



City Research Online

City, University of London Institutional Repository

Citation: Maini, Vincenzo (2012). Price and liquidity discovery, jumps and co-jumps using high frequency data from the foreign exchange markets. (Unpublished Doctoral thesis, City University London)

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/2382/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

PRICE AND LIQUIDITY DISCOVERY, JUMPS AND CO-JUMPS
USING HIGH FREQUENCY DATA FROM THE FOREIGN
EXCHANGE MARKETS

Vincenzo Ludovico Maini

Thesis submitted for the Degree of Doctor of Philosophy

Faculty of Finance

Cass Business School, City University

London, England

September 2012

CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	vii
ABSTRACT	1
ACKNOWLEDGMENTS	3
INTRODUCTION	4
1 MOVING FROM PRICE TO LIQUIDITY DISCOVERY	10
1.1 Introduction	10
1.2 The Model	17
1.2.1 The Price Discovery Model proposed by Hasbrouck (1991) and Dufour and Engle (2000)	17
1.2.2 Moving from Price to Liquidity Discovery	20
1.3 The Data	25
1.3.1 Preparation of the Data	27
1.3.2 Variables	28
1.3.3 Preliminary Analysis	32
1.4 Estimation and Results	34
1.4.1 The Relevance of Time in the VAR System	35
1.4.2 Analysis of the Model's Dynamics	40

1.4.3	Model Sensitivity to Microstructural Variables	44
1.5	Final Remarks	46
1.6	Appendix A: A Family of Models for Stochastic Trade Arrival Times	75
1.7	Appendix B: Data re-sampling - Matlab Code	79
1.8	Appendix C: Estimation Results for the EACD and the EXACD Model	81
2	A TESTING PROCEDURE FOR CO-JUMPS	85
2.1	Introduction	85
2.2	A Co-Jump Testing Procedure	87
2.2.1	Jump and Co-Jump Test Indicator Functions	90
2.3	Monte Carlo Experiment	96
2.3.1	Simulation Design	96
2.3.2	Monte Carlo Findings	99
2.3.2.1	An Evaluation of the Univariate Tests for Jumps	100
2.3.2.2	Co-Jump Tests I: re-sampled data	106
2.3.2.3	Co-Jump Tests II: tick-by-tick data	111
2.4	Final Remarks	114
2.5	Appendix: Simulation Design - Matlab Code	132
3	THE LIQUIDITY-PRICE TRANSMISSION MECHANISM	136
3.1	Introduction	136
3.2	The Structure of the Data	138
3.2.1	Preparation of the Data	141
3.3	Estimation and Results	146
3.3.1	Re-Sampled Data Set	146
3.3.2	Tick-by-Tick Data Set	151
3.4	Final Remarks	157
	CONCLUSIONS AND FURTHER RESEARCH	189

CONTENTS

iii

BIBLIOGRAPHY

194

LIST OF FIGURES

1.1	EUR/USD mid-quote price behavior.	64
1.2	Available liquidity in the order book.	65
1.3	Returns in liquidity and trading activity Asian opening.	66
1.4	Returns in liquidity and trading activity London opening.	67
1.5	Market, new limit and cancelled limit orders - Example.	68
1.6	Histogram of time durations.	69
1.7	Liquidity and transactional durations.	70
1.8	Liquidity time and threshold indicators.	71
1.9	Impulse response functions for available liquidity.	72
1.10	Impulse response functions: comparison across regimes.	73
1.11	Impulse response functions: comparison across regimes.	74
2.1	Jump and Co-Jump Tests for re-sampled and tick-by-tick data.	128
2.2	Simulated Prices for underlying process $S_{(1)}$ with and without Jumps.	129
2.3	Simulated Prices for underlying process $S_{(1)}$ and $S_{(2)}$ with and without Jumps.	130
2.4	Quadratic Variation and Covariation of Two Stochastic Price Processes $S_{(1)}$ and $S_{(2)}$	131
3.1	Normalized EUR/USD FX Mid-Quote Price Action.	174
3.2	Normalized Cumulative Available Liquidity and Quotation Time.	175
3.3	Normalized Liquidity Variable $\rho_{(2)}$	176
3.4	Normalized EUR/USD FX Mid-Quote Price and Liquidity Variable $\rho_{(2)}$	177
3.5	Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 30 obs).	178

3.6	Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 60 obs).	179
3.7	Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 90 obs).	180
3.8	Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 120 obs).	182
3.9	Auto-correlation Function (ACF) plot for weekly data (tick data, 30 obs)	183
3.10	Auto-correlation Function (ACF) plot for weekly data (tick data, 60 obs).	184
3.11	Auto-correlation Function (ACF) plot for weekly data (tick data, 90 obs.).	185
3.12	Auto-correlation Function (ACF) plot for weekly data (tick data, 120 obs.).	186
3.13	Illustrative Example of Data re-sampling for Test Computation.	187
3.14	Illustrative Example of Data re-sampling for Test Computation.	188

LIST OF TABLES

1.1	Summary statistics.	49
1.2	Estimated coefficients for trade equation.	50
1.3	Estimated coefficients for liquidity equation.	51
1.4	Autocorrelations of time durations.	52
1.5	Model 1 - Estimated coefficients for liquidity and trade equation.	53
1.6	Model 2 - Estimated coefficients for liquidity and trade equation.	54
1.7	Model 3 - Estimated coefficients for liquidity and trade equation.	55
1.8	The significance of time in the liquidity and trade activity equations.	56
1.9	TACD model estimation.	57
1.10	TACD first regime - misspecification tests.	58
1.11	TACD second regime - misspecification tests.	59
1.12	Autocorrelations of standardized time durations in TACD model.	60
1.13	Autocorrelations of squared standardized time durations in TACD model.	61
1.14	Impulse response functions.	62
1.15	Model 4 - estimated coefficients for liquidity and trade equation.	63
1.16	EACD Model Estimations.	82
1.17	EXACD Model Estimations.	83
2.1	Formulae for Jumps and Co-jump test indicators	116
2.2	Parameter Values for the Monte Carlo Simulations	116
2.3	Size and power of the univariate tests for jumps for varying jump intensity	117

2.4	Power of the univariate tests for jumps for varying jump size	118
2.5	Size of the univariate tests for jumps for varying microstructural noise	119
2.6	Power of the univariate tests for jumps for varying microstructural noise	120
2.7	Size and power of the univariate tests for jumps for varying rounding noise	121
2.8	Size and power of the first battery of co-jump tests for varying jump intensity	122
2.9	Power of the first battery of co-jump tests for varying correlation	123
2.10	Size and power of the first battery of co-jump tests for varying microstructural noise	124
2.11	Size and power of the second battery of co-jump tests for varying jump intensity	125
2.12	Power of the second battery of co-jump tests for varying correlation	126
2.13	Size and power of the second battery of co-jump tests for varying microstructural noise	127
3.1	Auto-correlations (ACF) of weekly data	160
3.2	Auto-correlations (ACF) of weekly data	161
3.3	Auto-correlations (ACF) of weekly data	162
3.4	Auto-correlations (ACF) of weekly data	163
3.5	Auto-correlations (ACF) of weekly data	164
3.6	Auto-correlations (ACF) of weekly data	165
3.7	Percentage of Jump Events for EUR/USD FX and Liquidity Variable $\rho_{(1)}$	166
3.8	Percentage of Jump Events for EUR/USD FX and Liquidity Variable $\rho_{(2)}$	167
3.9	Percentage of Co-jump Events for EUR/USD FX and Liquidity Variable $\rho_{(1)}$	168
3.10	Percentage of Co-jump Events for EUR/USD FX and Liquidity Variable $\rho_{(2)}$	169
3.11	Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $\rho_{(1)}$	170
3.12	Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $\rho_{(1)}$	171
3.13	Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $\rho_{(2)}$	172
3.14	Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $\rho_{(2)}$	173

ABSTRACT

The thesis provides a novel contribution to the literature of microstructural theory and discovery models. The main contributions are twofolds. First, we move from price to liquidity discovery and explicitly study the dynamic behavior of a direct measure of liquidity observed from the foreign exchange markets. We extend the framework presented by Hasbrouck (1991) and Dufour and Engle (2000) by allowing the coefficients of both liquidity and trade activity to be time dependent. We find that liquidity time is characterized by a strong stochastic component and that liquidity shocks tend to have temporary effects when transactional time is low or equivalently when trading volatility is high.

We then analyze the contribution of liquidity to systemic risk and contagion and, in particular, assess the price impact of liquidity shocks. We extend the approach in Dumitru and Urga (2012) and present a co-jump testing procedure, robust to microstructural noise and spurious detection, and based on a number of combinations of univariate tests for jumps. The proposed test allows us to distinguish between transitory-permanent and endogenous-exogenous co-jumps and determine a causality effect between price and liquidity. In the empirical application, we find evidence of contemporaneous and permanent co-jumps but little signs of exogenous co-jumps between

the price and the available liquidity of EUR/USD FX spot during the week from May 3 to May 7, 2010.

ACKNOWLEDGMENTS

I express my gratitude to the following people, who provided constant source of inspiration throughout the journey of my studies:

To my mentors, Vincenzo and Amelia who have constantly shown me the way.

To my parents, Carlo and Lilliana who have always supported me in good and difficult times.

To Prof. Giovanni Urga for his trust, advice and commitment.

To Federica Ferretti for her immense patience and understanding.

To Amedeo Ferri Ricchi for the constant push and spiritual training.

To my colleagues at Deutsche Bank, and in particular to William Hutchings, Giuseppe Nuti, Roel Oomen, Giovanni Pillitteri, Neehal Shah, Panayiotis Stergiou for the support, the useful ideas and discussions.

INTRODUCTION

The analysis and the composition of market liquidity plays an important role in price discovery and microstructural theory. Kyle (1985) provides a very general definition of market liquidity in terms of tightness, resiliency and depth. Tightness is typically measured as the spread between the bid and the ask price, directly observed from the market. This particular measure of liquidity is used in the empirical work of Amihud and Mendelson (1980), Bessembinder (1994) and Chordia et al. (2000). Resiliency and depth are often measured in terms of price impact functions. In particular, following a shock or an unexpected trade, impulse response functions are used to determine the speed of convergence of prices towards their pre-shock equilibrium level. This is the approach followed by Hasbrouck (1991), Dufour and Engle (2000) and Large (2007). Banti et al. (2012) propose a measure of liquidity based on the notion of expected return reversal as in Pastor and Stambaugh (2003). Order flow data, defined as difference between the number of buy and sell market transactions, is used instead as an indirect measure of market liquidity and related to the dynamics of asset pricing in Evans and Lyons (2002), Berger et al. (2008), Evans (2010) and Chen et al. (2012).

The purpose of the dissertation is not to provide a theoretical model for order driven markets alongside the work produced by Rosu (2009), Goettler et al. (2005), Foucault

et al. (2005), Cont et al. (2010), nor to present a framework in the context of optimal order execution strategies as in Harris and Hasbrouck (1996), Almgren and Chriss (1999, 2000), Kissell and Malamut (2005), Frey and Sandas (2008). Our motivation is to model the behavior of a direct measure of liquidity, not inferred nor derived from prices or order flow data and use the particular structure and the high information content of the limit order book. There are some important differences between the approach used here and the methodology followed by previous studies to measure market liquidity. First, we employ a direct measure of liquidity defined as the sum of the limit orders at a given distance from the best displayed price in a limit order book. We refer to this particular measure of liquidity as available liquidity throughout Chapter 1 and 3. Available liquidity on the bid (ask) side represents the real amount of orders that can be executed on the buy (sell) side of a limit order book. Second, from the behavior of this direct measure of liquidity we construct a number of additional liquidity variables. In particular, we define market orders and new limit and cancelled orders at different price levels in the order book. We also introduce a measure of order book resiliency and of net order imbalance to use the information content of asymmetric structure of the order book. We also define and study the behavior of transactional and liquidity duration measured, respectively, as the time between consecutive market orders and changes in available liquidity. Liquidity time, in particular, provides an additional, indirect, measure of order book and market liquidity. Finally, we present an alternative and more informative measure of the order flow indicator, commonly used in microstructural theory as an indirect measure of liquidity. In our model, we construct the order flow indicator as the sum of market and new limit orders. We refer to this variable as trading

activity throughout the dissertation. Trading activity on the buy (sell) side of the order book is measured as the sum of sell (buy) market orders and new incoming buy (sell) limit orders and used to capture important features of the behavior of market liquidity.

We then analyze the contribution of liquidity to systemic risk and contagion and, in particular, assess the price impact of liquidity shocks. Jumps and co-jumps are typically studied in relation to price dynamics and most of the literature on jumps focuses on the discontinuous and extremely large price returns caused by a trading or a macroeconomic shock. We instead propose a co-jump testing procedure used to distinguish between transitory-permanent and endogenous-exogenous jumps between liquidity and price. Finally, we use a data-set, collected at a very high frequency, motivated by the interest to understand the influence of high speed trading on liquidity and prices. The results presented are particularly relevant also in the context of the recent debate on high frequency trading and the need to impose a set of operational rules in the interest of financial stability.

The dissertation is divided in three chapters. In Chapter 1, we propose a comprehensive liquidity discovery model which is used to analyze the behavior of a direct measure of liquidity and the interaction of a number of microstructural variables. In particular, we extend the vector autoregressive representation of Hasbrouck (1991) and Dufour and Engle (2000) and establish an explicit relationship between changes in the available liquidity and a measure of trading activity observed from a representative EUR/USD FX spot limit order book during the week from May 3 to May 7, 2010. The dynamic behavior of liquidity is estimated by allowing time to have a deterministic and a stochastic component. Time plays a very important role and, in our study, we

distinguish between two very specific time measures. We first identify transactional time as the duration between consecutive market orders, and then define liquidity time as the interval between changes in the liquidity variable. The distinction is particularly relevant and enables to better understand the role of time in a high frequency trading framework. For this purpose, we allow the autoregressive conditional duration model to switch across different trading volatility regimes. The use of a regime switching model allows us to capture changes in the liquidity behavior and in the relationship with other microstructural variables due to time-varying trading intensities. Moreover, we use impulse response functions to measure the speed at which liquidity moves to the equilibrium level following an exogenous shock.

In Chapter 2, we introduce a co-jump testing procedure based on the combination of univariate tests for jumps as an extension of Dumitru and Urga (2012). The combination of tests allows us to address some of the issues with the existing tests for co-jumps and extend the notion of a co-jump event. In particular, the proposed tests are shown to be robust to microstructural noise and can be adjusted to take into account for non-synchronous trading. Also, in the common notion of a co-jump used in the literature two or more variables are characterized by a simultaneous and discontinuous path over a given time interval. The jumps are traditionally both exogenous and no causality between the two can be inferred. We identify instead a causality effect between different jumps observed over a fixed time horizon and distinguish between contemporaneous, permanent and lagged or exogenous co-jump events. Finally, we assess the behavior of the proposed co-jump tests under different levels of jump intensity, volatility, correlation and microstructural noise.

In Chapter 3, we present an empirical application where we explicitly assess the transmission mechanism between the price of EUR/USD FX spot and the associated liquidity displayed in a representative limit order book. Our estimates are based on observations measured at a very high frequency. We use the co-jump testing procedure, based on the combination of univariate tests for jumps presented in Chapter 2, to measure the impact and assess the effects of liquidity jumps on prices. We also use two distinct liquidity measures to take into account the information content of transactional time. Finally, we determine the type of co-jump event observed and establish an explicit causality effect between liquidity and the EUR/USD FX spot price during the week from May 3 to May 7, 2010.

CHAPTER 1

MOVING FROM PRICE TO LIQUIDITY DISCOVERY

1.1 Introduction

Following the abrupt events experienced by financial markets in recent years, the behavior of liquidity has started to play an important role in microstructural theory. Liquidity shocks and the associated systemic risks have led to an increasing interest from regulatory bodies, policymakers and market participants who are in the process of defining new rules and risk management practices with a specific focus on liquidity (Goodhart 2007, Farhi et al. 2007, Rochet 2008, Espinoza et al. 2008, Cao and Illing 2010, Perotti and Suarez 2011, Hafeez 2011). The introduction of precise liquidity requirements for financial institutions has also been prompted by the need to properly assess the risks related to the presence of liquidity externalities. Moreover, trading agents are defining execution rules to better manage and offer liquidity at an increasing high frequency and reduce the issues related to latency, defined as the time between the submission and the consequent execution of an order (Mahanti et al. 2008, Riordan and Storckenmaier 2011). Neoclassical economists have also shown that liquidity, order flows and trade activity have a deep influence on price discovery. Liquidity in particular is described

as the result of the interaction of a number of agents who reveal and exchange private information respectively through prices and trading activity.

The seminal contribution of Hasbrouck (1991) has generated a huge body of literature dealing with price discovery models using transactional data. A number of extensions and empirical applications have followed. Kempf and Korn (1998) measure the price impact of unexpected financial transactions using data on the DAX futures and highlight the informational content of trades of different size and type. Dufour and Engle (2000) use the autoregressive conditional duration (ACD) model of Engle and Russell (1998) to accommodate for stochastic transactional time and explain the role of time in the price formation process of a number of stocks quoted on the NYSE. Evans and Lyons (2002) introduce a microstructural model, based on a vector autoregressive representation, to study the relationship between order flows and exchange rate dynamics and to show the superior forecasting ability of microstructural variables as opposed to the more traditional fundamental models based on macro variables. The importance of order flows and trading volumes in price discovery models is also highlighted in Perraudin and Vitale (1996), Breedon and Vitale (2004), Moberg and Sucarrat (2007), and Berger et al. (2008). Payne (2003) studies the informational content of trades executed in foreign exchange markets and measures the relationship between trading activity, performed at different times of the day, and price changes. A similar approach is used by Floegel (2006) who identifies various levels of informational asymmetry among trades performed by different types of dealers.

A common feature to all these studies on price discovery is that liquidity is indirectly measured through the use of price impulse functions or through the notion of price

spread. The concept of a liquid market is either related to the speed at which prices converge to an equilibrium level following an informational shock in the system or to the absolute distance of bid and offer prices, with a lower (higher) spread being associated to higher (lower) levels of liquidity. Danielsson and Payne (2010) provide an important contribution to the analysis of the behavior of liquidity in foreign exchange markets. In their study, liquidity is measured both indirectly and directly through spreads and depth. However, most of the leading literature on discovery models has not focused specifically on liquidity dynamics and no direct measure had been proposed despite the recognition of the importance of liquidity in microstructural theory (Copeland and Galai 1983, Kyle 1985, Hasbrouck 1993, Burghardt et al. 2006, Hafeez 2007).

The main objective of Chapter 1 is to propose a comprehensive liquidity discovery model which allows us to analyze the behavior of a direct measure of liquidity and the interaction of a number of microstructural variables. Our contribution to the literature is three-fold. First, we move from price to pure liquidity discovery and study the dynamics of available liquidity. This is particularly relevant given the little reference that the literature on price discovery gives to a direct measure of liquidity. We define liquidity in terms of depth, as in Kyle (1985) and Danielsson and Payne (2010), using an high observation frequency. We measure liquidity as the cumulative units of a currency that a dealer can purchase or sell, in a limit order book, at a given distance from the best price. We refer to available liquidity throughout the chapter. Second, we extend the vector autoregressive representation of Hasbrouck (1991) and Dufour and Engle (2000) by allowing all the coefficients of the vector autoregressive model to be time dependent and establish an explicit relationship between changes in liquidity and

market transactions. The dynamic behavior of liquidity is estimated by allowing time to have a deterministic and a stochastic component. We also distinguish between two very specific time measures. We first identify transactional time as the duration between consecutive market orders, and then define liquidity time as the interval between changes in the liquidity variable. The distinction is particularly relevant and enables to better understand the role of time in a high frequency trading framework. For this purpose, we allow the ACD model to switch across different trading volatility regimes. The use of a regime switching model allows us to capture changes in the liquidity behavior and in the relationship with other microstructural variables due to time-varying trading intensities. Moreover, we use impulse response functions to measure the speed at which liquidity moves to the equilibrium level following an exogenous shock. The framework provides an important contribution to the analysis of high frequency trading. Third, we use a data-set which contains transactional information from the main reference cash market in foreign exchange, observed and re-sampled at a very high frequency. Our empirical study is based on one week of data observed from a representative EUR/USD FX spot limit order book and recorded at a millisecond level. Our dataset contains time stamped, tick-by-tick data on available liquidity, prices and quotation times from May 3 to May 7, 2010, subsequently re-sampled over a frequency of 10 seconds. Limit, market orders and a number of transactional variables and liquidity measures have also been constructed from the data. The analysis of the series at such high frequency enables us also to measure the impact of high speed trading on liquidity.

Our main results are as follows. In our analysis, we find strong predictability and feedback effects between available liquidity, changes in trading activity and time dura-

tions. As in Easley and O'Hara (1992) and in Dufour and Engle (2000), trade durations have a significant informational content and are characterized by a strong stochastic component. However, we also find that liquidity durations are statistically significant over a number of lags and that the impact of latency on the liquidity behavior varies across different trading regimes. The use of two distinct measures of stochastic times and the assumption of two trading regimes, that our model allows to identify, provides also additional insights on how liquidity is affected by a non-homogeneous trade intensity. In particular, we first identify an inventory effect of time, where, in the absence of external shocks, the available liquidity at k ticks from the best price tends to dry up with time and with an increase in market order activity. A greater time at which liquidity moves from one state to another is typically associated to a lower trading activity and to a smaller liquidity decay at a given price level. In our empirical application, we relate our measure of liquidity to the concept of inventory as described in Avellaneda and Stoikov (2008). In particular, we find that the available liquidity, measured as the sum of limit orders, can provide a direct measure of the inventory of a trading agent. We also find that a greater time between changes in available liquidity, which translates into a more static structure of the order book, is typically associated to a larger inventory or opportunity cost faced by a trading agent when a limit order is selected as opposed to a market order. This result highlights the issue of optimal order execution, and, in particular, the trade-off between the opportunity cost of a limit order and the transactional cost of a market order as also discussed in Harris and Hasbrouck (1996), Almgren and Chriss (1999, 2000), Kissell and Malamut (2005), Frey and Sandas (2008), Cont et al. (2010) and Kratz and Schoneborn (2011). The relationship between

liquidity, measured in terms of bid-ask spread, and trading inventory is also discussed in Bessembinder et al. (1996).

Moreover, we find that trade durations are negatively correlated with available liquidity and incoming limit orders. During times of intense trading activity the level of available liquidity at a given price is in fact negatively affected. These results are due to the fact that liquidity, offered by a dealer at a given price or distance from the best available price, is not infinite and decreases as market orders, on the opposite side of the book, are matched in the absence of fresh liquidity introduced in the order book. Second, we identify an informational effect of time, where, following an external shock, the rate of adjustment of liquidity is greater during times of high volatility or, equivalently, low transactional and liquidity durations. We also find evidence of a negative relationship between the information content of trades and both trade and liquidity durations. In particular, the initial and the final hours of the day tend to be characterized by lower volatility and market activity and hence by a worse information content, the duration between consecutive transactions tends to be higher, the liquidity impact of a trade more persistent and the liquidity decaying process weaker. This finding is consistent with the empirical result of Easley and O'Hara (1992), Lyons (1995), Dufour and Engle (2000), Payne (2003) who show that prices tend to adjust faster and that the informational content of trades is greater during periods of higher trading intensity. Moreover, during times characterized by a higher trading intensity, market making agents tend to tighten price spreads, defined as the difference between the ask and the bid price, and increase the amount of displayed liquidity. This result shows that high speed trading allows liquidity to adjust quicker to the equilibrium level following an

exogenous shock and that the level of information content provided by trading activity increases with a decrease in transactional time. In particular, high speed trading seems to facilitate the flow of private information in the market and to reduce the negative effects of unexpected trading activity. Further, there is evidence of temporal clustering and of a decaying pattern in the liquidity process and a strong relationship between trading intensity, volatility and liquidity. Liquidity is drained by matching market orders, on the opposite side of the book, with the drained liquidity not being immediately supplied at the same price level. We find no correlation between available liquidity or trading activity on different sides of the book and this highlights a different trading behavior and motivation among trading agents on the bid and ask side of the market. Finally, both trade and liquidity durations show a M-shaped pattern during the trading day similar to the realized volatility of the underlying asset price. The same pattern is highlighted in Engle and Russell (1998) and in Dufour and Engle (2000).

The chapter is organized as follows. In Section 1.2, we first introduce the price discovery model as in Hasbrouck (1991) and Dufour and Engle (2000), and then we present an alternative vector autoregressive model specification enriched with regime switching stochastic time. In Section 1.3, we describe the data and the main features of the trading venue and of order book structure. In Section 1.4, we present the estimation results and study the impact of both liquidity and transactional durations on liquidity and trade activity. Some final remarks and proposals for further research are discussed in Section 1.5.

1.2 The Model

In standard price discovery models, *vector autoregressive* (VAR) representations are used to capture the joint behavior of price changes and trades. Mid-quote prices, q_t , are measured as the average of the bid and the offer price at time t and just after a trade x_t . Given the non-stationary nature of the process, price changes, Δq_t , are considered in the dynamic model and studied jointly with the behavior of trades. Trading activity, x_t , is typically measured by the sign of the underlying transaction and expressed by a dummy variable, which takes a value equal to one if the initiated trade is a buy order or minus one if the initiated trade is a sell order instead.

1.2.1 The Price Discovery Model proposed by Hasbrouck (1991) and Dufour and Engle (2000)

Hasbrouck (1991) and Dufour and Engle (2000) propose the following VAR representation:

$$\begin{aligned}\Delta q_t &= \sum_{i=1}^{\infty} a_i \Delta q_{t-i} + \sum_{i=0}^{\infty} b_i x_{t-i} + v_{1,t} \\ x_t &= \sum_{i=1}^{\infty} c_i \Delta q_{t-i} + \sum_{i=1}^{\infty} d_i x_{t-i} + v_{2,t}\end{aligned}\tag{1.1}$$

to model the relationship between mid-quote price changes, Δq_t , and a trade indicator variable, x_t , which takes a value equal to one if a buy-trade is initiated or minus one otherwise. $v_{1,t}$ and $v_{2,t}$ are i.i.d processes measuring shocks, determined by public and private information, to the price and the trade equation respectively. Unlike in the original model set up described by Hasbrouck (1991), Dufour and Engle (2000) allow the trade coefficient to vary with time in both the price and the trade equations. Time

is considered, in this context, a variable with both deterministic and stochastic components. The coefficients of the trade variable in the price equation are time dependent and given by:

$$b_i = \gamma_i^q + \sum_{j=1}^J \lambda_{j,i}^q D_{j,t-i} + \delta_i^q \ln(T_{t-i}) \quad (1.2)$$

and, similarly, the coefficients of the trade equations are given by:

$$d_i = \gamma_i^x + \sum_{j=1}^J \lambda_{j,i}^x D_{j,t-i} + \delta_i^x \ln(T_{t-i}) \quad (1.3)$$

where $D_{j,t-i}$ represents a time-of-day dummy variable while γ , λ and δ the estimation coefficients. In particular, γ represents a constant component, λ and δ capture the impact of deterministic and stochastic components of the time process respectively. The duration between consecutive transactions, measured by T , enters non-linearly in the VAR representation. In a more general set-up, the equation for the mid-quote price change can be expressed as:

$$\Delta q_t = \sum_{i=1}^N \alpha_i \Delta q_{t-i} + \sum_{i=0}^N [\gamma_i^q + \mathbf{z}_{t-i} \delta_i^q] x_{t-i} + v_{1,t} \quad (1.4)$$

where \mathbf{z}_{t-i} represents a row vector of observations for different transactional variables and where δ_i provides the column vector of corresponding coefficients. The significance of time in the dynamic system (1.4) is verified by testing the statistical significance of the coefficients δ_i . In particular, intraday periodicities can be found when the δ_i^q are jointly zero and at least one of the λ_i^q is not zero at the same time.

Given the stochastic nature of the time process, ACD models are used to model the dynamic behavior of diurnally adjusted trade durations and to compute impulse response functions. The conditional duration mean is modelled using an ARMA-type specification:

$$\phi_t = \omega + \sum_{j=1}^p p_j \tilde{T}_{t-j} + \sum_{i=1}^q \zeta_i \phi_{t-i} + \lambda D_{t-1} \quad (1.5)$$

where \tilde{T} represents the diurnally adjusted duration between consecutive transactions, $\tilde{T} = T_t / \Phi_t(t-1)$, with $\Phi_t(t-1)$ being the deterministic component of time and D_{t-1} diurnal dummy variables which measure the time-of-the day effects while p , ζ and λ represent the estimation coefficients. The conditional density of \tilde{T} is given by $g(\tilde{T}) = \theta / \phi_t^\theta \tilde{T}^{\theta-1} \exp \left[- \left(\tilde{T} / \phi_t \right)^\theta \right]$ for $\theta, \phi_t > 0$.

In Hasbrouck (1991), conditional impulse response functions are represented as:

$$I_q = \Delta q_{t+k}^x = \sum_{i=0}^k \beta_i v_{2,t} \quad (1.6)$$

Dufour and Engle (2000) propose an extension assuming a joint distribution of marks and arrival times conditional on their past values and represented as follows:

$$I_q(k, v_t, T_t, \omega_{t-1}) = E[q_{t+k} | v_{2,t} = 1, T_t = \tau, \omega_{t-1}] = E_T[E(q, x | T)] \quad (1.7)$$

where ω_{t-1} indicates the information content of all the variables in the model up to time $t-1$.

The framework presented by Dufour and Engle (2000) allows to measure the impact of transactional time and also to capture diurnal *versus* stochastic components in the

time process. Not surprisingly, during times of high trading activity, prices converge to the full information value, after a shock, at a faster pace. Also, standardized trade durations are found to be non strongly exogenous and mainly dependent on past volume and past standardized durations. Authors finally show that the effect of time durations on price changes tends to be marginal if compared to the same effect produced by volume dynamics and spread when these are introduced as additional independent variables in a more general model specification.

In the traditional framework of price discovery models, liquidity is measured indirectly through the use of price impact functions, price spreads or as a function of some microstructural or trading variables. The concept of a liquid market is often related to the speed at which prices convergence to an equilibrium level following an informational shock in the system. In this chapter, however, we directly measure and construct a number of liquidity variables by looking at the available liquidity displayed on one side of a limit order book from the foreign exchange markets. Moreover, in our model, the available liquidity is characterized by a deterministic and a stochastic time component, and analyzed under different trading regimes.

1.2.2 Moving from Price to Liquidity Discovery

Let Δq_t^k express the logarithmic changes in the cumulative liquidity displayed at k ticks from the best price in a limit order book. The measure of liquidity is very similar to Kyle's (1985) definition of market depth and is computed as in Danielsson and Payne (2010).

In our model, we introduce and model different types of durations. First a trans-

actional duration which measures the time interval between consecutive market orders and a liquidity duration which measures the time interval at which changes in available liquidity occur. This important distinction enables us to isolate the role played by the time at which liquidity moves from one state to another.

In addition to available liquidity, we also study the behavior of a transactional variable, ω_t , which measures the overall trading activity on one side of order book and defined as the sum of market and new incoming limit orders. More specifically, trading activity on the bid (ask) side of the market is recorded when a buy (sell) limit order is matched by a sell (buy) market order, under the assumption that market orders can only be executed against the best limit orders available at a certain price (i.e. market orders cannot “walk the book”) and also when a new limit order (i.e. new additional available liquidity) is added to existing orders at the same price priority in the book. This particular measure of trading activity is more informative than the trade indicator variables used in previous studies on price discovery.

We propose the following VAR representation to model the joint behavior of changes in the available liquidity observed on one side of the order book and the proposed measure of trading activity:

$$\begin{aligned}\Delta q_t^k &= \sum_{i=1}^n [\gamma_{q,i}^q + \mathbf{z}_{t-i} \delta_{q,i}^q] \Delta q_{t-i}^k + \sum_{i=1}^n [\gamma_{\omega,i}^q + \mathbf{h}_{t-i} \delta_{\omega,i}^q] + v_{1,t} \\ \omega_t &= \sum_{i=1}^n [\gamma_{q,i}^\omega + \mathbf{z}_{t-i} \delta_{q,i}^\omega] \Delta q_{t-i}^k + \sum_{i=1}^n [\gamma_{\omega,i}^\omega + \mathbf{h}_{t-i} \delta_{\omega,i}^\omega] + v_{2,t}\end{aligned}\tag{1.8}$$

In the dynamic system described in (1.8), \mathbf{z}_{t-i} and \mathbf{h}_{t-i} are row vectors of observations and δ_i represents a column vector with the correspondent coefficients. Model (1.8) extends the original model (1.1) by allowing both the liquidity and the trade coefficients

to vary with time, where time is again characterized by a deterministic and a stochastic component. In the case where the variables in \mathbf{z}_{t-i} and \mathbf{h}_{t-i} are current and past time durations, with $D_{j,t-i}$ being a set of n time-of-day dummy variables, we can re-write the liquidity impact coefficients as:

$$\begin{aligned} a_i &= [\gamma_{q,i}^q + \delta_{q,i}^q \ln(T_{t-i}^q)] \\ c_i &= [\gamma_{q,i}^\omega + \delta_{q,i}^\omega \ln(T_{t-i}^q)] \end{aligned} \quad (1.9)$$

where T_{t-i}^q measures the observed time for a change in available liquidity (*liquidity duration*). Analogously, we can express the trade impact coefficients as:

$$\begin{aligned} b_i &= \left[\gamma_{\omega,i}^q + \sum_{j=1}^n \lambda_{j,\omega,i}^q D_{j,t-i} + \delta_{\omega,i}^q \ln(T_{t-i}^\omega) \right] \\ d_i &= \left[\gamma_{\omega,i}^\omega + \sum_{j=1}^n \lambda_{j,\omega,i}^\omega D_{j,t-i} + \delta_{\omega,i}^\omega \ln(T_{t-i}^\omega) \right] \end{aligned} \quad (1.10)$$

where T_{t-i}^ω measures the time between consecutive market orders on the opposite side of the order book (*transactional duration*).

Equations (1.9) and (1.10) nest a special case of the model presented by Dufour and Engle (2000) and referenced in (1.1)-(1.3). Model (1.8) is also flexible to accommodate other transactional variables in the \mathbf{z}_{t-i} and \mathbf{h}_{t-i} vectors like price spreads or net order imbalance for example. The framework presented above is also quite convenient from an estimation point of view. The liquidity and the trade equations are in fact modeled as a conditional process of both liquidity and transactional durations which allows to estimate the VAR system directly.

In order to capture state dependent trade intensities, i.e. regime shifts, in the series, the ARMA-type conditional expected durations are modeled using a regime

switching threshold ACD (TACD) model introduced by Russell et al. (2001). The model is represented as:

$$\begin{aligned} x_i &= \psi_i \varepsilon_i^k \\ \psi_i &= \omega^k + \sum_{j=1}^m \alpha_j^k x_{i-j} + \sum_{j=1}^q \beta_j^k \psi_{i-j} \end{aligned} \tag{1.11}$$

with ε_i^k being an i.i.d. vector with positive and regime specific intensities. The vector \mathbf{k} denotes the number of regimes with $0 = r_0 < r_1 < r_2 < \dots < r_k = \infty$ being the threshold values. The model allows us to deal with non-linearity and to accept a number of different trading volatility regimes. In our empirical application, we identify two regimes. A regime characterized by low trading intensity (or equivalently by high time durations) is denoted with $R_{k=1}$ while a regime characterized instead by a high trading intensity (or equivalently by low time durations) is denoted with $R_{k=2}$. The threshold indicators are computed as a function of the rolling mean and standard deviations of liquidity time. In particular, given the vector of liquidity time $\mathbf{x} = \{x_1, \dots, x_n\}$, the model moves from $R_{k=1}$ to $R_{k=2}$ when $x_i < \{x_m + \xi \sigma_m\}$ for $i = 1 \dots n$, where $x_m = (1/m) \sum_{i=1}^m (x_i)$ and $\sigma_m = \sqrt{1/(m-1) \sum_{i=1}^m (x_i - x_m)^2}$ with $m = [1, \dots, n]$. The choice of the number of threshold levels, the size of the rolling sample m and the form of the threshold function is obtained by fitting the threshold function to the empirical data. In particular, we find that $\xi = 6$ provides a good fit to the data set used in the empirical application when two trading regimes are considered.

The use of a TACD model in our empirical analysis is motivated by the observed behavior of liquidity and transactional durations and, in particular, by the M-shaped pattern displayed in proximity of the opening and closing hours of a trading session. The

regime switching model allows us to isolate the diurnal effects caused by the opening and closing hours of the trading day and highlight the different behavior observed during the central hours of the day which are characterized instead by high transactional volatility. De Luca and Zuccolotto (2006) also identify two trading regimes across which the shape of the underlying distribution can vary while Iordanis and Maher (2011) use a threshold switching model to allow the shape parameter of different distributions to vary across regimes characterized by non homogeneous trading intensities. In Appendix A, we report a comprehensive family of ACD models commonly used to estimate stochastic arrival times.

In the context of discovery models, it is particularly relevant to use *response functions* in order to measure the permanent or transitory impact on available liquidity of exogenous shocks generated by unexpected trading activity. The results can be very important to distinguish between informative and non-informative trade activity. We identify the informational shocks using two white noise processes, $v_{1,t}$ and $v_{2,t}$. In particular, $v_{1,t}$ is associated to the liquidity shocks caused by publicly available information, while $v_{2,t}$ measures the impact of private information mainly driven by unexpected trades. In order to compute impulse response functions we can follow two different methodologies. The first approach follows from Dufour and Engle (2000) in the case where $\sum_{i=1}^n \delta_{\omega,i} \neq 0$. Alternatively, in the case where $\sum_{i=1}^n \delta_{\omega,i} = 0$, we can directly compute the impulse functions from the VAR model as in Hasbrouck (1991). We use a modified representation of the impulse response function which is dependent on both liquidity and transactional durations:

$$I_{\Delta q}(\eta, v, T^q, T^\omega, I_{t-1}) = E(\Delta q_{t+k}^k | v_{2,t} = 1, T^q = \tau^q, T^\omega = \tau^\omega, I_{t-1}) \quad (1.12)$$

where T^q and T^ω measure the liquidity and the transactional time respectively.

In the empirical application presented in this chapter, we also assume that time durations are non strongly exogenous and that the arrival of trade shock hits the equilibrium level for liquidity and trading activity. In particular, we start from a steady-state equilibrium level where $\Delta\omega_t = 0$ and $\Delta q_t^0 = 0$ with $q_t^0 = q^*$ at $t = 0$. We then introduce a shock in the trade activity equation equal to $v_{2,t} = 1$ and measure the impact on the liquidity process through the impulse response function as in (1.12). Given that we model changes in liquidity, conditional on both transactional and liquidity durations, we need to compute the joint density of liquidity and durations in order to determine the output of the impulse response function. The joint density is, in this case, given by the product of the conditional density of liquidity times the marginal density of transactional durations which are modelled using the TACD specification as in (1.11). Finally, we use Monte Carlo simulations to estimate the impulse response function.

1.3 The Data

The data used in the empirical application consist of five time series: the available liquidity on the bid and the ask side of a representative EUR/USD FX spot limit order book, expressed in EUR millions; the bid and ask price associated to the available liquidity in the order book, expressed as units of US Dollars for 1 EUR; quotation time,

expressed in milliseconds. The data spans over an entire trading week, from May 3 to May 7, 2010. The time frame considered is particularly interesting given that, during this time horizon, EUR/USD FX spot displayed a large move, opening on Monday at around 1.3250 and closing the session on Friday just above 1.2750. Despite the short time interval, in our empirical application, we use data observed at a very high frequency to obtain a large sample.

The behavior of the mid-quote price for EUR/USD spot is shown in Figure 1.1. The vertical dotted line shows the end of the NY trading session and the open of the Australian trading session. Figures 1.2 shows the available liquidity observed, on the bid side of the order book, at $k = 0, 1$ and 2 ticks from the best displayed price in the book during the week from May 3 to May 7, 2010. The greatest amount of liquidity is observed during the central hours of the trading day, while liquidity tends to dry up in proximity to the opening and the closing of the trading session. Figures 1.3-1.4 show the returns of liquidity, market and new limit orders during the first hours of the trading session and the London opening respectively on May 3, 2010. The combination of market and new limit orders define the trading activity in the order book. We notice that the level trading activity becomes more intense and regular during the London opening hours compared to the initial hours of the Asian trading session.

[Insert Figures 1.1 - 1.4]

Summary statistics for the variables in the data are presented in Table 1.1. An average of 245,000 orders were actively traded on a daily basis during the reference period considered in the analysis, with almost twice as many limit than market orders. Approximately 17% of the orders have been inputted as new limit orders close to the

best price, around 15% of the total overall limit orders have been cancelled and almost one fourth of the total limit orders were filled completely. The size of both new limit and cancelled orders shows the tendency to decrease away from the best price. Limit and market orders appear also to be more frequent on the bid side of the market as opposed to new limit and cancelled orders.

[Insert Table 1.1]

In what follows, we present details of the data and how these have been adjusted and sampled, a description of the variables, and finally a preliminary empirical analysis based on descriptive statistics of the series.

1.3.1 Preparation of the Data

The data is initially recorded to the one thousandth of a second over irregular time intervals and subsequently resampled over equally spaced intervals of 10 seconds. The conversion from tick to calendar time is done by selecting a fixed time grid $\{t = 1, \dots, T\}$ of equally spaced time intervals and using the most recent mark at the end of the selected interval as the observation for that particular interval.

The decision to re-sample the data over fixed size time intervals is motivated by two main arguments. First, as discussed in Engle and Russell (2004), issues of extreme discreteness and micro-structural noise, typical of high frequency data, can lead to spurious estimations, impact the volatility of the series and generate a high degree of kurtosis in the data. Converting the data from an irregular space grid to regularly spaced intervals, combined with pre-averaging, can also reduce the autocorrelation in the series returns and limit the effects of temporal dependence. Second, re-sampling

over fixed intervals of 10 seconds can substantially reduce the computational efforts of the modelling framework used to produce the estimation results. Similar empirical results would have also been obtained using a re-sampling frequency of 5 seconds.

Given the virtual continuity of the currency markets, as opposed to equity and fixed income markets, we do not discard the opening and the overnight transactions, despite the lower liquidity and the consequent higher volatility in the series, but we do eliminate the reporting errors observed during the very last trading day in proximity of the NY closing. In order to take into account the so called diurnal effects in the series, we adjust both the transactional and the liquidity durations following the same procedure as in Engle and Russell (1998) and in Dufour and Engle (2000). First, we fit the time series of the unadjusted durations using a polynomial equation, we then compute the ratio between actual and fitted trade durations. To improve the fitting accuracy we transform the predictors, in this case represented by the unadjusted trade durations, by normalizing their center and scale. The predictors have a zero mean and standard deviation equal to one.

1.3.2 Variables

From the original data set we derive a number of microstructural variables. We define Δq_t^k as the change in the natural logarithm of cumulative available liquidity at $k = 2$ ticks from the best displayed price on the bid side of the order book. In the original dataset, the available liquidity in the order book is recorded at three different price levels: the available liquidity on the bid (ask) side at the best displayed bid (ask) price in the order book which represents the total amount of buy (sell) limit orders with the

highest priority in terms of execution, followed by the available liquidity on the bid (ask) side of the order book, observed at different price levels, and, in particular, at $k = 1$ and $k = 2$ ticks away from the best displayed price where the tick is measured as 0.0001 US Dollars per 1 EUR. Cumulative available liquidity is a direct measure of the displayed depth in the order book. In the empirical application, presented in Chapter 1, we take into account only the liquidity observed on the bid side of the market.

Unlike in previous studies, two measures of time durations are introduced here: a transactional duration, T^ω , which indicates the time interval between two consecutive market orders and a liquidity duration, T^q , which indicates instead the time at which available liquidity changes from previous levels. Both time durations are expressed in milliseconds.

The trading activity, ω_t , on the bid (ask) side of the order book is measured in terms of market and new limit orders. More specifically trading activity on the bid (ask) side of the order book is recorded when a buy (sell) limit order is matched by a sell (buy) market order, under the assumption that market orders can only be executed against the best limit orders available at a certain price (i.e. market orders cannot “walk the book”) and also when a new limit order (i.e. new additional available liquidity) is added in the order book. Market orders, m_t , are measured in EUR millions and represent the amount of liquidity drained from the market either on bid (market orders sell) or on the offer (market orders buy).

Cancelled and new limit orders, here indicated as $clim_t$ and $nlim_t$, are again expressed in EUR millions and represent respectively the amount of liquidity firstly displayed and subsequently withdrawn from the order book (i.e. not filled) or fresh new

liquidity inputed in the order book at k -ticks from the best available price.

Figure 1.5 shows the methodology used to derive market, cancelled and new limit orders from the original time series. In particular, the *first* box shows that a sell market order is identified either when the best displayed bid price in the order book decreases and computed as the difference between the available liquidity, on the bid side of the order book, over the time interval $[t - 1, t]$ or when the best displayed bid price remains constant but available liquidity drops. In particular, the best bid price in the book drops from 1.3310 to 1.3309 between 4th and 5th observation. Available liquidity drops from 22 to 7 over the same time interval. We identify a sell market order and we quantify the order in 22 units. The *second* box shows that a new limit order is identified when the best displayed bid price remains constant over the time interval $[t - 1, t]$ but available liquidity increases over the same time interval. In particular, the best bid price in the book remains constant between the 7th and the 8th observation but the available liquidity at $k = 0$ ticks from the best price increase by 1 unit. We identify a new limit order and quantify the order in 1 unit. Finally, the *third box* shows that a cancelled limit order is identified at $k = 1$ or $k = 2$ ticks from the best displayed price in the order book in the event where the price does not change over the time interval $[t - 1, t]$ but available liquidity drops. In particular, the bid price at $k = 2$ ticks from the best displayed price in the order book remains constant between the 15th and the 16th observation at 1.3306 but the available liquidity drops from 9 to 8 units. We identify a cancelled limit order and quantify the order in 1 unit. Sell market orders and cancelled orders will have a negative effect while new buy limit orders a positive effect on the available liquidity on the bid side of the order book.

[Insert Figure 1.5]

The spread, represented by the variable s_t , is measured as the difference between the best ask and the bid price and expressed in pips of the domestic currency (USD).

