Dynamique de carnets d'ordres boursiers : modèles stochastiques et théorèmes limites

Adrien de Larrard

Université Pierre et Marie Curie

École Doctorale Paris Centre

Thèse de doctorat

Discipline : Mathématiques

présentée par

Adrien de Larrard

Dynamique de carnets d'ordres boursiers : modèles stochastiques et théorèmes limites

---

Price dynamics in limit order markets : queueing models and limit theorems

---

dirigée par Rama Cont

Soutenue le 02 Octobre 2012 devant le jury composé de :

M. Rama Cont	Université Pierre et Marie Curie	directeur
M. Laurent Decreusefond	Telecom-Paris Tech	examinateur
M. Marc Hoffmann	CREST ENSAE	rapporteur
M. Charles-Albert Lehalle	Crédit Agricole Cheuvreux	examinateur
M. Michel Mandjes	University of Amsterdam	rapporteur
M. Mathieu Rosenbaum	Université Pierre et Marie Curie	examinateur
M. Alexander Schied	Universität Mannheim	examinateur

École doctorale Paris centre Case 188
4 place Jussieu
75 252 Paris cedex 05

A ma famille.

Contents

Résumé

Cette thèse propose un cadre mathématique pour la modélisation de la dynamique du prix et du flux d'ordres dans un marché électronique où les participants achetent et vendent un produit financier en soumettant des ordres limites et des ordres de marché à haute frequence à un *carnet d'ordres* centralisé. Nous proposons un modèle stochastique de carnet d'ordres en tant que système de files d'attente représentant la totalité des ordres d'achat et de vente au meilleur niveau de prix (bid/ask) et nous montrons que les principales caractéristiques de la dynamique du prix dans un tel marché peuvent être comprises dans ce cadre. Nous étudions en détail la relation entre les principales propriétés du prix et la dynamique du processus ponctuel décrivant l'arrivée et l'exécution des ordres, d'abord dans un cadre Markovien (Chapitre 2) puis, en utilisant des méthodes asymptotiques, dans le cadre plus général d'un processus ponctuel stationnaire dans sa limite *heavy traffic*, pour lequel les ordres arrivent fréquemment, comme c'est le cas pour la plupart des marchés liquides (Chapitres 3 et 4).

Le Chapitre 2 étudie un modèle Markovien de dynamique de carnet d'ordres, dans lequel l'arrivée d'ordres de marché, d'ordres limites et d'annulations est décrite à l'aide d'un processus de Poisson ponctuel. L'état du carnet d'ordres est décrit par une marche aléatoire changée de temps dans le quadrant positif et régénérée à chaque fois qu'elle atteint le bord. Ce modèle permet d'obtenir des expressions analytiques pour la distribution des durées entre changements de prix, la distribution et les autocorrélations des changements de prix, ainsi que la probabilité que le prix augmente, conditionnellement à l'état du carnet d'ordres. Nous étudions la limite de diffusion du prix et exprimons la volatilité des changements de prix à l'aide de paramètres décrivant l'intensité des ordres d'achat, de vente et d'annulations. Ces résultats analytiques permettent de mieux comprendre le lien entre volatilité du prix et flux d'ordres.

Le Chapitre 3 étudie un modèle plus général de carnet d'ordres pour lequel les arrivées d'ordres et les tailles d'ordres proviennent d'un processus ponctuel stationnaire très général. Nous obtenons un théorème central limite fonctionnel pour la dynamique jointe des files d'attente des ordres de vente et d'achat, et prouvons que, pour un marché liquide, dans lequel les ordres d'achat et de vente arrivent à haute fréquence, la dynamique du carnet d'ordres peut être approximée par un processus à sauts Markovien diffusant dans l'orthant et dont les caractéristiques peuvent être exprimées à l'aide de propriétés statistiques du flux d'ordres sous-jacent. Ce résultat permet d'obtenir des approximations analytiques pour plusieurs quantitiés d'interêt telles que la probabilité que le prix augmente ou la distribution de la durée avant le prochain changement de prix, conditionnellement à l'état du carnet d'ordres. Ces quantités sont exprimées en tant que solutions d'équations elliptiques, pour lesquelles nous donnons des solutions explicites dans certains cas importants. Ces résultats s'appliquent à une classe importante de modèles stochastiques, incluant les modèles basés sur les processus de Poisson, les processus auto-excitants ou la famille de processus ACD-GARCH.

Le Chapitre 4 est une étude plus detaillée de la dynamique du prix dans un marché où les ordres de marché, les ordres limites et les annulations arrivent à haute frequénce. Nous étudions d'abord la dynamique discrète du prix à l'échelle de la seconde et nous obtenons des relations analytiques entre les propriétés statistiques des changements de prix dans une journée -distribution des incréments du prix, retour à la moyenne et autocorrélations- et des propriétés du processus décrivant le flux d'ordres et la profondeur du carnet d'ordres. Ensuite nous étudions le comportement du prix à des fréquences

plus faibles pour plusieurs régimes asymptotiques -limites fluides et diffusives- et nous obtenons pour chaque cas la tendance du prix et sa volatilité en fonction des intensités d'arrivées d'ordres d'achat, de vente et d'annulations ainsi que la variance des tailles d'ordres. Ces formules permettent de mieux comprendre le lien entre volatilité du prix d'un côté et le flux d'ordres, décrivant la liquidité, d'un autre côté. Nous montrons que ces résultats sont en accord avec la réalité des marchés liquides.

Summary

This thesis proposes a mathematical framework for the modeling the intraday dynamics of prices and order flow in *limit order markets*: electronic markets where participants buy and sell a financial contract by submitting market orders and limit orders at high frequency to a centralized *limit order book*. We propose a stochastic model of a limit order book as a queueing system representing the dynamics of the queues of buy/sell limit orders at the best available (bid/ask) price levels and argue that the main features of price dynamics in limit order markets may be understood in this framework. We study in detail the relation between the statistical properties of the price and the dynamics of the point process describing the arrival and execution of orders, first in a Markovian setting (Chapter 2) then, using asymptotic methods, in a more general setting of a stationary point process in the *heavy traffic limit*, where orders arrive very frequently, as in most liquid stock markets (Chapters 3 and 4).

Chapter 2 studies a Markovian model for limit order book dynamics, in which arrivals of market order, limit orders and order cancelations are described in terms of a Poisson point process. The state of the order book is then described as a time-changed random walk in the positive quadrant regenerated at each hitting time of the boundary. This model allows to obtain analytical expressions for the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, *conditional* on the state of the order book, by mapping them into quantities related to hitting times of a random walk in $\mathbb{Z}_+^2$ killed at the boundary. We study the diffusion limit of the price process and express the volatility of price changes in terms of parameters describing the arrival rates of buy and sell orders and cancelations. These analytical results provide some insight into the relation between order flow and price dynamics in order-driven markets.

Chapter 3 studies a more general queueing model in which order arrivals and order sizes are described by a stationary point process, allowing for a wide range of distributional assumptions and temporal dependence structures in the order flow. We derive a functional central limit theorem for the joint dynamics of the bid and ask queues and show that, in a liquid market where buy and sell orders are submitted at high frequency, the intraday dynamics of the limit order book may be approximated by a Markovian jump-diffusion process in the positive orthant, whose characteristics are explicitly described in terms of the statistical properties of the underlying order flow. This result allows to obtain tractable analytical approximations for various quantities of interest, such as the probability of a price increase or the distribution of the duration until the next price move, conditional on the state of the order book. Both quantities are expressed in terms of the solution of elliptic equation in the positive orthant, for which solutions are given in important special cases. These results apply to a wide class of stochastic models proposed for order book dynamics, including models based on Poisson point processes, self-exciting point processes and models of the ACD-GARCH family.

Chapter 4 is a more detailed study of price dynamics in a limit order market where market orders, limit orders and order cancelations occur with high frequency according to a stationary marked point process. We first study the discrete, high-frequency dynamics of the price and derive analytical relations between the statistical properties of intraday price changes -distribution of increments, mean reversion and autocorrélation- and properties of the process describing the order flow and depth of the order book. We then study the behavior of the price process at lower frequencies under various heavy-

traffic limits –fluid limits and diffusion limits– and derive in each case the price trend and intraday volatility in terms of the arrival rates of buy and sell orders and cancelations and the variance of order sizes. These analytical formulae provide insights into the link between price volatility on one hand and high-frequency order flow and liquidity on the other hand and are shown to be in good agreement with high-frequency data for US stocks.

Remerciements

En premier lieu je souhaite exprimer ma profonde gratitude envers mon directeur de thèse Rama Cont qui a su me guider pendant ces trois années de doctorat. J'ai eu la chance de bénéficier de ses remarques pertinentes, de son immense culture scientifique et de sa vision profonde du domaine sur lequel nous avons travaillé.

Je remercie également Marc Hoffmann et Michel Mandjes d'avoir accepté d'être les rapporteurs de ma thèse. Merci à Laurent Decreusefond, Charles-Albert Lehalle, Mathieu Rosenbaum et Alexander Schied pour avoir examiné ma thèse et participé à ma soutenance. Je remercie tout particulièrement Charles Albert Lehalle de m'avoir fait découvrir le sujet passionnant et encore très largement inexploré de la microstructure des marchés financiers pendant mon stage à Cheuvreux.

Un immense merci à Xin Guo, qui m'a initié à la recherche en mathématiques appliquées et m'a accueilli pendant deux années formidables à UC Berkeley.

Je tiens à remercier les collègues du LPMA, qui ont rendu l'ambiance du laboratoire fort sympathique. Je remercie Amel, Andreea, Benjamin, Clément, Fabien Lakshithe et Laura avec lesquels j'ai partagé de très bons moments. Je n'oublierai pas tous mes amis qui ne sont pas du laboratoire et qui m'ont accompagné.

Je suis très reconnaissant envers mes professeurs de classes préparatoires du lycée Pierre de Fermat Michel Gonnord et Jean-Marie Mercier qui m'ont transmis leurs passions pour les sciences.

Je ne pourrai clôturer ces remerciements sans me tourner vers ma famille qui m'a constamment encouragé et conforté dans mes choix. Je les remercie chaleureusement pour leur soutien moral sans lequel je n'aurais pas pu mener à bien mes longues études.

Chapter 1

Introduction

High-frequency trading has experienced a significative increase in activity during the last decade. With recent change in trading rules and the improvement in computer's speed, high-frequency trading (e.g. trading initiated by fast computers) has become the main source of trading on financial markets. Traditionally, investors would trade on an exchange such as the New York Stock Exchange (NYSE) and keep their positions for a couple of days or months. These days, algorithms buy and hold shares for only a couple of seconds or milliseconds, benefiting from micro arbitrages or using high frequency trading strategies. Even long term investors use high frequency trading strategies to minimize their trading impact and optimize their costs. Financial markets have moved from *specialist-driven markets*, where a market maker or a designed specialist centralizes all buy and sell orders, to *order-driven markets*, where every investor can provide liquidity by quoting buy and sell orders. Simultaneously, many terabytes of financial time series have been recorded and could be used to understand the microstructure of current financial markets. In this thesis we propose a stochastic model of order driven markets and give a rigorous method for establishing relations between properties of the order flow - which describes liquidity - to the properties of the price process, such as volatility and mean reversion.

Limit order books

In an *order-driven market*, market participants can post two types of buy/sell orders: *market orders* and *limit orders*. A limit order is an order to trade a certain amount of a security at a given price. Limit orders are posted to an electronic trading system, and the state of outstanding limit orders can be summarized by stating the quantities posted at each price level: this is known as the *limit order book*. The lowest price for which there is an outstanding limit sell order is called the *ask price* and the highest buy price is called the *bid price*. At any time, the limit order book gathers all limit buy and sell orders available at different prices. Figure 1.1 shows such limit order book. The column on the left (resp. the right) collects all limit buy (resp. sell) orders. Every line in the left column (resp. the right column) displays the number of limit buy orders (resp. sell orders) available at a given price. For this limit order book, the bid price is 83.40 (resp. ask price is 83.41) and the total number of orders available at this price is 7 (resp. 3). The bid-ask spread equals 1cent.

A market order is an order to buy/sell a certain quantity of the asset at the best available price in the limit order book. When a market order arrives it is matched with the best available price in the limit order book, and a trade occurs. The quantities available in the limit order book are updated accordingly. For instance, if an investor buys 10 shares of IBM in the limit order book from figure 1.1 with market orders, he will consume 3 shares at 83.41 and 2 shares at 83.42. The ask price will immediately move from 81.41 to 81.43.

A limit order sits in the order book until it is either executed against a market order or canceled. A limit order may be executed very quickly if it corresponds to a price near the bid and the ask, but may take a long time if the market price moves away from the requested price or if the requested price

Figure 1.1: New York Stock Exchange open book for the stock IBM. Source: www.ndxtrading.com/interactivedemo

is too far from the bid/ask. Alternatively, a limit order can be canceled at any time. Recently, new orders, variants of limit and market orders, were created to make trader's life easier. For instance hidden orders are not displayed in the limit order books and pegged buy (resp. sell) orders remain at the bid (resp. ask) price until they are executed.

The literature related to limit order books modeling may be divided into two main categories:

- empirical studies, which focus on statistical features of quantities related to limit order books. Thanks to the availability of large databases of financial data being recently recorded, the statistical properties of high-frequency financial data have been widely studied. These studies (Bollerslev and Engle (1993) Pagan (1996) Cont (2001) Clark (1973) Silva and Yakovenko (2006) Kyle et al. (2010) Madhavan et al. (1997) Lillo and Farmer (2004) Bouchaud et al. (2002) Bouchaud et al. (2003) Gopikrishnan et al. (2000) Maslov and Mills (2001) Challet and Stinchcombe (2001) Mike and Farmer (2008) Alfonsi et al. (2010)) model, among other quantities, the average shape of limit order books, the volume of orders and the durations between orders.

- the theoretical models among which one can distinguish three main approaches:

  The econometric approach models the price dynamics without looking at the order flow, by incorporating *stylized facts*, such as the negative autocorrelation between price increments, correlation between trading intensity and volatility or volatility clustering in the time series.

  The literature in market microstructure (O'Hara (1997) Foucault et al. (2005), Garman (1976) Kyle (1985) Hasbrouck (2007)) links the behavior of investors with properties of price microstructure (bid-ask spread, price volatility, trading intensity, etc...). Market microstructure models were designed for *specialist based* markets, where a designed specialist would quote bid and ask prices on the New York Stock Exchange and would globally set the market price. Understanding the behavior of this specialist could give some insights on the price behavior. Historically, the main class of microstructure models, *agent based models* have modeled the behavior of this designed specialist and divided the trading population between *informed investors*, who would know the firm value, and *noise traders* who would trade for various reasons but without additional information.

  Eventually, stochastic models of limit order books deduce properties of the price process by assuming that the order flow is a random process. Stochastic models reproduce the properties of a limit order book without resorting to detailed behavioral assumptions about market participants or introducing unobservable parameters describing agent preferences, as in more detailed market microstructure models.

This thesis may be viewed as a first step towards bridging the gap between the theoretical approach, which links agents behavior with price dynamics, and the empirical approach, which proposes statistical models for financial time series. Our approach consist in modeling the flow of orders -limit orders, market orders and cancelations - as a stochastic point process for which we propose various levels of description at various time resolutions, from ultra-high frequency to daily. We show that in liquid markets where orders are frequent, *limit theorems* may be used to study the link between the behavior of prices and order flow at these various time scale. In particular, we give a rigorous method for establishing relations between properties of the order flow - which describes liquidity - to properties of the price process, such as volatility and mean-reversion.

The introductory chapter is organized as follows: in section 1.1, we review the recent change in the US financial market and describe the current financial landscape. We introduce algorithmic trading and eventually show how the *consolidated limit order book* has become the key object at the core of price formation. Section 1.2 reviews some empirical properties of limit order books. In section 1.3 we present theoretical models, mostly agent based models. Section 1.4 summarizes the original contributions of this thesis.

1.1 Old and new financial markets

At the end of the 1990s, several important changes in regulation shaped a new landscape of equity trading in the United States. In 1996, the Security and Exchange Commission (SEC) authorized trading on purely electronic exchange and in 2000, the SEC blew away the minimum price increments to one penny. These two modifications led to an immediate increase in electronic trading. The proportions of orders traded electronically increased form 18 percents in 2000, to more than 70 percents in 2011. Simultaneously, high frequency trading has become one of the most controversial topics in finance. Paul Krugman, wrote in the *New York Times*: 'High frequency trading probably degrades the stock market's function, because it is a kind of tax on investors who lack access to those superfast computers'. Moreover, it is widely believed that high frequency traders have amplified the effects of the *Flash crash* of May 2010. Meanwhile, many academics argue that high-frequency traders increase trading volume, reduce bid-ask spread and eventually reduce transaction costs for investors. In this introductive part we will briefly review the evolution of equity markets in the United States and describe how financial markets currently work.

1.1.1 Rise of algorithmic trading

Trading rules remained unchanged between the creation of the New York Stock Exchange (NYSE) in 1794 and the mid 1970s when the arrival of computers modified market microstructure. Until the 1970s the cost of trading was fixed to 0.25 percents and the minimum price increments was 1/8th of a dollar. Every day, an auction would take place and the highest bid would win. In 1971, the first electronic market, the NASDAQ, was created. Over the years, NASDAQ volume increased and developed its automated trading systems. NASDAQ was also the first stock market in the United States to start trading online. The increase of electronic trading led to a bigger competition between market makers and the bid-ask spread could be reduced. In June 1997, the minimum price increment on the NYSE dropped from 1/8th to 1/16th of a dollar. In 2000, the SEC decreased the high minimum spread and in early 2001, Nasdaq, Amex and NYSE moved to one cent for stocks whose value is bigger than one dollar.

The other change in regulation focused on transparency. In 1996, the SEC asked the market makers to display their orders. This new rule improved the price for investors and a National Best Bid Offer (NBBO) could be observed by investors at any time. In the 2000s, new venues were created and tried to attract liquidity. In order to attract liquidity, they started to offer *rebates* for liquidity providers and they decreased the fees of liquidity takers. With the spread decimalization and the fragmentation of liquidity across many venues, fewer orders are displayed on a specific location at the given time. From this time investors could no longer buy a large number of stocks at once. They now need to slice and dice orders.

Finally, in 2005, a new set of rules finalized the revolution of equity financial markets. The main changes are

- Order protection (rule 611): all trades have to occur at the NBBO (National Best Bid Offer). If one exchange does not have orders available at the NBBO, it has to rout market orders to an exchange with a better price.
- Sub-Penny rules (rule 610): Stocks, whose value is more than 1 dollar can not have a price increment less than one dollar. In practice, there is an exception for orders on dark pools and for hidden orders which allow mid-spread orders (half penny).

1.1.2 Impact of technological changes on equity markets

The consequences of this progressive change of microstructure environment are multiple (see e.g. Menkveld (2011a), Hendershott et al. (2012), Angel et al. (2010) for more details):

- Liquidity is now spread across many different financial exchanges. No individual venue accounts for more than 20 percent of overall trading volume. The market share of the New York Stock Exchange decreased from more than 90 percent in the beginning of the 1990s to less than 20 percent in 2011. Investors prefer to trade in new venues such as BATS or INET because these are fastest and offer better fees, both for liquidity suppliers and liquidity takers.
- The daily traded volume has increased (see figure 1.2). In the meantime, due to the slice and dice practice of investors, the average trade size has decreased (see figure 1.3).

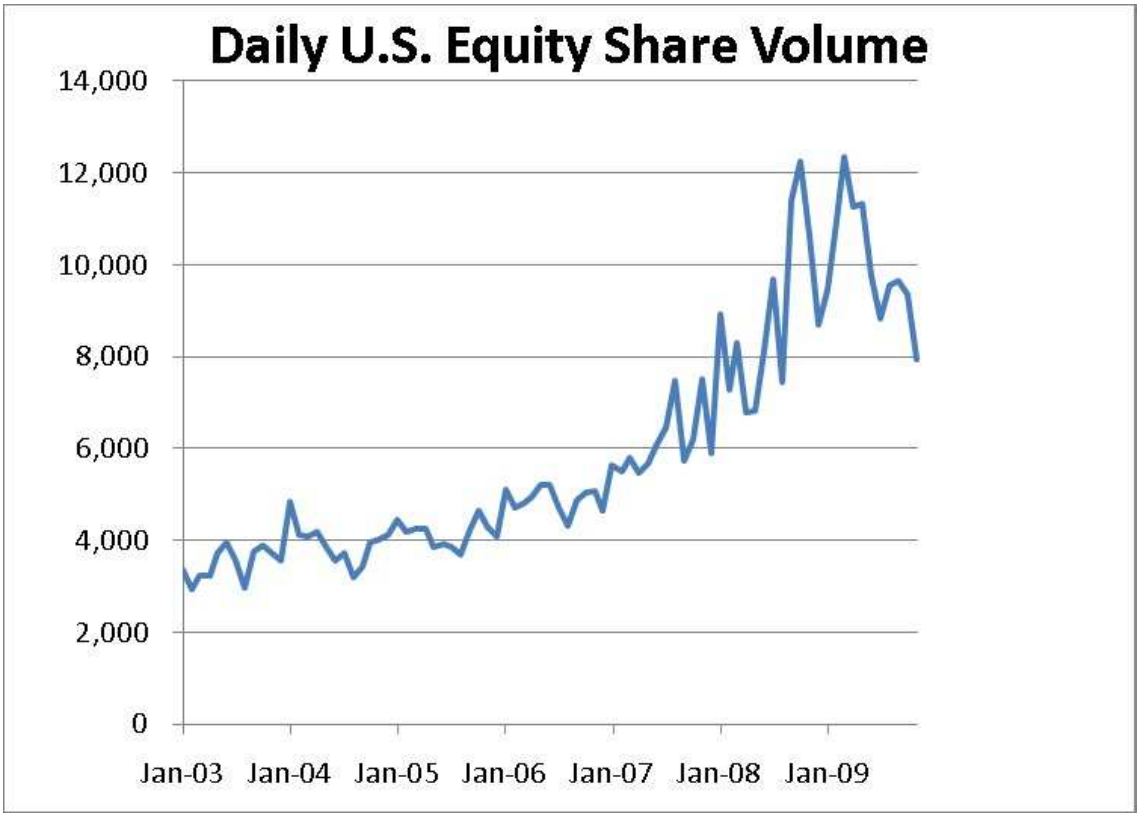

Figure 1.2: Average daily volume. Source: Barclays Capital Equity Research. Right.

- Bid-Ask spread of securities traded on electronic venues have substantially decreased. On figure 1.4 one can observe that the average spread has been divided by 5 in ten years.

Figure 1.3: Average trade size on the New York Stock Exchange. Source: NYSE-Euronext, nyx.com.

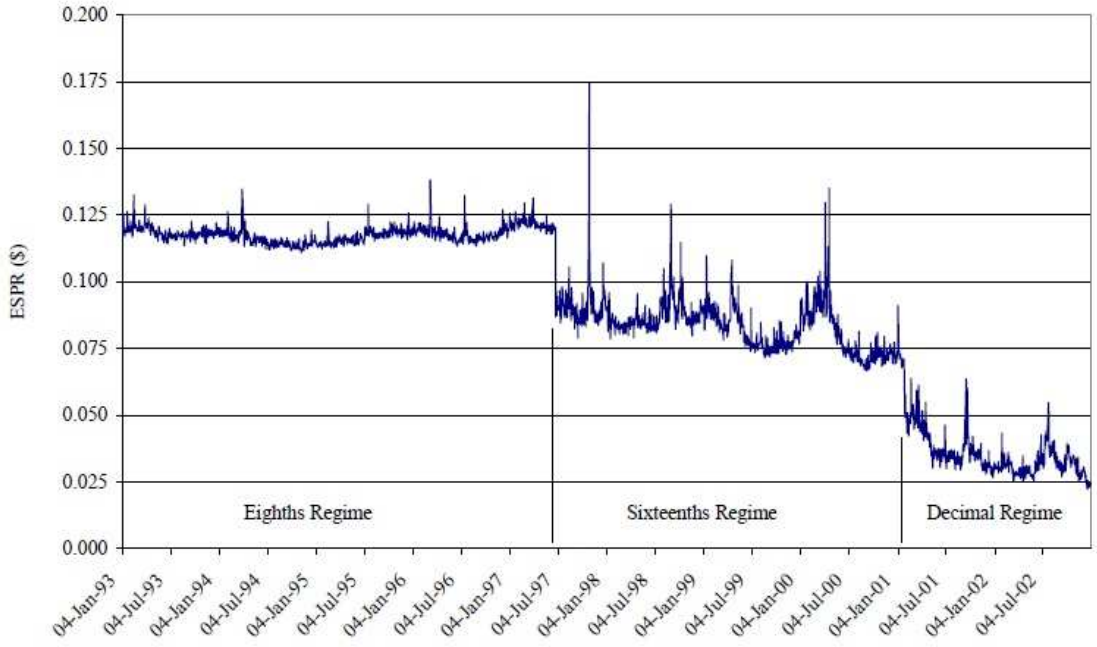

Figure 1.4: Average bid-ask spread for stocks listed on the NYSE. Source: Chordia et al. (2008).

- We have moved from context where professional human traders traded against each other to a context where algorithms, which can be run by anybody, play against each other.

- Speed and technology have become a key issue for a large number of investors. If, on one side, bid-ask spread and execution fees have substantially decreased, on the other side, investors need to use expensive computers and sophisticated techniques to minimize their trading impact and be executed at the best price.

1.1.3 High-frequency trading

High frequency traders account for more than 70 percent of the overall trading activity in the equity market in the United States (Brogaard (2010), Iati (2009)). These high frequency traders regroup a large pool of investors among which we can find:

Market makers

A market maker is an individual or a company which places both limit buy orders and limit sell orders on stocks. Market makers earn their living by matching buyers and sellers. If John wants to buy 200 shares of Microsoft at 30.25 dollars and Paul wants to sell 200 shares at 30.24, a market maker will be the intermediary that will deal with both John and Paul and eventually will earn 2 dollar on this transaction. The wider is the bid ask spread, the more profitable are market makers. The main source of risk of market makers comes from their inventory. Market makers quote limit orders on visible venues to earn the spread and generally use hidden pegged orders to liquidate their inventories. The behavior of market makers in electronic markets has recently been studied by Menkveld (2011b).

Since market makers always take positions in the opposite side of the market (they are selling when people buy and buying when people sell), they need to incorporate in their strategies all sort of information that may be correlated with forward returns. These pieces of information, generally called *signals*, are shared by all majors market makers. Hence one can view the whole population of market makers as a unique market maker, sometimes called the *omniscient market maker.*

Electronic Market makers

Some firms, who receive a flow of orders, are allowed to *internalize* these orders. Instead of routing the whole order flow to electronic venues, they can match buyers and sellers internally. These firms are called electronic market makers. In the United States, four electronic market makers, ATD, Citadel, UBS and Knight Capial, account for more than 60 percents of overall equity trading.

Institutional Investors

Institutional investors, who manage large portfolios, may sometimes have to buy or sell a significative proportion of a firm. In the 1980s, an institutional investor would call a broker and wait for a couple of minutes for the broker to decide the transaction price. During this time, the news of an investor trying to buy an important number of stock would leak and the stock price would increase. To avoid such a negative impact, investors decided in the 1990s to split their order in small parts and trade progressively. Several methods are well known to minimize trading impact, among which we can find TWAP, VWAP, POV (see e.g. Almgren and Chriss (1999) Almgren and Chriss (2001) Almgren and Chriss (2003) Predoiu et al. (2011) Gatheral (2010) Schied and Schoeneborn (2008) Guéant et al. (2011) Bouchard et al. (2011)) or more sophisticated strategies based on stochastic control. These strategies, which aim at minimizing the market impact of large trades, are executed over several hours or days of trading. At this time horizon, it is reasonable to model the price dynamics with a diffusion.

Execution traders

When an investor has decided to trade a certain amount of stocks during a short period of time, it still has the choice of using limit and market orders. Market orders are more expensive but are immediately executed, whereas limit orders carry the risk of not being executed. Investor can also chose between using hidden orders (orders that are not visible by others investors) and visible orders, trade on visible exchanges or dark pools, use limit or peg orders, and can go to about 30 different venues in the United States. Execution traders decide where and how to rout orders.

Statistical arbitrage/ Pure arbitrage

Algorithmic traders use short term *signals* to trade securities and hold positions for very brief periods of time. A signal is an information which is positively correlated with forward returns. For instance the following signals are used by both high frequency traders, market makers and arbitragers:

- Order book imbalance. As we will see in this thesis, the ratio between the number of orders at the ask and the number of orders at the bid is highly correlated with the probability that the next price move will be an increase.
- Order flow imbalance. In Cont et al. (2010a), the authors link the variation of price with what they call the order flow imbalance $OFI(\Delta T)$. The Order flow imbalance is the difference between all buy orders (including market orders and cancelations at the ask and limit orders at the bid) and all sell orders occurring during the time interval ΔT. Cont et al. (2010a) show a correlation close to one between variations of the order flow imbalance and price moves.
- Trade autocorrelations. It is widely known that the sequence of trade signs is highly correlated in time. With the sequence of trades, called the *tape*, one can easily create several signals.
- Many other signals can be built using latencies between venues, lead-lag effects, correlation across securities, etc..

Statistics arbitragers use these signals to buy or sell stocks for only a couple of seconds. For instance, if they find a strategy profitable 53 percent of the time, they can still earn their living by using this strategies many times, on non-correlated securities. Thanks to the law of large numbers, they hope to end up with a positive gain almost surely.

1.1.4 The consolidated limit order book

Every investor willing to trade a stock will have to chose between going to an electronic market maker, a visible venue or a dark pool. The global structure of equity markets is summarized on the graph 1.5

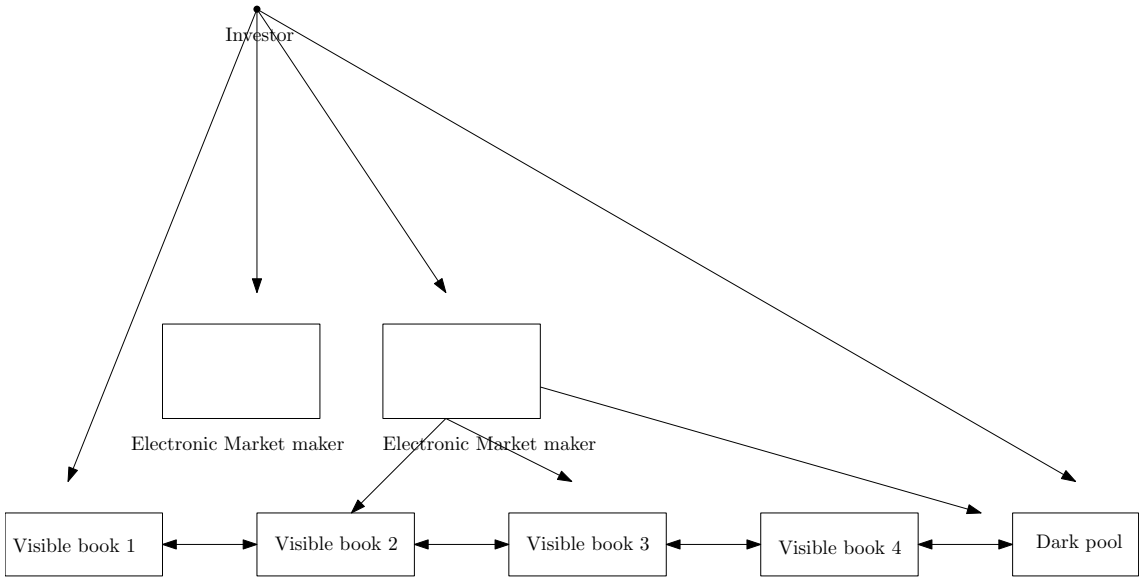

Figure 1.5: Actual market organization of Equity markets.

In practice, if a market order is sent to a venue that does not have orders available at the NBBO, this venue has to rout the orders to another venue with orders available at the NBBO. This rule links all limit order books together. Even if liquidity is fragmented geographically, it is reasonable to assume that all the liquidity is available in a virtual object called the *consolidated limit order book*.

Definition 1.1. *The consolidated limit order book is a virtual limit order book built with the aggregation of limit orders from all limit order books.*

The ask price can only increase when all limit orders at the ask level are consumed over all visible venues. It is equivalent to say that the ask price moves only when the ask queue is consumed on the consolidated limit order book. From a modeling point of view, the consolidated limit order book is the key object, at the core of price formation. In this thesis we will focus on the dynamics of this consolidated limit order book and neglect effects due to the fragmentation of equity markets.

Figure 1.6 displays a consolidated limit order book. One can observe on each line the venue where limit orders are displayed and the number of orders at the corresponding price.

Id	Hour	Qtty	Bid	Ask	Qtty	Hour	Id
BTRD	14:00:15	5	24.760	24.770	15	14:00:22	BTRD
ARCA	14:00:17	69	24.760	24.770	10	14:00:24	AMEX
INCA	14:00:19	14	24.760	24.780	1	14:00:19	PERT
CINN	14:00:21	78	24.760	24.780	1	14:00:23	COWN
SBSH	14:00:34	10	24.760	24.780	78	14:00:35	BRUT
MWRI	14:00:41	5	24.760	24.780	5	14:00:37	CINN
MLCO	13:59:01	1	24.750	24.790	250	13:57:37	FBCO
SIZE	13:59:31	10	24.750	24.790	15	14:00:30	TDCM
MADF	14:00:02	2	24.750	24.800	1	14:00:17	MLCO
MWSE	14:00:21	10	24.750	24.800	6	14:00:53	SBSH
SCHB	14:00:28	94	24.750	24.810	10	13:59:34	TMBR

Figure 1.6: A consolidated limit order book. Source: http://www.smartquant.com

1.2 Empirical studies

The availability of high frequency data on transactions, quotes and order flow in electronic order-driven markets has revolutionized data processing and statistical modeling techniques in finance and brought up new models based on statistical properties of financial time series. These models use the tools from time series analysis and econometrics to replicate several *stylized facts* observed on financial time series. In this section we briefly review some of the most well known empirical observations related to market microstructure.

1.2.1 Stylized facts of financial time series

This expression *stylized facts* refers to all empirical facts that arise in statistical studies of financial time series and that are persistent across various time periods, places, markets, assets, etc. One can find several of these facts in several reviews (e.g. Bollerslev and Engle (1993); Pagan (1996); Cont (2001)). A few stylized facts related to high frequency trading are:

- Negative autocorrelation of price return at a tick-by-tick time scale. On most stocks traded on electronic markets, consecutive increments of the price are negatively correlated. One generally says that the price mean reverts. On figure 1.7 we plot the autocorrelation of consecutive price increments for the stock Citigroup on the 26th of June, 08.

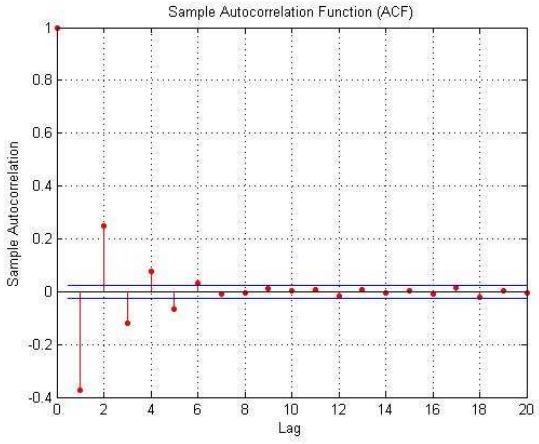

Figure 1.7: Autocorrelogram or the sequence of consecutive price increments for Citigroup, 26th June 08.

- Absence of autocorrelation of returns at a lower frequency. The autocorrelation of price increments at a lower frequency (e.g. a couple of minutes or more) disappears. As observed on figure 1.7, The autocorrelation function decays very rapidly to zero, even for a few lags of 1 minute. It is well known (see e.g. Pagan (1996); Cont (2001)), that the correlation between successive returns is close to zero.
- Correlation between the number of trades and volatility. It has been widely known for a long time that trading intensity is highly correlated with price volatility. Clark (1973) noticed, by observing cotton prices, a curvilinear relationship between trading volume and volatility. Silva and Yakovenko (2006) among others, show that the variance of log-returns after N trades, i.e. over a time period of N in trade time, is proportional to N.

- Microstructure invariants. Kyle et al. (2010) develop a theoretical framework to build microstructure invariants related to market depth, bid-ask spread, and order sizes.

1.2.2 Order size and volume

Intraday Seasonality

Activity on financial markets is not constant throughout the day. Every country has its deterministic pattern for trading intensity. In the United states, trading activity is intense after the opening at $9:30-10:30$am, then decreases during lunch time and eventually peaks again at the end of trading day between $3:00$pm and $4:00$pm. At the opening, the number of messages sent to electronic venues can be so high that the internal latency of the venues jumps from less than a millisecond to several hundreds of milliseconds. In Europe, the scenario is different, one can observe the same *U shape* of trading activity plus an increase of trading around $3:00$pm (French time), corresponding to the opening of Americans markets.

Autocorrelation of trade sides

Several empirical studies (Madhavan et al. (1997) Lillo and Farmer (2004) Bouchaud et al. (2002), Bouchaud et al. (2003)) highlight the high autocorrelation in the side of trades (e.g. buy or sell). Let $(\epsilon_i, i \geq 1)$ the sequence of trade signs. $\forall i \geq 1, \epsilon_i = 1$ when the i-th trade is a buy order, and $\epsilon_i = -1$ when it is a sell order. Lillo and Farmer (2004) Bouchaud et al. (2003) estimate the autocorrelogram of the sequence of trade signs and highlight the strong autocorrelation of these signs. They propose a model where:

$$\mathbb{E}[\epsilon_i \epsilon_{i+l}] = \frac{c_0}{l^\gamma} \quad \text{for} \quad \gamma < 1,$$

where c_0 is a constant. This high autocorrelation of trade signs may be explained by two factors. On one hand, most high frequency traders use the same signals to build their strategies and eventually trade in the same time periods. As we will show in this thesis, the order book pressure (ratio between bid and ask queue sizes) is a key indicator to the next price move. When the probability of a price increase becomes significant, many traders decide to take the remaining orders available at this price. A sudden flow of orders consumes the whole queue. On the other hand, we showed in section 1.1 that the overall liquidity available at a current time on a specific venue is small. Investors need to split orders and scatter them across time. By splitting orders and sending them one after another, investors increase correlation of trade signs. These two reasons explain why correlation in trade signs is strong.

Number of shares per order

Several studies have focus on the empirical distribution of trade sizes. Gopikrishnan et al. (2000) and Maslov and Mills (2001) observe a power law decay with an tail index between 2.3 and 2.7 for market orders and close to 2 for limit orders. Challet and Stinchcombe (2001) looked at the correlation between these trade sizes and noticed a clustering of orders. Figure 1.8 displays the sequence of signed order sizes for the stock Citigroup. Market orders, limit orders and cancelations are represented in these figures. Negative order sizes arise from market orders and cancelations whereas positive sizes stem from limit orders.

To our knowledge, no empirical study has emphasized the correlation between order sizes coming at the bid and at the ask. Another interesting fact observed on figure 1.9, is that the sequences of successive order sizes $(V_i^a, i \geq 1)$ are not correlated whereas the sequence of absolute value of order sizes are highly correlated.

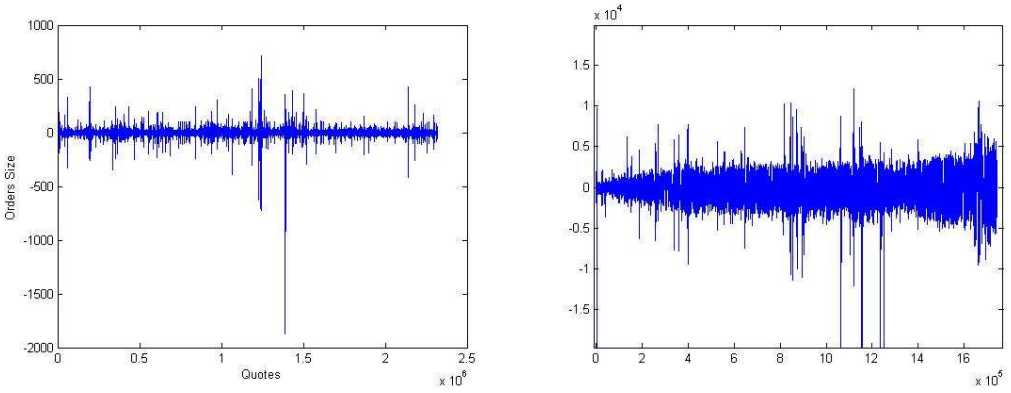

Figure 1.8: Number of shares per event for events affecting the ask (left) and the bid (right). The stock is Citigroup on the 26th of June 2008

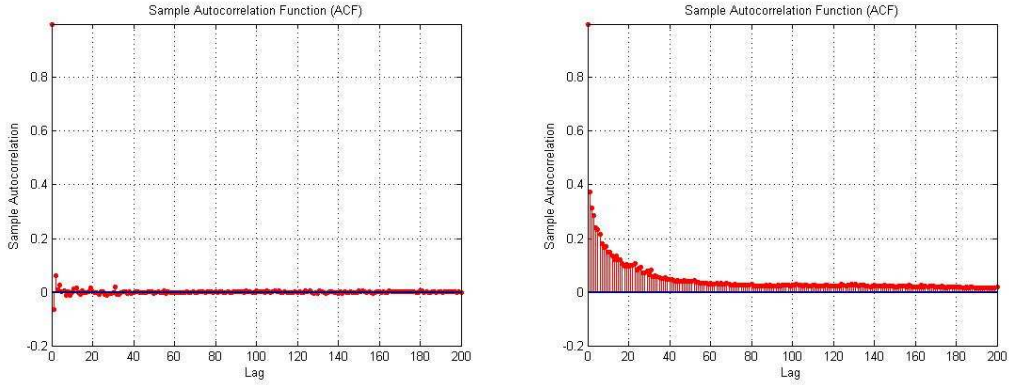

Figure 1.9: Autocorrelogram of order sizes (left) and absolute order sizes (right). The stock is Citigroup on the 26th of June 2008

We propose, in Chapter 3, a realistic model for the sequence of trade sizes $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ which captures empirical features of the sequence of order sizes:

$$V_i^b = \sigma_i^b z_i^b, \qquad V_i^a = \sigma_i^a z_i^a, \tag{1.1}$$

$$(\sigma_i^b)^2 = \alpha_0^b + \alpha_1^b (V_{t-1}^b)^2, \qquad (\sigma_i^a)^2 = \alpha_0^a + \alpha_1^a (V_{t-1}^a)^2, \quad \text{where} \qquad (z_i^b, z_i^a)_{i\geq 1} \overset{IID}{\sim} N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{1.2}$$

and $(\alpha_0^b, \alpha_1^b, \alpha_0^a, \alpha_1^a)$ are real coefficients satisfying

$$\alpha_0^b + \alpha_1^b < 1, \quad \text{and} \quad \alpha_0^a + \alpha_1^a < 1. \tag{1.3}$$

As shown by Bougerol and Picard (1992), the sequences of order sizes $(V_i^b, i \geq 1)$ and $(V_i^a, i \geq 1)$ are then well defined, stationary sequences of random variables with finite second-order moments. The model of order sizes given equation (1.1) is the equivalent of the ARCH model for returns. It captures the following features:

- The sequence of consecutive order sizes $(V_i^a, i \geq 1)$ (resp. $(V_i^b, i \geq 1)$) is not correlated.
- The sequence of squared order sizes $((V_i^a)^2, i \geq 1)$ (resp. $((V_i^b)^2, i \geq 1)$) is positively correlated.
- Order sizes at the ask and at the bid are correlated.

1.2.3 Average shape of the limit order book

The average shape of the limit order book, displayed figure 1.10 for the Japanese stock Sky, is the average number of orders posted at a fixed distance from the other side have been examined in various markets.

The average shape of the order book is observed to be hump shaped on each side, with a maximum close to the best quote (typically at one level from the best quote) and decreasing beyond the second level (Bouchaud et al. (2002) Bouchaud et al. (2003)).

Alfonsi et al. (2010) uses this order book shape to minimize trading costs in a limit order book.

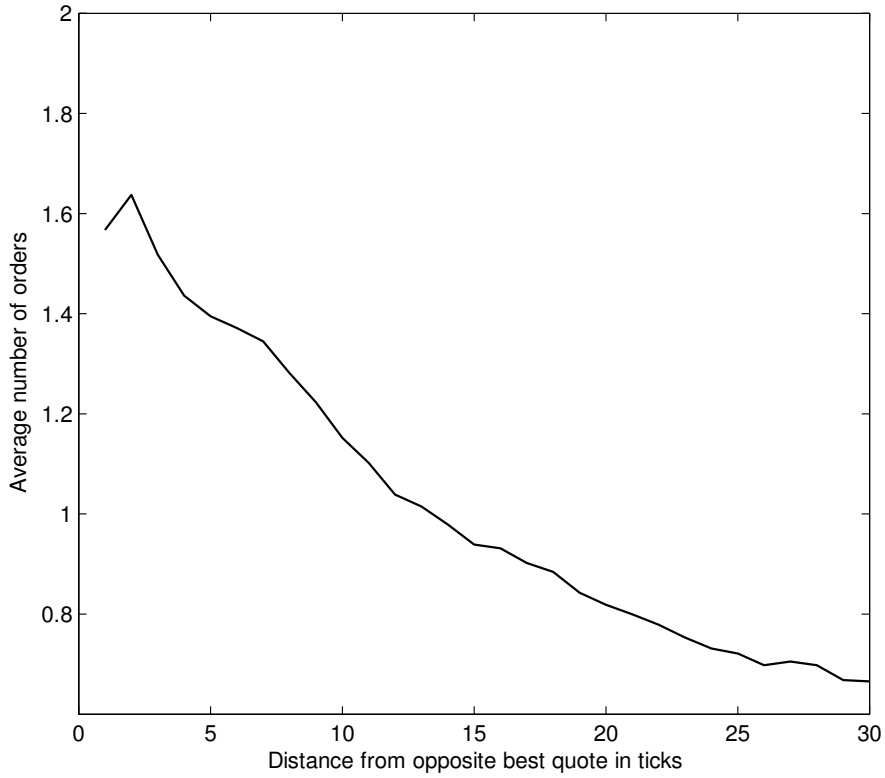

Figure 1.10: Average number of orders posted at a fixed distance from the opposite side for the Japanese stock Sky.

1.2.4 Market impact modeling

A controversial topic in financial microstructure is the shape of the market impact function. The market impact function is the price impact at time $t + T$ caused by an order of size V placed at time T. Many authors have conducted empirical studies to estimate the relation between market impact and order sizes. An exhaustive synthesis is proposed by Bouchaud et al. (2008), where different types of impacts, as well as some theoretical models are discussed. Generally, theoretical models assume two kinds of market impact, a temporary market impact which vanishes after a long time, and a permanent market impact. Gatheral (2010) proves that, only certain form of market impact are possible to avoid arbitrage strategies. More recently, Cont et al. (2010a) take a different approach on market impact. Instead of only looking at the impact of market orders, the authors estimate the impact of all orders affecting the order book (market orders, limit orders and cancelations) and show that the market impact is a linear function of net order flow.

1.3 Theoretical models

Market microstructure is a branch of economics and finance concerned with the details of how exchange occurs in financial markets. At the end on the 19th century, Walras noticed that the process of price formation could lead to unstable equilibrium. He was the first to offer a theoretical model for price equilibrium, where the price appeared after a process of 'tatonnement'. In perfect markets, Walrasian equilibrium prices reflect the competitive demand curves of all potential investors. Later, Garman (1976) developed in 1976 one of the earliest models of dealership and auction markets and went so far as to deduce the statistical properties of prices by simulating the order-arrival process. Modern market microstructure confronts microeconomic theory to the actual workings of markets. Amihud and Mendelson (1980), Kyle (1985), O'Hara (1997), Chakraborti et al. (2011) and Hasbrouck (2007) provide extensive overviews of the market microstructure literature. One can distinguish between three classes of theoretical models:

- the econometric literature, which models the price process without incorporating parameters of the order flow or utility functions of investors,
- the agent-based models, which deduct property of the price process from the behavior of investors, and
- the stochastic models of limit order books, which reproduce dynamical properties of limit order books and price process by modeling order flow as a stochastic process.

1.3.1 Econometric models of price dynamics

The high frequency dynamics of the mid-price (or bid/ask price) $(P_t, t \geq 0)$ is a piecewise constant stochastic process. The mid price only moves when either the bid or the ask queue empties or when a limit order is posted in the spread. Denote $(X_1, ..., X_n)$ the consecutive moves of the mid price.

$$\forall t \geq 0, \quad P_t = S_{N_t}, \quad \text{where} \quad S_n = X_1 + ... + X_n \quad \text{and} \quad (N_t, t \geq 0)$$

counts the number of price moves during $[0, t]$. Econometric studies model the distribution of the discrete price increments $(X_1, ..., X_n)$ and the durations between price changes. More generally, several econometric studies model durations between events.

Modeling durations between events

The observation of non-Poissonian arrival times (cf figure 4.1) has generated interest in modeling of durations between events. Engle and Russell (1998) and Engle and Lunde (2003) have introduced autoregressive condition duration or intensity models that may help modeling these processes of orders submission. See Hautsch (2004) for a textbook treatment.

Models based on Poisson point processes

The simplest way to model inter-events durations or price durations uses Poisson processes. In Chapter 2, we model the arrival rate of limit orders, market orders and cancelations with independent Poisson processes of parameters λ, μ and θ.

Self-exciting point processes

Several features, which are not captured in models based on Poisson processes, may be adequately represented by a multidimensional *self-exciting* point process (Andersen et al. (2010), Hautsch (2004)), in which the arrival rate $\lambda_i(t)$ of an order of type i is represented as a stochastic process whose value depends on the recent history of the order flow: each new order increases the rate of arrival for

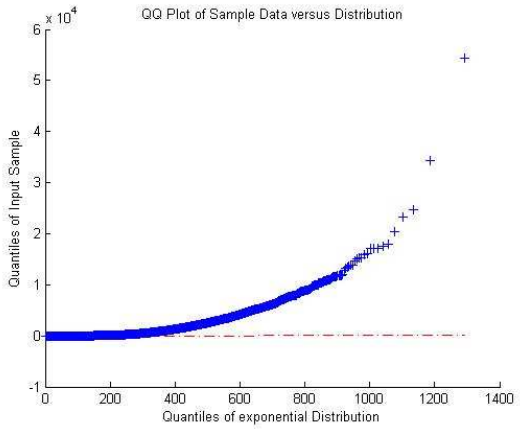

Figure 1.11: Quantiles of inter-event durations compared with quantiles of an exponential distribution with the same mean (Citigroup, June 2008). The dotted line represents the benchmark case where the observations are exponentially distributed, which is clearly not the case here.

subsequent orders of the same type (self-exciting property) and may also affect the rate of arrival of other order types (mutually exciting property):

$$\lambda_i(t) = \theta_i + \sum_{j=1}^{J} \delta_{ij} \int_0^t e^{-\kappa_i(t-s)} dN_j(s)$$

Here δ_{ij} measures the impact of events of type j on the rate of arrival of subsequent events of type i: as each event of type j occurs, λ_i increases by δ_{ij}. In between events, $\lambda_i(t)$ decays exponentially at rate κ_i. Maximum likelihood estimation of this model on TAQ data Andersen et al. (2010) show evidence of self-exciting and mutually exciting features in order flow: the coefficients δ_{ij} are all significantly different from zero and positive, with $\delta_{ii} > \delta_{ij}$ for $j \neq i$.

Autoregressive conditional durations

Models based on Poisson processes fail to capture serial dependence in the sequence of durations, which manifests itself in the form of clustering of order book events. One approach for incorporating serial dependence in event durations is to represent the duration duration τ_i between transactions $i-1$ and i as

$$T_i = \psi_i \epsilon_i,$$

where $(\epsilon_i)_{i\geq 1}$ is a sequence of independent positive random variables with common distribution and $\mathbb{E}[\epsilon_i] = 1$ and the *conditional duration* $\psi_i = \mathbb{E}[T_i | \psi_{i-j}, T_{i-j}, j \geq 1]$ is modeled as a function of past history of the process:

$$\psi_i = G(\psi_{i-1}, \psi_{i-2}, ..., ..; T_{i-1}, T_{i-2}, ..., ..).$$

Engle and Russell's Autoregressive Conditional Duration model Engle and Russell (1998) proposes an ARMA(p, q) representation for G:

$$\psi_i = a_0 + \sum_{i=1}^{p} a_k \psi_{i-k} + \sum_{i=1}^{q} b_q T_{j-k}$$

where $(a_0, ..., a_p)$ and $(b_1, ..., b_q)$ are positive constants. The ACD-GARCH model of Ghysels and Jasiak Ghysels and Jasiak (1998) combines this model with a GARCH model for the returns. Engle Engle (2000) proposes a GARCH-type model with random durations where the volatility of a price

change may depend on the previous durations. Variants and extensions are discussed in Hautsch (2004). Such models, like ARMA or GARCH models defined on fixed time intervals, have likelihood functions which are numerically computable.

Modeling price increments

Several econometric studies (Engle (1982) Engle and Russell (1998) Bollerslev (1986) Bollerslev and Engle (1993)) use time series analysis to model price dynamics. These models build time series that reproduce several stylized facts of price returns such as volatility clustering or the absence of autocorrelation of consecutive price increments to model price dynamics. This approach models price dynamics without looking at the order flow.

The ARCH(q) process, introduced by Engle (1982), is described as follows. Let $(\epsilon_n, n \geq 1)$ a sequence of IID Gaussian $\mathcal{N}(0,1)$ random variables and

$$\alpha_0 > 0 \quad \text{and} \quad \alpha_i \geq 0 \quad \text{for} \quad i \leq q \quad \text{with} \quad \alpha_0 + ... + \alpha_q < 1.$$

The sequence of price returns $(r_n, n \geq 1)$ defined as:

$$r_n = \sigma_n \epsilon_n, \quad \sigma_n^2 = \alpha_0 + \alpha_1 r_{n-1}^2 + \alpha_2 r_{n-2}^2 + ... + \alpha_q r_{n-q}^2$$

is an ARCH(q) process. This is a sequence of uncorrelated random variables with clustering volatility. Later, several econometric models such as GARCH, iGARCH generalize this ARCH time series to capture other features of price dynamics such as the skewness of returns and the fat tail of price returns.

1.3.2 Market microstructure models

Until the development of limit order markets at the end of the 1990s, the price discovery took place in the New York Stock Exchange (NYSE). In a 1995 paper, Hasbrouck (1985) used econometric methods to analyze the initial stages of the US stock market fragmenting into regional exchanges, and found that the price discovery mainly takes place in the New York Stock Exchange. At that time, most theoretical models aimed at modeling the behavior of the specialist of the NYSE, who was at the origin of price formation. These models are called *agent based models*.

All agent based models share the same framework: the demand for a risky asset comes from several agents with exogenously defined utility functions reflecting their preferences and risk aversions. Agent based models aim at linking the behavior of agents to the property of price microstructure (bid-ask spread, volatility, market impact, etc..). Some are capable of reproducing several *stylized facts* like the emergence of herding behavior Lux (1985), volatility clustering Lux and Marchesi (2000), or fat-tailed distributions of stock returns Cont and Bouchaud (2000), that are observed in financial data. The first agent based models include two classes of agents, informed traders who know the value of the traded security and noise traders. Noise traders buy and sell for various reasons such as hedging a position for instance. A recent review of agent based models can be found in Lebaron (2006) and Hasbrouck (2007). These theoretical models highlight that only a very small component of price volatility and trading volume is due to informed traders, with the most part produced by the noise traders.

In order to better understand the mechanism of agent based models we will detail two well-known models. The Kyle model focuses on price dynamics whereas Glosten's model highlights the relation between the bid-ask spread and the proportion of informed investors. These two models make very strong assumptions on the trading population and require the knowledge of parameters that cannot be estimated empirically. However, we will see that, despite their simplicity, these two models provide interesting insight on the relation between trading population and price microstructure.

Roll model (1984)

In 1984, Roll (1984) proposes a market microstructure model, for which the spread $s = 2c$ is constant, and trades arrive independently with constant order sizes. According to Roll's model, the market price $(p_t, t \geq 0)$ oscillates around an efficient price u_t:

$$\forall t \geq 0, \quad p_t = u_t + c \quad \text{if the trade } t\text{-th arrives at the ask and} \quad p_t = u_t - c \quad \text{if it occurs at the bid.}$$

One can easily show that:

$$\forall t \geq 0, \quad \mathbb{E}[\Delta p_t = 0], \quad \text{Var}[\Delta p_t] = \sigma_u^2 + 2c^2 \quad \text{and} \quad \text{Cov}(\Delta p_t, \Delta p_{t-1}) = -c^2,$$

where σ_u is the volatility of the efficient price.

This model, although being very simple, captures empirical properties of the price such as the negative autocorrelation of price increments (see table 4.4). We will see in this thesis that price mean-reversion holds under much more general assumptions and can be linked to some statistical properties of the order flow. The parametric model proposed in section 4.2.2 links this mean-reversion with the skewness of the distribution of the bid and ask queues after a price move.

Kyle model (1985)

Another well known model was introduced by Kyle in 1985 Kyle (1985). In Kyle's model, the traded security has a value v which follows a Gaussian distribution $\mathcal{N}(p_0, \sigma_0)$. The trading population consists of an informed trader, who knows the security value v and trades x shares, a liquidity trader (or noise trader) who submits a net order flow $l \sim \mathcal{N}(0, \sigma_l)$, independent of v, and a market maker or specialist who sets the clearing price p after observing the total order flow.

The informed trader conjectures that the market maker uses a linear price adjustment rule $p = \lambda(x + l) + \mu$, where λ is an inverse measure of liquidity. The expected profit of the informed traders is

$$\mathbb{E}[\pi] = x(v - \lambda x - \mu).$$

The informed traders chooses x to maximize $\mathbb{E}[\pi]$ yielding $x = (v - \mu)/2\lambda$. On the other side, market efficiencies requires $p = \mathbb{E}[v|l + x]$. A short computation shows that the equilibrium price is:

$$p = p_0 + \frac{1}{2}\frac{\sqrt{\sigma_0}}{\sigma_l}(l + x).$$

In a multi-period setting, the price evolution is given by the following equations

$$p_n = p_{n-1} + \lambda[x_n + l_n] \quad \text{and} \quad l_n = \beta_n[v - p_{n-1}], \tag{1.4}$$

where λ is a measure of liquidity, x_n is the quantity traded by the informed investor, l_n the number of orders traded by the noise traders.

In Kyle's model, price volatility is a multiple of the inverse measure of liquidity λ and the variance of trading intensity $x_n + l_n$. In Chapters 2 and 4, several results generalize this relation between a measure of liquidity and price volatility. Moreover we will see how to express λ with statistics from the order flow.

Glosten and Milgrom (1985)

One of the earliest model to focus on the bid-ask spread is given by Glosten and Milgrom (1985). Let's assume that there is one security with a payoff V which is either low $\underline{V}$ or high $\bar{V}$ with:

$$\mathbb{P}[V = \underline{V}] = \delta = 1 - \mathbb{P}[V = \bar{V}] \tag{1.5}$$

The trading population comprises informed and uninformed traders. Informed traders know the value outcome V. The proportion of informed traders in the population is μ. Given that the customer buys, the probability that $V = \underline{V}$ is

$$\mathbb{P}[V = \underline{V}|\text{Buy}] = \frac{\delta(1-\mu)}{1+\mu(1-2\delta)} \quad \text{and} \quad \mathbb{P}[V = \underline{V}|\text{Sell}] = \frac{\delta(1+\mu)}{1-\mu(1-2\delta)}. \tag{1.6}$$

The dealer's ask price A and bid price B are:

$$A = \mathbb{E}[V|\text{Buy}] = \frac{\underline{V}(1-\mu)\delta + \bar{V}(1-\delta)(1+\mu)}{1+\mu(1-2\delta)} \quad \text{and} \quad B = \mathbb{E}[V|\text{Sell}] = \frac{\underline{V}(1+\mu)\delta + \bar{V}(1-\delta)(1-\mu)}{1+\mu(1+2\delta)}. \tag{1.7}$$

The bid-ask spread is

$$A - B = \frac{4(1-\delta)\delta\mu(\bar{V} - \underline{V})}{1-(1-2\delta)^2\mu^2}.$$

After n trades, the bid-ask spread becomes approximately:

$$A_n - B_n = 4\delta_n(1-\delta_n)\mu(\bar{V} - \underline{V}),$$

where δ_n is the updated value of δ after n trades. We have neglected the term in μ^2, which should be small for realistic markets. This model is by construction compatible with a random walk for the midpoint. It also predicts that the bid-ask spread declines on average throughout the day, since the update rule drives δ_n either to zero or to one.

Agent based models of limit order markets

With the change of financial landscape described in section 1.1, the role of the specialist has weakened. Not only could investors chose the place were they would trade but they could also chose between using limit orders and market orders. Recently several agent based models (e.g. Parlour (1998), Foucault et al. (2005), Rosu (2009)) focused on this new type of financial markets. They have shown that the evolution of the price in such markets is rather complex and depends on the state of the order book. With the rise of algorithmic trading, Rosu (2009) has added a new population of traders called high frequency traders. Rosu (2009) links price volatility with trading intensity by the following relation:

$$\sigma = \frac{C}{\sqrt{\lambda}} \frac{i+1}{i} \sigma_u, \tag{1.8}$$

where λ is the total trading intensity, i is the proportion of informed traders, σ_u is a *fundamental* volatility and $C > 0$ is a constant. In this thesis, we show that this relation between order arrival intensity and price volatility holds under much more general assumptions, and may be derived without behavioral assumptions for market participants.

