Modelling empirical features and liquidity resilience in the Limit Order Book

Efstathios Panayi

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computing

University College London

May 15, 2015

I, Efstathios Panayi, confirm that the work presented in this thesis is my own, except for the parts indicated in my statement of conjoint work. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The contribution of this body of work is in developing new methods for modelling interactions in modern financial markets and understanding the origins of pervasive features of trading data. The advent of electronic trading and the improvement in trading technology has brought about vast changes in individual trading behaviours, and thus in the overall dynamics of trading interactions. The increased sophistication of market venues has led to the diminishing of the role of specialists in making markets, a more direct interaction between trading parties and the emergence of the Limit Order Book (LOB) as the pre-eminent trading system. However, this has also been accompanied by an increased fluctuation in the liquidity available for immediate execution, as market makers try to balance the provision of liquidity against the probability of an adverse price move, with liquidity traders being increasingly aware of this and searching for the optimal placement strategy to reduce execution costs.

The varying intra-day liquidity levels in the LOB are one of the main issues examined here. The thesis proposes a new measure for the resilience of liquidity, based on the duration of intra-day liquidity droughts. The flexible survival regression framework employed can accommodate any liquidity measure and any threshold liquidity level of choice to model these durations, and relate them to covariates summarising the state of the LOB. Of these covariates, the frequency of the droughts and the value of the liquidity measure are found to have substantial power in explaining the variation in the new resilience metric. We have shown that the model also has substantial predictive power for the duration of these liquidity droughts, and could thus be of use in estimating the time between subsequent tranches of a large order in an optimal execution setting.

A number of recent studies have uncovered a commonality in liquidity that extends across markets and across countries. We outline the implications of using the PCA regression approaches that have been employed in recent studies through synthetic examples, and demonstrate that using such an approach for the study of Euro-

pean stocks can mislead regarding the level of liquidity commonality. We also propose a method via which to measure commonality in liquidity resilience, using an extension of the resilience metric identified earlier. This involves the first use of functional data analysis in this setting, as a way of summarising resilience data, as well as measuring commonality via functional principal components analysis regression.

Trading interactions are considered using a form of agent-based modelling in the LOB, where the activity is assumed to arise from the interaction of liquidity providers, liquidity demanders and noise traders. The highly detailed nature of the model entails that one can quantify the dependence between order arrival rates at different prices, as well as market orders and cancellations. In this context, we demonstrate the value of indirect inference and simulation-based estimation methods (multi-objective optimisation in particular) for models for which direct estimation through maximum likelihood is difficult (for example, when the likelihood cannot be obtained in closed form). Besides being a novel contribution to the area of agent-based modelling, we demonstrate how the model can be used in a regulation setting, to quantify the effect of the introduction of new financial regulation.

Acknowledgements

I would like to start by offering my sincerest gratitude to my supervisor, Gareth Peters, who has consistently provided support in my efforts to familiarise myself with new concepts. If I am proud of the work presented in this thesis, it is in no small part due to the patience he has exhibited during our collaboration and his unique knowledge, as well as his constant belief in my abilities throughout.

I would also like to thank my second supervisor, Mark Harman, who is responsible for my first steps in academic research, and indeed, my first publication.

I would like to thank my family for their backing since the very first day of my undergraduate degree, and without them none of this would have been possible. I am truly grateful for their encouragement to pursue scientific endeavour.

Finally, I would like to dedicate this work to my fiancée, whom I met just prior to embarking on this journey. Chryso has always been confident in my ability to see this thesis through, and her support has been invaluable in difficult times. For this, and everything else that you do, I am indebted to you.

Publications and presentations

Accepted journal papers

1. Efstathios Panayi and Gareth W Peters. Liquidity commonality does not imply liquidity resilience commonality: A functional characterisation for ultra-high frequency cross-sectional lob data. *to appear, Quantitative Finance Special Issue on Big Data Analytics*, 2015b

Published conference and workshop papers

1. Efstathios Panayi and Gareth Peters. Survival models for the duration of bid-ask spread deviations. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pages 9–16. IEEE, 2014

2. Efstathios Panayi, Mark Harman, and Anne Wetherilt. Agent-based modelling of stock markets using existing order book data. In *Multi-Agent-Based Simulation XIII*, pages 101–114. Springer, 2013

Submitted journal papers and working papers

1. Efstathios Panayi and Gareth Peters. Stochastic simulation framework for the limit order book using liquidity motivated agents. *in review, Journal of Financial Engineering*, 2015a

2. Efstathios Panayi, Gareth W Peters, Jon Danielsson, and Jean-Pierre Zigrand. Market liquidity resilience. *London School of Economics Working Paper Series*, 2014

3. Gareth William Peters, Ariane Chapelle, and Efstathios Panayi. Opening discussion on banking sector risk exposures and vulnerabilities from virtual currencies: An operational risk perspective. *in review, Journal of Operational Risk*, 2014

Presentations

Systemic Risk Centre meeting, Coopers Hall, London . . . . . Oct 2014
Forecasting Financial Markets Conference, Marseille, France . . . . . May 2014
Computational Intelligence for Financial Engineering & Economics, London Feb 2014
Computational and Financial Econometrics, Senate House, London . . . . . Dec 2013
Recent advances in Algorithmic and High Frequency Trading, UCL, London May 2013
Multi-Agent-Based Simulation XIII International Workshop, Valencia, Spain July 2012

Posters

Theory of Big Data, UCL, London . . . . . Jan 2015

International visits

Institute of Statistical Mathematics, Tokyo, Japan . . . . . Jul-Aug 2013

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *"Why are you always in such a hurry, Mr. Lipwig?"*
> *"Because people don't like change. But make the change happen fast enough and you go from one type of normal to another."*
>
> — Terry Pratchett, *Making money*

The modern economics literature has established a strong, positive relationship between the development of financial systems and economic growth[1]. Financial markets arise in order to encourage the dissemination of information, and facilitate transactions by reducing transaction costs. In particular, the development of financial intermediaries who are better at ameliorating these frictions has been found to be causally related to the level of economic growth [Levine et al., 2000].

Markets enable specialisation, and aid in the allocation of resources to sectors of the economy where the social return is greatest [Greenwood and Smith, 1997]. Financial assets traded in markets become less risky, as markets allow investors to enter and exit positions swiftly and inexpensively, when they wish to alter their portfolios [Arestis et al., 2001]. Reducing the cost of mobilising investments is thus one of the primary benefits of a well-functioning financial system.

More recent work suggests that the finance-growth relationship that held until the late 80s is indeed waning [Rousseau and Wachtel, 2011]. The relationship appears also to be non-linear: while the first steps in the development of a financial system go hand in hand with growth, further financial development becomes detrimental to the economy [Cecchetti et al., 2012]. There have also traditionally been doubts about the causality of the relationship between financial

[1]There is a substantial literature that studies the relationship between financial market development and economic growth, starting more than 40 years ago with the work of Goldsmith [1969], McKinnon [1973] and more recently, with the analytical models of Levine [1997], Greenwood and Smith [1997].

sector development and the overall growth in the economy [Robinson, 1953], and results in developing economies support the view that finance follows enterprise [Ang and McKibbin, 2007].

Financial markets have experienced a number of crises in recent years, some of which can be attributed to the excessive deepening of the financial sector [Rousseau and Wachtel, 2011]. For the duration of crises, such as the 1997 East Asian crisis, or the more recent financial crisis of 2007-2009, the ability of investors to build or liquidate their positions without much effect on the market was diminished [Laeven and Valencia, 2010]. That liquidity co-moves with the market was already known, but recent work has also uncovered the bi-directionality of the relationship [Brunnermeier and Pedersen, 2009]. In this context, 'dry-ups' in liquidity have also been found to be a contributing factor in the amplification of small shocks into full-blown financial crises [Brunnermeier, 2008].

1.1 Electronic trading and the Limit Order Book

Financial markets were traditionally physical meetingplaces for buyers and sellers to interact and exchange financial assets. Stock market trading has been ongoing for more than 400 years, with the Dutch East India Company stock trading on the Amsterdam Capital Market since the early 17th century [Gelderblom and Jonker, 2004]. Even prior to this, there was known to be trading in governmental debt products in Venice, arising from an active money market, and becoming key to the development of the city as a financial centre [Mueller and Lane, 1997].

Perhaps the most iconic figure of physical marketplaces is the floor or 'pit' trading at the Chicago Board of Trade, where brokers wearing colourful suits signalled their trading intentions using a series of hand signals. While CME Group still maintains the pit (at a significant cost to them[2]) this is largely for historical reasons, as the vast majority of trading has moved to electronic exchanges. Of the exchanges that were predominantly floor-based in the past, the London Stock Exchange transitioned first, moving their trading from the floor to being performed by computer and telephone in dealing rooms in 1986[3], and other exchanges soon followed.

Electronic trading has now been introduced in the leading financial exchanges of more than 100 countries, and Jain [2005] documents this shift over the last 40 years. The relevant trading technology was first developed by the U.S. brokerage firm Instinet as early as 1969.

[2] http://www.bloomberg.com/news/2013-11-06/cme-spends-30-million-a-year-on-open-outcry-trading.html

[3] http://www.londonstockexchange.com/about-the-exchange/company-overview/our-history/our-history.htm

The Toronto Stock Exchange was the first leading exchange to introduce electronic trading in 1977, and the vast majority of venues are now either fully electronic, or combine electronic and floor trading. The automation of trading has been shown to be associated with several positive changes in financial markets and the economy, such as a reduction in the cost of equity, (a phenomenon that is more prevalent in developing markets), an increase the availability of liquidity and a positive price reaction following the switch.

The ever-increasing share of electronic trading has brought about a vast increase in order flow, while financial regulations (e.g. the 'Regulation National Market System' (Reg NMS) in the U.S., and the 'Markets in Financial Instruments Directive' (MiFID) in Europe), designed to foster competition amongst trading exchanges, have resulted in a proliferation of trading venues for equity markets. While both developments have undoubtedly had a positive effect on some aspects of equity trading, they have also made it more difficult for regulators to assess certain characteristics of markets, because of the speed and volume of transactions [Cont, 2011].

A development that has arguably contributed to the complexity of stock price and volume dynamics is the emergence of the Limit Order Book (LOB) as the pre-eminent trading mechanism. The LOB is the central matching mechanism used in the majority of equity trading venues today [Roşu, 2009]. It collects all the buying and selling interest in a particular stock and presents an aggregation of these orders to every market participant. Every order to buy (called a bid) or to sell (called an ask) has a specified size and a maximum or minimum execution price, respectively, and the LOB orders these by price first, then by time of submission. Limit orders that are not executed immediately enter the LOB and aggressive orders (called market orders) execute against these resting orders (orders in the queue).

Glosten and Milgrom [1985] suggests that limit orders are submitted at different levels, in order for market participants to protect themselves from adverse selection, that is, execution of their order against a trader with superior information. Modelling the short-term dynamics of the LOB is important for a number of reasons, including optimal execution of orders [Bertsimas and Lo, 1998, Almgren and Chriss, 2001, Avellaneda and Stoikov, 2008, Obizhaeva and Wang, 2013, Alfonsi et al., 2010], and quantifying market impact [Hautsch and Huang, 2012]. As such, there have been a number of empirical studies investigating determinants of certain aspects of the LOB, like the bid-ask spread [Chung et al., 1999], and market depth [Ahn et al., 2001].

1.2 Market liquidity

Features such as the bid-ask spread, market depth, volume, and others are typically studied under the umbrella term of 'market liquidity', a concept which measures the relative ease with

which one can take up or liquidate a position in the market, without much effect on market dynamics. In general, one would favour markets where one is able to enter or exit positions swiftly, and without incurring large costs. Liquidity is considered a desirable characteristic of financial markets, as in liquid markets we generally expect to observe fewer abrupt changes, or 'jumps', in terms of either the security price or the volume available for that security. While there is no universally agreed measure to uniquely capture the concept of what is market liquidity, proposed measures should reflect aspects like immediacy of execution (that is, there is trading interest throughout the trading day), tightness (or the ability to buy or sell at roughly the same price) and depth (the ability to enter the position at some size) [Von Wyss, 2004].

It is important to make the distinction between market liquidity and funding liquidity, with the latter being related to the ease of obtaining funding for trading. Although the two are interrelated [Brunnermeier and Pedersen, 2009], this thesis will focus on local LOB dynamics, for which market liquidity is a central concept.

Market liquidity is important for different stakeholders, including the firms whose assets are being traded, liquidity providers, asset managers, exchanges and regulators. For example, for firms looking to raise capital through an IPO, projected market liquidity is a factor in both the associated costs, as well as the pricing of the stock itself [Butler et al., 2005, Ellul and Pagano, 2006]. This is because many investors will be more reluctant to hold a position that will be difficult or expensive to liquidate swiftly.

Regulators also have an interest in liquidity fluctuations and reducing trading frictions, as part of their obligation to ensure orderly markets. Several trading rules have been introduced by the SEC in an attempt to improve liquidity levels, for example the Order Handling Rule, the Limit Order Display Rule[4] and others. These were largely beneficial, reducing bid-ask spreads substantially, but had the unintended consequence of reducing depth in lower volume stocks [Goldstein and A Kavajecz, 2000].

There has long been a distinction between market participants who supply and consume liquidity. Market makers supply 'immediacy', that is, they are willing to bear the risk of keeping a position until the final buyer or seller arrives to take that position [Grossman and Miller, 1988]. A large portion of their liquidity provision activity is against asset managers, who want to execute (buy or sell) a large trading volume for one or more assets, and who do not have a choice regarding the assets they invest in and little choice on the timing of when they can invest. Such investors have a strong vested interest to seek out markets that have a high level

[4]The text for both these rules can be found at `https://www.sec.gov/rules/other/34-38156.txt`.

of liquidity. That is, they care about liquidity, insofar as it affects their return on investment, as illiquid securities cost more to buy and can be sold for less [Foucault et al., 2013].

Liquidity is therefore important for risk and execution purposes, and this is true particularly in modern LOBs, where market participants have adapted their trading behaviour, in order to improve their execution. On the New York Stock Exchange (NYSE), for example, the average order size is one-eighth of that of fifteen years ago, in terms of number of shares, and one-third in dollar value [Chlistalla et al., 2011]. This indicates the partitioning of large orders into multiple smaller orders and traders taking advantage of the replenishing of liquidity. It also shows that a good understanding of the intra-day dynamics of liquidity is necessary for market participants. This 'hidden' liquidity constitutes one of the axes of this thesis, and the duration of time required for such liquidity to become visible in the LOB is studied in detail.

1.3 High-frequency trading effects and regulation

In the rapidly evolving environment described above, in which face-to-face interactions have been phased out in favour of electronic trading, there has been a vast increase in the speed of transactions. The majority of the activity occurs in the millisecond environment now, a few orders of magnitude faster than humans can process information. The fast response time to LOB activity is termed 'low latency' [Hasbrouck and Saar, 2013], while the volume of activity performed on the such timescales is usually referred to as 'high-frequency trading'.

In this setting, many of the more traditional models and tools that express the probability of events based on some historical distribution have proved unable to explain certain phenomena, like the May 6, 2010 'Flash Crash'. On this day major indices and futures products in the US dropped by more than 6% between 2:41 and 2:45 p.m., having already dropped by 4% in the preceding 5 hours. Prices then recovered with the same speed and intensity, almost reaching the day's opening prices within 15 minutes [Kirilenko et al., 2014]. It is estimated that for a brief period during this day, US markets had lost more than $1 trillion of their value [Easley et al., 2011]. The SEC/CFTC joint advisory committee investigation that followed [Securities et al., 2010] suggested that there was no fundamental reason for a rapid drop of such magnitude, in the sense that it was not caused by specific economic news.

The events that transpired on that day highlighted a number of important points: Firstly, that even in the presence of all the relevant data, it is not straightforward to pinpoint the initial cause of such large scale, and sudden, price movements and liquidity droughts, as evident by the disagreements between the SEC/CFTC report and those of other industry participants[5].

[5] http://www.nanex.net/FlashCrashFinal/FlashCrashSummary.html

Secondly, while the 'tipping point' may not be known, endogenous dynamics (feedback loops) certainly contribute to such phenomena. Thirdly, models of short term dynamics based on historical distributions (used, for example in Value At Risk models) may be ill equipped to handle such sudden intra-day volatility[Bookstaber, 2012].

One of the factors identified in the SEC report as contributing to this incident was the activity of 'high frequency traders' (HFT), who typically enter and exit positions very frequently. Their aim is not to profit from the appreciation or the depreciation of an asset over some period, but rather to benefit from temporary price discrepancies between markets, or from providing liquidity in a single market. In their latter role, they are similar to market makers, but unlike them, they do not have quoting obligations.

Some of the regulation that has been introduced or proposed in recent years, like the financial transaction tax, or minimum resting times for limit orders, is intended to curtail the growth of HFT. Unfortunately, this new regulation has often brought about unintended consequences: Colliard and Hoffmann [2013] describes the introduction of the financial transaction tax in Italian equity markets. This was designed to curtail high frequency trading, but has instead led to a sharp drop in turnover for this asset class, making it more difficult (or costly) for investors to enter and exit positions. There is therefore a pressing need to understand the processes that lead to such a reduction in liquidity, which originate either from endogenous dynamics in the LOB or from the introduction of financial regulation, as in the recent Italian case.

The second axis of this thesis is in modelling trading interactions in the LOB, in the hope that such a model can eventually be used as a 'dry-run' to test the effects of particular forms of trading activity, or the effect of the introduction of new regulation. Our interest is specifically in the behaviour of both stock prices and the volumes available for trading in the LOB over short horizons.

1.4 Thesis contributions

1.4.1 Processing high-frequency trading activity and developing an efficient LOB implementation

This thesis revolves around the modelling of trading interactions and features of the LOB, and thus obtaining an appropriate dataset to analyse and compare results against was an important consideration. Fortunately, we were able to obtain a very detailed dataset from the Chi-X venue, which contained all the trades and trading interest in around 1300 European assets for a period of 4 months in 2012. This dataset is uniquely suited to the type of empirical analysis performed

in this thesis, as it enabled us to evaluate whether certain results hold across countries and across industries. In addition, while it is not a primary exchange, it commands more than a quarter of the trading activity in the 3 most active European markets (UK, Germany and France). Therefore, insofar as the trading activity on the primary exchanges is similar to that on large secondary exchanges, the results we present here should hold for the primary European equity exchanges also.

The dataset consisted of a total of 82 event logs, one for each trading day in the four month period. Each of these was between 2 and 3GB in size (for a total of over 200GB of data), and contained information about tens of millions of events pertaining to the 1300 assets traded on Chi-X. The first challenge was to divide these logs into individual files detailing the trading activity for every individual asset. The purpose of this was two-fold: Firstly, in order to make the dataset more manageable, and secondly, because the models we are developing in this thesis are predominantly focused on a single asset at a time.

Following that, we developed an efficient LOB implementation, as the software used at Chi-X to replay daily activity is not made available externally. This entailed creating a matching engine (a virtual order book), that could read in every limit order arrival, execution and cancellation, recreating the state of the order book at each event timestamp. In this way, one is able to view the LOB at a transaction-by-transaction level (event time) and can also obtain information that is not otherwise available using the raw data (number/volume of orders at each level of the order book, the age of orders in the order book, the value of certain liquidity measures which require detailed LOB data etc).

Even though packages for visualising and retrieving data from an LOB exist in statistical software (e.g. in the statistical programming language R, package *orderbook*[6] by Kane et al. [2011]), these are limited, in that they are not able to capture the required details of every individual order that are used in the models developed in this thesis. In addition,the choice of programming language to create the limit order book matching engine had to be sufficiently rich and flexible to allow for high-throughput processing of massive chunks of data. In general, standard languages such as R and Matlab were unsuitable for such large scale tasks. For example, an R implementation of a complex data structure is bound to be to slow and plagued by memory management and RAM problems with such massive files, and is thus not particularly suited to processing daily activity of hundreds of thousands of transactions per day and per asset efficiently. This meant that an object-oriented language was required, and we provide in

[6]The package is available at `http://cran.r-project.org/web/packages/orderbook/index.html`

Chapter 2 a more detailed description of the implementation.

1.4.2 Modelling the resilience of LOB liquidity

We mentioned liquidity as an important notion in financial markets, and indeed for some investors it is a primary concern in deciding the markets and assets they would like to operate in. The traditional, static notion of liquidity, which has been the focus of several recent studies, is based on instantaneous snapshots of market variables. As such, they are perhaps less useful in indicating the timeframe in which a market should recover from a liquidity shock, coming perhaps from either a single, large order or from a series of smaller orders that remove orders resting on the LOB. This aspect has been termed 'resilience' [Kyle, 1985], and there have been a number of attempts to pin down the concept.

Some definitions have related resilience to price evolution, in the sense of indicating a swift return to some former level of the price of the asset [Kyle, 1985, Obizhaeva and Wang, 2013], or the difficulty in affecting a permanent change in the price [Harris, 2002]. Another interpretation of resilience is related to order replenishment, which would indicate the return of volume to the LOB after a shock [Garbade and Garbade, 1982], or the tightening of the spread [Foucault et al., 2005]. Despite the fact that this aspect of liquidity was introduced more than 30 years ago, there have been relatively few studies that have quantified resilience with a metric, and even fewer to propose a quantitative model for it, with the exception of Foucault et al. [2005] and Large [2007].

In this thesis, we focus on the second interpretation of liquidity resilience, as an indicator of order replenishment, as this is important for a number of stakeholders, including brokerage houses, that aim to minimise trading costs, and regulators, who aim to ensure orderly markets through exchange rules and management. Our first contribution is to propose a new definition for liquidity resilience, which captures the time required for liquidity to return to the LOB after a shock. The metric measures the duration of intra-day liquidity dry-ups, or droughts, relative to any threshold level of liquidity and for a user-specified liquidity measure. This makes it more general than previous metrics, which were tied to a particular liquidity measure or threshold level (e.g. Foucault et al. [2005]). It should therefore prove to be more useful to the stakeholders mentioned above, who would be interested in the fluctuation of liquidity at different timescales or different levels of interest.

Our modelling approach uses survival analysis to measure the duration of these droughts. In this context, we assess the explanatory power of survival time regression models for capturing these features of liquidity intra-daily, and we relate these durations to a number of instantaneous

and lagged covariates obtained from the LOB. This is achieved through the development of a survival time model, similar in structure to that used previously by Lo et al. [2002], and we present an analysis of the lifecycle of droughts in liquidity during the trading day.

We investigate the suitability of these models and determine the relevant covariates to include for various assets and under different market conditions. We also evaluate the predictive power of the models for the resilience of two different liquidity measures and interpret the structural features of the optimal models for each. We explore a very large number of model structures, across datasets for different companies and over an extended period of four months of detailed LOB data, in order to evaluate the relative importance of each LOB variable in the approach. We find that, in agreement with what one would expect, a larger difference of the liquidity from a threshold level would be associated with a longer deviation from that level. On the other hand, a larger frequency of such deviations from a liquidity threshold level would be associated with a swifter return to that level.

An obvious use for the liquidity resilience metric and the associated survival regression framework would be in improving one's estimates of expected execution times. Execution algorithms typically place orders when there is sufficient liquidity in the LOB, and this approach can produce the conditional mean of the period until that level of liquidity is achieved. Because of the flexibility of this approach, it can accommodate any liquidity measure and any liquidity level as a threshold. As such, it can be incorporated into a number of different execution algorithms, which use different liquidity measures as the basis for their decision making.

The resuls suggest that the resilience metric and associated model can have positive implications for market quality. Liquidity droughts have been identified as amplifying factors for market shocks, both when liquidity is considered over longer periods of time [Brunnermeier, 2008], and for intra-day periods [Easley et al., 2011]. In the latter case, this can be understood as liquidity providers not wanting to risk building positions which will be difficult or costly to unwind. Understanding that the duration of a liquidity drought is directly related to the state of the LOB can help inform liquidity providers that may be considering leaving the market. The continued presence of liquidity providers in the LOB will limit the extent of intra-day shocks, as the liquidity droughts which are central to most such shocks will be better understood and forecast.

Alternatively, in the case of a regulator or exchange, the state of the LOB can be informative about the expected intra-day liquidity drought durations. For stressed states, in which the LOB is expected to diverge from what are considered to be acceptable levels of liquidity for an extended period of time, a regulator could consider imposing quoting rules to liquidity

providers, or the exchange could provide incentives for them to improve the market.

1.4.3 Quantifying the commonality in liquidity and resilience

Recent studies involving different assets for data from both primary and secondary exchanges have shown that for a number of different liquidity measures, one may observe a high degree of commonality in liquidity. For equities, the observed commonality is significant at both the market and the industry level [Chordia et al., 2000, Huberman and Halka, 2001], while Brockman et al. [2009] also provide evidence of this, in both developed and emerging markets. The degree of commonality found across multiple different exchanges led the latter to comment that 'firm-level liquidity cannot be understood in isolation'. Liquidity co-movement has been found to be prevalent particularly during equity market breaks and debt market crises [Hasbrouck and Seppi, 2001].

Our first contribution in the liquidity commonality literature is to show that, at least in the equity space, the assumption that one can capture all features of liquidity commonality via a PCA regression approach, which is by far the most common way to quantify commonality, will not always be appropriate. In particular, using only second order moments will not capture heavy tailed features observed in the empirical distribution of the liquidity for certain assets. The outcome of using only PCA methods, which are based on second moments, is that the analysis is then driven by the most illiquid assets, which act as outliers in the cross-sectional dataset. We perform ICA (Independent Component Analysis), which incorporates higher order information, verifying that these assets correspond to the maximally non-Gaussian components.

In addition to the considerations regarding the appropriateness of the statistical assumptions for an analysis of commonality, one should note that existing liquidity commonality approaches only reflect the aspects of liquidity measure chosen. In the case of the spread, this would be the tightness, and in the case of the XLM, it would also reflect the depth. However, since such measures cannot quantify liquidity resilience, the associated commonality analysis will not reflect this aspect of liquidity either. We extend the analysis to determine if the liquidity commonality observed is also present when one incorporates notions of resilience.

The resilience metric proposed previously was indicative about the time one would expect to wait, for a particular liquidity measure to reach a user-defined threshold. In order to obtain a more complete picture about resilience, one could then calculate the expected time for liquidity to return to a number of different liquidity thresholds after a shock, where these thresholds could represent every possible value that a particular liquidity measure (e.g. the spread) can take during the day. Taken together, these expected durations would then form a curve, or a

resilience profile, indicating the expected duration of fluctuations from any threshold.

These profiles are informative about the level of LOB liquidity replenishment for each asset, therefore a commonality analysis can identify clusters of assets for which we would expect a swift return to high liquidity levels. Using these profiles as building blocks, we advance the literature on liquidity commonality by quantifying commonality in liquidity resilience for two common liquidity measures, namely, the inside spread and the XLM.

A notable contribution here was the first use of functional data analysis (FDA) in a financial LOB setting, in order to reduce the dimensionality of the liquidity resilience data and enable the comparison of functional data forms (the liquidity resilience profiles above). In particular, in an analogy with the use of PCA to extract market factors of daily liquidity, we use functional PCA, to obtain market factors contributing to the variation in daily liquidity resilience. Commonality can then be obtained as the explanatory power of these market factors for the resilience of each asset.

Our results suggest that these market factors for liquidity resilience can explain between 10% and 40% of the time required for the spread to return to a low threshold level after a shock. We interpret this figure as the degree of commonality in liquidity resilience for European stocks, and attribute it to market making activity in the LOB. One would expect that market makers use similar algorithms in replenishing LOB liquidity for different stocks, which would then manifest in common resilience behaviour in these stocks.

However, we also note that that the FPCA market factors are less explanatory about liquidity resilience at higher threshold levels, or for different liquidity measures, such as the XLM. Deviations from higher thresholds would indicate more extreme intra-day shocks. We have therefore suggested that the lower explanatory power of market factors at these levels may again be due to market makers, who are now less inclined to post liquidity in the LOB after these shocks. This is in order to avoid excess exposure in a market where it will be costly to unwind a position.

Contrasting these results with the liquidity commonality findings, we find that temporal commonality in the liquidity measures does not necessarily entail commonality in liquidity resilience. We would argue that this has positive implications for market quality, as it indicates that slow liquidity replenishment in certain assets is not necessarily contagious for the market.

1.4.4 Modelling trading activity in the LOB

While liquidity and resilience are certainly two important aspects affecting daily trading activity, they are only summaries of the complex stochastic process that is the LOB. There are several

other aspects of the LOB which are of interest to stakeholders, including the variation in the intensity of the trading activity, as well as the inter-dependence between limit order submission and cancellation activity at different levels. Variables that are central to LOB dynamics include the timing of arrival of orders, the placement and size of orders, as well as the shape of the LOB [Chakraborti et al., 2011]. Understanding these aspects can help inform trading decisions, but is also key to regulators considering interventions to help improve the quality of a particular market.

There have been two major approaches for modelling key features of the LOB, which can be broadly separated according to whether they consider the activity of the participants at the centre of the process. In the first approach, agent-based modelling (ABM) frameworks typically characterise the trading population in the LOB through a simple set of agent attributes that defines their trading behaviour. In models found in the economics literature, agents are assumed to act strategically, by maximising utility functions in order to determine their course of action, see for example [Parlour and Seppi, 2008, Roşu, 2009, Foucault et al., 2005]. However, many macro aspects of the LOB can also be obtained through even simpler agent characterisations, see for example Chen et al. [2012] for a discussion regarding the common characteristics of financial markets that can be explained through agent-based models of varying complexity. ABM has now also gained acceptance beyond the academic community, as Bookstaber [2012] has described it as a method that the US Office of Financial Research intends to use to improve their understanding of potential vulnerabilities in the financial markets.

The second approach abstracts away the market participant from the modelling process. Instead a stochastic modelling approach is taken, where the complex trading dynamics are distilled into a set of statistical assumptions. These models can capture key empirical properties of the processes comprising the LOB stochastic structure [Cont et al., 2010, Huang and Kercheval, 2012]. They also give rise to LOB simulation frameworks which feature these same properties, see for instance Daniels et al. [2003].

The final contribution of this thesis is a new form of LOB model, which keeps aspects of both agent-based modelling frameworks as well as the stochastic models approaches. The aim is to be able to capture key features of the observed LOB process (specifically related to the price and volume processes), but to also be able to interpret such features with regard to the agent behaviours the model characterises. The distinction between the activity of different agents is based on the liquidity motivations of the market participants. We develop two such types of agents in this framework, market makers (liquidity providers), and liquidity demanders, which are intended to encompass a stylised model for algorithmic traders, noise traders, trend

followers and other types of speculators. The model expresses abstractions of these traders' trading behaviours in a stochastic model framework, which is significantly more detailed than typical simple agent models.

In the past, there was some skepticism about ABMs, due in part to the 'perceived lack of robustness' [Windrum et al., 2007], as well as the lack of rigour in the calibration of the agent and model parameters. A novel development of this thesis is a rigorous and efficient statistical estimation procedure for the agent model parameters based on a combination of indirect inference, a simulation based likelihood procedure and multi-objective optimization. We calibrate our representative agent stochastic model to real high frequency data from Level 2 limit order book data from Chi-X. We show how such a procedure can be used to estimate the model such that the resulting simulations approximate real data in more than one aspect, in our case the behaviour of the intra-day price and volume processes.

One practical benefit of the agent-based modelling approach we develop is that under this model, one is able to estimate the effect of a regulatory intervention, as long as we can model the impact of a particular intervention on the order flow of the agent. For instance, it was recently shown that even the availability of information about trading interest in the LOB (pre-trade transparency) on its own is enough to affect trading behaviours [Boehmer et al., 2005, Bortoli et al., 2006], both in the size of the orders and the frequency of cancellations, while the effects on the market itself include reduced liquidity and increased volatility [Madhavan et al., 2005]. Under the agent stochastic model framework we are able to evaluate the effect of a 'quote-to-trade ratio' imposition, which has been discussed in the context of limiting high-frequency trading in the LOB. The empirical predictions of the model suggest that the imposition of such a ratio is, ceteris paribus, sufficient to limit extreme intra-day volatility in the price process.

1.5 Outline of thesis structure

Chapter 2 provides an overview of the operation of the LOB, the pre-eminent trading mechanism in modern stock markets and the matching mechanism of the exchange whose data we use in the empirical studies presented in this thesis. We also provide an overview of the LOB dataset used and then discuss how one can manipulate the text files containing the entire trading activity for a single day, in order to extract the data related to a single asset. We explain how this can be used to rebuild the entire LOB, and specify the data structures one can use to obtain an efficient implementation of the LOB. Finally, we provide some summary statistics regarding LOB activity.

Chapter 3 contains a review of the literature related to the work presented here. We discuss

some of the fundamental aspects of liquidity, and explain why it is an important concept for a number of stakeholders. We then list some of the most common measures of liquidity, according to the aspect of liquidity they reflect. We also provide some examples of the intra-day and long term variation of these measures when applied to this data and attribute the variation to market making activity and information asymmetry.

This is followed by a summary of empirical analyses of liquidity, starting with a discussion of liquidity commonality across markets and across regions, for different equity classes. Empirical studies for markets featuring different trading mechanisms have found that this can lead to different liquidity levels, and regulatory decisions and exchange rules and enhancements have also been shown to have an impact on liquidity.

The second part of the literature review is focused on financial market simulation models. The focus is particularly on financial ABMs, and the section starts with a short historical overview of the development of ABM, as well as a short overview of the different areas of application. This is then followed with the rationale of using ABM as an approach for economic and financial modelling, before a description of some of the most important economic ABM models. The financial ABM literature is then covered in depth, and we start with defining the axes along which we compare the models covered. It is not straightforward to compare ABMs, as they generally vary greatly, both in the level of abstraction of the trading mechanism, and in their goals.

We consider firstly the types of agent strategies used in models, and considering both simple, zero-intelligence formulations and more realistic strategies based on actual market behaviours. Then we discuss the mechanism for price determination used in each model, which separates models in which the price is determined endogenously and exogenously. Finally, we describe the use cases for such models, mainly in replicating real market features and in policy testing.

Chapter 4 introduces a new metric for the resilience of liquidity in the LOB. This is contrasted with existing definitions of resilience and we argue that the additional flexibility of the new metric makes it applicable to a wider range of situations. A quantitative framework to model the variation in the metric is also introduced, which is based on a survival regression approach. This approach can then encapsulate the effect of the state of the LOB on the resilience metric.

We identify the most important LOB covariates through extensive model selection over all variable combinations and over an extended period of time, so as to identify the variables that were found to be both significant and increase explanatory power. These covariates then

form the basis of models used for prediction. We also show how one can use obtain conditional quantile levels of the resilience metric, as a method to evaluate potential extreme levels.

In Chapter 5, we first review previous approaches on liquidity commonality and establish the disadvantages of using PCA regression as a method to extract common factors and quantify commonality. Empirical analysis of an equity dataset clearly demonstrates that such an approach will be driven by the most illiquid assets, and thus the commonality results will not be conclusive. We then show how one can address these issues with an Independent Component Analysis approach, which incorporates higher order information.

This is followed by a description of a model to quantify the commonality in the resilience of liquidity. We first introduce the financial data analysis framework to be utilised, and explain how it will be applied to the functional profiles of liquidity resilience. We summarise the variation in the daily market factors of resilience (extracted through functional principal components analysis) over time and utilise these in a functional regression setting. The explanatory power of these market factors over time for different assets is then interpreted as the level of commonality in liquidity resilience in the equity space.

In Chapter 6, we present an LOB simulation framework, where the distinction in agent trading activity is based upon their liquidity motivations. We first explain how this description is more appropriate for a modern financial market compared to previous approaches that had broadly divided activity into that originating from traders motivated by price fundamentals or from recent price activity. We then present a detailed stochastic model framework which has the flexibility to incorporate time-varying limit order, market order and cancellation intensities for the different levels of the LOB, as well as the heterogeneity in order sizes.

We also present a simulation-based approach to estimate this model, which is a combination of an indirect inference approach with multi-objective optimisation. We estimate a number of different versions of the model, starting with a basic model, where we make a series of assumptions, and then progressively relaxing those assumptions, in order to evaluate which elements of the model contribute most to capturing the required price and volume dynamics. Finally, we evaluate the effect of imposing increasingly restrictive quote-to-trade ratios on daily LOB activity.

Chapter 7 concludes, summarises the contributions of this thesis and sets out an agenda for future work.

Chapter 2

LOB construction and descriptive statistics

A challenge in financial econometrics is to summarise and study statistical features, or characteristics, of large-scale datasets, derived from unevenly-spaced observations at an ultra high-frequency arrival rate. The data structure outlining the buying and selling interest in an asset is known as the Limit Order Book (LOB). Market participants are typically allowed to place two types of orders on the venue: Limit orders, where they specify a price over which they are unwilling to buy (or a price under which they are unwilling to sell), and market orders, which are executed at the best available price. Market orders are executed immediately, provided there are orders of the same size on the opposite side of the book. Limit orders are only executed if there is trading interest in the order book at, or below (above), the specified limit price. If there is no such interest, the order is entered into the limit order book, where orders are displayed by price, then time priority.

The LOB mechanism is the most common method used in major stock exchanges to match the buying and selling interest in stocks [Jain, 2003]. As such, there have been a number of studies of order book flow [Ranaldo, 2004, Kavajecz, 1999, Biais et al., 1995], as well as attempts to model this flow using analytical models Roşu [2009], Cont et al. [2010], Cont and De Larrard [2013]. More recent studies of the order book focus mainly on the increasing representation of high frequency trading firms in the market [Hasbrouck and Saar, 2013], or the impact that their order flow has on market quality [Brogaard, 2010].

Figure 2.1 shows an example snapshot of the order book for a particular stock, as traded on the Chi-X exchange, at a particular instance of time. A market order to buy 200 shares would result in 3 trades: 70 shares at 2702, another 100 shares at 2702 and the remaining 30 at 2704. A limit order to sell 300 shares at 2705, on the other hand, would not be executed immediately,

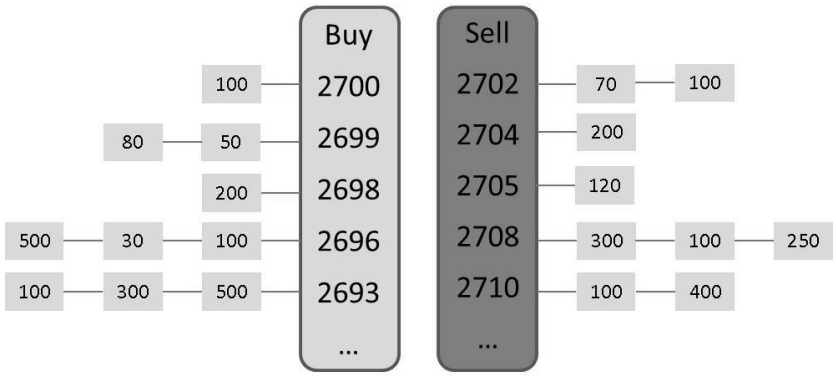


Figure 2.1: An example of the state of the Chi-X order book. The left hand side of the book is the buying interest, while the right hand side is the selling interest. The highest bid (order to buy) is for 100 shares at 2700 cents, while there are two lowest offers (orders to sell) for 70 and 100 shares at 2702. Orders are prioritised by price, then time.

as the highest order to buy is only at 2700 cents. It would instead enter the limit order book on the right hand side, second in priority at 2705 cents after the order for 120 shares which is already in the book.

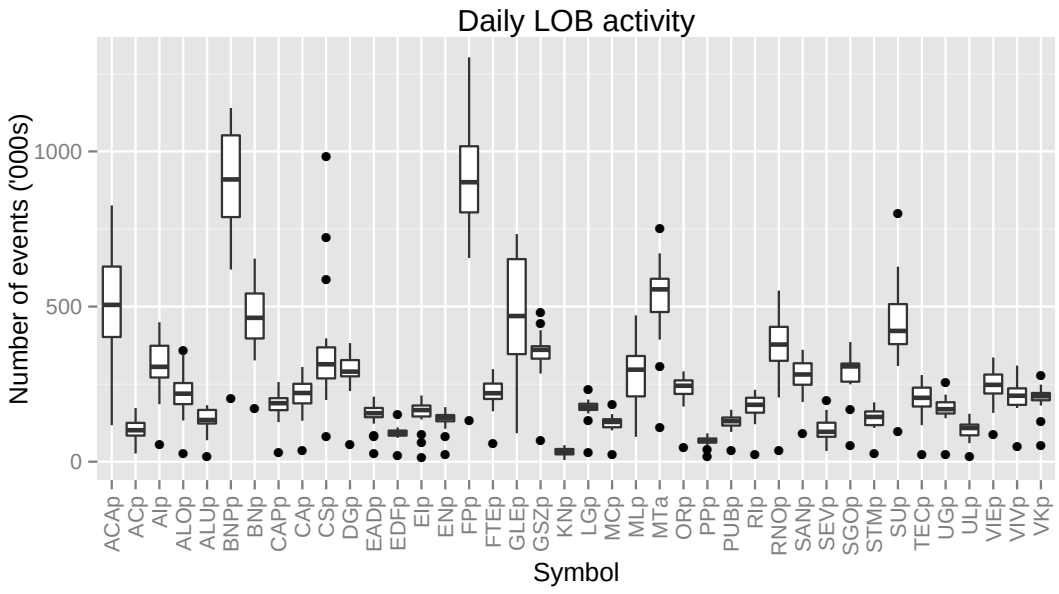


Figure 2.2: The variation in the total trading activity (including limit order submissions, executions and cancellations) in the LOBs of the stocks in the CAC 40, per day, for all trading days in January 2012.

2.1 Dataset and available order types

The data is provided by Chi-X, which (prior to its merger with BATS) was a pan-European multilateral trading facility (MTF). While it is not a primary exchange, it commands a large market share of the trading volumes for the assets traded on it. Indicatively, for the week starting the 2nd of January 2012, it had 25.5% of the CAC40 trading volume, 26.9% for the stocks on the DAX index and 33.4% of the FTSE100[1]. We note that Chi-X has not operated circuit breakers (temporary trading halts following sudden market moves) during the trading sample, which makes Chi-X data particularly clean for the purposes of our study given the number of complex circuit-breaker mechanisms operated by other exchanges (see, e.g. Brugler and Linton [2014] for a study of the efficacy of circuit-breakers on the London Stock Exchange).

The dataset consisted of over 1300 assets across a diverse range of sectors (e.g. financial, construction and telecommunications). The exchange has both a visible and a hidden order book and orders are routed to each book according to the type and size of the order: Limit orders, pegged orders and part of each iceberg order (explained below) are displayed in the visible book; orders meeting MiFID large in scale requirements are routed to the visible book, but remain non-displayed; orders in the hidden order book are executed at the mid-price. Chi-X operates a dark pool called Chi-Delta, which accounted for around 1% of trading volume for CAC40 stocks[2]. This data is not considered here.

Our dataset contains information about the visible book, which supports the submission of the following order types:

1. Limit order, with a specified price and size.
2. Market peg, which matches the best opposite price (e.g., the lowest sell price if it is a buy order). It can be priced, such that it does not move above (below) a particular limit.
3. Mid peg, which pegs to the price between the best available bid and offer.
4. Primary peg, which pegs to the best price on the same side.
5. Iceberg order, similar to a limit order, but only a portion of it is displayed.

These orders can have an attached maximum time in force, after which they are cancelled automatically:

[1] http://www.liquidmetrix.com/LiquidMetrix/Battlemap

[2] http://www.tagaudit.com/mydocuments/market_indicators_december_2011.pdf

- Day orders, which are valid for the day.
- Good till -, which are valid until a particular time.
- Execute and eliminate, which has a specified size and (possibly) a limit price. The portion of the order that is not traded immediately gets cancelled.
- Fill or kill, which is similar to the above, but either trades immediately in its entirety or gets cancelled.

The orders remaining on the LOB at the end of the day, regardless of the type or time in force option selected above, get automatically cancelled. This means that one only needs a single day's data in order to recreate the state of the LOB for that day.

It should be noted that our dataset only has information about limit order submissions, executions or cancellations, from which one would not be able to infer the order type or time in force. That is, a limit order submission may be the result of a mid peg, primary peg, iceberg order or limit order, while an execution may be the result of a market peg or limit order. A cancellation may be automatic (as a result of the fill or kill or execute and eliminate options), or as a result of a manual cancellation request. It may also be the result of a modification of the price of an order, which is represented in the data as a cancellation and resubmission with the same order ID. For the most part, we do not attempt to infer this information here, but rather propose models with fewer event types (limit orders, market orders and cancellations) in order to match the outcome of the interaction between these more complicated order types.

For more information about these order types, Prix et al. [2007] provide a breakdown of LOB activity on Xetra by order type and time in force, as well as the probability of every event sequence (from submission to full/partial cancellation, or to full/partial cancellation).

2.2 Manipulating 'flat' order files to obtain useful LOB data

The dataset consisted of the event logs from 82 trading days between the 2nd of January and the 27th of April 2012. These logs are in text format and the entire trading activity (limit orders, executions and cancellations) for every asset traded on Chi-X is included in a single file every day (these are termed 'flat' files, as they do not convey information about the structure of the LOB). The dataset therefore required some pre-processing, as these files typically range between 2 and 3 GB in size, and contain more than 50 million events (limit order additions, executions or cancellations).

In order to construct the LOB for a single asset on a particular day, it is then prudent (and faster) to separate out the events pertaining to that asset, before doing the matching. This presents some difficulties, as we will see in the data snapshot below that lines representing executions or cancellations omit details such as the stock symbol and whether the order being executed or cancelled is an order to buy or to sell.

A large initial undertaking was therefore to run through the entire dataset every day and separate out the orders for each asset in individual files. This was done in Python, which has the advantage of providing regular expression matching, and makes such an operation faster. The difficulty encountered above, where information from certain types of events was omitted, was overcome by using a machine with 32GB memory capacity and was able to keep every order in memory, so that subsequent cancellations and executions could identify which asset a particular event relates to.

The result of this process was a collection of more than 100,000 files containing the activity of all 1300 assets for every one of the 82 trading days. Along with the LOB implementation described in Chapter 2.3, we could now reconstruct the LOB and extract summary statistics for any asset and any day in the period under consideration very quickly.

The following is a snapshot from the dataset for the trading activity of German stock Siemens AG:

S28813122A 6539S 97SIEd 748700Y
S28813124X 6539 97
S28813129X 4610 97
S28813129A 4610B 97SIEd 747200Y
S28813130A 6973S 97SIEd 748600Y
S28813145X 6973 97
S28813147A 6982S 97SIEd 748600Y
S28813374A 7068S 76SIEd 748000Y
S28813374X 6982 97
S28813961A 7457S 316SIEd 747300Y
S28813962X 4610 97
S28813976E 7457 316 175
S28813977A 7467B 962SIEd 745200Y
S28814001A 7484B 97SIEd 747200Y
S28814437X 7068 76
S28815231X 7467 962

The 8 digits after the starting character 'S' represent the timestamp in milliseconds after midnight, London time. The first timestamp (28813122) then corresponds to 08:00:13.122. The last symbol in the first column is 'A', 'X' or 'E', representing an order addition, cancellation or execution respectively. The order types described above are thus all translated to limit orders by the matching engine, and market peg orders for example would then execute against these resting limit orders.

Order additions: In the second column we have the order ID, followed by 'B' for a limit order to buy, or 'S' for a limit order to sell. In the third column we have the number of shares and the stock symbol, and in the last column we have the limit price for the order (in one-hundredths of a cent) and a 'Y' to indicate the end of the line.

Order cancellations: In the second column we have the ID of the order being cancelled, and we have to look back to the original order to understand whether this is a cancellation of a buy or a sell order. In the third column, we have the number of shares being cancelled from the order, as Chi-X allows for partial cancellations.

Order executions: In the second column we have the ID of the order being executed, and we have to look back to the original order to understand whether this is an execution of a buy or a sell order. In the third column, we have the number of shares being traded from the order, and we may have a partial execution of a resting limit order, with the rest of the order remaining in the LOB. The final column is the trade ID, which is distinct from the order ID.

The Chi-X LOB data is undoubtedly a very valuable resource for studying intra-day LOB dynamics, but we should note that there are some potential issues with the data: For events within the same millisecond, the exchange does not guarantee that the ordering is correct. Identifying timing discrepancies can only be done in the case when a group of events breaks the price/time priority of the LOB, and even then, it is a manual process to construct the correct ordering.

In practice, having constructed the LOB for more than 100 stocks over a 4 month period, the number of events with obviously incorrect ordering (i.e. which broke price/time priority) were very few, as we observed such an event for every few days of data. In order to determine the effect of these timing issues on the evolution of the LOB, we observed the state of the LOB in the period immediately after these events. The effect was mainly on the size of executed or cancelled orders, and it generally dissipated after a few events, and in any case did not last for more than a few seconds. As such, we do not believe this data quality issue will have more than a minimal effect on the results presented here.

2.3 Building the LOB

The software used at Chi-X to reconstruct the LOB, in order to consider daily activity is not made available externally. For this reason, one of our first tasks was to create a matching engine (a virtual order book), that could read in every limit order arrival, execution and cancellation, recreating the state of the order book at each millisecond timestamp. In this way, one is able to view the LOB at a transaction-by-transaction level (event time) and can also obtain information that is not otherwise available using the raw data (number/volume of orders at each level of the order book, the age of orders in the order book, the value of certain liquidity measures that require detailed LOB data etc).

Constructing the LOB for an asset, in order to extract variables of interest and calculate descriptive statistics is a non-trivial task. Even though packages for visualising and retrieving data from an LOB exist in statistical software (e.g. in R Kane et al. [2011]), these are limited, in that they are not able to capture the required details of every individual order that are used in the models developed in this thesis. In addition, an R implementation of a complex data structure is bound to be slower, and thus not particularly suited to processing daily activity of hundreds of thousands of transactions for a single asset efficiently.

I therefore use the object-oriented programming language JAVA to construct the LOB, due to its library of data structures that can be leveraged to ensure an efficient implementation. My implementation is based on a number of base *classes*, as follows:

Order: Every limit order is represented as a class and contains the following information: Buy or sell order indicator, limit price, order size and order ID.

Limit: We use a class to represent the queue of limit orders at a particular level of the LOB. Every instance of the *Limit* class has an associated limit price and a *double-ended queue* of limit orders. This means that an order in the queue can be accessed in $O(N)$ time, where N is the number of orders at a particular level of the LOB. We do not generally observe large numbers of orders at any given level, therefore a more efficient data structure is not necessary.

OrderBook: The *OrderBook* is the class which processes every event (incoming limit order, execution or cancellation) read from the flat order file. At a high level, when a limit order arrives, the *OrderBook* instance checks to see whether there is already trading interest at that price (that is, an instance of the *Limit* class at that limit price, with at least one order in the queue). If it exists, an *Order* object containing the relevant information for that order is appended at the end of the queue. If it does not, an instance of the *Limit* class is created and the *Order* object becomes the sole object in the queue.

Of course, we will have limit orders at different limit prices and thus many different queues

of *Order* objects - these are instances of the *Limit* class. We will therefore need a data structure to hold these instances. For both the buy and sell side, we use the *TreeMap* class, where individual *Limit* queues are the objects mapped. The class is useful for this purpose because it is efficient in storing key (limit price)/value(Limit object) pairs in sorted order, and allows for fast retrieval of the objects (due to having both Tree and Map structures). The time complexity for adding, retrieving or removing elements from a TreeMap structure is $O(log(N))$, where N is the number of elements in the data structure. This compares favourably with the $O(N)$ time complexity for the same operations in a list structure. We thus have a considerable reduction in time spent for building the LOB using the structures above, compared to a naive implementation.

2.4 Descriptive statistics

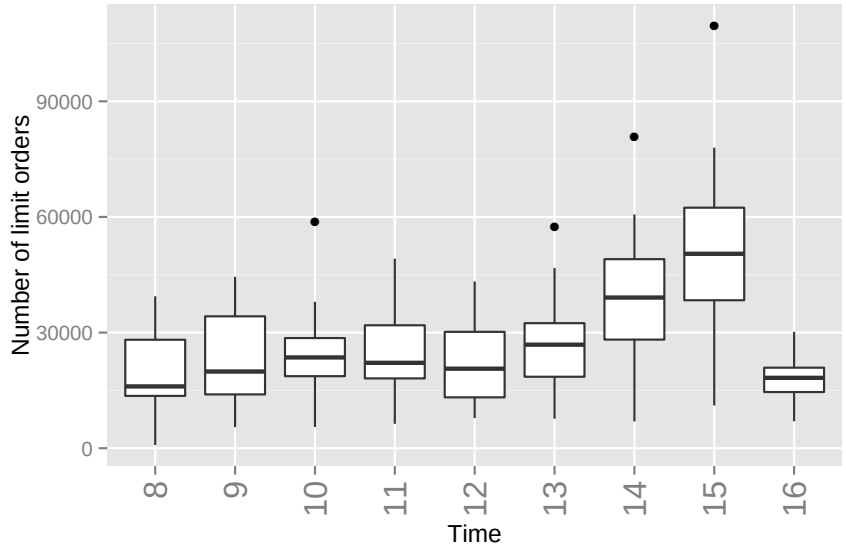


Figure 2.3: The variation in the hourly limit order submission rate for all trading days between January and April 2012 for Credit Agricole SA.

We considered data in the visible order book which contained the size (in shares), timestamp and individual ID of every limit order entered into the order book, along with the cancellations and executions of those orders, partial or in full. Using this dataset we can recreate the order book for any of the assets (stocks, ETFs and IDRs) traded on the venue on a particular day.

Electronic LOBs for liquid stocks are characterised by very high volumes of trading activity. Figure 2.2 shows the total activity in the LOB for the stocks in the main CAC40 index,

liquid French stocks per day, for all trading days in January 2012. In these figures, we include the number of order submissions, cancellations and executions, but at least some proportion of these would have been generated automatically, due to the submission of certain order types, such as the pegged orders described above. We observe that there are stocks for which trading activity is in excess of 1 million orders per day, for example BNP Paribas SA (stock symbol BNPp) and Total SA (stock symbol FPp). We find greater variation in the activity in some stocks - banking stocks BNP Paribas SA (BNPp) and Credit Agricole SA (ACAp) - than others, for example utility company GDF Suez SA (GSZp) and construction company Vinci SA (DGp).

We should note that while the continuous trading hours on Chi-X (08:00 to 16:30 London time) are not necessarily the same as those in the national exchanges where the assets trade, for some (like the French stocks Figure 2.2) the opening hours coincide. Hence, we do not have any additional considerations that would result from the sudden submission or withdrawal of liquidity from another exchange.

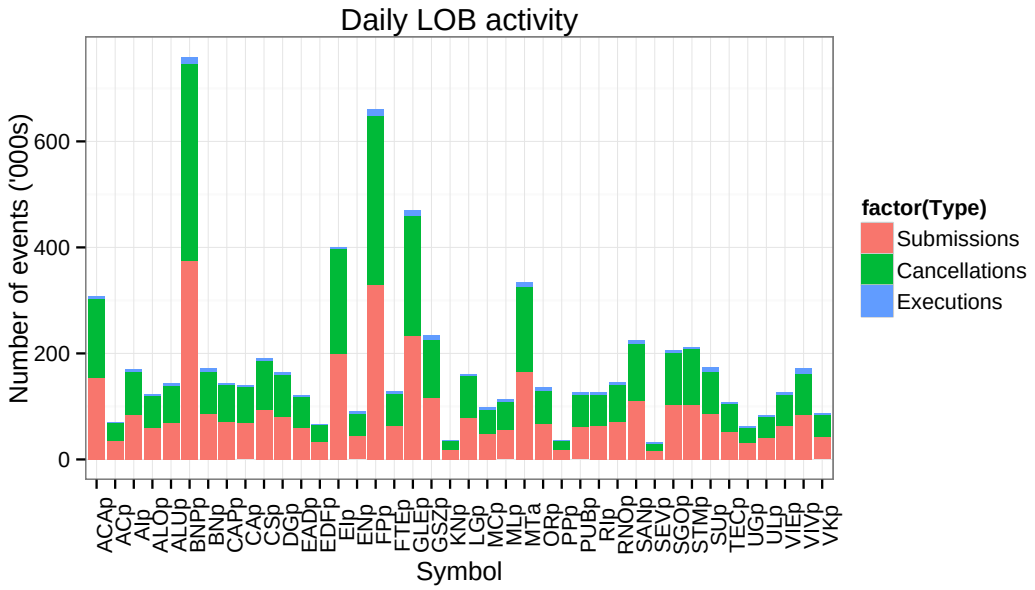


Figure 2.4: The trading activity for the CAC40 stocks on the 5th of March 2012, divided into limit order submissions, executions and cancellations.

The trading activity throughout the day is far from homogeneous - in European markets, there is a pronounced drop in activity close to mid-day, followed by a rise until the end of the day. The higher activity in the morning and afternoon has been documented in the past by Biais et al. [1995]. They also provide interpretations for this phenomenon, for example banks executing orders received before the market opens in the morning, or unwinding open positions in the afternoon. In Figure 2.3 we show the limit order submission activity per hour for the 4

month period in our dataset for Credit Agricole. We see that the submission rates are relatively stable for most of the morning, and then dip close to mid-day. The activity increases again after mid-day (when US markets open also) and continues to rise until the end of the trading day.

It is a feature of modern financial markets that trading activity is characterised by a relatively low proportion of trades, compared to the number of limit order submissions and cancellations. There are a number of factors contributing to this, for example the activities of market makers, who need to price their bids and offers appropriately, so as to avoid the risk of adverse selection [Glosten and Milgrom, 1985]. Some proprietary trading algorithms may also produce a large number of cancellations, due in part to 'chasing' the current market price for the asset, or seeking 'latent' liquidity (i.e. that is available, but not displayed) [Hasbrouck and Saar, 2009].

We note that unlike the London Stock Exchange[3], Chi-X does not charge a high usage surcharge or impose a quote-to-trade ratio. Participants therefore have few, if any, constraints restricting the behaviours above. In the subset of stocks we investigated, we found that executions (partial or full) typically account for between 1 and 5% of daily trading activity. Figure 2.4 splits the trading activity on a particular day into limit order submissions, executions and cancellations, from which it is clear that executions form only a small proportion of total trading activity.

[3]http://www.londonstockexchange.com/products-and-services/trading-services/pricespolicies/trading-services-price-list-effective-2-july-2012.pdf

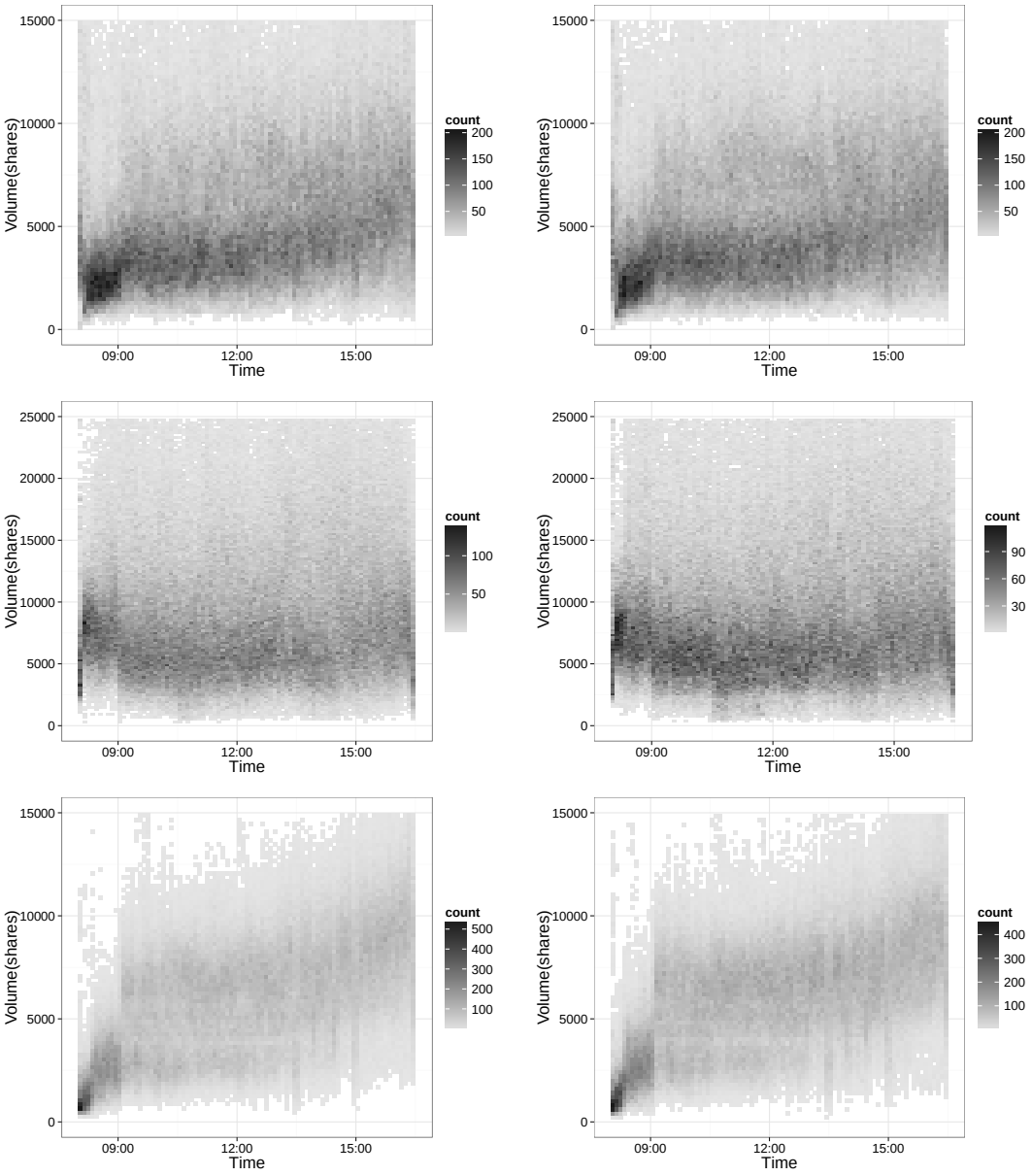


Figure 2.5: The variation in the volume (in numbers of shares) resting in the bid (left) and ask(right) side of the LOB throughout the day for stocks Peugeot (upper), Credit Agricole (middle) and Sanofi (lower). Volumes are extracted every 10 seconds throughout the period April-June 2012 and these 'heatmaps' aggregate information in 10 minute intervals. We find that volumes are low at the very start of the day, they increase rapidly and remain approximately constant for the rest of the day.

Chapter 3

LOB and liquidity related work

3.1 Importance of liquidity

We start with a brief, high-level definition of the two types of liquidity: funding and market liquidity. Funding liquidity is related to the ease of obtaining funding for trading, while market liquidity is related to the ease with which an asset can be traded. As such, they have different characteristics, and are usually measured at different timescales. However, Brunnermeier and Pedersen [2009] showed that there is an interrelation between funding and market liquidity, and changing margin and capital requirements are evident in the liquidity of financial assets. Jensen and Moorman [2010] confirm a link between monetary conditions and inter-temporal variation in liquidity.

Funding and market liquidity are central to macro and micro-economic structures respectively [Brunnermeier, 2008, Chordia et al., 2001]. For example, in macro-structure, liquid capital markets are essential for the efficiency of capital allocation, which results in low costs of capital for issuers, whilst at the micro-level, the liquidity of the market affects the interest in large numbers of trading interests. Hence, liquidity can ensure, on the micro-level, that investors can carry out their transactions at any time, allowing for a wide array of market participant behaviours. In this thesis we concentrate on trading interactions in the LOB, and we will thus focus on market liquidity.