We also introduce a very important transactional variable, often used in price discovery models and represented by the net order imbalance, denoted with noi_t , which is here measured as the excess amount of available liquidity displayed on the bid over the ask side of the order book. This variable is very similar to the net order imbalance in Cao et al. (2009). We can interpret the net order imbalance as a ex-ante measure of the trade indicator variable used by Hasbrouck (1991) and Evans and Lyons (2002) in their models. Both the net order imbalance and the trade indicator variables express the willingness of market participants to enter the buy or the sell side of the market. The net order imbalance can be used, in this context, to forecast trading activity with a higher number of buy (sell)-side trades being expected following an increase (decrease) in the net order book imbalance.

Finally, we introduce a variable that measures the market impact of trading activity, denoted as moi_t , that can be used as a proxy of market resilience. We compute this particular measure of liquidity as the ratio of relative mid-price variation over the level of previous trading activity defined as the sum of market and new limit orders at $k = 0$ ticks from the best displayed price in the order book.

In Appendix B, we report part of the Matlab code used to re-sample the data over fixed time intervals and to construct the micro-structural variables presented in the chapter.

1.3.3 Preliminary Analysis

In Tables 1.2 - 1.3, we report some preliminary analysis of the transactional variables introduced above. In line with previous studies, we find that both changes in trading activity and liquidity show a strong autocorrelation. In particular, the autocorrelation tends to increase at a lower distance from the best mid-point price.

[Insert Tables 1.2 - 1.3]

We also note negative autocorrelation in both the trade and liquidity equations. No signs of cross-correlations can be found when looking at the different sides of the order book, with changes in trading activity on the bid side being completely unrelated to changes in trading activity on the offer side of the order book. Not surprisingly changes in the liquidity displayed at the best price in the order book tend to have a greater impact on the overall liquidity dynamics. Strong autocorrelation is again detected in the series of the raw and adjusted time durations as shown in Table 1.4.

[Insert Tables 1.4]

The autocorrelations and the partial autocorrelations are far from zero and all the signs are positive. The Ljung-Box statistic is examined to test the null hypothesis that the first 15 autocorrelations are jointly equal to zero. We perform the test on the raw and the adjusted time series. The null is rejected in both cases with a chi-squared statistic of 126.70 (125.00) and 76.84 (58.54) for both raw (adjusted) liquidity and transactional durations respectively. Figure 1.6 reports a plot of the histogram and the distribution for the liquidity and the transactional durations respectively. Trade

durations also display a M-shaped pattern typical of many financial series. The initial and the final hours of the day tend to be characterized by lower volatility and trading activity and hence by a worse liquidity. The duration between consecutive transactions tends to be higher and the overall liquidity turnover lower as shown in Figure 1.7.

[Insert Figures 1.6 - 1.7]

Not surprisingly logarithmic trade durations are overall negatively correlated with the level of trade activity (both on the bid and on the offer side) and with the level of incoming limit orders which is probably driven by an informational asymmetry in the market. This result shows volatility clustering, where periods characterized by high trading intensity and volatility are usually followed by a number of consecutive periods of high and slowly decaying volumes. In line with Dufour and Engle (2000), we also find some evidence of a negative correlation between trade durations and available liquidity at the best price. The greater the time between two consecutive transactions and the lower the level of available liquidity displayed on one side of the market. The spread, defined as the difference between the ask and the bid price, is also negatively correlated with the current and the lag liquidity duration but shows signs of positive correlation with both the current and the lag transactional duration. A market maker tends to reduce the spread if available liquidity remains stable and does not change suddenly. Also, due to inventory effects and to the fact that liquidity, in the form of limit orders, has an intrinsic cost, spreads are tighten in order to give a greater incentive to trade and absorb non utilized quantities of liquidity. However, an increase in the time between consecutive trades gives a market making agent the incentive to widen the spread as the informational asymmetry increases and the price discovery process becomes less

transparent. Changes in available liquidity are also found to be associated with large shifts in trading activity. This very last result ties well with the observation of a M-shaped pattern in trade and liquidity durations during the trading day. An increase in market orders, associated with a lower trade duration, tends to drain an increasing amount of liquidity which again motivates market making agents to widen spreads and reduce the number of additional trades.

1.4 Estimation and Results

In this section, we report and comment the results from the estimation of the VAR system defined in (1.7)-(1.10) and study the impact of both liquidity and transactional durations on liquidity and trade dynamics. In this context, time durations can be characterized by intraday deterministic or stochastic components. In order to capture possible diurnal effects in time durations, we introduce a set of dummy variables, $D_{j,t-i}$, one for the first trading hour, one for the middle part of the day and a final one for the last trading hour of the day. A Wald test of the null that all the lagged diurnal dummies are jointly zero is performed. The null is rejected at a five percent confidence interval with a p -value of 0.0375. The Ljung-Box statistic is also examined to determine the proper lag structure in the VAR model specification and we identify a structure with five lags.

The final VAR model specification is:

$$\begin{aligned}
\Delta q_t^k &= \sum_{i=1}^5 [\gamma_{q,i}^q + \delta_{q,i}^q \ln(T_{t-i}^q)] \Delta q_{t-i}^k + \sum_{i=0}^5 [\gamma_{\omega,i}^q + \delta_{\omega,i}^q \ln(T_{t-i}^\omega)] \Delta \omega_{t-i} + \dots \\
&\quad + \sum_{j=1}^3 D_{j,t} \Delta \omega_t + v_{1,t} \\
\Delta \omega_t &= \sum_{i=1}^5 [\gamma_{q,i}^\omega + \delta_{q,i}^\omega \ln(T_{t-i}^q)] \Delta q_{t-i}^k + \sum_{i=1}^5 [\gamma_{\omega,i}^\omega + \delta_{\omega,i}^\omega \ln(T_{t-i}^\omega)] \Delta \omega_{t-i} + \dots \\
&\quad + \sum_{j=1}^3 D_{j,t-1} \Delta \omega_{t-1} + v_{2,t}
\end{aligned} \tag{1.13}$$

1.4.1 The Relevance of Time in the VAR System

The estimated VAR coefficients are shown in Tables 1.5-1.7 where the bold format is used to indicate the values of the coefficients that are statistically significantly different from zero at a five percent confidence level. We use White's *heteroscedasticity* and *autocorrelation consistent standard error* to compute the Wald and the *t*-statistics.

[Insert Tables 1.5 - 1.7]

We initially run three different models with slightly different specifications. In Model 1 (see Table 1.5) we study the joint behavior of cumulative liquidity changes and levels of trade activity. We notice again strong signs of autocorrelation in both the liquidity and the trade activity equations. In the liquidity equation, we also detect a negative relationship between changes in cumulative liquidity and the overall level of trading activity. Liquidity and transactional time are significant but only at the very first lags, while there seems to be no deterministic component in transactional time. In the trade equation, we immediately observe a positive relationship between trading activity and

liquidity. Time is again statistically relevant but only at the first lag showing, in this case, both a stochastic and a deterministic component. These very first results tend to confirm our initial findings. Due to inventory effects, liquidity tends to dry up with time and with an increase in trade activity. Also, the greater the time between consecutive transactions and the lower the associated liquidity. This result can be explained with the risk-averse behavior of a market making agent who, in the absence of fresh and incoming information, channelled through trades, is less inclined to provide liquidity. Looking at the size of the coefficient, liquidity time seem to have a stronger impact than transactional time in both the equations, being positively related to changes in liquidity and negatively related to trade activity. This again can be explained looking at the decaying-pattern of liquidity. A greater time at which liquidity moves from one state to another is typically associated to a lower trading activity which, at the same time, allows liquidity to accumulate at a faster pace at given price level.

Very similar results are obtained in Model 2 (see Table 1.6), where changes in cumulative liquidity are related also to changes in the liquidity displayed at the best price in the order book. The coefficients for the trade time durations are still negative, in both the equations, as in Dufour and Engle (2000), despite being statistically significant only at the first two lags. The negative correlation can be explained by the fact that liquidity dries up slowly as transactions become less frequent. A lower number of transactions may be associated with a higher degree of informational asymmetry in the market and this explains why the liquidity posted by a market maker is automatically reduced due to a higher risk aversion. The coefficients of the dummy variables in the trade equation are particularly significant unlike in the liquidity equation where no sign

of diurnal deterministic effects can be observed. The estimation results also point to a strong autocorrelation in both the available liquidity and the trade activity equations. The negative autocorrelation in the liquidity process can be explained through an inventory analysis of limit order books. The existence of feedback strategies can instead explain the positive autocorrelation in the trade activity equation. The presence of a greater number of informed trading agents, initially responsible for an increase in trading activity, may also be used to support this result.

In Model 3 (see Table 1.7), we relate changes in cumulative liquidity to variations in trading activity. Unlike in the previous cases, we immediately observe a negative autocorrelation in the trade process and a positive and statistically significant impact of trade durations in both the equations. We also notice a strong presence of diurnal effects in transactional time in the first equation, despite only in the first hour of the trading day, while again the three dummies seem to be all statistically significant in the trade equation.

We test the stochastic and diurnal effects of time and report the results in Table 1.8.

[Insert Table 1.8]

The first column of Table 1.8 shows Wald statistics for the hypothesis that the time coefficients, including the coefficients of the dummy variables, are jointly zero. The null is rejected in both the liquidity and the trade equations. In the second column, we test if the diurnal dummy alone is statistically significant and, also in this case, we reject the null of hypothesis in the case of the trade equation. Finally, in the third column, we test the significance of the stochastic component and we find strong stochastic effects

in the time process in the two equations considered in the analysis. The sum of the time coefficients is positive and also statistically different from zero.

The estimation results for the TACD model are presented in Table 1.9. Quasi-maximum likelihood is used under the assumption of an exponential distribution for the standardized durations. In particular, as discussed in Pacurar (2008), the standard exponential distribution is often used given that provides quasi-maximum likelihood (QML) estimators for the ACD parameters. Consistent and asymptotically normal estimates of the ACD model are obtained by maximizing the quasi-likelihood function, even if the distribution of the standardized durations $\tilde{x}_i = x_i/\psi_i\phi(t_{i-1})$ is not exponential. In the TACD model, we use two different regimes, $R_{k=1}$ and $R_{k=2}$, with the threshold indicators being a function of the rolling mean and standard deviation of both liquidity time. The first trading regime, $R_{k=1}$, is characterized by low trading volatility or, equivalently, by high time durations. Conversely, a high trading intensity and short time durations can be observed in the second trading regime, $R_{k=2}$. Figure 1.8 shows the threshold indicators during May 6, 2010. In particular, we plot the liquidity time for the available liquidity at $k = 0$ ticks from the best displayed price in the order book (dark blue line) together with the threshold indicators (light blue line). Low volatility, or alternatively, high liquidity durations can be observed during the first and the last hours of the trading session. The trading hours where liquidity displays the highest turnover are between 6:00 am and 9:00 am and between 12:00 and 15:00 pm during London time. Liquidity time shows the tendency to increase as we approach the London close and peaks at around 18:00 before moving lower again due to the higher liquidity exchanged during the NY hours.

[Insert Table 1.9 and Figure 1.8]

From the estimation results, we can observe a high statistical significance in the autoregressive parameter (beta coefficient) in all the different models used. Both the trade and liquidity time processes are weakly stationary given that $\sum_{i=1}^n \alpha_i + \sum_{i=1}^n \beta_i < 1$ despite being close to unity showing a certain degree of persistence. In Table 1.9 we also show that the autoregressive parameter is more significant and of greater relevance in the second regime characterized by a higher trading intensity. A number of misspecification tests for the TACD model are presented in Tables 1.10-1.11. We report a set of normality tests, a Box-Pierce test for serial correlation in the standardized residuals with lags equal to 5, 10, 20 and 50 and finally a Box-Pierce tests for serial correlation in the squared standardized residuals again with lags equal to 5, 10, 20 and 50. We find evidence of strong serial correlation in both the liquidity and trade duration equations and strong signs of non-normality across the different regimes.

[Insert Tables 1.10 - 1.11]

In Tables 1.12-1.13 we show the autocorrelations and the Ljung-Box tests for standardized and squared standardized time durations in the two trading regimes. Testing the assumption of independence in the standardized time durations provides an additional diagnostic check on the model. For the TACD (1,1), the Ljung-Box is much less than the statistic for the raw and adjusted durations shown in Table 1.13 across the different regimes. The statistics strongly exceeds the 5% critical value in the case of liquidity durations and in the second regime. The autocorrelation is otherwise very weak particularly in the first regime. Similar results are obtained in the case of squared

standardized durations where the autocorrelation is even weaker. These statistics show that, despite the initial misspecification tests, the TACD (1,1) model is still able to provide satisfactory results and account for the intertemporal dependence in transaction arrival times.

For completeness, in Appendix C, we report the results of the estimation for the general *exponential*ACD (EACD) and *EXponential* ACD (EXACD) models as shown in Tables 1.16-1.17 respectively where QML has also been used.

[Insert Tables 1.12 - 1.13]

1.4.2 Analysis of the Model's Dynamics

In this section, we measure the impact of an exogenous shock, due to unexpected trading activity, on available liquidity using impulse response functions. Given that $\sum_{i=1}^n \delta_{\omega,i} \neq 0$, we use a Montecarlo experiment to construct an estimator for the impulse response function where liquidity is conditional on both transactional and liquidity time durations. We follow a similar approach as in Dufour and Engle (2000) where we first filter out time-of-day effects from both the transactional and liquidity duration series by fitting a piece-wise linear spline to obtain a diurnally adjusted series with unit mean. We then fit an exponential TACD (1,1) model on the adjusted time series and estimate the conditional time durations. The estimated coefficients of the TACD model, used to construct the impulse response functions, are presented in the first columns of the bottom panel of Table 1.9. The QML estimates of the TACD coefficients are obtained using standard GARCH software where the dependent variable is set equal to $\sqrt{x(i)}$ and the conditional mean equal to zero. We use the exponential

TACD to simulate the conditional durations $n = 60$ steps forward, we then compute the impulse response function for each of the n steps using the estimated VAR coefficients from Table (1.5), we repeat the procedure over 5,000 times, and finally average the outcome of the impulse response function at each time step.

We consider three models based on alternative specifications for the impulse functions. We use the representation in (1.6), introduced by Hasbrouck (1991), where $v_{2,t} = 1$ represents an exogenous shock in the trade activity equation. We indicate the first model with the notation HS. We then use the same approach as in Dufour and Engle (2000) and compute impulse response functions as in (1.7) where trade durations are assumed non-constant. We indicate the second model with the notation DE. We finally run model (1.12) where both liquidity and time durations are assumed to be non-constant. We indicate the third model with the notation UM. We run the three models across different trading regimes and we report the results in Table 1.14. We first consider the entire trading day without any distinction between high and low trading volatility regime. We then analyze a regime, indicated with $R1$, characterized by low trading volatility and high trade durations. We finally consider a regime, indicated with $R2$, characterized by high trading volatility and low trade durations. In all these scenarios, we start from a steady-state equilibrium where $\Delta\omega_t = 0$ and $\Delta q_t = 0$ with $q_t^0 = q^*$ at time $t = 0$. We introduce a shock in the trade activity equation equal to $v_{2,t} = 1$ and measure the subsequent impact on the liquidity process.

[Insert Table 1.14]

The plot of the different function outputs are reported in Figure 1.9. The adjustment time for liquidity to converge to the equilibrium level tends to be quicker and smoother

when only the central and most liquid hours of the trading day are considered. In the first row of the graph we show the adjustment process of cumulative available liquidity adjusts when the entire trading sample is considered. It takes an average of 90 - 100 seconds for cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price to recover from the exogenous shock. The second row shows the adjustment process during the most illiquid times of the trading day, where transactions are less frequent and trading volatility low. The adjustment process is much more erratic and noisy and it takes an average of 250 seconds for liquidity to absorb the shock. The third row shows the liquidity adjustment process during the most liquid part of the day where it takes an average of 60 - 80 seconds for liquidity to move back towards equilibrium.

A comparison between impulse response functions across different trading volatility regimes is shown in Figure 1.10. We can observe that the liquidity adjustment process is much slower and erratic in the first regime (R1) characterized by lower trading volatility and higher time durations compared to the second regime (R2). Figure 1.11 shows instead a comparison between different models across the two trading regimes. In particular, we compare the model where only trading durations are non-constant (model DE), with the model where both the liquidity and the transactional time are non-constant (model UM). The impulse response functions obtained from the two models converge at the same time. However, model UM shows larger shock waves around the equilibrium level in the two trading regimes considered in the analysis. Not surprisingly the size of the liquidity jump is also bigger, on average, in the first trading regime due to the lower trading volatility and liquidity depth and in the second model due to fact that liquidity time is assumed to be non-constant. Moving across regimes, we can

observe that positive liquidity shocks are more persistent and of greater size in the second model.

[Insert Figures 1.9 - 1.11]

It is interesting to note that liquidity experiences an initial negative jump following unexpected trading activity only in the first model across the different trading regimes considered. A negative jump is recorded in the second and in the third model only in the first trading regime. However, in the second regime, the liquidity jump is positive for both model two and three. We believe that this liquidity behavior is due to a different risk aversion level across the two trading regimes considered. In the second regime a market making agent is clearly less risk-adverse and more inclined to provide liquidity even in the event of unexpected trading due to a higher level of market activity and greater transparency. The ability of the market maker to distinguish between informed vs. uninformed trading activity, in the first regime, is much lower due to the lack of flows and informational content. This explains a negative liquidity jump following an unexpected trade.

We confirm the previous estimation results presented as in Easley and O'Hara (1992), Lyons (1995), Dufour and Engle (2000) and Payne (2003) and conclude that, following an external shock, liquidity adjusts at a higher rate during times of high volatility and low transactional durations. The convergence towards an equilibrium level becomes slower during times of low volatility and high transactional durations. The negative relationship between information content of trades and durations indicates that, during times-of-day characterized by lower volatility and market activity,

the duration between consecutive transactions tends to be higher, the liquidity impact of a trade more persistent and the liquidity adjusting process slower.

1.4.3 Model Sensitivity to Microstructural Variables

A final important step in our analysis is to investigate the significance of a number of microstructural variables to address not only the possible additional contribution in explaining market liquidity but also to evaluate the degree of correlation with stochastic time.

We extend model (1.13) as follows:

$$\begin{aligned} b_i &= \left[\gamma_{\omega,i}^q + \sum_{j=1}^J \lambda_{j,\omega,i}^q D_{j,t-i} + \delta_{\omega,i}^q \ln(T_{t-i}^\omega) + \theta_{1,\omega,i}^q noi_{t-i} + \theta_{2,\omega,i}^q moi_{t-i} + \theta_{3,\omega,i}^q s_{t-i} \right] \\ d_i &= \left[\gamma_{\omega,i}^\omega + \sum_{j=1}^J \lambda_{j,\omega,i}^\omega D_{j,t-i} + \delta_{\omega,i}^\omega \ln(T_{t-i}^\omega) + \theta_{1,\omega,i}^\omega noi_{t-i} + \theta_{2,\omega,i}^\omega moi_{t-i} + \theta_{3,\omega,i}^\omega s_{t-i} \right] \end{aligned} \quad (1.14)$$

The extended model allows the time dependent coefficients to be a function of three additional transactional variables; noi_t represents the net-order imbalance given by the difference between new limit orders on the bid and on the ask side of the order book at $k = 0 - 2$ ticks from the best price, moi_t a market impact variable given by the ratio between price change and the level of trading activity at time $t - 1$; and s_t the price spread measured as the difference between the bid and the ask price at time t expressed in units of the domestic currency.

From a comparison of the results reported in Table 1.15 with those reported in Table 1.5, we notice that the coefficients, γ_i , δ_i and λ_i are substantially the same which favors

the conclusion that the three additional variables complement the explanatory power of the original stochastic time model. Strong autocorrelation at all lags is still evident in both the equations. The different set of coefficients in the four columns of Tables 1.5-1.15 are of the same size and sign. In particular we observe from the second column of the Tables that changes in trade activity negatively impact the available liquidity even if the effect tends to fade quickly soon after the first lags. Liquidity duration remains positively related to liquidity as shown in the third column of the Table 1.15 while there are still signs of a strong negative relationship between trade durations and both changes in trade activity and liquidity as shown in the four column of Table 1.15. The coefficients for the diurnal dummy are statistically significant only in the trading equation of both models as shown in the fifth column of Tables 1.5 and 1.15.

Looking at the three additional transactional variables, we observe a strong significance in the price spread in both the liquidity and the trading equations ($\theta_{3,i}$) while both the net order imbalance ($\theta_{1,i}$) and the market impact of a trade ($\theta_{2,i}$) show a poor explanatory power in both the equations. Not surprisingly an increase in price spread is responsible for a decrease in volumes and in the displayed liquidity especially at lags greater than one. Net order imbalance is statistically significant in the liquidity equation even if the net effect is close to zero, while it fails to show any significance in the trade activity equation. The sign of the market impact variable is in line with the one of the lagged trade activity and indicates the presence of a high degree of trading volatility clustering. Trading activity seems to become more frequent during periods of high market volatility and when the market impact of a trade is large potentially pointing to informed trading. This is particularly relevant looking at the sign and the

size of the coefficients of the trading equation in the seventh column of Table 1.15.

[Insert Table 1.15]

1.5 Final Remarks

In Chapter 1, we extended the vector autoregressive model with time dependent coefficients introduced by Hasbrouck (1991) and Dufour and Engle (2000) in order to study the relationship between available liquidity and trading activity in foreign exchange markets. The main novelties of our study can be summarized as follows. First, we use a specific measure of liquidity directly observed from the foreign exchange markets as opposite to a liquidity proxy measured by transactional volumes and trade frequency. Second, the dynamic behavior of liquidity is estimated by allowing time to have a deterministic and a stochastic component. Third, we make a further distinction by identifying a stochastic liquidity and transactional time. Finally, we model conditional expected durations using a regime switching threshold representation in order to incorporate a state dependent trading intensity. We use transactional, time stamped, data for EUR/USD spot over the trading week, from May 3 to May 7, 2010.

There is robust empirical evidence of a strong negative relationship between both the levels and the changes in trading activity and the changes in the amount of available liquidity in a limit order book, strong and persistent autocorrelation effects and a significant impact of time durations in both the liquidity and in the trade activity process especially at the first lags. No deterministic component could be found in the liquidity time process. Strong diurnal effects were instead found in the transactional time process. Impulse response functions were used to study the impact of exogenous

shocks, in the form of unexpected trade activity, on liquidity. We found that unexpected trade activity has an initial but only temporary negative impact. The use of regime switching threshold models allowed us to identify and capture different sensitivities of liquidity between two trading regimes. During times of high volatility and intense turnover, liquidity adjusted quickly to the equilibrium level reached in the previous state. During times of low volatility and poor market activity, instead, the adjustment process became slower and more erratic. In addition to time dependence, we have also evaluated the impact of a number of microstructural variables. A strong statistically significant negative relationship between price spreads and market activity and changes in displayed liquidity was observed, while other variables like net order imbalance and market impact showed weak if not insignificant relationships.

The results presented in Chapter 1 have interesting practical implications and can be particularly relevant in the analysis of dynamic order-allocation strategies or in the construction of high frequency algorithmic trading programmes. The most common issues for a market making trading agent are to identify at what price levels liquidity is available and for what size, to understand how liquidity evolves through time and under different trading regimes and to measure the speed at which liquidity can converge back to the equilibrium level following an external shock to the system. We believe that the framework presented in this study gives an important contribution to address these issues. The impulse response functions derived from the regime switching model also allow to distinguish between temporary and permanent shocks in the liquidity process and to isolate the information content of a trade. In this chapter we have also measured the impact of an exogenous shock, caused by unexpected trading activity, on an en-

ogenous liquidity process. However, we did not address the transmission mechanism between liquidity and prices and considered the causality effects of a liquidity shock. The significance of a liquidity jump can be tested using combinations of conventional nonparametric tests for jumps as shown in Dumitru and Urga (2012). This step will not only allow us to exploit the information content of genuine jumps but also the possibility to forecast and measure the price impact of a liquidity jump. This exercise is even more relevant when considering a data set characterized by high volatility. This line of research is developed in Chapter 2.

Table 1.1: Summary statistics.

The data are observed at a millisecond level and subsequently re-sampled over a 10 second time interval. Available Liquidity is the amount of liquidity, measured in EUR millions, displayed at k ticks from the best bid and the offer respectively. Average duration is the time, measured in milliseconds, between changes in the respective variables. Market activity is generated by market orders, new limit orders and cancelled orders. Market orders are measured in EUR millions and represent the amount of liquidity drained from the market either on the bid (market orders sell) or on the offer (market orders buy). Cancelled and new limit orders are again expressed in EUR millions and represent respectively the amount of liquidity firstly displayed and subsequently withdrawn from the order book or fresh new liquidity added to the order book at k ticks from the best available price, where $k = \{0, 1, 2\}$ with one tick being measured as 0.0001 US Dollars per 1 EUR. The spread is simply measured as the difference between the best ask and the bid price and expressed in pips of the domestic currency (USD). Gross market activity is given by the sum of the market activity on the bid and on the ask side of the book. Net order imbalance is measured as the excess amount of liquidity displayed on the bid over the ask side of the book. The market impact variable, measured on the bid and on the ask side of the market, is a proxy of market resilience and is given by the relative mid-price variation over the level of market activity on the bid (ask) respectively.

Variables	BID Side				OFFER Side			
	No Obs	Average Px	Average Size	Average Duration	No Obs	Average Px	Average Size	Average Duration
Available Liquidity	39960	1.2949	6.21	747	39960	1.2951	6.20	768
Market Orders	16875	1.2936	4.79	719	16889	1.2987	4.73	745
New Limit Orders L:1	6658	1.2995	3.76	1140	6346	1.2996	3.69	1060
New Limit Orders L:2	6220	1.2983	3.18	1068	5945	1.2990	2.94	1045
New Limit Orders L:3	5283	1.2979	2.31	963	5314	1.2978	2.36	920
Cancelled Orders L:2	5720	1.2986	3.32	1069	5759	1.2985	2.98	997
Cancelled Orders L:3	4720	1.2979	2.20	1054	4769	1.2981	2.15	1036
Spread	39960	-	1.55	-				
Market Activity	23533	1.2993	4.50	-				
Gross Market Activity	36278	1.2987	5.77	-				
Net Order Imbalance	39960	-	0.76	-				
Market Impact	39959	-	-4.0e-07	-				

Table 1.2: Estimated coefficients for trade equation.

The Table reports the coefficient estimates for the trade equation. $\Delta\omega_t^k$ represents changes in the trading activity, where k represents a binary variable assuming values equal to 0 if the trade activity is performed on the bid side of the market and values equal to 1 if the trade activity is performed on the offer side of the market. More specifically a transaction on the bid (offer) side of the market is recorded when (i) a buy (sell) limit order is matched by a sell (buy) market order, under the assumption that market orders can only be executed against the best limit orders available at a certain price and also (ii) when a new limit order is added to the order book. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

$$\Delta\omega_t^k = c + \sum_{i=1}^5 \alpha_i^0 \Delta\omega_{t-i}^0 + \sum_{i=1}^5 \alpha_i^1 \Delta\omega_{t-i}^1$$

Bid - Equation				Offer - Equation			
α_1^0	-0.7708	α_1^1	0.0036	α_1^0	-0.7851	α_1^1	0.0048
α_2^0	-0.6100	α_2^1	0.0125	α_2^0	-0.6232	α_2^1	-0.0009
α_3^0	-0.4583	α_3^1	0.0059	α_3^0	-0.4724	α_3^1	-0.0042
α_4^0	-0.3130	α_4^1	0.0072	α_4^0	-0.3157	α_4^1	-0.0104
α_5^0	-0.1656	α_5^1	0.0018	α_5^0	-0.1604	α_5^1	0.0192
Ljung-Box (5)	4727.2			Ljung-Box (5)	4896.6		
Sample Size	39960			Sample Size	39960		

Table 1.3: Estimated coefficients for liquidity equation.

The Table reports the coefficient estimates for the liquidity equation. Δq_t^k represents changes in cumulative available liquidity, observed on the bid side of the order book, at $k = 0 - 2$ ticks from the best displayed price. Δq_t^1 , Δq_t^2 and Δq_t^3 represent the available liquidity, observed on the bid side of the order book, at $k = 0, 1$ and 2 ticks respectively from the best displayed price in the book. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

$$\Delta q_t^k = c + \sum_{i=1}^5 \alpha_i^k \Delta q_{t-i}^k + \sum_{i=1}^5 \alpha_i^1 \Delta q_{t-i}^1 + \sum_{i=1}^5 \alpha_i^2 \Delta q_{t-i}^2 + \sum_{i=1}^5 \alpha_i^3 \Delta q_{t-i}^3$$

	Cumulative	1 Price Level	2 Price Level	3 Price Level
α_1^k	-0.3740	α_1^1 -0.0508	α_1^2 -0.0430	α_1^3 -0.0357
α_2^k	-0.2036	α_2^1 -0.0513	α_2^2 -0.0471	α_2^3 -0.0326
α_3^k	-0.1311	α_3^1 -0.0399	α_3^2 -0.0362	α_3^3 -0.0191
α_4^k	-0.0872	α_4^1 -0.0262	α_4^2 -0.0241	α_4^3 -0.0104
α_5^k	-0.0499	α_5^1 -0.0171	α_5^2 -0.0124	α_5^3 -0.0218
Ljung-Box (5)	166.15			
Sample Size	39960			

Table 1.4: Autocorrelations of time durations.

The Table reports the autocorrelations of both liquidity and transactional durations measured in millisecond. Liquidity duration is the time, measured in milliseconds, between changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price. Transactional duration is the time, measured in milliseconds, between two consecutive market orders. Bold denotes significance at the 5 percent level.

Lag Structure	Raw Durations		Adjusted Durations	
	Liquidity	Transactional	Liquidity	Transactional
Lag 1	0.0411	0.0369	0.0349	0.0363
Lag 2	0.0601	0.0535	0.0424	0.0298
Lag 3	0.0670	0.0177	0.0912	0.0144
Lag 4	0.0361	0.0258	0.0438	0.0204
Lag 5	0.0538	0.0647	0.0367	0.0401
Lag 6	0.0340	0.0242	0.0356	0.0359
Lag 7	0.0394	0.0179	0.0274	0.0212
Lag 8	0.0116	0.0276	0.0140	0.0309
Lag 9	0.0060	0.0306	0.0024	0.0359
Lag 10	0.0130	0.0160	0.0149	0.0188
Lag 11	0.0156	0.0227	0.0278	0.0288
Lag 12	0.0206	0.0252	0.0152	0.0292
Lag 13	0.0567	0.0474	0.0392	0.0302
Lag 14	0.0586	0.0288	0.0793	0.0349
Lag 15	0.0250	0.0224	0.0258	0.0253
Ljung-Box (15)	126.70	76.84	125.00	58.54
Sample Size	39960	39960	39960	39960

Table 1.5: Model 1 - Estimated coefficients for liquidity and trade equation.

The Table reports the coefficient estimates for both the liquidity and the trade activity equations as in 1.13 (Model 1). Δq_t represents changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price, ω_t represents instead the level of trading activity on the buy side of the order book. More specifically a transaction on the bid (ask) side of the market is recorded when (i) a buy (sell) limit order is matched by a sell (buy) market order, under the assumption that market orders can only be executed against the best limit orders available at a certain price and also (ii) when a new limit order is added in the order book. T_t^q is the time observed for changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price while the variable T_t^ω represents the duration between two consecutive market orders both expressed at a millisecond level. Finally, D_t represents the diurnal dummy variable used to indicate the time of the day. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

Constant	Lag Liquidity		Lag Trade Activity		Lag Liquidity		Lag Trade Activity		Lag Trade		R^2	
						Lag Duration		Lag Duration		Diurnal Dummy		
Panel A: Liquidity Equation												
c	0.0120	$\gamma_{q,1}$	-0.4970	$\gamma_{\omega,0}$	-0.0036	$\delta_{q,1}$	0.0219	$\delta_{\omega,0}$	-0.0034	λ_1	0.0029	0.24
		$\gamma_{q,2}$	-0.3365	$\gamma_{\omega,1}$	-0.0015	$\delta_{q,2}$	0.0063	$\delta_{\omega,1}$	0.0005	λ_2	-0.0003	
		$\gamma_{q,3}$	-0.2258	$\gamma_{\omega,2}$	-0.0014	$\delta_{q,3}$	0.0047	$\delta_{\omega,2}$	0.0000	λ_3	0.0027	
		$\gamma_{q,4}$	-0.1528	$\gamma_{\omega,3}$	-0.0008	$\delta_{q,4}$	-0.0018	$\delta_{\omega,3}$	-0.0001			
		$\gamma_{q,5}$	-0.0959	$\gamma_{\omega,4}$	-0.0014	$\delta_{q,5}$	0.0029	$\delta_{\omega,4}$	-0.0001			
				$\gamma_{\omega,5}$	-0.0013			$\delta_{\omega,5}$	-0.0001			
Panel B: Trade Activity Equation												
c	1.4482	$\gamma_{q,1}$	1.4856	$\gamma_{\omega,0}$	0.0743	$\delta_{q,1}$	-0.4605	$\delta_{\omega,0}$	-0.0230	λ_1	0.6748	0.15
		$\gamma_{q,2}$	1.3096	$\gamma_{\omega,1}$	0.0630	$\delta_{q,2}$	0.0021	$\delta_{\omega,1}$	-0.0065	λ_2	0.7182	
		$\gamma_{q,3}$	0.8410	$\gamma_{\omega,2}$	0.0527	$\delta_{q,3}$	-0.0710	$\delta_{\omega,2}$	-0.0051	λ_3	0.6719	
		$\gamma_{q,4}$	0.7140	$\gamma_{\omega,3}$	0.0539	$\delta_{q,4}$	0.0494	$\delta_{\omega,3}$	-0.0029			
		$\gamma_{q,5}$	0.3914	$\gamma_{\omega,4}$	0.0654	$\delta_{q,5}$	0.0332	$\delta_{\omega,4}$	0.0017			

Table 1.6: Model 2 - Estimated coefficients for liquidity and trade equation.

The Table reports the coefficient estimates for both the liquidity and the trade activity equations as in 1.13 (Model 2). Δq_t represents changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price, ω_t represents instead the level of trading activity on the buy side of the order book. More specifically a transaction on the bid (ask) side of the market is recorded when (i) a buy (sell) limit order is matched by a sell (buy) market order, under the assumption that market orders can only be executed against the best limit orders available at a certain price and also (ii) when a new limit order is added in the order book. Δq_t^1 represents changes in the available liquidity at the best displayed price in the order book also referred as liquidity at best. T_t^q is the time observed for changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price while the variable T_t^ω represents the duration between two consecutive market orders both expressed at a millisecond level. Finally, D_t represents the diurnal dummy variable used to indicate the time of the day. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

Constant	Lag Liquidity		Lag Trade Activity		Lag Liquidity At Best		Lag Trade Activity		Lag Trade		R^2	
						Lag Duration	Lag Duration		Diurnal Dummy			
Panel A: Liquidity Equation												
c	0.0119	$\gamma_{q,1}$	-0.5009	$\gamma_{\omega,0}$	-0.0031	$\delta_{q,1}$	0.0098	$\delta_{\omega,0}$	-0.0033	λ_1	0.0028	0.24
		$\gamma_{q,2}$	-0.3334	$\gamma_{\omega,1}$	-0.0016	$\delta_{q,2}$	0.0062	$\delta_{\omega,1}$	0.0005	λ_2	-0.0003	
		$\gamma_{q,3}$	-0.2262	$\gamma_{\omega,2}$	-0.0015	$\delta_{q,3}$	0.0032	$\delta_{\omega,2}$	0.0000	λ_3	0.0026	
		$\gamma_{q,4}$	-0.1508	$\gamma_{\omega,3}$	-0.0009	$\delta_{q,4}$	0.0007	$\delta_{\omega,3}$	-0.0001			
		$\gamma_{q,5}$	-0.0964	$\gamma_{\omega,4}$	-0.0014	$\delta_{q,5}$	0.0015	$\delta_{\omega,4}$	-0.0001			
				$\gamma_{\omega,5}$	-0.0014			$\delta_{\omega,5}$	-0.0001			
Panel B: Trade Activity Equation												
c	1.4293	$\gamma_{q,1}$	1.2158	$\gamma_{\omega,0}$	0.0808	$\delta_{q,1}$	-0.3686	$\delta_{\omega,0}$	-0.0198	λ_1	0.6686	0.15
		$\gamma_{q,2}$	1.0970	$\gamma_{\omega,1}$	0.0679	$\delta_{q,2}$	-0.1312	$\delta_{\omega,1}$	-0.0073	λ_2	0.7050	
		$\gamma_{q,3}$	0.8716	$\gamma_{\omega,2}$	0.0528	$\delta_{q,3}$	-0.0263	$\delta_{\omega,2}$	-0.0056	λ_3	0.6680	
		$\gamma_{q,4}$	0.6511	$\gamma_{\omega,3}$	0.0548	$\delta_{q,4}$	-0.0024	$\delta_{\omega,3}$	-0.0026			
		$\gamma_{q,5}$	0.3664	$\gamma_{\omega,4}$	0.0646	$\delta_{q,5}$	0.0032	$\delta_{\omega,4}$	0.0014			

Table 1.7: Model 3 - Estimated coefficients for liquidity and trade equation.

The Table reports the coefficient estimates for both the liquidity and the trade activity equations as in 1.13 (Model 3). Δq_t represents changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price, $\Delta \omega_t$ represents instead changes in trading activity on the buy side of the order book. More specifically a transaction on the bid (ask) side of the market is recorded when (i) a buy (sell) limit order is matched by a sell (buy) market order, under the assumption that market orders can only be executed against the best limit orders available at a certain price and also (ii) when a new limit order is added in the order book. Δq_t^1 represents changes in the available liquidity at the best displayed price in the order book also referred as liquidity at best. T_t^q is the time observed for changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price while the variable T_t^ω represents the duration between two consecutive market orders both expressed at a millisecond level. Finally, D_t represents the diurnal dummy variable used to indicate the time of the day. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

Constant		Lag Liquidity		Lag Trade Activity		Lag Liquidity At Best		Lag Trade Activity		Lag Trade		R^2
						Lag Duration		Lag Duration		Diurnal Dummy		
Panel A: Liquidity Equation												
c	-0.0006	$\gamma_{q,1}$	-0.5038	$\gamma_{\omega,1}$	-0.0011	$\delta_{q,1}$	0.0101	$\delta_{\omega,1}$	0.0003	λ_1	0.0044	0.23
		$\gamma_{q,2}$	-0.3351	$\gamma_{\omega,2}$	-0.0012	$\delta_{q,2}$	0.0064	$\delta_{\omega,2}$	0.0001	λ_2	0.0004	
		$\gamma_{q,3}$	-0.2291	$\gamma_{\omega,3}$	-0.0009	$\delta_{q,3}$	0.0034	$\delta_{\omega,3}$	0.0000	λ_3	0.0027	
		$\gamma_{q,4}$	-0.1527	$\gamma_{\omega,4}$	-0.0006	$\delta_{q,4}$	0.0005	$\delta_{\omega,4}$	0.0001			
		$\gamma_{q,5}$	-0.0998	$\gamma_{\omega,5}$	-0.0011	$\delta_{q,5}$	0.0018	$\delta_{\omega,5}$	-0.0003			
Panel B: Trade Activity Equation												
c	-0.1966	$\gamma_{q,1}$	1.4825	$\gamma_{\omega,0}$	-0.7808	$\delta_{q,1}$	-0.4146	$\delta_{\omega,0}$	0.0033	λ_1	0.6836	0.44
		$\gamma_{q,2}$	1.3076	$\gamma_{\omega,1}$	-0.5979	$\delta_{q,2}$	-0.1398	$\delta_{\omega,1}$	0.0100	λ_2	0.6214	
		$\gamma_{q,3}$	0.9823	$\gamma_{\omega,2}$	-0.4483	$\delta_{q,3}$	-0.0214	$\delta_{\omega,2}$	0.0084	λ_3	0.7042	
		$\gamma_{q,4}$	0.7385	$\gamma_{\omega,3}$	-0.2917	$\delta_{q,4}$	-0.0014	$\delta_{\omega,3}$	0.0127			
		$\gamma_{q,5}$	0.4139	$\gamma_{\omega,4}$	-0.1431	$\delta_{q,5}$	-0.0067	$\delta_{\omega,4}$	0.0114			

Table 1.8: The significance of time in the liquidity and trade activity equations.

The Table reports the statistical significance of the diurnal and the stochastic component in both the liquidity and transactional duration equations. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

<i>Equation</i>	Diurnal and Stochastic	Diurnal	Stochastic Component	
	Components	Dummy		
	$\lambda_j = \delta_i = 0 \ (i = 1, \dots, 5 - j = 3)$	$\lambda_j = 0 \ (j = 1, 2, 3)$	$\delta_i = 0 \ (i = 1, \dots, 5)$	$\sum \delta_i = 0$
<i>Liquidity</i>	749.84	3.6442	401.75	4.78
Transactional Time	396.66	-	372.68	69.66
Liquidity Time	108.18	-	11.91	5.71
<i>Trade Activity</i>	4373.2	3928.0	56.91	27.50
Transactional Time	3954.9	-	21.23	28.69
Liquidity Time	4324.2	-	31.82	17.48

Table 1.9: TACD model estimation.

The Table reports the parameter estimates for a regime switching threshold ACD (TACD) model on both liquidity and transactional durations, expressed in milliseconds, and after removing the diurnal effects through a polynomial interpolation. Conditional durations depend on m past durations and q past expected durations. The model allows to deal with non-linearity and to accept a number of regimes, indicated by k . A regime characterized by low trading intensity (or equivalently high time durations) is denoted with $k = 1$. A regime characterized instead by a high trading intensity (or equivalently low time durations) is denoted with $k = 2$. The threshold levels, $R_k = [r_{k-1}, r_k]$, are computed as a function of the rolling standard deviations of both liquidity and transactional time durations. The terms ϖ_μ and ϖ_σ refer to the constant terms in the conditional mean and variance equations respectively. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimators. AIC stands for Akaike's information criterion and BIC for Bayesian information criterion. Bold denotes significance at the 5 percent level.

First Regime		TACD (1,1) $k = 1$		TACD (2,2) $k = 1$		TACD (1,1) $k = 1$		TACD (2,2) $k = 1$	
Liquidity Duration		Coefficient		Trade Duration		Coefficient		Coefficient	
ϖ_μ		3.8328		3.8285		ϖ_μ		2.2579	2.2572
ϖ_σ		0.0328		0.0366		ϖ_σ		0.3258	0.4467
α_1		0.1057		0.2010		α_1		0.0798	0.1160
α_2		-		-0.0951		α_2		-	0.0000
β_1		0.8757		0.6757		β_1		0.8680	0.0000
β_2		-		0.1959		β_2		-	0.8117
<i>Likelihood</i>		-1398		-1394		<i>Likelihood</i>		-2173	-2170
<i>AIC</i>		2.9959		2.9917		<i>AIC</i>		4.6507	4.6502
<i>BIC</i>		3.0166		3.0228		<i>BIC</i>		4.6714	4.6812
Second Regime		TACD (1,1) $k = 2$		TACD (2,2) $k = 2$		TACD (1,1) $k = 2$		TACD (2,2) $k = 2$	
Liquidity Duration		Coefficient		Trade Duration		Coefficient		Coefficient	
ϖ_μ		0.5709		0.5714		ϖ_μ		0.5272	0.5268
ϖ_σ		0.0002		-		ϖ_σ		0.0001	0.0002
α_1		0.0163		0.0168		α_1		0.0113	0.0202
α_2		-		0.0146		α_2		-	0.0000
β_1		0.9832		-0.0006		β_1		0.9886	0.2494
β_2		-		0.9703		β_2		-	0.7304
<i>Likelihood</i>		-23849		-23870		<i>Likelihood</i>		-35095	-35073
<i>AIC</i>		1.2225		1.2236		<i>AIC</i>		1.7988	1.7978
<i>BIC</i>		1.2233		1.2247		<i>BIC</i>		1.7997	1.7991

Table 1.10: TACD first regime - misspecification tests.

The Table reports misspecification tests for the regime switching threshold ACD (TACD) model on both liquidity and transactional durations for the first regime. We present (a) a set of normality tests, (b) a Box-Pierce test for serial correlation in the standardized residuals with lags equal to 5, 10, 20 and 50, (c) a Box-Pierce test for serial correlation in the squared standardized residuals with lags equal to 5, 10, 20 and 50. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimators. Bold denotes acceptance of alternative hypothesis.

First Regime	TACD (1,1) k = 1	TACD (2,2) k = 1	TACD (1,1) k = 1	TACD (2,2) k = 1
Liquidity Duration	Coefficient	Coefficient	Trade Duration	Coefficient
Skewness	1.2974	1.252	Skewness	0.53077
Kurtosis	2.5514	2.3252	Kurtosis	-0.91128
Jarque-Bera	516.46	455.4	Jarque-Bera	76.335
Box-Pierce on SR			Box-Pierce on SR	77.626
Q(5)	73.1717	72.4261	Q(5)	4.60127
Q(10)	131.72	133.754	Q(10)	8.89278
Q(20)	200.937	202.904	Q(20)	22.3093
Q(50)	287.026	285.139	Q(50)	59.0725
Box-Pierce on SSR			Box-Pierce on SSR	57.2977
Q(5)	4.00851	1.16221	Q(5)	13.5235
Q(10)	12.5374	10.2309	Q(10)	18.5387
Q(20)	17.7873	16.176	Q(20)	22.7776
Q(50)	54.9287	54.2941	Q(50)	69.6866
				56.6998

Table 1.11: TACD second regime - misspecification tests.

The Table reports misspecification tests for the regime switching threshold ACD (TACD) model on both liquidity and transactional durations for the second regime. We present (a) a set of normality tests, (b) a Box-Pierce test for serial correlation in the standardized residuals with lags equal to 5, 10, 20 and 50, (c) a Box-Pierce test for serial correlation in the squared standardized residuals with lags equal to 5, 10, 20 and 50. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimators. Bold denotes acceptance of alternative hypothesis.

