-encadrant
Bence Tóth Capital Fund Management	Co-encadrant

Acknowledgments

First, I'm grateful to Damien Challet and Johannes Muhle-Karbe for their interest in my research and for agreeing to serve as my referees. I'm honored to have them, as well as Giulio Bottazzi, Albert Kyle and Sophie Laruelle in my jury.

In order of their appearances in my life, I warmly thank my three supervisors, Bence Tóth, Michael Benzaquen, and Iacopo Mastromatteo, who contributed in different and complementary ways to my achievements and my training. I thank Bence for showing me what it means to have a truly impressive insight, Michael for his methodological guidance, and Iacopo for showing me deep formal connections between physics and finance. I was extremely lucky for this perfect combination of insights, support and stimuli that fueled my creativity. I also thank Jean-Philippe Bouchaud, for the inspiring conversations I had the opportunity to have with him.

It is a pleasure for me to mention here Bruno Bassetti, who opened the doors of interdisciplinary science to me, and Andrea Gambassi who was the link that led me to discover the fascinating field to which I have contributed in the last three years.

The EconophysiX laboratory, managed with care as well as method by Michael, has allowed me to establish contacts with many brilliant and interesting people from many different backgrounds and who go in as many directions: Théo, Antoine, Federico, José, Armine, Valerio, Pierre, Mehdi, Samy, Cécilia, Jérome, Karl, Rudy, Léonard, Riccardo, Felipe, Anirudh, Elia, Salma, Natascha, Victor, Gianluca, Max, Swann and Nirbhay, thank you for the stimulating environment you spontaneously created. I thank LadHyX for allowing me to maintain genuinely physics-related discussions, especially with Tullio and Francesco. I am indebted to CFM for hosting me during these three years and for showing me how an effective company must also be an interactive community, as well as for providing me with proprietary datasets for my research. I also thank Gregory, and again Natascha and Cécilia for taking the time to help me in reading this thesis.

I would now like to thank the people who in these three years have been decisive for my serenity, accompanying me on this journey. I thank Edoardo, my neighbor in Paris, and again Gregory, who welcomed me in this wonderful city, in his cozy attic, which I will never forget.

The unconditional support from my family and friends has been invaluable. To all of them, I would like to say that they meant everything for a long time and that they will always be the key to what remains. Sometimes I think that my grandfather, Giorgio Lunghini, would have enjoyed having a dialogue, now that I am more aware of the complexity of one of the cornerstones of his life, which is economics.

Finally, I hug my partner, Valentina, who shows me every day what a better version of myself is and for giving me everything I need to close the gap with it.

To all of you, thank you.

Abstract

Price and volume dynamics in financial markets exhibit empirical regularities, called stylized facts. Statistical models capture the interplay between these stylized facts and are widely used to make quantitative predictions, but they do not explain why prices move in the first place. Microfounded models instead let the price dynamics emerge from the interactions between traders' strategies. The aim of this thesis is to partially bridge the gap between the literature on microfounded and statistical models. In particular, we explore how the predictions of a well-known microfounded model change if we relax some of its unrealistic assumptions. Interestingly, in doing so, we obtain microfoundations for two well-known statistical models, extending their predictive power.

We provide a microfoundation for the Transient Impact model, which is able to characterize the stationary interplay between the dynamics of orders and prices, solving the diffusivity puzzle. The microfoundation is achieved by generalizing the classic Kyle model of price formation to a stationary setting, assuming that the fundamental price is never made public. The stationary Kyle (s-Kyle) model that we propose is compatible with experimentally observed universal price diffusion in the short term, and non-universal mean-reversion on time scales at which correlations of fundamentals vanish. However, the s-Kyle model assumes strongly rational traders, i.e., each rational agent knows every other player's strategies and has unlimited computing power. While the Rational Expectation Hypothesis (REH) is in line with the Efficient Market Hypothesis (EMH), for which the price always reflects newly released fundamental innovations, it leads the s-Kyle model to make wrong predictions; namely, that price volatility is time-independent and smaller than the one related to fundamentals. The REH, therefore, prevents the s-Kyle model from solving the excess volatility puzzle if one does not assume an unrealistically high risk aversion of market actors. In order to improve that, we propose a second modification of the Kyle model, described below.

Following Shiller and the behavioral finance literature, we propose a behavioral Kyle (b-Kyle) setup by relaxing the REH. In doing so, we obtain a microfoundation of the Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) model. To do so, we assume that the market maker does not know the precise level of non-informed trading and of fundamental volatility; moreover, he updates his prior about fundamental volatility based on the realized market prices. The updating procedure is constructed such that future expectations match past outcomes, leading to tâtonnement dynamics reflecting the adaptive learning dynamics of traders' strategies. In this way, not only do we provide a micro-foundation for excess volatility, but also for the intermittent dynamics of price volatility. In fact, in an appropriate limit of the b-Kyle model, the dynamics becomes analytically tractable and we show that excess volatility follows a Kesten process, i.e., a stochastic multiplicative process repelled from zero. Accordingly, we provide a microfoundation of the class of GARCH models. The b-Kyle model is in line with the literature that challenges the EMH; in fact, it assumes that fundamental price volatility is constant, while it predicts intermittent price volatility. The explanation the b-Kyle model provides for price volatility clustering therefore agrees with the empirical finding that a large fraction of price jumps can not be explained

by fundamental innovations, but is instead caused by the self-exciting dynamics created by the interplay between traders' strategies.

We believe that the b-Kyle model can be useful for explaining why prices move, being parsimonious, yet realistic: it can help rationalize many puzzles tackled in the literature, ranging from price diffusivity to excess volatility and volatility clustering. Moreover, it can also interpolate from calm periods with highly fluctuating prices to fragile regimes with extremely probable flash crashes and liquidity crises.

Keywords: Price dynamics, stylized facts, microfoundations, price impact, volatility clustering, fundamental price, rational agents, adaptive agents

Résumé

Les prix sur les marchés financiers présentent une dynamique non triviale dont les régularités peuvent être résumées en un ensemble de faits stylisés. Alors que les modèles statistiques capturent l'interaction entre ces faits stylisés et sont utilisés pour faire des prédictions quantitatives, ils n'expliquent pas pourquoi les prix évoluent en premier lieu. En revanche, les modèles micro-fondés laissent la dynamique des prix émerger des interactions entre les stratégies des agents, fournissant des informations cruciales aux régulateurs et aux décideurs politiques. Cette thèse propose des micro-fondations pour deux modèles statistiques bien connus, étendant leur pouvoir prédictif.

Nous fournissons une explication microscopique au modèle à Propagateur, qui est un modèle statistique capable de caractériser la dynamique stationnaire des ordres et des prix, fournissant une solution au "puzzle de la diffusivité". La micro-fondation est obtenue en généralisant le modèle de Kyle à un cadre stationnaire, dans lequel le prix fondamental n'est jamais public. Le modèle stationnaire de Kyle (s-Kyle) que nous proposons est compatible avec la diffusion universelle des prix observée expérimentalement à court terme ainsi que le retour non universel à la moyenne pour des échelles de temps sur lesquelles les fluctuations des fondamentaux diminuent. Cependant, le modèle s-Kyle suppose des agents fortement rationnels. Alors que l'hypothèse d'attente rationnelle (REH) est conforme à l'hypothèse de marché efficient (EMH), elle conduit le modèle s-Kyle à faire de mauvaises prédictions, à savoir que la volatilité des prix est indépendante du temps et inférieure à celle liée aux fondamentaux. Le REH empêche donc le modèle s-Kyle de résoudre l'énigme de l'excès de volatilité dans la mesure où nous savons que les fluctuations de prix sont supérieures à celles liées aux fondamentaux grâce aux travaux de Shiller.

Suivant Shiller et la littérature sur la finance comportementale, nous proposons une model de Kyle comportementale (b-Kyle) en assouplissant REH. Nous supposons que l'agent qui contrôle le prix ne connaît pas le niveau précis des ordres non informés ni celui de la volatilité des fondamentaux et il met à jour son estimation de la volatilité des fondamentaux en se fondant sur l'historique des prix. La procédure de mise à jour conduit à une dynamique de tâtonnements qui reflète la dynamique d'apprentissage adaptatif des stratégies des agents. Nous fournissons non seulement une micro-fondation à la volatilité excessive, mais aussi à la dynamique intermittente de la volatilité des prix. En fait, dans une limite appropriée du modèle b-Kyle, nous montrons que l'excès de volatilité suit un processus de Kesten, c'est-à-dire un processus multiplicatif stochastique repoussé de zéro. En conséquence, nous fournissons une micro-fondation pour une généralisation des modèles d'hétéroscédasticité conditionnelle auto-régressive généralisée. Le modèle b-Kyle s'inscrit dans la littérature qui évalue la validité l'EMH; en fait, il suppose que la volatilité fondamentale des prix est constante, tout en prédisant une volatilité intermittente des prix. L'explication que le modèle b-Kyle fournit pour le regroupement de la volatilité des prix est donc en accord avec la conclusion empirique selon laquelle une grande partie des sauts de prix ne peut pas être expliquée par les innovations des fondamentaux, mais est plutôt causée par la dynamique auto-excitante créée par l'interaction entre stratégies des agents.

Nous pensons que le modèle b-Kyle peut être utile pour expliquer pourquoi les prix bougent, étant parcimonieux, mais réaliste: il peut aider à rationaliser de nombreuses questions abordées dans la littérature, allant de la diffusivité des prix à la volatilité excessive et au regroupement de la volatilité. De plus, il peut également interpoler des périodes calmes avec des prix très fluctuants à des régimes fragiles avec des crashs et des crises de liquidité extrêmement probables.

Mots clés: Dynamique des prix, faits stylisés, microfondations, impact prix, clustering de volatilité, prix fondamental, agents rationnels, agents adaptatifs

Papers and preprints

This manuscript comprises the research work that the author (Michele Vodret) has conducted during the last three years under the supervision of Iacopo Mastromatteo, Bence Tóth and Michael Benzaquen at the EconophysiX Lab and at the Laboratoire d'Hydrodynamique of the École Polytechnique de Paris (LadHyX), and includes the following works.

- Michele Vodret, Iacopo Mastromatteo, Bence Tóth and Michael Benzaquen.

  'A Stationary Kyle Setup: Microfounding propagator models'.

  In: *Journal of Statistical Mechanics: Theory and Experiment* (2021) 033410.

  DOI: `https://doi.org/10.1088/1742-5468/abe702`

- Michele Vodret, Iacopo Mastromatteo, Bence Tóth and Michael Benzaquen.

  'Do fundamentals shape the price response? A critical assessment of linear impact models'.

  Accepted for publication in: *Quantitative Finance*

  DOI: `https://doi.org/10.48550/arXiv.2112.04245`

- Michele Vodret, Iacopo Mastromatteo, Bence Tóth and Michael Benzaquen.

  'Microfounding GARCH Models and Beyond: A Kyle-inspired Model with Adaptive Agents'.

  Submitted to: *Journal of Economic Interaction and Coordination*

  DOI: `https://doi.org/10.48550/arXiv.2206.06764`

- Felipe Moret, Michele Vodret, Iacopo Mastromatteo, Bence Tóth and Michael Benzaquen.

  'On the origin of long-range correlations in trading activity'.

  in preparation.

Presentations at Conferences; Seminars and Schools

To maximise the potential impact of my research I participated, exposing my own work to two conferences, two seminars, and one summer school.

2022

- Speaker at WEMFIMAP Conference, at Centrale Supélec, Paris.
- Invited speaker at the Institute of Economics of Scuola Superiore Sant'Anna.
- Speaker at FMND2022 Conference, Paris.

2021

- Invited to give a seminar at the Quantitative Finance Research Group of Scuola Normale Superiore.
- Speaker at Behavioral Macro and Complexity Summer School, at Tinbergen Institute, Rotterdam.

Contents

List of Figures

List of Tables

Foreword

Audentes Fortuna iuvat.

Virgil, Aeneid

My understanding of science has been revolutionized by the encounter with statistical physics, a set of ideas and tools that describe how the properties of a large system emerge from the interactions of its constituents. It was immediately clear to me that the domain of applicability of this framework was not limited to the study of phase transitions and critical phenomena related to physical systems, although I focused my exciting undergraduate studies on these. Of course, I was only one of the many who got fascinated by this intuition: Boltzmann, one of the founders of statistical mechanics, wrote [1]

"This opens a broad perspective if we do not only think of mechanical objects. Let's consider applying this method to the statistics of living beings, society, sociology, and so forth."

It is also worth mentioning that the origin of the word statistics stems from the analysis of regularities in social data; one of the leading actors of these investigations was Quetelet who coined the term *social physics* [2]. With time, both physics and economics became more formal and rigid in their specializations, and the social origin of statistical physics was forgotten. The situation is well summarized by Ball [3]:

"Today physicists regard the application of statistical mechanics to social phenomena as a new and risky venture. Few, it seems, recall how the process originated the other way around, in the days when physical science and social science were the twin siblings of mechanistic philosophy and when it was not in the least disreputable to invoke the habits of people to explain the habits of inanimate particles."

Luckily, while working on my master's thesis, I became aware of econophysics; from that moment, the cloudy revolution that occurred a few years back became clearer: econophysics is the attempt to understand economic phenomena from a statistical physicist's point of view.

Of course, empirical data is the starting point for every physicist's attempt to describe reality; econophysics is by no means different from physics in this respect. Contrary to standard physics (but similarly to astrophysics, for example), it is very difficult to perform laboratory experiments in economic-related disciplines. For this reason, to study economic and financial systems one usually resorts to data related to real-world operating systems, such as financial markets, interbank networks, firms, and so on; the advent of computers led to the accumulation of an enormous amount of data which are nowadays intensively studied in academia, public institutions as well as in the private sector. In particular, high-quality financial data is recorded at the event scale every single day, which permits precise analysis

from intraday to yearly timescales. For this reason, the study of financial markets played a prominent role in economic-related research during the past 60 years.

The empirical analysis of financial data conducted over the past decades highlighted a set of statistical properties of price dynamics that are robust with respect to the type of security traded, market structure, geographic location, and time period; these robust statistical properties are named stylized facts. The robustness of these statistical properties begs an explanation. In the language of statistical physics, the stylized facts we observe in financial markets might be a signature of universality, i.e., the fact that details of the given stock, or the given market in which stocks are traded, are not essential to describe them. In physics, universality is displayed, for example, by systems close to their critical point, where they exhibit large fluctuations driven by their consituents' collective motion; the properties of these fluctuations are captured by a few properties, such as symmetries, system's spatial dimensionality and range of interactions. It is therefore extremely interesting, and natural for a statistical physicist, to try to understand if concepts such as universality and phase transition are applicable to financial markets and economic systems in general.

I was extremely lucky because, while I was doing my master's thesis, my former supervisor told me that a researcher in quantitative finance was giving lectures in the same building where I was working. I discovered the EconophysiX Chair, held by Michael Benzaquen, co-founded between Capital Fund Management and the LadHyX laboratory at École Polytechnique. Eventually, that researcher, i.e., Bence Tóth became one of my Ph.D. thesis supervisors.

The question that ignited my Ph.D. was: 'how do agents' interactions cause the coupled dynamics between trades and prices to emerge?', which was in line with my previous studies and research interests. Even though this question has been tackled by the discipline known as market microstructure, which started in the 1980s, no consensus around a given answer has been reached in the literature. Contrarily to what happens in other sciences, in economics it is usually harder to reach a consensus on these important subjects; one reason for this is that, as we recalled above, in economics, it is rare to have reproducible results since it is not possible to control the behavior of the system. Because of this, often ideologies replace rigorous scientific statements when we attempt to improve our economic systems. As a final note, the inertia of mainstream economic ideologies and the "Tarzan complex" of some physicists prevented ideas coming from different disciplines from interacting effectively.

Therefore, to try to provide an answer to the question above, I started from a well-known model in theoretical economics, the so-called Kyle model, instead of formulating a new model inspired by statistical physics. Kyle in 1985 proposed a minimalistic agent-based model to capture the interplay between trades and prices, starting from a description given in terms of rational traders possessing asymmetric information sets, interacting in a noisy environment. Agents' rationality implies that they know the model as well as the model builder, i.e., they construct their strategy in a deductive way, in line with the Rational Expectation (REH) and the Efficient Market Hypothesis (EMH), which were the main paradigms during those years. This very simple model is able to microfound the stylized fact named price impact, i.e., the fact that trades impact prices.

Fifteen years after Kyle's work, the focus in theoretical economics literature shifted towards descriptions of financial markets in terms of adaptive agents, rather than rational ones. Within the Adaptive Market Hypothesis (AMH), agents use inductive reasoning rather than deductive one. This paradigm shift is due to the difficulty encountered by models based on rational agents to capture stylized facts related to price dynamics, such as excess price volatility with respect to fundamentals, as well as price volatility clustering without resorting to ad-hoc, unrealistic assumptions.

My original work consists of two modifications of the Kyle model. The first one is

an attempt to reconcile the Kyle framework with the Transient Impact Model (TIM) (also known as the propagator model), a model used in quantitative finance. This is done without modifying the rationality assumption of traders in the Kyle model but by dropping the assumption that the fundamental price is revealed at some terminal time; in this way, a stationary version of the Kyle model can be mapped on the TIM. The second modification I propose is instead in line with the stream of literature related to the AMH and is able to recover heteroscedasticity, i.e., the fact that volatility of the price process changes over time, and excess volatility as a consequence of the ever-evolving agents' beliefs. In doing so, I show how the Kyle model with adaptive agents is able to provide a microfoundation for a particular Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) model. I also show how one can recover dynamics resembling real markets during liquidity crises and flash crashes.

The content of this manuscript is organized as follows. **Part I** includes a self-contained exposition of the notions useful for understanding my original work. **Chapter 1** is a historical excursus that starts from the first attempt to describe the price dynamics in financial markets by Bachelier in 1900, then goes through the debate surrounding the EMH, and ends with the review of the solution to the debate represented by the AMH. **Chapter 2** reviews several basic stylized facts related to price dynamics; whenever feasible, I recall statistical models able to capture them. In doing so, I review GARCH models. **Chapter 3** describes the interplay between prices and trades and I introduce the concept of price impact, i.e., the empirical fact that the very act of trading induces fluctuations in the price process. To rationalize these findings I review the TIM. The discussion about stylized facts and statistical models is amended with empirical analysis and models' calibrations against real-world data. Finally, the Kyle model, i.e., the starting point of my attempt of modeling the microstructure of financial markets, is introduced.

Parts II and **III** contain original works. The former is related to my contributions to the literature related to microfounded models with rational agents. First, **Chapter 4** presents the s-Kyle model, a modification of the Kyle framework which relaxes some of its original assumptions in order to recover a stationary price process. In doing so, I give a microfoundation to the TIM. Later, in **Chapter 5** I will calibrate the stationary Kyle model on empirical data, providing a piece of original evidence against the EMH. Finally, **Chapter 6** presents the b-Kyle model, which modifies the Kyle framework by relaxing the traders' rationality assumption in favor of a modeling approach in line with the AMH, providing a microfoundation to GARCH models.

The manuscript ends with a critical review of my work, highlighting promising future research directions [1].

[1] The end of each chapter contains a box in which take-home messages are condensed, for the reader's ease.

Part I

The theoretical minimum

Chapter 1

An historical excursus

nos esse quasi nanos gigantium humeris insidentes

quote attributed to Bernard de Chartres by Iohannes Saresberiensis, Metalogicon

Financial markets are (physical or electronic) places where traders meet and transactions take place. On the one hand, they massively contributed to human development, by allowing a better-performing resource-allocation mechanism. On the other hand, our societies are periodically victims of economic crises, sometimes triggered and/or amplified by the financial system. It is therefore important to study how financial markets work, and how they interact with the rest of the economy. This thesis treats the external economy as a given set of time-dependent parameters, and I will not consider the consequences that a given financial crisis can have on an economy. In this Chapter, I hope to convince the reader that financial markets are interesting in and of themselves.

Figure 1.1: The tulip folly, Jean-Léon Gérôme, 1882.

Keywords: Efficient Market, Rational Expectation, Adaptive Market

Contents

As anticipated in the Foreword, market microstructure literature stemmed from the Rational Expectation Hypothesis which was used to rationalize Fama's Efficient Market Hypothesis; in Section 1.1 I introduce these concepts, which ground the building block I use in my original modeling approach in Part II and III, i.e., the Kyle model introduced in Section 3.4.

In doing so, I describe the criticisms around the aforementioned assumptions introducing the so-called excess volatility puzzle, to which I propose a possible solution via the original model presented in Part III. this model is in line with the AMH introduced in Section 1.2, which has been recognized by the economic literature as a promising way to circumvent the drawbacks of the initial framework.

Finally, in Section 1.3, I briefly review how concepts and methods developed by physicists have been useful to this last research direction.

1.1 The Efficient Market Theory and its critiques

The study of financial markets can be traced back to Bachelier's Ph.D. thesis [4] published in 1900 under the title "The theory of speculation". Bachelier, under the supervision of Poincaré, provided a description of financial markets based on the process nowadays known as the random walk, 5 years before Einstein's description of Brownian motion [5]. Although Bachelier's work has been recognized as the foundation of financial mathematics only ~ 50 years later, a basic intuition he had is still highly debated even nowadays. In his opening paragraph, he recognizes that: "past, present and even discounted future events are reflected in market price, but often show no apparent relation to price changes". Bachelier's intuition about the link between price efficiency and price unpredictability was later taken seriously by economists once computers could quickly process large amounts of data.

Eventually, Fama confirmed that stock prices are not very predictable [6, 7]. Therefore, he postulated the so-called Efficient Market Hypothesis (EMH), which states that "the market effectively reflects into the market price all available information". This is a very optimistic statement, but it is not testable, as Fama highlighted in his Nobel lecture [8] when he explained the joint hypothesis problem. In short, one needs an asset pricing model to test the EMH, but, if the model fails to reproduce data, one does not know if either the model or the EMH is wrong. Therefore, in this thesis we adopt the usual approach taken by scientists, i.e., building more refined theories without being blind by preconceptions. Needless to say, Fama's statement is one of the most debated statements in the history of economic thought.

In 1965, the very same year in which Fama published his first empirical results about price unpredictability, Samuelson [9, 10] and Mandelbrot [11] argued that although prices are unpredictable, they do not need to be random walks. In particular, Mandelbrot introduced martingales, i.e., processes for which the best forecast, given past information, is the current value, symbolically expressed as:

$$\mathbb{E}[p_{t+T}|p_t] = p_t. \tag{1.1}$$

This martingale property is very compelling because it corresponds with the intuition that, if prices reflect all the available information and since future news is by definition unforecastable, the best prediction for the future price is the current one. According to the EMH, given the information set $\mathcal{I}_t$ and a given fundamental price p_t^{F} at time t, the price is constructed as follows:

$$p_t^{eff} = \mathbb{E}\left[p_t^{\mathrm{F}}|\mathcal{I}_t\right]. \tag{1.2}$$

Note that, although there are many proxies for the fundamental price, e.g. based on dividends, earnings and announcements, each of these is hard to justify. In modeling, dividends

are usually chosen to construct the fundamental price, which is given by the sum of (discounted) future dividends.

In his well-known review [12], Fama gives a classification of market efficiency, that distinguishes between three situations based on the information set reflected by the price:

> **weak:** $\mathcal{I}_t$ includes all current and past prices for the assets in the market.
>
> **semi-strong:** $\mathcal{I}_t$ includes all public information available to investors (e.g. news).
>
> **strong:** $\mathcal{I}_t$ comprises all information available to investors, including private information.

Fama argues that the EMH is quite well substantiated by financial data, but he also reported some anomalies.

In the 1970s it was somewhat common knowledge that the aggregated market, usually captured by indices, is a powerful tool to aggregate information, in line with the EMH. To rationalize this statement, it has been proposed [13] that the aggregate market behaves as a sort of representative trader that collects all available information and optimally reflects them into the market price. The Rational Expectation Hypothesis (REH) is nothing but the statement that financial markets can be described by representative agents who base their strategies on rational expectations, i.e., knowing the model as well as the model builder.

However, a major theoretical critique of this picture was given by Grossman and Stiglitz in their paper "On the Impossibility of Informationally Efficient Markets" published in 1980 [14]. They argue that perfectly informationally efficient markets are impossible for, if markets are perfectly efficient, there would be no incentive to gather information, in which case there would be little reason to trade and markets would eventually collapse.

Finally, Shiller realized in 1981 [15] (as also did LeRoy and Porter [16], independently), in his paper "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?", that the REH implies a bound on the price volatility, which is violated by empirical data. In fact, from the efficient market model given by Eq. (1.2), it follows that $p_t^{\mathrm{F}} = p_t^{\mathit{eff}} + U_t$, where U_t is a forecast error. The forecast error U_t must be uncorrelated with any information variable available at time t, otherwise, the forecast would not be optimal. Since the variance of the sum of two uncorrelated variables is the sum of their variances, it follows that the variance of p_t^{F} must be greater or equal to that of p_t^{eff}, i.e.:

$$(\sigma^{\mathrm{F}})^2 \geq (\sigma^{\mathit{eff}})^2. \tag{1.3}$$

By looking at the market price and the dividends related to the S&P-500 index (and many others) it has been found that in real markets the inequality given by the equation above is reversed; which means that the price in financial markets fluctuates much more ($\sim$10 times more!) than the underlying fundamental price. French and Roll in 1986 [17] documented a related phenomenon: stock return variances over weekends and exchange holidays are considerably lower than return variances over the same number of days when markets are open. This difference suggests that the very act of trading creates price volatility, which may well be a symptom of Kyle's [18] and Black's [19] noise traders.

Subsequent years saw much focus on the search for better and better models able to assess the validity of the EMH. For example, the initial model proposed by Shiller was later modified in order to account for the time-varying real discount rate or non-stationary dividends. Eventually, all these tests implied that the EMH in the strong form was rejected by empirical data. Nevertheless, because of the non-falsifiability of the EMH, these tests were not able to end the dispute. Eventually, in the 1990s a big part of the academic discussion shifted away from these econometric analyses. The bottom line of the stream of

literature related to testing the EMH is well condensed by Shiller's words that appeared in 2003 in a paper titled "From Efficient Markets Theory to Behavioral Finance" [20]:

> After all the efforts to defend the efficient markets theory, there is still every reason to think that, while markets are not totally crazy, they contain quite substantial noise, so substantial that it dominates the movements in the aggregate market.

1.2 Behavioral finance

Since the EMH in the version given by Eq. (1.2) is not able to account for the excess price volatility exhibited by real financial markets, in the 1990s economists started to look at other possible explanations. In particular, the compelling idea of 'bubbles' was eventually put on the table to explain the financial anomaly which led to the formulation of the excess-volatility puzzle.

One of the first documented bubbles dates back to the famous tulipmania in Holland in the 1630s, to which the picture at the beginning of this chapter alludes. In that picture, soldiers trample the tulip fields to limit their supply. In fact, after an exceptional increase of the tulip price in the period ranging from November 1636 to February 1637[1], there was a sudden collapse. The soldiers in the picture, by limiting the tulip supply, should have limited the drop in the price. This was not the case, and eventually many of the speculators who invested in tulips remained only with 5% of their initial investment.

To understand the causes of tulipmania, it is interesting to read a fictional conversation between two people, Gaergoedt and Waermondt, which was published by one anonymous observer in 1637 [21]:

> Gaergoedt: "You can hardly make a return of 10% with the money that you invest in your occupation [as a weaver], but with the tulip trade, you can make returns of 10%, 100%, yes, even 1000%.
>
> Waermondt: " ... But tell me, should I believe you?"
>
> Gaergoedt: "I will tell you again, what I just said."
>
> Waermondt: "But I fear that, since I would only start now, it's too late, because now the tulips are very expensive, and I fear that I'll be hit with the spit rod, before tasting the roast."
>
> Gaergoedt: "It's never too late to make a profit, you make money while sleeping. I've been away from home for four or five days, and I came home just last night, but now I know that the tulips I have increased in value by three or four thousand guilders; where do you have profits like that from other goods?"
>
> Waermondt: "I am perplexed when I hear you talking like that, I don't know what to do; has anybody become rich with this trade?"
>
> Gaergoedt: "What kind of question is this? Look at all the gardeners that used to wear white-gray outfits, and now they're wearing new clothes. Many weavers, that used to wear patched-up clothes, that they had a hard time putting on, now wear the glitteriest clothes. Yes, many who trade in tulips are riding a horse, have a carriage or a wagon, and during winter, an ice carriage, ...

[1] at its peak a bunch of tulips was worth a house.

Figure 1.2: Feedbacks and herding effects, a cartoon by K. Kallaugher, published on the Baltimore Sun.

An insight provided by these word-of-mouth communications is that feedback dynamics can be quite important in order to explain price booms and bursts (as also Fig. 1.2 tries to illustrate with a funny cartoon). As a consequence of a given price increase, traders might be tempted to believe that the price will continue to increase. In order to profit from this expected future price increase, more people tend to buy the asset at hand. Eventually, someone will recognize that prices exhibited an unrealistic increase in value, and they will start selling, causing the beginning of an abrupt burst. In short, price dynamics experience feedback mechanisms, which are nowadays considered key [22, 23] to understanding the excess price volatility that we talked about in the previous Section, and more generally the complex price dynamics which we shall investigate more closely in Chapter 2.

The importance of feedback mechanisms can be further substantiated by Cognitive Psychological research, which shows that humans often rely on representative heuristics while interacting with a complex environment, showing systemic biases. This stream of literature has been recognized with the Nobel Prize in economics to Kahneman and Tversky in 2002 and can be traced back to their seminal work that appeared in 1974 [24]. A consequence of their empirical findings is that people may tend to match stock price patterns in two broad categories, namely upward and downwards trends, thus leading to feedback dynamics, even if these categories may be rarely seen in the related fundamental price dynamics. Moreover, the findings of Kahneman and Tversky can be seen as a reason why price trajectories across centuries and across countries show similar empirical regularities; because they reflect universal parameters of human behavior.

One of the first critiques of the feedback mechanism which can be at the heart of large price swings is that, naively, the trends created by the feedback itself, strongly correlate with price changes over time. This is at first irreconcilable with Fama's empirical finding that price changes are not predictable. A simple model for feedback dynamics, proposed by Shiller in 1990 [25], shows that this critique does not hold; consider a model where the price p_t is determined by its own lagged changes and a forcing variable, which we refer to as the fundamental price p_t^{F} introduced in the previous Section. This model assumes the following price dynamics:

$$p_t = p_t^{\mathrm{F}} + c \int_{-\infty}^{t} e^{-k(t-\tau)} dp_\tau, \tag{1.4}$$

where $0 < c < 1$ and $k > 0$. The first condition implies that the price does not explode, while $k > 0$, meaning that it is given less weight to the more distant past, reflecting gradual memory loss. The equation above can be solved in order to give p_t in terms of p_t^{F}:

$$p_t = \frac{1}{1-c} p_t^{\mathrm{F}} - \frac{c}{1-c} \bar{p}_t^{\mathrm{F}}, \tag{1.5}$$

where

$$\bar{p}_t^{\mathrm{F}} = \lambda \int_{-\infty}^{t} e^{-\lambda(t-\tau)} p_\tau^{\mathrm{F}} d\tau, \quad \lambda = \frac{k}{1-c}. \tag{1.6}$$

This model implies that the price p_t is cointegrated with p_t^{F}, but amplifies its departure from a weighted average of its own lagged values. The price p_t will be higher than p_t^{F} if p_t^{F} has been increasing in recent years and lower if decreasing. Over short intervals of time, where $\bar{p}_t^{\mathrm{F}}$ is virtually constant, the price p_t has essentially the same short-term stochastic properties of p_t^{F} but amplified. The critique mentioned above regarding the serial correlation of price changes can thus be evaded by assuming that p_t^{F} is still approximatively a random walk. Thus, the approximate random walk character of stock prices is not a piece of evidence against feedback mechanisms.

Moreover, even if the feedback mechanism did imply some predictability in price changes, we can also note that the random walk character of stock prices is really not supported by

evidence anyway. In fact, there is a tendency for stock prices to continue in the same direction over intervals of six months to a year, but to revert over longer time intervals [26–30]. A similar pattern is therefore consistent with some combination of feedback effects together with a dependence on fundamental prices.

The methodological differences between mainstream and behavioral economics suggest that an alternative framework may be necessary to reconcile the EMH with its behavioral critics. One direction is to view financial markets from a biological perspective and, specifically, within an evolutionary framework in which markets, instruments, institutions, and investors interact and evolve dynamically according to the 'law' of economic selection. Under this view, financial agents compete and adapt, but they do not necessarily do so in sisan optimal fashion. The next section reviews some of the contributions that physicists provided to the research agenda related to adaptive markets.

1.3 Econophysics

This section collects some material useful to understand how physics benefitted other research areas, including economics and finance[2].

The set of technical tools and ideas in physics that bridges the micro-description to the macro behavior is called statistical mechanics; in this respect, it is similar to the way in which we study market microstructure in this thesis.

One of the first achievements of the field is the derivation of the ideal gas law, which describes the behavior of many different real gases, starting from a description given in terms of Newtonian dynamics. Real gases, however, do not follow the ideal gas law for every external parameter, i.e., pressure, temperature, or volume. In fact, as we all know, the gas phase is only one *phase* of matter: if we reduce the temperature of a system in the gas phase, eventually the molecules reorganize themselves into liquid or solid phases. This compelling empirical finding inspired many further achievements. In particular, it suggests that something crucial is missing in those economic models which predict only one possible time-independent equilibrium. In fact, economic systems experience crises, which can be thought of as abrupt changes between two equilibrium states [31, 32].

The ideal gas law is derived by assuming non-interacting molecules; this is a very good approximation for the gas phase but is obviously not valid in the liquid or in the solid phase, where the higher density implies crucial *interactions*. Analogously, economic models with only one static equilibrium are usually the outcome of dynamics originated by (non-interacting) rational representative agents. In the same way, as interactions between molecules are needed to build models accounting for different phases, we need to move from a rational agent paradigm to one grounded on adaptive interactive agents if we want to model economic dynamics displaying interesting phenomena, such as economic or financial crises [33, 34].

Phase transitions at the critical point, i.e., the point in the external parameter space where different phases co-exist, are characterized by large fluctuations that extend over the whole system size. Statistical mechanics studies these regimes starting from stylized micro descriptions which have been able not only to give qualitative but also quantitative accurate predictions regarding the properties of the fluctuation, for example, the so-called critical exponents. A major finding is that many different physical systems fall in so-called *universality classes*, i.e., classes that display the same set of universal quantities. The final message we think can be very useful in other disciplines, is that in a situation where the

[2]Although this section is useful to understand how a physicist thinks, it is not necessary to understand the content of my original work presented in Part II and III.

dynamics drives the system in a configuration where collective motions occur, there is no need to model every single detail of the model to reproduce important behavior, but only the relevant ones. For example, to model crises in financial markets, a modeling framework that allows herding can be very effective [34].

How can simple stylized models have such high predictive power? When the relevant motion happens on large spatial scales, as is the case for systems at their critical point, the micro description can be simplified by 'looking at the system at a distance '; since critical systems are usually scale-invariant, coarse-graining can be repeatedly applied, revealing some symmetries which were not encoded in the description of the system at the molecular level [35]. For instance, consider a crystal composed of atoms sitting on an infinite square lattice in a two-dimensional space with nearest-neighbor interactions. When we coarse-grain the system, one should take into account the interactions between the objects sitting at the nodes of the lattice. These interactions lead to new 'effective' terms in the energy function of the rescaled systems. As one applies the coarse-graining procedure over and over, old *symmetries* lead to new emergent ones; for example, translation and reflection symmetries of the square lattice combine into a rotational symmetry, leading to a unified description for many different physical systems, providing the key to the observed universality. In fact, Wilson and Fisher [36] proved that universality classes exist and that different systems lie in the same universality class if they share the same symmetries, spatial dimensionality, and range of interactions. We believe that the complex dynamics exhibited by prices in financial markets can be viewed as a different declination of the universality exhibited by physical systems: contrary to physics, the building blocks of economic systems are traders and, therefore, one should start from microfoundations that take into account ubiquitous biases which affect human beings and their heterogeneities, in line with the findings of Kahnemann and Tversky.

Further major progress obtained in physics, and in particular in statistical mechanics, is related to the work of Parisi, Nobel winner in 2022. Parisi extended the analysis of the universal signature of physical systems beyond the realm of those systems composed of identical parts and paved the way for applications to systems composed of interacting *heterogeneous* objects/agents [37]. This allowed, for example, to completely characterize the dynamics prescribed by the Minority Game, an archetypal model of financial markets, conceived by Challet and Zhang in 1997 [38].

1.3.1 The Minority game

The Minority game stems from the El Farol problem. The formulation of the problem came to Arthur [39], an economist from the Santa Fe Institute, who was inspired by a real-life problem: the El Farol bar has a limited number of seats, so the best thing for attendees is to go if there are empty seats. If you go to the bar while it is crowded you should have stayed home; on the other hand, if you stay at home while the bar is not crowded, you miss a nice evening. When Arthur proposed his problem to an audience in 1994, economists skipped the problem by saying that it could have been reformulated as a standard game-theoretic problem. The problem was instead simplified leading to the Minority game, which was eventually analytically solved by physicists [40], with the support of researchers from different backgrounds. The Minority Game provided a framework to understand how traders co-adapt and the relationship between information and market efficiency.

Let us sketch briefly the Minority game and the major outcomes. Consider a set of N agents who, at each finite time step, can choose between two possibilities, say ± 1. An agent wins if, after everyone has chosen, he is in the minority. In order to always have a minority, we suppose that N is odd. Each agent takes action based on the string μ of the

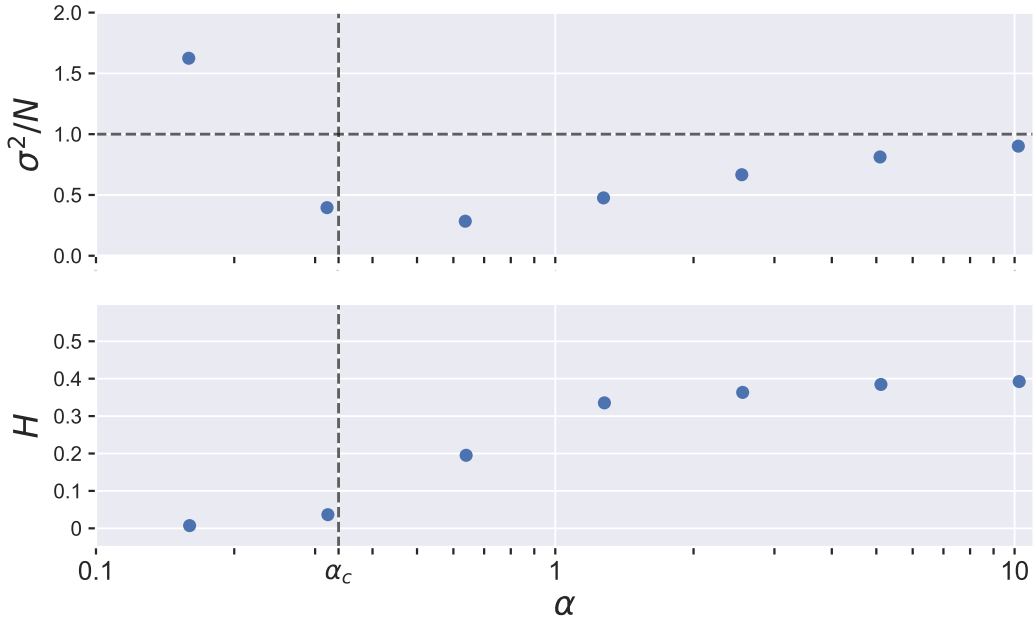


Figure 1.3: Variance per agent (top) and predictability H (bottom) as a function of the control parameter α.

last M outcomes. The number strings μ is 2^M. The number of possible strategies s, i.e., the maps that associate actions (± 1) to past histories (μ), is 2^{2^M}. Each agent is endowed with a fixed bag of strategies, which he ranks, as time goes by as if they were actually played; the strategy is rewarded if it predicts the outcome of the game. In the original Minority Game, the strategy s_t^i which agent i uses at time t is that with the highest score. Once every agent has fixed his best strategy, the attendance of the two sides is computed: if N_1 is the number of agents who took the choice 1, the difference in attendance of the two sides, a central quantity of interest, reads:

$$A = 2N_1 - N. \tag{1.7}$$

Symmetry arguments suggest that none of the two groups -1 or 1 will be systematically the minority one. This means that A_t will fluctuate around zero and $\langle A \rangle = 0$. The size of the fluctuations of A_t, instead, displays a remarkable non-trivial behavior. The variance

$$\sigma^2 = \langle A^2 \rangle \tag{1.8}$$

of A_t in the stationary state is a measure of how effective the system is at distributing resources. The smaller σ^2 is, the larger a typical minority group is. In other words, σ^2 is a reciprocal measure of the global efficiency of the system.

It turns out that the only relevant variables affecting the dynamics are the total number of strings μ and the number of agents, given respectively by 2^M and N. In particular, the only relevant parameter is the ratio given by $\alpha = 2^M/N$. When $\alpha \gg 1$ information is too complex and agents behave randomly. Indeed σ^2/N converges to one, i.e., the value it would take if agents were choosing their side by coin tossing. As α decreases, which means that M decreases or the number of agents N increases, σ^2/N decreases suggesting that agents manage to exploit the information in order to coordinate. But when agents become too numerous, σ^2/N starts increasing with N. The behavior for $\alpha \ll 1$ has been attributed to the occurrence of crowd effects. These findings are shown in Fig. 1.3

A further interesting observation is that depending on the value of α the model displays two different phases. In the asymmetric phase, $\mathbb{E}[A|\mu] \neq 0$ for at least one μ. Hence knowing the history $\mu(t)$ at time t, makes the sign of A_t statistically predictable. A measure of the degree of predictability is given by the function

$$H = \frac{1}{2^M} \sum_{\mu=1}^{2^M} \mathbb{E}[A|\mu]^2. \tag{1.9}$$

In the symmetric phase $\mathbb{E}[A|\mu] = 0$ for all μ and hence $H = 0$. If $\alpha > \alpha_c$ the game is predictable, meaning that conditionally on the given history, the winning choice is biased towards $+1$ or -1. The degree of predictability goes to zero as $\alpha \to \alpha_c^+$.

The overall picture offered by the minority game to interpret real financial market dynamics is that there is a regime in which markets are predictable and attract a larger amount of participants. In doing so, eventually, the market becomes crowded and efficient, leading to unpredictable outcomes. In short, the market *self-organizes* towards the critical efficient phase, which displays long-ranged fluctuations; for example, the time during which an agent keeps playing a given strategy is a power-law variable with an exponent $\mu = 1/2$, of infinite mean. This opens the door for a mechanism that generates a non-trivial temporal correlation in the total volume of activity, such as those exhibited by financial markets (we will come back to this point in Chapter 3).

Take home messages from Chapter 1

1. The Efficient Market Hypothesis (EMH), in its strong version, states that all information is reflected in the market price. Therefore, since by definition news is unexpected, it provides a theoretical foundation for price unpredictability observed in empirical data.

2. A major critique of the EMH is provided by volatility bounds. In particular, ever more refined tests show that the price p fluctuates much more than any proxy related to the fundamental price p^{F}, i.e., the price volatility σ is larger than that related to fundamentals σ^{F}, contrary to what the Rational Expectation Hypothesis implies.

3. The Adaptive Market Hypothesis (AMH) deviates from the rationality assumption, by assuming that agents may not have an omni-comprehensive information set. Moreover, even if this is the case, they may not exploit this information in a rational way, since they may rely on heuristics and biases.

4. Quite intriguingly, the empirical regularities exhibited by financial markets, the so-called stylized facts, are robust with respect to centuries, countries, and securities. In light of the findings of Kahnemann and Tversky, stylized facts might be a signature of the fact that humans are victims of the same biases; these biases may be the source of feedback mechanisms, which may amplify and induce large fluctuations in the price process, explaining anomalies such as the excess price volatility with respect to fundamentals.

5. Statistical Physics provided us with unifying concepts such as phase transitions, critical states, self-organized criticality, and universality, which turned out to be quite relevant for studying economic and financial systems, being valuable concepts to explain the ubiquitousness of stylized facts found in real-world data.

6. This thesis is aimed at investigating the relations between the assumptions related to traders' behavior and the predictive power of emerging price dynamics. In doing so, we will construct a bridge from the classic framework used in market microstructure, which usually relies on the assumption of agents' rationality, to the more flexible and predictive one used in the AMH literature. The resulting model stems from the classic Kyle framework (which will be recalled in Chapter 3), where rational traders are replaced by adaptive ones, and retrieves feedback effects that are key for reproducing several stylized facts.

Chapter 2

Price dynamics

If you are going to use probability to model a financial market, then you had better use the right kind of probability. Real markets are wild.

Benoît B. Mandelbrot, The (Mis)Behavior of Markets

Although the debate over whether financial markets are efficient or not is still partly ongoing, the knowledge related to price dynamics is increasing. To explain how it is possible to decouple the properties of the price from those related to the fundamental value, we shall invoke a basic argument that stems from Bachelier's [4] and Black's [19] analysis. The former argued that price fluctuations grow as the square root of time, while the latter argued that the price is always close to the fundamental price by a factor of 2. If this is the case, the anchor to fundamental values can only be felt on a time scale τ such that purely random fluctuations $\sigma\sqrt{\tau}$ reach the order of 50% the fundamental price, leading to $\tau = 6$ years for the stock market with a typical annual volatility of $\sigma = 20\%$; such long timescales suggest that the notion of fundamental price is secondary to understanding price dynamics at the scale of a few seconds to a few months.

Keywords: Stylized facts, Multiplicative dynamics, Heteroscedasticity, GARCH

Contents

This Chapter reviews salient properties related to price dynamics. First, I introduce in Section 2.1 the Limit Order Book, which is the standard way in which financial markets operate nowadays.

In Section 2.2 I present a set of daily empirical data related to the S&P-500 index, freely available from Yahoo! Finance, that I use in the following to amend the theoretical discussion with empirical analysis.

In Section 2.3, price diffusivity at high frequency is discussed. Note that the diffusive dynamics of prices at high frequency is a crucial piece of the so-called diffusivity puzzle, which is a reason why the Transient Impact model, presented in Section 3.3 and microfounded in Part II, has been proposed.

The remaining part of the Chapter deals with heteroscedasticity, i.e., the fact that price volatility fluctuates over time; this stylized fact, which is captured by the family of statistical models named Generalized Auto-Regressive Processes presented in Section 2.7, will be rationalized by means of an original microfounded model in Part III.

2.1 Limit order book

My original work does not focus on order book dynamics, but to understand some of the approximations we made and some arguments related to trading strategies, it is necessary to grasp the basics of the most common mechanism which rules trading nowadays.

Order books offer the possibility for an agent to trade passively or aggressively. With the first option, an agent can post a buy or a sell order, which consists of a number of shares at a given price, and wait for a counterpart. This passive order is called a limit order. Agents can cancel their passive orders if no counterpart has yet been found. The second option, called a market order, consists of matching existing limit orders. All these limit orders are stored in the limit order book. A schematic describing the interaction between supply and demand in financial markets, essential for understanding the price formation process, is displayed in Fig. 2.1. In the following, we refer to the bid (ask) side for the buy (sell) side. An important notion emerges from the order book, i.e., liquidity. It refers to the number of shares available in the order book at a given time. Thus, posting a limit order in the limit order book provides liquidity while posting a market order takes liquidity.

Let us focus on the quantities and concepts introduced in Fig. 2.1, which are the key elements for limit order book functioning. The prices on which market participants can place their orders are fixed: the quotation step is called the tick size and it varies across assets. For example, it is 0.01\$ for US stocks on NASDAQ and 1€ for the EUROSTOXX future contract. We call the best bid b_t (ask a_t) the best price on the bid (ask) side at time t. The bid-ask spread s_t is defined as the difference between the two:

$$s_t = a_t - b_t. \tag{2.1}$$

We can classify assets by looking at their bid-ask spread:

- **Large tick** stocks have an average bid-ask spread almost equal to one tick.
- **Small tick** stocks have an average bid-ask spread equal to a few ticks.

The properties of the order book are different for these types of stocks. While large tick stocks order books are full, small tick stock order books look sparse, i.e., some of the price levels are empty.

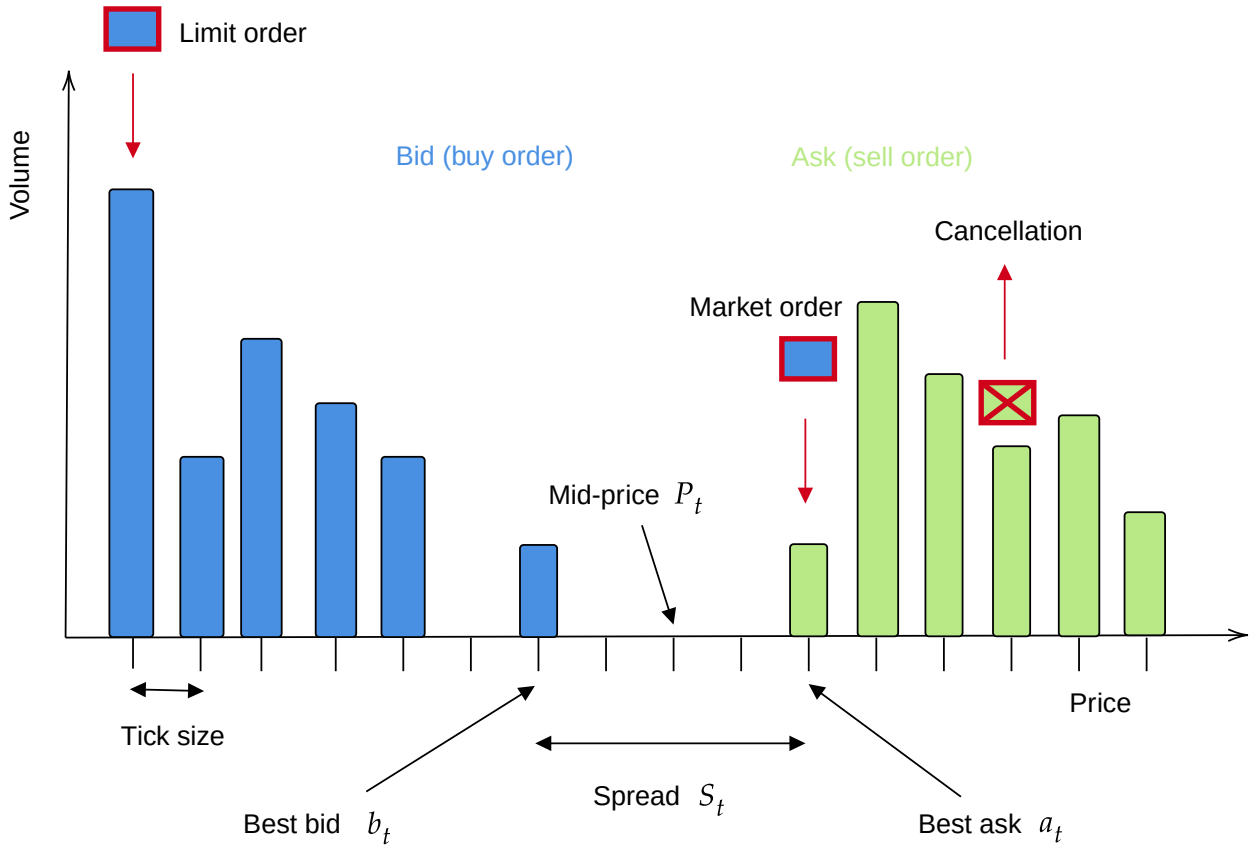


Figure 2.1: Sketch of the limit order book.