Market liquidity is considered a desirable characteristic of financial markets, as in liquid markets we generally expect to observe fewer abrupt changes, or 'jumps', in terms of either the security price or the volume available for that security. In such markets, participants can both build positions in these securities, and liquidate them, without incurring substantial execution costs. For some investors, liquidity is the most important decision-making criterion in selecting the markets and assets they would like to invest in, and is a central concept that quantifies the quality of particular securities markets. We will now detail the reasons why liquidity is an

important concept for different stakeholders, including the firms whose assets are being traded, liquidity providers, asset managers, exchanges and regulators.

3.1.1 Firms and cost of capital

The relation between market liquidity and the cost of capital for a firm is examined by Butler et al. [2005]. They relate the flotation costs (in particular, the investment banking fees of the underwriting syndicate assisting a firm in raising capital) to stock market liquidity and find an inverse relationship between them. This is explained through the risks assumed by the underwriting syndicate (inventory and adverse selection risks), as it would be more difficult to place an equity issue in an illiquid market.

Market liquidity may also be a consideration in the pricing of the initial public offering (IPO) itself. In order to entice investors who are uncertain regarding a security's value, IPOs are often underpriced (see, e.g. Ellul and Pagano [2006] and references within). The model of Ellul and Pagano [2006] suggests that if post-IPO market liquidity is expected to be lower, the IPO itself should be underpriced further. This is in order to compensate investors for the trading costs that they expect to incur as part of liquidating their holdings. The model predictions are supported by an empirical study of the pricing of British IPOs.

Amihud and Mendelson [2008] suggest that firms can reduce their cost of capital by increasing the liquidity of their stocks or bonds, through a number of methods. Firstly, they can substitute dividends for stock repurchases. Then, the firm can consider moving away from highly leveraged capital structures, which are associated with lower market liquidity. Finally, increasing the investor base with smaller, less informed investors can improve liquidity by reducing the extent of asymmetric information in trading.

3.1.2 Liquidity providers

The majority of early studies and models for liquidity relate to quote-driven (dealer) markets, where specialists post bid and ask prices for the securities they are trading. In such a market, the existence of the bid-ask spread is suggested to arise from asymmetric information and the risk of adverse selection (i.e. execution of one's order against a trader with superior information) by Glosten and Milgrom [1985]. In this model, asymmetric information may arise due to the propagation of insider information, which leaves traders making markets with an uncertainty about whether they are trading with a counterparty with superior information. Specialists recoup the losses incurred from trading with better informed counterparties by trading with (presumably less-informed) liquidity-seeking traders. The width of the bid-ask spread by the specialist therefore becomes one of the factors that affects their profitability. An example of the

adverse selection component in a liquidity measure will be presented in Section 3.2.2.1.

Market makers supply 'immediacy', that is, they are willing to bear the risk of keeping a position until the final buyer or seller arrives to take that position [Grossman and Miller, 1988]. In a market with a single market maker, the bid-ask spread would depend on the market maker's inventory [Amihud and Mendelson, 1980]. Market makers have limits to the risk they may assume (and thus the inventory they may hold) and Hendershott and Seasholes [2007] suggest that they are willing to accommodate traders as long as they can then unwind their positions in the future at favourable prices.

In a pure LOB, liquidity would not be dependent on a limited number of specialists, as any trader can post limit orders, and the state of the LOB results from the aggregation of the trading interest in the market (i.e. previously entered, but not executed, limit orders to buy or sell). In a relatively liquid LOB, one would expect that multiple market makers would be operating simultaneously for every asset, and therefore the market would not be severely affected by the departure of a single operator.

3.1.3 Asset managers

Asset managers routinely take the role of liquidity demanders, who want to execute (buy or sell) a large trading volume for one or more assets. They care about liquidity, insofar as it affects their return on investment, as illiquid securities cost more to buy and can be sold for less [Foucault et al., 2013]. Many asset managers do not have a choice regarding the assets they invest in, if they are running passive strategies, such as those associated with index-tracking funds. Rather, their goal is to ensure best execution, by achieving the minimum expected cost for their trades over a particular trading period. This has given rise to wide and varied literature in optimal execution/liquidation.

One of the first papers to address the issue was by Bertsimas and Lo [1998], who showed how one could obtain the optimal division of a block into several tranches, using stochastic dynamic programming. Almgren and Chriss [1999] incorporated risk considerations in their framework for executions, which they argued were necessary to produce different strategies, depending on the level of liquidity in the market.

There are various aspects of liquidity which come into play for optimal execution. Obizhaeva and Wang [2013] suggest that 'it is the dynamic properties of supply/demand such as its time evolution after trades [...] that are central to the cost of trading and the design of optimal strategy'. Liquidity is consumed by a trade, and the degree and speed with which it is replenished has an impact on the execution strategy. This dynamic aspect of liquidity will be

modelled in Chapter 4.

3.1.4 Exchanges

Recent regulation has opened up the competition for the trading of financial instruments across different venues. In Europe, the relevant regulation is the MiFID[1], while the US equivalent is Reg NMS[2], and both were implemented in 2007. This has led to fragmentation, a situation where a number of trading venues compete for order flow and a trader may experience increased 'search' costs as a result. For a market maker, for example, these search costs may include additional hardware located at every venue, in order to ensure the swift updating of quotes. For an institutional investor, these include 'smart routing' equipment, in order to ensure that market orders are routed to the venue offering the best prices for execution. Fragmentation is particularly prevalent in the equities markets, as the incumbent exchanges lost their monopolies, but it has not necessarily harmed market quality [O'Hara and Ye, 2011]. A measure of fragmentation in equities markets is provided by Fidessa, through their Fragementation Index[3].

Because of this fragmentation, exchanges have had to find ways to differentiate themselves from the competition, and constant, high-quality provision of liquidity is seen as an important way to attract custom. The NYSE website[4] suggests that 'today's NYSE is designed to maximize liquidity', with 'expert Designated Market Makers (DMMs) adding over 300 million shares of liquidity daily' along with Supplemental Liquidity Providers and Floor Brokers with algorithmic trading tools, to create more liquidity. The London Stock Exchange[5], on the other hand, lists the market makers active on the exchange, who provide 'continuous pricing and a high quality pool of liquidity'.

The cost of liquidity provision has various components, including adverse selection, order

[1] MiFID, available at `http://ec.europa.eu/internal_market/securities/isd/mifid/index_en.htm` introduced Multilateral Trading Facilities (MTFs), increasing competition between venues by offering the opportunity for pan-European trading through 'passporting' services across borders. However, it also led to an increase in fragmentation and dark pool trading, according to ESMA (`http://www.esma.europa.eu/system/files/09_355.PDF`).

[2] Regulation NMS, available at `http://www.sec.gov/rules/final/34-51808.pdf`, set out a number of rules designed to foster competition and fairness between exchanges. These included the Order Protection Rule, which aimed to ensure that investors receive best execution amongst immediately available quotes and the Access Rule, which aimed to improve linkage between NMS trading centres in order to promote fair and non-discriminatory access to quotations.

[3] `http://fragmentation.fidessa.com/`

[4] `http://www.nyse.com/equities/nyseequities/1166830723427.html`

[5] `http://www.londonstockexchange.com/traders-and-brokers/security-types/etfs/market-makers/market-makers.htm`

processing and inventory holding costs, some of which are reflected in the liquidity measures summarised in Section 3.2.2. An exchange that wants to attract liquidity providers has to understand whether any of these components dominate the others and direct their efforts towards reducing these costs. This may be through a change in trading system, if the aim is to reduce order processing costs, or improvements in disclosure, if adverse selection costs dominate [Foucault et al., 2013].

3.1.5 Regulators

Regulators, such as the US SEC, aim to 'protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation'[6]. It can be argued that liquidity is an intrinsic part of an orderly market, and as such, regulators have often enacted legislation in order to improve liquidity. As an example, following the introduction of the Common Cents Stock Pricing Act of 1997[7], the SEC either introduced or changed a number of trading rules in the NASDAQ, which included the Order Handling Rule as an amendment to the Quote Rule, the Limit Order Display Rule and others[8]. As a result of this, bid-ask spreads on the NASDAQ were reduced significantly.

Regulators focus interest in market liquidity additionally because of their third aim mentioned above, facilitating capital formation. We have identified a link between liquidity and the cost of capital identified above, and this was more than evident during the 2008 crisis. Foucault et al. [2013] describes the liquidity dry-up during this period, as well as the associated reduction in security issuance.

Market liquidity is therefore an important element that characterises the quality of financial markets, and its fluctuations have serious repercussions for a number of stakeholders. In the rest of this section, we will thus focus on describing the most important aspects of liquidity, listing common measures and summarising some of the most important empirical studies in the literature.

---

[6] `http://www.sec.gov/about/whatwedo.shtml`

[7] The Common Cents Stock Pricing Act of 1997, introduced to the U.S. Congress on 13/03/1997, 'amends the Securities Exchange Act of 1934 to instruct the Securities and Exchange Commission to require quotations in dollars and cents (decimals) for equity securities transactions and prescribe an implementation schedule' (`https://www.congress.gov/bill/105th-congress/house-bill/1053`). A summary of the bill can be found at the same location.

[8] The text for both these rules can be found at `https://www.sec.gov/rules/other/34-38156.txt`

3.2 LOB liquidity

The notion of LOB liquidity is considered a fundamental concept in high frequency financial modelling, but it has proven very difficult to capture via a single definitive measure. Common liquidity proxies reflect one or more of the following aspects [Von Wyss, 2004]:

1. **Trading Time/Immediacy**: The ability to execute a transaction immediately at the prevailing price. In its report, the IMF [Sarr and Lybek, 2002] calls this dimension 'immediacy', defining it as 'the speed with which orders can be executed, and, in this context, also settled, and thus reflects, amongst others, the efficiency of the trading, clearing and settlement systems'.

2. **Tightness**: The ability to buy and sell an asset at about the same price at the same time. This is commonly thought of as the bid-ask spread, although there are also other proxies to tightness that are used in the industry. The width of the bid-ask spread can be considered as part of the transaction cost, as it has to be crossed in order to enter and exit a position, when immediate order execution is required.

3. **Depth**: The ability to buy or sell a certain amount of an asset without (much) influence on the quoted price. The depth aspect relates firstly to whether immediate execution of a large order is possible, and secondly, whether it can be carried out at or close to the best bid or offer.

At the most basic level, one could consider the concept of liquidity to be the ability to convert shares into cash, and vice versa, at the lowest transaction costs. In Harris et al. [1991], a perfectly liquid market is defined to be one in which any amount of a given security can be instantaneously converted to cash and back into securities at no cost. Of course, in practice, this is unrealistic, and so the more realistic definition one may consider is that a liquid market is one in which the costs associated with the conversion are small.

These costs can be divided into explicit costs, which are the fixed costs associated with trading, and implicit costs, due to fluctuations in trading interest. Explicit costs could include brokerage fees, exchange trading fees (for example, many LOBs feature the maker-taker fee structure where aggressive traders pay a fee and liquidity providers receive refunds for making markets) and settlement/clearing fees. These fees are unavoidable, and even though they have been dropping steadily over the past few years [Foucault et al., 2013], we will not consider them in detail here.

Implicit costs, on the other hand, measure the difference between the ideal price one could obtain in the perfectly liquid market of Harris et al. [1991] and the market one operates in.

When market liquidity is high, one can buy or sell large volumes of an asset at a price close to the ideal price. As trading interest tends to vary throughout time, however, so do these implicit costs.

3.2.1 Fluctuation of liquidity

The fluctuation of market liquidity has been suggested to originate from the adverse selection problem faced by market makers more than four decades ago [Bagehot, 1971]. Since then, a rich literature has developed in single-asset liquidity, examining both the properties of liquidity measures in and of themselves, but also the effects of liquidity on asset pricing. As examples of the latter, we mention Amihud and Mendelson [1986], who model the effects of the spread on asset returns and find evidence of a 'liquidity premium', with assets with higher spreads commanding higher returns. Amihud [2002] confirm that a return-illiquidity relationship exists over time, but Constantinides [1986] finds less of an impact in multi-period models.

Since these earlier studies, access to massive high-frequency limit order data has allowed for significantly larger studies on liquidity across markets to be undertaken. This is a topical aspect of big data analysis in financial econometrics, and efforts to improve the understanding of liquidity evolution and co-evolution in the asset cross-section have recently been a focal issue of studies in equities, [Hasbrouck and Seppi, 2001, Karolyi et al., 2012, Riordan et al., 2013, Sklavos et al., 2013], commodities and futures [Frino et al., 2014, Marshall et al., 2013] and foreign exchange and bond markets [Holden et al., 2015].

3.2.2 Liquidity measures

Amihud [2002] suggests that liquidity is an elusive concept, one which cannot be captured with a single liquidity measure. As such, a number of theoretical aspects have been proposed, which capture different dimensions of liquidity. Certain aspects of liquidity have their basis in financial theory, and have been the subject of several studies.

The first is price impact and is is conceptually related to the price formation model of Kyle [1985]. It reflects the extent to which the price of an asset changes in response to a given order flow. Brennan and Subrahmanyam [1996] propose to measure price impact through the coefficient of signed order flow in the regression of price changes. Amihud [2002] obtains a low-frequency proxy as the ratio of average absolute returns to trading volume.

In an illiquid asset, part of the price impact is temporary [Mancini et al., 2013], and one may also consider the aspect of the reversal of the price following this temporary price impact. This is studied, e.g. in Stambaugh [2003]) and occurs following trading pressure that leads to excessive appreciation (depreciation) of an asset. It can then be considered as the subsequent

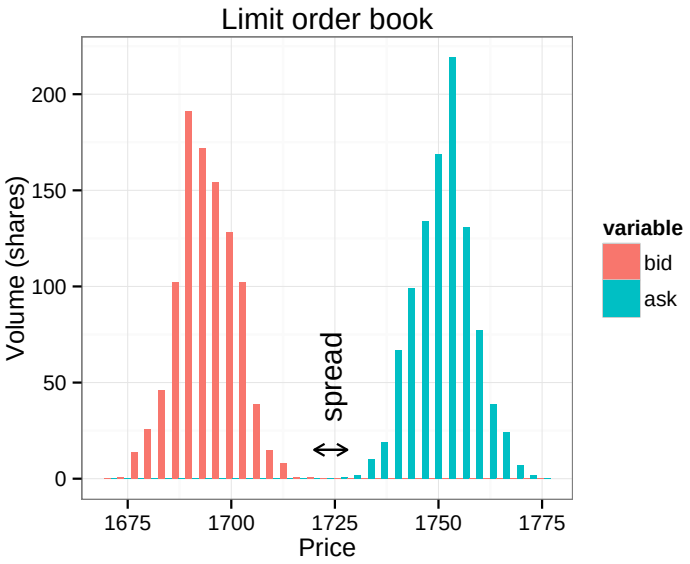


Figure 3.1: A possible state of the LOB and the inside spread.

reversal to the fundamental value [Campbell et al., 1993].

Even though such aspects of liquidity are driven by theoretical models, the measures above are not observed in real time and as such, are infrequently utilised in a trading environment. Instead, the availability of LOB data enables the provision of a richer representation of the trading interest for a particular asset and also, the quantification of ex-ante committed liquidity. To this end, one can identify the dimensions of **trading cost**, **quantity** and **time** as being relevant to practitioners [Holden et al., 2015].

Vayanos and Wang [2013] describes liquidity measures that do not fall under the category of theoretically-driven measures as being heuristic. However, they offer the advantage of being able to be evaluated in a very short period of time, and even through snapshots of the LOB. In addition, these are the measures that are most often reported by financial exchanges, which indicates their utility from a practitioner's point of view.

3.2.2.1 The cost dimension

The simplest measures of liquidity include variants of the popular family of spreads between levels of the bid and ask on the order book. The bid-ask spread, or the difference between the highest bid and lowest offered price in the LOB, represents the cost that an investor must incur in order to be guaranteed immediate execution, i.e. by crossing the spread with a market order. Figure 3.1 shows a possible state of the LOB, where volumes are aggregated per price level, and indicates the inside spread as the difference between the highest bid price $P_t^{b,1}$ and the lowest

ask price $P_t^{a,1}$

$$S_t = P_t^{a,1} - P_t^{b,1} \tag{3.1}$$

Further members of this family consider levels of the order book beyond the best bid or ask. For the n-th best price point (i.e. the n-th highest price level for the bid and n-th lowest price level for the ask for which there is volume resting in the LOB) we have

$$S_t^n = P_t^{a,n} - P_t^{b,n}$$

and we see in Figure 3.2 an example of how these spreads vary throughout the day for a selection of liquid European stocks.

The proportion of cancelled orders in the LOB (exhibited in Figure 2.4) shows that the instantaneous state of the LOB is not necessarily a good indicator of liquidity, as much of the volume posted on it will not be executed. It is easy to see the limitations of the bid-ask spread as a liquidity measure in this rapidly changing LOB. It is therefore necessary for us to consider this aspect of the LOB in any liquidity and liquidity resilience measure we propose.

Variants of the spread are discussed in detail by Goyenko et al. [2009], Holden et al. [2015] and include the simple percentage spread and log quoted spread

$$S_t^{perc} = \frac{P_t^{a,1} - P_t^{b,1}}{P_t^{mid}} \tag{3.2}$$

$$S_t^{log} = \ln(P^{a,1}) - \ln(P_t^{b,1}) \tag{3.3}$$

where $P_t^{mid} = \frac{P_t^{a,1}+P_t^{b,1}}{2}$. The percentage effective spread accounts also for trades that can happen inside and outside the best bid or offer[9], and this is updated after every trade

$$S_t^{eff} = 2\left|\ln(P_k) - \ln(P_t^{mid})\right| \tag{3.4}$$

where P_k is the price of the last trade at or before time t. It can be broken down into two components:

[9]Trading at prices other than the best bid or offer may occur when one buys or sells assets through a broker-dealer firm, rather than an exchange. Additionally, a market order submitted to the LOB may consume all the volume available at the best price and therefore 'walk up' or 'walk down' the book in search for additional volume. In this case, the execution price of the order can be considered as an average of the prices at which the order traded, weighted by the amounts traded at each price.

A trader submitting a market order may find their order executed against 'hidden' liquidity, for example an iceberg order, for which the trade price may fall within the bid-ask spread.

- The realised spread, which accounts for the temporary component of the effective spread

$$S_t^{real} = 2\gamma_k(\ln(P_k) - \ln(P_{t+\Delta t}^{mid}))$$

where

$$\gamma_k = \begin{cases} 1, & \text{if the liquidity demander submitted a buy order for the } k\text{-th trade} \\ -1, & \text{otherwise} \end{cases}$$

and Δt is a suitable time interval, typically 5 minutes [Goyenko et al., 2009].

- A price impact component, which accounts for the permanent component of the effective spread. This can be considered as the adverse selection component, as it measures the adverse change in the mid price for the liquidity provider.

$$S_t^{imp} = 2\gamma_k(\ln(P_{t+\Delta t}^{mid}) - \ln(P_t^{mid}))$$

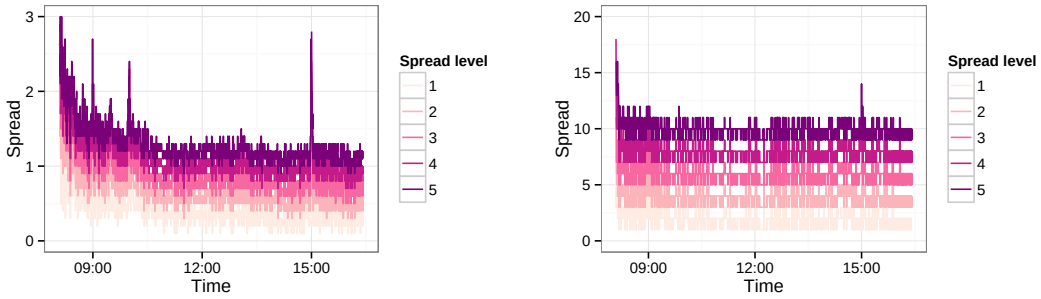


Figure 3.2: The variation in the spreads between the first 5 levels of the bid and ask for stocks Credit Agricole (left) and Sanofi (right). One can clearly see that the two stocks have different ticksizes (minimum possible changes in price), equal to 0.1 cents for Credit Agricole and 0.5 cents for Sanofi. The spreads measured here are $S_t^1 = P_t^{a,1} - P_t^{b,1}, \ldots, S_t^5 = P_t^{a,5} - P_t^{b,5}$. The spike at 15:00 corresponds to the time of an economic announcement in the US on that day (10 am ET).

3.2.2.2 The quantity dimension

The quantity dimension is related to depth, or 'the size of an order flow innovation required to change prices a given amount' [Kyle, 1985]. If $\boldsymbol{V}^{a,i}$ is the vector of sizes for the orders resting in the i-th level of the LOB and $\mathbf{1}$ is a vector of 1s then $TV^{a,i} = \mathbf{1}^T \cdot \boldsymbol{V}^{a,i}$ corresponds to the total volume (in number of shares) of orders at that level of the ask side of the book. Thus, in order to shift the mid price P_t^{mid} upwards, one requires a market buy order of at least $TV^{a,1}$

in size, while to shift the mid price downwards, a market sell order of at least $TV^{b,1}$ in size is required.

We can then can obtain 4 related depth measures [Von Wyss, 2004], the simple and log depth, the cumulative depth and dollar depth:

$$\begin{aligned} D_t^s &= TV^{a,1} + TV^{b,1}, \\ D_t^{log} &= \ln(TV^{a,1}) + \ln(TV^{b,1}), \\ D_t^{dollar} &= \frac{TV^{a,1}P_t^{a,1} + TV^{b,1}P_t^{b,1}}{2}, \\ D_t^{cum} &= \sum_{i=1}^{n} TV^{a,i} + \sum_{i=1}^{n} TV^{b,i}. \end{aligned}$$

3.2.2.3 Time dimension

The time dimension of liquidity is difficult to capture from the ex-ante committed volume, as it reflects the dynamic view of liquidity. As such, aspects of it can be modelled by the number of transactions per time unit, the number of orders submitted per time unit and the partial, or complete fill rates for orders submitted to the LOB [Von Wyss, 2004]. This dimension is related to resilience, or the speed of liquidity replenishment, which is considered in detail in Chapter 4.

3.2.2.4 Mixed measures

One recognises the need to extend the family of liquidity measures to consider the market impact and opportunity costs of trading when defining liquidity, especially from a large investor's perspective. To achieve this, the volume of resting orders in the LOB must be taken into consideration in the liquidity measure. Along these lines, Irvine et al. [2000] investigate properties of a particular class of measures of liquidity known generically as 'Cost of Round Trip Trade' (CRT) measures, where the round trip here refers to buying and immediately selling a certain amount of an asset. Such measures have the property that they summarise the structure of the LOB instantaneously for a given order size through a process of aggregation of key features of the LOB. By construction, they are intended to capture the ex-ante committed liquidity immediately available in the market.

The Xetra Liquidity Measure (XLM), proposed by Deutsche Boerse AG [Gomber and Schweickert, 2002], falls under the umbrella of CRT measures. Empirical studies using the XLM have compared liquidity costs across assets and also attempted to define and quantify

liquidity risk [Gomber et al., 2004, Ernst et al., 2012].

$$XLM_t(R) = \frac{\sum_{i=1}^{k} TV_t^{a,i}(P_t^{a,i} - P_t^{mid}) + (R - \sum_{i=1}^{k} TV_t^{a,i})(P_t^{a,k+1} - P_t^{mid})}{R} + \frac{\sum_{i=1}^{j} TV_t^{b,i}(P_t^{mid} - P_t^{b,i}) + (R - \sum_{i=1}^{j} TV_t^{b,i})(P_t^{mid} - P_t^{b,j+1})}{R} \tag{3.5}$$

with

$$k = \max\left(m \in \mathbb{N}; \sum_{i=1}^{m} TV_t^{a,i}(P_t^{a,i} - P_t^{mid}) \leq R\right)$$
$$j = \max\left(n \in \mathbb{N}; \sum_{i=1}^{n} TV_t^{b,i}(P_t^{mid} - P_t^{b,i}) \leq R\right).$$

This therefore returns the cost of a round trip weighted by the volume at each price, and a typical size is $R =$€25,000.

In the rest of this thesis, the empirical analysis focuses mainly on the inside spread (as the most commonly used liquidity measure) and the XLM (as the liquidity measure which captures most of the aspects of liquidity described above, in a single measure).

3.2.2.5 Liquidity approximations

It should be noted that in many studies, liquidity measures are obtained by approximation, due to the great expense required to obtain LOB data. An approximation for the spread, for example, can be obtained via the method of Roll [1984], using the first-order serial covariance of price changes. Where liquidity measures require a detailed breakdown of order volumes per LOB level, but limit order volume data is aggregated at different price levels, one could use the approach of Christensen et al. [2013] in order to uncover the most likely disaggregation into individual orders. In addition, in measuring commonality, it is also common to obtain lower frequency proxies for liquidity (Amihud [2002], Stambaugh [2003]), in order to reduce the data to a manageable size.

Goyenko et al. [2009] demonstrates the superiority of high-frequency liquidity benchmarks, compared to low-frequency proxies, while Mancini et al. [2013] argues for the use of good quality data as a necessity for measuring the determinants of liquidity. In this thesis, we use a millisecond-timestamped dataset to obtain liquidity data over a four month period, and through the reconstruction of the LOB, we do not need to rely on approximations of liquidity measures. This gives us the advantage of being able to obtain very accurate estimates of liquidity, and draw clear conclusions regarding aspects of liquidity.

3.3 Empirical analysis of liquidity

3.3.1 Temporal variation in liquidity

In the financial literature, early models focusing on quote-driven (dealer) markets had attributed the variation of liquidity on inventory holding costs [Amihud and Mendelson, 1980] and the risk of adverse selection [Easley and O'Hara, 1987], while Huang and Stoll [1997] also find a large order processing component. Affleck-Graves et al. [1994] then found differences in the breakdown of the spread into these components in quote and order-driven markets. Studying timeseries of variations in the bid-ask spread, Chordia et al. [2000] found that (long-term) liquidity is influenced by factors such as interest rates, market volatility and seasonal effects.

3.3.1.1 Intra-day variation

The intra-day variation of liquidity in markets has not been found to be consistent across markets. For example, while Chan et al. [1995] found a declining intra-day spread for NASDAQ securities (an L-shaped pattern), Wood et al. [1985] and Abhyankar et al. [1997] found a U-shape pattern (with larger spreads at the beginning and at the end of the day), for the NYSE and LSE respectively. Brockman and Chung [1999] found an inverted U-shaped pattern for the depth, which mirrors the U-shaped spread pattern (in that the peak of the depth and the trough of the spread both correspond to higher levels of liquidity).

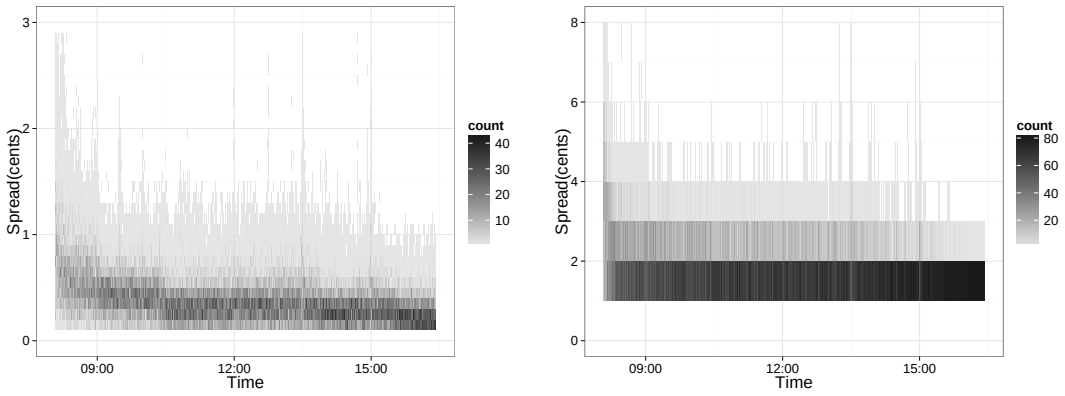


Figure 3.3: A heatmap of the intra-day value of the spread for Credit Agricole (left) and Sanofi (right) throughout the 4 month period.

In our Chi-X equity data, we identified an L-shaped pattern for both the spread and XLM liquidity measures, and we showed in Figure 3.2 an example of the spread of Credit Agricole and Sanofi on a particular trading day. We also show that this is consistent throughout time, with the heatmap in Figure 3.3. There are different reasons why the aforementioned L-shaped and U-shaped intra-day patterns may exist [Brockman and Chung, 1999].

The L-shaped pattern may arise to the existence of adverse selection, where a liquidity level would then reflect the level of asymmetric information in the market. Daily opening follows a period of non-trading, in which private information has accumulated (for example, when a trader has access to a report suggesting that earnings will be substantially different to what the market expects), but has not yet had time to be reflected in asset prices.

The U-shaped pattern can arise due to demand inelasticity and inventory management considerations, particularly close to the end of the day. Certain market participants need to trade to avoid being left with overnight positions. Liquidity suppliers can then exploit this demand with wider spreads.

3.3.1.2 Long-term liquidity variation

The theoretical model of Brunnermeier and Pedersen [2009] suggests that there should be a bi-directional relationship between market and funding liquidity. That is, the extent to which liquidity providers can provide market liquidity depends on the ease with which they can fund their positions. On the other hand, deteriorating market conditions can affect their funding, through destabilising margin requirements.

Empirical evidence of the significance of financing constraints as an impediment to high levels of liquidity is provided by Comerton-Forde et al. [2010]. The inventory positions of specialists on the NYSE are analysed over 11 years, alongside firm-level spreads. The authors find that these spreads widen after periods of poor performance, and this is consistent with specialists being more hesitant to take on inventory as their leverage ratios would already be stressed.

3.3.2 Commonality in liquidity

The majority of the literature discusses single asset liquidity and thus only captures individual variation in liquidity dynamics. Recently, however, there has been a burgeoning interest in studying the cross-sectional variation in liquidity in a number of assets over a period of time. One of the earliest studies to consider the co-movement of liquidity was in the work of Chordia et al. [2000]. This was achieved through a simple parametric model setting, by regressing liquidity changes for each asset against market or industry liquidity changes. The authors identified asset specific and aggregate market trading levels as being amongst the determinants of individual asset liquidity. Liquidity co-movement was assumed to result from the risk of maintaining inventory in the presence of institutional funds with correlated trading patterns.

Since this study, a number of asset specific and liquidity measure specific studies have been developed to quantify commonality. Hasbrouck and Seppi [2001] adopted a distinct frame-

work, using a combination of principal components analysis (PCA) and canonical correlation analysis to study the commonality in liquidity measures for the Dow 30. This uncovered the most important across-asset common factors in the price discovery/liquidity provision process in equity markets. They found that both returns and order flows are characterised by common factors. The liquidity measures they considered included variants of the spread, depth and the ex-ante trading cost. They found that commonality in order flows can account for roughly two thirds of the commonality in returns, but the common factors in the liquidity proxies above are relatively low. We provide an introduction to the dimensionality reduction techniques used in this study and discuss the implications for using PCA regression in the context of measuring commonality in Section 5.

Adding to the findings of Chordia et al. [2000], the study of Domowitz et al. [2005] also demonstrated that liquidity commonality in the Australian Stock Exchange may be induced by the co-movement in supply and demand, which materialises in the LOB as the cross-sectional correlation in order types (market and limit orders). The economic justification for this order type co-movement stems from traders' efforts to minimise execution costs, by submitting limit orders in an illiquid market, and market orders in a liquid one. They then demonstrated a linkage between liquidity commonality and return co-movement, which they argued is a key component of portfolio selection. Interestingly, they also argued that in contrast to liquidity commonality, return commonality is less affected by the correlation of order types, but is more related to the co-movement of aspects of the order flow, and specifically, order direction and size.

Due to the recent developments in big data analytics and the increasing availability of data, the processing of massive, multi-asset, multiple-day high frequency LOB datasets has become more tenable. Consequently, there has been an increasing interest in extending the smaller studies discussed above to encompass multiple days, assets and exchanges. Brockman et al. [2009] extend the model of Chordia et al. [2000] and consider 47 markets (exchanges) in 38 countries. In addition to the exchange-level commonality identified by Chordia et al. [2000], they find a global component in bid-ask spreads and depths, as well as regional components.

Similarly, an analysis of a dataset of more than 4000 firms over almost 20 years by Korajczyk and Sadka [2008] found that approximately 50% of the time-series variation in firm-level quoted and effective spreads can be explained by the first 3 principal components. More recently, Karolyi et al. [2012] considered daily equity data for 21,328 stocks in 40 developed and emerging countries between 1995 and 2004. They are able to demonstrate commonality in returns, liquidity and turnover and they explain this commonality through features of both the supply and demand sides of the market. On the supply side, they considered factors relating to

funding liquidity of financial intermediaries and on the demand side, factors related to investor protections, investor sentiment and trading behaviour of institutions.

In the FX space, utilising a dataset considering 40 FX rate liquidities over an extended period of 20 years, Karnaukh et al. [2013] found that commonality can explain an average of 36% of the variation in liquidity. However, this is higher in currencies in developed countries, as well as in times of market distress. The computational constraints of undertaking analysis across such an extended period of time are handled by extracting individual FX rate liquidity through PCA across low frequency liquidity proxies . The authors find that co-movements of FX rate liquidities are strong for at least the last 20 years, and, certainly, significantly stronger that in the equities asset class.

Understanding liquidity commonality is crucial for the success of strategies like the carry trade. For 9 currency pairs, Mancini et al. [2013] document strong contemporaneous comovements across exchange rate liquidities, and extract common information across 5 different liquidity measures. In common with the work presented in this thesis, they are able to utilise a high quality dataset and thus do not rely on approximating measures of liquidity to perform their analysis. They use both averaging (used previously by Chordia et al. [2000] and Stambaugh [2003]) and PCA (used by Hasbrouck and Seppi [2001] and Korajczyk and Sadka [2008]) to extract market-wide liquidity. They test for commonality by regressing individual liquidity measures against the first component for every exchange rate, and find that this explains between 70% and 90% of the variation in liquidity.

While liquidity may co-move across assets, the absolute level of liquidity for each asset will vary, and this will depend on a range of factors which may be different across asset classes. Stoll [1978] found that higher spreads were observed in more volatile stocks, while one can expect narrower spreads in larger stocks (in terms of market capitalisation) - see Vayanos and Wang [2013] and references within. In the bond markets, Chen et al. [2007] observed higher spreads for corporate bonds with lower rating or higher maturity.

3.3.3 Impact of trading mechanisms

Some markets (such as NYSE and NASDAQ) now operate a ‘hybrid’ system, which combines both the characteristics of quote and order driven markets. While such markets are predominantly order driven now, specialist market makers also feature, and are required to offer liquidity, if no one else will do so [Harris, 2002]. In return for assuming the obligation to offer liquidity, these specialists may be granted certain advantages, such as the ability to sell ‘short’ an asset without borrowing it (i.e. ‘naked’ short selling).

With regard to liquidity, there is evidence [Jain, 2003] that liquidity in hybrid systems is higher that in pure LOBs (order driven markets) or dealer systems (quote driven markets), because there are two sources of liquidity. The author reaches an even stronger conclusion that 'institutional features of an exchange are the major determinants of liquidity in its listed stocks', where these features include tick size choices, the presence of a designated market maker, automatic trade execution system, and order-flow centralization. In analysing the options markets, Mayhew [2002] found that the trading of assets on different exchanges results in a decrease in quoted spreads.

3.3.4 Impact of regulatory and exchange decisions on liquidity

With regard to individual exchange enhancements, Riordan and Storkenmaier [2012] document the effects of the reduction of system latency (or the amount of time required for a trader to submit an order and receive confirmation) on liquidity. Reduction in latency generally results from the offering of co-location services by the exchange, or improved market infrastructure. For the Deutsche Boerse, they found that technological upgrades, which were responsible for reducing latency from 50ms to 10ms, led to a decrease in quoted and effective spreads, although this was predominantly in small- and medium-sized stocks.

In recent years, high-frequency trading has begun to dominate trading activity across asset classes. HFTs were part of approximately 73% of equity market transactions, according to Hendershott et al. [2011]. The effect on the quality of equity markets is still unclear: Hendershott et al. [2011] find that HFT causally improve liquidity, while Zhang [2010] found that HFT activity increases volatility and also has a negative impact on the price discovery process. A comprehensive review of the effects of HFT on liquidity, and market quality more generally, can be found in Chakrabarty et al. [2013].

Both individual market regulators and supra-national organisations have been engaging in discussions in order to curtail high-frequency trading, producing recommendations and sets of guidelines. For the European, US and Australian markets, Chesini and Giaretta [2014] summarise the proposed regulation. Measures include imposing market making obligations (i.e. reguiring HFTs to make two-sided markets throughout the day), a minimum period in force for limit orders, in order to reduce excessive activity through limit order submissions and immediate cancellations, and order-to-trade ratios. The efficacy of the latter, which aims to improve the percentage of limit orders that are executed, will be evaluated in a simulation setting in Section 6.

One of the first developments that arguably had an impact on the activity of HFTs (par-

ticularly small and medium-sized HFTs) was the unfiltered (or naked) access ban by the SEC in 2010[10]. Chakrabarty et al. [2013] showed that the ban resulted in a drop in both quoting and trading activity, but liquidity actually improved. However, they also note that this was accompanied by a negative effect on short-term price discovery.

Another regulation that was considered for European markets was the Financial Transaction Tax (FTT). France[11] and Italy [Colliard and Hoffmann, 2013] enacted FTT regulation in 2012 and 2013 respectively. In the French case, this includes both a 0.2% tax on purchasing securities, as well as a tax on orders cancelled in the context of high-frequency trading. As a result of the regulation, France lost a significant share of European equity turnover, 23 percent in 2011 to an estimated 12.85 percent in 2013. In Italy, equity turnover of €101 bn in 2012 halved to €50 bn in 2013[12]. Therefore, as a result of acting before a broader consensus about HFT activity in the EU was achieved, liquidity in France and Italy was reduced significantly.

3.4 Financial market simulation models

We have seen that liquidity is an important consideration for market participants, and in the last part of this thesis, we will focus on explicitly modelling the connection between liquidity and aggregate market activity. This will take the form of an agent-based model of the LOB, in which agents act according to their liquidity motivations. In this section, we provide a review of the related agent-based modelling literature.

Research regarding the simulation of financial markets using agents can be traced back 30 years ago to the work of Cohen et al. [1983], who proposed a model for a stock exchange. They evaluated the impact of various stabilising policies on price, volatility and liquidity. They also introduced the concept of heterogeneous trading agents and an architecture for the limit order book, ideas which have been replicated in many forms since. More recently, a variety of approaches have been suggested, each drawing from a wide and varied literature, including Finance, Economics, Mathematics, Statistics and Physics. As a result, the area of financial simulation modelling has now grown to a size where it is impossible to consider every single

[10]The SEC 'ban' was essentially a rule to 'require brokers and dealers to have risk controls in place before providing their customers with access to the market', see for details `http://www.sec.gov/news/press/2010/2010-210.htm`.

[11]The relevant regulation for the French FTT case can be found in Article 235 ter ZD, relating to the French tax code, at `http://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006069577`.

[12]`http://www.thetradenews.com/news/Asset_Classes/Equities/Liquidity_dries_up_in_FTT_countries_-_report.aspx`

model. We can only hope to summarise some of the key contributions over the past three decades and provide a taxonomy of some of the more important models.

3.4.1 Agent design

The variety of approaches to financial market ABM is apparent from the design of agent behaviours. Descriptions of what constitutes typical agent activity in such models vary greatly, from simple, mechanistic agents operating under very simple rules through to adaptive agents, who change their strategy depending on market conditions and the recent profitability of each behaviour. As in many types of simulation models, there is a tension between model parsimony and expressiveness, or the ability to capture more complex dynamics. This is reflected in the two most frequently studied agent designs, i.e. variants of 'Zero-Intelligence' agents and chartist and fundamentalist agents. We outline the typical behaviour of both types of agents in the following.

3.4.1.1 Zero-intelligence traders

The simplest ABMs consist of a single, unsophisticated type of agent who essentially submits orders randomly, possibly subject to (very few) constraints, like budget considerations. An early example by Gode and Sunder [1993] showed that it was possible to achieve high allocative efficiency in a market (defined as the total profit actually earned by all the traders divided by the maximum total profit that could have been earned by all the traders, and used as a measure of the performance of the economy), using agents that submitted bids and offers with prices uniformly distributed within an interval. They argued that a minimum amount of discipline imposed by the market is sufficient for this, and that learning or intelligence are not necessary.

The results of Gode and Sunder [1993] are critiqued by Cliff et al. [1997], who suggest that the tendency of a market that consists of such zero intelligence (ZI) traders towards an equilibrium price is a result of the design of the market itself, rather than the traders. They extend the minimal ZI definition with a simple adaptive mechanism, whereby traders can adjust their profit margin according to the current price of the asset and whether they have traded their allocation for the period. They show that the results obtained using these enhanced zero-intelligence plus (ZIP) traders are closer to those observed in markets consisting of human experimental traders. Variants of ZI and ZIP strategies have subsequently been employed, either on their own or as part of a more diverse trading population, in the models of Tesauro and Bredin [2002], Niu et al. [2008] and, specifically in the financial ABM literature, by LiCalzi and Pellizzari [2003].

3.4.1.2 Chartist and fundamentalist traders

By far the most frequently studied variant of ABM consists of two types of agents, usually denoted 'chartists' and 'fundamentalists'. The earliest mention of the two terms is by Zeeman [1974], referring to conversations with Sharon Hintze, while an early model by Beja and Goldman [1980] also uses similar terminology to differentiate between trend followers and value investors. The strategies of such agents are generalisations of real trading strategies employed in a number of different markets.

Taylor and Allen [1992], surveying a number of London-based dealers, refer to fundamentalist traders as deriving their views from an economic analysis of the traded asset. Being a chartist dealer, on the other hand, involved 'providing forecasts or trading advice on the basis of largely visual inspection of past prices, without regard to any underlying economic or fundamental analysis'. In the context of an ABM, a fundamentalist trading strategy would result in a target price, and the agent would buy(sell) if the asset was undervalued(overvalued). Common chartist behaviours, on the other hand, include making decisions based on the price of the asset compared to its moving average in a particular period, or assuming that a short move in a certain direction will continue in the near future (a momentum strategy).

A simple formulation of the one-period forecasting functions of a fundamentalist ($\mathbb{E}_{f,t}\left[P_{t+1}|P_t^f, P_t\right]$) and chartist ($\mathbb{E}_{c,t}\left[P_{t+1}|P_t, P_{t-1}\right]$) trader at time t is given by Chen et al. [2012]:

$$\mathbb{E}_{f,t}\left[P_{t+1}|P_t^f, P_t\right] = P_t + a_f\left(P_t^f - P_t\right), 0 \leq a_f \leq 1 \tag{3.6}$$

$$\mathbb{E}_{c,t}\left[P_{t+1}|P_t, P_{t-1}\right] = P_t + a_c\left(P_t - P_{t-1}\right), 0 \leq a_c \leq 1 \tag{3.7}$$

The fundamentalist's view is that the fair value of the asset is P_t^f, and the price of will move in that direction with 'speed' a_f. A chartist's expectation simply extrapolates the last trend by a_c in the same direction. The stabilising and destabilising forces associated with the fundamentalist and chartist respectively are then clear in this formulation. If the coefficient a_c was instead negative, this would correspond to a contrarian, rather than a trend following strategy, which induces different dynamics in the asset price.

Turning now to the evolution of the fundamentalist's view of fair value of the asset, there have been a number of approaches in different models. For example, Farmer and Joshi [2002] and Westerhoff and Reitz [2003] use a simple random walk:

$$P_{t+1}^f = P_t^f + \eta_t$$

with $\{\eta_t\}$ a sequence of i.i.d. normal innovations. The value could also be obtained from the

application of an economic valuation model, such as (in the case of a company's stock price) estimating future dividends and discounting them to obtain the present value.

3.4.2 Market structure

We should mention briefly the main financial market structures available, which dictate the form of interaction between buyers and sellers. This should be considered carefully because different modelling approaches are appropriate for different market structures. Markets for financial securities generally fall under one of three types, namely quote driven, order driven and hybrid markets.

Quote driven (dealer) markets use specialists (dealers) to provide 2-way prices, i.e. prices at which one is able to buy or sell. These markets therefore do not display the trading interest of other participants besides the dealers and generally do not consolidate the quotes of these dealers. Traders in quote driven markets can only trade directly with dealers, who are not obliged to quote continuously in most cases. Quote driven markets are also known as price driven markets, and NASDAQ had started out using this exchange mechanism. This mechanism is now usually used in less liquid markets.

Order driven (Limit Order Book) markets display all limit orders from (potential) buyers and sellers. They are more transparent than quote driven markets, as market participants generally have access to more information about the level of buying and selling demand. In contrast with participants in quote driven markets, participants in order driven markets can trade directly with each other through the central matching mechanism. We have reviewed the operation of the Limit Order Book in Chapter 2 and will investigate continuous LOBs here (i.e. where the matching is continuous throughout the day), but there are also discrete equivalents of this structure where auctions are held periodically throughout the day.

Hybrid systems combine aspects of both mechanisms above, operating as LOBs, but also employing designated market makers, or dealers, for the less liquid stocks. The majority of exchanges operate either as a pure LOB, or as a hybrid system which features an LOB. Indicatively, the Helsinki, Hong Kong, Shenzhen, Swiss, Tokyo, Toronto, and Vancouver Stock Exchanges, together with Euronext and the Australian Securities Exchange operate as pure LOBs, while the New York Stock Exchange, NASDAQ, and the London Stock Exchange operate a hybrid LOB system [Gould et al., 2013]. The dataset considered here is from Chi-X, a secondary exchange that operates as a pure LOB also, but we will review models pertaining to both order and quote driven market structures in this section.

3.4.2.1 Quote driven market models

The models discussed here, which some authors refer to as Heterogeneous Agent Models (HAMs), are analytical in nature and simplified necessarily, in order to ensure tractability. One of the earliest models by Zeeman [1974] distinguished between two different types of market participants, chartists and fundamentalists. It does not model individual agent behaviour, but rather the result of the aggregate behaviours of the two heterogeneous groups, and thus can be considered as a predecessor to the more detailed models that were proposed later. It proposes a dynamical system linking the proportion of chartist and fundamentalists in the market to the rate of change of the level of the Dow-Jones index. Borrowing ideas from catastrophe theory, Zeeman envisioned a cusp catastrophe model, where fundamentalist and chartist proportions become controls for the state variable corresponding to the rate of change of the index.

As the various models that have been proposed differ greatly in their setup, we will focus mainly on the price determination mechanisms in the following discussion. Day and Huang [1990] include a market participant that is closer to the specialist(dealer) on the New York Stock Exchange. In this model, the dealer announces the market price of the asset at various times during the day and executes the orders of the other two types of investors (which have similar characteristics to the chartist and fundamentalist traders described above) at that price. Changes in the price of the asset result from the dealer reacting to the imbalance between demand and supply, as excess demand (or supply) results in a reduction (or increase) in the market maker's inventory.

The dealer is assumed to follow the simple price adjustment process

$$P_{t+1} = P_t + cD(P_t) \tag{3.8}$$

where $D(P_t) := \sum_i D_i(P_t)$ is the demand (or supply, if it's negative) for the asset at the announced price P_t. This is a sum of the demands of all traders, generically denoted by $D_i(P_t)$ for the i-th trader, and a concrete example for this demand function is given in Equation 3.9. Similar demand-driven price adjustment functions are found in the models of Lux [1995], Farmer and Joshi [2002] and Westerhoff and Reitz [2003], amongst others.

Some models make the assumption that the price will move such that the demand for the asset will equal the supply. This may be because it is set externally (for example, by a dealer), or because it converges to a market clearing price because of the actions of market participants. For example, Arthur et al. [1996] assume, in contrast to the model of Day and Huang [1990], that agents do not explicitly determine their own demand, but rather pass their demand parameters to their dealer, who determines a price such that total demand $\sum_i D_i(P_t)$ equals the number of

shares issued. For the i-th agent the demand is:

$$D_i(P_t) \propto \mathbb{E}_{i,t}[P_{t+1} + DIV_{t+1}] - P_t(1+r) \tag{3.9}$$

where r is the return on a risk-free asset, and the dividend process $\{DIV_t\}$ is assumed to follow an AR(1) process.

The price clearing mechanism used by Raberto et al. [2001] and Mannaro et al. [2008], on the other hand, is based on the intersection of the demand-supply curve. If the tuple $(v_i^b, p_i^b), i = 1 \dots B$ indicates the price and volume of the i-th buy order during a single timestep t of the model, while $(v_j^a, p_j^a), j = 1 \dots A$ indicates the price and volume of the j-th sell order, the demand curve f and supply curve g at time $t+1$ will be

$$f_{t+1}(p) = \sum_{i|p_i^b \geq p} v_i^b,$$
$$g_{t+1}(p) = \sum_{j|p_j^a \leq p} v_j^a.$$

The former indicates the total volume demanded at a price greater or equal to p, while the latter indicates the total volume supplied at a price less than or equal to p.

In a more complex example of a quote-driven agent-based model, Brock and Hommes [1998] develop a present discounted value asset pricing model with evolutionary dynamics, which they term Adaptive Belief Systems (ABS). They consider a model in which there are two available assets, a risky asset (a stock) and a risk-free asset, of which the latter is assumed to be available in infinite supply (and thus the traders' demand for it will not affect its price). The wealth equation for the i-th trader is

$$W_{i,t+1} = (1+r)W_{i,t} + (P_{t+1} + DIV_{t+1} - (1+r)P_t)z_{i,t}$$

where $z_{i,t}$ are the shares purchased at time t. The model assumes heterogeneous traders with common risk aversion a and beliefs about the return variance σ^2, and thus the variance of wealth for a trader $\mathbb{V}_{i,t}\left[W_{i,t+1}\right]$ will be $z_{i,t}^2$ times this quantity. Each one-period mean variance-optimising investor of type i solves

$$\arg\max_z \left(\mathbb{E}_{i,t}\left[W_{i,t+1}\right] - \frac{a}{2}\mathbb{V}_{i,t}\left[W_{i,t+1}\right]\right)$$

in order to determine their number of shares $z_{i,t}$, which gives

$$z_{i,t} = \frac{\mathbb{E}_{i,t}\left[P_{t+1} + DIV_{t+1} - (1+r)P_t\right]}{a\sigma^2}$$

and from this one can obtain the equilibrium pricing equation, according to the number of investor types, see details in Brock and Hommes [1998].

3.4.2.2 Order-driven market models

Order-driven market models usually model the trading interactions in the LOB, and allow for a more realistic description of the intra-day trading process. These models simulate the behaviour of individual market participants, usually based on the behaviour of various classes of real traders. The price of the traded financial asset is then determined from the limit and market orders submitted by these traders. Depending on the model, the instantaneous price is either considered to be the mid-point between the highest bid price and lowest ask price, or the last traded price.

Cohen et al. [1983] were the first to simulate the limit order book in this way, using a FORTRAN computer program. They implemented the price/time priority rules present in many exchanges, as well as the ability of market orders to 'walk up the book', i.e. execute against multiple limit orders. Later models, like those of Maslov [2000], Chiarella and Iori [2002] and LiCalzi and Pellizzari [2003] used the same market mechanism in their models with varying degrees of abstraction regarding trader behaviour and the lifetime of orders. For example, Maslov [2000] found that even a model with a minimal set of rules governing agent behaviour is sufficient to recreate realistic LOB features, such as fat tails in the distribution of returns. The agents in this model are again not considered explicitly, but instead a single trader is assumed to visit the market at each time step, placing a buy or sell order with equal probability. Since the model does not consider order inter-arrival times or the possibility of cancelled orders, it cannot, however, give rise to realistic simulations of intra-day LOB activity.

The model proposed by Chiarella and Iori [2002] assumes that agent behaviours have varying degrees of the chartist and fundamentalist influences introduced in Equations 3.6 and 3.7, which manifest in different expectations of short-term market returns $R_{t,t+1} = \frac{P_{t+1}-P_t}{P_t}$

$$\mathbb{E}_{i,t}\left[R_{t,t+1}|P_t^f, P_{t-L_i:t}\right] = g_1^i \frac{P_t^f - P_t}{P_t} + g_2^i \frac{1}{L_i}\sum_{j=1}^{L_i} \frac{P_{t-j} - P_{t-j-1}}{P_{t-j-1}} \tag{3.10}$$

where $g_1^i \sim N(0, \sigma_1)\mathbb{I}_{g_1^i>0}$ (where $\mathbb{I}_{g_1^i>0}$ denotes the indicator function) and $g_2^i \sim N(0, \sigma_2)$. The model does not provide a systematic study of the parameters, but rather an ad-hoc evaluation of the effect of average order lifetime on LOB liquidity and volatility. The model was further extended by Chiarella et al. [2009] and included different time horizon considerations for the different types of agent predictors. In deciding the amount of the asset they would like to hold, agents here are assumed to maximise a constant absolute risk aversion utility function

$$U(W_{i,t}, a_i) = 1 - e^{-a_i W_{i,t}} \tag{3.11}$$

where a_i is a risk aversion coefficient for the i-th agent, and $W_{i,t}$ is his/her wealth at time t,

as before. The possibility of different risk aversion coefficients shows precisely the ability of ABMs to model heterogeneity in a market, compared to representative agent methods found in the macroeconomics literature.

Several extensions have been proposed in recent years, in order to model specific features of the LOB that had been abstracted away in earlier models. For example, in order to capture the dependence in event activity, Toke [2011] uses a class of point processes, known as the Hawkes processes, in a simple agent-based model. Calculating the empirical distribution of times between successive orders, he finds that limit orders that follow market orders are more common. This can be interpreted as a sign of the market reacting, and replenishing liquidity, but not necessarily on the same side as the market order. A zero intelligence model is proposed, where agents submit limit orders and market orders according to Poisson processes, and cancellation times also follow a Poisson process. However, the introduction of dependence between event activity in a Hawkes processes model is shown to perform better than homogeneous Poisson processes, in producing a realistic shape for the distribution of the inside spread.

While the majority of ABMs model the trading of a single risky asset (like a stock, possibly along with a 'risk-free' asset), Consiglio et al. [2005] introduce a multiple asset framework. They impose a minimal set of rules governing agents' behaviour, namely budget constraints and a (exogenously assigned) target allocation of their wealth across the different assets. There are also considerations regarding the placement of their orders over time. They show that even with agent homogeneity, they are able to capture some of the statistical properties and temporal patterns of real markets. We will review some of these statistical properties in the following section.

3.4.3 Model aims

3.4.3.1 Replication of persistent features of financial markets

Many financial market ABMs aim to demonstrate the proximity of the model to real financial markets by reproducing certain statistical properties commonly found in financial data. These are most commonly related to the time series of an asset's returns, but can also include systematic observations about traded volume, volatility and liquidity. There is a large body of work describing stylised facts, and the main properties have been surveyed by Cont [2001] and Russell et al. [2010], for different data frequencies. Chen et al. [2012] identifies 30 such stylised facts, some of which are only present in either high or low frequency data. The main categories of these stylised facts pertain to properties of returns (absence of autocorrelations, fat tails in the returns distribution), trading duration (clustering of durations), transaction size (power law

in trade sizes) and the bid-ask spread (correlation with price moves). Of these, 12 have been replicated in ABMs, and in the 50 models surveyed the most common stylised facts are (in order):

- Fat tails in the distribution of price returns (R_t). More formally, the returns have a power-law tail with

$$F(|R_t| > x) \sim x^{-\alpha},$$

  and an exponent of 2-4 (see, e.g. Lux and Sornette [2002] and references within).

- Volatility clustering, or persistent volatility. This is the phenomenon where 'large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes' Mandelbrot [1997]. Formally

$$\text{corr}(|R_t|, |R_{t+\tau}|) > 0,$$

  for a range of different values of τ, see e.g. Cont [2007].

- Absence of autocorrelations in returns in general. In short time intervals (less than 20 minutes) such autocorrelations may be present, however, due to microstructure effects.

Cont [2001] suggests that 'albeit qualitative, these stylized facts are so constraining that it is not easy to exhibit even an (ad hoc) stochastic process which possesses the same set of properties and one has to go to great lengths to reproduce them with a model'. ABMs therefore try to reproduce a selection of these and demonstrate empirical validity of the model by relating the presence of these features to the parameters of the model. For example, Arthur et al. [1996] explained the appearance of price bubbles and crashes, fatter tails in the returns distribution and a larger trading volume to a single parameter, the frequency of exploration of the strategy space by the agents. Regarding the appearance of characteristics of financial prices that resemble particular scaling laws, Lux and Marchesi [1999] note that explaining this as an emergent property of the interaction of a number of agents would run contrary to the 'efficient market hypothesis'. However, observing this in their ABM, they are able explain it through the dynamic (switching) aspect of the agents' opinions.

A thorough study of the origin of the various stylised facts in financial market ABMs can be found in Chen et al. [2012]. The paper lists both the general category under which each model falls, as well as the particular parameters identified as explaining the presence of one or more stylised facts. For example, the family of ABMs deriving from the ant recruitment process

developed by Kirman [1993][13] is generally able to explain the absence of autocorrelations and fat tails in the price returns, as well as volatility clustering, through the parameter controlling the tendency of agents to switch strategy.

3.4.3.2 Policy testing

The nature of agent-based models means that they could be useful in 'providing a testbed for the study of policy questions targeting the medium and long run by basing the analysis on economic mechanisms that may unfold as a response to a policy intervention' [Dawid and Neugart, 2011]. There are a multitude of works in this area also, and the general approach is to:

1. Build an ABM model using some agents representative of the real trading population.

2. Calibrate the model and show that it can produce some of the stylised facts of financial markets, so that one can be confident that the simulated time series of returns/volume/volatility will have similar dynamics to real financial data.

3. Introduce the proposed regulatory mechanism, or exchange rule, and quantify the impact with respect to certain variables of interest - this usually includes measures volatility or liquidity.

An early application of ABM for policy testing was developed by Darley and Outkin [2007], where they tried to predict the impact of decimalisation (moving from trading 1/8ths and 1/16ths of a dollar to cents) on the NASDAQ. A number of different methods were used to model individual investor strategies, including complex reinforcement learning methods. The model predicted an increase in the inside spread as a result of the intervention, a prediction which subsequently came to pass.

Another proposed policy intervention that has received the ABM treatment in the recent past is that of a financial transaction tax. This is also termed a 'Tobin tax', after the economist that had originally suggested it as a way of curbing fluctuations in foreign exchange rates [Tobin, 1978]. The imposition of a small transaction tax in chartist and fundamentalist model by Westerhoff [2003] results in a decrease in volatility, while larger taxes were found to increase it. Part of the reason is because agents observe the effect of the transaction tax on the return of a strategy, and since large taxes (naturally) harm profitability, they choose not to trade. Similar results were obtained by Ehrenstein et al. [2005].

[13]The model by Kirman [1993] is a precursor to the chartist-fundamentalist agent design, and is presented in detail in Section 3.4.4.1. Papers deriving from this include that by Alfarano et al. [2005] and Gilli and Winker [2003]).

In both these models, however, agents are assumed to disregard the effect of the transaction tax prior to trading, and only respond (by changing their trading behaviour) to experiencing a change in their profitability. Mannaro et al. [2008] take a different approach, incorporating the tax into the traders' expectations. In contrast with the models above, they find that traded volumes decrease and price volatility increases as one increases the transaction tax.

In a novel extension, they also assumed the existence of two different markets, only one of which imposes a transaction tax. The traders then evaluate an attractor function to select which of the two markets they would like to operate in during the next period. Again, they find that the taxed market exhibits much greater volatility than the untaxed one.

Pellizzari and Westerhoff [2009] find that the impact of transaction taxes depends on the market system used in the model and therefore the predictions of the models above could all be correct, depending on the setting. They investigate the impact of transaction taxes for different market systems, namely a continuous double auction and a central dealership (examples of quote and order-driven market systems, respectively). They find that in an order-driven market, a transaction tax decreases trading but does not stabilise the market, as it simultaneously decreases liquidity, therefore every execution has a bigger effect. In a quote-driven market, a transaction tax does have the power to stabilise the market, but this is based on the assumption that the dealer has infinite (or at least abundant) liquidity.

As there have been various types of financial taxes suggested recently, Lengnick and Wohltmann [2013] use a hybrid of an economic model and a financial ABM to study the impact of both a Financial Activities Tax and a Financial Transaction Tax (FTT). With the dual objective of optimising revenues and stabilising the economy, their model suggests that the latter would be the better option, subject to certain constraints about the level of the tax.

Besides helping estimate the impact of transaction taxes, Westerhoff [2008] suggested that ABMs could be utilised to carry out experiments of other proposed regulatory mechanisms. This work uses modifications of a single model to investigate the effect of central bank interventions, in trying to control certain currency fluctuations, and trading halts, in trying to curb excess volatility. The simplicity of the model, however, limits the generalisation of the results. We present an example of an exchange intervention in a more realistic LOB setting in Section 6.4.

3.4.4 Methods for estimating ABM parameters

Early agent-based models were based on simple rules of interaction, and tried to explain macro behaviours through these rules. The models were not, however, generally estimated using rigor-

ous statistical methods, nor was there a systematic study of the effect of parameter changes on the results. This limited their use as explanatory tools [Janssen and Ostrom, 2006]. The increase in sophistication of these models in recent years, along with the availability of relevant data, has called for more rigorous approaches for estimation and calibration, and we review the most important methods in this section, along with examples of models which employ these methods. We note that the ABM presented in this thesis (in Chapter 6) is calibrated via simulation-based estimation, but we provide here a brief discussion and examples of direct estimation approaches also.

3.4.4.1 Direct estimation

Direct estimation is used in simpler models, when it is possible to derive an analytical expression containing all model parameters. Then, in the case of a regression model, for example, the model could be estimated by ordinary least squares, or by maximum likelihood estimation (MLE)

$$\mathcal{L}(\hat{\boldsymbol{\theta}}|X_{1:n}) = \sup_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|X_{1:n}) \tag{3.12}$$

if a likelihood function

$$\mathcal{L}(\boldsymbol{\theta}|X_{1:n}) = \prod_{i=1}^{n} f(X_i;\boldsymbol{\theta}) \tag{3.13}$$

can be derived. Alternative direct estimation methods include moment matching, quantile matching and loss function minimisation, although these are not considered here.

As examples of models estimated directly, it would be worthwhile to discuss the family of models deriving from the ant recruitment process developed by Kirman [1993], which has been very influential in the ABM literature. The model tried to explain the puzzling behaviour of ants in selecting either of two food sources, where there was invariably an 80/20 split between the two. Kirman [1993] proposed that the state of the system (the number k of a total of N ants at the first food source) would evolve according to the following simple Markov Chain:

$$k \to \begin{cases} k+1, \text{with } p_1 = (1-\frac{k}{N})(\epsilon + (1-\delta)\frac{k}{N-1}) \\ k-1, \text{with } p_2 = \frac{k}{N}(\epsilon + (1-\delta)\frac{N-k}{N-1}) \end{cases} \tag{3.14}$$

where ϵ is the probability of an independent change in behaviour, while $1-\delta$ is the contagion effect. For different values of these parameters, the model was able to reproduce either fluctuations around an equal split between the two sources, or 'herding', at one or the other.

In relating the recruitment effect of ants to financial markets, Kirman [1993] drew a parallel with trading agents that mimic other successful traders. In particular, he imagined the

trading population switching between a set of predefined strategies, according to the success of each group. The contagion effects of the Kirman ant model were formalised in a financial markets setting in a number of models, including by Lux [1995, 1997], Lux and Marchesi [1999], Alfarano et al. [2005, 2008], and Alfarano and Milakovic [2009].