1.3.3 Stochastic models of limit order books

The search for tractable models of limit order markets has led to the development of stochastic models which aim to retain the main statistical features of limit order books while remaining computationally manageable. Stochastic models reproduce the dynamics properties of a limit order book without resorting to detailed behavioral assumptions about market participants or introducing unobservable parameters describing agent preferences, as in more detailed market microstructure models.

Probabilistic agent based models

When more agents appeared in agent based models (chartists, trend followers, etc..) with various utility functions, the complexity of the mathematics became untractable. Several people noticed that the mathematical complexity of such agents based models could disappear after a suitable scaling of the price process. In this approach, the intensity of trading would depend on the whole price trajectory and the scaled price process could take very different dynamics depending on the agent populations. Let $(s_t, t \geq 0)$ the price process. Generally, this process is so complex that it is impossible to get some insights on its dynamics. However, depending on the structure of the trading population, one can show that a suitable scaling $(\phi(n), n \geq 1)$ leads to:

$$\left(\frac{s_{[nt]}}{\phi(n)}, t \geq 0\right) \Rightarrow (P_t, t \geq 0), \quad \text{as} \quad n \to \infty \quad \text{on} \quad (\mathcal{D}, J_1),$$

where the scaled, or *marcoscopic* price process $(P_t, t \geq 0)$ has a much simpler behavior. All the short term dependence disappears during the scaling. For instance Föllmer and Schweizer (1993) and Horst (2005) model asset prices as a sequence of temporary equilibrium prices in a random environment of investor sentiment and show that in a noise trader framework, the scaled price process follows an Ornstein-Uhlenbeck dynamics with random coefficients. In Follmer et al. (2005), the agents are allowed to use technical trading rules. This generates a feedback from past prices into the environment. For a more detailed discussion of probabilistic agent-based models, we refer to Bayraktar et al. (2006).

Analogies with particle systems

Several physicists have modeled the limit order book as a particle system, where each limit order is a particle moving in the limit order book according to a specific dynamics. Bak et al. (1997) propose a market similar the reaction-diffusion model $A + B \to 0$ in physics. In such a model, two types of particles are inserted at each side of a pipe and move randomly with steps of size 1. Each time two particles collide, they disappear and two new particles are inserted. The price lies at the frontier where the two particle collide. When more particles are inserted on the right, the collision frontier will slowly move to the left. The selling pressure is equivalent to the number of particles leaving the right side of the pipe. Maslov (2000) adds more parameters in this model. First, limit orders are submitted and stored in the model, without moving. Second, limit orders are submitted around the best quotes. Third, market orders are submitted to trigger transactions. Challet and Stinchcombe (2001) extend the work of Bak et al. (1997) and Maslov (2000), and develop the analogy between dynamics of limit order books and particle systems on an infinite dimensional grid. Bovier et al. (2004), Bovier and Černý (2007) compute the hydrodynamic limit of the whole system.

A Markovian model

In Cont et al. (2010b), the authors model the limit order book as a continuous time Markov process $X_t = (X_1(t), X_2(t), ...X_n(t))$, where $-X_p(t)$ (resp. $X_p(t)$) is the number of sell (resp. buy) orders available at price p, for $p \in \{1, ..., n\}$. Under this framework, it follows immediately that for all $t > 0$, the bid and ask prices are:

$$b_t = \sup\{p,\ X_p(t) < 0\} \quad \text{and} \quad a_t = \inf\{p,\ X_p(t) > 0\} \tag{1.9}$$

The authors assume that the intensity of limit orders at level p is $\lambda(p)$ and that these limit orders are canceled at rate $\theta(p)$. Market orders arrive according to a Poisson process of intensity μ. Under these assumptions, the whole process $X(t)$ is Markovian and ergodic, and several quantities of interest - transition probabilities for prices, distribution of durations - may be computed using Laplace transform techniques.

1.4 Summary of contributions

This thesis may be viewed as a step towards bridging the gap between the econometrics literature, and the more traditional temporary equilibrium models, which allow for analytic solutions but do not accurately capture the microstructure of automated trading systems.

The idea is to model the inflows of orders as a random process, whose statistical properties are chosen to match the observed ones, and to *derive* the dynamics of prices from the resulting interaction of this order flow with the limit order book. Contrary to equilibrium models, based on agents preference, we do not need to model the behavior of agents, nor to estimate the proportion of a given kind of participants to understand price dynamics.

Linking statistics of the order flow with the dynamics of the price is a complex task because the properties of the order flow (time autocorrelation, dependence across different price levels, heterogeneity of order sizes...) generate a multidimensional non-Markovian stochastic process for the limit order book dynamics. First, our approach consist in reducing the number of variables characterizing the limit order book. Once we have reduced the state of variables, we show how to approximate the dynamics of this reduced order book by a Markov process.

1.4.1 A reduced form model of the limit order book

As discussed in section Cont et al. (2010b), one can model a limit order book as a high dimensional queueing system (or point process); this approach requires specifying the arrivals of orders at all price levels and may lead to intensive computations even in simple settings. By contrast, in this thesis, we propose a simple reduced-form approach.

Empirical studies of limit order markets suggest that the major component of the order flow occurs at the (best) bid and ask price levels (see e.g. Biais et al. (1995)). All electronic trading venues also allow to place limit orders *pegged* to the best available price (National Best Bid Offer, or NBBO); market makers used these *pegged* orders to liquidate their inventories. Furthermore, studies on the price impact of order book events show that the net effect of orders on the bid and ask queue sizes is the main factor driving price variations (Cont et al. (2010a)). These observations, together with the fact that queue sizes at the best bid and ask of the consolidated order book are more easily obtainable (from records on trades and quotes) than information on deeper levels of the order book, motivate a reduced-form modeling approach in which we represent the state of the limit order book by:

- the bid price s_t^b and the ask price s_t^a,
- the size of the bid queue q_t^b representing the outstanding limit buy orders at the bid, and
- the size of the ask queue q_t^a representing the outstanding limit sell orders at the ask

Figure 1.12 summarizes this representation.

If the stock is traded in several venues, the quantities q^b and q^a represent the best bids and offers in the *consolidated* order book, obtained by aggregating over all (visible) trading venues. At every time t, q_t^b (resp. q_t^a) corresponds to all visible orders available at the bid price s_t^b (resp. s_t^a) across all exchanges.

The state of the order book is modified by *order book events*: limit orders (at the bid or ask), market orders and cancelations (see Cont et al. (2010b,a), Smith et al. (2003)). A limit buy (resp. sell) order of size x increases the size of the bid (resp. ask) queue by x, while a market buy (resp. sell) order decreases the corresponding queue size by x. Cancelation of x orders in a given queue reduces the queue size by x. Given that we are interested in the queue sizes at the best bid/ask levels, market orders and cancelations have the same effect on the queue sizes (q_t^b, q_t^a).

The bid and ask prices are multiples of the tick size δ. When either the bid or ask queue is depleted by market orders and cancelations, the price moves up or down to the next level of the order book. The price processes s_t^b, s_t^a is thus a piecewise constant process whose transitions correspond to hitting times of the axes $\{(0,y), y>0\} \cup \{(x,0), x>0\}$ by the process $q_t = (q_t^b, q_t^a)$.

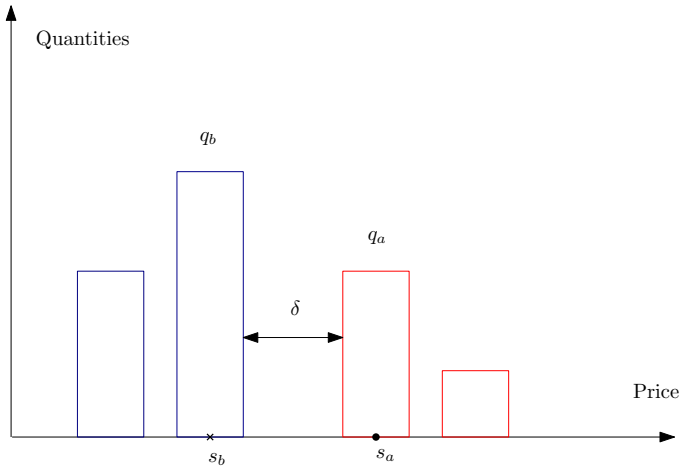

Figure 1.12: Simplified representation of a limit order book.

If the order book contains no 'gaps' (empty levels), these price increments are equal to one tick:

- when the bid queue is depleted, the (bid) price decreases by one tick.
- when the ask queue is depleted, the (ask) price increases by one tick.

If there are gaps in the order book, this results in 'jumps' (i.e. variations of more than one tick) in the price dynamics. We will ignore this feature in what follows but it is not hard to generalize our results to include it.

The quantity $s_t^a - s_t^b$ is the *bid-ask spread*, which may be one or several ticks. As shown in Table 1.1, for liquid stocks the bid-ask spread is equal to one tick for more than 98% of observations.

Bid-ask spread	1 tick	2 tick	$\geq$ 3 tick
Citigroup	98.82	1.18	0
General Electric	98.80	1.18	0.02
General Motors	98.71	1.15	0.14

Table 1.1: Percentage of observations with a given bid-ask spread (June 26th, 2008).

When either the bid or ask queue is depleted, the bid-ask spread widens immediately to more than one tick. Once the spread has increased, a flow of limit sell (resp. buy) orders quickly fills the gap and the spread reduces again to one tick. When a limit order is placed inside the spread, all the limit orders pegged to the NBBO price move in less than a millisecond to the price level corresponding to this new order. Once this happens, both the bid price and the ask price have increased (resp. decreased) by one tick.

The histograms in Figure 1.13 show that this 'closing' of the spread takes place very quickly: as shown in Figure 1.13 (left) the lifetime of a spread larger than one tick is of the order of a couple of milliseconds, which is negligible compared to the lifetime of a spread equal to one tick (Figure 1.13, right). In our model we assume that the second step occurs *infinitely fast*: once the bid-ask spread widens, it returns immediately to one tick. For the example of Dow Jones stocks (Figure 1.13), this is a reasonable assumption since the widening of the spread lasts only a few milliseconds. This simply means that we are not trying to describe/model how the orders flow inside the bid-ask spread at the millisecond time scale and, when we describe the state of the order book *after* a price change we have in mind the state of the order book *after* the bid-ask spread has returned to one tick.

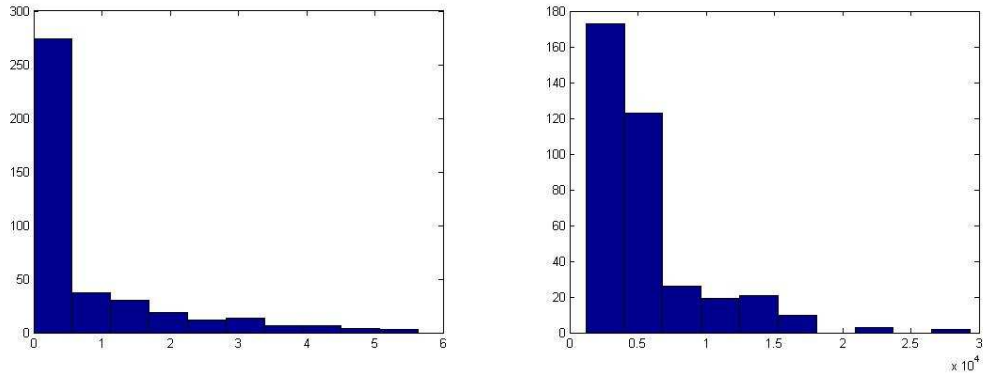

Figure 1.13: Left: average lifetime, in milliseconds of a spread larger than one tick for Dow Jones stocks. Right: average lifetime, in milliseconds of a spread equal to one tick.

Under this assumption, each time *one* of the queues is depleted, *both* the bid queue and the ask queues move to a new position and the bid-ask spread remains equal to one tick after the price change. Thus, under our assumptions the bid-ask spread is equal to one tick, i.e. $s_t^a = s_t^b + \delta$, resulting in a further reduction of dimension in the model.

The limit order book as a reservoir of liquidity

Once either the bid or the ask queue are depleted, the bid and ask queues assume new values. Instead of keeping track of arrival, cancelation and execution of orders at *all* price levels (as in Cont et al. (2010b), Smith et al. (2003)), we treat the queue sizes after a price change as a stationary sequence of random variables whose distribution represents the depth of the order book in a statistical sense. More specifically, we model the size of the bid and ask queues after a price increase by a stationary sequence $(R_k)_{k\geq 1}$ of random variables with values in $\mathbb{N}^2$. Similarly, the size of the bid and ask queues after a price decrease is modeled by a stationary sequence $(\tilde{R}_k)_{k\geq 1}$ of random variables with values in $\mathbb{N}^2$. The sequences $(R_k)_{k\geq 1}$ and $(\tilde{R}_k)_{k\geq 1}$ summarize the interaction of the queues at the best bid/ask levels with the rest of the order book, viewed here as a 'reservoir' of limit orders.

The variables R_k (resp. $\tilde{R}_k$) have a common distribution which represents the *depth* of the order book after a price increase (resp. decrease): Figure 1.14 shows the (joint) empirical distribution of bid and ask queue sizes after a price move for Citigroup stock on June 26th 2008.

The simplest specification could be to take $(R_k)_{k\geq 1}$, $(\tilde{R}_k)_{k\geq 1}$ to be IID sequences; this approach, used in chapter 2, turns out to be good enough for many purposes. But this IID assumption is not necessary and will be released in chapter 3

In summary, the state of the limit order book is thus described by a continuous-time process (s_t^b, q_t^b, q_t^a) which takes values in the discrete state space $\delta\mathbb{Z} \times \mathbb{N}^2$, with piecewise constant sample paths whose transitions correspond to the order book events. Denoting by $(t_i^a, i \geq 1)$ (resp. t_i^b) the event times at the ask (resp. the bid), V_i^a (resp. V_i^b) the corresponding change in ask (resp. bid) queue size, and $k(t)$ the number of price changes in $[0, t]$, the above assumptions translate into the following dynamics for (s_t^b, q_t^b, q_t^a):

- If an order or cancelation of size V_i^a arrives on the ask side at $t = t_i^a$,
  - if $q_{t-}^a + V_i^a \geq 0$, the order can be satisfied without changing the price;
  - if $q_{t-}^a + V_i^a < 0$, the ask queue is depleted, the price increases by one 'tick' of size δ, and the queue sizes take new values $R_{k(t)} = (R_{k(t)}^b, R_{k(t)}^a)$,

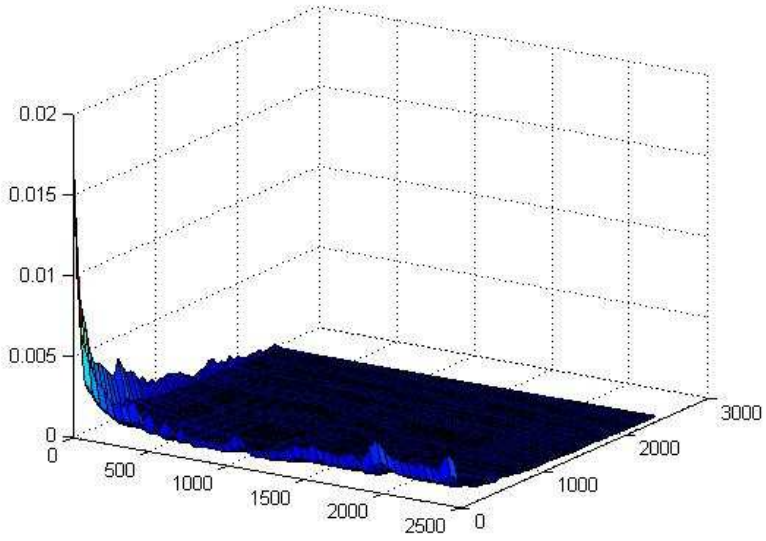

Figure 1.14: Joint density of bid and ask queue sizes after a price move (Citigroup, June 26th 2008).

$$(s^b_t, q^b_t, q^a_t) = (s^b_{t-}, q^b_{t-}, q^a_{t-} + V^a_i)1_{\{q^a_{t-} \geq -V^a_i\}} + (s^b_{t-} + \delta, R^b_{k(t)}, R^a_{k(t)})1_{\{q^a_{t-} < -V^a_i\}}, \tag{1.10}$$

- If an order or cancelation of size V^b_i arrives on the bid side at $t = t^b_i$,
  - if $q^b_{t-} + V^b_i \geq 0$, the order can be satisfied without changing the price;
  - if $q^b_{t-} + V^b_i < 0$, the bid queue gets depleted, the price decreases by one 'tick' of size δ and the queue sizes take new values $\tilde{R}_{k(t)} = (\tilde{R}^b_{k(t)}, \tilde{R}^a_{k(t)})$:

$$(s^b_t, q^b_t, q^a_t) = (s^b_{t-}, q^b_{t-} + V^b_i, q^a_{t-})1_{\{q^b_{t-} \geq -V^b_i\}} + (s^b_{t-} - \delta, \tilde{R}^b_{k(t)}, \tilde{R}^a_{k(t)})1_{\{q^b_{t-} < -V^b_i\}}. \tag{1.11}$$

Analogies with queueing systems

The dynamics of bid and ask queue sizes (q^b, q^a) may be viewed as a system of two queues evolving with the impact of limit orders, market orders and cancelations. For a general order book, the dependence between bid and ask order flows generates a correlation between bid and ask queues dynamics. This queueing system, for which the dynamics of queue lengths is correlated, is similar to tandem queues in dimension 2, for which every customer waits for a first server, then goes to a second server and eventually leaves the system. The waiting time of a customer entering such systems is the analogous of the duration until a limit buy order (resp. sell order) is executed. Similarly, the time when one server becomes idle corresponds to the time when the price moves in our system. Tandem queues have been widely studied. Harrisson (1978) computes the heavy-traffic approximation of the queues, and gives its stationary distribution. In dimension 2, a general method (Cohen (1988), Cohen and Boxma (1983)), based on the resolution of a Riemann-Hilbert boundary problem, gives the stationary distribution of queueing systems. A similar process, diffusing in the positive orthant with discontinuous reflection on the axis, have been studied by Baccelli and Fayolle (1987). The main difference between the dynamics of bid and ask queue sizes and queueing systems lies on the boundary. Most queueing systems have continuous reflection conditions at the boundaries, whereas our system is reinstated inside the orthant once it hits an axis. Moreover the price process $(s_t, t \geq 0)$ has no equivalent in queueing theory.

Quantities of interest

We will show in this thesis that this simplified approach allows to compute several quantities of interest such as:

- *Price durations.* The price moves when either the bid queue or the ask queue becomes empty. Let τ the distribution of the price move. We will see how to compute the unconditional distribution τ. We will also compute the distribution τ conditioned on observing at a given time x shares at the bid and y shares at the ask.
- *Probability of a price increase p^{up}.* The price increases if the ask queue empties before the bid queue. We will give an explicit expression of $p^{up}(x, y)$, the probability that the next price move is an increased, conditioned on observing x shares at the bid and y shares at the ask.
- *Autocorrelation of price increments.* Empirically, price mean reverts at a tick-by-tick time scale. We will see how to link this price mean reversion with parameters from the order flow. We will notice that this price mean reversion is closely linked to the skewness of the queue sizes (q^b, q^a) after a price move.
- *Price volatility.* The main achievement of this thesis is to relate price volatility with statistics of the order flow. We will give explicit expression for the volatility under very general assumptions and link price volatility with order flow statistics.
- *Drift of the price.* If the order flow is not completely symmetric at the bid and at the ask, the price drifts in one direction. We will express the price drift as a function of order flow parameters.

Comparison between our model and other microstructure models

Given the important number of market microstructure models reviewed in section 1.3, it is good to sum up the assumptions shared by these models (cf. table 1.2).

Assumptions/ Models	Constant spread	Constant order sizes	Existence of an efficient price	Specialist driven market	Agent-based model
Roll	×	×	×		
Glosten and Milgrom		×	×	×	×
Kyle			×	×	×
Rosu	×	×	×		×
Model proposed in this thesis	×				

Table 1.2: Assumptions made for each microstructure model

The approach developed in this thesis does not assume the existence of an unobservable efficient price, around which market prices would oscillate as it is for the Roll model. Neither do we split the trading population in various categories (e.g. patient buyers, informed traders, noise sellers, etc..). The properties of the price process (drift, volatility, distributions of marginals for instance) result directly from statistical properties of the order flow. Market participants - investors, market makers and arbitragers - generate a complex order flow that can lead to several regimes of the price process:

- When the order flow is symmetric at the bid and at the ask, we show that the daily behavior of the price is a driftless Brownian motion.
- When the order flow is asymmetric at the ask and at the bid, we prove that the price behavior is a Brownian motion with drift. We provide analytical expressions for both the drift and the volatility.
- A market crash can occur when the order of magnitude of buy orders is not comparable to the order of magnitude of sell orders. The distribution of price jumps may be linked to properties of the order flow.

Despite the weak number of assumptions (cf table 1.3), the framework described in this document allows to explain more stylized facts than most market microstructure models.

Stylized facts/ Models	Negative autocorrelation of price increments	Link between trading intensity and volatility	Distribution of price duration	Explicit expression of p^{up}
Roll	×			
Glosten and Milgrom				
Kyle	×	×		
Rosu (2009)		×		
Model from this thesis	×	×	×	×

Table 1.3: Stylized facts captured by each microstructure model

Where p^{up} is the probability that the price increases conditioned on the state of the limit order book. Analytical results given in this thesis are always expressed with observable quantity of the order flow (such as the average duration between order sizes or the variance of order sizes). It is not always the case for classical microstructure models which assume the existence of fixed proportions of agents that can not be easily estimated with market data.

On the other hand, the results obtained in this thesis may be applied to *agent based models* when the description of the agents is clear enough to fully characterize of the order flow. In Section 3.4 we give the example of a market with three agents that use limit and market orders in different ways: one market maker uses only limit orders and two agents use both market and limit orders. We provide a complete description of the price process, from the tick-by-tick price evolution to the daily dynamics of the price and we link the standard deviation of price increments with the proportion of each agent interacting in this market.

1.4.2 Heavy traffic approximations

For a general order flow, the dynamics of the limit order book $(q_t^b, q_t^a)_{t\geq 0}$ is complex and non-Markovian. The idea, to transform the process (q_t^b, q_t^a) into a tractable Markov process, consists in regrouping orders pack by pack and scale order sizes. When the size of these scaled batches of orders goes to infinity, we will prove in chapter 3 that the limit order book converges in distribution to a Markov process (Q^b, Q^a) whose generator can be related to order flow statistics. The limit process (Q^b, Q^a) is called the *Heavy traffic approximation* of (q^b, q^a). Heavy traffic approximations of non-Markovian stochastic processes have been widely studied in queueing theory for more than twenty years with applications in various domain of the industry such as call center optimization, bus scheduling or hospital appointment bookings. The mathematical tool behind these approximations is the functional central limit theorem. A functional central limit theorem is a central limit theorem applied at a sequence of stochastic processes. In this thesis we assume that the convergence of stochastic processes holds on the Skorokhod space $\mathcal{D}$ with the J_1 or the M_1 topology. A complete review of the J_1 and M_1 topologies can be found in Whitt (2002). Heavy traffic approximation of queuing networks have been widely studied (Iglehart and Whitt (1970), Harrisson (1973), Harrisson (1978), Reiman (1977), Dai and Nguyen (1994), Bramson and Dai (2001)).

Let n be the average number of orders coming at the bid or ar the ask per period of 10 seconds. We notice on table 1.4 that this number is very large. If one observes the limit order book every 10 seconds, the impact of one order is negligible. At this time scale, the order book has a much smoother behavior. The limit order book diffuses.

	Average no. of orders in 10s
Citigroup	4469
General Electric	2356
General Motors	1275

Table 1.4: Average number of orders in 10 seconds (June 26th, 2008).

Denote $(V_i^b, i \geq 1)$ (resp. $(V_i^a, i \geq 1)$) the signed number of shares of the i-th orders. For all $i \geq 1$, $V_i^b < 0$ (resp. $V_i^a < 0$) if the i-th order arriving at the bid side (resp. ask side) is either a market order or a cancelation. On the other hand, $V_i^b > 0$ (resp. $V_i^a > 0$) if the i-th order arriving at the bid side (resp. ask side) is a limit order. A suitable rescaling is a function $\phi = (\phi_n, n \geq 1)$ with the property that the following sequence of cumulated order sizes converges in distribution to a non-degenerate random variable (U, V), when properly scaled by $(\phi_n, n \geq 1)$:

$$\left(\frac{V_1^b + V_2^b, ..., V_n^b}{\phi_n}, \frac{V_1^a + V_2^a, ..., V_n^a}{\phi_n}\right) \Rightarrow^{\mathcal{D}} (U, V) \tag{1.12}$$

Fluid limit of the limit order book

When market orders and cancelations dominate limit orders, bid and ask queue sizes have a negative drift,

$$\mathbb{E}[V_i^a] < 0, \quad \text{and} \quad \mathbb{E}[V_i^b] < 0.$$

Therefore, when $(V_i^b, V_i^a)_{i\geq 1}$ is a stationary, ergodic sequence of random variables, the law of large numbers holds:

$$\left(\frac{V_1^b + ... + V_n^b}{n}, \frac{V_1^a + ... + V_n^a}{n}\right) \Rightarrow (\mathbb{E}[V_i^b], \mathbb{E}[V_i^a]).$$

One can prove, under minor assumptions, that the following sequence of processes converges in law to a Markov process $\tilde{Q} = (\tilde{Q}^b, \tilde{Q}^a)$ on the Skorokhod space $(\mathcal{D}, J_1)$:

$$\left(\frac{q^b_{[nt]}}{n}, \frac{q^a_{[nt]}}{n}\right)_{n\geq 1} \Rightarrow (\tilde{Q}^b_t, \tilde{Q}^a_t) \text{ as } n \to \infty \text{ on } (\mathcal{D}, J_1). \tag{1.13}$$

When $n >> 1$, the *Heavy traffic approximation* consists in approximating the dynamics of the limit order book process $(q^b_t, q^a_t)_{t\geq 0}$ by the Markov process $(\tilde{Q}^b_t, \tilde{Q}^a_t)_{t\geq 0}$.

Diffusion limit of the limit order book

On the other hand, when the average flow of limit orders equals the average flow of market orders and cancelations,

$$\mathbb{E}[V^a_i] = 0, \quad \text{and} \quad \mathbb{E}[V^b_i] = 0.$$

When $(V^a_i, V^b_i)_{i\geq 1}$ is a stationary, weakly dependent (α mixing for instance) sequence of random variables with finite second moment, the central limit theorem yields:

$$\left(\frac{V^b_1 + ... + V^b_n}{\sqrt{n}}, \frac{V^a_1 + ... + V^a_n}{\sqrt{n}}\right) \Rightarrow \mathcal{N}(0, \Sigma),$$

where $\mathcal{N}(0, \Sigma)$ is a centered Gaussian random variable whose covariance matrix can be expressed with the moments of the sequence of order sizes $(V^a_i, V^b_i)_{i\geq 1}$.

The functional central limit version of this limit theorem states, under some minor assumptions on the sequence of durations and the distribution of queue sizes after a price move, that the following sequence of processes converges to a Markov process (Q^b, Q^a):

$$\left(\frac{q^b_{[nt]}}{\sqrt{n}}, \frac{q^a_{[nt]}}{\sqrt{n}}, t \geq 0\right)_{n\geq 1} \Rightarrow (Q^b_t, Q^a_t)_{t\geq 0} \text{ as } n \to \infty \text{ on } (\mathcal{D}, J_1). \tag{1.14}$$

This process (Q^b, Q^a) is called the *heavy traffic approximation* of (q^b, q^a).

Fluid limit or diffusive limit ?

Properties of the order flow change with n. Trading intensity influences statistical properties of the order flow. This is why one needs to consider a sequence of order books $q^n = (q^{n,b}, q^{n,a})$ deriving from a sequence of order flows given by $(V^{n,b}_i, V^{n,a}_i, T^{n,b}_i, T^{n,a}_i)_{i\geq 1}$. If one assumes that the inter-event durations are finite (or asymptotically finite), the impact, on the net order flow, of a batch of n events at the ask is given by:

$$\frac{V^{n,a}_1 + V^{n,a}_2 + V^{n,a}_3 + ... + V^{n,a}_n}{\sqrt{n}} = \frac{(V^{n,a}_1 - \overline{V^{n,a}}) + (V^{n,a}_2 - \overline{V^{n,a}}) + ... + (V^{n,a}_n - \overline{V^{n,a}})}{\sqrt{n}} + \sqrt{n}\overline{V^{n,a}},$$

where $\overline{V^{n,a}} = \mathbb{E}[V^{n,a}_1]$. Under appropriate assumptions, this sum behaves approximately as a Gaussian random variable for large n:

$$\frac{V^{n,a}_1 + V^{n,a}_2 + V^{n,a}_3 + ... + V^{n,a}_n}{\sqrt{n}} \sim \mathcal{N}(\sqrt{n}\overline{V^{n,a}}, \sqrt{\text{Var}(V^{n,a}_1)}) \text{ as } n \to \infty. \tag{1.15}$$

Two regimes are possible, depending on the behavior of the ratio $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\text{Var}(V^{n,a}_1)}}$ as n grows:

- If $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\text{Var}(V^{n,a}_1)}} \to \infty$ as $n \to \infty$, the correct approximation is given by the *fluid limit*, which is deterministic.

- If $\lim_{n\to\infty} \frac{\sqrt{n}\overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_1^{n,a})}} < \infty$, the rescaled queue sizes behave like a diffusion process.

The fluid limit corresponds to the regime of law of large numbers, where random fluctuations average out and the limit is described by average queue size, whereas the diffusion limit corresponds to the regime of the (functional) central limit theorem, where fluctuations in queue size are asymptotically Gaussian.

Figure 1.15 displays the histogram of the ratio $\frac{\sqrt{n}\overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_1^{n,a})}}$ for stocks in the Dow Jones index, where n is chosen to represent the average number of order book events in a 10 second interval (typically $n \sim 100 - 1000$). This ratio is shown to be rather small at such intraday time scales, showing that the diffusion approximation, rather than the fluid limit, is the relevant approximation to use here.

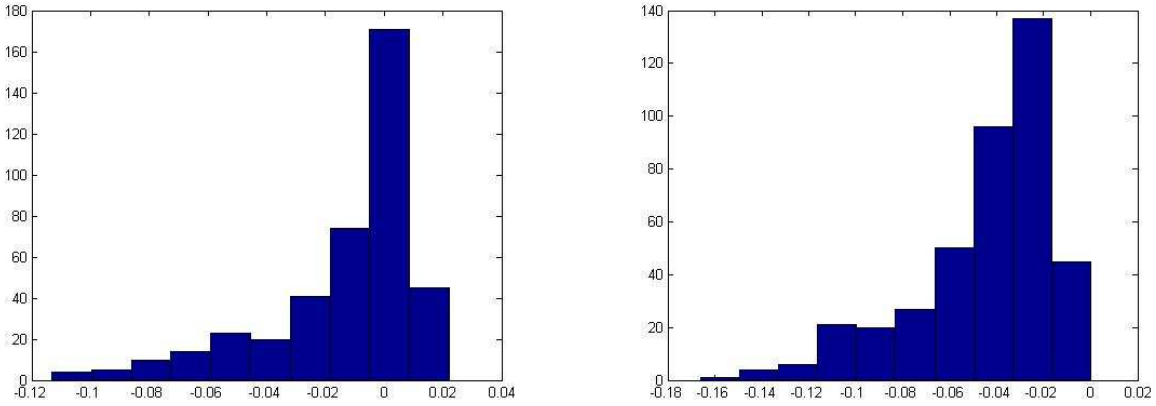

Figure 1.15: Empirical distribution of the ratio $\frac{\sqrt{n}\overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_1^{n,a})}}$ of stocks in the Dow-Jones index during June 08. Left: bid side. Right: ask side.

Indeed, the queue sizes (q_t^b, q_t^a) exhibit a diffusion-type behavior at such intraday time scales: Figure 1.16 shows the path of the cumulative order process

$$x_t = (q_0^b, q_0^a) + \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a \right) \tag{1.16}$$

sampled every second for CitiGroup stock on a typical trading day, where N_t^b (resp N_t^a) are processes counting the number of events occurring at the bid (resp. ask). For this example the average time between consecutive orders is $\lambda^{-1} \simeq 13ms \ll 1sec$ (therefore $n >> 1$). We observe that the process x behaves like a diffusion in the orthant with negative drift.

We will now show that this is a general result: under mild assumptions on the order flow process we will show that the (rescaled) for a sequence of order books $q^n = (q^{n,b}, q^{n,a})$, whose statistical properties depend on n, the queue sizes process

$$(Q_t^n)_{t\geq 0} := (\frac{q_{nt}^{n,b}}{\sqrt{n}}, \frac{q_{nt}^{n,a}}{\sqrt{n}})_{t\geq 0} \tag{1.17}$$

converges in distribution to a Markov process $(Q_t)_{t\geq 0}$ in the positive orthant, whose features will be described in terms of the statistical properties of the order flow.

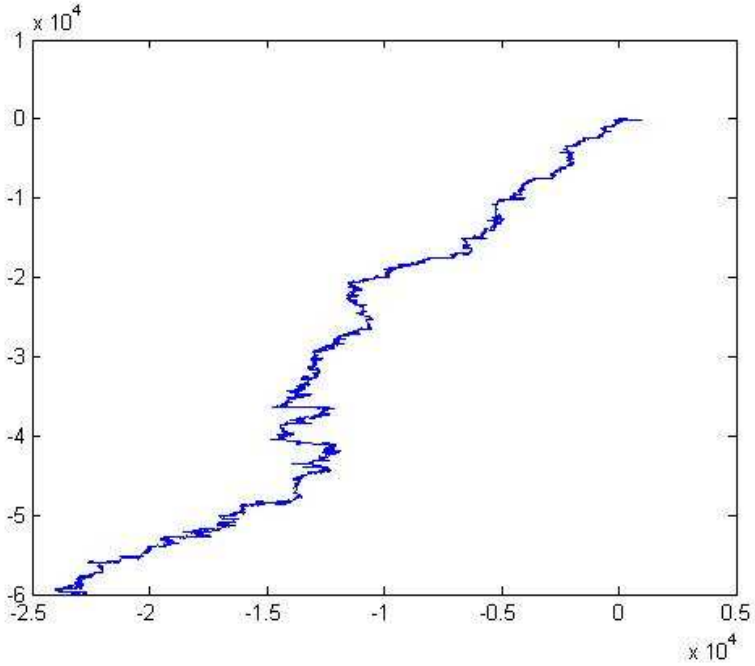

Figure 1.16: Evolution of the net order flow $x_t = (x_t^b, x_t^a)$ given in equation (1.16) for CitiGroup shares over one trading day (June 26, 2008).

1.4.3 Chapter 2: A markovian limit order book

Chapter 2 is dedicated to the study of a Markovian model where all the above quantities may be studied analytically. When orders arrive according to independent Poisson processes, and order sizes are constant, the limit order book becomes a Markov process in the orthant $\mathbb{R}^2_+$. Under these assumptions, it is possible to compute explicitly several useful quantities. We prove in proposition 2.2 that the probability of price increase conditioned on observing n shares at the ask and p shares at the bid is given by $p_1^{\text{up}}(n,p)$ where:

$$\forall (n,p) \in \mathbb{N}^2, \quad p_1^{\text{up}}(n,p) = \frac{1}{\pi}\int_0^\pi (2-\cos(t)-\sqrt{(2-\cos(t))^2-1})^p \frac{\sin(nt)\cos(\frac{t}{2})}{\sin(\frac{t}{2})}dt. \tag{1.18}$$

Despite the simplicity of the assumptions, we will see that this equation provides good fit on empirical data. In section 2.3.1 we compute the distribution τ of price durations:

$$\mathbb{P}[\tau > t | q_0^b = x,\ q_0^a = y] = \sqrt{(\frac{\mu+\theta}{\lambda})^{x+y}}\psi_{x,\lambda,\theta+\mu}(t)\psi_{y,\lambda,\theta+\mu}(t) \tag{1.19}$$

$$\text{where} \qquad \psi_{n,\lambda,\theta+\mu}(t) = \int_t^\infty \frac{n}{u} I_n(2\sqrt{\lambda(\theta+\mu)}u)e^{-u(\lambda+\theta+\mu)}du \tag{1.20}$$

and I_n is the modified Bessel function of the first kind.

We establish, in section 2.3.1, that the tail index of price durations is equal to one if the intensity of limit orders equals the intensity of market orders and cancelations. On the other hand, if market orders and cancelation dominate, the tail index of price durations equals two.

Empirically, the price has a mean-reverting behavior at high frequency. Denote $(X_1, ..., X_n)$ the sequence of price increments. In section 2.3.4 we define p_{cont} the probability to have two consecutive moves in the same direction:

$$p_{cont} = \mathbb{P}[X_2 = \delta | X_1 = \delta],$$

where δ is the tick size. We prove that the parameter p_{cont}, can be expressed with the function f:

$$p_{cont} = \sum_{i=1}^\infty \sum_{j=1}^\infty f(i,j) p_1^{\text{up}}(i,j),$$

where p_1^{up} is the function given in equation (1.18). The price dynamics follows a random walk if and only if $p_{cont} = 1/2$ which is obtained when

$$\forall (i,j) \in \mathbb{N}^2, \quad f(i,j) = f(j,i).$$

Eventually, we give a functional central limit theorem for the price process $(s_t, t \geq 0)$. When the intensity of limit orders λ equals the intensity of market orders and cancelations we show in theorem 2.1 the following functional central limit theorem:

$$\left(\frac{s_{[tn\log n]}}{\sqrt{n}},\ t \geq 0\right)_{n\geq 1} \Rightarrow \delta\sqrt{\frac{\pi\lambda}{D(f)}}(W_t, t \geq 0) \quad as \quad n \to \infty, \tag{1.21}$$

where the convergence is on the Skorokhod space $(\mathcal{D}, J_1)$. W is a standard Brownian motion and

$$\sqrt{D(f)} = \sqrt{\sum_{i=1}^\infty \sum_{j=1}^\infty ijf(i,j)} \tag{1.22}$$

is a measure of the average depth of bid and ask queues.

Equation (1.22) provides a very intuitive expression of price volatility. The variance of intraday price changes in a 'balanced' limit order market is given by the following simple relation:

$$\sigma_p^2 = \frac{\pi \delta^2 \lambda}{D(f)} \tag{1.23}$$

where δ is the 'tick size', λ is the intensity of order arrivals and $D(f)$ is a measure of market depth. These analytical results provide insights into the relation between order flow and price dynamics in order-driven markets.

Otherwise, when market orders and cancelations dominate limit orders, we prove in section 2.4.3 the functional central limit theorem:

$$\left(\frac{s_{[tn]}}{\sqrt{n}},\ t \geq 0 \right)_{n \geq 1} \overset{n \to \infty}{\Rightarrow} \frac{1}{\sqrt{\mathbb{E}[\tau]}} (W_t, t \geq 0), \tag{1.24}$$

where $\mathbb{E}[\tau]$ is the average time between two consecutive price move.

1.4.4 Chapter 3: Heavy traffic limit of the limit order book

Chapter 3 is devoted to the heavy traffic approximation of the limit order book (q_t^b, q_t^a). We prove a functional central limit theorem for the joint dynamics of the bid and ask queues when the intensity of orders becomes large, and use it to derive an analytically tractable jump-diffusion approximation for the intraday dynamics of the limit order book. Using empirical examples, we show that the assumptions behind our derivation are plausible for liquid US stocks and that the predictions of the model are validated by intraday data in such markets.

Let $q^n = (q^{n,b}, q^{n,a}, n \geq 1)$ a sequence of order books, whose statistical properties depend on n and satisfy;

Assumption 1.1. *There exist $\lambda^a > 0$ and $\lambda^b > 0$ such that*

$$\lim_{n \to \infty} \frac{T_1^{n,a} + T_2^{n,a} + ... + T_n^{n,a}}{n} = \frac{1}{\lambda^a} < \infty, \qquad \lim_{n \to \infty} \frac{T_1^{n,b} + T_2^{n,b} + ... + T_n^{n,b}}{n} = \frac{1}{\lambda^b} < \infty.$$

Assumption 1.2. *For all $n \geq 1$, the sequence $(V_i^{n,a}, V_i^{n,b})_{i \geq 1}$ is a stationary, uniformly mixing (Billingsley 1968, Ch. 4) sequence satisfying*

$$\sqrt{n}\mathbb{E}[V_1^{a,n}] \overset{n \to \infty}{\to} \overline{V^a}, \qquad \sqrt{n}\mathbb{E}[V_1^{b,n}] \overset{n \to \infty}{\to} \overline{V^b}, \qquad \text{and}$$

$$\lim_{n \to \infty} \mathbb{E}[(V_1^{n,a} - \mathbb{E}[V_1^{n,a}])^2] + 2 \sum_{i=2}^{\infty} Cov(V_1^{n,a}, V_i^{n,a}) = v_a^2 < \infty,$$

$$\lim_{n \to \infty} \mathbb{E}[(V_1^{n,b} - \mathbb{E}[V_1^{n,b}])^2] + 2 \sum_{i=2}^{\infty} Cov(V_1^{n,b}, V_i^{n,b}) = v_b^2 < \infty.$$

The following is a scaling assumption which basically states that, when grouping orders in batches of n orders, only batches whose size is $O(\sqrt{n})$ will have a non-negligible impact on the queue dynamics for large n:

Assumption 1.3. *There exist probability distributions F and $\tilde{F}$ on $\mathbb{R}_+^2$ such that*

$$\forall (x,y) \in \mathbb{R}_+^2, \quad n f_n(x\sqrt{n}, y\sqrt{n}) \overset{n \to \infty}{\Rightarrow} F(x,y) \quad \text{and} \quad n \tilde{f}_n(x\sqrt{n}, y\sqrt{n}) \overset{n \to \infty}{\Rightarrow} \tilde{F}(x,y).$$

On also assumes that the distributions F and $\tilde{F}$ have no mass on the axis.

Under assumptions 1.1, 1.2 and 1.3, we show in theorem 3.2 the following functional central theorem:

$$\left(\frac{q^n_{nt}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} (Q_t, t \geq 0) \quad \text{on} \quad (\mathcal{D}, J_1),$$

where the heavy traffic approximation (Q^b, Q^a) of the order book (q^b, q^a) is the unique Markov process on the state space $\mathbb{R}_+ \times \mathbb{R}_+$ with infinitesimal generator $\mathcal{G}$ given by:

$$\mathcal{G}h(x,y) = \lambda^a \overline{V^a} \frac{\partial h}{\partial x} + \lambda^b \overline{V^b} \frac{\partial h}{\partial y} + \frac{\lambda^a v_a^2}{2} \frac{\partial^2 h}{\partial x^2} + \frac{\lambda^b v_b^2}{2} \frac{\partial^2 h}{\partial y^2} + \rho \sqrt{\lambda^a \lambda^b} v_a v_b \frac{\partial^2 h}{\partial x \partial y}, \tag{1.25}$$

whose domain is the set $dom(\mathcal{G})$ of functions $h \in C^2(]0,\infty[\times]0,\infty[,\mathbb{R}) \cap C^0(\mathbb{R}^2_+, \mathbb{R})$ verifying the "reflection conditions"

$$h(x,0) = \int_{\mathbb{R}^2_+} h(g((x,0),(u,v)))F(du,dv), \qquad h(0,y) = \int_{\mathbb{R}^2_+} h(g((0,y),(u,v)))\tilde{F}(du,dv).$$

The function g appearing in the generator above characterizes the dependence between the state of the order book before and after a price move.

We provide in section 3.5.2 an analytical expression of this distribution τ between consecutive price moves:

$$\mathbb{P}[\tau > t | Q^b_0 = x, Q^q_0 = y] = \frac{2r_0}{\sqrt{2\pi t}} e^{-\frac{r_0^2}{4t}} \sum_{n=0}^{\infty} \frac{1}{(2n+1)} \sin \frac{(2n+1)\pi\theta_0}{\alpha} (I_{(\nu_n-1)/2}(r_0^2/4t) + I_{(\nu_n+1)/2}(r_0^2/4t)).$$

The probability of price increase $p^{up}(x,y)$, conditioned on observing x shares at the ask and y shares at the bid is computed in section 3.5.3:

$$p^{\text{up}}(x,y) = \frac{1}{2} - \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}} \frac{y-x}{y+x})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}. \tag{1.26}$$

We prove in section 3.5, the continuity of the price process as the trading intensity goes to infinity. For all $n \geq 1$, one defines $(S^n_t, t \geq 0)$ (resp. $(S_t, t \geq 0)$) the piecewise constant stochastic process which increases by one tick when the process Q^n (resp. Q) is replaced according to R^n_k for some k (resp. R_k) and decreases by one tick when Q^n (resp. Q) replaced according to $\tilde{R}^n_k$ for some k (resp. $\tilde{R}_k$). We will see that

$$S^n \Rightarrow S \quad \text{on} \quad (\mathcal{D}, J_1).$$

1.4.5 Chapter 4: Linking volatility and order flow

This chapter is devoted to the understanding of the price process $S = (S_t, t \geq 0)$. Through the analytical tractability of the Markov process (Q^b, Q^a), our model allows to obtain analytical expressions for various quantities of interest such as an explicit expression of the autocorrelation of consecutive price increments, the low-frequency dynamics of the price and the expression of the drift and the volatility of the price. In this chapter we extend the relation low frequency price volatility and microstructure parameters to a general order flow.

We show in section 4.2.1 that the sequence of consecutive price increments $(X_1, X_2, ..., X_n)$ of the price process $(S_t, t \geq 0)$ follows a homogenous Markov chain with parameters p^+ and p^-:

$$p^+ = \mathbb{P}[X_2 = \delta | X_1 = \delta] \text{ and } p^- = \mathbb{P}[X_2 = -\delta | X_1 = -\delta],$$

which are linked, in theorem 4.1, to the autocorrelation between consecutive price increments:

$$\mathrm{Corr}(X_1, X_2) = p_+ + p_- - 1. \tag{1.27}$$

In the same theorem 4.1, it is also proven that for all $n \geq 2$, the probability $p_n^{\mathrm{up}}(x, y)$ that the n-th price move is an increase, conditioned on observing x-shares at the bid and y shares at the ask, may be expressed with the parameters p^+, p^- and $p_1^{\mathrm{up}}(x, y)$ by:

$$\forall (x, y) \in \mathbb{R}_+^2, \quad p_n^{\mathrm{up}}(x, y) = \frac{1 - p-}{2 - p^+ - p^-}(1 - (p^+ + p^- - 1)^{n-1}) + (p^+ + p^- - 1)^{n-1} p_1^{\mathrm{up}}(x, y).$$

We propose in section 4.2.2 a parametric model for the density f (resp $\tilde{f}$) of the distribution F (resp. $\tilde{F}$), of order book size (Q^b, Q^a) after a price increase (resp. decrease):

$$f(r, \theta) = c^2 e^{-cr} \alpha (\frac{2}{\pi})^\alpha \theta^{\alpha - 1} \quad \textit{and} \quad \tilde{f}(r, \theta) = \tilde{c}^2 e^{-\tilde{c}r} \tilde{\alpha} (\frac{2}{\pi})^{\tilde{\alpha}} \theta^{\tilde{\alpha} - 1}, \tag{1.28}$$

where α (resp. $\tilde{\alpha}$ are parameters characterizing the bid-ask imbalance after a price increase (resp. decrease), and $1/c$ (resp $1/\tilde{c}$) measures the average *depth* of the order book after a price increase (resp. decrease). The parameter α (resp. $\tilde{\alpha}$), characterizing the skewness of the angular part of the distribution F (resp. $\tilde{F}$), are linked, in section 4.2.3, to the probabilities of 'continuation' p^+ and p^-.

For a general order flow, it is proven in theorems 4.2, 4.3 and 4.4 that over time scales γ_2 much larger than the interval between order book events, prices have diffusive behaviors and are modeled as such. At low frequencies, the price behaves like a Brownian motion with drift.

When the order flow is symmetric at the bid/ask, $p^+ = p^- := p_{cont}$, we prove in theorem 4.2 the following functional central limit theorem for the price process:

$$\left(\frac{s_{[nt]}}{\sqrt{n}},\ t \geq 0 \right) \Rightarrow \frac{1}{\sqrt{r_1(F)}} \sqrt{\frac{p_{cont}}{1 - p_{cont}}} \tag{1.29}$$

where $r_1(F)$ is the average time between two consecutive price move.

On the other hand, when the order flow is not symmetric at the bid/ask, $p^+ \neq p^-$ we show in theorem 4.3 that the price can have both a drift and a volatility. Let γ_0 be the average durations between two consecutive orders and $\gamma_2 >> \gamma_0$ a *macroscopic time scale* (e.g. daily time scale). At the time scale γ_2 we prove in theorem 4.3 and 4.4 that the price process $(P_t, t \geq 0)$ behaves like a Brownian motion with drift:

$$P_t = \frac{\gamma_2}{\gamma_0} \delta t d_p + \sqrt{\frac{\gamma_2}{\gamma_0}} \delta \sigma_p B_t,$$

where δ is the tick size, $(B_t, t \geq 0)$ is a standard Brownian motion, the drift of the price d_p is:

$$d_p = \frac{\dfrac{1}{1 - p^+} - \dfrac{1}{1 - p^-}}{(1 + \dfrac{p^+}{1 - p^+}) r_1(F) + (1 + \dfrac{p^-}{1 - p^-}) r_1(\tilde{F})}. \tag{1.30}$$

The price volatility σ_p is:

$$\sigma_p^2 = \frac{\dfrac{p^+(1 + p^+)}{(1 - p^+)^2} + \dfrac{p^-(1 + p^-)}{(1 - p^-)^2} - 2 \dfrac{p^+}{1 - p^-} \dfrac{p^-}{1 - p^-}}{(1 + \dfrac{p^+}{1 - p^+}) r_1(F) + (1 + \dfrac{p^-}{1 - p^-}) r_1(\tilde{F})}. \tag{1.31}$$

On equations (4.8) and (4.9), $r_1(F)$ (resp. $r_1(\tilde{F})$) is the average durations until the price moves after a price increase (resp. price decrease).

Eventually, in section 4.4, we link these parameters $p^+, p^-, r_1(F)$ and $r_1(\tilde{F})$ with the order flow and derive expressions of price volatility as a function of order flow statistics for several examples

The parameters of the the heavy traffic approximation of the order book (q^b, q^a) depend on properties of the order flow such as the symmetry between the bid and at the ask, the correlation between bid and ask queue sizes or the average order size. When the average order size is of order of magnitude $O(1/\sqrt{n})$, the diffusive limit of the order book is the proper rescaling whereas when the average order size is of order of magnitude $O(1)$ -for instance when marker orders and cancelations dominate limit orders- the heavy traffic limit of the order book is the fluid limit.

In section 4.4 we study several regimes of the order book and we link the price volatility with parameters of the order flow for all these regimes. Table 1.5 points to the sections where these regimes are studied.

Asymptotic regime of (q^b, q^a)	Bid/Ask symmetry	Bid/Ask correlation	Section
Fluid limit	symmetry	no correlation	Section 4.4.1
Diffusive limit	symmetry	no correlation	section 4.4.2
Diffusive limit	symmetry	negative correlation	section 4.4.2
Fluid limit	asymmetry		section 4.4.3

Table 1.5: Different regimes of order book and the sections where these order books are studied.

For instance, when the drift of the order book dominates over its volatility, under the assumption that the order flows are symmetric at the bid and at the ask and when F (resp. $\tilde{F}$) are given by equation (4.6), price volatility becomes:

$$\sigma_p^2 = \delta^2 \underbrace{\frac{(\pi/2)^\alpha}{(2^\alpha - 1)\int_0^{\pi/4} \sin(\theta)\theta^{\alpha-1}d\theta + \int_{\pi/4}^{\pi/2} \cos(\theta)\theta^{\alpha-1}d\theta}}_{\text{Bid-Ask asymetry}} \overbrace{\frac{c}{2}}^{\text{Order book depth}} \underbrace{(\mu - \lambda)}_{\text{Trading intensity}},$$

where δ is the tick size and α, c, λ, μ are parameters of the order flow.

On the other hand, when the order book dynamics (Q^b, Q^a) is a symmetric driftless Brownian motion, and when the density f of the distribution F follows a polar decomposition $f(r, \theta) = h(r)g(\theta)$, one can decompose price volatility as:

$$\sigma_p^2 = \underbrace{\delta^2}_{\text{tick}} \underbrace{\frac{\overbrace{\lambda^a v_a^2}^{\text{variance of order sizes/unit time}}}{\int_0^\infty h(r)r^3 dr \int_0^{\pi/2} g(\theta)\Gamma(\theta)d\theta}}_{\text{Order book depth}} \underbrace{\frac{p_{cont}}{1 - p_{cont}}}_{\text{mean reversion}}, \tag{1.32}$$

where δ is the tick size, λ^a the trading intensity, v_a^2 the variance of order sizes, p_{cont} the continuation probability ($p_{cont} = p^+ = p^-$ for this specific symmetric order book), ρ is the correlation between bid and ask queue dynamics, and $\Gamma(\rho, \theta)$ a function of bid-ask imbalance θ and the parameter ρ.

Equation (1.32) generalizes formula (1.23) and links low-frequency price volatility with parameters of the order flow for a general order book. Several components affect price volatility:

- The *variance of order sizes* $v_a^2 \lambda^a$
- The *skewness* of the order book after a price move through the parameter p_{cont}.
- The correlation between the order flow arriving at the ask and the order flow arriving at the bid.
- The *depth* of the order book $\int_0^\infty h(r) r^3 dr$.

Formula (1.32) gives some insights on the factors that influence price volatility. For instance, when the intensity of *all orders coming in the limit order book* is multiplied by the same factor x, the intensity of orders becomes $\lambda^a x$, the *limit order book depth* is multiplied by a factor x^2, and all other parameters are unchanged. So price volatility σ_p decreases by a factor $\sqrt{\frac{1}{x}}$.

Interestingly, Rosu (2009) shows the same dependence in $1/\sqrt{x}$ of price volatility using an equilibrium approach. We show through equation (1.32), that this relation between order arrival intensity and price volatility holds under much more general assumptions, and may be derived without behavioral assumptions for market participants.

Chapter 2

Price dynamics in a Markovian limit order market

2.1 Introduction

An increasing number of stocks are traded in electronic, order-driven markets, in which orders to buy and sell are centralized in a *limit order book* available to market participants and market orders are executed against the best available offers in the limit order book. The dynamics of prices in such markets are not only interesting from the viewpoint of market participants –for trading and order execution (Alfonsi et al. (2010), Predoiu et al. (2011))– but also from a fundamental perspective, since they provide a rare glimpse into the dynamics of supply and demand and their role in price formation (Cont 2011).

Equilibrium models of price formation in limit order markets (Parlour (1998), Rosu (2009)) have shown that the evolution of the price in such markets is rather complex and depends on the state of the order book. On the other hand, empirical studies on limit order books (Bouchaud et al. (2008), Farmer et al. (2004), Gourieroux et al. (1999), Hollifield et al. (2004), Smith et al. (2003)) provide an extensive list of statistical features of order book dynamics that are challenging to incorporate in a single model. While most of these studies have focused on unconditional/steady–state distributions of various features of the order book, empirical studies (Harris and Panchapagesan (2005), Cont et al. (2010a)) show that the state of the order book contains information on short-term price movements so it is of interest to provide forecasts of various quantities *conditional* on the state of the order book. Providing analytically tractable models which enable to compute and/or reproduce *conditional* quantities which are relevant for trading and intraday risk management has proven to be challenging, given the complex relation between order book dynamics and price behavior.

The search for tractable models of limit order markets has led to the development of stochastic models which aim to retain the main statistical features of limit order books while remaining computationally manageable. Stochastic models also serve to illustrate how far one can go in reproducing the dynamic properties of a limit order book without resorting to detailed behavioral assumptions about market participants or introducing unobservable parameters describing agent preferences, as in more detailed market microstructure models.

Starting from a description of order arrivals and cancelations as point processes, the dynamics of a limit order book is naturally described in the language of queueing theory (Cont 2011). Engle and Lunde (2003) model trade and quote arrivals as a bivariate point process. Cont et al. (2010b) model the dynamics of a limit order book as a tractable multiclass queueing system and compute various transition probabilities of the price conditional on the state of the order book, using Laplace transform methods.

2.1.1 Summary

We propose a Markovian model of a limit order market, which captures some salient features of the dynamics of market orders and limit orders and their influence on price dynamics, yet is even simpler than the model of Cont et al. (2010b) and enables a wide range of properties of the price process to be computed analytically.

Our approach is motivated by the observation that, if one is primarily interested in the dynamics of the price, it is sufficient to focus on the dynamics of the (best) bid and ask queues. Indeed, empirical evidence shows that most of the order flow is directed at the best bid and ask prices (Biais et al. (1995)) and the imbalance between the order flow at the bid and at the ask appears to be the main driver of price changes (Cont et al. (2010a)).

Motivated by this remark, we propose a parsimonious model in which the limit order book is represented by the number of limit orders (q_t^b, q_t^a) sitting at the bid and the ask, represented as a system of two interacting queues. The remaining levels of the order book are treated as a 'reservoir' of limit orders represented by the distribution of the size of the queues at the 'next-to-best' price levels.

Through its analytical tractability, the Markovian version of our model allows to obtain analytical expressions for various quantities of interest such as the distribution of the duration until the next price change, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, *conditional* on the state of the order book. Compared with econometric models of high frequency data Engle and Russell (1998), Engle and Lunde (2003) where the link between durations and price changes is specified exogenously, our model links these quantities in an endogenous manner, and provides a first step towards joint 'structural' modeling of high frequency dynamics of prices and order flow.

A second important observation is that order arrivals and cancelations are very frequent and occur at millisecond time scale, whereas, in many applications such as order execution, the metric of success is the volume-weighted average price (VWAP) so one is interested in the dynamics of order flow over a large time scale, typically tens of seconds or minutes. As shown in Table 3.2, thousands of order book events may occur over such time scales. This observation enables us to use asymptotic methods to study the link between price volatility and order flow in this model by studying the *diffusion limit* of the price process. In particular, we prove a functional central limit theorem for the price process and express the volatility of price changes in terms of parameters describing the arrival rates of buy and sell orders and cancelations. For example, we show (Theorem 2.1) that the variance of intraday price changes in a 'balanced' limit order market is given by the following simple relation:

$$\sigma^2 = \frac{\pi \delta^2 \lambda}{D(f)}$$

where δ is the 'tick size', λ is the intensity of order arrivals and $D(f)$ is a measure of market depth. These analytical results provide insights into the relation between order flow and price dynamics in order-driven markets. Comparison of these results with empirical data validates the main insights of the model.

	Average no. of orders in 10s	Price changes in 1 day
Citigroup	4469	12499
General Electric	2356	7862
General Motors	1275	9016

Table 2.1: Average number of orders in 10 seconds and number of price changes (June 26th, 2008).

2.1.2 Outline

The chapter is organized as follows. Section 2.2 introduces a reduced-form representation of a limit order book and presents a Markovian model in which limit orders, market orders and cancelations occur according to Poisson processes. Section 2.3 presents various analytical results for this model: we compute the distribution of the duration until the next price change (section 2.3.1), the probability of upward move in the price (section 2.3.2) and the dynamics of the price (section 2.3.4). In Section 2.4, we show that the price exhibits diffusive behavior at longer time scales and express the variance of price changes in terms of the parameters describing the order flow, thus establishing a link between volatility and order flow statistics.

2.2 A Markov model of limit order book dynamics

2.2.1 A stylized representation of a limit order book

Empirical studies of limit order markets suggest that the major component of the order flow occurs at the (best) bid and ask price levels (see e.g. Biais et al. (1995)). All electronic trading venues also allow to place limit orders *pegged* to the best available price (National Best Bid Offer, or NBBO); market makers used these *pegged* orders to liquidate their inventories. Furthermore, studies on the price impact of order book events show that the net effect of orders on the bid and ask queue sizes is the main factor driving price variations (Cont et al. (2010a)). These observations, together with the fact that queue sizes at the best bid and ask of the consolidated order book are more easily obtainable (from records on trades and quotes) than information on deeper levels of the order book, motivate a reduced-form modeling approach in which we represent the state of the limit order book by

- the bid price s_t^b and the ask price s_t^a
- the size of the bid queue q_t^b representing the outstanding limit buy orders at the bid, and
- the size of the ask queue q_t^a representing the outstanding limit sell orders at the ask

Figure 1 summarizes this representation.

If the stock is traded in several venues, the quantities q^b and q^a represent the best bids and offers in the *consolidated* order book, obtained by aggregating over all (visible) trading venues. At every time t, q_t^b (resp. q_t^a) corresponds to all visible orders available at the bid price s_t^b (resp. s_t^a) across all exchanges.

The state of the order book is modified by *order book events*: limit orders (at the bid or ask), market orders and cancelations (see Cont et al. (2010b,a), Smith et al. (2003)). A limit buy (resp. sell) order of size x increases the size of the bid (resp. ask) queue by x, while a market buy (resp. sell) order decreases the corresponding queue size by x. Cancelation of x orders in a given queue reduces the queue size by x. Given that we are interested in the queue sizes at the best bid/ask levels, market orders and cancelations have the same effect on the queue sizes (q_t^b, q_t^a).