Second Regime	TACD (1,1) $k = 1$	TACD (2,2) $k = 1$	Trade Duration	TACD (1,1) $k = 1$	TACD (2,2) $k = 1$
Liquidity Duration	Coefficient	Coefficient		Coefficient	Coefficient
Skewness	1.379	1.3645	Skewness	1.8245	1.8207
Kurtosis	2.4172	2.4888	Kurtosis	6.4724	6.4667
Jarque-Bera	21868	22180	Jarque-Bera	89767	89557
Box-Pierce on SR			Box-Pierce on SR		
Q(5)	1430.06	1495.01	Q(5)	1124.29	1114.94
Q(10)	2585.59	2705.37	Q(10)	1346.42	1336.32
Q(20)	4680.36	4901.22	Q(20)	1667.52	1656.19
Q(50)	10698.7	11219.8	Q(50)	2509.95	2495.11
Box-Pierce on SSR			Box-Pierce on SSR		
Q(5)	29.5228	33.4665	Q(5)	24.255	18.2018
Q(10)	33.4288	38.8055	Q(10)	29.0272	22.4539
Q(20)	41.5295	46.1851	Q(20)	35.933	29.6279
Q(50)	71.5135	75.2655	Q(50)	60.4597	54.6636

Table 1.12: Autocorrelations of standardized time durations in TACD model.

The Table reports the autocorrelations of standardized liquidity and trade durations measured in millisecond in the two regimes considered $R = [1, 2]$. Standardized durations are defined as $\tilde{x}_i = x_i / \psi_1 \phi(t_{i-1})$ where x_i represents either liquidity or transactional durations, ψ_1 the respective conditional durations and $\phi(t_{i-1})$ the diurnal adjustment factor. Liquidity duration is the time, measured in milliseconds, between changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price. Trade duration is the time, measured in milliseconds, between two consecutive market orders. Bold denotes significance at the 5 percent level.

Lag Structure	Standardized Durations R:1		Standardized Durations R:2	
	Liquidity	Transactional	Liquidity	Transactional
Lag 1	0.1824	0.0049	0.0424	0.0428
Lag 2	0.0492	-0.0132	0.0185	-0.0086
Lag 3	0.0718	0.0461	0.0184	0.0042
Lag 4	0.0194	0.0312	0.0146	0.0099
Lag 5	0.0358	0.0122	0.0179	0.0098
Lag 6	0.0341	0.0295	0.0152	-0.0062
Lag 7	0.0340	-0.0243	0.0122	-0.0012
Lag 8	0.0202	-0.0320	0.0178	-0.0012
Lag 9	0.0016	-0.0277	0.0074	0.0051
Lag 10	0.0340	0.0402	0.0117	0.0078
Lag 11	-0.0098	0.0217	0.0178	-0.0009
Lag 12	0.0242	0.0415	0.0116	-0.0007
Lag 13	0.0154	-0.0521	0.0111	-0.0032
Lag 14	0.0522	0.0152	0.0094	-0.0038
Lag 15	-0.0181	0.0276	0.0096	-0.0029
Ljung-Box (15)	5.219	0.911	15.28	5.755
Sample Size	919	919	38994	38994

Table 1.13: Autocorrelations of squared standardized time durations in TACD model.

The Table reports the autocorrelations of squared standardized liquidity and trade durations measured in millisecond in the two regimes considered $R = [1, 2]$. Standardized durations are defined as $\tilde{x}_i = x_i / \psi_1 \phi(t_{i-1})$ where x_i represents either liquidity or transactional durations, ψ_1 the respective conditional durations and $\phi(t_{i-1})$ the diurnal adjustment factor. Liquidity duration is the time, measured in milliseconds, between changes in cumulative available liquidity at $k = 0 - 2$ ticks from the best displayed price. Trade duration is the time, measured in milliseconds, between two consecutive market orders. Bold denotes significance at the 5 percent level.

Lag Structure	Sq Standardized Durations R:1		Sq Standardized Durations R:2	
	Liquidity	Transactional	Liquidity	Transactional
Lag 1	0.1136	0.0165	0.0292	0.0022
Lag 2	0.0344	-0.0220	0.0118	-0.0037
Lag 3	0.0366	0.1429	0.0116	-0.0020
Lag 4	-0.0042	0.0155	0.0083	0.0007
Lag 5	0.0211	0.0039	0.0116	0.0023
Lag 6	0.0112	-0.0205	0.0056	-0.0007
Lag 7	0.0136	-0.0278	0.0075	-0.0002
Lag 8	0.0249	-0.0391	0.0293	-0.0002
Lag 9	-0.0129	-0.0221	0.0050	0.0006
Lag 10	0.0185	0.0202	0.0028	0.0004
Lag 11	-0.0108	0.0467	0.0122	-0.0004
Lag 12	-0.0056	0.0309	0.0042	-0.0004
Lag 13	0.0014	-0.0429	0.0044	-0.0001
Lag 14	0.1092	0.0155	0.0020	-0.0006
Lag 15	-0.0266	0.0057	0.0060	-0.0002
Ljung-Box (15)	1.989	2.461	5.308	0.075
Sample Size	919	919	38994	38994

Table 1.15: Model 4 - estimated coefficients for liquidity and trade equation.

The Table reports the coefficient estimates for both the liquidity and the trade activity equations as in 1.13-1.14. The additional variables are represented by the *net-order imbalance* (*noi*), the *market impact of a trade* (*moi*) and the *price spread* (*s*). More in particular, the variable *noi* is given by the difference between new limit orders on the bid and on the ask side of the market at $k = 0 - 2$ ticks from the best price. The variable *moi* is given by the ratio between price changes and the level of market activity at time $t - 1$. Finally, the variable *s* is measured as the difference between the bid and the ask price at time t expressed in units of the domestic currency. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimator to construct Wald and t-statistics. Bold denotes significance at the 5 percent level.

$$b_i = \left[\gamma_{\omega,i}^q + \sum_{j=1}^J \lambda_{j,\omega,i}^q D_{j,t-i} + \delta_{\omega,i}^q \ln \left(T_{t-i}^\omega \right) + \theta_{1,\omega,i}^q noi_{t-i} + \theta_{2,\omega,i}^q moi_{t-i} + \theta_{3,\omega,i}^q s_{t-i} \right]$$

$$d_i = \left[\gamma_{\omega,i}^\omega + \sum_{j=1}^J \lambda_{j,\omega,i}^\omega D_{j,t-i} + \delta_{\omega,i}^\omega \ln \left(T_{t-i}^\omega \right) + \theta_{1,\omega,i}^\omega noi_{t-i} + \theta_{2,\omega,i}^\omega moi_{t-i} + \theta_{3,\omega,i}^\omega s_{t-i} \right]$$

Cons	Lag Liq	Lag Trade Act	Lag Liq	Lag Trade	Lag Trade	Lag Trade	Net Oda Imb	Lag Trade	Lag Trade	Lag Trade	Spread
Panel A: Liquidity Equation											
<i>c</i>	0.011	$\gamma_{\omega,0}$	$\delta_{q,1}$	$\delta_{\omega,0}$	λ_1	-0.002	$\theta_{1,0}$	0.00	$\theta_{2,0}$	$\theta_{3,0}$	121.60
R^2		$\gamma_{\omega,1}$	$\delta_{q,2}$	$\delta_{\omega,1}$	λ_2	0.001	$\theta_{1,1}$	-0.00	$\theta_{2,1}$	$\theta_{3,1}$	-50.47
	0.28	$\gamma_{\omega,2}$	$\delta_{q,3}$	$\delta_{\omega,2}$	λ_3	0.002	$\theta_{1,2}$	0.00	$\theta_{2,2}$	$\theta_{3,2}$	-38.82
		$\gamma_{\omega,3}$	$\delta_{q,4}$	$\delta_{\omega,3}$			$\theta_{1,3}$	0.00	$\theta_{2,3}$	$\theta_{3,3}$	-12.59
		$\gamma_{\omega,4}$	$\delta_{q,5}$	$\delta_{\omega,4}$			$\theta_{1,4}$	-0.00	$\theta_{2,4}$	$\theta_{3,4}$	-11.94
		$\gamma_{\omega,5}$		$\delta_{\omega,5}$			$\theta_{1,5}$	0.00	$\theta_{2,5}$	$\theta_{3,5}$	-9.91
Panel B: Trade Activity Equation											
<i>c</i>	1.538	$\gamma_{\omega,1}$	$\delta_{q,1}$	$\delta_{\omega,1}$	λ_1	0.666	$\theta_{1,1}$	0.00	$\theta_{2,1}$	$\theta_{3,1}$	-329.20
R^2		$\gamma_{\omega,2}$	$\delta_{q,2}$	$\delta_{\omega,2}$	λ_2	0.711	$\theta_{1,2}$	0.00	$\theta_{2,2}$	$\theta_{3,2}$	-251.05
	0.15	$\gamma_{\omega,3}$	$\delta_{q,3}$	$\delta_{\omega,3}$	λ_3	0.677	$\theta_{1,3}$	-0.00	$\theta_{2,3}$	$\theta_{3,3}$	-270.81
		$\gamma_{\omega,4}$	$\delta_{q,4}$	$\delta_{\omega,4}$			$\theta_{1,4}$	0.00	$\theta_{2,4}$	$\theta_{3,4}$	-323.87
		$\gamma_{\omega,5}$	$\delta_{q,5}$	$\delta_{\omega,5}$			$\theta_{1,5}$	-0.00	$\theta_{2,5}$	$\theta_{3,5}$	-327.31

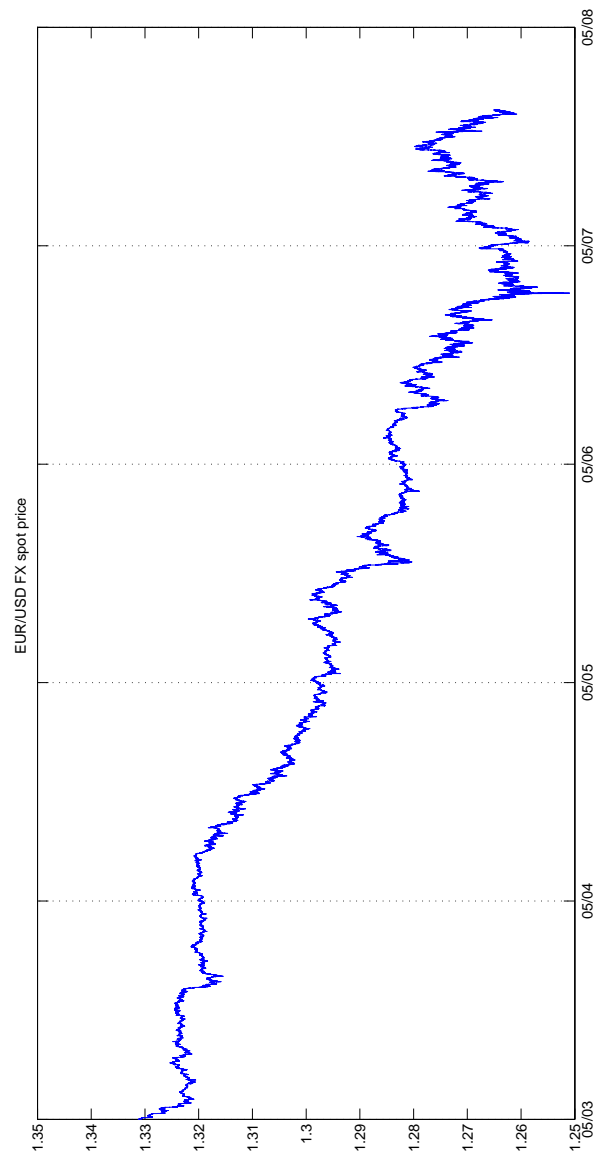


Figure 1.1: EUR/USD mid-quote price behavior.

The Figure shows the mid-quote price action of EUR/USD spot during the week from May 3 to May 7, 2010. In this time frame, EUR/USD spot moved from 1.3250 at the Asian opening to 1.2750 at the New York close on Friday. The vertical dotted line shows the end of the NY trading session and the open of the Australian trading session.

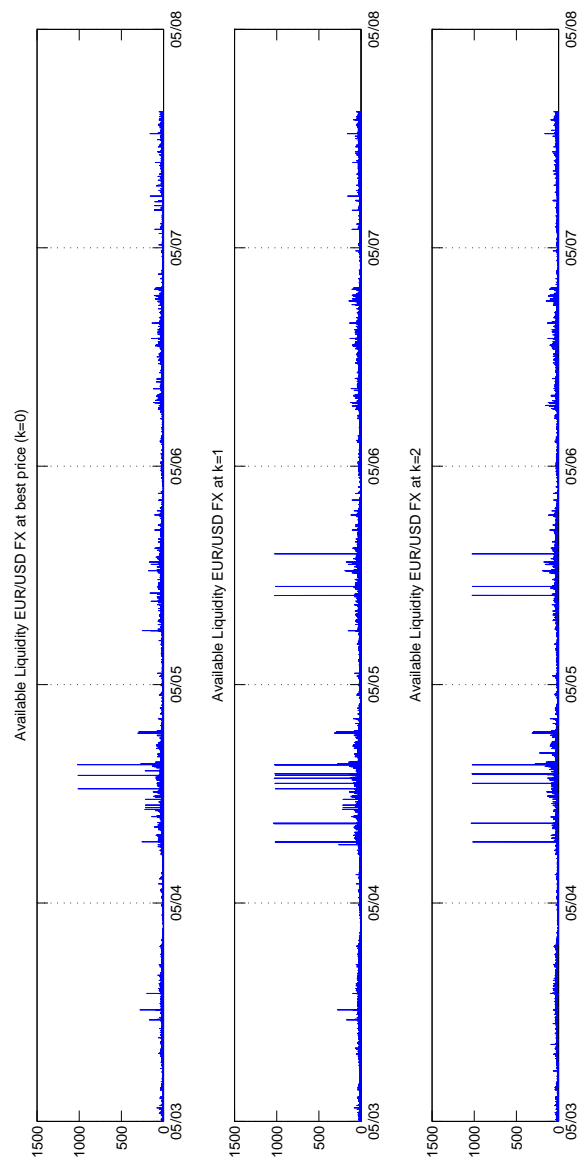


Figure 1.2: Available liquidity in the order book.

The Figures show the available liquidity observed at $k = 0$, 1 and 2 ticks respectively from the best displayed price on the bid side of the order book during the week from May 3 to May 7, 2010. Available liquidity is measured in EUR millions. The vertical dotted line shows the end of the NY trading session and the open of the Australian trading session.

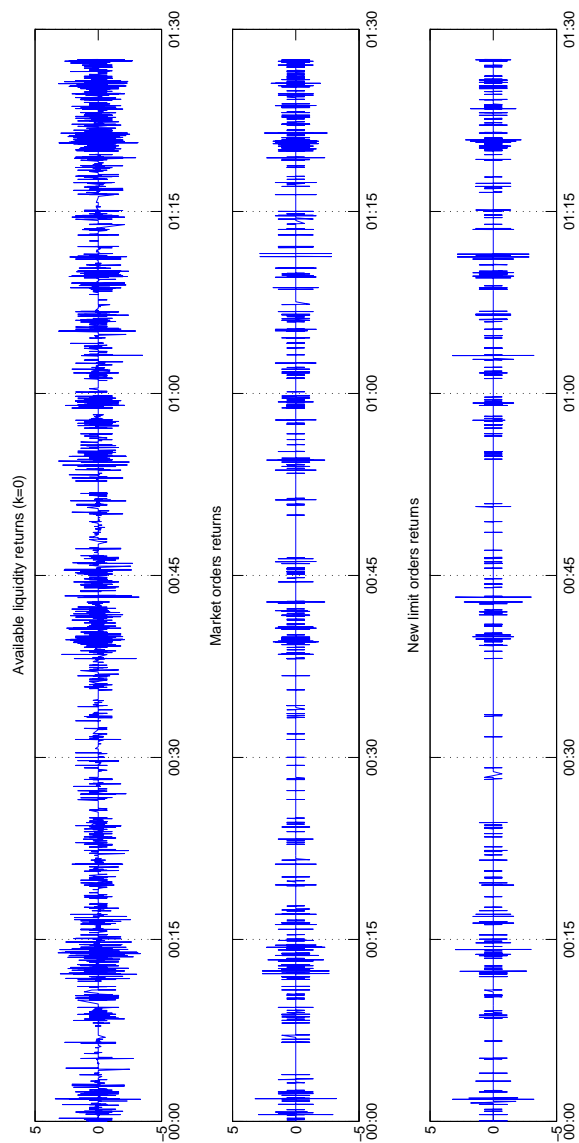


Figure 1.3: Returns in liquidity and trading activity Asian opening.

The Figures show the returns of available liquidity at $k = 0$ ticks from the best displayed price and the returns of market and new limit orders on the bid side of the order book during the Asian opening on May 3, 2010. The combination of market orders on the opposite side of the book and new limit orders define the trading activity on the bid side of the order book.

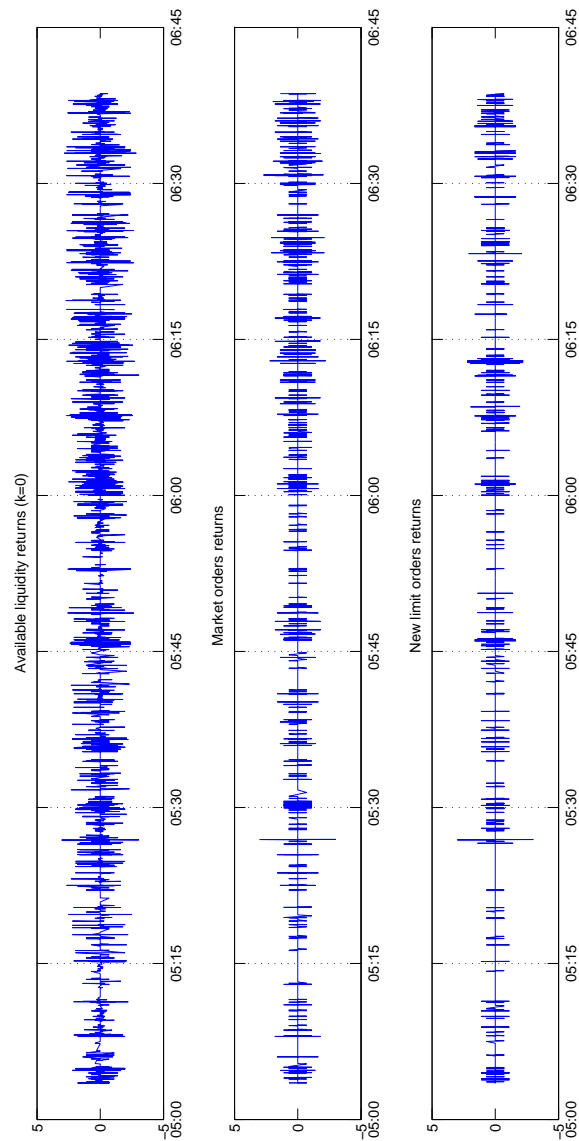


Figure 1.4: Returns in liquidity and trading activity London opening.

The Figures show the returns of available liquidity at $k = 0$ ticks from the best displayed price and the returns of market and new limit orders on the bid side of the order book during the London opening on May 3, 2010. The combination of market orders on the opposite side of the book and new limit orders define the trading activity on the bid side of the order book.

N	ObsTime	LiqTime	TransTime	px_bid_1	px_bid_2	px_bid_3	q_bid_1	q_bid_2	q_bid_3	mo_sell	clim_bid_2	clim_bid_3	nlim_bid_1
1	0	0.0011	0.0026	1.3311	1.3310	1.3309	4	26	13	0	0	0	0
2	104	0.0265	0.0026	1.3312	1.3311	1.331	1	3	26	0	0	0	0
3	385	0.0254	0.5248	1.3310	1.3309	1.3308	23	10	13	1	0	0	0
4	108	0.1027	0.0026	1.3310	1.3309	1.3308	22	10	13	1	0	0	0
5	200	0.2220	0.5164	1.3309	1.3308	1.3307	7	11	12	22	0	0	0
6	97	0.1137	0.2645	1.3307	1.3306	1.3305	7	10	14	7	0	0	0
7	699	0.1126	0.2617	1.3308	1.3307	1.3306	2	7	10	0	0	0	0
8	301	0.1225	0.0026	1.3308	1.3307	1.3306	3	9	10	0	0	0	1
9	204	0.2162	0.0026	1.3308	1.3307	1.3306	5	9	10	0	0	0	2
10	1696	1.7648	4.0978	1.3307	1.3306	1.3305	14	7	14	5	0	0	0
11	204	0.0011	20.9931	1.3307	1.3306	1.3305	14	7	14	0	0	0	0
12	1700	0.1235	0.0026	1.3309	1.3308	1.3307	1	30	16	0	0	0	0
13	104	0.3351	0.7771	1.3308	1.3307	1.3306	6	14	8	1	0	0	0
14	3918	0.6634	1.5379	1.3309	1.3308	1.3307	2	7	15	0	0	0	0
15	86	0.2170	0.5030	1.3308	1.3307	1.3306	7	14	9	2	0	0	0
16	103	0.2181	0.5052	1.3308	1.3307	1.3306	6	14	8	1	0	1	0
17	102	0.1167	0.2703	1.3309	1.3308	1.3307	1	8	15	0	0	0	0
18	99	0.6551	1.5163	1.3308	1.3307	1.3306	8	15	8	1	0	0	0

Figure 1.5: Market, new limit and cancelled limit orders - Example.

The Figure shows the methodology used to derive market, cancelled and new limit orders from the raw data. In particular, the *first* box shows that a sell market order is identified either when the best displayed bid price in the order book decreases and computed as the difference between the available liquidity, on the bid side of the order book, over the time interval $[t - 1, t]$ or when the best displayed bid price remains constant but available liquidity drops. The *second* box shows that a new limit order is identified when the best displayed bid price remains constant over the time interval $[t - 1, t]$ but the available liquidity increases over the same time interval. Finally, the *third* box shows that a cancelled limit order is identified at $k = 1$ or $k = 2$ ticks from the best displayed price in the order book in the event where the price does not change over the time interval $[t - 1, t]$ but the available liquidity drops.

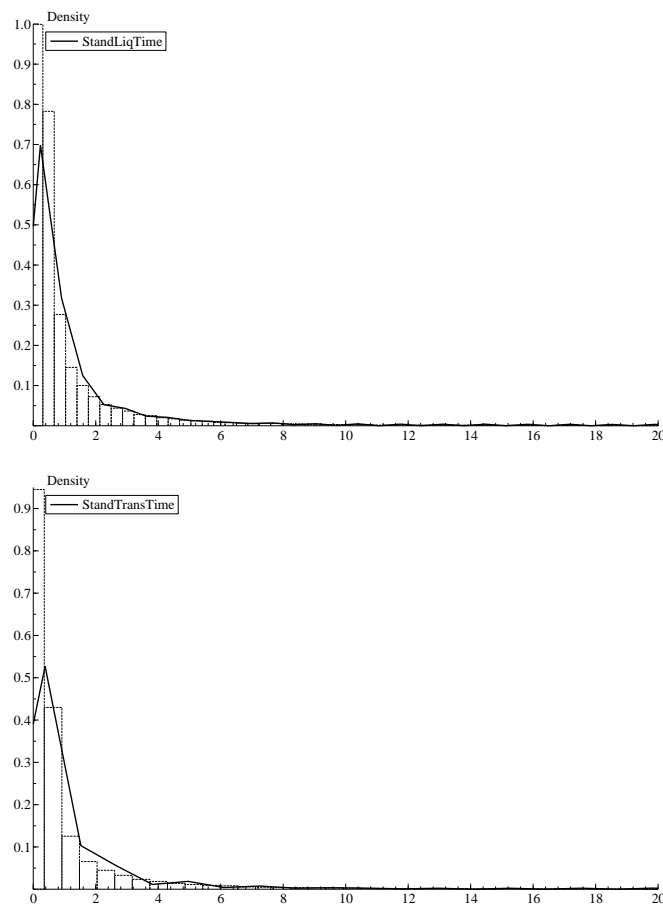


Figure 1.6: Histogram of time durations.

The Figures show the histogram and respective distribution of adjusted liquidity and transactional durations. Liquidity time is the time, measured in milliseconds, between changes in available liquidity at $k = 2$ ticks from the best displayed price. Transactional time is the time, measured in milliseconds, between two consecutive market orders.

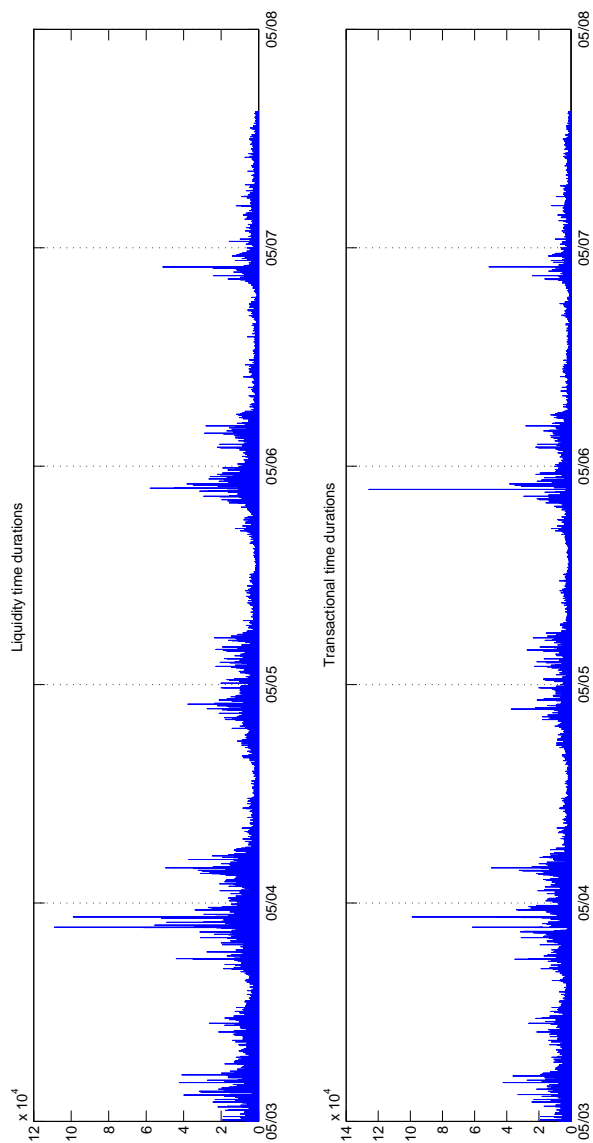


Figure 1.7: Liquidity and transactional durations.

The Figures show the M-shaped pattern of liquidity and transactional durations. Liquidity time is the time, measured in milliseconds, between changes in available liquidity at $k = 0$ ticks from the best displayed price in the order book. Transactional time is the time, measured in milliseconds, between two consecutive market orders. The vertical dotted line shows the end of the NY trading session and the open of the Australian trading session. In line with previous studies on price discovery, the beginning and the end of a trading day are characterized by a lower volatility, greater time duration between consecutive trades and a low liquidity turnover.

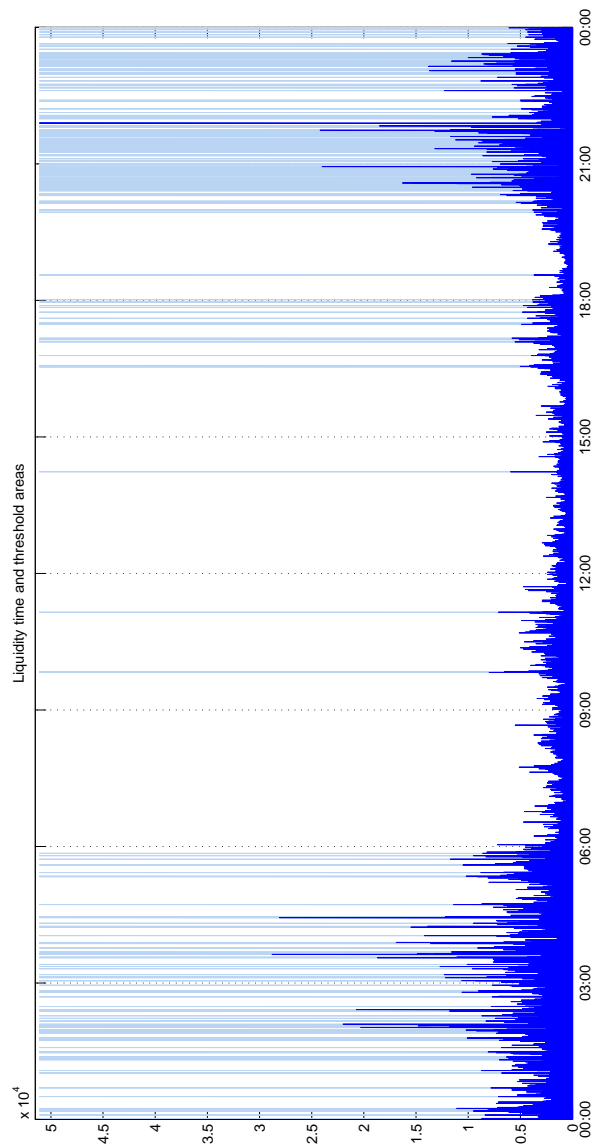


Figure 1.8: Liquidity time and threshold indicators.

The Figure shows the liquidity time, expressed in millisecond, between changes in available liquidity at $k = 0$ ticks from the best displayed price and the threshold areas during May 6, 2010. The threshold indicators, displayed in light blue, identify trading regimes characterized by low volatility or, alternatively, by high liquidity duration. Low volatility is observed during the very first and the closing hours of the trading day.

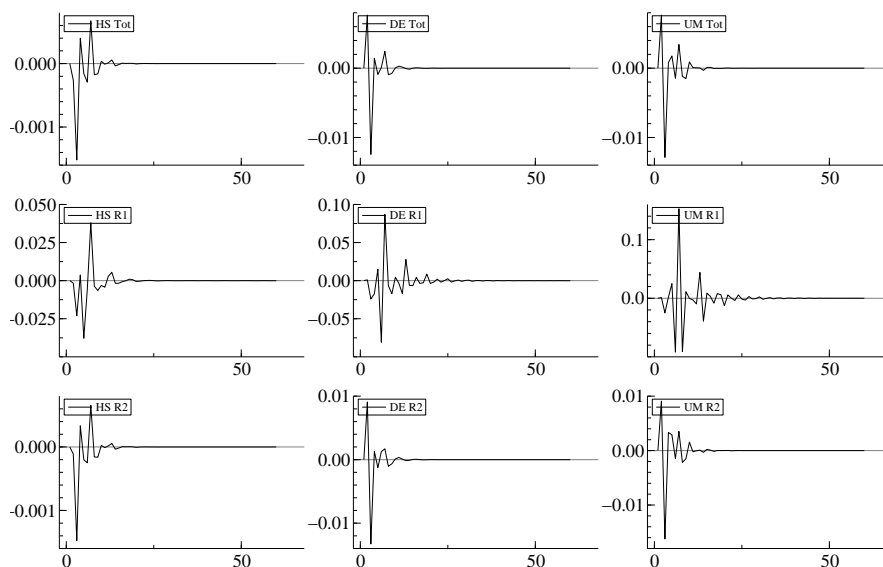


Figure 1.9: Impulse response functions for available liquidity.

The Figures show a number of impulse response functions for cumulative available liquidity at different point in time during the trading week and assuming different model specifications. We run three models under different market regimes. We start from the model presented by Hasbrouck (HS), we then add time variant coefficient as in Dufour and Engle (DE) and finally study the behavior of liquidity as described in this chapter (UM). We also analyze three different trading scenarios. We first consider the entire trading day without any distinction between high and low trading volatility regimes (M: Tot). We then analyze a regime characterized by low trading volatility and high time durations (M: R1). We then analyze a regime characterized instead by high trading volatility and low time durations (M: R2). In all the scenarios, we start from a steady-state equilibrium level where $\Delta\omega_t = 0$ and $\Delta q_t^0 = 0$ with $q_t^0 = q^*$ at $t = 0$. We then introduce a shock in the trade activity equation equal to $v_{2,t} = 1$ and measure the impact on the liquidity process through the impulse response functions.

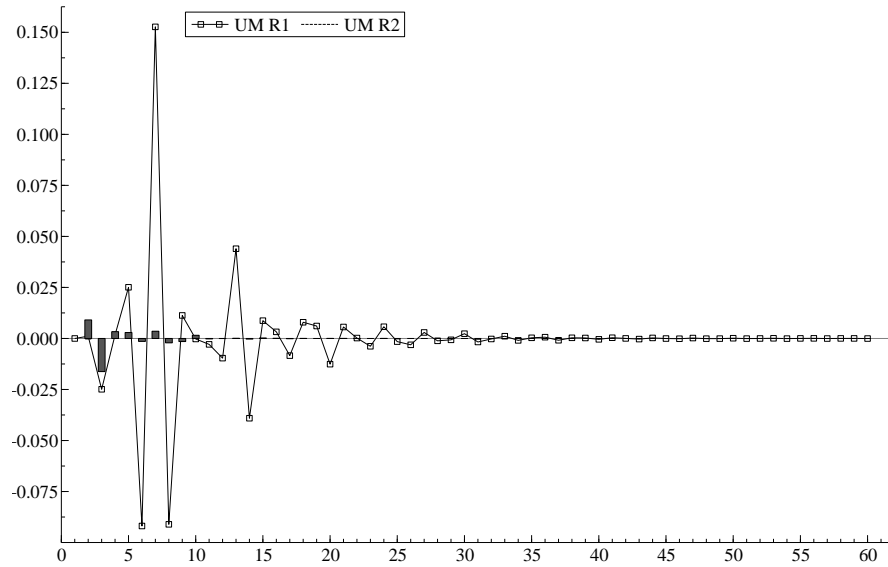


Figure 1.10: Impulse response functions: comparison across regimes.

The Figure shows a comparison between impulse response functions across different trading volatility regimes using model UM. We first analyze a regime characterized by low trading volatility and high time durations (M: R1). We then analyze a regime characterized instead by high trading volatility and low time durations (M: R2). In all the scenarios, we start from a steady-state equilibrium level where $\Delta\omega_t = 0$ and $\Delta q_t^0 = 0$ with $q_t^0 = q^*$ at $t = 0$. We then introduce a shock in the trade activity equation equal to $v_{2,t} = 1$ and measure the impact on the liquidity process through the impulse response functions.

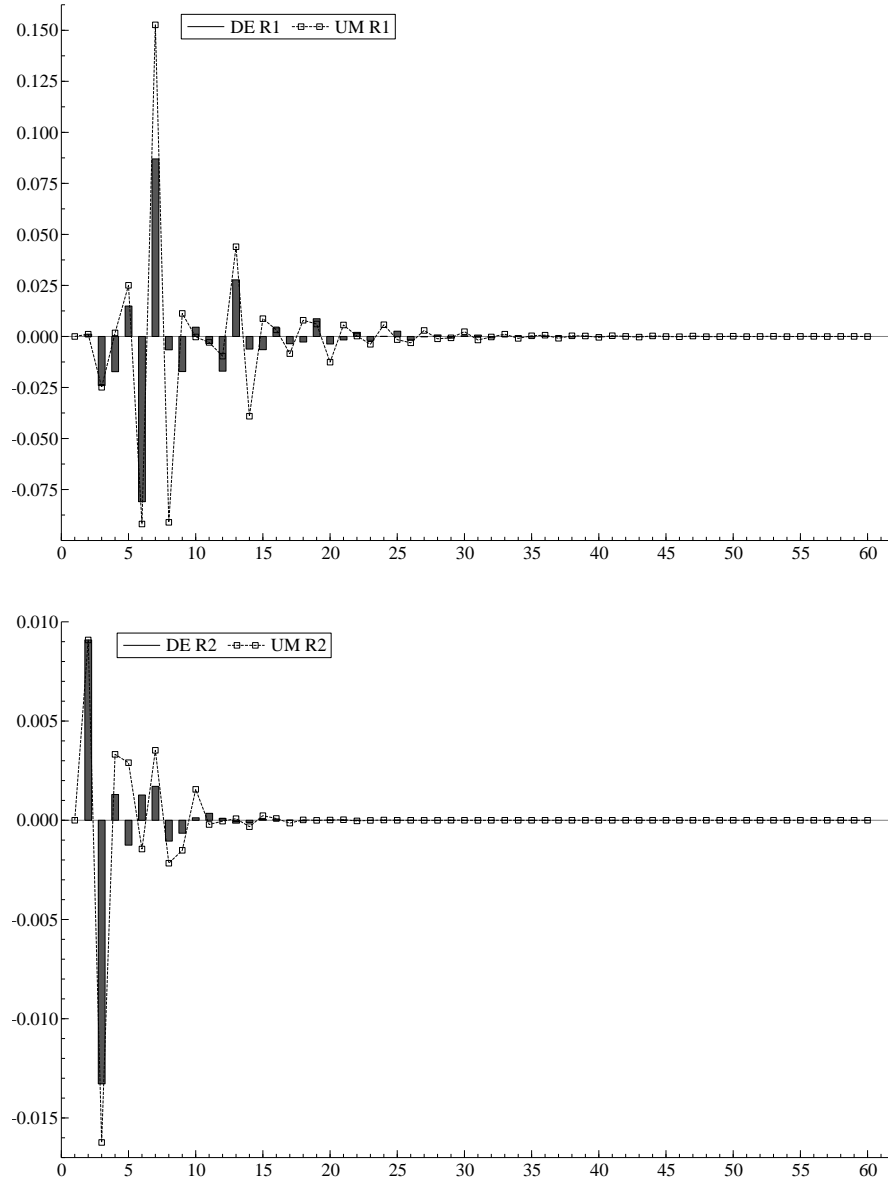


Figure 1.11: Impulse response functions: comparison across regimes.

The Figures show a comparison between impulse response functions obtained from different models and across two trading volatility regimes. The two models considered are the time variant coefficient model introduced by Dufour and Engle (DE) and the model presented in this chapter (UM). We analyze two different trading scenarios. We first analyze a regime characterized by low trading volatility and high time durations (M: R1). We then analyze a regime characterized instead by high trading volatility and low time durations (M: R2).

1.6 Appendix A: A Family of Models for Stochastic Trade Arrival Times

In Appendix A we provide a brief overview of the most common models for trade arrivals which can be considered a general extension of the ACD model introduced by Engle and Russell (1998) and Dufour and Engle (2000). A more comprehensive analysis of the theoretical and empirical literature on ACD models can be found in Pacurar (2008).

Given a series of trade durations $x_i = t_i - t_{i-1}$, where t_i represents the time at which an event has occurred, the conditional expected durations are represented by:

$$\psi_i = E(x_i | F_{i-1}) = \psi_i(x_{i-1}, z_{i-1}) \quad (1.15)$$

where x_{i-1} and z_{i-1} show the past value of the trade durations and trade marks respectively and F_{i-1} the information set available at time t_{i-1} . An important assumption of ACD models is that the standardized durations $\hat{\varepsilon}_i = x_i / \psi_i \phi(t_{i-1})$ are independent and identically distributed with $E(\varepsilon_i) = 1$ and higher order moments also independent.

Different specifications of the process for the expected durations ψ_i and different distributional assumptions for ε_i will produce a number of models for trade durations. In the original ACD model proposed by Engle and Russell (1998), conditional expected durations are described using a linear ARMA-type representation:

$$\psi_t = \varpi + \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{i=1}^q \beta_i \psi_{t-i} \quad (1.16)$$

where, in order to ensure positive conditional durations, sufficient but not necessary

conditions impose that $\omega > 0$, $\alpha > 0$ and $\beta > 0$. (1.16) can be rewritten as

$$x_t = \varpi_t + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) x_{t-j} - \sum_{j=1}^q \beta_j \eta_{t-j} + \eta_t \quad (1.17)$$

by letting $\eta \equiv x_t - \psi_t$. The sufficient condition in order to have a covariance-stationary stochastic process imposes that $\sum_{j=1}^m \alpha_j + \sum_{j=1}^q \beta_j < 1$. The density function $p(\varepsilon, \theta_\varepsilon)$ is instead defined on a non-negative support, the most common choice being either the Exponential (EACD) or the Weibull (WACD) distribution. The parameters of this first class of models, $\theta \equiv (\varpi, \alpha, \beta)$, are estimated by maximizing the following log-likelihood function:

$$L(\theta) = \sum_{i=1}^N [\log(\gamma/x_i) + \gamma \log(x_i/\psi_i) - (x_i/\psi_i)^\gamma] \quad (1.18)$$

which, in the Exponential case, when $\gamma = 1$, becomes:

$$L(\theta) = - \sum_{i=1}^N [x_i/\psi_i + \log \psi_i] \quad (1.19)$$

In the case of EACD models, consistent and asymptotically normal estimates of θ are obtained by maximizing the quasi-likelihood function described in (1.19) even if the distribution of ε is not exponential and provided that the conditional mean is correctly specified and that the standard errors are adjusted for heteroskedasticity and autocorrelation. An important limitation of EACD models relates, however, to a constant hazard function while WACD models allow an increasing (decreasing) hazard function by assuming $\gamma > 1$ ($\gamma < 1$).

Despite being the most common models for time durations, several extensions to the EACD and WACD representations have been recently proposed. The Fractionally Integrated ACD (FIACD) model has been introduced to handle the high persistence and the significant autocorrelation up to a good number of lags of many financial duration series. In the FIACD model the conditional expected duration is defined as

$$[1 - \beta(L)]\psi_t = \varpi + [1 - \beta(L) - [1 - \phi(L)](1 - L)^d]x_i \quad (1.20)$$

where $\phi(L) = \alpha(L) + \beta(L)$ and where $(1 - L)^d = \sum_{k=0}^{\infty} \Gamma(k - d)\Gamma(k + 1)^{-1}\Gamma(-d)^{-1}L^k$ with Γ being the gamma function. When $d = 1$ the FIACD model becomes an Integrated ACD (IACD).

In order to avoid negative expected durations Logarithmic ACD (LACD) models are instead introduced where the logarithm of the conditional expected duration is represented as:

$$\log(\psi_t) = \varpi + \sum_{j=1}^m \alpha_j \log(x_{i-j}) + \sum_{j=1}^q \beta_j \log(\psi_{i-j}) \quad (1.21)$$

or equivalently using the standardized durations:

$$\log(\psi_t) = \varpi + \sum_{j=1}^m \alpha_j \log(\varepsilon_{i-j}) + \sum_{j=1}^q (\beta_j - \alpha_j) \log(\psi_{i-j}) \quad (1.22)$$

where the following condition is imposed $|\alpha + \beta| < 1$ in order to ensure covariance stationary coefficients. The estimation is still performed by maximum likelihood (ML) with the Exponential or the Weibull distributions being the most used representations

for the density $p(\varepsilon; \theta_\varepsilon)$.

The EXponential ACD (EXACD) models are introduced to deal with a potential nonlinear dependence between the conditional duration and past trade durations and to measure the asymmetric impact of short vs. long durations. In particular, the conditional duration is modelled using an asymmetric function of past durations:

$$\log(\psi_t) = \varpi + \sum_{j=1}^m [\alpha_j \varepsilon_{t-j} + \delta_j (|\varepsilon_{t-j} - E(\varepsilon_{t-j})|)] + \sum_{i=1}^q \beta_i \log(\psi_{t-i}) \quad (1.23)$$

The regression slope in (1.23) depends on the trade duration being greater or lower than the conditional mean.

An alternative non-linear ACD modelling specification assumes the existence of a certain number of trading regimes each characterized by different dynamics and distributional features. A k -regime Threshold ACD (TACD) model is given by

$$\begin{aligned} x_i &= \psi_i \varepsilon_i^k \\ \psi_i &= \omega^k + \sum_{j=1}^m \alpha_j^k x_{i-j} + \sum_{j=1}^q \beta_j^k \psi_{i-j} \end{aligned} \quad (1.24)$$

where ε_i^k is an i.i.d vector with positive and regime specific intensities. The vector $k = \{1, 2, \dots, K\}$ denotes the number of regimes with $0 = r_0 < r_1 < \dots < r_K = \infty$ being the threshold values often identified with the presence of structural breaks in the time series. Again stationarity conditions impose the following restrictions on the parameters: $\omega^k > 0$, $\alpha_j^k > 0$ and $\beta_j^k > 0$.

1.7 Appendix B: Data re-sampling - Matlab Code

In Appendix B we show the Matlab code used to re-sample the original dataset and to create the microstructural variables from the liquidity data from the limit order book:

Calendar time conversion and resampling

```

while r < length(k) for i = 1:length(TRCN)
    if k(r) < TRCN(i)
        Q_bid_1(x+1) = Q1(i); P_bid_1(x+1) = L1(i);
        Q_bid_2(x+1) = Q2(i); P_bid_2(x+1) = L2(i);
        Q_bid_3(x+1) = Q3(i); P_bid_3(x+1) = L3(i);
        Time_T(x) = TRC(i); Time_T_R(x+1) = T1(i);
        r = r+1; x = x+1;
    elseif k(r) > TRCN(i)
        Q_bid_1(x+1) = Q1(i); P_bid_1(x+1) = L1(i);
        Q_bid_2(x+1) = Q2(i); P_bid_2(x+1) = L2(i);
        Q_bid_3(x+1) = Q3(i); P_bid_3(x+1) = L3(i);
        Time_T(x) = TRC(i); Time_T_R(x+1) = T1(i);
        r = r; x = x;
    end
end
end

Q_bid_T = Q_bid_1+Q_bid_2+Q_bid_3;

```

Determination of market orders and new limit orders for the bid side of the order book

```

for i = 1:dii-1
    if P_bid_1(i+1) == P_bid_1(i)
        if Q_bid_1(i+1) - Q_bid_1(i) < 0
            MO_Sell(i+1) = Q_bid_1(i) - Q_bid_1(i+1);
            NLim_bid_1(i+1) = 0;
        elseif Q_bid_1(i+1) - Q_bid_1(i) >= 0
            MO_Sell(i+1) = 0;
            NLim_bid_1(i+1) = Q_bid_1(i+1) - Q_bid_1(i);
        end
    elseif P_bid_1(i+1) < P_bid_1(i)
        MO_Sell(i+1) = Q_bid_1(i);
        NLim_bid_1(i+1) = 0;
    elseif P_bid_1(i+1) > P_bid_1(i)
        MO_Sell(i+1) = 0;
        NLim_bid_1(i+1) = 0;
    end
end
end

```

Determination of new limit orders and cancelled orders at second price level of the bid side of the order book

```

for i = 1:dii-1
    if P_bid_2(i+1) == P_bid_2(i)

```

```

if Q_bid_2(i+1) - Q_bid_2(i) < 0
    CLim_bid_2(i+1) = Q_bid_2(i) - Q_bid_2(i+1);
    NLim_bid_2(i+1) = 0;
elseif Q_bid_2(i+1) - Q_bid_2(i) >= 0
    CLim_bid_2(i+1) = 0;
    NLim_bid_2(i+1) = Q_bid_2(i+1) - Q_bid_2(i);
end
elseif P_bid_2(i+1) < P_bid_2(i)
    CLim_bid_2(i+1) = 0; NLim_bid_2(i+1) = 0;
elseif P_bid_2(i+1) > P_bid_2(i)
    CLim_bid_2(i+1) = 0;
    NLim_bid_2(i+1) = 0;
end
end
end

```

1.8 Appendix C: Estimation Results for the EACD and the EXACD Model

In Appendix C we provide the estimation results for the EACD and the EXACD Models respectively.