In particular, Alfarano et al. [2005] employed this herding mechanism in a model with fundamentalist traders (similar to the type described previously in this chapter) and noise traders. Through the transition probabilities between the two groups, which are similar to those in Equation 3.14, and using the reversibility of a Markov chain and the principle of conservation of probability, they obtain the so-called Master equation for $p_k(t)$, which denotes the probability of k agents (of a total trading population N) being of the first of the two types:

$$\frac{\Delta p_k(t)}{\Delta t} = \Sigma_{k'}(p_{k'}(t)\pi(k' \rightarrow k)) - (p_k(t)\pi(k \rightarrow k')). \tag{3.15}$$

This leads to a model for the proportion $\frac{n}{N}$ of traders of the first type that can be modelled akin to a stochastic volatility model, where the parameters are interpreted in terms of the propensity of each type of trader to switch. The simple structure of the model enables the derivation of the unconditional distribution of returns, which is fit to commodities and stock market return data, in order to estimate the behavioural parameters.

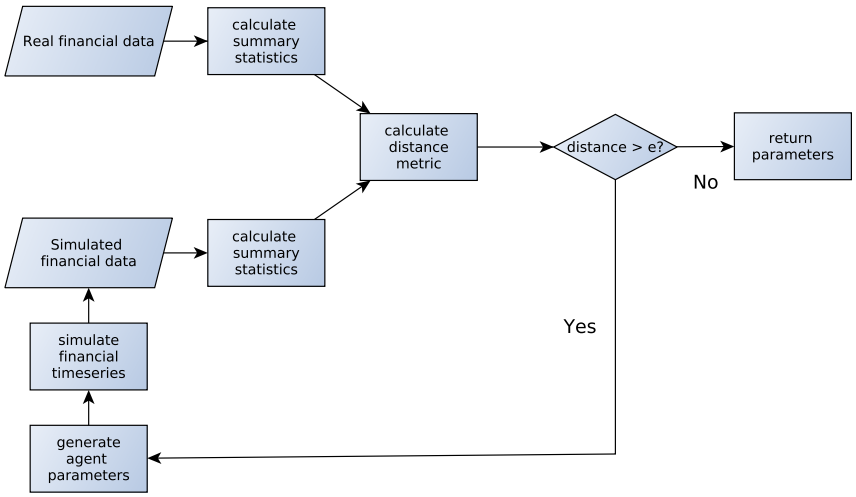
3.4.4.2 Indirect/simulation-based estimation

Indirect estimation of ABM parameters is used when an analytical expression for the theoretical model (for example, for the likelihood) is difficult or impossible to write down in closed form. However, in many cases the model can be specified in a generative fashion. That is, the proposed model is assumed to be able to generate/simulate a dataset comparable in form to the original dataset, given a certain set of parameters. The model is then not fit to the real data directly, but rather the optimisation aims to find the parameter vector that minimises some distance measure between a summary statistic of the real and simulated data, as illustrated in Figure 3.4. While the real data-generating process may be different, it is still desirable for the output of the model to resemble the observed data.

If we consider k summary statistics then two k-dimensional vectors $\mathbf{f}, \mathbf{g}$ can represent the statistics calculated using the real and simulated data respectively. There are a number of possible metrics that could be used to quantify the difference between them:

1. The L_p-norm distance, or Minkowski distance of order p, defined as:

$$\left(\sum_{i=1}^{k} |f_i - g_i|^p \right)^{\frac{1}{p}}$$



.

Figure 3.4: A flowchart of the indirect estimation procedure. The estimated parameters are returned once the distance between summary statistics computed on the real and simulated data is less than a tolerance e

where the most common distances are the L_1-norm, L_2-norm or L_∞-norm:

$$\sum_{i=1}^{k} |f_i - g_i|,$$

$$\left(\sum_{i=1}^{k} |f_i - g_i|^2 \right)^{\frac{1}{2}},$$

$$\lim_{p \to \infty} \left(\sum_{i=1}^{k} |f_i - g_i|^p \right)^{\frac{1}{p}} = \max(|f_1 - g_1|, |f_2 - g_2|, \ldots, |f_k - g_k|).$$

2. The Canberra distance, which is a weighted version of the L_1-norm:

$$\sum_{i=1}^{k} \frac{|f_i - g_i|}{|f_i| + |g_i|}.$$

It is obvious then, that this can only range between 0 and 1.

3. If the statistics are not of a comparable size, then a different weighting may be appropriate, leading to the Mahalanobis distance:

$$\frac{1}{I}(\boldsymbol{f} - \boldsymbol{g})' \boldsymbol{W} (\boldsymbol{f} - \boldsymbol{g}),$$

where $\boldsymbol{W}$ is a weighting matrix that takes into account the relative sizes. In addition, it may be chosen so that it has particular properties (for example, minimising the asymptotic covariance of the estimator).

As an example of a model estimated using a simulation-based approach, we revisit the Gilli and Winker [2003] version of the Kirman ant model, where the returns expectations of the two types of traders are as in Equation 3.6 and 3.7. The price determination process combines the two expectations and the perceived share of fundamentalist and chartist traders in the market:

$$\mathbb{E}_{m,t}\left[\Delta P_{t+1}|P_t^f, P_{t-1}^{mid}, P_t^{mid}\right] = \omega_t v(P_t^f - P_t^{mid}) + (1-\omega_t)(P_t^{mid} - P_{t-1}^{mid}).$$

where the perceived weight (since agents have a noisy signal) of fundamentalists, ω_t, is calculated as the probability $\omega_t = \mathbb{P}r(\widetilde{q}_t > \frac{1}{2})$, q_t is the actual proportion of fundamentalists and $\widetilde{q}_t \sim N(q_t, \sigma_q^2)$ and v is a correction term. The weight of chartists is then $1-\omega_t$. In every iteration of the simulation, the number of chartists and fundamentalists evolves according to Equation 3.14 and the price is computed from the market's expectation above, with the addition of some exogenous noise term.

The estimation proceeds by calculating the empirical kurtosis of the real and simulated price returns ($k^{emp}, \widetilde{k}^{abm}$ respectively) as a measure of the fat-tailedness, as well as the coefficient α of the ARCH(1) model

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2, \tag{3.16}$$

fit to the real and simulated log returns α^{emp} and $\widetilde{\alpha}^{abm}$ respectively. Then for a model parameter vector $\boldsymbol{\theta}$, the stochastic approximation to the objective function measures the distance between these summary statistics

$$\widetilde{f}(\theta) = |\widetilde{k}^{abm} - k^{emp}| + \lambda|\widetilde{\alpha}^{abm} - \alpha^{emp}|. \tag{3.17}$$

The optimisation algorithm used to minimise this distance is the Nelder-Mead simplex algorithm with a threshold accepting heuristic. A weakness with such an approach, however, is the ad hoc selection of the weighting factor λ based on the magnitudes of the two components. In Chapter 6, we will introduce an optimisation method that considers the components separately, so that such a weighting factor is not necessary.

In a follow-up paper, Winker et al. [2007] proposed the use of the method of simulated moments (MSM) for the calibration of this model. The MSM was proposed by McFadden [1989] and Pakes and Pollard [1989]. In the simplest case, we compare a single summary statistic calculated on, e.g., the mid-price in the real timeseries y_t, to the same statistic calculated on the simulated timeseries $y_t^s(\boldsymbol{\theta})$. The MSM estimator is the parameter vector $\hat{\boldsymbol{\theta}}$ which minimises the distance between them. For example, for the first moment estimator, we have:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}[\frac{1}{T}\sum_{t=1}^{T} y_i - \frac{1}{T}\sum_{t=1}^{T} y_i^s(\boldsymbol{\theta})]^2$$

3.4.4.3 Indirect inference

Indirect inference is a generalisation of the simulated method of moments. This method is again used when an economic model is intractable, but can be used to generate simulated data. Indirect inference employs an additional, auxiliary model, that does not necessarily describe the agent interactions accurately. The process entails that at each estimation step, we have a candidate parameter vector for the economic model, from which we generate a set of data. The auxiliary model is then fit to both the real data and the data simulated from the main economic model. A measure of the difference between the parameter vectors is then obtained, then the parameters of the economic model are iteratively adjusted, until the difference is below some threshold. We will describe this estimation method in detail in Chapter 6, as it is central to the estimation of the agent-based model proposed in this thesis.

Chapter 4

Liquidity and resilience of the LOB

In Chapter 3, we saw that a wide variety of liquidity measures have been proposed, reflecting aspects such as immediacy, tightness and depth. While these aspects are important in determining the cost of immediate execution of a very large order, such orders are becoming increasingly less common, as brokerage houses and large fundamentals traders aim to reduce execution costs. On the NYSE, the average order size is one-eighth of that of fifteen years ago, in terms of number of shares, and one-third in dollar value [Chlistalla et al., 2011], which indicates the partitioning of large orders into multiple smaller orders and traders taking advantage of liquidity replenishment. There is therefore a 'resilience' aspect to liquidity, which has been largely overlooked in the literature.

In this chapter, we define a new notion for liquidity resilience, captured through the concept of *threshold exceedance durations* (TEDs), that is, the durations of intra-day liquidity shocks. We exhibit the diurnal patterns of TED observations throughout the trading day and the effect of major economic announcements on the frequency of such events. We explain and forecast the level of the TED through a number of covariates summarising the state of the LOB, and this is achieved through the development of a survival time model, similar in structure to that used by Lo et al. [2002]. We show how the TED can accommodate different liquidity thresholds of interest, such that shocks can be defined as deviations from competitive levels of liquidity, or from more extreme levels. In our empirical analysis we show that the explanatory power of our model is slightly higher when considering the spread as the liquidity measure of choice, rather than the XLM (the Xetra Liquidity Measure, defined in Section 3.2.2), but both produce satisfactory results.

Our notion of liquidity resilience therefore extends the standard notion of resilience in two key dimensions, first by explicitly relating it to certain liquidity levels of interest, and secondly, by modelling and forecasting the time required for a market to recover following a liquidity

shock. We show that the state of the LOB is an important determinant of the level of liquidity resilience, while previous work by Kyle [1985] and Foucault et al. [2005] only considered exogenous factors to be relevant (informational asymmetries and waiting costs, respectively).

4.1 Modelling intra-day liquidity resilience

To start with, we should distinguish between the notion of liquidity, as a quality of a market that allows for relatively large orders to be placed without much effect on the market's dynamics, and resilience, as a quality of the market that allows for the dynamics to recover after a large order in a short period of time. Dong et al. [2007] note that resilience has received much less attention compared to other aspects of liquidity, citing the extensive research in depth and tightness. This may be because several frequently employed measures of depth and tightness can be readily computed from static views, or 'snapshots' of the LOB, as we have seen in our description of such measures in Section 3.2.2, whereas resilience is related to the change in the state of the LOB over time, and thus more difficult to capture. However, the embedding of simple notions of resilience into recent theoretical models of optimal execution [Alfonsi et al., 2010, Obizhaeva and Wang, 2013] means that resilience now needs to be considered in its own right, in order to pin down the concept and carry out empirical LOB analysis.

In the seminal paper of Kyle [1985], resilience is defined as "*the speed with which prices recover from a random, uninformative shock*". This is similar to the interpretation of Obizhaeva and Wang [2013], who suggest that in a resilient market is there is a swift convergence of the price of an asset to a new steady state, after a market order. Garbade and Garbade [1982] describes a resilient market as one in which "*new orders pour in promptly in response to a temporary order imbalance"*, whereas Harris [2002] suggests that in such a market, "*uninformed traders cannot change prices substantially"*. These interpretations of resilient markets differ somewhat, in that the former is related to order replenishment, while the latter concerns price evolution. Like liquidity itself, liquidity resilience does not have a universally accepted definition, but is only loosely understood as being related to the return to some former level of prices [Kyle, 1985], volumes [Garbade and Garbade, 1982] or a particular liquidity measure [Foucault et al., 2005].

One recent attempt to define and model resilience is given in the model of Foucault et al. [2005]. They analyse the determinants of liquidity resilience, which they define as the number of orders required for the spread to recover to a competitive level. They identify different liquidity resilience regimes for the LOB, i.e. regimes in which the spread returns to a competitive level after only a few orders (strong resilience) or after many orders (weak resilience). The

model relates these regimes to the proportion of patient and impatient traders (traders that predominantly submit limit or market orders, respectively). Specifically, they find that resilience increases with the proportion of patient traders, while resilience is reduced by a reduction in the tick size.

Several authors have also studied resilience empirically [Large, 2007, Gomber et al., 2011]. The resilience model of Large [2007] uses a parametric model which views limit orders, market orders and cancellations on either side of the LOB as a mutually-exciting ten-variate Hawkes point process. This formalises resilience in terms of the change in intensity of particular event types (e.g. limit order submissions on the bid side) after an instance of large trade is observed in the LOB. Gomber et al. [2011] define resilience as the change in the XLM in the period following large transactions. They find that the liquidity measure generally returns to close to its pre-trade level within 2-3 minutes of the large transaction.

While these previous attempts to capture liquidity resilience have exhibited the importance of the concept, they have generally tied the definitions to particular measures of liquidity, like the inside spread [Foucault et al., 2005], or the volume at the top of the LOB [Large, 2007]. The next section introduces a more general definition of resilience as the duration of deviations from a particular liquidity threshold level. Different liquidity measures and threshold levels of liquidity are appropriate for different applications (for example, an algorithmic execution setting would have different 'trigger' level to a setting where liquidity is monitored for systemic risk purposes). Our definition can accommodate any liquidity measure and threshold level of interest.

Compared to the papers above, which have developed theoretical models based on unobserved variables such as the relative proportions of patient and impatient traders in the LOB, our use of a regression model with covariates coming from the structure of the LOB enables us to assess the contribution of each covariate to the explanatory power of the model. It also gives us the ability to model different scenarios that had not previously occurred in the dataset, by modifying the covariate values. Finally, we can assess the predictive power of our model, and explain how such a forecasting model could then be readily incorporated into an execution model.

4.2 Defining liquidity resilience

The liquidity resilience notion introduced here is based on the idea of the *threshold exceedance duration*:

Definition 4.2.1. *The threshold exceedance duration (TED) is the length of time between the*

point at which a liquidity measure, $\mathcal{M}_t$, deviates from a threshold liquidity level, c, (in the direction of less liquidity), and the point at which it returns to at least that level again. The starting time and length of the TED are denoted by T_i and τ_i respectively, where i refers to the i^{th} exceedance. Formally:

$$\tau_i := \inf \{\tau \ : \ \mathcal{M}_{T_i+\tau} \leq c, \ \tau > 0\} \tag{4.1}$$

$$T_i := \inf \{t \ : \ \mathcal{M}_t > c, \ t > T_{i-1} + \tau_{i-1}\} \tag{4.2}$$

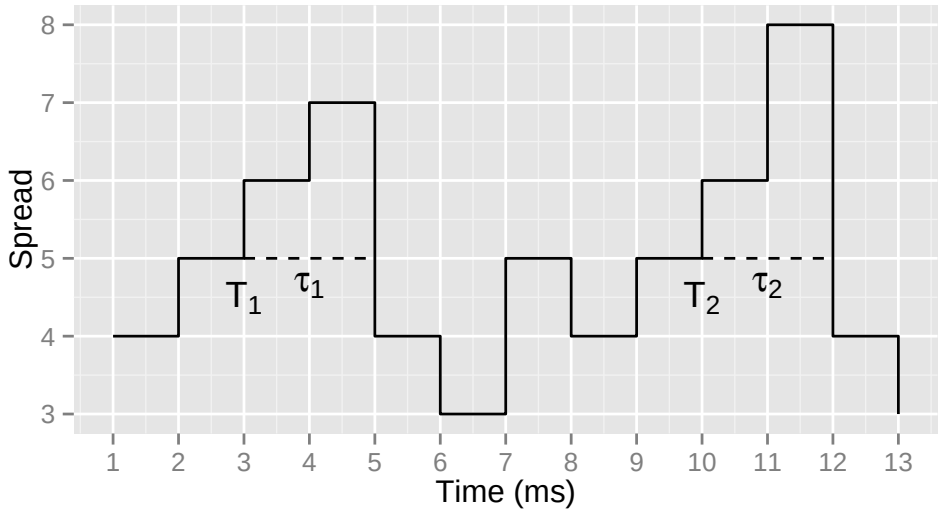


Figure 4.1: An example of the duration of exceedances over a liquidity threshold. Here, the spread is chosen as the liquidity measure, and the liquidity threshold c is chosen to be 5 cents.

To understand how such exceedance events are generated, in the case of the inside spread as the liquidity measure of interest, the initial deviation would have come either from a market order or from cancellations at the top of the book that had removed one or more levels of the bid or ask. The subsequent return to the threshold level would result from limit orders arriving inside the spread.

A resilient LOB would then be one in which TED durations are generally low, indicating that the market returns to a particular level of liquidity quickly after a shock. Thus, our main interest is in the length, rather than the frequency of exceedances. The reason is that in our view, the LOB is resilient when such exceedance events are short in duration, rather than when such exceedance events are rare, because their frequency pertains more to market depth and the amount of new information hitting the market, rather than to the resilience of the book and the quality of liquidity provision.

The definition here is quite flexible, since it is agnostic to the liquidity measure used and one could potentially use any of the measures discussed in Chapter 3.2.2. In the case of the XLM liquidity measure, for example, such a resilience proxy would provide information on the time to replenishment of the LOB volume and relative price structure, after an impact from a market order or cancellation, across multiple levels of the LOB.

There is also flexibility in the choice of the liquidity threshold, for example as a quantile of some historical distribution of the liquidity measure. We intentionally do not specify a threshold in the definition, as the recovery of a liquidity measure to, e.g. its median values may be important to a brokerage house or a large fundamentals trader executing a large order, while a regulator may be more interested in the duration of more infrequent events, where the spread reaches very high quantile levels of the empirical distribution.

4.2.1 Examples of TED liquidity resilience measures

We can now extract the TED observation random variables for two example liquidity measures. We will utilise these to illustrate the survival regression framework we develop. These liquidity measures are the inside spread and the XLM, which were defined in Equations 3.1(p. 55) and 3.5(p. 58), respectively.

In terms of the aspects of liquidity delineated in Section 3.2, the inside spread reflects the tightness aspect, while the XLM, as a Cost of Round Trip (CRT) measure, is also descriptive of the depth of volume at each level of the LOB. Whereas R is fixed in the definition of the measure on the Xetra exchange[1], we allowed it to vary, so that the measure was still defined when there is insufficient volume in the LOB. We set $R = \min(25000, \sum_i TV_t^{a,i}, \sum_i TV_t^{b,i})$, i.e. the minimum of 25000 of the local currency (EUR) and the volume available on either side of the LOB.

The TED observation random variable for the two liquidity measures is then defined as in Equation 4.1. The only difference being that the threshold c for the spread is in cents, whereas for the XLM it is in basis points, to be in the same units as the liquidity measure.

4.3 Features of the LOB data and TED observations

We use the 82 day trading sample (January 2nd to April 27, 2012) from the Chi-X equity dataset described in Chapter 2.1. This contains all limit order submissions, executions and cancellations in the visible order book for all stocks in the CAC40. Both limit order submissions

[1] `http://xetra.com/xetra/dispatch/en/xetraLiquids/navigation/xetra/300_trading_clearing/100_trading_platforms/100_xetra/600_xlm`

and executions in our dataset may be the result of pegged, limit or iceberg orders, however, the data only indicates the resulting submission of the limit order. In addition, a cancellation may be automatic (as a result of a time in force option), or as a result of a manual cancellation request, but this is not indicated in the data. We do not attempt to infer this information here, and in any case we have sufficient information to rebuild the LOB without it.

The TED observations considered are those occurring between 08:01 and 16:29 London time daily, to avoid market opening and closing effects. We also note that while the continuous trading hours on Chi-X are not necessarily the same as those in the national exchanges where the assets trade, for these French stocks the opening hours coincide. Hence, we do not have any additional considerations that would result from the sudden submission or withdrawal of liquidity from the primary exchange.

For both the descriptive statistics and the results, we select two stocks with which to perform further experiments, namely Credit Agricole (stock symbol ACAp) and Sanofi (stock symbol SANp). These stocks were chosen as being representative of stocks with low and high share prices, respectively, in order to examine any differences in the liquidity resilience behaviour of stocks due to share price. In addition, we provide an overview of the explanatory performance across all CAC40 stocks.

4.3.1 Intra-day variation in liquidity resilience

There are clear diurnal patterns in the frequency of the TED observations over our sample period. An exceedance is more likely to occur close to the start of the day, and we note a second concentration of exceedances around 13:30 London time, or 08:30 ET. While the concentration of observations in the morning may be due to well documented market open effects (see, e.g. Biais et al. [1995]), this does not explain the very distinct mid-day clustering. We postulate that this may be due to the release of a number of economic reports in the US (indicatively, weekly jobless claims, retail sales, core PPI, housing starts and non-farm payrolls are all released at this time, but on different dates), and empirical findings substantiate this.

In Figure 4.2, we focus on days with economic releases at 08:30 ET, in order to show the difference in TED occurrences from days with no announcements. It is clear then that the 13:30 clustering of TED observations can be attributed to these announcements. It is also very likely that a number of announcements scheduled for release at 10:00 ET are responsible for the second clustering of TEDs at 15:00 London time. While the announcement event has a clear effect on the frequency of TEDs around the time of publication, we have not found these events to have an effect on the daily model fits, and we will thus not consider this further.

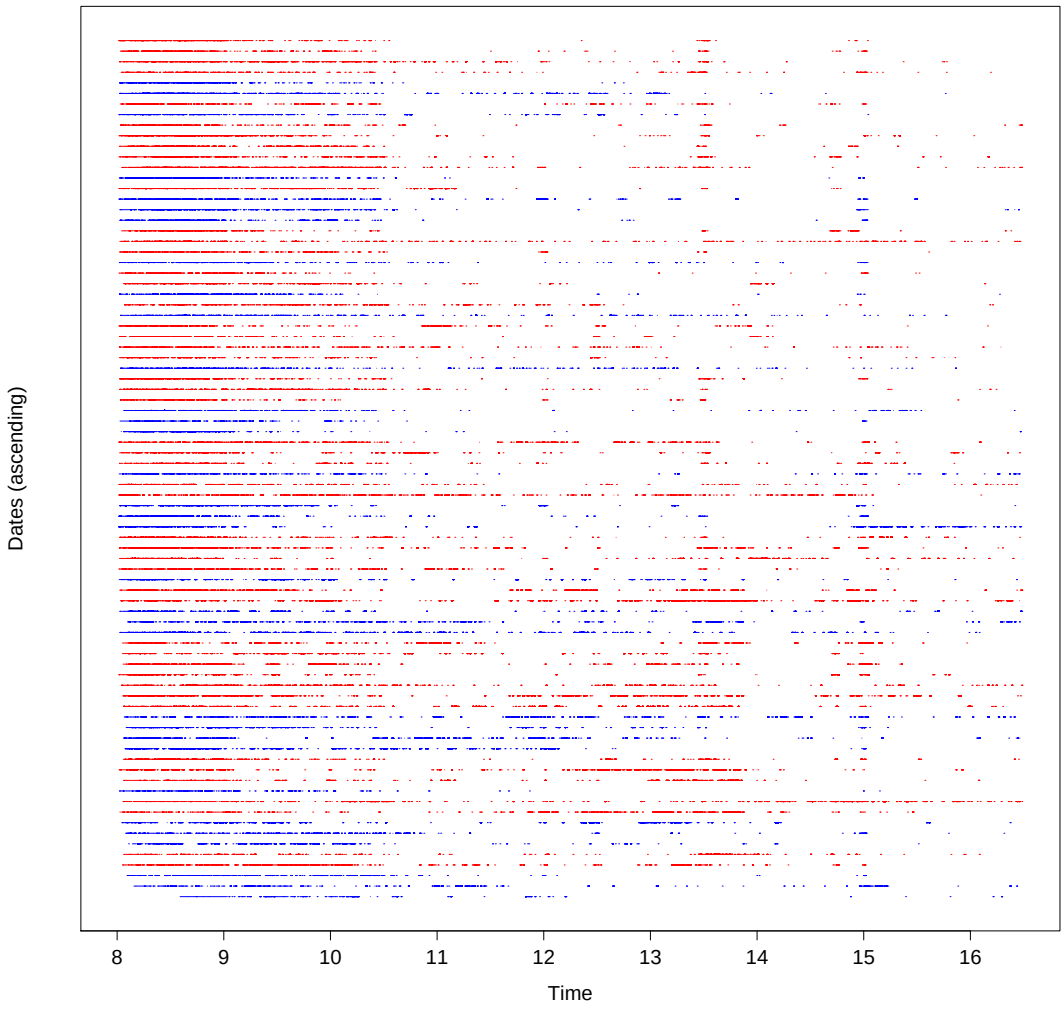


Figure 4.2: The duration of time during all the trading days in our dataset that the inside spread $\mathcal{M}$ of Credit Agricole (stock symbol ACAp) is above the 9th decile threshold. Time is on the x-axis starting at 8:01 am in the morning and ending at 16:29. Each day corresponds to one row on the y-axis. Rows are coloured blue if there is a US economic announcement at 13:30 London time (08:30 ET), and blue otherwise.

Liquidity droughts are generally longer and there is significantly more variation in their durations in the first two hours, compared to the rest of the day. In addition, the distribution of the durations on a log scale, for the examples we considered, tends to be symmetric. For periods where there are far fewer observations (such as mid morning and mid afternoon) we see a more skewed relationship in the log duration. Figures 4.3 and 4.14 show how this relationship varies throughout the day.

We next illustrate the number of TED observations one might expect every day, as we alter

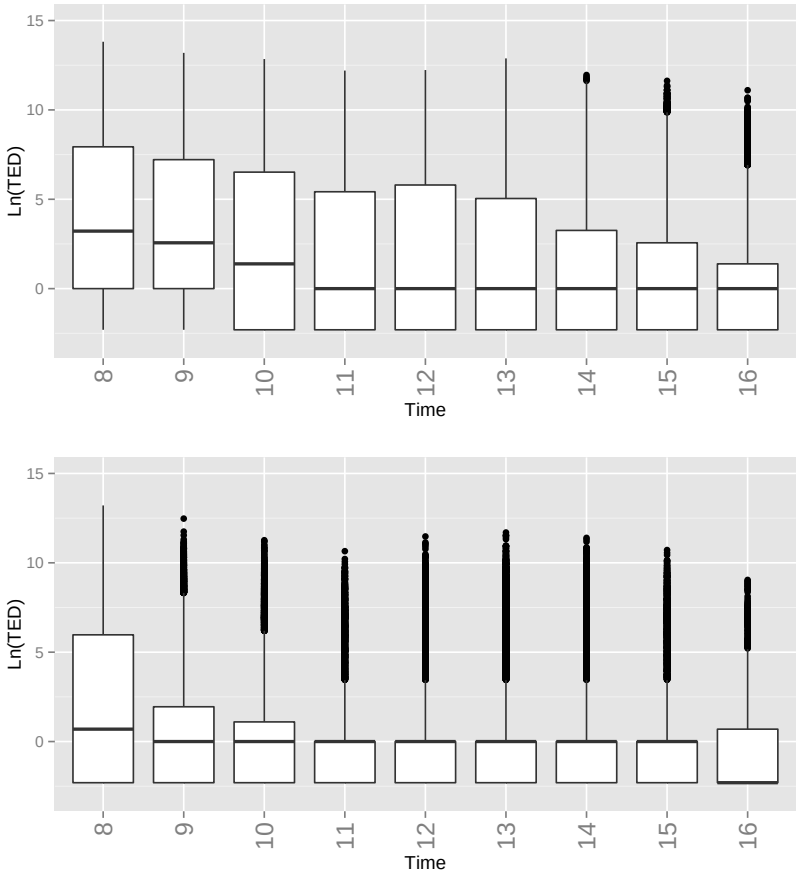


Figure 4.3: The duration of TED observations above the median (top) and 9th decile (bottom) spread value throughout the trading day for stock Credit Agricole for the liquidity measure $\mathcal{M}$ representing the inside spread. The boxplots show the 25th to 75th percentile range of the duration of these exceedances over the 4 month period. The black dots are beyond the whisker, which itself is 1.5 times the interquartile range from the upper or lower hinge.

the liquidity threshold. For the two liquidity thresholds corresponding to the 5th and 9th decile of the spread for every day, we show in Table 4.3(p. 118) the mean and minimum number of TED observations over the 82-day period. We note that for several assets, there are thousands of such observations, which is to be expected, as the daily activity is generally in the hundreds of thousands of events (limit orders, market orders and cancellations).

4.3.2 Data considerations and assumptions

We note that the granularity of the millisecond timestamped transactions in our dataset is such that we often have TED observations of value 0 (i.e. the liquidity returns to the threshold level in less than 1 ms). As this would cause problems with the estimation of our model, we set the

values of these observations to 0.1 ms, which is reasonable, as it corresponds approximately to the smallest round-trip time for messages to the exchange.

4.4 Liquidity resilience model formulation

In this section we will detail the statistical framework we adopt to model the TED random variables as a function of interpretable covariates (factors or transforms) derived from the LOB structure. Formally, we develop a survival regression framework that allows us to study the resilience of any chosen liquidity measure as a function of the LOB structure. To the best of our knowledge, there have been no previous studies that have adopted such a survival modelling framework to study resilience or liquidity dynamics. In fact, there have been relatively few previous studies incorporating a survival modelling regression framework into the study of features of the LOB [Chakrabarty et al., 2006, Lo et al., 2002], compared to the number of papers discussing the attributes of S.D.E., time-series and Markov chain-based models.

4.4.1 Survival analysis introduction

Survival analysis is a branch of statistics encompassing techniques to analyse positive-valued random variables, typically the lifetimes of mechanical components, or the time to response to a particular drug [Miller Jr, 2011]. Survival models can typically be incorporated into a regression framework, and we can thus explain some of the variation in the variable of interest through explanatory covariates. Harrell [2001] considers a wide range of regression modelling strategies in the context of survival analysis. The unique advantage of survival analysis, compared to the larger class of regression models, is that it can incorporate censored observations, which occur when one cannot obtain the true value of an observation because the terminating event (e.g. failure, response, death) has not occurred inside the observation window, and we will explain how such observations are treated separately in survival models in Section 4.4.2.

In the context of the LOB, survival regression frameworks have mainly been employed to explain the variation in the lifetimes of limit orders. Al-Suhaibani and Kryzanowski [2000] used such a framework for the Saudi Stock Market, assuming a Weibull distribution for the time to execution of limit orders, and considering cancelled or expired orders as censored. In the regression framework, they considered covariates related to aggressiveness and order size for the incoming limit order, as well as LOB covariates including the inside spread and total shares with higher priority in the LOB, order imbalance and the proportion of market orders. The inside spread was found to be related to an increase in the average limit order lifetime under the model, whereas increased order aggressiveness was naturally found to decrease the average

time to execution. A similar formulation was used by Cho and Nelling [2000] for the New York Stock Exchange.

Lo et al. [2002] used a more general formulation with a Generalised Gamma Distribution, but as limit order book data was not available, covariates are either approximations of the LOB state, or slower moving averages. Similar results to those of Al-Suhaibani and Kryzanowski [2000] are obtained regarding the effect of order aggressiveness in decreasing order lifetimes, but execution times were not found to be sensitive to the size of the limit order.

While in these studies, for which the datasets considered are in the late 1990s, the censored observations generally accounted for less than half of all observations, the changing nature of financial markets means that the vast majority of order traffic are now order cancellations. Thus, defining cancellations as censored observations in a survival model may no longer be appropriate, as it will affect the estimation of the model.

4.4.2 Survival model specification

Under the main assumption of survival modelling, the survival function that denotes the probability that the i-th threshold exceedance of the threshold c will have a duration that is beyond any positive time τ is given by:

$$S(\tau) = 1 - F(\tau; \boldsymbol{\beta}) = \mathbb{P}\mathrm{r}\left(\tau_i \geq \tau\right) \tag{4.3}$$

where $F(\tau; \boldsymbol{\beta})$ denotes the distribution function for the assumed probability model for the random survival times, and $\boldsymbol{\beta}$ is a vector of coefficients in the regression model, which parameterise the survival distribution.

If all observations are i.i.d and uncensored (i.e. the event of interest always happened within the observation window), we could simply estimate the model via standard maximum likelihood estimation, where for a given parameter vector $\boldsymbol{\beta}$, the likelihood function is

$$L(\boldsymbol{\beta}|\tau_1 \ldots \tau_n) = f(\tau_1 \ldots \tau_n|\boldsymbol{\beta}) = \prod_{i=1}^{n} f(\tau_i|\boldsymbol{\beta}) \tag{4.4}$$

However, for a given fixed threshold c, once the i-th exceedance at time T_i occurs, there is no guarantee that the liquidity process would ever return back through this threshold within the trading day. We do, however, assume that given enough time, the event of interest would eventually occur i.e. the liquidity process is mean reverting. Without this assumption, the density we specified for the survival times, which models the distributions of the durations, would be improper, as it would not normalise to unity on its support. We would then have to calculate the density conditioning on the event actually occurring.

Let each unit τ_i have a potential maximum observation time $T_D - T_i$, i.e. the time remaining until the end of the observation period T_D. For a censored observation, we only know that the lifetime τ_i exceeds the maximum observation time $T_D - T_i$, as censoring is non-informative (that is, the time of censoring is independent of the time of failure). The contribution to the likelihood of this event is then

$$L_i = S(T_D - T_i) \tag{4.5}$$

If we assume for these observations that T_i is independent of $T_d - T_i$, we can then obtain the likelihood function as follows:

$$L = \prod_{i=1}^{n} L_i = \prod_U f(\tau_i) \prod_C S(T_d - T_i)$$

where U and C are the sets of uncensored and censored observations, respectively.

4.4.3 Classes of survival models

The two most commonly used classes of survival models are the proportional hazards models and the accelerated failure time (AFT) models, for which a brief introduction is presented here, see Bradburn et al. [2003], Kalbfleisch and Prentice [2011] for a thorough discussion of both.

4.4.3.1 Cox proportional hazards models

With a Cox Proportional hazards model, a central concept is the hazard function

$$\begin{aligned} h(\tau) &:= \lim_{\delta\tau \to 0} \Pr\left(\tau \leq \tau_i < \tau + \delta\tau\right) \\ &= \frac{f(\tau)}{1 - F(\tau)} \end{aligned}$$

where $h(\tau)$ can be interpreted as the instantaneous rate of occurrence of the event of interest, given it has not occurred already. Integrating $h(\tau)$ to obtain the *cumulative hazard* $H(\tau)$

$$\begin{aligned} H(\tau) = \int_0^\tau h(t)dt &= -\ln(1 - F(t))|_0^\tau \\ &= -\ln(1 - F(\tau)) \\ &= -\ln(S(\tau)) \end{aligned}$$

and thus the survival function becomes

$$S(\tau) = e^{-H(\tau)} \tag{4.6}$$

One of the simplest possible survival models is for an exponentially distributed survival time, which can be obtained by the specification of a constant hazard function $h(\tau) = h$. The respective survival function is therefore

$$S(\tau) = e^{-h\tau} \tag{4.7}$$

In a proportional hazards model, the covariates affect the duration through the hazard function. For a given vector of covariates $\boldsymbol{x}_i$ and coefficients $\boldsymbol{\beta}$, the hazard function is

$$h(\tau|\boldsymbol{x}_i) = h_0(\tau)e^{\boldsymbol{x}_i'\boldsymbol{\beta}} \tag{4.8}$$

where $h_0(\tau)$ is a baseline hazard, which can take any form. One can see that changes in the covariates have a multiplicative effect on this baseline risk. In this case, the *hazard ratio* for two different individuals i, j with covariate vectors $\boldsymbol{x}_i, \boldsymbol{x}_j$ is

$$\frac{h(\tau|\boldsymbol{x}_i)}{h(\tau|\boldsymbol{x}_j)} = e^{(\boldsymbol{x}_i' - \boldsymbol{x}_j')\boldsymbol{\beta}} \tag{4.9}$$

which is useful for comparing the hazard functions of individuals in different groups, e.g. a placebo and treatment group in a medical study.

4.4.3.2 Accelerated Failure Time model

'Accelerated Failure Time' (AFT) models make the assumption that for two populations, the following holds for the respective survival functions $S_1(\tau)$ and $S_2(\tau)$

$$S_1(\tau) = S_2(\kappa\tau) \tag{4.10}$$

where κ is a constant. Thus the rate of 'aging' of the first population is κ times the rate of aging of the second population. The same relationship holds between the mean 'failure' times

$$\begin{aligned}
\mu_2 &= \int_{t=0}^{\infty} t f_2(t) dt \\
&= t\left(-S_2(t)\right)|_{t=0}^{\infty} + \int_{t=0}^{\infty} S_2(t) dt \\
&= -\lim_{t\to\infty} t S_2(t) + \int_{t=0}^{\infty} S_2(t) dt \\
&= \int_{t=0}^{\infty} S_2(t) dt \\
&= \kappa \int_{t=0}^{\infty} S_2(\kappa u) du \\
&= \kappa \int_{t=0}^{\infty} S_1(u) du \\
&= \kappa \mu_1
\end{aligned}$$

For the failure times, since we are modelling positive random variables, we can consider a log-linear formulation, incorporating model covariates as follows

$$\ln(\tau_i) = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{4.11}$$

where ε_i is a random error term. According to the distribution one chooses for τ, there is a corresponding distribution for ε_i, and vice-versa. A number of corresponding distribution pairs are listed in Table 4.1.

Distribution of ε_i	Distribution of τ
Normal	Lognormal
Logistic	Log-logistic
Log-gamma	Gamma
1 param extreme value	Exponential
2 param extreme value	Weibull

Table 4.1: Possible Accelerated Failure Time lifetime and associated error distributions.

From Equation 4.11 one can obtain the model for the failure time τ_i by exponentiating

$$\tau_i = \tau_{0,i} \exp\left\{\mathbf{x}_i'\boldsymbol{\beta}\right\} \tag{4.12}$$

where $\tau_{0,i} = \exp\{\varepsilon_i\}$. One can see that in an AFT model, the model covariates affect the duration by shifting the baseline distribution of τ, rather than the hazard function, as in the Cox proportional hazards model. A unit change in, say, $x_i^{(k)}$ to $x_i^{(k)} + 1$ will have a multiplicative effect of e^{β_k} on the failure time. The sign of the coefficient for a given covariate indicates the direction of the partial effect of this variable on the conditional probability that the duration of the deviation will exceed a time t.

We use the AFT approach in this chapter, as AFT models have a number of relevant advantages [Lambert et al., 2004]:

- The log-linear formulation of such models emphasizes that the roles of the regression parameters and dispersion parameters are clearly separated.
- The regression parameters in an AFT model are also robust towards neglected covariates.

For the second point, Hougaard [1999] examines the implications of not including all relevant covariates (for example, due to lack of knowledge). In the case of the log-linear AFT formulation, omitting covariates increases the variance of the error and its distribution may

fall outside the parametric family considered, but the regression part remains the same (and therefore, the model coefficients for the known covariates are unchanged). In contrast, in a proportional hazards case with constant hazard, such as the one considered earlier, omitting covariates leads to a distribution with a decreasing hazard.

4.4.4 AFT model estimation

The majority of the results presented in this section will be for lognormally distributed and Weibull-distributed exceedance times. These are both special cases of the log Generalised Gamma distribution, for which details are given in Appendix A. For the lognormal regression we assume for the error term $\varepsilon_i \overset{iid}{\sim} LN(\mu, \sigma)$ in Equation 4.11.

Each of the covariates in $\boldsymbol{x}$ is a transform from the LOB for which the liquidity measure is observed, and all covariates are described in Section 4.4.5. We note that we also considered models with interactions between the covariates, but interaction terms were not found to be significant in the majority of our models.

In the log-normal case, the observation random variables have the following distribution function and survival function:

$$f(t|\boldsymbol{x}_i) = \frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln(t) - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{2\sigma^2}\right]$$
$$S(t|\boldsymbol{x}_i) = 1 - F(t|\boldsymbol{x}_i) = \frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{\ln(t) - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sqrt{2\sigma^2}}\right)$$

Define $u = \frac{\ln(t) - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sqrt{2\sigma^2}}$. Then the log likelihood is:

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma) &= \ln L(\beta, \sigma) \\ &= \Sigma_{i\in U} \ln(f(\tau_i)) + \Sigma_{k\in C} \ln(S(\tau_k)) \\ &= \Sigma_{i\in U}\left[-\ln(\tau_i\sqrt{(2\pi(\sigma)^2)} - u_i^2\right] + \Sigma_{k\in C} \ln\left(\frac{1}{2} - \frac{1}{2}\text{erf}(u_k)\right) \end{aligned}$$

The partial derivatives with respect to β_j and σ are:

$$\frac{\partial l}{\partial \beta_j} = \Sigma_{i\in U}\left[2u_i \cdot \frac{x_{i,j}}{\sqrt{2(\sigma)^2}}\right] + \Sigma_{k\in C}\left[\frac{\frac{1}{\sqrt{\pi}}\exp(-u_k^2)\frac{x_{k,j}}{\sqrt{2(\sigma)^2}}}{\frac{1}{2} - \frac{1}{2}\text{erf}(u_k)}\right]$$
$$\frac{\partial l}{\partial \sigma} = \Sigma_{i\in U}\left[\frac{-1 + 2u_i^2}{\sigma}\right] + \Sigma_{k\in C}\left[-\frac{\frac{1}{\sqrt{\pi}}\exp(-u_k^2)(\frac{u_k}{\sigma})}{\frac{1}{2} - \frac{1}{2}\text{erf}(u_k)}\right]$$

The parameters can be estimated via MLE, with a Newton gradient descent method, using standard optimisation packages.

4.4.5 Model LOB Covariates

We consider the following categorisations of covariates in our nested model structures that are explored via a model search procedure. In evaluating these covariates for the construction of the regression design matrix, we consider the times of evaluation to match the times of the observed exceedance events above the specified threshold level, $t = T_i$. In the following, a 'level' of the LOB is defined as one in which there is at least 1 resting limit order. Thus the first 5 levels of the bid are the 5 levels closest to the quote mid-point, where there is available volume for trading. The covariates chosen pertain to the state of the limit-order book of a given stock. These are:

- The total number of sell limit orders in the first 5 levels of the LOB at time t, obtained according to $x_t^{(1)} = \sum_{i=1}^{5} \left| V_t^{a,i} \right|$ (where $|\cdot|$ is the number of orders at a particular level), and is denoted ask hereafter,
- The total number of buy limit orders in the first 5 levels of the LOB at time t, obtained according to $x_t^{(2)} = \sum_{i=1}^{5} \left| V_t^{b,i} \right|$, denoted bid,
- The total sell volume (in 1000s of shares) in the first 5 levels of the LOB at time t, obtained according to $x_t^{(3)} = \sum_{i=1}^{5} \frac{TV_t^{a,i}}{1000}$, denoted $askVolume$,
- The total buy volume (in 1000s of shares) in the first 5 levels of the LOB at time t, obtained according to $x_t^{(4)} = \sum_{i=1}^{5} \frac{TV_t^{b,i}}{1000}$, denoted $bidVolume$,
- The number of sell limit orders $x_t^{(5)}$ in the LOB that had received price or size revisions(and were thus cancelled and resubmitted with the same order ID), denoted by $askModified$,
- The number of buy limit orders $x_t^{(6)}$ in the LOB that had received price or size revisions, denoted by $bidModified$,
- The average age (in minutes) $x_t^{(7)}$ of sell limit orders in the first 5 levels at time t, denoted by $bidAge$.
- The average age (in minutes) $x_t^{(8)}$ of buy limit orders in the first 5 levels at time t, denoted by $askAge$.
- The instantaneous value of the spread at the point at which the i-th exceedance occurs, which is given by $x_t^{(9)} = P_t^{a,1} - P_t^{b,1}$ and denoted as $spreads$.
- For the nine previously defined covariates, we also include exponentially weighted lagged versions. For example, in the case of the $x_t^{(s)}$ covariate, the respective lagged co-

variate value is then given by $\sum_{n=1}^{d} w^n x_{t-n\Delta}^{(s)}$ where for a time t, we consider $w = 0.5$ is the weighting factor, $d = 5$ is the number of lagged values we consider and $\Delta = 1s$ is the interval between the lagged values. These covariates are hereafter denoted with the 'l' prefix.

- The number $x_t^{(10)}$ of previous TED observations in the interval $[t - \delta, t]$, with $\delta = 60s$, denoted by $prevexceed$.
- The time (in minutes) since the last exceedance, $x_t^{(11)}$, denoted by $timelast$.
- The average of the last 5 log TEDs, $x_t^{(12)}$, denoted by $prevTEDavg$.
- The activity in the entire CAC40 index (in 1000s of limit orders, cancellations, executions) in the previous second $x_t^{(13)}$, denoted by $indact$.
- A dummy variable indicating if the exceedance occurred as a result of a market order to buy, $x_t^{(14)}$, denoted by $mobuy$.
- A dummy variable indicating if the exceedance occurred as a result of a market order to sell, $x_t^{(15)}$, denoted by $mosell$.

Altogether we then have 24 covariates, 15 instantaneous and 9 lagged.

4.5 Results and Discussion

Let us now analyse our resilience metric defined with respect to the inside spread as the liquidity measure $\mathcal{M}$. We could assume that liquidity resilience was stationary over our 82-day sample period, in which case we could fit our model to the entire dataset. We demonstrate that this would not be a good assumption as the model fits produce varying coefficient values throughout the period, in Figures 4.4 and 4.15. Instead, we only assume that liquidity resilience is locally stationary (intra-day), and we fit the model to the dataset for every asset daily.

For the empirical evaluation of the explanatory power of our model regarding the variation in the TED observations, we adopt the AFT model formulation described in Section 4.4.2. We selected 2 example models that are special cases of the Generalised Gamma distribution (g.g.d.) family: The Lognormal model, for which we present the majority of our results, and the Weibull model in Section 4.5.5, to exhibit situations in which it may provide a better fit. We noted in Table 4.1 that under a log transform, the former gives a linear regression with additive Gaussian errors, while the latter gives us a linear regression with errors that have an extreme value distribution. We note that because of the censoring mechanism we used, we only had very

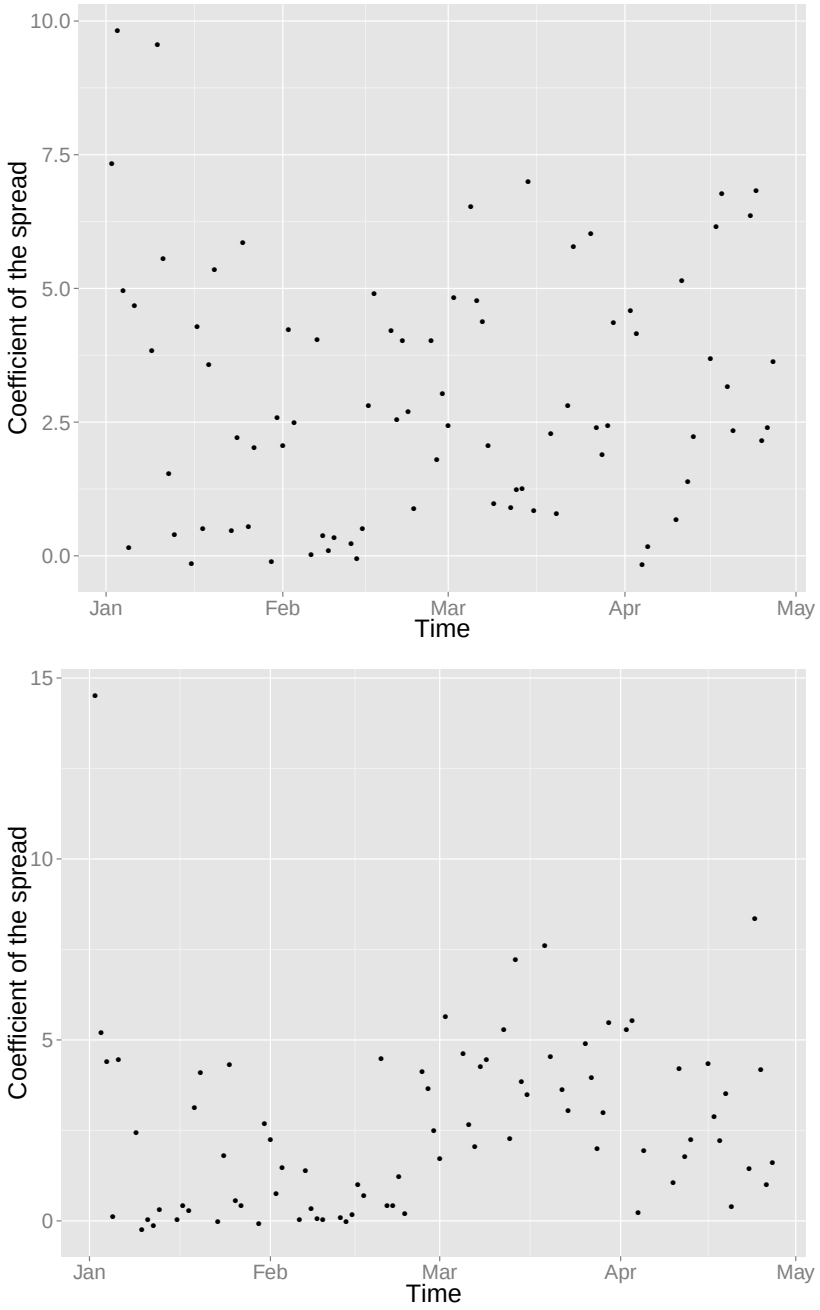


Figure 4.4: Coefficients of the value of the spreads, at the time of exceedance for every fitted daily model for stock Credit Agricole. The top graph is obtained using thresholds corresponding to the median spread, while the bottom graph uses thresholds corresponding to the 9th decile spread.

few censored observations. These were found to only have a minimal effect on the estimated coefficients and thus censoring was ignored to simplify estimation.

4.5.1 Explanatory power

In this section we assess the extent to which instantaneous and lagged variables in the LOB can explain the variation in our TED metric, and therefore become part of meaningful models to describe LOB liquidity resilience. We demonstrate that with a careful selection of the covariates, we can obtain substantial explanatory power for our liquidity resilience measure, which is consistently good over time, in different market conditions, and for a range of different assets.

We fit the lognormal regression model, in which all covariates explained in Section 4.4.5 are considered (together) each day, for the entire 4-month period of our dataset, for each asset in the CAC40. As there were very few censored observations, their effect is minimal, and thus these were generally ignored to simplify the estimation of the model. We evaluate the explanatory power of this model in terms of the proportion of the variation in the TED metric that can be explained by it, with the adjusted coefficient of determination (adjusted R^2), which we briefly explain here.

The coefficient of determination is

$$R^2 = \frac{SS_e}{SS_t}$$

which corresponds to the total variation explained by the regression model, where SS_e and SS_t are, respectively, the explained sum of squares and the total sum of squares. When introducing additional explanatory variables, we would always expect the R^2 value to increase. The adjusted R^2 is often used in its place, as it penalises larger models:

$$R^2_{adj} = 1 - (1 - R^2)\frac{N-1}{N-k-1}$$

We find that our model has substantial explanatory power, and we show in Figure 4.5 that the adjusted R^2 results obtained from fitting the model every day for the two stocks for which we are presenting detailed results. For the four month period these R^2 values are above 15% on many days, with scores above 20% on some days also.

Figure 4.6 shows the adjusted R^2 values obtained from fitting the full model, using as a threshold either the median or the 9th decile spread, obtained every day. We find that for many of the stocks, the median adjusted R^2 value is over 10%. For some stocks we find even more remarkable adjusted R^2 values of over 20%. We also note that for many assets, the explanatory power differs between the two threshold levels. This is an indication that resilience may exhibit different behaviours, when defined with respect to different liquidity thresholds, and we introduce the concept of the Liquidity Resilience Profile in Chapter 5 to obtain resilience results across multiple thresholds.

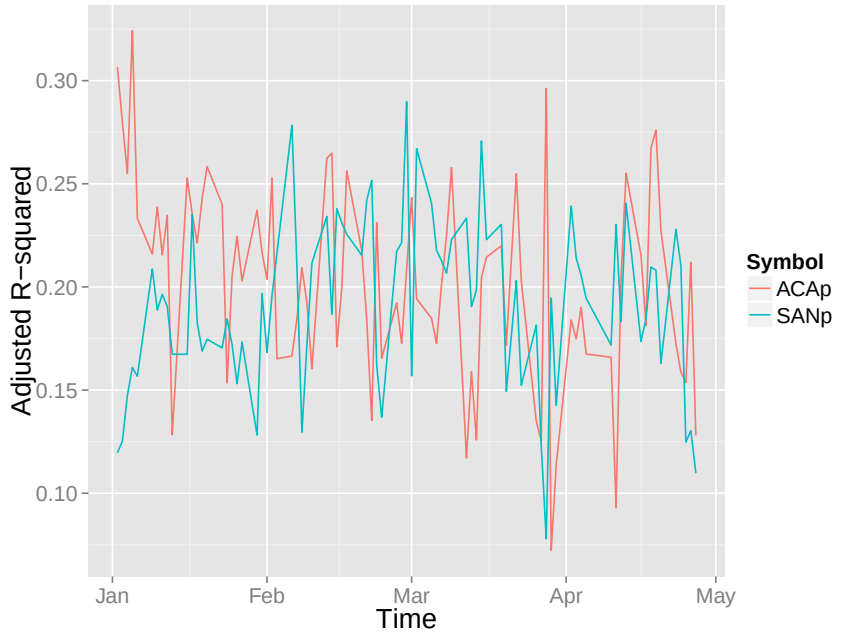


Figure 4.5: The adjusted R^2 values over time for stocks Credit Agricole (ACAp) and Sanofi (SANp).

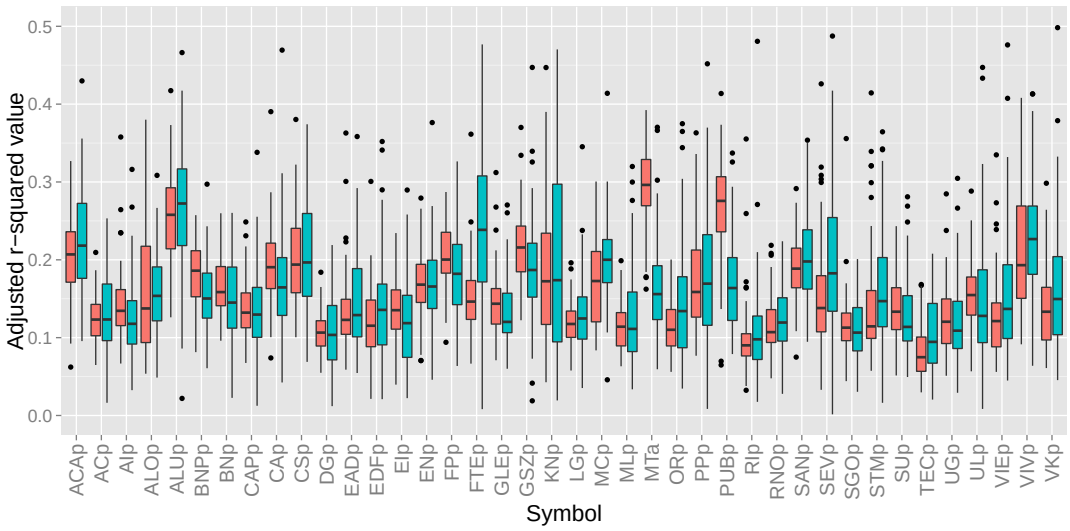


Figure 4.6: Boxplots of the adjusted R^2 value obtained from fitting the full survival model separately for each day in our dataset for both the threshold corresponding to the 5th decile of the spread (the median - red) and the 9th decile threshold (blue).

4.5.2 Model selection

In statistical modelling, one of the most prominent issues is finding the best regression equation, which entails choosing a subset of covariates that optimises some selection criterion [Gatu and Kontoghiorghes, 2006]. Including additional covariates always increases the explanatory power

of a model, but may result in overfitting, so a common approach used for model selection is to penalise the least squares or log likelihood scores, such that they take into account model size. This favours more parsimonious models and examples of criteria are Mallows' C_p, and Akaike's Information Criterion. The explanatory performance of our models was assessed in terms of the adjusted coefficient of determination (adjusted R^2), as above.

In Figures 4.5 and 4.6 we obtained results for the full model fit, but we should be able to improve on this result also, by selecting the subset that maximises the adjusted R^2. This, however, poses a computational problem. In a regression model with p covariates that can be included in a model, we have $2^p - 1$ possible models to choose from. As p increases, an exhaustive search of the entire space of possible models would thus be exponential in p. Although strategies to improve the efficiency of this search have been discussed, e.g. in Gatu and Kontoghiorghes [2006], for a large value of p, an exhaustive search through all possible models is prohibitive in terms of computational power.

In order to search through the model space, we thus employ a modification of the leaps package in R [Lumley, 2004], which uses an efficient version of the branch-and-bound algorithm first described in Furnival and Wilson [1974]. The algorithm can offer vast performance improvements, by eliminating large sections of the search space. It is guaranteed to terminate, yielding the subset that maximises our selection criterion.

A brief description of the general algorithm is as follows: For a given set of models in a partitioned model space, the algorithm proceeds by calculating upper and lower bounds for the selection criterion, for a supermodel and submodel of that set, respectively. If, during the search process, another model has been identified that has a higher selection criterion score than the upper bound, the given set can then safely be ignored, as it cannot give rise to a better performing model. Otherwise, the set is partitioned further. This process and partitioning is repeated until we have a singleton model, which is then evaluated.

In our case, for every model subspace $M_i, i = 1 \dots p$, where the M_i model subspace is the set of all possible models containing i covariates. For example, the full model contains all covariates and is the only model in its subspace, while the smallest model subspace is comprised of models that contain the intercept and any one of the possible covariates. Intermediate model subspaces are comprised of models with all combinations $i = 2 \dots p - 1$, where $p = n + m$ is the total number of covariates, n contemporaneous plus m lagged covariates. There are thus $C\binom{p}{i} = \frac{p!}{i!(p-i)!}$ models in total in each subspace - we are searching for the model that maximises the adjusted R^2 criterion.

Our modification to the leaps package is in the presentation of the results, so that it distin-

guishes between covariates that are selected to be part of the model ('Present'), and covariates that are significant in particular models ('Significant'), in Figure 4.7. In this way, we can identify which covariates are consistently present as we move across subspaces. This is interesting because it gives us a relative measure of the contribution of that covariate across different assumptions of parsimony for the model. Particularly for higher model subspaces, some of the covariates in each subset model are not significant, and we distinguish between the covariates that are significant or not, at a 5% level of significance.

To illustrate our findings we first present results for a given day of data for Credit Agricole in Figure 4.7 and for Sanofi in Figure 4.16, where for all model subspaces, we perform a branch-and-bound search, ranking the models in each subspace based on their adjusted R^2 score. We thus obtain the best combination of covariates, for each subspace and for each day of data. We can then identify the covariates that are consistently present as we move between model subspaces. This is interesting because it gives us a relative measure of the contribution of that covariate across different assumptions of parsimony for the model. Particularly for higher dimensional model subspaces, some of the covariates in each subset model are not significant, and we distinguish between the covariates that are significant or not, at the 5% level of significance.

The best models for each model subspace are ranked by the adjusted R^2 value, with the vertical lines in the graph representing covariates that are consistently part of the best model for every subspace. We observe that for both assets, the average (logarithm of the) TED over the past 5 exceedances (*prevTEDavg*), the instantaneous value of the spread (*spreads*) and the number of previous exceedances (*prevexceed*) are covariates that are generally selected to be part of the best fitting model across most model subspaces. This is true for durations of exceedances both over the median and the 9th decile spread threshold. Other covariates are also found to contribute to the explanatory power of the model, but either for selected assets or for selected liquidity thresholds. In practice, we also observe very small differences in the explanatory power of the best model for a number of subspaces (i.e. they only differ at most in the third decimal point of the score).

To get an indication of the time stability of these model structures (and identify covariates that are consistently selected in the model), we illustrate the relative frequency with which parameters appear in the best models of every subset. That is, for each model subspace, we count the number of times each covariate forms part of the model with the highest adjusted-R^2 value over the four month period. Figure 4.8 indicates that the covariates identified earlier as being important in explaining the variation in the TED for a single day (*prevTEDavg* and

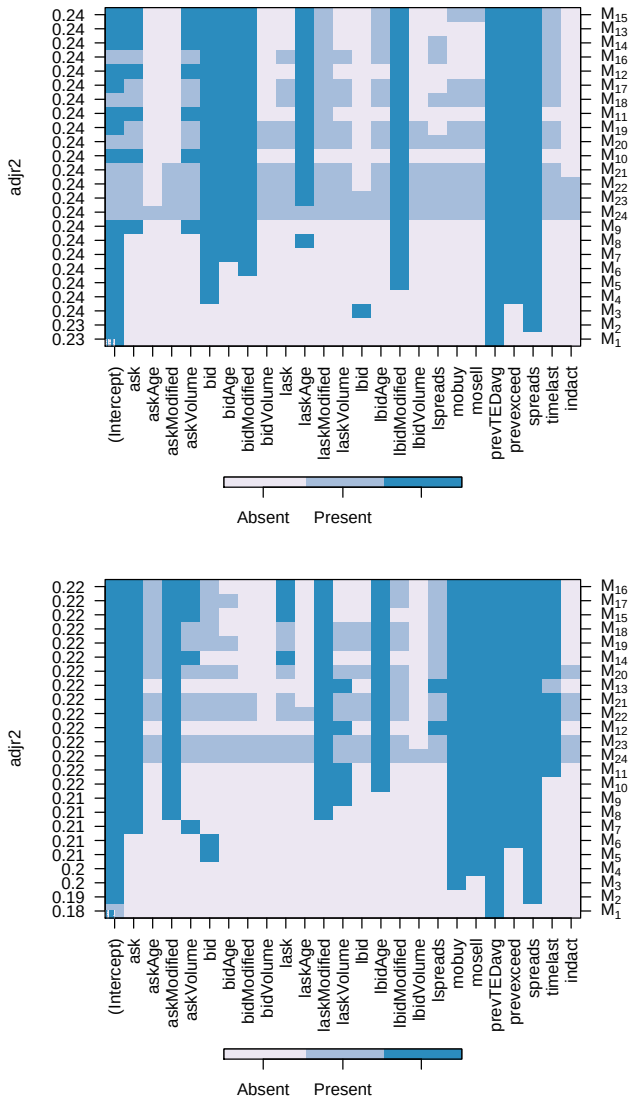


Figure 4.7: The adjusted R^2 values for the best models from each model subspace (where each subspace M_i contains all models with $i = 1 \ldots 24$ covariates) for a single trading day (the 17th of January 2012) for stock Credit Agricole, where the median spread (above) or the 9th decile (below) are used as the threshold. A dark shaded square indicates that a covariate has been included in the model and is statistically significant at the 5% level, with light squares not statistically significant. For instance, row M_3 corresponds to a specification with the following covariates: $intercept$, $lbid$, $prevTEDavg$ and $spreads$. The models are ranked by the best adjusted R^2 value, and we see that in this case, the best scoring model is obtained using a subset of 15 covariates, of which only 10 are found to be significant at the 5% level.

spreads) are also consistent features in models across time. However, *prevexceed* does not form part of the best model very frequently, except in higher model subspaces, possibly because it is less informative in the presence of the aforementioned covariates.

Besides the frequency of the presence of each covariate in the best fitting model of a given subspace, we also evaluate individual covariate significance over time via a partial t-test at the 5% level in Figure 4.9. At higher model subspaces, we find that several covariates are found to be statistically significant (i.e. reject a null hypothesis for a partial t-test) less frequently.

This is what one may expect, when covariates become less significant in the presence of other correlated covariates, i.e. collinearity in the factors of the LOB covariates takes effect.

4.5.3 Interpretation of covariates

Since we have obtained model fits for every model subspace, and for every day in our dataset, we can investigate the inter-day variation of the coefficients, as well as their magnitude and sign over time. In Figure 4.10 and Figure 4.19 we summarise these results for the best fitting model on each day. The plot demonstrates the following features: 1) the frequency with which each covariate appears in the best model over the four month period, 2) the variation in each coefficient for the days in which the respective covariate appeared in the model and 3) the coefficient sign, and thus its interpretation with regards to how it influences the resilience mean and quantile function surfaces, for a unit change in the value of the covariate.

