

**Econometrics of High Frequency Data and
Nonnegative Valued Financial Point Processes**

by

Yongdeng Xu

*A Thesis Submitted in Fulfilment of the Requirements for the Degree
of Doctor of Philosophy of Cardiff University*

Economics Section of Cardiff Business School, Cardiff University

January 2013



DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed (candidate)

Date

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of(insert MCh, Md, **MPhil**, PhD etc, as appropriate)

Signed (candidate)

Date

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references.

Signed (candidate)

Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date

Summary

Econometrics of high frequency data and nonnegative valued financial point process is addressed in an Autoregressive Conditional Duration (ACD) and Multiplicative Error Model (MEM). The basic idea is to model the nonnegative valued point process in terms of the product of a scale factor and an innovation process with nonnegative support. However, when extending such a model into a multivariate setting, the direct use of multivariate MEM model is restricted since conditional distributions for multivariate nonnegative valued random variables are often not available. A common strategy is to reduce the multivariate setting to a series of univariate problems by assuming: a) weak exogeneity. b) the independence of innovation terms. The objects of this thesis are to examine this strategy and develop a general form vector MEM. Three main Chapters have been developed.

We begin with the analysis of weak exogeneity. The independence of innovation terms is considered as a special case of weak exogeneity. The simulation study indicates that a failure of the weak exogeneity assumption implies not only inefficient but also biased estimate of the parameters. We then derive an LM test for weak exogeneity and the empirical results indicate that the weak exogeneity of duration is often rejected. Chapter 3 discusses the use of lognormal distribution for financial durations and we propose a lognormal ACD model. The empirical results show that lognormal ACD model is superior to Exponential and Weibull ACD model. It performs similarly to Burr or generalized gamma ACD model. In Chapter 4, we release weak exogeneity assumption and propose general form of vector MEM. Based on the results in Chapter 3, we further propose to use the multivariate lognormal distribution for the distribution of the vector MEM for which maximum likelihood is proved as a suitable estimation strategy. The model is then applied to the trade and quotes data from the New York Stock Exchange (NYSE) for the dynamics of trading duration, volume and price volatility. The empirical findings are generally consistent with market microstructure predictions.

Acknowledgements

I am especially grateful to my supervisor Prof. Garry Phillips and Prof. Nick Taylor for their insightful guidance and persistent encouragement through my work. In particular, I have benefitted incredibly from those regular and irregular meetings I have had with Prof. Garry Phillips.

I would also like to thank Dr Simone Manganelli in European Central Bank, for providing me his TAQ data and the Gauss programme for preparing the data.

Special thanks go to the financial sponsors that I have had. Cardiff Business School provides me one of its PhD grants to support my study and living in the UK. Prof. Patrick Minford provides me teaching opportunities. I am completely grateful to all of them.

I am also very grateful to my friends and colleagues at Cardiff Business School for making my time at university yet more rewarding. Among them, I shall mention especially, Panagiotis Tziogkidis, Xiaosui Wu, Thomas Wilkinson, Dr Yang Zhao, Fangzhou Huang, Kun Tian, Peng Zhou, Shun Jiang, Yu Sun.

Last but not least, I would like to express my great gratitude to my father and mother, for their continuing spiritual support and care. Special thanks and deep love go to my girlfriend, Wenna Lu, for her great support, understanding and encouragement in these years.

List of Contents

Chapter 1 Introduction and Main Contributions	1
1.1 <i>High Frequency Financial Data</i>	2
1.1.1 The Institutional Framework.....	2
1.1.2 Database and Trading Variables	5
1.1.3 Matching Trades and Quotes	7
1.1.4 Data Characteristics	7
1.2 <i>The Theoretical Background of Market Microstructure</i>	9
1.3 <i>The Empirical Modelling of High Frequency Data</i>	11
1.3.1 ACD models.....	11
1.3.2 Multiplicative Error Model	14
1.3.3 Multivariate MEM -- A Recursive Framework	15
1.4 <i>Motivations and main Contributions of this Thesis</i>	18
Chapter 2 Weak Exogeneity in the Financial Point Processes.....	21
2.1 <i>Introduction</i>	21
2.2 <i>Related Studies on Weak Exogeneity</i>	22
2.3 <i>Methodology</i>	24
2.3.1 Formal Definition of Weak Exogeneity.....	24
2.3.2 Different Types of Weak Exogeneity in Financial Point Processes	26
2.4 <i>Consequences of Incorrectly Assuming Weak Exogeneity- a Simulation Study</i>	32
2.5 <i>An LM Test for Weak Exogeneity in Financial Point Processes</i>	39
2.5.1 Testing Weak Exogeneity of Duration	41
2.5.2 Testing Joint Weak Exogeneity of Duration and Volume.....	42
2.5.3 Power of the Test	44
2.6 <i>Empirical Analysis</i>	47
2.6.1 Data	47