Table 1.16: EACD Model Estimations.

The Table below shows the parameter estimates for Exponential Autoregressive Conditional Duration (EACD) models on both liquidity and transactional durations, expressed in milliseconds, and after removing the diurnal effects through a polynomial interpolation. The terms ϖ_μ and ϖ_σ refer to the constant terms in the conditional mean and variance equations respectively. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimators. Bold denotes significance at the 5 percent level.

	EACD (1,1)	EACD (2,2)		EACD (1,1)	EACD (2,2)
Liquidity Duration	Coefficient		Trade Duration	Coefficient	
ϖ_μ	0.5739	0.5729	ϖ_μ	0.5316	0.5317
ϖ_σ	-	0.0004	ϖ_σ	0.0002	0.0004
α_1	0.0181	0.0204	α_1	0.0152	0.0150
α_2	-	0.0212	α_2	-	0.0152
β_1	0.9819	-0.0173	β_1	0.9848	-0.0143
β_2	-	0.9756	β_2	-	0.9841
<i>Likelihood</i>	-33150	-33089	<i>Likelihood</i>	-40376	-33150
<i>AIC</i>	1.6593	1.6564	<i>AIC</i>	2.0210	2.0210
<i>BIC</i>	1.6597	1.6574	<i>BIC</i>	2.0216	2.0220

Table 1.17: EXACD Model Estimations.

The Table below shows the parameter estimates for EXponential Autoregressive Conditional Duration (EXACD) models on both liquidity and transactional durations, expressed in milliseconds, and after removing the diurnal effects through a polynomial interpolation. Conditional durations depend on m past durations and q past expected durations. The model allows to measure the different impact on conditional duration of asymmetric time durations in the case where the durations are shorter or longer than the conditional mean. In the event where $\varepsilon_i < 1$ the slope factor equals $\theta_1 - \theta_2$ while $\varepsilon_i > 1$ is associated to a slope factor equal to $\theta_1 + \theta_2$. The terms ϖ_μ and ϖ_σ refer to the constant terms in the conditional mean and variance equations respectively. The standard errors are corrected by using White's heteroskedasticity consistent covariance estimators. Bold denotes significance at the 5 percent level.

	EXACD (1,1)	EXACD (2,2)		EXACD (1,1)	EXACD (2,2)
Liquidity Duration	Coefficient		Trade Duration	Coefficient	
ϖ_μ	0.5767	0.5765	ϖ_μ	-	0.5331
ϖ_σ	-	-	ϖ_σ	-	-
α_1	-0.5775	0.4254	α_1	-0.9025	0.2743
α_2	-	-0.5945	α_2	-	-0.7372
β_1	0.9975	-0.0018	β_1	1.0023	-0.0009
β_2	-	0.9969	β_2	-	0.9953
θ_1	0.0418	0.0424	θ_1	0.0558	0.0507
θ_2	0.1071	0.1090	θ_2	0.2206	0.1217
<i>Likelihood</i>	-33094	-33081	<i>Likelihood</i>	-53146	-40575
<i>AIC</i>	1.6566	1.6560	<i>AIC</i>	2.6601	2.0311
<i>BIC</i>	1.6577	1.6575	<i>BIC</i>	2.6610	2.0362

CHAPTER 2

A TESTING PROCEDURE FOR CO-JUMPS

2.1 Introduction

Chapter 2 presents a co-jump testing procedure based on the combination of univariate tests for jumps. Statistical tests that combine independent p-values using the union of rejections decision rule are discussed in Neuhauser (2003) and Harvey et al. (2009, 2011), while Loughin (2004) and Cheng and Sheng (2010) use combinations of p-values combinations. Dumitru and Urga (2012) propose a testing methodology, robust to spurious detection and microstructural noise, based on combinations of jump tests measured at different frequencies. The combination of tests allows to detect a lower number of spurious jumps or, equivalently, a greater percentage of truly-identified jumps. We extend this methodology to a multivariate context in order to identify common jumps between different stochastic processes.

The contribution of this chapter to the literature on co-jumps is threefold. First, we present a testing procedure alternative to the existing tests for co-jumps. The approach used here allows us to address some of the issues with the existing tests, to extend the notion of a co-jump event and to identify a lower number of spurious co-jumps.

The existing literature on non-parametric tests for co-jumps has only been recently developed mainly as an extension of the univariate tests for jumps. Barndorff-Nielsen and Shephard (2004a, 2004b) introduce the concept of realized bi-power covariation in order to test for co-jumps in multivariate price process as a natural evolution of the work on quadratic variation in the univariate case. Brandt and Diebold (2006) use a range based measure of the volatility of a portfolio of assets to estimate the covariance of the portfolio components. Bollerslev et al. (2008) focus on the covariance structure of intraday returns to construct a more robust test statistic for co-jumps where the notion of a cross product statistic, defined as normalized sum of individual returns, is used to measure the covariation of a portfolio of stocks. Bannouh et al. (2009) use a realized corange measure to estimate quadratic covariation which also captures important feature of microstructural models. Jacod and Todorov (2009) propose a test for co-jumps using a higher order power variation. Gobbi and Mancini (2012) allow both finite and infinite jump activity between two semimartingale processes and present an efficient and robust estimator of the diffusion part of the integrated covariation and of the co-jumps. Liao and Anderson (2011) introduce the notion of first-high-low-last which also provides a more efficient covariance estimator by using the full intraday price history of the asset price. In the common notion of a co-jump used in the literature two or more stochastic processes are characterized by a simultaneous and discontinuous path over a given time interval. The jumps are traditionally both exogenous and no causality between the two can be inferred. In this chapter, instead, we identify a causality effect between different jumps observed over a fixed time interval. We also present different types of co-jumps, and in particular we distinguish between *contemporaneous*,

permanent and *lagged* or *exogenous* co-jump events. We finally assess the performance of the proposed co-jump testing procedure under different levels of the jump intensity factor, jump size, correlation and microstructural noise. The issue of microstructural noise becomes particularly relevant and could severely affect the performance of the jump and co-jump tests especially when the data are collected and observed at a high frequency. We show that the proposed testing procedure is robust in power to different types of microstructural noise and can be easily adjusted to take into account the issue of non-synchronous trading highlighted in Bannough et al. (2009).

The chapter is organized as follows. Section 2.2 introduces the proposed co-jump testing procedure when the data are either re-sampled over equally spaced time intervals or observed at a tick-by-tick level. Section 2.3 describes the results of the Monte Carlo experiment used to evaluate the performance of a number of univariate tests for jumps and to analyze the size and the power of the various tests for co-jump. Section 2.4 concludes.

2.2 A Co-Jump Testing Procedure

The co-jump testing procedure presented in this chapter is based on a number of combinations of univariate tests for jumps for different stochastic processes and observation frequencies. In particular, we propose two distinct testing methodologies.

With the first set of tests, we combine univariate tests for jumps, computed at different observation frequencies, to define *contemporaneous*, *permanent* and *lagged* co-jump events. The tests can be used when the data are observed over regular time intervals of equal size and at different frequencies.

With the second set of tests, we combine univariate tests for jumps, computed over non-overlapping time intervals, to define *contemporaneous*, *permanent* and *exogenous* co-jump events. The tests can be used when the data are observed at a tick-by-tick level and not necessarily re-sampled over intervals of equal size.

Figure 2.1 shows the different jump and co-jump tests presented in the chapter for data re-sampled over equally spaced time intervals (left panel) and tick-by-tick data (right panel) respectively.

[Insert Figure 2.1]

Re-sampled data. In the presence of a data set, re-sampled over equally spaced time intervals, and in the case of three main re-sampling frequencies, where a time unit corresponds to one second, we identify three distinct *temporary* jump events, observed at a frequency of 1, 5 and 10 time units respectively over the time interval $[0, T]$. In a multivariate context, a *contemporaneous* co-jump event is observed when two or more jump events occur over the same time interval. In particular, we identify a contemporaneous co-jump event at a frequency of 1, 5 and 10 time units when two or more temporary jump events, observed at a frequency of 1, 5 and 10 time units respectively, occur simultaneously over the time interval $[0, T]$. A *permanent* co-jump event occurs when a co-jump, observed at a certain frequency over the time interval $[0, T]$, is also observed at a lower frequency over the same time interval. In particular, we identify a permanent co-jump event when a co-jump event at a frequency of 1 time unit is also observed at a frequency of 5 time units or alternatively when a co-jump event at a frequency of 5 time units is also observed at a frequency of 10 time units

over the same time interval. Finally, a *lagged* co-jump event occurs when the jump of one asset, observed at a certain frequency, is followed, over the same time interval $[0, T]$, by the jump of a second asset at a lower frequency under the condition of no contemporaneous co-jumps between the two assets. In particular, in the presence of two stochastic processes, we identify a lagged co-jump event when the following jump events are simultaneously observed over the same time interval $[0, T]$: a jump in asset one, at a frequency of 1 time units, is detected together with a jump in asset two, at a frequency of either 5 or 10 time units, but not at the same frequency of 1 time unit.

Tick-by-tick data. In the presence of a data set observed at a tick-by-tick level, we identify a *temporary* and a *permanent* jump event. In particular we define a temporary jump event, over two non overlapping time intervals $]T - t, T]$ and $]T, T + t]$, when a jump observed over the interval $]T - t, T]$ is not observed over the interval $]T, T + t]$. A temporary jump is also called *exogenous* as no temporal causality can be established between consecutive jumps. Alternatively, a permanent jump event is defined when a jump observed over the interval $]T - t, T]$ is also observed over the interval $]T, T + t]$. In this context, a permanent jump is also defined as *endogenous* as the likelihood to detect a jump over the time interval $]T, T + t]$ may be influenced by the presence of a jump over the time interval $]T - t, T]$. Similarly, in a multivariate context, a *contemporaneous (permanent)* co-jump event is identified when two temporary (permanent) jump events are observed over two non overlapping time intervals. Finally, we identify an *exogenous* co-jump event when two consecutive exogenous jumps are observed, over two non overlapping time intervals, or, alternatively, an *endogenous* co-jump event when two consecutive endogenous jumps are observed or when one endogenous jump occurs

together with an exogenous jump. It follows that a permanent co-jump will always also be endogenous and that an exogenous co-jump will always be contemporaneous.

In the next section we provide a more formal set of definitions for the different types of jump and co-jumps events together with the methodology used to construct the various tests proposed in this chapter and used in the empirical applications in Chapter 3.

2.2.1 Jump and Co-Jump Test Indicator Functions

Let \mathbf{X} be a random vector of returns generated from a probability distribution P_x . Let \mathbf{J} be a subsample of \mathbf{X} representing a random vector of extreme returns or jumps with $\mathbf{J} = [j_{(1)}, \dots, j_{(n)}]$. The null hypothesis $H_0 : j_{(i)} \in \Omega_0$ is tested against the alternative hypothesis of $H_1 : j_{(i)} \in \Omega_1$ for $i = 1 \dots n$ and with Ω_0 and Ω_1 being disjoint subsets of Ω . Ω_0 represents the subset of non-statistically significant jumps and Ω_1 the subset of statistically significant jumps in the jump space Ω . Alternatively Ω_0 defines the region of acceptance of H_0 while the subset Ω_1 defines the rejection or critical region. The univariate jump test statistics, $\varphi(j)$, determines whether to reject H_0 , accepting H_1 as true, or alternatively accept H_0 .

Definition 1. *Let $\varphi(j)$ be a univariate jump test statistic characterized by a known statistical distribution. We define $JT_{[0,T]}$ as a jump test indicator function which assumes values equal to one when the null of no-statistically significant jumps is rejected at a significance level α over the time interval $[0, T]$ and values equal to zero otherwise.*

More formally, for a pre-specified significance level α and critical value $c_{(\alpha)}$, we define

the *jump test indicator function* as:

$$\{JT_{[0,T]} := 1 \mid \Omega_1(\alpha) : \varphi(j)_{[0,T]} > c_{(\alpha)}\} \quad (2.1)$$

The definition provided in (2.1) can be used to identify both temporary and permanent jump events.

Moving to the multivariate case, we denote with $\mathbf{X}_{i,j}$ a matrix of correlated returns, generated from a multivariate probability distribution, and \mathbf{J}_i , \mathbf{J}_j and $\mathbf{CJ}_{i,j}$ sub-samples of $\mathbf{X}_{i,j}$ representing a random matrix of extreme returns or jumps for the stochastic process i and j and the extreme simultaneous returns or co-jumps respectively with $\mathbf{CJ} = [cj_{(1)} \dots cj_{(n)}]$. We define the null hypothesis as $H_0 : cj_{(i)} \in \Phi_0$ against the alternative hypothesis of $H_1 : cj_{(i)} \in \Phi_1$ for $i = 1 \dots n$ and with Φ_0 and Φ_1 being disjoint subsets of Φ . Φ_0 represents the subset of non-statistically significant co-jumps and Φ_1 the subset of statistically significant co-jumps in the co-jump space Φ . Alternatively, Φ_0 defines the region of acceptance of H_0 while the subset Φ_1 defines the rejection or critical region.

Definition 2. Let $\varphi_1(j)$ and $\varphi_2(j)$ be the univariate jump test statistics characterized by a known statistical distribution for the returns of two correlated stochastic processes $S_{(1)}$ and $S_{(2)}$. We define (a) $\varphi(cj)$ a multivariate co-jump test statistic characterized by a given statistical specification and given by the product of $\varphi_1(j)$ and $\varphi_2(j)$ and (b) $CJT_{[0,T]}$ a co-jump test indicator which assumes values equal to one when the null of no-statistically significant co-jumps is rejected at a significance level α over the

time interval $[0, T]$ and values equal to zero otherwise.

More formally, for a pre-specified significance level α and critical value $c_{(\alpha)}$ we define the *contemporaneous co-jump test indicator* as:

$$\{CJT_{[0,T]} := 1 \mid \Phi_1(\alpha) : \varphi(cj)_{m,[0,T]} > c_{(\alpha)}\} \quad (2.2)$$

In particular, for d -assets whose returns are computed over an observation frequency m , the *contemporaneous co-jump test indicator* is computed as:

$$CJT_{[0,T]} = \prod_{i=1}^d I(\varphi(j)_{i,m[0,T]}) \quad (2.3)$$

where $I(\cdot)$ is a unit step function which assumes values equal to one for positive defined arguments and m denotes the observation frequency. A similar definition is also used in Lahaye et al. (2011). The definition, provided in (2.2), is used to identify contemporaneous co-jump events when the data are both re-sampled over equally spaced time intervals and observed at a tick-by-tick level. The combination of (2.1) and (2.2) allows us to define also permanent and exogenous co-jump events. In particular:

Definition 3.1 *Let $\varphi(j)_{1,m(n)}$, $\varphi(j)_{2,m(n)}$ and $\varphi(cj)_{m(n)}$ be the univariate jump and co-jump test statistics respectively for the returns of two correlated stochastic processes $S_{(1)}$ and $S_{(2)}$ at two different observation frequencies m and n with $m < n$. We define $PCJT_{[0,T]}$ a co-jump test indicator which assumes values equal to one when the null of no-statistically significant permanent co-jumps is rejected at a significance level α over the time interval $[0, T]$ and values equal to zero otherwise.*

More formally, for a pre-specified significance level α and critical value $c_{(\alpha)}$ we define

the *permanent co-jump test indicator* as:

$$\{PCJT_{[0,T]} := 1 \mid \Phi_{1,m}(\alpha)\Phi_{1,n}(\alpha) : \varphi(cj)_{m,[0,T]} \cap \varphi(cj)_{n,[0,T]} > c_{(\alpha)}\} \quad (2.4)$$

with $m < n$. In particular, for d -assets whose returns are computed over observation frequencies m and n , when the data are re-sampled over equally spaced time intervals, the permanent co-jump test indicator is computed as:

$$PCJT_{[0,T]} = \prod_{z=m}^n \left\{ \prod_{i=1}^d I(\varphi(j)_{i,z[0,T]}) \right\} \quad (2.5)$$

In the case where the data are observed at a tick-by-tick level, we introduce the following definition:

Definition 3.2 *Let $\varphi(j)_{1,T_1(2)}$, $\varphi(j)_{2,T_1(2)}$, $\varphi(cj)_{T_1(2)}$ be the univariate jump and co-jump test statistics respectively for the returns of two correlated stochastic processes $S_{(1)}$ and $S_{(2)}$ over two non overlapping time intervals $T_1 =]T-t, T]$ and $T_2 =]T, T+t]$ where $[T_1, T_2] \in [0, T]$ with $T_2 > T_1$. We define $PCJT_{[T-t, T+t]}$ a co-jump test indicator which assumes values equal to one when the null of no-statistically significant permanent co-jumps is rejected at a significance level α over the time interval $[T-t, T+t]$ and values equal to zero otherwise.*

More formally, for a pre-specified significance level α and critical value $c_{(\alpha)}$ we define the *permanent co-jump test indicator* as:

$$\{PCJT_{[T-t, T+t]} := 1 \mid \Phi_{1,T_1}(\alpha)\Phi_{1,T_2}(\alpha) : \varphi(cj)_{T_1} \cap \varphi(cj)_{T_2} > c_{(\alpha)}\} \quad (2.6)$$

The permanent test, when the data are observed at a tick-by-tick level, is computed

as:

$$PCJT_{[T-t, T+t]} = \prod_{z=T_t}^{T_{t+1}} \left\{ \prod_{i=1}^d I(\varphi(j)_{i,z}) \right\} \quad (2.7)$$

Finally, we introduce the following definitions for the lagged and exogenous co-jump events respectively.

Definition 4.1 *Let $\varphi(j)_{1,m}$, $\varphi(j)_{2,n}$, $\varphi(cj)_{m,n}$ be the jump and the co-jump test statistics respectively for the returns of two correlated stochastic processes $S_{(1)}$ and $S_{(2)}$ at two different observation frequencies m and n with $m < n$. We define $LCJT_{[0,T]}$ a co-jump test indicator which assumes values equal to one when the null of no-statistically significant lagged co-jumps is rejected at a significance level α over the time interval $[0, T]$ and values equal to zero otherwise.*

More formally, for a pre-specified significance level α and critical value $c_{(\alpha)}$ we define the lagged co-jump test indicator as:

$$\{LCJT_{[0,T]} := 1 \mid \Phi_{1,m,n}(\alpha) : (\varphi(cj)_{m,n} \setminus \varphi(cj)_m > c_{(\alpha)})\} \quad (2.8)$$

In particular, for d -assets whose returns are computed over observation frequencies $\{m_1, \dots, m_n\}$ with $m_1 < m_n$, and when the data are re-sampled over equally spaced time intervals, the lagged co-jump test indicator is computed as:

$$LCJT_{[0,T]} = \max \left\{ 0, \left[\prod_{i=1}^d I(\varphi_{i,z(i)[0,T]}) \right]_{z=[m_1, \dots, m_n]} - \left[\prod_{i=1}^d I(\varphi_{i,z[0,T]}) \right]_{z=m, \dots, n} \right\} \quad (2.9)$$

where $z = [m_1, \dots, m_n]$ is a vector of size d . Similarly, when the data are observed at

a tick-by-tick level, the exogenous co-jump event is defined as:

Definition 4.2 *Let $\varphi_{1,T_1(2)}$, $\varphi_{2,T_1(2)}$, $\varphi(cj)_{T_1(2)}$ be the univariate jump and co-jump test statistics respectively for the returns of two correlated stochastic processes $S_{(1)}$ and $S_{(2)}$ over two non overlapping time intervals $T_1 =]T - t, T]$ and $T_2 =]T, T + t]$ where $[T_1, T_2] \in [0, T]$ with $T_2 > T_1$. We define $ECJT_{[T-t, T+t]}$ a co-jump test indicator which assumes values equal to one when the null of no-statistically significant exogenous co-jumps is rejected at a significance level α over the time interval $[T - t, T + t]$ and values equal to zero otherwise.*

More formally, for a pre-specified significance level α and critical value $c_{(\alpha)}$ we define the *exogenous co-jump test indicator* as:

$$\{ECJT_{[T-t, T+t]} := 1 \mid \Phi_1(\alpha) : \varphi(cj)_{T_2} \setminus \varphi(cj)_{T_2, T_1} > c_{(\alpha)}\} \quad (2.10)$$

In particular, for d -assets whose returns are computed over two non overlapping time intervals, the exogenous co-jump test indicator is computed as:

$$ECJT_{[T-t, T+t]} = \max \left\{ 0, \prod_{i=1}^d I(\varphi_{i, [T+t]}) - \left[\prod_{z=T_t}^{T_{t+1}} I(\varphi_{i, z}) \right]_{i=1, \dots, d} \right\} \quad (2.11)$$

The idea of an *endogenous* vs. *exogenous* jump together with the notion of a *temporary* vs. *permanent* co-jump between two stochastic processes is new and provides an interesting contribution to the literature. It also allows us to extend the common notion of co-jump from portfolio theory and to better understand the individual contribution of different stochastic processes to the co-jump event. For convenience, Table 2.1 reports the full set of jumps and co-jump test indicators. In particular, the top

panel shows the test indicators used when the data are re-sampled over equally spaced time intervals and for frequencies of 1, 5 and 10 time units. The bottom panel shows instead the test indicators used when the data are observed at a tick-by-tick level over two non-overlapping time intervals.

[Insert Table2.1]

2.3 Monte Carlo Experiment

2.3.1 Simulation Design

The reference model for all the simulations reported in this chapter is a jump diffusion model with compound Poisson jumps as described in Brigo et al. (2009). The stochastic differential equation (SDE) for the stochastic process $S(t)$ is given by:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) + S(t)dJ(t) \quad (2.12)$$

where $W(t)$ is a univariate Wiener process and $J(t)$ a univariate jump process which is represented as:

$$dJ(t) = (Y_{N(t)} - 1) dN(t) \quad (2.13)$$

where $N(T)$ represents a counting process which follows a homogeneous Poisson process characterized by an intensity factor λ and distributed like a Poisson distribution with parameter λT with Y_j being the size of the j -th jump. The Y_j 's are i.i.d. log-normal variables, distributed with mean μ_Y and variance σ_Y^2 , independent from

both the Brownian motion W and the Poisson process N . Applying Ito's lemma to $f(S) = \log(S)$ we can re-write (2.12) as:

$$d\log S(t) = \left(\mu + \lambda\mu_Y - \frac{1}{2}\sigma^2 \right) dt + \sigma dW(t) + [\log(Y_{N(t)}) dN(t) - \mu_Y \lambda dt] \quad (2.14)$$

The solution to (2.14) is given by:

$$S(T) = S(0) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T)\right) \prod_{j=1}^{N(T)} Y_j \quad (2.15)$$

We define $X(t) := \Delta \log(S(t))$ and apply a Euler-discretization of (2.15) to obtain:

$$X(t) = \Delta \log(S(t)) = \mu^* \Delta t + \sigma \sqrt{\Delta t} \varepsilon_t + \Delta J_t^* \quad (2.16)$$

where $\varepsilon \sim \mathcal{N}(0, 1)$, $\Delta J_t^* = \sum_{j=1}^{n_t} \log(Y_j) - \lambda\mu_Y \Delta t$ with $n_t = N_t - N_{t-\Delta t}$ and $\mu^* = (\mu + \lambda\mu_Y - 1/2\sigma^2)$.

In a multivariate set-up, where two correlated stochastic processes are modelled, we re-write (2.15) as:

$$\begin{aligned} S_{(1)}(T) &= S_{(1)}(0) \exp\left(\left(\mu_{(1)} - \frac{\sigma_{(1)}^2}{2}\right)T + \sigma_{(1)} W_{(1)}(T)\right) \prod_{j=1}^{N_{(1)}(T)} Y_{(1)j} \\ S_{(2)}(T) &= S_{(2)}(0) \exp\left(\left(\mu_{(2)} - \frac{\sigma_{(2)}^2}{2}\right)T + \sigma_{(2)} W_{(2)}(T)\right) \prod_{j=1}^{N_{(2)}(T)} Y_{(2)j} \end{aligned} \quad (2.17)$$

where the quadratic instantaneous covariation of the two Wiener processes is given by $\langle W_{(1)} W_{(2)} \rangle = \varrho(t) dt$ so that $W_{(2)}(t) = \varrho(t) W_{(1)} t + \sqrt{(1 - \varrho(t))} W_{(3)}$ with $W_{(1)}$ and $W_{(3)}$ being independent. As in the univariate case, $Y_{(1),j}$ and $Y_{(2),j}$ are i.i.d. log-normal variables, distributed with mean $\mu_{Y(1)}$ and $\mu_{Y(2)}$ and variances $\sigma_{Y(1)}^2$ and $\sigma_{Y(2)}^2$

independent from both the Brownian motion processes $W_{(1)}$ and $W_{(2)}$ and the Poisson processes $N_{(1)}(T)$ and $N_{(2)}(T)$. The processes $N_{(1)}(T)$ and $N_{(2)}(T)$ are characterized by a correlation structure implied by $\varrho(t)dt$.

In order to simulate two correlated stochastic processes, we use a Cholesky decomposition where we first compute $\mathbf{Z} \sim \mathcal{MN}(0, \mathbf{I})$ and then we set $\mathbf{V} = \mathbf{C}^T \mathbf{Z}$ with \mathbf{C} being the Cholesky decomposition of the variance-covariance matrix of the two processes denoted with Σ . The algorithm introduced by Yahav and Shmueli (2011) is used, instead, to simulate the two correlated Poisson processes, under the assumption of a constant rate vector where $\lambda_{(1)} = \lambda_{(2)}$. In particular, we first simulate two correlated and normally distributed Wiener processes $W_{(1)}^N(t)$ and $W_{(2)}^N(t)$, and then calculate the normal cumulative density functions (CDF) for each value $W_{(i)}^N$, with $i = 1, 2$. We denote the CDF with $\Phi(W_{(i)}^N)$. We finally compute the Poisson inverse CDF, with rate $\lambda_{(i)}$, as $W_{(i)}^P = \Xi \left\{ \Phi(W_{(i)}^N) \right\}$. The vector $W_{(i)}^P$ will be a two-dimensional Poisson vector with correlation matrix Σ . The Matlab code and the algorithm used in the simulation can be found in Appendix A.

We initially simulate $m = 2,000$ returns over a frequency of $dt = 1$ time unit, we then re-sample the simulated data over intervals of $dt = 5$ and 10 time units and finally re-run the simulation $n = 1,000$ times. In particular, we denote with $r_{(i),1}$, $r_{(i),5}$ and $r_{(i),10}$ the simulated returns at a frequency of one, five and ten time units for the stochastic process $S_{(i)}$ for $i = 1, 2$. The simulated time series consists of $r_1 = 2,000$, $r_5 = 400$ and $r_{10} = 200$ returns for each simulation. Table 2.2 reports the value of the parameters used in the Monte Carlo simulation.

[Insert Table 2.2]

Figure 2.2 shows the simulated path, across $n = 1,000$ simulations, of the stochastic process $S_{(1)}$ under the assumption of jumps (blue line) and zero jumps (red line). The jumps have been simulated using $\lambda_{(1)}dt = 0.25\%$, or, equivalently, $k = 5$ total number of jumps over the time horizon considered. Figure 2.3 shows instead the simulated path, across $n = 1,000$ simulations, of the stochastic processes $S_{(1)}$ and $S_{(2)}$ again under the assumption of jumps (blue line) and zero jumps (red line) and the combined path of the processes under the assumptions of jumps. The jumps have been simulated using $\lambda_{(1)}dt = \lambda_{(2)}dt = 0.05\%$ and $\varrho = 0.50$. Finally, the first two panels of Figure 2.4 show the quadratic variation of $S_{(1)}$ and $S_{(2)}$ respectively under the assumption of jumps (blue line) and zero jumps (red line) while the bottom panel shows the quadratic covariation of $S_{(1)}$ and $S_{(2)}$ again under the assumption of jumps (blue line) and zero jumps (red line). The jumps have been simulated using $\lambda_{(1)}dt = \lambda_{(2)}dt = 0.05\%$ and $\varrho_{(1,2)} = 0.50$. Not surprisingly, we notice that, in the presence of jumps, the quadratic variation and the quadratic covariation show a greater number of spikes and are both characterized by a higher level of clustering compared to the case where jumps are assumed to be zero.

[Insert Figures 2.2 - 2.4]

2.3.2 Monte Carlo Findings

We divide our Monte Carlo analysis in three distinct parts. First, we perform a comprehensive evaluation of the most common univariate tests for jumps. In particular, we estimate the linear and the ratio Barndorff-Nielsen and Shephard (2005) tests (LBNS and RBNS respectively), the Andersen et al. (2012) MinRV and MedRV test (MinRV

and MedRV respectively), the Jiang and Oomen (2008) test (JO), the Andersen, Bollerslev and Dobrev (2007, ABD) and Lee and Mykland (2008, LM) test (ABD-LM), the Corsi et al. (2010) test (CPR) and the Podolskij and Ziggel (2010) test (PZ). We compute the size and the size corrected power of the univariate tests under different levels of jump intensity, jump size and microstructural noise using a 5% significance level.

Second, we compute the first battery of co-jump tests based on combinations of univariate tests for jumps measured and estimated at different frequencies. In particular, we consider a frequency of 1, 5 and 10 time units over a fixed time interval of $m = 2,000$ observations. We evaluate the size and the power of the test presented again under different levels of jump intensity, jump size, correlation and microstructural noise using a 5% significance level.

Finally, we compute the second battery of co-jump tests on simulated data which is observed at a tick-by-tick level over non-overlapping time intervals and not re-sampled over a fixed observation frequency. As in the previous case, we evaluate the size and the power of the tests under different levels of jump intensity, jump size, correlation and microstructural noise using a 5% significance level.

2.3.2.1 An Evaluation of the Univariate Tests for Jumps

SIZE. In order to evaluate the size of the univariate tests for jumps we use the jump diffusion model specified in (2.12). We also indicate with JT_1 , JT_2 and JT_3 the jump test indicators measured at a frequency of 1, 5 and 10 time units. The top-left section of Table 2.3 reports the size of the univariate tests for jumps under the assumption of $\mu_{(1)} = \mu_{Y(1)} = 0.00$, $\sigma_{(1)} = \sigma_{Y(1)} = 0.10$ and zero microstructural noise ($\varpi = 0.00$). We

immediately notice a big size distortion in the case of the ABD-LM test at different sampling frequencies, with a size of 2.40% at a frequency of one, five and ten time units. We also find evidence of a slight oversize in the case of the LBNS, the RBNS, the MinRV and the PZ tests at a frequency of one and five time units. The MedRV and the JO test are slightly undersized especially at a frequency of five time units.

POWER. We use the jump diffusion model specified in (2.12) to evaluate the power of the tests by adding a continuous jump process characterized by a different intensity factor and jump size.

Varying jump intensity. Under the alternative hypothesis of discontinuous price paths, and in order to examine the performance of the univariate tests for jumps as the number of jumps increase, we allow $\lambda dt = \{0.05\%, 0.25\%, 0.50\%, 2.50\%, 5.00\%\}$ over the time frame considered in the analysis. The power refers, in this case, to the ability of the tests to detect a jump when the jump intensity factor is different from zero. For different values of the lambda factor, we consider a jump size that is normally distributed with $\mu_{Y(1)} = 0.00$ and $\sigma_{Y(1)} = 0.10$. The level of the microstructural noise is again set at zero ($\varpi = 0.00$). Table 2.3 reports the size corrected power of the univariate tests for jumps. We notice that the frequency of correctly identified jumps increases as the jump intensity raises. In particular, all the tests considered display a strong power at a frequency of 1 and 5 time units. A slightly lower power is observed as the sampling frequency decreases. This is particular evident in the case of the JO test and for high levels of lambda. The best tests in terms of power are the LBNS, the RBNS, the ABD-LM and the CPR while the PZ and the JO test display, on a average, a weaker power especially at lower frequencies. The LBNS, the RBNS and the

CPR tests are again ranked high in terms of power also at low frequencies. A similar behavior was also observed in Dumitru and Urga (2012).

[Insert Table 2.3]

Varying jump size. We also study the power of the jump tests by letting the jump size vary. In particular, we draw the jump size from a normal distribution with zero mean and a standard deviation that varies from $\sigma_{Y(1)} = 0.10$ to $\sigma_{Y(1)} = 1.25$. We assume that the jump intensity factor remains constant and equal to $\lambda dt = 0.25\%$. Table 2.4 reports the size corrected power of the univariate tests for jumps. The LBNS and the RBNS display again the best power together with the JO and the CPR tests. The MinRV and the MedRV tests display a lower power when the observation frequency decreases and the jump volatility increases. The PZ test is the worst in terms of power among all the tests considered in the analysis when the jump volatility is low. The ability of the test to detect jump increases and converges to the power of the other tests when the jump volatility increases. Finally, the ABD-LM test displays a very consistent powers across different levels of jump volatilities but ranks slightly lower compared to the LBNS, the RBNS and the JO tests.

[Insert Table 2.4]

Impact of Microstructural Noise. The size and the size corrected power of the univariate tests for jumps is also assessed under different levels of micro-structural noise. We consider two types of noise. First, as in Dumitru and Urga (2012), we simulate an i.i.d. microstructural noise normally distributed with mean zero and a varying variance.

We add the simulated noise to the jump diffusion model described in (2.12) and study the statistical properties of the jump tests when the variance of the noise increases. We then analyze the microstructural noise caused by rounding effects and discreteness of the observations. This type of noise can be particularly relevant when the data are conditionally heteroskedastic or serially correlated. The issue of heteroskedasticity and serial correlation becomes even more relevant when data are observed at a very high frequency.

Table 2.5 reports the size of the univariate tests for jumps under different sampling frequencies and noise variances for the first type of noise. In particular, we let the noise volatility σ_n vary from $\sigma_n = 0.10$ to $\sigma_n = 1.25$ under the assumption of $\sigma_{(1)} = \sigma_{Y(1)} = 0.10$, $\mu_{(1)} = \mu_{Y(1)} = 0.00$ and $\lambda_{(1)}dt = 0.00\%$. With the exception of the PZ test, the remaining tests suffer from undersize at a frequency of 1 time unit. While the size of the other tests tends to converge to the nominal size at lower frequencies, the JO and the ABD-LM tests are still affected by undersize at a frequency of 5 and 10 time units. The degree of undersize is particularly evident as the noise variance increases as shown with the ABD-LM test. The PZ test suffers from oversize when the noise variance moves to $\sigma_n = 0.50$ with the size distortion being particularly severe at high levels of noise and, in particular, at $\sigma_n = 1.25$. The oversize of the test shows also the tendency to increase with lower re-sampling frequencies. Table 2.6 reports the size corrected power of the tests under the presence of the first type of microstructural noise. In particular, we simulate the process under the assumption of $\sigma_{(1)} = \sigma_{Y(1)} = 0.10$, $\mu_{(1)} = \mu_{Y(1)} = 0.00$ and $\lambda_{(1)}dt = 0.25\%$ and a noise volatility again ranging from $\sigma_n = 0.10$ to $\sigma_n = 1.25$. The power of the various tests is only moderately affected by an increase in the noise

variance. The best performing test is the LBNS, followed by the RBNS and the ABD-LM. The MinRV and the MedRV also rank high. Among all the tests considered, the PZ test shows the smallest power especially at lower frequencies. However, the loss in power does not seem to be caused by an increase in noise volatility. The power of the test varies between 77% and 92% when $\sigma_n = 0.10$ and falls to a range of 77% to 89% when $\sigma_n = 1.25$. It is interesting to notice that, across the various tests, a decrease in power is associated to lower frequencies and tends to be more evident at high levels of noise. As a final remark, we observe that, while the power of the univariate tests for jumps seems to be only moderately affected by the presence of microstructural noise, the size distortion is particularly evident when data are re-sampled at a high frequency. Pre-averaging or noise reducing techniques may be necessary to limit the effects of microstructural noise especially when dealing with empirical data observed at a high frequency.

[Insert Tables 2.5 - 2.6]

Table 2.7 reports the size (top panel) and the size corrected power (bottom panel) of the univariate tests for jumps in the presence of noise caused by rounding effects. In particular, we use three different rounding rules and run the tests at different re-sampling frequencies. In order to evaluate the size of the tests, we simulate the process under the assumption of $\sigma_{(1)} = \sigma_{Y(1)} = 0.10$, $\mu_{(1)} = \mu_{Y(1)} = 0.00$ and $\lambda_{(1)}dt = 0.00\%$ and an imposed rounding of $\text{rnd} = 3, 2$ and 1 decimal places respectively. The biggest size distortion is observed in the case of the LBNS, the RBNS and the CPR tests when 1 decimal place is used in the rounding and at a frequency of 1 time unit. The ABD-LM test appears undersized in particular at a frequency of 1 time units and when 3

decimal places are used in the rounding. With the exception of the ABD-LM test, we observe, on average, a slight oversize of the various tests which increases as the number of decimal places used in the rounding decreases. The power of the tests is evaluated under the assumption of $\sigma_{(1)} = \sigma_{Y(1)} = 0.10$, $\mu_{(1)} = \mu_{Y(1)} = 0.00$ and $\lambda_{(1)}dt = 0.25\%$ and an imposed rounding of $\text{rnd} = 3, 2$ and 1 decimal places respectively. As in the case of a white noise with varying variance, we do not find evidence of major effects on the power of the tests when rounding is introduced. The best power is displayed by the LBNS, the RBNS, the MinRV and the MedRV tests followed by the JO test when 3 decimal places are used in the rounding. A similar ranking can be observed when 2 decimal places are used while the CPR test displays the best power after the LBNS and the RBNS when 1 decimal place is used. The power of the ABD-LM test is stable across frequencies and rounding levels but tends to rank lower compared to the other tests. The worst performance in terms of power is displayed by the PZ test. Not surprisingly, and in line with previous results, we observe a decreasing power with lower re-sampling frequencies.

[Insert Table 2.7]

Conclusions. The presence of microstructural noise can affect the size and the ability of the tests to detect jumps especially when the data are observed and measured at a high frequency. The best performing tests, under different types of microstructural noise, are the LBNS, the RBNS, the MinRV and the MedRV. The JO test also shows a good performance especially when the degree of discreteness in the data is not too extreme unlike the CPR test which tends to be robust in power also when data are

affected by rounding effects. Among the various tests the weaker performance is shown by the ABD-LM and the PZ test which is severely affected by the types of noise considered at lower frequencies. The power of the ABD-LM test tends to remain stable across different frequencies and increasing levels of noise but ranks lower than the other tests.

2.3.2.2 Co-Jump Tests I: re-sampled data

In order to compute the first battery of co-jump tests we consider data generated by the multivariate jump diffusion model specified in (2.17) and subsequently re-sampled at a frequency of 1, 5 and 10 time units. The log-returns, at different frequencies, are computed over the same time interval $[0, T]$.

We indicate with CJT_1 , CJT_2 and CJT_3 the contemporaneous co-jump test indicators measured at a frequency of 1, 5 and 10 time units. We indicate with $PCJT_1$, $PCJT_2$ and $PCJT_3$ the co-jump test indicators for permanent co-jumps observed respectively at a frequency of 1 and 5, 5 and 10 and 1 to 5 and 5 to 10 time units combined over the same time interval. We finally denote with $LCJT_1$, $LCJT_2$ and $LCJT_3$ the co-jump test indicator for lagged co-jumps observed respectively at a frequency of 1 to 5, 1 to 10, and 1 to 5 and 1 to 10 time units combined over the same time interval. The size and the power of the co-jump tests are evaluated at a 5% significance level.

SIZE. The top-left section of Table 2.8 reports the size of the co-jump test indicators CJT_1 , $PCJT_1$ and $LCJT_1$ under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and zero microstructural noise ($\varpi = 0.00$). The full set of results, for different observation frequencies, is available upon request. The best performing contemporaneous co-jump test, in terms of size, is the

PZ followed by the MinRV and the MedRV tests. The biggest size distortion is observed in the case of the RBNS, the ABD-LM and the CPR test where the average size is 6.8%. The size of the permanent co-jump tests are very close to the nominal size of 5.00% even we notice a slight undersize in the case of the PZ, the CPR and ABD-LM tests. The lagged co-jump tests show signs of oversize and in particular in the case of the MedRV, the LBNS and the ABD-LM tests. The best performing test is the CPR followed by the PZ test.

POWER. In order to study the power of the co-jump tests, we again consider the case of a varying jump intensity, correlation between the two stochastic processes and different types of microstructural noise.

Varying jump intensity. We consider, for the two correlated stochastic processes, a jump size that is normally distributed with mean zero and standard deviation equal to 0.10 and let the jump intensity λ vary. In particular, we let $\lambda dt = \{0.05\%, 0.25\%, 0.50\%, 2.50\%, 5.00\%\}$ and assume that the jump intensities of the two processes are the same, i.e. $\lambda_{(1)}dt = \lambda_{(2)}dt$. Table 2.8 reports the size corrected power of the co-jump test indicators CJT_1 , $PCJT_1$ and $LCJT_1$ under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and zero microstructural noise ($\varpi = 0.00$). The full set of results, for different observation frequencies, is available upon request. In line with the univariate case, we observe that the frequency of correctly identified co-jumps increases as the jump intensity factor raises. The best contemporaneous co-jump test in terms of power is the JO followed by the LBNS test. The corrected power for these tests is greater than 80% in the case where $\lambda_{(1)}dt = \lambda_{(2)}dt = 0.05\%$ (e.g. average number of 1 jump over the time interval

considered) and increases to an average of 99.9% when $\lambda_{(1)}dt = \lambda_{(2)}dt = 0.25\%$ (e.g. average number of 5 jumps over the time interval considered). The worst contemporaneous co-jump test in terms of power is the PZ, followed by the CPR and the MinRV. A similar ranking is also observed in the case of the permanent and lagged co-jump tests with the JO and the RBNS displaying the greatest power. The power of all the co-jump tests shows the tendency to converge to 1.00 as the jump intensity factor λdt moves to values equal to 0.50%. In particular, the best improvement, in terms of power, is shown by the PZ, the CPR, the MinRV and the ABD-LM tests.

[Insert Table 2.8]

Varying correlation factor. Table 2.9 reports the size corrected power of the co-jump test indicators CJT_1 , $PCJT_1$ and $LCJT_1$ under different levels of the correlation variable $\varrho_{(1,2)}$ and, in particular, when we allow $\varrho_{(1,2)}$ to vary from 0.00 to 0.95 under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$ and $\lambda dt = 0.05\%$. In the absence of correlation, we notice that the contemporaneous co-jump tests rank higher than the permanent and the lagged co-jump tests across the different test statistics used to construct the co-jump tests with the exception of the ABD-LM and the PZ tests. We also notice that, on average, all co-jump tests show the tendency to decrease in power when the correlation factor raises while the difference in power between contemporaneous, permanent and lagged is not as marked at higher levels of correlation compared to the case where correlation is zero. In the absence of correlation, the best performing test is the JO followed by the LBNS tests. A similar ranking is also observed at different correlation levels and across the different co-jump

tests. The worst performance is shown by the PZ test which shows an average corrected power of 67% compared to an average of 75% across the different co-jump tests.

Impact of Microstructural Noise. The top panel of Table 2.10 shows the size and the size corrected power of the co-jumps tests in the presence of an i.i.d. microstructural noise with mean zero and varying variance σ_n . In particular we allow two different noise regimes. In the first regime, σ_n assume values equal to 0.10 (low noise) while in the second regime $\sigma_n = 1.00$ (high noise). The first six columns of the Table report the size while the last six columns the size corrected power under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and $\lambda dt = \{0.00\%, 0.05\%\}$. We notice that the average size of the various co-jump tests tends to be much higher than the average size of the test in the absence of noise. This is also true when we compute the test under a regime characterized by low noise. In particular, we notice a big size distortion in the case of the contemporaneous co-jumps tests constructed from the LBNS, the MinRV and the MedRV tests. The size of the tests is very close to the nominal size of 5% in the case of the permanent co-jump tests even if we observe still a slight oversize in the case of the MinRV and the MedRV tests. Similar issues of oversize are also observed in the case of lagged co-jump tests and in particular when the PZ, the MinRV and the MedRV test are used. The size of the PZ test overshoots under a regime characterized by high noise. This result is in line with the univariate case where we also observed a very large size distortion from the PZ test for high levels of noise. The best performing test in terms of power and under a low noise regime is the JO test followed by the LBNS and the RBNS. The PZ and the CPR tests are instead characterized by a lower power especially in the case of high structural

noise. In particular, the lagged CPR test suffers from very little power together with the MedRV compared to the other tests.

The bottom panel of Table 2.10 shows the size and the size corrected power of the co-jump tests in the presence of noise caused by rounding effects. As in the previous case, we first introduce a regime where the noise is low and where we impose a rounding of $\text{rnd} = 3$ decimal places and, subsequently, introduce a regime where the noise is high and where we impose a rounding of $\text{rnd} = 1$ decimal place. The size and the size corrected power of the test is computed under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and $\lambda dt = \{0.00\%, 0.05\%\}$. As in the univariate case, rounding effects can affect the size of the co-jump tests as shown, in particular, when only 1 decimal place is used. We notice that the average size of the co-jump tests tends to be slightly higher in the case of low noise compared to the size of the test in the absence of noise. This is particularly relevant in the case of contemporaneous and lagged co-jump tests. The size of the LBNS, the RBNS, the MinRV and the CPR is severely affected by the rounding noise. Even if the size of the tests gets closer to the nominal level in the case of permanent co-jumps, we can still observe a high degree of oversize for the permanent LBNS and CPR tests and the lagged LBNS and the RBNS tests. The power of the co-jump tests seems to be less affected by noise. The best performance is shown by the contemporaneous JO test, followed by the LBNS and the MedRV tests. The power tends to decrease as we move from the contemporaneous to the lagged case. Also, with the exception of the LBNS, the RBNS and the CPR test we notice that the power of the various tests decreases slightly as we move from a low to a high noise regime.

[Insert Tables 2.9 - 2.10]

2.3.2.3 Co-Jump Tests II: tick-by-tick data

In order to compute the second battery of co-jump tests we generate tick-by-tick data and compute the log-returns over non overlapping time intervals of fixed size. The simulations are generated from the multivariate specification of the jump diffusion model described in (2.12). The size and the power of the co-jump tests are evaluated at a 5% significance level. In particular, we compute three co-jump test indicators in order to detect contemporaneous, permanent and exogenous co-jump events denoted with CJT , $PCJT$ and $ECJT$ respectively.