In this analysis, we recall that under an AFT framework, the sign of the coefficient for a given covariate indicates the direction of the partial effect of this variable, on the conditional probability that the resilience, as measured by the exceedance duration for a given threshold, will exceed a time t. Therefore we can interpret positive coefficient values as influencing the liquidity resilience of the LOB by slowing the return to a desirable level, whilst negative coefficients tend to result in a rapid return to the considered liquidity level, indicating higher resilience marginally, with respect to that covariate.

The *prevTEDavg* covariate, which is an average of the last 5 log TED observations, and generally has a positive coefficient, is thus associated with a slower return to the threshold liquidity level. Thus, our model indicates that the expected TED over a particular threshold will be larger, when the duration of similar exceedances in the near past has been longer. We also find that the instantaneous spread covariate (i.e. the value of the spread at the moment when it first exceeds the threshold) appears frequently in the best model and has a positive coefficient (and would also increase the expected TED). This results matches our intuition, as the wider the spread just after an event at time T_i+, the longer we would expect the spread exceedance to

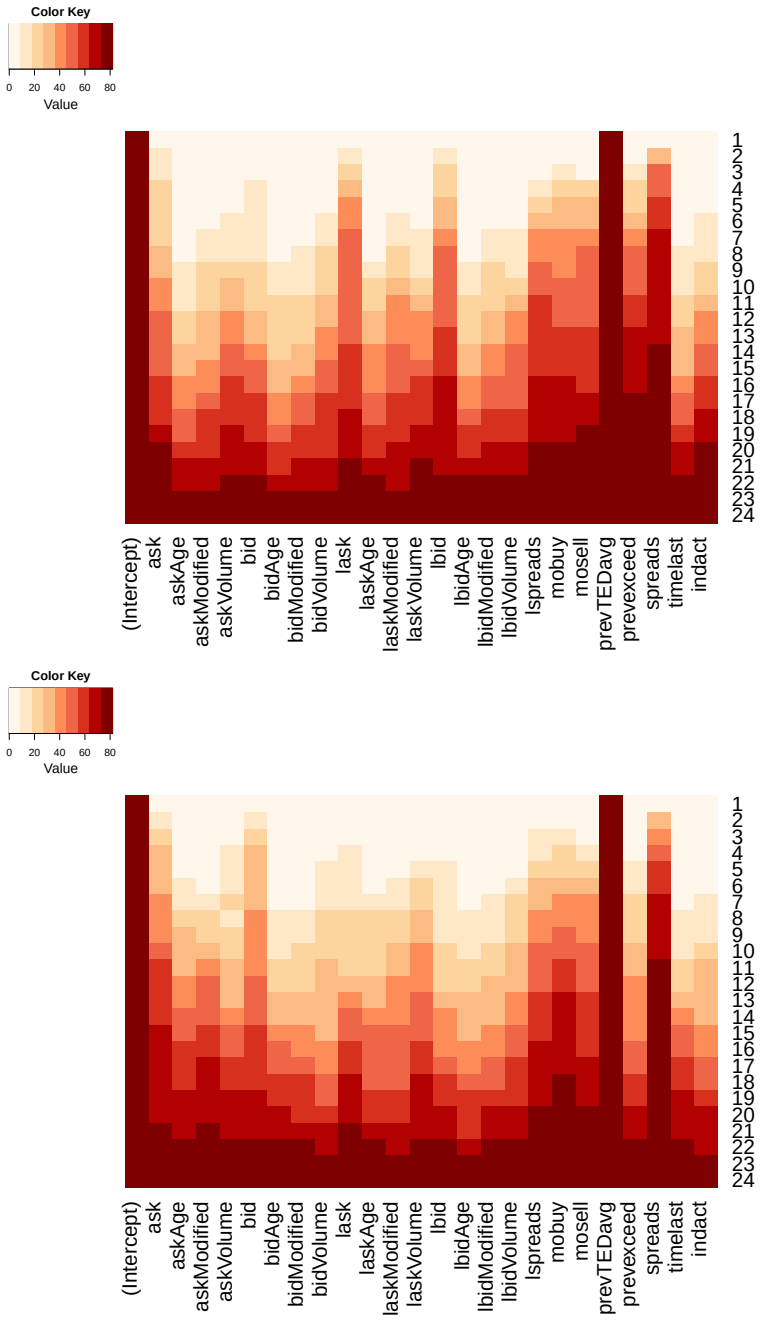


Figure 4.8: Heatmap of the relative frequency with which parameters appear in the best daily models of every subspace (frequency in terms of the number of daily models over the 82 day period) for the Credit Agricole dataset using the daily median (left) or the 9th decile (right) of the spread as the threshold value. So for instance, the element in row 10 and column *lask* indicates the relative frequency (in terms of the fraction of days over the 82 day period) by which the covariate *lask* has appeared in the best model with 10 covariates amongst all models with 10 covariates.

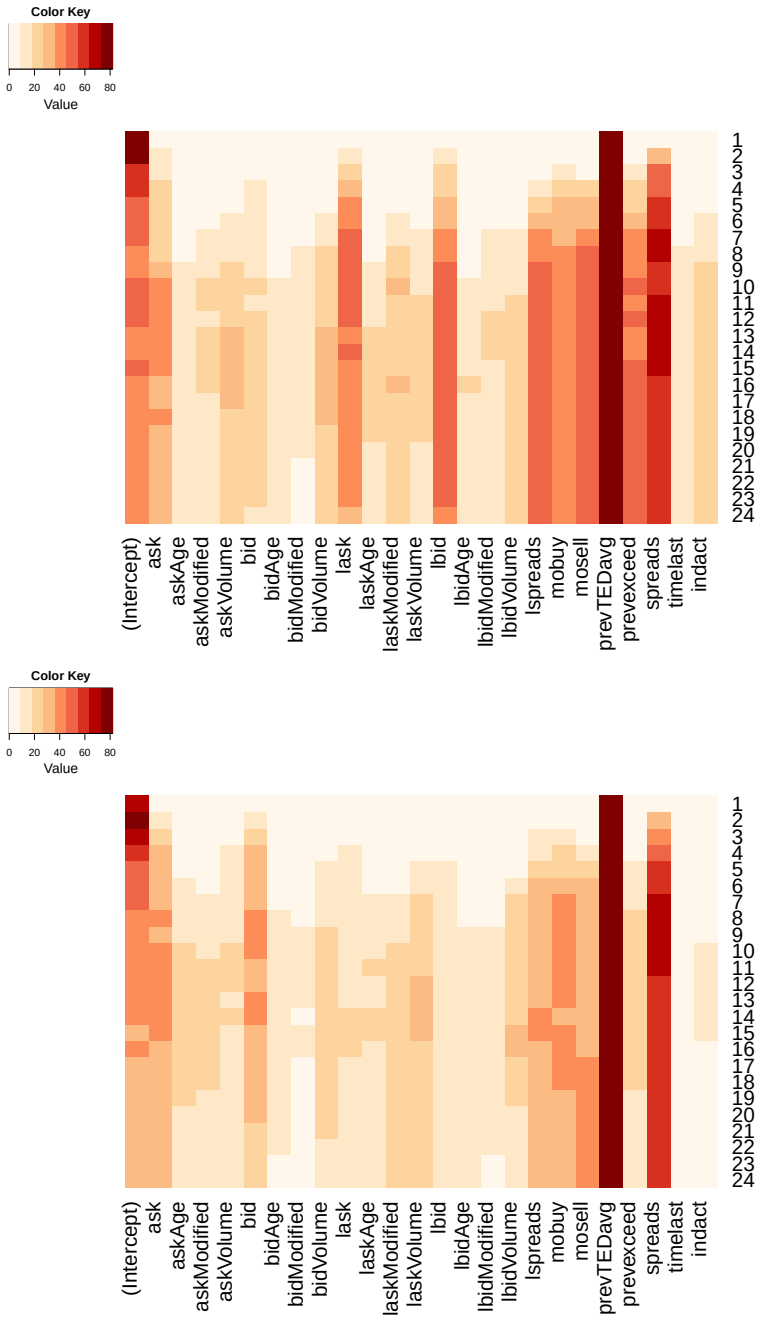


Figure 4.9: Heatmap of the relative frequency with which the parameters are found to be significant at the 5% level (frequency in terms of the number of daily models over the 82 day period) for the Credit Agricole dataset, using the median and 9th decile thresholds.

last, on average.

The *askModified* and *bidModified* coefficients (which measure the number of orders in the LOB that have had price or size revisions in the first 5 levels of the LOB, retaining the same order ID) have median coefficient values which are also positive. These orders can be interpreted as fleeting liquidity, as some LOB orders are modified a number of times before

they execute (either because of some strategy or because they are pegged to some price level), or they are cancelled altogether.

Of particular interest are the *mobuy* and *mosell* covariates, i.e. dummy variables indicating whether the exceedance resulted from a buy or sell market order respectively (if both are zero, then the exceedance was a result of a cancellation). For stock Credit Agricole, the coefficients are generally found to be positive, indicating that exceedances from market orders are associated with an increase in the expected TED, compared to cancellations. For stock Sanofi, on the other hand, the opposite effect is found. This may be due to different market making strategies being active in the aforementioned stocks.

4.5.4 Forecasting liquidity resilience

The TED survival regression framework lends itself naturally to out of sample forecasting. This would be a practical use in many areas, for example the optimal execution/liquidation of large orders on an exchange, in order to lower execution costs. For a specified volume and a specified partitioning of that volume into trading blocks, the model could help us determine the average time it would take for liquidity to recover after the immediate execution of each trade, for any state of the LOB. Alternatively, the model could help us determine the optimal size of the trading blocks, so that the time required for liquidity to recover is below some desired value, for a given LOB state. In this regard, it could also provide a means of determining 'optimal' LOB states in which to initiate execution, such that liquidity resilience is high under the model.

In our illustrative examples we focus on the first use case for the model, i.e. we forecast the average time for liquidity to recover to a certain threshold value post execution. For this, we estimate the model at both hourly and daily intervals time and use the estimated model parameters and the latest covariates to forecast the next TED.

$$\mathbb{E}\left[\ln(\tau_{i+1})|\boldsymbol{x}_{T_{i+1}}, \boldsymbol{\beta}_{\boldsymbol{T_h}}\right] = \boldsymbol{x}'_{T_{i+1}}\boldsymbol{\beta}_{\boldsymbol{T_h}}, \;\; T_h \leq T_{i+1}. \tag{4.13}$$

We show the results in Table 4.2, for both intraday, and one–day–ahead analysis. As way of comparison, we also compute the power of a naïve forecasting approach, which predicts that the next TED will be equal to the previous one. We evaluate forecast accuracy by the mean square prediction error (MSPE) and standard deviation of the prediction error, with one set the results for each time period and asset. Table 4.2 shows the aggregate results, averaged across all time periods and assets.

For intra-day forecasting, we quantify the predictive power of our model over hourly intervals. In particular, for every hour starting from 10:00, we estimate the best model on all

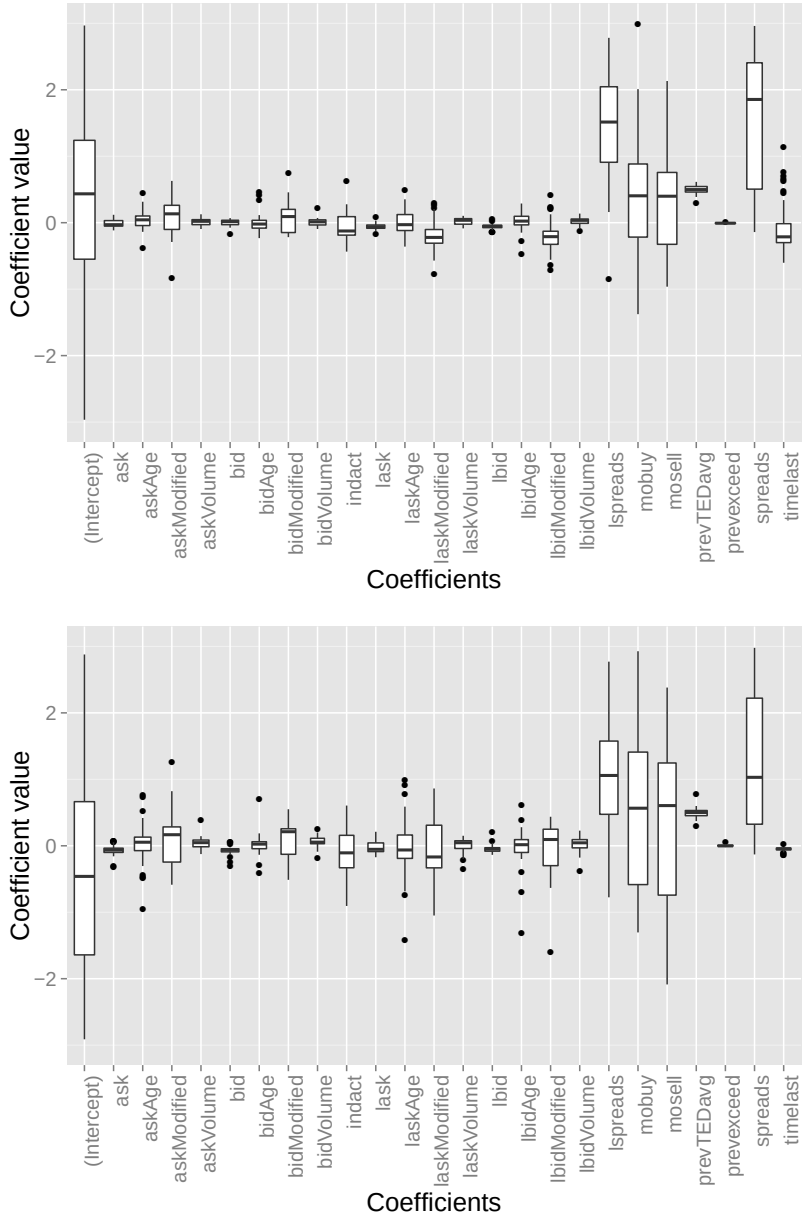


Figure 4.10: Coefficients of the best daily models (in terms of the adjusted R^2 values) for Credit Agricole for median threshold spread exceedances (left) and 9th decile spread exceedances (right). The width of every boxplot is proportional to the number of times that the covariate appears in the best model over the four month period. The hinges of the boxes correspond to the 25th and 75th percentiles, and whiskers extend to 1.5 times the interquartile range.

previous data that day and calculate the mean square prediction error (MSPE) of the model on the next hour. We also evaluate the predictive power of our models for 1 day ahead forecasting.

We note that the forecasting power of the model is substantially higher than the naïve approach for every hourly interval, and the standard deviations of the errors are lower. We find

that the forecasting power generally improves later in the day, and this is something one would expect, as the model later is estimated with more data. While the forecasting power improves for the naïve approach also, it remains much lower than that of the model. Similarly, the one day ahead forecasting with the model performs much better than the naïve forecast.

forecast	time	Model		Naïve	
horizon		MSPE	sd	MSPE	sd
one day ahead		21.1	4.2	26.9	5.2
hour by hour	10:00	19.6	4.3	28.6	5.3
	11:00	18.7	4.2	29.3	5.4
	12:00	18.5	4.2	29.3	5.4
	13:00	19.8	4.3	28.1	5.3
	14:00	16.2	3.9	25.5	5.0
	15:00	14.7	3.7	23.8	4.8
	16:00	13.1	3.4	21.6	4.6

The mean square hourly prediction error (MSPE) and standard deviation of the prediction error for the best daily model, as well as a naïve approach predicting the previous TED value, averaged across all days and all assets. Each line indicates the MSPE for a single hour, starting from 10:00–11:00, where the model is estimated from all daily data up to the start of each period.

Table 4.2: Forecast accuracy

4.5.5 Liquidity drought extremes

A second application of this model that we present here is as a regulatory tool for the monitoring of liquidity. One would expect that regulatory bodies are interested in ensuring uninterrupted liquidity, as it is an integral part of a fair and orderly market. However, they would probably focus on the extreme liquidity levels that occur, and the durations of these extreme events.

For this application, instead of the conditional mean response of the observation variable, we now consider conditional quantiles of the response. That is, if a TED event occurs in a (stationary) LOB regime, given covariates $\boldsymbol{x}$, we can make a prediction about the 90th quantile of the response: This is the duration of time such that there is a 90% probability under the model that liquidity will return to the threshold level in this period.

To complete the analysis we plot the change in the conditional quantile level of the TED, as covariate $prevTEDavg$ is allowed to vary. The TED is defined as before, using the spread

as the liquidity measure and the median of the empirical distribution as the threshold. The conditional quantile function for a given quantile level u in the Lognormal case is given by

$$\begin{aligned} Q\left(u;\boldsymbol{x}_t\right) &= F^{-1}(\tau_i;\boldsymbol{x}_t,u) \\ &= \exp\left(\beta_0 + \sum_{s=1}^{p} x_t^{(s)}\beta_s + \sigma\Phi^{-1}(u)\right) \end{aligned}$$

We obtain the conditional quantile levels for the Lognormal distribution in Figure 4.11 and the Weibull distribution in Figure 4.12. The conditional quantile function in the Generalised Gamma case is given in Appendix A.

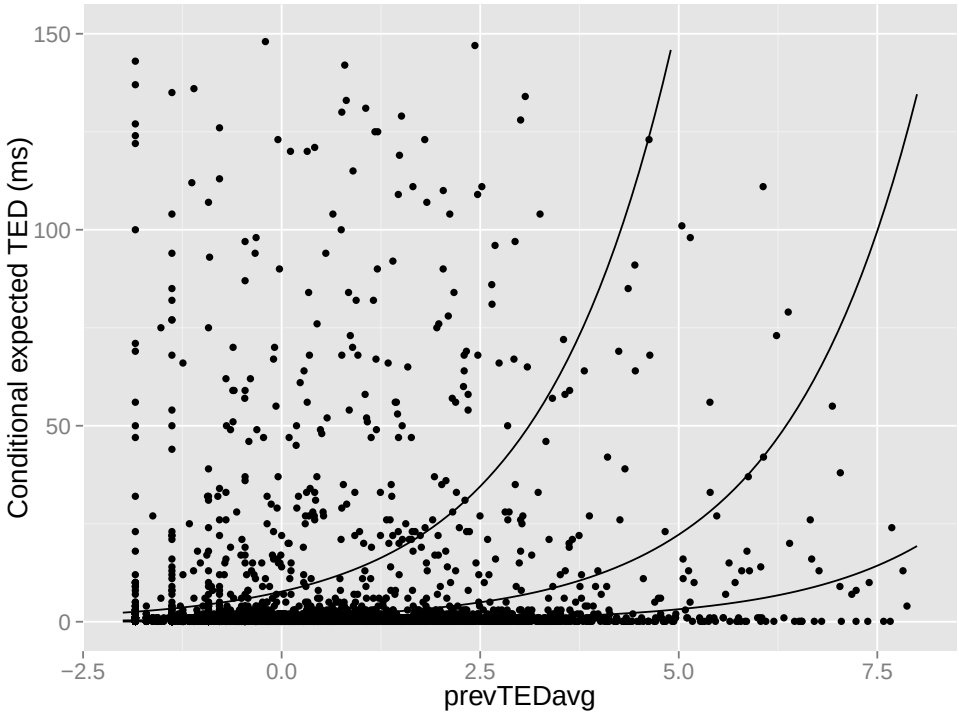


Figure 4.11: 25th (lower line), 50th (center) and 75th (upper) conditional quantile levels of the TED for different values of the prevTEDavg covariate for Credit Agricole on a single day. These are obtained using the quantile function of the Lognormal distribution. The dots are the realised value of the TED observations.

Such an analysis would be useful in understanding how extreme levels of particular covariates affect quantile levels of the TED. We can therefore understand how the different quantile surfaces for the TED behave for these extreme values of the spread and above. This enables regulators to identify which are the most important covariates associated with an increase in extreme periods of illiquidity (that is, where the liquidity measure remains above the threshold

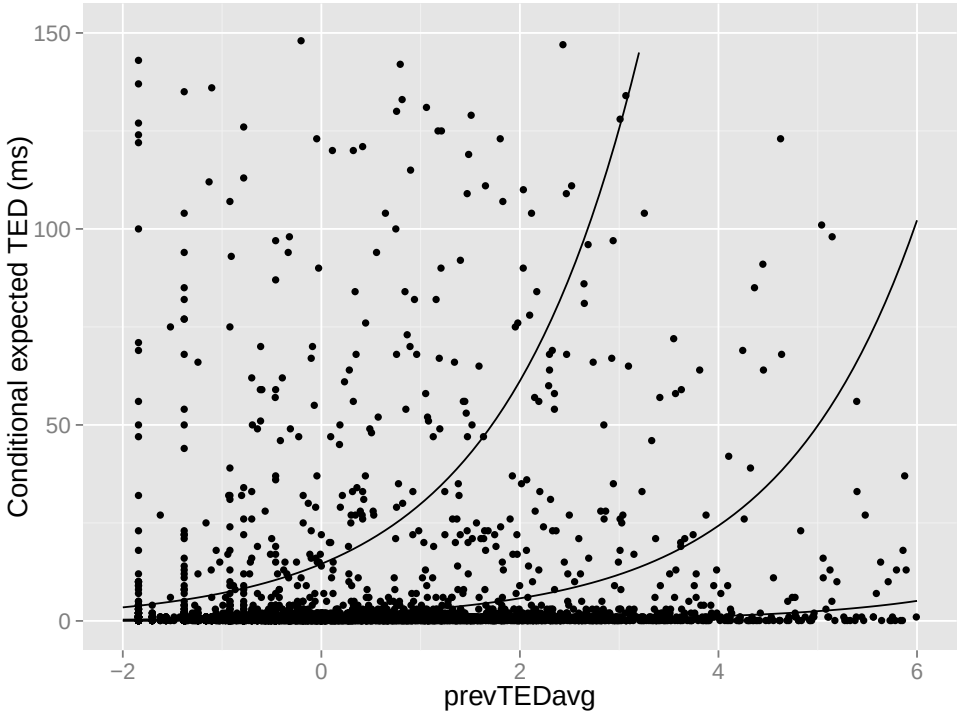


Figure 4.12: 25th (lower line), 50th (center) and 75th (upper) conditional quantile levels of the TED for different values of the prevTEDavg covariate for Credit Agricole on a single day. These are obtained using the quantile function of the Weibull distribution. The dots are the realised value of the TED observations.

for extended period of time). In addition, to the extent that a covariate taking extreme values is considered a scenario in which the LOB is stressed, a regulator can make inferences about the duration of relative illiquidity under such stressed conditions.

In non-stressed conditions, that is, where covariates take what would be considered to be 'normal' values, regulators may be interested in the range of probable values of the TED. Obtaining high quantile levels of the TED under the model could then help them identify situations which fall outside this range, which may be due to a change in the LOB regime or due to a particular event that will require their intervention.

Finally, one can use the model to help inform proposed regulation regarding interactions in the LOB, such as transaction taxes, minimum resting times or trade to quote ratios. To the extent that such regulation has an impact on particular covariates in the LOB, one can estimate the indirect effect it has on liquidity resilience, and therefore the ability of institutional investors and mutual funds to place orders without incurring a large execution cost.

4.5.6 Results for the XLM liquidity measure

In this section, we have thus far presented the results for the explanatory power of our model for the TED observations, when using the spread as the liquidity measure of choice. However, the framework we have proposed can incorporate any liquidity measure, and we have performed a limited set of experiments using the XLM, in order to briefly evaluate the performance of the model when using other measures. We chose to construct the XLM using a volume R=€25.000, as this is the standard volume used on the Xetra exchange. We again selected the thresholds as the median and 9th decile of the empirical distribution of observations of the XLM on every day.

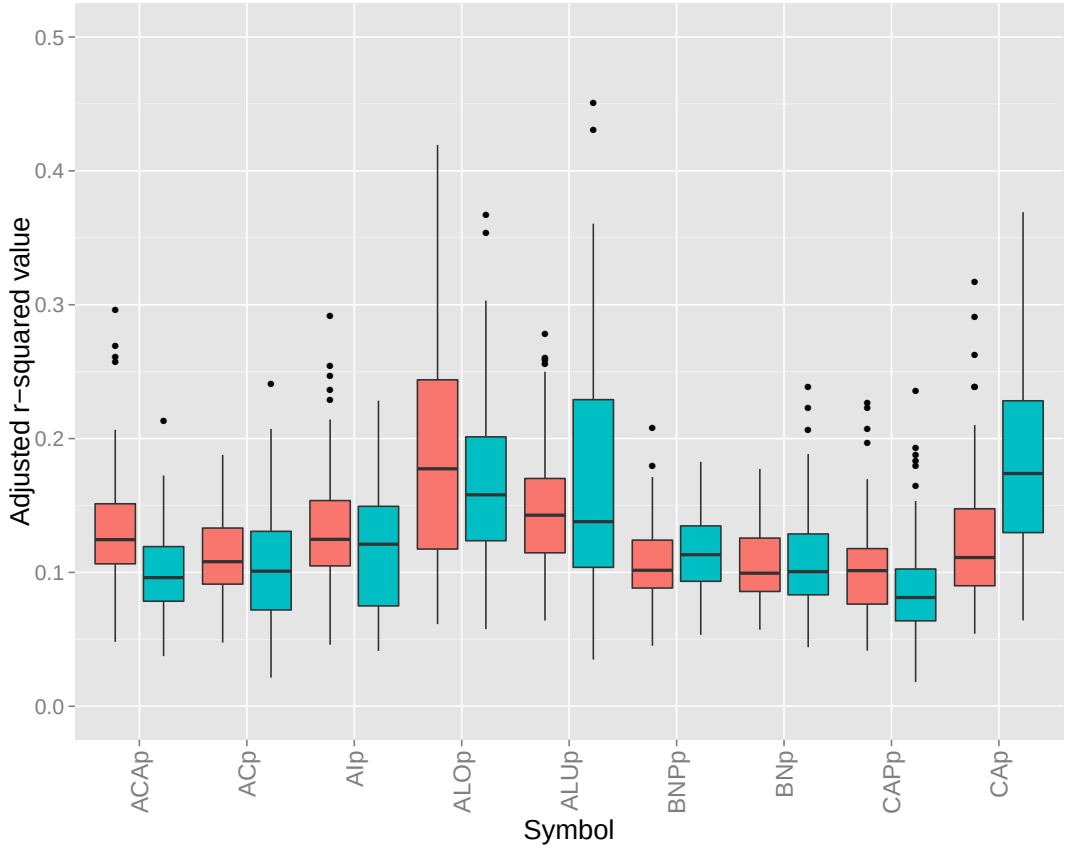


Figure 4.13: Boxplots of the adjusted R^2 value obtained from fitting the full survival model separately for each day in our dataset for both the threshold corresponding to the 5th decile of the XLM (the median - red) and the 9th decile threshold (blue), for a subset of assets in the CAC40.

With regard to the explanatory power of the model, Figure 4.13 shows that it is slightly lower than for the equivalent models using the spread as the liquidity measure in the TED definition. This is not unexpected, as the XLM is a much more informative liquidity measure, and

combines both spread and volume information at each level. We still find significant explanatory power of over 15% for some assets.

4.6 Discussion

We have defined formally the notion of liquidity resilience in terms of the duration of the deviation of a liquidity measure from a threshold level. We have demonstrated how a suitably defined survival regression framework can capture this notion, through consideration of the LOB structure at the start of this deviation. We have shown that such a model can be both interpretable and have good explanatory power in capturing resilience for different liquidity measures and threshold levels. In terms of its forecasting power for the duration of liquidity droughts, the model is also shown to be superior to naïve approaches, both in intra-day and in one day ahead prediction.

When we considered the inside spread as the liquidity measure in our model, we found that several covariates were both consistently selected in the estimation of the best fitting models in a range of model subspaces and were statistically significant at a level of 5%. They included instantaneous and lagged values of the spread, the average of previous TED observations, the number of previous threshold exceedance observations in the previous second, and the origin of the exceedance (cancellation or market order). While these covariates were chosen by our model selection process to be part of the best fitting daily models most frequently, we found that the covariate coefficients generally varied over our 82 day trading sample. This illustrates that liquidity (and liquidity resilience) is not stationary inter-daily.

A possible application area for our model would be in the field of optimal order execution. Gomber et al. [2011] have found that, for the submission of large orders, the timing is dependent on the liquidity level. Our model could thus help in estimating the time for a particular liquidity measure to reach the level required for an order submission. The model could also be helpful from a regulation perspective, in estimating how long extreme periods of illiquidity in the LOB are likely to last, by calculating (high) conditional quantile levels under the model, and we have presented an example of such an approach in this chapter.

In the next section we will introduce liquidity resilience profiles, which summarise the resilience behaviour across thresholds, given assumptions regarding market conditions. This allows for the comparison of resilience across assets in normal and stressed LOB scenarios and could inform a brokerage firm about the potential risk of trading an asset in certain conditions.

4.7 Additional figures and tables

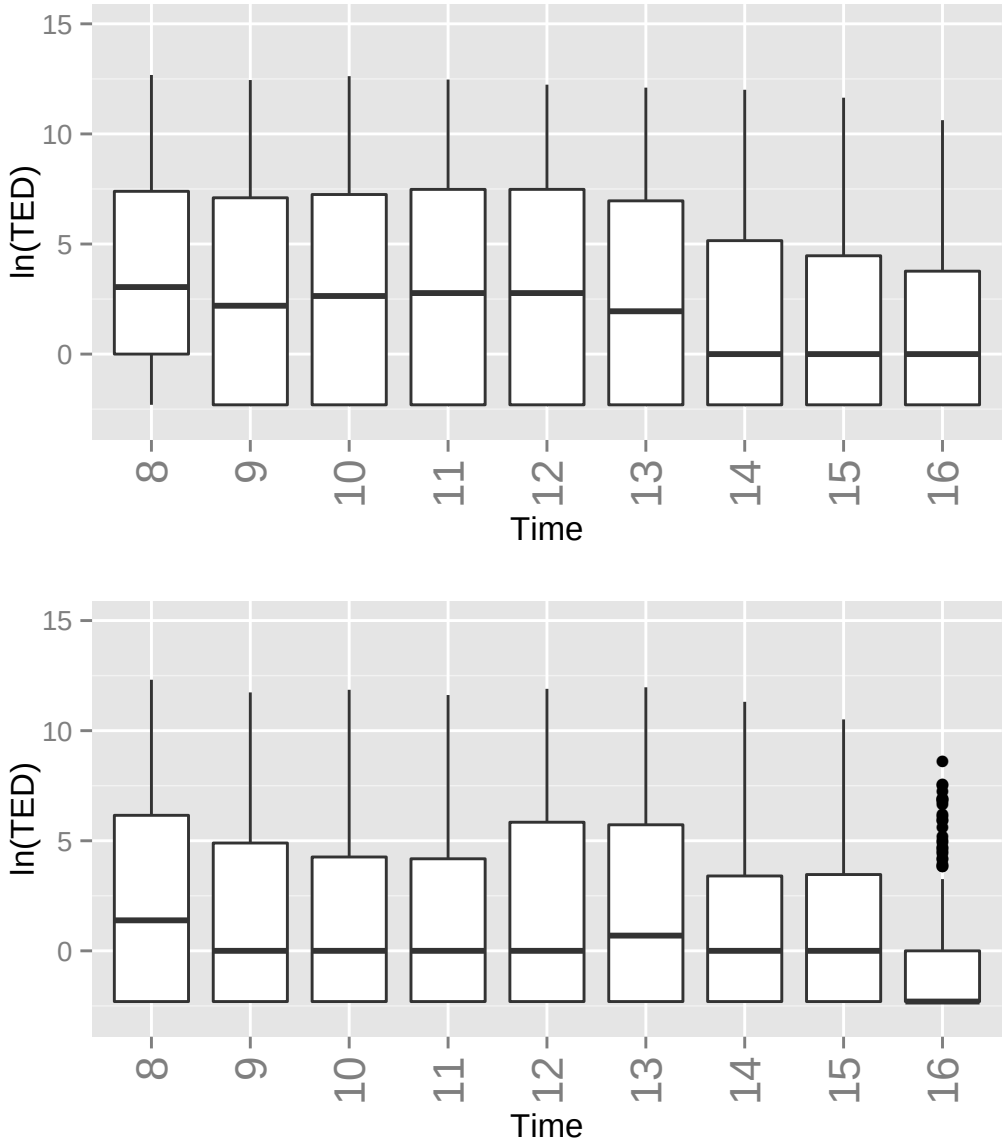


Figure 4.14: The duration of TED observations above the median (top) and 9th decile (bottom) spread value throughout the trading day for stock Sanofi for the illiquidity measure $\mathcal{M}$ representing the inside spread.

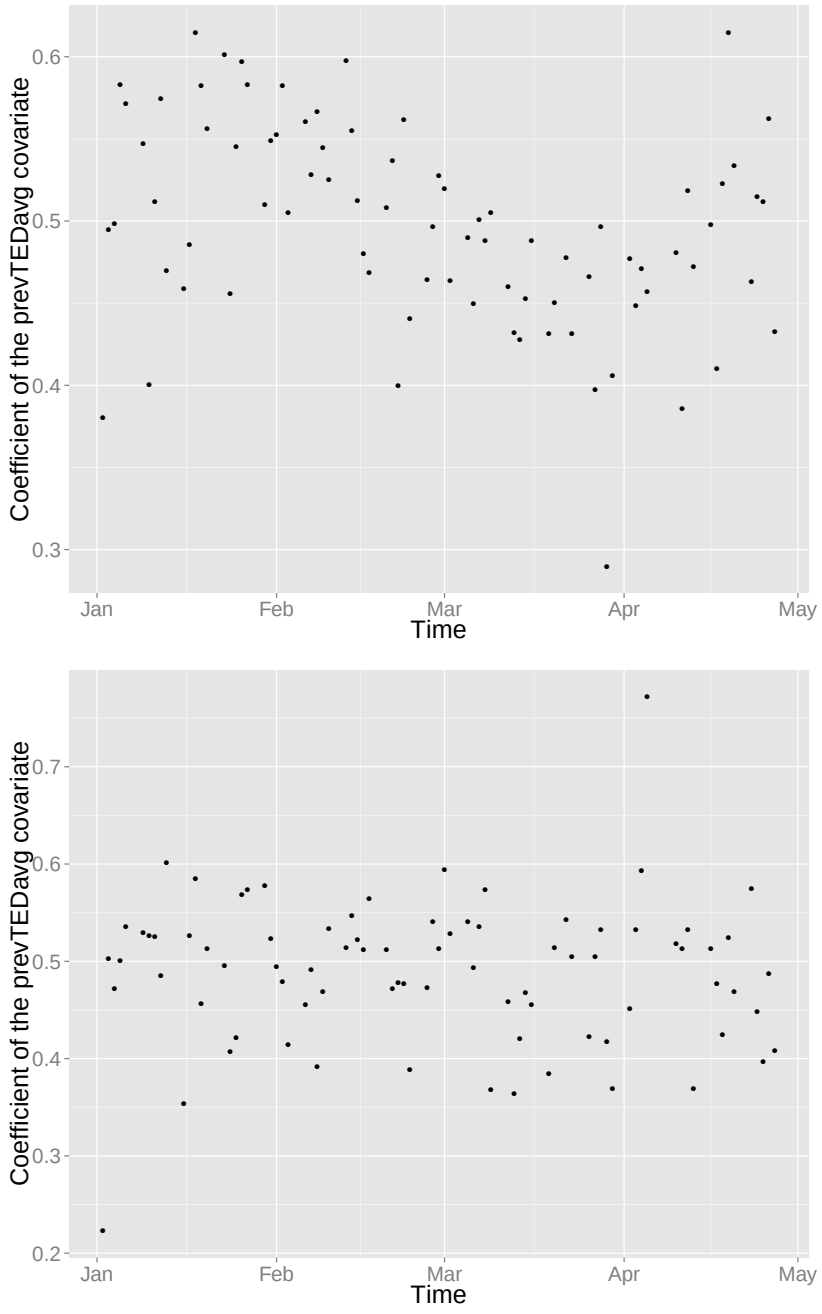


Figure 4.15: Coefficients of the prevTEDavg covariate, for every fitted daily model for stock Credit Agricole. The top graph is obtained using thresholds corresponding to the median spread, while the bottom graph uses thresholds corresponding to the 9th decile spread.

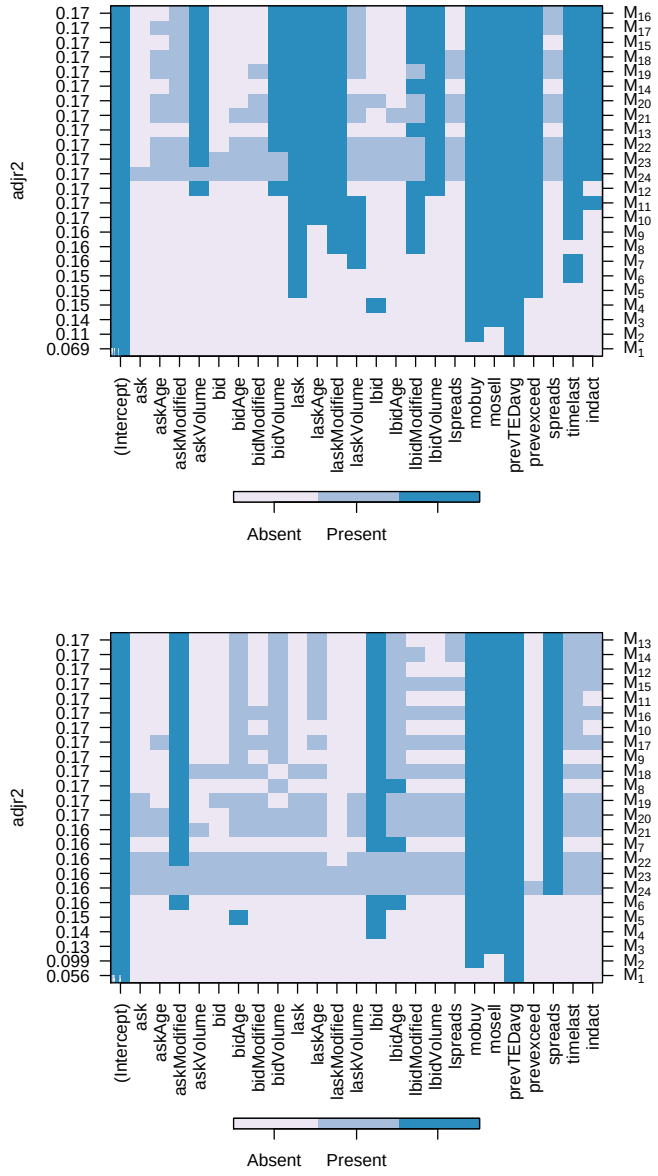


Figure 4.16: The adjusted R^2 values for models of using the best subsets of covariates (of size 1 to 24, in this case) for a single trading day (the 17th of January 2012) for stock Sanofi in the lognormal specification and the median spread (top) or the 9th decile (bottom) as the threshold.

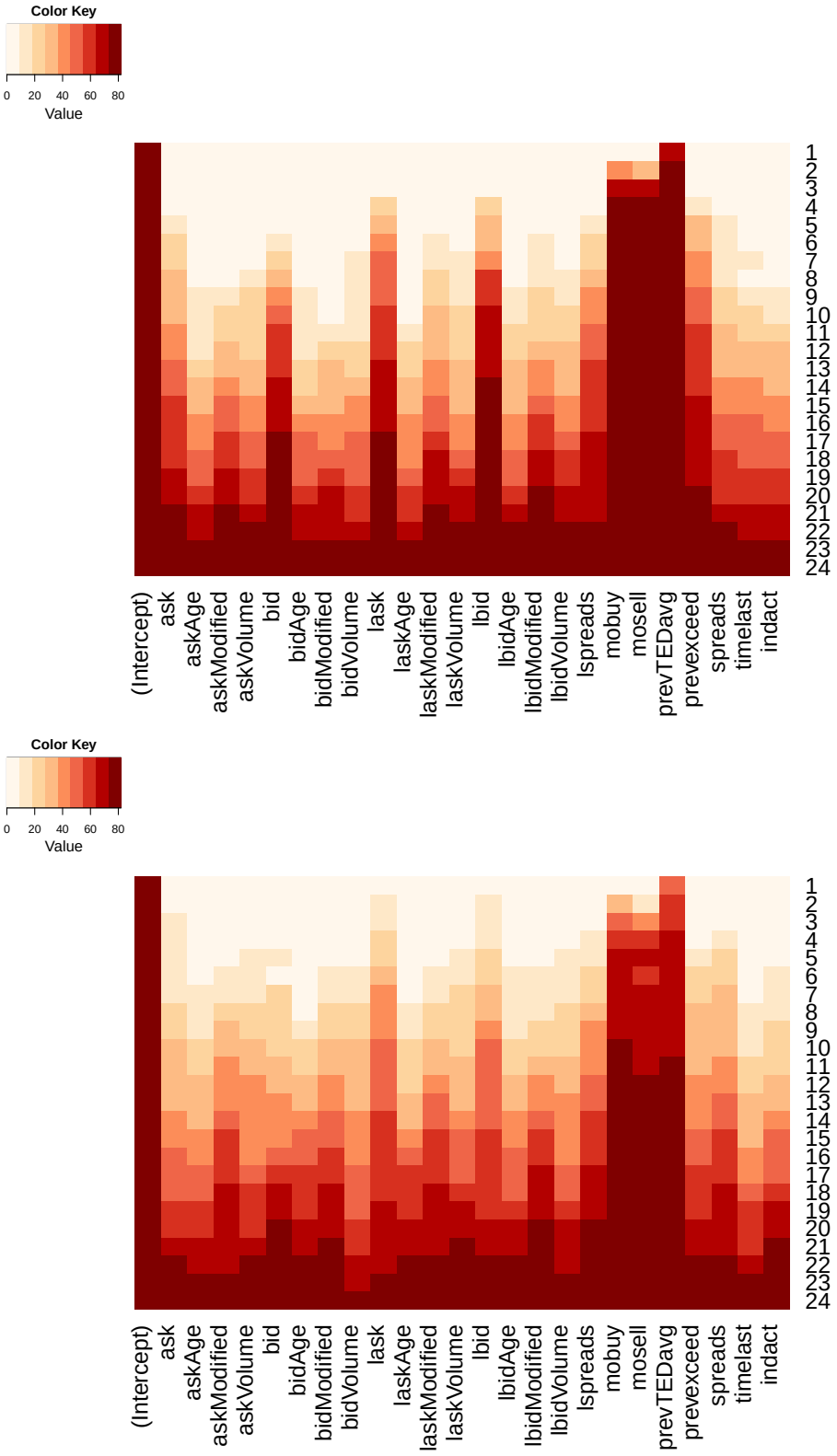


Figure 4.17: Heatmap of the relative frequency with which parameters appear in the best daily models of every subspace (frequency in terms of the number of daily models over the 82 day period) for the Sanofi dataset, using the daily median (top) or the 9th decile (bottom) of the spread as the threshold value.

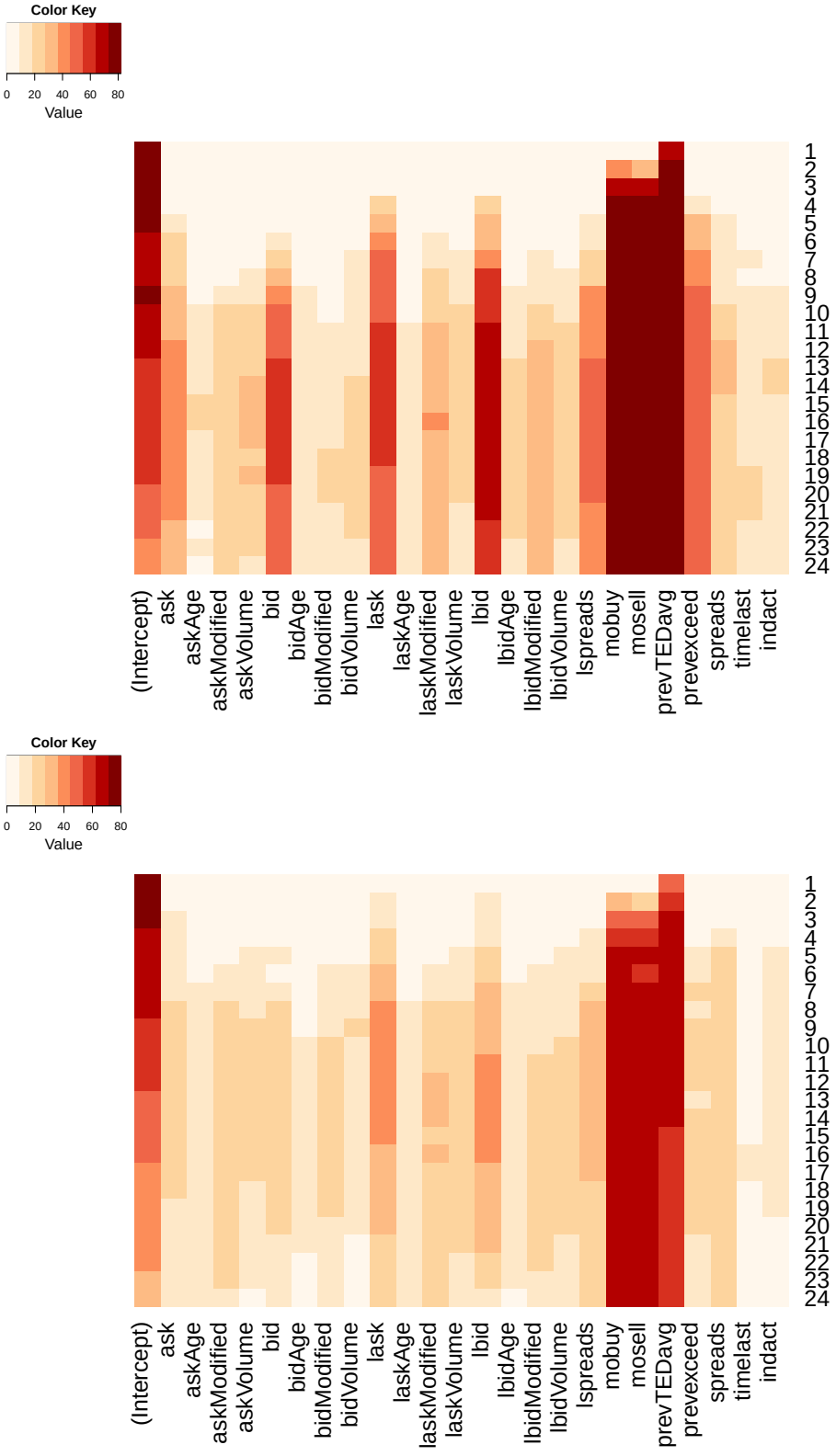


Figure 4.18: Heatmap of the relative frequency with which the parameters are found to be significant at the 5% level (frequency in terms of the number of daily models over the 82 day period) for the Sanofi dataset, using the median and 9th decile thresholds.

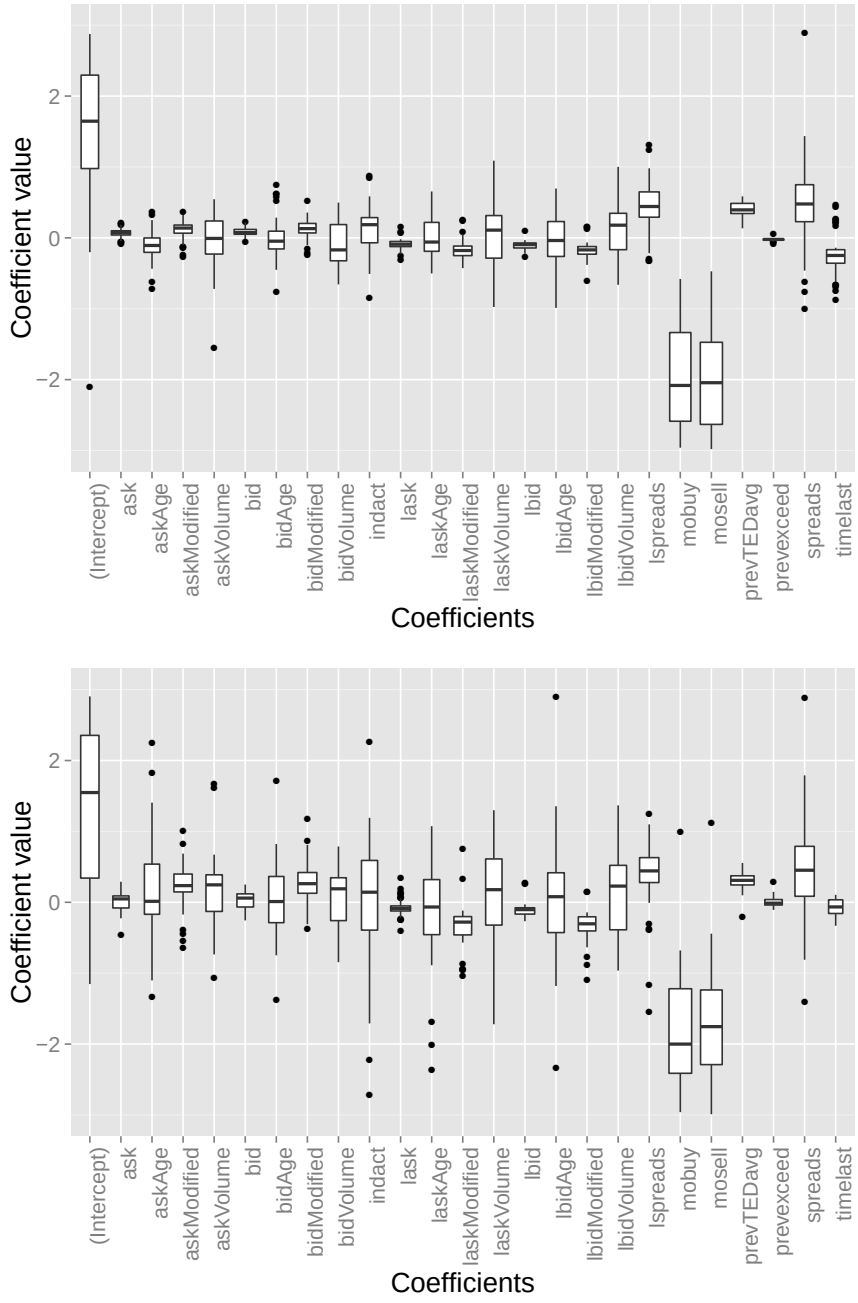


Figure 4.19: Coefficients of the best models of any size (in terms of the adjusted R^2 values) for Sanofi for median threshold spread exceedances (top) and 9th decile spread exceedances (bottom).

	* Mean #TEDs		* Min #TEDs	
	5th decile	9th decile	5th decile	9th decile
ACAp	5465	1963	1125	327
ACp	1702	535	846	210
AIp	3115	1213	131	50
ALOp	2642	1097	621	259
ALUp	1165	678	341	91
BNPp	11279	3331	3917	809
BNp	2786	729	665	124
CAPp	2098	627	431	270
CAp	1946	578	792	158
CSp	4027	1250	1167	157
DGp	2685	889	1275	267
EADp	1853	572	542	190
EDFp	1163	327	383	122
EIp	1214	424	276	47
ENp	1848	489	805	136
FPp	6086	1716	2123	426
FTEp	2167	282	352	63
GLEp	6279	2238	1189	531
GSZp	2387	575	589	92
KNp	1337	572	180	74
LGp	2895	935	589	318
MCp	1323	431	208	95
MLp	2673	952	1364	320
MTa	3802	530	1703	127
ORp	2600	1009	1201	346
PPp	563	165	250	19
PUBp	2923	679	1115	218
RIp	1751	664	838	208
RNOp	3106	1402	967	370
SANp	2325	678	756	132
SEVp	814	276	293	50
SGOp	2844	1039	1407	415
STMp	2007	818	561	178
SUp	6321	1837	2178	218
TECp	2252	868	935	316
UGp	2836	835	345	137
ULp	796	222	300	48
VIEp	1697	622	605	206
VIVp	1951	605	369	60
VKp	2251	702	963	158

Table 4.3: The mean and minimum number of TED observations for each CAC40 stock in the 82-day period under consideration in this study.

Chapter 5

Liquidity and resilience commonality

In this chapter, we first revisit the literature in liquidity commonality, which quantifies commonality through the explanatory power of the principal components of liquidity measures in a regression. We show that at least in the equity space, the assumption that one can capture all the features of liquidity commonality via a PCA regression approach will not always be appropriate. The outcome of using PCA methods, which are based on second moments, is that the analysis is driven by the most illiquid assets, which act as outliers in the cross-sectional dataset. We therefore utilise ICA (Independent Component Analysis) methods, which address this issue by incorporating higher order information, in order to first determine the assets which exhibit these heavy tailed features.

We then propose a model to quantify the commonality in the resilience of liquidity, an aspect which is not reflected in previous commonality approaches. Using the liquidity definition in the previous chapter, we construct a curve of the expected TEDs as a function of the threshold, which we will term the Liquidity Resilience Profile (LRP). Since LRPs are informative about the level of LOB liquidity replenishment for each asset, a commonality analysis can identify clusters of assets for which we would expect a swift return to a high liquidity levels after a shock.

The market factors contributing to the variation in daily liquidity resilience for 82 European stocks are then obtained through a functional principal component analysis (FPCA) of the LRP curves. We demonstrate that there is a consistency in the shape of the first three functional principal components (FPCs) over time and then regress the LRPs for individual assets against these FPCs, interpreting their explanatory power as a measure of commonality in the liquidity resilience profile of the given asset with the market factors. For the equities dataset under consideration, we found that the first 3 FPCs could explain between 10 and 40% of the variation in liquidity resilience at low liquidity thresholds. However, at more extreme liquidity exceedance

thresholds, the commonality between individual asset liquidity resilience profile behaviour and market factors diminishes significantly, and individual asset factors take effect.

5.1 Introduction to component analysis and dimensionality reduction

In the empirical analysis of multivariate data, there are various statistical procedures used to analyse the different variables, often with the purpose of compressing the data, by finding a set of variables which account for most of the information in the original variables. These new variables are usually a combination of the original variables, and are chosen so as to satisfy certain properties, which are different for every technique. This section introduces two well-known dimensionality reduction techniques, Principal Components Analysis (PCA) and Independent Components Analysis (ICA).

In this context, let us first introduce the concept of Singular Value Decomposition (SVD).

Definition 5.1.1. *The singular value decomposition (SVD) of an $N \times p$ matrix $\boldsymbol{X}$ is a factorisation*

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{5.1}$$

where $\boldsymbol{U}$ is an $N \times N$ orthogonal matrix, $\boldsymbol{\Sigma}$ is an $N \times p$ diagonal matrix, $\boldsymbol{V}^T$ is a $p \times p$ orthogonal matrix.

There is an intimate relationship between SVD and eigendecomposition, or spectral decomposition, which is used in PCA.

5.1.1 Principal Components Analysis

PCA was introduced by Pearson [1901], and independently by Hotelling [1933] in the more familiar approach in which it is presented here. The goal is to find linear combinations of the existing variables which are i) uncorrelated and ii) have maximal variance, and we will make this precise in the following. Let $\boldsymbol{Z} \in \mathbb{R}^{N \times p}$ be the original data matrix from which we subtract the column mean from every element to obtain $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & & x_{2p} \\ \dots & \dots & \dots \\ x_{N1} & \dots & x_{Np} \end{bmatrix} \tag{5.2}$$

where the rows of the matrix are the N observations of $\boldsymbol{x}_i = [x_{i1}, \ldots, x_{ip}]$ and the columns $\boldsymbol{x}_{(j)} = [x_{1j}, \ldots, x_{Nj}]'$ are the variables, or 'features' of the data, which PCA will try to project onto a different basis. In the transformed data

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W} \tag{5.3}$$

$\mathbb{V}(\boldsymbol{y}_{(1)})$ is maximised under the constraint $\left|\left|\boldsymbol{w}_{(1)}\right|\right| = 1$, and for each $j > 1$, $\mathbb{V}(\boldsymbol{y}_{(j)})$ is maximised under the constraint $\left|\left|\boldsymbol{w}_{(j)}\right|\right| = 1$ and the additional constraint $Cov(\boldsymbol{y}_{(i)}, \boldsymbol{y}_{(j)}) = 0, i = 1, \ldots, j-1$.

Let us denote by $\boldsymbol{S}_{\boldsymbol{X}} = \frac{1}{p-1}\boldsymbol{X}^T\boldsymbol{X}$ the covariance matrix of $\boldsymbol{X}$. The problem above can be restated as obtaining $\boldsymbol{W}$, such that the covariance matrix of $\boldsymbol{Y}$ is a diagonal matrix. In this case one can use eigendecomposition, or spectral decomposition, using the following theorem:

Theorem 5.1.2. *Spectral theorem. Any real square symmetric matrix $\boldsymbol{Z}$ can be decomposed as*

$$\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \tag{5.4}$$

where $\boldsymbol{U}$ is a matrix whose columns are the eigenvectors of $\boldsymbol{Z}$ and $\boldsymbol{D}$ is a diagonal matrix whose entries are the eigenvalues of $\boldsymbol{Z}$.

$\boldsymbol{S}_{\boldsymbol{X}}$ is a square positive semi-definite symmetric matrix. By the spectral theorem, it can be decomposed as

$$\boldsymbol{S}_{\boldsymbol{X}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \tag{5.5}$$

where $\boldsymbol{D}$ is a diagonal matrix whose entries are the eigenvalues $\lambda_1, \ldots, \lambda_p$ of $\boldsymbol{S}_{\boldsymbol{X}}$. These eigenvalues are non-negative, since $\boldsymbol{S}_{\boldsymbol{X}}$ is positive semi-definite. $\boldsymbol{U}$ is a matrix of the corresponding eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ and thus

$$\boldsymbol{S}_{\boldsymbol{X}}\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i, i = 1, \ldots, p \tag{5.6}$$

The trace $tr(\boldsymbol{S}_{\boldsymbol{X}})$ is equal to the sum of the variances, and this is also equal to the sum of the eigenvalues $\sum \lambda_i$. Then the eigenvector $\boldsymbol{u}_i$ corresponding to λ_i will explain a proportion

$$\frac{\lambda_i}{tr(\boldsymbol{S}_{\boldsymbol{X}})}$$

of the variation $\boldsymbol{X}$.

Let us assume that the eigenvalues are arranged so that $\lambda_1 > \lambda_2 > \ldots > \lambda_p$, and thus the highest variation in the data will be in direction $\boldsymbol{u}_1$. We can see that if we select $\boldsymbol{W} = \boldsymbol{U}$, i.e.

the matrix of eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$ in Equation 5.3, then this makes the covariance matrix of $\boldsymbol{Y}$ diagonal:

$$\begin{aligned}\boldsymbol{S_Y} &= \frac{1}{p-1}\boldsymbol{Y}^T\boldsymbol{Y} \\ &= \frac{1}{p-1}(\boldsymbol{XW})^T\boldsymbol{XW} \\ &= \frac{1}{p-1}\boldsymbol{W}^T\boldsymbol{X}^T\boldsymbol{XW} \\ &= \boldsymbol{W}^T\boldsymbol{S_X}\boldsymbol{W} \\ &= \boldsymbol{W}^T\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}^T\boldsymbol{W} \\ &= \boldsymbol{D}.\end{aligned}$$

These eigenvectors are called the principal components of $\boldsymbol{X}$. In practice, the principal components can be computed using SVD. Letting

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{5.7}$$

we can obtain the covariance as

$$\begin{aligned}\boldsymbol{S_X} &= \frac{\boldsymbol{X}^T\boldsymbol{X}}{p-1} \\ &= \frac{(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)}{p-1} \\ &= \frac{\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T}{p-1} \\ &= \frac{\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T}{p-1}\end{aligned}$$

and because $\boldsymbol{S_X}$ is symmetric,

$$\boldsymbol{S_X} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T. \tag{5.8}$$

Comparing this to the eigendecomposition, we find the correspondance that the eigenvectors of $\boldsymbol{S_X}$ are the same as $\boldsymbol{V}$. In addition, the eigenvalues of $\boldsymbol{S_X}$ can be obtained from the *singular values* $\sigma_i = \boldsymbol{\Sigma}_{i,i}, i = 1, \ldots, p$ as $\lambda_i = \frac{\sigma_i^2}{p-1}$.

The outcome of the PCA procedure is a number of these eigenvectors and the transformation of the data $\boldsymbol{Y}$. Since the objective of PCA is to reduce the dimensionality of the data, one would usually select a subset of these eigenvectors $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ such that the proportion of variation explained

$$\sum_{i=1}^{k} \frac{\lambda_1}{tr(\boldsymbol{S_X})} > v$$

where typical values of v are 0.8, 0.9 etc.

5.1.2 Independent Components Analysis

The objective of ICA is to find a linear transformation of the observed data into a set of components which are independent from each other. It was originally introduced for the purpose of blind source separation, i.e. identifying the (independent) unknown signals from a number of mixed signals, without knowledge regarding the mixing coefficients. It differs from other methods in that it searches for nongaussian components [Hyvärinen et al., 2004].

Let $\boldsymbol{X} \in \mathbb{R}^{N \times p}$, constructed from vector observations $\boldsymbol{x}$ be our data matrix as above. Then we can formulate the problem by assuming that $\boldsymbol{x}$ is generated according to

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} \tag{5.9}$$

where $\boldsymbol{y}$ is a vector whose components are assumed mutually independent, and in this exposition we will also assume they are identically distributed for simplicity. Then ICA aims to find the matrix $\boldsymbol{A}$, or more precisely, its inverse $\boldsymbol{A}^{-1}$ [Hyvarinen, 1999]. This assumed data-generating model is well-defined if the components $\boldsymbol{y}_i$ are non-gaussian [Hyvärinen et al., 2004]. The original components could then be obtained as

$$\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{x} \tag{5.10}$$

In practice, because we do not know $\boldsymbol{A}$, we need to find an appropriate estimator. Let u be a combination of the original components

$$u = \boldsymbol{b}^T\boldsymbol{x} = \boldsymbol{b}^T\boldsymbol{A}\boldsymbol{y} = \boldsymbol{q}^T\boldsymbol{y} \tag{5.11}$$

If $\boldsymbol{b}$ is one of the rows of $\boldsymbol{A}^{-1}$, then u would be equal to one of the independent components y_i, and all elements of $\boldsymbol{q}$ would be equal to 0, with a single 1.

With this formulation we observe that u is a linear combination of independent (and identically distributed) components $y_1, \ldots, y_p$. Because of the Central Limit Theorem this combination usually becomes 'more Gaussian' than the y_i and least Gaussian, when it equals one of the independent components y_i. The goal of ICA can then be stated as maximising nongaussianity, and we will refer to measures of nongaussianity in the following.

5.1.3 ICA procedure

Practically, ICA attempts to find components that are statistically 'as independent from each other as possible' [Hyvarinen, 1999], which can be achieved by minimising the mutual information of the transformed components, and we present such as procedure here. As before, we assume that the each column X_i has mean zero, and we additionally assume that it has been

'whitened', so that its variance is 1. We describe a simple whitening process here, starting with the previous decomposition of the covariance matrix

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \tag{5.12}$$

to obtain the orthogonal matrix of eigenvectors $\boldsymbol{U}$ and the diagonal matrix of eigenvalues $\boldsymbol{U}$. Using a linear whitening operator $\boldsymbol{V}$

$$\boldsymbol{V} = \boldsymbol{U}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}^T \tag{5.13}$$

one can obtain the whitened data

$$\boldsymbol{z} = \boldsymbol{V}\boldsymbol{x}. \tag{5.14}$$

ICA is not the same as whitening, however, as uncorrelatedness is a weaker condition than independence. Whitening can only give the ICs up to an orthogonal transformation (p.160, Hyvärinen et al. [2004]). It is usually performed as a pre-processing step in ICA, however, as it reduces the complexity of the problem as one only needs to estimate an orthogonal mixing with fewer degrees of freedom.

