CHAPTER 1

Introduction

> "One of the funny things about the stock market is that every time one person buys, another sells, and both think they are astute." – *William Feather, American publisher and author (1889–1981)*

What is a limit order book? It is a device that the vast majority of organized electronic markets (all equity, futures and other listed derivatives markets) use to store in their central computer the list of all the interests of market participants. It is essentially a file in a computer that contains *all* the *orders* sent to the market, with their characteristics such as the **sign** of the order (buy or sell), the **price**, the **quantity**, a **timestamp** giving the time the order was recorded by the market, and a host of various market-dependent information. In other words, the limit order book contains, at any given point in time, on a given market, the list of all the transactions that one could possibly perform on this market. Its evolution over time describes the way the market moves under the influence of its participants. In fact, the study of limit order books can provide deep insight into the understanding of the financial market, which is an excellent example of an evolving "complex system" where the different participants *collectively interact* to find the best price of an asset. Hence, this field attracts mathematicians, economists, statistical physicists, computer scientists, financial engineers, and many others, besides the practitioners.

A market in which buyers and sellers meet *via* a limit order book is called an **order-driven market**. In order-driven markets, buy and sell orders are matched as they arrive over time, subject to some priority rules. Priority is always based on price, and then, in most markets, on time, according to a *FIFO* (First In, First Out) rule. Such priority rules are enforced in the vast majority of financial markets, although there exist some notable exceptions or variants: For instance, the Chicago Mercantile Exchange (CME) uses for some order books a *prorata* rule in place of (or together with) time

priority. Several different market mechanisms have been studied in the microstructure literature, see for example, Garman (1976); Kyle (1985); Glosten (1994); O'Hara (1997); Biais et al. (1997); Hasbrouck (2007). We will not review these mechanisms in this book [except Garman (1976) in Chapter 5], and rather keep our focus on the almost universal standard of **price/time** priority.

Essentially, three types of orders can be submitted:

- *Limit order* An order to specify a price at which one is willing to buy or sell a certain number of shares, with their corresponding price and quantity, at any point in time;
- *Market order* An order to *immediately* buy or sell a certain quantity, at the best available opposite quote;
- *Cancellation order* An order to cancel an existing limit order.

In the literature dealing with limit order books or market microstructure, agents who submit limit orders are referred to as *liquidity providers*, while those who submit market orders are referred to as *liquidity takers*. In real markets, ever since the various deregulation waves hit the US markets in 2005 and the European markets in 2007 [see for instance, Abergel et al. (2014)], there is no such thing as a *pure* liquidity provider or taker, and this classification should be understood as a convenient shorthand rather than a realistic description of the behaviour of market participants.

It is to be noted that depending on the market under consideration, there exist many variations of the three basic types of orders described above. A catalogue of real order types one can encounter on financial markets is given in Appendix A. Needless to say, the list provided is not exhaustive, and will be expanding over time. In this book, for practical reasons related to the structure of the available datasets, and because we are mainly interested in understanding and modelling universal features of limit order books, the focus will be on a somewhat stylized view of the market where orders can be simply of the "market", "limit" or "cancellation" type.

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

$$S := P^A - P^B, \tag{1.1}$$

is always positive, and is called the *spread*. We define the *mid-price* as the average between the bid and the ask

$$P := \frac{P^A + P^B}{2}. \tag{1.2}$$

Prices are not continuous, but rather have a discrete resolution ΔP, the *tick size*, which represents the smallest quantity by which they can change.

Why study limit order books? It is clear that the study of the empirical properties, as well as the mathematical modelling and numerical simulation, of limit order books, is of paramount importance for the researcher keen on gaining a deep understanding of financial markets.

Traditionally in financial econometrics, the data consist in time series of prices of one or several assets, and models are based on the statistical properties of the various quantities one can build from these time series: Returns, volatility, correlation... However, in order-driven markets, the price dynamics is controlled by the interplay between the incoming order flow and the order book (Bouchaud et al. 2002). Figure 1.1 is a schematic illustration of this process, with the conventional representation of quantities on the bid side by non-positive numbers.