SIZE. The top-left section of Table 2.11 reports the size of the co-jump test indicators under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and zero microstructural noise ($\varpi = 0.00$). We notice a size distortion in the case of the contemporaneous co-jump tests which appears, on average, greater than the one observed with the first battery of tests. The oversize is particularly evident in the case of the contemporaneous PZ, the MinRV and the MedRV tests. The permanent co-jump tests are also affected by size issues with the LBNS, the RBNS, the CPR and the PZ tests displaying oversize while the JO and the ABD-LM tests being slightly undersized compared to the nominal size of the tests. The exogenous co-jump tests are also characterized by a higher size compared to the case of the first battery of tests. The biggest size issue can be observed in the case of the CPR and the PZ tests.

POWER. In order to study the power of the second battery of co-jump tests,

we again consider the case of a varying jump intensity, correlation between the two stochastic processes and different types of microstructural noise.

Varying jump intensity. Table 2.11 reports the size corrected power of the co-jump test indicators under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and zero microstructural noise ($\varpi = 0.00$) where $\lambda dt = \{0.05\%, 0.25\%, 0.50\%, 2.50\%, 5.00\%\}$. The rate of correctly identified co-jumps shows the tendency to increase as the jump intensity factor moves from 0.05% to 5.00%. In particular, we notice that, while the power of the second battery of contemporaneous co-jump tests is smaller than the power displayed by the first battery of tests, this is not the case for the permanent and the exogenous co-jump tests. The exogenous co-jump tests, computed from the second battery of tests, also show a greater power than the permanent co-jump tests. The power of tests converge to 99% – 100% as we move to $\lambda dt = 0.25\%$. The speed of convergence is slightly higher than in the case of the first battery of tests.

[Insert Tables 2.11]

Varying correlation factor. Table 2.12 shows the power sensitivity of the co-jump tests when we allow $\varrho_{(1,2)}$ to vary from 0.00 to 0.95 under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$ and $\lambda dt = 0.05\%$. The power of the various co-jump tests increases with higher correlation levels and also when we move from the contemporaneous to the exogenous case. The best contemporaneous co-jump test is the JO, followed by the LBNS and the RBNS tests. As in the previous case, the worst performance is shown by the PZ and the CPR tests.

Impact of Microstructural Noise. The top panel of Table 2.13 shows the size and the size corrected power of the co-jumps tests in the presence of an i.i.d. microstructural noise with mean zero and varying variance σ_n . In particular we allow two different noise regimes. In the first regime, σ_n assume values equal to 0.10 (low noise) while in the second regime $\sigma_n = 1.00$ (high noise). The first six columns of the Table report the size while the last six columns the size corrected power under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$, $\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and $\lambda dt = \{0.00\%, 0.05\%\}$. The size distortion, under a low noise regime, is not as severe as in the case of the first battery of tests. The size of the permanent co-jump tests is very close to the nominal value with the exception of the JO and the ABD-LM tests both affected by a lower size. The size of the exogenous co-jump test is very similar to the size of the contemporaneous co-jump tests especially under low levels of noise. Under a regime characterized by high noise, we observe a very big size distortion in the case of the PZ test, affected, as in previous case, by large oversize. The CPR and the JO tests are also affected by size issues. Finally, in terms of power, we do not detect any big impact moving from a low to a high noise regime. This confirms the results obtained with the first battery of tests.

The bottom panel of Table 2.13 shows the size and the size corrected power of the co-jump tests in the presence of noise caused by rounding effects. As in the previous case, we first introduce a regime where the noise is low and where we impose a rounding of $\text{rnd} = 3$ decimal places and, subsequently, introduce a regime where the noise is high and where we impose a rounding of $\text{rnd} = 1$ decimal place. The size and the size corrected power of the test is computed under the assumption of $\mu_{(1)} = \mu_{(2)} = \mu_{Y(1)} = \mu_{Y(2)} = 0$,

$\sigma_{(1)} = \sigma_{(2)} = \sigma_{Y(1)} = \sigma_{Y(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$ and $\lambda dt = \{0.00\%, 0.05\%\}$. Rounding is particularly relevant in this case as we observe a very severe size distortion with the LBNS, the RBNS, the MinRV and the CPR tests. Even if this result is in line with the findings obtained using the first battery of tests, the size distortion, in this particular case, seems more relevant. Finally, we notice a higher power, across the various tests, when we move from $rdn = 3.00$ to $rdn = 1.00$ decimal places. The jump in power is particularly visible in the case of the CPR and the MinRV permanent co-jump tests. The best performing test, under a low noise regime, is the JO test followed by the LBNS and the RBNS tests. The LBNS and the RBNS rank higher under a high noise regime followed by the CPR and the JO tests.

[Insert Tables 2.12 - 2.13]

2.4 Final Remarks

In Chapter 2, we considered different combinations of univariate tests for jumps and proposed a co-jump testing methodology in order to detect statistically significant common jumps between two correlated stochastic processes. In particular, we introduced a testing procedure in the case where the data are either re-sampled over equally spaced time intervals or observed at a tick-by-tick level. We also presented different tests to identify the presence of *contemporaneous*, *permanent* and *lagged* or *exogenous* co-jump events.

A Monte Carlo experiment was also presented to first assess the statistical properties of the univariate tests for jumps and, subsequently, to study the statistical properties of the proposed co-jump testing procedure under different levels of the jump intensity fac-

tor, jump size and microstructural noise. We found a strong sensitivity of the proposed co-jump testing procedure to the jump intensity variable λ and, in particular, we found that the rate of correctly identified co-jumps tends to increase as the jump intensity factor raises. While we found very little sensitivity of the first set of co-jump tests to changes in the correlation factor, the size corrected power of the second set of co-jump tests was positively affected by an increase in correlation. We also observed a big size distortion of the proposed co-jump testing procedure under different types of microstructural noise. In particular, we found that the noise caused by rounding effects can severely affect the size of the tests as shown in the case of the LBNS, the RBNS, the MinRV and the CPR tests. We also noticed that the proposed co-jump testing procedure is robust to different levels of noise as the power of the tests is not particularly affected when we move from a low to a high noise regime. Overall, the strongest performance, in terms of power, was displayed by the LBNS, the RBNS and the JO followed by the MedRV tests while the PZ and the CPR tests were the most affected by microstructural noise.

In the next chapter, we present an empirical application of the proposed co-jump testing procedure. In particular, we use ultra high frequency data observed from the EUR/USD FX spot market to relate liquidity shocks to EUR/USD FX spot price jumps during the week from May 3 to May 7, 2010.

Table 2.1: Formulae for Jumps and Co-jump test indicators

The table reports the formulae used to compute the jump and co-jump test indicators. In particular, the top panel shows the test indicators used when the data are re-sampled over equally spaced time intervals at an observation frequency of m and n time units. The bottom panel shows instead the test indicators used when the data are observed at a tick-by-tick level over two non-overlapping time intervals.

	Code	Formulae
Re-sampled data		
Contemporaneous co-jump	CJT	$\prod_{i=1}^d I(\varphi(j)_{i,m[0,T]})$
Permanent co-jump	PCJT	$\prod_{z=m}^n \left\{ \prod_{i=1}^d I(\varphi(j)_{i,z[0,T]}) \right\}$
Lagged co-jump	LCJT	$\max \left\{ 0, \left[\prod_{i=1}^d I(\varphi_{i,z(i)[0,T]}) \right]_{z=\{m_1, \dots, m_n\}} - \left[\prod_{i=1}^d I(\varphi_{i,z[0,T]}) \right]_{z=m, \dots, n} \right\}$
Tick-by-tick data		
Contemporaneous co-jump	CJT	$\prod_{i=1}^d I(\varphi(j)_{i[T-t, T+t]})$
Permanent co-jump	PCJT	$\prod_{z=T_t}^{T_t+1} \left\{ \prod_{i=1}^d I(\varphi(j)_{i,z}) \right\}$
Exogenous co-jump	ECJT	$\max \left\{ 0, \prod_{i=1}^d I(\varphi_{i,[T+t]}) - \left[\prod_{z=T_t}^{T_t+1} I(\varphi_{i,z}) \right]_{i=1, \dots, d} \right\}$

Table 2.2: Parameter Values for the Monte Carlo Simulations

The Table reports the values of the parameters of the data generating process used in the simulation exercise. We report the values for the correlated stochastic processes, $S_{(1)}$ and $S_{(2)}$, the jump processes, $J_{(1)}$ and $J_{(2)}$ and the microstructural noise denoted as ϖ . We initially simulate $m = 2,000$ returns, we then re-sample the simulated data over intervals of $dt = 1, 5$ and 10 time units and finally re-run the simulation $n = 1,000$ times. In particular, we denote with r_1 , r_5 and r_{10} the simulated returns at a frequency of one, five and ten time units for the stochastic process $S_{(i)}$ for $i = 1, 2$. The simulated time series consists of $r_1 = 2,000$, $r_5 = 400$ and $r_{10} = 200$ returns for each simulation.

Parameter	$S_{(1)}$	$S_{(2)}$	$J_{(1)}$	$J_{(2)}$	ϖ
nsim	1,000	1,000	-	-	-
nobs	2,000	2,000	-	-	-
dt	1/2,000	1/2,000	-	-	-
$\mu_{(i)}$	0.00	0.00	0.00	0.00	0.00
$\sigma_{(i)}$	0.10	0.10	0.10 - 1.25	0.10 - 1.25	0.10 - 1.25
$\varrho_{(i,j)}$	0.00 - 0.95	0.00 - 0.95	0.00 - 0.95	0.00 - 0.95	-
$\lambda_{(i)}$	-	-	0.00 - 5.00%	0.00 - 5.00%	-
$S_{(i),t=0}$	100	100	-	-	-

Table 2.3: Size and power of the univariate tests for jumps for varying jump intensity

The Table reports the size and the size corrected power of the univariate tests for jumps for a varying jump intensity factor. In particular, we estimate the linear and ratio Barndorff-Nielsen and Shephard (2005) tests (LBNS and RBNS respectively), the Andersen et al. (2012) MinRV and MedRV tests (MinRV and MedRV respectively), the Jiang and Oomen (2008) test (JO), the Andersen, Bollerslev and Dobrev (2007, ABD) and Lee and Mykland (2008, LM) tests (ABD-LM), the Corsi et al. (2010) test (CPR) and the Podolskij and Ziggel (2010) test (PZ). The test $JT1$, $JT2$ and $JT3$ are computed using simulated data re-sampled over a frequency of 1, 5 and 10 time units respectively. We report the results using a 5% significance level.

Procedure	JT1	JT2	JT3	JT1	JT2	JT3	JT1	JT2	JT3
$\lambda dt :$	0.00%	0.00%	0.00%	0.05%	0.05%	0.05%	0.25%	0.25%	0.25%
LBNS	0.054	0.056	0.048	0.564	0.528	0.494	0.986	0.968	0.960
RBNS	0.052	0.060	0.046	0.566	0.528	0.482	0.986	0.968	0.954
MinRV	0.058	0.060	0.048	0.542	0.532	0.462	0.970	0.942	0.906
MedRV	0.048	0.040	0.042	0.552	0.524	0.490	0.972	0.942	0.932
JO	0.046	0.042	0.048	0.574	0.552	0.504	0.984	0.966	0.934
ABD-LM	0.024	0.024	0.024	0.514	0.478	0.448	0.968	0.956	0.942
CPR	0.050	0.056	0.046	0.484	0.458	0.424	0.964	0.940	0.916
PZ	0.056	0.062	0.058	0.464	0.400	0.356	0.910	0.818	0.784
$\lambda dt :$	0.50%	0.50%	0.50%	2.50%	2.50%	2.50%	5.00%	5.00%	5.00%
LBNS	1.00	1.00	0.99	1.000	1.000	1.000	1.000	1.000	1.000
RBNS	1.00	1.00	0.99	1.000	1.000	1.000	1.000	1.000	1.000
MinRV	0.99	0.93	0.88	0.852	0.862	0.742	0.848	0.972	0.870
MedRV	0.99	0.95	0.94	0.972	0.994	0.976	0.998	1.000	0.954
JO	0.99	0.96	0.94	0.960	0.830	0.742	0.930	0.674	0.456
ABD-LM	1.00	1.00	1.00	1.000	1.000	1.000	1.000	1.000	0.998
CPR	1.00	0.99	0.99	1.000	1.000	1.000	1.000	1.000	1.000
PZ	0.95	0.90	0.86	0.964	0.904	0.894	0.964	0.918	0.922
	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10

Table 2.4: Power of the univariate tests for jumps for varying jump size

The Table reports the size corrected power of the univariate tests for jumps for a varying jump volatility. In particular, we estimate the linear and ratio Barndorff-Nielsen and Shephard (2005) tests (LBNS and RBNS respectively), the Andersen et al. (2012) MinRV and MedRV tests (MinRV and MedRV respectively), the Jiang and Oomen (2008) test (JO), the Andersen, Bollerslev and Dobrev (2007, ABD) and Lee and Mykland (2008, LM) tests (ABD-LM), the Corsi et al. (2010) test (CPR) and the Podolskij and Ziggel (2010) test (PZ). The test JT_1 , JT_2 and JT_3 are computed using simulated data re-sampled over a frequency of 1, 5 and 10 time units respectively. We report the results using a 5% significance level.

Procedure	JT1	JT2	JT3	JT1	JT2	JT3	JT1	JT2	JT3
$\sigma_{Y(1)} :$	0.10	0.10	0.10	0.25	0.25	0.25	0.50	0.50	0.50
LBNS	0.986	0.968	0.960	0.994	0.992	0.990	0.994	0.992	0.994
RBNS	0.986	0.968	0.954	0.994	0.992	0.990	0.994	0.992	0.994
MinRV	0.970	0.942	0.906	0.980	0.954	0.914	0.976	0.954	0.910
MedRV	0.972	0.942	0.932	0.984	0.950	0.922	0.976	0.942	0.870
JO	0.984	0.966	0.934	0.992	0.986	0.962	0.994	0.988	0.980
ABD-LM	0.968	0.956	0.942	0.980	0.980	0.978	0.976	0.974	0.968
CPR	0.964	0.940	0.916	0.980	0.974	0.972	0.976	0.972	0.970
PZ	0.910	0.818	0.784	0.984	0.960	0.950	0.988	0.986	0.984
$\sigma_{Y(1)} :$	0.75	0.75	0.75	1.00	1.00	1.00	1.25	1.25	1.25
LBNS	0.996	0.996	0.996	0.994	0.994	0.994	0.990	0.990	0.992
RBNS	0.996	0.996	0.996	0.994	0.994	0.994	0.990	0.990	0.992
MinRV	0.986	0.954	0.898	0.980	0.934	0.900	0.982	0.934	0.886
MedRV	0.980	0.932	0.876	0.976	0.934	0.880	0.980	0.918	0.862
JO	0.996	0.990	0.984	0.994	0.992	0.990	0.990	0.986	0.986
ABD-LM	0.984	0.982	0.980	0.984	0.984	0.984	0.974	0.976	0.972
CPR	0.984	0.984	0.980	0.984	0.986	0.984	0.972	0.972	0.974
PZ	0.996	0.992	0.996	0.996	0.992	0.992	0.990	0.990	0.990
	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10

Table 2.5: Size of the univariate tests for jumps for varying microstructural noise

The Table reports the size of the univariate tests for jumps in the presence of an i.i.d. microstructural noise with varying variance. In particular, we estimate the linear and ratio Barndorff-Nielsen and Shephard (2005) tests (LBNS and RBNS respectively), the Andersen et al. (2012) MinRV and MedRV tests (MinRV and MedRV respectively), the Jiang and Oomen (2008) test (JO), the Andersen, Bollerslev and Dobrev (2007, ABD) and Lee and Mykland (2008, LM) tests (ABD-LM), the Corsi et al. (2010) test (CPR) and the Podolskij and Ziggel (2010) test (PZ). The test JT_1 , JT_2 and JT_3 are computed using simulated data re-sampled over a frequency of 1, 5 and 10 time units respectively. We report the results using a 5% significance level.

Procedure	JT1	JT2	JT3	JT1	JT2	JT3	JT1	JT2	JT3
$\sigma_n :$	0.10	0.10	0.10	0.25	0.25	0.25	0.50	0.50	0.50
LBNS	0.038	0.062	0.054	0.038	0.044	0.056	0.052	0.068	0.044
RBNS	0.036	0.056	0.050	0.042	0.048	0.058	0.046	0.060	0.032
MinRV	0.038	0.062	0.054	0.022	0.072	0.066	0.048	0.064	0.048
MedRV	0.042	0.044	0.048	0.024	0.058	0.050	0.042	0.052	0.062
JO	0.026	0.048	0.038	0.028	0.026	0.030	0.024	0.036	0.036
ABD-LM	0.040	0.032	0.020	0.022	0.018	0.024	0.024	0.040	0.020
CPR	0.024	0.042	0.060	0.034	0.048	0.046	0.034	0.058	0.044
PZ	0.052	0.026	0.052	0.056	0.060	0.066	0.088	0.066	0.104
$\sigma_n :$	0.75	0.75	0.75	1.00	1.00	1.00	1.25	1.25	1.25
LBNS	0.048	0.060	0.056	0.042	0.056	0.048	0.026	0.064	0.042
RBNS	0.042	0.056	0.062	0.042	0.058	0.048	0.024	0.058	0.048
MinRV	0.032	0.076	0.046	0.030	0.074	0.048	0.026	0.076	0.052
MedRV	0.048	0.064	0.040	0.050	0.062	0.048	0.024	0.062	0.042
JO	0.040	0.034	0.050	0.028	0.046	0.036	0.028	0.036	0.042
ABD-LM	0.042	0.016	0.020	0.022	0.036	0.022	0.028	0.012	0.018
CPR	0.048	0.052	0.068	0.034	0.072	0.054	0.022	0.054	0.044
PZ	0.106	0.098	0.152	0.144	0.156	0.178	0.202	0.200	0.236
	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10

Table 2.6: Power of the univariate tests for jumps for varying microstructural noise

The Table reports the size corrected power of the univariate tests for jumps in the presence of an i.i.d. microstructural noise with varying variance. In particular, we estimate the linear and ratio Barndorff-Nielsen and Shephard (2005) tests (LBNS and RBNS respectively), the Andersen et al. (2012) MinRV and MedRV tests (MinRV and MedRV respectively), the Jiang and Oomen (2008) test (JO), the Andersen, Bollerslev and Dobrev (2007, ABD) and Lee and Mykland (2008, LM) tests (ABD-LM), the Corsi et al. (2010) test (CPR) and the Podolskij and Ziggel (2010) test (PZ). The test JT_1 , JT_2 and JT_3 are computed using simulated data re-sampled over a frequency of 1, 5 and 10 time units respectively. We report the results using a 5% significance level.

Procedure	JT1	JT2	JT3	JT1	JT2	JT3	JT1	JT2	JT3
$\sigma_n :$	0.10	0.10	0.10	0.25	0.25	0.25	0.50	0.50	0.50
LBNS	0.992	0.978	0.962	0.986	0.974	0.972	0.976	0.970	0.958
RBNS	0.992	0.978	0.962	0.986	0.974	0.970	0.976	0.966	0.958
MinRV	0.984	0.960	0.920	0.984	0.950	0.920	0.970	0.946	0.920
MedRV	0.986	0.966	0.926	0.984	0.952	0.930	0.976	0.948	0.910
JO	0.986	0.966	0.910	0.978	0.940	0.908	0.976	0.948	0.922
ABD-LM	0.982	0.974	0.962	0.980	0.962	0.944	0.962	0.946	0.928
CPR	0.982	0.960	0.944	0.970	0.938	0.934	0.948	0.938	0.918
PZ	0.916	0.844	0.774	0.908	0.786	0.796	0.896	0.858	0.752
$\sigma_n :$	0.75	0.75	0.75	1.00	1.00	1.00	1.25	1.25	1.25
LBNS	0.986	0.968	0.944	0.982	0.960	0.940	0.962	0.940	0.908
RBNS	0.986	0.964	0.942	0.982	0.958	0.940	0.962	0.936	0.904
MinRV	0.980	0.960	0.912	0.980	0.952	0.896	0.958	0.924	0.856
MedRV	0.980	0.950	0.912	0.984	0.954	0.930	0.964	0.930	0.878
JO	0.974	0.950	0.904	0.972	0.942	0.902	0.964	0.904	0.842
ABD-LM	0.964	0.946	0.918	0.966	0.932	0.902	0.960	0.944	0.898
CPR	0.952	0.938	0.904	0.948	0.908	0.886	0.940	0.894	0.854
PZ	0.912	0.808	0.794	0.910	0.808	0.778	0.886	0.798	0.770
	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10

Table 2.7: Size and power of the univariate tests for jumps for varying rounding noise

The Table reports the size (top panel) and the size corrected power (bottom power) of the univariate tests for jumps in the presence of noise caused by rounding effects. In particular, we estimate the linear and ratio Barndorff-Nielsen and Shephard (2005) tests (LBNS and RBNS respectively), the Andersen et al. (2012) MinRV and MedRV tests (MinRV and MedRV respectively), the Jiang and Oomen (2008) test (JO), the Andersen, Bollerslev and Dobrev (2007, ABD) and Lee and Mykland (2008, LM) tests (ABD-LM), the Corsi et al. (2010) test (CPR) and the Podolskij and Ziggel (2010) test (PZ). The test JT_1 , JT_2 and JT_3 are computed using simulated data re-sampled over a frequency of 1, 5 and 10 time units respectively. We report the results using a 5% significance level.

Procedure	JT1	JT2	JT3	JT1	JT2	JT3	JT1	JT2	JT3
rnd.	3.00	3.00	3.00	2.00	2.00	2.00	1.00	1.00	1.00
LBNS	0.068	0.048	0.072	0.048	0.054	0.066	0.508	0.088	0.066
RBNS	0.062	0.050	0.054	0.052	0.062	0.058	0.496	0.082	0.058
MinRV	0.060	0.064	0.050	0.048	0.052	0.058	0.166	0.086	0.066
MedRV	0.058	0.060	0.038	0.062	0.052	0.074	0.056	0.064	0.062
JO	0.036	0.034	0.040	0.052	0.038	0.040	0.038	0.046	0.040
ABD-LM	0.016	0.022	0.026	0.018	0.022	0.044	0.038	0.038	0.032
CPR	0.066	0.036	0.058	0.056	0.034	0.060	0.428	0.058	0.046
PZ	0.058	0.072	0.058	0.068	0.052	0.056	0.082	0.072	0.070
rnd.	3.00	3.00	3.00	2.00	2.00	2.00	1.00	1.00	1.00
LBNS	0.966	0.960	0.944	0.994	0.978	0.968	1.000	0.978	0.968
RBNS	0.966	0.958	0.940	0.992	0.978	0.970	1.000	0.978	0.964
MinRV	0.958	0.928	0.902	0.992	0.964	0.936	0.974	0.958	0.930
MedRV	0.954	0.936	0.906	0.990	0.972	0.946	0.982	0.962	0.942
JO	0.972	0.938	0.902	0.988	0.950	0.918	0.982	0.966	0.934
ABD-LM	0.958	0.944	0.926	0.988	0.974	0.966	0.978	0.968	0.948
CPR	0.946	0.930	0.920	0.984	0.960	0.940	0.990	0.942	0.934
PZ	0.898	0.818	0.776	0.924	0.840	0.798	0.926	0.804	0.784
	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10	dt: 1	dt: 5	dt: 10

Table 2.8: Size and power of the first battery of co-jump tests for varying jump intensity

The Table reports the size and the size corrected power of the first battery of co-jump tests for a varying jump intensity factor. The first battery of co-jump tests is computed using re-sampled data over the time interval $[0, T]$. The various tests are based on different combinations of univariate tests for jumps. We report the results using a 5% significance level.

Procedure	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1
λ_{dt} :	0.00%	0.00%	0.00%	0.05%	0.05%	0.05%	0.25%	0.25%	0.25%
LBNS	0.066	0.048	0.070	0.812	0.82	0.800	0.998	0.998	0.996
RBNS	0.068	0.052	0.062	0.798	0.816	0.778	0.998	0.996	0.994
MinRV	0.062	0.046	0.068	0.784	0.794	0.722	0.990	0.994	0.994
MedRV	0.062	0.054	0.082	0.800	0.82	0.674	1.000	1.000	0.986
JO	0.064	0.056	0.064	0.844	0.838	0.836	1.000	0.998	1.000
ABD-LM	0.068	0.046	0.070	0.788	0.788	0.778	0.998	0.998	0.996
CPR	0.068	0.042	0.054	0.742	0.738	0.658	0.996	0.998	0.976
PZ	0.054	0.04	0.060	0.682	0.68	0.624	0.978	0.990	0.980
λ_{dt} :	0.50%	0.50%	0.50%	2.50%	2.50%	2.50%	5.00%	5.00%	5.00%
LBNS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RBNS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MinRV	1.000	1.000	1.000	0.998	1.000	0.946	1.000	1.000	0.938
MedRV	0.998	1.000	1.000	1.000	1.000	0.980	1.000	1.000	0.998
JO	0.998	1.000	1.000	0.998	1.000	0.998	0.992	0.984	0.992
ABD-LM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CPR	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000
PZ	0.990	0.998	0.986	0.994	0.998	0.998	0.992	1.000	0.996

Table 2.9: Power of the first battery of co-jump tests for varying correlation

The Table reports the size corrected power of the first battery of co-jump tests for varying correlation factor $\varrho_{(1,2)}$. The first battery of co-jump tests is computed using re-sampled data over the time interval $[0, T]$. The various tests are based on different combinations of univariate tests for jumps. We report the results using a 5% significance level.

Procedure	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1
$\varrho_{(1,2)} :$	0.00%	0.00%	0.00%	0.50%	0.50%	0.50%	0.75%	0.75%	0.75%	0.95%	0.95%	0.95%
LBNS	0.832	0.806	0.796	0.812	0.82	0.800	0.792	0.778	0.762	0.788	0.778	0.778
RBNS	0.802	0.730	0.780	0.798	0.816	0.778	0.774	0.030	0.754	0.778	0.030	0.758
MinRV	0.808	0.788	0.740	0.784	0.794	0.722	0.764	0.758	0.704	0.770	0.760	0.716
MedRV	0.806	0.780	0.658	0.800	0.82	0.674	0.772	0.772	0.656	0.774	0.774	0.654
JO	0.852	0.834	0.852	0.844	0.838	0.836	0.804	0.792	0.808	0.810	0.798	0.790
ABD-LM	0.800	0.812	0.798	0.788	0.788	0.778	0.746	0.742	0.736	0.776	0.774	0.756
CPR	0.768	0.718	0.614	0.742	0.738	0.658	0.694	0.670	0.588	0.736	0.704	0.622
PZ	0.688	0.692	0.686	0.682	0.68	0.624	0.664	0.622	0.624	0.668	0.664	0.670

Table 2.10: Size and power of the first battery of co-jump tests for varying microstructural noise

The Table reports the size and the size corrected power of the first battery of co-jump tests for varying microstructural noise ϖ . We first add an i.i.d. noise with varying variance (top panel) and we then study the impact of rounding on the size and the power of the various tests (bottom panel). The first six columns of the Table show the size of the co-jump tests for different values of microstructural noise while the last six columns the size corrected power in the presence of jumps. The first battery of co-jump tests is computed using re-sampled data over the time interval $[0, T]$. The various tests are based on different combinations of univariate tests for jumps. We report the results using a 5% significance level.

Procedure	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1	CJT1	PCJT1	LCJT1
$\sigma_n :$	0.10%	0.10%	0.10%	1.00%	1.00%	1.00%	0.10%	0.10%	0.10%	1.00%	1.00%	1.00%
LBNS	0.102	0.052	0.074	0.078	0.034	0.082	0.852	0.824	0.814	0.740	0.714	0.708
RBNS	0.056	0.030	0.058	0.038	0.030	0.060	0.816	0.030	0.784	0.704	0.030	0.668
MinRV	0.110	0.060	0.110	0.068	0.040	0.086	0.808	0.816	0.744	0.736	0.692	0.656
MedRV	0.114	0.064	0.080	0.066	0.040	0.068	0.810	0.812	0.680	0.732	0.710	0.546
JO	0.064	0.058	0.060	0.076	0.054	0.082	0.860	0.844	0.838	0.770	0.746	0.766
ABD-LM	0.046	0.050	0.048	0.052	0.076	0.044	0.770	0.786	0.758	0.742	0.732	0.712
CPR	0.096	0.032	0.074	0.076	0.034	0.070	0.740	0.708	0.630	0.690	0.640	0.540
PZ	0.092	0.040	0.116	0.230	0.172	0.264	0.710	0.696	0.682	0.706	0.676	0.676
rnd	3.00	3.00	3.00	1.00	1.00	1.00	3.00	3.00	3.00	1.00	1.00	1.00
LBNS	0.096	0.038	0.074	0.728	0.330	0.488	0.812	0.782	0.782	0.932	0.840	0.888
RBNS	0.040	0.030	0.042	0.324	0.030	0.224	0.770	0.030	0.756	0.836	0.030	0.804
MinRV	0.104	0.040	0.088	0.278	0.112	0.196	0.784	0.760	0.710	0.786	0.746	0.724
MedRV	0.076	0.038	0.048	0.096	0.058	0.060	0.814	0.778	0.636	0.768	0.766	0.646
JO	0.062	0.056	0.070	0.036	0.044	0.052	0.838	0.820	0.824	0.804	0.774	0.790
ABD-LM	0.070	0.074	0.042	0.072	0.074	0.064	0.786	0.796	0.780	0.758	0.762	0.752
CPR	0.080	0.032	0.056	0.636	0.252	0.124	0.762	0.736	0.632	0.884	0.764	0.686
PZ	0.068	0.034	0.104	0.100	0.046	0.128	0.676	0.652	0.654	0.664	0.642	0.630

Table 2.11: Size and power of the second battery of co-jump tests for varying jump intensity

The Table reports the size and the size corrected power of the second battery of co-jump tests for a varying jump intensity factor. The second battery of co-jump tests is computed using tick-by-tick data over two non overlapping time intervals. The various tests are based on different combinations of univariate tests for jumps. We report the results using a 5% significance level.

Procedure	CJT	PCJT	ECJT	CJT	PCJT	ECJT	CJT	PCJT	ECJT
$\lambda_{dt} :$	0.00%	0.00%	0.00%	0.05%	0.05%	0.05%	0.25%	0.25%	0.25%
LBNS	0.064	0.062	0.068	0.754	0.876	0.930	1.000	1.000	1.000
RBNS	0.062	0.062	0.064	0.738	0.838	0.910	1.000	1.000	1.000
MinRV	0.070	0.056	0.066	0.718	0.832	0.880	0.996	1.000	0.998
MedRV	0.070	0.056	0.062	0.758	0.886	0.898	0.998	1.000	1.000
JO	0.062	0.030	0.052	0.806	0.934	0.962	1.000	1.000	1.000
ABD-LM	0.068	0.046	0.054	0.778	0.904	0.940	1.000	1.000	1.000
CPR	0.062	0.068	0.070	0.678	0.800	0.888	0.990	1.000	1.000
PZ	0.074	0.060	0.070	0.660	0.764	0.816	0.984	1.000	0.996
$\lambda_{dt} :$	0.50%	0.50%	0.50%	2.50%	2.50%	2.50%	5.00%	5.00%	5.00%
LBNS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RBNS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MinRV	1.000	1.000	1.000	1.000	1.000	0.974	1.000	1.000	0.954
MedRV	1.000	1.000	0.998	1.000	1.000	0.986	1.000	1.000	0.998
JO	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000	0.998
ABD-LM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CPR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PZ	0.994	1.000	0.998	0.996	0.998	1.000	0.998	1.000	1.000

Table 2.12: Power of the second battery of co-jump tests for varying correlation

The Table reports the size corrected power of the second battery of co-jump tests for varying correlation factor $\varrho_{(1,2)}$. The second battery of co-jump tests is computed using tick-by-tick data over two non overlapping time intervals. The various tests are based on different combinations of univariate tests for jumps. We report the results using a 5% significance level.

Procedure	CJT	PCJT	ECJT	CJT	PCJT	ECJT	CJT	PCJT	ECJT	CJT	PCJT	ECJT
$\varrho_{(1,2)} :$	0.00%	0.00%	0.00%	0.50%	0.50%	0.50%	0.75%	0.75%	0.75%	0.95%	0.95%	0.95%
LBNS	0.756	0.900	0.920	0.754	0.876	0.930	0.772	0.896	0.928	0.780	0.906	0.944
RBNS	0.750	0.886	0.900	0.738	0.838	0.910	0.764	0.862	0.908	0.766	0.892	0.934
MinRV	0.734	0.852	0.882	0.718	0.832	0.880	0.716	0.838	0.874	0.724	0.856	0.918
MedRV	0.754	0.866	0.900	0.758	0.886	0.898	0.754	0.884	0.912	0.746	0.910	0.924
JO	0.796	0.922	0.946	0.806	0.934	0.962	0.810	0.918	0.950	0.788	0.934	0.954
ABD-LM	0.764	0.892	0.912	0.778	0.904	0.940	0.748	0.902	0.938	0.758	0.904	0.922
CPR	0.680	0.836	0.860	0.678	0.800	0.888	0.694	0.810	0.858	0.698	0.820	0.872
PZ	0.648	0.749	0.786	0.660	0.764	0.816	0.634	0.739	0.798	0.636	0.784	0.836

Table 2.13: Size and power of the second battery of co-jump tests for varying microstructural noise

The Table reports the size and the size corrected power of the second battery of co-jump tests for varying microstructural noise ϖ . We first add an i.i.d. noise with varying variance (top panel) and we then study the impact of rounding on the size and the power of the various tests (bottom panel). The first six columns of the Table show the size of the co-jump tests for different values of microstructural noise while the last six columns the size corrected power in the presence of jumps. The second battery of co-jump tests is computed using tick-by-tick data over two non overlapping time intervals. The various tests are based on different combinations of univariate tests for jumps. We report the results using a 5% significance level.

Procedure	CJT	PCJT	ECJT	CJT	PCJT	ECJT	CJT	PCJT	ECJT	CJT	PCJT	ECJT
$\sigma_n :$	0.10%	0.10%	0.10%	1.00%	1.00%	1.00%	0.10%	0.10%	0.10%	1.00%	1.00%	1.00%
LBNS	0.068	0.062	0.074	0.058	0.054	0.066	0.766	0.908	0.934	0.746	0.888	0.916
RBNS	0.064	0.060	0.072	0.058	0.058	0.066	0.756	0.882	0.916	0.730	0.872	0.894
MinRV	0.066	0.066	0.084	0.040	0.026	0.050	0.710	0.854	0.892	0.700	0.828	0.868
MedRV	0.076	0.070	0.068	0.046	0.044	0.046	0.726	0.860	0.908	0.722	0.858	0.884
JO	0.052	0.028	0.054	0.062	0.036	0.052	0.774	0.928	0.952	0.794	0.906	0.936
ABD-LM	0.068	0.036	0.058	0.062	0.066	0.050	0.742	0.864	0.924	0.744	0.874	0.910
CPR	0.060	0.054	0.068	0.038	0.030	0.056	0.694	0.820	0.876	0.646	0.786	0.864
PZ	0.062	0.050	0.066	0.154	0.178	0.200	0.610	0.758	0.788	0.686	0.780	0.836
rnd	3.00	3.00	3.00	1.00	1.00	1.00	3.00	3.00	3.00	1.00	1.00	1.00
LBNS	0.058	0.052	0.070	0.658	0.754	0.579	0.792	0.930	0.938	0.914	0.980	0.974
RBNS	0.060	0.054	0.072	0.644	0.741	0.547	0.784	0.914	0.930	0.910	0.974	0.962
MinRV	0.046	0.048	0.048	0.234	0.261	0.158	0.740	0.858	0.880	0.792	0.928	0.910
MedRV	0.058	0.054	0.048	0.096	0.080	0.074	0.772	0.892	0.906	0.764	0.896	0.910
JO	0.066	0.044	0.058	0.056	0.032	0.040	0.834	0.946	0.960	0.796	0.938	0.950
ABD-LM	0.070	0.054	0.068	0.086	0.056	0.054	0.762	0.890	0.924	0.770	0.898	0.930
CPR	0.064	0.044	0.068	0.554	0.655	0.477	0.720	0.830	0.886	0.866	0.974	0.942
PZ	0.064	0.042	0.072	0.066	0.060	0.056	0.650	0.774	0.826	0.668	0.796	0.824

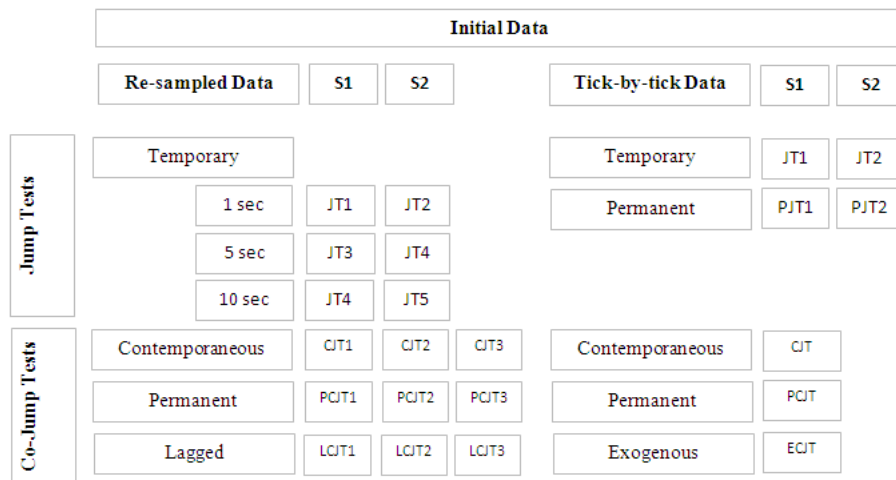


Figure 2.1: Jump and Co-Jump Tests for re-sampled and tick-by-tick data.

The Figure shows the different jump and co-jump tests presented in this chapter for re-sampled (left panel) and tick-by-tick (right panel) data for two correlated stochastic processes, $S_{(1)}$ and $S_{(2)}$.

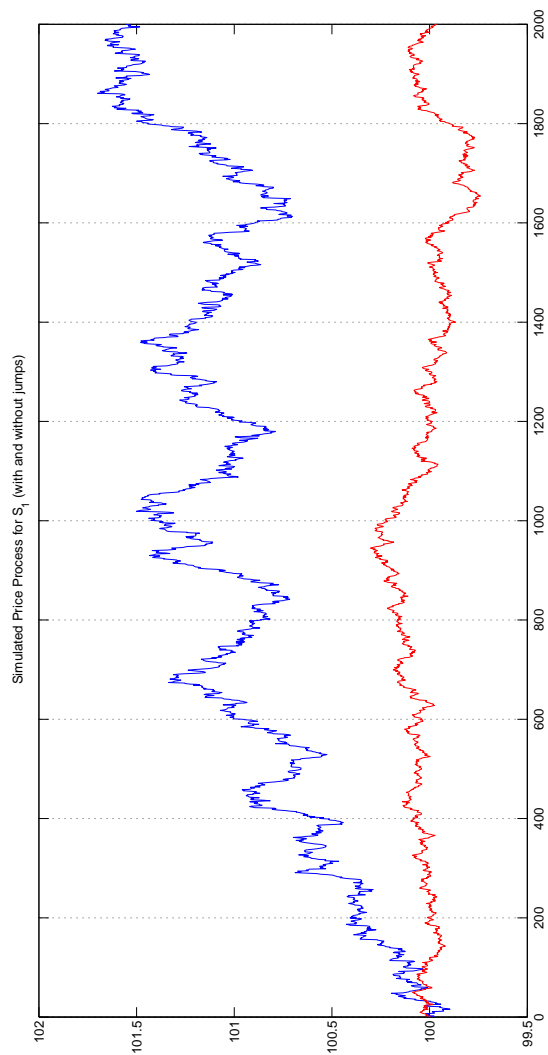


Figure 2.2: Simulated Prices for underlying process $S_{(1)}$ with and without Jumps.

The Figure shows the simulated price series for the underlying process $S_{(1)}$ under the assumption of jumps (blue line) and no-jumps (red line). The initial parameters used in the simulation are: $S_{(1),0} = 100$, $\mu_{(1)} = \mu_{J(1)} = 0.00$, $\sigma_{(1)} = \sigma_{J(1)} = 0.10$, $\lambda dt_{(1)} = 0.25\%$, and $\varpi = 0.00$.

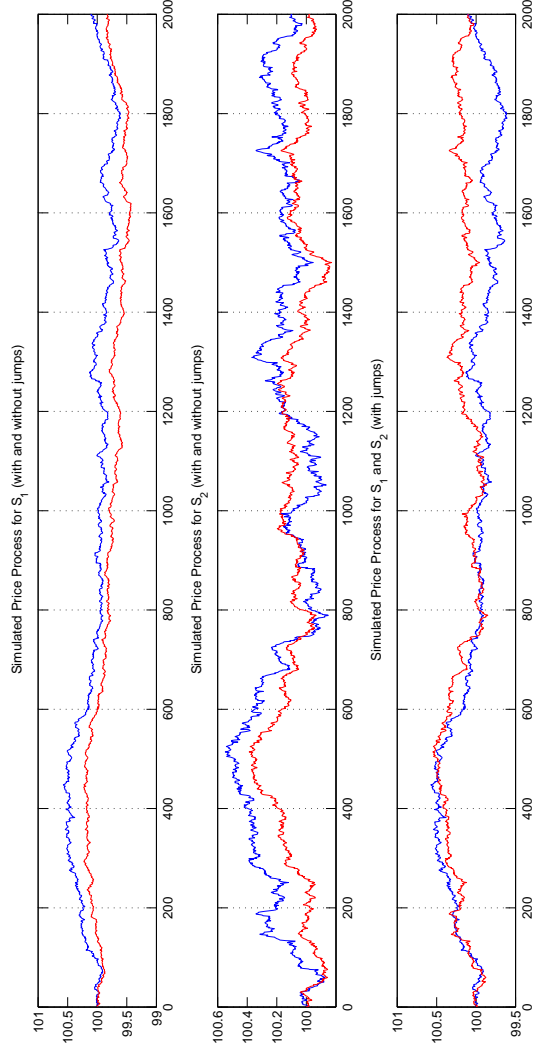


Figure 2.3: Simulated Prices for underlying process $S_{(1)}$ and $S_{(2)}$ with and without Jumps.

The Figures show the simulated price series for the underlying process $S_{(1)}$ (top panel), $S_{(2)}$ (mid panel) under the assumption of jumps (blue line) and no-jumps (red line) together with $S_{(2)}$ (red) under the assumption of non-zero jumps in the bottom panel. The initial parameters used in the simulation are: $S_{(1),0} = 100$, $\mu_{(1)} = \mu_{J(1)} = \mu_{(2)} = \mu_{J(2)} = 0.00$, $\sigma_{(1)} = \sigma_{J(1)} = \sigma_{(2)} = \sigma_{J(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$, $\lambda dt_{(1)} = \lambda dt_{(2)} = 0.05\%$, and $\varpi = 0.00$. The jumps intensities are not correlated.

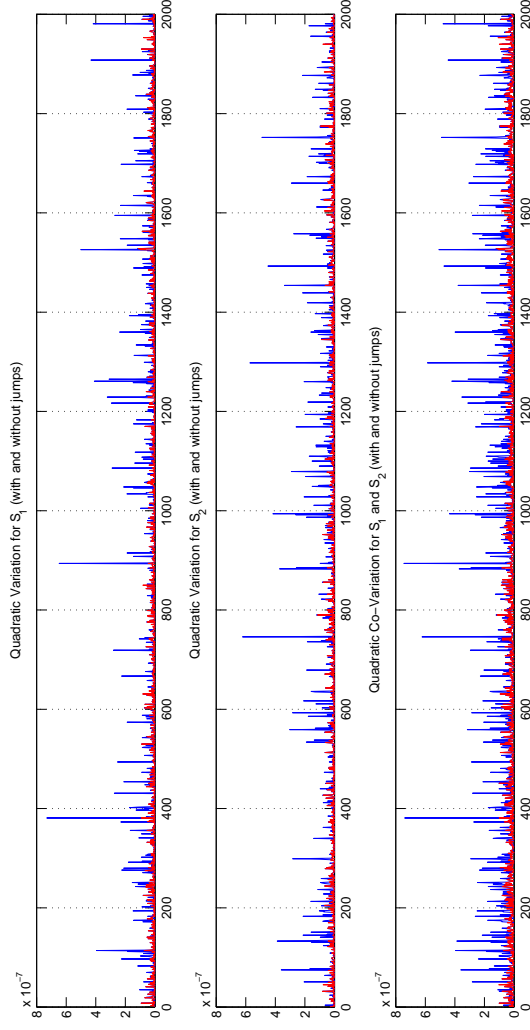


Figure 2.4: Quadratic Variation and Covariation of Two Stochastic Price Processes $S_{(1)}$ and $S_{(2)}$.

The Figures show the quadratic variation (top and middle panel) and the quadratic covariation (bottom panel) for two correlated stochastic processes $S_{(1)}$ and $S_{(2)}$ under the assumption of jumps (blue line) and no-jumps (red line). The initial parameters used in the simulation are: $S_{(1),0} = 100$, $\mu_{(1)} = \mu_{J(1)} = \mu_{(2)} = \mu_{J(2)} = 0.00$, $\sigma_{(1)} = \sigma_{J(1)} = \sigma_{(2)} = \sigma_{J(2)} = 0.10$, $\varrho_{(1,2)} = 0.50$, $\lambda dt_{(1)} = \lambda dt_{(2)} = 0.05\%$, and $\varpi = 0.00$. The jumps intensities are not correlated.

2.5 Appendix: Simulation Design - Matlab Code

In order to model multivariate Poisson random variables we use the algorithm presented by Yahav and Shmueli (2011). In particular:

1. We use a Cholesky decomposition of the variance-covariance matrix of a p -dimensional vector of price returns to simulate multivariate normal random vectors. The vector of multivariate normal random variables is denoted with X^N .
2. For each value of $X_{(i)}^N$ where $i \in 1, 2, \dots, p$ we calculate the Normal CDF $\Phi(X_{(i)}^N)$.
3. For each $\Phi(X_{(i)}^N)$ we calculate the Poisson inverse CDF (quantile) with a rate $\lambda_{(i)}$:

$$X_{(i)}^P = \Xi^{-1} \left\{ \Phi \left(X_{(i)}^N \right) \right\}.$$

The vector $X_{(i)}^P$ is a p -dimensional Poisson vector with correlation matrix R^P and rates Λ . The algorithm allows us to transform the Normal marginals of the vector X^N using the Normal CDF and obtain a p -dimensional vector with uniform marginals. The uniform variates are then subsequently transformed into Poisson variates using the Poisson inverse CDF.