The 'price' of the asset is formed within the order book. The most common price is the mid-price:

$$p_t = \frac{a_t + b_t}{2}. \tag{2.2}$$

Let us note that the mid-price changes when the best ask or the best bid changes. If the number of shares v_t^b at the best bid and best ask v_t^a are bigger than the average size of market orders, the time scale of price changes will be far bigger than the one between two order book events. If we need a price that evolves on a faster time scale, we can introduce a price that accounts for supply/demand interactions, such as $p_t = \left(v_t^a b_t + v_t^b a_t\right) / \left(v_t^b + v_t^a\right)$. Whatever definition of the price we choose, it is the basic information on the asset coming from the order book, that is available to all market participants. Its formation takes place through the order book, as a dynamic interaction of supply and demand.

2.2 Seventy years of S&P-500 data

The S&P-500 data analyzed in this Chapter are shown in Fig. 2.2. The top panel shows raw mid-price data; where one can see that the price exhibits an overall upward trend, sometimes interrupted by large downward swings. The middle panel shows price changes δp_t, defined as:

$$\delta p_t = p_t - p_{t-1}. \tag{2.3}$$

These fluctuations are clearly increasing with time, suggesting that multiplicative dynamics are well-suited to describe them. In fact, the trend can be completely removed if one

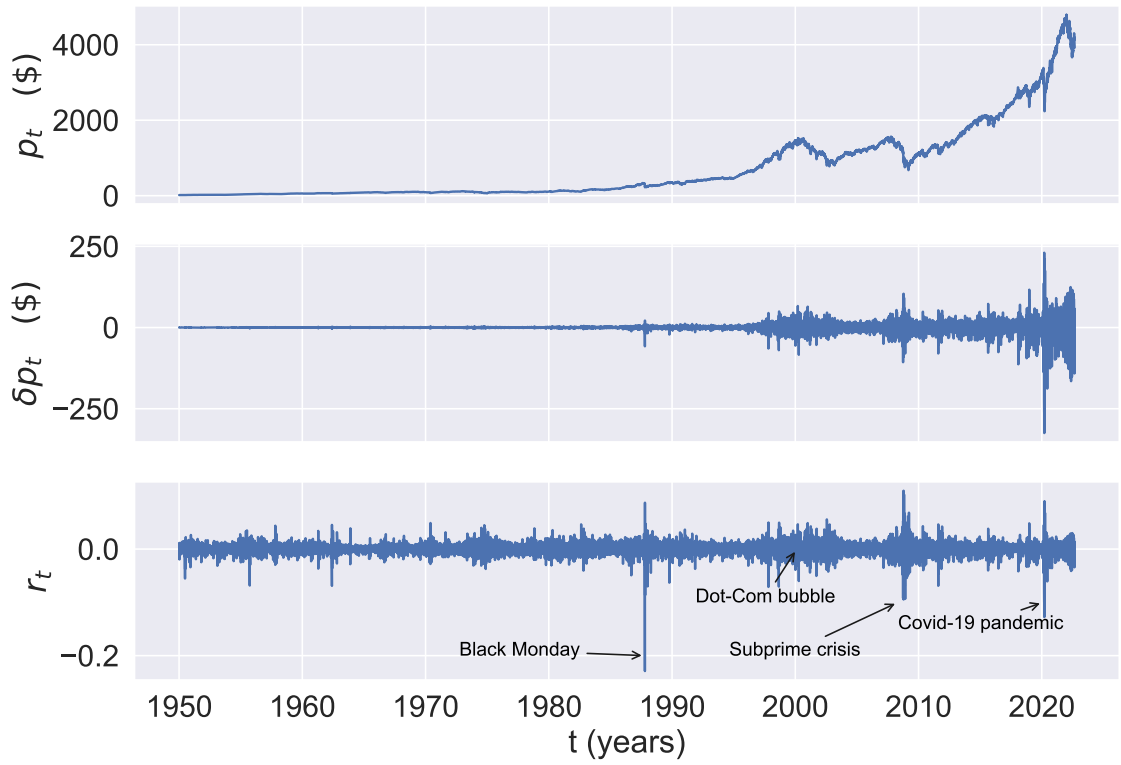


Figure 2.2: (Top) Price, (middle) price changes, and (bottom) returns for S&P-500 from 1970 to 2022.

considers returns, i.e., normalized price changes:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}. \tag{2.4}$$

Returns are shown in the bottom panel of Fig. 2.2, from which one can extract interesting pieces of information: from an aggregate point of view, one can see that fluctuations of returns are clustered in time and that large swings occur quite frequently; for instance, the mean daily volatility is of the order of 1%, while on the Black Monday of 1987, the registered daily return was −20% (!).

From the bottom panel, one can detect major financial crashes that happened in the past decades:

- Black Monday, 1987.
- Dot-Com bubble, 2000.
- Subprime crisis, 2008.
- Flash Crash, 2010.
- Covid-19 pandemic, 2020.
- Ukraine war, 2022.

Although the origins of these large price swings are very different, the feedback effects which amplified them are thought to be universal, as explained in Sec. 1.2.

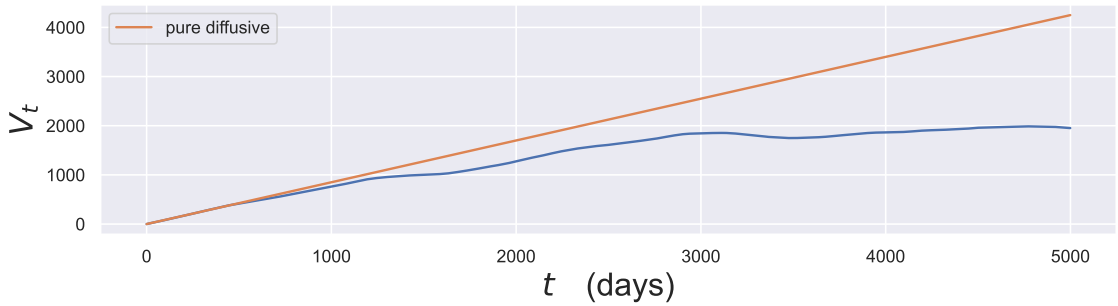


Figure 2.3: Variogram related to S&P-500 data.

2.3 Price correlation over multiple timescales

Bachelier [4] proposed modeling the price p_t in financial markets via what is known nowadays as the Random walk or the Brownian motion process, respectively, if in discrete or continuous time. The First Law of Bachelier states that the variogram V_τ, i.e., the average over squared price changes, grows linearly with time:

$$V_\tau = \langle (p_{t+\tau} - p_t)^2 \rangle = \sigma^2 \tau. \tag{2.5}$$

Bachelier modeled the price process as an additive process, given by:

$$p_t = p_{t-1} + \delta p_t, \tag{2.6}$$

where the price change δp_t is time-stationary with zero mean and variance σ^2. In order to compute Eq. (2.5), one needs to know the Auto-Covariance Function (ACF) of price changes $C_\tau^{\delta p} = \langle \delta p_t \delta p_{t+\tau} \rangle$. If $C_\tau^{\delta p} = \sigma^2 \delta_\tau$, then Eq. (2.6) describes the random walk model. If the price change ACF is positive, the model is a random walk with a positive trend, whereas if the ACF is negative the price process is a mean-reverting random walk.

A very important remark is that the volatility profile is itself dependent on the sampling scale used. For instance, if the price is sampled at the tick-by-tick, one can observe mean reversion at very high frequencies, due to the so-called bid-ask bounce[1]. This thesis will consider sampling scales at which these microstructural effects are unimportant.

The variogram profile grows linearly up to a time interval equal to a few months[2]. For longer timescales, one observes a mean-reverting behavior, i.e., a sub-linear increase, as one can see from Fig. 2.3; as argued by Black [19], price is always within a factor 2 from the price, implying mean-reversion over yearly timescales.

[1]Inside every market place there is a bid-ask spread. The bid is the top price that someone is willing to pay. The ask is the lowest price someone is willing to receive in exchange for a trade. In a stable market structure, there is usually a difference between the two. When someone decides that they need to buy a stock, and are willing to pay the price of the ask, they will 'cross' the bid-ask spread and 'lift' the ask. The price that is recorded from the transaction will be the price of the ask. Now, what happens if the next trade is someone who wants to sell a stock? They will cross the bid-ask spread again, and "hit" the bid. The price that is recorded will be the price of the ask. If we only look at the price, it will have jumped from the ask to the bid. Imagine that we do this a few more times, all of the trades being executed at the bid or ask price level. If we were to calculate the variance of price, an important statistic used for other financial calculations, we will see that the variance is quite high, even if the price never moved from the tiny range within the bid-ask window. In effect, the bid-ask bounce, if not addressed, will inflate the value of calculated variance.

[2]As noted in Chapter 1.2 a trend effect can be seen if a sufficiently large data-set is used. See, for example, Ref. [41]

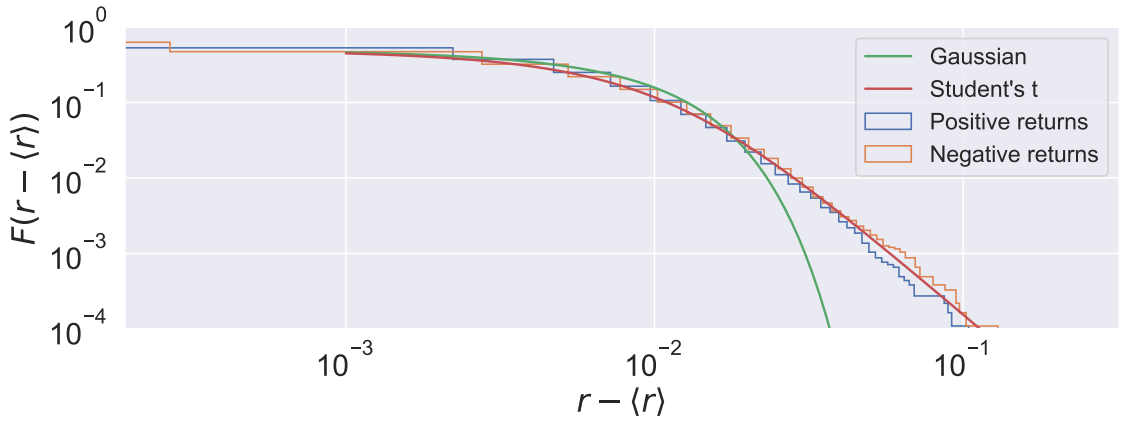


Figure 2.4: Cumulative distribution function of returns related to S&P-500 data. The power-law exponent is found to be equal to $\mu = 3.5$.

2.4 Statistical description of large price changes

If price changes are independent and identically distributed (iid) and Gaussian, the corresponding Probability Distribution Function (PDF) reads:

$$\mathbb{P}(\delta p, t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta p^2}{2\sigma^2}\right). \tag{2.7}$$

Accordingly, the PDF of the price p_t at time t, defined in Eq. (2.6), is Gaussian with mean zero and variance $\sigma^2 t$. Figure 2.4 shows a histogram of empirical returns related to the S&P-500 data. On top of the empirical histogram, we show the outcome of different fits obtained with different PDFs. As one can clearly see, the Gaussian one (green line) does not capture the fat tail of the empirical distribution, i.e., extreme events have a higher than Gaussian probability of occurrence. Another fit (red) line, is related to the Student's t distribution, given by:

$$\mathbb{P}(\delta p, t) = \frac{\Gamma(\frac{\mu+1}{2})}{a\sqrt{\mu\pi}\Gamma(\frac{\mu}{2})}\left(1 + \frac{\delta p^2}{a^2\mu}\right)^{-(\mu+1)/2}, \tag{2.8}$$

where Γ is the Gamma function. As one can see from the fit in Fig. 2.4, the Student's t distribution assigns a larger probability to extreme events.

In order to gauge the distance of a given PDF from the Gaussian one, it is useful to consider the so-called excess kurtosis, defined in terms of the fourth moment divided by the squared variance:

$$\kappa = \frac{\langle (\delta p - \langle \delta p \rangle)^4 \rangle}{\sigma^4} - 3. \tag{2.9}$$

Note that for Gaussian processes κ is equal to zero, while for the Student's t distribution, if $\mu \geq 4$, $\kappa = 6/(\mu - 4)$. Instead, if $2 < \mu \geq 4$, the kurtosis is infinite. In the remaining μ's range, the kurtosis is undefined. Note that an increasing κ, means that the probability of extreme events increases with respect to a Gaussian distribution. In Fig. 2.4 we calibrate a Gaussian and a Student's t distribution over S&P-500 data. Interestingly, the kurtosis found by means of the Student's t distribution turns out to be infinite, since the fitting procedure gives $\mu = 3.5$.

2.5 Multiplicative dynamics

When Bachalier's work was eventually re-discovered by economists in the 1950s, empirical analyses showed that a given stock's price change tends to be proportional to the price itself (as one can see from Fig. 2.2)[3]. A simple multiplicative model for the price process is given by:

$$p_{k+1} = p_k(1 + r_k). \tag{2.10}$$

The distinction between multiplicative and additive processes will become crucial when in Chapter 5 I analyze price trajectories over very long time spans.

Note that one can map the multiplicative equation above to the additive one (Eq. (2.6)), by taking the logarithms while considering $(1 + r_k) = \exp \eta_k$; accordingly, if η_k are iid Gaussian random variables with mean $\langle \eta \rangle$ and variance σ^2, one obtains a Gaussian distribution for the log price. Converting it back to the multiplicative case, one obtains the following so-called log-normal distribution for the price process in Eq. (2.10):

$$\mathbb{P}(p, t) = \frac{1}{p\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(\ln p - \langle \eta \rangle t)^2}{2\sigma^2 t}\right). \tag{2.11}$$

Note that the log-normal distribution is quite similar to the Gaussian one in the body of the distribution. However, as shown in Fig. 2.4, the Gaussian PDF underestimates the level of the tails of returns, while the log-normal predicts that positive price jumps are more frequent than negative ones; this is at odds with empirical findings showing that these jumps are rather symmetrical, and, more precisely, large negative draw-downs are more frequent than positive draw-ups.

A way to consider the tail of the log-normal distribution is to expand the argument of the exponential, obtaining:

$$\mathbb{P}(p, t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\langle r \rangle^2}{2\sigma^2} t\right) p^{-1-\mu(p,t)}, \quad \mu(p, t) = -\frac{\langle r \rangle}{\sigma^2} + \frac{\ln(p)}{2\sigma^2 t} \tag{2.12}$$

and so when $2\sigma^2 t \gg \ln p$ the distribution resembles a power-law with an exponential cut-off. A final note regarding the log-normal distribution is that it does not describe a stationary system. In fact, the mean and the variance diverge over time. To obtain a stationary distribution with exact power-law tails, one must enforce additional constraints. For example, it is sufficient to add a repulsive barrier, as shown in Sec. 2.7.

2.6 Heteroscedasticity

The study of price volatility is of paramount importance: optimal strategies in financial markets are the results of a trade-off between expected risk and return (see for example the Capital Asset Pricing Model [42]); in particular, the risk of a given strategy is related to the price volatility. Moreover, volatility can also be traded, for instance through options, where one needs to properly forecast it in order to price them correctly. Estimating price volatility has been therefore the subject of intensive studies in the 1980s. First, it was recognized that squared returns were correlated over time, as one can see from the middle panel of Fig. 2.5. Later, Engle proposed Auto-Regressive Conditional Heteroscedasticity

[3]Over a short time horizon, however, the market's microstructure comes into play and a description based on additive dynamics is better suited; in fact, in real markets, there is always the so-called thick size, which dictates the minimum possible price change.

(ARCH) models, defined in the following Section, which describe a system that is locally non-stationary but asymptotically stationary, meaning that the parameters which control the PDF of returns fluctuate. This phenomenon is called heteroscedasticity; we will propose a microfoundation for it in Part III.

Before introducing ARCH models, let us first rationalize the heteroscedasticity phenomenon by following Ref. [43]. If the returns distribution at time t ($\mathbb{P}(r,t)$) varies sufficiently slowly, one can compute an empirical estimate of the volatility over a time range way greater than the trading scale, therefore collecting many observations, but small compared to the time scale over which the PDF varies. However, if this is not the case, the PDF constructed from $\{r_t, r_{t+1}, \ldots, r_{t+T}\}$ appears non-Gaussian even if all the $\mathbb{P}(r,t)$ are. To explicitly show this, one can calculate the averaged $\mathbb{P}(r,t)$ over the distribution of volatility, given by:

$$\langle \mathbb{P}(r,t) \rangle_\sigma = \int \mathbb{P}(\sigma) \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{r^2}{2\sigma^2}\right) d\sigma. \tag{2.13}$$

Assuming that $\mathbb{P}(\sigma)$ decays for large σ as $\exp(-\sigma^c)$, $c > 0$. Through a saddle-point calculation one easily obtains:

$$\langle \mathbb{P}(r,t) \rangle_\sigma \propto -r^{\frac{2c}{2+c}}. \tag{2.14}$$

Since $c < 2 + c$, this asymptotic decay is always slower than the Gaussian case, which corresponds to $c \to \infty$. The case where the volatility itself has a Gaussian tail ($c = 2$) leads to an exponential decay of $\tilde{\mathbb{P}}(r)$.

Another signature of heteroscedasticity can be observed by looking at the squared return ACF. Let us consider the very simple model given by:

$$r_t = \eta_t \sigma_t, \tag{2.15}$$

where η_t are iid random variables with zero mean with unit variance, and σ_t is stochastic and can be correlated in time with the η_t's. Assuming that η_t and σ_t are not correlated, the correlation between returns is given by:

$$\langle r_t r_{t+\tau} \rangle = \langle \sigma_t \sigma_{t+\tau} \rangle_\sigma \langle \eta_t \eta_{t+\tau} \rangle_\eta = \delta_\tau \langle \sigma^2 \rangle_\sigma. \tag{2.16}$$

Hence returns are not correlated, but they are not independent, since higher-order, non-linear, correlation functions reveal a non-trivial structure. For example, the ACF of squared returns reads:

$$C_\tau^{r^2} = \langle r_t^2 r_{t+\tau}^2 \rangle - \langle r_t^2 \rangle \langle r_{t+\tau}^2 \rangle = \langle \sigma_t^2 \sigma_{t+\tau}^2 \rangle_\sigma - \langle \sigma_t^2 \rangle_\sigma \langle \sigma_{t+\tau}^2 \rangle_\sigma. \tag{2.17}$$

This ACF reveals a rich structure: for instance, as shown in Fig. 2.5 and 2.6, the ACF of squared returns is long-ranged. The kurtosis of the sum of returns, therefore, decays slower than predicted with an uncorrelated variance. Assuming that the above ACF decreases with τ, the kurtosis of the sum of N returns is given by:

$$\kappa(N) = \frac{1}{N}\left[\kappa(0) + (3 + \kappa(0)) g_0 + 6 \sum_{i=1}^{N} \left(1 - \frac{i}{N}\right) g_i\right], \tag{2.18}$$

where $\kappa(0)$ is the kurtosis of the variable η_t, and g_l the correlation function of the variance, defined as:

$$\langle \sigma_t^2 \sigma_{t+\tau}^2 \rangle_\sigma - \langle \sigma^2 \rangle_\sigma^2 = \langle \sigma^2 \rangle_\sigma^2 g_\tau. \tag{2.19}$$

It is interesting to see that for $N = 1$, the above formula gives:

$$\kappa(1) = \kappa(0) + (3 + \kappa(0)) g_0 > \kappa(0), \tag{2.20}$$

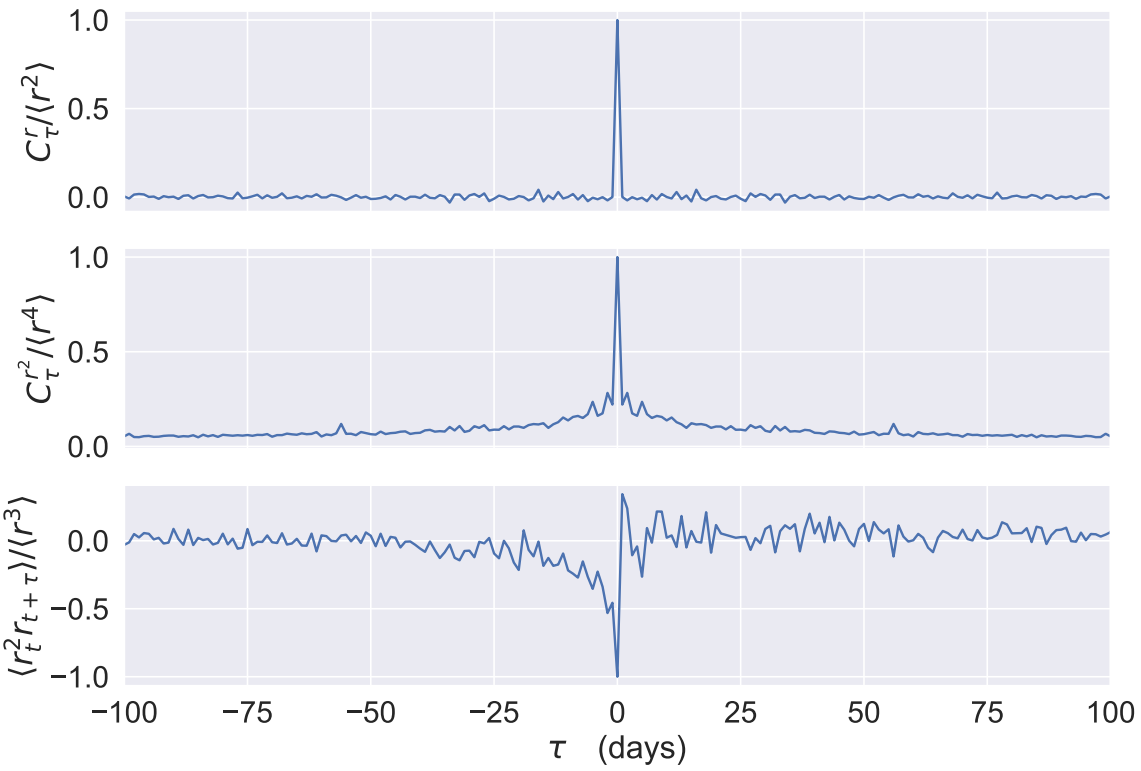


Figure 2.5: Statistical properties of returns related to S&P-500 data. (Top) ACF of returns, (middle) ACF of returns squared, and (bottom) cross-covariance between returns and returns squared.

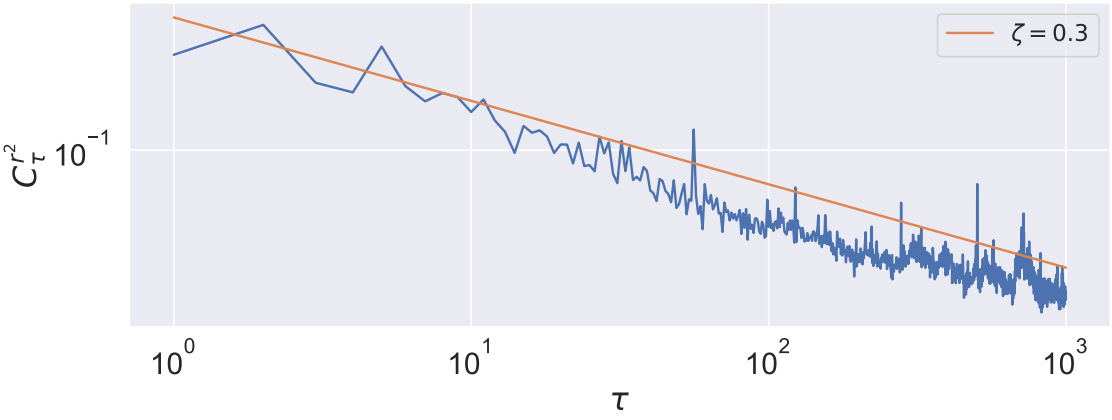


Figure 2.6: Long range correlation of squared returns related to S&P-500 data. (Orange) Power-law function with exponent ζ.

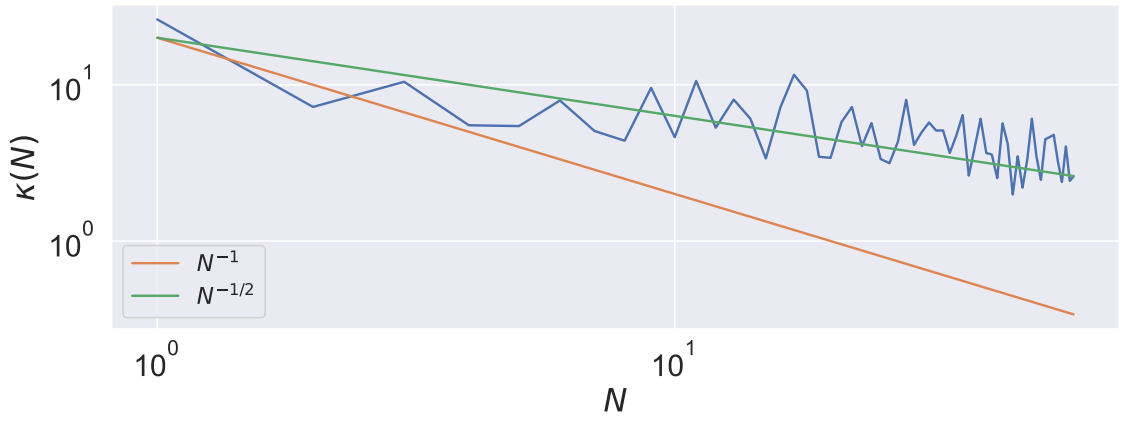


Figure 2.7: Kurtosis of the sum of N returns related to S&P-500 data.

which means that even if $\kappa(0) = 0$, fluctuating volatility is enough to produce some kurtosis. This argument can be viewed as a refinement of the one which led to Eq. (2.14). As can be seen from the analysis of empirical data in Fig. 2.7, the kurtosis of the sum of returns $\sum_{i=1}^{N} r_i$ decays slower than $1/N$, meaning that the variance ACF is correlated over time.

Empirical analysis reveals not only that volatility ACF decays over time as a power law, or as the sum of different exponential functions, but also that its PDF exhibits fat tails. The following Section introduces a multiplicative process that admits a stationary equilibrium with volatility correlations and a fat tail in the distribution of returns.

2.7 Stationary Multiplicative processes

As we have seen in Section 2.5, the multiplicative equation which defines the geometric Brownian motion (given by Eq. (2.10)) does not lead to a well-defined stationary distribution. In order to obtain a stationary distribution with power-law tails, one can consider a stochastic multiplicative process repelled from zero.

In what follows we first consider a model for price heteroscedasticity in discrete time, followed by a discussion of its continuous limit, for which the PDF exhibiting power-law tail can be characterized analytically in closed form[4]. In Chapter 6 I will develop a microfounded model with adaptive agents whose dynamics implies, in a specific limit, an equilibrium where

[4]In discrete time, one obtains again power-law tails, which can be characterized by the Kesten-Goldie theorem.

The starting point is to consider a multiplicative process repelled from zero X_t, which can be thought of as the price variance or volatility, that reads:

$$X_t = A_t X_{t-1} + B_t, \tag{2.21}$$

where the sequences A_t, B_t are supposed to be iid random variables. The process above has been studied extensively in the mathematical literature starting with the analysis performed by Kesten [44]. Sufficient conditions in order to obtain a stationary process are given by $\mathbb{E}[\log A] < 0$ and $\mathbb{E}[\log B] < \infty$. The condition $\mathbb{E}[\log A] < 0$ ensures that the sums that stem from the iteration of the equation above onto itself constitute a random walk with negative drift, and therefore the products $\Pi_k = A_t A_{t-1} \dots A_{t-k-1}$ decay to zero at an exponential rate. This implies the summability of the infinite series which stems from Eq. (2.21). The right tail of X is determined by the products Π_k as well. Indeed, if $\mathbb{P}(A > 1) > 0$, Π_k may exceed 1 finitely often with positive probability. A very important result is the Kesten-Goldie theorem which allows the tail of the stationary distribution of X to be precisely characterized. If

$$\mathbb{E}[A_t^{\mu}] = 1, \quad \mathbb{E}[B^{\mu}] < \infty \quad \mathbb{E}[A_t^{\mu+1}] < \infty, \tag{2.22}$$

the price undergoes a stationary multiplicative process of the GARCH family, presented below.

2.7.1 GARCH models

In 1982 Engle proposed a very simple model able to account for heteroscedasticity [45]. The so-called ARCH(p) model is defined as:

$$\sigma_t^2 = \sigma_0^2 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2, \tag{2.24}$$

where σ_0^2 and $\alpha_1, \ldots, \alpha_p$ are positive variables and r_t in Eq. (2.4) are iid Gaussian random variables. In order to overcome the problems related to the optimal determination of the $p+1$ parameters $\sigma_0^2, \alpha_1, \ldots, \alpha_p$ which best describe the time evolution of a given economic time series, Bollerslev generalized ARCH(p) models in 1986. The GARCH(p, q) model is defined by [46]:

$$\sigma_t^2 = \sigma_0^2 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_i^2. \tag{2.25}$$

Since most of the properties can be addressed considering the GARCH(1, 1) model, in what follows we analyze this simplified case; from the equation above evaluated with $p = q = 1$, the following mean price variance is:

$$\sigma^2 = \frac{\sigma_0^2}{1 - \alpha_1 - \beta_1}, \tag{2.26}$$

and the kurtosis is:

$$\kappa = \frac{6\alpha_1^2}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2}. \tag{2.27}$$

Moreover, from the equation defining the GARCH(1,1) model, it is easy to show that:

$$C_{t+1}^{\sigma^2} = (\alpha_1 + \beta_1) C_t^{\sigma^2}, \tag{2.28}$$

which is solved by an ACF given by:

$$C_t^{\sigma^2} = A e^{-t/\tau}, \tag{2.29}$$

where

$$\tau = |\ln(\alpha_1 + \beta_1)|^{-1}, \quad A = \frac{2\sigma_0^4 \alpha_1^2}{(1 - \alpha_1 - \beta_1)^2 (1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)}. \tag{2.30}$$

The GARCH(1,1) model can be easily calibrated, as described in App. A.1. The outcome of the calibration is shown in Fig. 2.8. For a general GARCH(p, q) process one obtains the variance ACF as a sum of exponentially decaying functions. Several modifications of GARCH models capture additional stylized facts. For instance, asymmetric GARCH models are able to capture the leverage effect[5] (see the bottom panel of Fig. 2.5), while FIGARCH models [48] describe the long-range volatility exhibited by real data (see Fig. 2.6).

---

then

$$\mathbb{P}(X) \sim_{X \gg 1} \frac{1}{X^{\mu}}. \tag{2.23}$$

[5]Note that the leverage effect can be interpreted as another signature of universal human biases [47].

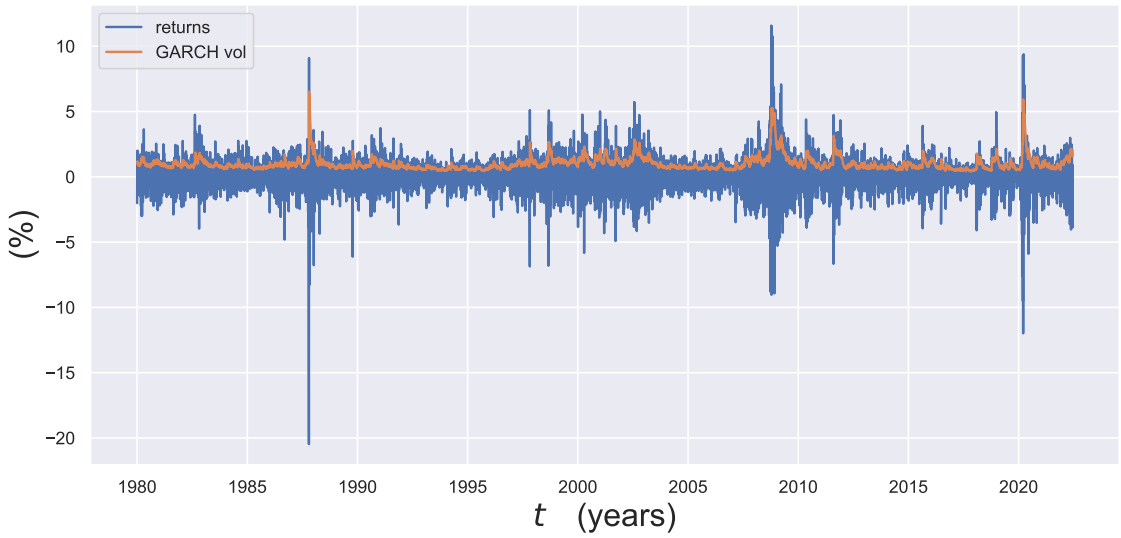


Figure 2.8: Estimation of GARCH model on S&P-500 data. The optimal set of parameters is given by: $\sigma_0^2 = 0.0111, \alpha_1 = 0.0861$ and $\beta_1 = 0.9071$.

2.7.2 GARCH as a learning process

GARCH models belong to the class of observation-driven models, i.e., models where the parameters evolve with past observations. Quite recently [49], a general class of observation-driven models had been proposed, the so-called Score-Driven (SD) one. SD models update parameters based on past observations.

In what follows, I show how a GARCH model can be rephrased as an SD model and draw conclusions from this mapping.

Consider returns given by:

$$r_t = \sigma_t \epsilon_t, \tag{2.31}$$

where

$$\epsilon_t \sim \mathcal{N}(0, 1). \tag{2.32}$$

Let us assume that the probability of observing a return r_t, conditional on observing a volatility σ_t is Gaussian, as in the GARCH model, where:

$$\mathbb{P}(r_t|\sigma_t) = \frac{1}{\sqrt{2\pi}\sigma_t} \, e^{-\frac{r_t^2}{2\sigma_t^2}}. \tag{2.33}$$

Note, that the GARCH(1,1) model can be seen as a SD model, defined by

$$\sigma_{t+1}^2 = \sigma_0 + B\sigma_t^2 + A\mathscr{I}^{-1/2}\nabla_t, \tag{2.34}$$

where the score ∇_t is given by:

$$\nabla_t = \frac{\delta \log \mathbb{P}(r_t|\sigma_t)}{\delta \sigma_t^2}, \tag{2.35}$$

and the weight $\mathscr{I}$ is given by:

$$\mathscr{I}_t = -\mathbb{E}\left[\frac{\delta^2 \log \mathbb{P}(r_t|\sigma_t)}{\delta^2 \sigma_t^2} \middle| \sigma_t^2 \right]. \tag{2.36}$$

The use of the score for updating σ_t^2 is intuitive. It defines the steepest ascent direction for improving the model's local fit in terms of the likelihood or density at time t given the

current position of the parameter σ_t^2. This provides the natural direction for updating the parameter. In addition, the score depends on the complete density, and not only on the first- or second-order moments of the observations r_t which distinguishes the SD framework from most other observation-driven approaches in the literature. By exploiting the full density structure, the SD model introduces new data transformations to update the time-varying parameter σ_t^2. The weight $\mathscr{I}$ is the Fisher information matrix, which is defined as the variance of the score ∇_t, which in turn is given by Eq. (2.36) under mild regularity assumptions [50].

Computing Eqs. (2.35) and (2.36) using the conditional PDF given by Eq.(2.33) and inserting the results in Eq. (2.34) one obtains:

$$\sigma_{t+1}^2 = \sigma_0^2 + \alpha_1 r_t^2 + \beta_1 \sigma_t^2, \tag{2.37}$$

which is exactly the GARCH(1,1) presented in the previous Section.

These results imply that the GARCH model could be seen as a predictive filter, instead of the generating process of the true volatility process. We will come back to this point in Chap. 6, where we will microfound the GARCH model via a stylized agent-based model with adaptive agents.

2.7.3 Continuous limit of GARCH processes

Kesten processes (see Eq. (2.21)) are widely used in finance to model the feedback dynamics where past high values of volatility influence present market activity, leading to tails in the probability distribution and to volatility correlations. Starting from returns r_t defined as in Eq. (2.4) where the variable η_k is iid with unit variance, one postulates that the present day volatility σ_k depends on how the market feels past market volatility. If past price variations happened to be high, the market interprets this as a reason to be more nervous and increases its activity, thereby increasing σ_t. One could therefore consider the following dynamical equation:

$$\sigma_{t+1} - \sigma_0 = (1-\phi)(\sigma_t - \sigma_0) + g\phi\sigma_t\xi_t, \quad 0 < \phi < 1, \tag{2.38}$$

by which the price volatility eventually relaxes towards the equilibrium value σ_0, but is continuously excited by the observation of the previous day's activity through the absolute value of r_t. The coefficient g measures the influence of the updated information on the volatility dynamics. Now, going to a continuous-time formulation, one finds that the volatility $\mathbb{P}(\sigma, t)$ obeys the following Fokker-Planck equation

$$\frac{\delta \mathbb{P}(\sigma,t)}{\delta t} = \phi \frac{\delta(\sigma - \sigma_0)\mathbb{P}(\sigma,t)}{\delta\sigma} + Dg^2\phi^2 \frac{\delta^2\sigma^2\mathbb{P}(\sigma,t)}{\delta\sigma^2}, \tag{2.39}$$

where D is the variance of the noise ξ_t. The equilibrium solution $\mathbb{P}_e(\sigma)$ is obtained by setting the left-hand side to zero, obtaining:

$$\mathbb{P}_e(\sigma) = \frac{\exp(-\sigma/\sigma_0)}{\Gamma[\mu]\sigma^{1+\mu}}, \tag{2.40}$$

with $\mu = 1 + (Dg^2\phi)^{-1}$. This is an inverse Gamma distribution, which rapidly goes to zero when $\sigma \to 0$, but has a long tail for large price volatilities. Note that using a saddle-point calculation, one finds that the tails of the distribution of σ are bequeathed to those of r. The distribution of price changes thus has power-law tails, with the same exponent μ. Interestingly, a short-memory market, corresponding to $\phi \sim 1$, has much wilder tails

than a long-memory market: in the limit, $\phi \to 0$, one indeed has $\mu \to \infty$. In other words, over-reactions are a potential causes for power law tails. Similarly, strong feedback (large g) decreases the value of μ.

Take home messages from Chapter 2

1. To a first approximation, prices are unpredictable on short timescales, e.g., from seconds to months, while on longer ones they are mean-reverting, reflecting an anchor to fundamental values.

2. Extreme returns are more probable than what is predicted by a Gaussian Probability Density Function (PDF). An important quantity useful to gauge the distance of the return distribution from a Gaussian one is the tail exponent μ related to the return PDF.

3. Even if returns are uncorrelated, they are not independent; in fact, the correlation between price volatility, called Heteroscedasticity, implies that squared returns are correlated in time.

4. Multiplicative processes are well suited to model the intermittent dynamics of price volatility. For instance, Generalized Auto-Regressive Conditional Heteroscedastic (GARCH) models allow for a description that predicts fat tails in the return PDF as well as volatility correlations.

5. What are the assumptions needed to rationalize from an agent-based perspective the feedback dynamics assumed in GARCH model to explain volatility clustering and large price fluctuations? In the final Section of the following chapter we review the Kyle model, which will be the starting point for the microfoundation we give to the GARCH model in Chapter 6.

Chapter 3

Order-driven price dynamics

Some economists, when thinking about long memory, are concerned that it undercuts the Efficient Market Hypothesis that prices fully reflect all relevant information; that the random walk is the best metaphor to describe such markets; and that you cannot beat such an unpredictable market. Well, the Efficient Market Hypothesis is no more than that, a hypothesis. Many a grand theory has died under the onslaught of real data.

Benoît B. Mandelbrot, The (Mis)Behavior of Markets

The interplay between trades and prices is key to understanding the price formation process in financial markets, which can be seen as a collective evaluation system where the 'fair' price of an asset is found by the aggregation of information dispersed across a large number of investors. Non-informed investors also participate in the process, by attempting to profit from a local imbalance between supply and demand, thus acting as counterparts when liquidity is needed, or trading for exogenous reasons. Order submission and trading are the means by which information is aggregated. This process creates statistical regularities in the price dynamics, which we shall review before discussing the traders' behavior that may ignite them.

Keywords: Price Impact, Long Memory, Diffusivity Puzzle, Market Microstructure

Contents

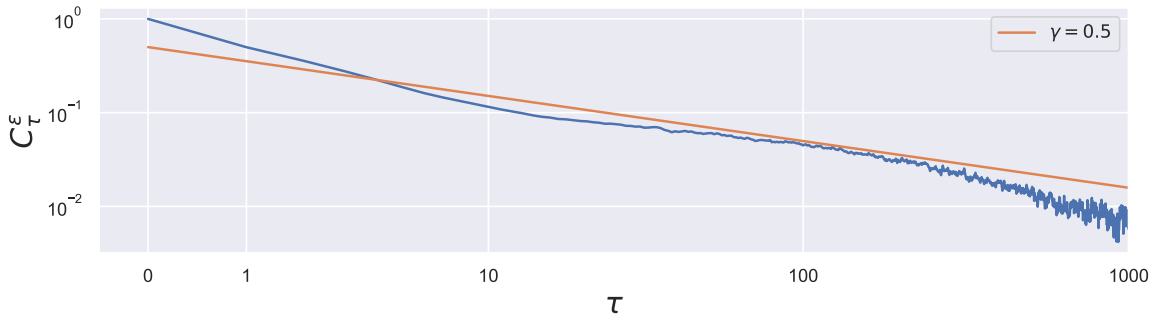


Figure 3.1: Covariance function of trade signs related to E-mini S&P 500 futures contract data versus event time.

First, in Section 3.1 and 3.2, we review the stylized facts that lead to the so-called diffusivity puzzle. In a nutshell, the unpredictability of price changes is, at a first glance, at odds with the empirical finding that buy-initiated or sell-initiated trades are persistent and the fact that there is a positive correlation between trade signs and future price changes: if you buy (sell) a stock now, on average the price will increase (decrease); this effect is called price impact. We then introduce in Section 3.3 the Transient Impact Model and in Section 3.4 the Kyle model, following Ref. [51] and [52]; the former is able to provide a solution to the diffusivity puzzle, while the latter causes the concept of price impact to emerge from traders' behavior. These models are the starting point for my original research presented in Part II.

3.1 Long-term correlation of trade signs

Here we investigate the correlations of trade signs by analyzing one month (January 2018) of trade-by-trade data for the E-mini S&P 500 futures contract. Note that in addition to data related to the price, this dataset also contains information regarding whether the trade was a buy or sell initiated trade. Let us define the sign of a trade ϵ_t in an operative way as follows: $\epsilon_t = -1$ if the trade was initiated by an aggressive sell market order (see Sec. 2.1), else $\epsilon_t = 1$. It has been found [53, 54] that the ACF of trade signs, defined as

$$C^\epsilon_\tau = \langle \epsilon_{t+\tau} \epsilon_t \rangle - \langle \epsilon_{t+\tau} \rangle \langle \epsilon_t \rangle \tag{3.1}$$

decays as a power law:

$$C^\epsilon_\tau = \frac{c_\infty}{\tau^\gamma}. \tag{3.2}$$

This can be seen in Fig. 3.1, where the orange line corresponds to a power law with exponent $\gamma = 0.5$; furthermore, assuming $c_\infty = 0.5$, if we observe a market order now, the probability that a market order 10^4 trades in the future is a buy order exceeds that for a sell order by more than 0.5%. In modern equities markets for liquid assets, a time lag of 10^4 trades corresponds to a few hours. Interestingly, this predictability does not imply price predictability: this is the core of the efficiency paradox.

Where does the predictability of trade signs come from? Two explanations have been presented in the literature, namely herding, in which different market participants submit orders with the same sign, and order splitting, in which market participants who wish to execute large trades split their intended volume into many smaller orders, which they then submit incrementally.

Herding can occur as a consequence of the arrival, for example, of a buy market order which may lead a group of market participants to infer that there is a trader able to forecast a price increase. As a consequence, the group of market participants may decide to create many more buy market orders. This behavior can cascade and thereby lead to long sequences of trades in the same direction.

Order splitting is instead related to the structure of the limit order book, presented in Section 2.1; suppose you want to buy 100 times more shares than those available at the best ask. Sending a market order whose volume is much larger than the volume at the best ask is a bad idea because it would mean paying a much worse price than the ask price, and could possibly wreak havoc in the market. Similarly, sending a very rapid succession of smaller market orders would presumably send a strong signal that there is a hurried buyer in the market, and would likely cause many sellers to increase their prices. Also placing a large limit order does not work because observing an unusually large limit order also signals a large buying interest. This influences both buyers (who are now tempted to buy at the ask price rather than hoping to achieve a better price by placing their own limit orders) and sellers (who think that it might be a bad idea to sell now if the price is likely to go up, as suggested by the new limit order arrival). In fact, limit orders - which are often described as passive because they provide liquidity - can impact prices considerably. In summary, a buyer who seeks to purchase a very large quantity of an asset does not have any other realistic choice than to split the desired trade into many small orders and execute them incrementally, over a period that might span several days or even months. Intuitively, the probability of revealing information increases with the size of a (limit or market) order, because smaller orders are more likely to go unnoticed while larger orders are more likely to attract attention. These actions are consistent with the idea that market participants' execution of large metaorders can cause long-range autocorrelations in the observed order flow.

Although both are likely to play a role, the influence of order-splitting is much stronger than herding. For instance, the relative importance of these effects has been investigated [55] using data where the identity of the trader sending the order is known (even if anonymized). The autocorrelation function of order flow can be exactly decomposed as $C_\tau^\epsilon = C_\tau^{split} + C_\tau^{herd}$ where the first (second) term is the contribution to the correlation considering only cases when the two market orders at time t and $t+\tau$ were placed by the same (different) trader(s). To measure the relative importance of the two components, the authors of Ref. [55] use brokerage data. Some exchanges provide data where each order contains the coded identity of the broker who sent the order.

An extensive investigation shows unambiguously that C_τ^{split} always explains more than 75% of C_τ^ϵ and, except for very short τ (one or two trades) the value is above 85%. This empirical finding strongly indicates that order splitting is the main driver of the correlated order flow. Similar results are obtained when using data with agents rather than brokers.

3.2 Response function

This Section investigates whether there is a relation between price changes and trades.

The response function, R_τ, is defined as the average of the time-dependent price change multiplied by the sign of the trade:

$$R_\tau = \langle (p_{t+\tau} - p_t)\epsilon_t \rangle. \tag{3.3}$$

Note that one can also define a response function conditioned with respect to the traded

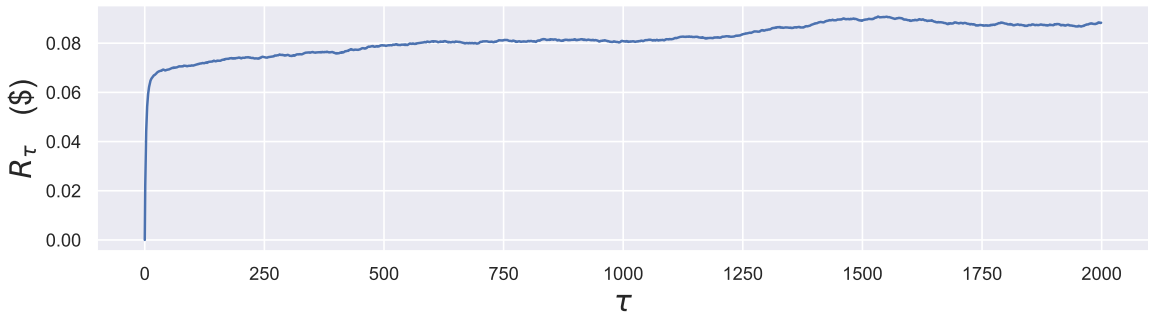


Figure 3.2: Response function related to E-mini S&P 500 futures contract data. τ measures time in event time.

volume at the transaction time t:

$$R_\tau^V = \langle (p_{t+\tau} - p_t)\epsilon_t \rangle|_{V_t=V} \,. \tag{3.4}$$

Empirical investigations show $R_1^V := R^V$ to be well fitted by a power law with a small exponent, or by a logarithmic function of the volume traded. Moreover, the dependence with respect to time and with respect to the traded volume is decoupled [54]

$$R_\tau^V \sim R_\tau R^V. \tag{3.5}$$

Fig. 3.2 shows the time-dependent part of the response function to be almost flat, meaning that a given trade, due to the sequence of prior correlated ones, permanently impacts the price.