The bid and ask prices are multiples of the tick size δ. When either the bid or ask queue is depleted by market orders and cancelations, the price moves up or down to the next level of the order book. The price processes s_t^b, s_t^a is thus a piecewise constant process whose transitions correspond to hitting times of the axes $\{(0,y), y \in \mathbb{N}\} \cup \{(x,0), x \in \mathbb{N}\}$ by the process $q_t = (q_t^b, q_t^a)$.

If the order book contains no 'gaps' (empty levels), these price increments are equal to one tick:

- when the bid queue is depleted, the (bid) price decreases by one tick.
- when the ask queue is depleted, the (ask) price increases by one tick.

If there are gaps in the order book, this results in 'jumps' (i.e. variations of more than one tick) in the price dynamics. We will ignore this feature in what follows but it is not hard to generalize our results to include it.

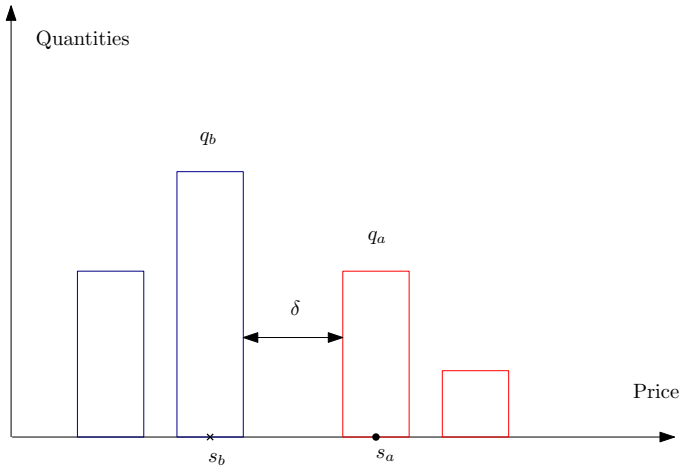

Figure 2.1: Simplified representation of a limit order book.

The quantity $s_t^a - s_t^b$ is the *bid-ask spread*, which may be one or several ticks. As shown in Table 2.2.1, for liquid stocks the bid-ask spread is equal to one tick for more than 98% of observations.

Bid-ask spread	1 tick	2 tick	$\geq$ 3 tick
Citigroup	98.82	1.18	0
General Electric	98.80	1.18	0.02
General Motors	98.71	1.15	0.14

Table 2.2: Percentage of observation with a given bid-ask spread (June 26th, 2008).

When either the bid or ask queue is depleted, the bid-ask spread widens immediately to more than one tick. Once the spread has increased, a flow of limit sell (resp. buy) orders quickly fills the gap and the spread reduces again to one tick. When a limit order is placed inside the spread, all the limit orders pegged to the NBBO price move in less than a millisecond to the price level corresponding to this new order. Once this happens, both the bid price and the ask price have increased (resp. decreased) by one tick.

The histograms in Figure 2.2 show that this 'closing' of the spread takes place very quickly: as shown in Figure 2.2 (left) the lifetime of a spread larger than one tick is of the order of a couple of milliseconds, which is negligible compared to the lifetime of a spread equal to one tick (Figure 2.2, right). In our model we assume that the second step occurs *infinitely fast*: once the bid-ask spread widens, it returns immediately to one tick. For the example of Dow Jones stocks (Figure 2.2), this is a reasonable assumption since the widening of the spread lasts only a few milliseconds. This simply means that we are not trying to describe/model how the orders flow inside the bid-ask spread at the millisecond time scale and, when we describe the state of the order book *after* a price change we have in mind the state of the order book *after* the bid-ask spread has returned to one tick.

Under this assumption, each time *one* of the queues is depleted, *both* the bid queue and the ask queues move to a new position and the bid-ask spread remains equal to one tick after the price change. Thus, under our assumptions the bid-ask spread is equal to one tick, i.e. $s_t^a = s_t^b + \delta$, resulting in a further reduction of dimension in the model.

Once either the bid or the ask queue are depleted, the bid and ask queues assume new values. Instead of keeping track of arrival, cancelation and execution of orders at *all* price levels (as in Cont et al. (2010b), Smith et al. (2003)), we treat the queue sizes after a price change as stationary variables

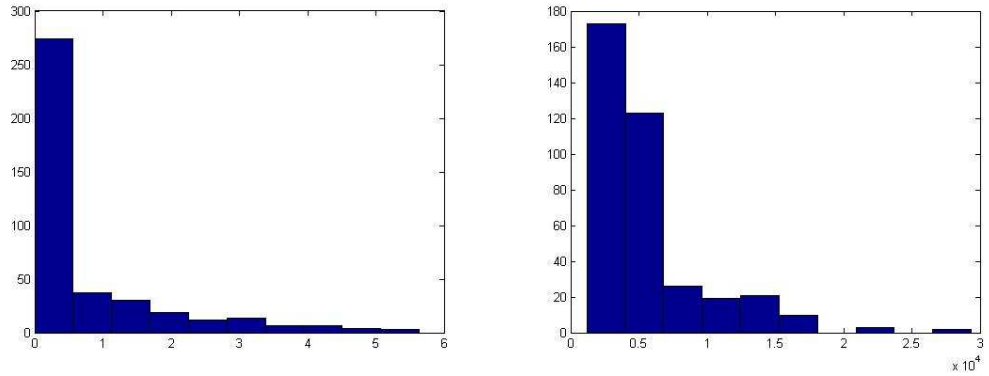

Figure 2.2: Left: Histogram of average lifetime, in milliseconds of a spread larger than one tick for Dow Jones stocks in June 2008. Right: Histogram of average lifetime, in milliseconds of a spread equal to one tick.

drawn from a certain distribution f on $\mathbb{N}^2$ which represents, in a statistical sense, the *depth* of the order book after a price change: $f(x,y)$ represents the probability of observing $(q_t^b, q_t^a) = (x,y)$ right after a price increase. Similarly, we denote $\tilde{f}(x,y)$ the probability of observing $(q_t^b, q_t^a) = (x,y)$ right after a price decrease. More precisely, denoting by $\mathcal{F}_t$ the history of prices and order book events on $[0,t]$,

- if $q_{t-}^a = 0$ then (q_t^b, q_t^a) is a random variable with distribution f, independent from $\mathcal{F}_{t-}$.
- if $q_{t-}^b = 0$ then (q_t^b, q_t^a) is a random variable with distribution $\tilde{f}$, independent from $\mathcal{F}_{t-}$.

The distributions f and $\tilde{f}$ summarize the interaction of the queues at the best bid/ask levels with the rest of the order book, viewed here as a 'reservoir' of limit orders. Figure 2.3 shows the (joint) empirical distribution of bid and ask queue sizes after a price move for Citigroup stock on June 26th 2008.

In summary, state of the limit order book is thus described by a continuous-time process $X_t = (s_t^b, q_t^b, q_t^a)$ which takes values in the discrete state space $\delta\mathbb{Z} \times \mathbb{N}^2$, with piecewise constant sample paths whose transitions correspond to the order book events. Denote by $(T_i^a, i \geq 1)$ (resp. T_i^b) the durations between two consecutive orders arriving at the ask (resp. the bid) and V_i^a (resp. V_i^b) the size of the associated change in queue size: $V_i^a > 0$ for a limit order at the ask (resp. $V_i^b > 0$ for a limit order at the bid), while market orders and cancellations correspond to negative values (and a decrease in queue size) of V_i^a (resp. V_i^b). The above assumptions translate into the following dynamics for $X_t = (s_t^b, q_t^b, q_t^a)$:

- If an order or cancelation arrives on the ask side at time T:

$$(s_T^b, q_T^b, q_T^a) = (s_{T-}^b, q_{T-}^b, q_{T-}^a + V_i^a)1_{\{q_{T-}^a + V_i^a > 0\}} + (s_{T-}^b + \delta, R_i^b, R_i^a)1_{\{q_{T-}^a + V_i^a \leq 0\}},$$

- If an order or cancelation arrives on the bid side i.e. $T \in \{T_i^b, i \geq 1\}$:

$$(s_T^b, q_T^b, q_T^a) = (s_{T-}^b, q_{T-}^b + V_i^b, q_{T-}^a)1_{\{q_{T-}^b + V_i^b > 0\}} + (s_{T-}^b - \delta, \tilde{R}_i^b, \tilde{R}_i^a)1_{\{q_{T-}^b + V_i^b \leq 0\}},$$

  and $(R_i)_{i\geq 1} = (R_i^b, R_i^a)_{i\geq 1}$ is a sequence of IID variables with (joint) distribution f, and $(\tilde{R}_i)_{i\geq 1} = (\tilde{R}_i^b, \tilde{R}_i^a)_{i\geq 1}$ is a sequence of IID variables with (joint) distribution $\tilde{f}$.

2.2.2 A Markovian model for order book dynamics

To give a complete statistical description of the dynamics of the limit order book, we need to describe the distributional properties of the sequences $T_i^a, T_i^b, V_i^a, V_i^b$ describing the timing and size of order book events.

We assume that these events occur according to independent Poisson processes:

- Market buy (resp. sell) orders arrive at independent, exponential times with rate μ,
- Limit buy (resp. sell) orders at the (best) bid (resp. ask) arrive at independent, exponential times with rate λ,
- Cancelations occur at independent, exponential times with rate θ.
- These events are mutually independent.
- Orders are equal in size (assumed to be 1 without loss of generality).

Denoting by $(T_i^a, i \geq 1)$ (resp. T_i^b) the durations between two consecutive queue changes at the ask (resp. the bid) and V_i^a (resp. V_i^b) the size of the associated change in queue size, the above assumptions translate into the following properties for the sequences $T_i^a, T_i^b, V_i^a, V_i^b$:

(i) $(T_i^a)_{i\geq 0}$ is a sequence of independent random variables with exponential distribution with parameter $\lambda + \theta + \mu$,

(ii) $(T_i^b)_{i\geq 0}$ is a sequence of independent random variables with exponential distribution with parameter $\lambda + \theta + \mu$,

(iii) $(V_i^a)_{i\geq 0}$ is a sequence of independent random variables with

$$\mathbb{P}[V_i^a = 1] = \frac{\lambda}{\lambda + \mu + \theta} \quad \textit{and} \quad \mathbb{P}[V_i^a = -1] = \frac{\mu + \theta}{\lambda + \mu + \theta}, \tag{2.1}$$

(iv) $(V_i^b)_{i\geq 0}$ is a sequence of independent random variables with

$$\mathbb{P}[V_i^b = 1] = \frac{\lambda}{\lambda + \mu + \theta} \quad \textit{and} \quad \mathbb{P}[V_i^b = -1] = \frac{\mu + \theta}{\lambda + \mu + \theta} \tag{2.2}$$

- These sequences are independent.

These assumptions constitute a simplification of the actual statistical properties of the order flow, which can include dependence in durations, non-exponential durations and heterogeneous order sizes (Engle and Russell 1998, Engle and Lunde 2003, Bouchaud et al. 2008, Cont 2011). However, they allow to account for features such as the rate of arrival of limit and market orders (and cancellations), random fluctuations in order book depth and the endogenous nature of price dynamics in a simple way which, as we shall illustrate, allows to obtain some analytical insights.

Under these assumptions $q_t = (q_t^b, q_t^a)$ is a Markov process, taking values in the orthant $\mathbb{N}^2$, whose transitions occur at the order book events $\{T_i^a, i \geq 1\} \cup \{T_i^b, i \geq 1\}$:

- At the arrival of a new limit buy (resp. sell) order the bid (resp. ask) queue increases by one unit. This occurs at rate λ.
- At each cancelation or market order, which occurs at rate $\theta + \mu$, either:
  (a) the corresponding queue decreases by one unit if it is > 1, or
  (b) if the ask queue is depleted then q_t is a random variable with distribution f.
  (c) if the bid queue is depleted then q_t is a random variable with distribution $\tilde{f}$.

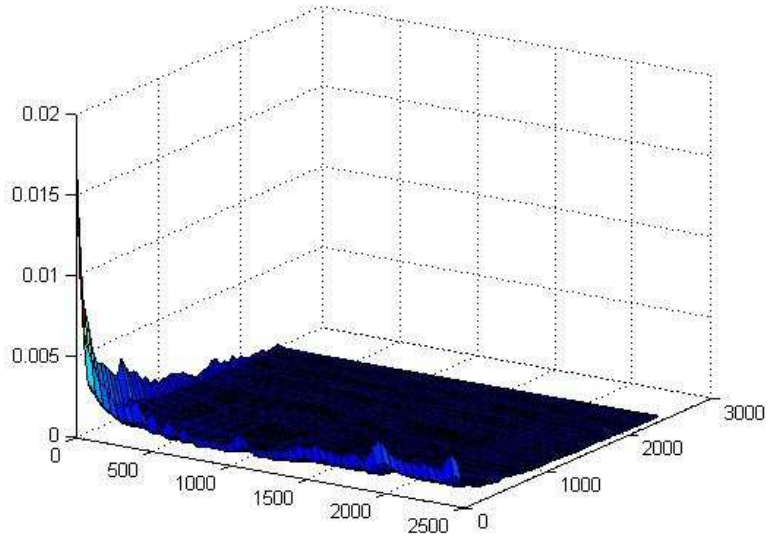

Figure 2.3: Joint density f of bid and ask queue sizes after a price increase (Citigroup, June 26th 2008).

The values of λ and $\mu + \theta$ are readily estimated from high-frequency records of trades and quotes (see Cont et al. (2010b) for a description of the estimation procedure). Table 2.3 gives examples of such parameter estimates for the stocks mentioned above. We note that in all cases $\lambda < \mu + \theta$ but that the difference is small: $|(\mu + \theta) - \lambda| \ll \lambda$.

	$\hat{\lambda}$	$\hat{\mu} + \hat{\theta}$	$\frac{\hat{\mu}+\hat{\theta}-\hat{\lambda}}{\hat{\lambda}}$
Citigroup	2204	2331	0.0576
General Electric	317	325	0.0252
General Motors	102	104	0.0196

Table 2.3: Estimates for the intensity of limit orders and market orders+cancellations, in number of batches per second (each batch representing 100 shares) on June 26th, 2008).

Remark 2.1 (Independence assumptions)**.** *The IID assumption for the sequences* $(R_n), (\tilde{R}_n)$ *is only used in Section 2.4. The results of Section 2.3 do not depend on this assumption.*

2.2.3 Quantities of interest

In applications, one is interested in computing various quantities that intervene in high frequency trading such as:

- the conditional distribution of the duration between price moves, given the state of the order book (Section 2.3.1),
- the probability of a price increase, given the state of the order book (Section 2.3.2),

- the dynamics of the price autocorrelations and distribution and autocorrelations of price changes (section 2.3.4), and
- the volatility of the price (section 2.4).

We will show that all these quantities may be characterized analytically in this model, in terms of order flow statistics.

2.3 Analytical results

The high-frequency dynamics of the price may be described in terms of *durations* between successive price changes and the magnitude of these price changes. It is of interest to examine what information the current state of the (consolidated) order book gives about the dynamics of the price. We now proceed to show how the model presented above may be used to compute the conditional distributions of durations and price changes, given the current state of the order book, in terms of the arrival rates of market orders, limit orders and cancelations. The results of this section do not depend on the assumptions on the sequences $(R_n), (\tilde{R}_n)$.

2.3.1 Duration until the next price change

We consider first the distribution of the duration until the next price change, starting from a given configuration (x, y) of the order book. We define

- τ_a the first time when the ask queue $(q^a_t, t \geq 0)$ is depleted,
- τ_b the first time when the bid queue $(q^b_t, t \geq 0)$ is depleted

Since the queue sizes are constant between events, one can express these stopping times as:

$$\tau_a = \inf\{T^a_1 + ... + T^a_i,\ q^a_{T^a_1+...+T^a_i-} + V^a_i = 0\} \qquad \tau_b = \inf\{T^b_1 + ... + T^b_i,\ q^b_{T^b_1+...+T^b_i-} + V^b_i = 0\}$$

The price $(s_t, t \geq 0)$ moves when the queue $q_t = (q^b_t, q^a_t)$ hits one of the axes: the duration until the next price move is thus

$$\tau = \tau_a \wedge \tau_b.$$

The following theorem gives the distribution of the duration τ, conditional on the initial queue sizes:

Proposition 2.1 (Distribution of duration until next price move)**.** *The distribution of τ conditioned on the initial state of the order book is given by:*

$$\mathbb{P}[\tau > t | q^b_0 = x,\ q^a_0 = y] = \sqrt{\left(\frac{\mu+\theta}{\lambda}\right)^{x+y}}\quad \psi_{x,\lambda,\theta+\mu}(t) \quad \psi_{y,\lambda,\theta+\mu}(t) \tag{2.3}$$

$$\text{where} \qquad \psi_{n,\lambda,\theta+\mu}(t) = \int_t^\infty \frac{n}{u} I_n(2\sqrt{\lambda(\theta+\mu)}u) e^{-u(\lambda+\theta+\mu)} du \tag{2.4}$$

and I_n is the modified Bessel function of the first kind. The conditional law of τ has a regularly varying tail

- *with tail exponent 2 if $\lambda < \mu + \theta$*
- *with tail exponent 1 if $\lambda = \mu + \theta$. In particular, if $\lambda = \mu + \theta$, $E[\tau | q^b_0 = x, q^a_0 = y] = \infty$ whenever $x > 0, y > 0$.*

Proof. Since $(q_t^a, t \geq 0)$ follows a birth and death process with birth rate λ and death rate $\mu + \theta$, $\mathcal{L}(s,x) := \mathbb{E}[e^{-s\tau_a}|q_0^a = x]$ satisfies:

$$\forall s > 0, \quad \mathcal{L}(s,x) = \frac{\lambda \mathcal{L}(s,x+1) + (\mu+\theta)\mathcal{L}(s,x-1)}{\lambda + \mu + \theta + s}.$$

The polynomial $P(X) = \lambda X^2 - (\lambda + \mu + \theta + s)X + \mu + \theta$ has two real roots; since $P(1) = -s < 0$, one root is > 1, the other is < 1; since $\mathcal{L}(s,0) = 1$ and $\lim_{x\to\infty} \mathcal{L}(s,x) = 0$,

$$\mathcal{L}(s,x) = \left(\frac{(\lambda+\mu+\theta+s) - \sqrt{((\lambda+\mu+\theta+s))^2 - 4\lambda(\mu+\theta)}}{2\lambda}\right)^x.$$

The symmetry assumption implies that $\mathbb{P}[\tau_b > t|q_0^b = x]$ is given by a similar expression. Using independence, $\mathbb{P}[\tau > t|q_0^b = x, q_0^a = y] = \mathbb{P}[\tau_b > t|q_0^b = x]\mathbb{P}[\tau_a > t|q_0^a = y]$,

$$\mathbb{P}[\tau > t|q_0^b = x, q_0^a = y] = \int_t^\infty \hat{\mathcal{L}}(u,x)du \int_t^\infty \hat{\mathcal{L}}(u,y)du.$$

This Laplace transform may be inverted (Feller 1971, XIV.7) and yields

$$\hat{\mathcal{L}}(t,x) = \frac{x}{t}\sqrt{(\frac{\mu+\theta}{\lambda})^x} \quad I_x(2\sqrt{\lambda(\theta+\mu)}t)e^{-t(\lambda+\theta+\mu)},$$

which gives us the expected result. This allows in particular to study the *tail* behavior of the conditional distribution of τ:

- If $\lambda < \mu + \theta$,

  $$\mathcal{L}(s,x) = \alpha(s)^x \underset{s\to 0}{\sim} 1 - \frac{x(\lambda+\mu+\theta)}{2\lambda(\mu+\theta-\lambda)}s$$

  so Karamata's Tauberian theorem (Feller 1971, XIII.5) yields

  $$\mathbb{P}[\tau_a > t|q_0^a = x] \underset{t\to\infty}{\sim} \frac{x(\lambda+\mu+\theta)}{2\lambda(\mu+\theta-\lambda)}\frac{1}{t}.$$

  The conditional law of the duration τ given $q_0 = (x,y)$ is thus regularly varying with tail index 2:

  $$\mathbb{P}[\tau > t|q_0^b = x, q_0^a = y] \underset{t\to\infty}{\sim} \frac{xy(\lambda+\mu+\theta)^2}{\lambda^2(\mu+\theta-\lambda)^2}\frac{1}{4t^2}. \tag{2.5}$$

- If the order flow is balanced i.e. $\lambda = \mu + \theta$ then

  $$\mathcal{L}(s,x) = \alpha(s)^x \underset{s\to 0}{\sim} 1 - \frac{x}{\sqrt{\lambda}}\sqrt{s},$$

  the law of τ_a is regularly-varying with tail index $1/2$ and

  $$\mathbb{P}[\tau_a > t|q_0^a = x] \underset{t\to\infty}{\sim} \frac{x}{\sqrt{\pi\lambda}}\frac{1}{\sqrt{t}}.$$

  The duration then follows a heavy-tailed distribution with infinite first moment:

  $$\mathbb{P}[\tau > t|q_0^b = x, q_0^a = y] \underset{t\to\infty}{\sim} \frac{xy}{\pi\lambda}\frac{1}{t}; \tag{2.6}$$

□

The expression given in (2.3) is easily computed by discretizing the integral in (2.4). Plotting (2.3) for a fine grid of values of t typically takes less than a second on a laptop. Figure 3 gives a numerical example, with $\lambda = 12$ sec^{-1}, $\mu + \theta = 13$ sec^{-1}, $q_0^a = 4, q_0^b = 5$ (queue sizes are given in multiples of average batch size).

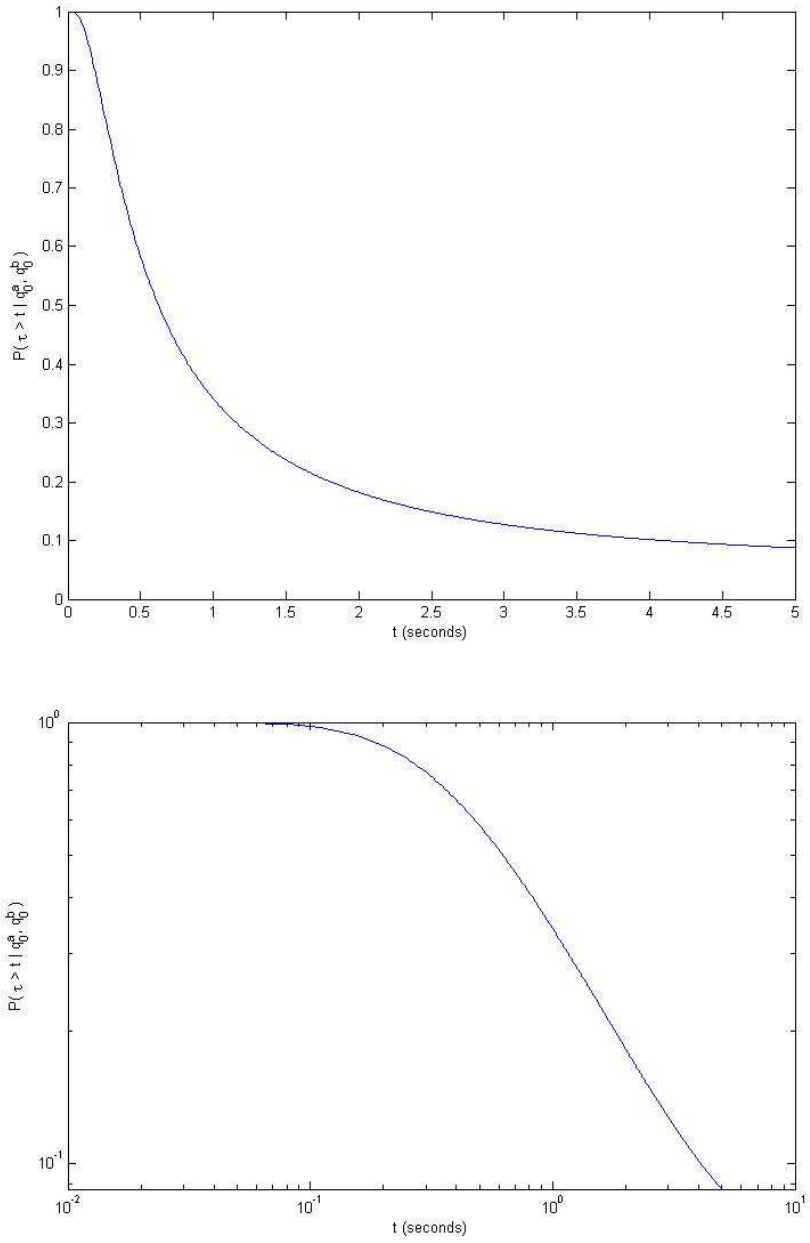

Figure 2.4: Above: $P(\tau > t | q_0^b = 4, q_0^a = 5)$ as a function of t for $\lambda = 12, \mu + \theta = 13$. Below: same figure in log-log coordinates. Note the Pareto tail which decays as t^{-2}.

2.3.2 Probability of price increase: balanced order flow

Starting from a given configuration of the limit order book, the probability that the next price move is an increase is given by the probability that the process (q_t^b, q_t^a) hits the x-axis before the y-axis When $\lambda = \mu + \theta$, i.e. when the flow of limit orders is balanced by the flow of market orders and cancelations, this probability can be computed analytically in terms of hitting time distributions of a random walk in the orthant:

Proposition 2.2. *For $(n,p) \in \mathbb{N}^2$, the probability $p_1^{up}(n,p)$ that the next price move is an increase, conditioned on having the n orders on the bid side and p orders on the ask side is:*

$$p_1^{up}(n,p) = \frac{1}{\pi}\int_0^\pi \left(2-\cos(t)-\sqrt{(2-\cos(t))^2-1}\right)^p \frac{\sin(nt)\cos(\frac{t}{2})}{\sin(\frac{t}{2})}dt. \tag{2.7}$$

Proof. Let us start by noting that $q_t 1_{t\leq\tau} = M_{N_{2\lambda t}}$ where $(M_n, n \geq 0)$ is a symmetric random walk in the positive orthant $\mathbb{Z}_+^2$ killed at the boundary and $(N_{2\lambda t}, t \geq 0)$ is a Poisson process with parameter 2λ. Hence the probability of an upward move in the price starting from a configuration $q_t^b = n, q_t^a = p$ for the order book is equal to the probability that the random walk M starting from (n,p) hits the x-axis before the y-axis. The generator of the bivariate random walk $(M_n, n \geq 1)$ is the discrete Laplacian so $p_1^{up}(n,p) = \mathbb{P}[\tau_a < \tau_b | q_{0-}^b = n, q_{0-}^a = p]$ satisfies, for all $n \geq 1$ and $p \geq 1$,

$$4p_1^{up}(n,p) = p_1^{up}(n+1,p) + p_1^{up}(n-1,p) + p_1^{up}(n,p+1) + p_1^{up}(n,p-1), \tag{2.8}$$

with the boundary conditions: $p_1^{up}(0,p) = 0$ for all $p \geq 1$ and $p_1^{up}(n,0) = 1$ for all $n \geq 1$. This problem is known as the *discrete Dirichlet* problem; solutions of (2.8) are called discrete harmonic functions. (Lawler and Limic 2010, Ch. 8) show that for all $t \geq 0$, the functions

$$f_t(x,y) = e^{xr(t)}\sin(yt), \qquad \text{and} \qquad \tilde{f}_t(x,y) = e^{-xr(t)}\sin(yt) \qquad \text{with} \quad r(t) = \cosh^{-1}(2-\cos t)$$

are solutions of (2.8). In (Lawler and Limic 2010, Corollary 8.1.8) it is shown that the probability that a simple random walk $(M_k, k \geq 1)$ starting at $(n,p) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ reaches the axes at $(x,0)$ is

$$\frac{2}{\pi}\int_0^\pi e^{-r(t)p}\sin(nt)\sin(tx)dt,$$

therefore

$$p_1^{up}(n,p) = \sum_{k=1}^\infty \frac{2}{\pi}\int_0^\pi e^{-r(t)p}\sin(tn)\sin(tk)dt.$$

Since

$$\sum_{k=1}^m \sin(kt) = \frac{\sin(\frac{mt}{2})\sin(\frac{(m+1)t}{2})}{\sin(t/2)} = \frac{\cos(\frac{t}{2}) - \cos((m+\frac{1}{2})t)}{2\sin(t/2)},$$

using integration by parts we see that the second term leads to the integral:

$$\int_0^\pi \underbrace{\frac{e^{-r(t)p}\sin(nt)}{\sin(t/2)}}_{g(t)}\cos((m+1/2)t)dt = -\frac{1}{m+\frac{1}{2}}\int_0^\pi g'(t)\sin((m+\frac{1}{2})t)dt \underset{m\to\infty}{\to} 0.$$

since g' is bounded. So finally

$$p_1^{up}(n,p) = \frac{1}{\pi}\int_0^\pi e^{-r(t)p}\sin(tn)\frac{\cos(\frac{t}{2})}{\sin(\frac{t}{2})}dt.$$

Noting that $e^{-r(t)} = 2-\cos(t)-\sqrt{(2-\cos(t))^2-1}$ we obtain the result. □

Note that the conditional probabilities (2.7) are, in the case of a balanced order book, independent of the parameters describing the order flow.

The expression (2.7) is easily computed numerically: Figure 2.5 displays the shape of the function p_1^{up}. Comparison with empirical data for CitiGroup stock (June 2008) shows good agreement between the theoretical value (2.7) and the empirical transition frequencies of the price conditional on the state of the consolidated order book.

2.3.3 Probability of price increase: asymmetric order flow

In this subsection we relax the symmetry assumptions above and allow the intensity of limit and market orders at the bid and the ask to be different; more precisely we assume that:

- Limit orders at the ask arrive at independent, exponential time with parameter λ^a
- Market orders and cancelations at the ask arrive at independent, exponential time with parameter $\mu^a + \theta^a$
- Limit orders at the bid arrive at independent, exponential time with parameter λ^b
- Market orders and cancelations at the bid arrive at independent, exponential time with parameter $\mu^b + \theta^b$

and that these Poisson processes are independent. The dynamics of bid and ask queues may be then represented as

$$q_t = M_{N_{\Lambda t}} \quad \text{for} \quad \Lambda = \lambda^a + \mu^a + \theta^a + \lambda^b + \mu^b + \theta^b,$$

where $N_{\Lambda t}$ is a Poisson process with intensity Λ and $(M_n, n \geq 0)$ is a random walk on $\mathbb{N}^2$ killed when it hits the boundary, whose the transition probabilities are:

$$p_{0,1} = \frac{\lambda^a}{\Lambda} \quad p_{1,0} = \frac{\lambda^b}{\Lambda} \quad p_{0,-1} = \frac{\mu^a + \theta^a}{\Lambda} \quad p_{-1,0} = \frac{\mu^b + \theta^b}{\Lambda}. \tag{2.9}$$

The following result generalizes Proposition 2.2 for an asymmetric order flow.

Proposition 2.3. *Given $(q^b, q^a) = (n, p)$, the probability $p_1^{up}(n,p)$ that the next price move is an increase is:*

$$p_1^{up}(n,p) = 1 - \frac{1}{\pi}\left(\frac{\mu^a + \theta^a}{\lambda^a}\right)^p \frac{2[\lambda^a(\mu^a+\theta^a)]^{1/2}}{\mu^a + \theta^a + \lambda^a} \times$$

$$\int_0^\pi dt \frac{2\lambda^b Z(t) - (\Lambda - 2[\lambda^a(\mu^a+\theta^a)]^{1/2}\cos(t))}{\frac{2[(\mu^a+\theta^a)\lambda^a]^{1/2}}{\lambda^b+\mu^b+\theta^b}\cos(t) - 1} \frac{Z(t)^n \sin(pt)\sin(t)}{\sqrt{(\Lambda - 2[(\mu^a+\theta^a)\lambda^a]^{1/2}\cos(t))^2 - 4(\mu^b+\theta^b)\lambda^b}},$$

where $(Z(t), t \geq 0)$ is defined by:

$$\forall t \geq 0, \quad Z(t) = \frac{\Lambda - 2[(\mu^a+\theta^a)\lambda^a]^{1/2}\cos(t) - \sqrt{(\Lambda - 2[(\mu^a+\theta^a)\lambda^a]^{1/2}\cos(t))^2 - 4\lambda^b(\theta^b+\mu^b)}}{2\lambda^a}.$$

Proof. Using results from Kurkova and Raschel (2011), it is shown in Raschel and van Leeuwaarden (2011) that the probability that M starting from (n,p) hits the x-axis before the y-axis is given by

$$1 - \frac{1}{\pi}\left(\frac{p_{0,-1}}{p_{0,1}}\right)^p \frac{2\sqrt{p_{0,1}p_{0,-1}}}{p_{0,1}+p_{0,-1}} \int_0^\pi Z(t)^n \sin(pt)\sin(t) \times$$

$$\frac{2p_{1,0}Z(t) - (1 - 2\sqrt{p_{0,1}p_{0,-1}}\cos(t))}{\frac{2\sqrt{p_{0,1}p_{0,-1}}}{p_{0,1}+p_{0,-1}}\cos(t) - 1} \frac{dt}{\sqrt{(1 - 2\sqrt{p_{0,1}p_{0,-1}}\cos(t))^2 - 4p_{1,0}p_{-1,0}}},$$

where $(Z(t), t \geq 0)$ is the function defined by:

$$\forall t \geq 0, \quad Z(t) = \frac{1 - 2[p_{0,1}p_{0,-1}]^{1/2}\cos(t) - \sqrt{(1 - 2[p_{0,1}p_{0,-1}]^{1/2}\cos(t))^2 - 4p_{1,0}p_{-1,0}}}{2p_{1,0}}.$$

Applying this formula to the random walk whose transition probabilities are given in Equation (2.9) we obtain that the probability of a price increase at the next price move, starting from a configuration $q_t^b = n, q_t^a = p$ of the queues is equal to the probability that the random walk M starting from (n, p) hits the x-axis before the y-axis; thus, it is given by equation (2.3). □

2.3.4 Dynamics of the price

The high-frequency dynamics of the price is described by a piecewise constant, right continuous process $(s_t, t \geq 0)$ whose jumps times correspond to times when the order book process $(q_t, t \geq 0)$ hits one of the axes. Denote by $(\tau_1, \tau_2, ...)$ the successive durations between price changes. The number of price changes that occur during $[0, t]$ is given by

$$N_t := \max\{\ n \geq 0, \quad \tau_1 + ... + \tau_n \leq t\ \}$$

At $t = \tau_i$, $s_{\tau_i} = s_{\tau_{i-}} + 1$ if $q_{\tau_{i-}^a} = 0$ and $s_{\tau_i} = s_{\tau_{i-}} - 1$ if $q_{\tau_{i-}^b} = 0$. $(X_1, X_2, X_3, ..., X_n, ...)$ are the successive moves in the price which, in general, are not independent random variables. We have

$$s_t = Z(N_t) \quad \text{where} \qquad Z(n) = \sum_{i=1}^{n} X_i$$

is the price after n price changes have occurred. Hence, for all $t \geq 0$, $s_t = Z(N_t)$. We are interested in the n-step ahead distribution of the price change:

$$p_n^{up}(x, y) = \mathbb{P}[X_n = +\delta \quad | \quad (q_0^b, q_0^a) = (x, y)] \tag{2.10}$$

For $n = 1$ this corresponds to the probability $p_1^{up}(x, y) = p_1(x, y)$ of an upward price move, computed in Theorem 2.2. To simplify the analysis we use, in this Section and the next one, the following symmetry assumption:

Assumption 2.1 (Bid-ask symmetry)**.** $\tilde{f}(x, y) = f(y, x)$.

This assumption means that the distribution of bid queue after a price decrease is the same as the distribution of the ask queue after a price increase; this is the case for example if the order flows at the first two levels on each side of the book have similar statistical features.

A key quantity for studying the dynamics of the price is the probability of two successive price changes in the same direction,

$$p_{cont} = \mathbb{P}[X_{k+1} = \delta | X_k = \delta] = \mathbb{P}[X_{k+1} = -\delta | X_k = -\delta] \tag{2.11}$$

which may be expressed in terms of the distribution of queue sizes f after a price change:

Proposition 2.4. *Let $p_{cont} = \mathbb{P}[X_2 = \delta | X_1 = \delta] = \mathbb{P}[X_2 = -\delta | X_1 = -\delta]$ be the probability of two successive price moves in the same direction.*

- $p_{cont} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i, j)p_1^{up}(i, j)$.
- $\forall k \geq 1, \quad \mathbb{E}[X_k | q_0^b = x, q_0^a = y] = (2p_1^{up}(x, y) - 1)(2p_{cont} - 1)^{k-1}$.
- $cov(X_1, X_2 | q_0^b = x, q_0^a = y) = \delta^2(2p_{cont} - 1)(1 - (2p_1^{up}(x, y) - 1)^2)$.

- *Conditional on the current state of the limit order book, the n-step ahead distribution of the price change is given by:*

$$p_n^{up}(x,y) := \mathbb{P}[X_n = \delta | q_0^b = x,\ q_0^a = y] = \frac{1 + (2p_{cont} - 1)^{n-1}(2p_1^{up}(x,y) - 1)}{2}. \tag{2.12}$$

Proof. First, let us prove that $\mathbb{P}[X_2 = \delta | X_1 = \delta] = \mathbb{P}[X_2 = -\delta | X_1 = -\delta]$:

$$\mathbb{P}[X_2 = \delta | X_1 = \delta] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j) p_1^{up}(i,j),$$

where p_1^{up} is given in 2.7, by symmetry of the bid and the ask, for all $(n,p) \in \mathbb{N}^2$, $p_1^{up}(n,p) = 1 - p_1^{up}(p,n)$. By assumption 2.1, for all $(i,j) \in \mathbb{N}^2$, $f(i,j) = \tilde{f}(j,i)$. Therefore,

$$\mathbb{P}[X_2 = \delta | X_1 = \delta] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)(1 - p_1^{up}(j,i)),$$

$$\mathbb{P}[X_2 = \delta | X_1 = \delta] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \tilde{f}(j,i)(1 - p_1^{up}(j,i)),$$

$$\mathbb{P}[X_2 = \delta | X_1 = \delta] = \mathbb{P}[X_2 = -\delta | X_1 = -\delta].$$

$p_n^{up} = p_n^{up}(x,y)$ defined by (2.12) is then characterized by the following relation:

$$\begin{pmatrix} p_n^{up} \\ 1 - p_n^{up} \end{pmatrix} = \begin{pmatrix} p_{cont} & 1 - p_{cont} \\ 1 - p_{cont} & p_{cont} \end{pmatrix} \begin{pmatrix} p_{n-1} \\ 1 - p_{n-1} \end{pmatrix},$$

hence

$$\begin{pmatrix} p_n^{up} \\ 1 - p_n^{up} \end{pmatrix} = \begin{pmatrix} p_{cont} & 1 - p_{cont} \\ 1 - p_{cont} & p_{cont} \end{pmatrix}^{n-1} \begin{pmatrix} p_1 \\ 1 - p_1 \end{pmatrix}.$$

The eigenvalues of this matrix are 1 and $2p_{cont} - 1$:

$$\begin{pmatrix} p_{cont} & 1 - p_{cont} \\ 1 - p_{cont} & p_{cont} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2p_{cont} - 1 \end{pmatrix} \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

Therefore

$$p_n^{up}(x,y) = \frac{1 + (2p_{cont} - 1)^{n-1}(2p_1(x,y) - 1)}{2}.$$

Moreover for all $n \geq 2$,

$$\mathbb{E}[X_n | q_0^b = x, q_0^a = y] = (2p_n^{up}(x,y) - 1) = (2p_{cont} - 1)^{n-1}(2p_1(x,y) - 1).$$

and the correlation between two consecutive price moves is given by:

$$\begin{aligned} \operatorname{cov}(X_1, X_2 | q_0^b = x, q_0^a = y) &= \mathbb{E}[X_1 X_2 | q_0^b = x, q_0^a = y] - \mathbb{E}[X_1 | q_0^b = x, q_0^a = y]\mathbb{E}[X_2 | q_0^b = x, q_0^a = y] \\ &= \delta^2(2p_{cont} - 1) - \delta^2(2p_1(x,y) - 1)(2p_2(x,y) - 1), \end{aligned}$$

$$\operatorname{cov}(X_1, X_2 | q_0^b = x, q_0^a = y) = \delta^2(2p_{cont} - 1) - (2p_1(x,y) - 1)^2(2p_{cont} - 1) = (2p_{cont} - 1)(1 - (2p_1(x,y) - 1)^2). \tag{2.13}$$

□

Remark 2.2 (Negative autocorrelation of price changes at first lag)**.** *It is empirically observed that high frequency price movements have a negative autocorrelation at the first lag; this observation is often attributed to the 'bid-ask' bounce of transaction prices, but in fact it also holds for the time series of bid or ask prices (Cont 2001). Our model links the value of this autocorrelation at first lag to the properties of the distribution f of order book depth. As observed from* (2.13)*, the sign of the* $\mathrm{cov}(X_1, X_2|q_0^b, q_0^a)$ *does not depend on the initial configuration* (q_0^b, q_0^a) *of the bid/ask queues, so* $\mathrm{cov}(X_k, X_{k+1}) = \mathrm{cov}(X_1, X_2) < 0$ *if and only if*

$$p_{cont} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)p_1^{up}(i,j) < 1/2$$

where f is the joint distribution of queue sizes after a price increase. This condition is satisfied for most high-frequency data sets of Dow Jones stocks we have examined. For example, for CitiGroup stock we find

$$p_{cont} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)p_1^{up}(i,j) \simeq 0.35$$

This asymmetry of f corresponds to the fact that, after an upward price move, the new bid queue is generally smaller than the ask queue since the ask queue corresponds to the limit order previously sitting at second best ask level, while the bid queue results from the accumulation of orders over the very short period since the last price move. Under this condition, high frequency increments of the price are negatively correlated: an increase in the price is more likely to be followed by a decrease in the price.

Remark 2.3. *The sequence of price increments* $(X_1, X_2, ...)$ *is uncorrelated if and only if*

$$p_{cont} = \sum_{i,j\geq 1} f(i,j)p_1^{up}(i,j) = 1/2$$

2.3.5 "Efficient" price

Various authors (see e.g. Robert and Rosenbaum (2011)) have considered models in which the evolution of transaction prices is based on a non-observed (semi)martingale $\hat{s}$, sometimes called the "efficient" price: the observed price is then either a noisy version of $\hat{s}$ or the value of $\hat{s}$ rounded to the nearest tick.

Given the probability $p_1^{up}(q_t^b, q_t^a)$ that the next price move is an "uptick" (Equation (2.7)), we can construct an auxiliary process $\hat{s}$ whose value $\hat{s}_t$ represents the expected value of the price after its next move:

$$\begin{aligned}\forall t \geq 0,\ \hat{s}_t &= (s_t + \delta)\, p_1^{up}(q_t^b, q_t^a) + (s_t - \delta)\left(1 - p_1^{up}(q_t^b, q_t^a)\right),\\ \hat{s}_t &= s_t + \delta(2p_1^{up}(q_t^b, q_t^a) - 1),\end{aligned}$$

$(\hat{s}_t, t \geq 0)$ is a continuous-time stochastic process with values between $s_t - \delta$ and $s_t + \delta$:

$$\forall t \geq 0, \quad s_t - \delta \leq \hat{s}_t \leq s_t + \delta.$$

The process $\hat{s}$ incorporates the information on the price s_t and the state of the order book (q^b, q^a) insofar as it affects the next price move. The following result shows

Proposition 2.5. *If* $p_{\mathrm{cont}} = 1/2$ *then* $(\hat{s}_t, t \geq 0)$ *is a martingale.*

Remark 2.4. *This condition is verified in particular if* $\forall (i,j) \in \mathbb{N}^2,\ f(i,j) = f(j,i)$ *but more generally if*

$$\sum_{i,j\geq 1} p_1^{up}(i,j)f(i,j) = \frac{1}{2}$$

Proof. Let $(\tau_1, \tau_2, ..., \tau_k)$ the sequence of times when the price s moves and $(X_1, ..., X_n)$ the sequence of consecutive price moves. Since $p_{cont} = 1/2$, $(X_1, ..., X_k, ...)$ is a sequence of I.I.D Bernoulli random variables with parameter $1/2$. Therefore we have the following property:

$$\forall (i,j) \in \mathbb{N}^2,\ i < j,\ \mathbb{E}[s_{\tau_j} | \mathcal{F}_{\tau_i}] = s_{\tau_i}.$$

The function p_1^{up}, from equation (2.7), satisfies the equation $Lp_1^{up} = 0$, where L is the generator of the process (q^b, q^a). Hence p_1^{up} is an harmonic function for the process (q^b, q^a), and the process $(p_1^{up}(q_t^b, q_t^a), t \geq 0)$ is a martingale. We proved that

$$\forall\ s \leq t < \tau_1,\ \ \mathbb{E}[\hat{s}_t | \mathcal{F}_s] = \hat{s}_s.$$

By recurrence on k, one can easily notice that

$$\forall\ k\ \geq 1,\ \forall \tau_k \leq t < \tau_{k+1},\ \ \mathbb{E}[\hat{s}_t | \mathcal{F}_{\tau_1}] = \hat{s}_{\tau_1}.$$

Assuming $s \leq \tau_1 \leq t$,

$$\begin{aligned}
\mathbb{E}[\hat{s}_t | \mathcal{F}_s] &= \mathbb{E}[\hat{s}_t | \mathcal{F}_s, X_1 = 1]\mathbb{P}[X_1 = 1 | \mathcal{F}_s] + \mathbb{E}[\hat{s}_t | \mathcal{F}_s, X_1 = -1]\mathbb{P}[X_1 = -1 | \mathcal{F}_s], \\
&= \mathbb{E}[\hat{s}_t | \mathcal{F}_s, X_1 = 1] p_1^{up}(q_t^b, q_t^a) + \mathbb{E}[\hat{s}_t | \mathcal{F}_s, X_1 = -1](1 - p_1^{up}(q_t^b, q_t^a)), \\
&= (s_s + \delta) p_1^{up}(q_s^b, q_s^a) + s_s(1 - p_1^{up}(q_s^b, q_s^a)), \\
&= s_s + \delta(2p_1^{up}(q_s^b, q_s^a) - 1) = \hat{s}_s,
\end{aligned}$$

which completes the proof. □

Contrarily to the 'latent price' models alluded to above, here $\hat{s}$ is a function of the state variables (s_t, q_t^b, q_t^a) and thus is observable, provided one observes trades and quotes.

Remark 2.5. *When $p_{cont} \neq 1/2$, the process $(\hat{s}_t, t \geq 0)$ fails to possess the martingale property. When $p_{cont} < 1/2$, the jump of $\hat{s}$ is negative after a price increase and positive after a price decrease.*

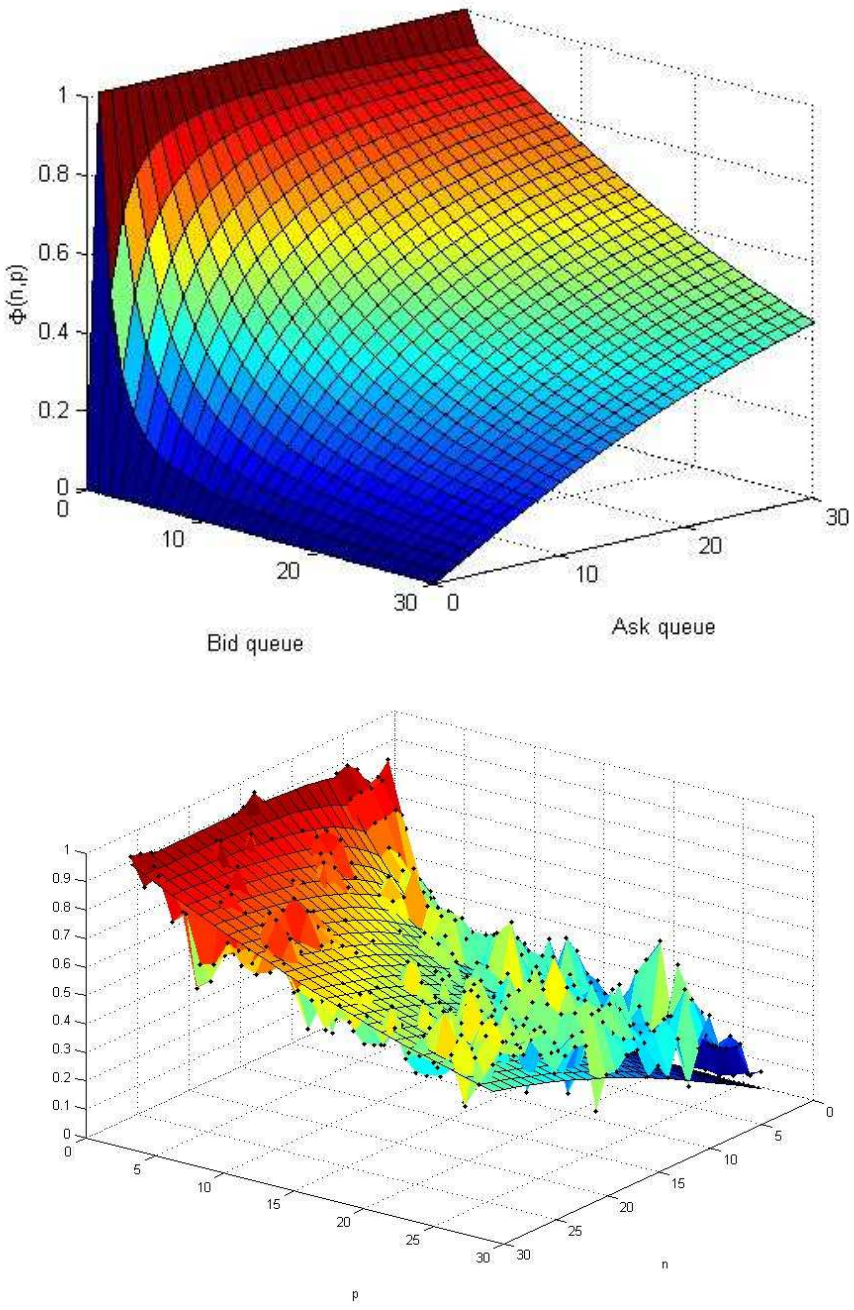

Figure 2.5: Above: conditional probability of a price increase, as a function of the bid and ask queue size. Below: comparison with empirical transition frequencies for CitiGroup stock price tick-by-tick data on June 26, 2008.

2.4 Diffusion limit of the price process

As discussed in Section 2.3.4, the high frequency dynamics of the price is described by a piecewise constant stochastic process $s_t = Z(N_t)$ where

$$Z(n) = X_1 + ... + X_n \quad and \quad N_t = \sup\{k;\ \tau_1 + ... + \tau_k \leq t\}$$

is the number of price changes during $[0, t]$.

However, over time scales much larger than the interval between individual order book events, prices are observed to have diffusive dynamics and modeled as such. To establish the link between the high frequency dynamics and the diffusive behavior at longer time scales, we shall consider a time scale t_n over which the average number of order book events is of order n and exhibit conditions under which the scaled price process

$$(s_t^n := \frac{s_{t_n}}{\sqrt{n}}, t \geq 0)_{n \geq 1}$$

satisfies a functional central limit theorem i.e. converges in distribution to a non-degenerate process $(p_t, t \geq 0)$ as $n \to \infty$. The choice of the time scale t_n is such that

$$\frac{\tau_1 + ... + \tau_n}{t_n}$$

has a well-defined limit: it is imposed by the distributional properties of the durations which, as observed in Section 2.3.1, are heavy tailed. In this section, we show that, under a symmetry condition, this limit can be identified as a diffusion process whose diffusion coefficient may be computed from the statistics of the order flow driving the limit order book.

Assume $\lambda + \theta \leq \mu$ and that the joint distribution f of the queue sizes after a price move satisfies:

$$D(f) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} ij f(i,j) < \infty \tag{2.14}$$

The quantity $D(f)$ represents a measure of market depth: $\sqrt{D(f)}$, quoted in number of shares, represents an average of the size of the bid and ask queues after a price change.

In this section we assume that f is symmetric: $\forall i, j \geq 0,\ f(i,j) = f(j,i)$. Together with Assumption 2.1, this implies that price increments form a sequence $(X_i, i \geq 1)$ of independent random variables with distribution:

$$\mathbb{P}[X_1 = \delta] = \mathbb{P}[X_1 = -\delta] = \frac{1}{2}.$$

We will show that the limit p is then a diffusion process which describes the dynamics of the price at lower frequencies. In particular, we will compute the volatility of this *diffusion limit* and relate it to the properties of the order flow.

In the following $\mathcal{D}$ denotes the space of right continuous paths $\omega : [0, \infty) \to \mathbb{R}^2$ with left limits, equipped with the Skorokhod topology J_1, and $\Rightarrow$ will designate weak convergence on $(\mathcal{D}, J_1)$ Billingsley (1968), Whitt (2002).

2.4.1 Balanced order flow

We first consider the case of a *balanced* order flow for which the intensity of market orders and cancelations is equal to the intensity of limit orders. The study of high-frequency quote data indicates that this is an empirically relevant case for many liquid stocks: as shown in Table 2.3, the imbalance between arrival of limit orders on one hand and market orders/ cancelations on the other hand is around 5% or less for these stocks.

For a balanced order flow, we proved in Section 2.3.1 that the distribution of price duration τ conditioned on observing i shares at the bid and j shares at the ask at $t = 0$ has a tail index 1:

$$\mathbb{P}[\tau > t | q_0^b = i, q_0^a = j] \sim \frac{ij}{\pi\lambda t}. \tag{2.15}$$

Similarly, the unconditional distribution of price durations has a tail index of one:

$$\mathbb{P}[\tau > t] \overset{t\to\infty}{\sim} \frac{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} ijf(i,j)}{\pi\lambda t} = \frac{D(f)}{\pi\lambda t}.$$

The sequence of durations between consecutive move of the price consecutive $(\tau_1, \tau_2, \tau_3, ...)$ is a sequence of IID random variables with tail index 1. The following lemma 2.1, which we include for completeness, gives a law of large numbers for this sequence of durations (see Samorodnitsky and Taqqu (1994)).

Lemma 2.1. *The sequence of durations* $(\tau_1, \tau_2, \tau_3, ...)$ *satisfies*

$$\frac{\tau_1 + \tau_2 + ... + \tau_k}{k \log k} \overset{k\to\infty}{\Rightarrow} \frac{D(f)}{\pi\lambda}.$$

Proof. $(\tau_k, k \geq 1)$ is a sequence of i.i.d, regularly varying random variables, with tail index equal to 1. Let $\mathcal{L}(s)$ be the Laplace transform of the distribution of τ_2:

$$\mathcal{L}(s) = 1 - s\int_0^{\infty} \exp(-st)\mathbb{P}[\tau > t]dt.$$

We have

$$\frac{1}{\log(n)}\int_0^n \exp(-\frac{st}{n\log n})\mathbb{P}[\tau > t]dt \overset{n\to\infty}{\to} \frac{D(f)}{\pi\lambda} \qquad \text{and} \qquad \frac{1}{\log(n)}\int_n^{\infty} \exp(-\frac{st}{n\log n})\mathbb{P}[\tau > t]dt \overset{n\to\infty}{\to} 0.$$

Therefore

$$\frac{s}{n\log(n)}\int_0^{\infty} \exp(-\frac{st}{n\log n})\mathbb{P}[\tau > t]dt = \frac{D(f)}{\pi\lambda}\frac{s}{n} + o(\frac{1}{n})$$

which implies

$$\left(\mathcal{L}(\frac{s}{n\log n})\right)^n = (1 + \frac{D(f)}{\pi\lambda} + o(\frac{1}{n}))^n \to \exp(-\frac{D(f)}{\pi\lambda}).$$

So one can conclude that

$$\frac{\tau_1 + \tau_2 + ... + \tau_n}{n\log n} \overset{\mathcal{D}}{\Rightarrow} \frac{D(f)}{\pi\lambda}.$$

□

Setting $t_n = tn\log n$, N_{t_n}, the number of price change in the interval $[0, nt\log n]$, satisfies

$$\frac{t_n}{N_{t_n}\log N_{t_n}} \overset{n\to\infty}{\sim} \frac{D(f)}{\pi\lambda} \qquad a.s.$$

Hence if one defines $\rho : (1,\infty) \to (1,\infty)$ as the inverse function of $t \mapsto t\log t$,

$$\forall t > 1, \quad \rho(t)\log\rho(t) = t, \tag{2.16}$$

then $\rho(t) \sim_{t\to\infty} \frac{t}{\log(t)}$ and the number of price changes occurring during $[0, t_n]$ verifies the following scaling relations

$$N_{t_n} \overset{n\to\infty}{\sim} \rho\left(\frac{\pi\lambda t_n}{D(f)}\right) \overset{n\to\infty}{\sim} \frac{\pi\lambda tn\log n}{D(f)\log(n\log n)} \sim \frac{\pi\lambda t}{D(f)}n \quad \text{a.s.} \tag{2.17}$$

The asymptotic relation (2.17) shows that the number of price moves occurring during $[0, tn\log n]$ is of order n for large n, and proportional to the ratio of the order intensity λ to the order book depth $D(f)$. Since each price change is $\pm\delta$, this factor $\frac{\pi\lambda}{D(f)}$ also shows up in the expression of price volatility:

Theorem 2.1. *If $\lambda = \mu + \theta$,*

$$\left(\frac{s_{tn\log n}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \left(\delta\sqrt{\frac{\pi\lambda}{D(f)}}W_t, t \geq 0\right)$$

where δ is the tick size, $D(f)$ is given by (2.14) *and W is a standard Brownian motion.*

Proof. Let $t_n = tn\log n$. One can decompose the process $(\frac{s_{tn\log n}}{\sqrt{n}}, t \geq 0)$ as

$$\frac{s_{tn\log n}}{\sqrt{n}} = \frac{Z([tn\pi\lambda/D(f)])\delta}{\sqrt{n}} + \left(\frac{Z(N_{t_n})\delta}{\sqrt{n}} - \frac{Z([tn\pi\lambda/D(f)])\delta}{\sqrt{n}}\right) \tag{2.18}$$

Since $(X_1, X_2, ...)$ is a sequence of IID random variables with mean zero, one can apply the Donsker's invariance principle to the sequence of processes to obtain

$$(\frac{Z([t\pi\lambda n/D(f)])\delta}{\sqrt{n}}, t \geq 0)_{n\geq 1} \overset{n\to\infty}{\Rightarrow} \left(\delta\sqrt{\frac{\pi\lambda}{D(f)}}W_t, t \geq 0\right).$$

As shown in (2.17),

$$N_{tn\log n} \overset{n\to\infty}{\sim} \frac{nt\pi\lambda}{D(f)},$$

therefore for $t \geq 0$,

$$\left(\frac{Z(N_{t_n})\delta}{\sqrt{n}} - \frac{Z([tn\pi\lambda/D(f)])\delta}{\sqrt{n}}\right) \overset{n\to\infty}{\Rightarrow} 0. \tag{2.19}$$

Hence the finite dimensional distributions of the sequence of processes

$$\left(\frac{Z(N_{t_n})\delta}{\sqrt{n}} - \frac{Z(t\pi\lambda/D(f))\delta}{\sqrt{n}}, t \geq 0\right)_{n\geq 1}$$

converge to a point mass at zero. Since this sequence of processes is tight on $(\mathcal{D}, J_1)$, this sequence of processes converges weakly to zero on $(\mathcal{D}, J_1)$ (see Whitt (2002)). So finally

$$\left(\frac{s_{tn\log n}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \delta\sqrt{\frac{\pi\lambda}{D(f)}} \quad W.$$

□

2.4.2 Empirical test using high-frequency data

Theorem 2.1 relates the 'coarse-grained' volatility of intraday returns at lower frequencies to the high-frequency arrival rates of orders. Denote by $\tau_0 = 1/\lambda$ the typical time scale separating order book events. Typically τ_0 is of the order of milliseconds. In plain terms, Theorem 2.1 states that, observed over a time scale $\tau_2 >> \tau_0$ (say, 10 minutes), the price has a diffusive behavior with a diffusion coefficient given by

$$\sigma_n = \delta\sqrt{\frac{n\pi\lambda}{D(f)}} \tag{2.20}$$

where δ is the tick size, n is an integer verifying $n\ln n \quad \tau_0 = \tau_2$ which represents the average number of orders during an interval τ_2 and $\sqrt{D(f)}$, which measures the average depth of the bid and ask queues after a price change.

The relation (2.20) links the variance of price changes, to statistical properties of the order flow: it yields an estimator for price volatility which may be computed from observations on quote durations.

It shows that, in two 'balanced' limit order markets with the same tick size and same arrival rate of orders at the bid/ask, the market with higher depth of the next-to-best queues will lead to lower price volatility.

More precisely, this formula shows that the microstructure of order flow affects price volatility through the ratio $\lambda/D(f)$ where λ is the rate of execution/cancelation of limit orders and $D(f)$, given by (2.14), is a measure of market depth: in fact, our model predicts a proportionality between the variance of price increments and this ratio. This is an empirically testable prediction.

Figure 2.6 compares, for stocks in the Dow Jones index, the standard deviation of 10-minute price increments with $\sqrt{\lambda/D(f)}$. We observe that, indeed, stocks with a higher value of the ratio $\lambda/D(f)$ have a higher variance, and standard deviation of price increments increases roughly proportionally to $\sqrt{\lambda/D(f)}$. The linear relation (2.20) explains 68% of the cross sectional variation iof intraday volatility across stocks.