We show below the Matlab code used to simulate the path of two correlated stochastic processes where the Poisson processes used to model the price jumps are also correlated.

```
m1 = 5; m2 = 10;
nsim = 1000; nsteps = 2000; dt = 1/nsteps;
drift1 = 0.00; drift2 = 0.00;
drift1J = 0.00; drift2J = 0.00;
```

```

vol1 = 0.10; vol2 = 0.10;
volJ1 = 0.10; volJ2 = 0.10;
rho = 0.50;
sigma = [vol1^2 vol1*vol2*rho; vol1*vol2*rho vol2^2];
lam1 = 0.00; lam2 = 0.00;
S01 = 100; S02 = 100;

for i = 1:nsim
    mnoise = normrnd(0,voln,2,nsteps);
    Rnd12 = randn(2,nsteps)+mnoise;
    VolMa = (ChoDe'*Rnd12);
    PhiVolMa = normcdf(VolMa);
    if lam1 > 0
        P = poissinv(PhiVolMa,lam1);
    elseif lam1 == 0
        P = zeros(2,nsteps);
    end
    jumpnb = poissrnd(P*dt);
    jump = normrnd(driftJ1*(jumpnb-lam1*dt),volJ1.*sqrt(jumpnb));
    M1(i,:) = (drift1+driftJ1*lam1-1/2*vol1.^2)*dt+sqrt(dt)*vol1*Rnd12(1,:)+jump(1,:);
    M2(i,:) = (drift2+driftJ2*lam2-1/2*vol2.^2)*dt+sqrt(dt)*vol2*Rnd12(2,:)+jump(2,:);
    M1nj(i,:) = (drift1-1/2*vol1.^2)*dt+sqrt(dt)*vol1*Rnd12(1,:);
    M2nj(i,:) = (drift2-1/2*vol2.^2)*dt+sqrt(dt)*vol2*Rnd12(2,:);
end

U1(1:nsim,1) = S01; U2(1:nsim,1) = S02;
U1nj(1:nsim,1) = S01; U2nj(1:nsim,1) = S02;

for i = 1:nsteps
    U1(1:nsim,i+1)=U1(1:nsim,i).*exp(M1(1:nsim,i));

```

```
U2(1:nsim,i+1)=U2(1:nsim,i).*exp(M2(1:nsim,i));  
U1nj(1:nsim,i+1)=U1nj(1:nsim,i).*exp(M1nj(1:nsim,i));  
U2nj(1:nsim,i+1)=U2nj(1:nsim,i).*exp(M2nj(1:nsim,i));  
end
```


CHAPTER 3

THE LIQUIDITY-PRICE TRANSMISSION MECHANISM

3.1 Introduction

The relationship between asset prices and liquidity has been analyzed, tested and widely accepted by academics, practitioners and most recently by regulators. The transmission mechanism that links changes in liquidity to the price evolution of an underlying asset has become even more relevant during times characterized by high volatility and market distress. Morris and Shin (2004) define the notion of a liquidity black hole occurring when a fall in prices is followed by an increase in the liquidation of the underlying asset, which, in the absence of fresh liquidity, can generate an additional adverse move in the price. A similar analysis is provided by Adalid and Detken (2007) who again relate shocks to liquidity to collapses in the asset prices. As discussed in Herring et al. (2008) a change in the perception of liquidity can cause a severe market correction even if the underlying price for that particular asset does not diverge from its fundamental values. The value of liquidity is affected by time, location and most importantly by price of the underlying asset as shown in Sparrow and Ilijanic (2010) who also quantify liquidity as a function of transactional costs and risk preferences. A liquidity crisis can also affect

the ability of market to properly function and, in some extreme cases, to survive. In particular, Allen et al. (2011) and Calem et al. (2011) show that the absence of liquidity associated with a financial shock can deeply affect the lending capacity of banks.

The analysis and the composition of market liquidity plays an important role also in price discovery as already discussed in Chapter 1. In their respective models, Kyle (1985) shows that private information is channeled through market liquidity, while Hasbrouck (1991), Engle and Russell (1998) and De Jong and Schotman (2010) relate price dynamics to different measures of liquidity. In particular, Kyle's sequential equilibrium model describes the behavior of noise, informed, and market making trading agents as a function of their access to liquidity. Market liquidity is defined and measured in terms of tightness, resiliency and depth. Depth and resiliency are also considered endogenous variables and signal the presence of insider and noise traders in the market. A greater number of noise traders have the tendency to increase available liquidity, while a negative correlation is observed between the level of private information and market depth.

Despite the relevance of liquidity shocks, jumps are often studied in relation to prices. Most of the literature on jumps and the empirical applications only focus on the discontinuous path of prices. Also, in the common notion of a co-jump, two or more variables are characterized by a discontinuous path over the same observation frequency but the jumps are both exogenous. In our empirical application, we combine the testing methodology based on combinations of univariate tests for jumps, measured either at different observation frequencies or over non-overlapping time intervals, to study the price impact of liquidity shocks. In particular, we relate jumps in liquidity to the

dynamics of the underlying asset price and establish a causality effect between liquidity and price. Moreover, the jump and co-jump testing procedure, presented Chapter 2, is used to identify a number of different jump and co-jump events and, in particular, we distinguish between contemporaneous, permanent and lagged or exogenous co-jumps.

The chapter is organized as follows. Section 3.2 introduces the data used in the empirical application. In Section 3.3 we present the results of the co-jump testing procedure when data are first re-sampled over fixed time intervals and, subsequently, observed at a tick-by-tick level. Section 3.4 concludes.

3.2 The Structure of the Data

The data used in the empirical application consist of five time series: the available liquidity on the bid and the ask side of the EUR/USD FX spot limit order book, expressed in EUR millions; the bid and ask price associated to the available liquidity in the order book, expressed as units of US Dollars for 1 EUR; quotation time, expressed in milliseconds. The data span over an entire trading week, from May 3 to May 7, 2010. The time frame considered is particularly interesting given that, during this time horizon, EUR/USD FX spot displayed a large move, opening on Monday at around 1.3250 and closing the session on Friday just above 1.2750. Despite the short time interval, in our empirical application, we use data observed at a very high frequency to obtain a large sample. From the raw data, we cannot recognize the identity and the characteristics of the market participants in that particular security.

The available liquidity in the order book is recorded at three different price levels: the available liquidity on the bid (ask) side at the best displayed bid (ask) price in the

order book represents the total amount of buy (sell) limit orders with the highest priority in terms of execution. We denote this particular measure of liquidity as $qx_{(1),t,bid(ask)}$. The available liquidity on the bid (ask) side of the order book is also observed at different price levels, and, in particular, at $k = 1$ and $k = 2$ ticks away from the best displayed price. We denote these measures of liquidity as $qx_{(2),t,bid(ask)}$ and $qx_{(3),t,bid(ask)}$ respectively. The total amount of available liquidity on the bid (ask) side of the order book and at $k = 2$ ticks from the best displayed price is denoted as $Qx_{t,bid(ask)} = \sum_{i=1}^3 qx_{(i),t,bid(ask)}$ and is referred to market depth throughout the chapter.

The bid (ask) price is the observed price associated to the available liquidity in the order book. In particular we denote with $px_{(1),t,bid(ask)}$, $px_{(2),t,bid(ask)}$ and $px_{(3),t,bid(ask)}$ the price associated to the available liquidity at $k = 0, 1$ and 2 ticks from the best displayed price on the buy (sell) side of the order book. As discussed in Chapter 1, the bid and the ask side of the order book seem to behave under different rules. For this particular reason, given that the focus of this chapter is to analyze the behavior of liquidity and price jumps rather than the dynamic structure of the limit order book, we will use a price variable computed as the mid-quote of the best bid and the best ask price in the order book. We denote the variable as $Px_t = (px_{(1),t,bid} + px_{(1),t,ask})/2$. Figure 3.1 shows the normalized EUR/USD FX mid-quote price behavior. In order to preserve data confidentiality, we show the price series in index form. The vertical dotted line shows the end of the NY trading session and the open of the Australian trading session.

Finally quotation time is defined as the time, expressed in milliseconds, between consecutive quotes on either side of the order book. Quotation time is denoted as trx_t .

Similar to the liquidity variable introduced by Frey (2000) and Esser and Moench (2005), denoted in this chapter as $\rho_{(1),t,bid(ask)} = 1 / \{Qx_{t,bid(ask)} * Px_t\}$, we introduce a second liquidity variable $\rho_{(2)}$ which is function of market depth and quotation time, $\rho_{(2),t,bid(ask)} = 1 / \{Qx_{t,bid(ask)} * trx_t\}$. Both the liquidity variables $\rho_{(1)}$ and $\rho_{(2)}$ are inverse functions of available liquidity. In particular, the liquidity variable $\rho_{(2)}$ increases when the available liquidity or the market depth on the bid (ask) side of the order book decreases or equivalently when the time between consecutive quotes decreases.

Figure 3.2 shows the behavior of cumulative available liquidity on the bid side of the order book and quotation time. We notice a common pattern between liquidity and time. Quotation time tends to spike during the first hours of the Australian trading session and the final hours of the NY trading session. The trading regime at those times is characterized by low transactional volatility and little liquidity. From the top section of Figure 3.2 we notice that, during those times, the cumulative available liquidity, observed from the order book, tends to be small compared to the central hours of the trading day. This is particularly evident during the 4th and the 5th of May where we also observe a few sudden spikes. The liquidity variable $\rho_{(2)}$ aims to capture this relationship between time and liquidity. Quotation time is used here as a proxy of transactional volatility. As in the case of the liquidity variable $\rho_{(1)}$, also the liquidity variable $\rho_{(2)}$ is expected to decrease when market transactions become more frequent, or equivalently, when the transactional time decreases or transactional volatility increases. Figure 3.3 shows the behavior of the liquidity variable $\rho_{(2)}$. We notice that the liquidity variable shows indeed the tendency to spike as we approach the

close of the NY trading session where we typically observe low transactional volatility and little liquidity.

[Insert Figures 3.1 - 3.3]

The relationship between price and the liquidity variable $\rho_{(2)}$ can also be studied further by looking at Figure 3.4 which shows the pattern of the normalized price series for the EUR/USD FX spot mid-quote price and the liquidity variable $\rho_{(2)}$ on May 6, 2010. This day appears particularly volatile when we look at the price path of EUR/USD FX over the entire trading week. Not surprisingly, we notice that large moves in prices are observed together with sudden spikes in the liquidity variable and close to the end of the trading session and in particular around 18:00 London time. Spikes in the liquidity variable are associated to periods of low volatility and little liquidity which again could explain the erratic behavior of the EUR/USD FX spot price during these times.

[Insert Figures 3.4]

3.2.1 Preparation of the Data

Two different methodologies are used to manage the data, initially observed at a tick-by-tick level over irregular time intervals. We first re-sample the data over equally spaced time intervals. We normalize the series and pre-average the re-sampled data in order to reduce the impact of micro-structural noise, subsequently divide the time series in blocks of equal size and, finally, select a fixed number of observations per block. The re-sampling methodology is used in Dumitru and Urga (2012) in order to compute

jump tests at different observation frequencies. The first testing procedure, presented in Chapter 2 and based on the combination of univariate tests for jumps across different frequencies, will be used on the first data-set.

We then use the raw tick-by-tick data, we again normalize the series, pre-average and finally divide the sample in blocks of equal size. The time series will not be re-sampled and we will benefit from the full information content of the original data-set. Tick-by-tick data are traditionally used in the jump and co-jump literature. The second testing procedure presented in Chapter 2 will be used on the second data set.

Re-sample over fixed intervals. The data is initially observed at a tick-by-tick level and subsequently re-sampled over a frequency of 1, 5 and 10 seconds. The re-sampling of the data allows us to reduce the impact of excessive discreteness in the time series and to better handle issues related to computational complexity. Furthermore, in order to preserve data confidentiality, we represent the three microstructural variables in index form rather than as actual levels. In particular, the time series is normalized such that the average liquidity, price and transactional time over the whole sample is set at 100. A similar convention was adopted in Chaboud et al. (2004) and Berger et al. (2008).

The continuously compounded returns are constructed from the natural logarithms of the time series. In particular, we denote the liquidity and price returns with $rLRI$ and $rPRI$ respectively while we use $rLV1$ and $rLV2$ to indicate the returns of the liquidity variables, $\rho_{(1)}$ and $\rho_{(2)}$, introduced in the previous section. We measure the returns using different observation frequencies of 1, 5 and 10 seconds.

Given the structure of our data-set, where the price and the associated liquidity from the order book are observed at the same time, we find no evidence of the so called

Epps-effect caused by non-synchronous trading and which allows covariance estimates to converge to zero as the observation frequency increases. We find instead a high degree of microstructural noise especially in the liquidity process and at a frequency of 1 second. Pre-averaging is commonly performed to reduce the impact of the noise on the various jump and co-jump tests traditionally used in the literature (see Jacod et al. 2009, Christensen et al. 2010, Hautsch and Podolskij 2010, Mykland and Zhang 2011). We perform pre-averaging on the normalized time series using different combinations of windows and sampling frequencies. In particular we denote the pre-averaging window with kp and use a window size of $kp = 30, 60, 90$ and 120 observations. We compute the auto-correlation function for lags from 0 to 20 and plot the correspondent correlogram. Figures 3.5 - 3.8 show the auto-correlation plot (ACF) for the continuously compounded logarithmic returns, computed over a 1 second re-sampling interval, of cumulative available liquidity $Qx_{t,bid}$, mid-quote price Px_t and the two liquidity variables $\rho_{(1)}$ and $\rho_{(2)}$ defined in the previous section for different values of the pre-averaging window size. The ACF for the normalized liquidity and price returns shows very little signs of auto-correlation for lags greater than 1. Similar results can also be observed with the normalized returns of the liquidity variable $\rho_{(2)}$ despite a slightly greater degree of auto-correlation especially at lags equal to 1 and 2. The liquidity variable $\rho_{(1)}$ instead seems to be affected by a higher degree of auto-correlation which is particularly severe when the pre-averaging window is set at 30 observations.

[Insert Figures 3.5 - 3.8]

Tables 3.1 - 3.4 report the auto-correlations at different lags, re-sampling frequencies and pre-averaging window size again for the compounded logarithmic returns of

cumulative available liquidity, mid-quote price and the two liquidity variables. The upper and lower bounds for auto-correlation have been computed at a significance level of $\alpha = 5\%$. From the Tables we highlight two important results. First, the lower the re-sampling frequency and the lower the auto-correlation detected in the series. The bold mark denotes auto-correlation values higher (lower) than the upper (lower) confidence bound, i.e. when the null hypothesis of no auto-correlation at and beyond a given lag is rejected. Across Tables 3.1 - 3.4 we can see that the rate of rejection of the null hypothesis decreases as we decrease the re-sampling frequency from 1 to 10 seconds. This result shows that the degree of microstructural noise is lower when the time series is sampled at a lower frequency or, equivalently, that a higher pre-averaging window is needed at higher sampling frequencies. Finally we also notice the optimal choice of the pre-averaging windows is a function of the re-sampling frequency. In particular we observe that when data are sampled over a frequency of 10 seconds an averaging window size of 30 observations allows us to dramatically reduce the degree of auto-correlation at lags greater than 1. The auto-correlation completely disappears with a window size of 90 to 120 observations. We find that a window size of at least 60 observations is needed instead when the data are re-sampled over a frequency of 5 seconds with virtually no sign of auto-correlation being detected when 90 to 120 observations are used in the pre-averaging exercise. Finally, when the data are re-sampled over a frequency of 1 second, at least 90 observations are needed to reduce the impact of auto-correlation.

[Insert Tables 3.1 - 3.4]

Tick-by-tick data set. The data used here is directly observed from the original time series and not re-sampled over a fixed observation frequency. As in the previous case,

where re-sampled data are used, we perform pre-averaging on the normalized time series using different combinations of observation windows. In particular we denote the pre-averaging window with kp and use a window size of $kp = 30, 60, 90$ and 120 observations. We compute the auto-correlation function for lags from 0 to 20 and plot the correspondent correlogram. Figures 3.9 - 3.12 show the auto-correlation plot (ACF) for the continuously compounded logarithmic returns of cumulative available liquidity $Qx_{t,bid}$, mid-quote price Px_t and the two liquidity variables $\rho_{(1)}$ and $\rho_{(2)}$ defined in the previous section for different values of the pre-averaging window size. From the ACF we notice that the liquidity, the price returns and the second liquidity variable $\rho_{(2)}$ seem to be well behaved also at a pre-averaging window of $kp = 30$ observations. We observe, instead, a high degree of auto-correlation at a number of lags in the first liquidity variable, $\rho_{(1)}$, when $kp = 30$ or 60 .

[Insert Figures 3.9 - 3.12]

Tables 3.5 - 3.6 report the auto-correlations at different lags and pre-averaging window size again for the compounded logarithmic returns of cumulative available liquidity, mid-quote price and the two liquidity variables. The upper and lower bounds for auto-correlation have been computed at a significance level of $\alpha = 5\%$. The bold mark denotes auto-correlation values higher (lower) than the upper (lower) confidence bound, i.e. when the null hypothesis of no auto-correlation at and beyond a given lag is rejected. From Tables 3.5 - 3.6 we see that the rate of rejection of the null hypothesis decreases as we increase the pre-averaging window and notice that the optimal choice of the window seems to be equal to $kp = 90$ observations when tick-by-tick data are used.

[Insert Tables 3.5 - 3.6]

As a final remark, and also given the virtual continuity of the currency markets as opposed to equity and fixed income markets, we do not discard the data observed at the open and close of a trading session but we do eliminate the very last observations in proximity of the NY close on the Friday due to a reporting error. The ability to recognize different trading regimes and patterns during the 24 hours trading day is critically important and highlights some interesting features of the data as discussed in the previous section.

3.3 Estimation and Results

3.3.1 Re-Sampled Data Set

In this section we first present the methodology used to test for the presence of jumps and co-jumps between liquidity and prices, when data are re-sampled over equally spaced intervals, and finally discuss the estimation results. The co-jump testing procedure is based on the combination of univariate tests for jumps introduced in Chapter 2.

Figure 3.13 shows an illustrative example of the data re-sampling exercise performed for the first day of the week. In particular, the data is initially observed at a tick-by-tick level and subsequently re-sampled over a frequency of 1, 5 and 10 seconds. In order to limit the impact of micro-structural noise, the data is firstly normalized so that the average over the whole sample is set at 100 and then pre-averaged using a window of $kp = 300, 60$ and 30 observations for data re-sampled at a frequency of 1, 5 and 10

seconds respectively. As shown in the previous section, the optimal selection for data re-sampled at a frequency of 5 and 10 seconds would be $kp = 60$ and 30 respectively. In order to have the same number of returns across frequencies we then select a window period of $kp = 300$ for data re-sampled at a frequency of 1 second. The re-sampled and pre-averaged data set consists of $n = 1331$ observations for the whole week from May 3rd to May 7th, 2010. In particular we have $n = 288$ re-sampled and pre-averaged observations for the first four days and $n = 179$ observations during the last day of the week. We subsequently divide the time series in blocks of equal size and select a number of $m = 28$ observations per block. The returns of the mid-quote price and the liquidity variables and the respective jump tests are then computed for each block.

[Insert Figure 3.13]

Test computation. In order to identify a jump or a co-jump event in our series, we use the jump and co-jump test indicator functions discussed in Chapter 2. The test indicator functions assign a value equal to one if a jump or a co-jump is detected over a block of size m or, alternatively, a value equal to zero if no jump or co-jump is identified at a significance level $\alpha = 5\%$. In particular, $JT_{(i)}$ represents a vector of signal variables (e.g. zeros and ones) which indicate the presence of a jump at each block for asset i . The size of the vector $JT_{(i)}$ will be equal to the number of blocks used. The proportion of identified jumps in both liquidity and price is given by the arithmetic average, across the different blocks, of the elements of the vector $JT_{(i)}$. We denote with $JT_{(1)}$ and $JT_{(2)}$ the test indicator vectors for the liquidity and the price processes respectively. Similarly, CJT represents a co-jump test indicator vector of

signal variables which indicate the presence of a co-jump at each block. In particular, we denote with CJT_1 the contemporaneous co-jump test indicator vector for liquidity and price. The contemporaneous co-jump test indicator vector is given by the product of the jump vectors $JT_{(1)}$ and $JT_{(2)}$. In the case where k observation frequencies are used, we define with $JT_{(1),k}$ and $JT_{(2),k}$ the jump indicator vectors for liquidity and price measured at a frequency of k seconds. In our empirical application we let $k = 1, 5$ and 10 seconds respectively. We also define as CJT_1, CJT_2 and CJT_3 the *contemporaneous co-jump* test indicator vector between liquidity and price at a frequency of $1, 5$ and 10 seconds respectively. We define as $PCJT_1$ and $PCJT_2$ the *permanent co-jump* test indicator vectors between liquidity and price at a frequency of $1-5$ and $5-10$ seconds while $PCJT_3$ measures the permanent co-jump at either a frequency of $1-5$ or $5-10$ seconds. We finally define as $LCJT_1$ and $LCJT_2$ the *lagged co-jump* test indicator vectors between the liquidity variable measured at a frequency of 1 second and the mid-price variable measured at 5 and 10 seconds respectively. The test $LCJT_3$ measures the lagged co-jump observed at a frequency of $1-5$ and $1-10$ seconds combined.

Analysis of the results. Tables 3.7 and 3.8 report the percentage of identified jumps at a significance level $\alpha = 5\%$ during the week from May 3rd to May 7th, 2010. In particular, the jump indicator vectors $JT_{(1),1}, JT_{(1),5}$ and $JT_{(1),10}$ provide a measure of liquidity jumps for the liquidity variables defined as $\rho_{(1)}$ and $\rho_{(2)}$ at a frequency of $1, 5$ and 10 seconds respectively, while the jump indicator vectors $JT_{(2),1}, JT_{(2),5}$ and $JT_{(2),10}$ provide a measure of mid-quote price jumps again at a frequency of $1, 5$ and 10 seconds respectively for $m = 28$ observations per block. The full set of results for different levels of m is available upon request.

We immediately notice, from Table 3.7, that the percentage number of identified jumps in both liquidity and price tends to decrease as we move from a high to a low observation frequency. Also, with the exception of the MinRV and the MedRV tests we also detect a lower number of liquidity than price jumps across all the observation frequencies used. In particular, we observe an average percentage of liquidity jumps equal to 30%, 13% and 10% at frequencies 1, 5 and 10 seconds compared to an average percentage of price jumps equal to 36%, 22% and 17%. The very high number of jumps detected by the MinRV, the MedRV, the CPR and the PZ test may be driven by the level of microstructural noise. In particular, the PZ test reports a suspiciously high number of jump events which we believe may be spurious. This result is in line with the findings from the simulation exercise from Chapter 2 where the PZ was shown to be the most affected univariate test in the presence of noise. The best performing test in terms of power is the LBNS, followed by the RBNS, the JO and the ABD-LM tests. Similar results would have been obtained using different levels of m .

Similar results can be observed when the liquidity variable $\rho_{(2)}$ is used in the analysis. Table 3.8 shows that the average number of liquidity jumps is lower than the average number of price jumps at all frequencies and for all the test methodologies used with the exception of the PZ test. Again, as in the previous case, we notice a big percentage of jumps detected by the PZ and the CPR tests. The best performance is shown by the LBNS, the RBNS, the JO and the ABD-LM tests. The JO test, however, together with the MedRV test fails to detect any sign of liquidity jumps at an observation frequency of ten seconds. Similar results would have been obtained using different levels of m .

[Insert Tables 3.7 - 3.8]

Table 3.9 reports the results of the first battery of co-jump tests when the first liquidity variable $\rho_{(1)}$ is used. We immediately notice that the average percentage of detected contemporaneous co-jumps between liquidity and prices is greater than the average percentage of permanent and the lagged co-jump events. We also notice that the percentage of contemporaneous co-jumps shows the tendency to decrease for higher observation frequencies. We observe an average percentage of contemporaneous co-jumps equal to 38% when the observation frequency is set at 1 second, and 20% and 15% when the observation frequencies are set at 5 and 10 seconds respectively. The MinRV, the MedRV, the CPR and the PZ tests show a high number of contemporaneous co-jump events at a frequency of 1 second. The result is not surprising given the high number of liquidity jumps identified by these tests. We believe however that a good portion of the identified contemporaneous co-jumps events is spurious and driven by noise. The best performing tests are again the LBNS, the RBNS, the JO and the ABD-LM. However, the LBNS and the RBNS tests fail to detect any sign of contemporaneous co-jump at a frequency of 10 seconds together with the MedRV test. We also find that most of the contemporaneous co-jumps detected are also permanent especially at a frequency of 1 and 5 seconds. The highest number of permanent co-jumps is identified at a frequency of 1 second by the MinRV test followed by the CPR and the MedRV tests. Finally, with the exception of the MinRV, the CPR and the PZ test, we find very little evidence of lagged co-jumps in particular at frequencies greater than 1-5 seconds. The result shows that jumps in liquidity observed at a frequency of 1 second are usually not followed by statistically significant price jumps at lower frequencies.

Table 3.10 reports the results of the second battery of co-jump tests when the second liquidity variable $\rho_{(2)}$ is used. We still observe a higher number of contemporaneous co-jumps at a frequency of 1 second with the PZ, the MedRV and the CPR still affected by power issues. On average, we find that the number of detected co-jumps is higher when the second liquidity variable is used and, in particular, we observe a higher percentage of contemporaneous co-jumps at a frequency of 5 seconds. The difference is evident when we look at the PZ, the LBNS, the RBNS and the MinRV tests. With the exception of the PZ test, we observe a slightly lower percentage of contemporaneous co-jumps at a frequency of 10 seconds in the case of $\rho_{(2)}$ compared to $\rho_{(1)}$. Most of the contemporaneous co-jumps are also found to be permanent especially at a frequency of 1 and 5 seconds. Finally, we again find very small number of lagged co-jumps with a lower average percentage of co-jumps detected when $\rho_{(1)}$ is used. In terms of performance, we find a similar ranking to the previous case, with the LBNS, the RBNS and the ABD-LM displaying the best power followed by the JO test.

[Insert Tables 3.9 - 3.10]

3.3.2 Tick-by-Tick Data Set

The tick-by-tick data are firstly normalized so that the average over the whole sample is set at 100 and then pre-averaged using a window of $kp = 90$ observations. The pre-averaged data set consists of $n = 13,593$ observations for the whole week from May 3rd to May 7th, 2010. We subsequently divide the time series in blocks of equal size and initially select a number of $m = 100$ observations per block. The returns of the mid-quote price and the liquidity variables and the respective jump and co-jump tests

are then computed for each block. The methodology used is very similar to the case where the data are re-sampled over equally space time intervals.

Test computation. A jump event is identified through the univariate tests for jumps introduced in Chapter 2. In particular we use a jump test indicator and assign a value equal to one (zero) if the null of no jump is rejected (accepted) at a significance level $\alpha = 5\%$. We use the vector $JT_{(i)}$ for $i = 1, 2$ to signal the presence of a jump at each block and, in particular, denote with $JT_{(1)}$ and $JT_{(2)}$ the jump test indicator vectors for liquidity and price respectively. As in the previous case, the proportion of identified jumps is given by the arithmetic average, across the different blocks, of the elements of the vector $JT_{(i)}$. When the data are observed at a tick-by-tick level we are able to identify different types of jump and co-jump events. In particular, we identify a *temporary jump* when we observe only one jump over two consecutive blocks. A temporary jump is also called *exogenous* as no temporal causality can be established between consecutive jumps. We also identify a *permanent jump* when we observe, over two consecutive blocks, a jump in either liquidity or mid-price. In this context, we also say that a permanent jump is *endogenous* as the state of a jump in one block is likely to be influenced or caused by the state of a jump in a consecutive block. We use the vector PJT to signal the presence of a permanent jump at each block. We use $PJT_{(1)}$ and $PJT_{(2)}$ to denote the permanent jump for liquidity and price respectively. A value equal to one (zero) is assigned to the vector when the null of no permanent jump is rejected (accepted). The percentage of identified permanent jumps is given by the arithmetic average, across the different blocks, of the elements of the vector PJT . We identify a *permanent co-jump* when the intersection of two *permanent jump* events

for two different processes yields a non-zero result. We finally identify an *endogenous co-jump* event when two consecutive endogenous jumps are observed or, alternatively, when one endogenous jump occurs together with an exogenous jump. In this context, a permanent co-jump will always also be endogenous. Similarly, we identify an *exogenous co-jump* when two consecutive exogenous jumps are observed. An exogenous co-jump will always be contemporaneous but the opposite may not be true. The vectors $PCJT$ and $ECJT$ are used to signal the presence of a permanent and an exogenous co-jump event respectively at each block. Figure 3.14 shows an illustrative example of the different jump and co-jump test constructions when tick-by-tick data are used. The first section of the Figure shows the presence of contemporaneous co-jumps denoted with CJT . The second section shows instead the presence of permanent jumps and in particular highlights a permanent jump in liquidity at block 5. The third section of the Figure shows the presence of a permanent and endogenous co-jump at block 5 while the bottom section of the Figure highlights the difference between endogenous and exogenous co-jumps. As an example, looking at block number 3, we identify two contemporaneous jump events with $JT_{(1),3} = JT_{(2),3} = 1$. None of the two jump events is permanent or endogenous as we find no sign of jumps in block 2. The co-jump is considered in this case both contemporaneous and exogenous. Moving to blocks number 5 and 6, we observe a contemporaneous jump in the liquidity process, i.e. $JT_{(1),6} = 1$, and a permanent or endogenous jump in the price process, i.e. $PJT_{(2),6} = 1$. In this case, we identify a contemporaneous co-jump in block 6, but we cannot consider the co-jump exogenous under the definition provided in Chapter 2.

[Insert Figure 3.14]

Analysis of the results. Tables 3.11 - 3.14 report the results of the jump and co-jump tests for the two liquidity variables used, $\rho_{(1)}$ and $\rho_{(2)}$ respectively, and the mid-quote price at a significance level $\alpha = 5\%$ and under different pre-averaging assumptions. In particular, Table 3.11 shows the percentage of identified jumps and co-jumps when $m = 100$, kp varies from $kp = 30$ to 120 in step of 30 observations and $\rho_{(1),t}$ is used as a liquidity variable. $JT1$ and $JT2$ indicate the contemporaneous jump test indicator vectors for liquidity and price, while CJT , $PCJT$ and $ECJT$ provide a measure of the number of identified contemporaneous, permanent and exogenous co-jumps respectively. With the exception of the MinRV, the MedRV and the JO tests we notice a greater number of contemporaneous jumps in the liquidity process. The difference in jump frequency between liquidity and price is particularly evident in the case of the PZ, the CPR and the ABD-LM tests. The greater dispersion in the number of identified liquidity jumps, when $\rho_{(1)}$ is used, can be driven by the high level of microstructural noise. The number of contemporaneous price jumps tends to converge at a faster rate when a higher pre-averaging window is used. This result reinforces the importance of pre-averaging as a way to reduce the level of microstructural noise which could be affect the power of the jump and co-jump tests. The PZ and the ABD-LM tests appear to be the most affected by the presence of noise while the LBNS, the RBNS and the JO tests display a robust behavior also when the pre-averaging window is short. With the exception again of the ABD-LM and the PZ, we find that the number of identified contemporaneous co-jumps tends to converge across the various testing procedures also under a smaller number of observations used in the pre-averaging window. We also observe that the contemporaneous CPR test displays little power especially at

lower levels of kp . Finally, the permanent and the exogenous co-jump tests display a similar behavior with an average percentage of identified co-jumps equal to 20%. We believe that the PZ procedure detects a higher number of spurious co-jumps due to the presence of microstructural noise. This result is in line with the findings from the Monte Carlo simulation presented in Chapter 2.

Table 3.12 reports the percentage of identified jumps and co-jumps when $kp = 90$ and m varies from 20 to 100 with $\rho_{(1)}$ being the liquidity variable used. The sensitivity analysis of the various jump and co-jump tests under different block sizes is particularly relevant as different levels of m would imply a different variance and covariance structure in the returns of both liquidity and price. We notice a substantial decrease in the percentage of identified liquidity and price jumps when we move from $m = 20$ to $m = 100$. This result is particularly evident when we exclude the PZ tests from the computation. The PZ test is in fact still affected by a high degree of noise and detects a very high number of (spurious) liquidity jumps and co-jumps. A similar pattern can also be observed in the case of the ABD-LM test which detects a higher number of jumps and co-jumps and displays a higher dispersion compared to the other tests as we move from low to high levels of m . The power of the MinRV, the MedRV and the CPR seems to be also affected by noise at low levels of m but the power of the tests shows the tendency to converge as m moves towards 100. The most robust performance, in terms of power, is again displayed by the LBNS, the RBNS and the JO tests across different levels of m .

[Insert Tables 3.11 - 3.12]

Table 3.13 reports the percentage of identified jumps and co-jumps when $m = 100$,

kp varies from 30 to 120 in step of 30 observations and $\rho_{(2)}$ is used as a liquidity variable. We immediately notice that the PZ test is still affected by microstructural noise as the number of identified liquidity jumps is much greater than the average number of jumps detected by the other tests. Unlike in the previous case, the ABD-LM appears well behaved also at lower levels of kp while the JO test detects a slightly higher number of jumps and co-jumps compared to the other tests and across different levels of kp . We finally notice a lower dispersion in the number of detected jumps and co-jumps across the different tests moving from a lower to a higher number of observations used in the pre-averaging window. The result is particularly evident in the case of the permanent and exogenous co-jump tests and partially confirms that the second liquidity variable $\rho_{(2)}$ is affected by a lower microstructural noise compared to $\rho_{(1)}$.

Table 3.14 reports the percentage of identified jumps and co-jumps when $kp = 90$ and m varies from 20 to 100 with $\rho_{(2)}$ being the liquidity variable used. We notice that, on average, the tests detect a lower number of jump and co-jump event when a greater number of observations is used at each block. We find that the average percentage of identified jumps, when we exclude the PZ test, equals 20% under the assumption of $m = 20$ and 9% when $m = 100$. The average number of co-jumps also falls from 26% to an average of 12% with the exogenous co-jumps showing the biggest drop moving from an average of 31% to 14%. The PZ test is still affected by microstructural noise as it detects a suspiciously high number of liquidity jumps. The high number of liquidity jumps also affects the computation of the contemporaneous, permanent and exogenous co-jump detection rate. The power of CPR test is also affected at low levels of m while it converges quickly to the average power of the other tests when $m = 100$ with the

JO test showing the opposite behavior and increasing the number of detected liquidity and price jumps and co-jumps as m raises. The most robust performance, in terms of power, is displayed by the LBNS, the RBNS and the JO tests.

[Insert Tables 3.13 - 3.14]

3.4 Final Remarks

In Chapter 3, we presented a framework that allowed us to explicitly assess the transmission mechanism between price and liquidity dynamics and relate liquidity shocks to price jumps in the EUR/USD FX spot market during the week from May 3 to May 7, 2010. The time frame considered was particularly relevant given that EUR/USD spot displayed a large move opening on Monday at around 1.3250 and ending the session on Friday just above 1.2750.

We divided the empirical exercise in two different parts. In the first part of the empirical application, we used the first battery of jump and co-jump tests on data re-sampled over equally spaced time intervals. We found that the percentage number of identified jumps in both liquidity and price tends to decrease as we move from a high to a low observation frequency. With the exception of the MinRV and the MedRV tests, we also detected a lower number of liquidity than price jumps across all the observation frequencies used. The high number of jumps detected by the MinRV, the MedRV, the CPR and the PZ test was probably affected by the level of microstructural noise. We also noticed a strong performance, in terms of ability to detect jumps, of the LBNS test, followed by the RBNS, the JO and the ABD-LM tests. The average percentage of contemporaneous co-jumps between liquidity and prices was found higher

than the average percentage of permanent and lagged co-jump events. We also found a high number of spurious co-jumps using the MinRV, the MedRV, the CPR and the PZ tests. Finally, very little evidence of lagged co-jumps was observed, in particular, at frequencies greater than 1-5 seconds. The result indicated that jumps in liquidity observed at a frequency of 1 second are usually not followed by statistically significant price jumps at lower frequencies.

In the second part of the empirical application, we used tick-by-tick data with no re-sampling and computed the second battery of tests presented in Chapter 2. The tests allowed us to distinguish between different jump and co-jump events and in particular to measure the number of contemporaneous and permanent jumps and co-jumps together with exogenous co-jumps between two different liquidity measures and the mid-quote spot price of EUR/USD FX. We overall noticed a greater number of contemporaneous jumps in the liquidity process. The difference in jump frequency between liquidity and price was particularly evident in the case of the PZ, the CPR and the ABD-LM tests. The greater dispersion in the number of identified liquidity jumps, when the first liquidity variable was used, could have been driven again by the high level of microstructural noise. When we let the pre-averaging window size increase, we noticed a lower dispersion in the number of contemporaneous price jumps detected. This result confirmed the importance of pre-averaging as a way to reduce the level of microstructural noise especially when the data is collected at a high frequency. The noise in the series affected the PZ and the ABD-LM tests in particular. As we considered a different number of observations per block, we noticed a substantial decrease in the percentage of detected liquidity and price jumps. The power of the MinRV, the MedRV

and the CPR tests was also found to be affected by noise and, in particular, when a low number of observations was used. Overall, as in the first empirical application, we observed a strong performance, in terms of power, of the LBNS, the RBNS and the JO tests.

There are a number of areas for a further extension of the empirical analysis. First, liquidity is shown to play an important role in the context of microstructural contagion and shocks to liquidity, especially when driven by informed trading, have a more permanent impact on prices. In our empirical analysis we have not distinguished between trading regimes or time-of-the-day effects. We would expect a higher number of lagged co-jumps to be detected in proximity of the opening or the closing of a trading session when both trading volatility and displayed liquidity are low as also observed from the data. It would then be particularly interesting to run the co-jump tests for different trading times during the day and isolate diurnal effects. Second, despite the large sample used in the empirical analysis, we would need to extend the time interval considered and measure the robustness of the proposed testing methodology to different trading cycles. It would also be interesting to assess if the transmission mechanism between liquidity and prices is stable across different trading regimes or if it can be affected by other microstructural variables. Finally, the transactional variables, used in the analysis, belong to one side of the limit order book. We would need to study the dynamics of the entire order book in order to detect an asymmetric response of the price to a liquidity shock. We leave these developments to future research.

Table 3.1: Auto-correlations (ACF) of weekly data

The Table reports the ACF of the weekly data, re-sampled over a frequency of 1, 5 and 10 seconds, and for a pre-averaging window of $kp = 30$ observations. Bold denotes values outside of the confidence interval using a significance level $\alpha = 5\%$. If the auto-correlation is higher (lower) than the upper (lower) bound, denoted with $c1$ ($c2$) respectively, the null hypothesis that there is no auto-correlation at and beyond a given lag is rejected at a significance level of α .

kp= 30													
Lag	rLRI	rPRI	rLV1	rLV2	rLRI	rPRI	rLV1	rLV2	rLRI	rPRI	rLV1	rLV2	rLV2
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	-0.172	0.042	-0.167	-0.076	-0.176	-0.015	-0.251	-0.072	-0.100	-0.121	-0.071	-0.026	-0.026
2	-0.041	0.008	-0.054	-0.021	0.040	-0.074	0.158	0.033	-0.005	-0.003	0.077	0.054	0.054
3	0.009	0.028	0.019	-0.014	-0.048	0.059	-0.083	0.000	-0.076	0.028	-0.159	-0.074	-0.074
4	-0.005	-0.022	-0.014	-0.014	0.034	-0.013	0.073	0.008	0.002	0.061	-0.018	0.031	0.031
5	-0.015	-0.045	-0.029	0.000	-0.002	-0.004	-0.029	-0.008	-0.035	-0.019	-0.007	0.026	0.026
6	-0.026	0.007	-0.016	-0.014	0.059	0.029	0.069	-0.024	-0.008	-0.081	-0.023	-0.030	-0.030
7	-0.006	-0.022	-0.030	0.016	-0.048	-0.029	-0.080	-0.017	-0.075	0.033	-0.065	0.008	0.008
8	-0.009	-0.037	0.004	0.021	0.032	0.003	0.074	-0.002	0.033	-0.027	-0.001	-0.071	-0.071
9	-0.012	0.035	-0.021	-0.029	-0.023	0.009	-0.076	0.005	0.048	0.057	0.100	0.053	0.053
10	-0.015	0.003	-0.032	-0.005	0.020	-0.060	0.029	0.001	0.038	-0.026	0.054	-0.129	-0.129
11	0.003	-0.014	-0.005	0.007	0.002	0.036	-0.002	-0.028	-0.018	-0.045	0.036	0.020	0.020
12	0.020	-0.045	0.050	-0.006	0.017	-0.046	0.032	-0.023	0.062	-0.001	0.008	-0.090	-0.090
13	0.005	0.010	0.006	-0.004	-0.010	-0.073	-0.043	0.066	-0.014	0.061	-0.020	0.008	0.008
14	-0.003	0.013	-0.004	-0.023	0.001	-0.036	0.014	-0.007	-0.043	-0.095	-0.043	0.000	0.000
15	-0.015	0.010	-0.026	0.014	-0.060	0.046	-0.035	-0.054	-0.103	-0.040	-0.104	0.007	0.007
16	-0.010	0.015	-0.009	-0.015	-0.029	-0.016	-0.028	-0.003	0.030	0.049	-0.005	-0.004	-0.004
17	-0.012	-0.007	-0.019	0.023	0.026	-0.004	0.045	0.001	0.035	-0.026	0.039	0.000	0.000
18	-0.007	-0.015	-0.036	-0.001	-0.022	0.020	-0.002	-0.014	-0.007	0.061	-0.020	0.081	0.081
19	0.017	0.007	0.007	-0.014	0.002	0.061	0.024	-0.010	-0.012	-0.009	0.033	-0.027	-0.027
20	0.019	-0.017	0.032	-0.003	0.001	-0.009	-0.006	-0.009	-0.041	0.006	-0.016	0.041	0.041
fr:	1 sec				5 secs				10 secs				
c1	0.0245				0.0548				0.0776				
c2	-0.0245				-0.0548				-0.0776				

Table 3.4: Auto-correlations (ACF) of weekly data

The Table reports the ACF of the weekly data, re-sampled over a frequency of 1, 5 and 10 seconds, and for a pre-averaging window of $kp = 120$ observations. Bold denotes values outside of the confidence interval using a significance level $\alpha = 5\%$. If the auto-correlation is higher (lower) than the upper (lower) bound, denoted with $c1$ ($c2$) respectively, the null hypothesis that there is no auto-correlation at and beyond a given lag is rejected at a significance level of α .

kp = 120																				
Lag	rLRI	rPRI	rLV1	rLV2	rLRI	rPRI	rLV1	rLV2	rLRI	rPRI	rLV1	rLV2	rLRI	rPRI	rLV1	rLV2				
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000				
1	-0.095	-0.022	-0.155	-0.045	0.029	0.003	0.100	0.078	0.042	-0.085	0.100	-0.140	0.133	-0.005	0.076	-0.011				
2	-0.077	-0.054	-0.056	0.007	-0.088	0.002	-0.079	-0.048	0.133	-0.005	0.076	-0.011	-0.201	-0.044	-0.232	-0.095				
3	0.015	-0.069	-0.006	0.013	0.069	-0.017	0.074	-0.009	-0.201	-0.044	-0.232	-0.095	-0.034	0.085	-0.061	0.007				
4	0.025	0.010	0.045	-0.028	0.092	-0.030	0.051	-0.084	-0.034	0.085	-0.061	0.007	-0.034	0.085	-0.061	0.007				
5	0.008	-0.038	0.046	0.013	-0.036	0.032	-0.057	-0.023	-0.056	-0.004	-0.032	0.000	-0.056	-0.004	-0.032	0.000				
6	-0.034	0.039	-0.075	-0.005	0.032	-0.108	0.018	-0.004	-0.016	-0.157	-0.013	0.006	-0.016	-0.157	-0.013	0.006				
7	-0.034	0.021	-0.022	0.001	0.034	-0.009	0.005	-0.066	0.004	-0.005	0.028	-0.003	0.004	-0.005	0.028	-0.003				
8	0.029	0.004	0.073	0.027	-0.003	0.026	-0.001	0.026	0.005	-0.038	-0.005	-0.009	0.005	-0.038	-0.005	-0.009				
9	0.018	-0.023	0.033	0.005	0.061	0.034	0.067	0.022	-0.072	0.047	-0.072	-0.067	-0.072	0.047	-0.072	-0.067				
10	-0.014	0.073	0.007	-0.001	-0.080	0.162	-0.050	-0.006	-0.038	0.041	-0.038	0.108	-0.038	0.041	-0.038	0.108				
11	-0.009	-0.039	-0.013	-0.002	-0.009	-0.013	-0.033	-0.017	-0.106	-0.115	-0.064	0.003	-0.106	-0.115	-0.064	0.003				
12	0.016	0.022	0.049	0.028	0.061	-0.060	0.032	-0.037	0.076	-0.034	0.053	-0.006	0.076	-0.034	0.053	-0.006				
13	-0.055	-0.015	-0.083	-0.032	0.049	-0.178	0.030	0.002	-0.024	-0.015	-0.003	-0.050	-0.024	-0.015	-0.003	-0.050				
14	0.036	0.024	0.066	0.003	-0.073	-0.015	-0.068	0.059	0.093	-0.015	0.059	-0.071	0.093	-0.015	0.059	-0.071				
15	-0.043	-0.025	-0.035	0.015	0.054	0.119	0.041	0.034	0.008	-0.073	-0.006	-0.014	0.008	-0.073	-0.006	-0.014				
16	-0.024	0.001	-0.016	-0.042	0.019	-0.042	0.028	-0.005	0.048	-0.063	0.042	-0.061	0.048	-0.063	0.042	-0.061				
17	0.045	-0.070	0.042	0.032	0.018	-0.001	0.000	-0.034	0.014	0.092	-0.021	0.013	0.014	0.092	-0.021	0.013				
18	0.003	-0.044	0.002	-0.006	-0.040	0.152	-0.034	0.015	0.090	-0.044	0.063	-0.019	0.090	-0.044	0.063	-0.019				
19	-0.029	0.036	-0.022	-0.006	-0.046	0.088	-0.039	-0.090	-0.038	0.022	-0.042	0.004	-0.038	0.022	-0.042	0.004				
20	0.025	0.068	0.030	-0.036	0.099	-0.019	0.076	0.043	0.079	-0.088	0.044	-0.020	0.079	-0.088	0.044	-0.020				
fr:	1 sec															10 secs				
c1	0.049															0.1552				
c2	-0.049															-0.1552				

Table 3.5: Auto-correlations (ACF) of weekly data

The Table reports the ACF of the weekly tick-by-tick data and for a pre-averaging window of $kp = 30$ and 60 observations. Bold denotes values outside of the confidence interval using a significance level $\alpha = 5\%$. If the auto-correlation is higher (lower) than the upper (lower) bound, denoted with $c1$ ($c2$) respectively, the null hypothesis that there is no auto-correlation at and beyond a given lag is rejected at a significance level of α .