The stylized fact captured by the response function can be explained by advancing a few alternative explanations:

> **Trades convey a signal about private information.** The arrival of new private information trigger trades, that cause other agents to update their valuations, which in turn change prices.
>
> **Agents successfully forecast short-term price movements and trade accordingly.** Thus, there might be a market impact even if these agents have absolutely no effect on prices. With Hasbrouck's words [56]: 'orders do not impact prices. It is more accurate to say that orders forecast prices'.
>
> **Random fluctuations in supply and demand.** Fluctuations in supply and demand can be completely unrelated to information, but the net effect on market impact is the same. In this sense impact is a completely mechanical (or statistical) phenomenon.

In the first two explanations, market impact is both friction and also the mechanism that let prices adjust to the arrival of new information. In the third explanation, instead, market impact is unrelated to information and may merely be a self-fulfilling prophecy that could even occur with no informed traders. Identifying the dominating mechanism in real markets is therefore of fundamental importance to understanding price formation.

3.3 The Transient Impact model

Why is it that markets are statistically efficient (i.e., the expected future price change is zero) while trade signs impact prices and display long-range correlations? This question is the core of the so-called diffusivity puzzle, also known as the efficiency paradox.

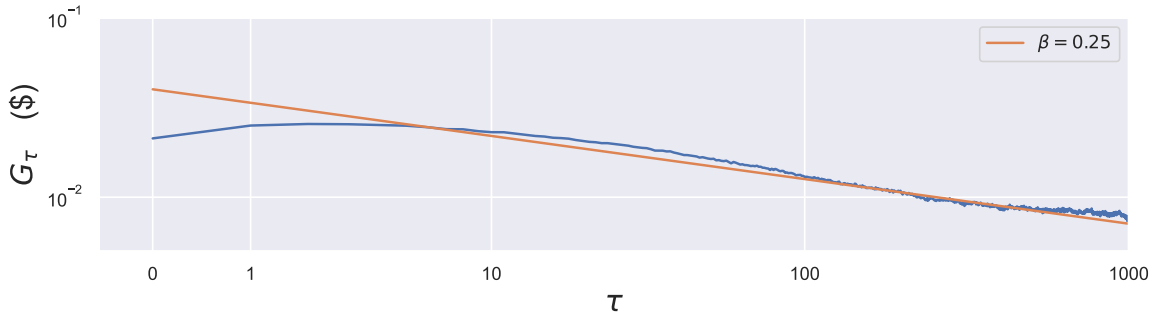


Figure 3.3: Propagator related to E-mini S&P 500 futures contract data versus event time.

In order to offer a unified picture of the stylized fact we discussed in the previous sections of this chapter, thereby solving the efficiency paradox, one expects that the reactional component of price impact is mostly transient such that it allows the price to undergo a diffusive dynamics, despite the long-range correlation of trades signs. This idea is formalized in the so-called Transient Impact model (TIM), also called the propagator model. The central equation of this model is the additive model for the price dynamics given by Eq. (2.6) where the price change δp_t is given by the sum of the impact of prior trades signs weighted by the decaying kernel K and an uncorrelated noise ξ representing exogenous news; therefore, the price change δp_t reads:

$$\delta p_t = \sum_{t' \leq t} K_{t-t'} \epsilon_{t'} + \xi_t, \tag{3.6}$$

where the propagator G_t, or the price impact function, is defined as

$$K_t = G_{t+1} - G_t \tag{3.7}$$

and where $G^0 = 0$.

With these equations, the additive price dynamics becomes:

$$p_t = p_0 + \sum_{t' \leq t} G_{t-t'} \epsilon_{t'} + \sum_{t' \leq t} \xi_{t'}. \tag{3.8}$$

In App. A.2 we show how one can calibrate the above model to real-world data.

Once the TIM is calibrated against real data, one obtains a shape for the impact function well described by a power-law decaying function:

$$G_t \sim_{t \gg 1} \frac{G_\infty}{t^\beta}, \quad \beta < 1, \tag{3.9}$$

as can be seen from Fig. 3.3. The propagator, therefore, decays to zero so slowly that its sum over all t is divergent.

Interestingly, in order to solve the diffusivity puzzle, a particular value of β has to be chosen. One can show this by plugging in the power-law ansatz for the propagator in the definition of the price, respectively given by the equation above and Eq. (3.8), in the definition of the variogram; in the asymptotic limit $t \gg 1$ one obtains:

$$V_t \sim \left(G_\infty^2 c_\infty I(\gamma, \beta) t^{1-2\beta-\gamma} + \Sigma^2 \right) t, \tag{3.10}$$

where $I(\gamma, \beta)$ is a finite integral. As the propagator decays more quickly (i.e., as β increases), super-diffusion is less pronounced until

$$\beta = \beta_c = (1 - \gamma)/2, \tag{3.11}$$

for which the long-term volatility is constant. Note that the constraint above is compatible with the results of the empirical analysis we show in Fig. 3.1 and 3.3.

Therefore, we see that in order to have diffusive prices, the decay of the price impact has to be precisely related to the decay of the trade signs ACF. With the equation above, one can analytically compute the long-term response function and, surprisingly, one can observe that the complicated integral related to the zeroth-order contribution vanishes if $\beta = \beta_c$, leaving a sub-leading term that saturates to a finite value when $t \to \infty$.

We finally note that in the study of the Brownian motion a very interesting result is the so-called Fluctuation Dissipation Theorem, which relates the intrinsic fluctuation of a stochastic system with the (small) deviation caused by an external force. A similar relation can be derived in the present framework, between the variogram and the squared response function, representing respectively the intrinsic fluctuation of the price and the deviation caused by a given trade. In particular, an affine relation exists between V_t/t and R_t^2 which can be investigated both empirically and analytically [54].

3.3.1 Interpretation in terms of agents

Here, following Ref. [54], we show what could be a possible interpretation of the above findings in terms of real agents' strategies. Traders in financial markets can be classified into two broad categories:

- One is that of 'liquidity takers' who submit aggressive orders to the market. These actors are motivated to trade by the fact that they possess (or think to possess) some private information about the future price of the stock. Since market orders are executed immediately, traders who believe they are informed can take advantage of the immediacy, at the expense of crossing the bid-ask spread (see Sec. 2.1).

- The other category is that of liquidity providers (or market makers), who, in the simplest instance, offer to buy or sell without taking bare positions and trying to earn the bid-ask spread s: the sell price is always slightly larger than the buy price, so that each round trip operation leads to a profit equal to the bid-ask spread s, at least if the midpoint has not changed in the meantime. More precisely, note that the average gain per share $\mathcal{G}$ can be computed [54]; assuming that market orders do not contain useful information but are the result of hedging, noise trading, misguided interpretation, errors, etc., then the price should not move up on the long run, and should eventually mean revert to its previous value, so that:

$$\mathcal{G} = s + R_0^V - R_\infty^V = s + \ln V[R_0 - R_\infty], \tag{3.12}$$

  where we used the decomposition of the response function given by Eq. (3.5) and we assumed a logarithmic behavior for the volume-dependent part, in line with empirical observations.

  From the equation above, one sees that it is in the interest of the market maker to mean revert the price, such that $R_0 > R_\infty$. However, this mean reversion cannot be too fast, otherwise, a real informed trader might buy stocks at a very modest price. Hence, this means reversion must be slow.

A very simple model that sheds some light on how to rationalize the propagator model

in terms of agents' strategies is [54]:

$$\frac{dp_t}{dt} = -\Omega(p_t - \bar{p}_t) + \eta_t, \tag{3.13}$$

$$\frac{d\bar{p}_t}{dt} = k(p_t - \bar{p}_t), \tag{3.14}$$

where η_t is the driving force due to trading, Ω is the mean-reversion timescale and k^{-1} is the memory timescale over which the exponential moving average $\bar{p}_t$ is computed. The solution is given by $p_t = \int^t dt' G_{t-t'} \eta_{t'}$, where the price impact function is given by:

$$G_t = G_\infty + (1 - G_\infty) \exp[-(\Omega + k)t]. \tag{3.15}$$

Therefore, the price impact function is given by a single decaying exponential towards a fixed long-term value given by $G_\infty = k/(\Omega + k)$. If the fundamental price was known to every market actor, then $k = 0$ and $G_\infty = 0$. In the opposite limit where $k \gg \Omega$ the last known price is taken as the reference price and $G_\infty \to 1$. A possible way to obtain an impact function that resembles a power law is to consider different market makers that use different timescales. The resulting price impact is a sum of decaying exponential functions with different timescales which can mimic a power-law behavior.

The message of the above model is quite compelling: it states that nobody knows with precision what the fundamental price is and that its best estimate is given by the price itself properly weighted on a given time frame, similar to the model given by Eq. (1.4).

3.3.2 History dependent price impact model

An alternative way to ensure statistical efficiency is to assume that the price impact of each order is permanent but history-dependent. According to the martingale hypothesis, only the surprise component of the next trade should have an impact on the price.

In order to rephrase the TIM, we define the conditional expectation $\hat{\epsilon}_t$ of the trade sign ϵ_t with respect to the last observed one:

$$\hat{\epsilon}_t = \mathbb{E}[\epsilon_t | \epsilon_{t-1}]. \tag{3.16}$$

We can then replace Eq. (3.6) with:

$$r_t = G_{1,t}(\epsilon_t - \hat{\epsilon}_t) + \xi_t. \tag{3.17}$$

Since neither the sign surprise $(\epsilon_t - \hat{\epsilon}_t)$ nor ξ_t can be predicted, it follows that for any immadiate impact $G_{1,t}$,

$$\mathbb{E}[r_t | \epsilon_{t-1}] = 0. \tag{3.18}$$

Therefore, prices in the model given by Eq. (3.17) are a martingale, even when the sign of the next trade is highly predictable.

By imposing the condition above, one can get insight on how the market reacts to different trades, depending on the surprise they convey. To evaluate it, we can decompose the expectation value as the sum of the different possibilities weighted by their probability. There are two possibilities, either the sign ϵ_t matches $\hat{\epsilon}_t$ or it does not. Let $\mathbb{E}_t^+[r_t]$ denote the ex-post value of the return given that ϵ_t matches $\hat{\epsilon}_t$ and $\mathbb{E}_t^-[r_t]$ the return ex-post value given that ϵ_t did not match $\hat{\epsilon}_t$. The absence of predictability imposes that:

$$\frac{1 + |\hat{\epsilon}_t|}{2} \mathbb{E}_t^+[r_t] + \frac{1 - |\hat{\epsilon}_t|}{2} \mathbb{E}_t^-[r_t] = 0, \tag{3.19}$$

where the factors that weigh the expectation values are the probabilities associated with a binary random variable. As a consequence, one can derive the following inequality:

$$\left|\frac{\mathbb{E}_t^+[r_t]}{\mathbb{E}_t^-[r_t]}\right| = \frac{1-|\hat{\epsilon}_t|}{1+|\hat{\epsilon}_t|} \leq 1, \tag{3.20}$$

i.e., the most likely outcome has the smallest impact. This mechanism, which is a crucial condition for market stability, is called asymmetric dynamical liquidity.

In order to prove the equivalence of the history-dependent impact model (HDIM) given by Eq. (3.17) to the TIM, let us assume that market order signs are well modeled by a discrete auto-regressive (DAR) process:

$$\hat{\epsilon}_t = \sum_{k=1}^{p} a_k \epsilon_{t-k}. \tag{3.21}$$

Accordingly, one can rearrange the equation that defines the propagator model, i.e., Eq. (3.8), to give:

$$p_t = p_{t-1} + \theta\epsilon_n + \sum_{t'=1}^{\infty}[G_{t'+1} - G_{t'}] + \eta_t, \quad \theta = G_1. \tag{3.22}$$

The equivalence is obtained with:

$$\theta a_t = G_{t+1} - G_t, \quad or \quad G_t = \theta[1 - \sum_{t'=1}^{t-1} a_{t'}]. \tag{3.23}$$

In particular, the power law relaxation with exponent β is obtained if $a_k \sim k^{-\beta-1}$. Therefore, in history-dependent impact models with a DAR hypothesis for the order flow, one recovers exactly the TIM.

3.3.3 The Propagator model as a reduced description of all the order flow

A possible argument against the propagator model is that it only takes into account market orders, and hence it disregards the effect of limit orders. We show that actually, this is not a valid argument against the propagator model since it incorporates the effect of limit orders in an effective way. To show this, let us assume that the propagator model is modified to:

$$p_t = p_0 + \sum_{0\leq t'<t} \tilde{G}_{t-t'}\epsilon_{t'} + \sum_{0\leq t'<t} H_{t-t'}\xi_{t'} + \sum_{0\leq t'<t} \eta_{t'}, \tag{3.24}$$

where the unobserved events (limit order placements and cancellations) ξ_t are correlated with the observed ϵ_t. Therefore, the unobserved ξ_t can be expressed in terms of the observed ϵ_t via an appropriate linear filter (see Ref. [51] for details). The fitted propagator is therefore given by

$$G_t = \tilde{G}_t + \sum_{t'=1}^{t} H_{t'}\Xi_{t-t'}. \tag{3.25}$$

What is interesting is that G_t can inherit a time dependence from Ξ even if the true propagators G and H are time-independent. In other words, the decay of a single market order's impact should be interpreted as a consequence of the interplay between market orders and limit orders.

3.4 The Kyle model

This chapter introduces the micro-founded model which will be the starting point of my original theoretical approach. The study of market microstructure sometimes allows deriving statistical properties observed in empirical market data from a description based on interacting agents. In order to obtain tractable models, the strong rationality assumption is usually enforced, meaning that agents know the model as well as the model builder, and they optimally construct their strategies. Therefore, agents are strategic: an important aspect of an agent's decision is the inference made from market statistics about others' information. The Kyle model is the first model which addresses strategic aspects of information. In particular, this model allows the explicit characterization of how an informed trader would choose to transact to maximize the value of private information. This provides a way to characterize how information is incorporated into security prices across time given the strategic use of information by an informed trader. The Kyle model is composed of two rational traders with asymmetric information, plus a third agent who injects noise into the system and allows for the overcoming of the no-trade theorem [14]. The two strategic traders are called the informed trader and the market maker. A single risk-neutral informed trader and an uninformed liquidity taker submit orders to a risk-neutral market maker who clears the excess demand while setting the price. In the Kyle model, it is the informed agent's conjecture about the market maker's pricing policy as well as the market maker's inference about the informed agent's information that play a crucial role in determining the nature (and even the existence) of the equilibrium. The resulting equilibrium price, in the simplest instance of the model, is given by the excess demand created by the liquidity takers multiplied by the so-called price impact function (introduced in the previous Section). In what follows we detail the Kyle model, which we later use to provide a micro-foundation for the TIM and for the GARCH$(1,1)$ model.

Before proceding, for clarity purposes, let me note that informational asymmetry is not the only way to microfound the notion of price impact. An alternative possibility relies on the interplay between inventory constraint and risk management [57–61]. The idea is that market makers immediately shift their price quotes after building up inventory in a given trade, but this price shift then gradually disappears as they unload their positions either to other clients or on the open market.

3.4.1 The single-auction setting

Initially, Kyle considers [18] a single trading period in which the informed trader submits his optimal order along with the orders submitted by uninformed traders. Kyle then temporally extends the model to consider the sequential-auction and continuous-action frameworks. As the intuition is the same, we focus on the simpler (single-period) version of the model.

The market maker aggregates the orders and clears all trades at a single price. Hence, Kyle's model does not allow for a bid-ask spread nor does it analyze the transaction price for individual trades. For this reason, the Kyle model can be useful to study empirical data after some aggregation.

Consider a single-asset market populated by three traders:

- At time $t = 0$, the informed trader discovers private information about the price p^{F} that the asset will have at $t = 1$. This liquidation value is assumed to be normally distributed with zero mean and variance $(\sigma^{\mathrm{F}})^2$. Based on his private information, the informed trader chooses a volume q^{IT} of the asset, in such a way as to maximize his expected profit.

- The noise trader trades for idiosyncratic reasons. In doing so, the noise trader generates a random order flow with an excess demand of q^{NT}, whose sign and amplitude is independent of p^{F}. The noise trades have zero mean and variance ω^2.

- The market maker clears the excess demand $q = q^{\mathrm{NT}} + q^{\mathrm{IT}}$, at a clearing price p. The market maker's choice for the price p is such that he breaks even on average.

At time $t = 1$ the price p^{F} is revealed and the informed trader's asset is exactly worth p^{F}. This assumes that he can buy or sell any quantity of the asset at price p^{F} without causing an impact.

In the tradition of theoretical economics, one then looks for an equilibrium between the informed trader and the market maker such that:

i) Profit maximization: given the market maker's price-clearing policy, the informed trader's demand q^{IT} maximizes his expected gain. He buys/sells at the clearing price p, but the asset will be worth p^{F}, so his gain is given by:

$$\mathcal{G} = q^{\mathrm{IT}}(p^{\mathrm{F}} - p). \tag{3.26}$$

ii) Market efficiency: the market maker's clearing price must be such that $p = \mathbf{E}[p^{\mathrm{F}}|q]$. This corresponds to a situation where the market maker breaks even on average, given an excess demand q.

Given that the informed trader knows all of the above, how should he choose q^{IT} to maximize his expected profit $\mathbf{E}[\mathcal{G}]$ at time $t = 1$? To choose an optimal value of q^{IT} to maximize $\mathcal{G}$, he must consider the market maker's clearing price. Specifically, the informed trader is aware of the market maker's price-clearing rule for choosing p as a function of q, so he must use this knowledge when deciding how to act.

On the other hand, the market maker observes the excess demand q, but does not know the value of p^{F}. Since orders are anonymous, he also does not know the values of q^{NT} or q^{IT}. However, the market maker knows that the insider knows his price-clearing rule, and also knows that he acts to optimize his gain $\mathcal{G}$ with the information at his disposal, given by p^{F}, which he will try to infer from the excess demand q.

The model also assumes the market maker knows that the excess demand executed by the noise trader is a Gaussian random variable with zero mean and variance equal to ω^2 and that the p^{F} is similarly a Gaussian random variable with zero mean (over time) and variance $(\sigma^{\mathrm{F}})^2$. The latter quantity is related to the typical amount of information available to insiders at the time $t = 0$ but not yet included in the price.

3.4.2 Linear equilibrium

Consider the case where the market maker's price-clearing rule is linear in the order imbalance, such that:

$$p = Gq, \tag{3.27}$$

for some impact parameter G. One can show that the solution to the problem presented above is self-consistent and can be fully determined. Indeed, profit maximization on the informed trader's part leads to:

$$q^{\mathrm{IT}} = \arg\max_{q^{\mathrm{IT}}} \mathbb{E}[\mathcal{G}], \tag{3.28}$$

$$\mathcal{G} = q^{\mathrm{IT}}(p^{\mathrm{F}} - Gq). \tag{3.29}$$

Since the expected value of the noise trade q^{NT} is zero, one has $\mathbb{E}[q] = q^{\mathrm{IT}}$. This leads to a quadratic maximization problem, the solution of which is:

$$q^{\mathrm{IT}} = \frac{1}{2}\frac{p^{\mathrm{F}}}{G}. \tag{3.30}$$

Since the market maker knows that the informed trader will do this, he attempts to estimate the value of q^{IT}, when he only observes q. Estimating q^{IT} allows him to guess, using the equation above, the value of p^{F} used by the informed trader, and chooses G in such a way that his clearing price p is an unbiased estimate of the future price.

Using Bayes' theorem, the conditional probability that the insider's volume is q^{IT}, given q, is:

$$\mathbb{P}_{\mathrm{MM}}(q^{\mathrm{IT}}|q) \propto \mathbb{P}(q|q^{\mathrm{IT}})\mathbb{P}(q^{\mathrm{IT}}) \tag{3.31}$$

$$\propto \exp\left[-\frac{(q-q^{\mathrm{IT}})^2}{2\omega^2}\right]\exp\left[\frac{-2(q^{\mathrm{IT}})^2G^2}{(\sigma^{\mathrm{F}})^2}\right], \tag{3.32}$$

where in the second exponential we have used the relation between q^{IT} and p^{F} (see Eq. (3.30)). By merging the two exponential terms, we see that the market maker inferred distribution for q^{IT} is Gaussian, with the following conditional mean:

$$\mathbb{E}[q^{\mathrm{IT}}|q] = \frac{(\sigma^{\mathrm{F}})^2}{(\sigma^{\mathrm{F}})^2 + 4G^2\omega^2}q. \tag{3.33}$$

This in turn converts into the market maker's best estimate of the insider's view on the future price; again using Eq. (3.30) one finds:

$$p = \mathbb{E}[p^{\mathrm{F}}|q] = 2G\frac{(\sigma^{\mathrm{F}})^2}{(\sigma^{\mathrm{F}})^2 + 4G^2\omega^2}q. \tag{3.34}$$

Identifying this expression with Eq. (3.27), it also follows that

$$G = \frac{2G(\sigma^{\mathrm{F}})^2}{(\sigma^{\mathrm{F}})^2 + 4G^2\omega^2}. \tag{3.35}$$

Simplifying this expression provides the following solution for the price impact function G:

$$G = \frac{\sigma^{\mathrm{F}}}{2\omega}. \tag{3.36}$$

By choosing this value of G, the market maker makes sure that the strategic behavior of the insider and the stochastic nature of the noise trader combines in such a way that the realized price p is an unbiased estimate of the fundamental price p^{F}, given the publicly available information at time $t = 0$.

3.4.3 Discussion

The Kyle model raises several interesting points of discussion. First, the impact of an order grows linearly with q: the Kyle model leads to a linear impact law. The price impact G grows with the typical amount of private information present in the market (measured by σ^{F}) but decreases with the typical volume of uninformed trades (measured by ω). This captures the basic intuition that market makers protect themselves against adverse selection from informed traders by increasing the cost of trading while benefitting from the presence of noise traders to reduce price impact.

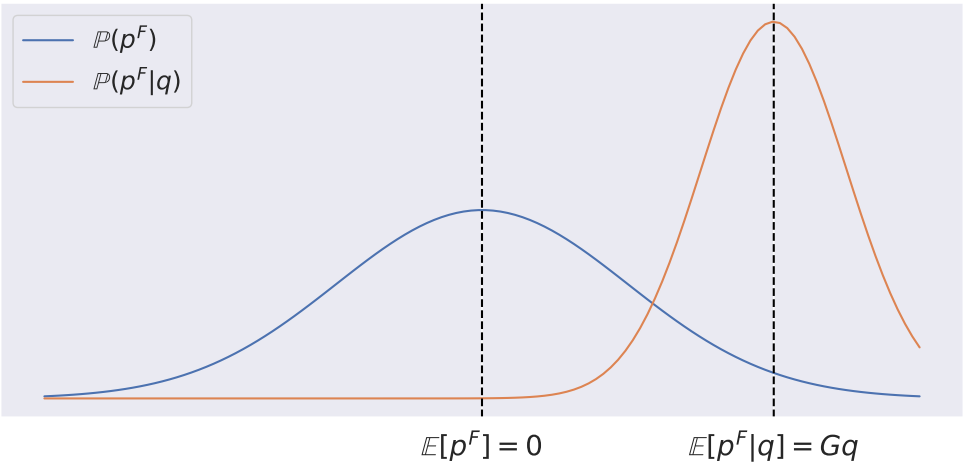


Figure 3.4: Example of prior (blue) and posterior (orange) beliefs of the market maker about the fundamental price.

Second, using the result for q^{IT} in Equation (3.30), the result for p in Eq. (3.34) and Eq. (3.36), it follows that the insider's gain is given by :

$$\mathbb{E}[q^{\mathrm{IT}}(p^{\mathrm{F}} - p)] = \frac{1}{2}\sigma^{\mathrm{F}}\omega. \tag{3.37}$$

Therefore, the conditional expectation of the insider's gain increases with the amount of private information and with the overall liquidity of the market, which is provided by the noise traders. For typical values of the predictor (i.e. for values of p^{F} of the order of σ^{F}), it follows that $q^{\mathrm{IT}} \sim \omega$, so the insider contributes a substantial fraction of the total traded volume. This is not very realistic: informed traders tend to limit their trading to a small fraction of the total volume; this is due to order splitting, as we have emphasized in Sec. 3.1. Finally, the pricing error at $t = 1$ can be measured as:

$$\langle (p - p^{\mathrm{F}})^2 \rangle = \frac{1}{2}(\sigma^{\mathrm{F}})^2. \tag{3.38}$$

Therefore, the market maker is only able to halve the variance of the uncertainty of the fundamental price known to the insider (see Fig.3.4). The Kyle model provides a clear picture of the origin of price impact. In the model, market makers fear that someone in the market is informed, and therefore react by increasing the price when they observe a surplus of buyers and decreasing the price when they observe a surplus of sellers. The model also illustrates that even though the model permits the insider to enter an infinitely large position, his private information only leads to bounded profits, because of impact-related costs.

Note that when moving slightly away from its core assumptions, the Kyle model becomes a self-fulfilling mechanism: suppose that the market maker overestimates σ^{F} (i.e. he overestimates the quality of information available to the insider). In reality, because there is no 'terminal time' when a 'true price' p^{F} is revealed, such market makers will overreact to the excess demand when setting the value of p. In the efficient-market picture, these pricing errors should self-correct through arbitrage from the informed trader (i.e. with a signal to trade in the opposite direction at the next time steps). This would lead to excess high-frequency volatility. However, variograms of empirical data are rather linear at high frequency (see Fig. 2.3), which instead suggests that the whole market shifts its expectations around the new traded price.

Take home messages from Chapter 3

1. Why is it that price is diffusive up to several weeks while the order flow signs are correlated in time and trades impact the price? This conundrum, also known as the diffusivity puzzle, can be solved via the Transient Impact Model (TIM).

2. By calibrating the TIM on real price data, one obtains a slow decaying price impact function. This finding can be substantiated by theoretical investigations if one assumes a power-law decaying Auto-Correlation Function of the order flow process and price diffusivity; in fact, one predicts a price impact function that decays at high frequency over time as a power law, with an exponent related to that of the order signs. The slow decay of the price impact function is able to counter-balance the persistence created by the persistence of the trades' signs, without preventing the long-time price response function from reaching a finite level.

3. The Kyle model is an agent-based model with three classes of agents. The informed trader has private information about the future price and chooses a trading volume to optimize his profit. The noise trader submits a random trade volume. The market maker clears the volumes submitted by the informed trader and the noise trader and chooses his clearing price to equal his expectation of the fundamental price such that his expected profit is zero. The informed trader and the market maker are rational, i.e., they know each other's strategy and use this information to construct their strategy.

4. In the Kyle model volumes impact the price because of their informational content. This exposes the market maker to adverse selection. By adjusting the price (negatively) to order-flow imbalance, the market maker ensures that on average, the impact compensates exactly for this adverse selection.

5. When all distributions are Gaussian, the price change scales linearly with the informed trader's volume. The proportionality coefficient is called price impact function. The value of the price impact function measures market liquidity: the larger the coefficient, the more a given volume impacts the price and the more expensive trading is, i.e., the less the market is liquid. The larger the number of noise traders, the more liquid the market is. In this sense, a market needs uninformed participants to function smoothly.

6. The Kyle model assumes that the fundamental price is revealed to all market participants at some terminal time. This is an unrealistic assumption, which we shall relax in Chapter 4. In doing so, we give a microfoundation for the Propagator model, which we calibrate against real data in Chapter 5.

7. The Kyle model is not able to explain excess volatility, because it assumes a perfectly rational market maker. Moreover, it is not able to capture the intermittency exhibited by price variations observed in real-world data. In Chapter 6 we replace rational traders with adaptive ones, microfounding the GARCH$(1,1)$ model reviewed in Chapter 2.

Part II

Asymmetrically informed rational agents

Chapter 4

The stationary Kyle model

Reality must take precedence over public relations, for nature cannot be fooled. It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiments, it's wrong.

Richard Feynman

We provide an economically sound microfoundation to linear price impact models, by deriving them as the equilibrium of a suitable agent-based system. In particular, we retrieve the so-called propagator model is the high-frequency limit of a generalized Kyle model, in which the assumption of a terminal time at which fundamental information is revealed is dropped. This allows describing a stationary market populated by asymmetrically-informed rational agents. We investigate the stationary equilibrium of the model and show that the setup is compatible with universal price diffusion at small times, and non-universal mean-reversion at time scales at which fluctuations in fundamentals decay. Our model suggests that at high frequency one should observe a quasi-permanent impact component, driven by slow fluctuations of fundamentals, and a faster transient one, whose timescale should be set by the persistence of the order flow.

Keywords: Market Microstructure, Price Impact, Statistical Inference

From:
A Stationary Kyle Setup: Microfounding propagator models
Michele Vodret, Iacopo Mastromatteo, Bence Tóth, Michael Benzaquen

Contents

4.1 Introduction

Financial markets are designed to achieve two seemingly unrelated goals: they allow market participants to find other agents with whom to transact (thereby solving a *liquidity* problem), and at the same time they allow to discover the price at which such transactions should take place (thereby solving an *information*-related task).

The Efficient Market Hypothesis (EMH) states that prices integrate all information that is publicly available [62]. If this is the case, there can be no forecastable structure in asset returns for agents in possession of public information only. Historically, the EMH was first rationalized theoretically with the introduction of the Rational Expectation Hypothesis (REH). According to the REH all agents are rational and perfectly informed about the other players' strategies. This hypothesis is appealing since it allows to build analytically tractable setups [63] in which financial markets are able to deliver the promise they were conceived for, once some exogenous source of dynamics is injected into the system, thus preventing no-trade theorems. It has also important drawbacks: for example, the REH implies that the value of a risky asset is completely determined by its fundamental price, equal to the present discounted value of the expected stream of future dividends. As already argued by Shiller [15], the excess volatility puzzle, i.e., the fact that the price deviates substantially from the fundamental value, cannot be explained by the REH. Nevertheless, the REH is still considered the main expectation formation paradigm in many economic circles [64].

An important class of REH models is the so-called Information-Based Models. These models typically involve the presence of agents that trade due to exogenous reasons (*noise traders*) that use financial markets in order to find counterparties for satisfying needs that come from *outside* of the market, and arbitrageurs that possess privileged information on the traded goods (*informed traders*) and thus choose to transact whenever they expect to use their informational advantage in their favor. From this perspective, informed traders provide a service (making prices informative) that noise traders can choose to pay in order to be granted access to liquidity. To lubricate this mechanism, dealers (*market makers*) are typically required: instead of letting noise traders and informed traders interact directly, market makers can temporarily incorporate the imbalance in the trading pressure, accepting to bear inventory risk for a limited time under the promise of some reward (bid-ask spread, rebate fees). Their activity allows deferring in time the moment at which the initial buyer and the final seller meet, thus enabling both informed and noise traders to find more easily possible counterparties.

A particularly successful class of models to describe statistical regularities in financial markets involves the notion of *propagator*, a linear kernel used in autoregressive models that couples price changes to past order flow imbalances. In this setting, the (discounted) price of a good at time t, which we denote p_t, can be expressed as a function of the past signed order flow imbalance q_t as:

$$p_t := \sum_{t'=-\infty}^{t} G_{t-t'} q_{t'} \,, \tag{4.1}$$

where the causal kernel G_t is the propagator.[1] Propagator models were originally proposed in order to solve the so-called *diffusivity puzzle*, namely the fact that price efficiency, and consequently price diffusion, can be achieved even if the order flow imbalances q_t display long-ranged correlation [54]. Moreover, variations of these models have proven to be effective

[1] Note that we are omitting from Eq. (4.1) a residual noise term, that can be easily restored in order to account for price changes that are not explained by the past order flow.

in order to paint an accurate picture of the market at high frequency [65, 66], in the sense that a large fraction of the price fluctuations can be explained by the past order flow [67].

On the other hand, the perspective taken in order to construct such models is quite distinct from the one preferred in theoretical economics. The propagator setup is not properly microfounded. In fact, it builds on statistical stylized facts, rather than on an economic rationale. The goal of this Chapter is to bridge this gap in an economically standard setting by showing how propagator-like models can be rationalized as the equilibrium resulting from a set of rational agents seeking to achieve optimality. Along this line, our work is closely related to the classic Kyle setting [18], in which the price discovery mechanism emerges as a linear equilibrium between three representative agents with asymmetric information. We think that our minimal model, as the Kyle model, is simple enough to be extended in several directions.

We establish a setting for an Information-Based Model that gives rise to a stationary market, where the equilibrium pricing rule is given by Eq. (4.1). A similar setting has been considered in Ref. [68] in the special case of a stationary Markovian system. Here, instead, we keep the model general, so to account for memory effects (order flows are strongly correlated in real markets), thus extending some of the results of the aforementioned investigation. Our work goes beyond the purely theoretical aspect since the framework we build allows to explicitly construct kernels G_t that ensure price efficiency under different circumstances.

The organization of the Chapter is as follows. Section 4.2 introduces the notations we use throughout the Chapter. In Section 4.3 we present the model. Section 4.4 is devoted to the study of the equilibrium of the model. Section 4.5 discusses the relation of our model with its building blocks, namely the original propagator and the Kyle model. In Section 4.6 we further investigate the model we propose in the paradigmatic Markovian case, whose tractable solution allows us to gain intuition on the system. In Section 4.7 we conclude and propose several extensions of our framework.

4.2 Notations

Throughout the Chapter, we will alternate between scalar notations, in which the time dependence of the variables is explicit (e.g. X_t), vector notations, and matrix notations. We will use bold symbols for vectors and Sans Serif symbols for matrices.

For convenience we introduce two types of vectors: $\boldsymbol{X}_t := \{X_{t'}\}_{t'=-\infty}^{t}$ and $\boldsymbol{X}_{/t} := \{X_{t'}\}_{t'=t}^{\infty}$. Further, for a given vector $\boldsymbol{X}_t$ we define the associated Toeplitz matrix as $\mathsf{X}_{t,t} := \{X_{t'-t''}\}_{t',t''=-\infty}^{t}$. In some cases, we will omit the time index for brevity. In situations where such omission would be ambiguous, we will restore time indices explicitly, e.g. to deal with matrices such as $\mathsf{X}_{/t,/t} = \{X_{t'-t''}\}_{t',t''=t}^{\infty}$ or $\mathsf{X}_{/t,t} = \{X_{t'-t''}\}$ with $t' \geq t$ and $t'' \leq t$. The transpose operation will be denoted by the superscript $\top$.

The identity matrix is denoted I, the vector with all components equal to one is written $\mathbf{1}$, and the upper triangular matrix with all non-zero entries equal to 1 is denoted by U. The Kronecker delta is represented by a vector $\mathbf{e}_t$ with components $e_t(t') = \delta_{t,t'}$. The lag operator, i.e., the operator that acts on an element of a time series to produce the previous element, is denoted L. In this way, we write $\boldsymbol{X}_{t-1} = \mathsf{L}\ \boldsymbol{X}_t$. Dimensionless quantities are signified with tildes.

4.3 A Simple Agent-Based Market Model

4.3.1 Setup of the Model

Consider a market in which agents exchange a risky asset (stock) against a safe asset (cash). The (discounted) transaction price of the risky asset at time t is denoted by p_t. Each unit of the risky asset entitles its owner to a stochastic payoff μ_t in cash (dividend) at each unit of time t. The dividend process μ_t is modeled as an exogenous, stationary, zero-mean Gaussian process with autocovariance function (ACF):

$$\Xi_\tau^\mu := \mathbb{E}[\mu_t \mu_{t+\tau}]\,. \tag{4.2}$$

The portfolio of each agent comprises a combination of risky and safe assets. The position of agent i in the risky asset at time t is given by Q_t^i, whereas his trades are denoted by $q_t^i := Q_t^i - Q_{t-1}^i$. With these conventions, the equations for the evolution of cash C_t^i, stock-position Q_t^i, and wealth W_t^i for each agent can be written down respectively as:

$$\begin{aligned} \Delta C_t^i &:= \mu_t Q_t^i - p_t q_t^i & (4.3)\\ \Delta Q_t^i &:= q_t^i & (4.4)\\ \Delta W_t^i &:= \Delta C_t^i + Q_t^i p_t - Q_{t-1}^i p_{t-1}. & (4.5) \end{aligned}$$

We consider an agent-based market model with asymmetric information akin to the well-known Kyle model [18], in which the agents take actions at discrete time steps t. A strategic agent possessing privileged information about the realizations of the stochastic dividend process (*informed trader*, or IT) trades with a non-strategic and non-informed trader (*noise trader*, or NT) that accesses the market for exogenous reasons. Both the IT and NT are modeled as liquidity takers. A liquidity provider (*market maker*, or MM) provides liquidity for both the NT and the IT and sets the transaction price p_t.

At the beginning of each time interval $[t, t+1]$ both the IT and the NT build a demand for the risky stock q_t^i (with $i \in \{\mathrm{IT}, \mathrm{NT}\}$). *After* the excess demand $q_t := q_t^{\mathrm{IT}} + q_t^{\mathrm{NT}}$ is formed, the MM clears the excess demand of the liquidity takers, executing a trade $q_t^{\mathrm{MM}} := -q_t$ and setting the transaction price p_t. The price p_t arises endogenously as the result of the action of the agents, described in what follows.

Before discussing the information sets and the strategies of the different agents, let us highlight that both the IT and the MM have exact knowledge of the statistical properties of the exogenous processes μ_t and q_t^{NT}, as well as each other's strategy. Past prices and excess demands are also public information.

Noise trader The NT acts in a purely stochastic fashion. His demand process q_t^{NT} is a zero-mean, stationary Gaussian process with ACF given by:

$$\Omega_\tau^{\mathrm{NT}} := \mathbb{E}[q_t^{\mathrm{NT}} q_{t+\tau}^{\mathrm{NT}}]\,. \tag{4.6}$$

Informed trader The IT is a strategic, risk-neutral (expected) utility maximizer. His access to privileged information about the dividend process is modeled by assuming that he observes past realizations of the process μ_t and uses such information to maximize his future expected wealth. Moreover, since realized past excess demand is public information,

the IT can trivially infer the NT's past trades. The information accessible to the IT at time t is thus given by:

$$\mathcal{I}_t^{\mathrm{IT}} = \left\{ \mathbf{q}_{t-1}, \mathbf{q}_{t-1}^{\mathrm{NT}}, \boldsymbol{\mu}_{t-1} \right\}. \tag{4.7}$$

So the IT builds his demand without exploiting equal-time information on either p_t nor on the decision of his peers. In order to maximize his wealth, the IT exploits privileged information on realized dividends.

Since the IT is risk-neutral and assuming that the price is a linear function of realized excess demands (we shall discuss why this is the case in a moment), his demand q_t^{IT} at time t is a *linear* function of his current information set $\mathcal{I}_t^{\mathrm{IT}}$:

$$\mathbf{q}_t^{\mathrm{IT}} \;=\; \mathsf{R}\mathbf{q}_{t-1} + \mathsf{R}^{\mathrm{NT}}\mathbf{q}_{t-1}^{\mathrm{NT}} + \mathsf{R}^{\mu}\boldsymbol{\mu}_{t-1}, \tag{4.8}$$

where we have introduced the *demand* kernels $(\mathsf{R}, \mathsf{R}^{\mathrm{NT}}, \mathsf{R}^{\mu})$. Let us give here a first description of these demand kernels. Since we discuss a market with multiple trading periods, the IT strategically takes into account past trades and past dividends in order to determine his demand. The demand kernel R accounts for the dependence on past order flow which arises from the price impact of past traded volumes. The kernel R^{NT} accounts for the dependence that comes from the price impact induced by expected future trades of the NT, while the kernel R^{μ} accounts for the dependence arising from expected future dividends. The demand kernels are the result of a Model Predictive Control (MPC) [69] strategy. Indeed, as soon as a new piece of information is available to the IT (i.e. at each time-step t), he will construct an updated long-term strategy, and he will trade accordingly. More details about IT's MPC strategy are provided in Sec. 4.3.3, with explicit expressions of the demand kernels.

Market maker The MM is risk-neutral and competitive. He sets a pricing rule that allows him to statistically break even on every trade, without controlling the inventory that he might accumulate while matching the demand. The past realization of the dividend process μ_t is unknown to the MM, and so is the proportion of the demand due respectively to the IT and the NT. Thus, the information set available to the MM at time t is solely given by realized aggregate excess demand:

$$\mathcal{I}_t^{\mathrm{MM}} := \{\mathbf{q}_t\}. \tag{4.9}$$

An important point is that the resulting excess demand q_t conveys information to the MM about the asset's fundamental value, via the information set used by the IT (Eq. (4.7)) to construct his trading schedule (Eq. (4.8)). Note also that the information set of the MM is *not* contained in the information set of the IT, due to the fact that the excess demand q_t is only available to the IT at time $t+1$.

Since the MM knows that the IT' s trading schedule is given by Eq. (4.8), from the total order flow he can infer information about past dividends, albeit this information is distorted by the presence of noise induced by the NT. From Eq. (4.8), the dynamics of the excess demand is given by:

$$\mathbf{q}_t = (\mathsf{I} - \mathsf{R}\mathsf{L})^{-1} \left[\left(\mathsf{I} + \mathsf{R}^{\mathrm{NT}}\mathsf{L} \right) \mathbf{q}_t^{\mathrm{NT}} + \mathsf{R}^{\mu}\boldsymbol{\mu}_{t-1} \right]. \tag{4.10}$$

Due to the Gaussian nature of both μ_t and q_t^{NT} and the risk-neutral nature of market participants, the choice of considering a linear (instead of a general) equilibrium implied by Eq.(4.1) appears natural. On the other hand, we do not have proof of the uniqueness of the linear stationary equilibrium. Given that such uniqueness holds in a framework similar

to ours (given by Ref. [70]), it is reasonable to expect that uniqueness should hold even in our case. Thus, the market can be modeled by the MM as a Linear Gaussian State-Space Model (LG-SSM) [71]. Actually, while the state of the market, i.e. realized dividends $\boldsymbol{\mu}_t$ and NT's trades $\mathbf{q}_t$, are not observable by the MM, he can infer these quantities, and in particular realized dividends, filtering them out from his information set. This procedure in the LG-SSM literature is referred to as Kalman filtering technique. More details about these important aspects of the model will be given in the following Section.

4.3.2 Competitive pricing rule

As anticipated above, we assume the MM to be competitive and risk neutral. Thus, by a Bertrand auction type of argument [18], we postulate a break-even condition for the MM for each T-period holding strategy built as follows: buy q_t units of stock by matching the demand at time t at a price p_t and sell them back at time $t+T$ at a price $\mathbb{E}[p_{t+T}|\mathcal{I}_t^{\mathrm{MM}}]$, earning the dividends in the meanwhile. Note that even though the MM cannot choose to execute with certainty at $t+T$, we can see T as the time lag at which the MM decides to mark-to-market his position, even if he might not be actually able to liquidate it. Imposing competitiveness of the MM, this trajectory should have zero payoff on average, leading us to postulate a pricing rule of the form:

$$p_t = \sum_{t'=t}^{t+T-1} \mathbb{E}[\mu_{t'}|\mathcal{I}_t^{\mathrm{MM}}] + \mathbb{E}[p_{t+T}|\mathcal{I}_t^{\mathrm{MM}}]. \tag{4.11}$$

Thus, the price at time t is given by the long-term sum of future dividends plus a boundary term which in general is non-zero.

Stationary dividends with zero mean

If the boundary term in Eq. (4.11) evaluated at $T=\infty$ is equal to zero, i.e., the transversality condition holds, one obtains the standard EMH fundamental rational expectation pricing rule:

$$p_t = \mathbb{E}\left[p_t^{\mathrm{F}}|\mathcal{I}_t^{\mathrm{MM}}\right], \quad \text{where} \;\; p_t^{\mathrm{F}} = \mathbf{1}_{/t}^{\top}\, \boldsymbol{\mu}_{/t}. \tag{4.12}$$

In the case of a mean-reverting dividends process with zero mean, the transversality condition is justified. We will investigate the model with this assumption, for simplicity reasons. Under this prescription, the job of the MM is to provide the optimal forecast of discounted future cash flows from infinity to the present time t, given his current information set. Notice that restoring a fundamental price with a non-zero mean would simply amount to a rigid (although, infinite) shift of the price process since the mean of the fundamental price is public information and so it is immediately incorporated into the price.

It will be interesting to compare the result of the MM's estimate, given by Eq. (4.12), with the one constructed by the IT, which is not distorted by the noise induced by the NT:

$$p_t^{\mathrm{IT}} = \mathbb{E}\left[p_t^{\mathrm{F}}\middle|\,\mathcal{I}_t^{\mathrm{IT}}\right]. \tag{4.13}$$

Let us note here that the dividends have to be predictable for the market to be nontrivial. In fact, if the dividend process is not correlated, i.e., $\Xi_\tau^\mu = \Xi_0^\mu \delta_\tau$, then $p_t^{\mathrm{IT}} = 0$, i.e., the IT does not have any informational advantage over the MM. Thus, in this case, the MM would simply set the price equal to zero.

With the pricing rule given by Eq. (4.12) the MM statistically breaks even for each buy or sell trade, if he waits enough time for the income due to the dividends to restore his cash account to zero. This local constraint is thus given by:

$$\mathbb{E}[\Delta C_t^{\mathrm{MM}}] = 0. \tag{4.14}$$

As a consequence $\mathbb{E}[\Delta C_t^{\mathrm{IT}}] + \mathbb{E}[\Delta C_t^{\mathrm{NT}}] = 0$, i.e., the gain of the IT is balanced by the losses of the NT. This is what typically happens in models where NT are uninformed and non-rational [63].

In the following we give the explicit expression of the pricing rule (4.12) in terms of the IT's trading schedule, i.e., in terms of the IT's demand kernels introduced in Eq. (4.8).

Dividends regression from observed excess demand

The pricing rule given by Eq. (4.12) prescribes that the MM should estimate the sum of future dividends by observing realized excess demand. This problem can be solved in two steps. First, the MM estimates realized dividends by applying a filter on realized excess demand. The optimal estimator of realized dividends is well known in the LG-SSM literature as the Kalman filter and it is linear in the measurements, i.e., the realized excess demand in our model. Then, the MM computes the expected sum of future dividends summing over the forecasts of future dividends. In the following, we detail these two steps.

The MM's estimate of realized dividends $\hat{\boldsymbol{\mu}}_t := \mathbb{E}[\boldsymbol{\mu}_t | \mathcal{I}_t^{\mathrm{MM}}]$ is given by $\hat{\boldsymbol{\mu}}_t = \mathsf{K}\mathbf{q}_t$, where we have implicitly defined the (steady-state) Kalman gain K. This matrix can be constructed in a standard way [71, 72] given the dynamics of the MM's measurements, i.e., Eq. (4.10). The Kalman gain K is proportional to the signal noise, i.e., Ξ^μ, and inversely proportional to the measurement noise, which is the ACF of the excess demand $\Omega_\tau := \mathbb{E}[q_t q_{t+\tau}]$ and it is explicitly given by:

$$\mathsf{K} = \Xi^\mu (\mathsf{J}^\mu)^\top \Omega^{-1}, \tag{4.15}$$

where[2]

$$\begin{aligned} \mathsf{J}^\mu &= (\mathsf{I} - \mathsf{RL})^{-1} \mathsf{R}^\mu \mathsf{L} \\ \Omega &= \mathsf{J}^\mu \Xi^\mu \left(\mathsf{J}^\mu\right)^\top + \mathsf{D}^{\mathrm{NT}}. \end{aligned} \tag{4.16}$$

J^μ is the matrix that multiplies the dividends in the r.h.s. of Eq. (4.10) and D^{NT} is the NT's dressed ACF, given by:

$$\mathsf{D}^{\mathrm{NT}} = (\mathsf{I} - \mathsf{RL})^{-1} \left(\mathsf{I} + \mathsf{R}^{\mathrm{NT}}\mathsf{L}\right) \Omega^{\mathrm{NT}} \left(\mathsf{I} + \mathsf{R}^{\mathrm{NT}}\mathsf{L}\right)^\top \left[(\mathsf{I} - \mathsf{RL})^{-1}\right]^\top. \tag{4.17}$$

The noise ACF is dressed since the noise (i.e., the NT's trade process) not only affects the excess demand dynamics by construction ($\mathbf{q}_t = \mathbf{q}_t^{\mathrm{IT}} + \mathbf{q}_t^{\mathrm{NT}}$), but also because the IT's optimal trading strategy depends upon past and future realizations of the noise (see Eq. (4.8)).

[2] Using the Woodbury identity on Eq. (4.15), one obtains the alternative expression of the gain matrix K:

$$\mathsf{K} = \left[(\Xi^\mu)^{-1} + (\mathsf{J}^\mu)^\top \left(\mathsf{D}^{\mathrm{NT}}\right)^{-1} \mathsf{J}^\mu\right]^{-1} (\mathsf{J}^\mu)^\top \left(\mathsf{D}^{\mathrm{NT}}\right)^{-1}.$$

This alternative expression gives a complementary interpretation of the gain matrix K: in fact the matrix inside the square bracket is the dividends posterior information matrix. This matrix is given by the dividends prior information matrix $(\Xi^\mu)^{-1}$ summed to the information added by the measurement, i.e., $(\mathsf{J}^\mu)^\top \left(\mathsf{D}^{\mathrm{NT}}\right)^{-1} \mathsf{J}^\mu$.

From estimated realized dividends $\hat{\boldsymbol{\mu}}_t$, the MM has to estimate the fundamental price p_t^{F}, defined in Eq. (4.12). To do so, he builds the forecast of future dividends as $\mathbb{E}[\boldsymbol{\mu}_{/t}|\hat{\boldsymbol{\mu}}_t] = \mathsf{F}^{\mu}\hat{\boldsymbol{\mu}}_t$, where we introduced the dividends forecast matrix F^{μ}. Since the dividends process is Gaussian with zero-mean, F^{μ} depends only on the ACF of the dividends Ξ^{μ}. Finally, by summing over the estimated future dividends we obtain the following equation for the price at time t:

$$p_t = \mathbf{1}_{/t}^{\top}\, \mathsf{F}^{\mu}\mathsf{K}\, \mathbf{q}_t. \tag{4.18}$$

Notice, that Eq. (4.18) explicitly gives the rule for the propagator G. In fact, in compact notation, the propagator model is given by $p_t = \mathbf{1}_{/t}^{\top}\, \mathsf{G}\, \mathbf{q}_t$[3].