We stated in Section 5.1.2 that the original components are assumed nongaussian. This is indeed a requirement for ICA, as uncorrelated (perhaps as a result of whitening) jointly gaussian variables are independent, and therefore would not contain any information about the mixing matrix $\boldsymbol{A}$. Nongaussianity therefore becomes an objective in estimating the ICA model.

For this purpose, we first introduce the notion of differential entropy H [Comon, 1994, Hyvarinen, 1999] of a random vector $\boldsymbol{u}$

$$H(\boldsymbol{u}) = -\int f(\boldsymbol{u}) \log(f(\boldsymbol{u}))\mathrm{d}\boldsymbol{u}. \tag{5.15}$$

Amongst variables of equal variance, gaussian variables have the largest differential entropy [Hyvärinen et al., 2004]. We therefore introduce the negentropy J, as a measure of nongaussianity [Comon, 1994]

$$J(\boldsymbol{u}) = H(\boldsymbol{v}) - H(\boldsymbol{u}) \tag{5.16}$$

where $\boldsymbol{v}$ is a Gaussian random vector with the same covariance vector as $\boldsymbol{u}$. Negentropy is non-negative and only becomes 0 when $\boldsymbol{u}$ is Gaussian.

$J(\boldsymbol{u})$ can be approximated by [Hyvarinen, 1999]

$$J(u_i) = c\left[E\left\{G(u_i)\right\} - E\left\{G(v)\right\}\right]^2 \tag{5.17}$$

where there are a number of options for the 'contrast function' $G(\cdot)$, such as

$$G_1(u) = \frac{1}{a_1} \log \cosh(a_1 u), \tag{5.18}$$

$$G_2(u) = -\exp\left(-\frac{u^2}{2}\right). \tag{5.19}$$

For a chosen contrast function G, $J_G(\cdot)$ can be considered as the objective function in the optimisation, i.e. for the first component $u_i = \boldsymbol{b}^T \boldsymbol{x}$ we have

$$\max \sum_{i=1}^{p} J_G(\boldsymbol{b}_i) \text{ wrt } \boldsymbol{b}_i, i = 1, \ldots, p \tag{5.20}$$

$$\text{s.t. } E\left\{(\boldsymbol{b}_j^T \boldsymbol{x})(\boldsymbol{b}_k^T \boldsymbol{x})\right\} = \delta_{jk}. \tag{5.21}$$

As formulated above, we can now approach this with a number of optimisation algorithms, and Hyvärinen et al. [2004] (p.185) provides an example with both the gradient descent method and the FastICA algorithm used in this chapter.

5.1.4 ICA component selection

In contrast to PCA, in ICA one cannot calculate the variances of individual components. Because in the assumed model in Equation 5.9 both $\boldsymbol{A}$ and $\boldsymbol{y}$ are unknowns, a multiplier in one of the columns of $\boldsymbol{A}$ could be cancelled by dividing the corresponding element of $\boldsymbol{y}$ by the same amount. There is therefore no equivalent way as in PCA to select a subset of the components which explain the majority in the variation.

When the problem is not one of blind source separation, where one aims to identify the original sources, but rather one of dimensionality reduction (as is the case in this chapter), one can aim to find 'interesting' projections of the multidimensional data. In the case of ICA, these projections are precisely those which maximise the negentropy approximations above. In this case, ICA is very closely related to another technique called projection pursuit, see Girolami and Fyfe [1996].

5.1.5 Implications of using PCA and ICA for data coming from different distributions

We will now demonstrate the effect of performing PCA and ICA and datasets that are generated so that they best demonstrate the difference between the two techniques. In Figure 5.1 we show heatmaps of data coming from a combination of Gaussian distributions and Student t distributions. For both we have generated 10000 points and for the Gaussian case we have for the original sources

$$\boldsymbol{s}_{(1)} \sim \mathcal{N}(10, 5)$$
$$\boldsymbol{s}_{(2)} \sim \mathcal{N}(5, 1)$$

and for the observed data

$$\boldsymbol{z}_{(1)} = 0.5\boldsymbol{s}_{(1)} + 0.5\boldsymbol{s}_{(2)}$$
$$\boldsymbol{z}_{(2)} = 0.2\boldsymbol{s}_{(1)} + 0.8\boldsymbol{s}_{(2)}$$

For the Student t case

$$\boldsymbol{s}_{(1)} \sim St(3)$$
$$\boldsymbol{s}_{(2)} \sim St(3)$$

and for the observed data

$$\boldsymbol{z}_{(1)} = 0.1\boldsymbol{s}_{(1)} + 0.9\boldsymbol{s}_{(2)}$$
$$\boldsymbol{z}_{(2)} = 0.8\boldsymbol{s}_{(1)} + 0.2\boldsymbol{s}_{(2)}$$

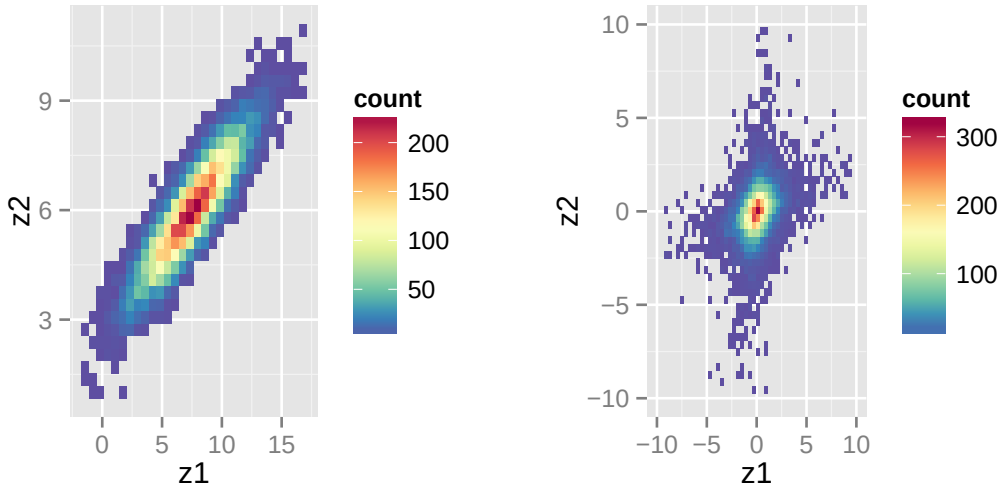


Figure 5.1: A heatmap of a 2-dimensional dataset, coming from linear combinations of Gaussian data (left) and Student's t distribution (right).

In Figure 5.2 we observe the effect of performing PCA and ICA on the Gaussian data. We note that PCA uncovers the directions of maximal variance, which do not necessarily correspond to the original sources. ICA cannot calculate variances of individual components, and

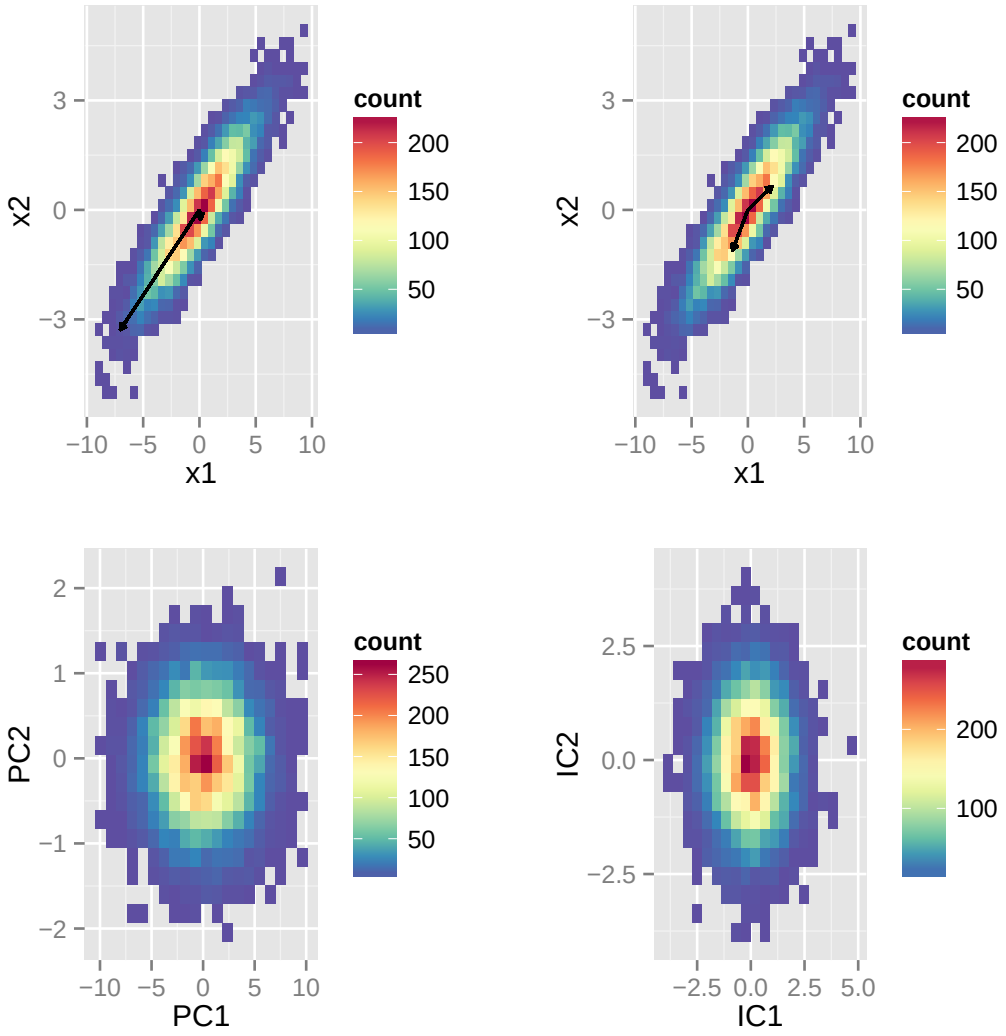


Figure 5.2: (Top): The Gaussian data in Figure 5.1, and the direction of maximal variance found by PCA (left) and the direction of maximal non-Gaussianity found by ICA (right). (Bottom) The corresponding transformed data. Note that PCA uncovers the directions of maximal variance, and thus the range of the x-axis in the left figure is different from the right.

thus only has the effect of whitening the data. We observe that ICA cannot be used for the purpose of separating Gaussian components.

In Figure 5.3 we observe the effect of performing PCA and ICA on the Student t data. PCA searches for directions of maximal variance that are orthogonal to each other, while ICA searches for the directions of maximum non-Gaussianity, and can thus uncover the original components. ICA is therefore much more effective than PCA at separating out the non-Gaussian components.

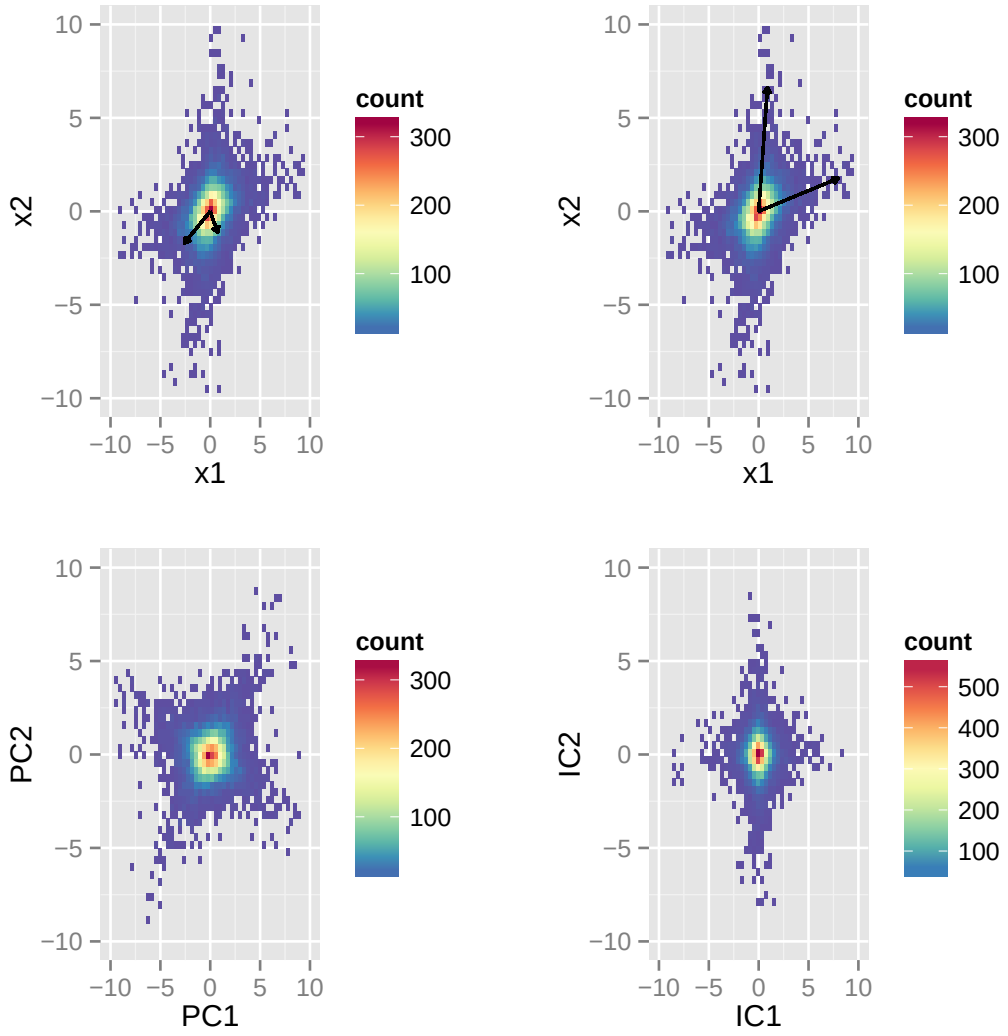


Figure 5.3: The Student t data in Figure 5.1 with the directions uncovered by PCA (left) and ICA(right), and the corresponding transformed data (bottom).

5.1.6 PCA and ICA regression

While PCA and ICA can be useful as standalone statistical techniques, they are also used in conjuction with other techniques. For example, PCA has been used extensively for regression analysis, particularly when data is multicollinear [Jolliffe, 2005]. The existence of multicollinearity in a set of data will have adverse effects for multiple linear regression, namely making the regression estimators unstable. Multicollinearity can exist due to very high correlation between two or more variables, but Mansfield and Helms [1982] provide examples to show that it may exist even when this is not the case.

The following description of PCA regression, will be based on the exposition of Jolliffe

[2005] (p. 167). Let us consider the standard regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5.22}$$

We can rewrite $\boldsymbol{X}\boldsymbol{\beta}$ as $\boldsymbol{X}\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{\beta}$, as $\boldsymbol{W}$ is an orthogonal matrix. Therefore we can rewrite Equation 5.22 as

$$\boldsymbol{y} = \boldsymbol{Y}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{5.23}$$

since $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}$, and where we denote $\boldsymbol{\gamma} = \boldsymbol{W}^T\boldsymbol{\beta}$.

Estimating γ using least squares, one can then obtain β as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{W}\hat{\boldsymbol{\gamma}} \tag{5.24}$$

PCA identifies the directions of maximal variance, and thus one can select a subset of the transformed variables which account for the majority of the variance in the data. We can therefore use those in the reduced regression model

$$\boldsymbol{y} = \boldsymbol{Y}_m\boldsymbol{\gamma}_m + \boldsymbol{\varepsilon}_m \tag{5.25}$$

where $\boldsymbol{Y}_m$ is the subset of the original variables of highest variance. Jolliffe [2005] explains how multicollinearities appear as PCs with very small variances, and thus very small corresponding eigenvalues in the covariance matrix.

PCA can help overcome multicollinearity issues by identifying such PCs to be omitted from the regression. However, there are also some caveats, as the process also introduces some bias, and for a further discussion regarding selecting components in PCA regression [Jolliffe, 2005]. In addition, since the variables in the regression have been transformed, it may now not be as straightforward to interpret the results.

5.2 Liquidity commonality in a secondary market (Chi-X): PCA, ICA and regression

Examples of the PCA regression approach outlined in Section 5.1.6 for the purposes of quantifying liquidity commonality have already been described in Chapter 3.3.2 (p. 60), and we revisit this approach for a selection of 82 of the most liquid stocks on Chi-X, from 3 countries (UK, France and Germany) and 10 different industries. We provide further information regarding these stocks in Tables 5.1 and 5.2. In order to perform such an analysis, one has to ensure

that the liquidity measurements for all assets are at regular intervals throughout the trading day. Since the LOB for each asset is an event-driven stochastic process, we need to sample the process to obtain evenly-spaced observations of liquidity, in order to perform a PCA analysis. We thus perform a pre-processing step, first constructing the LOB for each asset and then sampling to obtain liquidity measurements aligned at 1 second intervals. We thus obtain measurements of the spread and XLM every second, for all 82 assets, throughout the 4 month period under consideration.

We then perform PCA on the liquidity data every day, and regress the liquidity of individual assets against the first 3 PCs (which we consider to be the market factors). This is similar to the analysis undertaken, e.g. in Mancini et al. [2013] and other studies discussed in Section 3.3.2, and allows us to investigate temporal commonality in liquidity throughout a day. Performing a daily regression (rather than a regression over the entire period as in Mancini et al. [2013]) enables us to assess the fit over time and identify features that would otherwise be lost through time averaging and smoothing the signal.

Figures 5.4 and 5.5 show the R^2 scores of the regression for every asset on a randomly selected day, where the assets are broken down by country and by sector respectively. For most assets for both the spread and the XLM, the R^2 score is around 25%, although we notice that there are particular assets (e.g. NEXp, WCHd) which have very high R^2 scores.

Similar results are observed consistently throughout the 4 month period. A summary of the explanatory power of the PCA regression for every asset over time is provided in Figure 5.7. The high R^2 scores for certain assets imply that the first few 'market' PCs for the liquidity measures, which are obtained from the cross-section of all assets on a given day, essentially mirror the liquidity of these few particular assets. That is, the PCA and resulting commonality analysis is driven by those assets. This is a feature we would like to understand, as it has not been discussed previously in the literature.

To further investigate this feature, in Figure 5.6 we present a plot of the spread and XLM on a randomly selected day for one of the assets with high R^2 (Nexans SA, stock symbol NEXp) and contrast it with the same liquidity measures for a second randomly selected stock with low PCA regression coefficient of determination (Barlcays - stock symbol BARCl). When contrasting these liquidity profiles, we note the very distinct spikes for NEXp, in both liquidity measures, indicating a heavy tailed distribution for these liquidity measures for this asset. Heavy tailed features of liquidity were also observed in the other assets which had exceptionally high R^2 values in the PCA regression.

If one performs the standard PCA approach to extract market factors affecting liquidity,

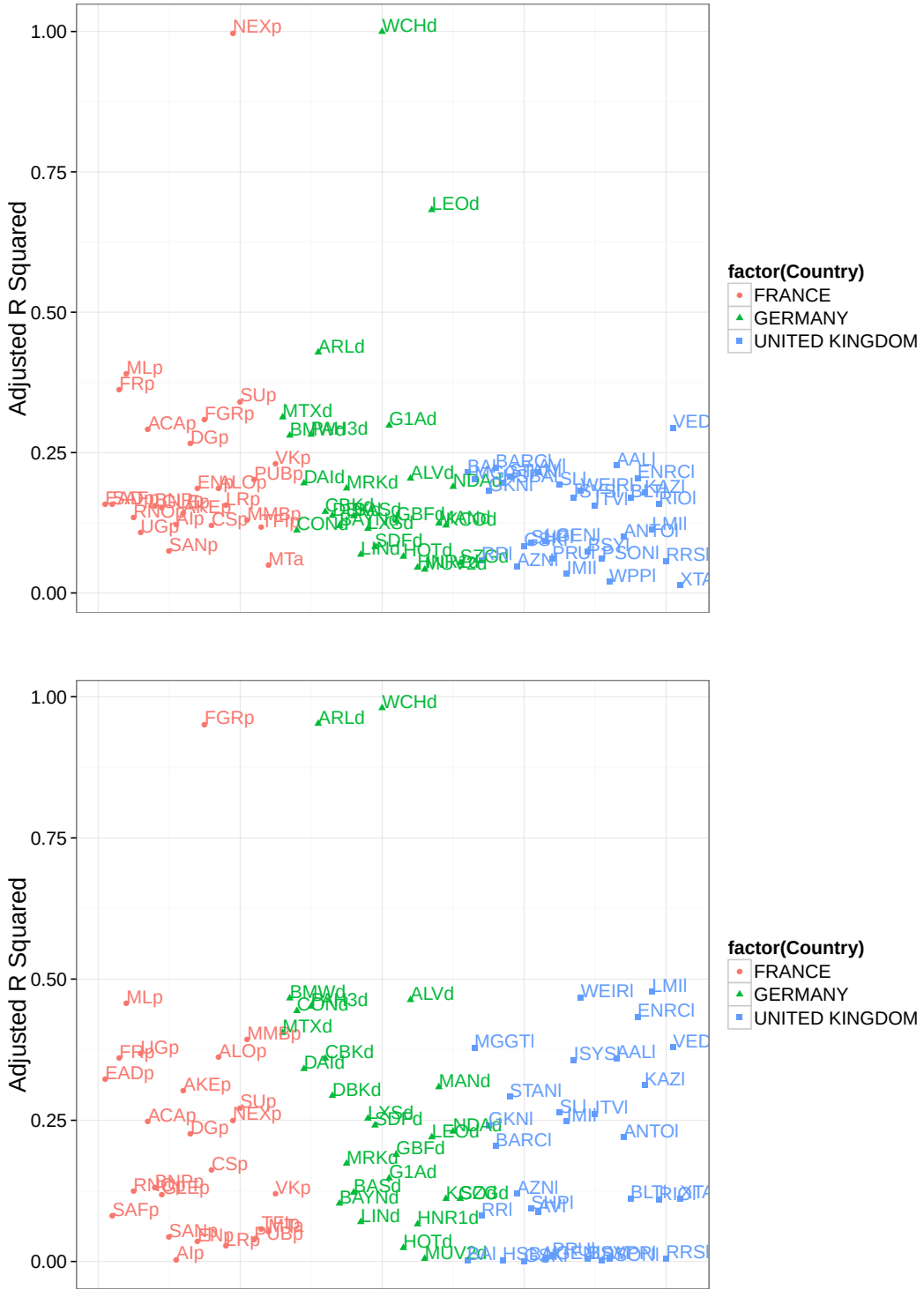


Figure 5.4: The R^2 values obtained from regressing individual asset liquidity against the first three PCs obtained across assets for the spread(top) and XLM(bottom), where assets are grouped by country. The labels indicate the Chi-X symbol of every asset.

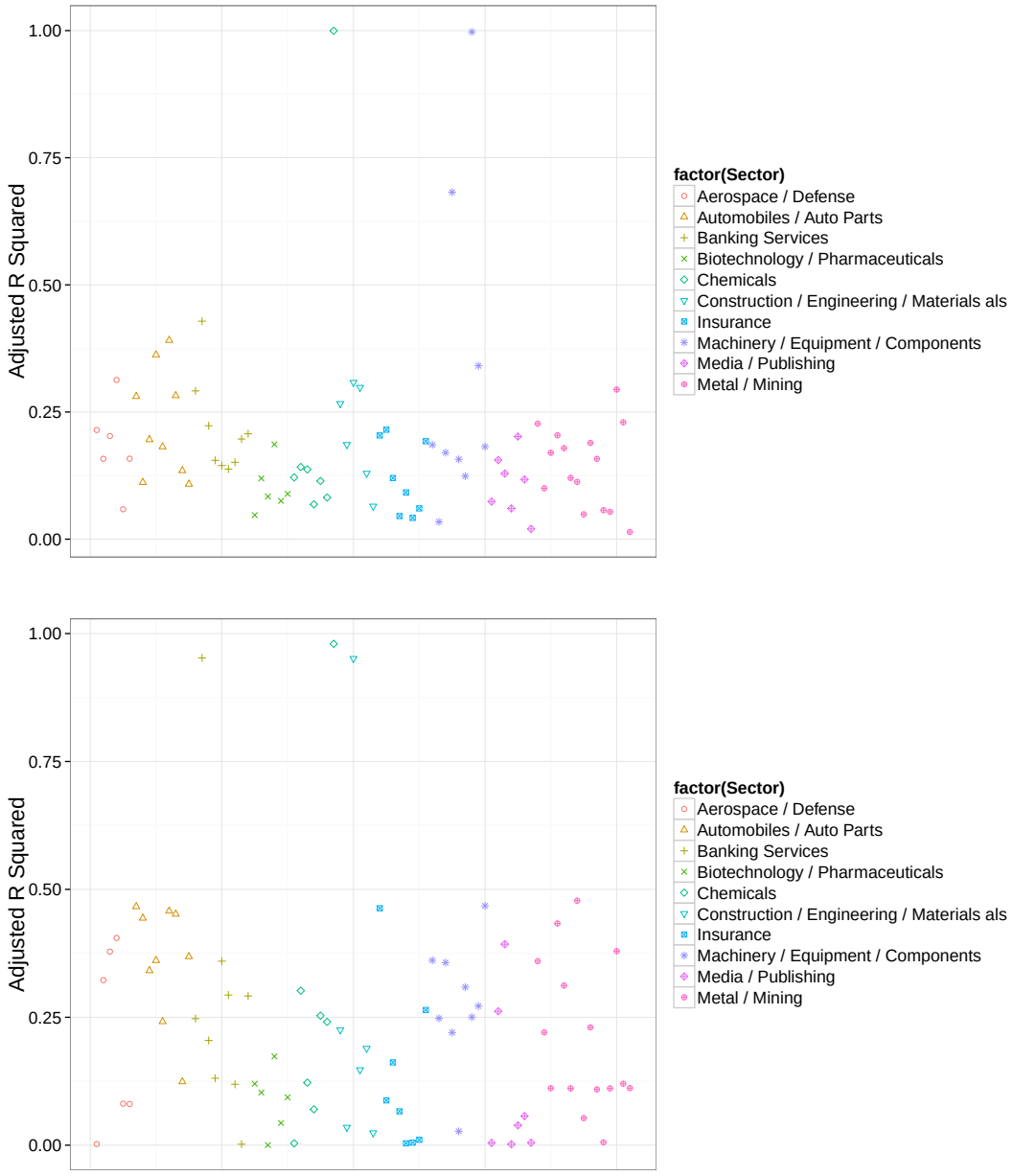


Figure 5.5: The R^2 values obtained from regressing individual asset liquidity against the first three PCs obtained across assets for the spread(top) and XLM(bottom), where assets are grouped by sector.

one would see that due to some relatively illiquid periods in the day for certain assets, these assets will dominate the PCA decomposition. This would therefore give a misleading picture of the market contribution to liquidity. If the PCs were then used in a regression, one would then expect that the explanatory power of the PCs for the assets which did not feature such illiquidity spikes would be fairly low, as is the case here.

Removing the two assets which regularly appear to drive the commonality does not solve

the issue, as other assets which contain heavy tailed features then become more prominent. In general, we find a large number of assets whose liquidity measures, to varying degrees, are heavy tailed. We note that these features were not reported in the work of Mancini et al. [2013], although liquidity in the foreign exchange markets which they investigate is generally much higher than in the equity markets, and therefore such heavy tails may not feature in the distributions of the liquidity measures they consider. Heavy tailed distributions in LOB depth are not specific to equities, however, as they have been identified and studied in Richards et al. [2012] also.

In light of these results, we would suggest that the explanatory power of PCs extracted through liquidity data with heavy tailed features may instead be interpreted as the degree of illiquidity commonality. This is because the leading PCs do not accurately reflect market liquidity, but rather the liquidity of the most illiquid assets. Our analysis reveals that the liquidity for most assets is poorly explained by these illiquid assets, and that there may therefore be, perhaps indirectly, a commonality in the liquidity of the remaining assets. However, this remains to still be studied. This is a different explanatory mechanism for the observed liquidity commonality features in the asset cross-section, compared to those discussed in Chordia et al. [2000], Domowitz et al. [2005] and Hasbrouck and Seppi [2001].

One important aspect of our analysis is in considering the appropriateness of the statistical techniques for large-scale datasets before routine application, see discussion on such matters in the PCA context in Candès et al. [2011]. We also suggest that the summary statistic or measure one selects for the datasets, in this case the liquidity measure, should be chosen appropriately so as to satisfy the assumptions of the statistical analysis being performed to assess commonality. Based on these findings we argue that it would be pertinent to therefore either consider alternative liquidity measures that don't demonstrate these statistical heavy tailed features so that PCA regressions may still be applied accurately, or to modify the approach adopted for the PCA to account for heavy tailed data, such as via the techniques discussed in Chen et al. [2009]. Alternatively, particularly for these large-scale high-frequency LOB liquidity measure datasets exhibiting marginal heavy tailed features, one could utilise Independent Component Analysis (ICA), in order to identify the assets which account for most of the non-Gaussian structure.

5.2.1 Independent Component Analysis

In this section we perform a commonality study of liquidity based on higher order moments. When liquidity measures in the asset cross-section are either heavy tailed, non-linear in the relationship with a market driving factor, or non-stationary, then one may resort to other forms

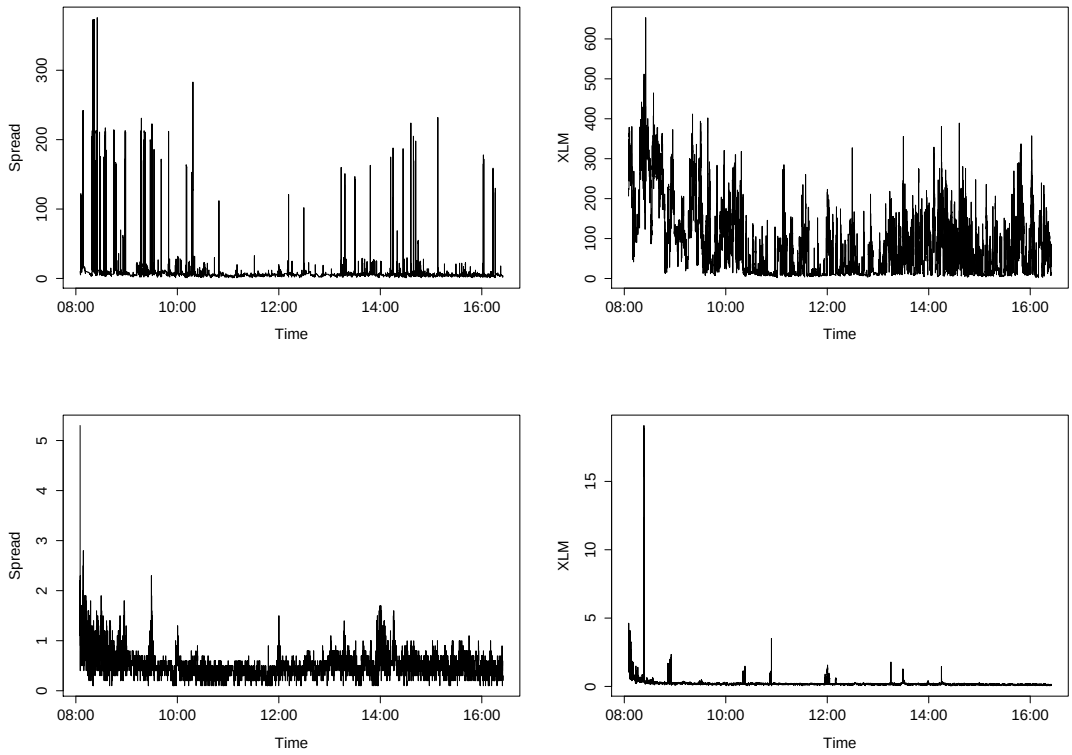


Figure 5.6: The daily evolution of the spread (left) and XLM (right) Nexans SA (symbol NEXp, top) and Barclays (symbol BARCl, bottom) on the 15th of February 2012

of decomposition, such as ICA, see discussion in Section 5.1.4.

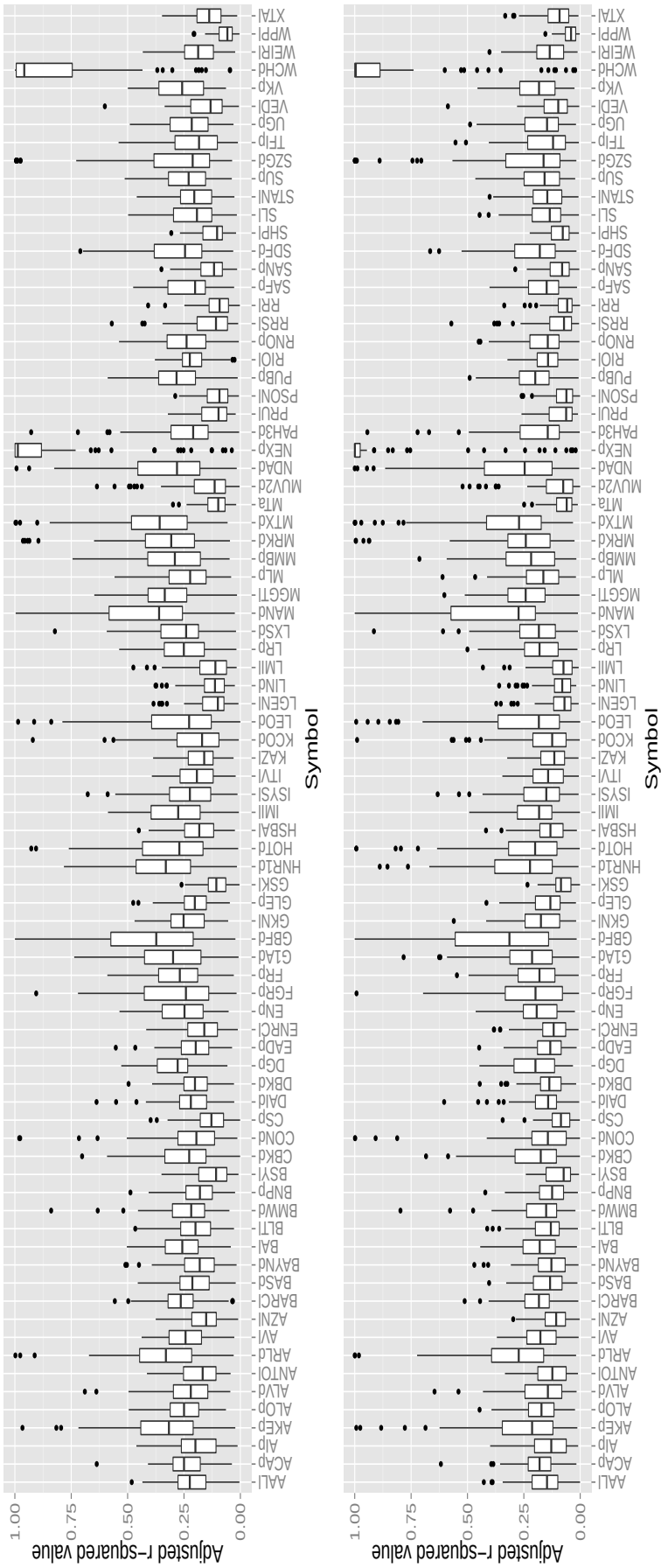


Figure 5.7: (Subplot 1): A summary of the R^2 values from the PCA regression using the leading 3 PCs across assets for the spread. These boxplots are obtained as follows: For every day in our dataset, we perform PCA on the spread data for all assets (sampled at 1 second intervals) and extract the first 3 PCs. For every asset, we then regress the intra-day spread against these PCs (which we consider to be the market factors), and obtain the R^2 coefficient of determination. Repeating this process for every day in our dataset, we have a series of R^2, which we summarise with boxplots for each asset. (Subplot 2):The R^2 values obtained from regressing individual asset liquidity against the first three ICs obtained across assets for the spread.

ICA methods, in contrast to the correlation-based transformations obtained in PCA for the market liquidity factors, not only de-correlate the liquidity measures in the asset cross-section each day, but also reduce higher-order statistical dependencies. Whereas PCA minimises the covariance of the data, ICA minimises higher-order statistics such as the fourth-order cumulant (or kurtosis), thus minimising the mutual information of the output. Specifically, PCA yields orthogonal vectors of high energy content in terms of the variance of the signals, whereas ICA identifies independent components for non-Gaussian signals.

After regressing each asset's liquidity against the three ICA components maximising non-Gaussianity, every day, we again obtained strong evidence to suggest that the assets dominating the PCA analysis due to Gaussianity violations, such as Nexans SA, also correspond to those that were very well explained in a linear projection by the ICA components. The coefficients of determination for the daily regressions for the other assets in the analysis are displayed in boxplots in Subplot 2 of Figure 5.7.

For the assets which are determined to coincide with the independent components from the ICA analysis, we have seen that these will dominate the PCA decomposition. We therefore suggest that ICA can be used as a preliminary step to identify whether these non-Gaussian features exist, and consequently, whether a PCA approach would be appropriate.

5.3 Liquidity resilience for high frequency data

In addition to the considerations regarding the appropriateness of the statistical assumptions for an analysis of commonality, we note that existing liquidity commonality approaches only reflect the aspects of liquidity measure chosen. In the case of the spread, this would be the tightness, and in the case of the XLM, it would also reflect the depth. However, since such measures cannot quantify liquidity resilience (which can be understood as the speed of liquidity replenishment), the associated commonality analysis will not reflect this aspect of liquidity either. Here, we extend the analysis to determine if the liquidity commonality observed is also present when one incorporates notions of resilience.

In Chapter 4, we introduced the Threshold Exceedance Duration (TED) and explained how one can use a parametric survival regression model to model the variation in the TEDs over time, where the duration variable of interest is denoted by τ. Using this formulation, one can obtain model based estimates of the expected (log) duration of an exceedance over a chosen threshold j of the liquidity measure, i.e. $\mathbb{E}[\ln(\tau^{[j]})|\hat{\boldsymbol{\beta}}, \boldsymbol{x}]$ given covariates $\boldsymbol{x}$ that are based on the LOB structure. They are a subset of those considered in the previous chapter, which were found to be more explanatory for the variation of the TED. They are as follows:

- The total number of asks in the first 5 levels of the LOB at time t, obtained according to $x_t^{(1)} = \sum_{i=1}^{5} \left| V_t^{a,i} \right|$ (where $|\cdot|$ is the number of orders at a particular level)
- The total number of bids in the first 5 levels of the LOB at time t, obtained according to $x_t^{(2)} = \sum_{i=1}^{5} \left| V_t^{b,i} \right|$
- The total ask volume in the first 5 levels of the LOB at time t, obtained according to $x_t^{(3)} = \sum_{i=1}^{5} TV_t^{a,i}$
- The total bid volume in the first 5 levels of the LOB at time t, obtained according to $x_t^{(4)} = \sum_{i=1}^{5} TV_t^{b,i}$
- The instantaneous value of the liquidity measure (spread or XLM) at the point at which the i-th exceedance occurs
- The number $x_t^{(6)}$ of previous TED observations in the interval $[t - \delta, t]$, with $\delta = 1s$
- The time since the last exceedance, $x_t^{(7)}$
- The average of the last 5 log TEDs, $x_t^{(8)}$
- A dummy variable indicating if the exceedance occurred as a result of a market order to buy, $x_t^{(9)}$
- A dummy variable indicating if the exceedance occurred as a result of a market order to sell, $x_t^{(10)}$

5.3.1 Summarising resilience behaviour

Once $\mathbb{E}[\ln(\tau^{[j]})|\hat{\boldsymbol{\beta}}, \mathbf{x}]$ is obtained for a large range of liquidity thresholds we can combine these to obtain the *Liquidity Resilience Profile* (LRP). The LRP is a summary of the expected resilience behaviour of an asset across different liquidity thresholds:

Definition 5.3.1. *The daily Liquidity Resilience Profile is a curve of the expected TEDs as a function of the liquidity threshold, given the state of the LOB, as quantified by the covariates characterising the LOB for the given asset.*

To facilitate comparison between the liquidity resilience behaviour of assets at different threshold levels, we present results for the logarithm of the expected TED, i.e. we calculate $\mathbb{E}[\ln(\tau^{[j]})|\hat{\boldsymbol{\beta}}^{[j]}, \boldsymbol{x}^{[j]}]$, where the j are threshold levels $j = 1, \ldots, 9$ corresponding to deciles of the empirical spread or XLM distribution. We can then identify the commonality in the

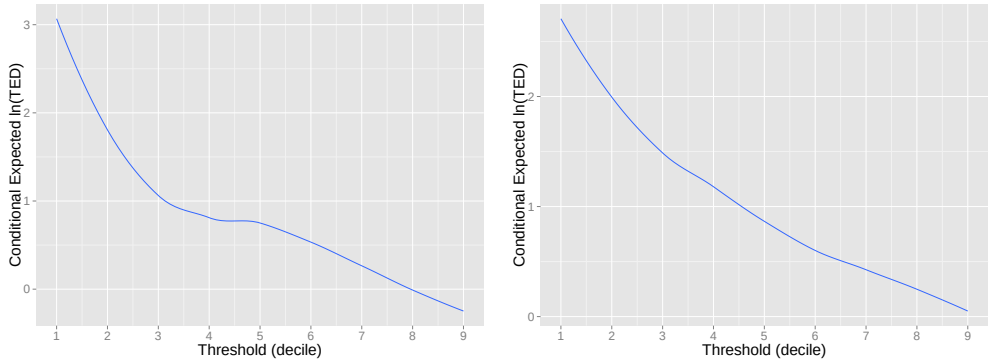


Figure 5.8: Liquidity resilience profile for Credit Agricole in the normal LOB regime, in which covariates take their median values. The x-axis represents the decile threshold used in the TED definition, and the curve is obtained by considering thresholds corresponding to deciles of the empirical distribution of the liquidity measure - in this case, the spread(left) and XLM(right).

expected TED over the median spread, or the 9th decile of the XLM, for example. We explain how the smooth functional representation is obtained in the following section.

We should note here that we will diverge somewhat from previous analyses, which only quantified the *temporal* commonality between the liquidity of individual assets. The temporal component of liquidity resilience commonality is captured through the similarity in the expected daily exceedance times over a threshold (measured through the TED metric). However, as we have obtained a representation of the expected exceedance time as a function of the level of the threshold, we can also quantify this commonality at different thresholds.

The functional representation enables us to then establish whether such commonality in liquidity resilience behaviour exists at any or all levels of the liquidity measure. The part of the curve corresponding to low thresholds of the spread or XLM indicates the expected time to return to a high level of liquidity, which would interest a brokerage house trying to minimise execution costs. The part of the curve corresponding to high thresholds, on the other hand, indicates the expected duration of periods with very low liquidity, which could be considered by a regulator as part of their efforts to ensure uninterrupted liquidity in financial markets.

We note that it would also be possible to extract such a functional form for the LRP directly, by performing the regression across all thresholds at once. However, it is more computationally efficient to do so in stages, by first fitting the survival regression model per threshold and then estimating the best fitting curve across the conditional expected TEDs using functional data analysis.

5.4 Functional data analysis characterisations of massive LOB data sets

In the analysis of financial data, one often has to deal with the issue of the high dimensionality of the data sets under consideration. We argue that there are significant advantages in summarising such high dimensional massive data sets under a functional characterisation. Functional data analysis (FDA) is a statistical approach that can be used to reduce the dimensionality of the problem, like a PCA analysis, but it allows one to capture additional features and perform analysis that are not possible in standard PCA approaches in the Euclidean space.

FDA has several advantages compared to multivariate analysis [Coffey et al., 2011]. It can achieve a parsimonious representation of an entire series of measurements from a single source, as a single functional entity, rather than a set of discrete values. It can also account for the ordering of the data (time-based or otherwise) through smoothing, as it is unlikely that adjacent values will vary by a large amount [Ramsay, 2006]. In addition, compared to multivariate analysis, it does not require that concurrent measurements are taken from every source of information. For these reasons, FDA would be highly appropriate for the analysis of unevenly spaced and high dimensional financial data. Detailed accounts of each aspect of FDA are provided in the text of Ramsay [2006].

Once a functional representation is obtained, one can explore functional equivalents of analyses performed in the multivariate space. For example, one can perform functional principal components analysis (FPCA) to extract the leading *eigenfunctions* characterising the functional dataset. Canonical correlation analysis can also be applied in the functional space, in order to investigate the modes of variability in two sets of functions that are most associated with one another. There are different ways in which one can build a functional linear model, with either a functional dependent variable, a set of functional covariates, or both [Ramsay, 2006]. We use the *concurrent model* in this chapter, involving a form of functions on functions regression, where we assume that the response is only affected by the dependent variables at the same point of the domain of the functions.

In the next section we detail how one can use FDA to obtain functional representations of LRPs, and FPCA to extract the dominant modes of variation every day for the asset cross-section. In addition, we will build a concurrent functional multiple regression model to quantify the explanatory power of the functional principal components (FPCs) for the LRPs of individual assets.

5.5 Functional data summaries: smoothed functional representations for LRPs

Functional data analysis is the study of functional data, where the domain of the function is usually time, but could be frequency, space, or in our case, thresholds of a liquidity measure. It differs from multivariate analysis methods such as time series modelling in its imposition of smoothness constraints:

$$\mathbf{y} = x(\mathbf{u}) + \boldsymbol{\epsilon} \tag{5.26}$$

where the $x(u)$ is considered to be a smooth functional data observed at certain points $\mathbf{u} = (u_1, \ldots, u_n)$, in the presence of noise, to get observations $\mathbf{y} = (y_1, \ldots, y_n)$. FDA then enables us to describe the variation in functional data, obtain derivatives and cluster curves according to their similarity.

In this chapter our interest is in the expected liquidity resilience behaviour of every asset for different thresholds. In this context, the dependent y_j is $\mathbb{E}[\ln(\tau^{[j]})|\hat{\boldsymbol{\beta}}^{[j]}, \mathbf{x}^{[j]}]$, where the j are threshold levels $j = 1, \ldots, 9$, defined as deciles of the empirical distribution of the liquidity measure. We will first explain how to obtain a smoothed representation $x(u)$ of the liquidity resilience profile of every asset, and then determine whether the dominant modes of variation over the different assets can be explanatory for the resilience of individual assets over the long term.

5.5.1 Defining a basis system for functional data representation

The first challenge is obtaining a functional data representation of discrete (and possibly noisy) observations of the daily LRPs for each asset. We can represent the LRP function $x(u)$ on a given day, for a given asset, using a basis expansion method, where a linear combination of the K basis functions $\phi_k(u)$ (with coefficients c_k) can approximate a smooth function for a sufficiently large K:

$$x(u) = \sum_{k=1}^{K} \phi_k(u) c_k. \tag{5.27}$$

If we then have N functions then

$$x_i(u) = \mathbf{c}_i^T \boldsymbol{\phi}(u), i = 1 \ldots N.$$

Common examples of bases used in FDA include Fourier bases, which are useful when data is periodic, and spline bases, of which several may be considered (B-splines, M-splines, I-splines etc.). Splines are piecewise polynomials, taking values in sub-intervals of the observation range. They are defined by:

- the range $[u_0, u_L]$ in which they take values;
- the order m of the spline, which is one higher than the highest degree polynomial;
- break points and knots, or the points which divide the observation range. Over a particular subinterval, the order of the polynomial is fixed. There can be several knots at a particular break point, if more than one basis function takes values in an adjacent subinterval.

We choose a B-spline basis here, and B-splines are defined recursively from lower order B-splines as follows:

$$B_{i,0}(u) = \begin{cases} 1 & \text{if } u_i \leq u < u_{i+1} \\ 0 & \text{elsewhere.} \end{cases}$$

$$B_{i,j+1}(u) = \alpha_{i,j+1}(u)B_{i,j}(u) + [1 - \alpha_{i,j+1}](u)B_{i+1,j}(u)$$

with $\sum_i B_{i,j}(u) = 1$ and

$$\alpha_{i,j}(u) = \begin{cases} \frac{u-u_i}{u_{i+j}-u_i} & \text{if } u_{i+j} \neq u_i \\ 0 & \text{otherwise.} \end{cases} \tag{5.28}$$

The spline function $S(u)$ is then defined as

$$S(u) = \sum_{k=1}^{m+L-1} c_k B_{k,m}(u). \tag{5.29}$$

This formulation means that a basis function is positive over at most m subintervals (this is called the compact support property), making estimation efficient. In our application, we obtain the LRP curve as a functional representation of the daily expected conditional log TED at each threshold. We achieve this via a cubic B-spline basis (i.e. the order $m = 4$). There is a continuity and smoothness restriction that adjacent polynomials (and first two derivatives) are constrained to be equal at the knots. In the range of observation thresholds $[u_0, u_L]$ we consider $L - 1$ interior knots, the interior knot sequence is generically denoted by $\mathbf{u} = (u_1, \ldots, u_{L-1})$. This produces a total of $m + L - 1$ basis functions for the function representation we adopt for the LRP of each asset each day.

We thus have to select the number of breakpoints (i.e. the value of L) and the number of basis functions K to use in the B-spline basis, although selecting a value for one will determine the other. As our curves are constructed over 9 thresholds of the liquidity measure, we select $L = 4$ in order to obtain a parsimonious representation of the LRP.

5.5.1.1 Estimation of functional representations of LRPs

We perform linear regression on the basis functions to obtain the coefficient vector $\mathbf{c}$, i.e. by minimising the sum of squared errors

$$SSE = \sum_{j=1}^{n}(y_j - \sum_{k=1}^{K} c_k \phi_k(u_j))^2 = (\mathbf{y} - \Phi\mathbf{c})'(\mathbf{y} - \Phi\mathbf{c}) \tag{5.30}$$

where Φ is an N by K matrix containing $\phi_j(t_k)$. The OLS estimate is

$$\hat{\mathbf{c}} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y} \tag{5.31}$$

and the vector of fitted values is

$$\hat{\mathbf{y}} = \Phi\hat{\mathbf{c}} = \Phi(\Phi^T\Phi)^{-1}\Phi^T\mathbf{y} \tag{5.32}$$

from which we can see that $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ acts as a simple linear smoother. This approximation is only appropriate if we assume i.i.d errors, but this is not often the case with functional data. In order to enforce smoothness, we can add a roughness penalty to the least squares criterion

$$PENSSE_\lambda = (\mathbf{y} - \Phi\mathbf{c})'(\mathbf{y} - \Phi\mathbf{c}) + \lambda J(x) \tag{5.33}$$

where λ is a tuning parameter and $J(x)$ measures roughness, for example through the curvature $J_2(x) = \int_u [D^2x(u)]^2du$, or, more generally, using any linear differential operator $J(x) = \int_u \sum_{k=1^m} \alpha_k D^k[(x(u))]du$. The D operator is used to denote derivatives, such that $D^0x(u) = x(u)$ and $D^mx(u) = \frac{d^mx(u)}{du}$.

We impose the J_2 roughness penalty in the estimation, as in areas where the function is highly variable, the square of the second derivative will be large. A theorem from De Boor [2001] shows that when choosing $J_2(x) = \int_u [D^2x(u)]^2du$, a cubic spline with knots at points u_j minimises $PENSSE_\lambda$. Spline smoothing with the roughness penalty above is still a linear operation, where the smoother is now $(\Phi^T\Phi + \lambda\mathbf{R})^{-1}\Phi^T$, where

$$\mathbf{R} = \int D^2\boldsymbol{\phi}'(u)D^2\boldsymbol{\phi}(u)du \tag{5.34}$$

see Ramsay [2006] for a derivation. This is usually computed by numerical integration.

The last point is choosing the smoothing parameter λ, and a widely used approach is the generalised cross-validation (GCV) method proposed by Craven and Wahba [1978]. One can use the generalised cross-validation measure, whereby minimisation of the criterion is a method to select λ. For each asset, on every day, we calculate the GCV value on a fine grid, and present in Figure 5.9 in the histogram of the values of λ corresponding to the lowest GCV. We find that these values concentrate at very low levels and very high levels. However, empirical

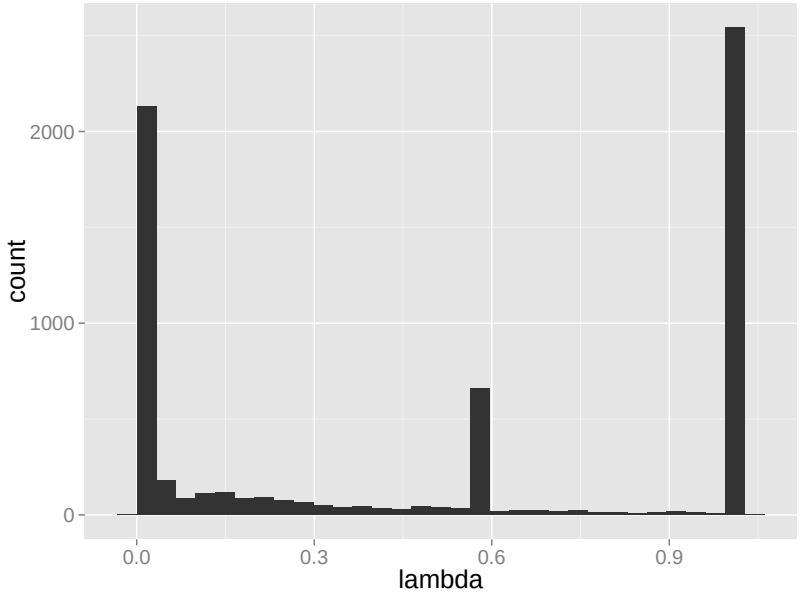


Figure 5.9: The optimal λ value calculated by the GCV procedure for every LRP fit (i.e. for every asset on every day).

analysis shows that using a large smoothing parameter (close to 1) leads to oversmoothing, and we lose some of the interesting features in the data. For this reason, we choose to use the same parameter $\lambda = 0.02$ for every asset, on every day. We summarise the result of this data preparation in Figure 5.10 for the LRPs for all assets for both the spread and the XLM, obtained using the B-spline basis and roughness penalisation method described above. There is variation in the LRPs of individual assets over time, and this justifies our choice of investigating liquidity resilience (and its commonality) daily first. The darker shaded area shows that there is a clustering of LRPs across assets and across time, and this is a first visual confirmation of commonality in liquidity resilience over the different (relative) liquidity measure thresholds. Now that we have estimated smoothed representations of the daily LRPs for every asset, we will treat them as the observed data that we will analyse to quantify any commonality.

5.6 Functional principal components analysis

We now evaluate the market factors characterising the asset cross-section functional LRP profiles each day. In functional principal components analysis (FPCA), we seek the dominant modes of variation over a set of curves. As an example of PCA in the multivariate space, Hasbrouck and Seppi [2001] and Korajczyk and Sadka [2008] characterise market-wide liquidity as the first principal component of individual FX rates. We focus on the functional equivalent, but want to characterise the resilience of liquidity, rather than liquidity itself. We then determine

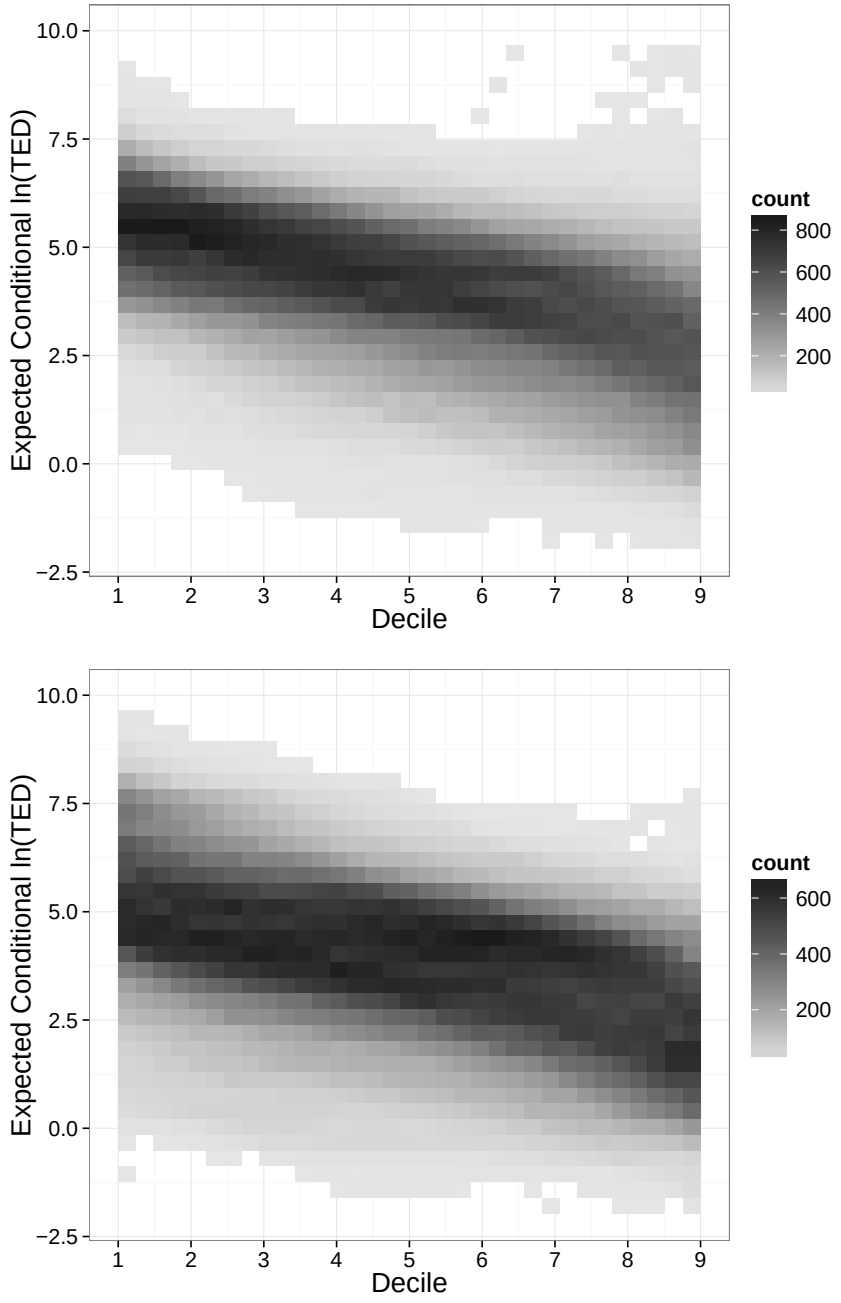


Figure 5.10: Projected LRPs of all 82 assets, for the entire 81 day period, onto a single grid for the spread (top) and XLM (bottom).

the extent to which these can be explanatory for individual asset resilience over time.

Specifically, given smoothed functional data $\{x_i(u)\}_{i\in 1:I}$, we are searching for the weight functions ξ, such that the corresponding scores

$$f_i = \int \xi(u) x_i(u) du \tag{5.35}$$

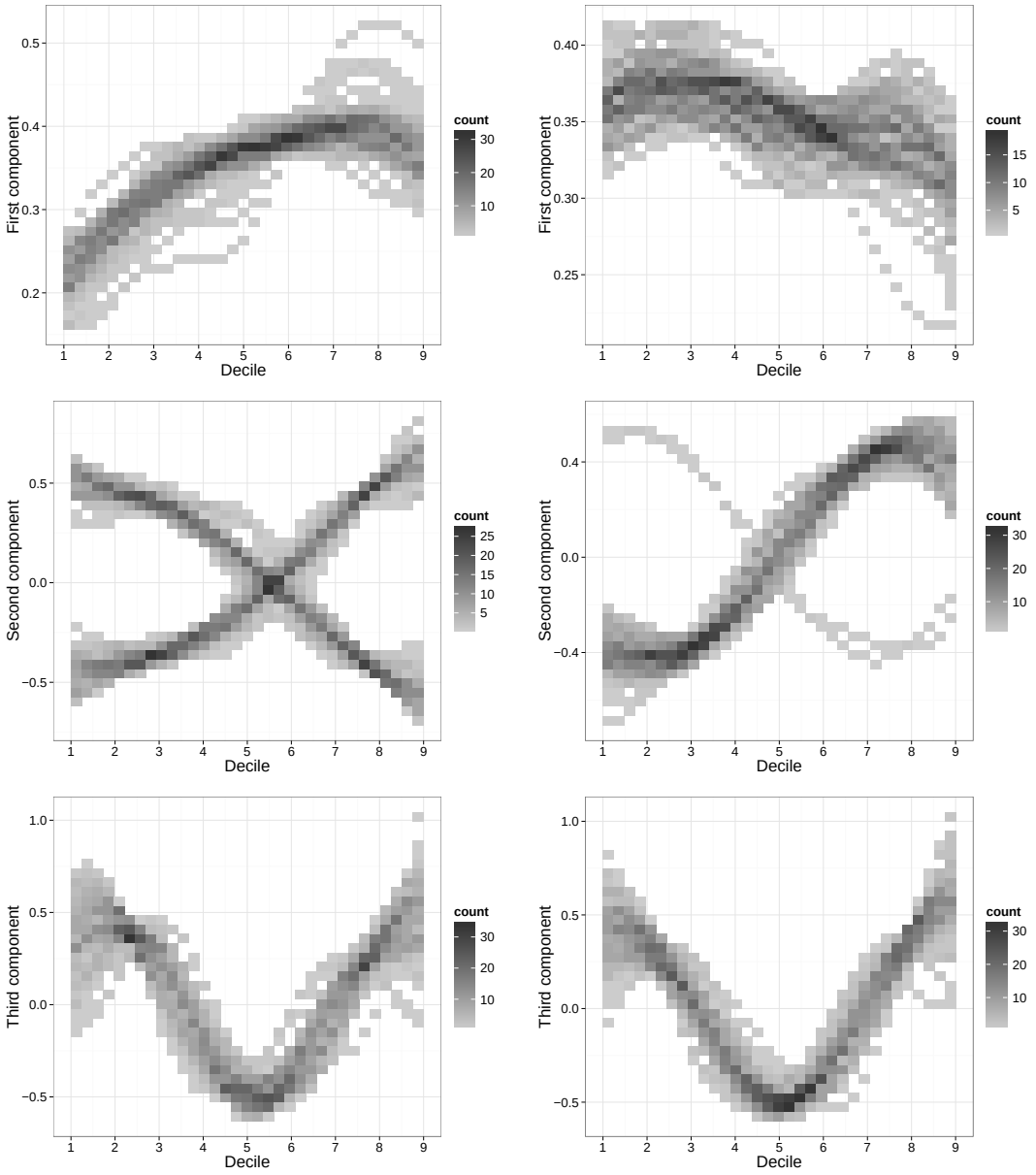


Figure 5.11: The first three functional PCs extracted from the LRP data every day, projected onto the same axes for the spread (left) and XLM (right)

have the largest possible variation. That is, weight function ξ_1 should maximise

$$\sum_i \left[\int \xi(u) x_i(u) du \right]^2 \tag{5.36}$$

subject to the constraint $\int \xi(u)^2 du = 1$. In this context, ξ_1 will be the most important functional component of the market-wide liquidity resilience profile and will correspond to the dominant model of functional variation. It will be represented by a linear combination of basis functions, just as individual LRPs.