Lag	kp = 30						kp = 60					
	rLRI	rPRI	rLV1	rLV2	rLRI	rLV2	rLRI	rPRI	rLV1	rLV2	rLRI	rLV2
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	-0.150	0.021	-0.175	-0.132	-0.170	-0.132	-0.170	0.048	-0.176	-0.142	-0.176	-0.142
2	-0.024	0.013	-0.019	-0.010	0.000	-0.010	0.000	0.000	-0.024	-0.031	-0.024	-0.031
3	-0.023	0.021	-0.014	-0.017	-0.019	-0.017	-0.019	0.007	-0.008	-0.012	-0.008	-0.012
4	-0.008	0.004	0.013	0.001	-0.004	0.001	-0.004	-0.016	-0.028	-0.016	-0.028	-0.016
5	-0.014	-0.003	-0.031	-0.013	0.014	-0.013	0.014	-0.016	0.032	0.025	0.032	0.025
6	-0.004	0.005	0.013	0.005	-0.005	0.005	-0.005	0.001	-0.008	-0.008	-0.008	-0.008
7	-0.006	0.003	0.014	0.016	-0.015	0.016	-0.015	-0.013	0.011	-0.003	-0.013	-0.003
8	0.000	0.005	-0.016	-0.008	-0.017	-0.008	-0.017	-0.010	-0.032	-0.022	-0.032	-0.022
9	-0.001	-0.007	-0.026	-0.013	0.029	-0.013	0.029	0.040	0.052	0.036	0.040	0.036
10	-0.007	-0.013	-0.007	-0.004	-0.020	-0.004	-0.020	0.006	0.007	0.021	0.006	0.021
11	0.004	0.002	0.004	0.005	0.007	0.005	0.007	-0.001	-0.029	-0.016	-0.001	-0.016
12	-0.002	0.010	-0.008	-0.008	0.017	-0.008	0.017	0.011	0.051	0.030	0.011	0.030
13	-0.012	-0.015	-0.021	-0.010	-0.018	-0.010	-0.018	-0.012	-0.054	-0.042	-0.012	-0.042
14	-0.003	-0.014	0.012	0.004	0.001	0.004	0.001	-0.011	0.022	0.020	-0.011	0.020
15	-0.013	-0.009	-0.013	-0.017	-0.013	-0.017	-0.013	0.008	0.003	0.000	-0.013	0.000
16	0.006	-0.009	-0.018	-0.010	0.003	-0.010	0.003	0.005	-0.017	-0.015	0.005	-0.015
17	0.011	0.002	0.025	0.011	-0.006	0.011	-0.006	-0.007	0.001	0.009	-0.007	0.001
18	0.000	0.008	-0.022	-0.009	-0.004	-0.009	-0.004	-0.012	-0.039	-0.025	-0.012	-0.039
19	-0.002	0.018	0.030	0.015	-0.012	0.015	-0.012	-0.004	0.011	-0.006	-0.004	-0.006
20	-0.005	-0.005	0.001	-0.006	0.003	-0.006	0.003	-0.003	0.003	0.006	-0.003	0.006
c1	0.014				0.020		0.020					
c2	-0.014				-0.020		-0.020					

Table 3.6: Auto-correlations (ACF) of weekly data

The Table reports the ACF of the weekly tick-by-tick data and for a pre-averaging window of $kp = 90$ and 120 observations. Bold denotes values outside of the confidence interval using a significance level $\alpha = 5\%$. If the auto-correlation is higher (lower) than the upper (lower) bound, denoted with $c1$ ($c2$) respectively, the null hypothesis that there is no auto-correlation at and beyond a given lag is rejected at a significance level of α .

Lag	kp = 90				kp = 120			
	rLRI	rPRI	rLV1	rLV2	rLRI	rPRI	rLV1	rLV2
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	-0.135	0.043	-0.130	-0.095	-0.141	0.030	-0.146	-0.113
2	-0.051	0.013	-0.021	-0.017	-0.042	-0.007	-0.072	-0.015
3	-0.008	-0.016	-0.013	-0.017	0.000	-0.027	0.030	0.046
4	-0.002	0.011	-0.014	-0.009	0.005	-0.014	0.000	-0.016
5	0.002	-0.042	0.025	-0.002	-0.013	0.007	0.032	0.014
6	-0.005	0.022	-0.002	-0.006	0.012	-0.011	0.016	0.017
7	0.003	-0.005	-0.005	-0.018	0.012	0.021	-0.014	-0.008
8	0.005	-0.001	0.040	0.014	-0.014	0.031	0.004	0.007
9	0.020	0.006	0.005	-0.016	0.016	-0.024	0.009	0.020
10	-0.012	0.004	-0.004	0.011	-0.020	-0.014	0.011	-0.013
11	0.018	0.000	-0.004	0.005	0.033	0.008	0.011	0.003
12	0.005	-0.015	-0.015	-0.004	-0.018	0.006	-0.020	-0.012
13	-0.017	0.015	-0.008	-0.019	-0.003	-0.009	0.029	0.006
14	-0.010	-0.014	-0.019	-0.002	-0.005	0.000	-0.015	-0.022
15	-0.010	0.008	0.037	0.016	0.006	0.008	-0.076	-0.033
16	0.016	0.003	0.024	0.016	-0.008	0.003	0.035	0.014
17	-0.010	-0.011	-0.001	-0.007	0.021	-0.020	0.007	-0.007
18	-0.003	-0.015	-0.005	0.006	-0.008	0.021	-0.012	-0.040
19	0.007	-0.003	0.012	-0.008	-0.016	0.025	0.028	0.052
20	-0.005	-0.013	-0.061	-0.034	-0.010	-0.017	-0.032	-0.054
c1	0.024				0.028			
c2	-0.024				-0.028			

Table 3.7: Percentage of Jump Events for EUR/USD FX and Liquidity Variable $\rho_{ho(1)}$

The Table reports the percentage of identified jumps at a significance level $\alpha = 5\%$ during the week from May 3rd to May 7th, 2010. In particular, the jump indicator vectors $JT_{(1),1}$, $JT_{(1),5}$ and $JT_{(1),10}$ provide a measure of liquidity jumps for the liquidity variable $\rho_{ho(1)}$ at a frequency of 1, 5 and 10 seconds respectively, while the jump indicator vectors $JT_{(2),1}$, $JT_{(2),5}$ and $JT_{(2),10}$ provide a measure of mid-quote price jumps again at a frequency of 1, 5 and 10 seconds respectively for $m = 28$ observations per block.

Procedure	JT1,1	JT2,1	JT1,5	JT2,5	JT1,10	JT2,10
m: 28						
LBNS	0.121	0.197	0.030	0.091	0.000	0.030
RBNS	0.061	0.136	0.030	0.106	0.015	0.045
MinRV	0.424	0.379	0.212	0.288	0.167	0.273
MedRV	0.439	0.424	0.030	0.061	0.000	0.030
JO	0.106	0.182	0.091	0.106	0.015	0.015
ABD-LM	0.076	0.076	0.061	0.045	0.106	0.076
CPR	0.409	0.636	0.061	0.439	0.030	0.333
PZ	0.773	0.833	0.530	0.606	0.455	0.576

Table 3.8: Percentage of Jump Events for EUR/USD FX and Liquidity Variable $\rho_{ho(2)}$

The Table reports the percentage of identified jumps at a significance level $\alpha = 5\%$ during the week from May 3rd to May 7th, 2010. In particular, the jump indicator vectors $JT_{(1),1}$, $JT_{(1),5}$ and $JT_{(1),10}$ provide a measure of liquidity jumps for the liquidity variable $\rho_{ho(2)}$ at a frequency of 1, 5 and 10 seconds respectively, while the jump indicator vectors $JT_{(2),1}$, $JT_{(2),5}$ and $JT_{(2),10}$ provide a measure of mid-quote price jumps again at a frequency of 1, 5 and 10 seconds respectively for $m = 28$ observations per block.

Procedure	JT1,1	JT2,1	JT1,5	JT2,5	JT1,10	JT2,10
m: 28						
LBNS	0.091	0.197	0.015	0.091	0.030	0.030
RBNS	0.076	0.136	0.030	0.106	0.030	0.045
MinRV	0.303	0.379	0.303	0.288	0.212	0.273
MedRV	0.439	0.424	0.000	0.061	0.000	0.030
JO	0.076	0.182	0.030	0.106	0.000	0.015
ABD-LM	0.061	0.076	0.000	0.045	0.061	0.076
CPR	0.379	0.636	0.061	0.439	0.061	0.333
PZ	0.879	0.833	0.788	0.606	0.606	0.576

Table 3.9: Percentage of Co-jump Events for EUR/USD FX and Liquidity Variable $rho_{(1)}$

The Table reports the percentage of identified co-jumps at a significance level $\alpha = 5\%$ during the week from May 3rd to May 7th, 2010. In particular, the co-jump indicator vectors CJT_1 , CJT_2 and CJT_3 provide a measure of contemporaneous co-jumps between liquidity and prices at a frequency of 1, 5 and 10 seconds respectively. The co-jump indicator vectors $PCJT_1$, $PCJT_2$ and $PCJT_3$ provide a measure of permanent co-jumps between liquidity and prices at a frequency of 1-5, 5-10 and 1-5 and 5-10 combined. Finally, the co-jump indicator vectors $LCJT_1$, $LCJT_2$ and $LCJT_3$ provide a measure of lagged co-jumps between liquidity and prices at a frequency of 1-5, 1-10 and 1-5 and 1-10 combined. The liquidity variable used is defined by $rho_{(1)}$.

Procedure	CJT1	CJT2	CJT3	PCJT1	PCJT2	PCJT3	LCJT1	LCJT2	LCJT3
m: 28									
LBNS	0.136	0.015	0.000	0.136	0.015	0.030	0.030	0.000	0.015
RBNS	0.106	0.030	0.000	0.152	0.030	0.061	0.015	0.015	0.030
MinRV	0.515	0.394	0.273	0.530	0.470	0.561	0.273	0.424	0.455
MedRV	0.500	0.061	0.000	0.364	0.045	0.076	0.076	0.030	0.045
JO	0.167	0.091	0.015	0.091	0.030	0.061	0.091	0.030	0.030
ABD-LM	0.136	0.015	0.091	0.045	0.091	0.076	0.045	0.091	0.076
CPR	0.621	0.288	0.197	0.515	0.273	0.364	0.258	0.288	0.288
PZ	0.879	0.697	0.636	0.909	0.742	0.803	0.833	0.848	0.924

Table 3.10: Percentage of Co-jump Events for EUR/USD FX and Liquidity Variable $\rho_{ho(2)}$

The Table reports the percentage of identified co-jumps at a significance level $\alpha = 5\%$ during the week from May 3rd to May 7th, 2010. In particular, the co-jump indicator vectors CJT_1 , CJT_2 and CJT_3 provide a measure of contemporaneous co-jumps between liquidity and prices at a frequency of 1, 5 and 10 seconds respectively. The co-jump indicator vectors $PCJT_1$, $PCJT_2$ and $PCJT_3$ provide a measure of permanent co-jumps between liquidity and prices at a frequency of 1-5, 5-10 and 1-5 and 5-10 combined. Finally, the co-jump indicator vectors $LCJT_1$, $LCJT_2$ and $LCJT_3$ provide a measure of lagged co-jumps between liquidity and prices at a frequency of 1-5, 1-10 and 1-5 and 1-10 combined. The liquidity variable used is defined by $\rho_{ho(2)}$.

Procedure	CJT1	CJT2	CJT3	PCJT1	PCJT2	PCJT3	LCJT1	LCJT2	LCJT3
m: 28									
LBNS	0.136	0.076	0.000	0.106	0.030	0.030	0.030	0.000	0.030
RBNS	0.106	0.091	0.015	0.106	0.045	0.076	0.045	0.045	0.076
MinRV	0.439	0.439	0.288	0.500	0.409	0.455	0.242	0.288	0.318
MedRV	0.591	0.015	0.000	0.348	0.000	0.030	0.030	0.000	0.000
JO	0.167	0.061	0.000	0.061	0.030	0.030	0.076	0.030	0.045
ABD-LM	0.106	0.015	0.045	0.045	0.030	0.045	0.015	0.030	0.030
CPR	0.591	0.273	0.152	0.470	0.212	0.288	0.303	0.242	0.348
PZ	0.985	0.818	0.682	0.955	0.864	0.909	0.955	0.848	0.879

Table 3.11: Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $rho_{(1)}$

The Table reports the percentage of identified jump and co-jump events using combinations of univariate tests for jumps computed at a significance level $\alpha = 5\%$ and under the assumption of $kp = 30, 60, 90$ and 120. The data is observed at a tick-by-tick level during the week from May 3rd to May 7th, 2010. In particular, the test indicator vectors JT1 and JT2 provide a measure of contemporaneous jumps while CJT, PCJT and ECJT of contemporaneous, permanent and exogenous co-jumps respectively. The liquidity variables used is defined by $rho_{(1)}$.

Procedure	JT1	JT2	CJT	PCJT	ECJT	Procedure	JT1	JT2	CJT	PCJT	ECJT
kp: 30											
LBNS	0.157	0.071	0.118	0.103	0.148	LBNS	0.148	0.059	0.108	0.059	0.134
RBNS	0.130	0.076	0.118	0.111	0.135	RBNS	0.123	0.030	0.094	0.064	0.124
MinRV	0.066	0.069	0.093	0.059	0.106	MinRV	0.099	0.039	0.084	0.050	0.099
MedRV	0.084	0.071	0.098	0.057	0.101	MedRV	0.103	0.059	0.084	0.064	0.119
JO	0.118	0.189	0.177	0.158	0.246	JO	0.084	0.197	0.148	0.178	0.198
ABD-LM	0.337	0.197	0.270	0.276	0.352	ABD-LM	0.369	0.172	0.296	0.292	0.416
CPR	0.147	0.005	0.066	0.037	0.163	CPR	0.123	0.005	0.059	0.025	0.198
PZ	0.789	0.007	0.386	0.426	0.916	PZ	0.798	0.025	0.394	0.446	0.896
kp: 90											
LBNS	0.170	0.067	0.119	0.090	0.142	LBNS	0.158	0.089	0.158	0.110	0.210
RBNS	0.119	0.074	0.104	0.097	0.104	RBNS	0.129	0.089	0.149	0.120	0.200
MinRV	0.015	0.074	0.074	0.022	0.090	MinRV	0.099	0.050	0.109	0.080	0.170
MedRV	0.022	0.059	0.052	0.052	0.060	MedRV	0.079	0.089	0.099	0.060	0.100
JO	0.119	0.200	0.133	0.164	0.261	JO	0.129	0.149	0.188	0.200	0.150
ABD-LM	0.415	0.163	0.304	0.366	0.440	ABD-LM	0.386	0.158	0.307	0.300	0.380
CPR	0.156	0.022	0.104	0.082	0.142	CPR	0.168	0.030	0.109	0.100	0.210
PZ	0.852	0.044	0.600	0.679	0.933	PZ	0.851	0.050	0.663	0.720	0.920

Table 3.12: Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $rho_{(1)}$

The Table reports the percentage of identified jump and co-jump events using combinations of univariate tests for jumps computed at a significance level $\alpha = 5\%$ and under the assumption of $m = 20, 50, 75$ and 100. The data is observed at a tick-by-tick level during the week from May 3rd to May 7th, 2010. In particular, the test indicator vectors JT1 and JT2 provide a measure of contemporaneous jumps while CJT, PCJT and ECJT of contemporaneous, permanent and exogenous co-jumps respectively. The liquidity variables used is defined by $rho_{(1)}$.

Procedure	JT1	JT2	CJT	PCJT	ECJT	Procedure	JT1	JT2	CJT	PCJT	ECJT
m: 20						m:50					
LBNS	0.157	0.071	0.118	0.103	0.148	LBNS	0.148	0.059	0.108	0.059	0.134
RBNS	0.130	0.076	0.118	0.111	0.135	RBNS	0.123	0.030	0.094	0.064	0.124
MinRV	0.066	0.069	0.093	0.059	0.106	MinRV	0.099	0.039	0.084	0.050	0.099
MedRV	0.084	0.071	0.098	0.057	0.101	MedRV	0.103	0.059	0.084	0.064	0.119
JO	0.118	0.189	0.177	0.158	0.246	JO	0.084	0.197	0.148	0.178	0.198
ABD-LM	0.337	0.197	0.270	0.276	0.352	ABD-LM	0.369	0.172	0.296	0.292	0.416
CPR	0.147	0.005	0.066	0.037	0.163	CPR	0.123	0.005	0.059	0.025	0.198
PZ	0.789	0.007	0.386	0.426	0.916	PZ	0.798	0.025	0.394	0.446	0.896
m: 75						m: 100					
LBNS	0.170	0.067	0.119	0.090	0.142	LBNS	0.158	0.089	0.158	0.110	0.210
RBNS	0.119	0.074	0.104	0.097	0.104	RBNS	0.129	0.089	0.149	0.120	0.200
MinRV	0.015	0.074	0.074	0.022	0.090	MinRV	0.099	0.050	0.109	0.080	0.170
MedRV	0.022	0.059	0.052	0.052	0.060	MedRV	0.079	0.089	0.099	0.060	0.100
JO	0.119	0.200	0.133	0.164	0.261	JO	0.129	0.149	0.188	0.200	0.150
ABD-LM	0.415	0.163	0.304	0.366	0.440	ABD-LM	0.386	0.158	0.307	0.300	0.380
CPR	0.156	0.022	0.104	0.082	0.142	CPR	0.168	0.030	0.109	0.100	0.210
PZ	0.852	0.044	0.600	0.679	0.933	PZ	0.851	0.050	0.663	0.720	0.920

Table 3.13: Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $rho_{(2)}$

The Table reports the percentage of identified jump and co-jump events using combinations of univariate tests for jumps computed at a significance level $\alpha = 5\%$ and under the assumption of $kp = 30, 60, 90$ and 120. The data is observed at a tick-by-tick level during the week from May 3rd to May 7th, 2010. In particular, the test indicator vectors JT1 and JT2 provide a measure of contemporaneous jumps while CJT, PCJT and ECJT of contemporaneous, permanent and exogenous co-jumps respectively. The liquidity variables used is defined by $rho_{(2)}$.

Procedure	JT1	JT2	CJT	PCJT	ECJT	Procedure	JT1	JT2	CJT	PCJT	ECJT
kp: 30											
LBNS	0.034	0.071	0.084	0.067	0.096	kp: 60	LBNS	0.030	0.059	0.054	0.094
RBNS	0.066	0.076	0.103	0.101	0.113		RBNS	0.034	0.030	0.089	0.104
MinRV	0.096	0.069	0.106	0.079	0.128		MinRV	0.079	0.039	0.108	0.064
MedRV	0.064	0.071	0.086	0.079	0.103		MedRV	0.094	0.059	0.113	0.129
JO	0.177	0.189	0.270	0.347	0.320		JO	0.148	0.197	0.266	0.287
ABD-LM	0.081	0.197	0.138	0.103	0.145	kp: 90	ABD-LM	0.108	0.172	0.138	0.153
CPR	0.059	0.005	0.039	0.025	0.108		CPR	0.069	0.005	0.039	0.030
PZ	0.926	0.007	0.565	0.608	0.985		PZ	0.906	0.025	0.655	0.743
kp: 120											
LBNS	0.037	0.067	0.074	0.067	0.075	LBNS	0.050	0.089	0.089	0.050	0.060
RBNS	0.044	0.074	0.081	0.090	0.097	RBNS	0.040	0.089	0.129	0.070	0.060
MinRV	0.096	0.074	0.096	0.097	0.090	MinRV	0.099	0.050	0.139	0.090	0.120
MedRV	0.052	0.059	0.089	0.060	0.097	MedRV	0.109	0.089	0.139	0.110	0.130
JO	0.163	0.200	0.244	0.276	0.269	JO	0.099	0.149	0.218	0.240	0.350
ABD-LM	0.133	0.163	0.126	0.134	0.216	ABD-LM	0.129	0.158	0.139	0.080	0.140
CPR	0.081	0.022	0.052	0.045	0.127	CPR	0.059	0.030	0.059	0.040	0.080
PZ	0.926	0.037	0.674	0.716	0.978	PZ	0.911	0.059	0.683	0.800	0.990

Table 3.14: Percentage of Jump, Co-Jump Events for EUR/USD FX and Liquidity Variable $rho_{(2)}$

The Table reports the percentage of identified jump and co-jump events using combinations of univariate tests for jumps computed at a significance level $\alpha = 5\%$ and under the assumption of $m = 20, 50, 75$ and 100. The data is observed at a tick-by-tick level during the week from May 3rd to May 7th, 2010. In particular, the test indicator vectors JT1 and JT2 provide a measure of contemporaneous jumps while CJT, PCJT and ECJT of contemporaneous, permanent and exogenous co-jumps respectively. The liquidity variables used is defined by $rho_{(2)}$.

Procedure	JT1	JT2	CJT	PCJT	ECJT	Procedure	JT1	JT2	CJT	PCJT	ECJT
m: 20											
LBNS	0.112	0.137	0.133	0.096	0.147	m: 50	LBNS	0.052	0.085	0.066	0.052
RBNS	0.075	0.103	0.099	0.081	0.122		RBNS	0.041	0.070	0.066	0.067
MinRV	0.387	0.402	0.507	0.553	0.560		MinRV	0.207	0.203	0.280	0.259
MedRV	0.436	0.415	0.507	0.572	0.566		MedRV	0.229	0.196	0.277	0.274
JO	0.071	0.178	0.141	0.136	0.202		JO	0.103	0.166	0.166	0.163
ABD-LM	0.078	0.062	0.059	0.038	0.084	ABD-LM	0.103	0.118	0.122	0.067	
CPR	0.330	0.031	0.191	0.158	0.456	CPR	0.155	0.037	0.107	0.078	
PZ	0.891	0.043	0.611	0.727	0.957	PZ	0.915	0.030	0.642	0.722	
m: 75											
LBNS	0.044	0.077	0.066	0.044	0.083	m: 100	LBNS	0.037	0.067	0.074	0.067
RBNS	0.039	0.055	0.061	0.056	0.094		RBNS	0.044	0.074	0.081	0.090
MinRV	0.144	0.077	0.127	0.117	0.167		MinRV	0.096	0.074	0.096	0.097
MedRV	0.105	0.116	0.155	0.106	0.139		MedRV	0.052	0.059	0.089	0.060
JO	0.149	0.204	0.238	0.239	0.267		JO	0.163	0.200	0.244	0.276
ABD-LM	0.122	0.116	0.133	0.089	0.144	ABD-LM	0.133	0.163	0.126	0.134	
CPR	0.061	0.033	0.055	0.039	0.144	CPR	0.081	0.022	0.052	0.045	
PZ	0.912	0.033	0.680	0.783	0.978	PZ	0.926	0.037	0.674	0.716	

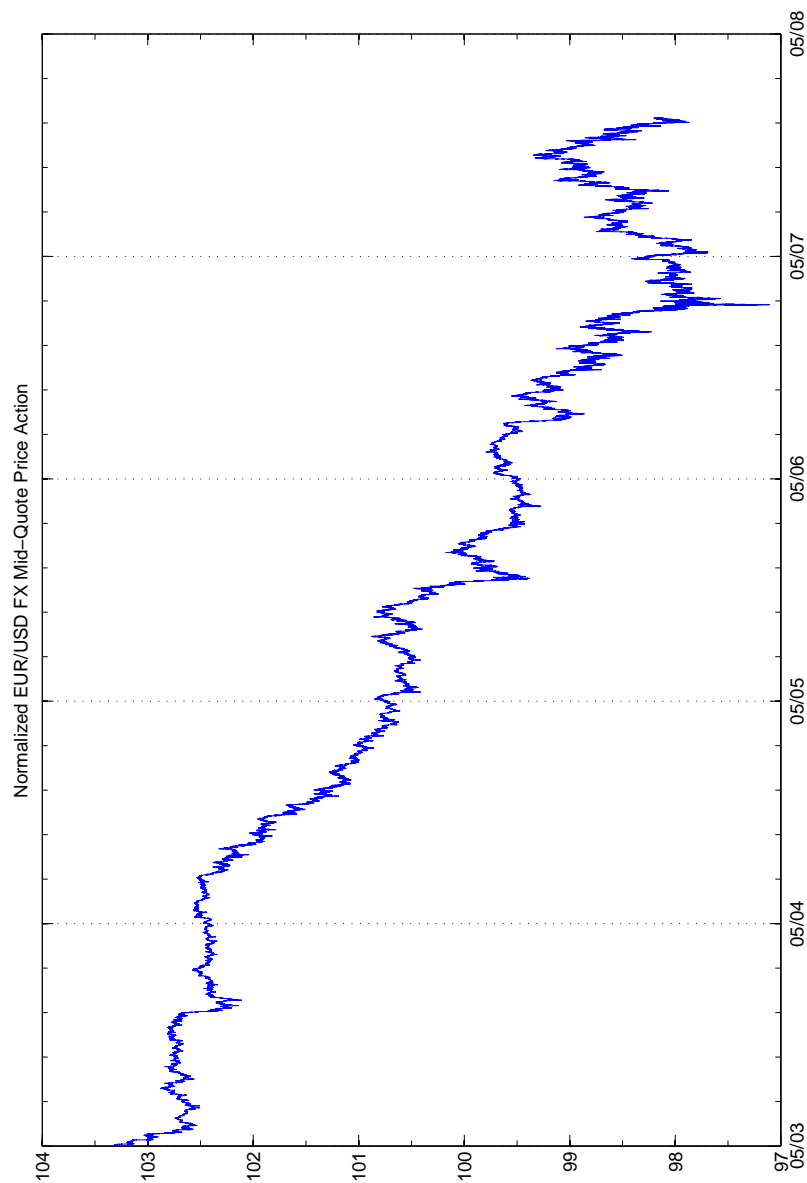


Figure 3.1: Normalized EUR/USD FX Mid-Quote Price Action.

The Figure shows the tick-by-tick normalized mid-quote price series of EUR/USD FX spot during the week from May 3 to May 7, 2010. To preserve data confidentiality, we show the price series in index form. The time series is normalized so that the average mid-quote price over the whole sample is set at 100. The vertical dotted line marks the end of the NY trading session and the open of the Australian trading session.

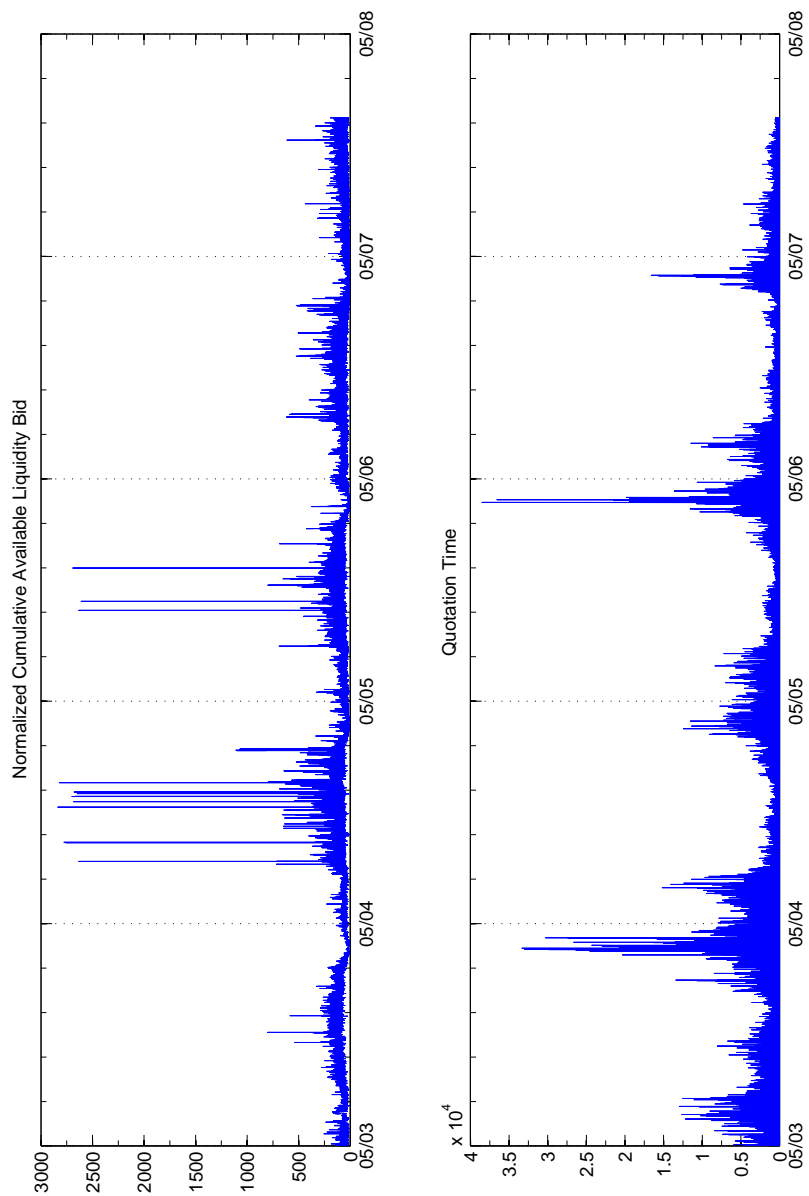


Figure 3.2: Normalized Cumulative Available Liquidity and Quotation Time.

The Figures show the tick-by-tick cumulative available liquidity variable $Qx_{t,bid}$ on the bid side of the order book and the quotation time trr_t during the week from May 3 to May 7, 2010. To preserve data confidentiality, both the liquidity and the time series have been considered in index form. The time series is normalized so that the average cumulative liquidity and the quotation time over the whole sample is set at 100. The vertical dotted line marks the end of the NY trading session and the open of the Australian trading session.

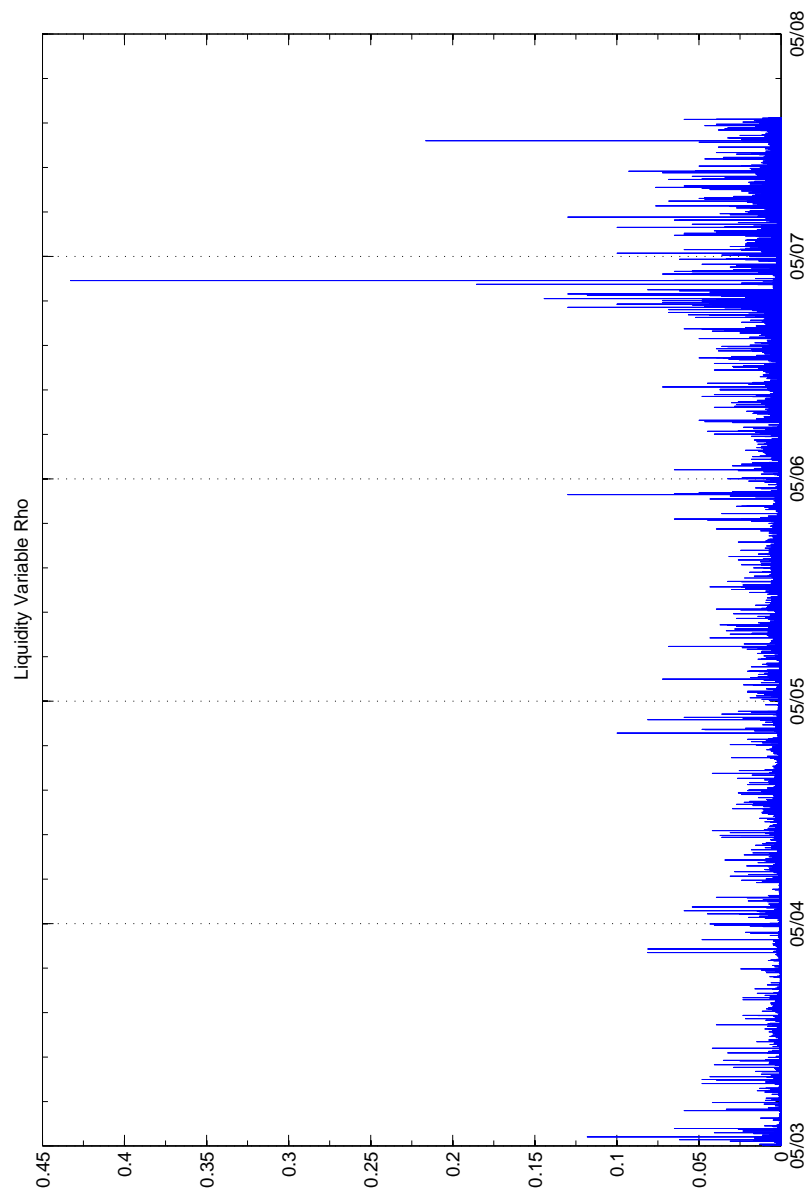


Figure 3.3: Normalized Liquidity Variable $\rho_{(2)}$.

The Figure shows the tick-by-tick liquidity variable $\rho_{(2)}$ for EUR/USD FX during the week from May 3 to May 7, 2010. The liquidity variable is given by $\rho_{(2)} = 1/(Q_{x_t, bid} * tx_t)$ where tx_t represents the quotation time, i.e. the time expressed in milliseconds between consecutive quotations on either side of the order book, and $Q_{x_t, bid(ask)}$ the cumulative available liquidity on the bid (ask) side of the order book at $k = 2$ ticks from the best displayed bid (ask) price. To preserve data confidentiality, the liquidity series has been considered in index form. The time series is normalized so that the average cumulative liquidity over the whole sample is set at 100. The vertical dotted line marks the end of the NY trading session and the open of the Australian trading session.

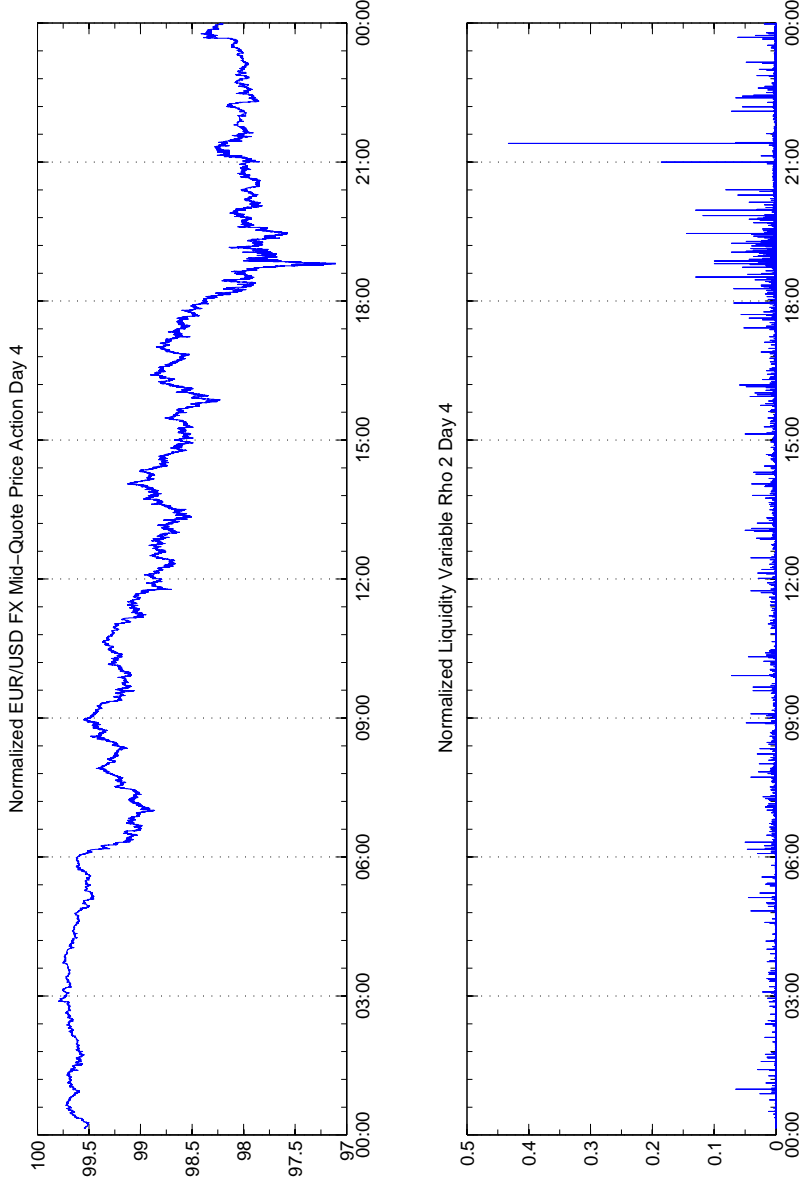


Figure 3.4: Normalized EUR/USD FX Mid-Quote Price and Liquidity Variable $\rho_{ho(2)}$.

The Figures show the tick-by-tick mid-quote price for EUR/USD FX spot and the liquidity variable $\rho_{ho(2)}$ during May 6, 2010. The normalized mid-quote price for EUR/USD FX spot is denoted as $Px_t = (px_{1,t,bid} + px_{1,t,ask})/2$ while the liquidity variable is given by $\rho_{ho(2)} = 1/(Qx_{t,bid} * trx_t)$ where trx_t represents the quotation time, i.e. the time expressed in milliseconds between consecutive quotations on either side of the order book, and $Qx_{t,bid(ask)}$ the cumulative available liquidity on the bid (ask) side of the order book at $k = 2$ ticks from the best displayed bid (ask) price. To preserve data confidentiality, the liquidity series has been considered in index form. The time series is normalized so that the average cumulative liquidity over the whole sample is set at 100. The vertical dotted line marks the end of the NY trading session and the open of the Australian trading session.

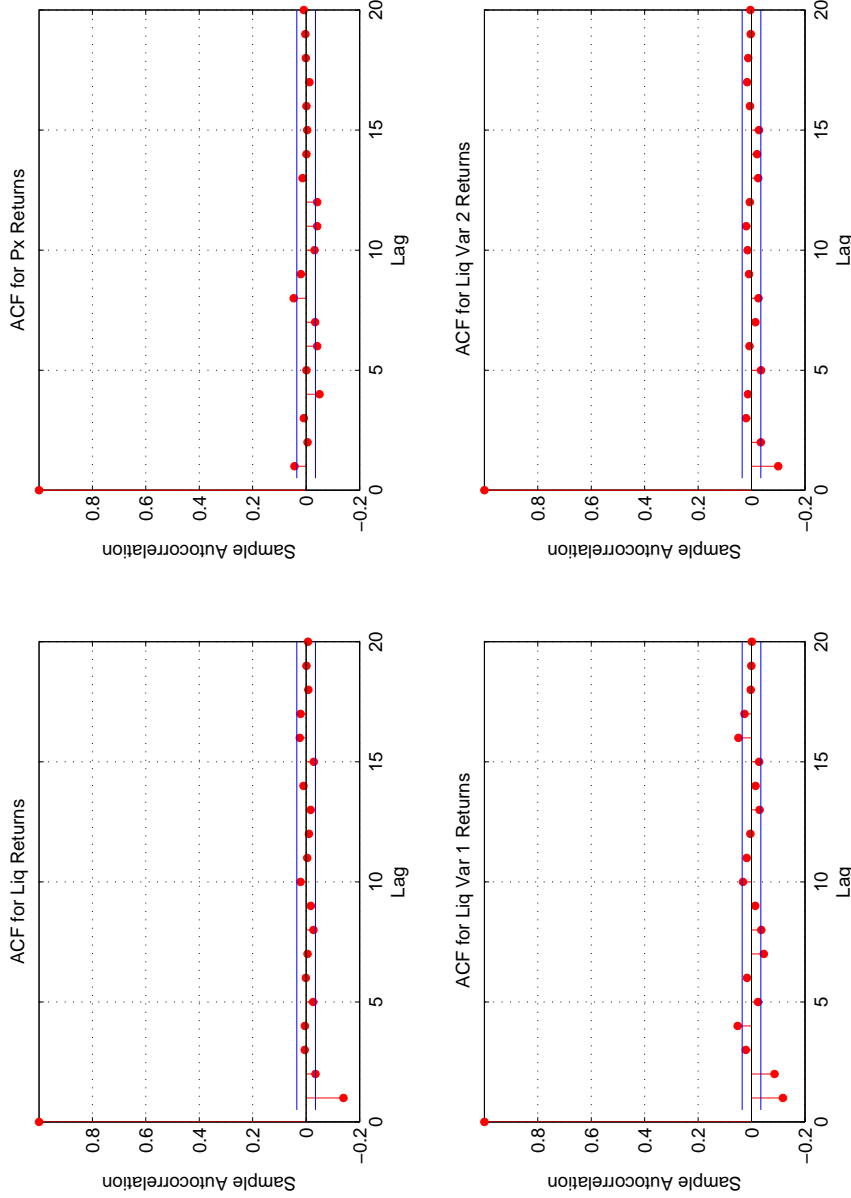


Figure 3.5: Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 30 obs).

The Figures show the plot for the ACF of weekly data, re-sampled over a frequency of 1 second, after pre-averaging is performed over a window of 30 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $rho_{(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $rho_{(2)}$.

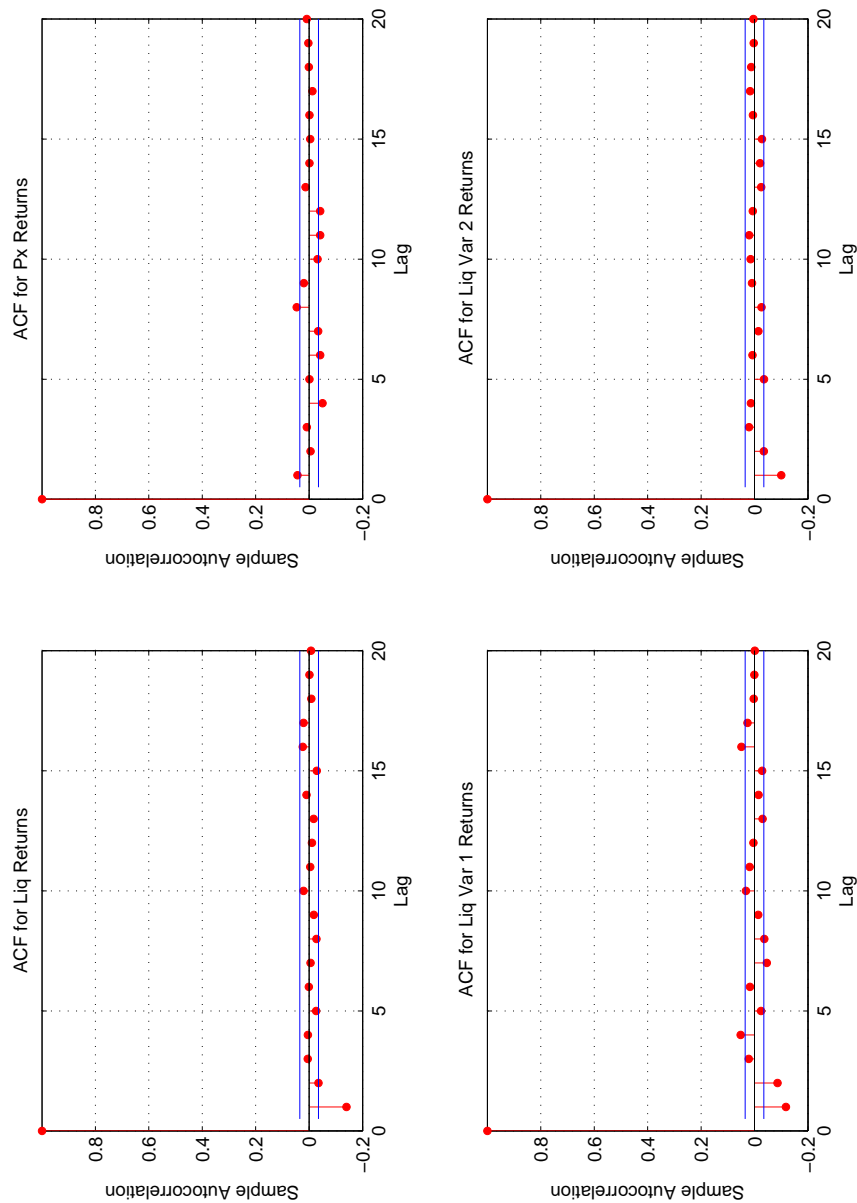


Figure 3.6: Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 60 obs).

The Figures show the plot for the ACF of weekly data, re-sampled over a frequency of 1 second, after pre-averaging is performed over a window of 60 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $rho_{(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $rho_{(2)}$.