In the following Section, we construct the IT's optimal trading strategy based on the maximization of his expected future wealth, as a function of the MM's pricing rule. This means that, as anticipated, the IT's demand kernels $(\mathsf{R}, \mathsf{R}^{\mathrm{NT}}, \mathsf{R}^{\mu})$ are functions of the propagator G introduced in Eq. (4.1), and so is the Kalman gain matrix K introduced in Eq. (4.16). Because of this, Eq. (4.18) will turn out to be a self-consistent equation for the propagator G.

4.3.3 Optimal insider trading

The utility function U_t^{IT}, whose expectation is maximized by the IT at each time step t, is defined by the value of his wealth account at a terminal time $t+T$ (where T is not related to that introduced in Sec. 4.3.2), given by W_{t+T}^{IT}, in which his position Q_t^{IT} in the risky asset is flattened. Thus, $U_t^{\mathrm{IT}} = W_{t+T}^{\mathrm{IT}}$ subject to the constraint $Q_{t'}^{\mathrm{IT}} = 0$ for $t' \geq t+T$.

At each time step t, the IT optimizes his expected utility function over the whole future trajectory $\mathbf{q}_{/t}^{\mathrm{IT}}$ given the information set at the current time $\mathcal{I}_t^{\mathrm{IT}}$ given by Eq. (4.7), and trades the first step of the optimal strategy. The IT's trade at time t is thus calculated as follows:

$$q_t^{\mathrm{IT}} = \mathbf{e}_t^{\top} \arg\max_{\mathbf{q}_{/t}^{\mathrm{IT}}} \mathbb{E}\left[U_t^{\mathrm{IT}} \middle| \mathcal{I}_t^{\mathrm{IT}}\right], \tag{4.19}$$

where $\mathbf{e}_t^{\top}$ explicits the fact that only the first step of the future trajectory is executed. Notice that the presence of a finite liquidation time does not break the assumption of the time-translational invariance of the model, because the terminal condition is also receding as time moves on. Indeed, the IT will in general hold a non-zero position Q_t^{IT} up to $t \to \infty$ despite the presence of the liquidation constraint. The constraint should then be seen as a device used by the IT in order to properly mark to market the value of his *current* stock positions at time t by taking into account the forecast of their *future* liquidation value p_{t+T}, rather than as a measure taken to prevent him from trading at large times.

In the following, we analyze the case in which $T = \infty$ with mean-reverting dividends.

Stationary demand kernels of the insider with infinite horizon

If $T = \infty$ in Eq. (4.19), the IT can neglect the round-trip constraint, since liquidation costs are pushed to the far-away future and, due to the assumptions of zero-mean and mean-reverting dividends, the expected price at infinity is zero. Because of this, the actual trading profile of the IT that we will consider in the following is given by Eq. (4.19) with $U_t^{\mathrm{IT}} = C_{\infty}^{\mathrm{IT}}$. In doing so, the maximization program is given by

$$q_t^{\mathrm{IT}} = \mathbf{e}_t^{\top} \arg\max_{\mathbf{q}_{/t}^{\mathrm{IT}}} \mathbb{E}\left[C_{\infty}^{\mathrm{IT}} \middle| \mathcal{I}_t^{\mathrm{IT}}\right], \quad \text{where} \quad C_{\infty}^{\mathrm{IT}} = C_{t-1}^{\mathrm{IT}} - \left(\mathbf{q}_{/t}^{\mathrm{IT}}\right)^{\top}\left(\mathbf{p}_{/t} - \mathbf{p}_{/t}^{\mathrm{F}}\right). \tag{4.20}$$

[3]Compare with Eq. (4.1), where the propagator model is introduced.

In order to keep the discussion simple, we consider the dividend process with integrable ACF, such that the IT's estimate of the fundamental price p_t^{F} is finite. One can in fact relax this hypothesis, with a suitable renormalization of the price and dividends process.

Notice that the introduction of a non-zero mean for the fundamental price does not affect the IT's strategy or the price impact function. In fact, since the expectation of the fundamental price is assumed to be public information, the MM could immediately incorporate it into the price, as discussed previously. Then, since the IT's gain in Eq. (4.20) is proportional to the difference between the price and the IT's estimate of the fundamental price, it follows that the IT's trading strategy does not depend on the mean of the fundamental price. To conclude, since the propagator given by Eq. (4.18) depends only on IT's demand kernels $(\mathsf{R}, \mathsf{R}^{\mathrm{NT}}, \mathsf{R}^{\mu})$ and ACFs of dividends and NT's trades $(\Xi^{\mu}, \Omega^{\mathrm{NT}})$ via the Kalman filter (Eq. (4.15)) and the dividend forecast matrix F^{μ}, it follows that the mean of the fundamental price is immaterial in shaping the price impact function.

The expression for the demand kernels $(\mathsf{R}, \mathsf{R}^{\mathrm{NT}}, \mathsf{R}^{\mu})$ at equilibrium can be determined as the solution of the quadratic optimization program defined by Eq. (4.20). The expected gain at infinity C_{∞}^{IT} depends on estimated future dividends (via $\mathbf{p}_{/t}^{\mathrm{F}}$) and on estimated future NT's trades (via $\mathbf{p}_{/t}$). Thus, in order to write it down explicitly, we need the dividends forecast matrix F^{μ} introduced in the previous Section, and the forecast matrix of NT's trades, F^{NT}, defined similarly by $\mathbb{E}[\mathbf{q}_{/t}^{\mathrm{NT}}|\mathbf{q}_t^{\mathrm{NT}}] = \mathsf{F}^{\mathrm{NT}}\mathbf{q}_t^{\mathrm{NT}}$.

Since $\mathbb{E}[C_{\infty}^{\mathrm{IT}}|\mathcal{I}_t^{\mathrm{IT}}]$ depends on past realizations and forecasts, we insert time subscripts over matrix symbols in order to avoid ambiguities. We obtain:

$$\begin{aligned}\mathbb{E}[C_{\infty}^{\mathrm{IT}}|\mathcal{I}_t^{\mathrm{IT}}] = &-\frac{1}{2}\left(\mathbf{q}_{/t}^{\mathrm{IT}}\right)^{\top}\mathsf{G}_{/t,/t}^{\mathrm{sym}}\mathbf{q}_{/t}^{\mathrm{IT}} \\ &-\left(\mathbf{q}_{/t}^{\mathrm{IT}}\right)^{\top}\left[\mathsf{G}_{/t,t-1}\mathbf{q}_{t-1} + \mathsf{G}_{/t,/t}\mathsf{F}_{/t,t-1}^{\mathrm{NT}}\mathbf{q}_{t-1}^{\mathrm{NT}} - \mathsf{U}_{/t,/t}\mathsf{F}_{/t,t-1}^{\mu}\boldsymbol{\mu}_{t-1}\right],\end{aligned} \tag{4.21}$$

where we dropped C_{t-1}^{IT}, since it does not depend on IT's future trades $\mathbf{q}_{/t}^{\mathrm{IT}}$, and we introduced the symmetric propagator $\mathsf{G}^{\mathrm{sym}} = (\mathsf{G}+\mathsf{G}^{\top})$ in order to write in a compact form the quadratic term in $\mathbf{q}_{/t}^{\mathrm{IT}}$. The quadratic term in $\mathbf{q}_{/t}^{\mathrm{IT}}$ of Eq. (4.21) is the cost term that the IT will face due to his own future market impact, while the linear term in $\mathbf{q}_{/t}^{\mathrm{IT}}$ is his signal term. The first term of the signal comes from price impact due to known order flow realizations, the second one comes from the expected price impact of future NT's trades, while the third one comes from his private information about $\mathbf{p}_{/t}^{\mathrm{F}}$.

The expression for the IT's demand kernels, defined in Eq. (4.8), can be obtained in terms of the propagator G and the forecast matrices F^{NT} and F^{μ} inserting Eq. (4.21) in Eq. (4.20):

$$\mathbf{R}_t = -\mathbf{e}_t^{\top}\left[\mathsf{G}_{/t,/t}^{\mathrm{sym}}\right]^{-1}\mathsf{G}_{/t,t-1}, \tag{4.22a}$$

$$\mathbf{R}_t^{\mathrm{NT}} = -\mathbf{e}_t^{\top}\left[\mathsf{G}_{/t,/t}^{\mathrm{sym}}\right]^{-1}\mathsf{G}_{/t,/t}\mathsf{F}_{/t,t-1}^{\mathrm{NT}}, \tag{4.22b}$$

$$\mathbf{R}_t^{\mu} = \mathbf{e}_t^{\top}\left[\mathsf{G}_{/t,/t}^{\mathrm{sym}}\right]^{-1}\mathsf{U}_{/t,/t}\mathsf{F}_{/t,t-1}^{\mu}. \tag{4.22c}$$

Finally, we have all the ingredients to write down explicitly the functional equation for the equilibrium pricing rule, which will be given in the following Section.

4.4 The linear equilibrium

4.4.1 Equilibrium condition and numerical solution

The linear equilibrium of the model can be found by self-consistently taking into account the competitive pricing rule of the MM and the strategy of the IT, given respectively in Eqs. (4.18) and (4.22). The self-consistent equation for the propagator, in scalar notation, reads:

$$G_{t-s} = \sum_{t'=t}^{\infty} \sum_{t''=-\infty}^{t} F^{\mu}_{t',t''} K_{t'',s}[G] \tag{4.23}$$

where we have made it explicit that the filter K is a function of the propagator G itself: in fact, K is given in terms of IT's demand kernels $(\mathsf{R}, \mathsf{R}^{\mathrm{NT}}, \mathsf{R}^{\mu})$ by Eq. (4.15), which depend on the propagator G via Eqs. (4.22).

The linear equilibrium equation (4.23) is a non-linear functional equation for the propagator G_t. As such it is not amenable for analytical treatment in the general case of arbitrary Gaussian, zero-mean and stationary dividends and NT's trades process. Nevertheless, we have been able to solve Eq. (4.23) iteratively, as illustrated in Appendix B.1. In two special cases, we have been able to validate the result of the iterative numerical solver by means of the analytical solution of Eq. (4.23) (see Appendix B.2).

Via an extensive analysis of the model based on the iterative numerical solver of Eq. (4.23) we found that the market at equilibrium exhibits some robust properties, that hold in case of an integrable and stationary ACF of the NT's trades and dividends, regardless of the exact structure of the ACFs. These properties are listed below.

4.4.2 Generic equilibrium properties

Return covariance The equilibrium is characterized by a return ACF $\Xi_\tau := \mathbb{E}[\Delta p_t \Delta p_{t+\tau}]$ with the same temporal structure as that related to the IT's price estimate p^{IT}_t, given by Eq. (4.13), which will be referred to as Ξ^{IT}_τ. In formula:

$$\Xi_\tau = \Xi_0 \tilde{\Xi}^{\mathrm{IT}}_\tau, \text{ with } \tilde{\Xi}^{\mathrm{IT}}_0 = 1. \tag{4.24}$$

The price distortion induced by the noise injected into the system by the NT is thus completely encoded in a scalar, the return variance Ξ_0.

The left panels of Figs. 4.1, 4.2 and 4.3 display numerical results that do confirm Eq. (4.24). In particular, in top panels, bullet points correspond to Ξ_τ/Ξ_0 obtained by means of the numerical solver of Eq. (4.23) and show a good collapse on the dashed line, which corresponds to $\Xi^{\mathrm{IT}}_\tau/\Xi^{\mathrm{IT}}_0$ calculated semi-analytically. In the bottom part of the panels instead, we show the relative cumulative absolute error between the two curves, defined as:

$$\mathrm{err}^{\Xi}_\tau = \frac{\sum_{i=0}^{\tau} |\Xi_i/\Xi_0 - \Xi^{\mathrm{IT}}_i/\Xi^{\mathrm{IT}}_0|}{\sum_{i=0}^{\tau} |\Xi_i/\Xi_0|}. \tag{4.25}$$

In Figs 4.1 and 4.2, where non-markovian ACFs are examined, these errors are larger than in Fig. 4.3, where ACFs decay exponentially. This is due to the fact that in the former case the forecast of future dividends suffers from finite size effects. The estimation of these effects is carried on in detail in Appendix B.1.1.

The inset of the left-top panels shows the variogram of the price, defined by $V_\tau := \mathbb{E}\left[(p_t - p_{t+\tau})^2\right]$, which, as expected, is linear at high frequencies and mean-reverting at low frequencies.

Excess demand covariance The equilibrium is characterized by an excess demand ACF $\Omega_\tau := \mathbb{E}[q_t q_{t+\tau}]$ with the same temporal structure as the one related to NT's trades, plus an extra contribution at lag 0. In formula:

$$\Omega_\tau = a(\tilde{\Omega}^{\rm NT}_\tau + \tilde{b}\delta_\tau), \text{ with } \tilde{\Omega}^{\rm NT}_0 = 1, \tag{4.26}$$

where the symbol δ_τ denotes the discrete delta function, while a and b are scalars. The excess demand variance is given by $\Omega_0 = a(1+\tilde{b})$.

Since IT's information at time t does not include the current trade of the NT $q^{\rm NT}_t$ (see Eq. (4.7)), the best that the IT can do in order to hide his trades is to create a trading strategy such that the excess demand ACF resembles that of the NT apart from the lag 0 term. Because of the distortion at lag 0, we call this property *quasi-camouflage strategy*[4]. Indeed, in order to prolong his informational advantage over the MM, the IT hides his trades in the excess demand process by creating a strategy that resembles that of the NT alone.

Right panels of Figs. 4.1, 4.2 and 4.3 display numerical results that confirm the quasi-camouflage property. In top panels bullet points correspond to Ω_τ/Ω_1 obtained by means of the numerical solver of Eq. (4.23) which show a good collapse for positive lags on the dashed line, which corresponds to $\Omega^{\rm NT}_\tau/\Omega^{\rm NT}_1$. It is clear, from the insets of the plots on the left, that the collapse is not reached at lag 0. As it will be shown in the next Section, this extra contribution at lag 0 depends in a nontrivial way on the ACFs of the dividends and NT's trades. On the bottom, the relative cumulative absolute error between the two curves is presented. In this case, it starts from lag 1, so:

$$\text{err}^\Omega_\tau = \frac{\sum_{i=1}^\tau |\Omega_i/\Omega_1 - \Omega^{\rm NT}_i/\Omega^{\rm NT}_1|}{\sum_{i=1}^\tau |\Omega_i/\Omega_1|}. \tag{4.27}$$

Again, these errors are larger in the case where non-markovian ACFs are examined.

From the properties given by Eqs. (4.24) and (4.26), together with the MM's break-even condition, one is in principle able to find the propagator. In fact, introducing the price ACF $\Sigma_\tau := \mathbb{E}[p_t p_{t+\tau}]$, from the definition of the propagator (4.1) follows that:

$$\Sigma_\tau = \sum_{t'=-\infty}^{t+\tau} \sum_{t''=-\infty}^{t} G_{t+\tau-t'} G_{t-t''} \Omega_{|t'-t''|}, \text{ with } \tau > 0, \tag{4.28}$$

where the price ACF Σ_τ can be computed from Eq. (4.24) and the excess demand ACF is given by Eq. (4.26). This program can be accomplished in the case of a Markovian system and it is described in full detail in Sec. 4.6. There, we shall provide semi-analytical results for all of the parameters introduced in the equations listed above, which do share qualitative features with the general non-Markovian case. An interesting finding of this analysis is given by the fact that as the predictability of the NT's trades process increases, the IT's camouflage becomes exact, allowing him to reduce the cost due to the price impact of his trading schedule.

But before discussing the Markovian case, let us highlight similarities and differences with respect to existing models.

[4]Camouflage is also called inconspicuous strategy in the economics literature [68, 73, 74]

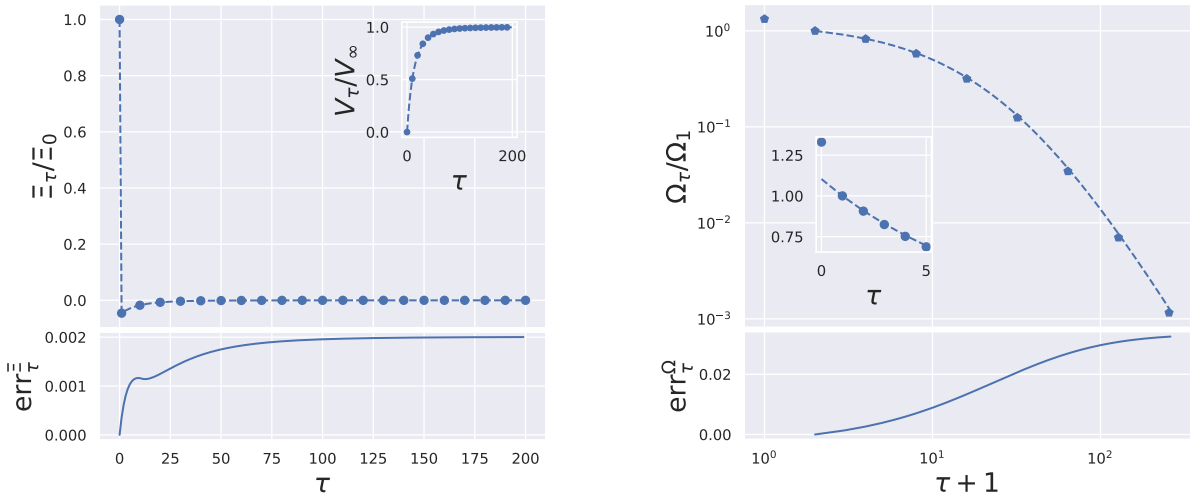


Figure 4.1: Numerical check of equilibrium properties with ACFs given by $(1+|\tau|/\tau_k)^{-\gamma_k}$ where $k=\{\mu,\mathrm{NT}\}$. We arbitrarily choose $(\tau_{\mathrm{NT}},\tau_\mu,\gamma_{\mathrm{NT}},\gamma_\mu)=(30,50,3,5)$. The numerical solver has been implemented with $T_{cut}=5\cdot 10^2$ and $T_{it}=200$. (Left) In the upper panel, we show the good collapse between Ξ_τ/Ξ_0 (bullet points) and $\Xi^{\mathrm{IT}}_\tau/\Xi^{\mathrm{IT}}_0$ (dashed line). The collapse between these two ACFs is quantified in the bottom panel, where the relative cumulative absolute error between the two curves is displayed. The inset in the top panel shows the collapse on the variogram. (Right) In the main top panel, we show the good collapse for positive lags between Ω_τ/Ω_1 (bullet points) and $\Omega^{\mathrm{NT}}_\tau/\Omega^{\mathrm{NT}}_1$ (dashed line), whereas in the inset we show that the collapse does not involve the lag 0 term. In the bottom panel the collapse between these two ACFs is quantified, calculating the relative cumulative absolute error starting from lag 1.

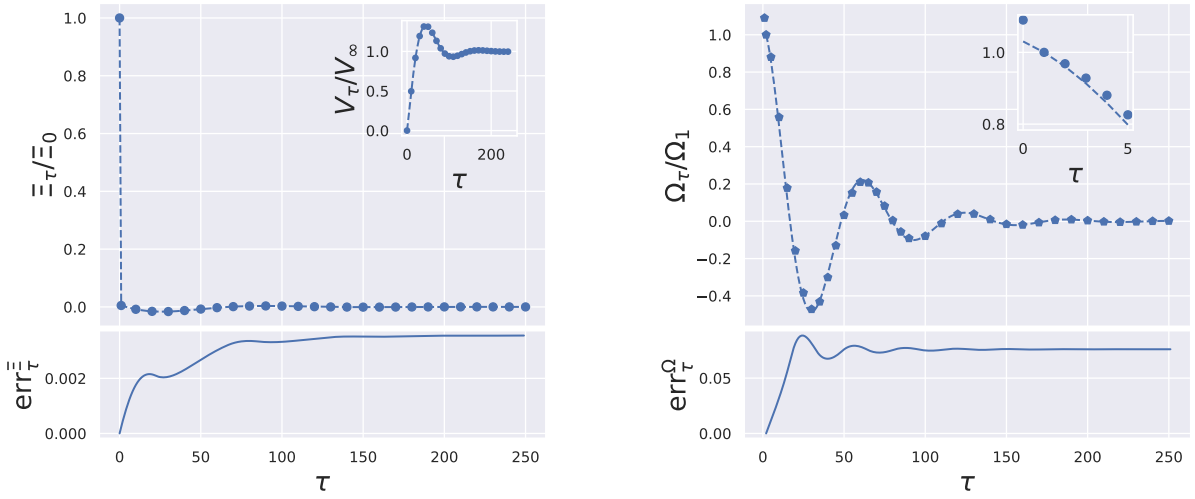


Figure 4.2: Numerical check of equilibrium properties with ACFs given by $\exp^{-\tau/\tau_{1,k}}\sin(x/\tau_{2,k}+\pi/2)$ where $k=\{\mu,\mathrm{NT}\}$. We arbitrarily choose $(\tau_{1,\mathrm{NT}},\tau_{1,\mu},\tau_{2,\mathrm{NT}},\tau_{2,\mu})=(40,40,20,10)$. The numerical solver has been implemented with $T_{cut}=10^3$ and $T_{it}=500$. (Left) In the upper panel, we show the good collapse between Ξ_τ/Ξ_0 (bullet points) and $\Xi^{\mathrm{IT}}_\tau/\Xi^{\mathrm{IT}}_0$ (dashed line). The collapse between these two ACFs is quantified in the bottom panel, where the relative cumulative absolute error between the two curves is displayed. The inset in the top panel shows the collapse on the variogram. (Right) In the main top panel, we show the good collapse for positive lags between Ω_τ/Ω_1 (bullet points) and $\Omega^{\mathrm{NT}}_\tau/\Omega^{\mathrm{NT}}_1$ (dashed line), whereas in the inset we show that the collapse does not involve the lag 0 term. In the bottom panel the collapse between these two ACFs is quantified, calculating the relative cumulative absolute error starting from lag 1.

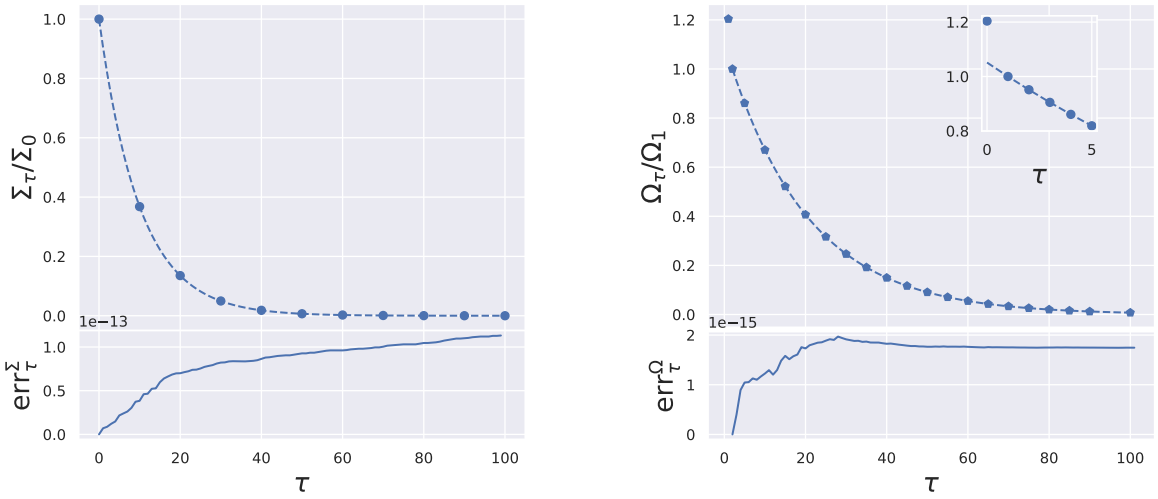


Figure 4.3: Numerical check of equilibrium properties for ACFs given by $e^{-\tau/\tau_k}$ where $k = \{\mu, \mathrm{NT}\}$. We arbitrarily fixed $(\tau_{\mathrm{NT}}, \tau_\mu) = (10, 20)$. The numerical solver has been implemented with $T_{cut} = 5 \cdot 10^2$ and $T_{it} = 200$. (Left) In the upper panel, we show the good collapse between Ξ_τ/Ξ_0 (bullet points) and $\Xi^{\mathrm{IT}}_\tau/\Xi^{\mathrm{IT}}_0$ (dashed line). The collapse between these two ACFs is quantified in the bottom panel, where the relative cumulative absolute error between the two curves is displayed. The inset in the top panel shows the collapse on the variogram. (Right) In the main top panel, we show the good collapse for positive lags between Ω_τ/Ω_1 (bullet points) and $\Omega^{\mathrm{NT}}_\tau/\Omega^{\mathrm{NT}}_1$ (dashed line), whereas in the inset we show that the collapse does not involve the lag 0 term. In the bottom panel the collapse between these two ACFs is quantified, calculating the relative cumulative absolute error starting from lag 1.

4.5 Relation with existing models

4.5.1 Kyle model

While strongly inspired by the one-period Kyle model, our model is quite different on several grounds. First, instead of exogenously postulating the presence of a fundamental price, in our setting it is the integrated-dividend process that plays the role of the fundamental price, mechanically relating it to the payoff of the asset. Second, we do not have explicit fundamental price revelation, thus allowing us to consider a stationary set in the model. Such a stationary regime is relevant in practice because in order to analyze the behavior of the market at short time scales (minutes, hours) one would like to abstract away the non-stationary effects potentially induced by the dynamics of the fundamental information (e.g, dividends, earnings announcements, scheduled news) at slower time scales. Third, we introduced (integrable) serial correlations both in the dividends – equivalently, in the fundamental price – and in the order flow.

Let us also point out how we can recover the Kyle model in our setting. Assuming that (i) the NT's trades are uncorrelated, (ii) the sum of future dividends p_t^{F} follows a random walk process, (iii) the IT knows the value of p_t^{F} at the beginning of each period and (iv) p_t^{F} becomes public information once the MM has set the price, we recover exactly an iterated version of the single period Kyle model.

4.5.2 Propagator model

Equation (4.28) is the cornerstone equation when dealing with propagator models. It is typically used in the literature in order to extract a propagator G_t from empirical data given the order flow correlation and the price volatility. Hence, our framework allows us to recover the propagator model in an economically standard setting, with three important caveats:

- The excess demand ACF Ω_τ function observed in real markets is typically non-integrable, due to the strongly persistent nature of the order flow [54, 55].
- The price process observed in real markets is close to be diffusive at high frequency.
- The propagator observed in real markets is found to be a slowly decaying function of time.

Let us address these empirical facts, showing how one can account for them within our stylized model.

First, the non-integrability of the excess demand ACF can be retrieved by extending our framework to the case in which the NT's trades ACF are themselves non-integrable. This is due to the fact that the camouflage condition relating excess demand and noise trading is also expected to extend to the setting of non-integrable NT's trades.

Second, price diffusivity also can be recovered in our model as the limiting regime in which dividends are much slower than any other time scale in the model. In order to prove this, note that the variogram of the price can be written in terms of the price ACF Σ_τ as follows:

$$V_\tau = V_\infty(1 - \tilde{\Sigma}_\tau), \text{ where } \tilde{\Sigma}_0 = 1, \tag{4.29}$$

where the first equality holds in stationary conditions, as the one described by the model introduced here. Thus, we do recover price diffusivity at high frequency if $\tilde{\Sigma}_\tau - 1 \propto \tau/\tau^*$ in the high frequency limit of the model, i.e., $\tau \ll \tau^*$, where τ^* is some typical timescale. Instead, in the opposite low-frequency limit $\tau \gg \tau^*$, because of the assumption of mean-reverting dividends, which translates into having a mean-reverting fundamental price, the price ACF decays to zero, i.e., $\Sigma_\tau \sim 0$, and we recover a flat variogram. For example, in the Markovian case described below, where the dividends ACF is an exponential decay function with timescale τ_μ, one has $\tau^* = \tau_\mu$. To wrap up, if the hypothesis of linear price ACF Σ_τ in the high-frequency limit holds, the price in our model interpolates between two very different situations: when the model is probed in its high-frequency limit it describes a market with diffusive price, while in the low-frequency limit the price is mean-reverting. This is very satisfactory since it is obtained with a single propagator, which is the solution of Eq. (4.23). At high frequency, where the dividend process appears highly persistent, price diffusivity stems from IT's surprises in dividends' variations: this is the universal mechanism that originates the diffusive behavior in our model. In fact, in this limit, the IT's estimate of the fundamental price is a martingale, and thus it is described by a diffusive process. From Eq. (4.24) follows that the price process itself is described by a diffusive process.

Since the first two properties can be retrieved, the third one follows from standard scaling arguments. Thus, in the high-frequency limit price impact has to be a slowly decaying function of time in order to ensure price diffusion while having a strongly correlated order flow process via Eq. (4.28).

It is interesting to notice that in order to observe any impact at all in the model, one is forced to introduce a non-trivial[5] dividend process: the introduction of fundamental information that gives the IT an informational advantage over the MM is enough in order to induce non-trivial dynamics into the price, and to typically induce a diffusive behavior of prices at high frequency. Hence, the price paid in order to micro-found the propagator model is the introduction of an auxiliary dividend process, whose detailed shape is inessential at high enough frequency, but whose fluctuations sets the scale of the price response.

4.6 Markovian case

Significant simplifications of the equilibrium condition (4.23) are possible in the case in which both the dividend and the NT flow are Markovian processes, where their ACFs are given by:

$$\Xi^\mu_\tau = \Xi^\mu_0 \alpha^\tau_\mu, \tag{4.30a}$$

$$\Omega^{\rm NT}_\tau = \Omega^{\rm NT}_0 \alpha^\tau_{\rm NT}. \tag{4.30b}$$

One of these simplifications comes from the fact that the price estimate $p^{\rm IT}_t$ given by Eq. (4.13) is proportional to the current dividend realization, i.e., $p^{\rm IT}_t = \mu_{t-1}\alpha_\mu/(1-\alpha_\mu)$. Thus, the price efficiency property given in Eq. (4.24) becomes:

$$\Xi_\tau = \Xi_0 \tilde{\Xi}^{\rm F}_\tau, \text{ with } \tilde{\Xi}^{\rm F}_0 = 1, \tag{4.31}$$

where $\Xi^{\rm F}$ is the return ACF of the fundamental price $p^{\rm F}_t$. From Eq. (4.31) follows that the ACF of the price process Σ_τ is a decaying exponential with timescale given by $\tau_\mu :=$

[5] See the brief discussion under Eq. (4.13)

$-1/\log(\alpha_\mu)$. As a result, the price process in the Markovian case is a discrete Ornstein-Uhlenbeck process with timescale $\tau_\mu := -1/\log(\alpha_\mu)$.

We validated the result of the iterative numerical solution exposed in the previous Section by solving explicitly the equilibrium condition in two peculiar Markovian cases: the case of non-correlated NT trades, obtained by replacing the equation for the NT's trades ACF by $\Omega_\tau^{\mathrm{NT}} = \Omega_0^{\mathrm{NT}}\delta_\tau$, and the case in which the ACF timescale of NT's trades is the same as the dividends' one, i.e., the case given by Eq. (4.30) with $\alpha_\mu = \alpha_{\mathrm{NT}}$. These findings are reported in Appendix B.2.

Furthermore, we found the explicit solution to the equilibrium condition by imposing the generic equilibrium properties listed in the previous Section, together with the MM's break-even condition given by Eq. (4.14). Details about the outcome of this procedure are given in the following Sections.

Let us point out that even though the choice of Markovian dividends and NT's trades processes are made in order to obtain analytical results and build an intuition about the system in a simple case, the main qualitative conclusions found in this Section do extrapolate to generic stationery, mean reverting processes with integrable ACFs.

4.6.1 Propagator

Non-correlated NT trades This case is particularly simple since the quasi-camouflage property given by Eq. (4.26) becomes exact. Eq. (4.28) is solved by an exponential decay propagator with the same timescale as the dividends ACF, i.e., τ_μ. The amplitude of the propagator is derived in App. B.2.1.

Correlated NT trades The solution of Eq. (4.23) is obtained in two steps. First, we build an ansatz based on the quasi-camouflage strategy property, i.e., Eq. (4.26) and the property about return ACF given by Eq. (4.24). Details about this are given in Appendix B.3.1. Then we fix the ansatz by imposing the MM's break-even condition (see Appendix B.3.2). The results of this procedure described below, do match the results of the iterative numerical solver of Eq. (4.23).

The propagator we find reads:

$$G_\tau = G_0 \left[\frac{\alpha_\mu - \alpha_{\mathrm{NT}}}{\alpha_\mu - \rho} \alpha_\mu^\tau + \left(1 - \frac{\alpha_\mu - \alpha_{\mathrm{NT}}}{\alpha_\mu - \rho} \right) \rho^\tau \right], \tag{4.32}$$

where a new timescale $\tau_\rho := -1/\log(\rho)$ appears. This new timescale is given, in the general Markovian case, by a non-linear combination of the two fundamental timescales τ_μ and $\tau_{\mathrm{NT}} := -1/\log(\alpha_{\mathrm{NT}})$ (the implicit expression for ρ and G_0 is obtained as illustrated in Appendix B.3.2). From the left panel of Fig. 4.4, it is clear that in the regime in which $\tau_\mu, \tau_{\mathrm{NT}} \gg 1$, τ_ρ approaches a value close to the time-step, i.e., $\tau_\rho \sim 1$, thus being much smaller than the two fundamental timescales. This finding and the one related to the case with non-correlated NT's trades indicate that when dividends are highly persistent ($\alpha_\mu \to 1$) the propagator exhibits a quasi-permanent component and a non-zero transient component. The former stems from the apparent persistency of the fundamental process when probed at high frequency, while the latter arises from the nontrivial predictability of the NT's trade process.

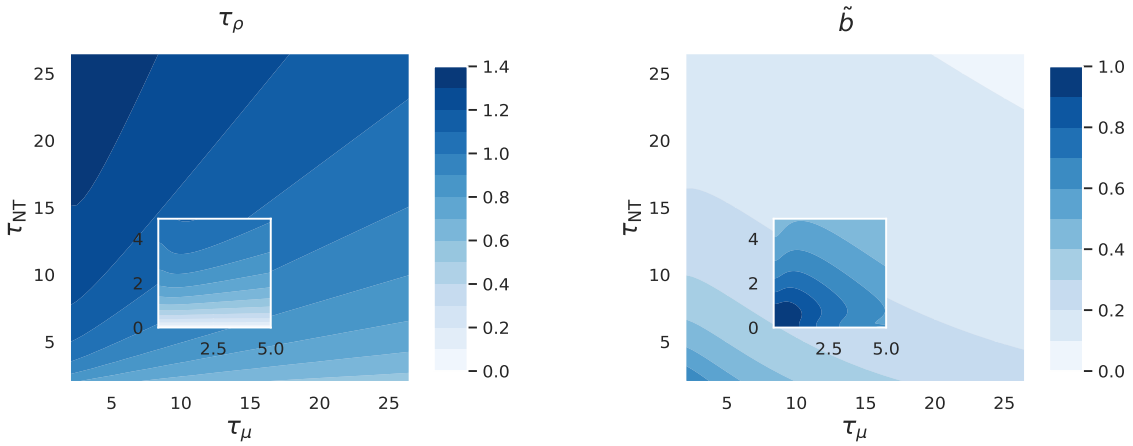


Figure 4.4: (Left) Endogenously generated timescale τ_ρ as a function of τ_μ and $\tau_{\rm NT}$. τ_ρ is never larger than ~ 2 time-step. (Right) Amplitude of the lag 0 contribution to Σ_τ, i.e., $\tilde{b}$ introduced in Eq. (4.26), as a function of τ_μ and $\tau_{\rm NT}$. As one can see in the inset, $\tilde{b}$ attains its maximum value for small timescales, while it decreases to zero as τ_μ and $\tau_{\rm NT}$ increase, thus recovering exact camouflage for the IT strategy.

As we shall see below, the large timescale behavior of τ_ρ is related to the behavior of the excess demand ACF distortion at lag 0, i.e., $\tilde{b}$, introduced in Eq. (4.26). In fact, in the derivation of Eq. (4.32) (see Appendix B.3.1) one finds:

$$\tilde{b} = \frac{\rho(1-\alpha_{\rm NT}^2)}{\alpha_{\rm NT}(1+\rho^2) - \rho(1+\alpha_{\rm NT}^2)}. \tag{4.33}$$

In the right panel of Fig. 4.4 we display $\tilde{b}$ as function of τ_μ and $\tau_{\rm NT}$. This amplitude is close to 1 in the limit of small dividends and NT's trades timescales and decreases to zero as these increase. Thus, the excess demand ACF temporal structure resembles more and more the NT's one as soon as the NT's trades or dividends are strongly correlated.

The interpretation of this finding is the following: the IT wants to hide his own trades in the excess demand process, by shaping the ACF to resemble the one of NT's trades. However, the IT knows only up to time $t-1$ the realization of the NT's trades process (see Eq. (4.7)). If this process is only weakly correlated, the IT's information about it does not allow a good prediction of NT's trade at time t. Therefore, the IT is not able to hide his current trade. Instead, if the NT's trades are strongly correlated, the IT's information about NT's past trades allows him to accurately predict the current NT's trade, and thus the IT is able to hide his current trade. Briefly, we find that:

$$\Omega_\tau \to \Omega_0 \tilde{\Omega}_\tau^{\rm NT} \text{ as } \alpha_{\rm NT} \to 1, \tag{4.34}$$

thus recovering an exact camouflage trading strategy of the IT, exhibited by many Kyle-like models [68, 74–76].

The limit $\alpha_{\rm NT} \to 1$ and $\alpha_\mu \to 1$ can be interpreted as the continuum limit of our discrete model. In this case, using $\Omega_\tau = \Omega_0 \tilde{\Omega}_\tau^{\rm NT}$, and Eq. (4.31) in continuous-time one can solve the continuous-time analog of Eq. (4.28), finding:

$$G_\tau = G_0 \left(\delta_\tau + \frac{\tau_\mu - \tau_{\rm NT}}{\tau_\mu \tau_{\rm NT}} e^{-\tau/\tau_\mu} \right). \tag{4.35}$$

From this equation, we can see that the term in the propagator that depends on the endogenously generated timescale (see Eq. (4.32)) approaches a Dirac delta function in the continuum limit of the model, as a result of the IT's exact camouflage strategy.

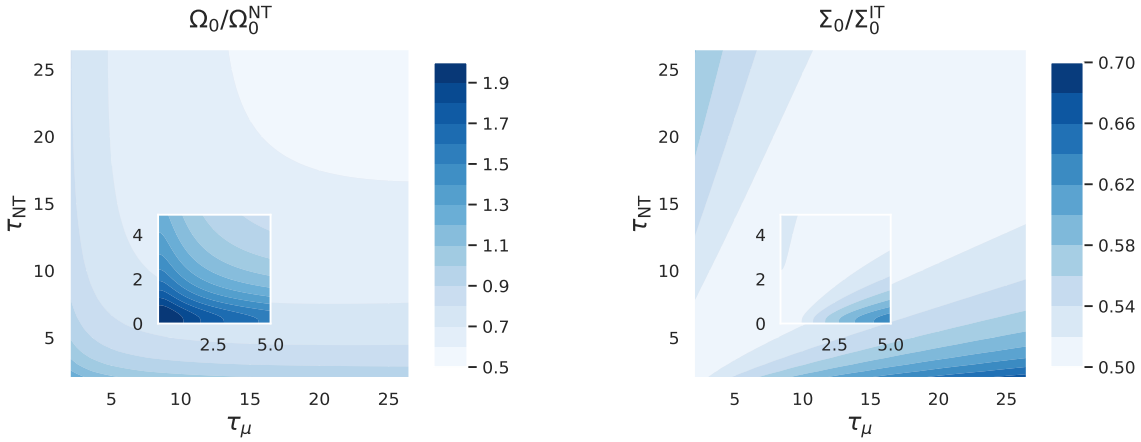


Figure 4.5: (Left) Ratio between the variance of the excess demand and variance of the NT's order flow as a function of τ_μ and $\tau_{\rm NT}$. When the timescale τ_μ and $\tau_{\rm NT}$ are small (inset), the excess demand is higher than the one of the NT's trades alone. Conversely, when the timescales τ_μ and $\tau_{\rm NT}$ are large, the excess demand variance is lower than the one of the NT's alone. (Right) Ratio between the variance of the price and the variance of the IT's fundamental price estimate as a function of τ_μ and $\tau_{\rm NT}$. The variance ratio, in this case, is very small when τ_μ is close to zero, while it increases as τ_μ increases.

4.6.2 Excess demand variance

The result for the ratio $\Omega_0/\Omega_0^{\rm NT}$ as a function of τ_μ and $\tau_{\rm NT}$ is presented in the left panel of Fig. 4.5. The variance ratio is bounded between 2, for small timescales, and 0.5, for large timescales. The increase of the ratio of variances, $\Omega_0/\Omega_0^{\rm NT}$, for small $\tau_{\rm NT}$ can be understood as follows. In this regime, the NT's current trade is almost unpredictable, thus the IT's current trade is independent of the current trade of the NT. As a consequence, the excess demand variance increases with respect to the NT's variance. As soon as the NT component of the order flow is predictable, the IT uses this information.

In particular, the IT's current trade is on average anti-correlated with the current NT trade. This enables the IT to move less the price, founding liquidity in the NT's trade and reducing the typical aggregate volume demanded by the MM. When the predictability of the NT's trades and dividends process increase, the current IT's trade is more anti-correlated with the current NT's trade, thus enabling him to lose less money due to price impact. The current IT's trade is instead positively correlated with the current dividend. Fig. 4.6 shows these findings.

4.6.3 Price variance

In our model price variance is directly linked to price efficiency, as argued below Eq. (4.24). As already noted by Shiller, in a Rational Expectation Model where the price is the expected fundamental price, using the principle from elementary statistics that the variance of the sum of two uncorrelated variables is the sum of their variances, one then has $\Sigma_0/\Sigma_0^{\rm IT} \leq \Sigma_0/\Sigma_0^{\rm F} \leq 1$, where $\Sigma_0^{\rm F}$ is the variance of the fundamental price.

We display the results for the ratio $\Sigma_0/\Sigma_0^{\rm IT}$ in the right panel of Fig. 4.5, as a function of the dividends and NT's timescales, which confirm the fundamental constraint exposed before. Moreover, we find that the ratio of variances strongly depends on τ_μ. In particular, if the dividends are weakly correlated the price variance poorly reflects the IT's price estimate

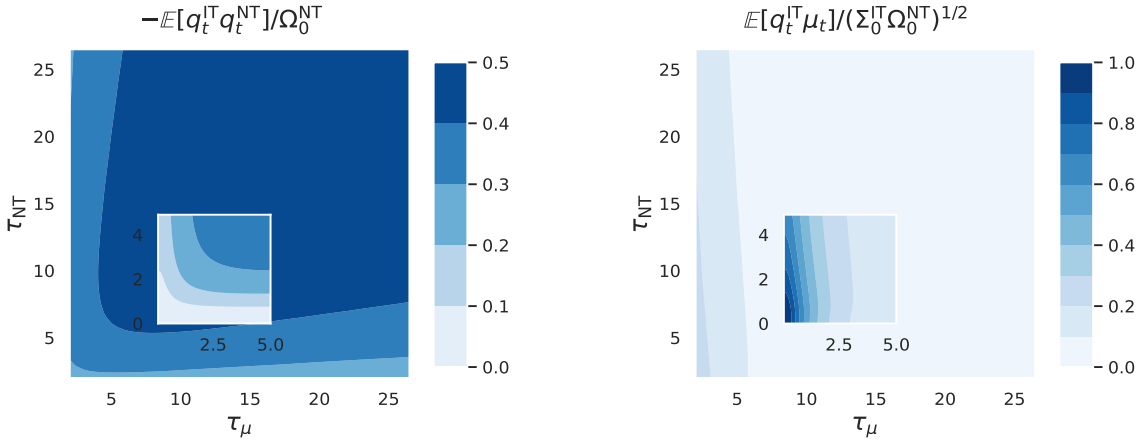


Figure 4.6: (Left) Ratio of the covariance between equal-time IT's and NT's trades, and the variance of NT's trades as a function of τ_μ and τ_{NT}. The IT's trades are anti-correlated with the equal-time NT's trades. (Right) Properly rescaled covariance of current IT's trade and dividend as a function of τ_μ and τ_{NT}. The IT's trades are on average positively correlated with the equal time dividend. When the predictability of the NT's trades and dividends process increase, the current IT's trade is more positively (negatively) correlated with the current dividend (NT's trade), thus enabling him to gain more (lose less).

variance Σ_0^{IT}. Instead, in the limit of large dividend timescales with respect to the one of the NT's trades, the price variance better reflects the IT's price estimate p_t^{IT}. In the regime of small τ_{NT} and large τ_μ the price variance accounts for all the variance of the IT's price estimate, Σ_0^{IT}, as indeed found analytically from the calculations reported in Appendix B.2.1.

4.6.4 Payoffs and market-making risk

As explained around Eq. (4.14), the payoff of the different agents is, on average, the following: the MM breaks even, the NT loses and the IT gains what the NT loses.

If the dividend process is completely unpredictable (but still stationary with zero-mean), then the price is set to zero by the MM; thus the IT won't trade anymore and the NT's losses are reduced to zero. When the τ_μ becomes large with respect to τ_{NT} (bottom right corner of main left panel of Fig. 4.7), the price is more and more efficient as we have seen in the previous Section. In this case, the IT's gains are lowered, as well as the NT's losses. These findings are reported in the left panel of Fig. 4.7, where we plot the ratio $-\mathbb{E}[\delta^{q_t^{\mathrm{NT}}} C_t^{\mathrm{NT}}]/(\Xi_0^\mu \Omega_0^{\mathrm{NT}})^{1/2}$, with $\delta^{q_t^{\mathrm{NT}}} C_t^{\mathrm{NT}} = -q_t^{\mathrm{NT}} \left(p_t - \sum_{t' \geq t} \mu_{t'}\right)$.

Another interesting quantity is the risk per trade experienced by the MM, i.e., $r_t^{\mathrm{MM}} = \mathbb{E}[(\delta^{q_t} C_t^{\mathrm{MM}})^2]$, where $\delta^{q_t} C_t^{\mathrm{MM}} = q_t \left(p_t - \sum_{t' \geq t} \mu_{t'}\right)$. We find:

$$r_t^{\mathrm{MM}} = \mathbb{E}[q_t^2] \; \mathbb{E}\left[\left(p_t - p_t^{\mathrm{IT}}\right)^2\right], \tag{4.36}$$

where we used the break-even condition (Eq. (4.14)) together with Wick's theorem to calculate higher-order correlations of a Gaussian process. The analytical solution is given in the right panel of Fig. 4.7. As we can see, the risk experienced by the MM is high when both the timescales of the two fundamental processes are small, while it decreases when both the dividends and the NT's trades become predictable.

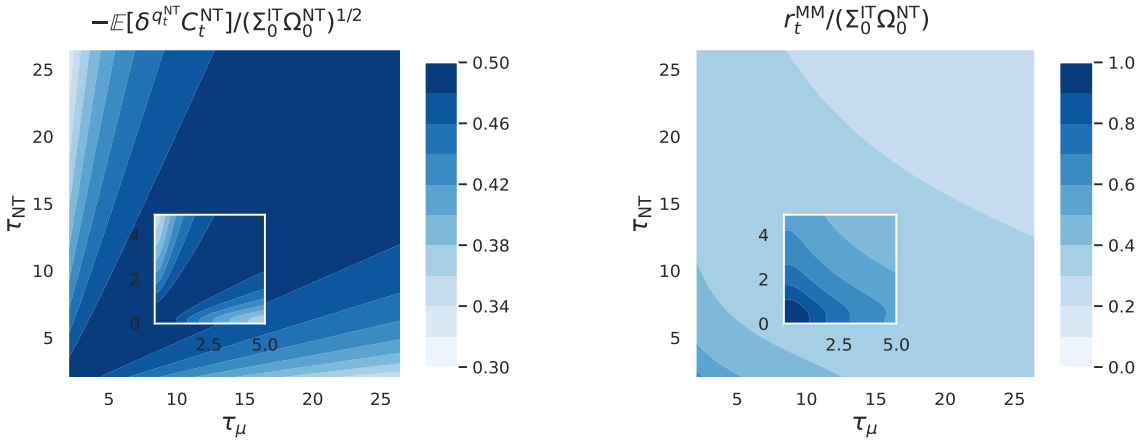


Figure 4.7: (Left) Properly rescaled loss per trade of the NT (or gain per trade of the IT) as a function of τ_μ and τ_{NT}. When τ_μ is close to zero (inset) the loss per trade of the NT is close to zero, while these increase as the predictability of the dividends and NT's trades increase. (Right) Properly rescaled MM's risk per trade as a function of τ_μ and τ_{NT}. From the inset, we can see that the risk is higher when the NT's trades and dividends are close to being unpredictable, whereas the risk is lower as the predictability increase.

4.7 Conclusion

The aim of this Chapter was to provide an economically standard microfoundation for linear price impact models, customarily used in the econophysics literature. To do so, we presented a multi-period Information-Based Model and we analyzed its equilibrium. The model is built by generalizing the seminal Kyle model, which constitutes a theoretical cornerstone of market microstructure. First, we removed the assumption of fundamental price revelation, assuming that a stock pays dividends to the owner but only the insider collects and exploits information about past dividends. Then, we modeled the dividends process and the noise trader trading schedule as stationary stochastic processes. In order to regularise the model, we assumed that the dividend ACF was integrable, to ensure a bounded fundamental price of the traded stock. The model appeared to exhibit a stationary equilibrium, which we have investigated in detail. A self-consistent equation for the pricing-rule set by the market-maker has been derived and solved numerically. Two robust properties have been found: the price ACF retains the same temporal structure as the insider's fundamental price estimate and the insider strategy respects a quasi-camouflage condition, i.e., the ACF of the excess demand retains the temporal structure of the noise trader's one apart from the lag 0 term.