Consider the mean and covariance functions for the functional LRPs denoted by $\mu(u) =$

$\mathbb{E}[x_i(u)]$ and $v(u,w)$

$$v(u,w) = \frac{1}{N-1}\sum_{i=1}^{N} x_i(u)x_i(w) \tag{5.37}$$

and the covariance operator V

$$V\xi = \int v(\cdot,u)\xi(u)du \tag{5.38}$$

This operator has orthonormal eigenfunctions $\xi_1,\ldots,\xi_K$, with eigenvalues $\rho_1 \geq \rho_2 \geq \ldots \geq \rho_K$, satisfying $V\xi_k = \rho_k\xi_k$

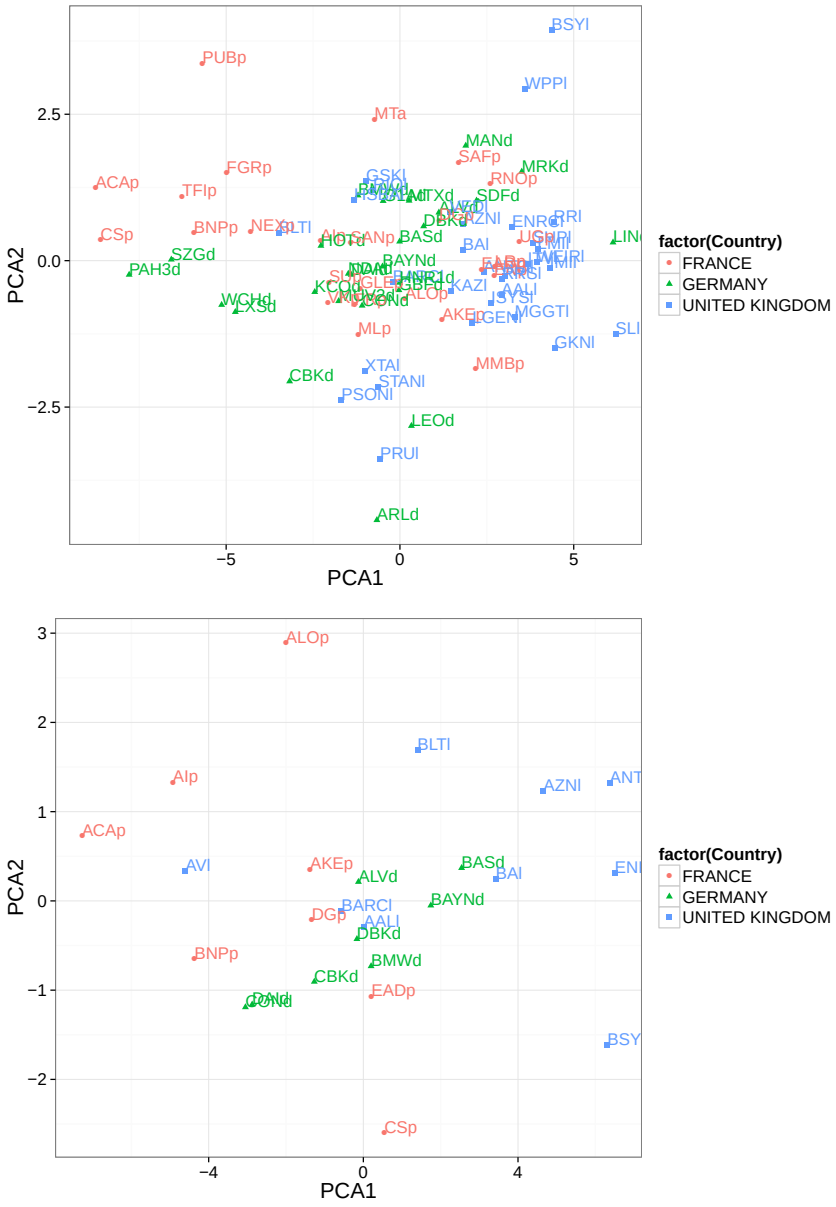


Figure 5.12: Scores for the first two PCs for every asset for the spread(top) and XLM(bottom), for single-day equity data.

In Figure 5.11 we project the first three leading market liquidity resilience factors from the

FPCA from every daily dataset onto the same axes, in order to understand whether the dominant modes of variation vary over time. In the vein of the liquidity commonality literature, we will call these the market factors of resilience. We note that the first FPC is fairly constant over time, and is greater at higher threshold levels for the spread. This indicates that the market component of resilience is important for explaining deviations from more extreme levels of the spread. Once we consider a liquidity measure which takes depth into account, however, such as the XLM, the opposite seems to apply: we observe that the contribution of the first FPC from the daily XLM liquidity resilience profiles tends to decrease at higher thresholds.

There are two distinct modes for the second FPC, which become almost identical, if one is flipped across the x-axis, and this is the case for both the spread and the XLM. We find that for some assets, multiplying the second PC by the score for that asset almost eliminates the second mode of variation and thus the effect of the second PC becomes relatively constant over time.

5.7 Functional principal component regression for LRPs

Recall that Mancini et al. [2013] regress individual exchange rate liquidity against the principal component obtained over all rates, and interpret the R^2 coefficient of determination for every asset as the degree of commonality. For our study of liquidity resilience commonality we perform a similar regression idea except we extend this in our case to the functional space setting. The functional principal components obtained every day that characterize the market liquidity resilience factors will now be used as functional covariates, in order to explain the variation in LRPs for individual assets inter-daily.

A linear regression model relating a functional response to a collection of functional covariates at the same points is called a concurrent model and is given as follows:

$$x_{i,t}(u) = \beta_0(u) + \sum_{j=1}^{q} \beta_j(u)\xi_{j,t}(u) + \epsilon_{i,t}(u) \tag{5.39}$$

where t is the day index, β_0 is the intercept function, and the β_j are coefficient functions of the covariate functions ξ_j i.e. the market functional PCs. β_0 could also be considered as being the product with a constant function whose value is always one. Let the t by q matrix $\mathbf{Z}$ contain the covariate functions $\xi_{j,t}$. In matrix notation, the concurrent functional linear model is given by

$$\boldsymbol{x}(u) = \mathbf{Z}(u)\boldsymbol{\beta}(u) + \boldsymbol{\epsilon}(u) \tag{5.40}$$

and the fitting criterion (if we also include a roughness penalty $J(\beta_j)$) becomes

$$LMSSE(\boldsymbol{\beta}) = \int \mathbf{r}(u)'\mathbf{r}(u)du + \sum_{j=1}^{q} \lambda_j \int J(\beta_j(u))du \tag{5.41}$$

where $\mathbf{r}(u) = \boldsymbol{x}(u) - \mathbf{Z}(u)\boldsymbol{\beta}(u)$. For the estimation method utilised see details in Ramsay [2006].

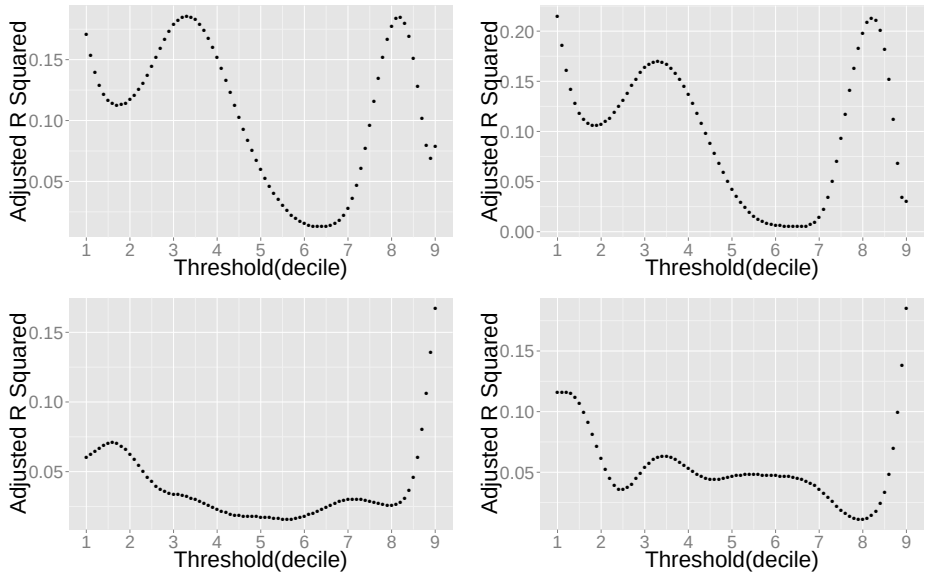


Figure 5.13: The R^2 functions obtained from regressing individual asset Liquidity Resilience Profiles against the first two PCs obtained from the daily LRP curves using the spread (top) and XLM(bottom) for stocks Nexans SA and Credit Agricole.

The entire functional PCA regression procedure is summarised in the following steps:

1. We first obtain functional representations of the LRP for every asset on every day.

2. We then extract the first 3 components from the LRPs every day, which will serve as our covariates.

3. We set up a basis for the coefficient functions $\beta_0, \beta_1, \beta_2, \beta_3$.

4. Finally, we take LRPs for a single asset over time (this will be the dependent variable) and run the regression.

Here β_0 will have a constant basis, while for $\beta_1, \beta_2, \beta_3$ we set up a cubic spline basis as before, but with 5 basis functions. We imposed the same J_2 roughness penalisation as before, in order

to avoid possible overfitting. We can assess the quality of the fit for asset i using the R^2 function

$$SS_{reg}(u) = \sum_t [\hat{x}_{i,t}(u) - \mu_t(u)]^2$$
$$SS_{res}(u) = \sum_t [\hat{x}_{i,t}(u) - x_{i,t}(u)]^2$$
$$R^2(u) = \frac{SS_{reg}(u)}{SS_{reg}(u) + SS_{res}(u)}$$

We should note that in our results we omit the intercept function β_0, as it did not increase the explanatory power of the regression. We present the R^2 function for two assets , where we can now observe the explanatory power of the regression at different threshold levels. That is at threshold level u, $R^2(u)$ denotes the proportion of variation in the LRP of an individual asset at threshold u that is explained by the first 3 components obtained from the FPCA analysis, again at level u. The R^2 varies as we alter u, and this is consistent across assets, as we can see in Figure 5.13. The advantage of this representation is that we can identify the ranges of u liquidity thresholds for which the principal components are more or less explanatory over time.

We note that in the case of the spread, the PCs are explanatory about the initial part of the curve, that is, where we consider deviations from relatively low spreads. Just as market factors (principal components) extracted from the spreads of individual assets can explain around 25%of the variation in the absolute level of the spread for these assets (as we had noted in Figure 5.7), PCs extracted from the LRPs can explain between 10 and 40% of the variation in the expected duration of spread deviations, at least at lower spreads. Once we consider deviations from higher spreads, the explanatory power of market factors drops sharply.

This indicates that there are additional factors that become important at higher liquidity levels, which are specific for each asset. A possible explanation for this would be a difference in the efficiency of market making between assets. As there are varying requirements for market making for assets in different jurisdictions, it is possible that in certain assets, market makers can stop operating in more illiquid LOB regimes, to avoid building any position that would then be costly to unwind. This would mean that for those assets, the expected duration of deviations above greater levels of the spread would be higher, as there would be fewer market participants willing to replenish the market after a shock.

In the case of the XLM, we find that the commonality in liquidity resilience - which we again measure through the explanatory power of the market factors, extracted through the functional PCA - is markedly lower than the corresponding temporal commonality of the XLM across assets. A possible reason for the lower adjusted R^2 levels for the XLM compared to the respective figures for the spread is that while market makers may act similarly in trying to

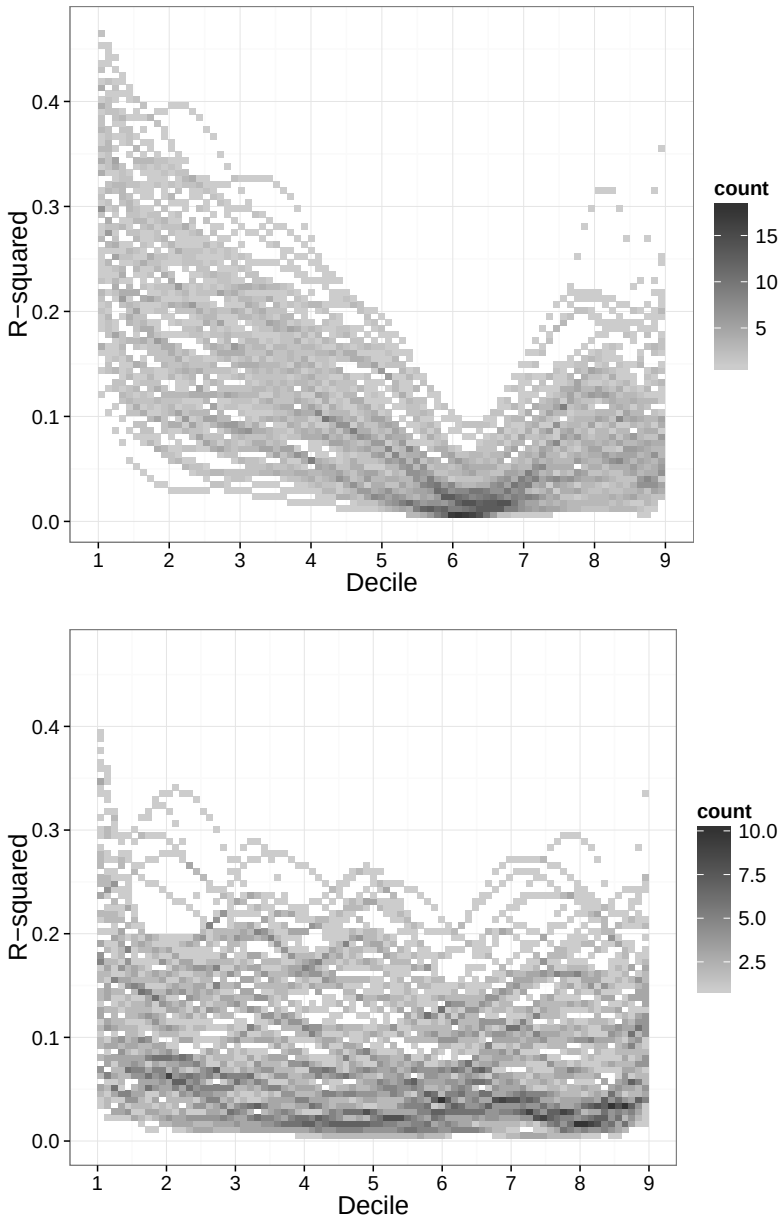


Figure 5.14: The R^2 functions obtained from regressing individual asset Liquidity Resilience Profiles against the first three PCs obtained from the daily LRP curves for the spread(top) and XLM(bottom).

tighten the spread after a shock, they are less inclined to post large volumes in the LOB for certain assets, in order to avoid excess exposure in a market where it will be costly to unwind a position.

5.8 Discussion

We have reviewed the performance of the standard approach for measuring liquidity commonality through principal components regression, on a massive dataset from a pan-European equity venue. We have shown that the assumption that one can capture the most important features of liquidity commonality, through methods which are based on second moments, will not always be appropriate. We therefore perform ICA (Independent Component Analysis), which addresses this issue by incorporating higher order information, to assess commonality in liquidity.

The standard approach to liquidity commonality fails to capture commonality in the resilience of liquidity, or the speed of replenishment of the LOB. We addressed this by proposing an approach to quantify the commonality in liquidity resilience (as characterised by 4) by first obtaining a functional representation of resilience and then measuring the explanatory power of market factors extracted from the asset cross-section. We have shown that functional data analysis can be very valuable for characterising features of massive datasets, such as those extracted from high-frequency LOB data, as it can vastly reduce the dimensionality of the data and make comparisons between assets possible.

Our results suggest that market factors for liquidity resilience (as captured by functional principal components analysis) can explain between 10 and 40% of the time required for the spread to return to a low threshold level after a shock. The same market factors are found to be much less explanatory if we consider higher threshold levels of the spread. Once we also consider a liquidity measure that takes depth into account, such as the XLM, the explanatory power diminishes significantly.

We have interpreted these results through the prism of market making activity in the LOB. While market makers may act similarly in trying to tighten the spread after a shock, they are less inclined to post large volumes in the LOB for certain assets, in order to avoid excess exposure. We also identified the possible absence of quoting obligations for certain assets to be a contributing factor to explaining these outcomes.

Contrasting these results with our liquidity commonality findings, we find that temporal commonality in the liquidity measures does not necessarily entail commonality in liquidity resilience. We would argue that this has positive implications for market quality, as it indicates that slow liquidity replenishment in certain assets is not necessarily contagious for the market. Future studies will further explore the economic ramifications of these findings in detail.

	Country	Name	Symbol	Sector
1	FRANCE	EADS (PAR)	EADp	Aerospace / Defense
2	FRANCE	SAFRAN	SAFp	Aerospace / Defense
3	FRANCE	VALEO	FRp	Automobiles / Auto Parts
4	FRANCE	MICHELIN	MLp	Automobiles / Auto Parts
5	FRANCE	RENAULT	RNOp	Automobiles / Auto Parts
6	FRANCE	PEUGEOT	UGp	Automobiles / Auto Parts
7	FRANCE	CREDIT AGRICOLE	ACAp	Banking Services
8	FRANCE	BNP PARIBAS	BNPp	Banking Services
9	FRANCE	SOCIETE GENERALE	GLEp	Banking Services
10	FRANCE	SANOFI	SANp	Biotechnology / Pharmaceuticals
11	FRANCE	AIR LIQUIDE	AIp	Chemicals
12	FRANCE	ARKEMA	AKEp	Chemicals
13	FRANCE	VINCI (EX SGE)	DGp	Construction / Engineering / Materials als
14	FRANCE	BOUYGUES	ENp	Construction / Engineering / Materials als
15	FRANCE	EIFFAGE	FGRp	Construction / Engineering / Materials als
16	FRANCE	AXA	CSp	Insurance
17	FRANCE	ALSTOM	ALOp	Machinery / Equipment / Components
18	FRANCE	LEGRAND	LRp	Machinery / Equipment / Components
19	FRANCE	NEXANS	NEXp	Machinery / Equipment / Components
20	FRANCE	SCHNEIDER ELECTRIC	SUp	Machinery / Equipment / Components
21	FRANCE	LAGARDERE GROUPE	MMBp	Media / Publishing
22	FRANCE	PUBLICIS GROUPE	PUBp	Media / Publishing
23	FRANCE	TF1 (TV.FSE.1)	TFIp	Media / Publishing
24	FRANCE	ARCELORMITTAL	MTa	Metal / Mining
25	FRANCE	VALLOUREC	VKp	Metal / Mining
26	GERMANY	MTU AERO ENGINES HLDG.	MTXd	Aerospace / Defense
27	GERMANY	BMW	BMWd	Automobiles / Auto Parts
28	GERMANY	CONTINENTAL	CONd	Automobiles / Auto Parts
29	GERMANY	DAIMLER	DAId	Automobiles / Auto Parts
30	GERMANY	PORSCHE AML.HLDG.PREF.	PAH3d	Automobiles / Auto Parts
31	GERMANY	AAREAL BANK	ARLd	Banking Services
32	GERMANY	COMMERZBANK	CBKd	Banking Services
33	GERMANY	DEUTSCHE BANK	DBKd	Banking Services
34	GERMANY	BAYER	BAYNd	Biotechnology / Pharmaceuticals
35	GERMANY	MERCK KGAA	MRKd	Biotechnology / Pharmaceuticals
36	GERMANY	BASF	BASd	Chemicals
37	GERMANY	LINDE	LINd	Chemicals
38	GERMANY	LANXESS	LXSd	Chemicals
39	GERMANY	K + S	SDFd	Chemicals
40	GERMANY	WACKER CHEMIE	WCHd	Chemicals
41	GERMANY	GEA GROUP	G1Ad	Construction / Engineering / Materials als
42	GERMANY	BILFINGER BERGER	GBFd	Construction / Engineering / Materials als

Table 5.1: Country and sector information about the first 42 of the 82 assets used in this study. Continued in Table 5.2.

	Country	Name	Symbol	Sector
43	GERMANY	HOCHTIEF	HOTd	Construction / Engineering / Materials als
44	GERMANY	ALLIANZ	ALVd	Insurance
45	GERMANY	HANNOVER RUCK.	HNR1d	Insurance
46	GERMANY	MUENCHENER RUCK.	MUV2d	Insurance
47	GERMANY	LEONI	LEOd	Machinery / Equipment / Components
48	GERMANY	MAN	MANd	Machinery / Equipment / Components
49	GERMANY	KLOECKNER & CO	KCOd	Metal / Mining
50	GERMANY	AURUBIS	NDAd	Metal / Mining
51	GERMANY	SALZGITTER	SZGd	Metal / Mining
52	UK	BAE SYSTEMS	BAl	Aerospace / Defense
53	UK	MEGGITT	MGGTl	Aerospace / Defense
54	UK	ROLLS-ROYCE HOLDINGS	RRl	Aerospace / Defense
55	UK	GKN	GKNl	Automobiles / Auto Parts
56	UK	BARCLAYS	BARCl	Banking Services
57	UK	HSBC HDG.	HSBAl	Banking Services
58	UK	STANDARD CHARTERED	STANl	Banking Services
59	UK	ASTRAZENECA	AZNl	Biotechnology / Pharmaceuticals
60	UK	GLAXOSMITHKLINE	GSKl	Biotechnology / Pharmaceuticals
61	UK	SHIRE	SHPl	Biotechnology / Pharmaceuticals
62	UK	AVIVA	AVl	Insurance
63	UK	LEGAL & GENERAL	LGENl	Insurance
64	UK	PRUDENTIAL	PRUl	Insurance
65	UK	STANDARD LIFE	SLl	Insurance
66	UK	IMI	IMIl	Machinery / Equipment / Components
67	UK	INVENSYS	ISYSl	Machinery / Equipment / Components
68	UK	WEIR GROUP	WEIRl	Machinery / Equipment / Components
69	UK	BRITISH SKY BCAST.GROUP	BSYl	Media / Publishing
70	UK	ITV	ITVl	Media / Publishing
71	UK	PEARSON	PSONl	Media / Publishing
72	UK	WPP	WPPl	Media / Publishing
73	UK	ANGLO AMERICAN	AALl	Metal / Mining
74	UK	ANTOFAGASTA	ANTOl	Metal / Mining
75	UK	BHP BILLITON	BLTl	Metal / Mining
76	UK	EURASIAN NATRES.CORP.	ENRCl	Metal / Mining
77	UK	KAZAKHMYS	KAZl	Metal / Mining
78	UK	LONMIN	LMIl	Metal / Mining
79	UK	RIO TINTO	RIOl	Metal / Mining
80	UK	RANDGOLD RESOURCES	RRSl	Metal / Mining
81	UK	VEDANTA RESOURCES	VEDl	Metal / Mining
82	UK	XSTRATA	XTAl	Metal / Mining

Table 5.2: Country and sector information about the remaining assets used in this study.

Chapter 6

Liquidity motivated agent-based model of the LOB

In the previous 2 chapters we offered both a theoretical contribution to the understanding of a particular aspect of LOB liquidity, namely resilience, as well as an empirical analysis of liquidity and resilience commonality in the equity markets. LOB liquidity is a very important determinant of market quality, and its intra-day variation plays a large part in defining the overall trading strategies of market participants. In this chapter we try to capture the central role of liquidity in the LOB by considering daily activity as part of a new form of agent-based model based on heterogeneous trading agents, whose motivations are liquidity-driven.

We develop two types of agents in our framework, liquidity providers (market makers), and liquidity demanders, who encompass a stylised representation of algorithmic traders, noise traders, trend followers and other types of speculators. These agents are abstractions of real market participants and their attributes, but they are expressed in a stochastic model framework, which characterises these behaviours in more detail than typical simple agent models. This places our model part way between a traditional agent-based model and a pure stochastic LOB model. The model can be of practical use, as it is readily interpreted with regard to the agent model dynamics, it can be easily estimated from data and it reproduces many important empirical properties of LOB dynamics.

We develop an efficient way to perform statistical calibration and estimation of the agent model parameters based on a combination of Indirect Inference (a simulation based likelihood procedure) and multi-objective optimisation. We calibrate our agent-based stochastic model to real high frequency data from 5 level market depth limit order book data from Chi-X. We then demonstrate how such an agent-based model is a valuable framework for testing exchange regulations, for brokerage decisions and other trading based scenarios.

6.1 New perspective: Stochastic agent-based models for the LOB

In this section we present the formal mathematical specification for each component of our stochastic agent-based model. This includes the stochastic models for limit order placements and cancellations by a liquidity provider agent and the stochastic models for market order placements by liquidity demanding representative agents. The stochastic ABM framework can model the non-linear dependence in intra-day LOB activity, where the dependence is considered both between different types of events (e.g. limit and market orders), but also the same type of events, but at different levels (e.g. cancellations at level 2 and level 5 of the ask side of the LOB). We make extensive use of the flexible multivariate skew-t distribution, which is unique in enabling the modelling of heavy tails, tail dependence, skew and clustering of volatility [Demarta and McNeil, 2005, Fung and Seneta, 2010].

6.1.1 Limit Order Book simulation framework

We consider the intra-day LOB activity in fixed intervals of time $\ldots, [t-1, t), [t, t+1), \ldots$. For every interval $[t, t+1)$, we allow the total number of levels on the bid or ask sides of the LOB to be dynamically adjusted as the simulation evolves. These LOB levels are defined with respect to two reference prices, equal to $P^{b,1}_{t-1}$ and $P^{a,1}_{t-1}$, i.e. the price of the highest bid and lowest ask price at the start of the interval. We consider these reference prices to be constant throughout the interval $[t-1, t)$ and thus, the levels on the bid side of the book are defined at integer number of ticks away from $P^{a,1}_{t-1}$, while the levels on the ask side of the book are defined at at integer number of ticks away from $P^{b,1}_{t-1}$.

This does not mean that we expect the best bid and ask prices to remain constant, just that we model the activity (i.e. limit order arrivals, cancellations and executions) according to the distance in ticks from these reference prices during this period. We note that it is of course possible that the volume at the best bid price is consumed during the interval, and that limit orders to sell are posted at this price, which would be considered at 0 ticks away from the reference price. To allow for this possibility, we actively model the activity at $-l_d + 1, \ldots, 0, \ldots, l_p$ ticks away from each reference price. Here, the p subscript will refer to passive orders, i.e. orders which would not lead to immediate execution, if the reference prices remained constant. d refers to direct, or aggressive orders, where it is again understood that they are aggressive with respect to the reference prices at the start of the period. Therefore, we actively model the activity at a total $l_t = l_p + l_d$ levels on the bid and ask, as indicated in Figure 6.1.

We assume that activity that occurs further away is uncorrelated with the activity close

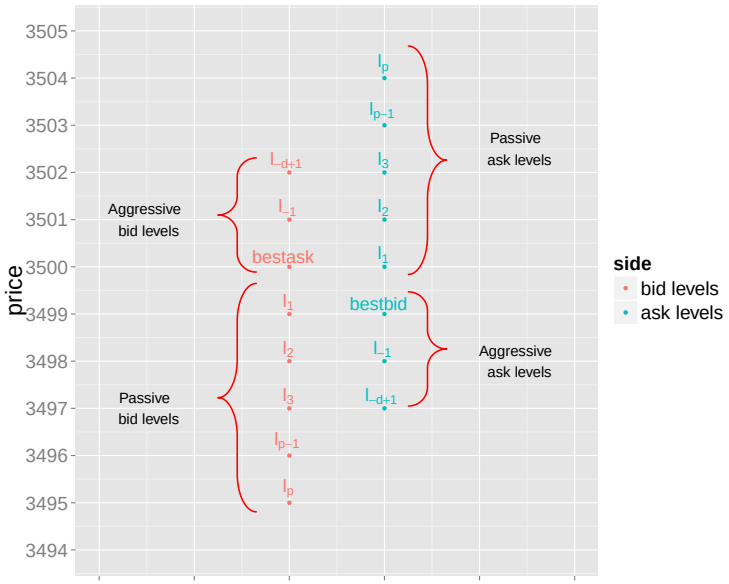


Figure 6.1: The actively modelled levels of the LOB in the agent-based model presented in this paper. There are a total of l_t levels on each side, where l_p are passive levels and l_d are direct, or aggressive levels (i.e. would lead to immediate execution). The levels of the ask are considered around the best bid price at the start of each interval, and likewise the levels of the bid side are considered around the best ask side at the start of each interval. In this figure, as in our model, we have $l_p = 5$ and $l_d = 3$.

to the top of the book (as is evident in Figure 6.2), and therefore unlikely to have much of an impact on price evolution and the properties of the volume process. Therefore, the volume resting outside the actively modelled LOB levels $(-l_d + 1, \ldots, 0, \ldots, l_p)$ on the bid and ask is assumed to remain unchanged until the agent interactions brings those levels inside the band of actively modelled levels.

To present the details of the simulation framework, including the stochastic model components for each agent, i.e. liquidity providers and liquidity demanders, we first define the following notation:

1. $\boldsymbol{V}_t^a = (V_t^{a,-l_d+1}, \ldots, V_t^{a,l_p})$ - the random vector for the number of orders resting at each level on the ask side at time t at the actively modelled levels of the LOB at time t

2. $\boldsymbol{N}_t^{LO,a} = (N_t^{LO,a,-l_d+1}, \ldots, N_t^{LO,a,l_p})$ - the random vector for the number of limit orders entering the limit order book on the ask side at each level in the interval $[t-1, t)$

3. $\boldsymbol{N}_t^{C,a} = (N_t^{C,a,1}, \ldots, N_t^{C,a,l_t})$ - the random vector for the number of limit orders cancelled on the ask side in the interval $[t-1, t)$

4. $N_t^{MO,a}$ - the random variable for the number of market orders submitted by liquidity demanders in the interval $[t-1, t)$

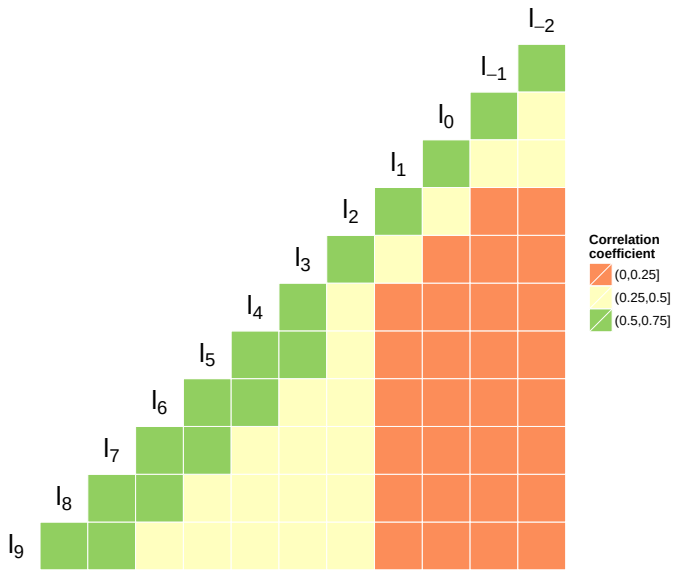


Figure 6.2: Correlation in the LOB limit order submission intensities on the bid side of the LOB in 10 second intervals, with the levels defined with respect to the best ask price. l_1 to l_9 denote passive levels (i.e. priced above the reference price) and l_0 to l_{-2} denote aggressive or direct levels (priced at or below the reference price, for immediate execution if the reference price had remained constant). The data set considered here is the daily LOB activity for stock BNP Paribas on 17/01/2012.

We consider the processes for limit orders and market orders, as well as cancellations to be linked to the behaviour of real market participants in the LOB. In the following, we model the aggregation of the activity of 2 classes of liquidity motivated agents, namely liquidity providers and liquidity demanders. As we model LOB activity in discrete time intervals, we process the aggregate activity at the end of each time interval in the following order:

1. Limit order arrivals - passive - by the liquidity provider agent.
2. Limit order arrivals - aggressive or direct - by the liquidity provider agent.
3. Cancellations by the liquidity provider agent.
4. Market orders by the liquidity demander agent.

The rationale for this ordering is that the vast majority of limit order submissions and cancellations is typically accounted for by the activity of high-frequency traders, and many resting orders are cancelled before slower traders can execute against them. In addition, such an ordering allows us to condition on the state of the LOB, so that we do not have more cancellations at a particular level than the orders resting at that level. We do not see this as a limitation, as the time interval we consider can be made as small as desired for a given simulation.

6.1.2 Stochastic agent representation: liquidity providers and demanders

We assume liquidity providers are responsible for all market-making behaviour (i.e. limit order submissions and cancellations on both the bid and ask side of the LOB). After liquidity is posted to the LOB, liquidity seeking market participants, such as mutual funds using some execution algorithm, can take advantage of the resting volume with market orders. For market makers, achieving a balance between volume executed on the bid and the ask side can be profitable; however, there is also the risk of adverse selection, i.e. trading against a trader with superior information, which may lead to losses if, e.g. a trader posts multiple market orders that consume the volume on several levels of the LOB. The risk of adverse selection as a result of asymmetric information is one of the basic tenets of market microstructure theory [O'hara, 1995]. To reduce this risk, market makers cancel and resubmit orders at different prices and/or different sizes.

Definition 6.1.1 (**Limit order submission process for the liquidity provider agent**). *Consider the limit order submission process of the liquidity provider agent to include both passive and aggressive limit orders on the bid and ask sides of the book, assumed to have the following stochastic model structure:*

1. *Let the multivariate path-space random matrix* $\boldsymbol{N}_{1:T}^{LO,k} \in \mathbb{N}_{+}^{l_t \times T}$ *be constructed from random vectors for the numbers of limit order placements* $\boldsymbol{N}_{1:T}^{LO,k} = \big(\boldsymbol{N}_1^{LO,k}, \boldsymbol{N}_2^{LO,k}, \ldots, \boldsymbol{N}_T^{LO,k}\big)$. *Furthermore, assume these random vectors for the number of orders at each level at time* t *are each conditionally dependent on a latent stochastic process for the intensity at which the limit orders arrive, given by the random matrix* $\boldsymbol{\Lambda}_{1:T}^{LO,k} \in \mathbb{R}_{+}^{l_t \times T}$ *and on the path-space by* $\boldsymbol{\Lambda}_{1:T}^{LO,k} = \Big(\boldsymbol{\Lambda}_1^{LO,k}, \boldsymbol{\Lambda}_2^{LO,k}, \ldots, \boldsymbol{\Lambda}_T^{LO,k}\Big)$. *In the following,* $k \in \{a, b\}$ *indicates the respective process on the ask and bid side.*

2. *Assume the conditional independence property for the random vectors*

$$\left[\boldsymbol{N}_s^{LO,k} | \boldsymbol{\Lambda}_s^{LO,k}\right] \perp\!\!\!\perp \left[\boldsymbol{N}_t^{LO,k} | \boldsymbol{\Lambda}_t^{LO,k}\right], \quad \forall s \neq t, \quad s, t \in \{1, 2, \ldots, T\}\,. \tag{6.1}$$

3. *For each time interval $[t-1,t)$ from the start of trading on the day, let the random vector for the number of new limit orders placed in each actively modelled level of the limit order book, i.e. the price points corresponding to ticks $(-l_d+1,\ldots,0,1,\ldots,l_p)$, as depicted in Figure 6.1, be denoted by $\boldsymbol{N}_t^{LO,k} = (N_t^{LO,k,-l_d+1},\ldots,N_t^{LO,k,l_p})$, and assume that these random vectors satisfy the conditional independence property*

$$\left[N_t^{LO,k,s}|\Lambda_t^{LO,k,s}\right] \perp\!\!\!\perp \left[N_t^{LO,k,q}|\Lambda_t^{LO,k,q}\right], \quad \forall s \neq q, \quad s,q \in \{-l_d+1,\ldots,0,1,\ldots,l_p\}.$$

4. *Assume the random vector $\boldsymbol{N}_t^{LO,k} \in \mathbb{N}_+^{l_t}$ is distributed according to a multivariate generalized Cox process with conditional distribution $\boldsymbol{N}_t^{LO,k} \sim \mathcal{GCP}\left(\boldsymbol{\lambda}_t^{LO,k}\right)$ given by*

$$\mathbb{P}\boldsymbol{r}\left(N_t^{LO,k,-l_d+1} = n_1,\ldots,N_t^{LO,k,l_p} = n_{l_t}\middle|\, \boldsymbol{\Lambda}_t^{LO,k} = \boldsymbol{\lambda}_t^{LO,k}\right) = \prod_{s=-l_d+1}^{l_p} \frac{\left(\lambda_t^{LO,k,s}\right)^{n_s}}{n_s!} \exp\left[-\lambda_t^{LO,k,s}\right] \tag{6.2}$$

5. *Assume the independence property for random vectors of latent intensities unconditionally according to*

$$\boldsymbol{\Lambda}_s^{LO,k} \perp\!\!\!\perp \boldsymbol{\Lambda}_t^{LO,k}, \quad \forall s \neq t, \quad s,t \in \{1,2,\ldots,T\}. \tag{6.3}$$

6. *Assume that the intensity random vector $\boldsymbol{\Lambda}_t^{LO,k} \in \mathbb{R}_+^{l_t}$ is obtained through an element-wise transformation of the random vector $\boldsymbol{\Gamma}_t^{LO,k} \in \mathbb{R}^{l_t}$, where for each element we have the mapping*

$$\Lambda_t^{LO,k,s} = \mu_0^{LO,k,s} F\left(\Gamma_t^{LO,k,s}\right) \tag{6.4}$$

where we have $s \in \{-l_d+1,\ldots,l_p\}$, baseline intensity parameters $\left\{\mu_0^{LO,k,s}\right\} \in \mathbb{R}_+$ and a strictly monotonic mapping $F : \mathbb{R} \mapsto [0,1]$.

7. *Assume the random vector $\boldsymbol{\Gamma}_t^{LO,k} \in \mathbb{R}$ is distributed according to a multivariate skew-t distribution $\boldsymbol{\Gamma}_t^{LO,k} \sim MSt(\boldsymbol{m}^k, \boldsymbol{\beta}^k, \nu^k, \Sigma^k)$ with location parameter vector $\boldsymbol{m}^k \in \mathbb{R}^{l_t}$, skewness parameter vector $\boldsymbol{\beta}^k \in \mathbb{R}^{l_t}$, degrees of freedom parameter $\nu^k \in \mathbb{N}_+$ and $l_t \times l_t$ covariance matrix Σ^k. Hence, $\boldsymbol{\Gamma}_t^{LO,k}$ has density function*

$$f_{\boldsymbol{\Gamma}_t^{LO,k}}\left(\boldsymbol{\gamma}_t; \boldsymbol{m}^k, \boldsymbol{\beta}^k, \nu^k, \Sigma^k\right) = \frac{cK_{\frac{\nu^k+l_t}{2}}\left(\sqrt{\left(\nu^k + Q(\boldsymbol{\gamma}_t, \boldsymbol{m}^k)\right)\left[\boldsymbol{\beta}^k\right]^T\left[\Sigma^k\right]^{-1}\boldsymbol{\beta}^k}\right)\exp\left(\boldsymbol{\gamma}_t - \boldsymbol{m}^k\right)^T\left[\Sigma^k\right]^{-1}\boldsymbol{\beta}^k}{\left(\sqrt{\left(\nu^k + Q(\boldsymbol{\gamma}_t, \boldsymbol{m}^k)\right)\left[\boldsymbol{\beta}^k\right]^T\left[\Sigma^k\right]^{-1}\boldsymbol{\beta}^k}\right)^{-\frac{\nu^k+l_t}{2}}\left(1+\frac{Q(\boldsymbol{\gamma}_t,\boldsymbol{m}^k)}{\nu^k}\right)^{\frac{\nu^k+l_t}{2}}} \tag{6.5}$$

where $K_v(z)$ is a modified Bessel function of the second kind given by

$$K_v(z) = \frac{1}{2}\int_0^\infty y^{v-1}e^{-\frac{z}{2}(y+y^{-1})}dy \tag{6.6}$$

and c is a normalisation constant. We also define the function $Q(\cdot,\cdot)$ as follows:

$$Q(\boldsymbol{\gamma}_t, \boldsymbol{m}^k) = (\boldsymbol{\gamma}_t - \boldsymbol{m}^k)^T \left[\Sigma^k\right]^{-1} (\boldsymbol{\gamma}_t - \boldsymbol{m}^k) \tag{6.7}$$

This model also admits skew-t marginals and a skew-t copula, see Smith et al. [2012] for details. Importantly, this stochastic model admits the following scale mixture representation,

$$\boldsymbol{\Gamma}_t^{LO,k} \overset{d}{=} \boldsymbol{m}^k + \boldsymbol{\beta}^k W + \sqrt{W}\boldsymbol{Z} \tag{6.8}$$

with Inverse-Gamma random variable $W \sim IGa\left(\frac{\nu^k}{2}, \frac{\nu^k}{2}\right)$ and independent Gaussian random vector $\boldsymbol{Z} \sim N\left(\boldsymbol{0}, \Sigma^k\right)$.

8. *Assume that for every element $N_t^{LO,k,s}$ of order counts from the random vector $\boldsymbol{N}_t^{LO,k}$, there is a corresponding random vector $\boldsymbol{O}_t^{LO,k,s} \in \mathbb{N}_+^{N_t^{LO,k,s}}$ of order sizes. We assume that the element $O_{i,t}^{LO,k,s}, i \in \left\{1,\ldots,N_t^{LO,k,s}\right\}$ is distributed as $O_{i,t}^{LO,k,s} \sim H(\cdot)$. Furthermore, we assume that order sizes are unconditionally independent $O_{i,t}^{LO,k,s} \perp\!\!\!\perp O_{i',t}^{LO,k,s}$ for $i \neq i'$, $s \neq s'$ and $t \neq t'$.*

We now define the second component of the liquidity provider agents, namely the cancellation process. The cancellation process has the same stochastic process model specification as the limit order submission process above, including a skew-t dependence structure between the stochastic intensities at each LOB level on the bid and ask. We therefore only specify the differences unique to the cancellation process relative to the order placement model definition in the below specification, to avoid repitition.

Definition 6.1.2 (Limit order cancellation process for liquidity provider agent). *Consider the limit order cancellation process of the liquidity provider agent to have an identically specified stochastic model structure as the limit order submissions. The exception to this pertains to the assumption that the number of cancelled orders in each interval at each level is right-truncated at the total number of orders at that level.*

1. *As for submissions, we assume for cancellations a multivariate path-space random matrix $\boldsymbol{N}_{1:T}^{C,k} \in \mathbb{N}_+^{l_t \times T}$ constructed from random vectors for the number of cancelled orders given by $\boldsymbol{N}_{1:T}^{C,k} = \left(\boldsymbol{N}_1^{C,k}, \boldsymbol{N}_2^{C,k}, \ldots, \boldsymbol{N}_T^{C,k}\right)$. Furthermore, assume for these random vectors for the number of cancelled orders at each of the l_t levels, the latent stochastic*

process for the intensity is given by the random matrix $\mathbf{\Lambda}^{C,k}_{1:T} \in \mathbb{R}^{l_t \times T}_{+}$ *and given on the path-space by* $\mathbf{\Lambda}^{C,k}_{1:T} = \left(\mathbf{\Lambda}^{C,k}_{1}, \mathbf{\Lambda}^{C,k}_{2}, \ldots, \mathbf{\Lambda}^{C,k}_{T}\right)$.

2. *Assume that for the random vector* $\tilde{\boldsymbol{V}}^{k}_{t}$ *for the volume resting in the LOB after the placement of limit orders we have* $\tilde{\boldsymbol{V}}^{k}_{t} = \boldsymbol{V}^{k}_{t-1} + \boldsymbol{N}^{LO,k}_{t}$, *and that the random vector* $\boldsymbol{N}^{C,k}_{t} \in \mathbb{N}^{l_t}_{+}$ *is distributed according to a truncated multivariate generalized Cox process with conditional distribution* $\boldsymbol{N}^{C,k}_{t} | \tilde{\boldsymbol{V}}^{k}_{t} = \underline{v} \sim \mathcal{GCP}\left(\boldsymbol{\lambda}^{C,k}_{t}\right) \mathbb{I}(\boldsymbol{N}^{C,k}_{t} < \underline{v})$ *(with* $\underline{v} = (v_{-l_d+1}, \ldots, v_{l_p})$*) given by*

$$\mathbb{P}r\left(N^{C,k,-l_d+1}_{t} = n_{-l_d+1}, \ldots, N^{C,k,l_p}_{t} = n_{l_p} \middle| \mathbf{\Lambda}^{C,k}_{t} = \boldsymbol{\lambda}^{C,k}_{t}, \tilde{\boldsymbol{V}}^{k}_{t} = \underline{v}\right) = \prod_{s=-l_d+1}^{l_p} \frac{\frac{(\lambda^{C,k,s}_{t})^{n_s}}{n_s!}}{\sum_{j=0}^{v_s} \frac{(\lambda^{C,k,s}_{t})^{j}}{j!}}. \tag{6.9}$$

3. *Assume that for the cancellation count* $N^{C,k,s}_{t}$, *the orders with highest priority are cancelled from level* s *(which are also the oldest orders in their respective queue). Assume also that cancellations always remove an order in full, i.e. there are no partial cancellations.*

We complete the specification of the representative agents by considering the specification of the liquidity demander agent.

Definition 6.1.3 (**Market order submission process for liquidity demander agent**). *Consider a representative agent for the liquidity providers to be composed of a* ***market order*** *component, which has the following stochastic structure:*

1. *Assume a path-space random vector* $N^{MO,k}_{1:T} \in \mathbb{N}^{1 \times T}_{+}$ *for the number of market orders constructed from the random variables for the number of market orders in each interval* $N^{MO,k}_{1:T} = \left(N^{MO,k}_{1}, N^{MO,k}_{2}, \ldots, N^{MO,k}_{T}\right)$. *Furthermore, assume that for these random variables the latent stochastic process for the intensity is given by random variable* $\Lambda^{MO,k}_{1:T} \in \mathbb{R}^{l_t \times T}_{+}$, *and given on the path-space by* $\Lambda^{MO,k}_{1:T} = \left(\Lambda^{MO,k}_{1}, \Lambda^{MO,k}_{2}, \ldots, \Lambda^{MO,k}_{T}\right)$.

2. *Assume the conditional independence property for the random variables*

$$\left[N^{MO,k}_{s} | \Lambda^{MO,k}_{s}\right] \perp\!\!\!\perp \left[N^{MO,k}_{t} | \Lambda^{MO,k}_{t}\right], \quad \forall s \neq t, \quad s,t \in \{1,2,\ldots,T\}. \tag{6.10}$$

3. *Assume that for the random variable* $\tilde{R}^{k}_{t}$ *for the volume resting on the opposite side of the LOB after the placement of limit orders and cancellations we have* $\tilde{R}^{k}_{t} =$

$\Sigma_{s=1}^{l_p}\left[\tilde{V}_{t-\Delta t}^{k',s} - N_t^{C,k',s}\right]$, *where* $k' = a$ *if* $k = b$, *and vice-versa, and that the random variable* $N_t^{MO,k} \in \mathbb{N}_+$ *is distributed according to a truncated generalized Cox process with conditional distribution* $N_t^{MO,k}|\tilde{R}_t^k = r \sim \mathcal{GCP}\left(\lambda_t^{MO,k}\right)\mathbb{I}(N_t^{MO,k} < r)$ *given by*

$$\mathbb{P}r\left(N_t^{MO,k} = n \middle| \Lambda_t^{MO,k} = \lambda_t^{MO,k}, \tilde{R}_t^k = r\right) = \frac{\frac{(\lambda_t^{MO,k})^n}{n!}}{\sum_{j=0}^{r}\frac{(\lambda_t^{MO,k})^j}{j!}}. \tag{6.11}$$

4. *Assume the independence property for random vectors of latent intensities unconditionally according to*

$$\Lambda_s^{MO,k} \perp\!\!\!\perp \Lambda_t^{MO,k}, \;\; \forall s \neq t, \;\; s,t \in \{1,2,\ldots,T\}. \tag{6.12}$$

5. *Assume that for each intensity random variable* $\Lambda_t^{MO,k} \in \mathbb{R}_+$ *there is a corresponding transformed intensity variable* $\Gamma_t^{MO,k} \in \mathbb{R}$ *and the relationship for each element is given by the mapping*

$$\Lambda_t^{MO,k} = \mu_0^{MO,k} F\left(\Gamma_t^{MO,k}\right) \tag{6.13}$$

for some baseline intensity parameter $\mu_0^{MO,k} \in \mathbb{R}_+$ *and strictly monotonic mapping* $F : \mathbb{R} \mapsto [0,1]$.

6. *Assume that the random variables* $\Gamma_t^{MO,k} \in \mathbb{R}$, *characterizing the intensity before transformation of the Generalized Cox-Process, are distributed in interval* $[t-1,t)$ *according to a univariate skew-t distribution* $\Gamma_t^{MO,k} \sim St(m_t^{MO,k}, \beta^{MO,k}, \nu^{MO,k}, \sigma^{MO,k})$.

7. *Assume that for every element* $N_t^{MO,k}$ *of market order counts, there is a corresponding random vector* $\boldsymbol{O}_t^{MO,k,s} \in \mathbb{N}_+^{N_t^{MO,k}}$ *of order sizes. We assume that the element* $O_{i,t}^{MO,k}, i \in \left\{1,\ldots,N_t^{MO,k}\right\}$ *is distributed according to* $O_{i,t}^{MO,k} \sim H(\cdot)$. *Assume also that market order sizes are unconditionally independent* $O_{i,t}^{MO,k} \perp\!\!\!\perp O_{i',t}^{MO,k}$ *for* $i \neq i'$ *or* $t \neq t'$.

We denote the LOB state for the real dataset at time t on a given day by the random vector $\boldsymbol{L}_t$, and this corresponds to the prices and volumes at each level of the bid and ask. Utilising the stochastic agent-based model specification described above, and given a parameter vector $\boldsymbol{\theta}$, which will generically represent all parameters of the liquidity providing and liquidity demanding agent types, one can then also generate simulations of intra-day LOB activity and arrive at the synthetic state $\boldsymbol{L}_t^*(\boldsymbol{\theta})$. The state of the simulated LOB at time t is obtained from the state at time $t-1$ and a set of stochastic components, denoted generically by $\boldsymbol{X}_t$, which are obtained from a single stochastic realisation of the following components of the agent-based models:

- Limit order submission intensities $\boldsymbol{\Lambda}_t^{LO,b}$, $\boldsymbol{\Lambda}_t^{LO,a}$, order numbers $\boldsymbol{N}_t^{LO,b}$, $\boldsymbol{N}_t^{LO,a}$, and order sizes $\boldsymbol{O}_{i,t}^{LO,a,s}, \boldsymbol{O}_{j,t}^{LO,b,s}$, where $s = -l_d + 1 \dots l_p, i = 1 \dots N_t^{LO,a,s}, j = 1 \dots N_t^{LO,b,s}$.

- Limit order cancellation intensities $\boldsymbol{\Lambda}_t^{C,b}$, $\boldsymbol{\Lambda}_t^{C,a}$ and numbers of cancellations $\boldsymbol{N}_t^{C,b}$, $\boldsymbol{N}_t^{C,a}$.

- Market order intensities $\boldsymbol{\Lambda}_t^{MO,b}$, $\boldsymbol{\Lambda}_t^{MO,a}$, numbers of market orders $\boldsymbol{N}_t^{MO,b}$, $\boldsymbol{N}_t^{MO,a}$ and market order sizes $\boldsymbol{O}_{i,t}^{MO,a}, \boldsymbol{O}_{j,t}^{MO,b}, i = 1 \dots N_t^{MO,a}, j = 1 \dots N_t^{MO,b}$.

These stochastic features are combined with the previous state of the LOB, $\boldsymbol{L}_{t-1}^*(\boldsymbol{\theta})$, to produce the new state $\boldsymbol{L}_t^*(\boldsymbol{\theta})$ for a given set of parameters $\boldsymbol{\theta}$, given by

$$\boldsymbol{L}_t^*(\boldsymbol{\theta}) = G(\boldsymbol{L}_{t-1}^*(\boldsymbol{\theta}), \boldsymbol{X}_t). \tag{6.14}$$

$G(\cdot)$ is a transformation that maps the previous state of the LOB and the activity generated in the current step into a new step, much the same way as the matching engine updates the LOB after every event. As we model the activity in discrete intervals, however, the LOB is only updated at the end of every interval, and the incoming events (limit orders, market orders and cancellations) are processed in the order specified in Section 6.1.1. Conditional then on a realization of these parameters $\boldsymbol{\theta}$, the trading activity in the LOB can be simulated according to the procedure described in Algorithm 1.

6.2 Simulation based likelihood calibration

A common attribute of all agent-based modelling frameworks is that they are able to generate realisations of the stochastic process they represent, in our case the LOB process. That is, given a set of specifications for the parameters of the agents, the simulation of the agent model is trivial and efficient. However, it is also commonly the case that there is either no direct tractable (to evaluate pointwise) likelihood model or the likelihood model is complex and computationally costly to evaluate. In these cases, traditional parameter estimation methods based on likelihood inference are not directly applicable, when calibrating such models to observed LOB data. There are, however, a range of methods, which have yet to be utilised widely in the agent-based modelling literature, that allow one to still perform calibration of models, i.e. parameter estimation, for models specified in a simulation based format.

The structure of our model ensures that we can capture features such as the non-linear dependencies between the activity at different LOB levels. This activity includes limit order submissions that can be passive or aggressive, cancellations and market orders, and can arise

procedure SIMULATE($\boldsymbol{\theta}, T$)
  for $t = 1 \dots T$ **do**
    ▷ ***Simulate Liquidity Provider Limit Orders Bid/Ask.***
    for $k = a, b$ **do**
      ▷ *Simulate dependent stochastic intensities for limit order submissions.*
      Sample $\boldsymbol{\Gamma}_t^{LO,k} = \boldsymbol{\gamma}_t^{LO,k} \sim MSt(\boldsymbol{m}^k, \boldsymbol{\beta}^k, \nu^k, \Sigma^k)$ via Equation 6.8.
      Apply transformation $\boldsymbol{\lambda}_t^{LO,k} = \boldsymbol{\mu}_0^k F(\boldsymbol{\gamma}_t^{LO,k})$ in Equation 6.4.
      ▷ *Simulate dependent limit order counts at each level bid/ask.*
      Sample $\boldsymbol{N}_t^{LO,k} = \boldsymbol{n}_t^{LO,k} \sim \mathcal{GCP}\left(\boldsymbol{\lambda}_t^{LO,k}\right)$ via Equation 6.2.
      ▷ *Simulate limit order sizes.*
      for $s = -l_d + 1, \dots l_p, i = 1 \dots N_t^{LO,k,s}$ **do**
        $\boldsymbol{O}_{i,t}^{LO,k,s} \sim H(\cdot)$
    ▷ ***Simulate Liquidity Provider Cancelled Limit Orders Bid/Ask.***
    for $k = a, b$ **do**
      ▷ *Evaluate total volumes at each level bid and ask.*
      $\tilde{\boldsymbol{V}}_t^{LO,k} = \boldsymbol{V}_{t-1}^{LO,k} + \boldsymbol{N}_t^{LO,k} = \tilde{\boldsymbol{v}}_t^{LO,k}$
      ▷ *Simulate dependent stochastic intensity for bid and ask cancellation counts.*
      Sample $\boldsymbol{\Gamma}_t^{C,k} = \boldsymbol{\gamma}_t^{C,k} \sim MSt(\boldsymbol{m}^{C,k}, \boldsymbol{\beta}^{C,k}, \nu^{C,k}, \Sigma^{C,k})$ via Equation 6.8.
      Apply transformation $\boldsymbol{\lambda}_t^{C,k} = \boldsymbol{\mu}_0^{C,k} F(\boldsymbol{\gamma}_t^{C,k})$ in Equation 6.4.
      ▷ *Simulate dependent limit order cancellation counts at each level of the bid/ask.*
      Sample $\boldsymbol{N}_t^{C,k} = \boldsymbol{n}_t^{C,k} \sim \mathcal{GCP}\left(\boldsymbol{\lambda}_t^{C,k}\right) \mathbb{I}(\boldsymbol{N}_t^{C,k} < \tilde{\boldsymbol{v}}_t^{LO,k})$ via Equation 6.9.
    ▷ ***Simulate Liquidity Demander Market Orders.***
    for $k = a, b$ **do**
      ▷ *Evaluate the current resting volumes on each level of the bid/ask.*
      $\tilde{R}_t^{LO,k} = \Sigma_{s=1}^{l_p}\left[\tilde{V}_t^{LO,k',s} - N_t^{C,k',s}\right] = \tilde{r}_t^{LO,k}$
      ▷ *Simulate stochastic intensities for market order submissions.*
      Sample $\gamma^{MO,k} \sim St(m_t^{MO,k}, \beta^{MO,k}, \nu^{MO,k}, \sigma^{MO,k})$ from skew-t distribution.
      Evaluate transformation $\lambda_t^{MO,k} = \mu_0^{MO,k} F(\gamma_t^{MO,k})$ in Equation 6.13.
      ▷ *Simulate market order counts.*
      Sample $N_t^{MO,k} | \tilde{r}_t^{LO,k} \sim \mathcal{GCP}\left(\boldsymbol{\lambda}_t^{MO,k}\right) \mathbb{I}(\boldsymbol{N}_t^{MO,k} < \tilde{r}_t^{LO,k})$ via Equation 6.11.
      ▷ *Simulate market order sizes.*
      for $i = 1 \dots N_t^{MO,k}$ **do**
        $\boldsymbol{O}_{i,t}^{MO,k} \sim H(\cdot)$
    $L_t \leftarrow \mathrm{G}(L_{t-1}, \boldsymbol{N}_t^{LO,a}, \boldsymbol{N}_t^{LO,b}, \boldsymbol{N}_t^{C,a}, \boldsymbol{N}_t^{C,b}, N_t^{MO,a}, N_t^{MO,a}, \boldsymbol{O}_t^{LO,a}, \boldsymbol{O}_t^{LO,b}, \boldsymbol{O}_t^{MO,a}, \boldsymbol{O}_t^{MO,b})$
  return $\boldsymbol{L} = \{L_1, \dots, L_T\}$

Algorithm 1: Stochastic agent-based LOB simulation

from two different classes of agents. Given this complexity, obtaining the distributional form of the likelihood will be impossible.

6.2.1 Background on Indirect Inference

There is a substantial body of academic work related to simulation-based likelihood inference, and we focus on the subclass known as Indirect Inference, introduced by Smith [1990, 1993] and Gourieroux et al. [1993] and covered extensively in Gouriéroux et al. [2010], Gallant and Tauchen [1996] and the book length coverage in Gourieroux and Monfort [1997]. At its most fundamental level, Indirect Inference is a technique for parameter estimation in simulation based stochastic models. These are models for which one cannot evaluate the density for the data generating model, but for which one can generate data given a set of parameters. One can then compare the simulated data with the observed data, and obtain a measure of fitness for a set of parameters based on this comparison.

To achieve this via Indirect Inference, one introduces a new model, called the 'auxiliary model', which is mis-specified and typically not even generative, but can generally be estimated easily via for instance maximum likelihood estimation. This auxiliary model has its own parameter vector $\boldsymbol{\beta}$, with point estimator $\widehat{\boldsymbol{\beta}}$. These parameters of the auxiliary model describe aspects of the distributions of the observations. The idea of Indirect Inference is then to simply try to match aspects of the estimated auxiliary model parameters on the observed data $\boldsymbol{y}$, given by $\widehat{\boldsymbol{\beta}}(\boldsymbol{y})$, and the estimated auxiliary model parameters on the simulated data $\boldsymbol{y}^*(\boldsymbol{\theta})$, which is obtained through simulation using parameters of the actual model $\boldsymbol{\theta}$, given by $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^*(\boldsymbol{\theta}))$.

One sees that Indirect Inference only requires that the model one wants to estimate can be simulated, and proceeds by fitting a simpler auxiliary model to both the simulated and the real data. Estimates of the model parameters are then obtained by minimising the difference between the parameter vectors of the auxiliary model fit to the simulated data and the real data.

When considering the choice of an auxiliary model, the simplest form one may consider involves a comparison formed between a single summary statistic calculated on the real observed data, say $\boldsymbol{y}$ and also on the simulated synthetic data $\boldsymbol{y}^*$. Alternatively, one may consider methods that consider the use of a vector of summary auxiliary parameters, such as in Winker et al. [2007] who consider minimization of a weighted L2 quadratic error function between the real data vector of estimated moments and the synthetic simulated data equivalents. Others who have adopted such methods include McFadden [1989] and Pakes and Pollard [1989] who each proposed a modification of the method of moments estimator, called the Method of Simulated Moments (MSM). Alternative simulation-based estimation techniques include the simulated

maximum likelihood (SML) and the method of simulated scores (MSS). Such techniques have been used in the estimation of a number of economic models, for example dynamic stochastic general equilibrium (DSGE) models Ruge-Murcia [2007] and Markov models of asset pricing Duffie and Singleton [1993].

In this section, the auxiliary models we consider are based on aspects of the LOB stochastic process that we analyze. The key features we consider include the variation in the price and the volume resting in the LOB. In particular, we would like to capture the clustering of volatility in intra-day log returns and the dynamic behaviour of total volume in the first n levels of the LOB.

In detail, the sequence for obtaining the Indirect Inference estimator is as follows:

1. Take the observed sequence of LOB states $\boldsymbol{L}_{1:T}$ and transform them to auxiliary model data set $\boldsymbol{y} = \mathcal{T}\left(\boldsymbol{L}_{1:T}\right)$.

2. Using observed auxiliary model data $\boldsymbol{y}$, estimate auxiliary model parameters $\widehat{\boldsymbol{\beta}}\left(\boldsymbol{y}\right)$.

3. Initialize parameter vector of stochastic agent LOB model, in our case liquidity provider and liquidity demander agent models parameters $\boldsymbol{\theta}^{(0)}$. Then simulate a synthetic realization of the LOB model $\boldsymbol{L}^{*}_{1:T}\left(\boldsymbol{\theta}^{(0)}\right)$ from the stochastic agent model.

4. Take the synthetic sequence of LOB states $\boldsymbol{L}^{*}_{1:T}\left(\boldsymbol{\theta}^{(0)}\right)$ and transform them to auxiliary model data set $\boldsymbol{y}^{*}(\boldsymbol{\theta}^{(0)}) = \mathcal{T}\left(\boldsymbol{L}^{*}_{1:T}\left(\boldsymbol{\theta}^{(0)}\right)\right)$.

5. Using synthetic auxiliary model data $\boldsymbol{y}^{*}(\boldsymbol{\theta}^{(0)})$, estimate auxiliary model parameters $\widehat{\boldsymbol{\beta}}_0\left(\boldsymbol{y}^{*}(\boldsymbol{\theta}^{(0)})\right)$.

6. Estimate Mahalanobis distance or Euclidean distances between auxiliary parameter vectors $\mathcal{D}\left(\widehat{\boldsymbol{\beta}}\left(\boldsymbol{y}\right), \widehat{\boldsymbol{\beta}}_0(\boldsymbol{y}^{*}(\boldsymbol{\theta}^{(0)}))\right)$

7. Set optimal parameter vector $\widehat{\boldsymbol{\theta}}^{opt} = \boldsymbol{\theta}^{(0)}$ with distance $\mathcal{D}_{\text{min}} = \mathcal{D}\left(\widehat{\boldsymbol{\beta}}\left(\boldsymbol{y}\right), \widehat{\boldsymbol{\beta}}_0(\boldsymbol{y}^{*}(\boldsymbol{\theta}^{(0)}))\right)$.

8. Repeat steps 3 to 7 with proposed parameter vector $\boldsymbol{\theta}^{(j)}$ until convergence or for J total iterations, with step (vii) applied conditionally on the event
$\mathcal{D}_{\text{min}} > \mathcal{D}\left(\widehat{\boldsymbol{\beta}}\left(\boldsymbol{y}\right), \widehat{\boldsymbol{\beta}}_j\left(\boldsymbol{y}^{*}\left(\boldsymbol{\theta}^{(j)}\right)\right)\right)$

Several theoretical properties are known about the estimators obtained from such a data generative procedure, see discussions in Smith [2008] and Genton and Ronchetti [2003]. Under certain assumptions it can be shown that the Indirect Inference procedure produces a point

estimator of the model parameters which is both consistent and asymptotically Normal under fairly unrestrictive regularity conditions (Gourieroux and Monfort [1997]):

1. The likelihood, which we maximise, in order to estimate the auxiliary model parameters $\boldsymbol{\beta}$, tends asymptotically to a non-stochastic limit.