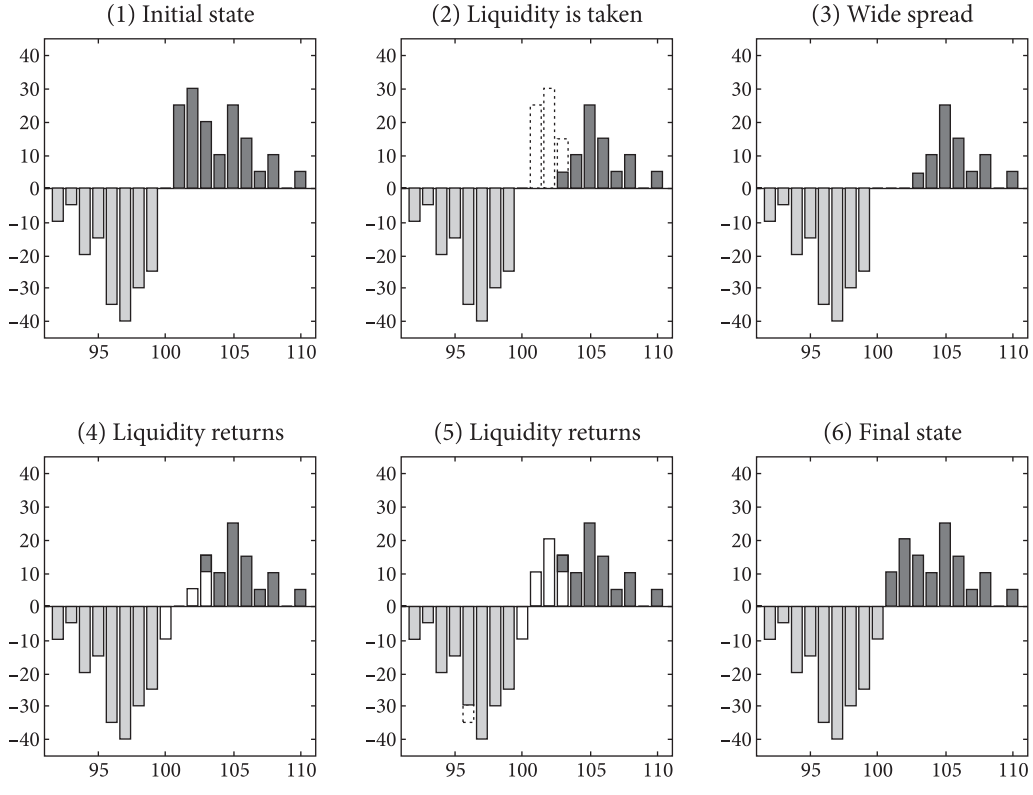

Fig. 1.1 A schematic illustration of the order book. A buy market order arrives and removes liquidity from the ask side, then, sell limit orders are submitted and liquidity is restored

The study of the limit order book therefore reveals, as a by-product, the price dynamics. One of our main motivations has been to understand the extent to which the mechanisms of the order book have an impact on the price dynamics at the *microstructure* level, and whether this impact remains visible at lower frequencies, i.e. when observing hourly or daily prices. Furthermore, the genuine scientific curiosity for this area of research has recently been very definitely enhanced by the rapid growth of *algorithmic trading* and *high frequency trading*. Market making strategies, optimal execution strategies, statistical arbitrage strategies, being executed at the individual order level, all require a perfect understanding of the limit order book. Some of the statistical properties presented in this book, in particular those pertaining to market imperfections, may be seen as building blocks of such an understanding.