As a consequence of these findings, we have been able to establish a precise correspondence between the propagator model and the Kyle model: the propagator model arises here as the high-frequency limit of a suitably stationarized Kyle model. The price impact function that is found in this regime displays a quasi-permanent component related to the timescale of variation of the fundamental information, and a transient one whose timescale is set by the persistence of the order flow.

The assumption of stationary dividends with integrable ACF translates into having a mean-reverting price process. Since price diffusivity can be retrieved in the high-frequency limit, the model is able to provide a stylized picture of what happens in real markets at high and low frequencies. The model alludes also to a relation between the diffusion constant of the price process and the timescale over which the fundamental price mean reverts. We leave the empirical check of this finding as an interesting follow-up of the present investigation.

The minimal model exposed here can be extended in several different ways.

First, further elements of realism (risk-aversion, spreads) could be integrated progressively, in order to see how much our main qualitative findings are impacted by these effects. Similarly, the non-linear, concave nature of impact [77] should be reconciled with our stylized, linear vision of the market.

The passive agent can be promoted to a rational agent that tracks a given target portfolio, introducing an element consistent with actual actions in real-world markets, which could create long-range correlated order flow.

Dividend revelation in real markets is infrequent, so another extension of the model proposed here is to take explicitly this fact into account, similarly to what is done in Ref. [78], where market closure is explicitly taken into account.

Finally, in order for our model to be able to account for the excess volatility puzzle, we need to relax the assumption on either rationality or information used by the agents that populate our universe. This can be done in different ways: for example, as in Ref. [79], we can relax the assumption of perfect structural knowledge: for example, we can assume that the agents do not know all the parameters that define the dividend process and they try to infer these starting from actual observation. Another interesting path would be, along the line of Ref. [80], to assume that the agents decide their demand according to a misspecified equation of motion for the price.

Take home messages from Chapter 4

1. It is possible to microfound the propagator model starting from a Kyle model in which the assumption of fundamental price revelation at some terminal time is relaxed.

2. The self-consistent equation for the price impact function which appears in the standard Kyle model is replaced by a functional fixed-point equation in the Stationary Kyle model.

3. The price impact function in the stationary equilibrium is nothing but the steady state Kalman filter associated with the market maker's filtering problem related to extracting the information about the fundamental price from the excess demand.

4. The functional fixed point equation for the price impact is analytically solved in the markovian case and numerically solved in the case of general ACF related to dividends and noise trades.

5. Irrespectively of the chosen ACF related to dividends and noise trades we found two robust properties related to the excess demand ACF, as well as the price ACF: the excess demand ACF has the same temporal structure as the ACF related to noise trades, meaning that the insider camouflages his trades such that he optimizes the profits stemming from his informational advantage over the market maker. On the other hand, the market maker is able to construct a price that has the same ACF's temporal structure as the one related to the fundamental price.

6. The coefficient between the price ACF and the fundamental price ACF is smaller than one, consistently with the hypothesis of rational traders and to Shiller's volatility bound (see Eq. (1.3)), but at variance with empirical analysis related to excess volatility.

7. Is it possible to build, within the Kyle framework, a model able to be consistent with excess volatility found in empirical data? We provide an answer to this question in Chapter 6, while in the following Chapter we calibrate the Stationary Kyle model against real data, providing an original piece of evidence against the Efficient Market Hypothesis.

Chapter 5

Where the strong rationality assumption fails

People know the price of everything and the value of nothing

Oscar Wilde, Lady Windemere's Fan, 1982

We compare the predictions of the stationary Kyle (s-Kyle) model, a microfounded multi-step linear price impact model in which market prices forecast fundamentals through information encoded in the order flow, with those of the propagator model, a purely data-driven model in which trades mechanically impact prices with a time-decaying kernel. We find that, remarkably, both models predict the exact same price dynamics at high frequency, due to the emergence of universality at small time scales. On the other hand, we find those models to disagree on the overall strength of the impact function by a quantity that we are able to relate to the amount of excess volatility in the market. We reveal a crossover between a high-frequency regime in which the market reacts sub-linearly to the signed order flow, to a low-frequency regime in which prices respond linearly to order flow imbalances. Overall, we reconcile results from the literature on market microstructure (sub-linearity in the price response to traded volumes) with those relating to macroeconomically relevant timescales (in which a linear relation is typically assumed).

Keywords: Market Microstructure, Price Impact, Calibration, Multi-Scale Analysis.

From:
Do fundamentals shape the price response? A critical assessment of linear impact models
Michele Vodret, Iacopo Mastromatteo, Bence Tóth, Michael Benzaquen

Contents

5.1 Introduction

Financial markets display very different dynamic properties depending on the time window at which they are probed. If markets are observed in a small enough time window, where the price exhibits diffusive dynamics, then the slow evolution of fundamentals plays a minor role: in this regime price dynamics is mostly driven by endogenous variables, such as past trades [81]. Conversely, by inspecting the market over larger time scales (e.g. months, years), fundamentals begin to play a more important role, and effects of mean-reversion to some notion of "fair value" start being visible [41, 82]. Although the literature provides several phenomenological models capable of providing a multi-scale view of financial markets [41, 83], no unique theoretical scenario is able to accommodate this crossover in a way that is both empirically accurate and economically sound.

Seminal contributions in the market microstructure literature [63] include noisy rational expectation models [18, 84] where the price partially reflects the underlying fundamental value of the asset, providing an explanation of trade-induced price impact based on a price discovery mechanism. Unfortunately, some assumptions customarily made in this type of models make them unsuitable for calibration in real markets, where several aspects diverge from their idealized counterparts: in real markets, no fundamental price revelation is provided at any time and the order flow process exhibits long-range correlations [53]. Some progress has been made in order to bridge the gap between microfounded price impact models and actual markets. In the Speculative Dynamics model (see [68]) the assumption of fundamental price revelation is dropped and a stationary price impact model is obtained. A further extension is provided by the s-Kyle model [85], which considers arbitrary Gaussian signal and noise processes, de-facto including the empirically relevant case of strongly correlated order flow. In Ref. [85] it is also highlighted that the stationarity property allows to capture qualitatively the different dynamical regimes discussed above.

Interestingly, in the high-frequency regime of the s-Kyle model, where fundamentals are slowly varying and the price is diffusive, one recovers a linear equilibrium formally equivalent to that customarily described by the propagator model [54] (see also Ref. [86]), which is an agnostic (i.e., not microfounded) model able to provide a statistically accurate picture of financial markets at high frequency. Nevertheless, as we shall see in this Chapter, from a practical point of view, these two models are very different. In fact, the excess-volatility puzzle [15] cannot be solved within the s-Kyle model. This puzzle is instead avoided by the propagator model, which has no *a-priori* on what the right price level should be, and is thus allowed to set the level of price response by calibrating the model directly with the empirical price. Although the s-Kyle model is not able to provide a solution to the excess-volatility puzzle, we shall argue that at small timescales, that has no effect on the qualitative price dynamics, because it only leads to an impact that is off by a multiplicative factor. Thus, the s-Kyle model provides a microfoundation for the structure of the price impact shape, even though it misses a magnitude component that relates to the excess-volatility puzzle.

The outline of this Chapter is the following. In Section 5.2 we show how the s-Kyle model captures the same shape of the price impact function as the propagator model, at high frequency. Section 5.3 is devoted to show how one can calibrate the stationary Kyle model against empirical data, while Sec. 5.4 contains the results of the calibration against 150 years of S&P-500 data. Finally, in Section 5.5 we conclude, suggesting an interesting way to reconcile the microfounded model with empirical price volatility[1].

[1]Since it differs from the focus of the Chapter, the reader will find in App. C a discussion related to the importance of the sampling scale at which markets are probed. In fact, it is known from the propagator

5.2 Price impact function in the stationary Kyle model

5.2.1 High- and low- frequency regimes

The price efficiency property (Eq. (4.24)) allows us to identify two different dynamical regimes according to the mean-reversion time scale related to the IT's estimate of the fundamental price, given by τ^{F}. In fact, the variogram $V_n = \mathbb{E}[(p_{n+m} - p_m)^2]$ calculated with a sampling scale τ can be expressed (in non-pathological cases) as

$$V_n = 2\Sigma_0 \left(1 - \frac{\Sigma_n^{\mathrm{IT}}}{\Sigma_0^{\mathrm{IT}}}\right) = \begin{cases} \sigma^2 n & n \ll \tau^{\mathrm{F}}/\tau \\ 2\Sigma_0 & n \gg \tau^{\mathrm{F}}/\tau \end{cases}, \tag{5.1}$$

for some positive σ, implying that the behavior of the price at time step n is diffusive when innovations in the fundamental price process are long-lived with respect to $n\tau$ and mean-reverting in the opposite case.

These two regimes can be equivalently characterized by the response function $R_n = \mathbb{E}[q_n(p_{m+n} - p_{m-1})]$, which is constant for $n \ll \tau^{\mathrm{F}}/\tau$ as a consequence of price diffusion, and decays for $n \gg \tau^{\mathrm{F}}/\tau$ as a consequence of price mean reversion. The model is thus consistent with the roughly flat empirical price response function reported in studies about high-frequency dynamics [54] for $n\tau$ smaller than a few days since τ^{F} is of the order of several months/few years [41].

Understanding the behavior of the price impact function G_n in the high-frequency regime is trickier, as its shape depends on the NT's order flow ACF. Let us consider two cases. (a) If the noise trader's order flow is uncorrelated, the informed trader also trades in an uncorrelated fashion (due to the camouflage, Eq. (4.26)). In this case, the shape of the response and impact functions is identical in the high-frequency regime, because only a permanent price impact function ensures diffusive prices in case of uncorrelated trade flow. (b) Now, let us consider a noise trader's order flow auto-correlated over a time window given by τ^{NT}. In this case, in order for the response function to be flat at high frequency, the impact function has to be a decreasing function [54], meaning that the price response to the first trade anticipates the correlated flow in the same direction. The impact function will then relax to a non-zero quasi-permanent value, even though the flat behavior of the response function will be preserved thanks to a stream of correlated orders of length $\sim \tau^{\mathrm{NT}}$. We call this impact quasi-permanent because it appears as a permanent one to any observer probing the market at scales $n \ll \tau^{\mathrm{NT}}/\tau \ll \tau^{\mathrm{F}}/\tau$. Obviously, at long times, i.e., $n \gg \tau^{\mathrm{F}}/\tau \gg \tau^{\mathrm{NT}}/\tau$, all impact functions will decay to zero by construction due to the mean-reverting nature of the IT's estimate of the fundamental price.

The qualitative picture in real markets at high frequency is very similar to case (b). In fact, the order flow displays long-term correlations. More precisely, the order flow ACF is not integrable (auto-covariance is measured to slowly decay even across days). Because of this, in order to have diffusive prices and flat response at high frequency, the impact function should slowly decay to zero in a precise non-integrable way in order to compensate for the persistence of the order flow [54]. Graphical details about cases (a) and (b) in the high and in the low-frequency regime are provided in Subsec. 5.2.3.

model literature, that if one studies financial markets at the trade by trade level, microstructural effects related to order book details (such as selective liquidity taking) are of high relevance. In particular, it is well-known [54] that a sub-linear price impact model gives higher predictive power than a linear one. We show that when analyzing coarse-grained data the opposite is true. In doing so, we provide a useful recipe to relate descriptions obtained with linear price impact models when the sampling scale is varied.

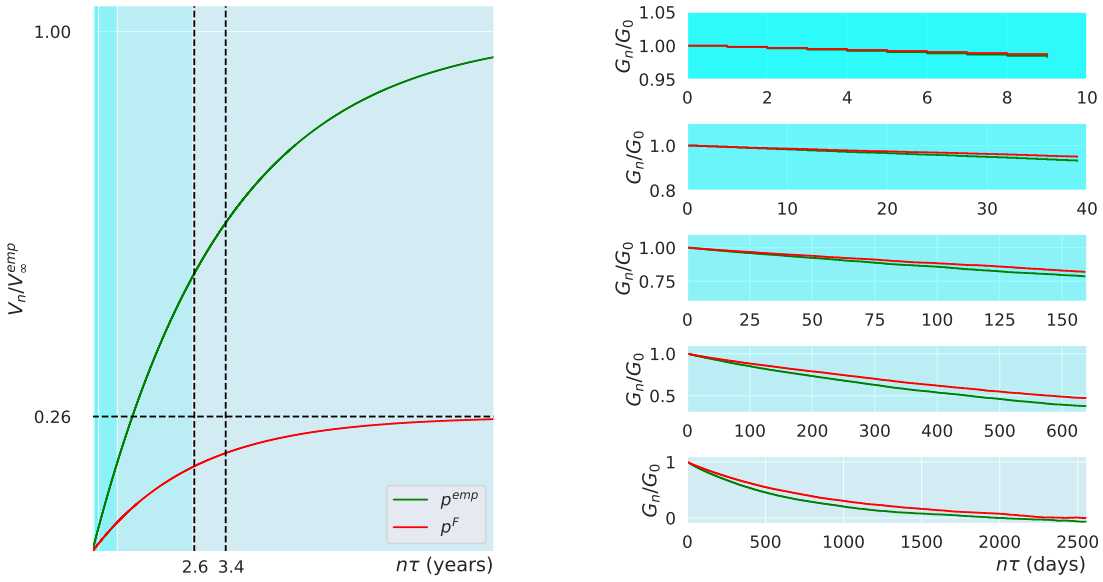


Figure 5.1: Universal time-dependence at high frequency and non-universal time-dependence at low frequency, for the case of a Markovian market with independent order flow. (Left) Variograms from synthetic data come from a Markovian market with parameters given in Table 1. The different time windows used in the calibration (right panel) are highlighted by different shades of blues.

5.2.2 Universality at high frequency

The discussion above hints at the fact that a kind of universality emerges in the s-Kyle model: the specific dynamics of the fundamental price process affects the short-term price dynamics only through the diffusion constant σ, but does not shape the short-term impact function.

This statement is tested with synthetic data as follows. We generate two synthetic datasets, with sampling scale $\tau = 1$ day, that mimic Markovian price dynamics (with mean-reversion timescales and price volatility in line with those that will be presented in Table 1, i.e., $\tau^{\text{emp}} = 250 \times 3.4$ and $\tau^{\text{F}} = 250 \times 2.6$ days and $V_\infty^{\text{F}}/V_\infty^{\text{emp}} = 0.26$), while order flow data are realizations of a non-correlated stochastic process. Variograms associated to these price processes are shown in the left panel of Figure 5.1. We then calibrate these data with different time windows, graphically identified by different blue bands: regions where the price undergoes pure diffusive dynamics are highlighted by an intense shade of blue, and as the dynamics becomes affected by price mean reversion, the blue band is lighter. One can observe from the right panel of Figure 5.1 that, modulo a global prefactor (absorbed by dividing the propagator function by its value at lag zero), the two price dynamics that we used (p^{F} and p^{emp}) induce the same behavior for $n\tau \ll \min\{\tau^{\text{emp}}, \tau^{\text{F}}\}$ given approximately by a constant (or quasi-permanent) price impact function. On the other hand, we see that as we approach a time window comparable with τ^{emp} or τ^{F}, price impact functions start to drift away from each other, because they become sensitive to the different mean-reversion price dynamics.

Even though the collapse between price impact functions at high frequency takes place by construction (apart from an amplitude related to price volatility), the non-trivial thing that we are able to predict is up to which point the collapse holds, and how it relates to effects linked to price mean-reversion dynamics at low frequencies. Therefore, the shape of

the price impact function at high frequency is only affected by the shape of the order flow ACF, i.e., it is completely determined by the dynamics of flow anticipation.

5.2.3 Linear price impact functions from different price dynamics

Here, analyzing synthetic data, we provide graphical details about cases (a) and (b) mentioned in Subsec. 5.2.1. Two calibrations with different sampling scales (τ_{short} and τ_{long}) are analyzed for each case. The calibration with τ_{long} is done on a time window that encompasses the low-frequency regime. Because of this, fundamental price mean reversion needs to be taken into account and we calibrate the s-Kyle model (with the numerical scheme detailed in Ref. [85]) starting from NT's order flow q^{NT} and it is fundamental price estimate p^{IT} ACFs. Conversely, the calibration with sampling scale τ_{short} is restricted only to the high-frequency regime (mimicking what is done usually when analyzing empirical data at high frequency). In this case, the numerical scheme used to solve the s-Kyle model cannot be applied because the model is not properly regularized, because ACFs do not have enough time to decay to zero. Because of this, we calibrate the model defined at sampling scale τ_{short} with the technique usually employed in the propagator model literature (details about it are given in App. C.3.1). Figure 5.2 shows that if the price and the order flow processes obtained at scale τ_{long} are compatible with those defined at sampling scale τ_{short}, the results of the two calibrations are compatible (via the coarse-graining argument explained in Subsec. C.1.1). In particular, the price impact functions calibrated with the two different sampling scales are compatible (see the bottom-left panels in Fig. 5.2). As a consistency check, note that calibrations related to case (b), where excess demand ACF's shape is a power law, validate the well-known constraint between the exponents of the power laws related to the excess demand ACF (β) and to the price impact functions (γ), given by $\gamma = (1-\beta)/2$ [54].

5.3 Datasets, detrending and calibration procedures

5.3.1 Low frequency

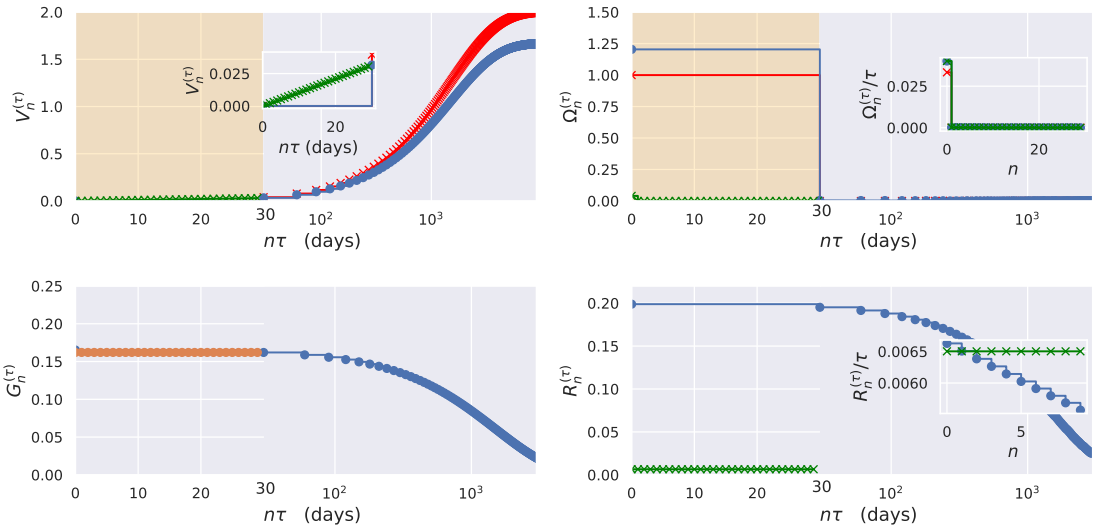
Presentation of the data. The dataset used for the calibration is about the S&P500 index and is publicly available online at github.com/datasets/s-and-p-500. These data include information about monthly ($\tau = 1$ month) prices (P_n^{emp}) and dividends (M_n) from January 1871 until March 2018, but do not include information about order flows.

From data about dividends, an estimate of the fundamental price can be constructed as $P_n^{\text{F}} = M_n \langle P_n^{\text{emp}}/M_n \rangle$[2]. The price and the fundamental price cannot be described by a stationary process, as one can see from the left panel of Figure 5.3: a clear trend is exhibited by both processes. Thus, we cannot calibrate the s-Kyle model on these raw data. One needs to de-trend them. The de-trending procedure's goal is to retrieve a stationary empirical price and fundamental price processes, so that the calibration of the s-Kyle model becomes possible.

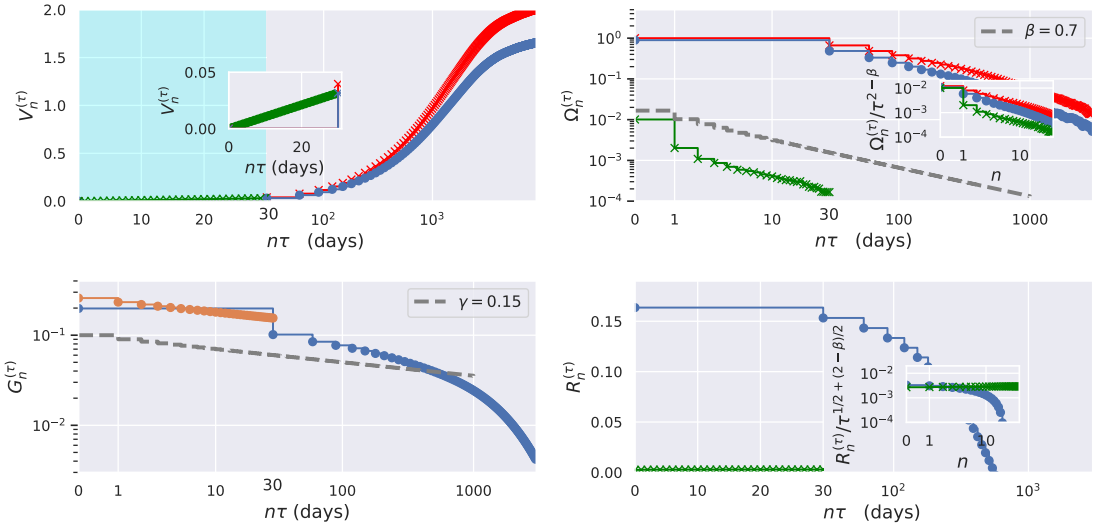
We define de-trended quantities as:

$$x_n = \frac{X_n}{X_0} \exp\left(- \sum_{m=0}^{n} \eta_m^X \right), \tag{5.2}$$

[2]Let us note that this prescription for the fundamental price is not causal since the mean price-dividend ratio $\langle P_n/M_n \rangle$ is calculated with all the data provided by the dataset.



(a) Independent order flow. Informed trader's estimates of the fundamental price are Markovian, with a mean-reversion time scale $\tau_{\mathrm{F}} = 50$ days. At sampling scale τ_{long}, the variance of the IT's estimate of the fundamental price is fixed to one, as well as the variance of the NT's order flow.



(b) Power law order flow. Informed trader's estimates of the fundamental price are Markovian, with mean-reversion time scale $\tau_{\mathrm{F}} = 50$ days and NT's order flow ACF is a combination of exponentials that mimics a decreasing power law function with exponent $\beta = 0.7$. At sampling scale τ_{long}, the variance of the IT's estimate of the fundamental price is fixed to one, as well as the variance of the NT's order flow. Dashed grey lines in top-right and bottom-left panels refer to decreasing power law functions with exponent β and γ, respectively.

Figure 5.2: Calibrations related to cases (a) and (b) mentioned in Subsec. 5.2.1, on synthetic datasets. For each case, we consider two models with different sampling scale, i.e., τ equal to $\tau_{\mathrm{short}} = 1$ day and $\tau_{\mathrm{long}} = 30$ days. We identify the high-frequency regime as the interval $[0, \tau_{\mathrm{long}}]$ (orange band). The set of input ACFs that specify the s-Kyle model at sampling scale τ_{long} are related to NT's trades and to it is fundamental price estimates (red lines). Details about them are given in sub-captions. Once the s-Kyle model is solved, we obtain $V_n^{(\tau_{\mathrm{long}})}, \Omega_n^{(\tau_{\mathrm{long}})}, G_n^{(\tau_{\mathrm{long}})}$ and $R_n^{(\tau_{\mathrm{long}})}$ (blue lines). The model with sampling scale τ_{short}, is calibrated with new input ACFs related to price and excess demand, and with the associated response function (green lines). Note that in this case the response function is not an output of the model and so it is again presented as a green line. These new ACFs are such that, after a proper rescaling (see App. C.1.1), the variogram, the excess demand ACF, and the response function match the one at low frequency, as shown in the insets. Results of calibrations with sampling scale τ_{short} are given by price impact functions (orange lines), i.e., $G_n^{(\tau_{\mathrm{short}})}$.

where the trend $\eta_n^X > 0$ is estimated in a causal way as follows.

De-trending. For each time series X_n we define the trend in a causal way, as follows:

$$\eta_n^X = \frac{1}{T/\tau} \log\left(\frac{X_n}{X_{n-T/\tau}}\right), \tag{5.3}$$

where we choose $T = 20$ years in order to be able to capture phenomena that occur on time scales as large as a few years, such as the mean-reversion of the fundamental price.

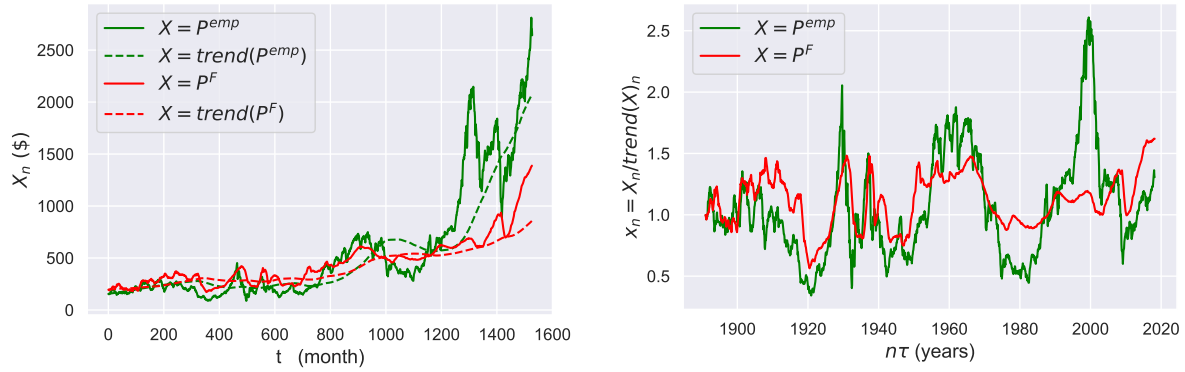


Figure 5.3: De-trending procedure applied to S&P500 index data. (Left) Empirical price and estimated fundamental price. In the legend, $trend(X)$ refers to the inverse of the exponential factor in Eq. 5.2. (Right) Stationary version of empirical price and estimated fundamental price.

The right panel of Figure 5.3 shows the de-trended version of price and fundamental price processes calculated using Eqs.(5.2) and (5.3). From the right panel of Fig. 5.3 one can see that the de-trended fundamental price errs less than the de-trended price. This is the cause of price excess-volatility, as emphasized in the main text.

Information about order flows is necessary to obtain a complete calibration, but in the dataset we are working with, no information is given about it. As we shall see in what follows, we can still partially calibrate the model, and obtain a complete description of the price process.

Calibration via the s-Kyle model. In this section, we detail the calibration procedure, given the de-trended time series we obtained above. Firstly, we remove the mean price level from the empirical de-trended price, assuming that it is common knowledge so that price impact it is insensitive to it.

Let us now detail how we can completely characterize the price process without information about the excess demand. Suppose that the signal and the noise ACFs are given by:

$$\Sigma_n^{\mathrm{IT}} = \Sigma_0^{\mathrm{IT}} \alpha^n \tag{5.4a}$$
$$\Omega_n^{\mathrm{NT}} = \Omega_0^{\mathrm{NT}} \delta_n \tag{5.4b}$$

where $0 < \alpha < 1$ and δ_n is the Kronecker's delta. The first equation states that the estimate of the IT of the (de-trended) fundamental price follows an auto-regressive process of order 1, which is a first approximation widely used in the literature (see, e.g., [41]). The second assumption states that on the scale of months the NT's order flow is not correlated, which is a good first approximation at very low frequencies.

We have been able to characterize analytically the stationary equilibrium that arises within the s-Kyle model in this case [85]. Interestingly, in the linear stationary equilibrium, the variance of the price set by the MM is given by:

$$\frac{\Sigma_0}{\Sigma_0^{\mathrm{IT}}} = \frac{1-\sqrt{1-\alpha^2}}{\alpha^2}, \tag{5.5}$$

which interpolates between the outcome of the original Kyle model if the signal becomes very short-lived ($\alpha \to 0$) and the case of a fully revealing price if the signal becomes permanent ($\alpha \to 1$). Moreover, in this case, information about the excess demand is not needed to specify completely the price process.

The calibration procedure goes as follows: from the fundamental price shown in the right panel (red line) of Fig. 5.3 one can calculate the associated empirical ACF. Then, we fit this empirical ACF with an exponential function, as prescribed by Eq. (5.4a). Although it is not possible to fix the amplitude of the price impact function G because we lack information about the excess demand's variance Ω_0, it is still possible to completely specify the price process by Eqs. (4.24), (5.4a) and (5.5). The results of this calibration are reported in Table 5.1.

5.4 Empirical results

5.4.1 Stationary Kyle model and excess volatility

Our first empirical question is the following: how literally should the s-Kyle model be taken? More specifically, our construction relates the price impact kernel with properties of the fundamental price, predicting that even at high frequency the magnitude of trade-induced price jumps should be related to some notion of fundamental information that is propagated all the way down from low frequencies to high frequencies. The s-Kyle model implies in particular that it should be possible to deduce the price impact function from a proxy of the fundamental price of a stock (e.g., via dividends, earnings) and the properties of the signed order flow, without ever measuring the market price.

We tested this approach on a sample of monthly data (i.e., with sampling scale $\tau = 1$ month), related to empirical prices (p^{emp}) and dividends (from which we constructed a proxy of the fundamental price p^{F}) of the S&P-500 index over ~ 150 years. The presentation of this dataset, the de-trending and the calibration procedures have been described in detail in Sec. 5.3. Note that we assume Markovian dynamics for the fundamental price process (defined by a decay time scale τ^{F} and an amplitude Σ_0^{F}) and an independent order flow, which are a good approximation for such a large sampling scale τ.

The left panel of Table 5.1 contains the estimations for the mean-reversion time scale of the different prices. We find that the mean-reversion time scale of the fundamental price τ^{F} is roughly in line with that deduced by the long-term behavior of the empirical market price τ^{emp}, although the empirical market price seems to be slightly more persistent with respect to the fundamental price. Note that the price efficiency condition (Eq. (4.24)) implies that the mean-reversion time scale of the price implied by the s-Kyle model τ is equal to that related to the fundamental price, i.e., $\tau = \tau^{\mathrm{F}}$. This means that the persistence amplification exhibited by the empirical price cannot be captured with the simple setting of the s-Kyle model.

τ^{emp} (years)	τ^{F} (years)
3.4 ± 0.3	2.6 ± 0.3

$V_\infty^{\text{F}}/V_\infty^{\text{emp}}$	$V_\infty/V_\infty^{\text{emp}}$
0.26 ± 0.02	0.15 ± 0.03

Table 5.1: Empirical results based on monthly recorded data about S&P-500 index. (Left) Mean-reversion time scale of de-trended empirical price, fundamental price, and price set by the MM in the s-Kyle setting. (Right) Price variance ratios as measures of explained empirical price volatility.

The second finding of this calibration is related to price volatility. The price volatility explained via the s-Kyle model is much smaller than that measured empirically as reported in the right panel of Table 5.1. We see that the squared fluctuations of both the fundamental price Σ_0^{F} and of the price set by the market maker Σ_0 in the s-Kyle model are much lower than the squared fluctuations of the empirical price Σ_0^{emp}. This should come as no surprise in light of well-known results on excess-volatility [15]. In fact, the market price exhibits higher volatility with respect to what would be implied by any reasonable proxy for price fundamentals. In our approach, the excess volatility reported in that body of work is automatically inherited by the price impact function, due to the fact that the price is the optimal estimator of the fundamental price (see Eq. 4.24). Moreover, as explained above, the price set by the risk-neutral MM in a noisy environment always reveals less information than the fundamental price so we have the following chain of inequalities: $\Sigma_0 < \Sigma_0^{\text{F}} < \Sigma^{\text{emp}}$.

This last result shows that if one was to literally believe to the s-Kyle model, the price predictions that it implies would only account for a rather small portion of the volatility of the empirical market price. That is at odds with many empirical results concerning the predictive power of the propagator model, which is able to account for a very large portion (up to 60-70% if calibrated using trade by trade data) of the empirical market price variation. The source of this difference relies on the fact that the propagator model's input is the empirical price and not the fundamental price as in the s-Kyle model. In this way, the propagator model is not affected by the excess-volatility puzzle, because there is not an explicit link to a notion of fundamental price. This implies that markets respond to the order flow as if it wasn't trying to anticipate the fundamental value of the risky asset, but something that would be much more in line with the market price.

5.5 Conclusions

We highlighted the impossibility to solve the excess-volatility puzzle in the s-Kyle model with an empirical analysis. This implies that the price impact function calibrated with the s-Kyle model will attain lower values than the one obtained by calibrating the propagator model. One can rephrase these findings by saying that the excess-volatility puzzle, in Kyle-like models are related to the excess price response puzzle. Nevertheless, the shape of the two obtained price impact functions is the same at high frequency, where the price is diffusive. This is explained by the microfounded model narrative by saying that the slow evolution of fundamentals does not shape the high-frequency dynamics of the price process, but only affects the magnitude of the price impact function via the diffusion constant related to the price.

Note that the excess-volatility puzzle can be solved in a setting with asymmetric information, like the s-Kyle model, if assumptions such as risk-neutrality and/or perfect structural

knowledge are relaxed. The accuracy of the microfounded model will increase if, for example, risk-aversion [87, 88] and/or learning dynamics [79, 80, 89, 90] are introduced.

Take home messages from Chapter 5

1. Price diffusivity exhibited by empirical data at high and medium frequency is recovered in the short time limit of the s-Kyle model.

2. We calibrated the s-Kyle model using 150 years of monthly S&P-500 data related to prices and dividends; we found that empirical prices have a variance 6 times higher than what the s-Kyle model predicts.

4. The calibration using the S&P-500 data fixes the mean-reversion timescale of the price that stems from the s-Kyle model to 2.6 years, which is in line with the one extracted by the market price, which is 3.4 years.

5. The assumption related to traders' rationality prevents the s-Kyle model to give an answer to the excess volatility puzzle.

6. In the following Chapter we present a modification of the Kyle model able to account for excess volatility without resorting to an irrealistic risk-aversion parameter. In doing so, we recover also price volatility clustering, without assuming that the fundamental price changes cluster in time.

Part III

Asymmetrically informed boundedly rational agents

Chapter 6

Microfounding GARCH Models and Beyond

In economics, there can never be a "theory of everything." But I believe each attempt comes closer to a proper understanding of how markets behave.

— Benoît B. Mandelbrot, The (Mis)Behavior of Markets

We relax the strong rationality assumption for the agents in the paradigmatic Kyle model of price formation, thereby reconciling the framework of asymmetrically informed traders with the Adaptive Market Hypothesis, where agents use inductive rather than deductive reasoning. Building on these ideas, we propose a stylized model able to account parsimoniously for a rich phenomenology, ranging from excess volatility to volatility clustering. While characterizing the excess-volatility dynamics, we provide a microfoundation for GARCH models. Volatility clustering is shown to be related to the self-excited dynamics induced by traders' behavior and does not rely on clustered fundamental innovations. Finally, we propose an extension able to account for the fragile dynamics exhibited by real markets during flash crashes.

Keywords: Adaptive Agents, Heteroscedasticity, Excess Volatility, Price Impact

From:
Microfounding GARCH Models and Beyond: A Kyle-inspired Model with Adaptive Agents
Michele Vodret, Iacopo Mastromatteo, Bence Tóth, Michael Benzaquen

Contents

6.1 Introduction

There exists a plethora of statistical models [4, 54, 91–96] to account for price dynamics and stylised facts in financial markets [97, 98]. While most of these models can be very useful for quantitative predictions, their formulation is not based on interacting agents, that is, they are not microfounded. An interesting research agenda consists in understanding how statistical models can be rationalized in terms of sound microfoundations. In particular, one asks what are the minimal hypotheses regarding agents' behavior needed to properly microfound a given statistical model; agent-based models are in fact well-known for their versatility [99].

During the early years of this research agenda [63], microfounded models relied on the Rational Expectation Hypothesis (REH), i.e., agents are endowed with perfect knowledge of the market model and have unlimited and cost-less computing power. While the REH allows for mathematical tractability and model interpretability, it has severe limitations regarding actual predictive power [15, 16, 21, 100]. Moreover, the deductive reasoning implied by the REH contrasts with the inductive reasoning on which humans often rely in complex situations [24, 39]. In face of uncertainty, in fact, humans rely on pattern recognition, hypothesis formation, deduction using currently held hypotheses, and replacement of hypotheses if needed; this leads to feedback effects, and consequently booms and bursts. Eventually, the Efficient Market Hypothesis (EMH), implied by the REH, is replaced with the formulation of the Adaptive Market Hypothesis (AMH) [25, 100, 101]. Although in the long-time limit of some games with adaptive agents the REH is recovered [102–104], this is not always the case [105] and, moreover, there is growing evidence that many stylized facts in financial markets cannot be captured within a framework where the REH holds [106–108]. Nevertheless, classic microfounded models can serve as a useful starting point for building more refined pictures of how financial markets work.

In his pioneering work, Kyle [18] proposed a highly stylized microfounded model for price formation: asymmetrically informed traders use rational expectations while interacting in the presence of noise trading. The corresponding rational equilibrium brings price impact, namely the fact that trades induce price jumps [54]. Notwithstanding, the Kyle model predicts that price volatility is smaller than that of the fundamental price and it is time-independent: if the fundamental price is interpreted as the efficient price, these findings are in strong contrast with empirical observations; price volatility is actually much larger than that related to the fundamental price by a factor ~ 5 [15, 16, 41], and exhibits intricate statistics with clustering and power law tails [97, 109]. Note that while the linear Kyle framework can be modified by considering a risk-averse market maker [87], an unrealistically high risk-aversion parameter [107] is needed in order to match empirical estimates.

The intermittent dynamics of price volatility can be accounted for by means of statistical descriptions, such as GARCH [94] models. While GARCH models are compatible with volatility clustering and power law tails, they are not order-driven models [63] and hence they cannot account for price impact. Moreover, no connection with the concept of fundamental price is given, and thus excess volatility is hardly definable. As a final note, since these statistical models are not microfounded, they leave the question of 'why do large price fluctuations cluster in time?' without a formal answer. The classic explanation for volatility clustering is stated by Engle, the doctoral advisor of Bollerslev, the author of GARCH models, in his Nobel prize lecture [110]: 'So at a basic level, financial price volatility is due to the arrival of new information. Volatility clustering is simply clustering of information arrivals. [...]'. This explanation is in line with EMH, which assumes that all the available information is encoded in the price. However, starting from the work of Cutler, Poterba

and Summers in 1988 [111], there is growing evidence that fundamental innovations only account for a small fraction of price jumps [67, 112], and clusters of jumps [113]. Thus, an alternative explanation is needed for volatility clustering. The REH accounts neither for excess volatility nor for the volatility clustering that is unrelated to fundamental innovations.

Here we provide a microfounded explanation for excess volatility and volatility clustering without the need to assume highly risk-averse agents or that fundamental innovation breaks in clusters. Instead, we consider agents in a Kyle setup who adapt their strategies over time [39, 80], in line with the AMH. We suppose that traders adapt their strategy because the environment is changing, namely, the noise trade level fluctuates over time. In this way, we recover a stationary regime in which the market maker sets the price according to temporarily fulfilled expectations or beliefs, which, in turn, give way to different beliefs when they cease to be fulfilled, and so on and so forth. As we shall see, these minimal ingredients introduce feedbacks and thus account for extreme events such as flash crashes. In doing so, we provide a microfoundation for GARCH models as well. We mention that a framework similar to ours has been recently proposed in Ref. [114] and [115], and previously in Ref. [116]; however, to our knowledge, this is the first model that gives a full microfoundation for volatility clustering, excess volatility and price impact; therefore the b-Kyle model seems a useful tool to understand the interplay between trades and prices, and eventually to investigate the nature of liquidity crises and flash crashes.

The Chapter is organized as follows. In Section 6.2 we present the model. Section 6.3 contains our main results, which is an analytical characterization of the dynamics in a simplified yet realistic limit, together with a numerical investigation of the intermittent price dynamics. In Section 6.4 we present two interesting extensions of the model, accounting for a risk-averse liquidity provider, and a cost-averse liquidity taker. Finally, in Section 6.5 we discuss our findings.

6.2 Model – Evolving market conditions and adaptive agents

Consider three agents, two liquidity takers, and one liquidity provider. They trade a single security over multiple trading rounds. At each trading round $t = 1, 2, \dots$, the liquidity takers first build their own demands: the informed trader (IT) knows the fundamental price p_t^{F} of the security, and exploits this private information to make profits, while the noise trader (NT) trades for exogenous reasons. The liquidity provider, therein called the market maker, filters the information about the fundamental price from the excess demand created by the liquidity takers, reflecting it into the price p_t.

The market conditions are specified by the statistics related to the signal and the noise process, respectively the fundamental price p_t^{F} and the noise trader order flow q_t^{NT}. These are modeled as Gaussian processes with zero mean and variance respectively given by ω^2 and ω_t^2, where

$$\omega_t^2 = \omega^2 + \delta\omega_t^2. \tag{6.1}$$

The fluctuations $\delta\omega_t^2$ follow an Auto-Regressive process of order one (AR(1)) with zero mean, volatility $\delta\omega^2 \ll \omega^2$ and typical timescale τ_{NT}; in formulas:

$$\delta\omega_t^2 = \exp(-1/\tau_{\mathrm{NT}})\delta\omega_{t-1}^2 + \eta_t\sqrt{1-\exp(-2/\tau_{\mathrm{NT}})}, \tag{6.2}$$

where η_t is a Gaussian process with zero mean and volatility given by $\delta\omega^2$. Note that although the fluctuations $\delta\omega_t^2$ can be negative, we are considering them small enough

($\delta\omega^2 \ll \omega^2$) such that the overall noise trade variance ω_t^2 is always positive. Other choices can be made for the structure of noisy order flow volatility fluctuations $\delta\omega_t$ that will not change qualitatively our results on intermittent volatility dynamics. Moreover, keeping the fundamental price volatility σ^{F} fixed while considering a time-varying noisy order flow volatility ω_t, is a matter of choice for the modeler, and will not change our results; in fact, the crucial point to obtain intermittent dynamics for price volatility, as we shall see, is that the market conditions evolve through time. However, we are implicitly assuming that the volatility of the fundamental price varies slower than that of noisy order flow. This seems to be a sound choice: fundamental price volatility varies slowly, consistent with the slow dynamics of the fundamental price, while noisy order flow fluctuations change rapidly, reflecting the fast, yet persistent, 'sentiment' dynamics in real markets.

The excess demand is the sum of the noise trader's and informed trader's demands:

$$q_t = q_t^{\mathrm{NT}} + q_t^{\mathrm{IT}}. \tag{6.3}$$

It is cleared by the market maker, who sets the price p_t of the security reflecting the unknown fundamental price p_t^{F}. This is done by filtering out from the only observable the market maker has access to, specifically, the excess demand q_t, the information about the fundamental price injected by the informed trader. To do so, the market maker needs a prior about the liquidity takers' strategies. The outcome of the market maker's decisions is a price p_t given, at each timestep t, by:

$$p_t = G_t q_t, \tag{6.4}$$

where G_t is the price impact function. In what follows, we explain how Eqs (6.3) and (6.4) are microfounded in terms of agents' strategies.

The informed trader knows that the price is set at each step by Eq. (6.4). Moreover, past prices and excess demands are public. This implies that at the beginning of each trading round t, the informed trader knows the price impact function realized at the previous step. Accordingly, the informed trader adapts his strategy over time by observing the evolving market conditions, which he captures via the evolving price impact function. We assume that the informed trader's best estimate for the one-step-ahead price impact function is the last observed one, which is a plausible heuristic rule. The informed trader is modeled as a risk-neutral utility maximizer, implying that his strategy, at each trading round t, reads [18]:

$$q_t^{\mathrm{IT}} \;=\; \frac{p_t^{\mathrm{F}}}{2G_{t-1}}. \tag{6.5}$$

The market maker knows about the liquidity takers' strategies, but he can only observe the realized excess demand q_t. Therefore, the market maker does not know the volatility of the fundamental price and of non-informed (or noise) trades: the market maker believes that these are $\widehat{(\sigma^{\mathrm{F}})}_t$ and $\widehat{\omega}$, respectively. We are therefore assuming that the market maker does not update his belief about noisy order flow volatility $\widehat{\omega}$, while he updates his belief about fundamental price volatility.[1] Note that the market maker's beliefs are denoted by hatted symbols, at variance with the ground-truth parameters that characterize the market conditions. The market maker is modeled as a risk-neutral, expected utility maximizer. This

[1]The idea, which will be formalized in what follows, is that the market maker revises his own belief about fundamental price volatility such that the price volatility expectation matches the price volatility estimate constructed from past price history. We shall see that this implies a feedback loop between past and future price volatility leading to the volatility clustering effect observed in empirical data.

implies that the price at step t is set to be the optimal estimator of the fundamental price p_t^{F}, given the beliefs $\{\widehat{(\sigma^{\mathrm{F}})}_t, \widehat{\omega}\}$ and the functional form of the informed trader's strategy q_t^{IT}. Accordingly, the price impact at each trading round t reads:

$$G_t = 2G_{t-1}\frac{\widehat{(\sigma^{\mathrm{F}})}_t^{\,2}}{\widehat{(\sigma^{\mathrm{F}})}_t^{\,2} + 4G_{t-1}^2\widehat{\omega}^2}. \tag{6.6}$$

Note that the equation above has the same structure as the standard solution for the risk-neutral Nash equilibrium of the single-step Kyle model [18, 51]; however, the price impact function G_t does not correspond to the real Nash equilibrium, due to the imperfect knowledge of the market maker about fundamental and noise trade volatilities.

We assume that $\widehat{(\sigma^{\mathrm{F}})}_t$ is a slowly varying function of t, meaning that the market maker's belief varies on timescales larger than that at which trading occurs. As we shall see, the dynamics allows for timescale separation: fast dynamics, which take place on timescales over which the beliefs $\widehat{(\sigma^{\mathrm{F}})}_t$ do not change, and slow dynamics, which take place on timescales over which the market maker revises his own beliefs. In the next subsections, we describe these dynamics.

6.2.1 Dynamics with constant belief

Consider a market maker's model that remains constant: $\widehat{(\sigma^{\mathrm{F}})}_t = \widehat{\sigma^{\mathrm{F}}}$ for every t. In Fig. 6.1 we show schematically the dynamics with constant beliefs of the market maker. The price impact dynamics, given by Eq. (6.6), admits a fixed point G_∞ regardless of the initial condition G_0, and implicitly defines a relaxation timescale, given τ_{fast}, such that if $t \gg \tau_{\mathrm{fast}}$, then $G_t \sim G_\infty$. The choice for the subscript 'fast' will become apparent shortly. The price impact and the expected price variance at the fixed point read:

$$G_\infty = \frac{\widehat{(\sigma^{\mathrm{F}})}}{2\widehat{\omega}}, \tag{6.7}$$

$$\hat{\sigma}_\infty^2 = \frac{\widehat{(\sigma^{\mathrm{F}})}^{\,2}}{2}, \tag{6.8}$$

which are reminiscent of the standard result of the Kyle model [18, 51], albeit these are calculated with market maker's beliefs. The relaxation timescale τ_{fast} is obtained from a standard dynamical system argument and it is found to be equal to one trading round.[2] Therefore, the subscript 'fast' relates to the fast equilibration dynamics of the price impact function.

6.2.2 Dynamics with belief revision

Consider a periodic updating procedure of the market maker's model, with period τ_{rev}. We refer to $k = 0, 1, 2, \ldots$ to denote the k'th update. In formulas, starting from $t = 1$ and $k = 0$, the fundamental price volatility belief at time t reads:

$$\widehat{(\sigma^{\mathrm{F}})}_t = \widehat{(\sigma^{\mathrm{F}})}_{k\tau_{\mathrm{rev}}}, \quad \text{for} \quad k\tau_{\mathrm{rev}} < t \leq (k+1)\tau_{\mathrm{rev}}. \tag{6.9}$$

[2] Let us mention here that τ_{fast} can have a more interesting behavior if the noise trader is cost-averse. We shall come back to this in Sec. 6.4.2

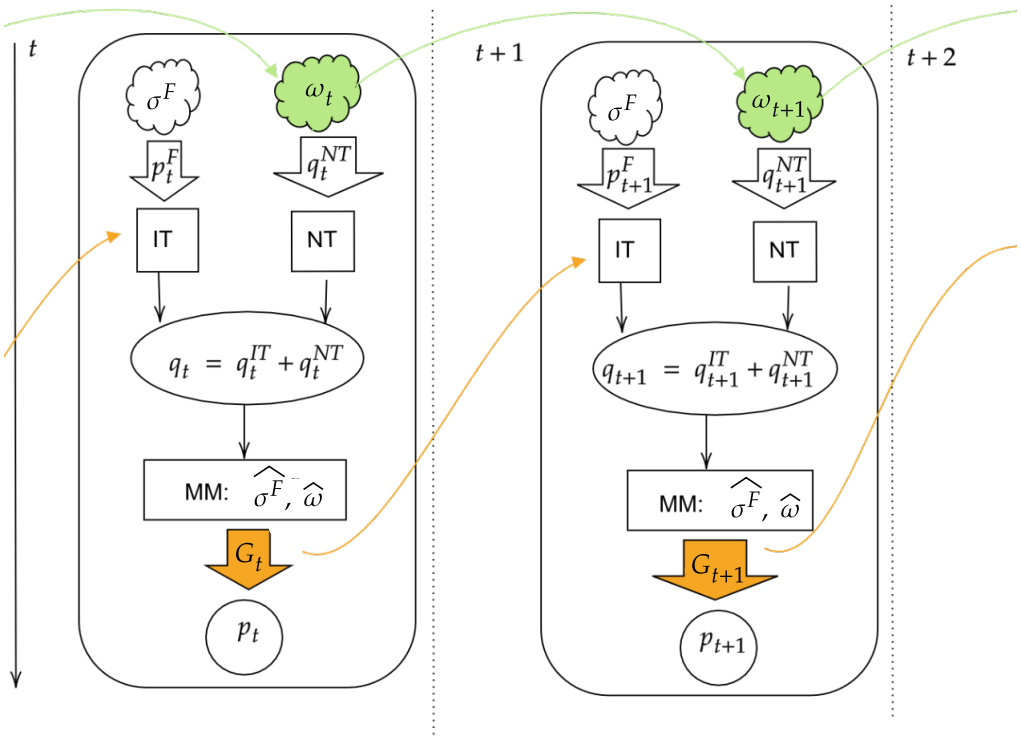


Figure 6.1: Dynamics with constant beliefs of the market maker, given by $\widehat{\omega}$ and $\widehat{(\sigma^{\mathrm{F}})}$. Two trading rounds are explicitly shown, i.e., t and $t+1$. The arrow of time in each trading round flows from the top to the bottom. Consider the trading round t. First, a realization of the fundamental price p_t^{F} and of the noise trade q_t^{NT} are obtained from two independent Gaussian processes with zero mean and variance respectively given by (σ^{F}) and ω_t. Then the liquidity takers create the excess demand q_t. Finally, the market maker clears the excess demand setting the price p_t with a price impact function G_t, which becomes available information to the informed trader at the trading round $t+1$. Green and orange arrows refer to the dynamics related to the noise trade variance and to the price impact function, respectively given by Eqs. (6.1),(6.2) and (6.6).