2.6.2 Testing for Weak Exogeneity - Empirical Results	50
2.7 Conclusion	52
Chapter 3 The Lognormal Autoregressive Conditional Duration (LNACD)	
Model and a Comparison with Alternative ACD Models.....	53
3.1 Introduction.....	53
3.2 ACD Models.....	55
3.2.1 Model Specification	55
3.2.2 Density Assumption.....	56
3.3 Methodology.....	58
3.3.1 Lognormal ACD Model.....	58
3.3.2 ARMA Specification of Lognormal ACD Model	60
3.3.3 The Shape of Hazard Function of Lognormal Distribution – Numerical Analysis	61
3.4 Specification Tests.....	64
3.4.1 Testing Financial Duration Model via Density Forecasts.....	64
3.4.2 Non-parametric Testing of Conditional Duration Model	65
3.5 Empirical Application	67
3.5.1 Data	67
3.5.2 Estimation Results	69
3.5.3 Specification Test Results.....	72
3.6 Conclusion	79
Chapter 4 The Dynamics of Trading Duration, Volume and Price Volatility – a Vector MEM Model	80
4.1 Introduction.....	80
4.2 Relevant Theoretical and Empirical studies	82
4.2.1 Market Microstructure Predictions	82
4.2.2 Empirical Studies	87
4.3 Methodology.....	90
4.3.1 Duration, Volume and Price Volatility --- a Recursive Framework	90

4.3.2 Econometric Concerns	92
4.3.3 Vector MEM	94
4.3.4 Specification of \mathcal{E}_t	95
4.3.5 Impulse Response Function	99
4.3.6 Vector ARMA Representation.....	100
4.4 Empirical Analysis	101
4.4.1 Data	101
4.4.2 Empirical Model	106
4.4.3 Empirical Results	110
4.4.4 Impulse Response Analysis	112
4.5 Conclusion	115
Chapter 5 General Conclusion.....	117
References	121
Appendices.....	129
Appendix 1 Proof of Case 1 Weak Exogeneity	129
Appendix 2 Proof of Case 2 Weak Exogeneity	130
Appendix 3 Typology of ACD Models	131
Appendix 4 Survive and Hazard Rate.....	132
Appendix 5 Hazard Rate of Lognormal Distribution	133
Appendix 6 The Visual Diagnostic Checks for other Three Stocks	134
Appendix 7 Lognormal Distribution.....	138
Appendix 8 Proofs of Impulse Response Function.....	141

List of Tables

Table 1-1: TAQ data recorded on trades for AIRGAS on January 02, 1998.....	6
Table 1-2: TAQ data recorded on quotes for AIRGAS on January 02, 1998.....	7
Table 2-1: Case 1 simulation summary statistics. Estimated parameters	34
Table 2-2: Case 2 simulation summary statistics. Estimated parameters	35
Table 2-3: Case 3 simulation summary statistics. Estimated parameters	36
Table 2-4: Power of the test	45
Table 2-5: Stock used in this analysis	47
Table 2-6: Summary statistics for the 10 stocks	50
Table 2-7: Weak Exogeneity Test -- LM Test Statistics.....	51
Table 3-1: Descriptive statistics for the price durations.	68
Table 3-2: ACD model estimates results	71
Table 3-3: Specification test results.....	78
Table 4-1: Summary of the related market microstructure literature	86
Table 4-2: Stocks used in the analysis	102
Table 4-3: Summary statistics for the 10 stocks	103
Table 4-4: Estimation results and diagnostics: frequently traded stocks.	108
Table 4-5: Estimation results and diagnostics: infrequently traded stocks.....	109
Table 4-6: Time (in hours) it takes to absorb shocks to the long term equilibrium variances.....	114

List of Figures

Figure 2-1: Power of the test.....	46
Figure 2-2: Nonparametric estimate of daily pattern of transaction duration and return square.	49
Figure 3-1: Density and Hazard of lognormal and Burr distribution.....	59
Figure 3-2: Hazard of a lognormal distribution with different σ	62
Figure 3-3: Non-parametric and parametric densities: BA Out sample results	74
Figure 3-4: Non-parametric and parametric densities: XON Out sample results	74
Figure 3-5: Histogram of Z: BA Out sample results.....	75
Figure 3-6: Histogram of Z, XON out of sample results.	76
Figure 3-7: Autocorrelation of z for BA out sample.....	77
Figure 3-8 : Autocorrelation of z for XON out of sample	77
Figure 4-1: A comparison of parametric densities.....	97
Figure 4-2: A comparison of parametric density and non-parametric densities--LSB	105
Figure 4-3: Impulse response function for TRN	113
Figure 4-4: Impulse response function for ABG	113