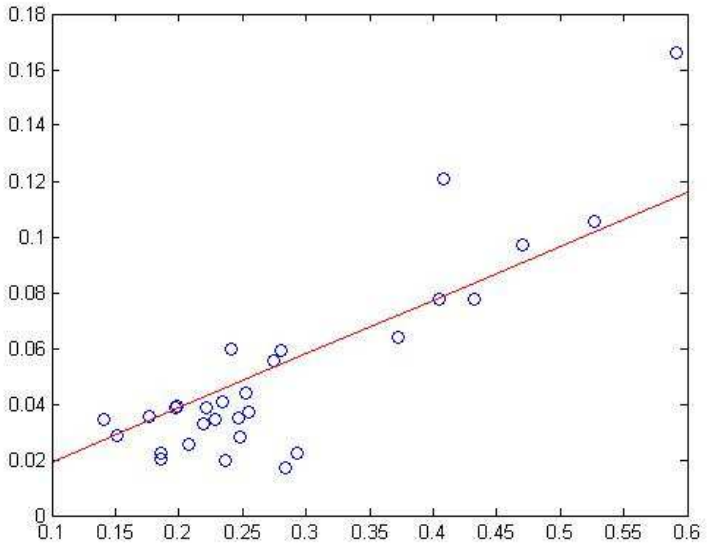

Figure 2.6: $\sqrt{\lambda/D(f)}$, estimated from tick-by-tick order flow (vertical axis) vs standard deviation of 10-minute price increments (horizontal axis) for stocks in the Dow Jones Index, estimated from high frequency data on June 26, 2008. Each point represents one stock. Red line indicates the best linear approximation.

Remark 2.6. *When the intensity of all orders coming in the limit order book is multiplied by the same factor x,*

- *The intensity of limit orders becomes λx*
- *The intensity of market orders and cancelations becomes $(\mu + \theta)x$*
- *The depth of the limit order book increases by a factor x, so $D(f)$ increases by a factor x^2.*

Substituting in the above formula, we then see that price volatility is decreased by a factor $\sqrt{\frac{1}{x}}$. Interestingly, Rosu (2009) shows the same dependence in $1/\sqrt{x}$ of price volatility using an equilibrium approach.

2.4.3 Case when market orders and cancelations dominate

We now consider the case in which the flow of market orders and cancelations dominates that of limit orders: $\lambda < \theta + \mu$. In this case, price changes are more frequent since the order queues are depleted by market orders or cancelations at a faster rate than they are replenished by limit orders. We also obtain a diffusion limit for the price process, but with a different scaling:

Theorem 2.2. *Let $\lambda < \theta + \mu$ and assume that the probability distribution f satisfies*

$$m(\lambda, \theta+\mu, f) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} m(\lambda, \theta+\mu, i, j) f(i,j) < \infty,$$

where for $(x, y) \in (\mathbb{N}^*)^2$,

$$m(\lambda, \theta+\mu, x, y) = \left(\frac{\mu+\theta}{\lambda}\right)^{\frac{x+y}{2}} \int_0^{\infty} dt \quad \psi_{x,\lambda,\mu+\theta}(t)\psi_{y,\lambda,\mu+\theta}(t)$$

and $\psi_{x,\lambda,\mu+\theta}$ is given by (2.4)*. Then*

$$\left(\frac{s_{nt}}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \left(\frac{\delta}{\sqrt{m(\lambda, \theta+\mu, f)}} W_t, t \geq 0\right)$$

where W is a standard Brownian motion.

Proof. $(\tau_2, \tau_3, ...)$ is a sequence of i.i.d random variables with finite mean given by

$$E[\tau_k] = m(\lambda, \theta+\mu, f) = \sum_{(i,j)\in\mathbb{Z}_+^2} f(i,j) m(\lambda, \theta+\mu, i, j).$$

We can apply the (strong) law of large numbers to this sequence:

$$\frac{\tau_1 + \tau_2 + ... + \tau_n}{n} \overset{n\to\infty}{\to} m(\lambda, \theta+\mu, f)$$

where $m(\lambda, \theta+\mu, f)$ is given as above. Therefore,

$$\forall t \geq 0, \qquad N_{tn} \overset{n\to\infty}{\sim} [\frac{tn}{m(\lambda, \theta+\mu, f)}].$$

The rest of the proof follows the lines of Theorem 2.1. We start by decomposing $\frac{s_{[nt]}}{\sqrt{n}}$ into

$$\frac{s_{[nt]}}{\sqrt{n}} = \frac{\delta Z([nt/m(\lambda, \theta+\mu, f)])}{\sqrt{n}} + (\frac{\delta Z(N_{nt})}{\sqrt{n}} - \frac{\delta Z([nt/m(\lambda, \theta+\mu, f)])}{\sqrt{n}}). \tag{2.21}$$

By Donsker's theorem,

$$\left(\frac{\delta Z([nt/m(\lambda, \theta+\mu, f)])}{\sqrt{n}}, t \geq 0\right) \overset{n\to\infty}{\Rightarrow} \left(\frac{1}{\sqrt{m(\lambda, \theta+\mu, f)}} \delta W_t, t \geq 0\right). \tag{2.22}$$

The second term of the decomposition converges to zero:

$$(\frac{\delta Z(N_{nt})}{\sqrt{n}} - \frac{\delta Z([nt/m(\lambda, \theta+\mu, f)])}{\sqrt{n}}, t \geq 0) \overset{n\to\infty}{\Rightarrow} 0, \tag{2.23}$$

which concludes the proof. □

Variance of price change at intermediate frequency Theorem 2.2 leads to an expression of the variance of the price at a time scale $\tau \gg \tau_0$, where $\tau_0 (\sim \text{ms})$ is the average interval between order book events:

$$\sigma^2 = \frac{\tau}{\tau_0} \frac{1}{m(\lambda, \theta + \mu, f)} \delta^2 \tag{2.24}$$

Here, $m(\lambda, \theta + \mu, f)$ represents the expected hitting time of the axes by the Markovian queuing system q with parameters $(\lambda, \theta + \mu)$ and random initial condition with distribution f.

This equation relates the variance of price changes (over a time scale τ_2) to the tick size and the statistical properties of the order flow.

2.4.4 Conclusion

We have exhibited a simple model of a limit order market in which order book events are described in terms of a Markovian queueing system. The analytical tractability of our model allow us to compute various quantities of interest such as

- the distribution of the duration until the next price change,
- the distribution of price changes, and
- the diffusion limit of the price process and its volatility.

in terms of parameters describing the order flow. These results provide analytical insights into the relation between price dynamics and order flow, in particular the relation between liquidity and volatility, in a limit order market.

We view this stylized model as a first step in the analytical study of realistic stochastic models of order book dynamics. Yet, comparison with empirical data shows that even our simple modeling set-up is capable of yielding useful analytical insights into the relation between volatility and order flow, worthy of being further pursued. Moreover, the connection with random walks in the orthant and two-dimensional multiclass queueing systems allows to use the rich analytical theory developed for these systems (Cohen and Boxma 1983, Fayolle et al. 1999, Kurkova and Raschel 2011) to further extend our study. We hope to pursue further some of these ramifications in future work.

A relevant question is to examine which of the above results are robust to departures from the model assumptions and whether the intuitions conveyed by our model remain valid in a more general context where one or more of these assumptions are dropped. This issue is further studied in Cont and de Larrard (2011) where we explore a more general queueing model relaxing some of the assumptions above.

Chapter 3

Heavy traffic limits and diffusion approximations

3.1 Introduction

An increasing proportion of financial transactions -in stocks, futures and other contracts- take place in electronic markets where participants may submit limit orders (for buying or selling), market orders and order cancelations which are then centralized in a *limit order book* and executed according to precise time and price priority rules. The limit order book represents, at each point in time, the outstanding orders which are awaiting execution: it consists in queues at different price levels where these orders are arranged according to time of arrival. A new limit buy (resp. sell) order of size x increases the size of the bid (resp. ask) queue by x. Market orders are executed against limit orders at the best available price: a market order decreases of size x the corresponding queue size by x. Limit orders placed at the best available price are executed against market orders.

The availability of high-frequency data on limit order books has generated a lot of interest in statistical modeling of order book dynamics, motivated either by high-frequency trading applications or simply a better understanding of intraday price dynamics (see Cont (2011) for a recent survey). The challenge here is to develop statistical models which capture salient features of the data while allowing for some analytical and computational tractability.

Given the discrete nature of order submissions and precise priority rules for their execution, is quite natural to model a limit order book as a queueing system; early work in this direction dates back to Mendelson (1982). More recently, Cont, Stoikov and Talreja Cont et al. (2010b) have studied a Markovian queueing model of a limit order book, in which arrivals of market orders and limit orders at each price level are modeled as independent Poisson processes. Cont and de Larrard (2010) used this Markovian queueing approach to compute useful quantities (the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, conditional on the state of the order book) and relate the volatility of the price with statistical properties of the order flow.

However, the results obtained in such Markovian models rely on the fact that time intervals between orders are independent and exponentially distributed, orders are of the same size and that the order flow at the bid is independent from the order flow at the ask. Empirical studies on high-frequency data show these assumptions to be incorrect (Hasbrouck (2007), Bouchaud et al. (2002, 2008), Andersen et al. (2010)). Figure 3.1 compares the quantiles of the duration between order book events for CitiGroup stock on June 26, 2008 to those of an exponential distribution with the same mean, showing that the empirical distribution of durations is far from being exponential. Figure 3.9 shows the autocorrelation function of the inverse durations: the persistent positive value of this autocorrelation shows that durations may not be assumed to be independent. Finally, as shown in

Figure 3.2 which displays the (positive or negative) changes in queue size induced by successive orders for CitiGroup shares, there is considerable heterogeneity in sizes and clustering in the timing of orders.

Other, more complex, statistical models for order book dynamics have been developed to take these properties into account (see Section 3.2.3). However, only models based on Poisson point processes such as Cont et al. (2010b), Cont and de Larrard (2010) have offered so far the analytical tractability necessary when it comes to studying quantities of interest such as durations or transition probabilities of the price, conditional on the state of the order book. It is therefore of interest to know whether the conclusions based on Markovian models are robust to a departure from these simplifying assumptions and, if not, how they must be modified in the presence of other distributional features and dependence in durations and order sizes.

The goal of this work is to show that it is indeed possible to restore analytical tractability without imposing restrictive assumptions on the order arrival process, by exploiting the *separation of time scales* involved in the problem. The existence of widely different time scales, from milliseconds to minutes, makes it possible to obtain meaningful results from an asymptotic analysis of order book dynamics using a *diffusion approximation* of the limit order book. We argue that this diffusion approximation provides relevant and computationally tractable approximations of the quantities of interest in liquid markets where order arrivals are frequent.

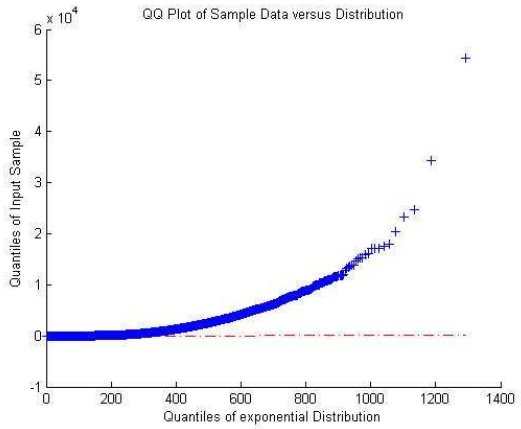

Figure 3.1: Quantiles of inter-event durations compared with quantiles of an exponential distribution with the same mean (Citigroup, June 2008). The dotted line represents the benchmark case where the observations are exponentially distributed, which is clearly not the case here.

Regime	Time scale	Issues
Ultra-high frequency (UHF)	$\sim 10^{-3} - 0.1$ s	Microstructure, Latency
High Frequency (HF)	$\sim 1 - 100$ s	Trade execution
"Daily"	$\sim 10^3 - 10^4$ s	Trading strategies, Option hedging

Table 3.1: A hierarchy of time scales.

As shown in Table 3.1, most applications involve the behavior of prices over time scales an order of magnitude larger than the typical inter-event duration: for example, in optimal trade execution the benchmark is the Volume weighted average price (VWAP) computed over a period which may

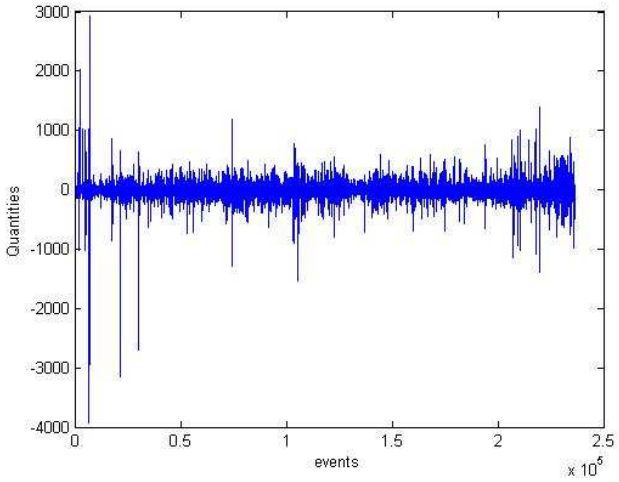

Figure 3.2: Number of shares per event for events affecting the ask. The stock is Citigroup on the 26th of June 2008

	Average no. of orders in 10s	Price changes in 1 day
Citigroup	4469	12499
General Electric	2356	7862
General Motors	1275	9016

Table 3.2: Average number of orders in 10 seconds and number of price changes (June 26th, 2008).

range from 10 minutes to a day: over such time scales much of the microstructural details of the market are averaged out. Second, as noted in Table 3.2, in liquid equity markets the number of events affecting the state of the order book over such time scales is quite large, of the order of hundreds or thousands. The typical duration τ_L (resp. τ_M) between limit orders (resp. market orders and cancelations) is typically $0.001 - 0.01 \ll 1$ (in seconds). These observations show that it is relevant to consider *heavy-traffic limits* in which the rate of arrival of orders is large for studying the dynamics of order books in liquid markets.

In this limit, the complex dynamics of the discrete queueing system is approximated by a simpler system with a continuous state space, which can be either described by a system of ordinary differential equations (in the 'fluid limit', where random fluctuations in queue size vanish) or a system of stochastic differential equations (in the 'diffusion limit' where random fluctuations dominate) (Iglehart and Whitt (1970), Harrison and Nguyen (1993), Whitt (2002)). Intuitively, the fluid limit corresponds to the regime of law of large numbers, where random fluctuations average out and the limit is described by average queue size, whereas the diffusion limit corresponds to the regime of the central limit theorem, where fluctuations in queue size are asymptotically Gaussian. When order sizes or durations fail to have finite moments of first or second order, other scaling limits may intervene, involving Lévy processes (see Whitt (2002)) or fractional Brownian motion Araman and Glynn (2011). As shown by Dai and Nguyen (1994), there are also cases where such a 'heavy traffic limit' may fail to exist. The relevance of each of these asymptotic regimes is, of course, not a matter of 'taste' but an empirical question which depends on the behavior of high-frequency order flow in these markets.

Using empirical data on US stocks, we argue that for most liquid stocks, while the rate of arrival of

market orders and limit orders is large, the *imbalance* between limit orders, which increase queue size, and market orders and cancels, which decrease queue size, is an order of magnitude smaller: over, say, a 10 minute interval, one observes an imbalance ranging from 1 to 10 % of order flow. In other words, over a time scale of several minutes, a large number N of events occur, but the bid/ask imbalance accumulating over the same interval is of order $\sqrt{N} \ll N$. In this regime, random fluctuations in queue sizes cannot be ignored and it is relevant to consider the *diffusion limit* of the limit order book.

In this chapter we study the behavior of a limit order book in this diffusion limit: we prove a functional central limit theorem for the joint dynamics of the bid and ask queues when the intensity of orders becomes large, and use it to derive an analytically tractable jump-diffusion approximation. More precisely, we show that under a wide range of assumptions, which are shown to be plausible for empirical data on liquid US stocks, the intraday dynamics of the limit order book behaves like as a planar Brownian motion in the interior of the positive orthant, and jumps to the interior of the orthant at each hitting time of the boundary.

This jump-diffusion approximation allows various quantities of interest to be computed analytically: we obtain analytical expressions for various quantities such as the probability that the price will increase at the next price change, and the distribution of the duration between price changes, conditional on the state of the order book.

Our results extend previous analysis of heavy traffic limits for such auction processes (Kruk (2003), Bayraktar et al. (2006), Cont and de Larrard (2010)) to a setting which is relevant and useful for quantitative modeling of limit order books and provide a foundation for recently proposed diffusion models for order book dynamics Avellaneda et al. (2011).

Outline. The chapter is organized as follows. Section 3.2 describes a general framework for the dynamics of a limit order book; various examples of models studied in the literature are shown to fall within this modeling framework (Section 3.2.3). Section 3.3 reviews some statistical properties of high frequency order flow in limit order markets: these properties highlight the complex nature of the order flow and motivate the statistical assumptions used to derive the diffusion limit. Section 3.4 contains our main result: Theorem 3.2 shows that, in a limit order market where orders arrive at high frequency, the bid and ask queues behaves like a Markov process in the positive quadrant which diffuses inside the quadrant and jumps to the interior each time it hits the boundary. We provide a complete description of this process, and use it to derive, in Section 3.4.3, a simple jump-diffusion approximation for the joint dynamics of bid and ask queues, which is easier to study and simulate than the initial queueing system.

In particular, we show that in this asymptotic regime the price process is characterized as a piecewise constant process whose transition times correspond to hitting times of the axes by a two dimensional Brownian motion in the positive orthant (Proposition 3.1). This result allows to study analytically various quantities of interest, such as the distribution of the duration between price moves and the probability of an increase in the price: this is discussed in Section 3.5.

3.2 A model for the dynamics of a limit order book

3.2.1 Reduced-form representation of a limit order book

Empirical studies of limit order markets suggest that the major component of the order flow occurs at the (best) bid and ask price levels (see e.g. Biais et al. (1995)). All electronic trading venues also allow to place limit orders *pegged* to the best available price (National Best Bid Offer, or NBBO); market makers used these *pegged* orders to liquidate their inventories. Furthermore, studies on the price impact of order book events show that the net effect of orders on the bid and ask queue sizes is the main factor driving price variations (Cont et al. (2010a)). These observations, together with the fact that queue sizes at the best bid and ask of the consolidated order book are more easily obtainable (from records on trades and quotes) than information on deeper levels of the order book, motivate a

reduced-form modeling approach in which we represent the state of the limit order book by

- the bid price s_t^b and the ask price s_t^a
- the size of the bid queue q_t^b representing the outstanding limit buy orders at the bid, and
- the size of the ask queue q_t^a representing the outstanding limit sell orders at the ask

Figure 3.3 summarizes this representation.

If the stock is traded in several venues, the quantities q^b and q^a represent the best bids and offers in the *consolidated* order book, obtained by aggregating over all (visible) trading venues. At every time t, q_t^b (resp. q_t^a) corresponds to all visible orders available at the bid price s_t^b (resp. s_t^a) across all exchanges.

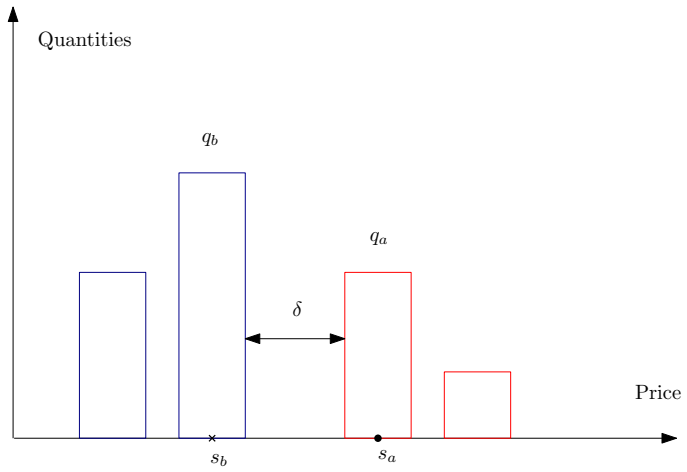

Figure 3.3: Simplified representation of a limit order book.

The state of the order book is modified by *order book events*: limit orders (at the bid or ask), market orders and cancelations (see Cont et al. (2010b,a), Smith et al. (2003)). A limit buy (resp. sell) order of size x increases the size of the bid (resp. ask) queue by x, while a market buy (resp. sell) order decreases the corresponding queue size by x. Cancelation of x orders in a given queue reduces the queue size by x. Given that we are interested in the queue sizes at the best bid/ask levels, market orders and cancelations have the same effect on the queue sizes (q_t^b, q_t^a).

The bid and ask prices are multiples of the tick size δ. When either the bid or ask queue is depleted by market orders and cancelations, the price moves up or down to the next level of the order book. The price processes s_t^b, s_t^a are thus piecewise constant processes whose transitions correspond to hitting times of the axes $\{(0,y), y>0\} \cup \{(x,0), x>0\}$ by the process $q_t = (q_t^b, q_t^a)$.

If the order book contains no 'gaps' (empty levels), these price increments are equal to one tick:

- when the bid queue is depleted, the (bid) price decreases by one tick.
- when the ask queue is depleted, the (ask) price increases by one tick.

If there are gaps in the order book, this results in 'jumps' (i.e. variations of more than one tick) in the price dynamics. We will ignore this feature in what follows but it is not hard to generalize our results to include it.

The quantity $s_t^a - s_t^b$ is the *bid-ask spread*, which may be one or several ticks. As shown in Table 3.3, for liquid stocks the bid-ask spread is equal to one tick for more than 98% of observations.

When either the bid or ask queue is depleted, the bid-ask spread widens immediately to more than one tick. Once the spread has increased, a flow of limit sell (resp. buy) orders quickly fills the gap

Bid-ask spread	1 tick	2 tick	$\geq$ 3 tick
Citigroup	98.82	1.18	0
General Electric	98.80	1.18	0.02
General Motors	98.71	1.15	0.14

Table 3.3: Percentage of observations with a given bid-ask spread (June 26th, 2008).

and the spread reduces again to one tick. When a limit order is placed inside the spread, all the limit orders pegged to the NBBO price move in less than a millisecond to the price level corresponding to this new order. Once this happens, both the bid price and the ask price have increased (resp. decreased) by one tick.

The histograms in Figure 3.4 show that this 'closing' of the spread takes place very quickly: as shown in Figure 3.4 (left) the lifetime of a spread larger than one tick is of the order of a couple of milliseconds, which is negligible compared to the lifetime of a spread equal to one tick (Figure 3.4 , right). In our model we assume that the second step occurs *infinitely fast*: once the bid-ask spread widens, it returns immediately to one tick. For the example of Dow Jones stocks (Figure 3.4), this is a reasonable assumption since the widening of the spread lasts only a few milliseconds. This simply means that we are not trying to describe/model how the orders flow inside the bid-ask spread at the millisecond time scale and, when we describe the state of the order book *after* a price change we have in mind the state of the order book *after* the bid-ask spread has returned to one tick.

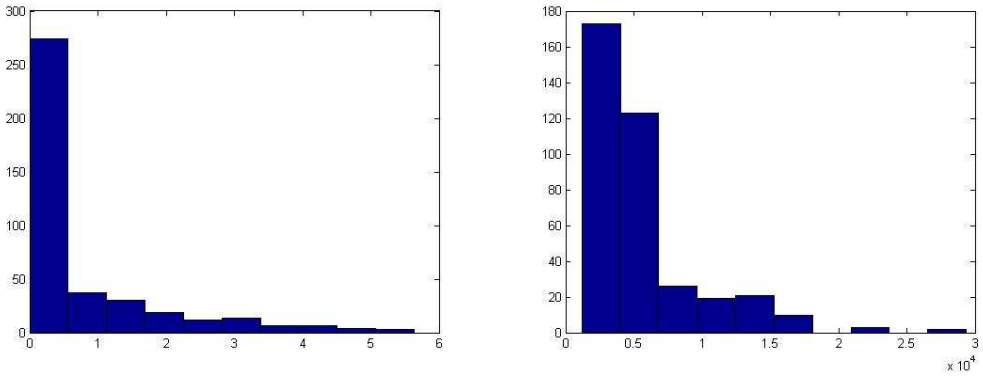

Figure 3.4: Left: Average lifetime, in milliseconds of a spread larger than one tick for Dow Jones stocks. Right: Average lifetime, in milliseconds of a spread equal to one tick.

Under this assumption, each time *one* of the queues is depleted, *both* the bid queue and the ask queue move to a new position and the bid-ask spread remains equal to one tick after the price change. Thus, under our assumptions the bid-ask spread is equal to one tick, i.e. $s_t^a = s_t^b + \delta$, resulting in a further reduction of dimension in the model.

Once either the bid or the ask queue are depleted, the bid and ask queues assume new values. Instead of keeping track of arrival, cancelation and execution of orders at *all* price levels (as in Cont et al. (2010b), Smith et al. (2003)), we treat the queue sizes after a price change as a stationary sequence of random variables whose distribution represents the depth of the order book in a statistical sense. More specifically, we model the size of the bid and ask queues after a price increase by a stationary sequence $(R_k)_{k\geq 1}$ of random variables with values in $\mathbb{N}^2$. Similarly, the size of the bid and ask queues after a price decrease is modeled by a stationary sequence $(\tilde{R}_k)_{k\geq 1}$ of random variables with values in

$\mathbb{N}^2$. The sequences $(R_k)_{k\geq 1}$ and $(\tilde{R}_k)_{k\geq 1}$ summarize the interaction of the queues at the best bid/ask levels with the rest of the order book, viewed here as a 'reservoir' of limit orders.

The variables R_k (resp. $\tilde{R}_k$) have a common distribution which represents the *depth* of the order book after a price increase (resp. decrease): Figure 3.5 shows the (joint) empirical distribution of bid and ask queue sizes after a price move for Citigroup stock on June 26th 2008.

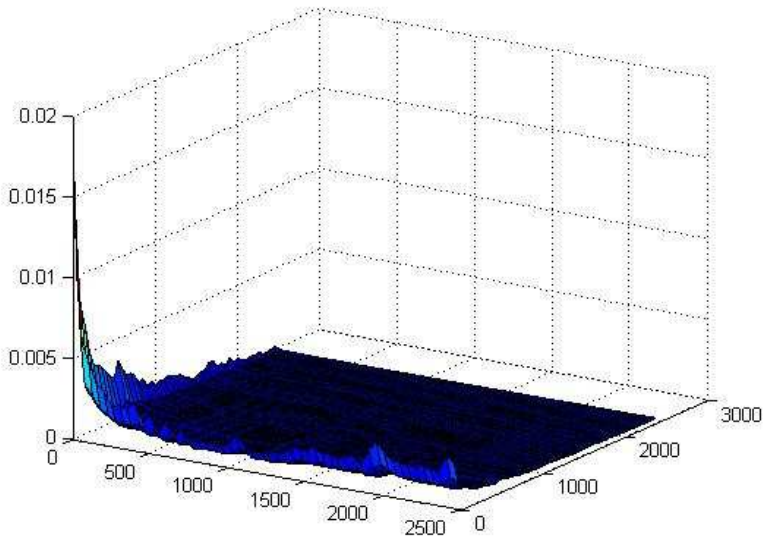

Figure 3.5: Joint density of bid and ask queue sizes after a price move (Citigroup, June 26th 2008).

The simplest specification could be to take $(R_k)_{k\geq 1}$, $(\tilde{R}_k)_{k\geq 1}$ to be IID sequences; this approach, used in Cont and de Larrard (2010), turns out to be good enough for many purposes. But this IID assumption is not necessary; in the next section we will see more general specifications which allow for serial dependence.

In summary, state of the limit order book is thus described by a continuous-time process (s_t^b, q_t^b, q_t^a) which takes values in the discrete state space $\delta\mathbb{Z} \times \mathbb{N}^2$, with piecewise constant sample paths whose transitions correspond to the order book events. Denoting by $(t_i^a, i \geq 1)$ (resp. t_i^b) the event times at the ask (resp. the bid), V_i^a (resp. V_i^b) the corresponding change in ask (resp. bid) queue size, and $k(t)$ the number of price changes in $[0, t]$, the above assumptions translate into the following dynamics for (s_t^b, q_t^b, q_t^a):

- If an order or cancelation of size V_i^a arrives on the ask side at $t = t_i^a$,
  - if $q_{t-}^a + V_i^a \geq 0$, the order can be satisfied without changing the price;
  - if $q_{t-}^a + V_i^a < 0$, the ask queue is depleted, the price increases by one 'tick' of size δ, and the queue sizes take new values $R_{k(t)} = (R_{k(t)}^b, R_{k(t)}^a)$,

$$(s_t^b, q_t^b, q_t^a) = (s_{t-}^b, q_{t-}^b, q_{t-}^a + V_i^a)1_{\{q_{t-}^a \geq -V_i^a\}} + (s_{t-}^b + \delta, R_{k(t)}^b, R_{k(t)}^a)1_{\{q_{t-}^a < -V_i^a\}}, \tag{3.1}$$

- If an order or cancelation of size V_i^b arrives on the bid side at $t = t_i^b$,
  - if $q_{t-}^b + V_i^b \geq 0$, the order can be satisfied without changing the price;
  - if $q_{t-}^b + V_i^b < 0$, the bid queue gets depleted, the price decreases by one 'tick' of size δ and the queue sizes take new values $\tilde{R}_{k(t)} = (\tilde{R}_{k(t)}^b, \tilde{R}_{k(t)}^a)$:

$$(s_t^b, q_t^b, q_t^a) = (s_{t-}^b, q_{t-}^b + V_i^b, q_{t-}^a)1_{\{q_{t-}^b \geq -V_i^b\}} + (s_{t-}^b - \delta, \tilde{R}_{k(t)}^b, \tilde{R}_{k(t)}^a)1_{\{q_{t-}^b < -V_i^b\}}. \quad (3.2)$$

3.2.2 The limit order book as a 'regulated' process in the orthant

As in the case of reflected processes arising in queueing networks, the process $q_t = (q_t^b, q_t^a)$ may be constructed from the **net order flow process**

$$x_t = (x_t^b, x_t^a) = \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a \right)$$

where N_t^b (resp. N_t^a) is the number of events (i.e. orders or cancelations) occurring at the bid (resp. the ask) during $[0,t]$. $x_t = (x_t^b, x_t^a)$ is analogous to the 'net input' process in queuing systems Whitt (2002): x_t^b (resp. x_t^a) represents the cumulative sum of all orders and cancelations at the bid (resp. the ask) between 0 and t.

$q = (q_t^b, q_t^a)_{t \geq 0}$ which takes values in the positive orthant, may be constructed from x by reinitializing its value to a a new position inside the positive orthant according to the rules (3.1)–(3.2) each time one of the queues is depleted: every time $(q_t)_{t \geq 0}$ attempts to exit the positive orthant, it jumps to a a new position inside the orthant, taken from the sequence $(R_n, \tilde{R}_n)$.

This construction may be done path by path, as follows:

Definition 3.1. *Let $\omega \in D([0,\infty), \mathbb{R}^2)$ be a right-continuous function with left limits (i.e. a cadlag function), $R = (R_n)_{n \geq 1}$ and $\tilde{R} = (\tilde{R}_n)_{n \geq 1}$ two sequences with values in $\mathbb{R}_+^2$. There exists a unique cadlag function $\Psi(\omega, R, \tilde{R}) \in D([0,\infty), \mathbb{R}_+^2)$ such that*

- *For $t < \tau_1$, let $\Psi(\omega, R, \tilde{R})(t) = \omega(t)$ where*

$$\tau_1 = \inf\{t \geq 0,\ \omega(t).(1,0) < 0 \text{ or } \omega(t).(0,1) < 0\}.$$

  is the first exit time of ω from the positive orthant.

- *$\Psi(\omega, R, \tilde{R})(\tau_1) = R_1$ if $\Psi(\omega, R, \tilde{R})(\tau_1 -).(0,1) < 0$, and $\Psi(\omega, R, \tilde{R})(\tau_1) = \tilde{R}_1$ if $\Psi(\omega, R, \tilde{R})(\tau_1 -).(1,0) < 0$.*

- *For $k \geq 1$,*
  $\Psi(\omega, R, \tilde{R})(t + \tau_k) = \Psi(\omega, R, \tilde{R})(\tau_k) + \omega(t + \tau_k) - \omega(\tau_k)$ for $0 \leq t < \tau_{k+1} - \tau_k$, where

$$\tau_{k+1} = \inf\{t \geq \tau_{k-1},\ \Psi(\omega, R, \tilde{R})(\tau_k) + \omega(t + \tau_k) - \omega(\tau_k) \notin \mathbb{R}_+^2\}$$

  is the first exit time of $(\Psi(\omega, R, \tilde{R})(t), t \geq \tau_k)$ from the positive orthant.

- *$\Psi(\omega, R, \tilde{R})(\tau_k) = R_k$ if $\Psi(\omega, R, \tilde{R})(\tau_k -).(0,1) < \Psi(\omega, R, \tilde{R})(\tau_{k-1}).(1,0)$ and $\Psi(\omega, R, \tilde{R})(\tau_k) = \tilde{R}_k$ otherwise.*

The path $\Psi(\omega, R, \tilde{R})$ is obtained by "regulating" the path ω with the sequences $(R, \tilde{R})$: in between two exit times, the increments of $\Psi(\omega, R, \tilde{R})$ follow those of ω and each time the process attempts to exit the positive orthant by crossing the x-axis (resp. the y-axis), it jumps to a a new position inside the orthant, taken from the sequence $(R_n)_{n \geq 1}$ (resp. from the sequence $(\tilde{R}_n)_{n \geq 1}$).

Unlike the more familiar case of a continuous reflection at the boundary, which arises in heavy-traffic limits of multiclass queueing systems (see Harrison (1978), Harrison and Nguyen (1993), Whitt (2002), Ramanan and Reiman (2003) for examples), this construction introduces a discontinuity by pushing the process into the interior of the positive orthant each time it attempts to exit from the axes.

To study the continuity properties of this map, we endow $D([0,\infty),\mathbb{R}^2)$ with Skorokhod's J_1 topology Billingsley (1968), Lindvall (1973) and the set $\left(\mathbb{R}_+^2\right)^{\mathbb{N}}$ with the topology induced by 'cylindrical' semi-norms, defined as follows: for a sequence $(R^n)_{n\geq 1}$ in $\left(\mathbb{R}_+^2\right)^{\mathbb{N}}$

$$R^n \overset{n\to\infty}{\to} R \in \left(\mathbb{R}_+^2\right)^{\mathbb{N}} \iff \left(\forall k \geq 1, \quad \sup\{|R_1^n - R_1|, ..., |R_k^n - R_k|\} \overset{n\to\infty}{\to} 0\right).$$

$D([0,\infty),\mathbb{R}^2) \times \left(\mathbb{R}_+^2\right)^{\mathbb{N}} \times \left(\mathbb{R}_+^2\right)^{\mathbb{N}}$ is then endowed with the corresponding product topology.

Theorem 3.1. *Let $R = (R_n)_{n\geq 1}, \tilde{R} = (\tilde{R}_n)_{n\geq 1}$ be sequences in $]0,\infty[\times]0,\infty[$ which do not have any accumulation point on the axes. If $\omega \in C^0([0,\infty),\mathbb{R}^2)$ is such that*

$$(0,0) \notin \Psi(\omega, R, \tilde{R})([0,\infty)\ ). \tag{3.3}$$

Then the map

$$\Psi : D([0,\infty),\mathbb{R}^2) \times \left(\mathbb{R}_+^2\right)^{\mathbb{N}} \times \left(\mathbb{R}_+^2\right)^{\mathbb{N}} \quad \to \quad D([0,\infty),\mathbb{R}_+^2) \tag{3.4}$$

is continuous at $(\omega, R, \tilde{R})$.

Proof: See Section 3.6.2 in the Appendix.

This construction may be applied to any cadlag stochastic process: given a cadlag process X with values in $\mathbb{R}^2$ and (random) sequences $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R}_n)_{n\geq 1}$ with values in $\mathbb{R}_+^2$, the process $\Psi(X, R, \tilde{R})$ is a cadlag process with values in $\mathbb{R}_+^2$.

It is easy to see that the order book process $q_t = (q_t^b, q_t^a)$ may be constructed by this procedure:

Lemma 3.1. *The queue size process $q = (q_t^b, q_t^a)_{tgeq0}$ is related to the net order flow by*

$$q = (q^b, q^a) = \Psi(x, R, \tilde{R})$$

where

- $x_t = (x_t^b, x_t^a) = \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a\right)$ *is the net order flow at the bid and the ask,*
- $R = (R_n)_{n\geq 1}$ *is the sequence of queue sizes after a price increase, and*
- $\tilde{R} = (\tilde{R}_n)_{n\geq 1}$ *is the sequence of queue sizes after a price decrease.*

One can thus build a statistical model for the limit order book by specifying the joint law of x and of the regulating sequences $(R, \tilde{R})$. This approach simplifies the study of the (asymptotic) properties of $q_t = (q_t^b, q_t^a)$.

Example 3.1 (IID reinitializations)**.** *The simplest case is the case where the queue length after each price change is independent from the history of the order book, as in Cont and de Larrard (2010). $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R}_n)_{n\geq 1}$ are then IID sequences with values in $]0,\infty[^2$. Figure 3.5 shows an example of such a distribution for a liquidly traded stock (NYSE: CitiGroup).*

The law of the process $Q = \Psi(x, R, \tilde{R})$ is then entirely determined by the law of the net order flow x and the distributions of R_n, $\tilde{R}_n$: it can be constructed from the concatenation of the laws of $(x_t, \tau_k \leq t < \tau_{k+1})$ for $k \geq 0$ (where we define $\tau_0 := 0$).

Example 3.2 (Pegged limit orders)**.** *Most electronic trading platforms allow to place limit orders which are pegged to the best quote: if the best quote moves to a new price level, a pegged limit order moves along with it to the new price level. The presence of pegged orders leads to positive autocorrelation and dependence in the queue size before/after a price change. The queue size after a price change may be modeled as*

- $q_{\tau_n} = R_n = (\epsilon_n^b + \beta q_{\tau_n -}^b, \epsilon_n^a)$ *if the price has increased, and*

- $q_{\tau_n} = \tilde{R_n} = (\tilde{\epsilon}^b_n, \tilde{\epsilon}^a_n + \tilde{\beta} q^a_{\tau_n -})$ *if the price has decreased*

where $\epsilon_n = (\epsilon^b_n, \epsilon^a_n), \tilde{\epsilon}_n = (\tilde{\epsilon}^b_n, \tilde{\epsilon}^a_n)$ *are IID sequences. Empirically, one observes a correlation of* $\sim 10\% - 20\%$ *between the queue lengths before and after a price change, which suggests an order magnitude for the fraction of pegged orders.*

As in the previous example, the law of of the process $q = \Psi(x, R, \tilde{R})$ *is determined by the law of the net order flow* x*, the coefficients* $\beta, \tilde{\beta}$ *and the distributions of* ϵ,$\tilde{\epsilon}$*: it can be constructed from the concatenation of the laws of* $(x_t, \tau_k \leq t < \tau_{k+1})$ *for* $k \geq 0$*.*

More generally, one could consider other extensions where the queue size after a price move may depend in a (nonlinear) way on the queue size before the price move and a random term ϵ_n representing the inflow of new orders after the n-th price change:

$$q_{\tau_n} = g(q_{\tau_n -}, \epsilon_n). \tag{3.5}$$

The results given below hold for this general specification although the examples 3.1 and 3.2 above are sufficiently general for most applications.

3.2.3 Examples

The framework described in Section 3.2.1 allows a wide class of specifications for the order flow process, and contains as special cases various models proposed in the literature. Each model involves a specification for the (random) sequences $(t^a_i, t^b_i, V^a_i, V^b_i)_{i\geq 1}$, $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R_n})_{n\geq 1}$ or, equivalently, $(T^a_i, T^b_i, V^a_i, V^b_i)_{i\geq 1}$, $R = (R_n)_{n\geq 1}$ and $\tilde{R} = (\tilde{R_n})_{n\geq 1}$ where $T^a_i = t^a_{i+1} - t^a_i$ (resp. $T^b_i = t^b_{i+1} - t^b_i$) are the durations between order book events on the ask (resp. the bid) side.

Models based on Poisson point processes

Cont and de Larrard (2010) study a stylized model of a limit order market in which market orders, limit orders and cancelations arrive at independent and exponential times with corresponding rates μ, λ and θ, the process $q = (q^b, q^a)$ becomes a Markov process. If we assume additionally that all orders have the same size, the dynamics of the reduced limit order book is described by:

- The sequence $(T^a_i)_{i\geq 0}$ is a sequence of independent random variables with exponential distribution with parameter $\lambda + \theta + \mu$,
- The sequence $(T^b_i)_{i\geq 0}$ is a sequence of independent random variables with exponential distribution with parameter $\lambda + \theta + \mu$,
- The sequence $(V^a_i)_{i\geq 0}$ is a sequence of independent random variables with

$$\mathbb{P}[V^a_i = 1] = \frac{\lambda}{\lambda + \mu + \theta} \quad \textit{and} \quad \mathbb{P}[V^a_i = -1] = \frac{\mu + \theta}{\lambda + \mu + \theta},$$

- The sequence $(V^b_i)_{i\geq 0}$ is a sequence of independent random variables with

$$\mathbb{P}[V^b_i = 1] = \frac{\lambda}{\lambda + \mu + \theta} \quad \textit{and} \quad \mathbb{P}[V^b_i = -1] = \frac{\mu + \theta}{\lambda + \mu + \theta}.$$

- All these sequences are independent.

It is readily verified that this model is a special case of the framework of Section 3.2.1: $(q_t)_{t\geq 0}$ may be constructed as in Definition 3.1, where the unconstrained process x_t is now a compound Poisson process.

Self-exciting point processes

Empirical studies of order durations highlight the dependence in the sequence of order durations. This feature, which is not captured in models based on Poisson processes, may be adequately represented by a multidimensional *self-exciting* point process Andersen et al. (2010), Hautsch (2004), in which the arrival rate $\lambda_i(t)$ of an order of type i is represented as a stochastic process whose value depends on the recent history of the order flow: each new order increases the rate of arrival for subsequent orders of the same type (self-exciting property) and may also affect the rate of arrival of other order types (mutually exciting property):

$$\lambda_i(t) = \theta_i + \sum_{j=1}^{J} \delta_{ij} \int_0^t e^{-\kappa_i(t-s)} dN_j(s)$$

Here δ_{ij} measures the impact of events of type j on the rate of arrival of subsequent events of type i: as each event of type j occurs, λ_i increases by δ_{ij}. In between events, $\lambda_i(t)$ decays exponentially at rate κ_i. Maximum likelihood estimation of this model on TAQ data Andersen et al. (2010) shows evidence of self-exciting and mutually exciting features in order flow: the coefficients δ_{ij} are all significantly different from zero and positive, with $\delta_{ii} > \delta_{ij}$ for $j \neq i$.

Autoregressive conditional durations

Models based on Poisson process fail to capture serial dependence in the sequence of durations, which manifests itself in the form of clustering of order book events. One approach for incorporating serial dependence in event durations is to represent the duration T_i between transactions $i-1$ and i as

$$T_i = \psi_i \epsilon_i,$$

where $(\epsilon_i)_{i\geq 1}$ is a sequence of independent positive random variables with common distribution and $\mathbb{E}[\epsilon_i] = 1$ and the *conditional duration* $\psi_i = \mathbb{E}[T_i | \psi_{i-j}, T_{i-j}, j \geq 1]$ is modeled as a function of past history of the process:

$$\psi_i = G(\psi_{i-1}, \psi_{i-2}, ..., ..; T_{i-1}, T_{i-2}, ..., ..).$$

Engle and Russell's Autoregressive Conditional Duration model Engle and Russell (1998) propose an ARMA(p, q) representation for G:

$$\psi_i = a_0 + \sum_{i=1}^{p} a_k \psi_{i-k} + \sum_{i=1}^{q} b_q T_{j-k}$$

where $(a_0, ..., a_p)$ and $(b_1, ..., b_q)$ are positive constants. The ACD-GARCH model Ghysels and Jasiak (1998) combine this model with a GARCH model for the returns. Engle (2000) proposes a GARCH-type model with random durations where the volatility of a price change may depend on the previous durations. Variants and extensions are discussed in Hautsch (2004). Such models, like ARMA or GARCH models defined on fixed time intervals, have likelihood functions which are numerically computable. Although these references focus on transaction data, the framework can be adapted to model the durations $(T_i^a, i \geq 1)$ and $(T_i^b, i \geq 1)$ between order book events with the ACD framework (Hautsch 2004).

A limit order market with patient and impatient agents

Another way of specifying a stochastic model for the order flow in a limit order market is to use an 'agent-based' formulation where agent types are characterized in terms of the statistical properties of the order flow they generate. Consider for example a market with three types of traders:

- *impatient* traders who only submit market orders:

- *patient* traders who use *only limit orders*: this is the case for example of traders who place stop loss orders or engage in strategies such as mean-reversion arbitrage or pairs trading which are only profitable with limit orders.
- other traders who use *both limit and market orders*; we will assume these traders submit a proportion γ of their orders as limit orders and $(1-\gamma)$ as market orders, where $0 < \gamma < 1$.

Denote by m (resp. l) the proportion of orders generated by impatient (resp. patient) traders:

$$\begin{aligned} \forall i \geq 1, \quad \mathbb{P}[i-\text{th trader uses only market orders}] &= m, \\ \mathbb{P}[i-\text{th trader uses only limit orders}] &= l, \\ \mathbb{P}[i\text{th trader uses both limit and market orders}] &= 1-l-m. \end{aligned}$$

Assume that the sequence $(T_i, i \geq 1)$ of duration between consecutive orders is a stationary ergodic sequence of random variables with $\mathbb{E}[T_i] < \infty$, that each trader has an equal chance of being a buyer or a seller and that the type of trader (buyer or seller) is independent from the past:

$$\mathbb{P}[i-\text{th trader is a buyer}] = \mathbb{P}[i-\text{th trader is a seller}] = \frac{1}{2}$$

Trader i generates an order of size V_i, where $(V_i, i \geq 1)$ is an IID sequence with:

$$\begin{aligned} \mathbb{P}[(V_i^b, V_i^a) = (V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, V_i)] &= \frac{m}{2}, \\ \mathbb{P}[(V_i^b, V_i^a) = (-V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, -V_i)] &= \frac{l}{2}, \\ \mathbb{P}[(V_i^b, V_i^a) = (\gamma V_i, -(1-\gamma)V_i)] = \mathbb{P}[(V_i^b, V_i^a) = (-(1-\gamma)V_i, \gamma V_i)] &= \frac{1-l-m}{2}. \end{aligned}$$

3.3 Statistical properties of high-frequency order flow

As described in Section 3.2.1, the sequence of order book events –the *order flow*– is characterized by the sequences $(T_i^a, i \geq 1)$ and $(T_i^b, i \geq 1)$ of durations between orders and the sequences of order sizes $(V_i^b, i \geq 1)$ and $(V_i^a, i \geq 1)$. In this section we illustrate the statistical properties of these sequences using high-frequency quotes and trades for liquid US stocks –CitiGroup, General Electric, General Motors– on June 26th, 2008.

3.3.1 Order sizes

Empirical studies Bouchaud et al. (2002, 2008), Gopikrishnan et al. (2000), Maslov and Mills (2001) have shown that order sizes are highly heterogeneous and exhibit heavy-tailed distributions, with Pareto-type tails:

$$\mathbb{P}(V_i^a \geq x) \sim Cx^{-\beta}$$

with tail exponent $\beta > 0$ between 2 and 3, which corresponds to a series with finite variance but infinite moments of order ≥ 3. The tail exponent $\beta > 0$ is difficult to estimate precisely, but the Hill estimator Resnick (2006) can be used to measure the heaviness of the tails. Table 3.4 gives the Hill estimator of the tail coefficient of order sizes for our samples. This estimator is larger than 2 for both the bid and the ask; this means that the sequence of order sizes have a finite moment of order two.

The sequences of order sizes $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ exhibit insignificant autocorrelation, as observed on Figure 3.6. However, they are far from being independent: the series of squared order sizes $((V_i^b)^2, i \geq 1)$ and $((V_i^a)^2, i \geq 1)$ are positively autocorrelated, as shown in Figure 3.7.

	Bid side	Ask side
Citigroup	[0.42, 0.46]	[0.29, 0.32]
General Electric	[0.42, 0.45]	[0.41, 0.46]
General Motors	[0.36, 0.42]	[0.44, 0.51]

Table 3.4: 95-percent confidence interval of the Hill estimator of the sequence of order sizes. When the Hill estimator is < 0.5, the estimated tail index is large than 2 and the distribution has finite variance.

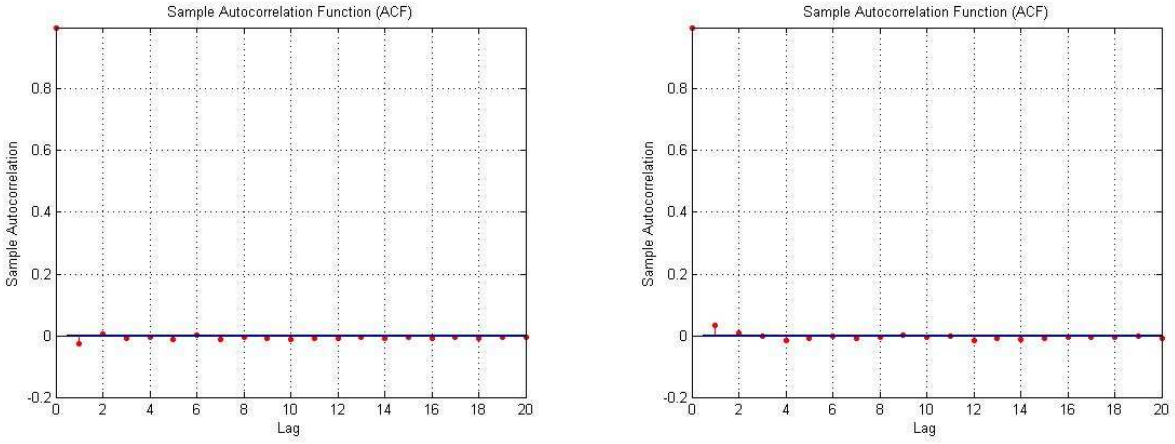

Figure 3.6: Autocorrelogram of the sequence of order sizes. Order coming at the ask on the left and at the bid on the right.

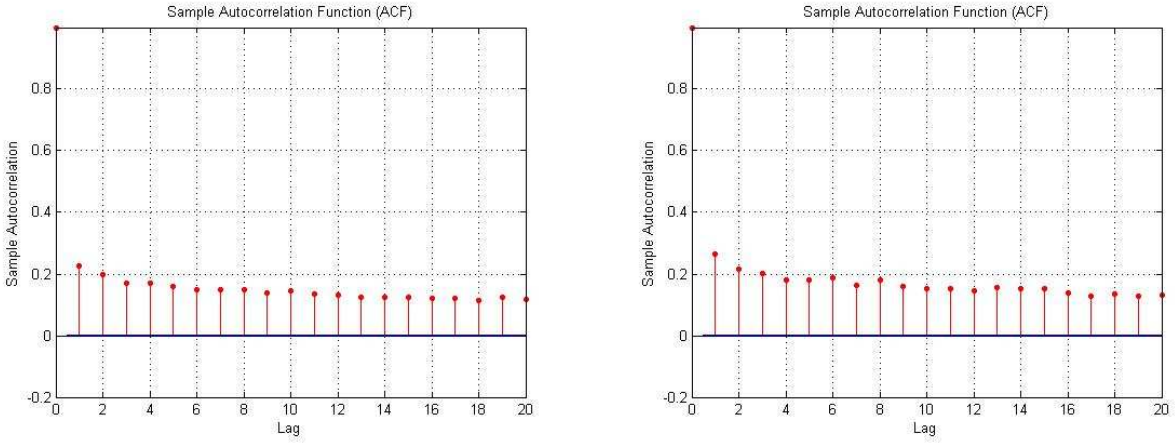

Figure 3.7: Autocorrelogram of the sequence of absolute order sizes. Order coming at the ask on the left and at the bid on the right.

Finally, the sequences $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ may be negatively correlated. This stems from the fact that a buyer can simultaneously use market orders on the ask side (which correspond to negative

values of V_i^a) and limit orders on the bid side (which correspond to positive values of V_i^b); the same argument holds for sellers (see Section 3.2.3).

These properties of the sequence $(V_i^a, V_i^b)_{i\geq 1}$ may be modeled using a bivariate ARCH process:

$$V_i^b = \sigma_i^b z_i^b \qquad V_i^a = \sigma_i^a z_i^a$$

$$(\sigma_i^b)^2 = \alpha_0^b + \alpha_1^b (V_{t-1}^b)^2, \qquad (\sigma_i^a)^2 = \alpha_0^a + \alpha_1^a (V_{t-1}^a)^2, \qquad \text{where} \qquad (z_i^b, z_i^a)_{i\geq 1} \overset{IID}{\sim} N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

and $(\alpha_0^b, \alpha_1^b, \alpha_0^a, \alpha_1^a)$ are positive coefficients satisfying

$$0 < \alpha_0^b + \alpha_1^b < 1, \quad \text{and} \quad 0 < \alpha_0^a + \alpha_1^a < 1. \tag{3.6}$$

As shown by Bougerol and Picard (1992), under the assumption (3.6), the sequence of order sizes $(V_i^b, V_i^a)_{i\geq 1}$ is then a well defined, stationary sequence of random variables with finite second-order moments, satisfying the properties enumerated above.

3.3.2 Durations

The timing of order book events is describe by the sequence of durations $(T_i^b, i \geq 1)$ at the bid and $(T_i^a, i \geq 1)$ at the ask. These sequences have zero autocorrelation (see Figure 3.8) but are not independence sequences: for example, as shown in Figure 3.9, the sequence of inverse durations $(1/T_i^b, i \geq 1)$ and $(1/T_i^b, i \geq 1)$ has significant autocorrelations.

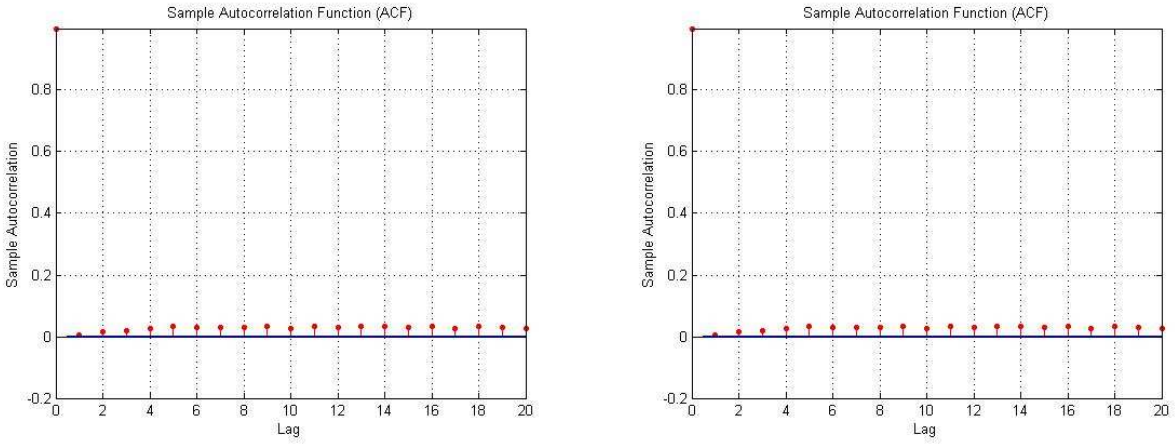

Figure 3.8: Autocorrelogram of the sequence of durations for events at the ask (left) and the bid (right).

Figure 3.10 represents the empirical distribution functions $\mathbb{P}[T^a > u]$ and $\mathbb{P}[T^b > u]$ in logarithmic scale. Both empirical distributions exhibit thin, exponential-type tails (which implies in particular that T^a and T^b have finite expectation).

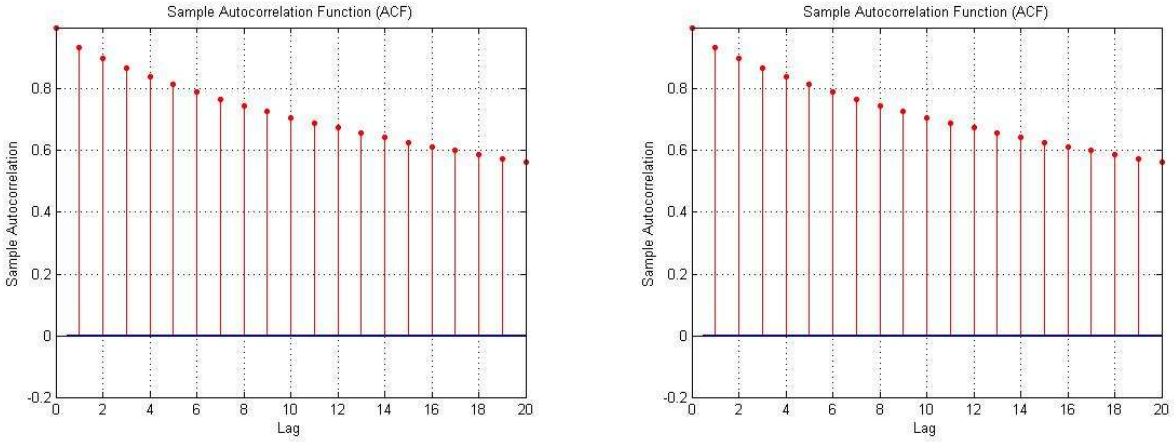

Figure 3.9: Autocorrelogram of the sequence of inverse durations for events at the ask (left) and the bid (right).

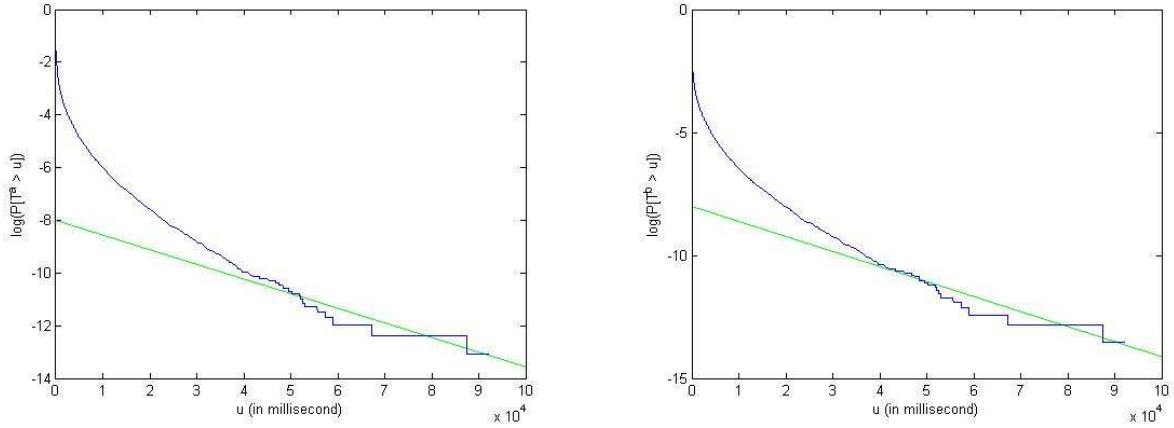

Figure 3.10: Logarithm of the empirical distribution function of durations for events at the ask (left) and the bid (right).

3.4 Heavy Traffic limit

At very high frequency, the limit order book is described by a two-dimensional piecewise constant process $q_t = (q_t^b, q_t^a)_{t\geq 0}$, whose evolution is determined by the flow of orders. The complex nature of this order flow –heterogeneity and serial dependence in order sizes, dependence between orders coming at the ask and at the bid– described in section 3.3, makes it difficult to describe q_t in an analytically tractable manner which would allow the quantities of interest to be computed either in closed form or numerically in real time applications. However, if one is interested in the evolution of the order book over time scales much larger than the interval between individual order book events, the (coarse-grained) dynamics of the queue sizes may be described in terms of a simpler process Q, called the *heavy traffic approximation* of q. In this limit, the complex dynamics of the discrete queueing system is approximated by a simpler system with a continuous state space, which can be either described by a system of ordinary differential equations (in the 'fluid limit', where random fluctuations in queue sizes vanish) or a system of stochastic differential equations (in the 'diffusion limit' where random fluctuations dominate). This idea has been widely used in queueing theory to obtain useful analytical insights into the dynamics of queueing systems Harrison and Nguyen (1993), Iglehart and Whitt (1970), Whitt (2002).

We argue that the heavy traffic limit is highly relevant for the study of limit order books in liquid markets, and that the correct scaling limit for the liquid stocks examined in our data sets is the "diffusion" limit. This *heavy traffic* limit is then derived in Theorem 3.2 and described in Section 3.4.3.