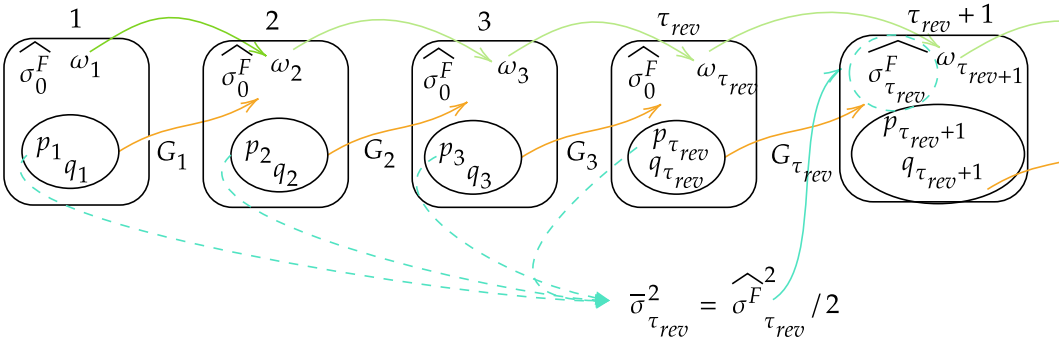


Figure 6.2: Dynamics with belief revision, starting from $t = 1$. For the first $\tau_{\rm rev}$ trading rounds the dynamics is represented, within each trading rounds, by boxes, as in Fig. 6.1. The updating procedure at the end of step $\tau_{\rm rev}$ is the only new feature: from the string of past $\tau_{\rm rev}$ prices, the market maker computes the empirical estimate $\bar{\sigma}^2_{\tau_{\rm rev}}$. From this estimate the market maker updates his belief about the fundamental price variance $\widehat{(\sigma^{\rm F})}^2_{\tau_{\rm rev}}$ accordingly to Eq. (6.10). In addition to the green and orange arrows, already present in Fig. 6.1, we show cyan arrows that represent the belief revision dynamics.

According to the equation above, at the end of the trading rounds $t = k\tau_{\rm rev}$ with $k \geq 1$, the market maker revises his belief about the fundamental price volatility.

The revision procedure comprises two steps. First, the market maker calculates an estimate of price volatility $\bar{\sigma}_{k\tau_{\rm rev}}$ taking into account the last $\tau_{\rm rev}$ recorded prices. The updating timescale $\tau_{\rm rev}$ controls the measurement error of the price volatility estimate: the larger $\tau_{\rm rev}$, the smaller the measurement error, that is the more precise is the estimate $\bar{\sigma}_{k\tau_{\rm rev}}$. Then, the market maker updates his fundamental price volatility belief $\widehat{(\sigma^{\rm F})}_{k\tau_{\rm rev}}$ such that his current price volatility estimate $\bar{\sigma}_{k\tau_{\rm rev}}$ equals the new long-time expected price volatility. Note that the relation between the expected long-time price volatility and fundamental price volatility belief is given by Eq. (6.8). Accordingly, the updated belief about the fundamental price variance satisfies the following condition:

$$\frac{\widehat{(\sigma^{\rm F})}^2_{k\tau_{\rm rev}}}{2} = \bar{\sigma}^2_{k\tau_{\rm rev}}. \tag{6.10}$$

The dynamics with belief revision is shown schematically in Fig. 6.2.

The heuristic rule given above coincides with $\hat{\sigma}_{k\tau_{\rm rev}} = \bar{\sigma}_{k\tau_{\rm rev}}$, if $\tau_{\rm rev} \gg \tau_{\rm fast}$ so that $G_t = G_\infty$ given by Eq. (6.8). In what follows, we shall refer to this regime as the sticky expectation one. In this regime the market maker is conservative, meaning that he slowly updates his belief; therefore, he constructs precise price volatility estimates, given that he uses a large number of past prices.

Now we have all the ingredients needed to understand why we choose to consider a time-varying fundamental price volatility belief $\widehat{(\sigma^{\rm F})}_t$: this choice allows to create a feedback loop between past prices and future price volatility. Specifically, past prices affect fundamental price volatility beliefs (see Eq. (6.10)), which in turn affect prices via the price impact function (see Eqs. (6.4) and (6.6)). This general feedback is well-known to be a feature of real markets [25], where price volatility exhibits intermittent dynamics [97].

We anticipate here that the slow dynamics of updating beliefs can reach a stationary regime, which will be investigated in the next section. In this case, given an initial belief $\widehat{(\sigma^{\rm F})}_0$, there will be an associated relaxation timescale with which the stationary regime is

achieved, denoted by $\tau_{\text{slow}} = k_{\text{slow}}\tau_{\text{rev}}$. Note that we used the subscript 'slow' to distinguish the slow relaxation timescale, from the fast relaxation one denoted by 'fast', related to the market maker's belief and the price impact function, respectively.

6.3 Results

6.3.1 Sticky expectation regime

In this section we show that in the sticky expectation regime ($\tau_{\text{rev}} \gg \tau_{\text{fast}}$) the model simplifies, allowing for analytical characterization of price volatility dynamics. Later, we elucidate the connection to GARCH models.

Kesten dynamics

The timescale with which the price impact relaxes to its long-time limit τ_{fast} is of order one, as stated in Sec. 6.2.1; therefore, we construct the sticky expectation regime as follows:

$$\frac{\tau_{\text{fast}}}{\tau_{\text{rev}}} \to 0, \quad \frac{\tau_{\text{NT}}}{\tau_{\text{rev}}} = r. \tag{6.11}$$

The first condition implies that the updating timescale τ_{rev} is way larger than the equilibrium timescale τ_{fast} of the price impact dynamics. This allows to compute analytically the price volatility estimate $\bar{\sigma}_{k\tau_{\text{rev}}}$, thanks to two simplifications. First, it allows to replace, in the calculation for price volatility estimate $\bar{\sigma}_{k\tau_{\text{rev}}}$, the price impact function with its fixed point value given by Eq. (6.7); note that one has to replace in Eq. (6.7) $(\widehat{\sigma^{\text{F}}})$ with its time-dependent value according to Eq. (6.9). Second, since the price volatility estimate is constructed with a large number of observations ($\tau_{\text{rev}} \to \infty$), we can neglect the associated measurement error. Accordingly, we simplify the notation of price volatility expectation $\bar{\sigma}_{k\tau_{\text{rev}}}$ by removing the bar symbol. The second condition in Eq. (6.11) implies that noisy order flow volatility varies in the intervals $(k\tau_{\text{rev}}, (k+1)\tau_{\text{rev}}]$ if $r < \infty$. In the remaining part of the section, we measure time in units of τ_{rev}, leaving only the index k, introduced in Eq. (6.9) to denote the trading rounds at which market maker's beliefs are updated.

The excess volatility $\sigma_k/(\sigma^{\text{F}})$ dynamics can be analytically characterised in the sticky expectation regime starting from Eq. (6.10) and using Eqs. (6.3), (6.4), (6.5) and (6.7). With the simplified notation detailed above, the dynamics for the excess variance reads:

$$\frac{\sigma_k^2}{(\sigma^{\text{F}})^2} = \frac{1}{4} + \frac{\omega_k^2}{2\widehat{\omega}^2}\frac{\sigma_{k-1}^2}{(\sigma^{\text{F}})^2}. \tag{6.12}$$

Therefore, the excess variance is a Kesten process [44], i.e., a stochastic multiplicative process repelled from zero. The Kesten dynamics implies the possibility of having intermittent dynamics for price volatility together with large price volatility fluctuations captured by power law behavior. In fact, the dynamics depends crucially on the mean value of the stochastic multiplicative factor $\omega_k^2/(2\widehat{\omega}^2)$. In the case where $\delta\omega_k^2$ are iid, if $\langle\omega_k^2/(2\widehat{\omega}^2)\rangle > 1$, the Kesten process diverges and no stable distribution is reached. Conversely, if $\langle\omega_k^2/(2\widehat{\omega}^2)\rangle < 1$, the Kesten process is stable and approaches a limiting distribution for large times.[3] This condition has an implication for the market maker's belief about noisy order flow volatility: the

[3]We will clarify what we mean by large times below when we analyze the dynamics more precisely.

stationary regime can be reached if and only if the market maker's underestimation of the noise trade variance $\widehat{\omega}^2$ is smaller than a critical threshold $\widehat{\omega}_c^2$. More details about this and about the generalization to the case of correlated $\delta\omega_k^2$ are given below. The equation above, after some manipulation involving Eq. (6.1), can be rewritten as:

$$\sigma_k^2 = \langle\sigma^2\rangle + \frac{\omega^2}{2\widehat{\omega}^2}\left(\sigma_{k-1}^2 - \langle\sigma^2\rangle\right) + \frac{\delta\omega_k^2}{2\widehat{\omega}^2}\sigma_{k-1}^2, \tag{6.13}$$

Accordingly, price volatility dynamics is the sum of three terms: a constant long-time contribution $\langle\sigma^2\rangle$, a deterministic mean-reverting contribution, and a stochastic one, which represents the update based on the last empirical observation, which in turn reflects the evolving market conditions, i.e., the noise trades fluctuations $\delta\omega_k^2$.

Comparison to GARCH models

In the quantitative finance literature, GARCH models [94] are very well-known [43, 117]. These are statistical models constructed explicitly in order to capture the intermittent dynamics of price volatility. In these models, price changes δp_k are modeled as the product between the equal time volatility and Gaussian iid random variables ξ_k with zero mean and unit volatility. The simplest model of this class is completely characterised by Eq. (6.13), with the substitutions $\frac{\omega^2}{2\widehat{\omega}^2} = \alpha$ and $\frac{\delta\omega_k^2}{2\widehat{\omega}^2} = g(\delta p_{k-1}^2 - 1)$, following the notation in Ref. [43], where $\alpha < 1$ and $g > 0$. This model is coined GARCH$(1,1)$, since only the previous time price volatility (σ_{k-1}) and price changes (δp_{k-1}) are taken into account. Although the sticky expectation regime of our model is characterized by a GARCH-like structure, the model we set up has a richer structure, highlighted below.

- In GARCH models no connection between market price and fundamental price is provided, at variance with our Kyle-inspired model. In our model, excess volatility is the variable of interest, and not price volatility by itself (see Eq. (6.12)).

- Although the price volatility in the sticky expectation regime of our model is of the GARCH type, our model is able to provide a microscopic interpretation for each of the three terms which appear in Eq. (6.13). The first one, i.e., the long-time price variance $\langle\sigma^2\rangle$ in our model is related to the distance between the market condition and market maker's beliefs, as we shall see in Sec. 6.3.2. Similarly, the second, i.e., the coefficient of mean-reversion is related to the ratio $\omega^2/(2\widehat{\omega}^2)$. Finally, the variability of the interaction term, in the sticky expectation regime of our model, is not due to the measurement error of price volatility, as in the GARCH$(1,1)$ model, but rather to the time variability of noisy order flow volatility: in the GARCH$(1,1)$ model, the price volatility estimate (δp_k) is calculated only with the previous price change, implying a sensible measurement error; conversely, in the sticky expectation regime of our model, price volatility estimate is constructed from the past $\tau_{\text{rev}} \to \infty$ prices, obtaining an estimate without measurement error.

- Individual returns, in the sticky expectation regime of our model, are not explicitly modeled, because only their volatility and the volatility of the noise trades affect the excess-volatility dynamics (see Eq. (6.12)). This is not the case for the GARCH$(1,1)$ model, where the kurtosis of returns can be calculated [43, 117]. In order to address directly this quantity, one must resort to the full model described in Sec. 2.

- The stochastic multiplicative factor δp_k^2 of GARCH models is uncorrelated, whereas the one in our model ($\delta\omega_k^2$) it is an AR(1) process. Accordingly, in the GARCH$(1,1)$ model, the Auto-Correlation function (ACF) of price variance is a single decaying exponential [117], with correlation timescale $\tau_{\rm ACF} = 1/|\log(\alpha)|$. The sticky expectation regime of our model recovers this result if $r \to 0$; as we shall see below, in the generic case of finite r, the ACF of price volatility is more complicated than that predicted by GARCH model.
- The model we built allows also to make predictions on a non-observable quantity, namely the ratio between informed and non-informed orders. Further details on this matter are given in Sec. 6.3.2.

6.3.2 Excess volatility

In the following, we characterize the excess volatility. First, we consider the case with fixed noise trade variance, then we consider the case where it fluctuates. For each investigation, first, the sticky expectation regime is considered as it is easily interpretable and manageable; later, we highlight the differences with the simulation of the model presented in Sec. 6.2, which are affected by the measurement error on the price volatility estimate of the market maker. In Appendix D.1 we present a pseudo-code for the simulation.

Static market conditions

Consider the case where the noise trade volatility does not fluctuate over time $\omega_k = \omega$. In what follows we refer to this case as the Mean Field (MF) regime. In this regime, the excess-volatility dynamics in the sticky expectation regime of our model, given by Eq. (6.12), becomes a deterministic dynamical equation. In the stable regime ($\omega^2/(2\widehat{\omega}^2) < 1$), the excess volatility converges to a finite fixed point. The relaxation is exponentially fast with a timescale

$$\tau_{\rm slow} = \tau_{\rm rev}\left(1 - \frac{\widehat{\omega}_c^2|_{\rm MF}}{\widehat{\omega}^2}\right)^{-1}, \tag{6.14}$$

where $\widehat{\omega}_c^2|_{\rm MF} = \omega^2/2$. Therefore, the timescale with which the stationary regime is achieved $\tau_{\rm slow}$ diverges as the market maker underestimation $\widehat{\omega}^2$ approaches the critical value $\widehat{\omega}_c^2|_{\rm MF}$. The mean excess volatility can be easily computed from Eq. (6.12), obtaining:

$$\left\langle \frac{\sigma^2}{(\sigma^{\rm F})^2} \right\rangle_{\rm MF} = \frac{1}{4}\left(1 - \frac{\widehat{\omega}_c^2|_{\rm MF}}{\widehat{\omega}^2}\right)^{-1}. \tag{6.15}$$

The equation above implies that the more the market maker underestimates the level of noise trading, the more the fixed point for the mean excess volatility grows. This can be easily explained as follows: if the market maker underestimates the mean value of noisy order flow volatility, he will overestimate the price impact function, leading to excessively volatile prices. In the case where the fluctuations $\delta\omega_k$ are iid, we recover the MF result, because $\langle\omega_k^2\sigma_{k-1}^2\rangle = \omega^2\langle\sigma^2\rangle$.

Interestingly, our model allows predicting the ratio between informed and non-informed trades. In fact, in the MF limit of the sticky expectation regime of our model, one finds:

$$\left\langle \frac{\omega^2}{\omega_{\rm IT}^2} \right\rangle_{\rm MF} = \frac{\omega^2}{\widehat{\omega}^2}\left(2 - \frac{\omega^2}{\widehat{\omega}^2}\right)^{-1}, \tag{6.16}$$

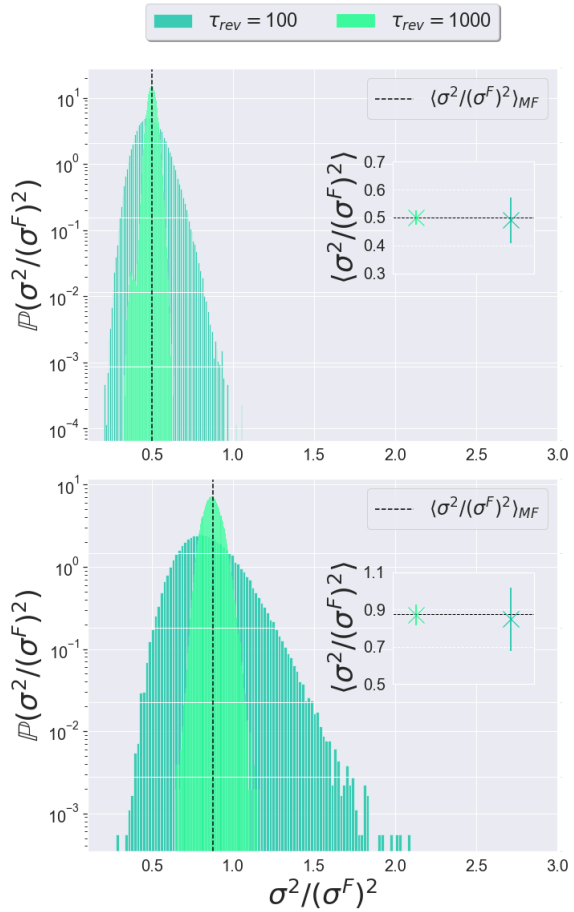


Figure 6.3: PDFs of excess variance with constant noise trade variance fluctuations ($\delta\omega^2 = 0$), varying the updating timescale $\tau_{\rm rev}$. The more the updating timescale becomes large, the more the excess-variance PDF is peaked on the mean value. (Top) Noise trade variance belief is equal to the true value. (Bottom) Noise trade variance belief lower than true value: $\widehat{\omega}^2/\omega^2 = 0.7$. Black dashed vertical lines show the MF level of excess variance. Insets compare the MF level of excess variance with simulation outcomes (with error bars).

where $\omega^2_{\rm IT}$ is the informed trade variance.

Numerical simulations First, we analyze the case where the market maker knows perfectly the statistical properties of the noisy order flow volatility, i.e., $\widehat{\omega} = \omega$; this is the case in which the REH holds. In the top panel of Fig. 6.3 we show results about the Probability Distribution Function (PDF) of excess variance for different values of the updating timescale $\tau_{\rm rev}$. There, one sees that the larger $\tau_{\rm rev}$ is, the more the excess variance is peaked around the strong rationality equilibrium value $\langle \sigma^2/(\sigma^{\rm F})^2 \rangle = 1/2$ [18, 51]. Fluctuations are induced by measurement errors that affect the price variance estimate. Note also that the larger $\tau_{\rm rev}$, the more the distribution is peaked since more past prices are considered to obtain the estimate for price variance. The inset shows in a clearer way that the mean excess variance agrees with the MF value of the sticky expectation regime that we analyzed in the previous section.

Next, we consider the case where the market maker underestimates the level of noisy order flow volatility, i.e., $\widehat{\omega} < \omega$. In this case, the excess variance is shifted to larger values, as we show in the bottom panel of Fig. 6.3. Note that fluctuations of the excess variance are more pronounced since underestimating the noise trade level boosts the effect of the

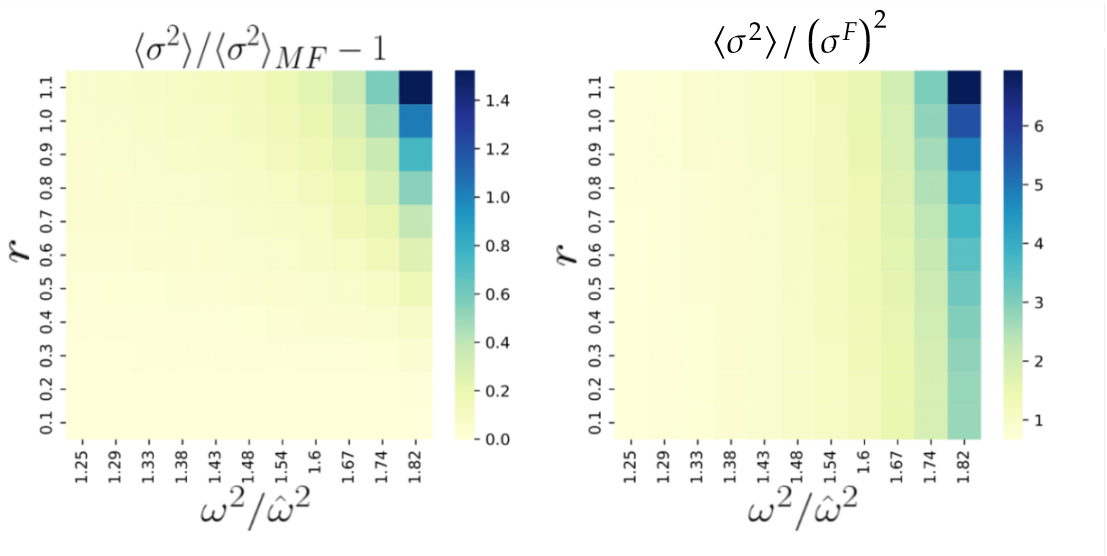


Figure 6.4: Mean excess variance as a function of the timescale ratio r and of market maker's underestimation of noise trade variance, measured by $\omega^2/\widehat{\omega}^2$. We set the variance of noise trade fluctuations to $\delta\omega^2/\omega^2 = 0.1$. (Left) Normalized by the mean-field level. One can observe the departure from the mean-field description, due to the combined effect of fluctuations and of market maker's underestimation of the noise trade variance. (Right) Mean excess variance. One can observe that a given level of excess variance can be obtained with different combinations of the parameters.

noise trade variance fluctuations. The inset shows again that the mean excess variance is in agreement with the MF value of the sticky expectation regime (see Eq. (6.15)).

Dynamic market conditions

Consider a fluctuating noise trade variance modeled as an AR(1) process with positive and finite correlation timescale $\tau_{\rm NT}$, as the noise trade variance we defined in Eq. (6.1). In principle, we can compute the mean value of excess variance in the sticky expectation regime of our model starting from Eq. (6.12). However, the correlation of the noise term complicates the analysis. Therefore, we consider the regime of fast vanishing ACF of noise trade variance fluctuations ($r \ll 1$) (or small overall level of fluctuations, i.e., $\delta\omega^2 \ll 1$). In fact, since ω_k^2 is Gaussian, all high-order auto-correlations boil down to terms proportional to products of two-time correlation functions between fluctuation $\delta\omega_k^2$ terms, which are proportional to $\delta\omega^2 \exp(-1/r) \ll 1$. The summation to first order in $\delta\omega^2 \exp(-1/r)$ leads to a modified mean excess variance. In particular, the mean excess variance is still of the form given by Eq. (6.15), but the critical value of market maker's belief about noise trade variance is modified to:

$$\widehat{\omega}_c^2 = \widehat{\omega}_c^2\big|_{\rm MF} + \frac{\delta\omega^2}{2}\exp(-1/\tau_{\rm NT}). \tag{6.17}$$

Similarly, the slow relaxation timescale $k_{\rm slow}$ is given by Eq. (6.14), where one has to replace the MF critical parameter with its 'fluctuations aware' version given above. These findings imply that a given level of mean excess variance can be obtained with a smaller underestimation of the market maker about noise trade variance if fluctuations are present, as we show in Figure 6.4. There, in the left panel, one can see that mean excess variance is equal to its MF value if r is small (bottom part of the plot) or if $\widehat{\omega}^2$ is negligible (left part of the plot). Conversely, as $\widehat{\omega}^2$ gets closer to the critical value (right part of the plot), one can

see a sharp multiplicative increase as the fluctuations are increased (top-right corner in the plot). The increase of the mean excess variance as fluctuations are more persistent (larger r) can be interpreted by saying that $1 - \widehat{\omega}_c^2/\widehat{\omega}^2$ is getting smaller, implying an increase of the overall mean excess variance (see Eq. (6.15) properly modified by Eq. (6.17)). The right panel of Fig. 6.4 shows that a given level of price variance can be obtained with different choices of market maker's underestimation of noise trade variance level and the persistence of noise trade variance fluctuations.

While in Fig. 6.4 a high mean excess variance requires in any case an underestimation of noise trade variance level, note that high excess variance can be obtained also if the market maker is right about the mean level value of fundamental price variance, but noise trade variance fluctuations are large in magnitude and are persistent. In this case, in fact, the situation our model describes is qualitatively similar to that encountered in economic models where quasi-non-ergodicity is taken into account [118]; quasi-non-ergodicity occurs when a stochastic process is ergodic at very long-time horizons, but where ergodicity breaks down on a time scale at which realizations from the process might realistically be observed by a human agent. This is exactly the situation the market maker faces if $r < \infty$; in this case, in fact, his belief about the fundamental price variance is sensitive to noise trade variance fluctuations.

Numerical simulations We analyze simulations with fixed updating timescale τ_{rev} and different timescales of fluctuations of noisy order flow volatility τ_{NT}. In the left panel of Fig. 6.5, one sees that the more persistent the fluctuations (large r), the more the PDF of excess variance is skewed toward large values (the shift is related to the underestimation of the noise trade's level, as before). The reason why this occurs can be understood from Eq. (6.12): the more the noisy order flow volatility is serially correlated, the more the feedback dynamics on price volatility persists in the same direction, leading to large fluctuations of excess volatility. In the following section, we characterize analytically the tail behavior in the sticky expectation regime. The inset shows that the more persistent the fluctuations of noisy order flow volatility, the higher the mean excess volatility, consistent with the results presented in Fig. 6.4.

6.3.3 Intermittent volatility dynamics

In the following, we characterize the intermittent volatility dynamics. As in the previous section, first, the sticky expectation regime is considered, and then our findings are further substantiated by the outcomes of the simulation of the model.

The intermittent dynamics of excess volatility can be characterized, at a first approximation, by the tail exponent of the Cumulative Distribution Function (CDF) of price volatility, and by the temporal decay of the ACF of the price variance, respectively given by μ and τ_{ACF} (actually, to provide an accurate description of empirical findings, more than one timescale is needed to characterize the ACF of the price variance [43]). Results regarding the CDF's power law tail and the structure of the ACF of price variance are available for Kesten processes with iid multiplicative noise in the stable regime, where $\langle \omega_k^2/(2\widehat{\omega}^2) \rangle < 1$. Below, we recall these important results, and we highlight how they change when an AR(1) process is considered.

[4] Note that we anticipated the MF result regarding the ACF of excess variance in Sec. 6.3.1.

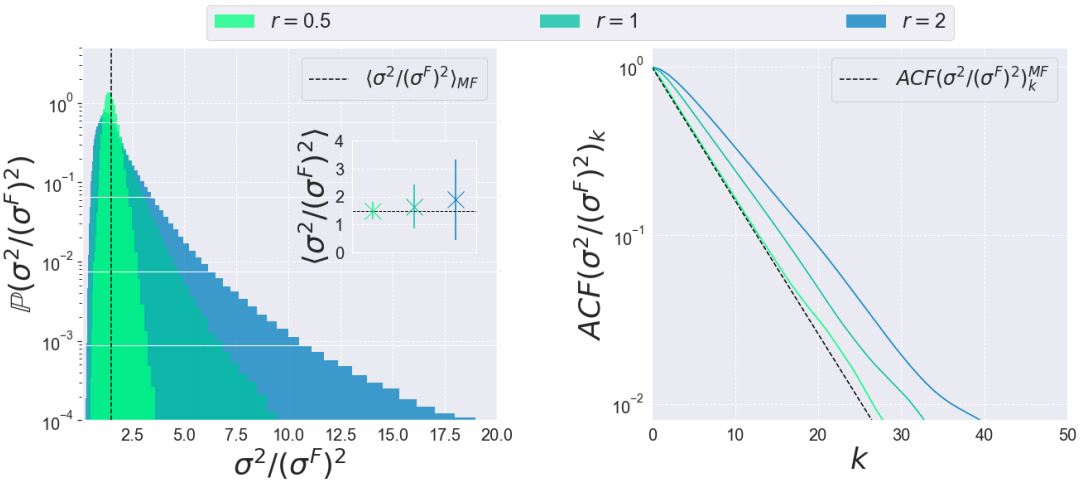


Figure 6.5: Simulations varying the timescale ratio r, while keeping fixed the updating time to $\tau_{\text{rev}} = 1000$. We set $\delta\omega^2/\omega = 0.2$ and $\widehat{\omega}^2/\omega^2 = 0.6$. (Left) PDF of excess variance. Inset: MF level of excess variance compared with simulations outcomes (with error bars). (Right) ACF of excess variance. Black dashed lines are related to MF results in the sticky expectation regime of our model.[4]

If the noise trade variance ω_k^2 is iid, the CDF of excess volatility decays asymptotically as a power law with exponent μ, which has the following form [119, 120]:

$$\left\langle \left(\frac{\omega_k^2}{2\widehat{\omega}^2} \right)^{\mu/2} \right\rangle = 1. \tag{6.18}$$

Accordingly, the tail of the price volatility CDF is thicker, i.e., μ is smaller, and the more the market maker underestimates the mean level of noisy order flow volatility. Equation (6.18) implies that the power law tail of the probability distribution shape is robust with respect to the underlying distribution of the multiplicative term $\omega_k^2/(2\widehat{\omega}^2)$. This insensitivity to micro-structural details justifies Kesten processes as an effective description of the universal intermittent dynamics exhibited by price volatility. If ω_k^2 are realizations of an AR(1) process, from numerical simulations we observe that the excess volatility has again an exponent μ which gets smaller with increasing persistence in noisy order flow volatility fluctuations.

If ω_k^2 are iid, the long-time ACF of price volatility is a single decaying exponential function. If $\mu > 2$, which is the case for real markets [43], the correlation timescale τ_{ACF} of price volatility writes [120]:

$$\tau_{\text{ACF}} = \frac{8}{\mu - 1} \left(\frac{\widehat{\omega}}{\delta\omega} \right)^4 . \tag{6.19}$$

Note that an interesting relation for μ can be obtained in the case of uncorrelated noisy order flow volatility fluctuations by comparing the equation above for τ_{ACF} with the one given in Sec. 6.3.1. According to the equation for τ_{ACF} given there, τ_{ACF} increases when $\widehat{\omega}$ approaches the critical value ω_c. From the equation above instead we conclude that τ_{ACF} increases if the level of fluctuations of noisy order flow volatility $\delta\omega$ decreases, recovering the MF regime we analyzed in the previous section in the limit case where $\delta\omega = 0$. In the case

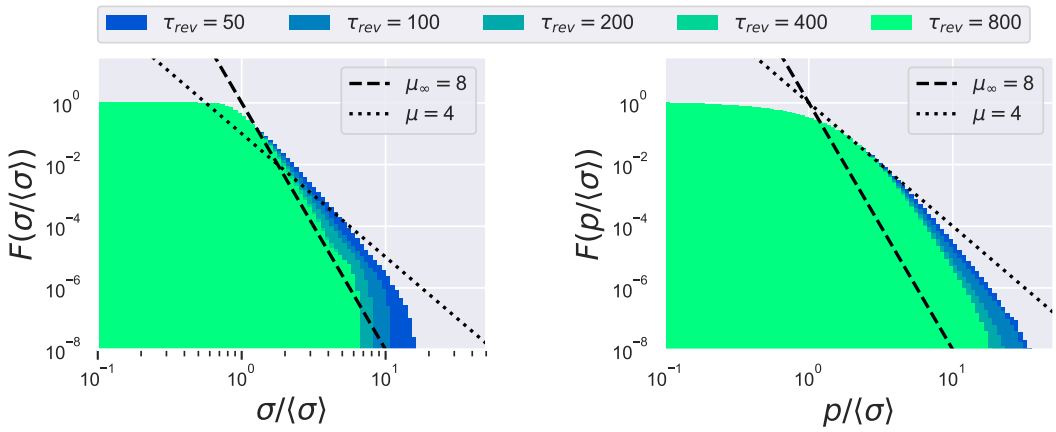


Figure 6.6: CDF of price volatility and price obtained from simulations with $r = 1$, varying the updating timescale $\tau_{\rm rev}$. The other parameters are chosen to be $\delta\omega^2 = 0.1$, $\widehat{\omega}^2 = 0.52$. (Left) CDF of normalized price volatility. (Right) CDF of normalized price. The dotted line refers to the results of the related Kesten process, while the dashed line is superposed for illustrative purposes.

where ω_k^2 is an AR(1) process, the more persistent the noise trader's volatility fluctuations, the more correlated price volatility, and the larger $\tau_{\rm ACF}$.

Empirical analysis conducted on the ACF of price volatility shows that at least two timescales are needed in order to capture its temporal structure [43]. In the case where $\delta\omega_k^2$ is an AR(1) process, the ACF of price variance obtained with finite r is captured by two decaying exponential functions, as we shall see below from numerical simulations.

Numerical simulations Here we compare the results of the simulation of the model we presented in Sec. 6.2 with the prediction of the Kesten process given by Eq. (6.18). To do so, we run different simulations with a fixed ratio $r = 1$ approaching the limit $\tau_{\rm rev} \to \infty$, where the measurement error on the price volatility estimate vanishes and the Kesten dynamics is recovered by construction. We set the model parameters in line with empirical predictions about mean excess volatility: $\langle\sigma/(\sigma^{\rm F})\rangle \sim 3.5$ is obtained if $\widehat{\omega}^2/\omega^2 = 0.52$ and $\delta\omega^2/\omega^2 = 0.1$. Interestingly, from Eq. (6.16), for every stock bought/sold by an informed trader, there are ~ 25 stocks bought/sold by the noise trader, which means that more than 90% of the overall liquidity is made of non-informed trades. In the left panel of Fig. 6.6, we show the CDF of normalized price volatility. A clear result is that the larger the updating timescale $\tau_{\rm rev}$, the thinner the tail. This agrees with intuition: the lower the measurement error on the price volatility estimate is, the lower the fluctuations of market maker's belief. The power law tail related to CDF of excess volatility obtained with $\tau_{\rm rev} = 800$ is well approximated by $\mu_\infty = 8$, according to the result obtained from the simulation of the Kesten process, where the measurement error on the price volatility estimate is neglected since $\tau_{\rm rev} \to \infty$. In the right panel of Fig. 6.6, we show that the power law tail of the price CDF does not change with respect to that related to price volatility, as expected. Let us mention here that one can obtain a thicker tail assuming a more realistic excess-demand process, specifically a long-range correlated process, typically with a power-law tail with exponent 1/2 [53]. Finally, the analysis of the ACF of price variance is done in the right panel of

Fig. 6.5. One sees that the more persistent the noise trader fluctuations, that is, the larger r, the slower the decay of the ACF of price variance. This result can be explained again by recalling the analysis of the sticky expectation regime given in the previous section: the more correlated the multiplicative factor in Eq. (6.12), the more correlated price volatility as well. In the opposite limit, the results of the simulations match the mean-field result of the sticky expectation regime we analyzed in the previous section, as expected. Note that the ACF for $r = 2$ is clearly not captured by a single decaying exponential, as is the case for empirical data [43].

6.4 Extensions

6.4.1 Risk-averse market maker

It is well-known that risk-aversion of the market maker implies in Kyle-like frameworks higher price volatility with respect to the risk-neutral case; see for example Ref. [87], where the case of an absolute relative risk-averse (CARA) market maker is considered. However, in order to explain the excess volatility encountered in empirical data an unrealistically high risk-aversion parameter has to be chosen [107]. It is interesting to evaluate the importance of market maker's risk aversion in driving the mean price volatility in our framework. To do so, we modify the market maker's model behavior, so that risk-aversion is enforced with CARA while the linearity of Eq. (6.4) is retained. The task of the market maker, in this case, is to choose G_t such that:

$$\mathbb{E}[U_t^{\mathrm{MM}}|q_t, \widehat{(\sigma^{\mathrm{F}})}_t, \widehat{\omega}] = \mathbb{E}[q_t(p_t - p_t^{\mathrm{F}})|q_t, \widehat{(\sigma^{\mathrm{F}})}_t, \widehat{\omega}] - \rho_t \mathrm{var}[q_t(p_t - p_t^{\mathrm{F}})|q_t, \widehat{(\sigma^{\mathrm{F}})}_t, \widehat{\omega}] = 0. \quad (6.20)$$

The risk-averse market maker's strategy depends now also on a risk-aversion parameter ρ_t, in addition to the beliefs about noise trades and fundamental price volatility. Accordingly, the self-consistent equation for the price impact function writes [87]:

$$G_t = \frac{2G_{t-1}\widehat{(\sigma^{\mathrm{F}})}_t^2}{\widehat{(\sigma^{\mathrm{F}})}_t^2 + 4G_{t-1}^2\widehat{\omega}^2}(1 + 2G_{t-1}\rho_t\widehat{\omega}^2). \quad (6.21)$$

It is standard in the quantitative finance literature to express the degree of risk-aversion in term of the Sharpe ratio, defined as the ratio between the expected gain and the square root of the risk associated to a given strategy. We define the Sharpe ratio per period as:

$$S = \frac{\mathbb{E}[\mathbb{E}[q_t(p_t - p_t^{\mathrm{F}})|q_t]]}{\sqrt{\mathbb{E}[\mathrm{var}[q_t(p_t - p_t^{\mathrm{F}})|q_t]]}} \quad (6.22)$$

In order to have a constant Sharpe ratio per period, the market maker has to choose a risk-aversion coefficient $\rho_t = S/(\widehat{\omega}\widehat{(\sigma^{\mathrm{F}})}_t)$.

As in the case of a risk-neutral market maker, the timescale needed for the price impact to reach the fixed point is still of order one, i.e., $\tau_{\mathrm{fast}} = 1$. The fixed point of Eq. (6.21) can be again computed, as we did in Sec. 6.2.1: assuming market maker's belief about the fundamental variance to be constant, $\widehat{(\sigma^{\mathrm{F}})}_t = \widehat{(\sigma^{\mathrm{F}})}$, the price impact and the expected price variance are respectively given by:

$$G_\infty = \frac{S + \sqrt{1+S^2}}{2}\frac{\widehat{(\sigma^{\mathrm{F}})}}{\widehat{\omega}}, \quad (6.23)$$

$$\hat{\sigma}_\infty^2 = \frac{\widehat{(\sigma^{\mathrm{F}})}^2}{2}\left[1 + S(S + \sqrt{1+S^2})\right]. \quad (6.24)$$

The risk-aversion of the market maker increases the value of the price impact, and, consequently, it increases the value of the expected price volatility, leading, as we shall see, to an increase in the actual price volatility.

The slow dynamics of excess variance in the sticky expectation regime defined by Eq. (6.11) is again of the Kesten type. In fact, following the same steps which led to Eq. (6.12), one finds:

$$\frac{\sigma_k^2}{(\sigma^{\mathrm{F}})^2} = \frac{1}{4} + \frac{\left(S + \sqrt{1+S^2}\right)^2}{2(1 + S(S + \sqrt{1+S^2}))} \frac{\omega_k^2}{\widehat{\omega}^2} \frac{\sigma_{k-1}^2}{(\sigma^{\mathrm{F}})^2}. \tag{6.25}$$

The MF version of the equation above obtained with $\omega_t = \omega$, admits a positive finite fixed point only if $\widehat{\omega}^2 > \omega_{c,S}^2|_{\mathrm{MF}} = \omega^2 \left(S + \sqrt{1+S^2}\right)^2 / [2(1 + S(S + \sqrt{1+S^2}))]$. Note that $\omega_{c,S}^2 \geq \omega_{c,S=0}^2$: therefore, if the market maker is risk-averse, the maximum error he can make (without preventing the stationary regime to establish) on the mean noisy order flow volatility is lower than that of the risk-neutral case. The contribution to the overall price volatility due to the risk-aversion of the market maker is qualitatively similar to what we investigated in Sec. 6.3.2, where we compared the critical value of market maker's belief about noisy order flow volatility in the presence or in absence of noisy order flow volatility fluctuations.

In the following we consider the risk-averse case with realistic values of the parameters, simulating the dynamics in the sticky expectation regime equilibria with uncorrelated noisy order flow volatility. We do so by setting an equal value for the mean excess volatility compatible with empirical results reported in the literature [15, 41], i.e., $\langle \sigma/(\sigma^{\mathrm{F}}) \rangle \sim 2.5$. For example, this level of excess volatility can be achieved in the risk-averse case with the choice of parameters $S = 0.1, \omega^2/\widehat{\omega}^2 = 1.75$ and $\delta\omega^2/\omega^2 = 0.15$. We obtain a price volatility correlation timescale of $\tau_{\mathrm{ACF}} \sim 25$ and an exponent $\mu \sim 5$ for the tail of the price volatility CDF. Since in empirical works the ACF timescale of price volatility is of the order of months, when asking for a realistic value of the annualized Sharpe ratio (in a competitive market one would expect it to be of order 1), one obtains plausible values for the daily Sharpe ratio S of the order of ~ 0.1. This illustrative example highlights the fact that in order to describe a reasonably competitive market ($S \sim 0.1$), excess volatility cannot be accounted for solely on the basis of risk aversion, and needs to be justified by a large negative bias in the estimation of the average noise trade level $\widehat{\omega}^2/\omega^2$, possibly boosted by the effect of the fluctuations. Indeed, one can explore the excess volatility for different values of r as we did in Sec. 6.3.2. As it is shown in Fig. 6.7, the mean excess volatility increases if the noise trade variance is correlated over time.

We thus provided a mechanism to generate a realistic mean excess-volatility level while recovering intermittent dynamics for price volatility which does not rely primarily on the risk-aversion of the market maker, providing a possible solution, in line with the AMH, for a long-lasting problem in economic literature [107].

6.4.2 Cost-averse noise trading, liquidity crises, and flash crashes

Up to now, we considered a passive noise trader, namely a noise trader who does not modify his trading intensity even if the price impact increases, resulting in higher trading costs. Moreover, the passive noise trader hypothesis implies that the level of liquidity is always finite and bounded from below by the constant liquidity provided by the noise trader. In reality, instead, the overall level of liquidity fluctuates over time; the market can experience liquidity crises, that is, situations in which the overall liquidity vanishes, while the price

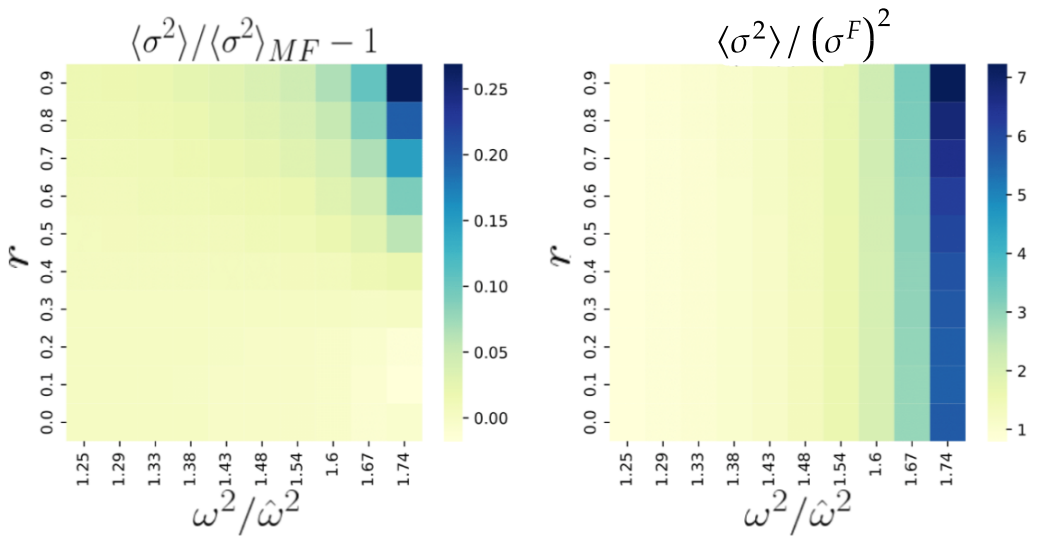


Figure 6.7: Mean excess variance as a function of the timescale ratio r and of market maker's underestimation of noise trade variance, measured by $\omega^2/\widehat{\omega}^2$ in the presence of a risk-averse market maker with $S = 0.1$. We set the variance of noise trade fluctuations to $\delta\omega^2/\omega^2 = 0.15$. (Left) Normalized by the mean-field level. One can observe the departure from the mean-field description, due combined effect of fluctuations and of market maker's underestimation of the noise trade variance. (Right) Mean excess variance. One can observe that a given level of excess variance can be obtained with different combinations of the parameters. The example analyzed in the main text corresponds to the set of parameters identified by the bottom-right corners.

impact goes virtually to infinity. In what follows we consider a cost-averse noise trader and we show how this modification can account for the fragility exhibited by financial markets.

The noise trader demand q_t^{NT} minimizes the expectation of a cost function C_t^{NT} made of two terms: the first is the usual profit and loss term multiplied by the cost-aversion parameter $\phi > 0$, while the second is the squared difference between the actual demand q_t^{NT} and a (moving) trading target q_t^{tgt}:

$$C_t^{\mathrm{NT}} = (q_t^{\mathrm{NT}} - q_t^{\mathrm{tgt}})^2 - \phi q_t^{\mathrm{NT}} p_t, \tag{6.26}$$

where q_t^{tgt} are realizations of a Gaussian process with zero mean and time-varying volatility ω_t which we assume to be known by the cost-averse noise trader. The time-varying volatility ω_t has an AR(1) structure. We assume that, as was the case for the informed trader, the noise trader can infer past price impact functions and use the last known value in order to construct his strategy. Accordingly, the trading strategy of the cost-averse noise trader reads:

$$q_t^{\mathrm{NT}} = \frac{q_t^{\mathrm{tgt}}}{1 + \phi G_{t-1}}, \tag{6.27}$$

We can relate the cost-aversion parameter ϕ with a tracking error $\xi > 0$ which measures how much the noise trader can afford to be off with respect to his trading target. The squared tracking error ξ^2 is defined as follows:

$$\xi^2 = \frac{\langle (q_t^{\mathrm{NT}} - q_t^{\mathrm{tgt}})^2 \rangle_\infty}{\omega^2} \tag{6.28}$$

where the subscript ∞ means that the noise trader sets his tracking error ξ assuming that the price impact is equal to its fixed point value, which we assume can be computed by

the noise trader. The link between the cost-aversion parameter ϕ and the tracking error parameter ξ is given by $\phi = 2\xi\widehat{\omega}/(\widehat{\sigma^{\mathrm{F}}})$.

We suppose that the risk-neutral market maker does not know the volatility of the noise trader's target, but he has a prior about it, namely, $\widehat{\omega}$. Moreover, we assume that the market maker knows the cost-aversion parameter ϕ (or, equivalently, ξ) of the noise trader, for simplicity. The dynamics of the price impact function is therefore again given by Eq. (6.6) where one has to take into account the new expression for the noise trade variance belief which stems from Eq. (6.27); this amounts to employ the substitution $\widehat{\omega}^2 \to \widehat{\omega}^2/(1+\phi G_{t-1})^2$ in Eq. (6.6). Accordingly, the long-time price impact function reads

$$G_\infty = \frac{1}{2(1-\xi)}\frac{\widehat{\sigma^{\mathrm{F}}}}{\widehat{\omega}}, \tag{6.29}$$

while the expected price volatility is given again by Eq. (6.8). Note that the price volatility is not affected by the noise trader's cost aversion ϕ (or ξ) in the case where the market maker knows with absolute precision this parameter. We assume that these fixed points are known by the noise trader, who uses such information to construct his strategy based on Eq. (6.28). The noise trade variance at the fixed point is $(1-\xi)^2\omega_t^2$ and vanishes if the noise trader's cost aversion is high, i.e., in the limit $\xi \to 1$; accordingly, also the overall liquidity scales with $(1-\xi)^2$. Therefore, the higher the cost-aversion of the noise trader, the smaller the overall level of liquidity in the market. At the same time, the price impact function diverges as $\xi \to 1$ such that the price volatility remains constant.

Interestingly, the cost-aversion of noise trades results in a kind of friction force that delays the approach of the price impact function to the fixed point. In fact, we find that the relaxation time related to the price impact dynamics with a constant belief is given by:

$$\tau_\xi = \frac{1}{1-\xi}. \tag{6.30}$$

An important difference with the case of the passive noise trader is that the relaxation timescale τ_ξ is now ξ-dependent. In particular, it diverges as the noise trader becomes extremely cost-averse ($\xi \to 1$), resulting in a market that evolves in a strongly out of equilibrium regime, where the liquidity vanishes, the price impact diverges, while price volatility remains bounded.

The overall liquidity fluctuates much more than what was implied by the first version of our model, where the strategy of the noise trader was cost-independent. We present this finding in Fig. 6.8, where we show the results of the modified model simulation in the sticky expectation regime given by Eq. (6.11). One can see in blue the results of simulations where the noise trader is cost neutral, while in orange one sees the results where the noise trader is cost averse. It is clear that the orange lines represent a regime where the liquidity and the price impact dynamics are more intertwined. In particular, the peak in the price impact function (orange line in the top panel) corresponds to the period in which the overall liquidity is small. Regarding the price volatility in presence of the cost-averse noise trader (orange line in bottom panel), while it rises in the proximity of the liquidity crises, the values attained are still comparable with those in absence of the noise trader's cost aversion. In fact, as we highlighted above, the equation for the price volatility does not change with respect to the case of a cost-neutral noise trader if the market maker knows exactly the cost aversion parameter of the noise trader.