How to model limit order books? There are several steps to take when modelling limit order books. Probably, the first one is to select a mechanistic description of the way incoming orders are stored and market orders are executed. This prerequisite is achieved, at least in a stylized form, in all the mathematical models of limit order books, and plays an important role in the simulation of limit order books, for which realistic *matching engines* must be developed in order to study trading strategies. The second step, at a more conceptual and scientifically more fundamental stage, involves choosing a mechanism for the arrival of orders, that is, for the submission of an order of a particular type at a specific date and time. Regarding this aspect, two main approaches have been successful in capturing key properties of the order book—at least to some extent. The first one, led by economists, models the interactions between *rational* agents who act strategically: The agents choose their trading decisions as solutions to individual utility maximization problems [see e.g., Parlour and Seppi (2008), and references therein]. In the second approach, proposed by econophysicists[1], agents are described statistically. In the simplest form along this line of research, the agents are supposed to act *randomly*. This approach is sometimes referred to as *zero-intelligence* order book modeling, in the sense that the arrival times and placements of orders of various types are *random and independent*, the focus being primarily on the "mechanistic" aspects of the continuous double auction rather than the strategic interactions between agents. Despite this apparently unrealistic simplification, statistical models of the order book do capture many salient features of real markets, and exhibit interesting, non-trivial mathematical properties that form the basis of a thorough understanding of limit order books. It is however necessary to depart from this overly simplified approach and study models were agents do interact, at least in a statistical way. Although, there exists a rather vast, fascinating literature on models of financial markets with interacting or competing agents (see e.g., Brock and Hommes

[1] Scholars who work in the interdisciplinary field of "Econophysics", comprised of the two fields economics and physics, using ideas and tools from both areas to study complex socio-economic systems.

(1998) Lux and Marchesi (2000)), very little is concerned with order-driven markets. Some recent results in this direction, based on a statistical approach using mutually exciting arrival processes, are presented in this book.

What is in this book? Our approach has been to start with the limit order book data, trying to assess their statistical properties. Hence, Chapter 2 is a survey of *stylized facts* on limit order books, Chapter 3 focuses on the shape of the order book and its relation to the size of incoming orders, whereas Chapter 4 is concerned with experimental evidence of the interaction between liquidity providers and takers on order-driven markets. We then moved on to the mathematical models: Chapter 5 is a survey of early works on limit order book modelling; Chapters 6 and 7 present an in-depth, rigourous mathematical theory of zero-intelligence models. In Chapter 8, we review some more advanced agent-based models, and present recent results on limit order books driven by interacting and competing statistical agents. We then provide in Chapter 9 a framework for simulations, and analyze and discuss some numerical results. Finally, in Chapter 10, we return to empirical studies, but with a different, more practical motivation, that of the profitability of trading strategies in order-driven markets.

What is not in this book? Obviously, so many things.

For the sake of consistency, we have deliberately left out several alternative approaches to order-driven markets modelling. Whether one actually requires to understand the motivation of the agents in order to obtain a faithful description of their behaviour is an open debate, and we are happy to participate in it with our systematic statistical approach.

Also of great importance is the study of *market impact*. This subject is definitely an important topic, with great practical implications, and although limit order book models obviously offer various possible mechanisms for market impact, we do not address this specific question.

Also connected to market impact, the systematic study of *trading strategies* in order-driven markets in only touched upon in the fourth part of this book, and should be studied at greater lengths.

We could keep on extending this list of regrets. It is clear that progresses must be made in the study of limit order books. Some are already in the making, and we certainly hope that this book will lend an impetus to many others.

PART ONE
EMPIRICAL PROPERTIES OF ORDER-DRIVEN MARKETS

CHAPTER 2

Statistical Properties of Limit Order Books: A Survey

2.1 Introduction

The computerization of financial markets in the second half of the 1980s provided empirical scientists with easier access to extensive data on order books. Biais et al. (1995) is an early study of the data flows on the newly (at that time) computerized Paris Bourse. Many subsequent papers offer complementary empirical findings and modelling perspectives, e.g., Gopikrishnan et al. (2000), Challet and Stinchcombe (2001), Maslov and Mills (2001), Bouchaud et al. (2002), Potters and Bouchaud (2003). In this chapter, we present a summary of some fundamental empirical facts. Basic statistical properties of limit order books, which can be observed from real data, are described and studied. Many variables crucial to a fine modelling of order flows and dynamics of order books are studied: Time of arrival of orders, placement of orders, size of orders, shape of order books, etc.