Chapter 1 Introduction and Main Contributions

High frequency data is also known as *transaction data*, *(ultra) high frequency data*, or *tick data*. It has become widely available in economics and finance over the past decade. As a result of the availability of these data sets and the rapid advance in computing power, there is a growing interest in models based on high-frequency financial data. Statistically speaking, high frequency data are realizations of point processes, that is, the arrival of the observations is random. This, jointly with other unique features of financial data (long memory, strong skewness, and kurtosis) implies that new methods and new econometric models are needed. This has created a new body of literature which is often referred to as "the econometrics of (ultra-) high-frequency finance" or "high-frequency econometrics". Bauwens, Pohlmeier et al. (2008) , in a book covering recent developments, illustrate high frequency financial econometrics as a combination of observed high frequency data, market microstructure theory, and econometric modelling. The three aspects form a system in which each component nicely dovetails with the others. Market microstructure theory deals with models explaining price and agent's behaviour in a market governed by certain rules. On the other hand, empirical analysis deals with the study of market behaviour using real financial data. For example, what are the relationships between traded volume, trade time, and price variations? How does the trade activity reflect information content in the fundamental asset price?

In this chapter, we discuss the three aspects of high frequency econometrics as an introduction of this thesis and then outline the motivations and main contributions. The remainder of this chapter is organized as follows. Section 1 gives an overview of institutional background of financial market and discusses the types of high frequency financial data. Section 2 provides a compact overview of major branches of market microstructure theory. We briefly explain the main principles of the information based model and inventory based models. Section 3 reviews the econometric approach to model the dynamics of high frequency data. In particular, ACD and MEM models are discussed in this section. Section 4 is the motivations and main contributions of this thesis.

1.1 High Frequency Financial Data

This section gives an overview of institutional background of financial market and different types of high frequency financial data. We first introduce to the institutional framework of trading. And then we discuss the different types of high frequency data and illustrate its time series patterns as well as the problems which should be taken into account in empirical analysis.

1.1.1 The Institutional Framework

When introducing the institutional framework of trading, we limit our discussion to those aspects that are closely related to empirical part of this thesis, particularly the types of markets, types of traders, and types of orders.

(1) Types of Markets

Based on the existence of market makers, two types of markets are identified: *quote driven (dealer) markets* and *order driven markets*.

In a *quote driven (dealer) market*, trades are only executed by market makers (dealers). The market makers quote the bid and ask prices by standing on the opposite side of the market. In a pure quote driven market, the traders are not allowed to trade themselves but must execute their trades by market makers. The market makers earn profit from the bid-ask spread and provide liquidity to the market. Despite the high availability of automation and electronic trading system, most markets, including very active ones such as the foreign exchange market, rely on market makers to act as intermediaries. Other examples are the NASDAQ stock market, New York Stock Exchange and London Stock Exchange.

In an *order driven market*, traders trade directly with each other. Since there are no market makers serving as the intermediaries, trading occurs according to specific rules. It is commonly structured as an automatic *limit order book market*. With a limit order, an investor associates a price with every order such that the order will be executed only if the investor receives that price or better. Effectively, the limit order providers supply liquidity. The studies by Harris and Hasbrouck (1996), and Foucault (1999), among others provide the knowledge of liquidity provision in the limit order book. In markets where dealers are also present, limit orders directly compete with them and serve as a check on their market power. On the NYSE, for example, the specialist can only trade after all limit orders at the best bid or offer order have been fulfilled.

(2) Types of Traders

Based on asymmetric information, three types of traders are identified: *informed traders*, *uninformed traders* and *market makers*. *Informed traders* are usually defined as a corporate officer with private information. *Uninformed traders* are mainly liquidity motivated, who simply behave as their belief of current information. *Market*

makers are assumed uninformed. Informed traders hope to get profit from their information while the market maker loses to informed traders on average. But the market makers are specialists and they can access the information by learning the signals in the market such as trading directions and volumes, thus recouping these losses on noise traders. Glosten and Milgrom (1985) assume that the bid-ask spread is increasing in information asymmetry. And different trader type's behavior is reflected in the bid price and ask price.

(3) Type of Orders

An order represents all the relevant trade information, such as what to trade, when to trade and how much to trade. A *bid (ask) order* reflects a trader's willingness to buy (sell) and contains the respective price and quantity the trade will accept. *Bid and ask prices* are the price that the trades are willing to trade. The highest (lowest) bid (ask) price available is called the *best bid (ask) price* or *best bid (ask) quote*. A *market quotation* gives the best bid and offer (ask) in the market and is called *best bid and offer (BBO)*. In US, the best bid and offer across consolidated markets for National Market System (NMS) stock is called the *National Best Bid and Offer (NBBO)*. The difference between the best ask and bid is called the *bid-ask spread*.