To obtain a volatility profile that reacts to liquidity crises one has to relax the assumption that the market maker knows the true cost-aversion parameter of the noise trader ξ. With

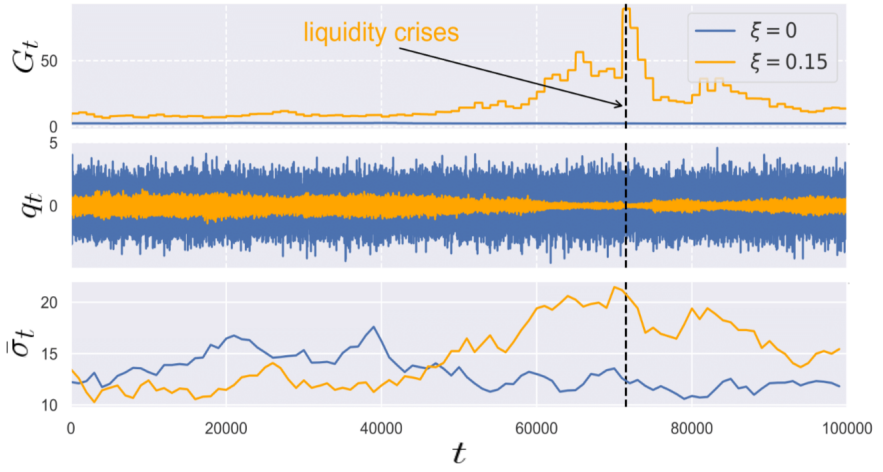


Figure 6.8: Comparison between sticky expectation regime dynamics with cost-averse and cost-neutral noise trader, with parameters $\widehat{\omega}^2/\omega^2 = 0.52, \delta\omega^2/\omega^2 = 0.15$ and $\tau_{\mathrm{NT}} = 0$. (Top) price impact dynamics. (Middle) Total excess-demand dynamics. (Bottom) Price volatility estimate. Blue lines are related to cost-neutral noise trading, while orange lines are related to the cost-averse case. The black dashed line points at the regime where, in the case of a cost-averse noise trader, the liquidity vanishes.

this modification, in fact, the resulting excess-volatility dynamics in the sticky expectation regime reads

$$\frac{\sigma_k^2}{(\sigma^{\mathrm{F}})^2} = \frac{1}{4}\left[1 + \left(\frac{1-\xi_k^2}{1-\hat{\xi}^2}\right)^2 \frac{\omega_k^2}{2\widehat{\omega}^2}\frac{\sigma_{k-1}^2}{(\sigma^{\mathrm{F}})^2}\right], \tag{6.31}$$

where $\hat{\xi}$ denotes the market maker's belief about the true tracking error parameter of the noise trader. The equation above shows that if the market maker overestimates the noise trader's cost aversion, and if $\hat{\xi} \to 1$, then the fixed point price volatility diverges, simulating a flash crash where the liquidity vanishes while the price volatility diverges.

6.5 Conclusion and outlook

Let us summarise what we have achieved. The aim of this Chapter was to modify the classic framework of asymmetrically informed agents in presence of noise in order to capture the excess volatility and volatility clustering exhibited by real financial markets, without resorting to unrealistic risk-averse agents [107] nor fundamental innovation clustering. We proposed a modification of the paradigmatic Kyle model of price formation, where agents adapt to the ever-evolving market conditions. Accordingly, each trader has its own model of reality, which does not generally match that of other traders; for example, we assumed that the market maker sets prices, while he does not know the precise level of signals and noise, that is, of the fundamental price and the non-informed trades variance. Moreover, we allowed the market maker to update his belief about the unobservable fundamental price variance by comparing the price variance expectations with empirical estimates. The market maker, therefore, acts based on a system of temporarily fulfilled expectations [39]. We analytically characterized the model in a realistic limit: the resulting stationary dynamics of excess volatility is of the Kesten type [44], i.e., a stochastic multiplicative process repelled from zero [119], which

exhibits intermittent dynamics and power law tails. Interestingly, power law behavior is robust against changes in the way market conditions evolve over time. This important finding is in line with the idea that adaptive behavior in the presence of noise, being a universal feature of human behavior, can be reflected in the universality of price dynamics across time and across markets. As a side, yet compelling, result, the microfoundation we propose is able to rationalize GARCH models while taking into account the relation of the price with fundamentals; as a consequence, the parameters that define the GARCH process are directly linked with non-observable quantities such as volatility of fundamentals and of noise trades. Hence, the microfoundation we propose enhances the interpretability as well as the predictive capacity of GARCH models. Our model predicts that some excess volatility can be accounted for by a mechanism based on quasi-non-ergodicity which has been recently proposed as a way to overcome the classic strong rationality paradigm [118]; in fact, we have shown that excess volatility is higher in situations where the updating timescale of the market maker is of the same order of magnitude as the timescale with which the market conditions vary.

The microfoundation we propose seems suitable to describe the large fraction of price jumps not directly linked to fundamental innovations' arrivals. In fact, it predicts symmetric price jumps, in line with empirical findings [113]; therefore, it points at the fact that symmetric GARCH models are more prone to describe price jumps not related to external fundamental innovations, at odds with the EMH story which relies on fundamental innovation clustering.

The present framework is versatile. For instance, in Sec. 6.4.2 we chose to analyze the case of a cost-averse noise trader, and showed that the more cost-aversion there is, the more fragile and illiquid the market is. If the market maker does not know precisely the noise trader's cost aversion, flash crashes can occur. Yet, we analyzed only the case of constant cost aversion. It could be of real interest to couple the cost-aversion of the noise trader and the updating timescale of the market maker's model with the current price volatility level. This coupling should lead to a more realistic description of the fragile and highly intertwined endogenous dynamics able to account for flash crashes in financial markets [121]. Another relevant extension, that may interest researchers dealing with financial contagion [122, 123], is to consider a multi-asset generalization. In fact, by assuming a network of interdependent fundamental prices, one can explore the extra fragility due to traders' interactions. Unlike Ref. [124], in which the effect of transaction costs is investigated, we consider the limit case in which spread costs are negligible, mimicking a market in which price movements are primarily driven by informational effects rather than by microstructural ones. Integrating transaction costs and, more generally, microstructural effects in our framework is an interesting extension that should be addressed along those lines. Another interesting extension is to consider several traders for each category, namely, informed traders and market makers (and cost-averse noise traders); this is expected to introduce additional effects, e.g., competition among market makers and the front-running of liquidity takers [125] and fat tails in the correlation volatility ACF [126–128]. Moreover, a description similar to that of Mean-Field Games [129] could emerge in the limit of infinitely many agents; note, however, that we don't expect the technical machinery of Mean Field Games to apply as it is since we are considering heuristic rules and not purely strategic behaviors. Finally, one could think of an extension with a correlated excess-demand process, in line with empirical observations [53]. See Refs. [85, 130, 131] for inspiration on how to tackle such ideas.

Take home messages from Chapter 6

1. The Kyle model has two major drawbacks if it is employed to describe real price dynamics. First, it is not able to account for the level of excess volatility found in empirical data without resorting to an unrealistically high risk-aversion parameter. Second, it is not able to capture the price volatility dynamics, which we investigated in Chapter 2.

2. We decided to break the assumption of perfect agents' rationality in the Kyle model to circumvent the above inconsistencies with empirical data. To do this, we replaced rational agents who use deductive reasoning to adaptive agents who use inductive reasoning.

3. We assumed that the external world parameters have dynamics on their own. We modeled this feature by assuming that the level of noise trading fluctuates over time.

4. We assumed that the market maker does not know with absolute precision the variance related to noise trader, nor the variance related to the fundamental price.

5. The learning dynamics of the market maker couples past prices with future ones, creating a feedback mechanism that induces a dynamics in the price volatility.

6. The dynamics of the price volatility is of the GARCH type. The assumption of periodic updates of the market maker's belief leads to a microfoundation of the GARCH$(1,1)$ model, which predicts price volatility dynamics affected by one single timescale.

7. Excess volatility is recovered without assuming a high risk-aversion parameter of the market maker but stems only from the mismatch between the market maker's beliefs about the true noise trade variance.

8. The model can be easily extended, by assuming a cost-averse noise trader, to account for liquidity crises, i.e., regimes in which the liquidity goes to zero, while the market fragility increases.

Conclusion and future research

My goal was to provide a bridge between the standard theoretical economics literature and well-known statistical models, having in mind that the modeler's ultimate aim is to explain the ubiquitous stylized facts that we observe in real financial data. To this end, I proposed two models rooted in the well-known framework proposed by Kyle in 1985, where rational agents with different information sets trade in the presence of noise [18]. This choice is motivated by two reasons: first, the Kyle model accounts for the trading mechanism in a stylized but effective way, which implies flexibility while allowing for tractability; second, my aim was to propose new models to a large audience including economists, and a model known to everyone who studies market microstructure sounded to me like a promising starting point. I tackled several drawbacks of the Kyle model.

In **Part II**, I dropped the unrealistic assumption of fundamental price revelation at a given terminal time. This allowed me to obtain the stationary Kyle (s-Kyle) model, presented in **Chapter 4**. In doing this, I provided a microfoundation for the Transient Impact model (TIM) [54], well known in econophysics, which allows for solving the diffusivity puzzle; interestingly, the TIM is recovered in a universal way in the high-frequency regime, where the price is diffusive; in the slow frequency regime, price dynamics exhibits a non-universal mean reversion as is the case in real financial markets. Although a very similar model was already obtained in the literature [68], I extended it to general non-markovian settings, which are needed to describe real markets [53]. I analyzed the model analytically in the markovian setting and numerically in the generic non-markovian one. I found two robust properties of the model, concerning the dynamics of excess demands and prices: the Auto-Correlation Function (ACF) of the excess demand has a temporal structure identical to the one resulting only from noise trading, which is reminiscent of standard results in multi-step Kyle frameworks and undergoes the name of insider's inconspicuousness. The price ACF reflects the temporal structure of the underlying fundamental price, albeit with a smaller amplitude; this last result is a consequence of the rationality assumptions and implies that the price volatility is smaller than the one related to the fundamental price, in stark contrast with empirical results [15].

Then, I presented in **Chapter 5** the calibration of the s-Kyle model against $\sim$ 150 years of S&P-500 data that included prices as well as dividends, from which I derived a fundamental price proxy. The results of the calibration highlighted several drawbacks of the s-Kyle model, inherited by the traders' rationality assumption: first, the empirical price variance was found to be ~ 6 times larger than the one reconstructed via the s-Kyle model; in light of the mapping to the TIM this amounts to saying that the predicted price impact function is qualitatively consistent in shape, but it is off in magnitude.

Having touched by hand the drawbacks of the traders' rationality assumption, I decided to modify the Kyle framework to be able to capture phenomena ranging from excess price volatility to price volatility clustering. I did this without resorting to assumptions such as extreme risk aversion of the market maker or high-frequency clustering of fundamental

innovations, which are not substantiated by empirical analyses. To accomplish this task, the discovery of the work by Hommes named "Behavioral Learning Equilibria" [106] was key: in that work, the author derives excess price volatility and excess price correlation persistence with respect to fundamentals by assuming that agents do not know the real underlying dynamics, which they try to infer by periodically updating a misspecified model. In this way, traders' rationality is broken assuming that the structural knowledge assumption does not hold, i.e., agents do not know the model as well as the model builder. The Kyle setup with adaptive agents, or behavioral Kyle (b-Kyle) model, presented in **Part III**, is built on similar assumptions but, contrary to Ref. [106], takes explicitly into account the coupled dynamics between prices and trades.

In **Chapter 6**, I gave a microfounded explanation of excess volatility and volatility clustering based on an endogenous feedback mechanism. I assume that traders do not know the exact parameters of the external world, which vary over time. The external world is parametrized by the variance of the noise trades and of the fundamental price. I assumed that the first varies over time, while the second is constant, mimicking the dynamics at high/medium frequency, where the sentiment dynamics, modeled by the noise trade variance, is important, while the fundamentals' one does not change on average. The agents have a wrong prior about the parameters of the external world and/or about the strategies of other players, and they adapt their own strategy as time goes by, reducing the gap between their priors and the actual price history. Since their model is wrong, this adaptive dynamics never settles down. Eventually, a realistic stationary equilibrium arises from the interaction between the traders' learning dynamics and the noisy environment, which affects and is affected by agents' beliefs. Contrary to the stationary equilibrium of the s-Kyle model, price volatility fluctuates forever, as in real markets. I characterized analytically the dynamics of the b-Kyle model in the regime of slow prior updating, giving a microfoundation to the well-known GARCH$(1,1)$ model. Therefore, we obtained a stationary equilibrium with power-law tails in the price and price volatility processes, in line with empirical analysis. I showed how to modify the b-Kyle model to model liquidity crises: by considering a cost-averse noise trader, I coupled order flow and price dynamics, so that when the price impact is high, the liquidity is small. The b-Kyle model is quite an effective attempt to provide a flexible microfoundation to financial markets. Figure 6.9 shows how I enriched the state of the art in market microstructure literature.

There are a few extensions of the b-Kyle model which I think can be very interesting as well as relevant.

- **Microfounded order splitting:** as we have seen in Chapter 4, I accounted for a non-markovian order flow in the stationary Kyle model only in an effective way. In other words, I did not propose a mechanism able to generate power-law-decaying ACF for the order flow directly from traders' interaction. A preliminary investigation [132], showed that it was possible to recover, within the stationary Kyle model, a correlated ACF of the order flow by considering a cost-averse noise trader, who needed to maintain his inventory close to a stochastic trading target. Quite intriguingly, the model exhibits a phase transition to an unstable market if the noise trader is too cost averse; in this case, the market maker is unable to find a finite positive value for the price impact function (a situation that closely resembles the market instability exhibited by the Glosten-Milgrom model [51]).

- **Microfounded mechanism to explain flash crashes:** as we have seen in Chapter 6, I was able to account for liquidity crises in the Kyle model with adaptive agents,

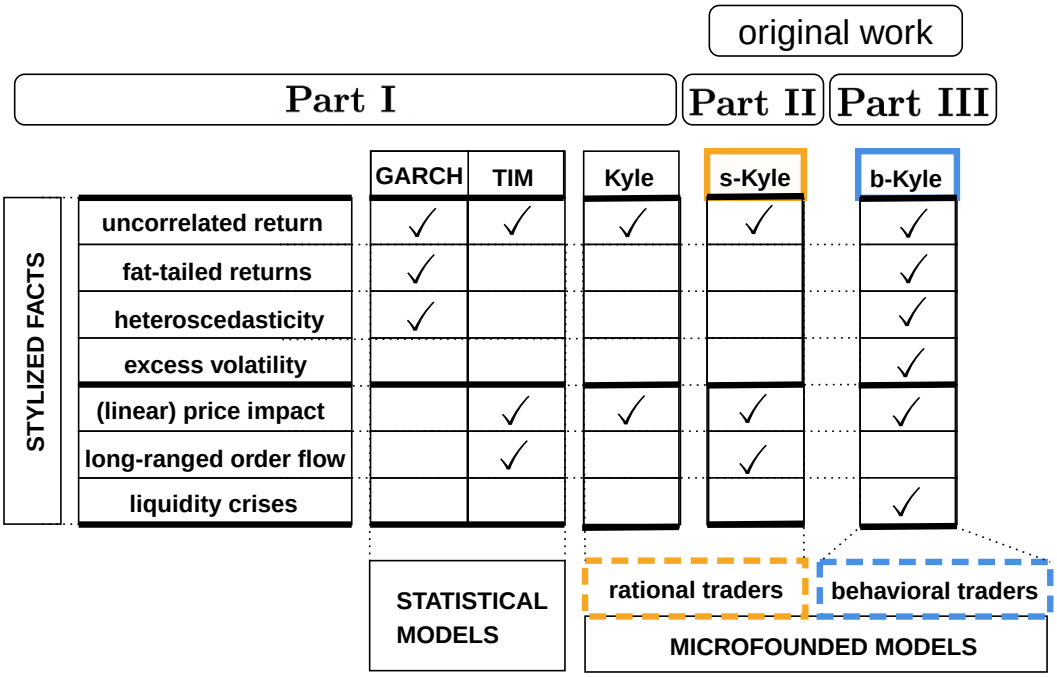


Figure 6.9: Thesis' roadmap. Part I contains useful material to understand the original work in the remaining parts of the thesis, with a particular focus on stylized facts and models able to rationalize (some) of them. Note how the original models I proposed in this manuscript enriched the predictive power with respect to the Kyle model, encompassing several stylized facts captured by ad-hoc statistical models.

by considering a cost-averse noise trader. To model a flash crash, i.e., a regime where not only the liquidity vanishes but the price volatility increases as well, an ingredient is missing: if the market maker does not know precisely the noise trader's cost-aversion, flash crashes can occur. Yet, we analyzed only the case where the market maker knows the constant level of cost aversion of the noise tader. It could be of real interest to couple the cost-aversion of the noise trader and the updating timescale of the market maker's model with the current price volatility level. This coupling should lead to a more realistic description of the fragile and highly intertwined endogenous dynamics able to account for flash crashes in financial markets [121].

- **Microfounded GARCH with long ranged ACF of price volatility:** as we have seen in Chapter 6, I was able to microfound the GARCH$(1,1)$ model in a specific limit of the b-Kyle model. In generic situations, the b-Kyle model is able to generate a price volatility correlation that decays as the sum of two exponentials, but it does not recover the realistic regime, where it decays as a power-law function. To reproduce the fact that price volatility decays in such a complex way, an ad hoc assumption has been put on the table, the so-called heterogeneous market hypothesis [133–135]. This assumption states that the market is populated by heterogeneous agents with different updating timescales. In contrast, our model assumes that the market maker uses a periodic updating rule.

  I think that a very interesting research question is: how the heterogeneous market assumption can be microfounded? Inspired by recent works on (market) ecology [136–139], I think a promising framework consists in thinking about traders in financial markets as animals in an environment with limited resources, i.e., as ecological com-

munities. How can the b-Kyle fit such a framework? The resources come from the breaking of the structural knowledge assumption; if agents do not know the parameters of the external world, (profit) space is available for an agent who understands before others the state of the world. In this setup, the heterogeneous market hypothesis can be microfounded by supposing that different market makers with different updating timescales exploit the strategy space as animals exploit different food niches. The ecological equilibrium should arise as the outcome of Darwinian dynamics where market makers change their strategies' time span by exploration/exploitation dynamics while undergoing a selection process (see, for example, Ref. [140]). Eventually, one should obtain a network of interacting traders, whose properties might be studied with tools and ideas already developed in theoretical ecology (see, for example, Ref. [141]).

I believe that my work demonstrated how a well-established framework in theoretical economics can be enriched to account for a more realistic phenomenology. Market microstructure is an incredibly complex field, where not only economic and mathematical backgrounds are useful; one needs also to account for notions coming from cognitive psychology, sociology as well as ecology. I truly hope that market microstructure can serve as a catalyst able to foster interactions between different research communities. I think that this interaction is the key to overcoming the biases that each discipline develops, and, therefore, to develop more and more realistic and useful models regarding how human actions affect and are affected by the environment in which they operate, be these financial markets, macro economies, or the planet Earth's biosphere.

Bibliography

[1] Victor M. Yakovenko and J. Barkley Rosser. Colloquium: Statistical mechanics of money, wealth, and income. *Rev. Mod. Phys.*, 81:1703–1725, Dec 2009.

[2] Massimo Borlandi. Expliquer les régularités sociales durkheim critique et continuateur de quetelet. *Durkheimian Studies / Études Durkheimiennes*, 14:9–20, 2008.

[3] Philip Ball. *Critical Mass: How One Thing Leads to Another*. Macmillan, 2003.

[4] Louis Bachelier. Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 3e série, 17:21–86, 1900.

[5] Albert Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 322(8):549–560, 1905.

[6] Eugene F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38 (1):34–105, 1965.

[7] Eugene F. Fama. Random walks in stock market prices. *Financial Analysts Journal*, 21(5):55–59, 1965.

[8] Eugene F. Fama. Two pillars of asset pricing. Technical Report 2013-8, Nobel Prize Committee, 2013.

[9] Paul A. Samuelson. Proof that properly discounted present values of assets vibrate randomly. *Industrial Management Review*, 6(2):41–49, 1965.

[10] Paul A. Samuelson. Rational theory of warrant pricing. *Industrial Management Review*, 6(2):13–39, 1965.

[11] Benoît Mandelbrot. Forecasts of future prices, unbiased markets, and "martingale" models. *The Journal of Business*, 39, 1965.

[12] Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[13] John F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335, 1961.

[14] Sanford J. Grossman and Joseph E. Stiglitz. On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408, 1980.

[15] R. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71:421–36, 01 1981.

[16] Stephen LeRoy and Richard D. Porter. The present-value relation: Tests based on implied variance bounds. *Econometrica*, 49(3):555–74, 1981.

[17] French. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, 17(1):5–26, 1986.

[18] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–35, 1985.

[19] Fischer Black. Noise. *The Journal of Finance*, 41(3):528–543, 1986.

[20] Robert J. Shiller. From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1):83–104, March 2003.

[21] Robert J. Shiller. Speculative Asset Prices. Technical Report 2013-6, Nobel Prize Committee, December 2013.

[22] Jean-Philippe Bouchaud. Economics need a scientific revolution. *Nature*, 455:1181, 11 2008.

[23] Jean-Philippe Bouchaud and Damien Challet. Why have asset price properties changed so little in 200 years. pages 3–17, 2017.

[24] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[25] Robert Shiller. Market volatility and investor behavior. *American Economic Review*, 80(2):58–62, 1990.

[26] John Y. Campbell, Andrew W. Lo, and A.Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[27] Andrew C. Szakmary, Qian Shen, and Subhash C. Sharma. Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking and Finance*, 34(2):409–426, 2010.

[28] Andrew Clare, James Seaton, Peter N. Smith, and Stephen Thomas. Trend following, risk parity and momentum in commodity futures. *International Review of Financial Analysis*, 31:1–12, 2014.

[29] Yves Lemperiere, Cyril Deremble, P. Seager, Marc Potters, and Jean-Philippe Bouchaud. Two centuries of trend following. *The Journal of Investment Strategies*, 3, 04 2014.

[30] Christof Schmidhuber. Trends, reversion, and critical phenomena in financial markets. *Physica A: Statistical Mechanics and its Applications*, 566:125642, 2021.

[31] Federico Guglielmo Morelli, Michael Benzaquen, Marco Tarzia, and Jean-Philippe Bouchaud. Confidence collapse in a multihousehold, self-reflexive dsge model. *Proceedings of the National Academy of Sciences*, 117(17):9244–9249, 2020.

[32] Federico Guglielmo Morelli, Michael Benzaquen, Jean-Philippe Bouchaud, and Marco Tarzia. Crisis propagation in a heterogeneous self-reflexive dsge model. *Plos one*, 16 (12):e0261423, 2021.

[33] Alan P. Kirman. Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives*, 6(2):117–136, Spring 1992.

[34] Jean-Philippe Bouchaud. Crises and collective socio-economic phenomena: simple models and challenges. Papers 1209.0453, arXiv.org, September 2012.

[35] Mehran Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007.

[36] Kenneth G. Wilson and Michael E. Fisher. Critical exponents in 3.99 dimensions. *Phys. Rev. Lett.*, 28:240–243, Jan 1972.

[37] Giorgio Parisi. Order parameter for spin-glasses. *Phys. Rev. Lett.*, 50:1946–1948, Jun 1983.

[38] Damien Challet and Yi-Cheng Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(3): 407–418, 1997.

[39] Brian W. Arthur. Inductive reasoning and bounded rationality. *American Economic Review*, 84(2):406–11, 1994.

[40] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. *Minority Games: Interacting agents in financial markets*. Oxford University Press, 2004.

[41] Jean-Philippe Bouchaud, Stefano Ciliberti, Yves Lempérière, Adam Majewski, Philippe Seager, and Kevin Sin Ronia. Black was right: Price is within a factor 2 of value. *SSRN Electronic Journal*, 11 2017.

[42] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

[43] Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press, 2003.

[44] Harry Kesten. Random difference equations and Renewal theory for products of random matrices. *Acta Mathematica*, 131(none):207 – 248, 1973.

[45] Robert Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[46] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[47] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[48] James Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22(1): 16–29, 2004.

[49] Drew Creal, Siem Jan Koopman, and André Lucas. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795, 2013.

[50] Erich L. Lehmann and George Casella. *Theory of Point Estimation*. Springer-Verlag, New York, NY, USA, second edition, 1998.

[51] Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, quotes and prices: financial markets under the microscope.* Cambridge University Press, 2018.

[52] Fabrizio Lillo. Order flow and price formation. Papers 2105.00521, arXiv.org, May 2021.

[53] Fabrizio Lillo and J. Doyne Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics Econometrics*, 8(3):1–35, 2004.

[54] 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*, 4(2):176–190, 2004.

[55] Bence Tóth, Imon Palit, Fabrizio Lillo, and J. Doyne Farmer. Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51(C):218–239, 2015.

[56] Joel Hasbrouck. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press, 2007.

[57] SANFORD J. GROSSMAN and MERTON H. MILLER. Liquidity and market structure. *The Journal of Finance*, 43(3):617–633, 1988. doi: https://doi.org/10.1111/j.1540-6261.1988.tb04594.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1988.tb04594.x`.

[58] Robin Greenwood. Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics*, 75(3):607–649, 2005. URL `https://EconPapers.repec.org/RePEc:eee:jfinec:v:75:y:2005:i:3:p:607-649`.

[59] Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340, 2013.

[60] Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516, 2016.

[61] Peter Bank, Ibrahim Ekren, and Johannes Muhle-Karbe. Liquidity in competitive dealer markets. *Mathematical Finance*, 31(3):827–856, 2021.

[62] Burton G. Malkiel. The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82, March 2003.

[63] Maureen O'Hara. *Market Microstructure Theory.* Blackwell Publishing Ltd, 1998.

[64] Jean-Philippe Bouchaud. Economics need a scientific revolution. *Nature*, 455:1181, 11 2008.

[65] Damien Eduardo Taranto, Giacomo Bormetti, Jean-Philippe Bouchaud, Fabrizio Lillo, and Bence Tóth. Linear models for the impact of order flow on prices. i. history dependent impact models. *Quantitative Finance*, 18(6):903–915, 2018.

[66] Michael Benzaquen, Iacopo Mastromatteo, Zoltan Eisler, and Jean-Philippe Bouchaud. Dissecting cross-impact on stock markets: an empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, 2017.

[67] Arman Joulin, Augustin Lefevre, Daniel Grunberg, and Jean-Philippe Bouchaud. Stock price jumps: News and volume play a minor role. *Wilmott Mag.*, 2008.

[68] Dan Bernhardt, Peter Seiler, and Bart Taub. Speculative dynamics. *Economic Theory*, 44(1):1–52, 2010.

[69] Lukas Hewing, Kim P. Wabersich, Marcel Menner, and Melanie N. Zeilinger. Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):269–296, 2020.

[70] D. Bernhardt and J. Miao. Informed trading when information becomes stale. *The Journal of Finance*, 59(1):339–390, 2004.

[71] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT press, 2012.

[72] Eric Benhamou. Kalman filter demystified: From intuition to probabilistic graphical model to real case in financial markets. *SSRN Electronic Journal*, 2018.

[73] Umut Çetin and Albina Danilova. Markovian nash equilibrium in financial markets with asymmetric information and related forward–backward systems. *Ann. Appl. Probab.*, 26(4):1996–2029, 2016.

[74] Kyung-Ha Cho. Continuous auctions and insider trading: Uniqueness and risk aversion. *Finance and Stochastics*, 7:47–71, 2003.

[75] Umut Çetin. Financial equilibrium with asymmetric information and random horizon. *Finance and Stochastics*, 22(1):97–126, 2018.

[76] Cheng Li and Hao Xing. Asymptotic glosten–milgrom equilibrium. *SSRN Electronic Journal*, 6, 2013.

[77] Jonathan Donier, Julius Bonart, Iacopo Mastromatteo, and Jean-Philippe. Bouchaud. A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*, 15, 2014.

[78] Alex Boulatov and Dmitry Livdan. Strategic trading with market closures. *Society for Economic Dynamics, 2006 Meeting Papers*, 01 2006.

[79] Allan G. Timmermann. Excess Volatility and Predictability of Stock Prices in Autoregressive Dividend Models with Learning. *The Review of Economic Studies*, 63(4): 523–557, 10 1996.

[80] Cars Hommes and Mei Zhu. Behavioral learning equilibria. *Journal of Economic Theory*, 150(C):778–814, 2014.

[81] Xavier Gabaix and Ralph S. J. Koijen. In search of the origins of financial fluctuations: The inelastic markets hypothesis. Working Paper 28967, National Bureau of Economic Research, 2021.

[82] James M. Poterba and Lawrence H. Summers. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22(1):27–59, 1988.

[83] Adam Majewski, Stefano Ciliberti, and Jean-Philippe Bouchaud. Co-existence of trend and value in financial markets: Estimating an extended Chiarella model. *Journal of Economic Dynamics and Control*, 112(C), 2020.

[84] Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

[85] Michele Vodret, Iacopo Mastromatteo, Bence Tóth, and Michael Benzaquen. A stationary kyle setup: microfounding propagator models. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(3):033410, 2021.

[86] 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.

[87] Avanidhar Subrahmanyam. Risk Aversion, Market Liquidity, and Price Efficiency. *Review of Financial Studies*, 4(3):416–441, 1991.

[88] Ming Guo and Albert S. Kyle. Dynamic strategic informed trading with risk-averse market makers. Working paper, Duke University, 2005.

[89] Allan G. Timmermann. How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics*, 108(4):1135–1145, 1993.

[90] William A. Branch. Imperfect knowledge, liquidity and bubbles. *Journal of Economic Dynamics and Control*, 62(C):17–42, 2016.

[91] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54, 1973.

[92] Benoit Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36:394–394, 1963.

[93] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[94] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[95] Emmanuel Bacry, J. Delour, and Jean-François Muzy. Multifractal random walk. *Phys. Rev. E*, 64:026103, Jul 2001.

[96] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[97] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[98] Gilles Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5):505–515, 2009.

[99] Mitja Steinbacher, Matthias Raddant, Fariba Karimi, Eva Cuena, Simone Alfarano, Giulia Iori, and Thomas Lux. Advances in the agent-based modeling of economic and social behavior. *SN Business and Economics*, 1, 07 2021.

[100] Andrew W. Lo. Efficient markets hypothesis. *The New Palgrave Dictionary of Economics*, 2:1–17, 2008.

[101] Andrew W. Lo. The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5):15–29, 2004.

[102] Lawrence Blume and David Easley. The market organism: Long-run survival in markets with heterogeneous traders. *Journal of Economic Dynamics and Control*, 33(5): 1023–1035, 2009.

[103] Pietro Dindo and Filippo Massari. The wisdom of the crowd in dynamic economies. *Theoretical Economics*, 15(4):1627–1668, 2020.

[104] Daniele Giachini. Rationality and asset prices under belief heterogeneity. *Journal of Evolutionary Economics*, 31(1):207–233, 2021.

[105] Giulio Bottazzi and Daniele Giachini. Far from the madding crowd: collective wisdom in prediction markets. *Quantitative Finance*, 19(9):1461–1471, 2019.

[106] Cars Hommes. Behavioral and experimental macroeconomics and policy analysis: A complex systems approach. *Journal of Economic Literature*, 59(1):149–219, March 2021.

[107] Stephen F. Leroy. Can risk aversion explain stock price volatility? *FRBSF Economic Letter*, 2013.

[108] Jean-Philippe Bouchaud. Economics need a scientific revolution. *Nature*, 455:1181, 11 2008.

[109] Salvatore Miccichè, Giovanni Bonanno, Fabrizio Lillo, and Rosario N Mantegna. Volatility in financial markets: stochastic models and empirical results. *Physica A: Statistical Mechanics and its Applications*, 314(1-4):756–761, 2002.

[110] Robert Engle. The nobel memorial prize for robert f. engle. *The Scandinavian Journal of Economics*, 106(2):165–185, 2004.

[111] David Cutler, James Poterba, and Lawrence Summers. What moves stock prices? NBER Working Papers 2538, National Bureau of Economic Research, Inc, 1988.

[112] Xiongfei Jiang, Tingting Chen, and Biaowen Zheng. Time-reversal asymmetry in financial systems. *Physica A Statistical Mechanics and its Applications*, 392, 08 2013.

[113] Riccardo Marcaccioli, Jean-Philippe Bouchaud, and Michael Benzaquen. Exogenous and endogenous price jumps belong to different dynamical classes. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(2):023403, feb 2022.

[114] Guido Ascari and Yifan Zhang. Limited memory, time-varying expectations and asset pricing. *working paper*, 2022.

[115] Sabiou M. Inoua. News-driven expectations and volatility clustering. *Journal of Risk and Financial Management*, 13(1), 2020.

[116] Hans Follmer, Ulrich Horst, and Alan Kirman. Equilibria in financial markets with heterogeneous agents: a probabilistic perspective. *Journal of Mathematical Economics*, 41(1-2):123–155, February 2005.

[117] Rosario N Mantegna and H Eugene Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

[118] Roger Farmer and Jean-Philippe Bouchaud. Self-fulfilling prophecies, quasi non-ergodicity and wealth inequality. Working Paper 28261, National Bureau of Economic Research, December 2020.

[119] Rama Cont and Didier Sornette. Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *Journal De Physique I*, 7, 09 1996.

[120] Jean-Philippe Bouchaud and Marc Mézard. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, 282(3):536–545, 2000.

[121] Antoine Fosset, Jean-Philippe Bouchaud, and Michael Benzaquen. Endogenous Liquidity Crises. *Journal of Statistical Mechanics: Theory and Experiment*, 2020. 21 pages, 11 figures, 1 table.

[122] Stanislao Gualdi, Giulio Cimini, Kevin Primicerio, Riccardo Di Clemente, and Damien Challet. Statistically validated network of portfolio overlaps and systemic risk. *Scientific Reports*, 6, 12 2016.

[123] Sebastian Poledna, Serafín Martínez-Jaramillo, Fabio Caccioli, and Stefan Thurner. Quantification of systemic risk from overlapping portfolios in the financial system. *Journal of Financial Stability*, 52:100808, 2021.

[124] Charles-Albert Lehalle, Eyal Neuman, and Segev Shlomov. Phase transitions in kyle's model with market maker profit incentives. *SSRN Electronic Journal*, 01 2021.

[125] Ioanid Roşu. Fast and slow informed trading. *Journal of Financial Markets*, 43:1–30, 2019.

[126] François Ghoulmie, Rama Cont, and Jean-Pierre Nadal. Heterogeneity and feedback in an agent-based market model. *Journal of Physics: Condensed Matter*, 17(14): S1259, mar 2005. doi: 10.1088/0953-8984/17/14/015. URL `https://dx.doi.org/10.1088/0953-8984/17/14/015`.

[127] Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007.

[128] Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, 2022.

[129] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2:229–260, 03 2007.

[130] Michele Vodret, Iacopo Mastromatteo, Bence Tóth, and Michael Benzaquen. Do fundamentals shape the price response? A critical assessment of linear impact models. *arXiv preprint arXiv:2112.04245*, December 2021.

[131] Francesco Cordoni and Fabrizio Lillo. Transient impact from the nash equilibrium of a permanent market impact game. *arXiv preprint arXiv:2205.00494*, 2022.

[132] Felipe Moret, Michele Vodret, Iacopo Mastromatteo, Bence Tóth, and Michael Benzaquen. On the origin of long-range correlations in trading activity. *in preparation*.

[133] Ulrich A. Müller, Michel M. Dacorogna, Rakhal D. Davé, Olivier V. Pictet, Richard B. Olsen, and J. Robert Ward. Fractals and intrinsic time: A challenge to econometricians. *Unpublished manuscript, Olsen & Associates, Zürich*, 130, 1993.

[134] Ulrich A. Muller, Michel M. Dacorogna, Rakhal D. Dave, Richard B. Olsen, Olivier V. Pictet, and Jacob E. von Weizsacker. Volatilities of different time resolutions – Analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2-3): 213–239, June 1997.

[135] Fulvio Corsi. A simple approximate long-memory model of realized volatility. *The Journal of Financial Econometrics*, 7(2):174–196, 2009.

[136] J. Doyne Farmer. Market Force, Ecology, and Evolution. Computing in Economics and Finance 1999 651, Society for Computational Economics, March 1999.

[137] Federico Musciotto, Luca Marotta, Jyrki Piilo, and Rosario N. Mantegna. Long-term ecology of investors in a financial market. *Palgrave Communications*, 4:1–12, 2018.

[138] Maarten P. Scholl, Anisoara Calinescu, and J. Doyne Farmer. How market ecology explains market malfunction. *Proceedings of the National Academy of Sciences*, 118 (26):e2015574118, 2021.

[139] Simon A. Levin and Andrew W. Lo. Introduction to pnas special issue on evolutionary models of financial markets. *Proceedings of the National Academy of Sciences*, 118(26): e2104800118, 2021.

[140] Juan Keymer, Miguel Fuentes, and Pablo Marquet. Diversity emerging: From competitive exclusion to neutral coexistence in ecosystems. *Theoretical Ecology*, 5, 05 2012.

[141] Brian Fath. Network mutualism: Positive community-level relations in ecosystems. *Ecological Modelling*, 208(1):56–67, 2007. Special Issue on Ecological Network Theory.

[142] Linda S. L. Tan. Explicit inverse of tridiagonal matrix with applications in autoregressive modelling. *IMA Journal of Applied Mathematics*, 84(4):679–695, 2019.

[143] David M. Cutler, James M. Poterba, and Lawrence H. Summers. What moves stock prices? Working Paper 2538, National Bureau of Economic Research, March 1988.

[144] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, pages 5–39, 2001.

[145] Dimitris Bertsimas and Andrew W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[146] Álvaro Cartea and Sebastian Jaimungal. Incorporating order-flow into optimal execution. *Mathematics and Financial Economics*, 10(3):339–364, 2016.

[147] Jim Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10 (7):749–759, 2010.

[148] Aurélien Alfonsi, Antje Fruth, and Alexander Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

[149] Iacopo Mastromatteo, Bence Tóth, and Jean-Philippe Bouchaud. Agent-based models for latent liquidity and concave price impact. *Phys. Rev. E*, 89:042805, 2014.

Appendices

Appendix A

Calibration of statistical models

A.1 Calibration of the GARCH$(1,1)$ model

The outcome of the calibration is given by: $\theta = \{\sigma_0, \alpha_1, \beta_1\}$. Given a string of returns $\{r_t\}_{t=1}^T$, we are interested in the joint distribution:

$$\mathbb{P}(\epsilon_0, \cdots, \epsilon_T; \theta) = \mathbb{P}(r_0; \theta)\mathbb{P}(r_1, \cdots, r_t | r_0; \theta) \tag{A.1}$$

$$= \mathbb{P}(r_0; \theta) \prod_{t=1}^{T} \mathbb{P}(r_t | r_{t-1}, \cdots, r_0; \theta) \tag{A.2}$$

$$= \mathbb{P}(r_0; \theta) \prod_{t=1}^{T} \mathbb{P}(r_t | r_{t-1}; \theta) \tag{A.3}$$

$$= \mathbb{P}(r_0; \theta) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{r_t^2}{2\sigma_t^2}\right). \tag{A.4}$$

Dropping $\mathbb{P}(r_0; \theta)$ and taking logs, we obtain the conditional log-likelihood function:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \frac{1}{2}\left[-2\log 2\pi - \log\sigma_t^2 - \frac{r_t^2}{\sigma_t^2}\right], \tag{A.5}$$

which is maximized by the optimal set of parameters θ.

A.2 Calibration of the Propagator model

In principle, since the response function and the trade signs ACF is empirically measurable, one could obtain the price impact function given the following equation:

$$R_t = G_t + \sum_{0<t'<t} G_{t-t'} C_{t'}^{\epsilon} + \sum_{t'>0} [G_{t+t'} - G_{t'}] C_{t'}^{\epsilon}. \tag{A.6}$$

However, this direct method is very sensitive to finite-size effects, providing a poor estimate of G_t.

An alternative calibration procedure consists in using the lagged sign-return correlation, defined as

$$S_\tau = \langle r_{t+\tau} \epsilon_t \rangle = R_{t+1} - R_t. \tag{A.7}$$

By substituting Eq. (3.6) into the equation above one obtains:

$$S_t = \sum_{t' \geq 0} C_{|t'-t|} K_{t'}, \tag{A.8}$$

which can be rewritten in matrix form as

$$\boldsymbol{S} = \mathsf{C}\boldsymbol{K}. \tag{A.9}$$

Inverting the above equation provides an explicit expression for K_t:

$$K_t = \sum_{t'=0}^{T} (\mathsf{C}^{-1})_{t,t'} S_{t'}. \tag{A.10}$$

Appendix B

A stationary Kyle setup

B.1 Numerical solver

The iterative numerical scheme is as follows:

- Choose a maximum time $T_{cut} - 1$ which is the maximum time lag at which the propagator can be evaluated. In doing so the propagator is a vector of T_{cut} elements.
- Choose a "seed" propagator.
- Plug this seed in the r.h.s. of Eq. (4.23).
- Insert the result obtained with this procedure in the r.h.s. for a number of iterations equal to T_{it}, checking for convergence.

The only issue of this procedure is the following: as one can see from the first of Eqs. (4.22), in order to compute $\mathbf{R}_t$ one has to evaluate the block matrix given by $\mathsf{G}_{/t,t}$. This matrix has entries that cannot be calculated, due to the truncation constraint of our numerical procedure. Nevertheless, because of the mean-reverting assumption of the dividends, we know that the propagator should decay to zero at large times, so the large lags terms in $\mathsf{G}_{/t,t}$ can be simply set to zero.

B.1.1 Convergence

In this Section we give further details about the convergence of the results of the iterative numerical solution of Eq. (4.12).

In Fig. B.1 we show results about the relative cumulative absolute error for the price ACF Σ_τ and the excess demand ACF Ω_τ. The first one is calculated as in Eq. (4.25), while the second one is given by (4.27).

We choose $T_{it} = 100$, power law ACFs for dividends and NT's trades. We plot the result for $T_{cut,i} = \Delta t \times i$, for different i. The plots on the left are obtained with a power law ACF that decays faster than the one used to obtain the plots on the right. We can see, as expected, that the slower the decay of the power law, the slower the convergence.

We have investigated the behavior of the error for higher T_{it}, but we didn't find quantitative differences.

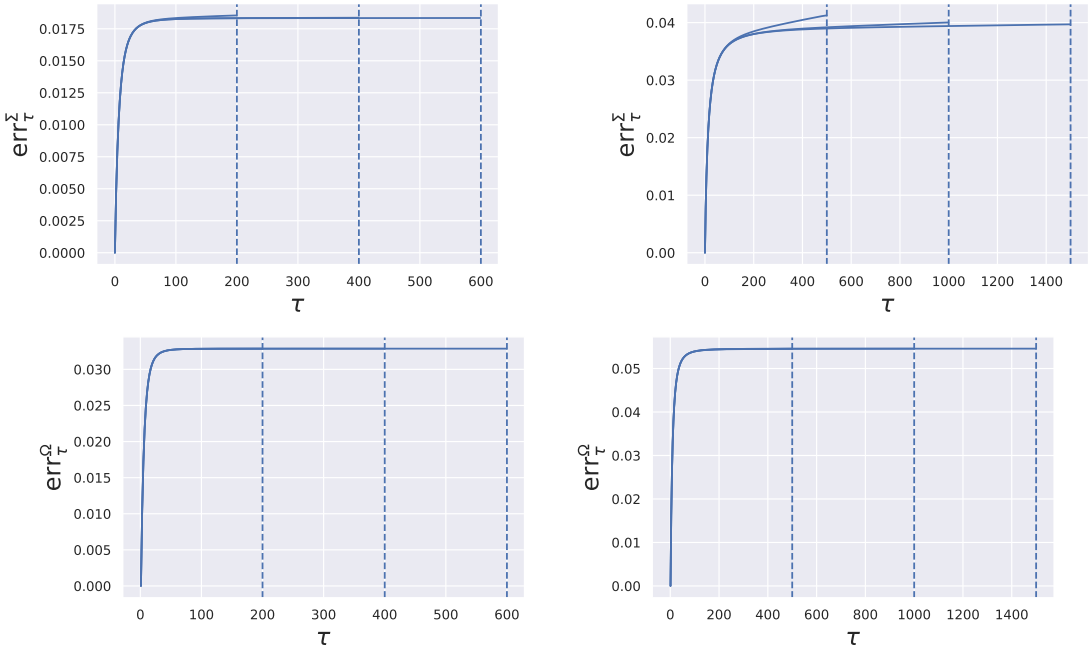


Figure B.1: Numerical check of equilibrium properties with ACFs given by $(1+|\tau|/\tau_k)^{-\gamma_k}$ where $k=\{\mu,\mathrm{NT}\}$. We arbitrarily choose $\tau_{\mathrm{NT}}=\tau_\mu=10$. (Left panels) $\gamma_{\mathrm{NT}}=\gamma_\mu=5$ and $\Delta t=200$. (Right panels) $\gamma_{\mathrm{NT}}=\gamma_\mu=3.5$ and $\Delta t=500$.

B.2 Particular solutions of equilibrium condition in the Markovian case

B.2.1 The case of non-correlated Noise

In the case of non-correlated NT's trades, the IT's forecast of future NT's trades is zero, and so the demand kernel R^{NT}, explicitly given in Eqs. (4.22), is zero. Since we are dealing with a Markovian dividend process, the IT's forecast at time t of future dividends relies only on the last known dividend, i.e., μ_{t-1} and so $\mathsf{R}^{\mu}=R^{\mu}\mathsf{I}$, where R^{μ} is a scalar.