The markets we are dealing with are order-driven markets with no official market maker, in which orders are submitted in a double auction and executions follow price/time priority. In order to make the results we present both self-contained and reproducible, the statistics have been computed directly using our own database. The set of data that we have used in this chapter is detailed in Appendix B, which contains the precise description of all the data sets used throughout this book.

2.2 Time of Arrivals of Orders

We compute the empirical distribution for interarrival times – or durations – of market orders for the stock BNP Paribas using the data set described in Appendix B.2. The results are plotted in Figs 2.1 and 2.2, both in linear and log scale. It is clearly observed

that the exponential fit is not a good one. We check however that the Weibull distribution fit is potentially a very good one. Weibull distributions have been suggested for example in Ivanov et al. (2004). Politi and Scalas (2008) also obtain good fits with q-exponential distributions.

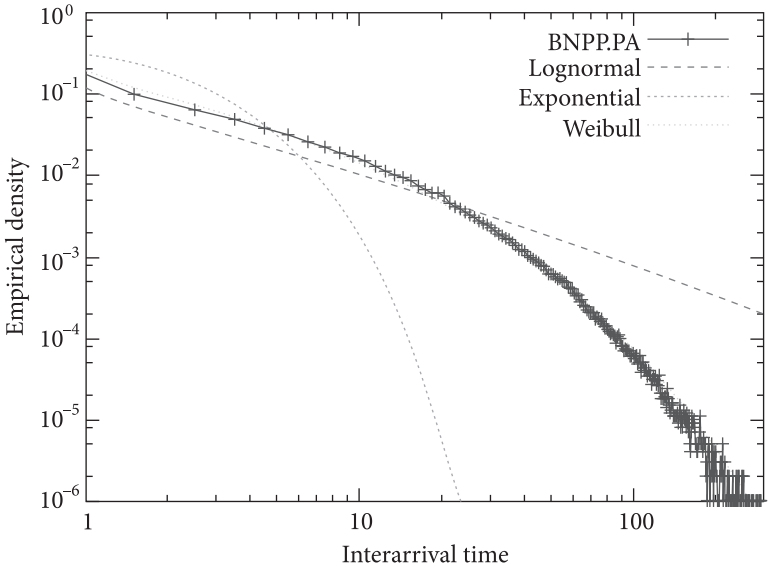

Fig. 2.1 Distribution of interarrival times for stock BNPP.PA in log-scale. Extracted from Chakraborti et al. (2011a)

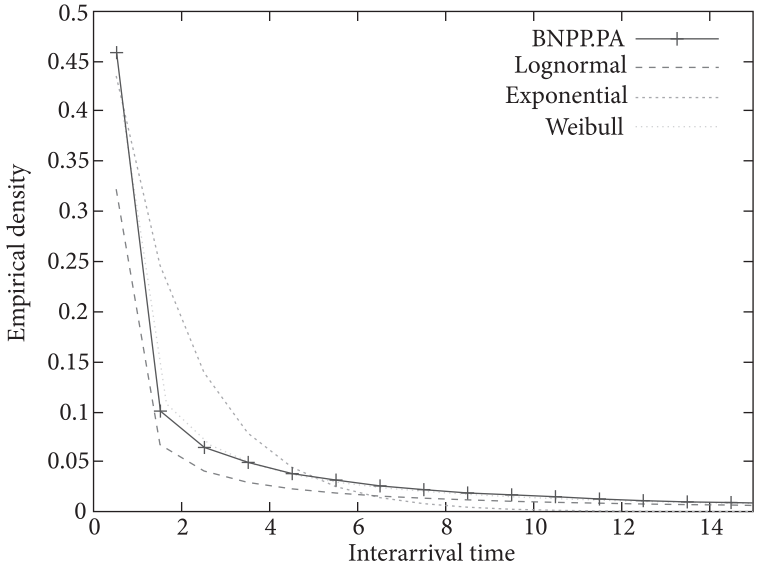

Fig. 2.2 Distribution of interarrival times for stock BNPP.PA (Main body, linear scale). Extracted from Chakraborti et al. (2011a)