A *market order* is an order that trades immediately at the best price currently available in the market. The corresponding price at which the order is executed is called *transaction price*. Market order traders "pay" the bid-ask spread as long as the order is filled with the offered quantity at the best ask or bid price. If the size of the market order is larger than the quantity offered at the best ask or bid, the trader must move prices and thus has to pay an extra premium ("*price concession*"). Then, buyers (sellers) have to bid prices up (down) in order to find a counter-party who is willing to take the other side a large trade. The resulting price movements are called

(instantaneous) market impacts or price impacts and naturally increase with the order size and are the dominant part of the trading costs (on top of the bid-ask spread). These trading costs induced by a potential market impact and the execution price uncertainty are often referred to as the price traders have to pay to obtain priority in the market, i.e., the “price of immediacy”.

A *limit order* is a trade instruction to trade at a price which is no worse than the so-called *limit price* specified by the trader. As the corresponding limit price is not necessarily offered on the other side of the market, a limit order faces execution risk. If no one is willing to take the opposite side at the required limit price, the order is not executed and is placed in the *limit order book* where all non-executed limit orders are queued according to price and time priority. Correspondingly, the larger the distance between the limit order and the best quote, the worse is the order’s position in the queue and the lower is its execution probability in given time. A limit order with a limit price at or above (below) the best ask (bid) price in case of a buy (sell) order is executed immediately and, if necessary, filled until the limit price level is reached. Such an order is called a *marketable limit order* corresponding to a market order where the trader limits the potential price impact (by correspondingly setting the limit price). Finally, a *market-to-limit order* is a market order, which is executed at the best ask/bid quote in the order book. Any unfilled part of a market-to-limit order automatically enters the order book.

1.1.2 Database and Trading Variables

The high frequency data are the data that are recorded whenever a trade, quote or a limit order occurs. This data is also called *transaction data*, *(ultra) high frequency data*, or *tick data*. High frequency data is widely used in the analysis of market microstructure theory. The most popular and widely used database is the Trades and

Quotes (TAQ) dataset released by the NYSE. It contains detailed information of the intraday trade and quote process at NYSE, NASDAQ and other local exchanges in the US. They are all quote driven markets. The TAQ database consists of two parts: the first reports the trade data, while the second lists the quote data posted by the market maker. The trade dataset contains trade volume, trade price, and the exact trade time (to the second). And the quote dataset contains bid (offer) price, bid (offer) size and the exact quote time(to the second). Table 1-1 and Table 1-2 show extracts of raw files from the “Trade and Quote” (TAQ) database released by the NYSE.

Table 1-1: TAQ data recorded on trades for AIRGAS on January 02, 1998

SYMBOL	DATE	EX	TIME	PRICE	SIZE	CORR
ARG	980102	N	93858	1412500	1500	0
ARG	980102	M	93900	1412500	200	0
ARG	980102	N	93904	1412500	1000	0
ARG	980102	T	94220	1412500	800	0
ARG	980102	N	94257	1425000	500	0
ARG	980102	N	94319	1425000	500	0
ARG	980102	N	94346	1431250	2000	0
ARG	980102	N	94357	1431250	2000	0
ARG	980102	N	94536	1431250	100	0
ARG	980102	N	94618	1437500	1000	0
ARG	980102	N	94627	1437500	1000	0
ARG	980102	N	95403	1437500	400	0

SYMBOL: stock symbol, *DATE*: trade date, *EX*: *exchange* on which the trade occurred, *TIME*: trade time, *PRICE*: transaction price, *SIZE*: trade size, *CORR*: correction indicator of correctness of a trade

Table 1-2: TAQ data recorded on quotes for AIRGAS on January 02, 1998

SYMBOL	DATE	EX	TIME	BID	BID SZ	OFFER	OFF SZ	MOOD
ARG	980102	N	93915	1406250	50	1425000	10	10
ARG	980102	X	93916	1350000	1	1450000	1	12
ARG	980102	M	93918	1300000	4	1625000	5	12
ARG	980102	M	93918	1387500	1	1437500	1	12
ARG	980102	T	93918	1400000	1	1437500	1	12
ARG	980102	T	93918	1400000	1	1437500	1	12
ARG	980102	P	93919	1393750	1	1437500	1	12
ARG	980102	B	93920	1387500	1	1437500	1	12
ARG	980102	N	94311	1406250	50	1437500	10	12
ARG	980102	X	94313	1350000	1	1462500	1	12
ARG	980102	T	94314	1400000	1	1450000	1	12
ARG	980102	T	94314	1400000	1	1450000	1	12