The self-consistent equilibrium condition given by Eq. (4.23) for the dimensionless propagator is given by:

$$\tilde{G}_{t-t'}=\frac{1}{1-\alpha_\mu}\mathbf{e}_t^{\top}\tilde{\Gamma}(\mathsf{I}-\mathsf{RL}), \tag{B.1}$$

where

$$\tilde{\Gamma}=\left[(\tilde{\Xi}^{\mu})^{-1}+(\tilde{\mathsf{R}}^{\mu}\mathsf{L})^{\top}\tilde{\mathsf{R}}^{\mu}\mathsf{L}\right]^{-1}(\tilde{\mathsf{R}}^{\mu}\mathsf{L})^{\top}. \tag{B.2}$$

The solution of Eq. (B.1) is constructed in three steps. i) First we analyze the vector $\mathbf{e}_t^{\top}\tilde{\Gamma}$ and we show that it is related to the inverse of a tri-diagonal matrix with modified corner elements, for which the explicit expression is known [142]. Then, ii) we prove that a single exponential propagator solves Eq. (B.1) and we identify the amplitude and the timescale of the propagator in terms of α_μ and R^{μ}. iii) Finally, we can calculate the expression of R^{μ} in terms of α_μ from its general expression given in Eqs. (4.22). In this way, we fix completely the shape of the propagator only in terms of α_μ.

i) Since in the Markovian case R^{μ} is proportional to the identity matrix, from Eq. (B.2) we obtain:

$$\mathbf{e}_t^{\top}\tilde{\Gamma}=(a_t,\mathbf{b}_{t-1})(\tilde{\mathsf{R}}^{\mu}\mathsf{L})^{\top}=\tilde{R}^{\mu}\mathbf{b}_{t-1}, \tag{B.3}$$

where the vector $\mathbf{b}_{t-1}$ can be found by means of the block matrix inverse formula applied to the matrix inside the square brackets of Eq. (B.2), given by:

$$M = (\tilde{\Xi}^{\mu})^{-1} + (\tilde{\mathsf{R}}^{\mu}\mathsf{L})^{\top}\tilde{\mathsf{R}}^{\mu}\mathsf{L} = \left[\begin{array}{c|c} a_t & \mathbf{B}_{t-1} \\ \hline \mathbf{B}_{t-1}^{\top} & \mathsf{C} \end{array}\right]. \tag{B.4}$$

In particular, using the block inverse formula, the vector $\mathbf{b}_{t-1}$ is given by

$$\mathbf{b}_{t-1} = -a_t^{-1}\mathbf{B}_{t-1}^{\top}(M/a_t)^{-1} = \alpha_{\mu}\mathbf{e}_{t-1}^{\top}(M/a_t)^{-1}, \tag{B.5}$$

where the last equality has been obtained with the following property (checked by direct inspection of Eq. (B.4)): $\mathbf{B}_{t-1}^{\top} \propto \mathbf{e}_{t-1}^{\top}$. (M/a_t) is the Schur's complement of M with respect to a_t, which is given by

$$(M/a_t) = \mathsf{C} - \mathbf{B}_{t-1}^{\top}a_t^{-1}\mathbf{B}_{t-1} = (\tilde{\Xi}^{\mu})^{-1} + (\tilde{R}^{\mu})^2\mathsf{I}. \tag{B.6}$$

(M/a_t) is a tri-diagonal matrix with modified corner elements. Thus, the inverse of the Schur's complement of M with respect to a_t can be calculated explicitly (see Ref.[142]). The explicit expression of Eq. (B.5) is given by a single decaying exponential:

$$\mathbf{b}_{t-1} = b_0\{\gamma^{\tau}\}_{\tau=0}^{\infty}, \tag{B.7}$$

where

$$b_0 = \alpha_{\mu}\frac{(\tilde{R}^{\mu})^2 - g}{(\tilde{R}^{\mu})^4}, \quad \gamma = \frac{g\alpha_{\mu}}{(\tilde{R}^{\mu})^2} \tag{B.8}$$

and g is given by:

$$g = \frac{\beta - \sqrt{\beta^2 - 4}}{2(\tilde{R}^{\mu})^{-2}\alpha_{\mu}}, \quad \beta = \frac{(\tilde{R}^{\mu})^{-2} + 1 + (\tilde{R}^{\mu})^{-2}\alpha_{\mu}^2 - \alpha_{\mu}^2}{(\tilde{R}^{\mu})^{-2}\alpha_{\mu}}, \tag{B.9}$$

so that $\mathbf{b}_{t-1}$ is completely specified by α_{μ} and $\tilde{R}^{\mu}$.

ii) We are going to prove that an ansatz for the propagator given by a decaying exponential towards zero actually solves Eq. (B.1). The ansatz for the propagator reads:

$$G_{t-t'} = G_0\rho^{t-t'}. \tag{B.10}$$

As a preliminary result, from this ansatz, one can compute the elements of the vector $\mathbf{R}_t$, which appear in Eq. (B.1), by means of the first equation in Eqs. (4.22) . This is given by:

$$R_{t-t'} = -R_0\rho^{t-t'}, \tag{B.11}$$

where

$$R_0 = 1 - g_s, \quad g_s = \frac{1 - \sqrt{1-\rho^2}}{\rho^2}. \tag{B.12}$$

Equipped with this result, together with Eq. (B.7) one can easily show that Eq. (B.1) is solved with the ansatz given by Eq. (B.10). The ansatz is constraint to satisfy the following equations:

$$\tilde{G}_0 = \frac{b_0}{1-\alpha_{\mu}}, \ \rho = \frac{\gamma}{(1-R_0)}. \tag{B.13}$$

iii) Since we proved that G is of the exponential form we are now able to compute the explicit form of R_μ, starting from its definition in Eq. (4.22). The explicit expression for $\tilde{R}_\mu$, which completely specifies R^μ, is given by:

$$R^\mu = \frac{1}{(1-\alpha)G_0}\left(\alpha_\mu g_s - \frac{\alpha_\mu^2}{\rho}\frac{1-(2-\rho^2)g_s}{1-g_s\rho\alpha_\mu}\right). \tag{B.14}$$

Now, we can use insert in the above equation the expression for g , g_s, G_0, ρ given respectively by Eqs. (B.9), (B.12) and (B.13) and solve for $\tilde{R}^\mu$. In doing so we find

$$\tilde{R}^\mu = 1. \tag{B.15}$$

Finally, reintroducing the variance terms, i.e., using $\Omega_\tau^{\rm NT} = \Omega_0^{\rm NT}\delta_\tau$ and $\Xi_\tau^\mu = \Xi_0^\mu \alpha_\mu^\tau$, then the solution to Eq. (B.1) is given by

$$G_\tau = \left(\frac{\Xi_0^\mu}{\Omega_0^{\rm NT}}\right)^{1/2}\frac{\alpha_\mu}{1-\alpha_\mu}\left(1-\frac{1-\sqrt{1-\alpha_\mu^2}}{\alpha_\mu^2}\right)\alpha_\mu^\tau. \tag{B.16}$$

B.2.2 The case of Noise and Signal with equal autocovariance timescales

In this section we deal with the Markovian case specified by Eqs. (4.30) with $\alpha_\mu = \alpha_{\rm NT}$. A difference with the previous case is given by the fact that now $\mathsf{R}^{\rm NT} = R^{\rm NT}\mathsf{I}$, where $R^{\rm NT}$ is a nonzero scalar. The solution of the self-consistent equilibrium condition (4.23) is akin to the one exposed in the previous section, due to a simplification induced by the assumption given by $\alpha_\mu = \alpha_{\rm NT}$. In order to show this we define $\mathsf{E}^{\rm NT}$ as:

$$\mathsf{E}_t^{\rm NT} = \left(\mathsf{I} + \mathsf{R}^{\rm NT}\mathsf{L}\right)\Omega^{\rm NT}\left(\mathsf{I} + \mathsf{R}^{\rm NT}\mathsf{L}\right)^\top. \tag{B.17}$$

The simplification is the following:

$$\left\{\left[(\Xi^\mu)^{-1} + (\mathsf{R}^\mu\mathsf{L})^\top(\mathsf{E}^{\rm NT})^{-1}\mathsf{R}^\mu\mathsf{L}\right]^{-1}(\mathsf{R}^\mu\mathsf{L})^\top(\mathsf{E}^{\rm NT})^{-1}\right\}_{t,t'} = \alpha_\mu\mathsf{R}^\mu\mathsf{L}\left\{\left[\mathsf{E}^{\rm NT}(\Xi^\mu)^{-1} + (R^\mu)^2\mathsf{I}\right]^{-1}\right\}_{t,t'}, \tag{B.18}$$

where the matrix inside the square bracket on the r.h.s. is a tri-diagonal matrix with modified corner elements, for which, as seen before, analytical results are available. Thus, akin to the previous case, a propagator given by a single exponential decay term given by Eq. (B.10) is a solution of the self-consistent equation for the propagator (4.23). The result of the calculation that we do not report here is given by

$$R^{\rm NT} : (R^{\rm NT})^4 - 3(R^{\rm NT})^2\alpha_\mu^2 + R^{\rm NT}\left(2\alpha_\mu^3 + 2\alpha_\mu\right) - \alpha_\mu^2 = 0, \tag{B.19}$$

where one has to retain the only positive real solution. Then,

$$R^\mu = \sqrt{\frac{\Omega_0^{\rm NT}}{\Xi_0^\mu}}\sqrt{1 + (R^{\rm NT})^2 + 2R^{\rm NT}\alpha_\mu}, \tag{B.20}$$

$$\rho = \frac{R^{\rm NT}}{1 + (R^{\rm NT})^2 + R^{\rm NT}\alpha_\mu} \tag{B.21}$$

and finally

$$G_0 = \sqrt{\frac{\Xi_0^\mu}{\Omega_0^{\mathrm{NT}}}} \frac{\alpha_\mu \sqrt{(R^{\mathrm{NT}})^2 - 2\alpha_\mu R^{\mathrm{NT}} + 1}}{(1-\alpha_\mu) R^{\mathrm{NT}} \left(-3\alpha_\mu + R^{\mathrm{NT}} \left(2 - \frac{1}{((R^{\mathrm{NT}})^2 - \alpha_\mu R^{\mathrm{NT}} + 1)\left(\sqrt{1 - \frac{(R^{\mathrm{NT}})^2}{((R^{\mathrm{NT}})^2 - \alpha_\mu R^{\mathrm{NT}} + 1)^2}} + 1 \right)} \right) + \frac{2}{R^{\mathrm{NT}}} \right)}. \tag{B.22}$$

B.3 Solution of the Markovian case

B.3.1 Construction of the Ansatz

In this section we prove the results presented in Sec. 4.6.1, in particular Eqs. (4.32) and (4.33). i) First, we rewrite the property exposed in Eq. (4.31) in expectation form. ii) Then we inject in this form the quasi-camouflage property and we find a simple finite-difference equation for the propagator whose solution gives the formulas presented in Eqs. (4.32) and (4.33).

i) If the price ACF is exponentially decaying with the dividends timescale, as found by means of the numerical solver, then the following relation holds:

$$\mathbb{E}[p_{t+1} | \mathcal{I}_t^{\mathrm{MM}}] = \alpha_\mu p_t. \tag{B.23}$$

Equation (B.23) gives us a relation between the excess demand ACF and the propagator. In fact using the equation that defines the propagator model, i.e., $p_t = \sum_{t' \leq t} G_{t-t'} q_{t'}$, it can be rewritten as:

$$G_0 \mathbb{E}[q_{t+1} | \mathcal{I}_t^{\mathrm{MM}}] = \alpha_\mu \sum_{t'=-\infty}^{t} G_{t-t'} q_{t'} - \sum_{t'=-\infty}^{t} G_{t+1-t'} q_{t'}. \tag{B.24}$$

This equation is particularly interesting and it holds regardless of the structure of the NT's trades auto-covariance.

Let us give a first example of how the above equation can be used in order to derive the result about non-correlated NT's trades. The camouflage is exact in this case, so the excess demands are uncorrelated, i.e., the l.h.s. of the above equation is zero, then we can see that G decays itself exponentially with the dividends time-scale. This is precisely what happens if the noise trades are not correlated, where the propagator is given by Eq. (B.16).

ii) In the following we deal with the case of arbitrary Markovian NT's trades process. Using the expression of the general forecast matrix of a Gaussian process with zero mean, we can rewrite Eq. (B.24), as

$$\left[(\tilde{\Omega})_0^{-1} \right]^{-1} (\tilde{\Omega})_{t+1-t'}^{-1} = \tilde{G}_{t+1-t'} - \alpha_\mu \tilde{G}_{t-t'}, \quad \tilde{G}_\tau = G_\tau / G_0. \tag{B.25}$$

Since we found that in generic situations an approximate camouflage relation holds, we know that the structure of the excess demand ACF matrix is given by Eq. (4.26). The inverse of the excess demand ACF can be computed, and it is given by:

$$(\tilde{\Omega})_0^{-1} = \frac{\omega - \sqrt{\omega^2 - 4}}{2\tilde{b}\alpha_{\mathrm{NT}}}, \qquad \omega = \frac{\tilde{b} + 1 + \tilde{b}\alpha_{\mathrm{NT}}^2 - \alpha_{\mathrm{NT}}^2}{\tilde{b}\alpha_{\mathrm{NT}}}, \tag{B.26}$$

and

$$(\tilde{\Omega})^{-1}_{t+1-t'} = -\frac{1}{\alpha_{\rm NT}\tilde{b}}\left(1-(1+\tilde{b}-\alpha^2_{\rm NT})(\tilde{\Omega})^{-1}_0\right)\left[(\tilde{\Omega})^{-1}_0\alpha_{\rm NT}\tilde{b}\right]^{t-t'}. \tag{B.27}$$

Then, we can rewrite Eq. (B.25) as

$$\tilde{G}_{t+1-t'} = \alpha_\mu \tilde{G}_{t-t'} + P\rho^{t-t'}, \tag{B.28}$$

where we defined

$$P = -\frac{1}{\alpha_{\rm NT}\tilde{b}}\left(1-(1+\tilde{b}-\alpha^2_{\rm NT})(\tilde{\Omega})^{-1}_0\right)\left[(\tilde{\Omega})^{-1}_0\right]^{-1}, \qquad \rho = (\tilde{\Omega})^{-1}_0\alpha_{\rm NT}\tilde{b}. \tag{B.29}$$

The solution of Eq. (B.28) is Eq. (4.32) introduced in the main text. Moreover the second equation in Eqs. (B.29) gives Eq. (4.33).

B.3.2 Solving the ansatz

In this Appendix, we present the calculations which allowed us to obtain the results presented in the figures of Secs. 4.6.2, 4.6.3 and 4.6.4.

From the expression of the propagator given by Eq. (4.32), one is able to derive the inverse of the symmetrized propagator, which is given by

$$(\tilde{\mathsf{G}}^{\rm sym})^{-1}_{t,t'} = \Gamma_1\gamma_1^{t-t'} + \Gamma_2\gamma_2^{t-t'} + \delta(t-t'), \tag{B.30}$$

where Γ_1 and Γ_2 are the solution of the following set of equations:

$$\begin{aligned}
\Gamma_1\frac{\alpha_\mu}{\alpha_\mu-\gamma_1} + \Gamma_2\frac{\alpha_\mu}{\alpha_\mu-\gamma_2} + 1 = 0,\\
\Gamma_1\frac{\rho}{\rho-\gamma_1} + \Gamma_2\frac{\rho}{\rho-\gamma_2} + 1 = 0,
\end{aligned} \tag{B.31}$$

whereas γ_1 and γ_2 are the two real positive solution of the equation below:

$$\frac{\alpha_\mu-\alpha_{\rm NT}}{\alpha_\mu-\rho}\left(\frac{1}{1-\alpha_\mu\gamma_1}-\frac{\alpha_\mu}{\alpha_\mu-\gamma_1}\right)+\left(1-\frac{\alpha_\mu-\alpha_{\rm NT}}{\alpha_\mu-\rho}\right)\left(\frac{1}{1-\rho\gamma_1}-\frac{\rho}{\rho-\gamma_1}\right)+1=0. \tag{B.32}$$

With the explicit expression of G^{sym} given above one is able to calculate the IT's demand Kernels given by Eqs. (4.22). These are given by

$$\begin{aligned}
R_{t-t'} &= -\alpha^{t-t'}\frac{\alpha_\mu-\alpha_{\rm NT}}{\alpha_\mu-\rho}\left(\frac{\Gamma_1}{1-\gamma_1\alpha_\mu}+\frac{\Gamma_2}{1-\gamma_2\alpha_\mu}+1\right)\\
&\quad -\rho^{t-t'}\left(1-\frac{\alpha_\mu-\alpha_{\rm NT}}{\alpha_\mu-\rho}\right)\left(\frac{\Gamma_1}{1-\gamma_1\alpha_\mu}+\frac{\Gamma_2}{1-\gamma_2\alpha_\mu}+1\right)\\
R^{\rm NT}_{t-t'} &= \delta_{t'-t}R^{\rm NT}\\
R^{\mu}_{t-t'} &= \delta_{t'-t}R^{\mu}
\end{aligned} \tag{B.33}$$

where

$$\begin{aligned}
R^{\rm NT} =& -\alpha_{\rm NT}\left[\frac{\alpha_\mu-\alpha_{\rm NT}}{\alpha_\mu-\rho}\left(\frac{\Gamma_1}{(1-\alpha_\mu\gamma_1)(1-\alpha_{\rm NT}\gamma_1)}+\frac{\Gamma_2}{(1-\alpha_\mu\gamma_2)(1-\alpha_{\rm NT}\gamma_2)}\right)\right.\\
&\left.+\left(1-\frac{\alpha_\mu-\alpha_{\rm NT}}{\alpha_\mu-\rho}\right)\left(\frac{\Gamma_1}{(1-\alpha_{\rm NT}\gamma_1)(1-\rho\gamma_1)}+\frac{\Gamma_2}{(1-\alpha_{\rm NT}\gamma_2)(1-\rho\gamma_2)}\right)+1\right],\\
R^{\mu} =& \frac{\alpha_\mu}{G_0(1-\alpha_\mu)}\left(\frac{\Gamma_1}{1-\gamma_1\alpha_\mu}+\frac{\Gamma_2}{1-\gamma_2\alpha_\mu}+1\right).
\end{aligned} \tag{B.34}$$

Moreover, by a careful inspection of previous formulas and numerical solver results of Eq. (4.23) in the markovian case, one realizes that the following property holds:

$$R^{\mu} = \sqrt{\frac{\Omega_0^{\mathrm{NT}}}{\Xi_0^{\mu}}}\sqrt{(R^{\mathrm{NT}})^2 + 2\alpha_{\mathrm{NT}}R^{\mathrm{NT}} + 1}. \tag{B.35}$$

From the equation above one is able to deduce the expression of G_0, by inverting the previous equation for R^{μ}.

Finally, imposing the break-even condition per trade of the MM given by Eq. (4.14), one is able to derive the following identity:

$$\Omega_0 = \Xi_0^{\mu}(R^{\mu})^2\frac{\alpha_{\mu}\rho}{\gamma_1\gamma_2}\left(\tilde{b} + \frac{\alpha_{\mu}-\alpha_{\mathrm{NT}}}{\alpha_{\mu}-\rho}\frac{1}{1-\alpha_{\mu}\alpha_{\mathrm{NT}}} + \left(1 - \frac{\alpha_{\mu}-\alpha_{\mathrm{NT}}}{\alpha_{\mu}-\rho}\right)\frac{1}{1-\alpha_{\mathrm{NT}}\rho}\right). \tag{B.36}$$

In order to close the ansatz on itself, we have to compute the total order flow ACF. To do this, we need to calculate the first row of the inverse $(\mathsf{I}-\mathsf{RL})^{-1}$ which appear in Eq. (4.10). This is given by

$$\{(\mathsf{I}-\mathsf{RL})^{-1}\}_{t-t'} = \frac{\{(\mathsf{G}^{sym})^{-1}\}_{t,t'}}{\{(\mathsf{G}^{sym})^{-1}\}_{t,t}} = \frac{\alpha_{\mu}\rho}{\gamma_1\gamma_2}\{\tilde{\mathsf{G}}^{sym}\}_{t-t'}. \tag{B.37}$$

The explicit expression of the excess demand at time t is given by

$$\begin{aligned} q_t = \frac{\alpha_{\mu}\rho}{\gamma_1\gamma_2}\Bigg\{&\left[q_t^{\mathrm{NT}} + \sum_{t'=-\infty}^{t}\left(\Gamma_1\gamma_1^{t-t'} + \Gamma_2\gamma_2^{t-t'}\right)q_{t'}^{\mathrm{NT}}\right] \\ &+ R_{\mathrm{NT}}\left[q_{t-1}^{\mathrm{NT}} + \sum_{t'=-\infty}^{t-1}\left(\Gamma_1\gamma_1^{t-t'-1} + \Gamma_2\gamma_2^{t-t'-1}\right)q_{t'}^{\mathrm{NT}}\right] \\ &+ R_{\mu}\left[\mu_{t-1} + \sum_{t'=-\infty}^{t-1}\left(\Gamma_1\gamma_1^{t-t'-1} + \Gamma_2\gamma_2^{t-t'-1}\right)\mu_{t'}\right]\Bigg\}. \end{aligned} \tag{B.38}$$

With this equation, one is able to compute explicitly the excess demand ACF. In particular, by comparing the lag-0 term of it with the functional form given in Eq. (4.26) and using Eqs. (4.33) and (B.36) one is able to compute an implicit very complicated equation for ρ, fixing completely the ansatz given by Eq. (4.32).

The figures presented in Sec. 4.6 have been obtained by fitting the result of the numerical solver with Eq. (4.32), obtaining numerical values for ρ which have been cross-validated using the aforementioned analytical implicit equation for ρ, and then using the equations exposed in this section to compute the other quantities of interest.

Appendix C

Do fundamentals shape the price response?

C.1 Linear models for price impact

C.1.1 The Transient Impact Model revisited

While the s-Kyle model is microfounded in terms of noisy heterogeneous agents where the price is the optimal estimator of the *fundamental price* (p_n^{F}), the Transient Impact model is meant to explain the *empirical price* (p_n^{emp}) in terms of past order flows (see Sec. 3.3). The equation that defines the linear propagator model is simply given by

$$p_n^{\mathrm{P}} = \sum_{m \leq n} G_{n-m}^{\mathrm{P}} q_m. \tag{C.1}$$

One finds that, besides its simplicity, the propagator model is able to explain a large fraction of price volatility, meaning that a large fraction of price moves are not related to exogenous shocks. In order to test the stationary Kyle model, in the following, we confront it with the propagator model.

The calibration of the stationary Kyle and of the propagator model requires different input processes. The input processes needed for the calibration of the stationary Kyle model are given by ACFs related to signal and noise, i.e., Σ_n^{IT} and Ω_n^{NT}. On the other hand, the propagator model, which is not microfounded, is calibrated using only publicly observable processes. In particular, the excess demand ACF Ω_n and the response function $R_n^{\mathrm{emp}} = \mathbb{E}[q_n(p_{m+n}^{\mathrm{emp}} - p_{m-1}^{\mathrm{emp}})]$ are needed [54].

In the regime where the price is diffusive, the shape of the price impact function is the same as that given by the s-Kyle model, since it solely depends on the total order flow ACF (see Sec. 5.2). The only difference in terms of prediction of these two models, if calibrated in a regime where the price is diffusive, is given by the magnitude of the price impact function. In particular, $G_0^{\mathrm{P}} \geq G_0$, since empirical prices are far more volatile than fundamental values [15, 143] , as further discussed in Subsec 5.4.1.

As a final remark, note that the propagator model proposed in Ref. [54] is not implemented in the linear form given by Eq. (C.1). In fact, the calibration in that work was made using trade-by-trade data, and the predictive power was maximized using a sub-linear function of signed volumes of past trades. In Subsec. C.1.2, where aggregated data on sampling scale $\tau \geq 1$ minute are analyzed, we calibrate assuming the propagator to be a linear

function of signed volumes of past trades in Eq. (C.1), showing how the predictive power increases with respect to a propagator model extremely concave in past total traded volumes (i.e., linear in the *sign* of the trades). In order to conduct this analysis, it will be interesting to compare linear price impact models' outcomes when the sampling scale of the dataset varies. In the following, we detail how one can relate predictions coming from calibrations conducted with data aggregated over different time windows.

Scaling relations as the sampling scale varies In order to relate the prediction of a discrete model with well-known time-dependent quantities (such as the mean-reversion time scale of the de-trended fundamental price), we need to specify the real-time counterpart τ of 1 lag, i.e., the discretization time step (or the *sampling scale*). In practice, this timescale is set by the time resolution τ of the dataset at hand. For this reason, it will be convenient to introduce a notation that makes this detail explicit. For example, if the price comes from a dataset with a one-minute resolution, $\tau = 1$ minute, then we define $p_{n=1}^{(\tau)}$ as the price at time 1 minute. This notation will be handy when comparing predictions coming from calibrations performed on different coarse-grained versions of the same dataset. In the following, we will be interested in comparing objects, like the price impact function, constructed from processes defined at different sampling scales. In those cases, the dependence on the sampling scale τ will be made explicit. If in a given equation, no mention of a scale τ is made, it means that all the objects that appear are constructed with the same sampling scale τ.

Suppose we have a dataset related to prices and total excess demand with a given sampling scale τ_{short}. We can calibrate the linear propagator model on it. Then, starting from the original dataset, one can construct a coarse-grained version of it with a sampling scale $\tau_{\text{long}} = r\tau_{\text{short}} > \tau_{\text{short}}$. The relations between excess demand and price variables at different sampling scales are given by:

$$q_n^{(\tau_{\text{long}})} = \sum_{m=(n-1)r+1}^{nr} q_m^{(\tau_{\text{short}})}, \tag{C.2a}$$

$$p_n^{(\tau_{\text{long}})} = p_{nr}^{(\tau_{\text{short}})}. \tag{C.2b}$$

Below we discuss the relations between total order flow ACF and response function calibrated at different sampling scales. Taking into account Eq. (C.2a), a power law decay with exponent β for the excess demand ACF (similar to the one empirically found in [53]) and starting from the definition of excess demand auto-covariance, one obtains the following relation between total order flow ACFs at different sampling scales:

$$\Omega_n^{(\tau_{\text{long}})} \approx \Omega_n^{(\tau_{\text{short}})} (\tau_{\text{long}}/\tau_{\text{short}})^{(2-\beta)}. \tag{C.3}$$

In the diffusive regime, a similar calculation for the empirical response function gives:

$$R_n^{(\tau_{\text{long}})} \approx R_n^{(\tau_{\text{short}})} (\tau_{\text{long}}/\tau_{\text{short}})^{1/2+(2-\beta)/2}. \tag{C.4}$$

In the following, for the sake of argument, we shall consider also total order flow processes following a non-correlated noise, instead of a strongly correlated one. Note that in this case the two equations above still holds, using $\beta = 1$. Due to the linearity of the price impact models we are working with, it is easy to understand the price impact function behavior when coarse-graining. We find that (see App. C.2)

$$G_n^{(\tau_{\text{long}})} \approx \frac{1}{r} \sum_{m=(n-1)r+1}^{nr} G_m^{(\tau_{\text{short}})}, \tag{C.5}$$

i.e., the price impact function on sampling scale τ_{long} is roughly the price impact function at sampling scale τ_{short} averaged over a time window of length τ_{long}.

C.1.2 Predictive power of agnostic linear price impact models

Here we investigate the predictive power of the linear version of the propagator model given by Eq. (C.1), showing that for sampling scale $\tau \geq 1$ minute the results are satisfactory. Our dataset contains order flow and price time-indexed data at the trade resolution for a variety of stocks for a time range of 8 years (see details about data, de-trending and calibration procedures can be found in App. C.3.1).

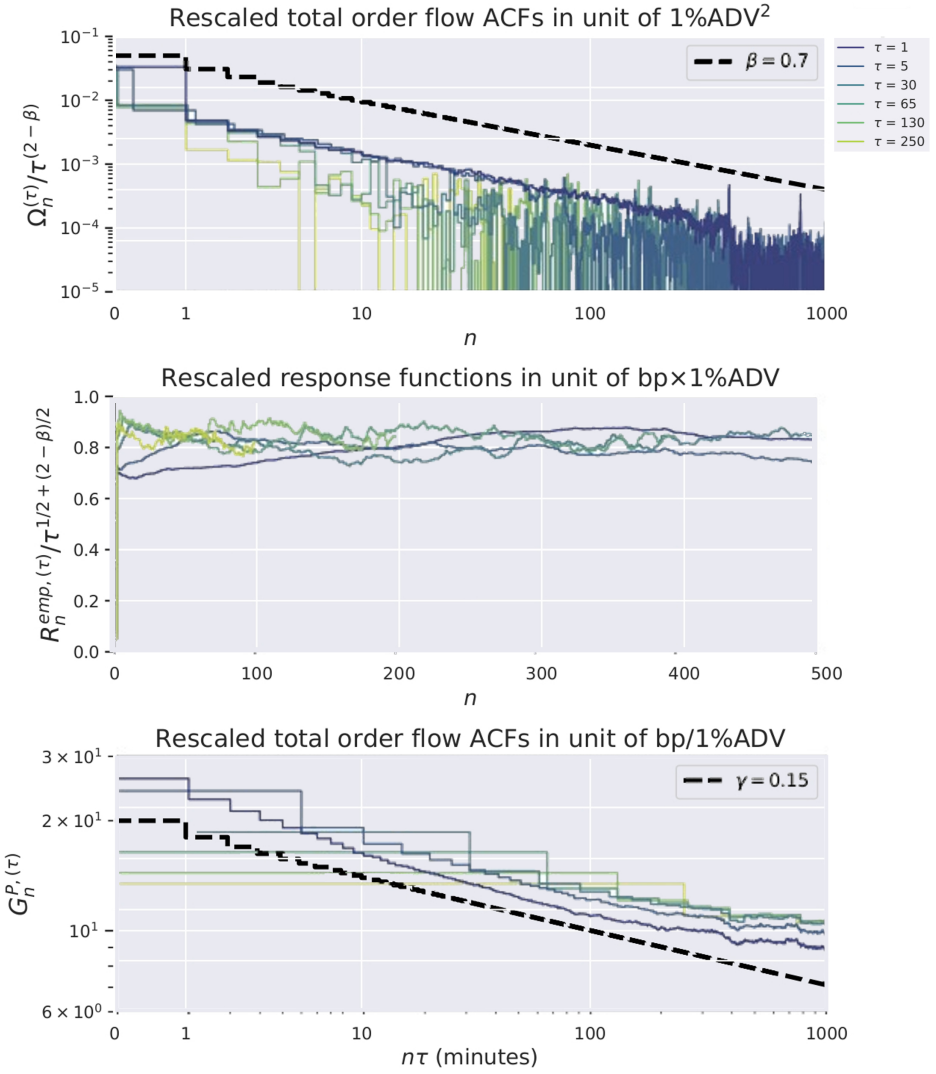


Figure C.1: Rescaled inputs and output of propagator model calibrations with different coarse-grained high-frequency data, averaged across stocks. Sampling scales τ related to different coarse-grained data are given in the legend, in units of minutes. Volumes are measured in units of 1% of average daily volume (ADV), while prices are measured in basis points (bp). (Top) Signed order flow ACFs, rescaled according to Eq. (C.3) and taking as a reference the total order flow ACF with sampling scale $\tau = 1$ minute. (Middle) Response functions, rescaled according to Eq. (C.4) and taking as a reference the response function with sampling scale $\tau = 1$ minute. (Bottom) Price impact functions. The dashed black lines in the middle and bottom panels refer to decreasing power law function with exponents β and γ, respectively, whose values are reported in the legends.

In Figure C.1 we show the order flow covariance, the price response, and the price impact function at sampling scales of 1, 5, 30, 65, 130, or 250 minutes. Note that for each

calibration, the analyzed time window is such that the price undergoes diffusive dynamics. Order flow ACFs show a slow decay, compatible with a power law, with an exponent lower than one. We find the power law exponent to be $\beta \approx 0.7$. A good collapse of the curves is obtained using the scaling given by Eq. (C.3). The unit for the covariance is a fraction of the average daily volume (ADV) squared. As known from the literature, the price response has a step-like shape, with a very quick increasing period and a plateau afterward. This flattening out of the response is a sign of the diffusivity of the prices: no trivial future price prediction can be made from observing the trade flow. A good collapse of the curves is obtained using the scaling given by Eq. (C.4). The unit for the price response is basis points of the price times fraction of the average daily volume (ADV). For the propagators at different sampling scales, we do not apply any rescaling, consistent with Eq. (C.5). We see that the propagator curves are comparable and quite similar at different sampling, validating the scaling found above. Furthermore, the relation between the exponents related to power laws that fits the order flow ACF (β) and the propagator (γ) [54] holds, i.e., $\gamma = (1-\beta)/2 \approx 0.15$. The units are basis points of the price for a fraction of the average daily volume (ADV).

The price impact function at lag zero G_0^{P} suggests that a trade flow imbalance of 1% of the average daily volume (ADV) leads to a price move of 26, 24, 18, 16, 14, or 13 basis points if executed in 1, 5, 30, 65, 130 or 250 minutes respectively. This decay of G_0^{P} for increasing τ is of importance. In fact, even though the total exchanged volume dominates the overall price impact of a sequence of trades, the way in which the order flow is realized at the trade-by-trade level has a measurable effect, since the impact function is in fact decaying (albeit, at a small rate) and thus the concentration of trades plays a role. The problem of optimal order execution (see, e.g., Refs. [144–148]) deals precisely with the minimization of price impact induced costs under an exogenous constraint for the total size of the trade. It is known in that context that it is possible to decrease transaction costs by exploiting the decay of the propagator function and that such a decrease is moderate whenever the decay of the impact function is slow. In case of a power-law decreasing $G_n \propto n^{-\gamma}$, the price impact related cost of an execution at constant rate ϕ during a period T scales as $T^{-\gamma}$ [149]. This is explicitly shown in Fig. C.2, where $G_0^{\mathrm{P},(\tau)}$ is taken as a proxy for execution costs with different execution times (τ playing the role of T).

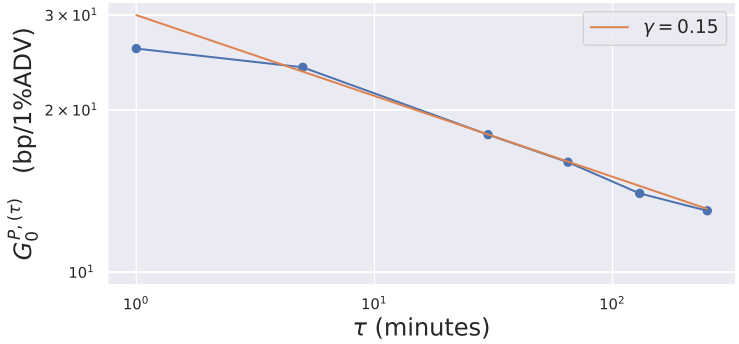


Figure C.2: Instantaneous price move in bp for a trade of 1%ADV executed in a time-window of size τ.

In the remaining part of the section, we study how much price volatility can be explained by the linear version of the propagator model calibrated at different sampling scales. We define the ratio between predicted and empirical price variograms as

$$\Phi_n^{\mathrm{P}} = \frac{V_n^{\mathrm{P}}}{V_n^{\mathrm{emp}}}. \tag{C.6}$$

Figure C.3 shows the metric $\Phi^{\mathrm{P},(\tau)}$ at a lag corresponding to $\sim$ 4 days. We show two curves. The blue one (q) corresponds to a calibration using the sum of signed trade flow in the bin as before. The orange one (ϵ) corresponds to a calibration using the sum of the signs of the trades in a bin. Let's first concentrate on the blue curve (q). We find that the explained variance increases initially as the bin size increases, flattening out for longer sampling scales of a few tens of minutes. Most importantly, for all values of τ the explained variance is well above the ratio of $\Sigma_0^{\mathrm{F}}/\Sigma_0^{\mathrm{emp}} = 0.26$ that can be found in Table 5.1. This means that necessarily we have excess price response, which implies excess volatility. This carries a very important message: the total trade flow imbalance is indeed of high relevance to explaining price moves when considering aggregated data. Let's now look at the orange curve (ϵ), for which, interestingly, one observes a better explanatory power than the (q) model for short τ scales. Indeed, at the microstructural level, the actual traded volume is very much conditioned on the available liquidity in the limit order book, and the sign of the trade is more informative than the trade itself. In fact, it is rare that a trade penetrates more than one price level. This means that agents condition the size of their transactions on liquidity, making large transactions when liquidity is high and small transactions when it is low. Thus, the information content is not related to the volume traded, but rather to the trade's sign. This effect undergoes the name of selective liquidity taking. When one considers aggregated data, however, the relation flips: as τ increases, the information contained in q becomes more valuable. A heuristic argument for this phenomenon is that the easier to manipulate a market variable, the less it should carry information: manipulating q, the exchanged volume means taking an actual risk. For ϵ, on the other hand, manipulation is much less risky: it can be done for example by placing many small trades in the market instead of a few large trades.

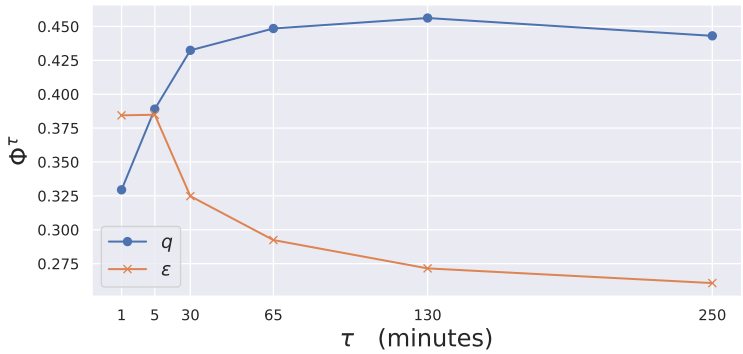


Figure C.3: Ratio between predicted and empirical variograms at lag $\sim$ 4 days, as a function of the sampling scale τ. Two linear propagator models are analyzed: price impact linear in signed order flows (blue line) and linear in the sum of trades signs (orange line).

We stress-tested the linearity assumption of the analyzed price impact models, demonstrating that the Propagator model has a high predictive power if the sampling scale is larger than the typical trading timescale because effects related to order book dynamics (such as selective liquidity taking) can be neglected. In order to check the robustness of our findings, we applied scaling arguments to obtain relationships between price impact functions at different sampling scales.

C.2 Price impact function scaling varying the sampling scale

Consider a slowly varying kernel $G_n^{(\tau_{\text{short}})}$ (assumption which will be realized empirically at small enough sampling scales τ_{short}). One can 'zoom-out' in time, by defining a new coarse-grained model with sampling scale $\tau_{\text{long}} = r\tau_{\text{short}} > \tau_{\text{short}}$. We show how the impact function changes when the sampling scale is changed. The argument goes as follows:

$$\begin{aligned} p_{nr}^{(\tau_{\text{short}})} &= \sum_{m\leq n}\left(G_{(n-m)r}^{(\tau_{\text{short}})}q_{mr}^{(\tau_{\text{short}})} + G_{(n-m)r-1}^{(\tau_{\text{short}})}q_{mr-1}^{(\tau_{\text{short}})} + \cdots + G_{(n-m)r-r+1}^{(\tau_{\text{short}})}q_{mr-r+1}^{(\tau_{\text{short}})}\right) \\ &\approx \sum_{m\leq n}\underbrace{\frac{1}{r}(G_{(n-m)r}^{(\tau_{\text{short}})} + G_{(n-m)r-1}^{(\tau_{\text{short}})} + \cdots + G_{(n-m)r-r+1}^{(\tau_{\text{short}})})}_{G_{n-m}^{(\tau_{\text{long}})}}\underbrace{(q_{mr}^{(\tau_{\text{short}})} + q_{mr-1}^{(\tau_{\text{short}})} + \cdots + q_{mr-r+1}^{(\tau_{\text{short}})})}_{q_m^{(\tau_{\text{long}})}} \\ &= p_n^{(\tau_{\text{long}})}. \end{aligned}$$

Equations above indicate that when downsampling the model to scale τ_{long}, the price is linear -at a first approximation- in the total imbalance over bins of size equal to τ_{long}. The price impact function at scale τ_{long} is the arithmetic mean over a time window equal to τ_{long} of the price impact function defined at scale τ_{short}. Even though one has exact equality only in the case of a perfectly constant kernel or a perfectly constant order flow imbalance, the approximation will still retain a good explanatory power as long as $G_n^{(\tau_{\text{short}})}$ is smooth enough.

C.3 Datasets, detrending and calibration procedures

C.3.1 High frequency

Presentation of the data We analyzed data about some of the most traded stocks, during the period January 2013 - December 2020.

The tick size of all the stocks is 0.01 USD. We reshaped the data removing effects coming from stock splitting. The bid-ask spread of a large tick stock is most of the time equal to one tick, whereas small tick stocks have spreads that are typically a few ticks. There exist also a number of stocks in the intermediate region between large and small tick stocks, which have the characteristics of both types. We choose 5 different stocks and we placed them all in the same pool. Data are indexed by a time label with precision at the microsecond ($\tau = 1\mu s$), where information about the precise timing of the transaction is stored.

The empirical mid-price P_n^{emp} is calculated as the mean between the bid and the ask. At the transaction level, we constructed order signs by labeling trades for which the transaction price is above the mid-price by $\epsilon_n = +1$ and all trades below as $\epsilon_n = -1$. Trades exactly at the mid-price were discarded. Signed order flow q_n are then constructed by multiplying the trade's sign by the quantity traded. Data about volumes are normalized by a rolling mean of the total daily volume exchanged over a time window ~ 50 days.

The empirical price exhibits a positive trend (as one can see from the left panel of Figure C.4), thus a de-trending procedure has to be implemented in order to meet the assumptions on which the stationary Kyle model is constructed. The de-trending procedure is discussed in detail below.

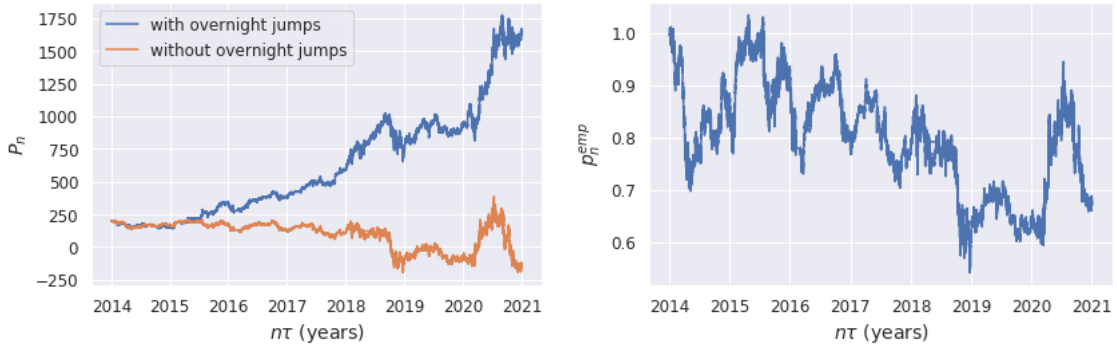


Figure C.4: High-frequency empirical data about AMZN stock. The sampling scale is equal to $\tau = 1\mu s$. (Left) The raw empirical price is shown (blue line). The trend in the mean of the raw empirical price is removed if we remove overnight jumps (orange line). (Right) We show the mid after the multiplicative de-trending procedure. The de-trended price doesn't exhibit a trend in the mean, nor in the volatility.

De-trending data at high frequency In order to proceed further, let us note that the main contribution to the trend in the price level is realized overnight, as the left panel of Figure C.4 shows. Removing overnight jumps is not enough to make the price stationary. In fact, as one can see from the orange line in the left panel of Fig. C.4, we still have a trend in volatility. In order to deal with this, we start from the price with overnights and we apply a logarithmic transformation. In doing this we obtain:

$$\log\left(P_n^{\text{emp}}\right) = \log(P_0^{\text{emp}}) + \sum_{m=0}^{n} \eta_m + \log(p_n^{\text{emp}}), \tag{C.7}$$

where p_n^{emp} is the residual part of the price, after the trend η_n is removed. After this logarithmic transformation, we remove the overnight jumps, removing -effectively- the trending component which depends on η_n. We do this by considering the trading activity in the period 9:30-16:00 on all days under analysis. For each stock, we concatenate the data on different trading days. Then, we apply an exponential transformation in order to get back to the de-trended price. The result of this operation on the AMZN stock is shown in the right panel of Figure C.4.

Appendix D

Microfounding GARCH Models and Beyond

D.1 How to simulate the model

Below we present a pseudo-code to perform simulations of the model presented in Sec. 6.2.

Consider a simulation which starts from $t = 1$ and $k = 0$, with initial conditions $\{\Lambda_0, \hat{\omega}_0, \delta\epsilon_1\}$ and total duration $T = \tau_{\rm rev} K$, with K a positive integer.

The other parameters needed to run the simulations are $\{\omega, \epsilon, \delta\epsilon, \hat{\epsilon}, \tau_{\rm NT}, \tau_{\rm rev}\}$. We set $\omega = \epsilon = 1$.

for $t \leq T$ **do**

Build the excess demand $q_t = q_t^{\rm IT} + q_t^{\rm NT}$: the informed trade $q_t^{\rm IT}$ is given by Eq. (6.5), i.e., it is the ratio between a realization of the fundamental price $p_t^{\rm F}$, obtained from a Gaussian process with zero mean and volatility ω, and the last observed price impact function Λ_{t-1}; the noise trade $q_t^{\rm NT}$ is a realization of a Gaussian process with zero mean and volatility ϵ_t.

Then, the price p_t is given by Eq. (6.4) and Eq. (6.6), with fundamental price and noisy order flow volatility beliefs respectively given by $\hat{\omega}_{k\tau_{\rm rev}}$ and $\hat{\epsilon}$; in doing this last step, one updates the price impact function.

if $t \neq (k+1)\tau_{rev}$ **then**

$t++$.

Update noisy order flow volatility ϵ_t according to Eqs. (6.1) and (6.2).

end

else

First, construct a price volatility estimate $\bar{\sigma}_{k\tau_{\rm rev}}$ from the past $\tau_{\rm rev}$ prices p_t.

Then, update the fundamental price volatility $\omega_{k\tau_{\rm rev}}$ according to Eq. (6.10). $t++, k++$.

end

end

Note that the simulation is characterized by four (+1 timescale, if the market is stable) timescales:

1. t, over which trading takes place.

2. τ_{NT}, over which noisy order flow volatility fluctuates. This is a parameter that the modeler has to fix.

3. τ_{fast}, over which the fast dynamics of price impact reaches the stationary regime. This parameter controls the fast dynamics of price impact, given by Eq. (6.6), which has been analyzed in Sec. 6.2.1.

4. τ_{rev}, over which the market maker updates his belief about fundamental price volatility. This is a parameter that the modeler has to fix.

5. τ_{slow}, over which the belief of the market maker converges in distribution if the market is stable. In this case, this parameter controls the slow dynamics of market maker's belief, given by Eq. (6.10) analyzed in Sec. 6.2.2.

Titre : Microfondation théorique des mécanismes de formation des prix

Mots clés : Dynamique des prix, faits stylisés, microfondations, impact prix, clustering de volatilité, prix fondamental, agents rationnels, agents adaptatifs

Résumé :

Les prix sur les marchés financiers présentent une dynamique non triviale dont les régularités peuvent être résumées en un ensemble de faits stylisés. Alors que les modèles statistiques capturent l'interaction entre ces faits stylisés et sont utilisés pour faire des prédictions quantitatives, ils n'expliquent pas pourquoi les prix évoluent en premier lieu. En revanche, les modèles micro-fondés laissent la dynamique des prix émerger des interactions entre les stratégies des agents, fournissant des informations cruciales aux régulateurs et aux décideurs politiques. Cette thèse propose des microfondations pour deux modèles statistiques bien connus, étendant leur pouvoir prédictif.

Nous fournissons une explication microscopique au modèle à Propagateur, qui est un modèle statistique capable de caractériser la dynamique stationnaire des ordres et des prix, fournissant une solution au "puzzle de la diffusivité". La micro-fondation est obtenue en généralisant le modèle de Kyle à un cadre stationnaire, dans lequel le prix fondamental n'est jamais public. Le modèle stationnaire de Kyle (s-Kyle) que nous proposons est compatible avec la diffusion universelle des prix observée expérimentalement à court terme ainsi que le retour non universel à la moyenne pour des échelles de temps sur lesquelles les fluctuations des fondamentaux diminuent. Cependant, le modèle s-Kyle suppose des agents fortement rationnels. Alors que l'hypothèse d'attente rationnelle (REH) est conforme à l'hypothèse de marché efficient (EMH), elle conduit le modèle s-Kyle à faire de mauvaises prédictions, à savoir que la volatilité des prix est indépendante du temps et inférieure à celle liée aux fondamentaux. Le REH empêche donc le modèle s-Kyle de résoudre l'énigme de l'excès de volatilité dans la mesure où nous savons que les fluctuations de prix sont supérieures à celles liées aux fondamentaux grâce aux travaux de Shiller.

Suivant Shiller et la littérature sur la finance comportementale, nous proposons une model de Kyle comportementale (b-Kyle) en assouplissant REH. Nous supposons que l'agent qui contrôle le prix ne connaît pas le niveau précis des ordres non informés ni celui de la volatilité des fondamentaux et il met à jour son estimation de la volatilité des fondamentaux en se fondant sur l'historique des prix. La procédure de mise à jour conduit à une dynamique de tâtonnements qui reflète la dynamique d'apprentissage adaptatif des stratégies des agents. Nous fournissons non seulement une micro-fondation à la volatilité excessive, mais aussi à la dynamique intermittente de la volatilité des prix. En fait, dans une limite appropriée du modèle b-Kyle, nous montrons que l'excès de volatilité suit un processus de Kesten, c'est-à-dire un processus multiplicatif stochastique repoussé de zéro. En conséquence, nous fournissons une micro-fondation pour une généralisation des modèles d'hétéroscédasticité conditionnelle auto-régressive généralisée. Le modèle b-Kyle s'inscrit dans la littérature qui évalue la validité l'EMH ; en fait, il suppose que la volatilité fondamentale des prix est constante, tout en prédisant une volatilité intermittente des prix. L'explication que le modèle b-Kyle fournit pour le regroupement de la volatilité des prix est donc en accord avec la conclusion empirique selon laquelle une grande partie des sauts de prix ne peut pas être expliquée par les innovations des fondamentaux, mais est plutôt causée par la dynamique auto-excitante créée par l'interaction entre stratégies des agents.

Nous pensons que le modèle b-Kyle peut être utile pour expliquer pourquoi les prix bougent, étant parcimonieux, mais réaliste : il peut aider à rationaliser de nombreuses questions abordées dans la littérature, allant de la diffusivité des prix à la volatilité excessive et au regroupement de la volatilité. De plus, il peut également interpoler des périodes calmes avec des prix très fluctuants à des régimes fragiles avec des crashs et des crises de liquidité extrêmement probables.

Title : Microfounded theories of price formation

Keywords : Price dynamics, stylized facts, microfoundations, price impact, volatility clustering, fundamental price, rational agents, adaptive agents

Abstract :

Price and volume dynamics in financial markets exhibit empirical regularities, called stylized facts. Statistical models capture the interplay between these stylized facts and are widely used to make quantitative predictions, but they do not explain why prices move in the first place. Microfounded models instead let the price dynamics emerge from the interactions between traders' strategies. The aim of this thesis is to partially bridge the gap between the literature on microfounded and statistical models. In particular, we explore how the predictions of a well-known microfounded model change if we relax some of its unrealistic assumptions. Interestingly, in doing so, we obtain microfoundations for two well-known statistical models, extending their predictive power.

We provide a microfoundation for the Transient Impact model, which is able to characterize the stationary interplay between the dynamics of orders and prices, solving the diffusivity puzzle. The microfoundation is achieved by generalizing the classic Kyle model of price formation to a stationary setting, assuming that the fundamental price is never made public. The stationary Kyle (s-Kyle) model that we propose is compatible with experimentally observed universal price diffusion in the short term, and non-universal mean-reversion on time scales at which correlations of fundamentals vanish. However, the s-Kyle model assumes strongly rational traders, i.e., each rational agent knows every other player's strategies and has unlimited computing power. While the Rational Expectation Hypothesis (REH) is in line with the Efficient Market Hypothesis (EMH), for which the price always reflects newly released fundamental innovations, it leads the s-Kyle model to make wrong predictions ; namely, that price volatility is time-independent and smaller than the one related to fundamentals. The REH, therefore, prevents the s-Kyle model from solving the excess volatility puzzle if one does not assume an unrealistically high risk aversion of market actors. In order to improve that, we propose a second modification of the Kyle model, described below.

Following Shiller and the behavioral finance literature, we propose a behavioral Kyle (b-Kyle) setup by relaxing the REH. In doing so, we obtain a microfoundation of the Generalized Auto-Regressive Conditional Heteroscedasticity model. To do so, we assume that the market maker does not know the precise level of non-informed trading and of fundamental volatility ; moreover, he updates his prior about fundamental volatility based on the realized market prices. The updating procedure is constructed such that future expectations match past outcomes, leading to tâtonnement dynamics reflecting the adaptive learning dynamics of traders' strategies. In this way, not only do we provide a micro-foundation for excess volatility, but also for the intermittent dynamics of price volatility. In fact, in an appropriate limit of the b-Kyle model, the dynamics becomes analytically tractable and we show that excess volatility follows a Kesten process, i.e., a stochastic multiplicative process repelled from zero. Accordingly, we provide a microfoundation of the class of Generalized Auto-Regressive Conditional Heteroscedasticity models. The b-Kyle model is in line with the literature that challenges the EMH ; in fact, it assumes that fundamental price volatility is constant, while it predicts intermittent price volatility. The explanation the b-Kyle model provides for price volatility clustering therefore agrees with the empirical finding that a large fraction of price jumps can not be explained by fundamental innovations, but is instead caused by the self-exciting dynamics created by the interplay between traders' strategies.

We believe that the b-Kyle model can be useful for explaining why prices move, being parsimonious, yet realistic : it can help rationalize many puzzles tackled in the literature, ranging from price diffusivity to excess volatility and volatility clustering. Moreover, it can also interpolate from calm periods with highly fluctuating prices to fragile regimes with extremely probable flash crashes and liquidity crises.

Institut Polytechnique de Paris
91120 Palaiseau, France