3.4.1 Fluid limit or diffusion limit?

Let $(V_i^{n,a}, i \geq 1)$ the sequence of order sizes, whose properties depend on the index n. One way of viewing the heavy traffic limit is to view the limit order book at a lower time resolution, by grouping together events in batches of size n. Since the inter-event durations are finite, this is equivalent to rescaling time by n. The impact, on the net order flow, of a batch of n events at the ask is

$$\frac{V_1^{n,a} + V_2^{n,a} + V_3^{n,a} + ... + V_n^{n,a}}{\sqrt{n}} = \frac{(V_1^{n,a} - \overline{V^{n,a}}) + (V_2^{n,a} - \overline{V^{n,a}}) + ... + (V_n^{n,a} - \overline{V^{n,a}})}{\sqrt{n}} + \sqrt{n} \quad \overline{V^{n,a}},$$

where $\overline{V^{n,a}} = \mathbb{E}[V_1^{n,a}]$. Under appropriate assumptions (see next section), this sum behaves approximately as a Gaussian random variable for large n:

$$\frac{V_1^{n,a} + V_2^{n,a} + V_3^{n,a} + ... + V_n^{n,a}}{\sqrt{n}} \sim N(\sqrt{n}\ \overline{V^{n,a}}, \mathrm{Var}(V_1^{n,a})) \ \text{ as } \ n \to \infty. \tag{3.7}$$

Two regimes are possible, depending on the behavior of the ratio $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_1^{n,a})}}$ as n grows:

- If $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_i^{n,a})}} \to \infty$ as $n \to \infty$, the correct approximation is given by the *fluid limit*, which describes the (deterministic) behavior of the average queue size.
- If $\lim_{n\to\infty} \frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_i^{n,a})}} < \infty$, the rescaled queue sizes behave like a diffusion process.

The fluid limit corresponds to the regime of law of large numbers, where random fluctuations average out and the limit is described by average queue size, whereas the diffusion limit corresponds to the regime of the (functional) central limit theorem, where fluctuations in queue size are asymptotically Gaussian.

Figure 3.11 displays the histogram of the ratio $\frac{\sqrt{n}\ \overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_i^{n,a})}}$ for stocks in the Dow Jones index, where for each stock n is chosen to represent the average number of order book events in a 10 second interval (typically $n \sim 100 - 1000$). This ratio is shown to be rather small at such intraday time scales,

showing that the diffusion approximation, rather than the fluid limit, is the relevant approximation to use here.

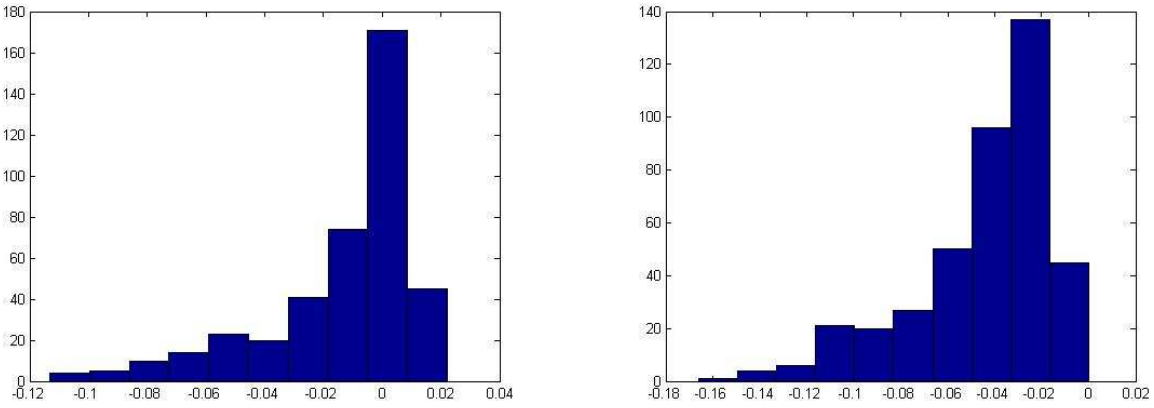

Figure 3.11: Empirical distribution of the ratio $\frac{\sqrt{n}\overline{V^{n,a}}}{\sqrt{\mathrm{Var}(V_1^{n,a})}}$ showing the relative importance of average change vs fluctuations in queue size, for stocks in the Dow-Jones index during June 08 (see Section 3.4.1). Low values of the ratio indicate that intraday changes in bid/ask queue size are dominated by fluctuations, rather than the average motion of the queue. Left: bid side. Right: ask side.

Bid and ask queue sizes (q_t^b, q_t^a) exhibit a diffusion-type behavior at such intraday time scales. Figure 3.12 shows the path of the net order flow process

$$x_t = (q_0^b, q_0^a) + \left(\sum_{i=1}^{N_t^b} V_i^b, \sum_{i=1}^{N_t^a} V_i^a \right) \tag{3.8}$$

sampled every second for CitiGroup stocks on a typical trading day. In this example, for which the average time between consecutive orders is $\lambda^{-1} \simeq 13\ ms \ll 1$ second, we observe that the process X behaves like a diffusion in the orthant with negative drift: the randomness of queue sizes does not average out at this time scale.

We will now show that this is a general result: under mild assumptions on the order flow process, we will show that the (rescaled) queue size process

$$(\frac{q_{nt}^{n,b}}{\sqrt{n}}, \frac{q_{nt}^{n,a}}{\sqrt{n}})_{t \geq 0} \tag{3.9}$$

converges in distribution to a Markov process $(Q_t)_{t \geq 0}$ in the positive orthant, whose features we will now describe in terms of the statistical properties of the order flow.

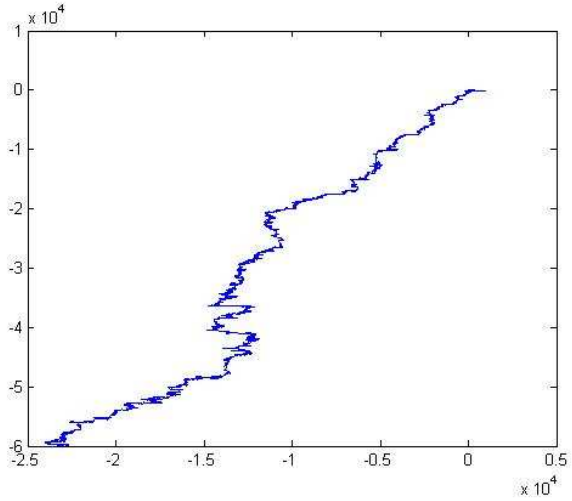

Figure 3.12: Evolution of the net order flow $X_t = (X_t^b, X_t^a)$ given by Eq. (3.8) for CitiGroup shares over one trading day (June 26, 2008). The starting point is taken to be (0,0) at the marke open. Note the irregular, diffusive feature of the path of X and its negative drift.

3.4.2 A functional central limit theorem for the limit order book

Consider now a sequence $q^n = (q_t^n)_{t\geq 0}$ of processes, where q^n represents the dynamics of the bid and ask queues in the limit order book at a time resolution corresponding to n events (see discussion above). The dynamics of q^n is characterized by the sequence of order sizes $(V_i^{n,b}, V_i^{n,a})_{i\geq 1}$, durations $(T_i^{n,b}, T_i^{n,a})_{i\geq 1}$ between orders and the fact that, at each price change

- $q_{\tau_k}^n = R_k^n = g(q_{\tau_k -}^n, \epsilon_k^n)$ if the price has increased, and
- $q_{\tau_k}^n = \tilde{R}_k^n = g(q_{\tau_k -}^n, \tilde{\epsilon}_k^n)$ if the price has decreased,

where $(\epsilon_k^n, k \geq 1)$ is an IID sequence with distribution f_n, and $(\tilde{\epsilon}_k^n, k \geq 1)$ is an IID sequence with distribution $\tilde{f}_n$. Note that this specification includes Examples 3.1 and 3.2 as special cases.

We make the following assumptions, which allow for an analytical study of the heavy traffic limit and are sufficiently general to accommodate high frequency data sets of trades and quotes such as the ones described in Section 3.3:

Assumption 3.1. *$(T_i^{n,a}, T_i^{n,b})_{i\geq 1}$ is a stationary array of positive random variables whose common distribution has a continuous density and satisfies*

$$\lim_{n\to\infty} \frac{T_1^{n,a} + T_2^{n,a} + ... + T_n^{n,a}}{n} = \frac{1}{\lambda^a} < \infty, \qquad \lim_{n\to\infty} \frac{T_1^{n,b} + T_2^{n,b} + ... + T_n^{n,b}}{n} = \frac{1}{\lambda^b} < \infty.$$

λ^a (resp. λ^b) represents the arrival rate of orders at the ask (resp. the bid).

Assumption 3.2. *$(V_i^{n,a}, V_i^{n,b})_{i\geq 1}$ is a stationary, uniformly mixing array of random variables satisfying*

$$\sqrt{n}\mathbb{E}[V_1^{n,a}] \overset{n\to\infty}{\to} \overline{V^a}, \qquad \sqrt{n}\mathbb{E}[V_1^{n,b}] \overset{n\to\infty}{\to} \overline{V^b}, \tag{3.10}$$

$$\lim_{n\to\infty} \mathbb{E}[(V_i^{n,a} - \overline{V^a})^2] + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{n,a}, V_i^{n,a}) = v_a^2 < \infty, \qquad \text{and}$$

$$\lim_{n\to\infty} \mathbb{E}[(V_i^{n,b} - \overline{V^b})^2] + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{n,b}, V_i^{n,b}) = v_b^2 < \infty.$$

The assumption of uniform mixing (Billingsley 1968, Ch. 4) implies that the partial sums of order sizes verify a central limit theorem, but allows for various types of serial dependence in order sizes. The scaling assumptions on the first two moments corresponds to the properties of the empirical data discussed in Section 3.4.1. Under Assumption 4.3, one can define

$$\rho := \lim_{n\to\infty} \frac{1}{v_a v_b} \left(2\max(\lambda^a, \lambda^b)\mathrm{cov}(V_1^{n,a}, V_1^{n,b}) + 2\sum_{i=1}^{\infty} \lambda^a \mathrm{cov}(V_1^{n,a}, V_i^{n,b}) + \lambda^b \mathrm{cov}(V_1^{n,b}, V_i^{n,a}) \right). \tag{3.11}$$

$\rho \in (-1, 1)$ may be interpreted as a measure of 'correlation' between event sizes at the bid and event sizes at the ask.

These assumptions hold for the examples of Section 3.2.3. In the case of the Hawkes model, Assumption 4.2 was shown to hold in Bacry et al. (2010). Also, these assumptions are quite plausible for high frequency quotes for liquid US stocks since, as argued in Section 3.3:

- The tail index of order sizes is larger than two, so the sequences $(V_i^b, i \geq 1)$ and $(V_i^a, i \geq 1)$ have a finite second moment.
- The sequence of order sizes is uncorrelated i.e. has statistically insignificant autocorrelation. Therefore the sum of autocorrelations of order sizes is finite (zero, in fact).
- The sequence of inter-event durations has a finite empirical mean and is not autocorrelated.

These empirical observations support the plausibility of Assumptions 4.2 and 4.3 for the data sets examined.

Assumption 4.3 has an intuitive interpretation: if orders are grouped in batches of n orders, then Assumption 4.3 amounts to stating that the variance of batch sizes should scale linear with n. This assumption can be checked empirically, using a variance ratio test for example: Figure 3.13 shows that this linear relation is indeed verifies for the data sets examined in Section 3.3.

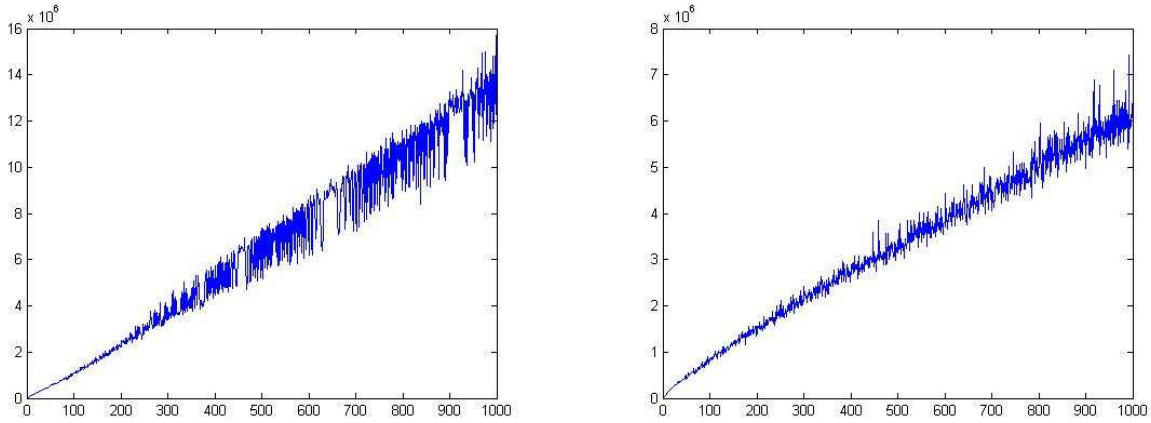

Figure 3.13: Variance of batch sizes of n orders, for General Electric shares, on June 26th, 2008. Left: ask side. Right: bid side.

The following scaling assumption states that, when grouping orders in batches of n orders, a good proportion of batches should have a size $O(\sqrt{n})$ (otherwise their impact will vanish in the limit when n becomes large):

Assumption 3.3. *There exist probability distributions $F, \tilde{F}$ on the interior $(0,\infty)\times(0,\infty)$ of the positive orthant, such that*

$$nf_n(\sqrt{n}\ .)\overset{n\to\infty}{\Rightarrow} F \quad \text{and} \quad n\tilde{f}_n(\sqrt{n}\ .)\overset{n\to\infty}{\Rightarrow}\tilde{F}.$$

Assumption 3.4. *$g\in C^2(\mathbb{R}^2_+\times\mathbb{R}^2_+,]0,\infty[^2)$ and*

$$\exists\alpha>0,\forall(x,y)\in\mathbb{R}^2_+\times\mathbb{R}^2_+,\qquad \|g(x,y)\|\geq\ \alpha\min(y_1,y_2).$$

Finally, we add the following condition for the initial value of the queue sizes:

$$\left(\frac{q_0^{n,b}}{\sqrt{n}},\frac{q_0^{n,a}}{\sqrt{n}}\right)\overset{n\to\infty}{\to}(x_0,y_0)\in]0,\infty[\times]0,\infty[\tag{3.12}$$

The following theorem, whose proof is given in the Appendix, describes the joint dynamics of the bid and ask queues in this heavy traffic limit:

Theorem 3.2 (Heavy traffic limit)**.** *Under Assumptions 4.1, 4.2, 4.3 and 3.4, the rescaled process*

$$(Q^n_t,t\geq 0)=\left(\frac{q^n_{nt}}{\sqrt{n}},t\geq 0\right)$$

converges weakly, on the Skorokhod space $(D(]0,\infty[,\mathbb{R}^2_+),J_1)$,

$$Q^n\quad\overset{n\to\infty}{\Rightarrow}Q$$

to a Markov process $(Q_t)_{t\geq 0}$ with values in $\mathbb{R}^2_+$, initial value $Q_0=(x_0,y_0)$ given in (3.12) and infinitesimal generator $\mathcal{G}$ given, for $x>0,y>0$, by

$$\mathcal{G}h(x,y)=\lambda^a\overline{V^a}\frac{\partial h}{\partial y}+\lambda^b\overline{V^b}\frac{\partial h}{\partial x}+\frac{\lambda^a v_a^2}{2}\frac{\partial^2 h}{\partial y^2}+\frac{\lambda^b v_b^2}{2}\frac{\partial^2 h}{\partial x^2}+\rho\sqrt{\lambda^a\lambda^b}v_a v_b\frac{\partial^2 h}{\partial x\partial y} \tag{3.13}$$

$$\mathcal{G}h(x,0)=\int_{]0,\infty[^2}\mathcal{G}h(g((x,0),(u,v)))F(du,dv),\quad \mathcal{G}h(0,y)=\int_{]0,\infty[^2}\mathcal{G}h(g((0,y),(u,v)))\tilde{F}(du,dv),$$

and whose domain is the set $dom(\mathcal{G})$ of functions $h\in C^2(]0,\infty[\times]0,\infty[,\mathbb{R})\cap C^0(\mathbb{R}^2_+,\mathbb{R})$ verifying the Wentzell boundary conditions

$$\forall x>0,\qquad h(x,0)=\int_{\mathbb{R}^2_+}h(g((x,0),(u,v)))F(du,dv), \tag{3.14}$$

$$\forall y>0,\qquad h(0,y)=\int_{\mathbb{R}^2_+}h(g((0,y),(u,v)))\tilde{F}(du,dv).$$

Proof. We outline here the main steps of the proof. The technical details are given in the Appendix. Define the counting processes

$$N^{a,n}_t=\sup\{k\geq 0,\ T^{a,n}_1+...+T^{a,n}_k\leq t\}\quad \textit{and}\quad N^{b,n}_t=\sup\{k\geq 0,\ T^{b,n}_1+...+T^{b,n}_k\leq t\} \tag{3.15}$$

which correspond to the number of events at the ask (resp. the bid), and the net order flow

$$X^n_t=\left(\sum_{i=1}^{N^{b,n}_{nt}}\frac{V^{b,n}_i}{\sqrt{n}},\sum_{i=1}^{N^{a,n}_{nt}}\frac{V^{a,n}_i}{\sqrt{n}}\right)$$

Then, as shown in Proposition 3.3 (see Appendix), X^n converges in distribution on $(D([0,\infty[,\mathbb{R}^2),J_1)$ to a two-dimensional Brownian motion with drift

$$(X^n_t)_{t\geq 0}\overset{n\to\infty}{\Rightarrow}\left(Z_t+t(\lambda^b\overline{V^b},\lambda^a\overline{V^a})\right)_{t\geq 0}$$

where Z is a planar Brownian motion with covariance matrix

$$\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b}v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b}v_a v_b & \lambda^a v_a^2 \end{pmatrix}.$$

Under assumption 4.1, using the Skorokhod representation theorem, there exist IID sequences $((\epsilon_k^n, n \geq 1), (\tilde{\epsilon}_k^n, n \geq 1), \epsilon_k, \tilde{\epsilon}_k)_{k\geq 1}$ and a copy X of the process

$$\left((x_0, y_0) + Z_t + t(\lambda^b\overline{V^b}, \lambda^a\overline{V^a})\right)_{t\geq 0}$$

on some probability space $(\Omega_0, \mathcal{B}, \mathbb{Q})$ such that $\epsilon_k^n \sim f_n, \tilde{\epsilon}_k^n \sim \tilde{f}_n, \epsilon_k \sim F, \tilde{\epsilon}_k \sim \tilde{F}$ and

$$\mathbb{Q}\left(X^n \overset{n\to\infty}{\to} X \quad ;\forall k \geq 1, \frac{\epsilon_k^n}{\sqrt{n}} \overset{n\to\infty}{\to} \epsilon_k, \quad \frac{\tilde{\epsilon}_k^n}{\sqrt{n}} \overset{n\to\infty}{\to} \tilde{\epsilon}_k\right) = 1.$$

Using the notations of Appendix 3.6.2, denote by

- $\tau_1^n = \tau(X^n)$ the first exit time of X^n from the positive orthant $\mathbb{R}_+^2$ and
- τ_k^n the first exit time of $\Psi_{k-1}(X^n, Q_{\tau_1^n}^n, ..., Q_{\tau_{k-1}^n}^n)$ from $\mathbb{R}_+^2$.

We can now construct the process Q by an induction procedure. Let $\tau_1 = \tau(X)$ be the first exit time of X from the orthant. Let $Q_t = X_t$ for $t < \tau_1$ and, by continuity of the first-passage time map and the last-evaluation map at a first passage time (Whitt 2002, Sec. 13.6.3),

$$(\tau_1^n, Q_{\tau_1^n -}^n) \overset{n\to\infty}{\to} (\tau_1, Q_{\tau_1 -}) \ \mathbb{Q} - a.s.$$

We now set

$$Q_{\tau_1} = g(X_{\tau_1 -}, \epsilon_1)1_{X_{\tau_1}.(0,1)<0} + g(X_{\tau_1 -}, \tilde{\epsilon}_1)1_{X_{\tau_1}.(1,0)<0}.$$

Since X is a Brownian motion,

$$\liminf_{r\downarrow 0}(X_{\tau_1 + r} - X_{\tau_1}).(1,0) < 0$$

therefore $\mathbb{P}\left(1_{X_{\tau_1}.(1,0)<0} = 1_{X_{\tau_1}.(1,0)\leq 0}\right) = 1$ so we can also write

$$Q_{\tau_1} = g(Q_{\tau_1 -}, \epsilon_1)1_{X_{\tau_1}.(0,1)\leq 0} + g(Q_{\tau_1 -}, \tilde{\epsilon}_1)1_{X_{\tau_1}.(1,0)\leq 0}.$$

X is a continuous process and the probability thats its path crosses the origin is zero, so by Lemma 3.2, X lies with probability 1 in the continuity set of the map $G : \omega \to 1_{\omega_{\tau(\omega)}.(0,1)<0}$. So using the continuity of $g(.,.)$, we can apply the continuous mapping theorem (Billingsley 1968, Theorem 5.1), to conclude that

$$Q_{\tau_1^n}^n \overset{n\to\infty}{\to} Q_{\tau_1} \quad \mathbb{Q} - a.s.$$

Let us now assume that we have defined Q on $[0, \tau_{k-1}]$ and shown that

$$(\tau_1^n, .., \tau_{k-1}^n, Q_{\tau_1^n}^n, ..., Q_{\tau_{k-1}^n}^n) \overset{n\to\infty}{\to} (\tau_1, .., \tau_{k-1}, Q_{\tau_1}, ..., Q_{\tau_{k-1}}) \quad \mathbb{Q} - a.s.$$

Since $\mathbb{Q}((0,0) \notin \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})([0,\infty)) \ ) = 1$, Lemma 3.4 implies that $(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})$ lies with probability 1 in the continuity set of Ψ_k, so by the continuous mapping theorem

$$\Psi_k(X^n, Q_{\tau_1^n}^n, ..., Q_{\tau_{k-1}^n}^n) \overset{n\to\infty}{\to} \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}) \quad \mathbb{Q} - a.s.$$

Define now τ_k as the first exit time of $\Psi_k(X, Q_{\tau_1}, ..., Q^n_{\tau_{k-1}})$ from the positive orthant $[0,\infty[\times[0,\infty[$. As before, by continuity of the first-passage time map and the last-value map at a first passage time (Whitt 2002, Sec. 13.6.3),

$$(\tau^n_k, Q^n_{\tau^n_k -}) \overset{n\to\infty}{\to} (\tau_k, Q_{\tau_k -}) \quad \mathbb{Q} - a.s.$$

We can now extend the definition of Q to $[0, \tau_k]$ by setting

$$Q_t = \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_k})(t) \qquad \text{for} \quad t < \tau_k, \qquad \text{and}$$

$$Q_{\tau_k} = g(Q_{\tau_k -}, \epsilon_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(0,1)<0} + g(Q_{\tau_k -}, \tilde{\epsilon}_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(1,0)<0}$$

As above, using the continuity properties of Ψ_k from Lemma 3.4 we conclude that $Q^n_{\tau^n_k} \to Q_{\tau_k}$ a.s. and using the Brownian property of X we can show that

$$Q_{\tau_k} = g(Q_{\tau_k -}, \epsilon_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(0,1)\leq 0} + g(Q_{\tau_k -}, \tilde{\epsilon}_k) 1_{\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}}).(1,0)\leq 0} \quad a.s.$$

So finally, we have shown that

$$\forall k \geq 1, \qquad (\tau^n_1, .., \tau^n_k, Q^n_{\tau^n_1}, ..., Q^n_{\tau^n_k}) \overset{n\to\infty}{\to} (\tau_1, .., \tau_k, Q_{\tau_1}, ..., Q_{\tau_k}) \qquad \mathbb{Q} - a.s.$$

We can now construct the sequences $R, \tilde{R}$ by setting

- $R_k = Q_{\tau_k}$ if $\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})(\tau_k -).(0,1) < 0$,
- $\tilde{R}_k = Q_{\tau_k}$ if $\Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})(\tau_k -).(1,0) < 0$.

Then $Q = \Psi(X, R, \tilde{R})$ where Ψ is the map defined in Definition 3.1. Let us now show that $(X, R, \tilde{R})$ lies with probability 1 in the J_1-continuity set of Ψ, in order to apply the continuous mapping theorem. X is a continuous process whose paths lie in $C^0([0,\infty), \mathbb{R}^2 - \{(0,0)\})$ almost surely. Since F and $\tilde{F}$ have zero mass on the axes, with probability 1 the sequences $(\epsilon_k)_{k\geq 1}, (\tilde{\epsilon}_k)_{k\geq 1}$ do not have any accumulation point on the axes. Assumption 3.4 then implies that the sequences $(R_k)_{k\geq 1}, (\tilde{R}_k)_{k\geq 1}$ do not have any accumulation point on the axes. From the definition of Ψ (Definition 3.1), Q jumps at each hitting time of the axes and, in between two jumps, its increments follow those of the planar Brownian motion X. Since $F, \tilde{F}$ have no mass at the origin and planar Brownian paths have a zero probability of hitting isolated points, with probability 1 the graph of $Q = \Psi(X, R, \tilde{R})$ does not hit the origin :

$$\mathbb{Q}\Big(\ (0,0) \notin \Psi(X, R, \tilde{R})([0,\infty) \ ) \ \Big) = 1. \tag{3.16}$$

So the triplet $(X, R, \tilde{R})$ satisfies the conditions of Theorem 3.1 almost-surely i.e. Ψ is continuous at $(X, R, \tilde{R})$ with probability 1. We can therefore apply the continuous mapping theorem (Billingsley 1968, Theorem 5.1) and conclude that

$$Q^n = (X^n, R^n, \tilde{R}^n) \overset{n\to\infty}{\Rightarrow} Q = \Psi(X, R, \tilde{R}).$$

The process $Q = \Psi(X, R, \tilde{R})$ can be explicitly constructed from the planar Brownian motion X and the sequences $R, \tilde{R}$: Q follows the increments of X and is reinitialized to R_n or $\tilde{R}_n$ at each hitting time of the axes. Lemma 3.5 in Appendix 3.6.4 uses this description to show that Q is a Markov process whose infinitesimal generator is given by (3.13)- (3.14). □

Remark 3.1 (Lévy process limits)**.** *The diffusion approximation inside the orthant fails when order sizes do not have a finite second moment. For example, if the sequence (V^a_i, V^b_i) is regularly varying with tail exponent $\alpha \in (0,2)$ (see Resnick (2006) for definitions), the heavy-traffic approximation Q is a pure-jump process in the positive orthant, constructed by applying the map Ψ to a two-dimensional α-stable Lévy process L:*

$$Q = \Psi(L, R, \tilde{R}),$$

i.e. by re-initializing it according to (3.5) *at each attempted exit from the positive orthant. We do not further develop this case here, but it may be of interest for the study of illiquid limit order markets, or those where order flow is dominated by large block trades.*

3.4.3 Jump-diffusion approximation for order book dynamics

Theorem 3.2 implies that, when examined over time scales much larger than the interval between order book events, the queue sizes q^b and q^a are well described by a Markovian jump-diffusion process $(Q_t)_{t\geq 0}$ in the positive orthant $\mathbb{R}^2_+$ which behaves like a a planar Brownian motion with drift vector

$$(\lambda^b\overline{V^b}, \lambda^a\overline{V^a}) \tag{3.17}$$

and covariance matrix

$$\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b}v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b}v_a v_b & \lambda^a v_a^2 \end{pmatrix}. \tag{3.18}$$

in the interior $]0,\infty[^2$ of the orthant and, at each hitting time τ_k of the axes, jumps to a new position

- $Q_{\tau_k} = R_k = g(Q_{\tau_k -}, \epsilon_k)$ whenever $Q^a_{\tau_{k-}} = 0$,
- $Q_{\tau_k} = \tilde{R}_k = g(Q_{\tau_k -}, \tilde{\epsilon}_k)$ whenever $Q^b_{\tau_{k-}} = 0$,

where the ϵ_k are IID with distribution F and the $\tilde{\epsilon}_k$ are IID with distribution $\tilde{F}$. We note that similar processes in the orthant were studied byBaccelli and Fayolle (1987) with queueing applications in mind, but not in the context of heavy traffic limits.

This process is analytically and computationally tractable and allows various quantities related to intraday price behavior to be computed (see next section).

If $\gamma_0 = (\mathbb{E}[T_1^a] + \mathbb{E}[T_1^b])/2$ is the average time between order book events, ($\gamma_0 \leq 100$ milliseconds), and $\gamma_1 \gg \gamma_0$ (typically, $\gamma_1 \sim$ 10-100 seconds) then Theorem 3.2 leads to an approximation for the distributional properties of the queue dynamics in terms of Q_t:

$$q_t \simeq^d \sqrt{N} \quad Q_{t/N} \qquad \text{where} \qquad N = \frac{\gamma_1}{\gamma_0}$$

So, under Assumptions 4.2, 4.3, 4.1 and 3.4 the queue sizes $(q^b_t, q^a_t)_{t\geq 0}$ can be approximated at the time scale γ_1 by a Markov process which

- behaves like a two-dimensional Brownian motion with drift (μ_b, μ_a) and covariance matrix Λ on $\{x > 0\} \cap \{y > 0\}$ with

$$\mu_a = \sqrt{N}\lambda^a\overline{V^a}, \qquad \mu_b = \sqrt{N}\lambda^b\overline{V^b}, \qquad \Lambda = N\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b}v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b}v_a v_b & \lambda^a v_a^2 \end{pmatrix} \tag{3.19}$$

  and,

- jumps to a new value $g(q_{t-}, \sqrt{N}\epsilon_k)$ if $q^a_{t-} = 0$,
- jumps to a new value $g(q_{t-}, \sqrt{N}\tilde{\epsilon}_k)$ if $q^b_{t-} = 0$,

where $\epsilon_k \sim F$, $\tilde{\epsilon}_k \sim \tilde{F}$ are IID.

This gives a rigorous justification for modeling the queue sizes by a diffusion process at such intraday time scales, as proposed in Avellaneda et al. (2011). The parameters involved in this approximation are straightforward to estimate from empirical data: they involve estimating first and second moments of durations and order sizes.

Example 3.3. *Set for instance $\gamma_1 = 30$ seconds and $\gamma_0 = (\mathbb{E}[T^a] + \mathbb{E}[T^b])/2$. The following table shows the parameters* (3.19) *estimated from high frequency records or order book events for three liquid US stocks.*

In particular we observe that the order of magnitude of the standard deviation of queue lengths is an order of magnitude larger than their expected change.

	Std deviation of Bid queue	Std deviation of Ask queue	μ_b	μ_a	ρ
Citigroup	6256	4457	-1033	-2467	0.07
General Electric	2156	2928	-334	-1291	0.03
General Motors	578	399	+78	-96	- 0.04

Table 3.5: Parameters for the heavy-traffic approximation of bid / ask queues over a 30-second time scale. The unit is a number of orders per period of 30 seconds.

Example 3.4. *Theorem 3.2 may also be used to derive jump-diffusion approximations for the limit order book in theoretical models such as the ones presented in Section 3.2.3. Let us illustrate this in the case of the heterogeneous trader model of Section 3.2.3.*

Let $(T_i, i \geq 1)$ be the sequence of duration between consecutive orders. We assume that this sequence is a sequence of stationary random variables with $\mathbb{E}[T_1] < \infty$. We also assume that every trader has an equal chance of being a buyer or a seller and that the type of trader (buyer or seller) is independent from the past:

$$\mathbb{P}[i-th\ trader\ is\ a\ buyer] = \mathbb{P}[i-th\ trader\ is\ a\ seller] = \frac{1}{2}$$

Finally the sequence of number of orders $(V_i, i \geq 1)$ is a stationary sequence of orders traded by the i-th trader with the property that $\mathbb{E}[V_1^2] < \infty$.

This order flow given by $(T_i, i \geq 1)$, $(V_i, i \geq 1)$, and the sequence of type (buyers or sellers, using limit orders, market orders or both) generates a sequence of durations $(T_i^a, i \geq 1)$, $(T_i^b; i \geq 1)$ and order sizes $(V_i^a, i \geq 1)$ and $(V_i^b, i \geq 1)$ which satisfy assumptions 4.2 and 4.3.

The sequence of durations $(T_i^a, i \geq 1)$ and $(T_i^b, i \geq 1)$ are two stationary sequences of random variables with finite mean:

$$\forall i \geq 0, \quad T_i = T_i^a = T_i^b. \quad therefore \quad \mathbb{E}[T_i] = \mathbb{E}[T_i^a] = \mathbb{E}[T_i^b] < \infty.$$

The sequence of order sizes $((V_i^b, V_i^a), i \geq 1)$ is a sequences of IID random variables with

$$\mathbb{P}[(V_i^b, V_i^a) = (V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, V_i)] = \frac{m}{2}, \tag{3.20}$$

$$\mathbb{P}[(V_i^b, V_i^a) = (-V_i, 0)] = \mathbb{P}[(V_i^b, V_i^a) = (0, -V_i)] = \frac{l}{2}, \tag{3.21}$$

$$\mathbb{P}[(V_i^b, V_i^a) = (\gamma V_i, -(1-\gamma)V_i)] = \mathbb{P}[(V_i^b, V_i^a) = (-(1-\gamma)V_i, \gamma V_i)] = \frac{1-l-m}{2}. \tag{3.22}$$

Theorem 3.2 then shows that (Q^b, Q^a) is a Markov process which behaves like a two-dimensional Brownian motion with drift (μ_b, μ_a) and covariance matrix Λ inside the positive orthant $\{x > 0\} \cap \{y > 0\}$ where:

$$\mu_b = \mu_a = \frac{\overline{V}}{2\mathbb{E}[T_1]}\left(2m + 2\gamma(1-l-m) - 1\right), \quad \Lambda = v^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad where \tag{3.23}$$

$$v^2 = \frac{\mathbb{E}[T_1]\mathbb{E}[V_1^2]}{4}\left(m + l + \frac{\gamma^2 + (1-\gamma)^2}{2}(1-l-m)\right) \ and \ \rho = -\frac{(1-l-m)^2\gamma(1-\gamma)}{1 + (1-l-m)(\gamma^2 - \gamma - 1/2)} < 0. \tag{3.24}$$

Figure 3.14 displays the value of the correlation ρ in different scenarios as a function of γ and the proportion $1-(l+m)$ of traders submitting orders of both types.

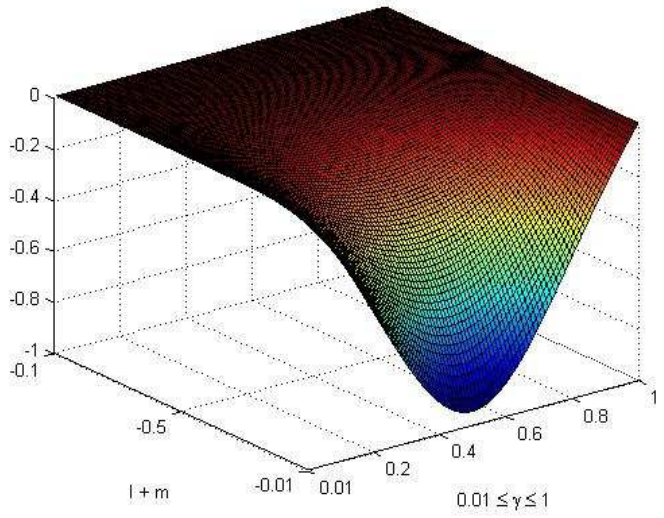

Figure 3.14: Correlation ρ between bid and ask queue sizes for different scenario. 1 - $(l+m)$ represents the proportion of traders using both market and limit orders, γ the proportion of limit orders and $(1-\gamma)$ the proportion of market orders.

3.5 Price dynamics

3.5.1 Price dynamics in the heavy traffic limit

Denote by $(s_t^n, t \geq 0)$ the (bid) price process corresponding to the limit order book process $(q_t^n)_{t\geq 0}$. As explained in Section 3.2, s^n is a piecewise constant stochastic process which

- increases by one tick at each event $(t_i^{a,n}, V_i^{a,n})$ at the ask for which $q^{a,n}(t_i^{a,n}) + V_i^{a,n} < 0$,
- decreases by one tick at each event $(t_i^{b,n}, V_i^{b,n})$ at the bid for which $q^{b,n}(t_i^{b,n}) + V_i^{b,n} < 0$.

Due to the complex dependence structure in the sequence of order durations and sizes, properties of the process s^n are not easy to study, even in simple models such as those given in Section 3.2.3. The following result shows that the price process converges to a simpler process in the heavy traffic limit, which is entirely characterized by hitting times of the two dimensional Markov process Q:

Proposition 3.1. *Under the assumptions of Theorem 3.2,*

$$(s_t^n, t \geq 0) \overset{n\to\infty}{\Rightarrow} S, \quad on \quad (D([0,\infty[,\mathbb{R}), M_1), \qquad \text{where}$$

$$S_t = \delta \left(\sum_{0\leq s\leq t} \mathbf{1}_{Q^a_{s-}=0} - \sum_{0\leq s\leq t} \mathbf{1}_{Q^b_{s-}=0} \right). \tag{3.25}$$

S is a piecewise constant cadlag process which

- *increases by one tick at t if $Q^a_{t-} = 0$ and*
- *decreases by one tick at t if $Q^b_{t-} = 0$.*

Proof. We refer the reader to Whitt (2002) or Whitt (1980) for a description of the M_1 topology. The price process s^n (rescaled in time) can be expressed as

$$s_t^n = \sum_{\tau_k^n \leq t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau^n_{k-1}})(\tau_k^n).(0,1)<0} - 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau^n_{k-1}})(\tau_k^n).(1,0)<0}.$$

where $\tau_k^n, Q^n_{\tau_k^n}$ are defined in the proof of Theorem 3.2, where it was shown that

$$\forall k \geq 1, \quad (X^n, \tau_1^n, .., \tau_k^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau_k^n}) \overset{n\to\infty}{\Rightarrow} (X, \tau_1, .., \tau_k, Q_{\tau_1}, ..., Q_{\tau_k}).$$

As shown in the proof of Theorem 3.2, $(X, Q_{\tau_1}, ..., Q_{\tau_k,...})$ lies, with probability 1, in the set of continuity points of Ψ_k for each $k \geq 1$ so

$$\Psi_k(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau_k^n}) \overset{n\to\infty}{\Rightarrow} \Psi_k(X, Q_{\tau_1}, ..., Q_{\tau_k}).$$

Applying Lemma 3.2 and the continuous mapping theorem (Billingsley 1968, Theorem 5.1) then shows that

$$1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau^n_{k-1}})(\tau_k^n-).(0,1)<0} \overset{n\to\infty}{\Rightarrow} 1_{\Psi_{k-1}(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})(\tau_k-).(0,1)<0}$$

The sequences of processes $\sum_{\tau_k^n\leq t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau^n_{k-1}})(\tau_k^n).(0,1)<0}$ and $\sum_{\tau_k^n\leq t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau^n_{k-1}})(\tau_k^n).(1,0)<0}$ belong to $D_\uparrow([0,\infty[,\mathbb{R}_+)$, the set of increasing cadlag trajectories. The convergence for the M_1 topology of sequences in $D_\uparrow$ reduces to the convergence on a dense subset including zeros. Therefore

$$\sum_{\tau_k^n\leq t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, ..., Q^n_{\tau^n_{k-1}})(\tau_k^n).(0,1)<0} \Rightarrow \sum_{\tau_k\leq t} 1_{\Psi_{k-1}(X, Q_{\tau_1}, ..., Q_{\tau_{k-1}})(\tau_k).(0,1)<0}, \quad \text{and}$$

$$\sum_{\tau_k^n \le t} 1_{\Psi_{k-1}(X^n, Q^n_{\tau_1^n}, \ldots, Q^n_{\tau^n_{k-1}})(\tau_k^n).(1,0)\le 0} \Rightarrow \sum_{\tau_k \le t} 1_{\Psi_{k-1}(X, Q_{\tau_1}, \ldots, Q_{\tau_{k-1}})(\tau_k).(1,0)\le 0}.$$

On the other hand, since the set of discontinuities of $\sum_{\tau_k \le t} 1_{\Psi_k(X, Q_{\tau_1}, \ldots, Q_{\tau_{k-1}})(\tau_k).(0,1)<0}$ and $\sum_{\tau_k \le t} 1_{\Psi_k(X, Q_{\tau_1}, \ldots, Q_{\tau_{k-1}})(\tau_k).(1,0)<0}$ have an intersection which is almost surely void, one can apply (Whitt 1980, Theorem 4.1) and (Whitt 2002, Theorem 12.7.1) and

$$(s_t^n, t \ge 0) \overset{n \to \infty}{\Rightarrow} S. \quad \text{on} \quad (D([0, \infty[, \mathbb{R}_+), M_1).$$

□

S is thus the difference between the occupation time of the y axis and the occupation time of the x axis by the Markov process Q. In particular, this result shows that, in a market where order arrivals are frequent, distributional properties of the price process s^n may be approximated using the distributional properties of the limit S. We will now use this result to obtain some analytical results on the distribution of durations between price changes and the transition probabilities of the price.

3.5.2 Duration between price moves

Starting from an initial order book configuration $Q_0 = (x, y)$,

- the next price increase occurs at the first hitting time of the x-axis by $(Q_{t-})_{t\ge 0}$:

$$\tau_a = \inf\{t \ge 0, Q^a_{t-} = 0\}$$

- the next price decrease occurs at the first hitting time of the y-axis by $(Q_{t-})_{t\ge 0}$:

$$\tau_b = \inf\{t \ge 0, \quad Q^b_{t-} = 0\}.$$

The *duration* τ until the next price changes is then given by

$$\tau = \tau_a \wedge \tau_b,$$

which has the same law as the first exit time from the positive orthant of a two-dimensional Brownian motion with drift. Using the results of Metzler (2010), Zhou (2001) we obtain the following result which relates the distribution of this *duration* to the state of the order book and the statistical features of the order flow process, for a balanced order flow where $\overline{V^a} = \overline{V^b} = 0$.

Proposition 3.2 (Conditional distribution of duration between price changes)**.** *In the case of balanced order flow where $\overline{V^a} = \overline{V^b} = 0$ the distribution of the duration τ until the next price change, conditional on the current state of the bid and ask queues, is given by*

$$\mathbb{P}[\tau > t | Q^b_0 = x, Q^a_0 = y] = \sqrt{\frac{2U}{\pi t}} e^{-\frac{U}{4t}} \sum_{n=0}^{\infty} \frac{1}{(2n+1)} \sin\frac{(2n+1)\pi\theta_0}{\alpha} (I_{(\nu_n - 1)/2}(\frac{U}{4t}) + I_{(\nu_n+1)/2}(\frac{U}{4t})),$$

where $\nu_n = (2n+1)\pi/\alpha$, I_n is the nth Bessel function,

$$U = \frac{(\frac{x}{\lambda_a v_a^2})^2 + (\frac{y}{\lambda_b v_b^2})^2 - 2\rho \frac{xy}{\lambda_a \lambda_b v_a^2 v_b^2}}{(1-\rho)}, \qquad \text{and}$$

$$\alpha = \begin{cases} \pi + \tan^{-1}(-\frac{\sqrt{1-\rho^2}}{\rho}) & \rho > 0 \\ \frac{\pi}{2} & \rho = 0 \\ \tan^{-1}(-\frac{\sqrt{1-\rho^2}}{\rho}) & \rho < 0 \end{cases} \quad \text{and} \quad \theta_0 = \begin{cases} \pi + \tan^{-1}(-\frac{y\sqrt{1-\rho^2}}{x-\rho y}) & x < \rho y \\ \frac{\pi}{2} & x = \rho y \\ \tan^{-1}(-\frac{y\sqrt{1-\rho^2}}{x-\rho y}) & x > \rho y \end{cases} \quad (3.26)$$

In particular, τ is regularly varying with tail index $\dfrac{\pi}{2\alpha}$.

Proof. When $\overline{V^a} = \overline{V^b} = 0$, the process Q behaves like a two-dimensional Brownian motion Z with covariance matrix given by (4.3) up to the first hitting time of the axes, so the distribution of the duration τ has the same law as the first exit time of Z from the orthant:

$$\tau \stackrel{d}{=} \inf\{t \geq 0, Q^a_{t-} < 0 \quad \text{or } Q^b_{t-} < 0\}$$

Using the results of Iyengar (1985), corrected by Metzler (2010) for the distribution of the first exit time of a two-dimensional Brownian motion from the orthant we obtain the result.

A result of Spitzer (1958) then shows that

$$\mathbb{E}[\tau^\beta | Q^b_0 = x, Q^a_0 = y] = \int_0^\infty t^{\beta-a}\mathbb{P}[\tau > t | Q^b_0 = x, Q^a_0 = y]dt < \infty$$

if and only if $\beta < \pi/2\alpha$, where α is defined in (3.26). Therefore the tail index of τ is $\dfrac{\pi}{2\alpha}$. This result does not depend on the initial state (x, y). □

- If $\rho = 0$, the two components of the Brownian Motion are independent and τ is a regularly-varying random variable with tail index 1. This random variable does not have a moment of order one.
- If $\rho < 0$, $\dfrac{\pi}{2\alpha} > 1$ and τ has a finite moment of order one. In practice, $\rho \approx -0.7$; this means that if $\mu_a = 0$ and $\mu_b = 0$, the tail index of τ is around 2.
- When $\rho > 0$, $\dfrac{\pi}{2\alpha} < 1$. The tail of τ is very heavy; τ does not have a finite moment of order one.

For all high frequency data sets examined, the estimates for μ_a, μ_b are negative (see Section 3.4.3); the durations then have finite moments of all orders.

Remark 3.2. *Using the results of Zhou (2001) on the first exit time of a two-dimensional Brownian motion with drift, one can generalize the above results to the case where $(\overline{V^b}, \overline{V^a}) \neq (0, 0)$: we obtain in that case*

$$\mathbb{P}[\tau > t | Q^b_0 = x, Q^a_0 = y] = \frac{2e^{a_1 x + a_2 y + a_t t - r_0^2/2t}}{\alpha t} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi\theta_0}{\alpha}\right) \int_0^\alpha \sin\left(\frac{n\pi}{\alpha}\right) g_n(\theta) d\theta \tag{3.27}$$

where θ_0, α are defined as above, $r_0 = \sqrt{U}$ and

$$g_n(\theta) = \int_0^\infty r e^{-r^2/2t} e^{d_1 r \sin(\theta-\alpha) - d_2 r \cos(\theta-\alpha)} I_{n\pi/\alpha}(\frac{r r_0}{t}) dr,$$

$$d_1 = \left(a_1\sqrt{\lambda^a}v_a + \rho a_2 v_b\sqrt{\lambda^b}\right), \quad d_2 = \left(\rho a_1\sqrt{\lambda^a}v_a + a_2 v_b\sqrt{\lambda^b}\right) \tag{3.28}$$

$$a_1 = -\frac{\mu_a\sqrt{\lambda^b}v_b + \mu_b\rho v_a\sqrt{\lambda^a}}{(1-\rho^2)\sigma_a^2\lambda^a\sqrt{\lambda^b}v_b}, \qquad a_2 = -\frac{\rho\mu_a\sqrt{\lambda^b}v_b + \mu_b v_a\sqrt{\lambda^a}}{(1-\rho^2)\sigma_b^2\lambda^b\sqrt{\lambda^a}v_a}, \tag{3.29}$$

$$\text{and} \qquad a_t = \left(a_1\frac{\lambda^a v_a^2}{2} + a_2\frac{\lambda^b v_b^2}{2} + 2\rho a_1\sqrt{\lambda^a\lambda^b}v_a a_2 v_b\right) - a_1\mu_a - a_2\mu_b. \tag{3.30}$$

3.5.3 Probability of a price increase

A useful quantity for short-term prediction of intraday price moves is the probability $p_1^{up}(x,y)$ that the price will increase at the next move given x orders at the bid and y orders at the ask; in our setting this is equal to the probability that the ask queue gets depleted before the bid queue.

In the heavy traffic limit, this quantity may be represented as the probability that the two-dimensional process $(Q_t, t \geq 0)$, starting from an initial position (x,y), hits the horizontal axis before hitting the vertical axis:

$$p_1^{up}(x,y) = \mathbb{P}[\tau_a < \tau_b | (Q_0^b, Q_0^a) = (x,y)].$$

Since this quantity only involves the process Q up to its first hitting time of the boundary of the orthant, it may be equivalently computed by replacing Q by a two-dimensional Brownian motion with drift and covariance given by (4.2)–(4.3).

However, when $\overline{V^a} = \overline{V^b} = 0$, one has a simple analytical solution which only depends on the size x of the bid queue, the size y of the ask queue and the correlation ρ between their increments:

Theorem 3.3. *Assume $\overline{V^a} + \overline{V^b} \leq 0$. Then $p_1^{up} : \mathbb{R}_+^2 \to [0,1]$ is the unique bounded solution of the Dirichlet problem*

$$\frac{\lambda^a v_a^2}{2}\frac{\partial^2 p_1^{up}}{\partial y^2} + \frac{\lambda^b v_b^2}{2}\frac{\partial^2 p_1^{up}}{\partial x^2} + \rho\sqrt{\lambda^a\lambda^b}\sigma^a\sigma^b\frac{\partial^2 p_1^{up}}{\partial x \partial y} + \lambda^a\overline{V^a}\frac{\partial p_1^{up}}{\partial y} + \lambda^b\overline{V^b}\frac{\partial p_1^{up}}{\partial x} = 0 \quad \text{for} \quad x > 0, \quad y > 0 \tag{3.31}$$

with the boundary conditions

$$\forall x > 0, \quad p_1^{up}(x,0) = 1 \quad \text{and} \quad \forall y > 0, \quad p_1^{up}(0,y) = 0. \tag{3.32}$$

When $\overline{V^a} = \overline{V^b} = 0$, $p_1^{\mathrm{up}}(x,y)$ is given by

$$p_1^{\mathrm{up}}(x,y) = \frac{1}{2} - \frac{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\,\frac{\frac{y}{\sqrt{\lambda_a}v_a} - \frac{x}{\sqrt{\lambda_b}v_b}}{\frac{y}{\sqrt{\lambda_a}v_a} + \frac{x}{\sqrt{\lambda_b}v_b}}\right)}{2\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)}, \tag{3.33}$$

where $\lambda^a, \lambda^b, v_a$ and v_b are defined in Assumptions 4.2 and 4.3.

Proof. Using the results of Yoshida and Miyamoto (1999), the Dirichlet problem (3.31)–(3.32) has a unique positive bounded solution $u \in C^2(]0,\infty[^2, \mathbb{R}_+) \cap C_b^0(\mathbb{R}_+^2, \mathbb{R}_+)$. Application of Ito's formula to $M_t = u(Q_t^b, Q_t^a)$ then shows that the process M^τ stopped at τ is a martingale, and conditioning with respect to $(Q_0^b, Q_0^a) = (x,y)$ gives $u(x,y) = p^{\mathrm{up}}(x,y)$

Assume now $\overline{V^a} = \overline{V^b} = 0$. Using a change of variable $x \mapsto x\sqrt{\lambda^b}v_b$ and $y \mapsto y\sqrt{\lambda^a}v_a$, one only needs to consider the case where $\sqrt{\lambda^b}v_b = \sqrt{\lambda^a}v_a$.

Up to the first hitting time of the axes, $(Q_t, t \geq 0)$ is identical in law to $Q = AB$ where

$$A = \begin{pmatrix} \cos(\beta) & \sin(\beta) \\ \sin(\beta) & \cos(\beta) \end{pmatrix},$$

with β satisfying $\rho = \sin(2\beta)$, $\beta \leq \pi/4$ and B a standard planar Brownian Motion with identity covariance. Using polar coordinates $(x,y) = (r\cos\theta, r\sin\theta)$ we have

$$\phi(r,\theta) := p_1^{up}(r\ A^{-1}(\cos(\theta), \sin(\theta)) = p_1^{up}\left(\frac{r}{\cos^2(\beta) - \sin^2(\beta)}(\cos(\beta+\theta), \sin(\theta-\beta))\right)$$

is a solution of the Dirichlet problem

$$\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial \phi}{\partial r}) + \frac{1}{r^2}\frac{\partial^2 \phi}{\partial \theta^2} = 0 \tag{3.34}$$

in the cone $C = \{(r,\theta), \qquad r > 0, \theta \in]-\beta, \frac{\pi}{2} - \beta[\}$, with the boundary conditions

$$\forall r > 0, \qquad \phi(r, -\beta) = 1 \qquad \phi(r, \frac{\pi}{2} - \beta) = 0 \tag{3.35}$$

A positive bounded solution, which in this case does not depend on r, is given by

$$\phi(r,\theta) = \frac{1}{\pi/2 + \arcsin\rho}(-\theta + \pi/2 + \arcsin(\rho)/2),$$

where ρ is the correlation coefficient between the bid and ask queues. By (Yoshida and Miyamoto 1999, Theorem 3.2), the Dirichlet problem (3.34)–(3.35) has a unique bounded solution, so finally

$$p_1^{up}(x,y) = \frac{1}{\pi/2 + \arcsin\rho}\left(\pi/2 + \arcsin(\rho)/2 - \arctan(\frac{\sin(\arctan(y/x) - \beta)}{\cos(\beta + \arctan(y/x))})\right).$$

□

Remark 3.3. *When $\sqrt{\lambda^a}v_a = \sqrt{\lambda^b}v_b$, the probability $p_1^{up}(x,y)$ only depends on the ration y/x and on the correlation ρ*

$$p_1^{\text{up}}(x,y) = \frac{1}{2} - \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{y-x}{y+x})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}, \tag{3.36}$$

and when $\rho = 0$ (which is the case for some empirical examples, see Section 3.4.3),

$$p_1^{up}(x,y) = \frac{2}{\pi}\arctan(\frac{y}{x}).$$

Figure 3.15 displays the dependence of the uptick probability p_1^{up} on the bid-ask imbalance variable $\theta = \arctan(y/x)$ for different values of ρ.

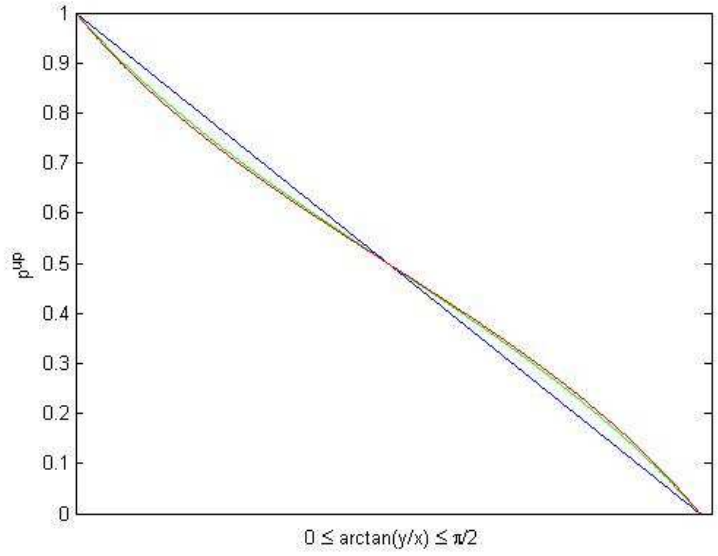

Figure 3.15: p_1^{up} as a function of the bid-ask imbalance variable $\theta = \arctan(y/x)$ for $\rho = 0$ (blue line), $\rho = -0.7$ (green line) and $\rho = -0.9$ (red line).

3.6 Appendix: Technical Proofs

3.6.1 A J_1-continuity property

Lemma 3.2. *Let $\tau : D([0,\infty),\mathbb{R}^2) \mapsto [0,\infty[$ be the first exit time from the positive orthant. The map*

$$\begin{aligned} G : (D([0,\infty),\mathbb{R}^2), J_1) &\to \mathbb{R} \\ \omega &\to 1_{\omega(\tau(\omega)).(0,1)<0}. \end{aligned} \tag{3.37}$$

is continuous on the set $\{\omega \in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\}), \qquad \tau(\omega) < \infty\}$.

When $\tau(\omega) < \infty$, $G(\omega) = 1$ indicates that ω first exits the orthant by crossing the x-axis. To prove this property, first note that

$$C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\}) = \bigcup_{n\geq 1} C_0([0,\infty),\mathbb{R}^2\backslash B(0,1/n)).$$

Let $\omega_0 \in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$. There exists $n \in \mathbb{N}$ such that $\omega_0 \notin B(0,1/n)$. Let $\epsilon > 0$ such that $\epsilon + \eta_{\omega_0}(\epsilon) + \eta_{\omega_0\circ\lambda}(\epsilon) < 1/n$, where η_ω is the modulus of continuity of ω. Let $\omega' \in D([0,\infty),\mathbb{R}^2)$ with $d_{J_1}(\omega_0,\omega') \leq \epsilon$. There exists $\lambda : [0,T] \to [0,T]$ increasing such that

$$||\omega_0 \circ \lambda - \omega||_\infty \leq \epsilon \quad \text{and} \quad ||\lambda - e||_\infty \leq \epsilon.$$

Without loss of generality, one can also assume, by J_1-continuity of τ at ω_0, that

$$|\tau(\omega_0) - \tau(\omega)| \leq \epsilon.$$

Now, we will show that $|\omega_0(\tau(\omega_0)) - \omega'(\tau(\omega'))| \leq \epsilon + \eta_{\omega_0}(\epsilon) + \eta_{\omega_0\circ\lambda}(\epsilon)$:

$$|\omega_0(\tau(\omega_0))-\omega'(\tau(\omega'))| = |\omega_0(\tau(\omega_0))-\omega_0\circ\lambda(\tau(\omega'))+\omega_0\circ\lambda(\tau(\omega'))-\omega_0\circ\lambda(\tau(\omega_0))+\omega_0\circ\lambda(\tau(\omega_0))-\omega'(\tau(\omega'))|,$$

therefore

$$\begin{aligned} |\omega_0(\tau(\omega_0)) - \omega'(\tau(\omega'))| &\leq ||\omega_0 \circ \lambda - \omega'||_\infty + |\omega_0 \circ \lambda(\tau(\omega')) - \omega_0 \circ \lambda(\tau(\omega_0))| + |\omega_0 \circ \lambda(\tau(\omega_0)) - \omega_0(\tau(\omega_0))| \\ &\leq \epsilon + \eta_{\omega_0}(\epsilon) + \eta_{\omega_0\circ\lambda}(\epsilon). \end{aligned}$$

Since $\epsilon + \eta_{\omega_0}(\epsilon) + \eta_{\omega_0\circ\lambda}(\epsilon) < 1/n$ and $\omega_0 \notin B(0,1,n)$, $1_{\tau(\omega_0).(0,1)<0} = 1_{\tau(\omega').(0,1)<0}$, which completes the proof of the continuity of the map G on the space $C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$.

3.6.2 Continuity of Ψ: proof of Theorem 3.1

To study the continuity of the map Ψ, we endow $D([0,\infty),\mathbb{R}^2)$ with Skorokhod's J_1 topology (see Lindvall (1973), Whitt (1980)). Let Λ_T the set of continuous, increasing functions $\lambda : [0,T] \to [0,T]$ and e the identical function on $[0,T]$. Recall that the following metric

$$d_{J_1}(\omega_1,\omega_2) = \inf_{\lambda\in\Lambda} \left(||\omega_2 \circ \lambda - \omega_1||_\infty + ||\lambda - e||_\infty\right).$$

defined for $\omega_1, \omega_2 \in D([0,T],\mathbb{R}^2)$, induces the J_1 topology on $D([0,T],\mathbb{R}^2)$, and $\omega_n \to \omega$ in $(D([0,\infty),\mathbb{R}^2), J_1)$ if for every continuity point T of ω, $\omega_n \to \omega$ in $(D([0,T],\mathbb{R}^2), J_1)$.