SYMBOL: stock symbol, *DATE*: quote date, *EX*: exchange on which the trade occurred, *TIME*: quote time, *BID*: bid price, *BID SZ*: bid size in number of round lots (100 shares), *OFFER*: offer (ask) price, *OFF SZ*: offer size in number of round lots (100 shares), *MODE*: quote condition

1.1.3 Matching Trades and Quotes

As in many other quote driven markets, the trade and quote data in the TAQ dataset are recorded separately, which raises the issue of appropriately matching the two datasets. The matching procedure is necessary whenever the analysis has to link the trade characteristics, like trade sizes and trade time, to the prevailing quote updating process (for example, Engle (2000), Manganelli (2005)). For NYSE data, Lee and Ready (1991) propose a 5 second rule to reduce the potential mismatching problem. Specifically, a trade is associated with a quote posted at least 5 seconds before the trade, since the quotes can be posted more quickly than trades are recorded. Lee and Ready (1991) show that this rule leads to the lowest rates of mismatching. This procedure becomes a standard rule in microstructure studies.

1.1.4 Data Characteristics

The high frequency data, commonly of most interest, are time stamps of trades, the best bid/ask quote updates, the traded volume, and the best bid-ask price. They

share some common properties, such as being irregularly spaced in time, non-negatively valued, discreteness of price change, temporal dependence, and intraday seasonality.

(1) Irregularly spaced in time

The transactions data are inherently irregularly spaced in time. Perhaps this is the most important property. As we can see from Table 1-1 and Table 1-2, some transactions appear to occur only seconds apart while others, for example at 9:54 in trade data, may be five or ten minutes apart. The result is that the commonly used econometric models, which are specified for fixed intervals, are not applicable for this analysis. One possibility is to interpolate the irregularly spaced data over fixed intervals. Alternatively, if the time interval itself is of interest, then its stochastic property needs to be taken into account.

(2) Discreteness of price change

For transaction data, institutional rules often restrict minimum price change, which is called a tick. The transaction price change must fall on multiples of ticks. In a market for an actively traded stock it is not generally common for the price to move a large number of ticks from one transaction to another. US stocks have undergone a transition from trading in 1/16th of a dollar to decimalization. For example, NYSE permitted 1/16th prices. This discreteness has an impact on many aspects of the market; for example, market liquidity, measuring volatility, or any characteristic of prices that is small relative to the tick size.

(3) Intraday seasonality

It is well known that intraday data, such as duration, volume and volatility exhibit strong intraday periodic components, with a high trading activity at the beginning and

end of the day. For most stock market's volatility, the frequency of trades, volume, and spreads all typically exhibit a U-shaped pattern over the course of the day.

(4) Temporal dependence

Unlike their lower frequency counterparts, high frequency financial returns data typically display strong dependence. The dependence is largely the result of price discreteness and bid-ask bounce.

Econometric frameworks and models to capture these specific properties are discussed in section 1.3. Before that, the theoretical background is discussed in the following section.

1.2 The Theoretical Background of Market Microstructure

In this section, we give a compact overview of the market microstructure literatures. As stated in O'Hara (1995) and Madhavan (2000), market microstructure deals with the topics, such as price discovery, inventory, liquidity, information diffusion and dissemination in market, the behaviour of the market participants. It provides theoretical explanation on the models of high frequency data. Typically, the market microstructure literature explain the trading activity using two types of models: asymmetric information based and inventory based models. Specifically, trading occurs either for information motivated or liquidity motivated reasons. The predictions of the relations between duration, volume and price volatility differ.

In the information-based model, three types of traders are assumed: informed traders; uninformed traders; and market makers. Informed traders are usually defined as corporate officers with private information, while uninformed traders are liquidity motivated and simply behave according to their current information. Market makers are also assumed to be uninformed. Apparently, the different traders have asymmetric information. Informed traders hope to obtain profits from their information so, on

average, the market makers lose out to the informed traders. Market makers are specialists and can access information by reading the signals in the market, such as trading intensity and volume, and can thus recoup any losses as uninformed traders. Their activities are covered by the sequential trade model (Glosten and Milgrom 1985; Diamond and Verrecchia 1987) and the strategic trade model ((Kyle 1985; Admati and Pfleiderer 1988; Easley and O'Hara 1992).)

In the sequential trade framework, the market maker and market participants behave competitively. Trades take place sequentially, with only one trader allowed to transact at any given point in time. Informed traders would like to trade as much (and as often) as possible. So the market maker would quickly (perhaps instantly) adjust prices to reflect this information. It is apparent that trading volume is positively (perhaps contemporaneously) correlated with price volatility. The strategic model allows the agents to act strategically. For example, in order to make full use of their private information, the informed traders may conceal their trading type by timing their trades carefully or choosing their trade sizes (Kyle 1985; Easley and O'Hara 1992). Uninformed traders may also learn by observing the actions of informed traders. In particular, Admati and Pfleiderer (1988) distinguish two types of uninformed traders in addition to informed traders: non-discretionary traders are similar to liquidity traders in the previous model; while discretionary traders, while uninformed, trade strategically. Discretionary traders choose the timing of their trades. They usually select the same period of transaction in an attempt to minimize adverse selection costs, and informed traders follow the pattern introduced by discretionary traders.