2. This limit is continuous in the simulation model parameters $\boldsymbol{\theta}$.

3. The so-called *binding function* linking the parameters of the auxiliary model to the parameters of the actual model we are trying to estimate is one-to-one and its derivative with respect to the auxiliary model parameters is of full column rank.

In addition, Indirect Inference can be shown to be asymptotically efficient when the model is correctly specified for the observed data.

6.2.2 Multi-objective Indirect Inference for simulation-based model calibration

To perform estimation of our model, we develop a novel extension of simulation-based estimation procedures which combines two key ideas: simulation-based likelihood inference based on Indirect Inference, and multi-objective optimisation methods, typically utilised in genetic search algorithms. We denote the resulting class of estimation methods as Multi-objective-II. The proposed Multi-objective-II estimation framework, unlike standard indirect inference, is designed to allow one to utilise multiple auxiliary models, each capturing different features of the LOB stochastic process. In this sense, this is a multi-objective extension of standard Indirect Inference procedures, which will naturally allow us to explore relevant features of the target stochastic process given by the LOB.

To proceed with the specification of the multi-objective-II estimation methodology, in addition to the LOB simulation framework described in Section 6.1, we need to specify

- The auxiliary model(s), each parameterised by a set of parameter vectors, generically denoted by $\boldsymbol{\beta}$, which are determined according to the features of the observed data stochastic process we would like to approximate with our model.

- The objective function quantifying the difference in the auxiliary model(s) parameters fit to the real data (for which we will use the shorthand $\widehat{\boldsymbol{\beta}}$ to represent $\widehat{\boldsymbol{\beta}}(\boldsymbol{y})$) and the auxiliary model(s) parameter fits to the synthetically generated data (where we will use the shorthand $\widehat{\boldsymbol{\beta}}^*(\boldsymbol{\theta})$ for $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^*(\boldsymbol{\theta}))$

- The search method that will explore the parameter space of the stochastic agent-based model when performing simulation based optimization for model calibration.

6.2.2.1 The auxiliary models

The auxiliary model(s), sometimes known as the estimating function(s), serve to capture aspects of the real data that we want reflected in our simulation, i.e. they do not necessarily have to correspond closely to the data generating process, but each should capture some relevant features that will inform estimation of the stochastic simulation model parameters. In standard Indirect Inference methods, there is only one auxiliary model utilised, which usually comes from a relatively simple class of models, for guidelines relating to selection see Heggland and Frigessi [2004].

In our framework, for a given candidate parameter vector $\boldsymbol{\theta}$ we generate M realisations of trajectories of the LOB process, i.e. $\left\{\boldsymbol{L}_t^{*,m}(\boldsymbol{\theta})\right\}_{t>0,m\in\{1,2,\ldots,M\}}$, from the stochastic agent-based LOB model. Then for each auxiliary model, parameterised by some vector, generically denoted by $\boldsymbol{\beta}$, we utilise the simulated data to obtain estimates of the auxiliary model parameters, for instance via a maximum likelihood framework:

$$\widehat{\boldsymbol{\beta}}^{*}\left(\boldsymbol{\theta}\right)=\arg\max_{\boldsymbol{\beta}}\sum_{m=1}^{M}\sum_{i=1}^{T}\log(f(\mathcal{T}(\boldsymbol{L}_t^{m}(\boldsymbol{\theta}))|\mathcal{T}(\boldsymbol{L}_{t-1}^{*,m}(\boldsymbol{\theta}));\boldsymbol{\beta})). \tag{6.15}$$

In principle, one can adopt as many auxiliary models as is deemed desirable for a particular application. However, several authors have explored the effect of the number of objective functions K on the estimation performance under a multi-objective optimization framework. For instance, Purshouse and Fleming [2003] and Hughes [2005] suggest that Pareto-ranking based methods, such as the one used in this paper, scale poorly with the number of objectives. Köppen et al. [2005] explains that an increase in the number of objectives may have a detrimental effect on the optimisation because the probability of dominance in a Pareto optimality based multi-objective framework will go to zero. A second issue with having a large number of objectives is the difficulty in comparing the results qualitatively, since in a task with K objectives, a set of solutions lies in a $K-1$ hyperspace. Based on this guidance, we focus on capturing two core features of LOB stochastic process, related to the evolution of the price and the properties of the volume resting near the top of the book.

Auxiliary Model 1 - Price features: If we denote the mid-price as $P_t^{mid}=\frac{P_t^{a,1}+P_t^{b,1}}{2}$ then the log return is defined as

$$R_t=\ln\frac{P_t^{mid}}{P_{t-\Delta_t}^{mid}}$$

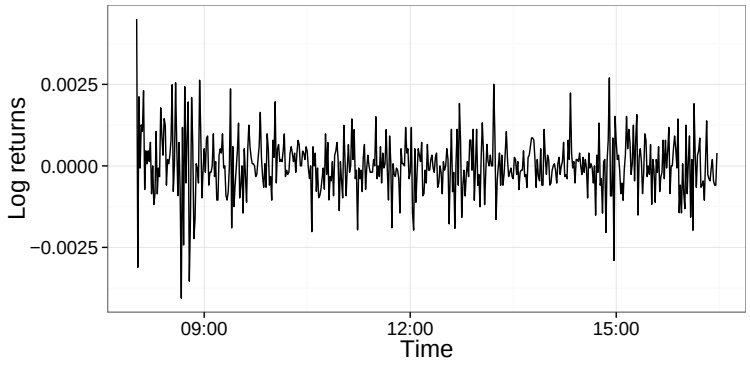


Figure 6.3: One-minute log returns for stock BNP Paribas on a typical day.

where Δ_t is a suitable interval, in our case 1 minute. The timeseries of log returns for a typical day for an illustrative stock GDF Suez is presented in Figure 6.3.

This illustrative timeseries displays typical features of mid price dynamics, such as heteroskedasticity. The presence of ARCH effects was formally confirmed by an ARCH-LM test. Hence, the volatility $\sigma_t = \sqrt{Var(R_t|R_{t-1}, \ldots)}$ is not constant, and can be captured with a generalised autoregressive conditionally heteroskedastic model, or GARCH(p,q) model, where with $R_t = \sigma_t \eta_t$ and $\eta_t \sim N(0,1)$, we have for the squared volatility

$$\sigma_t^2 = a_0 + a_1 R_{t-1}^2 + \ldots + a_p R_{t-p}^2 + b_1 \sigma_{t-1}^2 + \ldots + b_q \sigma_{t-q}^2$$

where $a_i \geq 0$, $b_j \geq 0$ for all $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, q\}$. For simplicity of the auxiliary model we utilise a GARCH(1,1) model for this aspect of the data, parameterized by $\boldsymbol{\beta}_1 = (a_0, a_1, b_1)$.

Auxiliary Model 2 - Volume features: In Figure 6.4 we demonstrate an example of the volume on the bid and ask side for a typical day for stock GDF Suez. We fit an ARIMA model to this data, in order to capture the time series structure of the LOB volumes. We will err on the side of parsimony during model identification, as we would like to obtain an auxiliary model with few parameters in our Indirect Inference procedure.

We first remove observed linear trends present in the LOB volume timeseries throughout the day by taking first differences, see Figure 6.4. The resulting sample ACF and PACF is given in Figure 6.4 and it indicates that an MA(1) model is appropriate. Hence, we fit an ARIMA(0,1,1) model to the volume data.

6.2.2.2 Combining multi-objective optimisation and Indirect Inference

Thus far, for a given set of parameters in our stochastic LOB agent model, we have simulated the order book process. This simulated data was then utilised to construct a framework in which

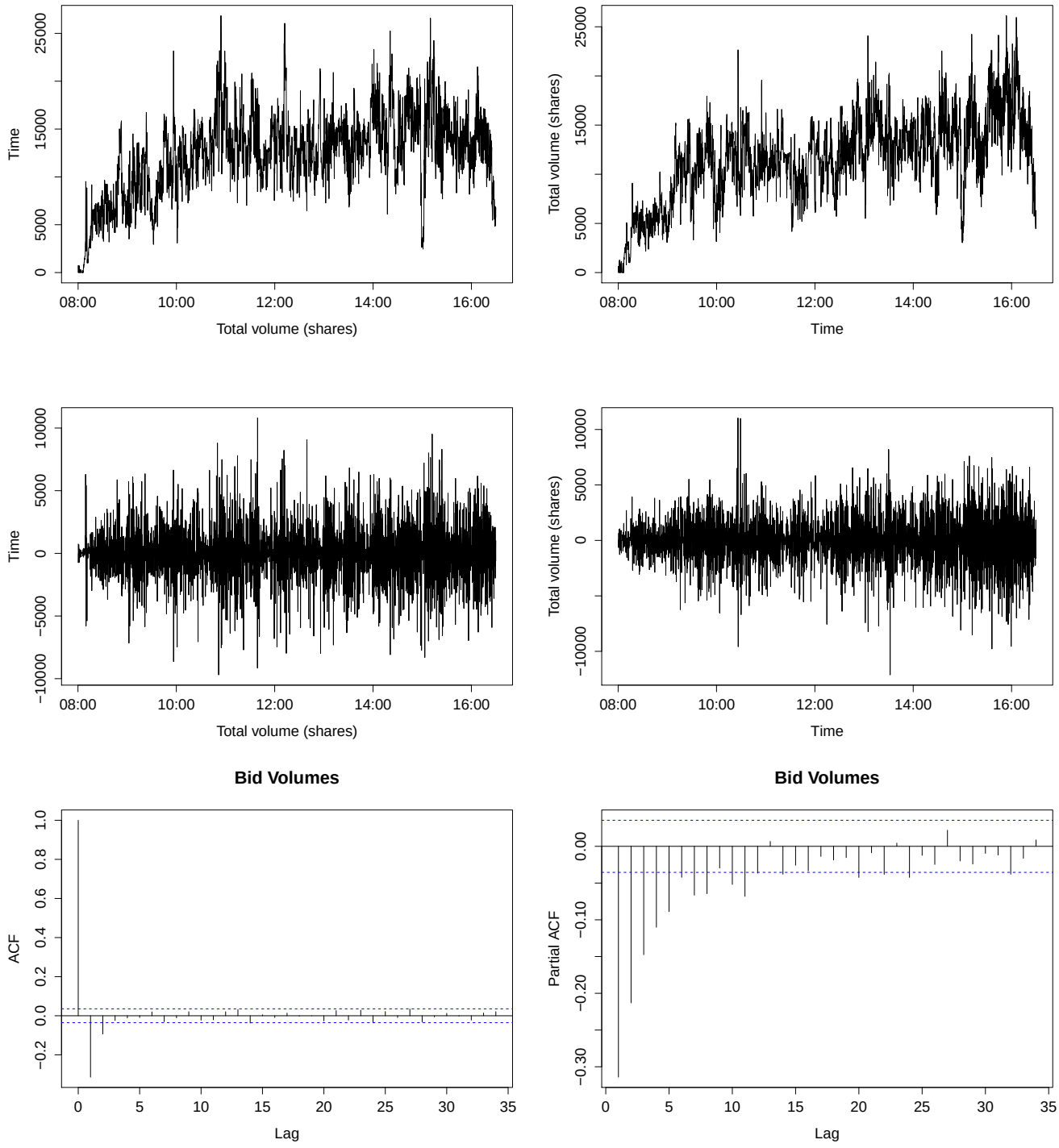


Figure 6.4: Top row subplots: Total sell volume resting in the LOB in the first 5 ticks away from the best bid (left), total buy volume resting in the first 5 ticks away from the best ask (right) for stock GDF Suez on a typical day. Middle Row Subplots: First differences of figures above. Bottom Row Subplots: Sample ACF and PACF.

we obtained multiple fitted parameter vectors, one for each auxiliary model considered. We now need to consider how to judge the suitability of the model parameter vector in capturing the true observed LOB stochastic process dynamics.

In standard Indirect Inference based frameworks, one would concatenate all the auxiliary model output parameter vector estimates into a single vector of auxiliary model parameters, in order to produce a single distance measure or discrepancy between the simulated data and actual data. This concatenation induces a loss in information, as for instance some auxiliary parameter model discrepancies may be on different scales to others. Therefore, if a naïve concatenation is applied, this often results in domination of a select few criteria, rather than considering each component in its own right.

We overcome this issue through the introduction of a multi-objective optimization framework. Such methods naturally adapt the simulation-based estimation to allow for competing criteria when assessing the suitability of the stochastic agent LOB model parameters via a collection of auxiliary model fits. The multi-objective optimisation method thus enables us to consider multiple distance measures, of discrepancy scores, as separate objective functions.

In this framework, the fitness, or suitability, of a parameter vector $\boldsymbol{\theta}$ for the stochastic agent LOB model is measured by simulating from the generative model and quantifying the difference between each auxiliary model's parameters. Each auxiliary model is fit to both the tranformation of the observed data to obtain ($\widehat{\boldsymbol{\beta}}_k$) and to the transformation of the simulated LOB data ($\widehat{\boldsymbol{\beta}}_k^*(\boldsymbol{\theta})$), for which a discrepancy score is calculated by measuring the distance between the two. The most commonly utilised distance measures are based on some form of weighted or unweighted norm, such as the L_p-norm, or Minkowski distance of order p, of which the L_∞-norm, the L_1-norm and the L_2-norm are frequently used in practice. We adopted the L_2-norm to measure discrepancies for both the price and volume-based auxiliary models we considered, given for the k-th auxiliary model by

$$\mathcal{D}_k(\boldsymbol{\theta}) = \mathcal{D}\left(\widehat{\boldsymbol{\beta}}_k, \widehat{\boldsymbol{\beta}}_k^*(\boldsymbol{\theta})\right) = \sum_{i=1}^{q_k}\left(\left[\widehat{\boldsymbol{\beta}}_k\right]_i - \left[\widehat{\boldsymbol{\beta}}_k^*(\boldsymbol{\theta})\right]_i\right)^2. \tag{6.16}$$

for each q_k-dimensional auxiliary model, $k = 1, \ldots, K$.

6.2.2.3 Multi-objective optimisation and the role of Pareto optimality

When our search is for an optimal parameter vector $\boldsymbol{\theta}$ that should satisfy multiple objective functions, in a vector $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) := [\mathcal{D}_1(\boldsymbol{\theta}), \ldots, \mathcal{D}_K(\boldsymbol{\theta})]$ to be minimised, there are many cases where there will not be a global minimum with respect to each individual objective. In this case, one can consider as an alternative to the single optimal value produced by an optimisation method, the notion of *Pareto optimality*, in reference to the Pareto efficient frontier. Informally,

this is the search for solutions such that there is no solution in the search space that can unilaterally improve a single criterion (objective function) without worsening another criterion, and this is formally defined in Definition 6.2.1 for the case of our estimation framework.

Definition 6.2.1 (Pareto Optimal Dominance of Parameter Solutions)**.** *Consider the set of K auxiliary models producing parameter vectors $\{\boldsymbol{\beta}_k\}_{k\in\{1,2,\ldots,K\}}$, each based on an underlying parameter vector $\boldsymbol{\theta} \in \Omega$, that produce, for selected objective functions, the values $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) := [\mathcal{D}_1(\boldsymbol{\theta}), \ldots, \mathcal{D}_K(\boldsymbol{\theta})]$. Then the selection of $\boldsymbol{\theta} \in \Omega$ is called Pareto-optimal or (non-dominated) with respect to the set of solutions in the feasible region Ω, if*

$$\nexists \widetilde{\boldsymbol{\theta}} \in \Omega \;\; s.t \;\; \boldsymbol{\mathcal{D}}(\widetilde{\boldsymbol{\theta}}) \prec \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}), \tag{6.17}$$

where we say that $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta})$ dominates $\boldsymbol{\mathcal{D}}(\widetilde{\boldsymbol{\theta}})$, denoted by $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) \prec \boldsymbol{\mathcal{D}}(\widetilde{\boldsymbol{\theta}})$, if

$$\mathcal{D}_k(\boldsymbol{\theta}) \leq \mathcal{D}_k(\widetilde{\boldsymbol{\theta}}) \;\; \forall k \in \{1, 2, \ldots, K\} \;\; \textit{and} \;\; \exists k \;\; s.t. \;\; \mathcal{D}_k(\boldsymbol{\theta}) < \mathcal{D}_k(\widetilde{\boldsymbol{\theta}}). \tag{6.18}$$

From this, we can then state the overall objective, incorporating all K auxiliary models and a common selection of $L2$-norm objective functions for the parameter vector $\boldsymbol{\theta}$ of the stochastic agent-based model as follows

$$\begin{aligned}
\widehat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}\in\Omega} \{\mathcal{D}_1(\boldsymbol{\theta}), \ldots, \mathcal{D}_K(\boldsymbol{\theta})\} \\
&= \arg\min_{\boldsymbol{\theta}\in\Omega} \left\{\mathcal{D}\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_1^*(\boldsymbol{\theta})\right), \ldots, \mathcal{D}\left(\hat{\boldsymbol{\beta}}_K, \hat{\boldsymbol{\beta}}_K^*(\boldsymbol{\theta})\right)\right\} \\
&= \arg\min_{\boldsymbol{\theta}\in\Omega} \left\{\sum_{i=1}^{q_1}\left(\left[\hat{\boldsymbol{\beta}}_1\right]_i - \left[\hat{\boldsymbol{\beta}}_1^*(\boldsymbol{\theta})\right]_i\right)^2, \ldots, \sum_{i=1}^{q_K}\left(\left[\hat{\boldsymbol{\beta}}_K\right]_i - \left[\hat{\boldsymbol{\beta}}_K^*(\boldsymbol{\theta})\right]_i\right)^2\right\} \\
&\quad \text{subject to} \quad \theta_{1_L} \leq \theta_1 \leq \theta_{1_U}, ..., \theta_{n_L} \leq \theta_n \leq \theta_{n_U}
\end{aligned} \tag{6.19}$$

where it is understood that this is a joint minimisation of the objective functions and that comparing two parameter vectors $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ is in terms of their relative domination, as per Equation 6.18. Here $[\theta_{i_L}, \theta_{i_U}]$, for all i, denote the boundaries of the feasible region Ω.

To complete the specification of the multi-objective Indirect Inference simulation based estimation framework we propose, we require a method to search the constrained parameter space Ω for feasible and Pareto optimal solutions. A variety of stochastic search methods are available for use in this context, see discussion in Coello et al. [2007].

We propose the use of an evolutionary genetic search method for this purpose, known in the literature as Multi-Objective Evolutionary Algorithms (MOEAs). We develop a version of such a stochastic search framework which combines the widely utilised NSGA-II genetic search algorithm by Deb et al. [2002], which is a Pareto-ranking based method, with an additional mutation kernel we designed specifically for a covariance matrix mutation operator, based on the

framework developed in Peters et al. [2012]. This additional mutation component is combined with the framework of NSGA-II, to ensure that the proposed covariance matrices in the stochastic agent LOB model, which are proposed at each step of the search, remain positive definite and symmetric. Details of this genetic search algorithm are provided below.

6.2.3 Adaptive genetic evolutionary search for multi-objective optimisation

A search strategy is also required to explore the parameter space in seeking Pareto optimal sets of parameters for the agents, i.e. liquidity provider and liquidity demander parameter vectors in the stochastic LOB model. In this regard, one may consider a multi-objective evolutionary algorithm (MOEA) framework. Such approaches have been the focus of extensive study over the past 15 years (see, e.g. Zhou et al. [2011], Eiben and Smith [2003], and references within) and would be particularly applicable to the problem at hand. There are several reasons for their popularity: they are inherently parallel, they feature operators to combine and mutate candidate solutions to rapidly arrive at improved solutions and are able to capture multiple Pareto-optimal solutions during the optimisation [Zitzler et al., 2000], which can be spread out across the Pareto front. In addition, there has been recent advances to better understand the relationship between such optimisation search frameworks and stochastic genetic search methods, see for instance discussions in Emmerich et al. [2013]. In this paper, we explore the utilisation of adaptive mutation kernels in the simulation based Multi-objective-II framework to efficiently explore the parameter space, where our approach merges traditional genetic search algorithms with adaptive Markov kernels utilised in adaptive MCMC methods, such as those studied in Haario et al. [2006], Roberts and Rosenthal [2009] and Andrieu et al. [2006].

The MOEA used in this paper is based on the NSGA-II (Non-dominated Sorting Genetic Algorithm II), developed by Deb et al. [2002]. This is an *elitist* MOEA, and in every iteration, combines the best parent solutions with the best offspring to produce a new family of candidate solutions. It produces a *diverse* Pareto-optimal front (i.e. the solutions are well-spread out across the front, due to the algorithm's use of a crowding distance operator) with low computational requirements ($O(mN^2)$ computational complexity, where m is the number of objectives, and N is the population size).

The algorithm is perhaps the most popular MOEA and is frequently used as a performance benchmark for other algorithms [Coello et al., 2007]. It has been used in various applications, including the generation expansion planning problem in power systems [Kannan et al., 2009] and for balancing objectives in groundwater monitoring designs [Reed and Minsker, 2004].

In addition, it has been been further developed in a Bayesian setting, in order to solve discrete multi-objective decomposable problems (see, e.g. Khan [2003], Laumanns and Ocenasek [2002], Khan et al. [2002]). Within this algorithm, we extend the features by also incorporating an adaptive global and local mutation kernel for a subset of the stochastic agent-based LOB model parameters $\boldsymbol{\theta}$. We first present an overview of the optimisation algorithm structure:

1. First, a family, or population, of N candidate solutions is initialised randomly from the feasible region.

2. For each solution, the objective functions are calculated and a rank is obtained reflecting Pareto dominance. That is, solutions are sorted into fronts, with the first front consisting of solutions that are not dominated by any other solutions, the second consisting of solutions that are only dominated by a single solution, and so on. Solutions are also assigned a crowding distance value, indicating the Euclidean distance from other solutions on the same front.

3. From this family of solutions, the crowding comparison operator is applied, and chooses the best solutions according to their rank, and in the case of ties, according to the crowding distance value.

4. Then, one or more evolutionary operators (detailed in the following section) are applied to evolve the selected set of solutions.

5. The new solutions are combined with the current family of solutions and the process is repeated from the second step, for a set number of iterations.

The algorithm outputs the non-dominated set of solutions with the highest ranking. We provide details about the operators used in multi-objective Indirect Inference procedure in the following section.

6.2.4 Algorithm settings and evolutionary operators

Details of a large number of evolutionary operators used in MOEAs can be found in Coello et al. [2007]. In NSGA-II, one has to first select the size of the population of candidate solutions for every iteration of the algorithm, in addition to the number of iterations (called generations in the MOEA nomenclature). In our optimisation, we use a population size of $N = 40$ parameter sets, and run the optimisation for a total of 40 generations.

We referred to a number of operators used to evolve and choose amongst the set of solutions, and we provide further information here about their function:

- **Selection operator:** From the second iteration of the algorithm onwards, there will be $2N$ sets of candidate solutions in step 3. The best N solutions are chosen based on a) dominance and b) crowding distance, or the distance of the solution from its neighbours. If the number of solutions on the first front is less than N, they are all selected, and the remainder are taken from further fronts. In the case where one must select fewer solutions than the number of solutions on a particular front, the solutions with the highest crowding distance value are chosen.

- **Crossover operator:** The Simulated Binary Crossover (SBX) operator is used. From two candidate solutions θ_1, θ_2, two new solutions θ'_1, θ'_2 are formed, where the k-th elements are as follows:

$$\theta'_{1,k} = \frac{1}{2}[(1-\bar{\beta})\theta_{1,k} + (1+\bar{\beta})\theta_{2,k}] \tag{6.20}$$

$$\theta'_{2,k} = \frac{1}{2}[(1+\bar{\beta})\theta_{1,k} + (1-\bar{\beta})\theta_{2,k}]. \tag{6.21}$$

Here, $\bar{\beta}$ is a random sample from a distribution with density

$$\bar{\beta} = \begin{cases} (\alpha u)^{\frac{1}{\eta_c+1}} & \text{if } u \leq \frac{1}{\alpha} \\ (\frac{1}{2-\alpha u})^{\frac{1}{\eta_c+1}} & \text{otherwise} \end{cases}$$

where $u \sim U(0,1)$ and $\alpha = 2 - \beta^{-(\eta_c+1)}$, with

$$\beta = 1 + \frac{2}{\theta_{2,k} - \theta_{1,k}} \min\left[(\theta_{1,k} - \theta_{k_L}), (\theta_{k_U} - \theta_{2,k})\right]. \tag{6.22}$$

The new solutions are guaranteed then to remain within the solution bounds $[\theta_{k_L}, \theta_{k_U}]$. We use the crossover operator with probability $p_c = 0.7$ and a distribution index $\eta_c = 5$.

- **Mutation operator:** The polynomial mutation operator is used. The mutation operator perturbs elements of the solution, according to the distance from the boundaries.

$$\theta'_k = \theta_k + \bar{\delta}(\theta_{k_U} - \theta_{k_L})$$

where we have for $\bar{\delta}$

$$\bar{\delta} = \begin{cases} \left[2\gamma + (1-2\gamma)(1-\delta)^{\eta_m+1}\right]^{\frac{1}{\eta_m+1}} - 1 & \text{if } \gamma < 0.5 \\ 1 - \left[2(1-\gamma) + 2(\gamma-0.5)(1-\delta)^{\eta_m+1}\right]^{\frac{1}{\eta_m+1}} & \text{if } \gamma \geq 0.5 \end{cases}$$

with

$$\delta = \min\left[(\theta_k - \theta_{k_L}), (\theta_{k_U} - \theta_k)\right].$$

Here, $\gamma \sim U(0,1)$ and the distribution index $\eta_m = 10$. The polynomial mutation operator is used with probability 0.2.

Covariance matrix mutation and sampling: The NSGA-II algorithm discussed above is only able to produce binary, integer, or real encodings for the output solution vectors. However, the stochastic process for the limit order submission activity by liquidity providers requires the specification of a positive definite and symmetric covariance matrix for the generation of intensities from a multivariate skew-t distribution. We cannot naively extend the evolutionary operators above (crossover and mutation) to produce new sets of covariance matrix candidate solutions as this would not guarantee the positive definiteness and symmetry constraints of the covariance matrix are preserved. We thus propose an extension to the MOEA, effectively another operator that will generate candidate solutions for the covariance matrices, such that every new generation remains in the manifold of positive definite matrices. This operator will generate new candidate covariance matrices once the evolutionary operators discussed previously have been applied.

To ensure that the optimisation algorithm searches the space of feasible solutions efficiently and does not get stuck in a suboptimal region of the space of possible solutions, our covariance matrix sampling operator has two components to undertake exloration and exploitation type moves. The mutation kernel is comprised of a mixture of Inverse Wishart distributions with different parameters, as per the proposal of Peters et al. [2012], one mixture component to provide global search (exploration) and a second mixture component to provide local searches (exploitation). The density of the Inverse Wishart distribution is

$$f(X;\Psi,p) = \frac{|\Psi|^{\frac{v}{2}}}{2^{\frac{pd}{2}}\Gamma_d(\frac{p}{2})}\,|X|^{-\frac{p+d+1}{2}} e^{-\frac{1}{2}tr\left(\Psi X^{-1}\right)}$$

where $X, \Psi \in \mathbb{R}^{dxd}$ and positive definite, and Γ_d is the d-variate gamma function. For the efficiency of the covariance mutation, we carry out an adaptive learning strategy for the specification of the local mixture component. In this case, the algorithm will explore the local region with high probability, but make potentially larger moves with smaller probability.

We now describe one complete covariance mutation step. In the n-th generation of the MOEA, we generate $\{\Sigma_{n,i}\}, i = 1 \ldots N$ from a mixture distribution $q(\Sigma_{n,i})$ defined as follows:

$$q(\Sigma_{n,i}) = (1 - w_1)\mathcal{IW}(\Psi_n, p_1) + w_1\mathcal{IW}(\Psi, p_2)$$

where p_1, p_2 are degrees of freedom parameters with $p_2 < p_1$, and where w_1 is small so that sampling from the second distribution happens infrequently. Here Ψ denotes an uninformative positive definite matrix, with the effect that sampling from the second distribution leads to moves away from the local region being explored. Ψ_n is also a positive definite matrix, fitted based on moment matching to the sample mean of the successfully proposed candidate solutions

in the previous stage of the Multi-Objective optimisation as follows:

$$\Psi_n = \frac{1}{\sum_{t=1}^{n} w^t} \sum_{t=1}^{n} w^t \frac{1}{\sum_{i=1}^{N} \frac{1}{r_{t,i}}} \sum_{i=1}^{N} \frac{1}{r_{t,i}} \tilde{\Sigma}_{t,i}$$

where $r_{t,i}$ is the non-domination rank of the i-th solution in the t-th generation, and w^t with $w < 1$ is an exponential weighting factor.

6.3 Stochastic agent LOB model assessment and calibration to real LOB data

We have provided a description of the stochastic agent-based LOB model we have developed for modelling trading interactions and their dependency. In addition, we have developed a method for the calibration of model parameters to observed LOB data. In this section, we illustrate the results of this calibration on real data, through a sequence of studies which aim to practically assess the importance of each component of the stochastic agent LOB model specification. To achieve this, we make a number of model simplifications and progressively relax these simplifying assumptions, in order to provide an understanding of the role each feature of our proposed model plays in the simulation framework. The **reference model** is the basic framework against which we compare the more detailed versions of the model, as detailed below.

6.3.1 Reference LOB model

In the stochastic agent-based LOB model, the liquidity provider agent has limit order submission and cancellation components which each require the specification of four independent l_t-dimensional multivariate skew-t distributions for the bid and ask sides, with $l_p = 5$ 'passive' levels and $l_d = 3$ 'direct', or aggressive levels for a total of $l_t = 8$ actively modelled levels for each side of the book. For each of these stochastic model components we require the estimation of the parameters: $\boldsymbol{m} \in \mathbb{R}^d$, the location for the mean intensity vector; $\boldsymbol{\gamma} \in \mathbb{R}^d$, the skewness of the stochastic intensity vector; $\nu \in \mathbb{R}^+$ which directly influences the heavy-tailedness of the stochastic intensity vector and $\Sigma \in \mathbb{R}^{d \times d}$ the covariance matrix of the stochastic intensity vector for order arrivals. We consider aggregate activity in 10 second intervals, and for the 8.5 hour trading days for the asset under consideration here, we have $T = 3060$ intervals in the day. The basic reference model is characterised by the following model assumptions:

- We assume that the associated limit order submission distributions for the bid and ask have common parameter value settings. In addition, market order submission distributions for the bid and ask are also assumed to have common parameter value settings. This

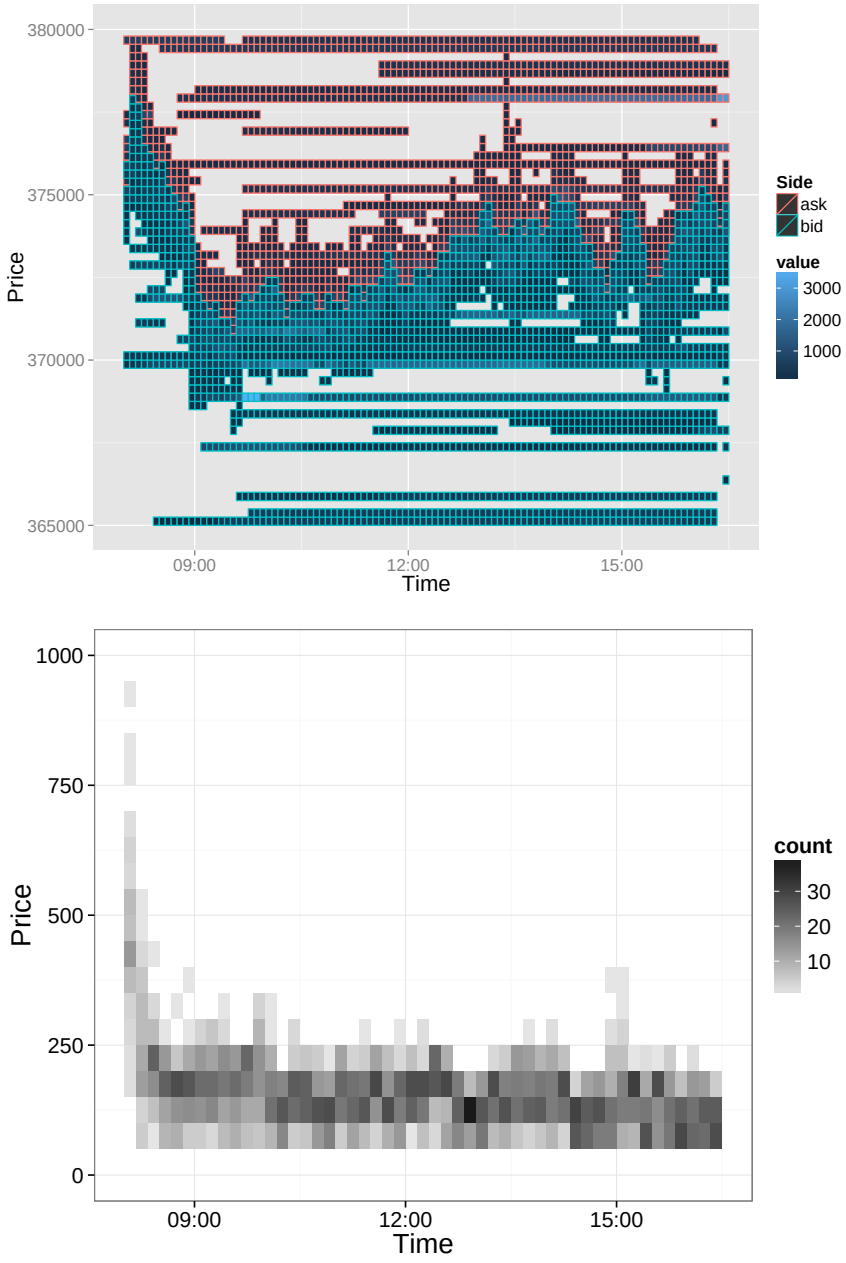


Figure 6.5: For stock BNP Paribas, the intensity of the volume process on either side, where the shading of each bin indicates the average number of shares available at those prices in that period. The plot on the bottom shows the evolution of the spread throughout the trading day.

is reasonably consistent with empirical observations for a large number of assets when observing the submission activity on either side of the LOB throughout the trading day.

- Since the vast majority of orders get cancelled prior to execution, we consider the parameters of the distribution of cancellations to also match the distribution of limit order placements.

- We also set $\boldsymbol{m} = \mathbf{0}$ and consider the skewness vector, $\boldsymbol{\gamma}$, to take a common value in all levels of the bid and ask such that $\boldsymbol{\gamma} = \gamma_0 \mathbf{1}$, where $\mathbf{1}$ is a vector of ones.

- The monotonic mapping $F(\cdot)$, transforming the random variables $\Gamma^{LO,k,s}, \Gamma^{C,k,s}, \Gamma^{MO,k}$ into intensity random variables $\Lambda^{LO,k,s}, \Lambda^{C,k,s}, \Lambda^{MO,k}$ is set as the CDF of the standard Normal. This transformation is necessary in order to ensure that intensities are positive, and to bound the event counts.

- For the baseline intensities of limit order activity at each level, we assume that they will be the same for the 'passive' limit orders on both sides, i.e. $\mu_0^{LO,a,1} = \ldots = \mu_0^{LO,a,l_p} = \mu_0^{LO,b,1} = \ldots = \mu_0^{LO,b,l_p} = \mu_0^{LO,p}$, while 'aggressive' limit orders will have a different limit order intensity, i.e. $\mu_0^{LO,a,0} = \ldots = \mu_0^{LO,a,-l_d+1} = \mu_0^{LO,b,0} = \ldots = \lambda_0^{b,-l_d+1} = \mu_0^{LO,d}$. Market order baseline intensities are also equal on either side, i.e. $\mu_0^{MO,a} = \mu_0^{MO,b} = \mu_0^{MO}$. The cancellation baseline activity will be the same as the submission baseline activity.

- Finally, we assume constant order sizes, i.e. $O_{i,t}^{LO,k,s} = c = O_{j,t}^{MO,k}$ for all $i \in \left\{1, \ldots, N_t^{LO,k,s}\right\}$, $j \in \left\{1, \ldots, N_t^{MO,k}\right\}$, $k \in \{a, b\}$, $s \in \{-l_d + 1, \ldots, l_p\}$ and $t \in \{1, \ldots, T\}$.

Hence, the basic reference model has the following parameter vector $\left\{\mu_0^{LO,p}, \mu_0^{LO,d}, \mu_0^{MO}, \gamma_0, \nu, \sigma^{MO}\right\}$, as well as the covariance matrix Σ to be estimated.

The cancellations are modelled by a dynamically evolving volume process, i.e. the Cox process is truncated to the available number of orders at each level, as specified in the model by $N_t^{C,k,s} | \left\{ \tilde{V}_t^{LO,k,s} = v \right\} \sim Po(\lambda_t^{C,k,s}) \mathbb{I}(N_t^{C,k,s} < v)$ where we denote by $V_{t-1}^{LO,k,s}$ the volume at level L_i at the start of the $[t-1, t)$ interval and $\tilde{V}_t^{LO,k,s}$ is the volume available after the arrival of the limit orders at time t, but before the cancellations and executions. One can simulate from the model, in order to obtain the state of the LOB at time t, L_t^*, and thus the available volume v, so that one can then draw from a truncated Poisson distribution with a truncation limit of v.

Before we begin the study of the stochastic agent-based LOB model and its calibration and simulation behaviour, we first show for a representative trading day, the evolution of the spread,

as well as the intensity of the volume process around the top of the book, for one of the most liquid stocks in the CAC40, namely BNP Paribas, in Figure 6.5. This provides an illustration of the LOB dynamics we should aim to recover with the model once accurately calibrated. We estimate the model on the data from this day, as an illustration of the calibration procedure.

6.3.2 Reference model: Calibration

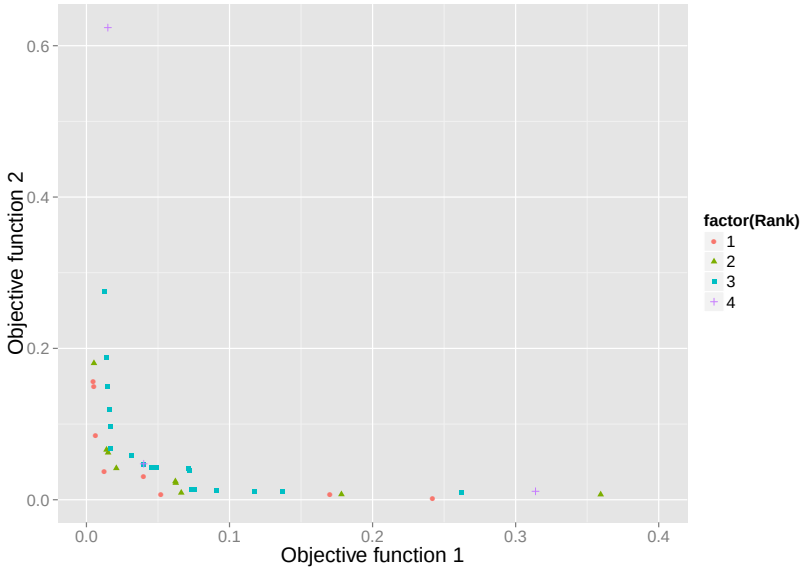


Figure 6.6: Objective function values for the parameter vectors produced by the multi-objective II method. These are grouped by non-domination rank, with a rank of 1 indicating non-dominated vectors, a rank of 2 indicating vectors dominated only by a single other vector and so on. Note that the points in each group form a Pareto front, a feature of the optimisation.

We present in Table 6.1 the results of the estimation using the multi-objective II approach proposed in this paper. There are 8 non-dominated solutions spread out accross the Pareto optimal front, each of which also has an associated covariance matrix, which has not been included here due to space considerations, instead we provide the trace as a summary. In the table, we also present a further 4 solutions with a non-domination rank of 2, i.e. parameter vectors which were dominated in both objective functions by only one other parameter vector. We present the non-domination rank, as well as the objective function values of the entire final parameter population in Figure 6.6. We note that in terms of the 2 objective function values associated with these parameter vectors, these are spread out across the Pareto front.

We assess the fit by a qualitative comparison of the simulations produced with the estimated parameters. In Figure 6.7 we present, for the first 2 Pareto optimal solutions of the

parameter vectors in Table 6.1, summaries of the price process for repeated simulations, as well as an example of the LOB evolution throughout the day. We see that the two Pareto optimal solution parameter vectors produce a broad variety of different price trajectories over repeated simulations. In particular, some points on the Pareto front of solutions for this basic reference model produce a time series of simulated prices which replicates a trading day with relatively volatile trade activity, whilst other points on the Pareto front favour more constrained trading simulated price activities. To understand how this may occur, we note that this is likely to be due to the relatively high baseline rate of market orders compared to baseline limit order rates in the first set of Pareto optimal solutions, compared to the second.

	$\mu_0^{LO,p}$	$\mu_0^{LO,d}$	μ_0^{MO}	γ_0	ν_0	σ_0^{MO}	$\mathrm{Tr}(\Sigma)$
1	30.84	8.16	4.75	-0.18	33.70	1.78	7.11
2	31.16	5.13	4.41	9.96	28.07	9.95	4.60
3	31.16	5.13	4.41	9.96	21.74	9.95	5.52
4	29.82	5.24	4.45	-0.52	20.57	4.70	4.26
5	46.87	7.42	4.77	0.64	28.34	8.81	6.82
6	22.05	8.18	8.13	-1.68	24.65	1.83	5.70
7	19.83	5.41	0.68	-0.28	28.85	2.15	5.25
8	12.95	3.12	2.93	2.35	35.25	3.58	6.01
9	30.84	8.16	4.75	-0.18	33.70	1.78	5.14
10	31.16	5.13	4.41	9.96	28.07	9.95	5.20
11	31.16	5.13	4.41	9.96	21.74	9.95	7.30
12	29.82	5.24	4.45	-0.52	20.57	4.70	4.32

Table 6.1: Non-dominated solutions after 40 iterations, with a population size of 40.

In Section 6.3.4, we provide further calibration results for the reference model, for multiple assets, over an extended period of 15 trading days. Summarising these results, we show that within the set of solutions produced by our estimation procedure, there is very commonly a subset which produce simulations which are similar to real trading observations in terms of their price and volume behaviour.

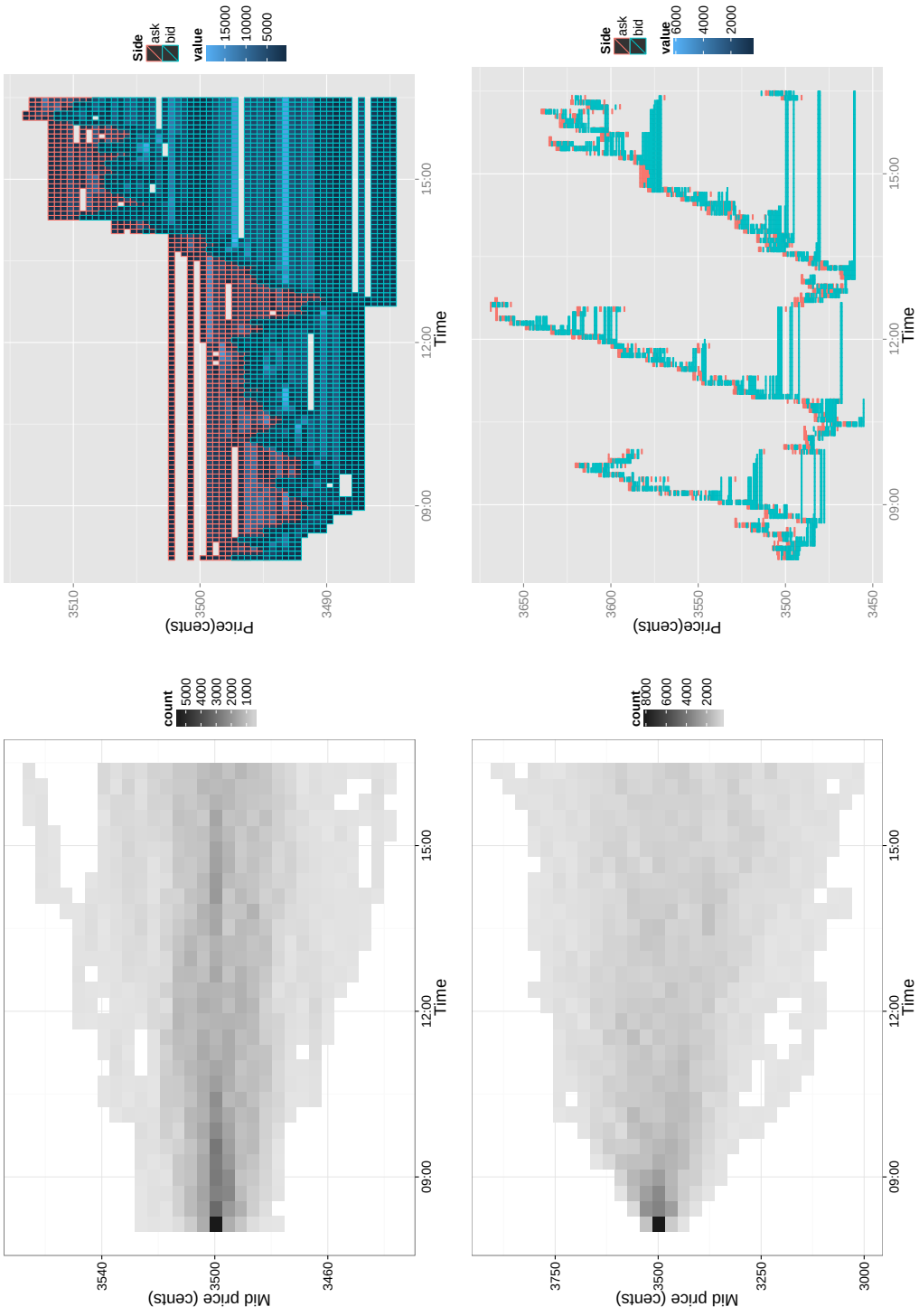


Figure 6.7: Simulations using 2 of the non-dominated parameter vectors resulting from estimating the basic model with NSGA-II. The figures on the left are heatmaps of the asset mid price over 100 simulations, while the figures on the right represent the state of the LOB over a single simulation.

6.3.3 Relaxing assumptions of the reference stochastic agent LOB model

The baseline model results are encouraging, however we still need to determine what influence the simplifying statistical model assumptions made in the reference model specification have on the calibration performance. This will now be assessed by progressively relaxing the assumptions. Our criterion for improvement relative to the reference model will be a reduction in the values of the objective functions of the solutions on the Pareto optimal front. We will only suggest that particular features should be relaxed if we observe such an improvement.

6.3.3.1 Incorporating an order size distribution

In our basic reference model, we assumed that orders sizes are constant, i.e. all limit order submissions, cancellations and executions were from an equal number of shares. This is similar to the model of Cont et al. [2010], which assumed that all orders are of unit size, which they set to correspond to the average size of limit orders observed for the asset. Abstracting away the order size aspect is an approximation one can make in order to simplify the model. However, such a simplifying assumption is not likely to be supported by the data, as we illustrate in Figure 6.8. Clearly, one observes that there is a range of distribution shapes for the order sizes of different assets.

It is clear that the distribution of order sizes will be affected by features such as minimum order sizes on an exchange (in number of shares, lots, or weight, depending on what is being traded). We observe empirically that for a range of equities traded in a number of countries, the distribution of order sizes has clear peaks at round figures - see Figure 6.8 for evidence of clustering order volumes at multiples of 100 shares, for example. This seems to be independent of the level at which they are submitted, whether it is a buy or a sell order, as well as the intensity of the order submissions in that period.

Therefore, we present a case study where we relax the assumption of a fixed order size, by considering instead a stochastic model where we assume that the order size is drawn from a mixture of distributions. In this case, we assume that both the limit and market order sizes are obtained by sampling from the following Gamma mixture

$$O_{i,t}^{LO,k,s} \sim w\, Gamma(\kappa_1, \theta_1) + (1-w)\, Gamma(\kappa_2, \theta_2), \forall i, t, k, s \tag{6.23}$$

where

$$Gamma(O; \kappa, \theta) = \frac{1}{\Gamma(\kappa, \theta^{\kappa})} O^{\kappa-1} \exp\left[-\frac{O}{\theta}\right];\ \ O \in \mathbb{R}^{+}, \tag{6.24}$$

with positive shape parameters κ_1, κ_2 and positive scale parameters θ_1, θ_2. We set $\kappa_1 = 1, \kappa_2 =$

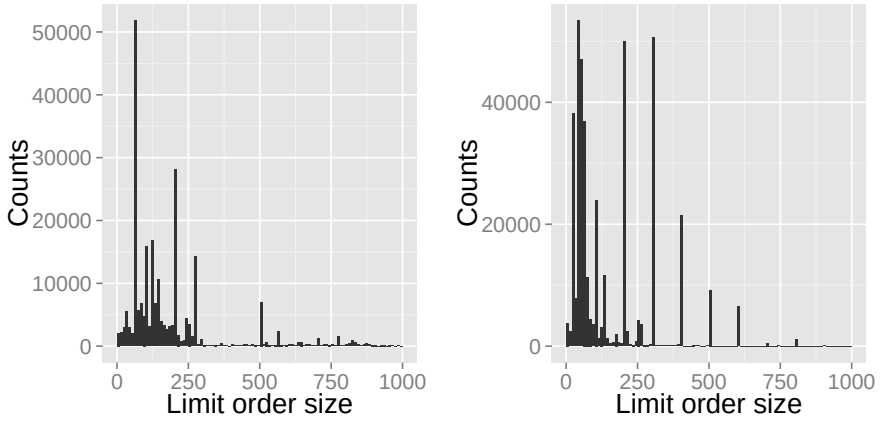


Figure 6.8: Histograms of order sizes for 2 CAC40 stocks - ACAp(left) and BNPp (right)

2 as we observed there was a mode present in the empirical distributions of order sizes, and we estimated the scale parameters for each mixture component to place the mode in the appropriate locations. Hence, we additionally estimate the parameters θ_1, θ_2 and the mixture weight w.

We run the stochastic optimization framework using the same settings (a parameter population of 40 candidate solutions and an evolution over 40 generations) and calibrate the relaxed reference model with the stochastic model for the order sizes to the same data set used in the reference model fit, i.e. the LOB data for BNP Paribas over an entire day. We obtain a Pareto optimal front which again contained multiple parameter vector solutions which were spread out over the Pareto front, indicating a successful exploratory search by the genetic search framework. Importantly, as shown in Figure 6.9 we observe the realized objective function values for the relaxed reference model, which we observe are clear improvements on the objectives achieved by the comparison basic reference model case in which the order sizes were fixed.

Figure 6.14 shows the intensity of the volume process and the evolution of the spread for a simulated trading day for 2 of these parameter vectors selected from the Pareto optimal front. Similarly to the reference model, the price and volume trajectories are still quite flexible between the different feasible, Pareto optimal solutions obtained for this calibration.

6.3.3.2 Introducing asymmetry and skewness to Limit Order intensity by depth

In the reference model, we assumed that the skewness parameter vector $\boldsymbol{\gamma}$ for the multivariate skew-t distribution assumed for the number of limit orders and cancellations at each level of the LOB were fixed to a common skew. This parsimonious choice was encoded in the model by the reference model assumption $\boldsymbol{\gamma}^{LO,a} = \boldsymbol{\gamma}^{LO,b} = \boldsymbol{\gamma} = \gamma_0 \mathbf{1}$ and $\gamma^{MO} = \gamma_0$, i.e. there was only

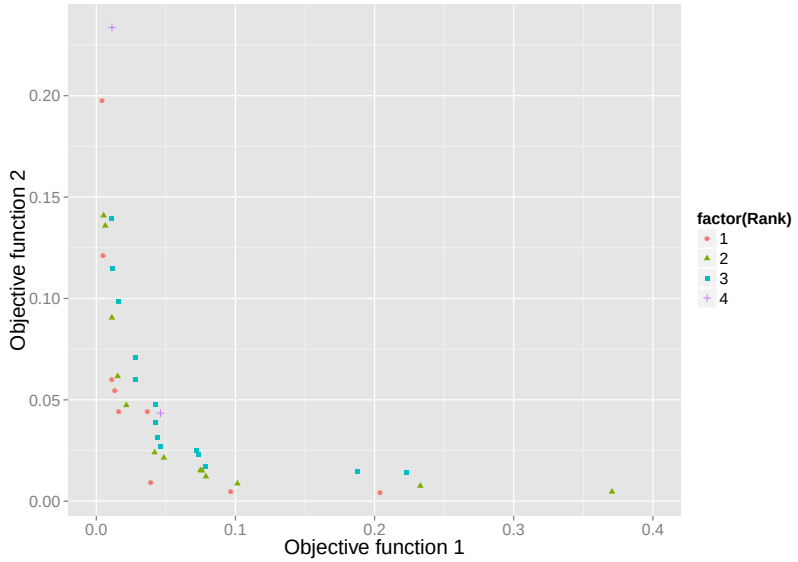


Figure 6.9: Objective function values for the parameter vectors produced by the multi-objective II method, in the case where we assume that order sizes are drawn from a mixture of Gamma distributions.

one skewness parameter which was common to all levels on both the bid and ask. The effect of this assumption on the price and volume dynamics in the reference model is now assessed by relaxing this feature and performing calibration of a relaxed version of the reference model to the same day of data from BNP Paribas.

We now allow $\boldsymbol{\gamma}^{LO,a} = \boldsymbol{\gamma}^{LO,b} = \boldsymbol{\gamma} = \{ \gamma^{LO,-l_d+1}, \ldots, \gamma^{LO,l_p} \} = \boldsymbol{\gamma}^{C,a} = \boldsymbol{\gamma}^{C,b}$, in order to gain additional flexibility in modelling the skewness in the multivariate counts for limit order and cancellation data. We also allow $\gamma^{MO,a} = \gamma^{MO,b} = \gamma_0^{MO}$ to enable the skewness of the market order data to be modelled separately. This will entail estimating an additional $l_d + l_p$ parameters. Again, we assess whether the Pareto optimal solutions improve in minimizing the objective functions under this relaxation of the constraints in the reference model assumptions.

Table 6.2 shows that in none of the parameter vectors produced by the multi-objective II estimation method are the elements of the skewness vector close to being equal to one another, which indicates that the use of the skew vectors with different skew at each level of the LOB for the bid and ask, in the Multivariate Skew-t distribution, is appropriate for the calibration to real data. As expected, incorporating these features improves the model power and suitability, measured by the objective function values achieved by the solutions in the Pareto optimal front, for the simulated stochastic agent LOB model realizations, when compared to the reference model.

	$\mu_0^{LO,p}$	$\mu_0^{LO,d}$	μ_0^{MO}	γ_0^{MO}	ν_0	σ_0^{MO}	$\gamma_0^{LO,-2}$	$\gamma_0^{LO,-1}$	$\gamma_0^{LO,0}$	$\gamma_0^{LO,1}$	$\gamma_0^{LO,2}$	$\gamma_0^{LO,3}$	$\gamma_0^{LO,4}$	$\gamma_0^{LO,5}$	$\mathrm{Tr}(\Sigma)$
1	39.35	4.00	0.54	-7.36	46.24	7.89	4.32	-7.30	1.89	-4.49	-7.86	4.51	4.78	-6.72	7.97
2	38.48	3.98	5.81	-1.35	8.63	8.21	7.41	4.35	7.47	-6.86	-2.29	1.16	4.74	-5.77	5.82
3	39.54	3.39	0.54	-6.46	46.24	7.89	4.32	-7.30	3.13	-4.49	-7.86	2.67	4.78	-6.72	6.41
4	11.33	2.56	2.53	6.90	3.14	1.98	-3.32	-3.55	-8.42	-3.32	-5.53	-4.10	-4.58	4.40	5.53
5	37.61	3.98	1.16	-1.35	2.59	8.21	4.40	-7.50	-6.64	-7.61	-7.56	1.20	4.78	-5.11	7.41
6	18.25	4.00	1.05	-2.34	2.52	8.21	-0.05	-8.93	-3.35	-7.37	-7.43	3.67	2.81	-2.61	5.67
7	13.40	5.97	6.14	-1.25	22.02	8.69	6.11	-1.65	-6.36	-8.16	-2.75	3.34	8.76	6.81	6.42
8	39.35	4.00	0.71	-6.44	5.67	2.25	-3.25	-7.34	1.89	-4.47	-7.86	4.51	4.78	-7.18	4.42

Table 6.2: Non-dominated solutions for the model where the elements of the skewness vector are allowed to vary.

6.3.4 Further results

In the previous part of this section we presented results for both the calibration of the reference model and models where we relaxed certain assumptions. This calibration was performed using the data from a single asset (BNP Paribas) over one day, in order to be able to present detailed results regarding objective function values, LOB evolution over individual simulations using individual solutions on the Pareto front, as well as summaries of repeated simulations. In this section, we repeat the calibration of the reference model for 5 assets (BNP Paribas, Credit Agricole, Total SA, Technip SA and Sanofi) every trading day between 01/02/2012 and 21/02/2012. The stocks were chosen from the French CAC40 stocks, and are therefore amongst the most liquid stocks in the country. Specifically, we chose assets that are representative of different industries (banking, energy and pharmaceutical) and have different ticksizes (minimum price increments) and market capitalisations, as these are some of the main factors that affect daily trading activity.

We summarise the results as follows: We first calibrate the reference model for each day and each asset individually, from which we obtain a set of J solutions (i.e. non-dominated solutions on the Pareto front) every time. For each solution (parameter vector $\hat{\boldsymbol{\theta}}_j, j \in 1 \dots J$), we simulate the LOB model N=50 times and fit the auxiliary models to the simulated data to obtain N auxiliary model parameter vectors $\beta_1^{i,j,*}$ and $\beta_2^{i,j,*}, i \in 1 \dots N$. The former are the ARIMA model parameters fit to the volume process on the bid and ask side, and the latter are the GARCH model parameters fit to the log returns.

We can then construct the empirical distribution for each parameter in these vectors, and determine the 95% confidence interval. From this, we can determine whether the parameter coefficients of the auxiliary model fit to the real data lie within this range, for each asset on the Pareto front. In Figures 6.11 and 6.12, we show for each day, each asset and each auxiliary

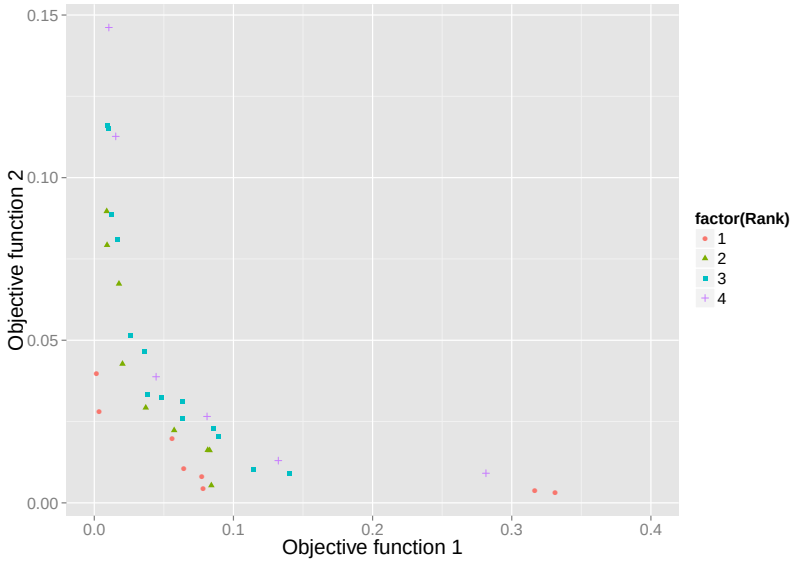


Figure 6.10: Objective function values for the parameter vectors produced by the multi-objective II method, in the case where we relax the assumption that the elements of the skewness vector in the Multivariate Skew-t distribution are equal.

model parameter, the proportion of solutions on the Pareto front for which the coefficients of the auxiliary model fit to the real data lie within the 95% confidence interval of the coefficients of the auxiliary model fit to the simulated data.

We note that the proportion varies over time, as one would expect, as not all solutions on the Pareto front will give rise to LOB dynamics that closely reflect those observed in real data. However, we note that this proportion is generally more than 25% for most parameters and most days. Thus, within the set of solutions produced by our estimation procedure, there is a subset which produce simulations which are similar to real trading observations in terms of their price and volume behaviour, which are the summaries of the LOB which our auxiliary models related to.

6.4 Regulatory interventions via the stochastic agent-based LOB model

In contrast to Westerhoff [2003], in our model the agents' strategy is not dependent on profitability. This is because of the division of our trading agents according to their liquidity considerations: Traders often consume liquidity due to considerations other than profit, such as rebalancing the weights of their holdings in a fund. They cannot simply choose to become liquidity providers because of the superior profitability of these agents, for a number of reasons.

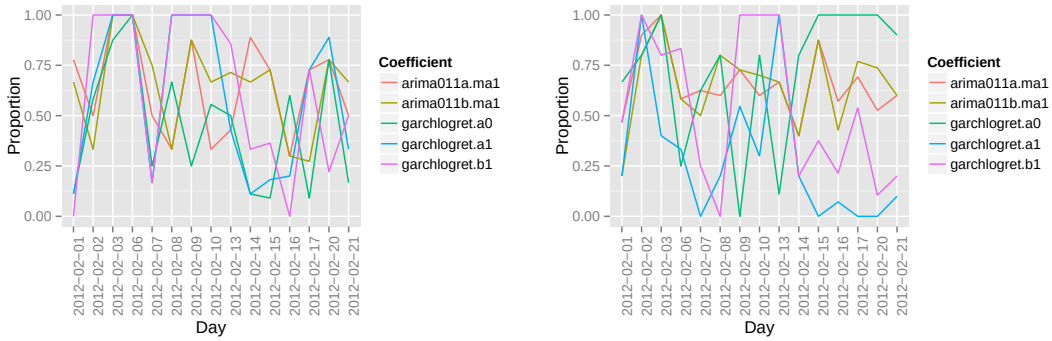


Figure 6.11: The proportion of solutions on the Pareto front for which the coefficients of the auxiliary model fit to the real data lie within the 95% confidence interval of the coefficients of the auxiliary model fit to the simulated data, for each trading day between 01/02/2012 and 21/02/2012 for 2 stocks. (Left): BNP Paribas. (Right): Credit Agricole.