The set $(\mathbb{R}^2_+)^{\mathbb{N}}$ is endowed with the topology induced by 'cylindrical' semi-norms, defined as follows: for a sequence $(R^n)_{n\geq 1}$ in $(\mathbb{R}^2_+)^{\mathbb{N}}$

$$R^n \overset{n\to\infty}{\to} R \quad \iff \forall k \geq 1, \quad \sup\{|R^n_1 - R_1|, ..., |R^n_k - R_k|\} \overset{n\to\infty}{\to} 0.$$

$D([0,\infty),\mathbb{R}^2)\times(\mathbb{R}^2_+)^{\mathbb{N}}\times(\mathbb{R}^2_+)^{\mathbb{N}}$ is then endowed with the corresponding product topology. The goal of this section is to characterize the continuity set of the map

$$\Psi : D([0,\infty),\mathbb{R}^2)\times(\mathbb{R}^2_+)^{\mathbb{N}}\times(\mathbb{R}^2_+)^{\mathbb{N}} \mapsto D([0,\infty),\mathbb{R}^2_+)$$

introduced in Definition 3.1. Let us introduce $C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$ be the space of continuous planar paths avoiding the origin:

$$C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\}) = \bigcup_{n\geq 1} C_0([0,T],\mathbb{R}^2\backslash B(0,1/n)).$$

Lemma 3.3. *Let $\omega\in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$. Then the map*

$$\begin{aligned}\Psi_1 : D([0,\infty),\mathbb{R}^2)\times\mathbb{R}_+\times\mathbb{R}_+ &\to D([0,\infty),\mathbb{R}^2) \\ (\omega,R_1,\tilde{R}_1) &\mapsto \omega + 1_{[\tau(\omega),\infty)}\left(1_{\sigma_b(\omega)=\tau(\omega)}(R_1-\omega_{\tau(\omega)}) + 1_{\sigma_a(\omega)=\tau(\omega)}(\tilde{R}_1-\omega_{\tau(\omega)})\right),\end{aligned} \tag{3.38}$$

where

$$\sigma_b(\omega)=\inf\{t\geq 0,\ \omega_t.(0,1)<0\},\quad \sigma_a(\omega)=\inf\{t\geq 0,\ \omega_t.(1,0)<0\}\quad \textit{and}\quad \tau(\omega)=\sigma_b(\omega)\wedge\sigma_a(\omega).$$

is continuous at ω with respect to the following distance on $(D([0,\infty],\mathbb{R}^2)\times\mathbb{R}_+\times\mathbb{R}_+)$:

$$d((\omega,R_1,\tilde{R}_1),(\omega',R_1',\tilde{R}_1')) = d_{J_1}(\omega,\omega') + |R_1-R_1'| + |\tilde{R}_1-\tilde{R}_1|$$

Proof. Let $(\omega_0,R_1,\tilde{R}_1)\in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})\times\mathbb{R}^2_+$, $(\omega',R_1',\tilde{R}_1')\in D(0,\mathbb{R}^2)\times\mathbb{R}^2_+$. Since $\omega_0\in C([0,\infty),\mathbb{R}^2\backslash\{(0,0)\})$, there exists $n>0$ such that $\omega_0\notin B(0,1/n)$. Let $0<\epsilon<1/n$ such that

$$d((\omega_0,R_1,\tilde{R}_1),(\omega',R_1',\tilde{R}_1'))<\epsilon.$$

Since $d_{J_1}(\omega_0,\omega')<\epsilon$, there exists $\lambda:[0,T]\to[0,T]$, non-decreasing such that:

$$||\lambda-e||_\infty<\epsilon,\quad \text{and}\quad ||\omega_0\circ\lambda-\omega||_\infty<\epsilon.$$

By continuity of τ for the J_1 topology Whitt (2002)[Theorem 13.6.4] at ω_0 (since ω_0 is continuous, the J_1 and M_1 topologies are identical at this point), one can also assume, without loss of generality, that

$$|\tau(\omega_0\circ\lambda)-\tau(\omega')|\leq\epsilon.$$

Moreover, since the graph of ω_0 does not intersect with $\overline{B(0,1/n)}$ and $\epsilon<1/n$, $1_{\tau(\omega_0)=\sigma_a(\omega_0)} = 1_{\tau(\omega')=\sigma_a(\omega')}$. Now define λ^ϵ by

$$\begin{aligned}\lambda^\epsilon : [0,T] &\to [0,T] \\ t &\mapsto \frac{\tau(\omega')}{\tau(\omega_0\circ\lambda)}\lambda_t.\end{aligned} \tag{3.39}$$

Then

$$\begin{aligned}||\lambda^\epsilon-e||_\infty = ||\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}\lambda-e||_\infty &\leq ||\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}\lambda-\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}e||_\infty + ||\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}e-e||_\infty \\ &\leq \epsilon\frac{\tau(\omega)}{\tau(\omega_0\circ\lambda)}+\frac{\epsilon}{\tau(\omega_0\circ\lambda)}.\end{aligned}$$

On the other hand

$$||\omega_0\circ\lambda^\epsilon-\omega||_\infty = ||\omega_0\circ\lambda^\epsilon-\omega_0\circ\lambda+\omega_0\circ\lambda-\omega||_\infty \leq ||\omega_0\circ\lambda^\epsilon-\omega_0\circ\lambda||_\infty+\epsilon\leq\eta_{\omega_0\circ\lambda}(\epsilon)+\epsilon,$$

where $\eta_{\omega_0\circ\lambda}$ is the modulus of continuity modulus of $\omega_0 \circ \lambda$. Therefore, since $1_{\tau(\omega_0\circ\lambda^\epsilon)} = 1_{\tau(\omega')}$ by definition of λ^ϵ and

$$\Psi_1(\omega_0, R_1, \tilde{R}_1) \circ \lambda^\epsilon - \Psi_1(\omega', R_1', \tilde{R}_1')) = \omega_0 \circ \lambda^\epsilon - \omega' + 1_{\tau(\omega_0\circ\lambda^\epsilon)} \left(1_{\tau(\omega')=\sigma_a}(R_1' - R_1) + 1_{\tau(\omega')=\sigma_b}(\tilde{R}_1' - \tilde{R}_1)\right).$$

Thus λ^ϵ satisfies $||\lambda^\epsilon - e|| \leq \epsilon(\frac{\tau(\omega')+1}{\tau(\omega_0\circ\lambda)})$ and

$$||\Psi_1(\omega_0, R_1, \tilde{R}_1) \circ \lambda^\epsilon - \Psi_1(\omega', R_1', \tilde{R}_1'))||_\infty \leq \eta_{\omega_0\circ\lambda}(\epsilon) + \epsilon + 2\epsilon$$

which proves that $(\omega_0, R_1, \tilde{R}_1)$ is a continuity point for Ψ_1. □

For $k \geq 2$, define recursively the maps

$$\begin{aligned} \Psi_k : D([0,\infty), \mathbb{R}^2) \times \mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}} &\to D([0,\infty), \mathbb{R}^2) \\ (\omega, (R_i, \tilde{R}_i)_{i\geq 1}) &\mapsto \Psi_1(\Psi_{k-1}(\omega, (R_i, \tilde{R}_i)_{i=1..k-1}), R_k, \tilde{R}_k). \end{aligned} \tag{3.40}$$

To simplify notation we will denote the argument of Ψ_k as $(\omega, R, \tilde{R})(= (\omega, (R_i, \tilde{R}_i)_{i\geq 1})$ although it is easily observed from (3.40) that Ψ_k only depends on the first k elements $(R_i, \tilde{R}_i)_{i=1..k})$ of $R, \tilde{R}$.

Lemma 3.4. *If* $(\omega, R, \tilde{R}) \in C([0,\infty), \mathbb{R}^2\backslash\{(0,0)\}) \times \mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}}$ *such that*

$$(0,0) \notin \Psi_k(\omega, R, \tilde{R})([0,\infty)\ ) \tag{3.41}$$

then Ψ_k *is continuous at* $(\omega, R, \tilde{R})$.

Proof. Let $(R_i, \tilde{R}_i)_{i\geq 1}$, $(R_i', \tilde{R}_i')_{i\geq 1}$, two sequences of random variables on $\mathbb{R}_+^2$ and define

$$\Omega_k(R, \tilde{R}) = \cap_{j=0}^k \Psi_j(C([0,\infty), \mathbb{R}^2\backslash\{(0,0)\}), R, \tilde{R})$$

where we have set $\Psi_0 = Id$. Consider $\omega_0 \in \Omega_k(R, \tilde{R})$, and $\omega \in D([0,T], \mathbb{R}_+^2)$, such that:

$$d_{J_1}(\omega_0, \omega) + \sup_{i=1..k} |R_i - R_i'| + \sup_{i=1..k} |\tilde{R}_i - \tilde{R}_i'| \leq \epsilon.$$

An application of the triangle inequality yields

$$\begin{aligned} &d_{J_1}(\Psi_k(\omega_0, (R_i, \tilde{R}_i)), \Psi_k(\omega', (R_i', \tilde{R}_i'))) \\ &\leq d_{J_1}(\Psi_k(\omega_0, (R_i, \tilde{R}_i)), \Psi_k(\omega', (R_i, \tilde{R}_i))) + d_{J_1}(\Psi_k(\omega', (R_i, \tilde{R}_i)), \Psi_k(\omega', (R_i', \tilde{R}_i'))) \end{aligned}$$

where the last term converges to zero when ϵ goes to zero by continuity of Ψ_1. □

We can now prove Theorem 3.1.

Proof. Proof of Theorem 3.1. Since ω is continuous, the jumps of $\Psi(\omega, R, \tilde{R})$ correspond to the first exit times from the orthant of the paths $\Psi_k(\omega, R, \tilde{R})$. Therefore, if $(R_n)_{n\geq 1}, (\tilde{R}_n)_{n\geq 1}$ have no accumulation points on the axes, the paths $\Psi(\omega, R, \tilde{R})$ only has a finite number of discontinuities on $[0,T]$ for any $T > 0$. So, for any $T > 0$, there exists $k(T)$ such that $\Psi = \Psi_{k(T)}$. Then thanks to Lemma 3.4, Ψ is continuous on the set of continuous trajectories whose image has a finite number of discontinuities and does not contain the origin. □

3.6.3 Functional central limit theorem for the net order flow

Proposition 3.3. *Let $(T_i^{a,n}, T_i^{b,n})_{i\geq 1}$ and $(V_i^{a,n}, V_i^{b,n})_{i\geq 1}$ be stationary arrays of random variables which satisfy Assumptions 4.2 and 4.3. Let $(N_t^{a,n}, t\geq 0)$ and $(N_t^{b,n}, t\geq 0)$ be the counting processes defined in* (3.15)*. Then*

$$\left(\sum_{i=1}^{N_{nt}^{a,n}} \frac{V_i^{a,n}}{\sqrt{n}}, \sum_{i=1}^{N_{nt}^{b,n}} \frac{V_i^{b,n}}{\sqrt{n}}\right)_{t\geq 0} \underset{n\to\infty}{\overset{J_1}{\Rightarrow}} \left(\Sigma B_t + t(\lambda^a \overline{V^a}, \lambda^b \overline{V^b})\right)_{t\geq 0} \tag{3.42}$$

where B is a standard planar Brownian motion and

$$\Sigma^t\Sigma = \begin{pmatrix} \lambda^a v_a^2 & \rho\sqrt{\lambda^a\lambda^b} v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b} v_a v_b & \lambda^b v_b^2 \end{pmatrix}, \tag{3.43}$$

Proof: First we will prove that the sequence of processes

$$\left(\sum_{i=1}^{[\lambda^a t]} \frac{V_i^{a,n}}{\sqrt{n}}, \sum_{i=1}^{[\lambda^b t]} \frac{V_i^{b,n}}{\sqrt{n}}\right)_{t\geq 0} \underset{n\to\infty}{\overset{J_1}{\Rightarrow}} \left(\Sigma B_t + t(\lambda^a \overline{V^a}, \lambda^b \overline{V^b})\right)_{t\geq 0}$$

weakly converges in the J_1 topology. Using the Cramer-Wold device, it is sufficient to prove that for $(\alpha,\beta)\in\mathbb{R}^2$,

$$\left(\alpha \sum_{i=1}^{[\lambda^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t]} \frac{V_i^{b,n}}{\sqrt{n}}\right)_{t\geq 0} \underset{n\to\infty}{\Rightarrow} (\alpha\lambda^a\overline{V^a} + \beta\lambda^b\overline{V^b}))t + \sqrt{(\alpha^2\lambda^a v_a^2 + \beta^2\lambda^b v_b^2 + 2\rho\alpha\beta v_a v_b\sqrt{\lambda^a\lambda^b})} B_t$$

If $\lambda^a\in\mathbb{Q}$ and $\lambda^b\in\mathbb{Q}$, it is possible to find λ such that $\lambda^a/\lambda\in\mathbb{N}$ and $\lambda^b/\lambda\in\mathbb{N}$. Let for all $(i,n)\in\mathbb{N}^2$,

$$W_i^n = \alpha\left(V_{(\lambda^a/\lambda)(i-1)+1}^{a,n} + V_2^{a,n} + ... + V_{\lambda^a i/\lambda}^{a,n}\right) + \beta\left(V_{(\lambda^b/\lambda)(i-1)+1}^{b,n} + V_2^{b,n} + ... + V_{\lambda^b i/\lambda}^{b,n}\right),$$

then for all $t>0$,

$$\alpha\sum_{i=1}^{[\lambda^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta\sum_{i=1}^{[\lambda^b t]} \frac{V_i^{b,n}}{\sqrt{n}} = \sum_{i=1}^{[\lambda t]} \frac{W_i^n}{\sqrt{n}}.$$

For all $n>0$, $(W_i^n, i\geq 1)$ is a sequence of stationary random variables. Therefore, thanks to (Jacod and Shiryaev 2003, Chap.VIII, Thm 2.29, p.426), and the fact that

$$\mathrm{var}(W_1^n) + 2\sum_{i=2}^{\infty} cov(W_1^n, W_i^n) \overset{n\to\infty}{\to} \sigma^2, \tag{3.44}$$

the sequence of processes $\left(\sum_{i=1}^{[\lambda nt]} \frac{W_i^n}{\sqrt{n}},\ t\geq 0\right)_{n\geq 1}$ converges weakly to a Brownian motion with volatility $\sqrt{\lambda}\sigma$. If $(\lambda^a,\lambda^b)\notin\mathbb{Q}^2$, there exists $(\lambda_n^a,\lambda_n^b)_{n\geq 1}$ such that

$$\lambda_n^a, \lambda_n^b \in \mathbb{Q} \quad and \quad |\lambda_n^a - \lambda^a| \leq \frac{1}{n}, \quad |\lambda_n^b - \lambda^b| \leq \frac{1}{n}.$$

As above, one can define an integer λ_n such that $\frac{\lambda_n^a}{\lambda_n}\in\mathbb{Q}$ and $\frac{\lambda_n^b}{\lambda_n}\in\mathbb{Q}$. Let for all $(i,n)\in\mathbb{N}^2$,

$$W_i^n = \alpha\left(V_{(\lambda_n^a/\lambda_n)(i-1)+1}^{a,n} + V_2^{a,n} + ... + V_{\lambda_n^a i/\lambda_n}^{a,n}\right) + \beta\left(V_{(\lambda_n^b/\lambda_n)(i-1)+1}^{b,n} + V_2^{b,n} + ... + V_{\lambda_n^b i/\lambda_n}^{b,n}\right),$$

One has for all $t > 0$,

$$\alpha \sum_{i=1}^{[\lambda^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t]} \frac{V_i^{b,n}}{\sqrt{n}} = \sum_{i=1}^{[\lambda_n t]} \frac{W_i^n}{\sqrt{n}} + \alpha \sum_{i=1}^{[\lambda^a t - \lambda_n^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t - \lambda_n^b t]} \frac{V_i^{b,n}}{\sqrt{n}}.$$

Moreover

$$\left(\alpha \sum_{i=1}^{[\lambda^a t - \lambda_n^a t]} \frac{V_i^{a,n}}{\sqrt{n}} + \beta \sum_{i=1}^{[\lambda^b t - \lambda_n^b t]} \frac{V_i^{b,n}}{\sqrt{n}} \right)_{t \geq 0} \Rightarrow^{J_1} 0,$$

therefore the convergence above holds even if λ^a or λ^b are not rationals. On one hand,

$$\begin{aligned} \mathrm{var}(W_i^n) = \quad & \mathrm{var}\left(\alpha(V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + ... + V^{a,n}_{\lambda_n^a i/\lambda_n}) + \beta(V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + V_2^{b,n} + ... + V^{b,n}_{\lambda_n^b i/\lambda_n}) \right) \\ = \quad & \alpha^2 \mathrm{var}\left(V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} ... + V^{a,n}_{\lambda_n^a i/\lambda_n}) \right) + \beta^2 \mathrm{var}\left(V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} ... + V^{b,n}_{\lambda_n^b i/\lambda_n}) \right) \\ + \quad & 2\alpha\beta \mathrm{cov}\left(V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} ... + V^{a,n}_{\lambda_n^a i/\lambda_n}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} ... + V^{b,n}_{\lambda_n^b i/\lambda_n}) \right). \end{aligned}$$

On the other hand, for all $i \geq 2$,

$$\begin{aligned} \mathrm{cov}(W_1^n, W_i^n) = \quad & \alpha^2 \mathrm{cov}\left(V_1^{a,n} + ... + V^{a,n}_{(\lambda_n^a/\lambda_n)}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + ... + V^{a,n}_{\lambda_n^a i/\lambda_n} \right) \\ + \quad & \beta^2 \mathrm{cov}\left(V_1^{b,n} + ... + V^{b,n}_{(\lambda_n^b/\lambda_n)}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + ... + V^{b,n}_{\lambda_n^b i/\lambda_n} \right) \\ + \quad & \alpha\beta \mathrm{cov}\left(V_1^{a,n} + ... + V^{a,n}_{(\lambda_n^a/\lambda_n)}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + ... + V^{b,n}_{\lambda_n^b i/\lambda_n} \right) \\ + \quad & \alpha\beta \mathrm{cov}\left(V_1^{b,n} + ... + V^{b,n}_{(\lambda_n^b/\lambda_n)}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + ... + V^{a,n}_{\lambda_n^a i/\lambda_n} \right). \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{var}(W_1^n) + 2\sum_{i=2}^{\infty} \mathrm{cov}(W_1^n, W_i^n) = \quad & \mathrm{var}(V_1^{a,n}) \frac{\lambda_n^a}{\lambda_n} + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{a,n}, V_i^{a,n}) \frac{\lambda_n^a}{\lambda_n} \\ + \quad & \mathrm{var}(V_1^{b,n}) \frac{\lambda_n^b}{\lambda_n} + 2\sum_{i=2}^{\infty} \mathrm{cov}(V_1^{b,n}, V_i^{b,n}) \frac{\lambda_n^b}{\lambda_n} \\ + \quad & 2\alpha\beta \mathrm{cov}\left(V_1^{a,n} ... + V^{a,n}_{\lambda_n^a/\lambda_n}, V_1^{b,n} ... + V^{b,n}_{\lambda_n^b/\lambda_n} \right) \\ + \quad & 2\alpha\beta \sum_{i=2}^{\infty} \mathrm{cov}\left(V_1^{a,n} ... + V^{a,n}_{\lambda_n^a/\lambda_n}, V^{b,n}_{(\lambda_n^b/\lambda_n)(i-1)+1} + ... + V^{b,n}_{\lambda_n^b i/\lambda_n} \right) \\ + \quad & 2\alpha\beta \sum_{i=2}^{\infty} \mathrm{cov}\left(V_1^{b,n} ... + V^{b,n}_{\lambda_n^b/\lambda_n}, V^{a,n}_{(\lambda_n^a/\lambda_n)(i-1)+1} + ... + V^{a,n}_{\lambda_n^a i/\lambda_n} \right) \end{aligned}$$

A simple calculation shows that

$$\begin{aligned}
2\alpha\beta\mathrm{cov}\left(V_1^{a,n}...+V_{\lambda_n^a/\lambda_n}^{a,n},V_1^{b,n}...+V_{\lambda_n^b/\lambda_n}^{b,n}\right)& \\
&+\quad 2\alpha\beta\sum_{i=2}^{\infty}\mathrm{cov}\left(V_1^{a,n}...+V_{\lambda_n^a/\lambda_n}^{a,n},V_{(\lambda_n^b/\lambda_n)(i-1)+1}^{b,n}+...+V_{\lambda_n^b i/\lambda_n}^{b,n}\right)\\
&+\quad 2\alpha\beta\sum_{i=2}^{\infty}\mathrm{cov}\left(V_1^{b,n}...+V_{\lambda_n^b/\lambda_n}^{b,n},V_{(\lambda_n^a/\lambda_n)(i-1)+1}^{a,n}+...+V_{\lambda_n^a i/\lambda_n}^{a,n}\right)\\
=2\alpha\beta\max(\frac{\lambda_n^a}{\lambda_n},\frac{\lambda_n^b}{\lambda_n})\mathrm{cov}(V_1^{a,n},V_1^{b,n})+&\quad 2\alpha\beta\sum_{i=2}^{\infty}\frac{\lambda_n^a}{\lambda_n}\mathrm{cov}(V_1^{a,n},V_i^{b,n})+\frac{\lambda_n^b}{\lambda_n}\mathrm{cov}(V_1^{b,n},V_i^{a,n}).
\end{aligned}$$

Therefore

$$\lim_{n\mapsto\infty}\mathrm{var}(W_1^n)+2\sum_{i=2}^{\infty}\mathrm{cov}(W_1^n,W_i^n)=\alpha\frac{\lambda^a}{\lambda}v_a^2+\beta\frac{\lambda^b}{\lambda}v_b^2+2\rho\sqrt{\alpha\beta}\frac{\sqrt{\lambda^a\lambda^b}}{\lambda}v_av_b,$$

where ρ is given in (4.1) and

$$\lim_{n\mapsto\infty}\mathbb{E}[W_i^n]=\alpha\frac{\lambda^a}{\lambda}\overline{V^a}+\beta\frac{\lambda^b}{\lambda}\overline{V^b},$$

which completes the proof of the convergence in (3.44). The law of large numbers for renewal processes implies that the following sequence of processes converges to zero in the J_1 topology Iglehart and Whitt (1971):

$$(N_{nt}^{a,n})_{t\geq 0}\overset{n\to\infty}{\Rightarrow}([\lambda^a t])_{t\geq 0},\quad and\quad (N_{nt}^{b,n})_{t\geq 0}\overset{n\to\infty}{\Rightarrow}([\lambda^b t])_{t\geq 0},$$

$$\left(\sum_{i=[\lambda^a t]}^{N_{nt}^{a,n}}\frac{V_i^{a,n}}{\sqrt{i}},\sum_{i=[\lambda^b t]}^{N_{nt}^{b,n}}\frac{V_i^{b,n}}{\sqrt{i}}\right)_{t\geq 0}\Rightarrow 0\qquad\text{in the } J_1 \text{ topology.}$$

3.6.4 Identification of the heavy traffic limit

Lemma 3.5. *The process Q is a Markov process with values in $\mathbb{R}_+^2$ and infinitesimal generator $(\mathcal{G},\mathrm{dom}(\mathcal{G}))$ given by* (3.13) *and*

$$\begin{aligned}
\mathrm{dom}(\mathcal{G})=\{h\in C^2(]0,\infty[\times]0,\infty[,\mathbb{R})\cap C^0(\mathbb{R}_+^2,\mathbb{R}),&\qquad\forall x>0,\qquad\forall y>0,\\
h(x,0)=\int_{]0,\infty[^2}h(g((x,0),(u,v)))F(du,dv),&\qquad h(0,y)=\textstyle\int_{]0,\infty[^2}h(g((0,y),(u,v)))\tilde{F}(du,dv)\}
\end{aligned}$$

Proof. To identify the infinitesimal generator of the process, we note that $h\in\mathcal{C}^0(\mathbb{R}_+^2)$ is in the domain of the infinitesimal generator if for all $(x,y)\in\mathbb{R}_+^2$

$$\lim_{t\to 0}\frac{\mathbb{E}[h(Q_t)-h(Q_0)|Q_0=(x,y)]}{t}<\infty.$$

For $x>0$, and $y>0$, a classical computation shows that if $h\in\mathcal{C}^2(]0,\infty[\times]0,\infty[)$,

$$\mathbb{E}[h(Q_t)|Q_0=(x,y)]=h(x,y)+t\left(\lambda^a\overline{V^a}\frac{\partial h}{\partial x}+\lambda^b\overline{V^b}\frac{\partial h}{\partial y}+\frac{\lambda^a v_a^2}{2}\frac{\partial^2 h}{\partial x^2}+\frac{\lambda^b v_b^2}{2}\frac{\partial^2 h}{\partial y^2}+\rho\sqrt{\lambda^a\lambda^b}v_av_b\frac{\partial^2 h}{\partial x\partial y}\right)+o(t),$$

which leads to equation (3.13). To examine whether the operator $\mathcal{G}$ is closable on $\mathbb{R}^2_+$ we note that, for $h \in C^2(]0,\infty[\times]0,\infty[) \cap C^0([0,\infty),\mathbb{R}^2)$ and $(x,y) \in \mathbb{R}^2_+$,

$$\begin{aligned}
\mathbb{E}[h(Q_t)|Q_0 = (x,0)] &= \int_{]0,\infty[^2} \mathbb{E}[h(Q_t)|Q_{0+} = g((x,0),(u,v))]F(du,dv) \\
&= \int_{]0,\infty[^2} \left(\mathbb{E}[h(Q_t)|Q_{0+} = g((x,0),(u,v))] - h(g((x,0),(u,v))) + h(g((x,0),(u,v)))\right) F(du,dv) \\
&= \int_{\mathbb{R}^2_+} \left(t\mathcal{G}h(g((x,0),(u,v))) + h(g((x,0),(u,v)))\right) F(du,dv) + o(t).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[h(Q_t)|Q_0 = (0,y)] &= \int_{]0,\infty[^2} \mathbb{E}[h(Q_t)|Q_{0+} = g((0,y),(u,v))]\tilde{F}(du,dv) \\
&= \int_{]0,\infty[^2} \left(\mathbb{E}[h(Q_t)|Q_{0+} = g((0,y),(u,v))] - h(g((0,y),(u,v))) + h(g((0,y),(u,v)))\right) \tilde{F}(du,dv) \\
&= \int_{]0,\infty[^2} \left(t\mathcal{G}h(g((0,y),(u,v))) + h(g((0,y),(u,v)))\right) \tilde{F}(du,dv) + o(t).
\end{aligned}$$

so as $t \to 0$, we have

$$\begin{aligned}
\frac{\mathbb{E}[h(Q_t)|Q_0 = (x,0)] - h(x,0)}{t} &= \int_{]0,\infty[^2} \mathcal{G}h(g((x,0),(u,v)))F(du,dv) \\
&\quad + \frac{1}{t}\int_{]0,\infty[^2} \left(h(g((x,0),(u,v))) - h(x,0)\right) F(du,dv) + o(1).
\end{aligned}$$

$$\begin{aligned}
\frac{\mathbb{E}[h(Q_t)|Q_0 = (0,y)] - h(0,y)}{t} &= \int_{]0,\infty[^2} \mathcal{G}h(g((0,y),(u,v)))\tilde{F}(du,dv) \\
&\quad + \frac{1}{t}\int_{]0,\infty[^2} \left(h(g((0,y),(u,v))) - h(0,y)\right) \tilde{F}(du,dv) + o(1).
\end{aligned}$$

Thus the limit $t \to 0$ is well defined only if h verifies, for $x > 0, y > 0$,

$$h(x,0) = \int_{]0,\infty[^2} h(g((x,0),(u,v)))F(du,dv), \qquad h(0,y) = \int_{]0,\infty[^2} h(g((0,y),(u,v)))\tilde{F}(du,dv), \tag{3.45}$$

This is a Wentzell boundary condition (Taira 1991) which corresponds to a jump to the interior whenever the process reaches the boundary of the quadrant. $\mathcal{G}$ is thus closable on the set

$$\mathrm{dom}(\mathcal{G}) = \{h \in \mathcal{C}^2(]0,\infty[\times]0,\infty[) \cap \mathcal{C}^0(\mathbb{R}^2_+), \qquad h \quad \text{verifies} \quad (3.14)\}$$

and, for $h \in \mathrm{dom}(\mathcal{G})$ we have

$$\mathcal{G}h(x,0) = \int_{]0,\infty[^2} \mathcal{G}h(g((x,0),(u,v)))F(du,dv), \qquad \mathcal{G}h(0,y) = \int_{]0,\infty[^2} \mathcal{G}h(g((0,y),(u,v)))\tilde{F}(du,dv).$$

The elliptic operator defined by the Laplacian on $(0,\infty)^2$ with Wentzell boundary conditions (3.45) thus admits a closure $(\mathcal{G}, \mathrm{dom}(\mathcal{G}))$ on $\mathbb{R}^2_+$ and verifies the assumptions of Galakhov and Skubachevskii (2001)[Theorem 3.1]. Galakhov and Skubachevskii (2001)[Theorem 3.1] then implies the existence of a $\mathbb{R}^2_+$–valued Feller process Q, unique in law, whose infinitesimal generator $(\mathcal{G}, \mathrm{dom}(\mathcal{G}))$. The limit process Q is thus a $\mathbb{R}^2_+$–valued Markov process associated with this semigroup. □

Chapter 4

Linking volatility and order flow.

4.1 Introduction

An increasing number of financial instruments are traded in *limit order markets* where market orders and limit orders are electronically submitted to a *limit order book* and executed according to well-defined time and price priority rules. In such markets traders seeking to transact immediately can buy or sell at the best available price using *market orders*, while other traders may seek to transact at a better price by submitting *limit orders*. The complex interaction between the orders submitted by different market participants then determines the dynamics of the order book and, consequently, the dynamics of market prices.

Understanding the dynamics of limit order books and their link to price dynamics is useful for analyzing and modeling high frequency data and provides a setting for application such as intraday risk management, price impact modeling and optimal trade execution. Various recent studies have considered such problems (see for example Alfonsi et al. (2010), Predoiu et al. (2011), Malo and Pennanen (2010), Bayraktar and Ludkovski (2011)), but a common features of these studies has been to model the dynamics of the price through an *exogenous* stochastic process. Yet, in reality, the dynamics of the price is not independent from the order flow but rather *generated* as an outcome of the interaction of limit and market orders and can be, in principle, *derived* from these elements in a model for the dynamics of the limit order books. In fact, any dynamic model for the limit order book implies some dynamics for bid and ask prices; however, for most such models the resulting price process is too complex to be analytically tractable and understanding the relation between price dynamics and order flow has proved challenging Bouchaud et al. (2008), even in stylized market microstructure models Parlour (1998).

Some recent studies, in which the limit order book is modeled as a Markovian queueing system Cont et al. (2010b), Cont and de Larrard (2010) in which arrivals of market orders and limit orders at each price level are independent Poisson processes, allow for sufficient tractability obtain analytical results on the dynamics of the price process. In Cont and de Larrard (2010), we used this Markovian queueing approach to compute useful quantities (the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, conditional on the state of the order book) and relate the volatility of the price with statistical properties of the order flow. However, the results obtained in these Markovian models rely on the fact that time intervals between orders are independent and exponentially distributed, orders are of the same size and that the order flow at the bid is independent from the order flow at the ask. Unfortunately, empirical studies on high-frequency data have consistently rejected such simplifying assumptions Engle and Russell (1998), Engle (2000), Bouchaud et al. (2002, 2008), Andersen et al. (2010), Cont (2011). Figure 4.1 compares the quantiles of the duration between order book events for CitiGroup stock on June 26, 2008 to those of an exponential distribution with the same mean,

showing that the empirical distribution of durations is far from being exponential. Finally, there is considerable heterogeneity in the sizes of orders: Figure 4.2 shows the (positive or negative) changes in queue size induced by successive orders for CitiGroup.

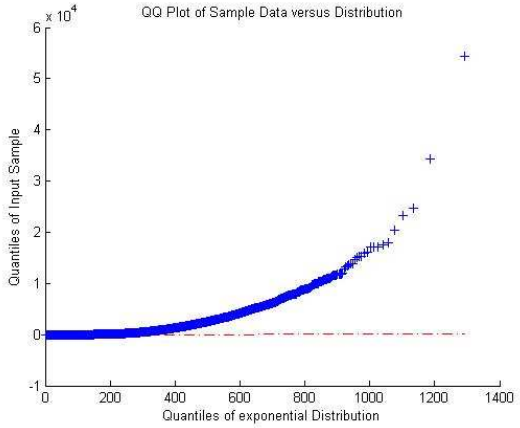

Figure 4.1: Quantiles of inter-event durations compared with quantiles of an exponential distribution with the same mean (Citigroup, June 2008). The dotted line represents the benchmark case where the observations are exponentially distributed, which is clearly not the case here.

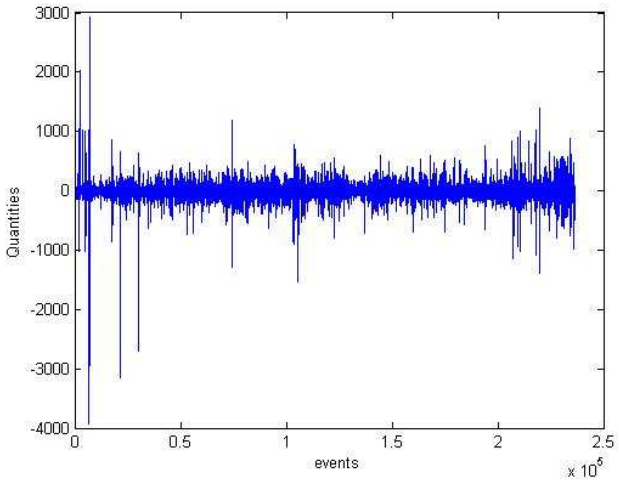

Figure 4.2: Number of shares per event for events affecting the ask. The stock is Citigroup on the 26th of June 2008

In a recent work Cont and de Larrard (2011), we proved that when the frequency of order arrivals is large, the intraday dynamics of the limit order book may be approximated by a Markovian jump-diffusion process in the positive quadrant, whose characteristics are explicitly described in terms of the statistical properties of the underlying order flow. In this paper, we pursue this asymptotic approach to study intraday price dynamics in a limit order market.

We first study the discrete, high-frequency dynamics of the price and derive analytical relations between the statistical properties of intraday price changes -distribution of increments, mean reversion and autocorrelation- and properties of the process describing the order flow and depth of the order

book. We then show that the behavior of the price process at lower frequencies is described by a diffusion limit and express the trend and volatility of this diffusion process in terms of the arrival rates of buy and sell orders and cancelations. These analytical results extend those obtained in Cont and de Larrard (2011) in a Markovian setting and apply to a wide class of order book models, including Poisson point process models, self-exciting point processes, and order flow models based on ACD-GARCH processes. Our results provide analytical insights into the link between price volatility on one hand and high-frequency order flow and liquidity on the other hand. Comparison with high frequency data for liquids US stocks confirm the validity of these insights.

4.1.1 Diffusion models of order book dynamics in liquid markets

Empirical studies of limit order markets suggest that the major component of the order flow occurs at the (best) bid and ask price levels (see e.g. Biais et al. (1995)). All electronic trading venues also allow to place limit orders *pegged* to the best available price (National Best Bid Offer, or NBBO); market makers used these *pegged* orders to liquidate their inventories. Furthermore, studies on the price impact of order book events show that the net effect of orders on the bid and ask queue sizes is the main factor driving price variations (Cont et al. (2010a)). These observations, together with the fact that queue sizes at the best bid and ask of the consolidated order book are more easily obtainable (from records on trades and quotes) than information on deeper levels of the order book, motivate a reduced-form modeling approach in which we represent the state of the limit order book by

- the bid price s_t^b and the ask price s_t^a
- the size of the bid queue q_t^b representing the outstanding limit buy orders at the bid, and
- the size of the ask queue q_t^a representing the outstanding limit sell orders at the ask

Figure 4.3 summarizes this representation.

If the stock is traded in several venues, the quantities q^b and q^a represent the best bids and offers in the *consolidated* order book, obtained by aggregating over all (visible) trading venues. At every time t, q_t^b (resp. q_t^a) corresponds to all visible orders available at the bid price s_t^b (resp. s_t^a) across all exchanges.

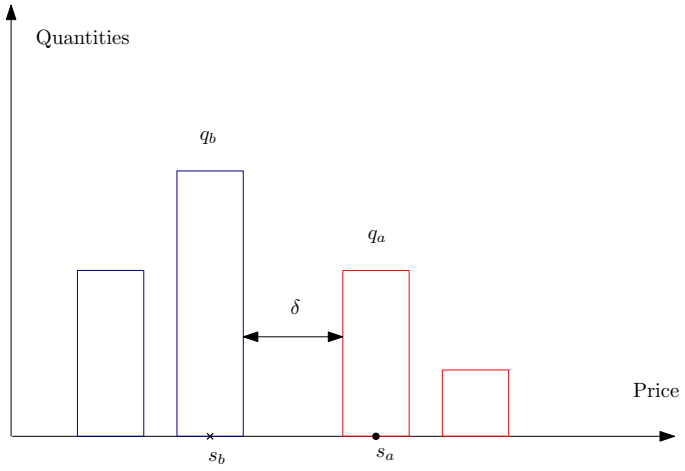

Figure 4.3: Simplified representation of a limit order book.

The state of the order book is modified by *order book events*: limit orders (at the bid or ask), market orders and cancelations (see Cont et al. (2010b,a), Smith et al. (2003)). A limit buy (resp.

sell) order of size x increases the size of the bid (resp. ask) queue by x, while a market buy (resp. sell) order decreases the corresponding queue size by x. Cancelation of x orders in a given queue reduces the queue size by x. Given that we are interested in the queue sizes at the best bid/ask levels, market orders and cancelations have the same effect on the queue sizes (q_t^b, q_t^a).

The bid and ask prices are multiples of the tick size δ. When either the bid or ask queue is depleted by market orders and cancelations, the price moves up or down to the next level of the order book. The price processes s_t^b, s_t^a is thus a piecewise constant process whose transitions correspond to hitting times of the axes $\{(0,y), y>0\} \cup \{(x,0), x>0\}$ by the process $q_t = (q_t^b, q_t^a)$.

If the order book contains no 'gaps' (empty levels), these price increments are equal to one tick:

- when the bid queue is depleted, the (bid) price decreases by one tick.
- when the ask queue is depleted, the (ask) price increases by one tick.

If there are gaps in the order book, this results in 'jumps' (i.e. variations of more than one tick) in the price dynamics. We will ignore this feature in what follows but it is not hard to generalize our results to include it.

The quantity $s_t^a - s_t^b$ is the *bid-ask spread*, which may be one or several ticks. As argued in Section 1.4.1, for many liquid stocks it is reasonable to assume that at time scales more than 10 ms, the spread is constant, equal to one 'tick': $\forall t \geq 0, s_t^a = s_t^b + \delta$. With this approximation, the limit order book may be described using the three variables (s_t^b, q_t^b, q_t^a).

Regime	Time scale	Issues
Ultra-high frequency (UHF)	$\sim 10^{-3} - 0.1$ s	Microstructure, Latency
High Frequency (HF)	$\sim 1 - 100$ s	Trade execution
"Daily"	$\sim 10^3 - 10^4$ s	Trading strategies, Option hedging

Table 4.1: A hierarchy of time scales.

	Average no. of orders in 10s	Price changes in 1 day
Citigroup	4469	12499
General Electric	2356	7862
General Motors	1275	9016

Table 4.2: Average number of orders in 10 seconds and number of price changes (June 26th, 2008).

As shown in Table 4.1, most applications involve the behavior of prices over time scales an order of magnitude larger than the typical inter-event duration: for example, in optimal trade execution the benchmark is the Volume weighted average price (VWAP) computed over a day or a an hour: over such time scales much of the microstructural details of the market are averaged out. Second, as noted in Table 4.2, in liquid equity markets the number of events affecting the state of the order book over such time scales is quite large, of the order of hundreds or thousands. The typical duration between limit orders (resp. market orders and cancelations) is typically $0.001 - 0.01 \ll 1$ (in seconds). These observations show that it is relevant to consider *heavy-traffic limits* in which the rate of arrival of orders is large Whitt (2002) for studying the dynamics of order books in liquid markets.

In Cont and de Larrard (2010), we proved, under assumptions 4.1, 4.2 and 4.3, a functional central limit theorem for the process (s^b, q^b, q^a) when the intensity of orders becomes large, and use it to derive

an analytically tractable jump-diffusion approximation for the intraday dynamics of the limit order book (S, Q^b, Q^a).

We consider a sequence $q^n = (q^n_t)_{t\geq 0}$ of processes, where q^n represents the dynamics of bid and ask queues in the limit order book at time resolution corresponding to n events. The dynamics of q^n is characterized by the sequence of order sizes $(V^{n,b}_i, V^{n,a}_i)_{i\geq 1}$, durations $(T^{n,b}_i, T^{n,a}_i)$ between price change and the fact that, at each price change,

- $q^n_{\tau^n_k} = R^n_k$ if the price has increased, and
- $q^n_{\tau^n_k} = \tilde{R}^n_k$ if the price has decreased.

Assumption 4.1. *Bid and ask queue sizes $(R^n_k)_{k\geq 1}$ (resp. $(\tilde{R}^n_k)_{k\geq 1}$) after a price increase (resp. decrease) are independent from the history of the order book before the price change, and follow a bivariate distribution F_n (resp $\tilde{F}_n$) verifying a scaling assumption: As n goes to infinity, $nf_n(\sqrt{n}.,\sqrt{n}.)$ (resp. $n\tilde{F}_n(\sqrt{n}.,\sqrt{n}.)$) converges weakly to a limit distribution F (resp. $\tilde{F}$):*

$$nF_n(\sqrt{n}.,\sqrt{n}.) \overset{n\to\infty}{\Rightarrow} F \quad \text{and} \quad n\tilde{F}_n(\sqrt{n}.,\sqrt{n}.) \Rightarrow \tilde{F}$$

where F and $\tilde{F}$ are two probability distributions on $]0,\infty[^2$ verifying

$$D(F) = \int_{]0,\infty[^2} xyF(dx\ dy) < \infty \qquad D(\tilde{F}) = \int_{]0,\infty[^2} xy\tilde{F}(dx\ dy) < \infty$$

Assumption 4.2. *Denote by $(T^{n,b}_i, i \geq 0)$ (resp. $(T^{n,a}_i, i \geq 0)$) the sequence of durations between two orders arriving and the bid queue (resp. ask queue). We assume that there exist $\lambda^a > 0$ and $\lambda^b > 0$ such that*

$$\lim_{n\to\infty} \frac{T^{n,a}_1 + T^{n,a}_2 + ... + T^{n,a}_n}{n} = \frac{1}{\lambda^a} < \infty, \qquad \lim_{n\to\infty} \frac{T^{n,b}_1 + T^{n,b}_2 + ... + T^{n,b}_n}{n} = \frac{1}{\lambda^b} < \infty.$$

Assumption 4.3. *For all $n \geq 1$, the sequence of order sizes arriving at the bid side $(V^{n,b}_i, i \geq 0)$ (resp. the ask side $(V^{n,a}_i, i \geq 0)$) is a stationary, uniformly mixing (Billingsley 1968, Ch. 4) sequence satisfying*

$$\sqrt{n}\mathbb{E}[V^{n,a}_1] \overset{n\to\infty}{\to} \overline{V^a}, \qquad \sqrt{n}\mathbb{E}[V^{n,b}_1] \overset{n\to\infty}{\to} \overline{V^b}, \qquad \text{and}$$

$$\lim_{n\geq 1} \mathbb{E}[(V^{n,a}_1 - \mathbb{E}[V^{n,a}_1])^2] + 2\sum_{i=2}^{\infty} Cov(V^{n,a}_1, V^{n,a}_i) = v_a^2 < \infty,$$

$$\lim_{n\geq 1} \mathbb{E}[(V^{n,b}_1 - \mathbb{E}[V^{n,b}_1])^2] + 2\sum_{i=2}^{\infty} Cov(V^{n,b}_1, V^{n,b}_i) = v_b^2 < \infty.$$

Under Assumption 4.3, there exists $\rho \in (-1, 1)$ such that

$$\lim_{n\to\infty} \frac{1}{v_a v_b} \left(2\max(\lambda^a, \lambda^b) Cov(V^{n,a}_1, V^{n,b}_1) + 2\sum_{i=1}^{\infty} \lambda^a Cov(V^{n,a}_1, V^{n,b}_i) + \lambda^b Cov(V^{n,b}_1, V^{n,a}_i) \right) = \rho. \tag{4.1}$$

ρ may be interpreted as a measure of 'correlation' between event sizes at the bid and event sizes at the ask.

Under these assumptions Cont and de Larrard (2011) show that the intraday dynamics of the limit order book may be approximated by a Markov process $Q = (Q^b_t, Q^a_t)_{t\geq 0}$ in the positive quadrant $\mathbb{R}^2_+$ which behaves like a a planar Brownian motion with drift vector

$$(\lambda^b \overline{V^b}, \lambda^a \overline{V^a}) \tag{4.2}$$

and covariance matrix

$$\begin{pmatrix} \lambda^b v_b^2 & \rho\sqrt{\lambda^a\lambda^b}v_a v_b \\ \rho\sqrt{\lambda^a\lambda^b}v_a v_b & \lambda^a v_a^2 \end{pmatrix}. \tag{4.3}$$

in the interior $]0,\infty[^2$ of the quadrant and jumps to a new position with distribution F (resp. $\tilde{F}$) at each hitting time of the x-axis (resp. the y-axis).

Q is a Markov process with infinitesimal generator

$$\mathcal{G}h(x,y) = \left(\lambda^a\overline{V^a}\frac{\partial h}{\partial y} + \lambda^b\overline{V^b}\frac{\partial h}{\partial x} + \frac{\lambda^a v_a^2}{2}\frac{\partial^2 h}{\partial y^2} + \frac{\lambda^b v_b^2}{2}\frac{\partial^2 h}{\partial x^2} + \rho\sqrt{\lambda^a\lambda^b}v_a v_b\frac{\partial^2 h}{\partial x\partial y}\right), \tag{4.4}$$

defined on the domain $dom(\mathcal{G})$ of functions $h \in C^2(]0,\infty[\times]0,\infty[,\mathbb{R}) \cap C^0(\mathbb{R}_+^2,\mathbb{R})$ verifying the "boundary conditions"

$$\int_{\mathbb{R}_+^2}(h(u,v)-h(0,y))F(du,dv) = 0, \qquad \int_{\mathbb{R}_+^2}(h(u,v)-h(x,0))\tilde{F}(du,dv) = 0.$$

A typical path for the dynamics of bid and ask queues in a liquid limit order market is shown in figure 4.4: the queue sizes follow a diffusion-type dynamics in between two price changes and jumps at each price change.

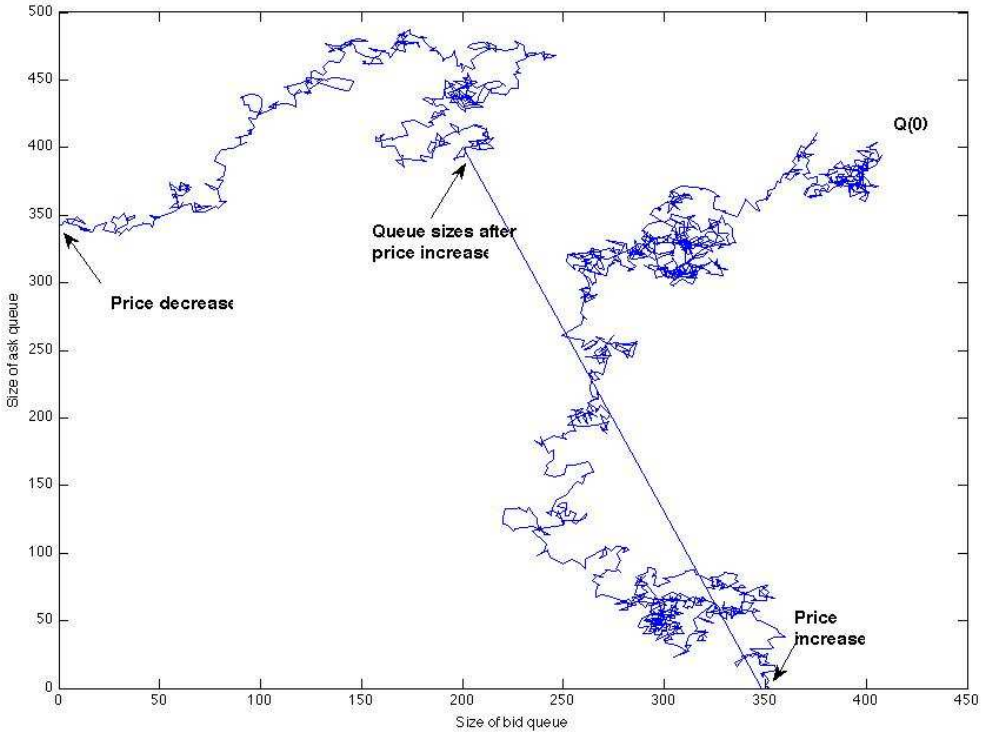

Figure 4.4: Evolution in time of the bid and ask queues: the queue sizes follow a diffusion-type dynamics in between two price changes and jumps at each price change.

In this regime, the price process is a piecewise constant process $S = (S_t)_{t\geq 0}$ which:

- increases by one tick when $Q = (Q^b, Q^a)$ hits the x-axis, and
- decreases by one tick when $Q = (Q^b, Q^a)$ hits the y-axis.

As observed on tables 4.1 and 4.2, one can define three time scales

$$\gamma_0 << \gamma_1 << \gamma_2.$$

The role played by these three time scaled is summarized on Figure 4.5. At the microscopic time scale γ_0 ($\sim 1 - 100$ millisecond) both the price $(s_t)_{t\geq 0}$ and the order book (q^b, q^a) are piecewise constant stochastic processes. At the 'mesoscopic' time scale γ_1 ($\sim$ 1 sec-1 min), the order book behaves as a Markov process (Q^b, Q^a) in the quadrant and the price S is a piecewise constant process.

Over lower frequencies, at the macroscopic time scale $\gamma_2 \sim$ 1 hour- 1 day, the price process S diffuses to a continuous state stochastic process $P = (P_t)_{t\geq 0}$.

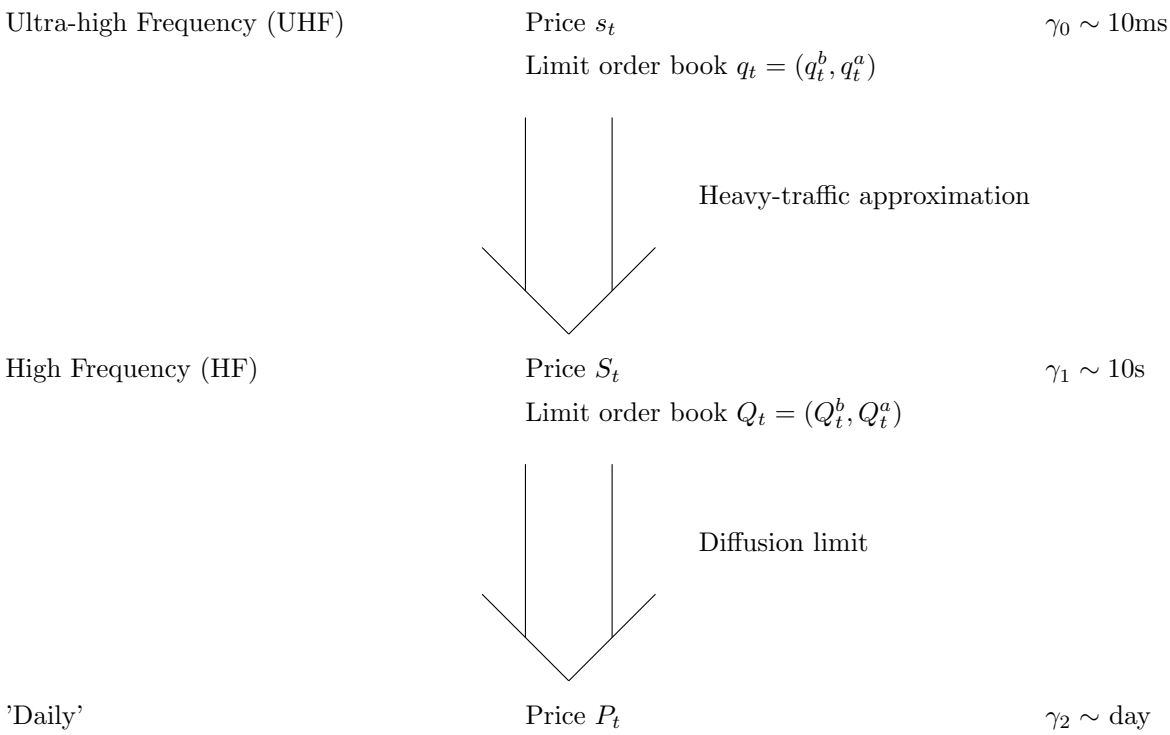

Figure 4.5: Three time scales

When the frequency of limit orders arrival is large ($\sim$ 1000/min), as it is for Dow-Jones stocks (cf table 4.2), it is reasonable to approximate the order book $(s_t, q_t^b, q_t^a)_{t\geq 0}$ by its heavy-traffic limit $(S_t, Q_t^b, Q_t^a)_{t\geq 0}$. One advantage is that the process $S = (S_t)_{t\geq 0}$ is much simpler to study than $(s_t)_{t\geq 0}$. In this paper we examine properties of this price process $S = (S_t)_{t\geq 0}$ such as the correlation of price increments, the distribution of price durations or the volatility of its heavy traffic limit.

4.1.2 Summary

This paper is devoted to the understanding of the price process $S = (S_t)_{t\geq 0}$. Through the analytical tractability of the Markov process (Q^b, Q^a), our model allows to obtain analytical expressions for various quantities of interest such as an explicit expression of the autocorrelation of consecutive price

increments, the low-frequency dynamics of the price and the expression of the drift and the volatility of the price.

We show in section 4.2.1 that the sequence of consecutive price increments $(X_1, X_2, ..., X_n)$ follows a homogenous Markov chain with parameters p^+ and p^-:

$$p^+ = \mathbb{P}[X_2 = \delta | X_1 = \delta] \text{ and } p^- = \mathbb{P}[X_2 = -\delta | X_1 = -\delta],$$

where δ is the tick size. Theorem 4.1 links these parameters to the autocorrelation between consecutive price increments

$$\text{corr}(X_1, X_2) = p^+ + p^- - 1. \tag{4.5}$$

and shows that the probability $p_n^{\text{up}}(x, y)$ that the n-th price move is an increase, conditioned on observing x-shares at the bid and y shares at the ask, may be expressed in terms of the parameters p^+, p^- and $p_1^{\text{up}}(x, y)$:

$$\forall (x, y) \in \mathbb{R}_+^2, \quad p_n^{\text{up}}(x, y) = \frac{1 - p-}{2 - p^+ - p^-}(1 - (p^+ + p^- - 1)^{n-1}) + (p^+ + p^- - 1)^{n-1} p_1^{\text{up}}(x, y).$$

We propose, in section 4.2.2, a parametric distribution for the probability distributions F (resp. $\tilde{F}$), which represents the distribution of the order book (Q^b, Q^a) after a price increase (resp. decrease), in terms of their densities:

$$f(r, \theta) = c^2 e^{-cr} \alpha (\frac{2}{\pi})^\alpha \theta^{\alpha - 1} \quad \textit{and} \quad \tilde{f}(r, \theta) = \tilde{c}^2 e^{-\tilde{c}r} \tilde{\alpha} (\frac{2}{\pi})^{\tilde{\alpha}} \theta^{\tilde{\alpha} - 1}, \tag{4.6}$$

where α (resp. $\tilde{\alpha}$ are parameters characterizing the bid-ask imbalance after a price increase (resp. decrease), and $1/c$ (resp $1/\tilde{c}$) measures the average *depth* of the order book after a price increase (resp. decrease). The parameter α (resp. $\tilde{\alpha}$), characterizing the skewness of the angular part of the distribution F (resp. $\tilde{F}$), are linked to the probabilities of 'continuation' p^+ and p^-.

For a stationary order flow, it is proven in theorems 4.2, 4.3 and 4.4 that over time scales γ_2 much larger than the interval between order book events, price dynamics may be described as a diffusion process driven by a Brownian motion; this results links the discrete high-frequency dynamics of prices to their diffusive dynamics at lower frequencies.

When the order flow is symmetric at the bid/ask, $p^+ = p^- := p_{cont}$, we prove in Theorem 4.2 the following functional central limit theorem for the price process:

$$\left(\frac{s_{[nt]}}{\sqrt{n}} \right)_{t \geq 0} \overset{n \to \infty}{\Rightarrow} \left(\frac{1}{\sqrt{r_1(F)}} \sqrt{\frac{p_{cont}}{1 - p_{cont}}} B_t \right)_{t \geq 0} \tag{4.7}$$

where B is a Brownian motion and $r_1(F)$ is the average time between two consecutive price move.

On the other hand, when the order flow is not symmetric at the bid/ask, $p^+ \neq p^-$ we show in theorem 4.3 that the price can have both a drift and a volatility. Let γ_0 be the average durations between two consecutive orders and $\gamma_2 >> \gamma_0$ a *macroscopic time scale* (e.g. daily time scale). At the time scale γ_2 we prove in theorem 4.3 and 4.4 that the price process $(P_t)_{t \geq 0}$ behaves like a Brownian motion with drift:

$$P_t = \frac{\gamma_2}{\gamma_0} \delta t d_p + \sqrt{\frac{\gamma_2}{\gamma_0}} \delta \sigma_p B_t,$$

where δ is the tick size. $(B_t)_{t \geq 0}$ is a standard Brownian motion. The drift of the price d_p is:

$$d_p = \frac{\frac{1}{1 - p^+} - \frac{1}{1 - p^-}}{(1 + \frac{p^+}{1 - p^+}) r_1(F) + (1 + \frac{p^-}{1 - p^-}) r_1(\tilde{F})}, \tag{4.8}$$

and price volatility σ_p is given by

$$\sigma_p^2 = \frac{\frac{p^+(1+p^+)}{(1-p^+)^2} + \frac{p^-(1+p^-)}{(1-p^-)^2} - 2\frac{p^+}{1-p^-}\frac{p^-}{1-p^-}}{(1+\frac{p^+}{1-p^+})r_1(F) + (1+\frac{p^-}{1-p^-})r_1(\tilde{F})}. \tag{4.9}$$

On equations (4.8) and (4.9), $r_1(F)$ (resp. $r_1(\tilde{F})$) is the average durations until the price moves after a price increase (resp. price decrease).

Eventually, in section 4.4, we link these parameters $p^+, p^-, r_1(F)$ and $r_1(\tilde{F})$ with the order flow and derive expressions of price volatility as a function of order flow statistics for several examples.

The parameters of the the heavy traffic approximation of the order book (q^b, q^a) depend on the properties of the order flow such as the symmetry between the bid and at the ask, the correlation between bid and ask queue sizes or the average order size. When the average order size is of order of magnitude $O(1/\sqrt{n})$, the diffusive limit of the order book is the proper rescaling whereas when the average order size is of order of magnitude $O(1)$ -for instance when marker orders and cancelations dominate limit orders- the heavy traffic limit of the order book is the fluid limit.

In section 4.4 we study several regimes of the order book and we link the price volatility with parameters of the order flow for all these regimes. Table 4.3 points to the sections where these regimes are studied.

Fluid limit or Diffusive limit of (q^b, q^a)	Bid/Ask symmetry	Bid/Ask correlation	Section
Fluid limit	symmetry	no correlation	Section 4.4.1
Diffusive limit	symmetry	no correlation	section 4.4.2
Diffusive limit	symmetry	negative correlation	section 4.4.2
Fluid limit	asymmetry		section 4.4.3

Table 4.3: Different regimes of order book and the sections where these order books are studied.

For instance, when the average evolution of the order book dominates over its volatility, under the assumption that the order flows are symmetric at the bid and at the ask and when F (resp. $\tilde{F}$) are given by equation (4.6), price volatility becomes:

$$\sigma_p^2 = \delta^2 \underbrace{\frac{(\pi/2)^\alpha}{(2^\alpha - 1)\int_0^{\pi/4}\sin(\theta)\theta^{\alpha-1}d\theta + \int_{\pi/4}^{\pi/2}\cos(\theta)\theta^{\alpha-1}d\theta}}_{\text{Bid-Ask asymetry}} \overbrace{\frac{c}{2}}^{\text{Order book depth}} \underbrace{(\mu-\lambda)}_{\text{Trading intensity}},$$

where δ is the tick size and α, c, λ, μ are parameters of the order flow.

On the other hand, when the order book dynamics (Q^b, Q^a) is a symmetric driftless Brownian motion, and when the distribution F has a density f with a polar decomposition $f(r,\theta) = g(r)h(\theta)$, we obtain the following representation for the variance per unit time of price changes

$$\sigma_p^2 = \underbrace{-4\rho}_{\text{Bid-Ask correlation}} \underbrace{\delta^2}_{\text{tick size}} \underbrace{\frac{\overbrace{\lambda^a v_a^2}^{\text{variance of order sizes/unit time}}}{\int_0^\infty h(r)r^3dr \int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta}}_{\text{Order book depth}} \underbrace{\frac{p_{cont}}{1-p_{cont}}}_{\text{mean reversion}}. \tag{4.10}$$

where δ is the tick size, λ^a the trading intensity, v_a^2 the variance of order sizes, p_{cont} the continuation probability ($p_{cont} = p^+ = p^-$ in the symmetric case) and ρ is the correlation between order sizes at bid and the ask. Equation (4.10) generalizes formula (1.23) and links low-frequency price volatility with parameters of the order flow for a general order book. Several components affect price volatility:

- The *variance of order sizes* $v_a^2\lambda^a$
- The *skewness* of the order book after a price move through the parameter p_{cont}.
- The correlation ρ between the order flow arriving at the ask and the order flow arriving at the bid.
- The *depth* of the order book $\int_0^\infty h(r)r^3dr$.

Formula (4.10) gives some insights on the factors that influence price volatility. For instance, when the intensity of *all orders coming in the limit order book* is multiplied by the same factor x, the intensity of orders becomes $\lambda^a x$, the *limit order book depth* is multiplied by a factor x^2, all other parameters being unchanged. So price volatility σ_p decreases by a factor $\sqrt{\frac{1}{x}}$.

Interestingly, Rosu (2009) shows the same dependence in $1/\sqrt{x}$ of price volatility using an equilibrium approach. We show through equation (4.10), that this relation between order arrival intensity and price volatility holds under much more general assumptions, and may be derived without behavioral assumptions for market participants.

4.1.3 Outline

The paper is organized as follows. In Section 4.2, we show that the price mean reverts at a ultra high-frequency time scale and we link this mean reversion property with the skewness of the distributions F and $\tilde{F}$, which represent the state of the limit order book after a price move. Section 4.3 is devoted to the computation of the fluid and the diffusion limit of the price process $(S_t)_{t\geq 0}$. It is showed that the price process diffuses to Brownian motion with drift. The expression of both the drift and the volatility of the price are consistent with empirical data. Finally, in Section 4.4 we explore two particular case, first when the drift of (Q^b, Q^a) dominates over its volatility, second when the dynamics of bid and ask queues behaves as a driftless symmetric Brownian motion. For these two cases, the expression of the price volatility is explicitly derived as a function of order arrival statistics. This formula helps to understand which factors contributes to price volatility. Finally, Section 4.5 is devoted to the proof of technical theorems.