In inventory based models, the trading process is effectively motivated by the market makers' desire to keep their inventory position at some specific level. Based on their inventory position and uncertainty about order flow, dealers alter their bid and

ask prices to elicit the desired imbalance of buy and sell orders thereby moderating deviations in order flow. The dealer's action in the market is simply independent of information. It only depends on trading costs, the dealer's previous position and net demand to the dealer (Ho and Stoll 1981; O'Hara and Oldfield 1986).

These types of model generally induce patterns of various trade characteristics, such as timing, price and volume. These factors contain information and reflect trade behaviour in the market.

1.3 The Empirical Modelling of High Frequency Data

The transactions data are inherently irregularly spaced in time. The result is that the commonly used econometric models, which are specified for fixed intervals, are not applicable for this analysis. One possibility is to interpolate the irregularly spaced data over fixed intervals, but some important information may lose. Alternatively, if the time interval itself is of interest, then its stochastic property needs to be taken into account. This, jointly with other unique features (such as long memory; strong skewness; and kurtosis) implies that new methods and new econometric models are needed. It was first addressed first addressed, by Engle and Russell (1998) in the context of an ACD model for the dynamics of transaction time and then extended by Engle (2002) and Manganelli (2005) in the context of an Multiplicative Error Model for the dynamics of other nonnegative valued financial point processes(for example, trade volume, bid-ask spread, different measurement of price volatility).

1.3.1 ACD models

The basic reference ACD model is proposed by Engle and Russell (1998) whose explicit objective is the modelling of times between events. There are at least two reasons to model transaction time in the ACD literature. First, the high frequency data

are naturally irregularly spaced in time. As long as transaction time is measured, the other trading variables (for example, trade volume, and price) can be modelled associated with transaction time. Second, the time interval itself is of interest. Market microstructure literature that is based on asymmetric information (Easley and O'Hara (1992) and Easley, Kiefer et al. (1997)) argue that the transaction time convey information and have a deep impact on the behaviour of market agents, thus should be modelled as well.

Let d_t be the time duration between events occurring at time t_i and t_{i-1} , such that $d_t = t_i - t_{i-1}$. The basic idea of ACD model is that duration can be modelled as product of its conditional expectation and an innovation term with nonnegative support,

$$d_t = \psi_t \varepsilon_t \quad (1.1)$$

where $\psi_t = E(d_t | \mathcal{F}_{t-1})$ and \mathcal{F}_{t-1} denotes the information available up to period t_{i-1} .

The ACD model is further characterized by the assumption that the innovation terms ε_t are independently and identically distributed (*i.i.d*). The second equation of ACD model is that the conditional expected duration is modelled as a linear function of past duration and past expected duration:

$$\psi_t = \omega + \alpha d_{t-1} + \beta \psi_{t-1}. \quad (1.2)$$

To ensure positivity of the conditional expected duration, common restrictions on the coefficients are that $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. Bauwens and Giot (2000) also propose a logarithmic version of the ACD model to guarantee the positivity. Two specifications are considered, referred to as Log-ACD1 and Log-ACD2, respectively:

$$\log \psi_t = \omega + \alpha \log d_{t-1} + \beta \log \psi_{t-1}, \quad (1.3)$$

$$\log \psi_t = \omega + \alpha \varepsilon_{t-1} + \beta \log \psi_{t-1}. \quad (1.4)$$

Since the ACD model is very similar to the GARCH model, it is not surprising that the linear ACD model can be extended in several ways. A flexible specification is the augmented ACD (AACD) model by Fernandes and Grammig (2006). It is obtained by a Box-Cox transformation and permits an asymmetric response to small and large shocks. The first-order parameterization is given by:

$$\psi_t^\lambda = \omega + \alpha \psi_{t-1}^\lambda [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^v + \beta \psi_{t-1}^\lambda . \quad (1.5)$$

Other specifications of the ACD model can be found in a summary paper Pacurar (2008); for instance, the Zhang, Russell et al. (2001)'s threshold ACD model and Bauwens and Veredas (2004)'s stochastic conditional duration model. One can also incorporate additional regressors in the ACD model to examine the microstructure effect.