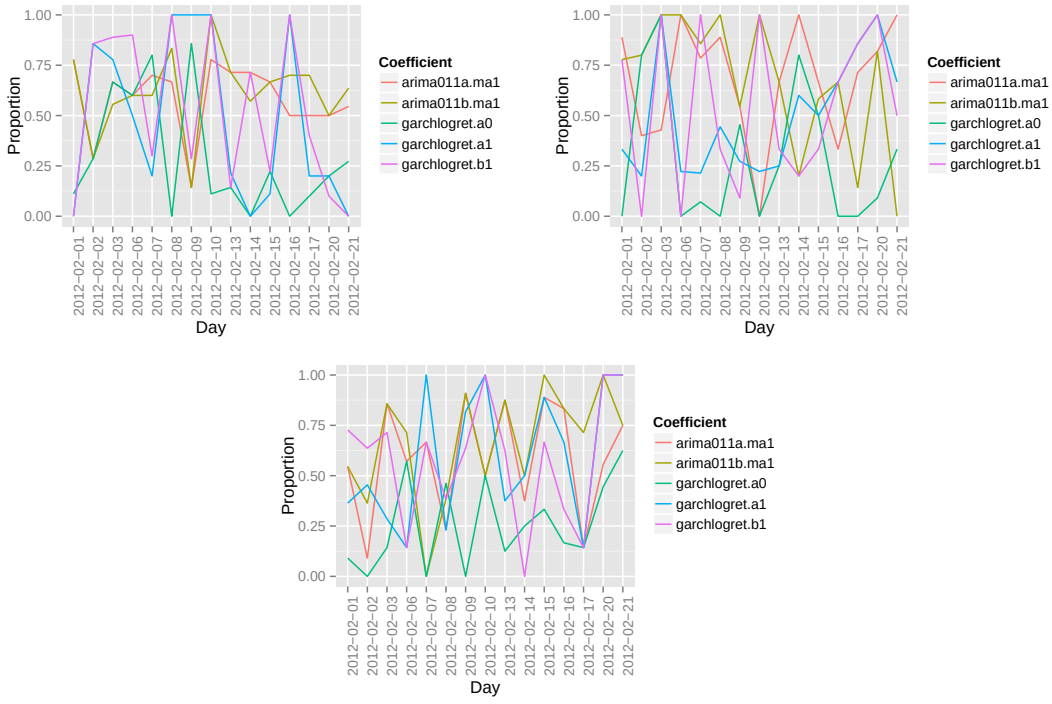


Figure 6.12: The proportion of solutions on the Pareto front for which the coefficients of the auxiliary model fit to the real data lie within the 95% confidence interval of the coefficients of the auxiliary model fit to the simulated data, for each trading day between 01/02/2012 and 21/02/2012 for 3 stocks. (Left) Total SA. (Right) Technip SA. (Bottom): Sanofi.

These include the investment in technology required to be able to carry out such a strategy in the millisecond environment, the inventory they will be required to hold, and, possibly, regulatory or exchange obligations they will have to adhere to.

Our model simulates the activity of a stock on a single LOB, on a single day. The introduction of MiFID has increased competition and allowed for the trading of stocks in pan-European multilateral trading facilities (MTFs). The trading on one venue will undoubtedly affect the trading interest in another, through the activity of cross-market arbitragers. In addition, there is the possibility that regulation can be imposed on one market, but not another, which will have implications for the efficacy of the regulation itself. Both Mannaro et al. [2008] and Westerhoff and Dieci [2006] have considered this in the context of an ABM, but with simpler models than the one considered here, which do not take into account the liquidity considerations of the agents.

6.4.1 Quote-to-trade ratio

The intervention we will consider here, as an example of the type of experiment that can be performed using our model, is the imposition of a quote-to-trade ratio. This ratio is already considered in certain exchanges, such as the LSE, which allows for 500 quotes per trade. Further quotes are allowed in the case of the LSE, but are subject to a 5 pence surcharge for every order[1]. In our model, we have made the assumption that the baseline limit order submission (or quote) intensity at every level $\lambda_0^{LO,a,i}$ is equal to the baseline cancellation intensity $\lambda_0^{C,a,i}$. That is, potentially all orders submitted in an interval can be cancelled prior to execution.

Given the setup of our model, it is more convenient to enforce a stochastic limitation for excessive trading, rather than a hard limit of (say) 100 limit orders to 1 market order. For a quote-to-trade ratio $q = \frac{100}{1}$, we impose the limit by specifying that for the cancellation activity $\lambda_t^{C,a,i} = (1 - \frac{1}{q})\lambda_t^{LO,a,i}$. This is an approach also taken by Aït-Sahalia and Saglam [2013], who, rather than enforcing a strict minimum resting time of 500 milliseconds, instead subject every order to a random minimum resting time that is exponentially distributed, but with the same mean.

We evaluate the outcome of such an intervention in our simulated LOB for 3 different quote-to-trade ratios, i.e. $q \in \left\{\frac{500}{1}, \frac{100}{1}, \frac{20}{1}\right\}$. Figure 6.13 shows the effect of the regulation on individual realisations of daily activity, as well as the price process in repeated realisations. We

[1] http://www.londonstockexchange.com/products-and-services/trading-services/pricespolicies/tradingservicespricelisteffective2december2013.pdf

have chosen one of the parameter vectors from the estimation of the basic model which generally showed excessive volatility. We note that, in our model, increasing q (and thus, reducing the relative number of cancellations) has the effect of constraining the mid-price process, and thus, curbing excess volatility.

While one cannot draw definite conclusions about the effect of such an intervention through an ABM simulation, it is a step a regulator may consider, particularly when comparing different approaches. For example, even in the implementation of a quote-to-trade ratio, the regulator may have a number of choices, for example, regarding the period over which they consider the ratio. We argue that our model can be informative for such considerations, and, given its flexibility, can give rise to a large number of computational experiments.

6.5 Discussion

We have presented a new form of agent-based model, in order to capture features of the complex stochastic process that is the Limit Order Book. The agent types we considered are representative of the classes of market participants in modern financial markets: In electronic LOBs, traders can be broadly separated according to their liquidity requirements, into liquidity providers and liquidity demanders. This is certainly more representative of the motivation for trading activity, compared to the chartist and fundamentalist models considered in the past (e.g. Frankel and Froot [1988], Kirman [1993], De Grauwe [1994], Farmer and Joshi [2002], Westerhoff and Reitz [2003], Manzan and Westerhoff [2007]).

We have not modelled the behaviour of individual agents, but rather the activity resulting from the entire class of agents. This has enabled us to directly model the dependence in event (limit order submission, cancellation and market order) activity between the different levels of the LOB, which would not have been possible by considering simpler formulations for individual agent strategies. We have employed a flexible Multivariate Skew-t model for the event intensities, which is unique for its ability to capture asymmetric and heterogeneous dependence, and its scalability in high dimensions. This has resulted in a very general formulation of the ABM, which also enables one to model the heterogeneity in order sizes.

Unsurprisingly, for such a model, it is difficult to write down the likelihood for estimation purposes, and a further contribution of this paper is a new simulation-based estimation approach. We termed this the Multi-Objective Indirect Inference estimation method, and in common with related approaches, our method only requires that the model is simulable and will produce parameter estimates, such that the simulated data approximates the real data in certain aspects. In our extension to the standard Indirect Inference approach, however, we considered

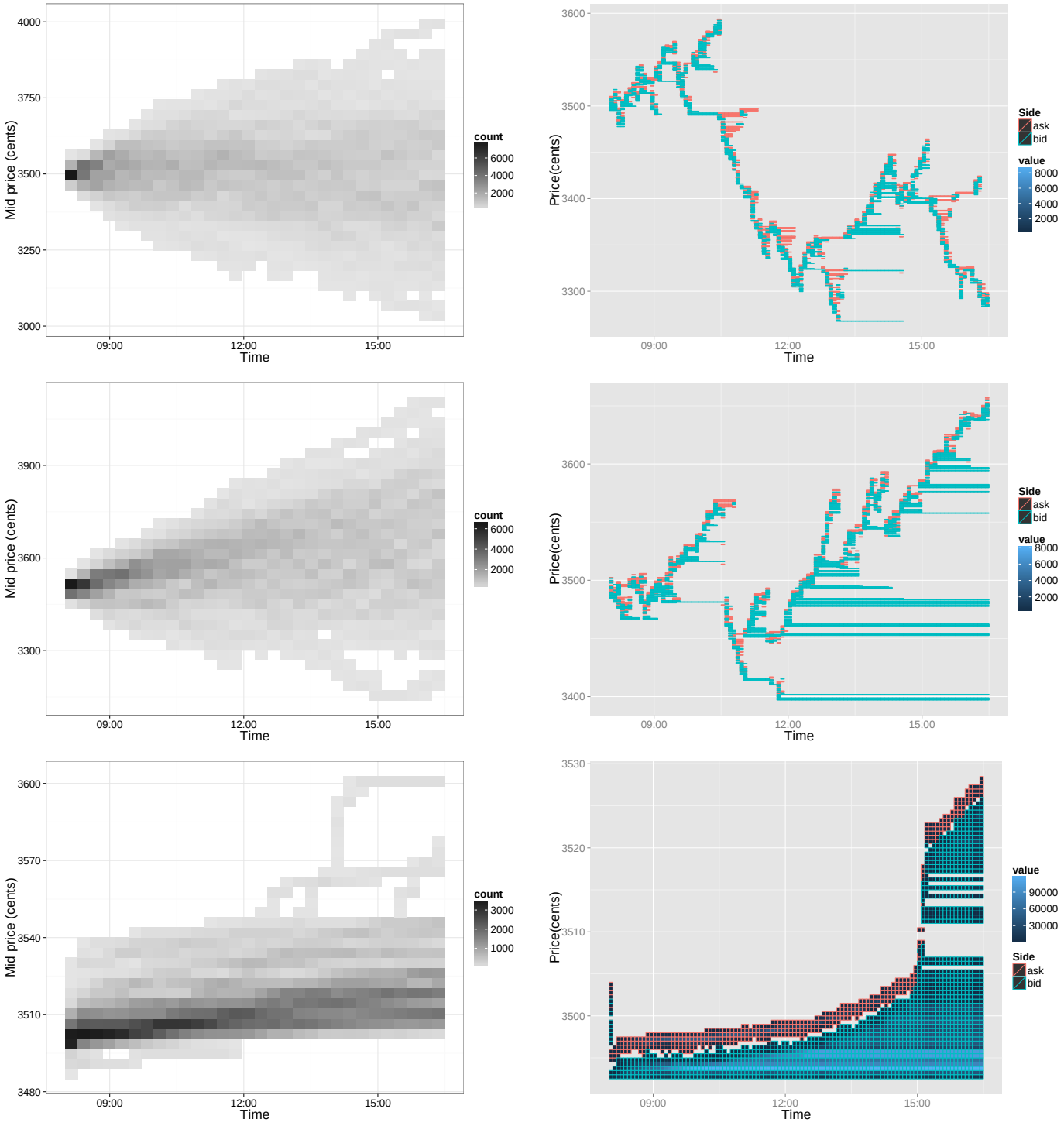


Figure 6.13: Simulations of the basic model, with the addition of a 'quote-to-trade ratio' regulatory intervention. The mid-price process and daily LOB volumes with a quote-to-trade ratio of $q = \frac{500}{1}$ (top), $\frac{100}{1}$ and $\frac{20}{1}$ (bottom).

auxiliary models relating to both the price and the volume process, and treated the estimation problem as a multi-objective problem.

Even with the simulation-based estimation method, the general formulation of the model requires the estimation of a very large number of parameters, and we therefore estimate a series of models of increasing complexity. We have shown that even the basic version, which assumes fixed order sizes and no heterogeneity in the skewness of the distribution, is still able to generate realistic simulations of simulated data. However, certain parameter vectors produce daily price dynamics that are more extreme than those observed in reality. Relaxing these assumptions generally leads to an improvement in the model estimates, in terms of their ability to produce simulations that closely reflect the price and volume dynamics observed in real data.

Our objective was to produce a realistic simulator of daily LOB activity, with the aim of utilising it to test the potential effects of regulatory interventions. We therefore evaluated the effect of implementing a quote-to-trade ratio, where we enforced a stochastic limitation on the number of cancellations, relative to the submitted limit orders. We found that altering the basic model to enforce regulation that limits the rate of cancelled to submitted orders to either 99.8% (similarly to the LSE) or 99%, had little effect on the price process. However, once we limited this percentage 95%, we observed a more constrained price process, where the excessive volatility observed for particular parameter vectors was reduced.

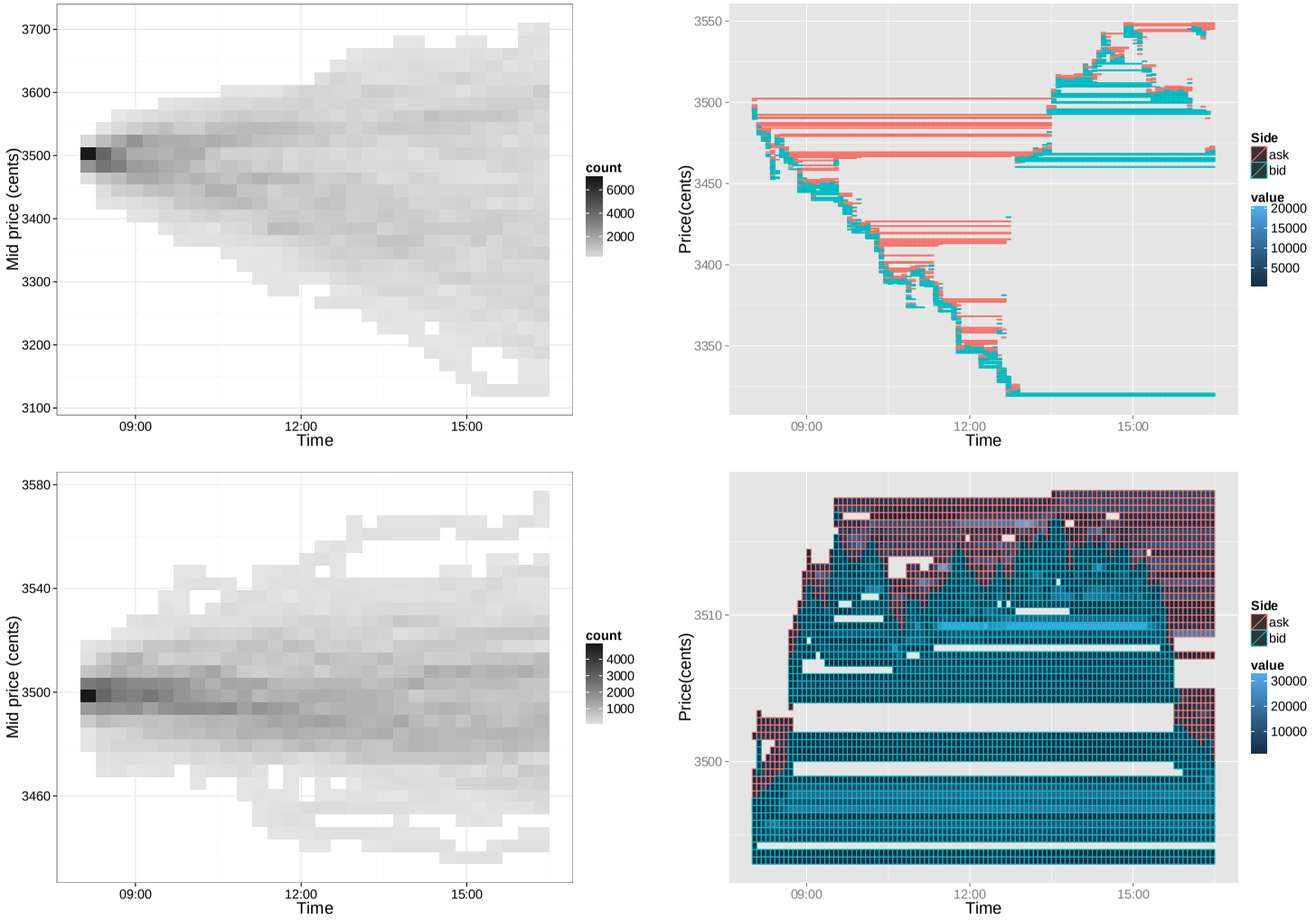


Figure 6.14: Simulations using 2 of the non-dominated parameter vectors resulting from estimating the basic model with NSGA-II, but assume that order sizes are drawn from a mixture of gamma distributions.

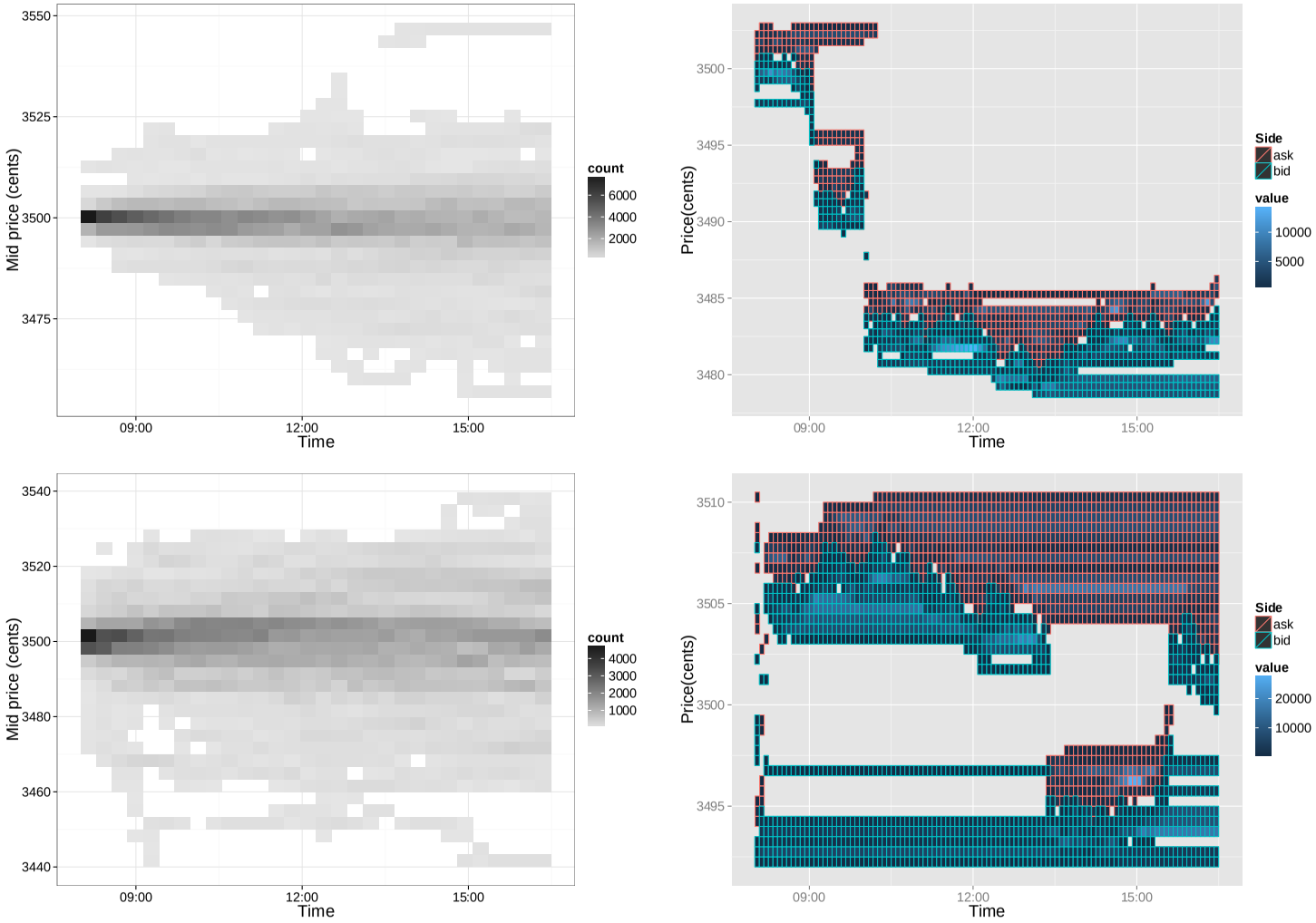


Figure 6.15: Simulations using 2 of the non-dominated parameter vectors resulting from estimating the basic model with NSGA-II, but relaxing the assumption that the elements of the skewness vector in the Multivariate Skew-t distribution are equal.

Chapter 7

Conclusion

> *It is well known that a vital ingredient of success is not knowing that what you're attempting can't be done.*
>
> — Terry Pratchett, *Equal rites*

7.1 Summary and contributions

The research work presented in this thesis constitutes an important contribution towards modelling various features of the Limit Order Book, the pre-eminent trading mechanism in modern financial markets. The thesis is practically relevant, as the empirical analysis improved immeasurably due to the availability of a vast, and very recent equities dataset from Chi-X, a European multilateral trading facility. The challenges inherent in leveraging such a dataset for academic research were the first hurdle, as the tools required to recreate the LOB from event logs were not available at the start of this thesis, or were not sufficiently detailed to capture the LOB aspects of interest. A set of tools was therefore created, and we have provided a guideline regarding the various software design options that have resulted in an efficient implementation.

7.1.1 Liquidity resilience

The first contribution of this thesis is a theoretical one, and is in response to the incomplete picture provided about liquidity by current definitions. In particular, while aspects such as the depth of the LOB or the tightness of the spread have been the subject of both theoretical and empirical studies over more than 50 years, the aspect of resilience has been rather neglected. In the electronic LOB, the availability of direct, high speed access to exchanges has enabled traders to partition large orders for staggered execution throughout the trading day, in order to reduce the costs of immediate trading. In this environment, the rate of order replenishment after a liquidity shock, which is closely related to the aspect of liquidity resilience, is central to the success of such algorithms.

The thesis established a new definition for liquidity resilience, so that it is informative to a variety of stakeholders. For the first time, this aspect of liquidity was formally defined for any choice of liquidity measure, as well as any choice of threshold, over which one would be interested in liquidity fluctuations. The 'Threshold Exceedance Duration', or TED, captures the time required for liquidity to return to the LOB after a shock and, as such, is related to the rate of volume replenishment.

The variation in this quantity was found to be related to the state of the LOB during the day. Through a survival time regression framework, we were able to relate these short intra-day liquidity droughts to LOB covariates, such as the spread, the volume on either side of the LOB, as well as various indicators of local activity. Due to the large number of possible covariates, as well as their changing importance in explaining the TED, the model was fit to the data every day. Through an extensive model selection procedure, considering a large number of model structures, and for a large number of liquid stocks, we were able to identify the covariates which were more likely to be explanatory in the regression. The empirical analysis suggested that, ceteris paribus, the LOB generally took longer to recover from larger deviations from a threshold of liquidity. On the other hand, frequent fluctuations were associated with a swifter return to the same threshold level.

The analysis presented regarding the effect of the LOB covariates enables the identification of 'regimes' of the LOB, in which liquidity resilience is high (low) and thus the duration of TEDs is short (long). In this way, the model could help determine the optimal size of the trading blocks, so that the time required for liquidity to recover is below some desired value, for a given LOB state. We have shown that the model has substantial predictive power for the duration of intra-day LOB droughts, when compared to naïve approaches, and the incorporation of the approach into an optimal execution framework will thus be a natural extension for the future.

7.1.2 Liquidity and resilience commonality

In recent years there has been a burgeoning interest measuring liquidity commonality, and it has been shown to be prevalent in the equities asset class, across industries and across countries. This can arise due to the presence of funds with correlated trading patterns, and is prevalent particularly during equity market breaks and debt market crises. This therefore creates a risk for market maker holding inventory in such assets.

The thesis furthered the state of the art in this area by showing that existing approaches, which employ Principal Components Analysis (PCA) regression to quantify commonality, are

not always appropriate. Firstly, through synthetic examples, it was shown that PCA is not particularly effective at separating out non-Gaussian components, compared to, for example, the ICA (Independent Components Analysis) method. Secondly, using liquidity data extracted from real trading data for 82 assets from 3 countries, we showed that the principal factors track the most illiquid assets, which act as outliers in the original dataset.

We also proposed a method to capture commonality in the new liquidity resilience metric introduced in this thesis. The initial step was to extend the metric from a single threshold to a series of thresholds, so that one obtains a 'liquidity resilience profile', which indicates the expected duration of an exceedance above any threshold. As with the original liquidity resilience study, this analysis was performed for two liquidity measures, the insider spread and the Xetra Liquidity Measure (XLM), a cost-of-round-trip type measure. The task was then to quantify the commonality in these curves across stocks, as a measure of liquidity resilience commonality.

An empirical contribution of this thesis is the use of functional data analysis (FDA) in this setting for the first time. This enabled a reduction in the dimensionality of the liquidity resilience data, as well as a comparison of functional data forms. Once these liquidity resilience profiles were converted to functional forms, we were able to use functional PCA, in order obtain the market factors contributing to liquidity resilience. Analogously to the PCA regression approach, we then quantified the explanatory power of these market factors for the resilience of each asset. The results suggest that these market factors are more explanatory about deviations from low thresholds, and for relatively simple liquidity measures, such as the inside spread.

7.1.3 Stochastic agent-based LOB modelling

The LOB is a complicated, multivariate, event-driven stochastic process, resulting from the combination of buy and sell orders in a multi-level queueing framework. Its dynamics are both of interest in the trading community, but also an active area of research, and recent academic contributions include efforts to explain statistical properties of stock prices, the shape of the LOB, and models of optimal execution. The final chapter of this thesis therefore aimed to capture pervasive features of the LOB and explore the utility of simulation models as testbeds for financial regulation.

We have presented a new form of simulation model for the limit order book, in which the distinction between the agent types was based on their liquidity motivations. Contrary to previous approaches, our approach modelled the aggregate trading activity of the two agent types (liquidity providers and liquidity demanders), rather than individual strategies. The model also

stands out because of its flexibility, as the trading behaviours are expressed in a stochastic model framework that is significantly more detailed than that typically observed in ABM studies. The unique advantage of this model is its ability to capture the dependence in limit order, market order and cancellation activity at different LOB levels, while still being interpretable with regard to agent motivations.

For the estimation of the model, a new method was proposed, which combined a simulation-based likelihood procedure called indirect inference with multi-objective optimisation. In this way, we were able to calibrate the model so that the resulting simulations exhibited price and volume dynamics approximating those observed in the real data. More complex versions of the model were also estimated, allowing for, e.g., heterogeneity in order sizes, and were found to bring simulated dynamics closer to those observed in real data.

The objective of this approach was to create a realistic simulator of daily LOB activity, so that one could determine the effect that a regulatory intervention would have. In the final part of this contribution, we showed how one could impose a 'quote-to-trade ratio', one of the recently discussed interventions aimed towards curbing excessive trading activity and excessive numbers of cancellations in particular. In the context of the model, such an intervention was found to be useful in reducing intra-day volatility and limiting the probability of extreme price moves.

7.2 Future research directions

The research presented in this thesis lies at the intersection of Computer Science, Statistics and Finance and has benefited greatly from collaborations with people in these fields. We anticipate that future work will continue to bring together aspects from different research areas, for financial applications predominantly. In this section, we will set out the research agenda for the near future, using the work presented here as a starting point.

Firstly, the natural extension of the liquidity resilience model would be in an optimal execution setting. In modern financial markets, traders wishing to buy or sell a large amount of an asset would typically do so in stages, in order to take advantage of order replenishment, thereby incurring smaller costs in the process. Recent models of optimal execution (e.g. Obizhaeva and Wang [2013], Alfonsi et al. [2010]) have considered the effect of liquidity resilience on the total cost of execution, but typically consider resilience to be constant or have a simple parametric form. The expected resilience of LOB liquidity was shown to be related to the state of the LOB in this thesis. Given how central the rate of order replenishment is to optimal execution, incorporating the heterogeneity of intra-day liquidity resilience into such an optimal execution

model would most certainly be an important contribution.

Secondly, the liquidity commonality study was performed over assets trading on a single venue. The proliferation of trading venues in the last few years, however, as a result of regulation aiming at increasing competition, means that considering only single-venue trading activity will only give an incomplete picture about asset liquidity. Large algorithmic traders now employ smart order routing algorithms, which divert trading volume to exchanges based on a variety of different factors. These algorithms are very likely to create commonality in liquidity, as well as liquidity resilience, in different venues. Once a sufficiently rich dataset is obtained, quantifying this commonality would be a relatively straightforward extension to the work presented here.

Finally, the agent-based model developed here allows for a large array of experimental studies, including estimating the effects of regulatory interventions. The thesis outlined a brief case study of the effect of introducing a quote-to-trade ratio, but there are various other interventions considered, including minimum resting times, transaction taxes, and switching to call auction mechanisms after a shock, the effects of which are still unclear. Exploring the possible effects of these interventions on the dynamics of the ABM simulations would thus be informative to a regulator trying to identify whether the interventions will achieve their objectives.

Appendix A

Generalised Gamma distribution for TED random variables

For a more flexible distributional form for the TED random variables, we propose a three parameter distributional family, namely the Generalised Gamma distribution (hereafter g.g.d.). We assume that the TED random variables are conditionally independent, given the LOB covariates:

$$\tau_i^{\text{TED}} \overset{i.i.d}{\sim} F(\tau; k, a, b) = \frac{\gamma\left(k, \left(\frac{\tau}{a}\right)^b\right)}{\Gamma(k)} \tag{A.1}$$

with the incomplete gamma function defined as:

$$\gamma(x, y) = \int_0^y \tau^{x-1} e^{-\tau} dt \tag{A.2}$$

The g.g.d. family includes as sub-families several popular parametric survival models: the exponential model ($b = k = 1$), the Weibull distribution (with $k = 1$), the Gamma distribution (with $b = 1$) and the Lognormal model as a limiting case (as $k \to \infty$).

The resulting density for the generalised gamma distribution is analytic and given by

$$f_\tau(\tau; k, a, b) = \frac{b}{\Gamma(k)} \frac{\tau^{bk-1}}{a^{bk}} \exp\left(-\left(\frac{\tau}{a}\right)^b\right) \tag{A.3}$$

with parameter ranges $k > 0$, $a > 0$ and $b > 0$ and a support of $\tau \in (0, \infty)$. We note that for this class of model one can write the survival function explicitly in closed form.

We now wish to relate this statistical model assumption to a set of explanatory variables (covariates) from lagged values of the LOB. To achieve this, it will be beneficial to work on the log scale with $\ln(\tau)$, i.e. with the log-generalized gamma distribution (hereafter l.g.g.d.), as this parameterisation improves identifiability and estimation of parameters. Discussions on this point are provided in significant detail in Lawless [1980].

Under the AFT framework, the regression structure we adopt for the l.g.g.d. model involves constant and nonstochastic terms k and b as well as the following loglinear form for the time-varying location coefficient $a(\boldsymbol{x}_t)$:

$$a(\boldsymbol{x}_t, \boldsymbol{z}_t) = \exp\left(\beta_0 + \sum_{s=1}^{p} x_t^{(s)} \beta_s\right). \tag{A.4}$$

with p covariates $\boldsymbol{x}_t = \left\{x_t^{(s)}\right\}_{s=1}^{p}$ measured instantaneously at the point of exceedance $t = T_i$. Each of the covariates is a transform from the LOB for which the liquidity measure is observed, and all covariates are described in Section 4.4.5. We note that we also considered models with interactions between the covariates, but interaction terms were not found to be significant in the majority of our models.

Under this AFT model with this location regression structure, we observe that the conditional mean of the survival times is also related directly to this linear structure where for the i-th exceedance of the threshold, we have

$$\mathbb{E}\left[\tau_i^{\text{TED}} | \boldsymbol{x}_{T_i}\right] = a(\boldsymbol{x}_{T_i}) \left(\frac{1}{k}\right)^{\frac{1}{b}} \frac{\Gamma\left(k + \frac{1}{b}\right)}{\Gamma(k)} \tag{A.5}$$

see details in Lo et al. [2002].

In addition, the conditional quantile function for a given quantile level u, is obtained through the transformation of the analytic closed-form quantile function of a Gamma random variable, denoted by G^{-1}, with shape u and scale k, which gives the conditional expression

$$\begin{aligned} Q(u; \boldsymbol{x}_t) &= F^{-1}(\tau_i; \boldsymbol{x}_t, u) \\ &= \exp\left(\beta_0 + \sum_{s=1}^{p} x_t^{(s)} \beta_s\right) \left[\left(\frac{1}{k}\right) G^{-1}(u; u, k)\right]^{\frac{1}{b}} \end{aligned}$$

Bibliography

Abhay Abhyankar, Dipak Ghosh, Eric Levin, and RJ Limmack. Bid-ask spreads, trading volume and volatility: Intra-day evidence from the London Stock Exchange. *Journal of Business Finance & Accounting*, 24(3):343–362, 1997.

John Affleck-Graves, Shantaram P Hegde, and Robert E Miller. Trading mechanisms and the components of the bid-ask spread. *The Journal of Finance*, 49(4):1471–1488, 1994.

Hee-Joon Ahn, Kee-Hong Bae, and Kalok Chan. Limit orders, depth, and volatility: Evidence from the stock exchange of Hong Kong. *The Journal of Finance*, 56(2):767–788, 2001.

Yacine Aït-Sahalia and Mehmet Saglam. High frequency traders: Taking advantage of speed. Technical report, National Bureau of Economic Research, 2013.

Mohammad Al-Suhaibani and Lawrence Kryzanowski. An exploratory analysis of the order book, and order flow and execution on the Saudi stock market. *Journal of Banking & Finance*, 24(8):1323–1357, 2000.

Simone Alfarano and Mishael Milakovic. Network structure and n-dependence in agent-based herding models. *Journal of Economic Dynamics and Control*, 33(1):78–92, January 2009.

Simone Alfarano, Thomas Lux, and Friedrich Wagner. Estimation of agent-based models: the case of an asymmetric herding model. *Computational Economics*, 26(1):19–49, 2005.

Simone Alfarano, Thomas Lux, and Friedrich Wagner. Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach. *Journal of Economic Dynamics and Control*, 32(1):101–136, 2008.

Aurélien Alfonsi, Antje Fruth, and Alexander Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

Robert Almgren and Neil Chriss. Value under liquidation. *Risk*, 12(12):61–63, 1999.

Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

Yakov Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56, 2002.

Yakov Amihud and Haim Mendelson. Dealership market: Market-making with inventory. *Journal of Financial Economics*, 8(1):31–53, 1980.

Yakov Amihud and Haim Mendelson. Asset pricing and the bid-ask spread. *Journal of financial Economics*, 17(2):223–249, 1986.

Yakov Amihud and Haim Mendelson. Liquidity, the value of the firm, and corporate finance. *Journal of Applied Corporate Finance*, 20(2):32–45, 2008.

Christophe Andrieu, Éric Moulines, et al. On the ergodicity properties of some adaptive MCMC algorithms. *The Annals of Applied Probability*, 16(3):1462–1505, 2006.

James B Ang and Warwick J McKibbin. Financial liberalization, financial sector development and growth: evidence from Malaysia. *Journal of Development Economics*, 84(1):215–233, 2007.

Philip Arestis, Panicos O Demetriades, and Kul B Luintel. Financial development and economic growth: the role of stock markets. *Journal of Money, Credit and Banking*, pages 16–41, 2001.

W. Brian Arthur, John Holland, Blake LeBaron, Richard Palmer, and Paul Tayler. Asset pricing under endogenous expectations in an artificial stock market. *Available at SSRN 2252*, 1996.

Marco Avellaneda and Sasha Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

Walter Bagehot. The only game in town. *Financial Analysts Journal*, 27(2):12–14, 1971.

Avraham Beja and M Barry Goldman. On the dynamic behavior of prices in disequilibrium. *The Journal of Finance*, 35(2):235–248, 1980.

Dimitris Bertsimas and Andrew W Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

Ekkehart Boehmer, Gideon Saar, and Lei Yu. Lifting the veil: An analysis of pre-trade transparency at the NYSE. *The Journal of Finance*, 60(2):783–815, 2005.

Richard Bookstaber. Using agent-based models for analyzing threats to financial stability. *Office of Financial Research Working Paper series*, 2012.

Luke Bortoli, Alex Frino, Elvis Jarnecic, and David Johnstone. Limit order book transparency, execution risk, and market liquidity: Evidence from the Sydney Futures Exchange. *Journal of Futures Markets*, 26(12):1147–1167, 2006.

Mike J Bradburn, Taane G Clark, Susan B Love, and Doug G Altman. Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods. *British journal of cancer*, 89(3):431, 2003.

Michael J Brennan and Avanidhar Subrahmanyam. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of financial economics*, 41(3): 441–464, 1996.

William A Brock and Cars H Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic dynamics and Control*, 22(8-9):1235–1274, 1998.

Paul Brockman and Dennis Y Chung. An analysis of depth behavior in an electronic, order-driven environment. *Journal of Banking & Finance*, 23(12):1861–1886, 1999.

Paul Brockman, Dennis Y Chung, and Christophe Pérignon. Commonality in liquidity: A global perspective. *Journal of Financial and Quantitative Analysis*, 44(04):851–882, 2009.

Jonathan Brogaard. High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, 2010.

James Brugler and Oliver Linton. Single stock circuit breakers on the London Stock Exchange: do they improve subsequent market quality? Technical report, Centre for Microdata Methods and Practice working paper, 2014.

Markus K Brunnermeier. Deciphering the liquidity and credit crunch 2007-08. Technical report, National Bureau of Economic Research, 2008.

Markus K Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *Review of Financial studies*, 22(6):2201–2238, 2009.

Alexander W Butler, Gustavo Grullon, and James P Weston. Stock market liquidity and the cost of issuing equity. *Journal of Financial and Quantitative Analysis*, 40(02):331–348, 2005.

John Y Campbell, Sanford J Grossman, and Jiang Wang. Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4):905–939, 1993.

Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

Stephen Giovanni Cecchetti, Enisse Kharroubi, Stephen Giovanni Cecchetti, and Stephen Giovanni Cecchetti. Reassessing the impact of finance on growth. Technical report, Bank for International Settlements, 2012.

B. Chakrabarty, K. Tyurin, Z. Han, and X. Zheng. A competing risk analysis of executions and cancellations in a limit order market. 2006.

Bidisha Chakrabarty, Pankaj K Jain, Andriy Shkilko, and Konstantin Sokolov. Quote intensity and market quality: Effects of the sec naked access ban. *Available at SSRN 2328231*, 2013.

Anirban Chakraborti, Ioane Muni Toke, Marco Patriarca, and Frederic Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

Kalok C Chan, William G Christie, and Paul H Schultz. Market structure and the intraday pattern of bid-ask spreads for NASDAQ securities. *Journal of Business*, pages 35–60, 1995.

Long Chen, David A Lesmond, and Jason Wei. Corporate yield spreads and bond liquidity. *The Journal of Finance*, 62(1):119–149, 2007.

Shu-Heng Chen, Chia-Ling Chang, and Ye-Rong Du. Review: Agent-based economic models and econometrics. *The Knowledge Engineering Review*, 27(2):187–219, 2012.

Tao Chen, Elaine Martin, and Gary Montague. Robust probabilistic PCA with missing data and contribution analysis for outlier detection. *Computational Statistics & Data Analysis*, 53 (10):3706–3716, 2009.

Giusy Chesini and Elisa Giaretta. Regulating high-frequency trading: An examination of European, US and Australian equity market structures. *Financial Systems, Markets and Institutional Changes*, page 152, 2014.

Carl Chiarella and Giulia Iori. A simulation analysis of the microstructure of double auction markets. *Quantitative Finance*, 2(5):346–353, 2002.

Carl Chiarella, Giulia Iori, and Josep Perelló. The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, 33(3): 525–537, 2009.

Michael Chlistalla, Bernhard Speyer, Sabine Kaiser, and Thomas Mayer. High-frequency trading. *Deutsche Bank Research*, pages 1–19, 2011.

Jin-Wan Cho and Edward Nelling. The probability of limit-order execution. *Financial Analysts Journal*, pages 28–33, 2000.

Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Commonality in liquidity. *Journal of Financial Economics*, 56(1):3–28, 2000.

Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Market liquidity and trading activity. *The Journal of Finance*, 56(2):501–530, 2001.

Hugh L Christensen, Richard E Turner, Simon I Hill, and Simon J Godsill. Rebuilding the limit order book: Sequential Bayesian inference on hidden states. *Quantitative Finance*, 13(11): 1779–1799, 2013.

Kee H Chung, Bonnie F Van Ness, and Robert A Van Ness. Limit orders and the bid–ask spread. *Journal of Financial Economics*, 53(2):255–287, 1999.

Dave Cliff, Janet Bruten, et al. Zero not enough: On the lower limit of agent intelligence for continuous double auction markets. *HP Laboratories Technical Report HPL*, 1997.

Carlos Coello Coello, Gary B Lamont, and David A Van Veldhuizen. *Evolutionary algorithms for solving multi-objective problems*. Springer, 2007.

Norma Coffey, Andrew J Harrison, Orna Donoghue, and Kevin Hayes. Common functional principal components analysis: A new approach to analyzing human movement data. *Human movement science*, 30(6):1144–1166, 2011.

Kalman J Cohen, Steven F Maier, Robert A Schwartz, and David K Whitcomb. A simulation model of stock exchange trading. *Simulation*, 41(5):181–191, 1983.

Jean-Edouard Colliard and Peter Hoffmann. Sand in the chips: Evidence on taxing transactions in an electronic market. Technical report, Working paper, European Central Bank, 2013.

Carole Comerton-Forde, Terrence Hendershott, Charles M Jones, Pamela C Moulton, and Mark S Seasholes. Time variation in liquidity: The role of market-maker inventories and revenues. *The Journal of Finance*, 65(1):295–331, 2010.

Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3): 287–314, 1994.

Andrea Consiglio, Valerio Lacagnina, and Annalisa Russino. A simulation analysis of the microstructure of an order driven financial market with multiple securities and portfolio choices. *Quantitative Finance*, 5(1):71–87, 2005.

George M Constantinides. Capital market equilibrium with transaction costs. *The Journal of Political Economy*, pages 842–862, 1986.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007.

Rama Cont. Statistical modeling of high-frequency financial data. *Signal Processing Magazine, IEEE*, 28(5):16–25, 2011.

Rama Cont and Adrien De Larrard. Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.

Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.

Peter Craven and Grace Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, 1978.

Marcus G Daniels, J Doyne Farmer, László Gillemot, Giulia Iori, and Eric Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical review letters*, 90(10):108102, 2003.

Vincent Darley and Alexander V. Outkin. *Nasdaq Market Simulation: Insights on a Major Market from the Science of Complex Adaptive Systems*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2007.

Herbert Dawid and Michael Neugart. Agent-based models for economic policy design. *Eastern Economic Journal*, 37(1):44–50, 2011.

Richard H Day and Weihong Huang. Bulls, bears and market sheep. *Journal of Economic Behavior & Organization*, 14(3):299–329, 1990.

Carl De Boor. Calculation of the smoothing spline with weighted roughness measure. *Mathematical Models and Methods in Applied Sciences*, 11(01):33–41, 2001.

Paul De Grauwe. Exchange rates in search of fundamental variables. Technical report, CEPR Discussion Papers, 1994.

Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, 2002.

Stefano Demarta and Alexander J McNeil. The t copula and related copulas. *International statistical review*, 73(1):111–129, 2005.

Ian Domowitz, Oliver Hansch, and Xiaoxin Wang. Liquidity commonality and return co-movement. *Journal of Financial Markets*, 8(4):351–376, 2005.

Jiwei Dong, Alexander Kempf, and Pradeep Yadav. Resiliency, the neglected dimension of market liquidity: Empirical evidence from the New York Stock Exchange. *Available at SSRN 967262*, 2007.

Darrell Duffie and Kenneth J Singleton. Simulated moments estimation of markov models of asset prices. *Econometrica*, 61(4):929–952, 1993.

David Easley and Maureen O'Hara. Price, trade size, and information in securities markets. *Journal of Financial economics*, 19(1):69–90, 1987.

David Easley, Marcos López de Prado, and Maureen O'Hara. The microstructure of the flash crash: flow toxicity, liquidity crashes and the probability of informed trading. *The Journal of Portfolio Management*, 37(2):118–128, 2011.

Gudrun Ehrenstein, Frank Westerhoff, and Dietrich Stauffer. Tobin tax and market depth. *Quantitative Finance*, 5(2):213–218, 2005.

Agoston E Eiben and James E Smith. *Introduction to evolutionary computing*. Springer, 2003.

Andrew Ellul and Marco Pagano. Ipo underpricing and after-market liquidity. *Review of Financial Studies*, 19(2):381–421, 2006.

Michael Emmerich, André Deutz, Oliver Schütze, Thomas Bäck, Emilia Tantar, Alexandru-Adrian Tantar, Pierre del Moral, Pierrick Legrand, Pascal Bouvry, and Carlos Coello Coello. Evolve-a bridge between probability, set oriented numerics, and evolutionary computation iv. Springer, 2013.

Cornelia Ernst, Sebastian Stange, and Christoph Kaserer. Measuring market liquidity risk-which model works best? *Journal of Financial Transformation*, 35:133–146, 2012.

J Doyne Farmer and Shareen Joshi. The price dynamics of common trading strategies. *Journal of Economic Behavior & Organization*, 49(2):149–171, 2002.

Thierry Foucault, Ohad Kadan, and Eugene Kandel. Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4):1171–1217, 2005.

Thierry Foucault, Marco Pagano, Ailsa Roell, and Ailsa Röell. *Market Liquidity: Theory, Evidence, and Policy*. Oxford University Press, 2013.

Jeffrey A. Frankel and Kenneth A. Froot. Chartists, fundamentalists, and the demand for dollars. *Greek Econom. Rev.*, 10:49–102, 1988.

Alex Frino, Vito Mollica, and Zeyang Zhou. Commonality in liquidity across international borders: Evidence from futures markets. *Journal of Futures Markets*, 2014.

Thomas Fung and Eugene Seneta. Tail dependence for two skew t distributions. *Statistics & probability letters*, 80(9):784–791, 2010.

George M Furnival and Robert W Wilson. Regressions by leaps and bounds. *Technometrics*, 16(4):499–511, 1974.

A. Ronald Gallant and George Tauchen. Which moments to match? *Econometric Theory*, 12 (04):657–681, 1996.

Kenneth D Garbade and K Garbade. *Securities markets*. McGraw-Hill New York, 1982.

Cristian Gatu and Erricos John Kontoghiorghes. Branch-and-bound algorithms for computing the best-subset regression models. *Journal of Computational and Graphical Statistics*, 15 (1):139–156, 2006.

Oscar Gelderblom and Joost Jonker. Completing a financial revolution: The finance of the dutch east india trade and the rise of the amsterdam capital market, 1595–1612. *The Journal of Economic History*, 64(03):641–672, 2004.

Marc G Genton and Elvezio Ronchetti. Robust indirect inference. *Journal of the American Statistical Association*, 98(461):67–76, 2003.

Manfred Gilli and Peter Winker. A global optimization heuristic for estimating agent based models. *Computational Statistics & Data Analysis*, 42(3):299–312, 2003.

Mark Girolami and Colin Fyfe. Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms. In *A. C, Back A (eds.), NIPS-96 Blind Signal Separation Workshop*, volume 8. Citeseer, 1996.

Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

Dhananjay K Gode and Shyam Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, pages 119–137, 1993.

Raymond William Goldsmith. *Financial structure and development*, volume 1. Yale university press, New Haven, 1969.

Michael A Goldstein and Kenneth A Kavajecz. Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. *Journal of Financial Economics*, 56(1): 125–149, 2000.

Peter Gomber and Uwe Schweickert. The market impact-liquidity measure in electronic securities trading. *Die Bank*, 7:485–489, 2002.

Peter Gomber, Uwe Schweickert, and Erik Theissen. Zooming in on liquidity. In *EFA 2004 Maastricht Meetings Paper*, number 1805, 2004.

Peter Gomber, Uwe Schweickert, and Erik Theissen. Liquidity dynamics in an electronic open limit order book: An event study approach. Technical report, CFR working paper, 2011.

Martin D Gould, Mason A Porter, Stacy Williams, Mark McDonald, Daniel J Fenn, and Sam D Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

Christian Gourieroux and Alain Monfort. *Simulation-based econometric methods*. Oxford University Press, 1997.

Christian Gourieroux, Alain Monfort, and Eric Renault. Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118, 1993.

Christian Gouriéroux, Peter CB Phillips, and Jun Yu. Indirect inference for dynamic panel models. *Journal of Econometrics*, 157(1):68–77, 2010.

Ruslan Y Goyenko, Craig W Holden, and Charles A Trzcinka. Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92(2):153–181, 2009.

Jeremy Greenwood and Bruce D Smith. Financial markets in development, and the development of financial markets. *Journal of Economic Dynamics and Control*, 21(1):145–181, 1997.

Sanford J Grossman and Merton H Miller. Liquidity and market structure. *the Journal of Finance*, 43(3):617–633, 1988.

Heikki Haario, Marko Laine, Antonietta Mira, and Eero Saksman. Dram: efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006.

Frank E Harrell. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer, 2001.

Larry Harris. *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press, USA, 2002.

L.E. Harris, Salomon Brothers Center for the Study of Financial Institutions, and Leonard N. Stern School of Business. *Liquidity, Trading Rules, and Electronic Trading Systems*. Monograph series in finance and economics. New York University Salomon Center, 1991.

Joel Hasbrouck and Gideon Saar. Technology and liquidity provision: The blurring of traditional definitions. *Journal of financial Markets*, 12(2):143–172, 2009.

Joel Hasbrouck and Gideon Saar. Low-latency trading. *Journal of Financial Markets*, 2013.

Joel Hasbrouck and Duane J. Seppi. Common factors in prices, order flows, and liquidity. *Journal of financial Economics*, 59(3):383–411, 2001.

Nikolaus Hautsch and Ruihong Huang. The market impact of a limit order. *Journal of Economic Dynamics and Control*, 36(4):501–522, 2012.

Knut Heggland and Arnoldo Frigessi. Estimating functions in indirect inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):447–462, 2004.

Terrence Hendershott and Mark S. Seasholes. Market maker inventories and stock prices. *The American Economic Review*, 97(2):pp. 210–214, 2007. ISSN 00028282.

Terrence Hendershott, Charles M Jones, and Albert J Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33, 2011.

Craig W Holden, Stacey E Jacobsen, and Avanidhar Subrahmanyam. The empirical analysis of liquidity. *Forthcoming, Foundations and Trends in Finance*, 2015.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

Philip Hougaard. Fundamentals of survival data. *Biometrics*, 55(1):13–22, 1999.

He Huang and Alec N Kercheval. A generalized birth–death stochastic model for high-frequency order book dynamics. *Quantitative Finance*, 12(4):547–557, 2012.

Roger D Huang and Hans R Stoll. The components of the bid-ask spread: A general approach. *Review of Financial Studies*, 10(4):995–1034, 1997.

Gur Huberman and Dominika Halka. particularly. *Journal of Financial Research*, 24(2):161–178, 2001.

Evan J Hughes. Evolutionary many-objective optimisation: many once or one many? In *The 2005 IEEE Congress on Evolutionary Computation, 2005*, volume 1, pages 222–227. IEEE, 2005.

Aapo Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, 10(3):626–634, 1999.

Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

Paul Irvine, George Benston, and Eugene Kandel. Liquidity beyond the inside spread: Measuring and using information in the limit order book. Technical report, 2000.

Pankaj Jain. Institutional design and liquidity at stock exchanges around the world. *Available at SSRN 869253*, 2003.

Pankaj K Jain. Financial market design and the equity premium: Electronic versus floor trading. *The Journal of Finance*, 60(6):2955–2985, 2005.

Marco A Janssen and Elinor Ostrom. Empirically based, agent-based models. *Ecology and Society*, 11(2):37, 2006.

Gerald R Jensen and Theodore Moorman. Inter-temporal variation in the illiquidity premium. *Journal of Financial Economics*, 98(2):338–358, 2010.

Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

David Kane, Andrew Liu, and Khanh Nguyen. Analyzing an electronic limit order book. *The R Journal*, 2(64-68):1, 2011.

S Kannan, S Baskar, James D McCalley, and P Murugan. Application of nsga-ii algorithm to generation expansion planning. *Power Systems, IEEE Transactions on*, 24(1):454–461, 2009.

Nina Karnaukh, Angelo Ranaldo, and Paul Söderlind. Understanding fx liquidity. *University of St. Gallen, School of Finance Research Paper*, (2013/15), 2013.

G Andrew Karolyi, Kuan-Hui Lee, and Mathijs A Van Dijk. Understanding commonality in liquidity around the world. *Journal of Financial Economics*, 105(1):82–112, 2012.

Kenneth A Kavajecz. A specialist's quoted depth and the limit order book. *The Journal of Finance*, 54(2):747–771, 1999.

Nazan Khan. *Bayesian optimization algorithms for multiobjective and hierarchically difficult problems*. PhD thesis, University of Illinois at Urbana-Champaign Urbana, IL, 2003.

Nazan Khan, Nazan Khan, David E Goldberg, David E Goldberg, Martin Pelikan, and Martin Pelikan. Multi-objective bayesian optimization algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 2002.

Andrei A Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: The impact of high frequency trading on an electronic market. 2014.

Alan Kirman. Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, 108(1): 137–156, 1993.

Mario Köppen, Raul Vicente-Garcia, and Bertram Nickolay. Fuzzy-pareto-dominance and its application in evolutionary multi-objective optimization. In *Evolutionary Multi-Criterion Optimization*, pages 399–412. Springer, 2005.

Robert A Korajczyk and Ronnie Sadka. Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics*, 87(1):45–72, 2008.

Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.

Mr Luc Laeven and Fabian Valencia. *Resolution of Banking Crises: The Good, the Bad, and the Ugly*. Number 10-146. International Monetary Fund, 2010.

Philippe Lambert, Dave Collett, Alan Kimber, and Rachel Johnson. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in medicine*, 23(20):3177–3192, 2004.

Jeremy Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

Marco Laumanns and Jiri Ocenasek. Bayesian optimization algorithms for multi-objective optimization. In *Parallel Problem Solving from Nature VII*, pages 298–307. Springer, 2002.

Jerry F Lawless. Inference in the generalized gamma and log gamma distributions. *Technometrics*, 22(3):409–419, 1980.

Matthias Lengnick and Hans-Werner Wohltmann. Agent-based financial markets and new Keynesian macroeconomics: A synthesis. *Journal of Economic Interaction and Coordination*, 8 (1):1–32, 2013.

Ross Levine. Financial development and economic growth: views and agenda. *Journal of economic literature*, 35(2):688–726, 1997.

Ross Levine, Norman Loayza, and Thorsten Beck. Financial intermediation and growth: Causality and causes. *Journal of monetary Economics*, 46(1):31–77, 2000.

Marco LiCalzi and Paolo Pellizzari. Fundamentalists clashing over the book: a study of order-driven stock markets. *Quantitative Finance*, 3(6):470–480, 2003.

Andrew W Lo, A Craig MacKinlay, and June Zhang. Econometric models of limit-order executions. *Journal of Financial Economics*, 65(1):31–71, 2002.

Thomas Lumley. The leaps package. *The R project for statistical computation*, 2004.

Thomas Lux. Herd behaviour, bubbles and crashes. *The economic journal*, pages 881–896, 1995.

Thomas Lux. Time variation of second moments from a noise trader/infection model. *Journal of Economic Dynamics and Control*, 22(1):1–38, 1997.

Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

Thomas Lux and Didier Sornette. On rational bubbles and fat tails. *Journal of Money, Credit, and Banking*, 34(3):589–610, 2002.

Ananth Madhavan, David Porter, and Daniel Weaver. Should securities markets be transparent? *Journal of Financial Markets*, 8(3):265–287, 2005.

Loriano Mancini, Angelo Ranaldo, and Jan Wrampelmeyer. Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance*, 68(5): 1805–1841, 2013.

Benoit B Mandelbrot. *The variation of certain speculative prices*. Springer, 1997.

Katiuscia Mannaro, Michele Marchesi, and Alessio Setzu. Using an artificial financial market for assessing the impact of Tobin-like transaction taxes. *Journal of Economic Behavior & Organization*, 67(2):445–462, 2008.

Edward R Mansfield and Billy P Helms. Detecting multicollinearity. *The American Statistician*, 36(3a):158–160, 1982.

Sebastiano Manzan and Frank H Westerhoff. Heterogeneous expectations, exchange rate dynamics and predictability. *Journal of economic behavior & Organization*, 64(1):111–128, 2007.

Ben R Marshall, Nhut H Nguyen, and Nuttawat Visaltanachoti. Liquidity commonality in commodities. *Journal of Banking & Finance*, 37(1):11–20, 2013.

Sergei Maslov. Simple model of a limit order-driven market. *Physica A: Statistical Mechanics and its Applications*, 278(3):571–578, 2000.

Stewart Mayhew. Competition, market structure, and bid-ask spreads in stock option markets. *The Journal of Finance*, 57(2):931–958, 2002.

Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026, 1989.

Ronald I McKinnon. *Money and capital in economic development*. Brookings Institution Press, 1973.

Rupert G Miller Jr. *Survival analysis*, volume 66. John Wiley & Sons, 2011.

Reinhold C Mueller and Frederic Chapin Lane. *The Venetian money market: banks, panics, and the public debt, 1200-1500*, volume 2. Johns Hopkins University Press Baltimore, 1997.

Jinzhong Niu, Kai Cai, Simon Parsons, Enrico Gerding, and Peter McBurney. Characterizing effective auction mechanisms: Insights from the 2007 TAC market design competition. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 1079–1086. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

Maureen O'hara. *Market microstructure theory*, volume 108. Blackwell Cambridge, MA, 1995.

Maureen O'Hara and Mao Ye. Is market fragmentation harming market quality? *Journal of Financial Economics*, 100(3):459–474, 2011.

Ariel Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, 57:1027–1057, 1989.

Efstathios Panayi and Gareth Peters. Survival models for the duration of bid-ask spread deviations. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pages 9–16. IEEE, 2014.

Efstathios Panayi and Gareth Peters. Stochastic simulation framework for the limit order book using liquidity motivated agents. *in review, Journal of Financial Engineering*, 2015a.

Efstathios Panayi and Gareth W Peters. Liquidity commonality does not imply liquidity resilience commonality: A functional characterisation for ultra-high frequency cross-sectional lob data. *to appear, Quantitative Finance Special Issue on Big Data Analytics*, 2015b.

Efstathios Panayi, Mark Harman, and Anne Wetherilt. Agent-based modelling of stock markets using existing order book data. In *Multi-Agent-Based Simulation XIII*, pages 101–114. Springer, 2013.

Efstathios Panayi, Gareth W Peters, Jon Danielsson, and Jean-Pierre Zigrand. Market liquidity resilience. *London School of Economics Working Paper Series*, 2014.

Christine A Parlour and Duane J Seppi. Limit order markets: A survey. *Handbook of financial intermediation and banking*, 5:63–95, 2008.

Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

Paolo Pellizzari and Frank Westerhoff. Some effects of transaction taxes under different microstructures. *Journal of Economic Behavior & Organization*, 72(3):850–863, 2009.

Gareth W Peters, Alice XD Dong, and Robert Kohn. A copula based bayesian approach for paid-incurred claims models for non-life insurance reserving. *arXiv preprint arXiv:1210.3849*, 2012.

Gareth William Peters, Ariane Chapelle, and Efstathios Panayi. Opening discussion on banking sector risk exposures and vulnerabilities from virtual currencies: An operational risk perspective. *in review, Journal of Operational Risk*, 2014.

Johannes Prix, Otto Loistl, and Michael Huetl. Algorithmic trading patterns in Xetra orders. *The European Journal of Finance*, 13(8):717–739, 2007.

Robin C Purshouse and Peter J Fleming. Evolutionary many-objective optimisation: An exploratory analysis. In *The 2003 Congress on Evolutionary Computation*, volume 3, pages 2066–2073. IEEE, 2003.

Marco Raberto, Silvano Cincotti, Sergio M Focardi, and Michele Marchesi. Agent-based simulation of a financial market. *Physica A: Statistical Mechanics and its Applications*, 299(1): 319–327, 2001.

James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.

Angelo Ranaldo. Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74, 2004.

Patrick M Reed and Barbara S Minsker. Striking the balance: long-term groundwater monitoring design for conflicting objectives. *Journal of Water Resources Planning and Management*, 130(2):140–149, 2004.

Kylie-Anne Richards, Gareth W Peters, and William Dunsmuir. Heavy-tailed features and empirical analysis of the limit order book volume profiles in futures markets. *arXiv preprint arXiv:1210.7215*, 2012.

Ryan Riordan and Andreas Storkenmaier. Latency, liquidity and price discovery. *Journal of Financial Markets*, 15(4):416–437, 2012.

Ryan Riordan, Andreas Storkenmaier, Martin Wagener, and S Sarah Zhang. Public information arrival: Price discovery and liquidity in electronic limit order markets. *Journal of Banking & Finance*, 37(4):1148–1159, 2013.

Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

Joan Robinson. *Rate of Interest and Other Essays*. Magmillan And Co. Ltd, 1953.

Richard Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4):1127–1139, 1984.

Ioanid Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11): 4601–4641, 2009.

Peter L Rousseau and Paul Wachtel. What is happening to the impact of financial deepening on economic growth? *Economic Inquiry*, 49(1):276–288, 2011.

Francisco J Ruge-Murcia. Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, 31(8):2599–2636, 2007.

Jeffrey R Russell, RF Engle, Y Ait-Sahalia, and LP Hansen. Analysis of high-frequency data. In *Handbook of Financial Econometrics: Tools and Techniques*, volume 1 of *Handbooks in Finance*, pages 383 – 426. North-Holland, San Diego, 2010.

Abdourahmane Sarr and Tonny Lybek. Measuring liquidity in financial markets. *IMF Working Paper*, 2002.

US Securities, Exchange Commission, the Commodity Futures Trading Commission, et al. Findings regarding the market events of may 6, 2010. *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues.*, 2010.

Konstantinos Sklavos, Lammertjan Dam, and Bert Scholtens. The liquidity of energy stocks. *Energy Economics*, 38:168–175, 2013.

Anthony A. Smith. *Three essays on the solution and estimation of dynamic macroeconomic models*. PhD thesis, Duke University, 1990.

Anthony A. Smith. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8(S1):S63–S84, 1993.

Anthony A. Smith. Indirect inference. In *The New Palgrave Dictionary of Economics, 2nd Edition*. 2008.

Michael S Smith, Quan Gan, and Robert J Kohn. Modelling dependence using skew t copulas: Bayesian inference and applications. *Journal of Applied Econometrics*, 27(3):500–522, 2012.

Robert F Stambaugh. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3), 2003.

Hans R Stoll. The pricing of security dealer services: An empirical study of NASDAQ stocks. *The Journal of Finance*, 33(4):1153–1172, 1978.

Mark P Taylor and Helen Allen. The use of technical analysis in the foreign exchange market. *Journal of international Money and Finance*, 11(3):304–314, 1992.

Gerald Tesauro and Jonathan L Bredin. Strategic sequential bidding in auctions using dynamic programming. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*, pages 591–598. ACM, 2002.

James Tobin. A proposal for international monetary reform. *Eastern Economic Journal*, 4(3): 153–159, 1978.

Ioane Muni Toke. Market making in an order book model and its impact on the spread. In *Econophysics of Order-Driven Markets*, pages 49–64. Springer, 2011.

Dimitri Vayanos and Jiang Wang. Market liquidity - theory and empirical evidence. volume 2, Part B of *Handbook of the Economics of Finance*, pages 1289 – 1361. Elsevier, 2013.

Rico Von Wyss. *Measuring and predicting liquidity in the stock market*. PhD thesis, St. Gallen University, 2004.

Frank Westerhoff. Heterogeneous traders and the tobin tax. *Journal of Evolutionary Economics*, 13(1):53–70, 2003.

Frank Westerhoff. The use of agent-based financial market models to test the effectiveness of regulatory policies. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 228(2):195–227, 2008.

Frank H Westerhoff and Roberto Dieci. The effectiveness of Keynes–Tobin transaction taxes when heterogeneous agents can trade in different markets: a behavioral finance approach. *Journal of Economic Dynamics and Control*, 30(2):293–322, 2006.

Frank H Westerhoff and Stefan Reitz. Nonlinearities and cyclical behavior: the role of chartists and fundamentalists. *Studies in Nonlinear Dynamics & Econometrics*, 7(4), 2003.

Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10 (2):8, 2007.

Peter Winker, Manfred Gilli, and Vahidin Jeleskovic. An objective function for simulation based inference on exchange rate data. *Journal of Economic Interaction and Coordination*, 2(2):125–145, 2007.

Robert A Wood, Thomas H McInish, and J Keith Ord. An investigation of transactions data for NYSE stocks. *The Journal of Finance*, 40(3):723–739, 1985.

E Christopher Zeeman. On the unstable behaviour of stock exchanges. *Journal of Mathematical Economics*, 1(1):39–49, 1974.

Frank Zhang. High-frequency trading, stock volatility, and price discovery. *Available at SSRN 1691679*, 2010.

Aimin Zhou, Bo-Yang Qu, Hui Li, Shi-Zheng Zhao, Ponnuthurai Nagaratnam Suganthan, and Qingfu Zhang. Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, 1(1):32–49, 2011.

Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary computation*, 8(2):173–195, 2000.