4.2 High-frequency price dynamics

In this section we highlight the mean reversion property of the price process $(S_t)_{t\geq 0}$, which can be linked to the skewness of the distributions F and $\tilde{F}$, used to reinitialize the order book process Q when it hits the boundary of the positive quadrant $\mathbb{R}_+^2$.

4.2.1 Mean reversion of prices at high frequency

Denote by $(X_1, X_2, ..., X_n)$ the successive moves of the price process $S = (S_t)_{t\geq 0}$. Empirically (cf table 4.4), this sequence of random variables is not a sequence of IID random variables. Price dynamics exhibits mean reversion at a high frequency time scale. One defines two parameters p^+ and p^- by:

$$p^+ = \mathbb{P}[X_2 = \delta | X_1 = \delta] \text{ and } p^- = \mathbb{P}[X_2 = -\delta | X_1 = -\delta],$$

where δ is the tick size. The following table 4.4 gives the empirical value of these parameters for stocks belonging to the Dow Jones index.

	p^+	p^-	$1-p^+-p^-$
American Airlines	0.4134	0.3974	0.19
General Electric	0.3740	0.3630	0.26
General Motors	0.4087	0.39	0.21
Hewlett Packard	0.4661	0.4531	0.08
American International Group	0.44472	0.4411	0.11

Table 4.4: Mean reversion parameters p^+ and p^- estimated on the trajectory of the price (June 26th, 2008).

For all stocks shown in Table 4.4, the parameters p^+ and p^- are smaller than 0.5. After a price move, the price is more likely to return to its previous position. The price process $(S_t)_{t\geq 0}$ is not a time-changed random walk: price moves depend on previous ones. The following theorem links the distribution of the nth price move X_n to the initial state (Q_0^b, Q_0^a) of the bid and ask queues and the conditional probabilities p^+ and p^-:

Theorem 4.1. *Given x shares at the bid and y shares at the ask, the probability $p_n^{up}(x,y)$ that the n-th price move is an increase,*

$$p_n^{up}(x,y) = \mathbb{P}[X_n = \delta | Q_0^b = x, Q_0^a = y].$$

is given by

$$\forall (x,y) \in \mathbb{R}_+^2, \quad p_n^{\text{up}}(x,y) = \frac{1-p^-}{2-p^+-p^-}(1-(p^++p^--1)^{n-1}) + (p^++p^--1)^{n-1} p_1^{\text{up}}(x,y). \quad (4.11)$$

In particular, when $p_1^{up}(x,y) = 1/2$,

$$\text{corr}(X_1, X_2) = p^+ + p^- - 1, \quad (4.12)$$

Proof Let $(x,y) \in \mathbb{R}_+^2$. Since the process Q is regenerated at each exit time from the interior of the orthant, the sequence $(p_n^{\text{up}}(x,y), n \geq 1)$ satisfies the following recurrence relation:

$$p_n^{up}(x,y) = p^+ p_{n-1}^{up}(x,y) + (1-p^-)(1-p_{n-1}^{up}(x,y)).$$

Hence

$$\begin{pmatrix} p_n^{up}(x,y) \\ 1-p_n^{up}(x,y) \end{pmatrix} = \begin{pmatrix} p^+ & 1-p^- \\ 1-p^+ & p^- \end{pmatrix}^{n-1} \begin{pmatrix} p_1^{up}(x,y) \\ 1-p_1^{up}(x,y) \end{pmatrix}.$$

The eigenvalues of this matrix are λ_1 and λ_2, given by

$$\begin{pmatrix} p^+ & 1-p^- \\ 1-p^+ & p^- \end{pmatrix} = \frac{1}{(1-p^-)(2-p^+-p^-)} \begin{pmatrix} 1-p^- & 1-p^- \\ \lambda_1 - p^+ & \lambda_2 - p^+ \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_2 - p^+ & p^- - 1 \\ p^+ - \lambda_1 & 1-p^- \end{pmatrix},$$

where

$$\lambda_1 = p^+ + p^- - 1 \quad \text{and} \quad \lambda_2 = 1$$

Therefore

$$p_n^{\text{up}}(x,y) = \frac{1-p^-}{2-p^+-p^-}(1-(p^++p^--1)^{n-1}) + (p^++p^--1)^{n-1} p_1^{\text{up}}(x,y).$$

We can now compute the covariance of price changes at first lag

$$\text{cov}(X_1, X_2 | Q_0^b = x, Q_0^a = y) = \mathbb{E}[X_1 X_2 | Q_0^b = x, Q_0^a = y] - \mathbb{E}[X_1 | Q_0^b = x, Q_0^a = y] \mathbb{E}[X_2 | Q_0^b = x, Q_0^a = y]$$

$$= p_1^{\text{up}}(x,y)p_+ + (1-p_1^{\text{up}}(x,y))p_- - p_1^{\text{up}}(x,y)(1-p_+) - (1-p_1^{\text{up}}(x,y))(1-p_-) - (2p_1^{\text{up}}(x,y)-1)(2p_2(x,y)-1)$$
$$= 2p_1^{\text{up}}(x,y)(p^+ - p^-) + 2p^- - 1 - (2p_1^{\text{up}}(x,y) - 1)(1 + 2p_1^{\text{up}}(x,y)(p_+p_- - 1) - 2p_-).$$

Hence when $p_1^{\text{up}}(x,y) = 1/2$, $\text{cov}(X_1, X_2 | Q_0^b = x, Q_0^a = y)$ does not depend on (x,y) and is equal to the (unconditional) correlation:

$$\text{corr}(X_1, X_2) = p^+ + p^- - 1$$

which concludes the proof.

Remark 4.1. *Formula* (4.11) *generalizes the result given in Cont and de Larrard (2011) for the particular case when* $p^+ = p^- := p_{cont}$*:*

$$p_n^{\text{up}}(x,y) = \frac{1 + (2p_{cont} - 1)^{n-1}(2p_1^{\text{up}}(x,y) - 1)}{2}.$$

Figure 4.6 shows an empirical test, which compares the correlation of consecutive increments to its theoretical counterpart (4.12) : we observe very good agreement with the data.

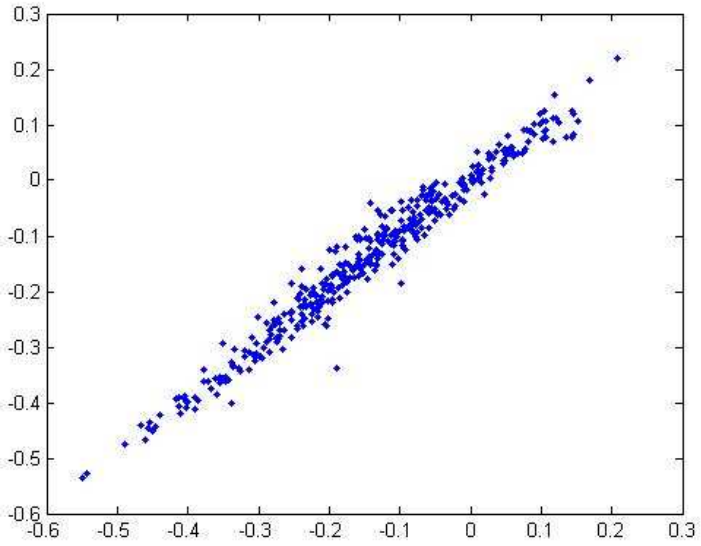

Figure 4.6: Correlation of consecutive increments of the price as a function of $p^+ + p^- - 1$ for Dow Jones stocks. Every point corresponds to a day of trading. The regression gives an R^2 of 0.96.

4.2.2 A parametric model for order book depth

We propose in this section a simple parameterizations for the joint distribution F (resp. $\tilde{F}$) of the bid and ask queue sizes $Q = (Q^b, Q^a)$ after a price increase (resp. decrease). First, we parameterize these distributions in terms of

- a radial component $\sqrt{|Q^b|^2 + |Q^a|^2}$, which measures the depth of the order book, and
- an angular component $\arctan(Q^a/Q^b) \in [0, \pi/2]$ which measures the *imbalance* between outstanding buy and sell orders.

If we now assume for simplification that these two variables $(R, \Theta) = \left(\sqrt{|Q^b|^2 + |Q^a|^2}, \arctan(Q^a/Q^b)\right)$ are independent, then F and $\tilde{F}$ take a product form:

$$F(x,y) = H(\sqrt{x^2+y^2})G\left(\arctan(\frac{y}{x})\right) \quad \text{and} \qquad \tilde{F}(x,y) = \tilde{H}(\sqrt{x^2+y^2})\tilde{G}\left(\arctan(\frac{y}{x})\right) \tag{4.13}$$

where $H, \tilde{H}$ are probability distributions on $\mathbb{R}_+$ and $G, \tilde{G}$ are probability distributions on $[0, \pi/2]$. As we shall observe below, this separable form (4.13) allows more analytical tractability and a transparent interpretation of the results.

The function H (resp. $\tilde{H}$) is the distribution of the *depth* $\sqrt{(Q^b)^2 + (Q^a)^2}$ of the limit order book (Q^b, Q^a) after a price increase (resp. decrease). Figure 4.7 displays the sample estimators for $\log(1 - H(r))$ and $\log(1 - \tilde{H}(r))$ as a function of r.

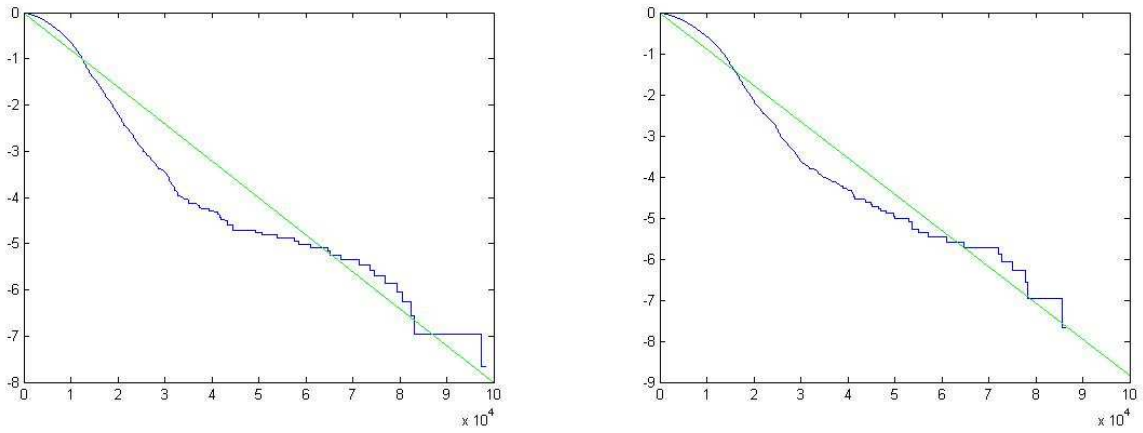

Figure 4.7: Logarithm of the empirical cumulative distribution function for H (left) and $\tilde{H}$ (right) for the stock Citi on the 26th of June 08. The (green) line is the best exponential approximation.

Figure 4.7 suggests that the radial densities h and $\tilde{h}$ may be modeled by exponential functions

$$\forall r > 0, \quad h(r) = c^2 e^{-cr}, \quad \text{and} \quad \tilde{h}(r) = \tilde{c}^2 e^{-\tilde{c}r},$$

where c and $\tilde{c}$ are positive constants.

The functions G (resp. $\tilde{G}$) represent the distribution of the bid-ask imbalance of the order book after a price increase (resp. decrease), as measures by the angular variable $\frac{2}{\pi}\arctan(\frac{Q^a}{Q_b})$. Price mean reversion, highlighted in Table 4.4 and Figure 4.8, suggests that after a price increase, the price

is more likely to decrease. The state of the order book (Q^b, Q^a) has more chance to be on the part $Q^a < Q^b$ than on the part $Q^a \geq Q^b$ after a price increase. After a price increase, the ask queue is more likely to be shorter than the bid queue. This translates into an asymmetry of the angular distribution G and $\tilde{G}$. This indicates that the function G is not a uniform distribution over $[0, \pi/2]$ but skewed towards large angles, whereas $\tilde{G}$ is skewed towards zero. We propose to parameterize G and $\tilde{G}$ through their probability densities $g, \tilde{g}$ as

$$\forall \theta \in [0, \pi/2], \quad g(\theta) = \alpha(\frac{2}{\pi})^{\alpha}\theta^{\alpha-1} \quad \text{and} \quad \tilde{g}(\theta) = \tilde{\alpha}(\frac{2}{\pi})^{\tilde{\alpha}}\theta^{\tilde{\alpha}-1}. \tag{4.14}$$

The parameter α (resp. $\tilde{\alpha}$) measures the skewness of the distribution G (resp $\tilde{G}$). When $\alpha = 1$, G is the uniform distribution over $[0, \pi/2]$. If $\alpha > 1$, the distribution G is skewed towards large angles. This corresponds to a mean reverting regime for the price, in which the price is more likely to come back to its previous position after each price change.

Figure 4.8 compares the empirical cumulative distribution of the angular variable $\Theta = \arctan(Q^a/Q^b)$ to the parametric model (4.14): we observe that the parametric form (4.14) is flexible enough to fit the empirical distributions. The estimated parameters are $\alpha = 2.15$ for the distribution G and $\tilde{\alpha} = 0.28$ for $\tilde{G}$.

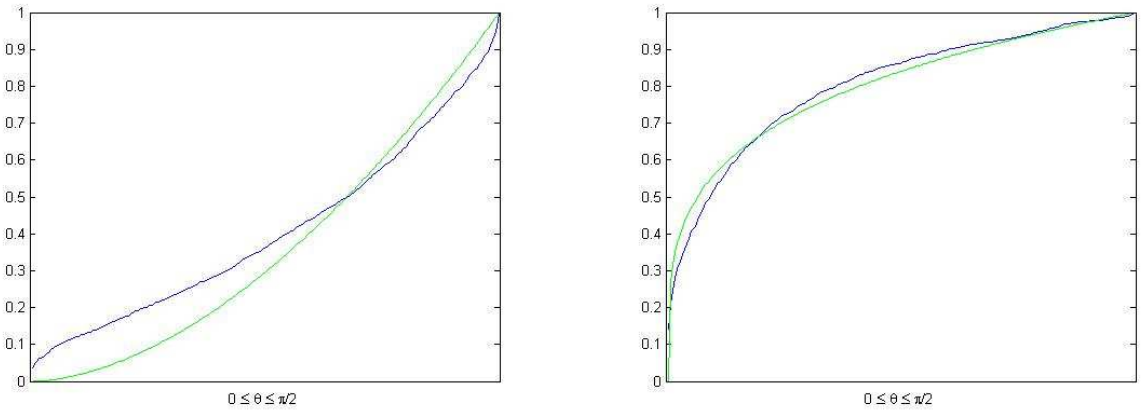

Figure 4.8: Left: empirical cumulative distribution function (in blue) of the bid-ask imbalance variable $\Theta = \arctan(Q^a/Q^b)$ after a price increase (left) and the best fit (green) using the parametric form (4.14). Right: empirical cumulative distribution function (in blue) of the bid-ask imbalance variable $\Theta = \arctan(Q^a/Q^b)$ after a price decrease (left) and the best fit (green) using the parametric form (4.14).

These assumptions lead to the following parametric representation in polar coordinates for the densities $f, \tilde{f}$ of $F, \tilde{F}$:

$$f(r,\theta) = c^2 e^{-cr}\alpha(\frac{2}{\pi})^{\alpha}\theta^{\alpha-1} \quad \text{and} \quad \tilde{f}(r,\theta) = \tilde{c}^2 e^{-\tilde{c}r}\tilde{\alpha}(\frac{2}{\pi})^{\tilde{\alpha}}\theta^{\tilde{\alpha}-1}, \tag{4.15}$$

where α, $\tilde{\alpha}$, c and $\tilde{c}$ are constants. Here $\frac{1}{c}$, $\frac{1}{\tilde{c}}$ represent the average *depth* of the order book and $\alpha > 0$ (resp $\tilde{\alpha} > 0$) characterize the bid/ask imbalance after a price increase (resp decrease).

Remark 4.2. *Another way to model the angular part of the distributions would be to use a Beta(α, β) distribution:*

$$\forall \theta \in [0, \pi/2], \quad g(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{2}{\pi}\right)^{\alpha+\beta} \theta^{\alpha} \left(\frac{\pi}{2} - \theta\right)^{\beta},$$

where $\alpha > 0$, $\beta > 0$, *and* Γ *is the Gamma function. The Beta distribution, which generalizes* (4.14), *is more flexible and allows* G *to be concentrated in any set of angles* $[\theta_0, \theta_1] \in [0, \pi/2]$.

4.2.3 Relation between price mean reversion and bid-ask asymmetry

In section 4.2.1, we defined two parameters p^+ and p^-, which characterize the autocorrelation of price increments. These parameters p^+ and p^- may be related to the distributions F and $\tilde{F}$ via

$$p^+ = \int_{R_+^2} p^{up}(x,y)F(dx,dy) \quad \text{and} \quad p^- = \int_{R_+^2} (1-p^{up}(x,y))\tilde{F}(dx,dy), \tag{4.16}$$

where $p^{\text{up}}(x,y)$ $(= p_1^{\text{up}})$ is the probability that the next price move is an increase. It is thus intuitive that the skewness in the order book depth distributions F and $\tilde{F}$ leads to mean reversion in the price. The following result makes this intuition precise, when $F, \tilde{F}$ is given by the parametric form (4.15), in a *balanced* order flow for which the trend (4.2) of queue sizes is zero:

Proposition 4.1. *If F and $\tilde{F}$ follows the polar decomposition* (4.15)*, and when the drift μ of (Q^b, Q^a), defined in* (4.2)*, is zero the mean reversion of the price can be linked with the skewness α (resp. $\tilde{\alpha}$) of the limit order book after a price increase (resp. decrease):*

$$p^+ = \frac{1}{2} - \alpha(\frac{2}{\pi})^\alpha \int_0^{\pi/2} \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}\theta^{\alpha-1}d\theta. \tag{4.17}$$

$$p^- = \frac{1}{2} + \tilde{\alpha}(\frac{2}{\pi})^{\tilde{\alpha}} \int_0^{\pi/2} \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}\theta^{\tilde{\alpha}-1}d\theta. \tag{4.18}$$

In particular when $\rho = 0$

$$p^+ = \frac{1}{1+\alpha}, \quad p^- = \frac{\tilde{\alpha}}{1+\tilde{\alpha}}. \tag{4.19}$$

Proof. In Cont and de Larrard (2011) $\mu^a = 0$ and $\mu^b = 0$, the function p^{up} only depends on the bid-ask imbalance variable θ and the correlation ρ between events sizes at the bid and the ask:

$$p^{up}(\theta) = \frac{1}{2} - \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}.$$

Therefore

$$\begin{aligned}
p^- &= \int_{R_+^2} (1-p^{up}(x,y))\tilde{F}(x,y)dxdy \\
&= \int_0^\infty \int_0^{\pi/2} \left(\frac{1}{2} + \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})} \right) \tilde{c}^2 e^{-\tilde{c}r}\tilde{\alpha}(\frac{2}{\pi})^{\tilde{\alpha}}\theta^{\tilde{\alpha}-1} \\
&= \frac{1}{2} + \tilde{\alpha}(\frac{2}{\pi})^{\tilde{\alpha}} \int_0^{\pi/2} \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})}\theta^{\tilde{\alpha}-1}d\theta.
\end{aligned}$$

Similarly

$$\begin{aligned} p^+ &= \int_{R_+^2} p^{up}(x,y)F(x,y)dxdy \\ &= \int_0^\infty \int_0^{\pi/2} \left(\frac{1}{2} - \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})} \right) c^2 e^{-cr} \alpha (\frac{2}{\pi})^\alpha \theta^{\alpha-1} \\ &= \frac{1}{2} - \alpha(\frac{2}{\pi})^\alpha \int_0^{\pi/2} \frac{\arctan(\sqrt{\frac{1+\rho}{1-\rho}}\frac{\tan(\theta)-1}{\tan(\theta)+1})}{2\arctan(\sqrt{\frac{1+\rho}{1-\rho}})} \theta^{\alpha-1} d\theta. \end{aligned}$$

□

These mean reversion parameters p^+ (resp p^-) depend only on the correlation between variations in bid and ask queue sizes and the skewness of F (resp $\tilde{F}$). In particular, when $\rho = 0$, we retrieve (4.19).

When $\alpha = \tilde{\alpha} = 1$, mean reversion disappears and $p^+ = p^- = 1/2$: the sequence $(X_1, ..., X_n)$ then becomes a sequence of IID Bernoulli random variables and the price S follows a time-changed symmetric random walk. Otherwise, when $\alpha > 1$, $p^+ < 1/2$ and when $\alpha < 1$, $p^+ > 1/2$.

Remark 4.3. *As observed on Figure 4.2.3, and Figure 4.2.3, the main contribution to p^+ (resp p^-) is clearly α (resp $\tilde{\alpha}$). Numerically, when $\rho < 0$, the approximation $p^+(\alpha, \rho) \approx p^+(\alpha, 0) = \frac{\alpha}{1+\alpha}$ leads to a relative error less than 10%. This approximation shows that, to leading order, the correlation between consecutive increments of the price depends only on the asymmetry parameters α and $\tilde{\alpha}$:*

$$\text{corr}(X_k, X_{k+1}) = p^+ + p^- - 1 \approx \frac{\tilde{\alpha} - \alpha}{(\alpha+1)(\tilde{\alpha}+1)}. \tag{4.20}$$

The sequence of price increments $(X_1, ..., X_n)$ is uncorrelated if $\alpha = \tilde{\alpha}$ and is a sequence of IID Bernoulli random variables only when $\alpha = \tilde{\alpha} = 1$.

4.2.4 "Martingale price"

Various authors have defined, in the context of high frequency price modeling, an auxiliary price process as an average of the bid and ask price. Burghardt et al. (2006) define what they call a '*true price*' as:

$$\hat{P}_t = \frac{Q_t^a}{Q_t^a + Q_t^b} S_t^b + \frac{Q_t^b}{Q_t^a + Q_t^b} S_t^a,$$

where S^b is the bid price, and S^a the ask price. Robert and Rosenbaum (2011) consider a (non-observed) "efficient" price and argue that trades can only occur when the efficient price is close to the grid $\delta\mathbb{Z}$, where δ is the tick size. When the efficient price is too far from the grid, the uncertainty on the next price value dissuades people from trading.

As in Cont and de Larrard (2010), we define an "efficient" price based on the price value and the state of the limit order book (q^b, q^a): given the probability $p^{up}(Q_t^b, Q_t^a)$ that the next price move is an "uptick", we can construct an auxiliary process $\hat{S}$ whose value $\hat{S}_t$ represents the expected value of the price after its next move:

$$\begin{aligned} \forall t \geq 0,\ \hat{S}_t &= (S_t + \delta)\, p^{up}(Q_t^b, Q_t^a) + (S_t - \delta)\left(1 - p^{up}(Q_t^b, Q_t^a)\right), \\ \hat{S}_t &= S_t + \delta(2p^{up}(Q_t^b, Q_t^a) - 1), \end{aligned}$$

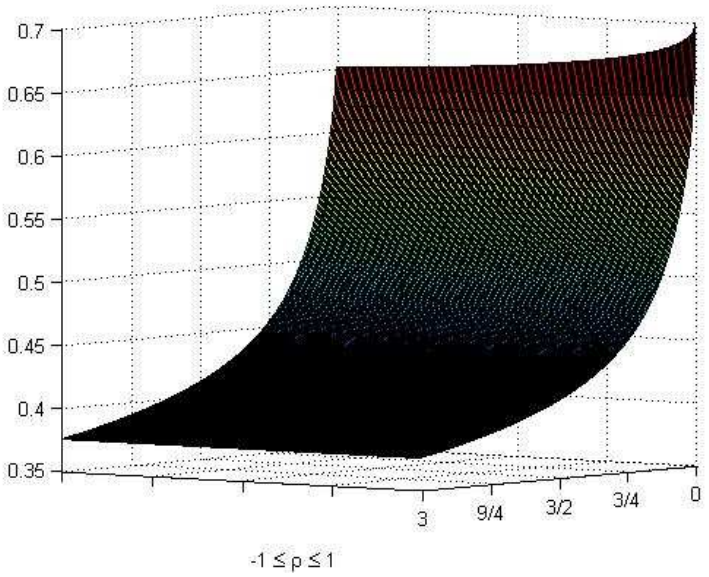

Figure 4.9: p^- as a function of ρ and α for $-1 \leq \rho \leq 1$ and $0 \leq \alpha \leq 3$.

$(\hat{S}_t)_{t\geq 0}$ is a continuous-time stochastic process with values between $S_t - \delta$ and $S_t + \delta$:

$$\forall t \geq 0, \quad S_t - \delta \leq \hat{S}_t \leq S_t + \delta.$$

The process $\hat{S}$ incorporates the information on the price S_t and the state of the order book Q insofar as it affects the next price move. The following result shows

Proposition 4.2. *If $p^+ = p^- = 1/2$, then $(\hat{S}_t, t \geq 0)$ is a martingale.*

Proof. Let $(\tau_1, \tau_2, ..., \tau_k)$ the sequence of times when the price S moves and $(X_1, ..., X_n)$ the sequence of consecutive price moves. Since $p^+ = p^- = 1/2$, $(X_1, ..., X_k, ...)$ is a sequence of I.I.D bernoulli random variables with parameter $1/2$. Therefore we have the following property:

$$\forall (i,j) \in \mathbb{N}^2, \ i < j, \ \mathbb{E}[S_{\tau_j} | \mathcal{F}_{\tau_i}] = S_{\tau_i}.$$

The function p^{up} satisfies the equation $Lp^{up} = 0$, where L is the generator of the process $Q = (Q^b, Q^a)$. Hence p^{up} is an harmonic function for the process (Q^b, Q^a), and the process $(p^{up}(Q^b_t, Q^a_t))_{t\geq 0}$ is a martingale. We proved that

$$\forall \ s \leq t < \tau_1, \quad \mathbb{E}[\hat{S}_t | \mathcal{F}_s] = \hat{S}_s.$$

By recurrence on k, one can easily notice that

$$\forall \ k \ \geq 1, \ \forall \tau_k \leq t < \tau_{k+1}, \quad \mathbb{E}[\hat{S}_t | \mathcal{F}_{\tau_1}] = \hat{S}_{\tau_1}.$$

Assuming $s \leq \tau_1 \leq t$,

$$\begin{aligned}
\mathbb{E}[\hat{S}_t | \mathcal{F}_s] &= \mathbb{E}[\hat{S}_t | \mathcal{F}_s, X_1 = 1]\mathbb{P}[X_1 = +\delta | \mathcal{F}_s] + \mathbb{E}[\hat{S}_t | \mathcal{F}_s, X_1 = -\delta]\mathbb{P}[X_1 = -\delta | \mathcal{F}_s], \\
&= \mathbb{E}[\hat{S}_t | \mathcal{F}_s, X_1 = \delta]\phi(Q^b_t, Q^a_t) + \mathbb{E}[\hat{S}_t | \mathcal{F}_s, X_1 = -\delta](1 - p^{up}(Q^b_t, Q^a_t), \\
&= (S_s + \delta)p^{up}(Q^b_t, Q^a_t) + s_s(1 - p^{up}(Q^b_t, Q^a_t), \\
&= S_s + \delta(2p^{up}(Q^b_t, Q^a_t) - 1) = \hat{S}_s,
\end{aligned}$$

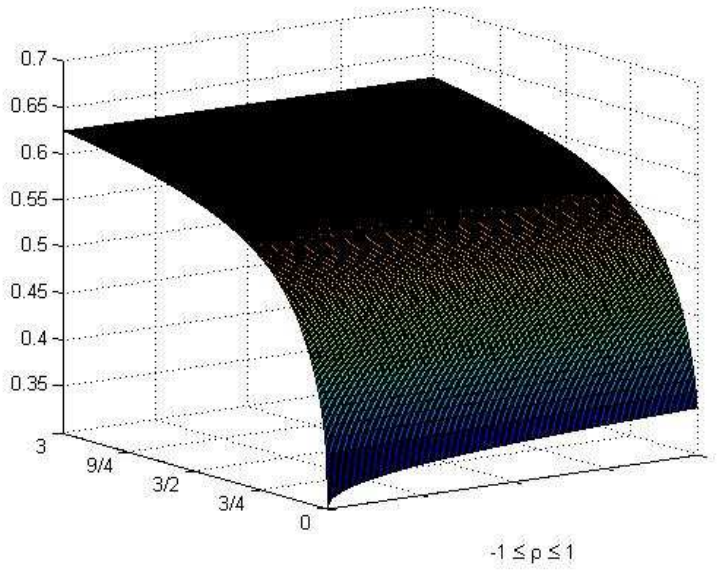

Figure 4.10: p^+ as a function of ρ and α for $-1 \leq \rho \leq 1$ and $0 \leq \alpha \leq 3$.

which completes the proof. □

Contrarily to the 'latent price' models alluded to above, here $\hat{S}$ is a function of the state variables (S_t, Q_t^b, Q_t^a) and thus is observable, provided one observes trades and quotes.

In the case where the correlation between the bid and the ask queues equals -1, $\hat{S}$ coincides with the definition given in Burghardt et al. (2006) since

$$p^{\text{up}}(Q_t^b, Q_t^a) = \frac{Q_t^a}{Q_t^a + Q_t^b}.$$

Remark 4.4. *When $p^+ \neq 1/2$, or $p^- \neq 1/2$, the martingale property of the process $(\hat{S}_t)_{t\geq 0}$ disappears. The jump times $\tilde{S} = (\tilde{S}_t)_{t\geq 0}$ are the hitting times of the axes by the process $Q = (Q^b, Q^a)$. When $p^+ < 1/2$ (resp. $p^+ > 1/2$), the jump is negative (resp. positive) after a price increase. Similarly, when $p^- < 1/2$ (resp. $p^- > 1/2$), the jump is positive (resp. negative) after a price decrease.*

4.3 Price dynamics at lower frequencies: fluid limits and diffusion limits

Let $(\tau_1, ..., \tau_k, ...)$ the sequence of times when the price S moves and $(X_1, ..., X_n, ...)$ the sequence of consecutive price moves. The high frequency dynamics of the price is described by a piecewise constant stochastic process $S_t = R_{N_t}$ where

$$R_n = X_1 + ... + X_n, \quad and \quad N_t = \sup\{k;\ \tau_1 + ... + \tau_k \leq t\} \tag{4.21}$$

is the number of price moves during $[0, t]$.

Over time scales much larger than the interval between individual order book events, prices are observed to have diffusive dynamics and modeled as such. To establish the link between the high frequency dynamics and the diffusive behavior at longer time scales, we shall consider a time scale $t\zeta(n)$ where $\zeta(n) \to \infty$ and exhibit conditions under which the rescaled price process

$$S^n := \left(\frac{S_{t\zeta(n)}}{\sqrt{n}}\right)_{t\geq 0}$$

satisfies a functional central limit theorem i.e. converges in distribution to a non-degenerate process $(P_t)_{t\geq 0}$ as $n \to \infty$:

$$(S^n_t, t \geq 0)_{n\geq 1} \Rightarrow (P_t, t \geq 0) \quad on \ \ (\mathcal{D}, J_1) \ \ \text{as} \ \ n \to \infty.$$

In this section we show that the mean reversion parameters p^+ and p^-, defined equation (4.2.1) play a critical role in the dynamics of the price at lower frequencies. More precisely we prove that:

- If $p^+ = p^- = 1/2$, the price process $(S_t)_{t\geq 0}$ is simply a time changed random walk, whose diffusive limit $(P_t)_{t\geq 0}$ is a Brownian motion with no drift.
- Similarly, when $p^+ = p^- := p_{cont}$ (*symmetric case*), we show that the low-frequency price dynamics is a Brownian motion whose volatility is a multiple of $\sqrt{\frac{p_{cont}}{1-p_{cont}}}$.
- When $p^+ \neq p^-$ (*asymmetric case*), the low-frequency price process $P = (P_t)_{t\geq 0}$ is a Brownian motion with drift. Both the drift and the volatility component of $(P_t)_{t\geq 0}$ can be related to the parameters p^+ and p^-.

4.3.1 Representation in terms of a Continuous Time Random Walk

Given a sequence of bivariate IID random variables $(Y_n, Z_n)_{n\geq 1}$, a *continuous time random walk* (CTRW) $(W_t)_{t\geq 0}$ is a piecewise constant stochastic process, for which the n-th price increment is the random variable Z_n. The durations between the $n-1$th and the n-th move is the random variable Y_n. A continuous time random walk is similar to a regular random walk, but has the property that the distribution of the waiting time between two increments and the distribution of the increments are not independent. CTRWs have been widely studied in physics, to model anomalous diffusions by incorporating a random waiting time between jumps.

When $p^+ \neq 1/2$ or $p- \neq 1/2$, the process $(R_n)_{n\geq 1}$ defined by

$$\forall n \geq 1, \quad R_n = X_1 + ... + X_n$$

is not a Markov chain therefore one can not use the classical central limit theorems. However, as observed on Figure 4.11, one can split the trajectory of $(R_n, n \geq 1)$ in a sequence of IID pieces of trajectory (or 'excursions') of lengths $(Y_n)_{n\geq 1}$ and heights $(Z_n)_{n\geq 1}$, with the property that the sequence

$(Y_n, Z_n)_{n\geq 1}$ is a sequence of IID random variables. This means that one can build a *continuous time random walk* from the sequence of price increments $(X_1, ..., X_n)$.

Let us first define two sequences of random variables $(\gamma_n^1,\ n \geq 1)$ and $(\gamma_n^2,\ n \geq 1)$ recursively by

$$\gamma_1^1 = \sup\{p \geq 1,\ X_1 = X_2 = ... = X_p\} \ \ and \ \ \gamma_1^2 = \sup\{p \geq \gamma_1^1,\ X_{\gamma_1^1+1} = ... = X_p\},$$

and for $k \geq 2$,

$$\gamma_k^1 = \sup\{p \geq \gamma_{k-1}^2,\ X_{\gamma_{k-1}^2+1} = ... = X_p,\} \ \ and \ \ \gamma_k^2 = \sup\{p \geq \gamma_k^1,\ X_{\gamma_1^k+1} = ... = X_p\}.$$

We call the sequence $(X_{\gamma_{k-1}^2+1}, ..., X_{\gamma_k^1}, X_{\gamma_1^k+1}, ..., X_{\gamma_k^2})$ a price *excursion* (more precisely, this is the k-th price excursion).

Proposition 4.3. *The sequence* $((Y_n, Z_n), n \geq 1)$*, defined by*

$$Y_n = \gamma_{n+1}^2 - \gamma_n^2 \ \ and \ \ Z_n = X_{\gamma_{n+1}^2} - X_{\gamma_n^2}. \tag{4.22}$$

is an IID sequence with the following distribution:

$$\mathbb{P}(Y_1 = k+q, Z_1 = k-q) = (p^+)^{k-1}(1-p^+)(p^-)^{q-1}(1-p^-).$$

Proof. Assume $X_1 = \delta$. The probability that $X_2 = ... = X_k = \delta$, $X_{k+1} = ... = X_{k+q} = -\delta$ and $X_{k+q+1} = \delta$ is clearly:

$$\mathbb{P}[X_2 = ... = X_k = \delta,\ X_{k+1} = ... = X_{k+q} = -\delta,\ X_{k+q+1} = \delta] = (p^+)^{k-1}(1-p^+)(p^-)^{q-1}(1-p^-).$$

This is exactly the probability of observing k price increases followed by q price decreases. i.e. the probability that the *price excursion* has a length of $k+q$ and a height of $k-q$:

$$\mathbb{P}[Y_1 = k+q, Z_1 = k-q] = (p^+)^{k-1}(1-p^+)(p^-)^{q-1}(1-p^-).$$

Assume that $X_1 = \delta$ and let $(k_1, q_1, k_2, q_2) \in \mathbb{N}^2$,

$$\begin{aligned}
&\mathbb{P}[Y_1 = k_1+q_1, Z_1 = k_1-q_1, Y_1 = k_2+q_2, Z_2 = k_2-q_2] = \\
&\mathbb{P}[X_2 = ... = X_{k_1+1} = \delta, X_{k_1+2} = ... = X_{k_1+q_1+2} = -\delta, \\
&X_{k_1+q_1+3} = ... = X_{k_1+q_1+3+k_2} = \delta, X_{k_1+q_1+k_2+4} = X_{k_1+q_1+k_2+q_2+5} - \delta] = \\
&(p^+)^{k_1-1}(1-p^+)(p^-)^{q_1-1}(1-p^-)(p^+)^{k_2-1}(1-p^+)(p^-)^{q_2-1}(1-p^-) = \\
&\mathbb{P}[Y_1 = k_1+q_1, Z_1 = k_1-q_1]\mathbb{P}[Y_2 = k_2+q_2, Z_2 = k_2-q_2]
\end{aligned}$$

Therefore, (Y_1, Z_1) and (Y_2, Z_2) are independent random variables. Replacing δ by $-\delta$ one can show similarly the independence of (Y_1, Z_1) from (Y_2, Z_2) given $X_1 = -\delta$. A reasoning by induction shows that $(Y_k, Z_k)_{k\geq 1}$ is an IID sequence of random variables, which completes the proof. □

Now denote by τ_+ (resp. τ_-) the duration until the next price move after a price increase (resp. price decrease). The duration of the i-th *price excursion* T_i is then given by

$$T_i \stackrel{d}{=} \underbrace{\tau_+^{\gamma_{i-1}^2+1} + ... + \tau_+^{\gamma_i^1}}_{\gamma_i^1 - \gamma_{i-1}^2} + \underbrace{\tau_-^{\gamma_i^1+1} + ... + \tau_+^{\gamma_i^2}}_{\gamma_i^2 - \gamma_i^1}$$

where $(\tau_+^j, j \geq 1)$ (resp. $(\tau_-^j, j \geq 1)$) are IID copies of τ_+ (resp. τ_-). Now define the following processes

$$\forall t \geq 0, \ \ M_t = \sup\{k;\ \ T_1 + ... + T_k \leq t\}, \ \ \text{and} \ \ W_t = \sum_{i=1}^{M_t} Z_i. \tag{4.23}$$

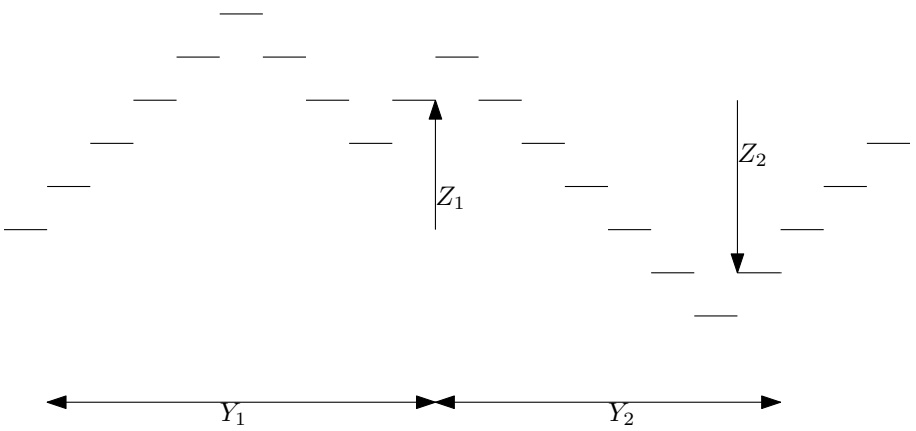

Figure 4.11: Decomposition of $(X_n, n \geq 1)$ on excursions of lengths $(Y_n, n \geq 1)$ and heights $(Z_n, n \geq 1)$

The process $(W_t, t \geq 0)$ is a also a Continuous time random walk, associated with the sequence of IID random variables $(T_i, Z_i)_{i\geq 1}$. The price process $(S_t)_{t\geq 0}$ can be decomposed as

$$\forall t \geq 0, \quad S_t = W_t + \sum_{Y_1+\ldots+Y_{M_t}+1}^{N_t} X_i,$$

where M_t is the process counting the number of price 'excursions' and N_t the process counting the number of price moves. This decomposition of the price process $S = (S_t)_{t\geq 0}$ as a continuous time random walk $W = (W_t)_{t\geq 0}$ plus a residual $\sum_{Y_1+\ldots+Y_{M_t}+1}^{N_t} X_i$ will be very useful to prove functional central limit theorems for the price $S = (S_t)_{t\geq 0}$.

4.3.2 Bid-ask symmetry: case when $p^+ = p^-$.

In all this section we assume that the distribution τ of the duration between two consecutive moves of the price has a finite moment of order two:

Assumption 4.4 (Finite second moment of durations)**.**

$$\forall (x, y) \in \mathbb{R}_+^2, \quad \mathbb{E}[\tau^2 | Q_0^b = x, Q_0^a = y] < \infty$$

This assumption holds in particular in two important cases:

- $\mathbb{E}[V_1^a] + \mathbb{E}[V_1^b] < 0$, where $(V_i^a, i \geq 1)$ (resp. $(V_i^b, i \geq 1)$) is the sequence of order sizes coming at the ask (resp. the bid). This assumption is satisfied if there are more market orders and cancelations than limit orders.
- $\mathbb{E}[V_1^a] = \mathbb{E}[V_1^b] = 0$: in this case, τ is the exit time from the positive quadrant of a planar Brownian motion without drift. τ is then heavy-tailed and fails to have a finite second moment if $\rho \geq 0$ where ρ is defined in equation (4.1). When this correlation is negative, (Cont and de Larrard 2012, Proposition 2) shows that

$$\mathbb{E}[\tau^2 | Q_0^b = x, Q_0^a = y] < \infty \quad \Longleftrightarrow \quad \rho < -\frac{1}{\sqrt{2}}.$$

We also make the following assumption, which holds when the statistical properties of the order flows at the bid and the ask are identical, but is in fact a weaker assumption since it only pertains to the quantities p^+ and p^-:

Assumption 4.5 (Up-down symmetry)**.**

$$p^+ = p^- := p_{cont}, \tag{4.24}$$

where p^+ and p^- are defined in equation (4.2.1)*.*

The following result, whose proof is given Appendix 4.5.2, describes the diffusion limit of the price process under these symmetry assumptions:

Theorem 4.2. *Let*

$$r_1(F) = \int_{R_+^2} F(dx, dy)\mathbb{E}[\tau | Q_0^b = x, Q_0^a = y], \tag{4.25}$$

be the expected duration until the next price change after a price increase and

$$r_1(\tilde{F}) = \int_{R_+^2} \tilde{F}(dx, dy)\mathbb{E}[\tau | Q_0^b = x, Q_0^a = y]. \tag{4.26}$$

be the expected duration until the next price change after a price decrease. Under assumptions 4.4 and 4.5,

$$\left(\frac{S_{nt}}{\sqrt{n}}\right)_{t\geq 0} \overset{n\to\infty}{\Rightarrow} \left(\delta\sqrt{\frac{2}{r_1(F)+r_1(\tilde{F})}\frac{p_{cont}}{1-p_{cont}}}B_t\right)_{t\geq 0}$$

on the Skorokhod space $(\mathcal{D}, J_1)$, where δ is the tick size and $B = (B_t)_{t\geq 0}$ is a Brownian motion.

Remark 4.5. *The proof given in Section 4.5.2 is based on the decomposition of S as a continuous time random walk W plus a residual. Alternatively, one can use the fact that the sequence $(X_n, n \geq 1)$ is a reversible random walk on the state $\{-\delta, \delta\}$, with transition matrix:*

$$\begin{pmatrix} p_{cont} & 1-p_{cont} \\ 1-p_{cont} & p_{cont} \end{pmatrix}. \tag{4.27}$$

Using (Aldous and Fill 2002, Ch 4,Prop. 29), since $(X_n, n \geq 1)$ is a reversible Markov chain on $\{-\delta, \delta\}$ satisfying $\mathbb{E}(X_n) = 0$,

$$\left(\sum_{i=1}^{[nt]} \frac{X_i}{\sqrt{n}}\right)_{t\geq 0} \overset{n\to\infty}{\Rightarrow} (\sigma B_t, t \geq 0)$$

on the Skorokhod space $(\mathcal{D}, J_1)$, where

$$\sigma^2 = 2(\frac{g_1^2}{\lambda_1} + \frac{g_2^2}{\lambda_2}) \quad with \quad g_m = \sum_{i\in\{\delta,-\delta\}} i\sqrt{\pi_i}u_{i,m}$$

where π is the stationary distribution of the Markov chain X and $u_{i,m}$ is the i-th component of the m-th eigenvector of the transition matrix (4.27)*. The expressions for the eigenvalues and the eigenvectors of the transition matrix* (4.27) *computed in the proof of Theorem* 4.1 *then allow to conclude.*

Remark 4.6. *In most examples (see Table 4.4 and Figure 4.6), price changes exhibit negative autocorrelation at first lag at high frequency which corresponds to the case $p_{cont} < 1/2$. In this case, the multiplicative term $\frac{p_{cont}}{1-p_{cont}} < 1$ decreases price volatility. Otherwise, when $p_{cont} > 1$, the price has a 'trend-following' behavior and the term $\frac{p_{cont}}{1-p_{cont}}$ increases price volatility*

4.3.3 Asymmetric order book: case when $p^+ \neq p^-$

When the skewness of F and the skewness of $\tilde{F}$ are not equal, or when the order flow has different statistical properties at the bid and at the ask, the mean reversion parameters can be different:

$$p^+ \neq p^-.$$

Empirically (cf table 4.4 above), these mean reversion parameters are slightly different. We will see in this section that this difference generates a drift in the price that can be related to these mean reversion parameters p^+ and p^-.

Let $r_k(F)$ be, for $k \in \{1, 2\}$, the k-th moment of the duration between price moves after a price increase

$$r_k(F) = \int_{R_+^2} F(dx, dy)\mathbb{E}[\tau^k | Q_0^b = x, Q_0^a = y], \tag{4.28}$$

and $r_k(\tilde{f})$ the k-th moment of the duration between price moves after a price decrease:

$$r_k(\tilde{F}) = \int_{R_+^2} \tilde{F}(dx, dy)\mathbb{E}[\tau^k | Q_0^b = x, Q_0^a = y]. \tag{4.29}$$

Theorem 4.3. *Under assumption 4.4,*

$$\left(\frac{S_{nt}}{n}, t \geq 0\right)_{n\geq 1} \Rightarrow (\delta t d_p, t \geq 0) \quad on \quad (\mathcal{D}, J_1), \tag{4.30}$$

where δ is the tick size and

$$d_p = \frac{\dfrac{1}{1-p^+} - \dfrac{1}{1-p^-}}{(1 + \dfrac{p^+}{1-p^+})r_1(F) + (1 + \dfrac{p^-}{1-p^-})r_1(\tilde{F})} \tag{4.31}$$

Proof. See section 4.5.3. □

Theorem 4.4. *Under assumption 4.4,*

$$\left(\frac{S_{nt} - d_p nt}{\sqrt{n}}, t \geq 0\right)_{n\geq 1} \Rightarrow^{\mathcal{D}} (\delta \sigma_p B_t, t \geq 0) \quad on \quad (\mathcal{D}, J_1), \tag{4.32}$$

where δ is the tick size, $(B_t)_{t\geq 0}$ is a Brownian motion and:

$$\sigma_p^2 = \frac{A + B + C}{(1 + \dfrac{p^+}{1-p^+})r_1(F) + (1 + \dfrac{p^-}{1-p^-})r_1(\tilde{F})} \quad with$$

$$A = \frac{(1+p^+)}{(1-p^+)^2} + \frac{(1+p^-)}{(1-p^-)^2} - \frac{2}{(1-p^+)(1-p^-)},$$

$$B = -2d_p\left(\frac{r_1(F)}{(1-p^+)^2} - \frac{r_1(\tilde{F})}{(1-p^-)^2} + \frac{r_1(F) - r_1(\tilde{F})}{(1-p^+)(1-p^-)}\right),$$

$$C = d_p^2\left(\frac{r_2(F)(1+p^+)}{(1-p^+)^2} + \frac{r_2(\tilde{F})(1+p^-)}{(1-p^-)^2} + \frac{2r_1(F)r_1(\tilde{F})}{(1-p^+)(1-p^-)}\right).$$

Proof: See section 4.5.3.

Remark 4.7. *If $d_p << 1$,*

$$A+B+C = \frac{(1+p^+)}{(1-p^+)^2} + \frac{(1+p^-)}{(1-p^-)^2} - 2\frac{1}{(1-p^+)(1-p^-)} + d_p\left(\frac{r_1(F)}{(1-p^+)^2} - \frac{r_1(\tilde{F})}{(1-p^-)^2} + \frac{r_1(F)-r_1(\tilde{F})}{(1-p^+)(1-p^-)}\right) + o(d_p).$$

In practice, the error made when using the simplified formula:

$$\sigma_p^2 \simeq \frac{(1+p^+)}{(1-p^+)^2} + \frac{(1+p^-)}{(1-p^-)^2} - 2\frac{1}{(1-p^+)(1-p^-)}$$

is less than 10^{-4}.

Remark 4.8. *Roginsky (1994) estimates the speed of convergence of this process to its Brownian limit as a function of the moments of Z_is. When $\mathbb{E}[Z_1^3] < \infty$, it is proven that the oscillation between $(\frac{S_{nt}}{\sqrt{n}})_{t\geq 0}$ and its Brownian limit $(P_t)_{t\geq 0}$ are bounded by $O(\frac{\log(n)}{\sqrt{n}})$. This result can be applied here, all moments of Z_i being finite.*

Remark 4.9. *The heavy-traffic approximation of theorem 4.4 is valid only when $\mathbb{E}[\tau^2] < \infty$. If, for instance, the average duration between two consecutive price moves becomes infinite, the rescaling given in theorem 4.4 is no longer valid. The decomposition of the price process $(S_n, n \geq 1)$ as a sequence of independent excursions of heights Z_n and durations T_n would remain true but T_n would not a finite moment of order two. Becker-Kern et al. (2004) provides a way to compute the heavy traffic limit of a general continuous time random walk. Interestingly, the set of possible limits is much wider than the set of stable processes. In particular, the heavy-traffic approximation of a continuous time random walk can be non-Markovian.*

4.3.4 Empirical test using high-frequency data

Theorems 4.3 and 4.4 relate the 'coarse-grained' drifts and volatility of intraday returns at lower frequencies to the parameters p^+, p^- and the average price durations $r_1(F)$ and $r_1(\tilde{F})$. Denote by γ_0 the average time between two orders. Typically γ_0 is of the order of milliseconds. In plain terms, Theorems 4.3 and 4.4 state that, observed over a time scale $\gamma_2 >> \gamma_0$ (say, 10 minutes), the price behaves like a Brownian motion with drift:

$$P_t \approx n\delta t d_p + \sqrt{n}\delta\sigma_p B_t.$$

where δ is the tick size, $n = \frac{\gamma_2}{\gamma_0}$ represents the average number of orders during an interval γ_2, and σ_p, d_p are given in equations (4.32), (4.31).

Figure 4.12, compares, for stocks in the Dow Jones index, the standard deviation of 10-minute price increments with

$$\delta\sqrt{\frac{\frac{(1+p^+)}{(1-p^+)^2} + \frac{(1+p^-)}{(1-p^-)^2} - \frac{2}{(1-p^+)(1-p^-)}}{(1+\frac{p^+}{1-p^+})r_1(F) + (1+\frac{p^-}{1-p^-})r_1(\tilde{F})}}.$$

We observe that the correlation between the theoretical expression of σ_p and the standard deviation of 10-minute price increments is high. A regression gives slope of 0.9664 and a R^2 of 0.77. Figure 4.13

compares, for the same stocks, the closing price minus its opening price (the "realized daily trend") with its model-based estimator

$$\delta \frac{\frac{1}{1-p^+} - \frac{1}{1-p^-}}{(1+\frac{p^+}{1-p^+})r_1(F) + (1+\frac{p^-}{1-p^-})r_1(\tilde{F})},$$

here p^+ and p^- are the mean reversion parameters defined and $r_1(F)$ (resp. $r_1(\tilde{F})$) is the average price duration after an increase (resp. decrease) of the price. Regression of the estimator d_p on the observed daily price trend yields a slope of 0.997 and a R^2 of 0.97.

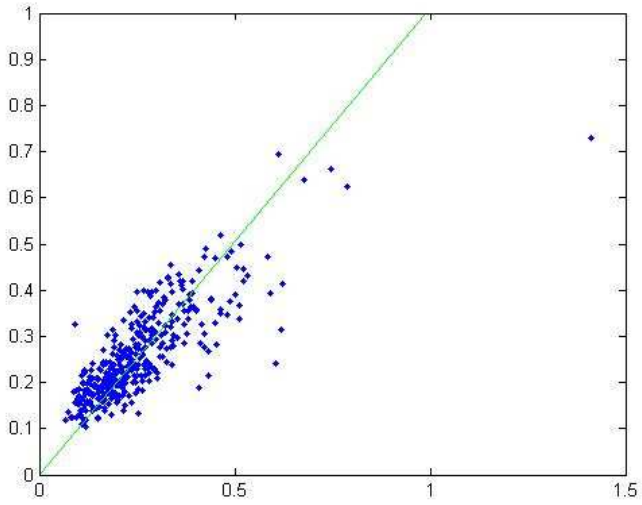

Figure 4.12: Annualized intraday price volatility estimated using formula (4.32) (vertical axis) vs annualized 10-minutes realized volatility estimated using high-frequency price changes (x-axis) for stocks in the Dow Jones Index. Each point corresponds to one stock on a given trading day.

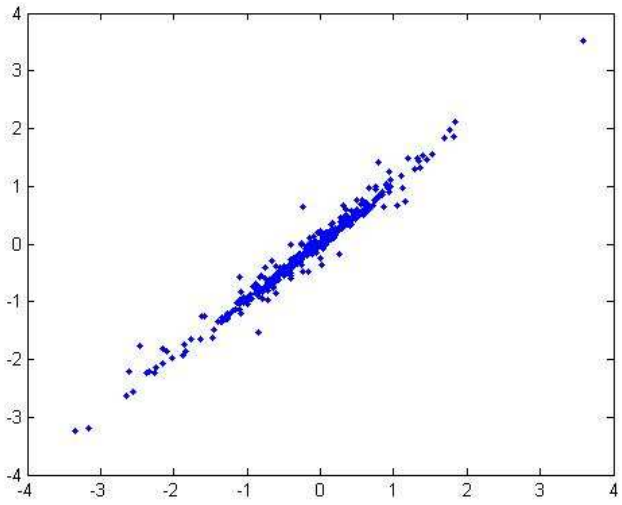

Figure 4.13: Daily drift of the price estimated with formula (4.31) (vertical axis) vs empirical daily drift of the price estimated using the closing mid-price minus the opening mid price (x-axis) for stocks in the Dow Jones Index. Each point corresponds to one stock on a given trading day.

4.4 Link between volatility and order flow

In the previous section, we linked the price volatility with the conditional probabilities parameters p^+, p^- and the average price duration after a price increase (resp. decrease) $r_1(F)$ (resp. $r_1(\tilde{F})$). In particular, when the order flow is symmetric at the bid and at the ask, the drift of the price is null and the price volatility is:

$$\sigma_p^2 = \delta^2 \frac{1}{r_1(F)} \frac{p_{cont}}{1 - p_{cont}}. \tag{4.33}$$

This equation (4.33) does not directly link the price volatility with the order flow. It is not clear how the trading intensity, the variance of order sizes, or the probability distributions F and $\tilde{F}$ affect price volatility. If one can express both parameters p_{cont} and $r_1(F)$ (resp. $r_1(\tilde{F})$), with order flow statistics, we could have an explicit relation between the order flow and σ_p.

The average price duration $r_1(F)$ may be computed using:

$$r_1(F) = \int_0^\infty \int_0^\infty \mathbb{E}[\tau | Q_0^b = x, Q_0^a = y] F(dx, dy);$$

As shown by Klein (1952), when (Q^b, Q^a) is a continuous Markov process with generator $\mathcal{G}$,

$$h(x, y) := \mathbb{E}[\tau | Q_0^b = x, Q_0^a = y]$$

is a solution to the elliptic partial differential equation

$$\forall x > 0, \quad y > 0, \quad \mathcal{G}h(x, y) = -1, \qquad \text{and} \qquad h(x, 0) = 0, \quad h(0, y) = 0. \tag{4.34}$$

The 'continuation probability' p_{cont} satisfies:

$$p_{cont} = \int_0^\infty \int_0^\infty p^{\text{up}}(x, y) F(dx, dy),$$

where p^{up} is the probability of a price increase conditioned on observing x shares at the bid and y shares at the ask. In Cont and de Larrard (2011) we proved that, when (Q^b, Q^a) is a Markovian process, with generator $\mathcal{G}$, p^{up} follows a partial differential equation:

$$\forall x > 0, \quad y > 0, \quad \mathcal{G}p^{\text{up}}(x, y) = 0, \qquad \text{and} \qquad p^{\text{up}}(x, 0) = 1, \quad p^{\text{up}}(0, y) = 0. \tag{4.35}$$

In summary, σ_p is linked to the order flow parameters via the solutions of the two elliptic equations (4.34) and (4.35). In this section we will solve these partial differential equations for particular regimes of the limit order book, when market orders and cancelations dominate the flow of limit orders and when $\overline{V^a} = \overline{V^b}$.

For a general regime of limit order book, when the order book Q has both a Brownian component and a non null drift, one can either solve numerically these partial differential or compute complex analytical expressions of p^{up} and h by solving Laplace equations and Helmholtz equations in the quadrant (see Spence (2010) for a review of these methods).

4.4.1 Case when market orders and cancelations dominate

In Cont and de Larrard (2011), we derived diffusion approximations for a limit order book under the condition that the imbalance between limit orders ion one hand and market orders and cancelations on the other hand is small with respect to the arrival intensity of these orders:

$$- \frac{\sqrt{n}\mathbb{E}[V_1^a]}{\mathbb{E}[T_1^a]\sqrt{\text{Var}(V_1^a)}} << 1, \tag{4.36}$$

where n denotes the number of orders arriving in the order book per unit time. Under this conditions, the bid/ask queue sizes (q^b, q^a) may be approximated by their diffusion limit (Q^b, Q^a) whose generator is given equation (4.4). However, when the flow of market orders and limit orders dominates the flow of limit orders,

Assumption 4.6.

$$-\frac{\sqrt{n}\mathbb{E}[V_1^a]}{\mathbb{E}[T_1^a]\sqrt{Var(V_1^a)}} >> 1.$$

we recover another asymptotic regime, the *fluid limit*: under Assumption 4.6, the sequence of processes

$$\left(\frac{q^b_{nt}}{n}, \frac{q^a_{nt}}{n}, t \geq 0\right)_{n\geq 1} \tag{4.37}$$

converges to a piecewise deterministic process $\tilde{Q} = (\tilde{Q}^b, \tilde{Q}^a)$ whose generator $\tilde{\mathcal{G}}$ is given by

$$\tilde{\mathcal{G}}h(x,y) = \frac{\mathbb{E}[V_1^a]}{\mathbb{E}[T_1^a]}\frac{\partial h}{\partial y} + \frac{\mathbb{E}[V_1^b]}{\mathbb{E}[T_1^b]}\frac{\partial h}{\partial x}, \tag{4.38}$$

whose domain is the set $dom(\tilde{\mathcal{G}})$ of functions $h \in C^1(]0,\infty[\times]0,\infty[,\mathbb{R}) \cap C^0(\mathbb{R}^2_+,\mathbb{R})$ verifying the "boundary conditions"

$$\int_{\mathbb{R}^2_+}(h(u,v)-h(0,y))F(du,dv) = 0, \qquad \int_{\mathbb{R}^2_+}(h(u,v)-h(x,0))\tilde{F}(du,dv) = 0.$$

In this section we will assume that the order flow is symmetric at the bid and at the ask:

Assumption 4.7 (bid-ask symmetry)**.** *F and $\tilde{F}$ have densities f and $\tilde{f}$ with*

$$\forall x>0, \quad \forall y>0, \qquad f(x,y) = \tilde{f}(y,x) \quad \textit{and} \quad \frac{\mathbb{E}[V_1^a]}{\mathbb{E}[T_1^a]} = \frac{\mathbb{E}[V_1^b]}{\mathbb{E}[T_1^b]} := \frac{\mathbb{E}[V_1]}{\mathbb{E}[T_1]}.$$

Under this assumption one can solve the partial differential equations (4.35)-(4.34) with the generator (4.54):

Proposition 4.4. *Under assumption 4.7, $p^{up}(x,y) = 1_{x<y}$ and*

$$\mathbb{E}[\tau|(\tilde{Q}^b_0,\tilde{Q}^a_0) = (x,y)] = x\frac{\mathbb{E}[V_1]}{\mathbb{E}[T_1]} \quad \textit{if} \quad y>x \tag{4.39}$$

$$= y\frac{\mathbb{E}[V_1]}{\mathbb{E}[T_1]} \quad \textit{if} \quad x>y \tag{4.40}$$

Proof. The process $\tilde{Q}$ is deterministic between two hitting times of the axes. Starting from $(\tilde{Q}^b_0,\tilde{Q}^a_0) = (x,y)$, the process $\tilde{Q}$ evolves vertically (resp. horizontally) and will reaches the x-axis (resp. y-axis) after a duration $y\frac{\mathbb{E}[V_1]}{\mathbb{E}[T_1]}$ (resp. $x\frac{\mathbb{E}[V_1]}{\mathbb{E}[T_1]}$) when $y<x$ (resp. $y>x$). Therefore $p^{up}(x,y) = 1_{x<y}$ and, since τ is deterministic conditional on $x<y$ or $x>y$, we obtain the result. □

We observe in particular that the fluid limit, advocated in some studies as a valid approximation such as optimal order execution, leads to trivial 'predictions' for price dynamics such as the one given in proposition 4.4. Therefore results based on such fluid limits should be examined with care.