To close the model, the specification of the conditional distribution of innovation terms is needed. By definition, ε_t is a random variable with probability density function defined over a nonnegative support. Engle and Russell (1998) initially consider the exponential distribution for the error ε_t with density

$$f(\varepsilon_t) = \exp(-\varepsilon_t), \quad \varepsilon_t \geq 0 . \quad (1.6)$$

The exponential distribution has a flat hazard function, which is too restrictive. Engle and Russell (1998) also consider the Weibull distribution for the error, which nests the exponential distribution as special case ($\gamma = 1$). The standard Weibull¹ density function is:

$$f(\varepsilon_t | \gamma) = \gamma (\varepsilon_t)^{\gamma-1} \exp(-\varepsilon_t^\gamma), \quad \varepsilon_t \geq 0 . \quad (1.7)$$

¹ We call it as “standard Weibull” because the scale parameter in this distribution is normalized as 1.

The Weibull distribution is flexible and permits both an increasing hazard function if $\gamma > 1$ or a decreasing hazard function if $\gamma < 1$.

Grammig and Maurer (2000) and Lunde (1999) further propose the use of the Burr distribution and generalized gamma (GG) distribution. Both the Burr and the GG distributions have two shape parameters which allow for hump-shaped hazard functions, thereby more flexible than exponential or Weibull distribution. Moreover, the Weibull and exponential distributions can be nested in the Burr and GG distributions.

1.3.2 Multiplicative Error Model

Other high frequency data, commonly of most interest, are the trading volume, bid-ask spread, and the price volatility. They share some common properties as duration. For example, they are irregularly spaced in time, nonnegatively valued, and persistently clustered over time. As an extension of GARCH (Bollerslev 1986) and ACD (Engle and Russell 1998) approach, Engle (2002) propose a class of models, named Multiplicative Error Model (MEM), which are particularly suitable for the dynamics of such nonnegative valued financial point process. The basic idea is to model the nonnegative valued process in terms of the product of a (conditional autoregressive) scale factor and an innovation process with nonnegative support.

Let x_t be a discrete time process defined on $[0, +\infty)$, and let \mathcal{F}_{t-1} the information available up to period $t-1$. $\{x_t\}$ follows a MEM if it can be specified as the product of an autoregressive scale factor and an *i.i.d* innovation term.

$$x_t = \mu_t \varepsilon_t \tag{1.8}$$

where $\mu_t = E(x_t | \mathcal{F}_{t-1})$, ε_t is a random variable with probability density function (pdf) defined over a $[0, +\infty)$ support. Typically, the flexible assumption of a Gamma pdf

with unit mean *i.i.d* ε_t terms is adopted and μ_t assumes following a GARCH-type process. The properties of ACD model can be applied to MEM. The base (1,1) specification of μ_t is:

$$\mu_t = \omega + \alpha x_{t-1} + \beta \mu_{t-1} . \quad (1.9)$$

The logarithmic version is also adopted to ensure positivity of the conditional expectation of x_t ,

$$\log \mu_t = \omega + \alpha \log x_{t-1} + \beta \log \mu_{t-1} . \quad (1.10)$$

Other extension for the conditional mean is also possible in the literature, for example, by adding predetermined variables or incorporating asymmetric effects in the model. The estimation of the parameters in μ_t is consistent by the Quasi maximum likelihood derivation.

The first two moment conditions of the MEM are also given by:

$$\begin{aligned} E(x_t | \mathcal{F}_{t-1}) &= \mu_t , \\ \text{Var}(x_t | \mathcal{F}_{t-1}) &= \mu_t^2 \text{Var}(\varepsilon_t) . \end{aligned} \quad (1.11)$$

The ACD model by Engle and Russell (1998) is a special case of MEM, but absolute return, trading volume, bid-ask spread and number of trades in a certain interval can be modelled with MEMs. Empirical results show a good performance of these models in capturing the stylized facts of the observed series (see, for example, Manganelli (2005); Hautsch (2008)).

1.3.3 Multivariate MEM -- A Recursive Framework

There are many instances in which the joint consideration of several nonnegative valued financial point processes is of interest. Example are joint modelling the

dynamics of trading duration, volume and price volatility for the same asset (Manganelli 2005). This motivates the multivariate extension of the univariate MEM.

It is notable that a completely parametric formulation of the multivariate MEM requires a full specification of the conditional distribution of multivariate nonnegative valued random variables. As a first step one may attempt to generalize the univariate gamma (or other exponential) distribution to a suitable multivariate version. But this is frustrated by the limitations of the multivariate Gamma distribution. Cipollini, Engle et al. (2007) find the only useful multivariate Gamma distribution only admits positive correlation, which is too restrictive.