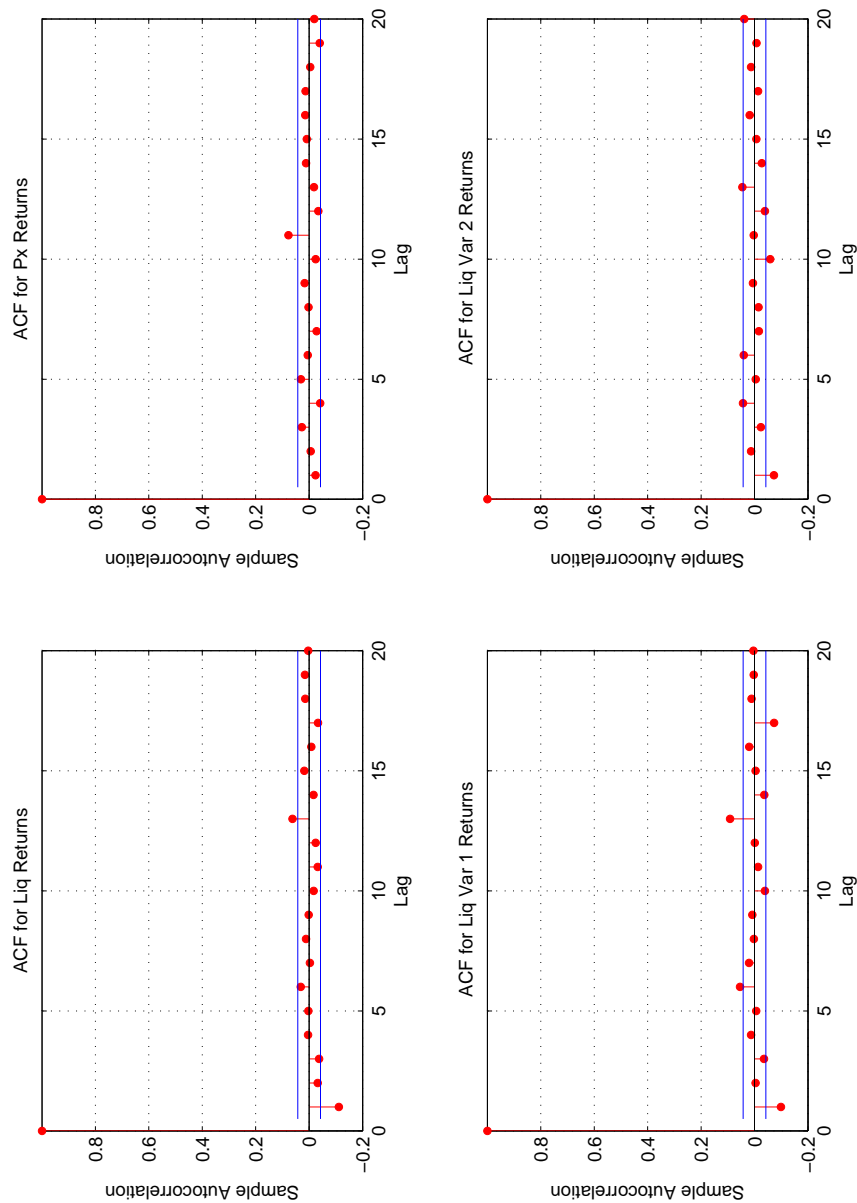


Figure 3.7: Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 90 obs).

The Figures show the plot for the ACF of weekly data, re-sampled over a frequency of 1 second, after pre-averaging is performed over a window of 90 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $rho_{(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $rho_{(2)}$.

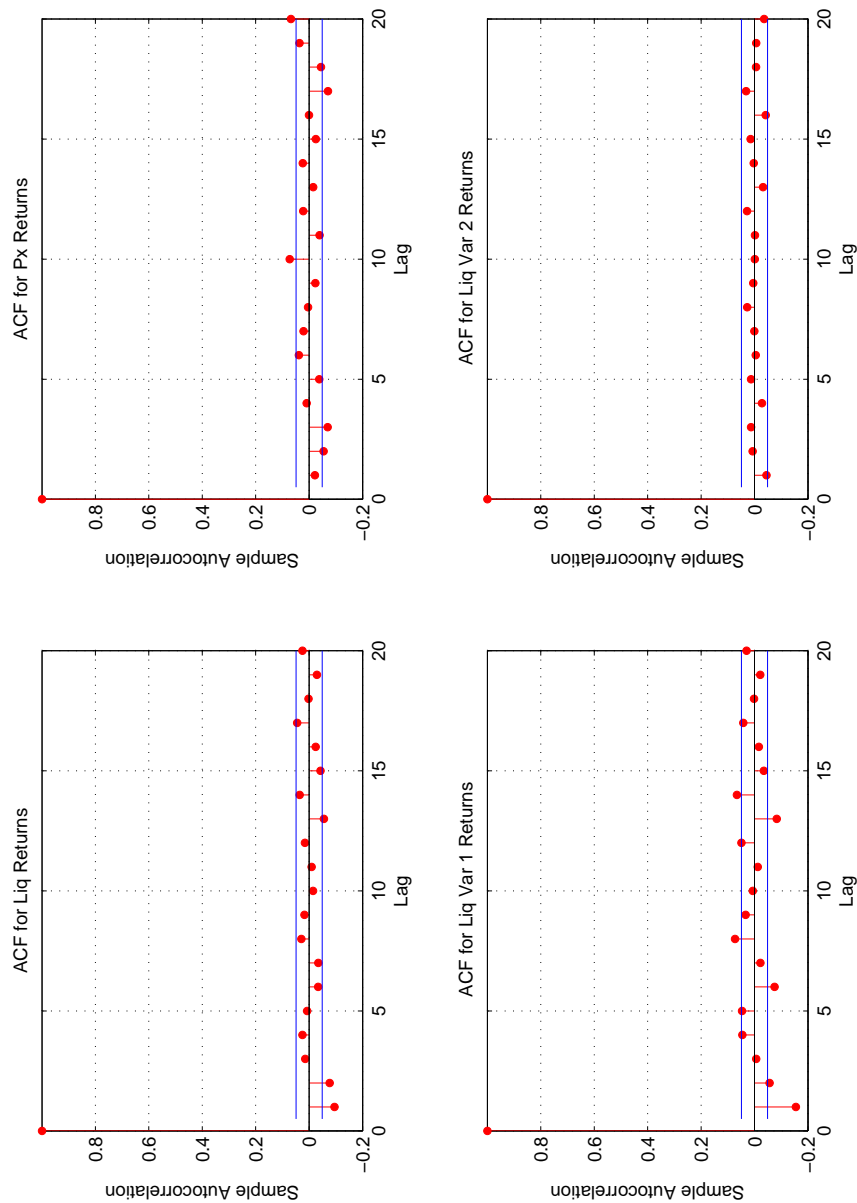


Figure 3.8: Auto-correlation Function (ACF) plot for weekly data (1 sec. re-sample, 120 obs).

The Figures show the plot for the ACF of weekly data, re-sampled over a frequency of 1 second, after pre-averaging is performed over a window of 120 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $rho_{(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $rho_{(2)}$.

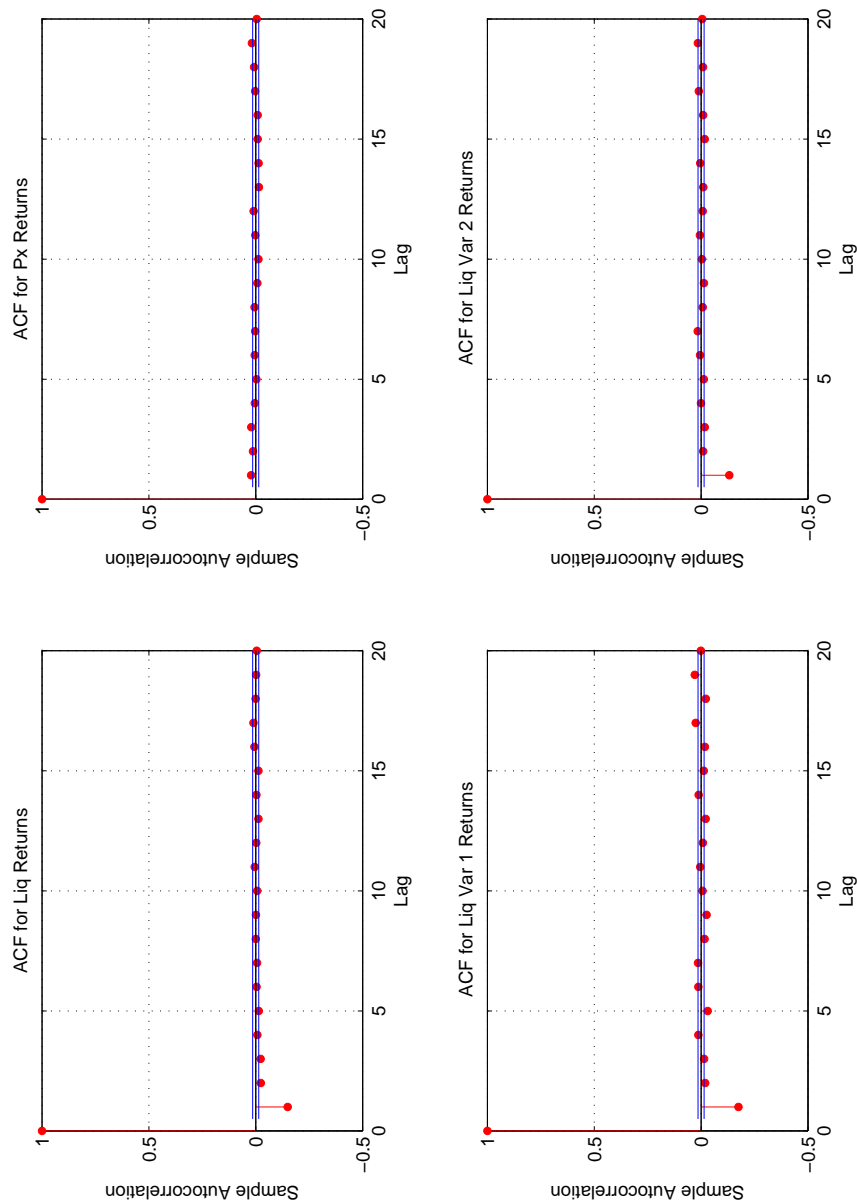


Figure 3.9: Auto-correlation Function (ACF) plot for weekly data (tick data, 30 obs)

The Figures show the plot for the ACF of weekly tick-by-tick data after pre-averaging is performed over a window of 30 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $rho_{(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $rho_{(2)}$.

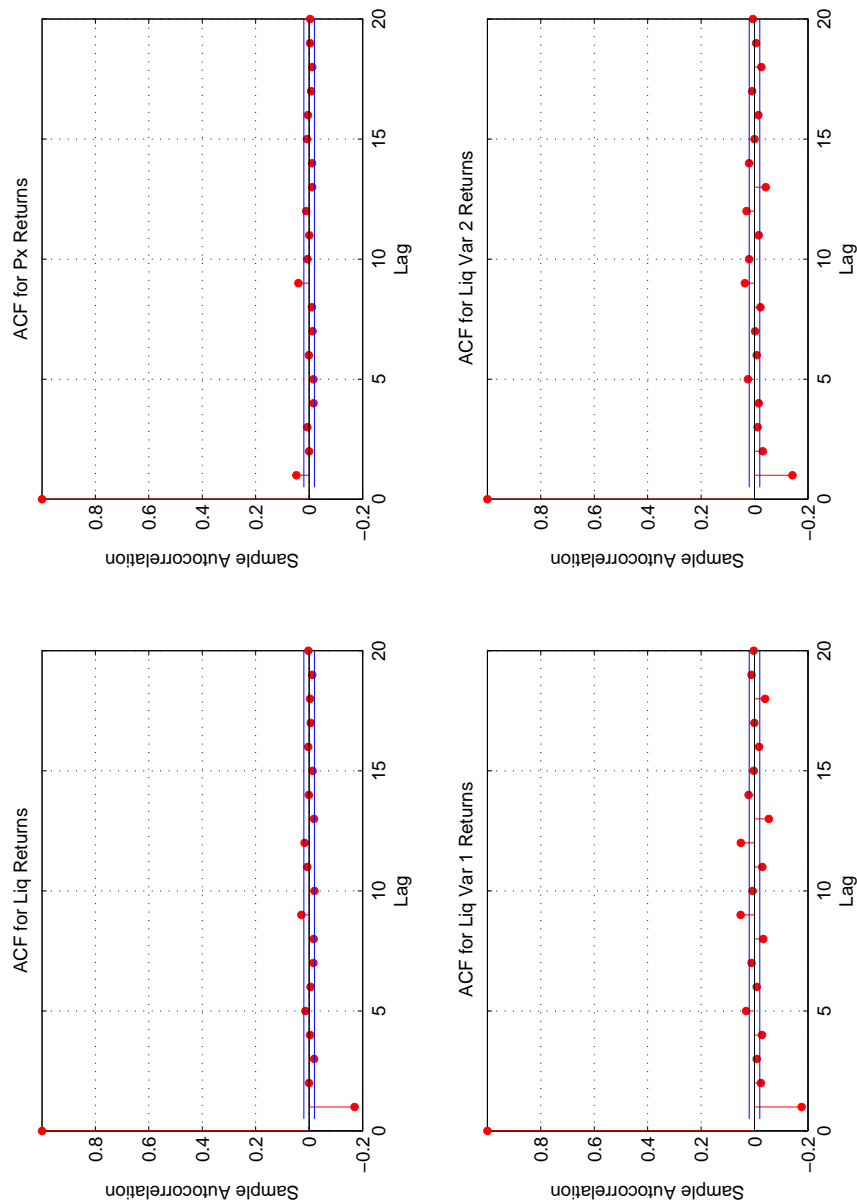


Figure 3.10: Auto-correlation Function (ACF) plot for weekly data (tick data, 60 obs).

The Figures show the plot for the ACF of weekly tick-by-tick data after pre-averaging is performed over a window of 60 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $\rho_{ho(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $\rho_{ho(2)}$.

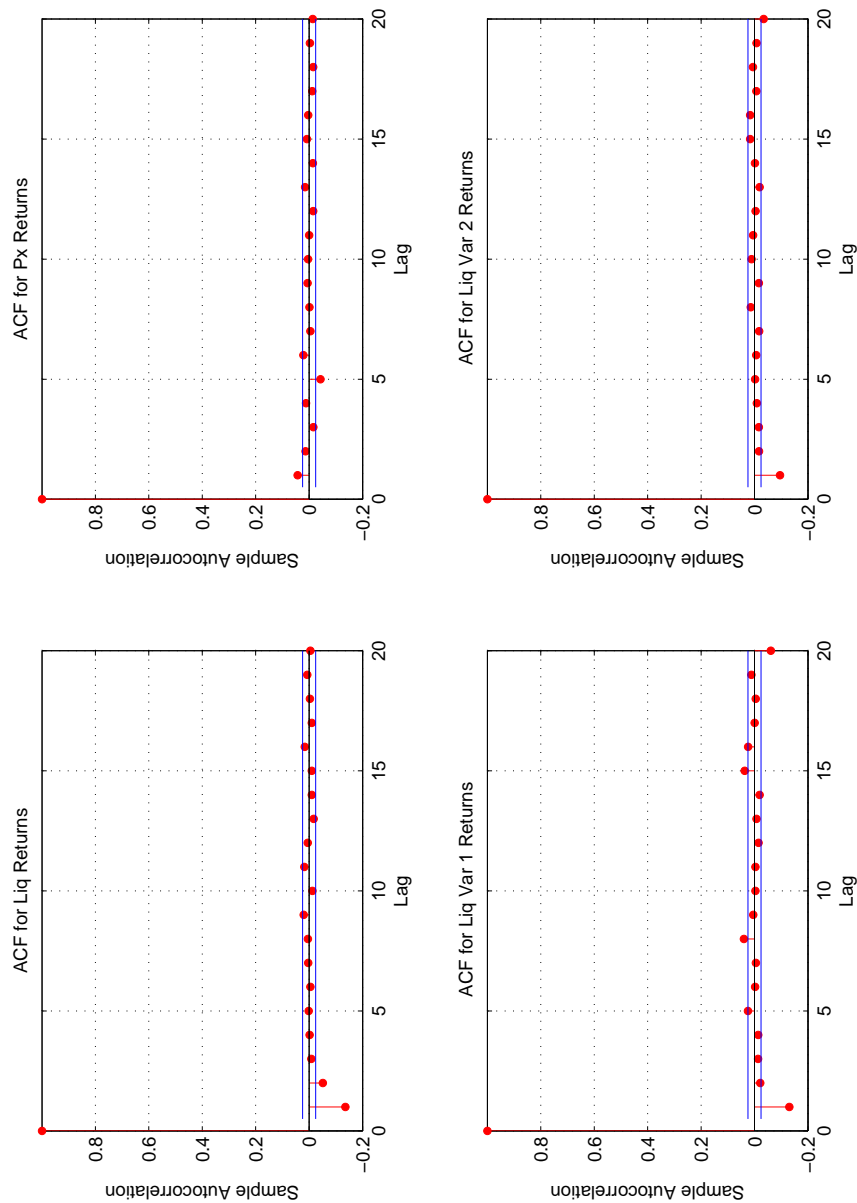


Figure 3.11: Auto-correlation Function (ACF) plot for weekly data (tick data, 90 obs.).

The Figures show the plot for the ACF of weekly tick-by-tick data after pre-averaging is performed over a window of 90 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $\rho_{ho(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $\rho_{ho(2)}$.

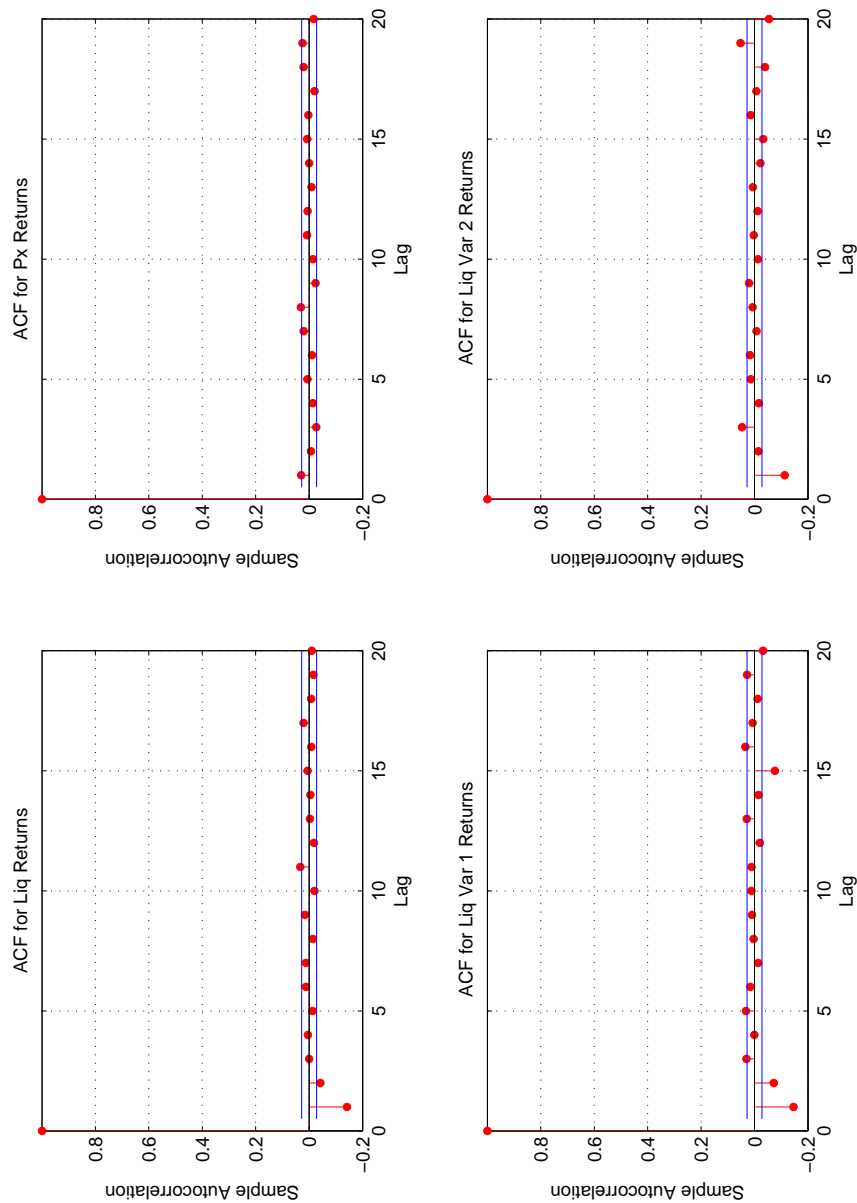


Figure 3.12: Auto-correlation Function (ACF) plot for weekly data (tick data, 120 obs.).

The Figures show the plot for the ACF of weekly tick-by-tick data after pre-averaging is performed over a window of 120 observations. The variable “Liq. Returns” represents the logarithmic returns of the normalized liquidity variable defined as the cumulative depth of the order book at $k = 2$ ticks from the best displayed price. The variable “Px Returns” represents the logarithmic returns of the normalized price variable defined as the best displayed mid-quote price in the order book. The variable “Liq Var 1 Returns” represents the logarithmic returns of the liquidity variable defined as in Esser and Moench (2005) and denoted as $\rho_{ho(1)}$ while the variable “Liq Var 2 Returns” represents the logarithmic returns of the liquidity variable introduced in this chapter and denoted as $\rho_{ho(2)}$.

Day 1	Block							
Obs	1	2	3	4	5	10
1	-	-	-	-	-	-	-	-
2	rtn1,1							...
3	rtn2,1
4
5
...
...
...
...
28	rtn27,1	rtn27,10
JT1	0	0	0	0	0	0	1	1
JT2	1	1	0	0	0	0	0	1
CJT1	0	0	0	0	0	0	0	1

Figure 3.13: Illustrative Example of Data re-sampling for Test Computation.

The Figure shows an illustrative example of the data re-sampling exercise performed for the first day of the week. The data is initially observed at a tick-by-tick level and subsequently re-sampled over a frequency of 1, 5 and 10 seconds. In order to limit the impact of micro-structural noise, the data is firstly normalized so that the average over the whole sample is set at 100 and then pre-averaged using a window of $kp = 300, 60$ and 30 observations for frequencies of 1, 5 and 10 seconds respectively. The re-sampled and pre-averaged data set consists of $n = 1331$ observations for the whole week from May 3rd to May 7th, 2010. In the example we assume that the data are re-grouped in blocks of equal size containing $m = 27$ returns. In particular we have $n = 288$ re-sampled and pre-averaged observations for the first four days and $n = 179$ observations during the last day of the week.

Jumps and Co-Jump Event							
Blocks	JT1	JT2	CJT	PJT1	PJT2	PCJT	ECJT
1	0	0	0	-	-	-	-
2	0	1	0	0	0	0	0
3	0	0	0	0	0	0	0
4	1	1	1	0	0	0	1
5	1	1	1	1	1	1	0
6	0	0	0	0	0	0	0
...
135	0	0	0	0	0	0	0

Permanent Jump Event							
Blocks	JT1	JT2	CJT	PJT1	PJT2	PCJT	ECJT
1	0	0	0	-	-	-	-
2	0	1	0	0	0	0	0
3	0	0	0	0	0	0	0
4	1	1	1	0	0	0	1
5	1	1	1	1	1	1	0
6	0	0	0	0	0	0	0
...
135	0	0	0	0	0	0	0

Permanent Jumps and Co-Jump Event							
Blocks	JT1	JT2	CJT	PJT1	PJT2	PCJT	ECJT
1	0	0	0	-	-	-	-
2	0	1	0	0	0	0	0
3	0	0	0	0	0	0	0
4	1	1	1	0	0	0	1
5	1	1	1	1	1	1	0
6	0	0	0	0	0	0	0
...
135	0	0	0	0	0	0	0

Exogenous Co-Jump Event							
Blocks	JT1	JT2	CJT	PJT1	PJT2	PCJT	ECJT
1	0	0	0	-	-	-	-
2	0	0	0	0	0	0	0
3	1	1	1	0	0	0	1
4	1	1	1	1	1	1	0
5	0	1	0	0	1	0	0
6	1	1	1	0	1	0	0
...
135	0	0	0	0	0	0	0

Figure 3.14: Illustrative Example of Data re-sampling for Test Computation.

The Figures show an illustrative example of the jump and co-jump test construction when tick-by-tick data are used. The data is observed at a tick-by-tick level, subsequently pre-averaged over a window of $kp = 90$ observations and re-grouped in blocks of $m = 100$ observations. $JT_{(1)}$ and $JT_{(2)}$ are jump test indicator functions for liquidity and price respectively as defined in Chapter 2. A value equal to one indicates the rejection of the null of no jumps at a significance level $\alpha = 5\%$. CJT is a co-jump test indicator function which indicates the presence of a contemporaneous co-jump. The signal vectors PJT_1 and PJT_2 assume values equal to one when the null of no permanent jump is rejected and values equal to zero otherwise. Finally, the vectors $PCJT$ and $ECJT$ indicate the presence of a permanent and an exogenous co-jumps.

CONCLUSIONS AND FURTHER RESEARCH

The main motivation of our research was to explore liquidity discovery models and analyze the behavior of a direct measure of liquidity derived from the foreign exchange markets. Our main goal was also to measure the price impact of liquidity shocks and to construct a robust testing methodology to distinguish between transitory-permanent and exogenous-endogenous co-jumps in price and liquidity in the context of ultra high frequency data. In Chapter 1, we used a specific measure of liquidity, defined as available liquidity, and directly observed from the foreign exchange markets as opposite to a liquidity proxy measured by transactional volumes or inferred from trading frequency or other microstructural variables. The dynamic behavior of liquidity was estimated by allowing time to have a deterministic and a stochastic component. We made a further distinction by identifying a stochastic liquidity and transactional time and modelled conditional expected durations using a regime switching threshold representation in order to incorporate a state dependent trading intensity. We found a robust empirical evidence of a strong negative relationship between both the levels and the changes in trading activity and the changes in the amount of available liquidity in a limit order book, strong and persistent autocorrelation effects and a significant impact of time durations in both the liquidity and in the trade activity process especially at the first

lags. Impulse response functions were used to study the impact of exogenous shocks, in the form of unexpected trade activity, on liquidity. We also found that unexpected trade activity has an initial but only temporary negative impact. During times of high volatility and intense turnover, liquidity adjusted quickly to the equilibrium level reached in the previous state. During times of low volatility and poor market activity, instead, the adjustment process became slower and more erratic. In addition to time dependence, we have also evaluated the impact of a number of microstructural variables. A strong statistically significant negative relationship between price spreads and market activity and changes in displayed liquidity was observed, while other variables like net order imbalance and market impact showed weak if not insignificant relationship.

In Chapter 2, we considered different combinations of univariate tests for jumps and proposed a co-jump testing methodology in order to detect statistically significant common jumps between two correlated stochastic processes. In particular, we introduced a testing procedure in the case where the data are either re-sampled over equally spaced time intervals or observed at a tick-by-tick level and, accordingly, proposed different tests to identify the presence of *contemporaneous*, *permanent* and *lagged* or *exogenous* co-jump events. A Monte Carlo experiment assessed the statistical properties of the univariate tests for jumps and, subsequently, evaluated the statistical properties of the proposed co-jump testing procedure under different levels of the jump intensity factor, jump size, correlation and microstructural noise. In our simulation exercise, we found a strong sensitivity of the proposed co-jump testing procedure to the jump intensity variable λ and, in particular, a positive relationship between the number of correctly identified co-jumps and the jump intensity factor. While we found very little sensitivity

of the co-jump tests for re-sampled data to changes in the correlation factor, the size corrected power of the co-jump tests for tick data was positively affected by an increase in correlation. We also observed a big size distortion of the proposed co-jump testing procedure under different types of microstructural noise. In particular, we found that the noise caused by rounding effects can severely affect the size of the tests as shown in the case of the LBNS, the RBNS, the MinRV and the CPR tests. We also reported that the proposed co-jump testing procedure was robust to different levels of noise as the power of the tests was not particularly affected. Overall, the strongest performance, in terms of power, was displayed by the LBNS, the RBNS and the JO followed by the MedRV tests while the PZ and the CPR tests were the most affected by microstructural noise.

In Chapter 3, we analyzed the contribution of liquidity shocks to systemic risk and contagion and, in particular, explicitly assessed the transmission mechanism between the EUR/USD FX spot price and the liquidity dynamics observed from a representative order book during the week from May 3 to May 7, 2010. The time frame considered was particularly relevant given that EUR/USD spot displayed a large move opening on Monday at around 1.3250 and ending the session on Friday just above 1.2750. The empirical exercise was divided in two main parts. In the first part, we used the first battery of jump and co-jump tests on data re-sampled over equally spaced time intervals and found that the percentage number of identified jumps in both liquidity and price tends to decrease as we move from a high to a low observation frequency. With the exception of the MinRV and the MedRV tests, we also detected a lower number of liquidity than price jumps across all the observation frequencies used. The high

number of jumps detected by the MinRV, the MedRV, the CPR and the PZ test was probably affected by the level of microstructural noise. The LBNS test, followed by the RBNS, the JO and the ABD-LM tests were characterized by a higher ability to detect contemporaneous jumps and co-jumps. Very little evidence of lagged co-jumps was observed, in particular, at frequencies greater than one to five seconds. The result indicated that jumps in liquidity observed at a frequency of one second are usually not followed by statistically significant price jumps at lower frequencies. In the second part of the empirical application, we used tick-by-tick data with no re-sampling and computed the second battery of tests. The tests allowed us to distinguish between different jump and co-jump events and in particular to measure the number of contemporaneous and permanent jumps and co-jumps together with exogenous co-jumps between two different liquidity measures and the mid-quote spot price of EUR/USD FX. The PZ and the ABD-LM were particularly affected by microstructural noise and, in particular, when a small pre-averaging window was used. As we considered a different number of observations per block, we also noticed a substantial variation in the percentage of detected liquidity and price jumps.

There are a number of areas for further research. In Chapter 1, we studied the behavior of available liquidity at k -ticks from the best displayed price in the order book. However, it would be interesting to evaluate the relationship between price dynamics and changes in available liquidity also using a more extensive data-set or to relate the bid to the ask side of the order book. In Chapter 2, we presented a co-jump testing procedure based on the combination of univariate tests for jumps. In particular, we used intersections and re-unions to combine a number of test statistics characterized

by a known statistical distribution and construct the co-jump test indicator functions. It would be interesting to study the statistical properties of the co-jump test indicator function and to create a co-jump statistic based on combinations of p -values where some form of dependence is allowed. Finally, in Chapter 3 we have not distinguished between trading regimes or time-of-the-day effects. Liquidity is shown to play an important role in the context of microstructural contagion and shocks to liquidity, especially when driven by informed trading, tend to have a more permanent impact on prices. We would expect a higher number of endogenous co-jumps to be detected in proximity of the opening or the closing of a trading session when both trading volatility and displayed liquidity are low as also observed from the data. It would then be particularly interesting to run the co-jump testing procedure across different trading times during the day and isolate diurnal effects. Also, despite the large sample used in the empirical analysis, we would need to extend the time interval considered and measure the robustness of the proposed testing methodology to different trading cycles. Moreover, it would be interesting to assess if the transmission mechanism between liquidity and prices is stable across different trading regimes or if it can be affected by other microstructural variables. Finally, the transactional variables, used in the analysis, belong to one side of the limit order book. We would need to study the dynamics of the entire order book in order to detect an asymmetric response of the bid vs. offer price to a liquidity shock. We leave these developments to future research.

BIBLIOGRAPHY

- [1] Adalid, Ramon and Carsten Detken, 2007, Liquidity shocks and asset price boom/bust cycles, ECB Working Paper Series 732.
- [2] Allen, Franklin, Aneta Hryckiewicz, Oskar Kowalewski and Gunseli Tumer-Alkan, 2011, Transmission of bank liquidity shocks in loan and deposit markets: the role of interbank borrowing and market monitoring, Wharton Financial Institutions Center Working Paper 10-28.
- [3] Almgren, Robert and Neil Chriss, 1999, Value under liquidation, *Risk* 12, 61-63.
- [4] Almgren, Robert and Neil Chriss, 2000, Optimal execution of portfolio transactions, *Journal of Risk* 3, 5-39.
- [5] Amihud, Yakov and Haim Mendelson, 1980, Dealership market: Market-making with inventory, *Journal of Financial Economics* 8, 31-53.
- [6] Andersen, Torben G., Tim Bollerslev and Dobrislav Dobrev, 2007, No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: theory and testable distributional implications, *Journal of Econometrics* 138, 125-180.

- [7] Andersen, Torben G., Dobrislav Dobrev and Ernst Schaumburg, 2012, Jump robust volatility estimation using nearest neighbor truncation, *Journal of Econometrics* 169, 75-93.
- [8] Avellenada, Marco and Sasha Stoikov, 2008, High-frequency trading in a limit order book, *Quantitative Finance* 8, 217-224.
- [9] Bannouh, Karim, Dick van Dijk and Martin Martens, 2009, Range-based covariance estimation using high-frequency data: the realized co-range, *Journal of Financial Econometrics* 7, 341-372.
- [10] Banti, Chiara, Kate Phylaktis and Lucio Sarno, 2012, Global liquidity risk in the foreign exchange market, *Journal of International Money and Finance* 31, 267-291.
- [11] Barndorff-Nielsen, Ole E. and Neil Shephard, 2004a, Power and bipower variation with stochastic volatility and jumps (with discussion), *Journal of Financial Econometrics* 2, 1-48.
- [12] Barndorff-Nielsen, Ole E. and Neil Shephard, 2004b, Measuring the impact of jumps in multivariate price processes using bipower covariation, Discussion paper, Nuffield College, Oxford University.
- [13] Barndorff-Nielsen, Ole E. and Neil Shephard, 2005, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics* 4, 1-30.

- [14] Berger, David W., Alain P. Chaboud, Sergey V. Chernenko, Edward Howorka, and Jonathan H. Wright, 2008, Order flow and exchange rate dynamics in electronic brokerage system data, *Journal of International Economics* 75, 93-109.
- [15] Bessembinder, Hendrik, 1994, Bid-ask spreads in the interbank foreign exchange markets, *Journal of Financial Economics* 35, 317-348.
- [16] Bessembinder, Hendrik, Kalok Chan and Paul J. Seguin, 1996, An empirical examination of information, differences of opinion and trading activity, *Journal of Financial Economics* 40, 105-134.
- [17] Bhaduri, Ranjan, Gunter Meissner and James Youn, 2007, Hedging liquidity risk: potential solutions for hedge funds, *Journal of Alternative Investments* 10, 80-90.
- [18] Bollerslev, Tim, Tzuo H. Law and George Tauchen, 2008, Risk, jumps and diversification, *Journal of Econometrics* 144, 234-256.
- [19] Brandt, Michael, W. and Francis X. Diebold, 2006, A no-arbitrage approach to range-based estimation of return covariances and correlations, *Journal of Business* 79, 61-74.
- [20] Breedon, Francis and Paolo Vitale, 2004, An empirical study of liquidity and information effects of order flow on exchange rates, ECB Working Paper Series 424.
- [21] Brigo, Damiano, Antonio Dalessandro, Matthias Neugebauer and Fares Triki, 2009, A stochastic processes toolkit for risk management. Mean reverting processes and jumps, *Journal of Risk Management in Financial Institutions* 2, 365 - 393.

- [22] Burghardt, Galen, Jerry Hanweck, and Lauren Lei, 2006, Measuring market impact and liquidity, *Journal of Trading* 1, 70-84.
- [23] Calem, Paul, Francisco Cavas and Jason Wu, 2011, The impact of a liquidity shock on bank lending: the case of the 2007 collapse of the private-label RMBS market, Federal Reserve Board Working Paper Series.
- [24] Cao, Charles, Oliver Hansch and Xiaoxin Wang, 2009, The information content of an open limit-order book, *Journal of Futures Markets* 29, 16-41.
- [25] Cao, Jin and Gerhard Illing, 2010, Regulation of systemic liquidity risk, *Journal of Financial Markets and Portfolio Management* 24, 31-48.
- [26] Chaboud, Alain, P., Sergey V. Chernenko, Edward Howorka, Raj S. Krishnasami Iyer, David Liu and Jonathan H. Wright, 2004, The high-frequency effects of U.S. macroeconomic data releases on prices and trading activity in the global interdealer foreign exchange market, *International Finance Discussion Paper*, 823.
- [27] Chen Shikuan, Chih-Chung Chien, and Ming-Jen Chang, 2012, Order flow, bid–ask spread and trading density in foreign exchange markets, *Journal of Banking and Finance* 36, 597-612 .
- [28] Cheng, Lan and Xuguang Sheng, 2010, Combination of “combinations of p-values”, Working Paper.
- [29] Chordia, Tarun, Richard Roll and Avanidhar Subrahmanyam, 2000, Commonality in liquidity, *Journal of Financial Economics* 56, 3-28.

- [30] Christensen, Kim, Roel C. Oomen and Mark Podolskij, 2010, Realised quantile-based estimation of the integrated variance, *Journal of Econometrics* 159, 74-98.
- [31] Cont, Rama, Sasha Stoikov and Rishi Talreja, 2010, A stochastic model for order book dynamics, *Operations Research* 58, 549-563.
- [32] Copeland, Thomas and Dan Galai, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457-69.
- [33] Corsi, Fulvio, Davide Pirino and Roberto Reno, 2010, Threshold bipower variation and the impact of jumps on volatility forecasting, *Journal of Econometrics* 159, 276-288.
- [34] Danielsson, Jon and Richard Payne, 2010, Liquidity determination in an order driven market, forthcoming in the *European Journal of Finance*.
- [35] De Jong, Frank and Peter C. Schotman, 2010, Price discovery in fragmented markets, *Journal of Financial Econometrics* 8, 1-28.
- [36] De Luca, Giovanni and Paola Zuccolotto, 2006, Regime switching Pareto distributions for ACD models, *Computational Statistics & Data Analysis* 51, 2179-91.
- [37] Dufour, Alfonso and Robert F. Engle, 2000, Time and the price impact of a trade, *Journal of Finance* 6, 2467-98.
- [38] Dumitru, Ana-Maria and Giovanni Urga, 2012, Identifying jumps in financial assets: a comparison between nonparametric jump tests, *Journal of Business and Economic Statistics* 30, 242-255.

- [39] Easley, David and Maureen O'Hara, 1992, Time and process of security price adjustment, *Journal of Financial Economics* 47, 905-927.
- [40] Engle, Robert F. and Jeffrey R. Russell, 1998, Autoregressive conditional duration: a new model for irregularly spaced transaction data, *Econometrica* 66, 1127-62.
- [41] Engle, Robert F. and Jeffrey R. Russell, 2004, Analysis of high frequency financial data, in *Handbook of Financial Econometrics* by Yacine Ait-Sahalia, Lars P. Hansen, and Jose A. Scheinkman Eds.
- [42] Espinoza, Raphaël, Charles Goodhart, and Dimitrios Tsomocos, 2008, State prices, liquidity and default, *Journal of Economic Theory* 39, 177-194.
- [43] Esser, Angelika and Burkart Moench, 2005, Modelling feedback effects with stochastic liquidity, *Lecture Notes in Economics and Mathematical Systems* 553, 9-46.
- [44] Evans, Martin D.D. and Richard K. Lyons, 2002, Order flow and exchange rate dynamics, *Journal of Political Economy* 110, 170-180.
- [45] Evans, Martin D.D., 2010, Order flows and the exchange rate disconnect puzzle, *Journal of International Economics* 80, 58-71.
- [46] Farhi, Emmanuel, Mikhail Golosov, and Aleh Tsyvinski, 2007, A theory of liquidity and regulation of financial intermediation, NBER Working Paper 12959.
- [47] Floegel, Volker, 2006, *The Microstructure of European Bond Markets*, European Business School, Deutscher Universitaets-Verlag (DUV) Edition.

- [48] Foucault, Thierry, Ohad Kadan and Eugene Kandel, 2005, Limit order book as a market for liquidity, *Review of Financial Studies* 18, 1171-1217.
- [49] Frey, Ruediger, 2000, Market illiquidity as a source of model risk in dynamic hedging, in *Model Risk* ed. by R. Gibson, RISK Publications.
- [50] Frey, Stefan and Patrick Sandas, 2008, The impact of hidden liquidity in limit order books, *CFR Working Papers* 09-06.
- [51] Gobbi, Fabio and Cecilia Mancini, 2012, Identifying the brownian covariation from the co-jumps given discrete observations, *Econometric Theory* 28, 249-273.
- [52] Goettler, Ronald, Christine A. Parlour and Uday Rajan, 2005, Equilibrium in a dynamic limit order market, *Journal of Finance* 60, 2149-2192.
- [53] Goodhart, Charles, 2007, Liquidity risk management, *Financial Markets Group Special Papers*.
- [54] Hafeez, Bilal, 2007, Currency markets, is money left on the table? *Deutsche Bank Guide to Currency Indices* 1, 9-16.
- [55] Hafeez, Bilal, 2011, How regulation will reshape FX markets, *Deutsche Bank Exchange Rate Perspective* 6, 10-21.
- [56] Harris, Lawrence and Joel Hasbrouck, 1996, Market vs limit orders: the superdot evidence on order submission strategy, *Journal of Financial and Quantitative Analysis* 31, 213-231.

- [57] Harvey, David I., Stephen J. Leybourne and A.M. Robert Taylor, 2009, Unit root testing in practice: dealing with uncertainty over the trend and initial condition, *Econometric Theory* 25, 587-636.
- [58] Harvey, David I., Stephen J. Leybourne and A.M. Robert Taylor, 2011, Testing for unit roots testing in the presence of uncertainty over both the trend and initial condition, Forthcoming in *Journal of Econometrics*.
- [59] Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.
- [60] Hasbrouck, Joel, 1993, Assessing the quality of a security market: a new approach to transaction-cost measurement, *Review of Financial Studies* 6, 191-212.
- [61] Hautsch, Nikolaus and Mark Podolskij, 2010, Pre-Averaging based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence, CREATES Research Papers 2010-29, School of Economics and Management, University of Aarhus.
- [62] Herring, Richard J., Chitru S. Fernando and Avanidhar Subrahmanyam, 2008, Common liquidity shocks and market collapse: lessons from the market for Perps, *Journal of Banking and Finance* 32, 1625-1635.
- [63] Iordanis, Kalaitzoglou and Ibrahim B. Maher, 2011, Does order flow in the european carbon allowances market reveal information ?, Durham Business School Working Paper.

- [64] Jacod, Jean and Viktor Todorov, 2009, Testing for common arrivals of jumps for discretely observed multidimensional processes, *Annals of Statistics* 37, 1792-1838.
- [65] Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij and Mathias Vetter, 2009, Microstructure noise the continuous case: the pre-averaging approach, *Stochastic Processes and their Applications* 119, 2249–2276.
- [66] Jiang, George J. and Roel C. Oomen, 2008, Testing for jumps when asset prices are observed with noise-a “swap variance” approach, *Journal of Econometrics* 144, 352-370.
- [67] Kempf, Alexander and Olaf Korn, 1998, Market depth and order size - an analysis of permanent price effects of DAX futures trades, *Journal of Financial Markets* 2, 29-48.
- [68] Kissell, Robert and Roberto Malamut, 2005, Understanding the profit and loss distribution of trading algorithms, *Institutional Investor Guide to Algorithmic Trading* 1, 41-49.
- [69] Kratz, Peter and Torsten Schoneboern, 2011, Optimal liquidation in dark pools, SFB 649 Discussion Paper 2011, Humboldt University Berlin.
- [70] Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-36.
- [71] Lahaye, Jerome, Sebastien Laurent and Christopher J. Neely, 2011, Jumps, co-jumps and macro announcements, *Journal of Applied Econometrics* 26, 893-921.

- [72] Large Jeremy, 2007, Measuring the resiliency of an electronic limit order book, *Journal of Financial Markets* 10, 1-25.
- [73] Lee, Suzanne S. and Per A. Mykland, 2008, Jumps in financial markets: a new nonparametric test and jump dynamics, *Review of Financial Studies* 21, 2535-2563.
- [74] Liao, Yin and Heather M. Anderson, 2011, Testing for co-jumps with high frequency financial data: an approach based on first-high-low-last prices, *Monash Econometrics and Business Statistics Working Paper* 9/11.
- [75] Loughin, Thomas M., 2004, A systematic comparison of methods for combining p-values from independent tests, *Computational Statistics and Data Analysis* 47, 467-485.
- [76] Lyons, Richard K., 1995, Tests of microstructural hypothesis in the foreign exchange market, *Journal of Financial Economics* 39, 321-351.
- [77] Mahanti, Sriketan, Amrut Nashikkar, Marti G. Subrahmanyam, George Chacko, and Gaurav Mallik, 2008, Latent liquidity: a new measure of liquidity with an application to corporate bonds, *Journal of Financial Economics* 88, 272-298.
- [78] Mykland, Per and Lan Zhang, 2011, Between data cleaning and inference: pre-averaging and other robust estimators of the efficient price, *Working Paper*, Department of Statistics, The University of Chicago.
- [79] Moberg, Jan-Magnus and Genaro Sucarrat, 2007, Stock market return, order flow and financial market linkages, available at SSRN: <http://ssrn.com/abstract=965649>.

- [80] Morris, Stephen and Hyun Song Shin, 2004, Liquidity black holes, *Review of Finance* 8, 1-18.
- [81] Neuhauser, Markus, 2003, Tests for genetic differentiation, *Biometrical Journal* 8, 974-984.
- [82] Pacurar, Maria, 2008, Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature, *Journal of Economic Surveys* 22, 711-751.
- [83] Pastor, Lubos and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.
- [84] Payne, Richard, 2003, Informed trade in spot foreign exchange markets: an empirical investigation, *Journal of International Economics* 61, 307-329.
- [85] Perotti, Enrico and Javier Suarez, 2011, A Pigovian approach to liquidity regulation, *Tinbergen Institute Discussion Papers* N. 11-040.
- [86] Perraudin, William and Paolo Vitale, 1996, Interdealer trade and information flows in a decentralized foreign exchange market, in *The Microstructure of Foreign Exchange Markets*, Ed. by J. Frankel, G. Galli and A. Giovannini. University of Chicago Press.
- [87] Podolskij, Mark and Daniel Ziggel, 2010, New tests for jumps in semimartingales models, *Statistical Inference for Stochastic Processes* 13, 15-41.
- [88] Riordan, Ryan and Andreas Storkenmaier, 2011, Latency, liquidity and price discovery, *KIT Working Paper Series*.

- [89] Rochet, Jean-Charles, 2008, Liquidity regulation and the lender of last resort, Banque de France Financial Stability Review 11, 45-52.
- [90] Rosu, Ioanid, 2009, A dynamic model of the limit order book, Review of Financial Studies 22, 4601-4641.
- [91] Russell, Jeffrey, Ruey S. Tsay and Michael Y. Zhang, 2001, A nonlinear autoregressive conditional duration model with applications to financial transaction data, Journal of Econometrics 104, 179-207.
- [92] Sparrow, Chris and Denis Ilijanic, 2010, The value of liquidity, The Journal of Trading 5, 10-15.
- [93] Yahav, Indal and Galit Shmueli, 2011, On generating multivariate Poisson data in management science applications, Applied Stochastic Models in Business and Industry 28, 91-102.