Proposition 4.5. *Under the assumptions 4.6 and 4.7,*

$$\left(\frac{S_{[nt]}}{\sqrt{n}}\right)_{t\geq 0} \overset{n\to\infty}{\Rightarrow} (\sigma_p B_t)_{t\geq 0},$$

where B is a Brownian motion, $p_{cont} := p^+ = p^-$, δ is the tick size and

$$\sigma_p^2 = \delta^2 \frac{1}{r_1(F)} \frac{p_{cont}}{1-p_{cont}}$$

and $r_1(F)$ is the average time between two consecutive price moves.

Proof. Under assumption 4.7, it is clear that $p^+ = p^- := p_{cont}$. Therefore one can apply Theorem 4.4. □

Proposition 4.5 allows to retrieve the results of Cont and de Larrard (2010):

Example 4.1 (Markovian limit order book)**.** *In Cont and de Larrard (2010), we studied a simple example of Markovian limit order book in which*

- *Market buy (resp. sell) orders and cancelations arrive at independent, exponential times with rate μ,*
- *Limit buy (resp. sell) orders at the (best) bid (resp. ask) arrive at independent, exponential times with rate λ,*
- *These events are mutually independent.*
- *All orders sizes are equal (assumed to be 1 without loss of generality).*
- *F has a density f which satisfy $f(x,y) = f(y,x)$.*

Under this last assumption, $p^+ = p^- = 1/2$ and Proposition 4.5 implies that when $\lambda < \mu$, the diffusion limit of the price process has a variance per unit time given by

$$\sigma_p^2 = \delta^2 \frac{1}{r_1(F)},$$

which is consistent with the formula given in Cont and de Larrard (2010). Proposition 4.5 allows to generalize this formula to the case where the angular part of F is skewed: then p_{cont} is in general different from $1/2$ and the formula for price volatility takes the more general form

$$\sigma_p^2 = \delta^2 \frac{1}{r_1(F)} \frac{p_{cont}}{1-p_{cont}}.$$

When F has a density f which is given by the parametric form (4.15):

$$f(r,\theta) = c^2 e^{-cr} (\frac{2}{\pi})^\alpha \alpha \theta^{\alpha-1}. \tag{4.41}$$

one can further express the parameters p_{cont} and $r_1(F)$ appearing in proposition 4.5 as a function of statistical properties of the order flow.

Proposition 4.6. *Under the assumptions 4.6 and 4.7, when F follows the parametric representation from equation (4.15), $p_{cont} = 2^{-\alpha}$ and the variance per unit time of price changes is given by*

$$\sigma_p^2 = \delta^2 \ \frac{(\pi/2)^\alpha}{(2^\alpha - 1)\int_0^{\pi/4} \sin(\theta)\theta^{\alpha-1} d\theta + \int_{\pi/4}^{\pi/2} \cos(\theta)\theta^{\alpha-1} d\theta} \ \frac{c}{2} \ (\mu - \lambda)$$

where δ is the tick size,

- $\frac{p_{cont}}{1-p_{cont}} = \frac{1}{2^\alpha - 1}$ *is the factor due to high-frequency mean-reversion,*
- $c/2$ *is the average depth of the limit order book after a price change.*
- $\mu - \lambda$ *is the average growth rate of the queue sizes, i.e. the difference between the rate at which orders leave the order book (either by cancelation or execution) and the rate at which they enter the bid/ask queues.*

Proof. As shown in Proposition 4.4, $p^{up}(r,\theta) = 1_{\theta \in [0,\pi/4]}$. Therefore,

$$p_{cont} = \int_0^\infty \int_0^{\pi/2} f(r,\theta) p^{up}(r,\theta) r dr d\theta = (2/\pi)^\alpha \alpha \int_0^{\pi/4} \theta^{\alpha-1} d\theta = (\frac{1}{2})^\alpha.$$

An immediate computation then shows that

$$\frac{p_{cont}}{1-p_{cont}} = \frac{(\frac{1}{2})^\alpha}{1-(\frac{1}{2})^\alpha} = \frac{1}{2^\alpha - 1}.$$

To apply theorem 4.4, we just need to compute $r_1(F)$, the average time between two consecutive price moves:

$$r_1(F) = \int_0^{\pi/2} \int_0^\infty c^2 e^{-cr} \alpha (2/\pi)^\alpha \theta^{\alpha-1} \mathbb{E}[\tau | (\tilde{Q}_0^b, \tilde{Q}_0^a) = (r\cos(\theta), r\sin(\theta))] r dr d\theta.$$

The process $\tilde{Q} = (\tilde{Q}^b, \tilde{Q}^a)$, starting from $(\tilde{Q}_0^b, \tilde{Q}_0^a) = (r\sin(\theta), r\cos(\theta))$ will hit the x-axis if $\theta < \pi/2$ at the time $\frac{r\sin(\theta)}{\mu-\lambda}$. If $\theta > \pi/4$, the process $\tilde{Q}$ will hit the y-axis at the time $\frac{r\cos(\theta)}{\mu-\lambda}$. Hence,

$$r_1(F) = \frac{1}{\mu-\lambda} \int_0^\infty c^2 r^2 e^{-cr} \alpha (2/\pi)^\alpha \left(\int_0^{\pi/4} \sin(\theta)\theta^{\alpha-1} d\theta + \int_{\pi/4}^{\pi/2} \cos(\theta)\theta^{\alpha-1} d\theta \right),$$

$$r_1(F) = \frac{1}{\mu-\lambda} \frac{2}{c} \alpha (2/\pi)^\alpha \left(\int_0^{\pi/4} \sin(\theta)\theta^{\alpha-1} d\theta + \int_{\pi/4}^{\pi/2} \cos(\theta)\theta^{\alpha-1} d\theta \right).$$

□

Remark 4.10. *The function* $\frac{(\pi/2)^\alpha}{(2^\alpha-1)\left(\int_0^{\pi/4} \sin(\theta)\theta^{\alpha-1} d\theta + \int_{\pi/4}^{\pi/2} \cos(\theta)\theta^{\alpha-1} d\theta\right)}$*, displayed in Figure 4.14, decreases with* α *until* $\alpha = 1.4$*. After* $\alpha = 1.4$*, the skewness increases price volatility. Intuitively, there is a tradeoff between the term* $p_{cont}/(1-p_{cont})$*, which decreases the volatility when the skewness* α *increases and the tern* $1/r_1(F)$ *which increases with the parameter* α*.*

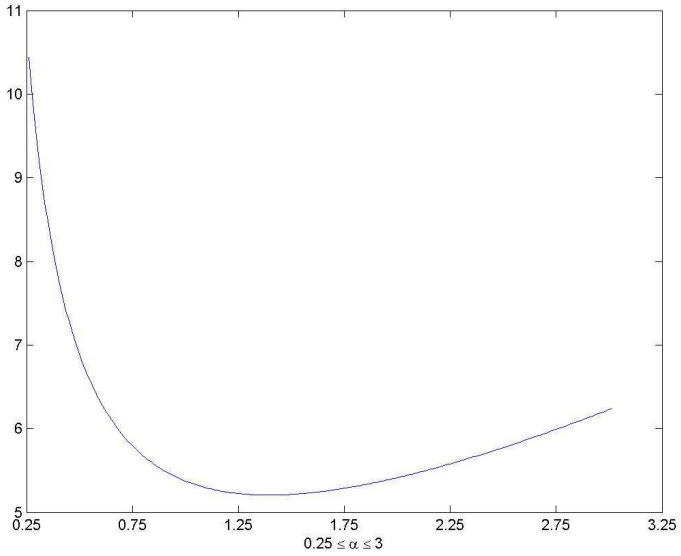

Figure 4.14: Price volatility as function of bid/ask imbalance parameter α in model 4.6.

4.4.2 A symmetric driftless limit order book

When the dynamics of the order book behaves as a symmetric driftless Brownian motion, one will see in this section that it is possible to explicitly link the volatility of the price to the statistical properties of the order flow.

Case when bid and ask queue sizes are independent

In Cont and de Larrard (2010), we studied a simple example of Markovian limit order book in which:

- Market buy (resp. sell) orders and cancelations arrive at independent, exponential times with rate μ,
- Limit buy (resp. sell) orders at the (best) bid (resp. ask) arrive at independent, exponential times with rate λ,
- These events are mutually independent.
- All orders sizes are equal (assumed to be 1 without loss of generality).
- The price does not mean revert: $p^+ = p^- = 1/2$.

We proved that, when $\lambda = \mu$, the diffusion limit of the price process is a Brownian motion:

$$\left(\frac{s_{n\log n\ t}}{n}, t \geq 0\right)_{n\geq 1} \Rightarrow \left(\delta\sqrt{\frac{\pi\lambda}{D(F)}}B_t, t \geq 0\right) \quad \text{on} \quad (\mathcal{D}, J_1), \tag{4.42}$$

where $D(F)$, whose square root is the *geometric mean* of the bid and ask queue sizes after a price change, characterizes the depth of the order book (q^b, q^a) after a price move:

$$D(F) = \int_{\mathbb{R}_+^2} xyF(dx, dy) \tag{4.43}$$

The Markovian order flow process described in Cont and de Larrard (2010), verifies assumptions 4.2 and 4.3 with the characteristics:

$$\mu^a = \mu^b = 0, \quad \lambda^a v_a^2 = \lambda^a v_a^2 = 2\lambda, \quad \rho = 0 \quad \text{and} \quad p^+ = p^- = 1/2.$$

We now show that the same diffusion limit can be derived for a more general order flow, without making use of the Markov property. We consider the case where the parameter describing the diffusion limit (Q^b, Q^a) of the order book satisfy:

Assumption 4.8.

$$\lambda^a\sigma_a^2 = \lambda^b\sigma_b^2, \quad \mu_a = \mu_b = 0 \quad \textit{and} \quad \rho = 0$$

Theorem 4.5. *Under assumption 4.8,*

$$\left(\frac{S_{n\log n\ t}}{\sqrt{n}}, t \geq 0\right)_{n\geq 1} \Rightarrow \left(\delta\sqrt{\frac{p_{cont}}{1-p_{cont}}\frac{\pi\lambda^a\sigma_a^2}{2D(F)}}B_t, t \geq 0\right) \quad \textit{on} \quad (\mathcal{D}, J_1), \tag{4.44}$$

where B is a Brownian motion and $D(F)$ is given by (4.43).

Proof. When $\rho = 0$, the two components Q^a and Q^b of the diffusion limit are independent Brownian motions. Since Q^a is a Brownian motion with volatility $\sqrt{\lambda^a}\sigma_a$, for $r > 0$ the process $M^r = (M_t^r)_{t\geq 0}$ defined as:

$$\forall t \geq 0, \qquad M_t^r = e^{r\sqrt{\sqrt{\lambda^a}\sigma_a}Q_t^a - r^2\sqrt{\lambda^a}\sigma_a t}$$

is a martingale. Therefore if one defines σ^a the first time Q^a hits 0:

$$\sigma^a = \inf\{t \geq 0,\ Q_t^a = 0\},$$

one has that for $r \geq 0$,

$$\mathbb{E}[e^{-r\sigma^a}|Q_0^a = x] = e^{-\frac{\sqrt{2r}}{\sigma_a\sqrt{\lambda^a}}x};$$

and by the Tauberian theorem of Karamata-Littlewood Feller (1971),

$$\mathbb{P}[\sigma^a > t|Q_0^a = x] \sim_{t\to\infty} \frac{\sqrt{2}x}{\sqrt{\pi}\sqrt{\sqrt{\lambda^a}\sigma_a}\sqrt{t}}.$$

Hence

$$\mathbb{P}[\tau > t|Q_0^a = x, Q_0^b = y] \sim_{t\to\infty} \frac{2xy}{\pi\sqrt{\lambda^a}\sigma_a t}.$$

Using (Cont and de Larrard 2010, Lemma 1) then implies

$$\frac{\tau_1 + \tau_2 + ... + \tau_n}{n\log n} \to \frac{2D(F)}{\pi\lambda^a\sigma_a^2} \quad \text{as} \quad n \to \infty$$

The rest of the proof is very similar to the proof of (Cont and de Larrard 2010, Theorem 2). □

The decomposition $f(r,\theta) = g(\theta)h(r)$ of the density of F allows to express price volatility as

$$\sigma_p^2 = \underbrace{\delta^2}_{\text{tick}} \frac{4}{\pi} \frac{\overbrace{\lambda^a v_a^2}^{\text{variance of order sizes/unit time}}}{\underbrace{\int_0^\infty h(r)r^3dr \int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta}_{\text{Order book depth}}} \underbrace{\frac{p_{cont}}{1-p_{cont}}}_{\text{mean reversion}}. \tag{4.45}$$

As argued in Section 4.2.2 a flexible and empirically plausible parameterization of F is

$$F(dr, d\theta) = c^2 e^{-cr}(\frac{2}{\pi})^\alpha \alpha\theta^{\alpha-1}drd\theta \tag{4.46}$$

where $c > 0$ represents the depth of the order book (Q^b, Q^a) after a price move and $\alpha > 0$ parameterizes the asymmetry of F. In the next proposition, we express the price volatility when F has this parametric form:

Proposition 4.7. *If the density of F follows the parametric form $f(r,\theta) = c^2e^{-cr}(\frac{2}{\pi})^\alpha\alpha\theta^{\alpha-1}$, price volatility becomes:*

$$\sigma_p^2 = \delta^2 \frac{c^2}{6} \lambda^a\sigma_a^2 \frac{\alpha+1}{2\alpha\ {}_1H_2(\frac{\alpha+1}{2}; \frac{3}{2}, \frac{\alpha-1}{2}; -\frac{\pi^2}{4})},$$

where ${}_1H_2$ is the generalized hypergeometric function.

Proof. One can easily compute all the components in equation (4.45):

- The mean reverting component $\frac{p^+}{1-p^+} = \frac{1}{\alpha}$
- The average order book depth $\int_0^\infty r^3 c^2 e^{-cr} dr = \frac{6}{c^2}$
- The integral $\frac{1}{2}\int_0^{\pi/2} (\frac{2}{\pi})^\alpha \theta^{\alpha-1} \sin(2\theta) d\theta = \frac{\alpha+1}{\alpha} \frac{1}{2\ {}_1H_2(\frac{\alpha+1}{2}; \frac{3}{2}, \frac{\alpha-1}{2}; -\frac{\pi^2}{4})}$

□

As observed in Figure 4.14, price volatility decreases with α until $\alpha = 1.45$. After $\alpha = 1.45$, price volatility increases with the parameter α parameterizing the bid-ask asymmetry. Intuitively, there is a tradeoff between the term $p_{cont}/(1-p_{cont})$, which decreases the volatility when the skewness α increases and the term $1/r_1(F)$ which increases with the parameter α.

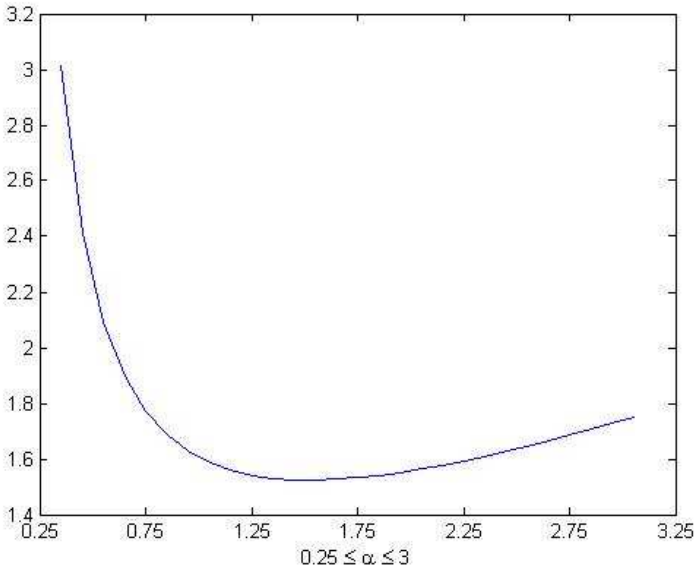

Figure 4.15: Price volatility as function of bid/ask imbalance parameter α in model from proposition 4.7

Case when bid and ask queue sizes are negatively correlated

In this subsection we consider the (frequently observed) case where the correlation between bid and ask queue sizes is negative.

Assumption 4.9. *F and $\tilde{F}$ have densities f, $\tilde{f}$ with:*

$$f(x,y) = \tilde{f}(y,x) \quad \textit{and} \quad \lambda^a \sigma_a^2 = \lambda^b \sigma_b^2, \quad \mu_a = \mu_b = 0, \quad \rho < 0.$$

The previous results show that the computation of $r_1(F)$ and $r_1(\tilde{F})$ are critical to estimate both the drift and the volatility of the price using statistics of the order flow. In Cont and de Larrard (2011), we gave an explicit expression of the distribution of the price durations τ. However, the numerical computation of $r_1(F)$ and $r_1(\tilde{F})$ using this formula seems challenging. The following proposition gives a simple expression for this quantity.

Proposition 4.8. *Under assumption 4.9,*

$$\mathbb{E}[\tau|Q_0^b = x, Q_0^a = y] = -\frac{xy}{2\rho\lambda^a\sigma_a^2}, \tag{4.47}$$

Proof. As shown by Klein (1952), $u : \mathbb{R}_+^2 \mapsto \mathbb{R}_+$ defined by

$$u(x,y) := \mathbb{E}[\tau|Q_0^b = x, Q_0^a = y].$$

is a positive solution to the boundary value problem

$$Lu = \lambda^a\sigma_a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + 2\rho\frac{\partial^2 u}{\partial x \partial y}\right) = -1, \tag{4.48}$$

in the positive quadrant $\mathbb{R}_+^2$ with the boundary conditions

$$\forall (x,y) \in \mathbb{R}_+^2, \qquad u(0,x) = u(0,y) = 0. \tag{4.49}$$

The scaling properties of Brownian motion imply that $u(\lambda x, \lambda y) = \lambda^2 u(x,y)$ or, in polar coordinates, $u(r,\theta) = r^2 u(1,\theta)$ (Spitzer (1958)). It is easily verified that $v(x,y) = -xy/2\rho\lambda^a\sigma_a^2$ is a positive solution of this boundary value problem, which grows polynomially at rate $r^2 = (x^2+y^2)$ as $(x,y) \to \infty$. The function $w = u - v$ is then a harmonic function in the orthant, solution of

$$\forall (x,y) \in \mathbb{R}_+^2, \qquad Lw(x,y) = 0 \qquad\qquad w(0,x) = w(0,y) = 0.$$

with $r^{-3}w(r,\theta) \to 0$ as $r \to \infty$. (Yoshida and Miyamoto 1999, Theorem 3.2) then implies that w is proportional to $r^2\sin(4\theta)$. So finally, we have shown that, in polar coordinates

$$\exists c \in \mathbb{R}, \qquad u(r,\theta) = v(r,\theta) + cr^2\sin(4\theta).$$

Since the second term changes sign, the only positive solution is given by $c = 0$ i.e. $u = v$. □

This following result follows immediately from Proposition 4.8:

Corollary 4.1. *Under Assumption 4.9, price volatility becomes*

$$\sigma_p^2 = \delta^2\frac{\lambda^a\sigma_a^2}{D(F)}\frac{p_{cont}}{1-p_{cont}}, \quad \textit{where} \qquad D(F) = \int_{\mathbb{R}_+^2} xyF(dx,dy).$$

Proposition 4.9. *Under assumption 4.9, if the density f of F has the polar decomposition:*

$$\forall (r,\theta) \in \mathbb{R}_+ \times [0,\pi/2], f(r,\theta) = h(r)g(\theta),$$

then the average time between two price moves $r_1(F)$ is given by

$$r_1(F) = r_1(\tilde{F}) = \frac{-1}{4\lambda^a\sigma_a^2\rho}\int_0^\infty r^3h(r)dr\int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta, \tag{4.50}$$

Proof. Assume that the density f of F follows a polar decomposition $f(r,\theta) = h(r)g(\theta)$. The average price duration $r_1(F)$ (resp. $r_1(\tilde{F})$) after a price increase (resp. decrease) is

$$\begin{aligned} r_1(F) = r_1(\tilde{F}) &= \int_{\mathbb{R}_+^2}\mathbb{E}[\tau|(x,y) = (r\cos(\theta), r\sin(\theta))]h(r)g(\theta)rdrd\theta \\ &= \frac{1}{\lambda^a\sigma_a^2}\int_0^\infty\int_0^{\pi/2}\left(\frac{-r^2\cos(\theta)\sin(\theta)}{2\rho}\right)h(r)g(\theta)rdrd\theta. \end{aligned}$$

Therefore the average time between two consecutive price move can be expressed as:

$$r_1(F) = r_1(\tilde{F}) = \frac{-1}{4\lambda^a\sigma_a^2\rho}\int_0^\infty r^3h(r)dr\int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta, \tag{4.51}$$

□

With the decomposition $F(r,\theta) = h(r)g(\theta)$ given in section 4.2.2, price volatility becomes

$$\sigma_p^2 = \underbrace{-4\rho}_{\text{Bid-Ask correlation}} \underbrace{\delta^2}_{\text{tick}} \frac{\overbrace{\lambda^a v_a^2}^{\text{variance of order sizes/unit time}}}{\underbrace{\int_0^\infty h(r)r^3dr}_{\text{Order book depth}} \int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta} \underbrace{\frac{p_{cont}}{1-p_{cont}}}_{\text{mean reversion}} . \tag{4.52}$$

The terms appearing in formula (4.52) have an intuitive interpretation:

- $\lambda^a v_a^2$ is the variance of the order book (Q^b, Q^a). λ^a is the *intensity* of orders arriving at the ask and v_a^2 is the variance of order sizes.
- $\frac{p_{cont}}{1-p_{cont}}$ comes from the *mean reversion* of the price. When the price mean-reverts, $p_{cont} < 1/2$ and $\frac{p_{cont}}{1-p_{cont}} < 1$. The price mean reversion decreases its volatility. On the other hand, when the price has a *trend following* behavior, $p_{cont} > 1$ and the volatility is increased by a factor $\frac{p_{cont}}{1-p_{cont}} > 1$. When price increments are not correlated, this parameter equals 1.
- $\int_0^\infty h(r)r^3dr$ measures the *average depth* of the order book (on both sides)
- $\int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta$: given that g represents the distribution of bid-ask asymmetry Q^a/Q^b after a price change, this term captures the influence of bid-ask imbalance on volatility.

Example 4.2. *When F has the parametric form* (4.15) *one can easily compute all the terms in Equation* (4.52)*:*

- *the mean reversion component* $\frac{p^+}{1-p^+} = \frac{1}{\alpha}$.
- *the average order book depth* $\int_0^\infty r^3 c^2 e^{-cr}dr = \frac{6}{c^2}$.
- *the term* $\int_0^{\pi/2} g(\theta)\sin(2\theta)d\theta = (2/\pi)^\alpha \alpha \int_0^{\pi/2} \theta^{\alpha-1}\sin(2\theta)d\theta$.

so the standard deviation of price changes is given by

$$\sigma_p^2 = \delta^2 \ \lambda^a v_a^2(-4\rho) \ \frac{c^2}{6} \frac{(\pi/2)^\alpha}{\alpha^2 \int_0^{\pi/2} \theta^{\alpha-1}\sin(2\theta)d\theta}.$$

Remark 4.11. *As it is shown on figure 4.16, price volatility is decreasing for $\alpha \leq \alpha^*$, and increasing in $\alpha \geq \alpha^*$ where $\alpha^* \simeq 1.85$.*

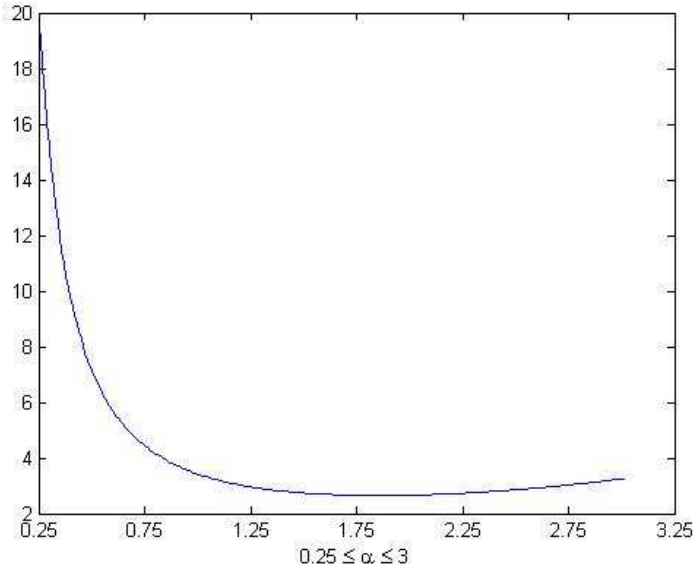

Figure 4.16: Volatility of price increments as a function of bid-ask imbalance parameter α in the parametric model of section 4.2.2.

4.4.3 "Flash crash": when sell orders overwhelm buy orders

In sections 4.4.1 and 4.4.2, we derived asymptotic expressions for price volatility in situations where the order flows at the bid and at the ask have comparable orders of magnitude.

However, when the selling pressure, represented by flow of market orders and cancelations at the bid plus the limit orders at the ask, dominate the flow of buy orders, as may be the case in a market crash, the magnitude of the order flow at the bid and the ask may be quite different. Such regimes occur in situations where the equilibrium between buyers and sellers is temporarily disrupted. We can represent such a situation using a two-scale model in which, the bid queue is depleted at rate which is an order of magnitude (say, $O(n^\beta)$ where n is the number of order book events per unit time) larger than the rate of change of the ask queue.

Assumption 4.10. *[Market crash]*

$$\left(\mathbb{E}[V_1^{n,b}], n^\beta \mathbb{E}[V_1^{n,a}]\right) \stackrel{n\to\infty}{\to} \left(\Pi^b, \overline{V^a}\right) \quad \text{with} \quad \Pi^b < 0 \quad \text{and} \quad \overline{V^a} \geq 0,$$

$$\frac{T_1^{n,b} + ... + T_n^{n,b}}{n} \to \frac{1}{\lambda^b}, \quad \frac{T_1^{n,a} + ... + T_n^{n,a}}{n} \to \frac{1}{\lambda^a},$$

$$n^2 \tilde{f}_n(n.,n.) \Rightarrow \tilde{F}.$$

Under assumption 4.10, one can prove the following functional central limit on the Skorokhod space $(\mathcal{D}, J_1)$:

$$\frac{q^n_{[nt]}}{n} \Rightarrow Q, \tag{4.53}$$

where Q is a piecewise-deterministic Markov process with infinitesimal generator

$$\mathcal{G}h(x,y) = \lambda^b \Pi^b \frac{\partial h}{\partial x} + \lambda^a \Pi^a \frac{\partial h}{\partial y}, \tag{4.54}$$

for $x > 0, y > 0$, where

$$\Pi^a = 0 \quad \text{if} \quad \beta > 0 \quad \text{and} \quad \Pi^a = \overline{V}^a \geq 0 \quad \text{when} \quad \beta = 0,$$

and whose domain is the set $dom(\mathcal{G})$ of functions $h \in C^1(]0,\infty[\times]0,\infty[,\mathbb{R}) \cap C^0(\mathbb{R}^2_+,\mathbb{R})$ verifying the boundary conditions

$$\int_{\mathbb{R}^2_+} (h(u,v) - h(0,y))\tilde{F}(du,dv) = 0$$

In a disequilibrium situation such as the one described by Assumption 4.10, one can no longer assume that the bid-ask spread is constant. The bid-ask spread may widen as the bid price falls under the influence of selling pressure, while the ask queue may or may not remove because of a temporary shortage in demand. Since sellers dominate buyers, the intensity of limit orders at the ask dominate the intensity of market orders and cancelations at the ask and the ask queue never reaches zero whereas the bid price decreases.

In between two price moves, Q is deterministic and moves horizontally to the left at speed $\lambda^b\Pi^b$ until it reaches the y axis, then jumps to a random position inside the quadrant with distribution $\tilde{F}$ and price decreases at each hitting time. If one assumes that the price decreases by one tick at every move, then we obtain the following description of the 'free fall' of the price:

Proposition 4.10. *Under Assumption 4.10, the fluid limit of the bid price process is a deterministic process with negative drift*

$$\left(\frac{S^b_{[nt]}}{n}, t \geq 0\right) \stackrel{n\to\infty}{\Rightarrow} (d_p t, t \geq 0), \quad \text{where} \quad d_p = \frac{\lambda^b \Pi^b}{\int_{\mathbb{R}^2} y\tilde{F}(dx,dy)} < 0. \tag{4.55}$$

Proof. The sequence $(\tau_1, ..., \tau_n, ...)$ of bid price durations is a sequence of IID random variables with common distribution given by:

$$\mathbb{P}[\tau_1 > t] = \int_0^\infty \int_{t\Pi^b\lambda^a}^\infty \tilde{F}(dy, dt).$$

Hence $\mathbb{E}[\tau] = \frac{1}{\lambda^a \overline{V^a}} \int_{\mathbb{R}^2} yF(dx, dy)$. On the other hand, the bid price S_t^b is the counting process associated to the sequence of durations $(\tau_1, ..., \tau_k)$. Thanks to (Whitt 2002, Chapter 13), the counting process $(S_t^b)_{t\geq 0}$ satisfies the following functional limit theorem:

$$\left(\frac{S^b_{[nt]}}{n}, t \geq 0\right) \Rightarrow \left(\frac{\lambda^b \Pi^b}{\int_{\mathbb{R}^2} y\tilde{F}(dx, dy)} t, t \geq 0\right).$$

□

Thus, when sell orders exceed buy orders by an order of magnitude, the price acquires a negative trend and drops linearly and this the deterministic trend of the price dominates price volatility.

This proposition actually gives an upper bound on the free fall of the price: in a selling panic, traders may also cancel limit orders which populate the lower levels of the order book, thus accelerating the descent of the price, since this can lead to gaps in the order book which lead to negative price changes which are larger than one tick. We do not model such effects here but it is clear than incorporating such phenomena can lead to a free fall in the price at a faster than linear rate, as has been observed in market crashes.

4.5 Proof of technical lemmas

4.5.1 Some useful lemmas

Lemma 4.1. *The sequence of IID random variables* $(T_i, i \geq 1)$*, defined equation* (4.3.1)*, satisfies:*

$$\mathbb{E}[T_1] = \frac{r_1(F)}{1-p^+} + \frac{r_1(\tilde{F})}{1-p^-}.$$

Proof.

$$\begin{aligned}
\mathbb{E}[T_2] &= \sum_{i=1}^\infty \sum_{j=1}^\infty \mathbb{E}[T_2 | \gamma_2^1 - \gamma_1^2 = i, \gamma_2^2 - \gamma_2^1 = j] \mathbb{P}[\gamma_2^1 - \gamma_1^2 = i, \gamma_2^2 - \gamma_2^1 = j] \\
&= \sum_{i=1}^\infty \sum_{j=1}^\infty \mathbb{E}[T_2 | \gamma_2^1 - \gamma_1^2 = i, \gamma_2^2 - \gamma_2^1 = j] (p^+)^i (1-p^+)(p^-)^j (1-p^-) \\
&= \sum_{i=1}^\infty \sum_{j=1}^\infty \left(i r_1(F) + j r(\tilde{F})\right) (p^+)^i (1-p^+)(p^-)^j (1-p^-) \\
&= \sum_{i=1}^\infty i r_1(F)(p^+)^i (1-p^+) + \sum_{j=1}^\infty j r_1(\tilde{F})(p^-)^j (1-p^-) = \frac{r_1(F)}{1-p^+} + \frac{r(\tilde{F})}{1-p^-}.
\end{aligned}$$

□

Theorem 4.6. *Let* W *be the process defined in equation* (4.23)*. If*

$$\mathbb{E}[Z_i^2] < \infty, \qquad \mathbb{E}[\tau_1^2] < \infty \qquad \text{and} \qquad \mathbb{E}[Y_i^2] < \infty, \qquad \text{then}$$

- *When* $\mathbb{E}[Z_1] = 0$,
$$(\frac{W_{nt}}{\sqrt{n}})_{t\geq 0} \Rightarrow (\sqrt{\frac{Var(Z_1)}{\mathbb{E}[T_1]}}B_t)_{t\geq 0} \quad on \quad (\mathcal{D}, J_1).$$

- *When* $\mathbb{E}[Z_1] \neq 0$,
$$(\frac{W_{[nt]}}{n})_{t\geq 0} \Rightarrow (\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}t)_{t\geq 0},$$
$$(\frac{W_{[nt]} - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}nt}{\sqrt{n}})_{t\geq 0} \Rightarrow (\sigma B_t)_{t\geq 0}$$
where
$$\sigma = \sqrt{\frac{Var(Z_1 - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}T_1)}{\mathbb{E}[T_1]}}.$$

Proof. When $\mathbb{E}[Z_1] = 0$, the sequence $(Z_1, Z_2, ...)$ is a stationary sequence of random variable with finite moment of order two. Therefore, one can apply the Donsker theorem and the following functional central limit theorem holds:

$$\left(\frac{Z_1 + ... + Z_{[nt]}}{\sqrt{n}}, t \geq 0\right) \Rightarrow \left(\sqrt{Var(Z_1)}B_t, t \geq 1\right) \quad on \quad (\mathcal{D}, J_1),$$

where $W = (B_t, t \geq 1)$ is a standard Brownian Motion. On the other hand

$$\frac{M_{[nt]}}{n} \Rightarrow \frac{1}{\mathbb{E}[T_1]}.$$

Therefore, by application of (Whitt 2002, Theorem 13.2.3),

$$\left(\frac{W_{[nt]}}{\sqrt{n}}, t \geq 0\right) \Rightarrow \left(\sqrt{\frac{Var(Z_1)}{\mathbb{E}[T_1]}}B_t, t \geq 0\right) \quad on \quad (\mathcal{D}, J_1).$$

Assume $\mathbb{E}[Z_1] \neq 0$. First, one can easily notice that

$$\left(\frac{W_{[nt]}}{n}, t \geq 0\right) \Rightarrow \left(\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}t, t \geq 0\right) \quad on \quad (\mathcal{D}, J_1).$$

Define, for $i \geq 1$, $\tilde{Z}_i := Z_i - \mathbb{E}[Z_i]$. The sequence $(\tilde{Z}_i, i \geq 1)$ is a sequence of IID random variables with finite moment of order two, therefore,

$$\left(\frac{\tilde{Z}_1 + ... + \tilde{Z_{[nt]}}}{\sqrt{n}}, \frac{M_{[nt]}}{n}, t \geq 0\right) \Rightarrow \left(Var(\tilde{Z}_1)B_t, \frac{t}{\mathbb{E}[T_1]}, t \geq 0\right) \quad on \quad (\mathcal{D}(\mathbb{R}^2), J_1).$$

Hence, by (Whitt 2002, Theorem 13.2.3),

$$(\frac{W_{[nt]} - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}nt}{\sqrt{n}})_{t\geq 0} \Rightarrow (\sqrt{\frac{Var(Z_1 - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}T_1)}{\mathbb{E}[T_1]}}B_t)_{t\geq 0},$$

which concludes the proof. □

4.5.2 Proof of Theorem 4.2: case when $p^+ = p^-$

First, when the mean reversion parameters are equal ($p^+ = p^- := p^{cont}$), thanks to theorem 4.6,

$$(\frac{W_{nt}}{\sqrt{n}})_{t\geq 0} \Rightarrow (\sqrt{\frac{Var(Z_1)}{\mathbb{E}[T_1]}}B_t)_{t\geq 0} \quad on \quad (\mathcal{D}, J_1).$$

Lemma 4.2. *When $p^+ = p^- := p_{cont}$, $Var(Z_1) = 2\frac{p_{cont}}{(1-p_{cont})^2}$.*

Proof. The variance of Z_1 is

$$\begin{aligned}\mathrm{Var}(Z_1) = \mathbb{E}[Z_1^2] - (\mathbb{E}[Z_1])^2 &= \sum_{k=0}^{\infty} k^2 \mathbb{P}[Z_1 = k] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(i-j)^2\mathbb{P}[Z_1 = i-j] \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(i-j)^2 p_{cont}^{i-1}(1-p_{cont})p_{cont}^{j-1}(1-p_{cont}) \\ &= 2\sum_{i=1}^{\infty} i^2 p_{cont}^{i-1}(1-p_{cont}) - 2\left(\sum_{i=1}^{\infty} i p_{cont}^{i-1}(1-p_{cont})\right)^2 = 2\frac{p_{cont}}{(1-p_{cont})^2}.\end{aligned}$$

□

To summarize,

$$\mathbb{E}[Z_1] = 0, \quad \mathbb{E}[T_1] = \frac{r_1(F) + r(\tilde{F})}{1 - p_{cont}}, \quad Var(Z_1) = 2\frac{p_{cont}}{(1-p_{cont})^2},$$

and

$$\frac{Var(Z_1)}{\mathbb{E}[T_1]} = \frac{2}{r_1(F) + r(\tilde{F})}\frac{p_{cont}}{1-p_{cont}}.$$

Denote, for $n \in \mathbb{N}$, by $(\zeta_t^n, t \geq 0)$ the stochastic process defined by

$$\forall t \geq 0, \quad \zeta_t^n = \sum_{i=Y_1+\ldots+Y_{M_{nt}+1}}^{N_{nt}} \frac{X_i}{\sqrt{n}}$$

$\zeta_{n,t}$ is a random variable with the property that

$$\mathbb{P}[\sup_{s\leq t}|\zeta_s^n| > u] \leq (p^+)^{\sqrt{n}u} + (p^-)^{\sqrt{n}u}.$$

Hence for $t \geq 0$, $\zeta_t^n \to 0$ a.s. Therefore

$$\left(\sum_{i=Y_1+\ldots+Y_{M_{nt}+1}}^{N_{nt}} \frac{X_i}{\sqrt{n}}\right)_{t\geq 0} \overset{n\to\infty}{\Rightarrow} 0$$

on (D, J_1). Moreover, for $n \geq 1$, $t \geq 0$

$$\frac{S_{nt}}{\sqrt{n}} = \frac{W_{nt}}{\sqrt{n}} + \left(\sum_{i=Y_1+\ldots+Y_{M_{nt}+1}}^{N_{nt}} \frac{X_i}{\sqrt{n}}\right)$$

which proves Theorem 4.2.

4.5.3 Proof of Theorems 4.3 and 4.4: case where $p^+ \neq p^-$

In this subsection, we prove theorems 4.3 and 4.4. First, thanks to theorem 4.6,

$$(\frac{W_{[nt]}}{n})_{t\geq 0} \Rightarrow (\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}t)_{t\geq 0}, \qquad on \qquad (\mathcal{D}, J_1) \qquad \text{and}$$

$$(\frac{W_{[nt]} - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}nt}{\sqrt{n}})_{t\geq 0} \Rightarrow (\sigma B_t)_{t\geq 0} \qquad on \qquad (\mathcal{D}, J_1)$$

where

$$\sigma = \sqrt{\frac{Var(Z_1 - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}T_1)}{\mathbb{E}[T_1]}}.$$

The parameters $\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}$ and $\sqrt{\frac{Var(Z_1 - \frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}T_1)}{\mathbb{E}[T_1]}}$ are computed in the next lemma. Moreover

$$\left(\sum_{i=Y_1+\ldots+Y_{M_{nt}+1}}^{N_{nt}} \frac{X_i}{\sqrt{n}}, t \geq 0\right)_{n\geq 1} \Rightarrow 0 \qquad on \qquad (\mathcal{D}, J_1),$$

which concludes the proof of theorems 4.4 and 4.3.

Lemma 4.3. *When $p^+ \neq p^-$,*

$$\mathbb{E}[Z_1] = \frac{1}{1-p^+} - \frac{1}{1-p^-}, \text{ and } \quad Var[Z_1 - \frac{Z_1}{T_1}T_1] = A + B + C,$$

where A, B and C are given in theorem 4.4.

Proof.

$$\begin{aligned}
\mathbb{E}[Z_1] &= \sum_{k=1}^{\infty} k\mathbb{P}[Z_1 = k] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(i-j)\mathbb{P}[Z_1 = i-j] \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(i-j)(p^+)^{i-1}(1-p^+)(p^+)^{j-1}(1-p^+) \\
&= \sum_{i=1}^{\infty} i(p^+)^{i-1}(1-p^+) - \sum_{j=1}^{\infty} j(p^-)^{j-1}(1-p^-) = \frac{1}{1-p^+} - \frac{1}{1-p^-}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}[Z_1^2] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(i-j)^2(p^+)^{i-1}(1-p^+)(p^-)^{j-1}(1-p^-) \\
&= \sum_{i=1}^{\infty} i^2(p^+)^{i-1}(1-p^+) + \sum_{j=1}^{\infty} j^2(p^+)^{j-1}(1-p^+) - 2\left(\sum_{i=1}^{\infty} i(p^+)^{i-1}(1-p^+)\right)\left(\sum_{j=1}^{\infty} j(p^+)^{j-1}(1-p^+)\right) \\
&= \frac{1+p^+}{(1-p^+)^2} + \frac{1+p^-}{(1-p^-)^2} - \frac{2}{(1-p^+)(1-p^-)}.
\end{aligned}$$

$$\begin{aligned}\mathbb{E}[T_2^2] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{E}[T_2^2|\gamma_2^1-\gamma_1^2=i,\gamma_2^2-\gamma_2^1=j]\\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left(i^2r_2(F)+j^2r_2(\tilde{F})+2ijr_1(F)r_1(\tilde{F})\right)(p^+)^{i-1}(1-p^+)(p^-)^{j-1}(1-p^-)\\ &= \frac{r_2(F)(1+p^+)}{(1-p^+)^2}+\frac{r_2(\tilde{F})(1+p^-)}{(1-p^-)^2}+\frac{2r_1(F)r_1(\tilde{F})}{(1-p^+)(1-p^-)}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[T_2Z_2] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{E}[T_2Z_2|\gamma_2^1-\gamma_1^2=i,\gamma_2^2-\gamma_2^1=j]\\ &= \sum_{i=1}^{\infty}i^2(p^+)^{i-1}(1-p^+)r_1(F)-\sum_{j=1}^{\infty}j^2(p^-)^{j-1}(1-p^-)r_1(\tilde{F})\\ &+\left(\sum_{i=1}^{\infty}i(p^+)^{i-1}(1-p^+)\right)\left(\sum_{j=1}^{\infty}j(p^+)^{j-1}(1-p^+)\right)\left(r_1(\tilde{F})-r_1(F)\right)\\ &= \frac{r_1(F)}{(1-p^+)^2}-\frac{r_1(\tilde{F})}{(1-p^-)^2}+\frac{r_1(F)-r_1(\tilde{F})}{(1-p^+)(1-p^-)}, \qquad \text{and}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[Z_1-\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}Z_1] &= \mathbb{E}[\left(Z_1-\frac{\mathbb{E}[Z_1]}{\mathbb{E}[T_1]}T_1\right)^2]\\ &= \mathbb{E}[Z_1^2]+\frac{\mathbb{E}[Z_1]^2\mathbb{E}[T_1^2]}{\mathbb{E}[T_1]^2}-2\frac{\mathbb{E}[Z_1]\mathbb{E}[Z_1T_1]}{\mathbb{E}[T_1]}=A+B+C,\end{aligned}$$

where the terms A,B and C are given in Theorem 4.4. □

Bibliography

Aldous, David, James Allen Fill. 2002. *Reversible Markov Chains and Random Walks on Graphs*. University of California, Berkeley.

Alfonsi, A., A. Schied, A. Schulz. 2010. Optimal execution strategies in limit order books with general shape function. *Quantitative Finance* **10** 143–157.

Almgren, Robert, Neil Chriss. 1999. Value under liquidation. *Risk* .

Almgren, Robert, Neil Chriss. 2001. Optimal execution of portfolio transactions. *J.Risk 3* .

Almgren, Robert, Neil Chriss. 2003. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied mathematical finance* .

Amihud, Yakov, Haim Mendelson. 1980. Dealership markets: Market-making with inventory. *Journal of Financial Economics 8* 31–53.

Andersen, Allan, Rama Cont, Ekaterina Vinkovskaya. 2010. A point process model for the high-frequency dynamics of a limit order book. Financial Engineering Report 2010-08, Columbia University.

Angel, James J., Thomas M.Joyce, Fred Tomczyk, Chester S. Spatt, Gary Katz, Brett F. Mock, Daniel Mathisson, Duncan L. Niederauer, Jennifer E. Bethel, Erik R. Sirri, John C. Giesea Jamil Nazarali, Arthur Levitt, Lawrence E. Harris. 2010. Current perspectives on modern equity markets. *Knight Capital Group, Inc* .

Araman, Victor, Peter Glynn. 2011. Fractional brownian motion with $H < 1/2$ as a limit of scheduled traffic. *Working Paper* **to appear**.

Avellaneda, Marco, Sasha Stoikov, Josh Reed. 2011. Forecasting prices from Level-I quotes in the presence of hidden liquidity. *Algorithmic Finance* **1** 35–43.

Baccelli, François, Guy Fayolle. 1987. Analysis of models reducible to a class of diffusion processes in the positive quarter plane. *SIAM Journal on Applied Mathematics* **47**(6) 1367–1385.

Bacry, Emmanuel, Sylvain Delattre, Marc Hoffman, Jean-Francois Muzy. 2010. Scaling limits for Hawkes processes and application to financial statistics. Working paper.

Bak, P., M. Paczuski, M. Shubik. 1997. Price variations in a stock market with many agents. *Physica A: Statistical and Theoretical Physics* **246**(3-4) 430–453.

Bayraktar, Erhan, Ulrich Horst, Ronnie Sircar. 2006. A limit theorem for financial markets with inert investors. *Mathematics of Operations Research* **2006** 33–54.

Bayraktar, Erhan, Michael Ludkovski. 2011. Liquidation in limit order books with controlled intensity. *Mathematical Finance* **to appear**.

Becker-Kern, Peter, Mark M. Meerschaert, Hans Peter Scheffler. 2004. Limit theorems for coupled continuous time random walks. *Annals of Probability* **32** 730–756.

Biais, Bruno, Pierre Hillion, Chester Spatt. 1995. An empirical analysis of the order flow and order book in the Paris Bourse. *Journal of Finance* **50**(5) 1655–1689.

Billingsley, P. 1968. *Convergence of Probability Measures*. Wiley.

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3) 307–327.

Bollerslev, Tim, Robert F Engle. 1993. Common persistence in conditional variances. *Econometrica* **61**(1) 167–86.

Bouchard, Bruno, Ngoc-Minh Dang, Charles-Albert Lehalle. 2011. Optimal control of trading algorithms: A general impulse control approach. *SIAM Journal on Financial Mathematics* **2** 404–438.

Bouchaud, J. P., M. Mezard, M. Potters. 2002. Statistical properties of stock order books: empirical results and models. *Quantitative Finance* **2** 251.

Bouchaud, Jean-Philippe, Doyne Farmer, Fabrizio Lillo. 2008. How markets slowly digest changes in supply and demand. T. Hens, K. Schenk-Hoppe, eds., *Handbook of Financial Markets: Dynamics and Evolution*. Elsevier: Academic Press, 57–160.

Bouchaud, Jean-Philippe, Yuval Gefen, Marc Potters, Matthieu Wyart. 2003. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance* **4**(2) 176–190.

Bougerol, Philippe, Nico Picard. 1992. Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **4** 1714–1730.

Bovier, Anton, Jiri Cerny, Ostap Hryniv. 2004. The opinion game: Stock price evolution from microscopic market modelling URL `http://econpapers.repec.org/RePEc:arx:papers:cond-mat/0401422`.

Bovier, Anton, Jiří Černý. 2007. Hydrodynamic limit for the $A + B \to \emptyset$ model. *Markov Processes and Related Fields* **13**(3) 543–564.

Bramson, Maury, J. G. Dai. 2001. Heavy traffic limits for some queueing networks. *Annals of Applied Probability* **11** 49–88.

Brogaard, Jonathan A. 2010. High frequency trading and its impact on market liquidity. *Working paper* .

Burghardt, Galen, Jerry Hanweck, Lauren Lei. 2006. Measuring market impact and liquidity. *AlternativeEdge Research Note, Newedge* .

Chakraborti, Anirban, Ioane Muni Toke, Marco Patriarca, Frederic Abergel. 2011. Econophysics: Empirical facts and agent-based models. *Quantitative Finance* 991–1012.

Challet, Damien, Robin Stinchcombe. 2001. Analyzing and modelling 1+1d markets. Tech. rep.

Chordia, Tarun, Richard Roll, Avanidhar Subrahmanyam. 2008. Liquidity and market efficiency. *Journal of Financial Economics* **87**(2) 249–268.

Clark, Peter K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**(1) 135–55.

Cohen, Jacob Willem, O. J. Boxma. 1983. *Boundary value problems in queueing system analysis*, *North-Holland Mathematics Studies*, vol. 79. North-Holland Publishing Co., Amsterdam.

Cohen, J.W. 1988. Boundary value problems in queueing theory. *Queueing systems* (3) 97–128.

Cont, R., A. de Larrard. 2012. Price dynamics in limit order markets: linking volatility with order flow. Working paper.

Cont, Rama. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**(2) 223–236.

Cont, Rama. 2011. Statistical modeling of high-frequency financial data. *IEEE Signal processing* **28**(5) 16–25. URL `http://ssrn.com/abstract=1748022`.

Cont, Rama, Jean-Philipe Bouchaud. 2000. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic Dynamics* **4**(02) 170–196.

Cont, Rama, Adrien de Larrard. 2010. Price dynamics in a markovian limit order market. Working paper. URL `http://ssrn.com/abstract=1735338`.

Cont, Rama, Adrien de Larrard. 2011. Order book dynamics in liquid markets: limit theorems and diffusion approximations. Working paper. URL `http://ssrn.com/abstract=1757861`.

Cont, Rama, Arseniy Kukanov, Sasha Stoikov. 2010a. The price impact of order book events. Working paper, SSRN. URL `http://ssrn.com/abstract=1712822`.

Cont, Rama, Sasha Stoikov, Rishi Talreja. 2010b. A stochastic model for order book dynamics. *Operations Research* **58** 549–563.

Dai, J. G., Vin Nguyen. 1994. On the convergence of multiclass queueing networks in heavy traffic. *Annals of Applied Probability* **4** 26–42.

Engle, R., J. Russell. 1998. Autoregressive conditional duration: a new model for irregularly-spaced transaction data. *Econometrica* **66** 1127–1162.

Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50**(4) 987–1007.

Engle, Robert F. 2000. The econometrics of ultra-high-frequency data. *Econometrica* **68**(1) 1–22.

Engle, Robert F., Asger Lunde. 2003. Trades and Quotes: A Bivariate Point Process. *Journal of Financial Econometrics* **1**(2) 159–188.

Farmer, J. Doyne, László Gillemot, Fabrizio Lillo, Szabolcs Mike, Anindya Sen. 2004. What really causes large price changes? *Quantitative Finance* **4** 383–397.

Fayolle, G., R. Iasnogorodski, V. Malyshev. 1999. *Random walks in the quarter-plane*. Springer, New York.

Feller, William. 1971. *An Introduction to Probability Theory and Its Applications, Vol. 2*. Wiley.

Follmer, Hans, Ulrich Horst, Alan Kirman. 2005. Equilibria in financial markets with heterogeneous agents: a probabilistic perspective. *Journal of Mathematical Economics* **41**(1-2) 123–155.

Föllmer, Hans, Martin Schweizer. 1993. A microeconomic approach to diffusion models for stock prices. *Mathematical Finance* **3**(1) 1–23. URL `http://ideas.repec.org/a/bla/mathfi/v3y1993i1p1-23.html`.

Foucault, T., O. Kadan, E. Kandel. 2005. Limit order book as a market for liquidity. *Review of Financial Studies* **18**(4) 1171–1217.

Galakhov, E.I., A.L. Skubachevskii. 2001. On feller semigroups generated by elliptic operators with integro-differential boundary conditions. *Journal of Differential Equations* **176**(2) 315–355.

Garman, Mark. 1976. Market microstructure. *Journal of Financial Economics 3* 344–366.

Gatheral, Jim. 2010. No-dynamic-arbitrage and market impact. *Quantitative finance* **10** 749–759.

Ghysels, Eric, Joanna Jasiak. 1998. GARCH for irregularly spaced financial data: the ACD-GARCH model. *Studies in Nonlinear Dynamics and Econometrics* **2**(4) 133–149.

Glosten, Lawrence R., Paul R. Milgrom. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* **14**(1) 71–100.

Gopikrishnan, Parameswaran, Vasiliki Plerou, Xavier Gabaix, H. Eugene Stanley. 2000. Statistical properties of share volume traded in financial markets. *Phys. Rev. E* **62** R4493–R4496.

Gourieroux, Christian, Joanna Jasiak, Gaelle Le Fol. 1999. Intra-day market activity. *Journal of Financial Markets* **2**(3) 193–226.

Guéant, O., J. Fernandez-Tapia, C.-A. Lehalle. 2011. Dealing with the inventory risk. *Technical report* .

Harris, Lawrence E., Venkatesh Panchapagesan. 2005. The information content of the limit order book: evidence from NYSE specialist trading decisions. *Journal of Financial Markets* **8**(1) 25–67.

Harrison, J. Michael. 1978. The diffusion approximation for tandem queues in heavy traffic. *Advances in Applied Probability* **10**(4) 886–905.

Harrison, J. Michael, Vien Nguyen. 1993. Brownian models of multiclass queueing networks: Current status and open problems. *Queueing Systems* **13** 5–40.

Harrisson, J. M. 1973. The heavy traffic approximation for single server queues in series. *Journal of Applied Probability* **10** 613–629.

Harrisson, J. M. 1978. The diffusion approximation for tandem queues in heavy traffic. *Advances in Applied Probability* **10**(4) 886–905.

Hasbrouck, Joel. 1985. One security, many markets: Determining the contributions to price discovery. *Journal of Finance* **50** 1175–1199.

Hasbrouck, Joel. 2007. *Empirical Market Microstructure*. Oxford University Press.

Hautsch, Nikolaus. 2004. *Modelling Irregularly Spaced Financial Data*. Springer.

Hendershott, Terrence, Charles Jones, Albert Menkveld. 2012. Does algorithmic trading improve liquidity? *Journal of Finance* .

Hollifield, B., R. A. Miller, P. Sandas. 2004. Empirical analysis of limit order markets. *Review of Economic Studies* **71**(4) 1027–1063.

Horst, U. 2005. Financial price fluctuations in a stock market model with many interacting agents. Sonderforschungsbereich 373 2001-36, Humboldt Universitaet Berlin. URL `http://ideas.repec.org/p/wop/humbsf/2001-36.html`.

Iati, Rob. 2009. The real story of trading software espionage. *AdvancedTrading.com* .

Iglehart, Donald L., Ward Whitt. 1970. Multiple channel queues in heavy traffic. I. *Advances in Appl. Probability* **2** 150–177.

Iglehart, Donald L., Ward Whitt. 1971. The equivalence of functional central limit theorems for counting processes and associated partial sums. *The Annals of Mathematical Statistics* **42**(4) pp. 1372–1378.

Iyengar, Satish. 1985. Hitting lines with two-dimensional brownian motion. *SIAM Journal on Applied Mathematics* **45** 983–989.

Jacod, Jean, Albert N. Shiryaev. 2003. *Limit theorems for stochastic processes*. Springer, Berlin.

Klein, G. 1952. Mean first-passage times of Brownian motion and related problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **211**(1106) 431–443.

Kruk, Łukasz. 2003. Functional limit theorems for a simple auction. *Math. Oper. Res.* **28**(4) 716–751.

Kurkova, Irina, Kilian Raschel. 2011. Random walk in $\mathbb{Z}_+^2$ with non-zero drift absorbed at the axes. *Bulletin de la Société Mathématique de France* **139** 341–387.

Kyle, Albert S. 1985. Continuous auctions and insider trading. *Econometrica* **53** 1315–1336.

Kyle, Albert S., Robert H. Smith, Anna A. Obizhaeva. 2010. Market microstructure invariants.

Lawler, Gregory F., Vlada Limic. 2010. *Random Walk: A Modern Introduction*, vol. 123. Cambridge Studies in Advanced Mathematics.

Lebaron, Blake. 2006. Agent-based computational finance. *in Handbook of Computational Economics, Agent-based Computational Economics*. II North-Holland, 166–209.

Lillo, Fabrizio, J. Doyne Farmer. 2004. The long memory of the efficient market. Quantitative Finance Papers cond-mat/0311053, arXiv.org.

Lindvall, Torgny. 1973. Weak convergence of probability measures and random functions in the function space $D(0, \infty)$. *Journal of Applied Probability* **10**(1) 109–121.

Lux, T., M. Marchesi. 2000. Volatility clustering in financial markets: A micro-simulation of interacting agents. Discussion Paper Serie B 437, University of Bonn, Germany. URL `http://ideas.repec.org/p/bon/bonsfb/437.html`.

Lux, Thomas. 1985. Herd behavior, bubbles and crashes. *The Economic Journal* **105**(431) 881–896.

Madhavan, Ananth, Matthew Richardson, Mark Roomans. 1997. Why do security prices change? a transaction-level analysis of nyse stocks. *Review of Financial Studies* **10**(4) 1035–64.

Malo, Pekka, Teemu Pennanen. 2010. Reduced form modeling of limit order markets. *Quantitative Finance* **0**(0) 1–12. doi:10.1080/14697688.2011.589402.

Maslov, S., M. Mills. 2001. Price fluctuations from the order book perspective - empirical facts and a simple model. *PHYSICA A* **299** 234.

Maslov, Sergei. 2000. Simple model of a limit order-driven market. *Physica A* **78** 571–578.

Mendelson, Haim. 1982. Market behavior in a clearing house. *Econometrica* 1505–1524.

Menkveld, Albert J. 2011a. Electronic trading and market structure. *UK Government Foresight Driver Review* .

Menkveld, Albert J. 2011b. High frequency trading and the new-market makers. *EFA 2011 paper* .

Metzler, Adam. 2010. On the first passage problem for correlated brownian motion. *Statistics & Probability Letters* **80**(5-6) 277–284.

Mike, Szabolcs, J. Doyne Farmer. 2008. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control* **32**(1) 200–234.

O'Hara, Maureen. 1997. *Market Microstructure Theory*. Blackwell Business, Cambridge, MA.

Pagan, Adrian. 1996. Simple model of a limit order-driven marke. *Journal of empirical finance* **3** 15–102.

Parlour, Christine A. 1998. Price dynamics in limit order markets. *Review of Financial Studies* **11**(4) 789–816.

Predoiu, S., G. Shaikhet, S. Shreve. 2011. Optimal execution of a general one-sided limit-order book. *SIAM Journal on Financial Mathematics* **2** 183–212.

Ramanan, Kavita, Martin I. Reiman. 2003. Fluid and heavy traffic diffusion limits for a generalized processor sharing model. *Ann. Appl. Probab.* **13**(1) 100–139.

Raschel, Kilian, Johan S.H. van Leeuwaarden. 2011. Random walks reaching against all odds the other side of the quarter plane. *arXiv:1104.3034v1 [math.PR]* .

Reiman, M. 1977. Queueing networks in heavy traffic. *Ph.D. Dissertation, Department of Operation Research, Standford University* .

Resnick, Sidney. 2006. *Heavy-tail phenomena: Probabilistic and Statistical Modeling*. Springer.

Robert, Christian Y., Mathieu Rosenbaum. 2011. A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics* **9**(2) 344–366.

Roginsky, Allen L. 1994. A central limit theorem for cumulative processes. *Advances in Applied Probability* **26**(1) 104–121.

Roll, Richard. 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance* **39**(4) 1127–39.

Rosu, Ioanid. 2009. A dynamic model of the limit order book. *Review of Financial Studies* **22** 4601–4641.

Samorodnitsky, Gennady, Murad S. Taqqu. 1994. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York.

Schied, Alexander, Torsten Schoeneborn. 2008. Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. MPRA Paper 7105, University Library of Munich, Germany.

Silva, A. Christian, Victor M. Yakovenko. 2006. Stochastic volatility of financial markets as the fluctuating rate of trading: an empirical study. Quantitative finance papers, arXiv.org. URL `http://econpapers.repec.org/RePEc:arx:papers:physics/0608299`.

Smith, E., J. D. Farmer, L. Gillemot, S. Krishnamurthy. 2003. Statistical theory of the continuous double auction. *Quantitative Finance* **3**(6) 481–514.

Spence, Euan Alastair. 2010. Boundary value problems for linear elliptic pdes. Ph.D. thesis, Trinity Hall, Cambridge.

Spitzer, Frank. 1958. Some theorems concerning 2-dimensional brownian motion. *Transactions of the American Mathematical Society* **87** 187–197.

Taira, Kazuaki. 1991. *On the existence of Feller semigroups with Boundary conditions*. American Mathematical Society.

Whitt, Ward. 1980. Some useful functions for functional limit theorems. *Math. Oper. Res.* **5**(1) 67–85.

Whitt, Ward. 2002. *Stochastic Process Limits*. Springer Verlag.

Yoshida, Hidenobu, Ikuko Miyamoto. 1999. Harmonic functions in a cone which vanish on the boundary. *Math. Nachr.* **202** 177–187.

Zhou, C. 2001. An analysis of default correlations and multiple defaults. *Review of Financial Studies* **14** 555–576.