

# PART IV

## Clustering and Correlations



## Introduction

So far, we have restricted our attention to models with simple statistical properties, such as Brownian motions and time-homogeneous Poisson processes. These models have assumed that all events are independent (once they are conditioned on the current state of the world), and have produced simple, stationary outputs.

This picture is far from the truth, for at least two reasons. First, real order flow exhibits complex feedback loops. As empirical measurements demonstrate, a large fraction of market activity is generated endogenously, so assuming that order flows are conditionally independent is clearly problematic. Whether stabilising or destabilising, feedback loops in financial markets can (and do) fundamentally change the resulting global behaviour. Second, order flows show long-range autocorrelations, to the extent that future order flows are often highly predictable, given observations of the past.

In this part, we introduce a collection of tools to model and understand the complex dependencies between market events. As with the previous parts, we start this discussion by using empirical evidence to address several important questions on this topic. To what extent do past events influence future events? What part of market activity is triggered by exogenous events, like the arrival of news? And what part is triggered endogenously by past activity?

The first part of our discussion will focus on the important role played by the time-clustering of events. We will introduce a flexible class of processes called Hawkes processes, which are the continuous-time analogue of a discrete ARCH process. As we will discuss, the Hawkes framework is a simple yet powerful tool for estimating and modelling events that cluster in time. By fitting Hawkes processes to LOB data, we will show that a large fraction of market activity appears to be endogenous. This will help us to illustrate how clustering might be responsible for the fat-tailed returns distributions that we discussed in Section 2.2.1.

The second part of our discussion will focus on an important puzzle that we will attempt to solve throughout the remaining parts of the book: the highly predictable character of order-flow direction. Specifically, we will show that the sign of market orders is strongly autocorrelated in time, such that observing a buy market order now significantly increases the probability of observing another buy market order far in the future.

This last point deserves detailed investigation. Where do such autocorrelations come from? Do traders exhibit herding behaviour, such that they all buy or sell aggressively at the same time? Or do these regularities emerge from the behaviour of individual agents? As usual, we will use empirical data to settle the argument. We will show that although herding behaviours exist, they are actually quite weak. Instead, what really causes order-flow autocorrelations is the *order-splitting*

behaviour of traders, who slice and dice so-called *metaorders* into many smaller *child orders*, which they execute incrementally over time periods of days or even weeks. Why do traders behave in this way? The answer lies in the ever-present shortage of liquidity in financial markets. This concept, which we will discuss extensively, will come to form one of the most important concepts of this book.