To simplify the estimation procedure, a common strategy in this study is to reduce the multivariate estimation to a series of univariate problems. Among them, Engle (2000) proposed a recursive framework, in which the joint density of several financial point processes are decomposed into the product of the marginal density of one process and conditional densities of other processes. For example, Engle (2000) express the joint density of duration and volatility as the product of the marginal density of the duration times and the conditional density of the volatility, given the duration. Under the assumption of weak exogeneity, the duration process and price volatility process can be estimated separately. This model is further extended by Manganelli (2005) to incorporate trading volume. The recursive framework of Engle (2000) and Manganelli (2005) reduces the complexity of the model, and are widely adopted in the existing empirical literature (see, for example, Engle (2000), Dufour and Engle (2000), Manganelli (2005); Russell and Engle (2005) and Engle and Sun (2007)). This is also one of research object of this thesis. We take Manganelli (2005) model for example to illustrate this recursive framework in detail.

Define $\{d_t, v_t, r_t\}, t=1, \dots, T$ as the three-dimensional time series associated with intraday trading duration, trading volume and the return process, respectively. In particular, duration is defined as the time elapsing between consecutive trades, volume is the trade size associated with each transaction and return is measured as the mid-quote change. The trivariate trading process - duration, volume and return volatility - can be modelled as follows:

$$\{d_t, v_t, r_t\} \sim f(d_t, v_t, r_t | \mathcal{F}_{t-1}; \theta) \quad (1.12)$$

where θ is a vector incorporating the parameters of interest.

In the recursive model, the joint distribution is decomposed into the product of three components: marginal density of durations, the conditional density of volumes given durations and the conditional density of the return volatility given durations and volumes. Specially,

$$\{d_t, v_t, r_t\} \sim g(d_t | \mathcal{F}_{t-1}; \theta_d) \cdot h(v_t | d_t, \mathcal{F}_{t-1}; \theta_v) \cdot k(r_t | d_t, v_t, \mathcal{F}_{t-1}; \theta_r). \quad (1.13)$$

Manganelli (2005) considers the following MEMs for duration, volume and volatility:

$$\begin{aligned} d_t &= \psi_t(\theta_d; \mathcal{F}_{t-1})u_t, \quad u_t \sim i.i.d.(1, \sigma_u^2) \\ v_t &= \phi_t(\theta_v; d_t, \mathcal{F}_{t-1})\eta_t, \quad \eta_t \sim i.i.d.(1, \sigma_\eta^2) \\ \hat{r}_t &= \sqrt{h_t(\theta_r; d_t, v_t, \mathcal{F}_{t-1})}\xi_t, \quad \xi_t \sim i.i.d.(0, 1) \\ \text{or } \hat{r}_t^2 &= h_t(\theta_r; d_t, v_t, \mathcal{F}_{t-1})\xi_t, \quad \xi_t \sim i.i.d.(1, \sigma_\xi^2) \end{aligned} \quad (1.14)$$

where \hat{r}_t^2 is the proxy for volatility², (ψ_t, ϕ_t, h_t) are the conditional expectations of duration, volume and volatility, respectively, and $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ is a vector of s

² In order to obtain a price change sequence which is free of the bid-ask bounce that affects price, we follow Ghysels, et al. (1998) and \hat{r}_t is obtained as the residuals of an ARMA(1,1) process of return series. See also in Hautsch (2008). One advantage of using \hat{r}_t is that it avoids the problem of exact zero values in r_t .

parameters of interest. Manganelli (2005) adopts the univariate exponential distribution for the innovations in this specification.

To capture the causal and feedback effect among these variables, he specifies the following first order autoregressive conditional model:

$$\begin{aligned}\psi_t &= w_1 + (a_{11}d_{t-1} + a_{12}v_{t-1} + a_{13}\hat{r}_{t-1}^2) + (b_{11}\psi_{t-1} + b_{12}\phi_{t-1} + b_{13}h_{t-1}), \\ \phi_t &= w_2 + (a_{21}d_{t-1} + a_{22}v_{t-1} + a_{23}\hat{r}_{t-1}^2) + (b_{21}\psi_{t-1} + b_{22}\phi_{t-1} + b_{23}h_{t-1}) + a_0^{12}d_t, \\ h_t &= w_3 + (a_{31}d_{t-1} + a_{32}v_{t-1} + a_{33}\hat{r}_{t-1}^2) + (b_{31}\psi_{t-1} + b_{32}\phi_{t-1} + b_{33}h_{t-1}) + a_0^{13}d_t + a_0^{23}v_t.\end{aligned}\tag{1.15}$$

Under the restrictions of weak exogeneity and independence of the innovations terms, the three components are estimated separately. One additional advantage of the recursive model is that the contemporaneous information is also incorporated.

1.4 Motivations and main Contributions of this Thesis

Similar to Manganelli (2005), we are initially interested in the modelling of nonnegative valued financial point processes, particularly the dynamics of trading duration, volume and price volatility. However, when extending ACD/MEM model into a multivariate setting, the full specification requires the joint probability distribution of nonnegative valued random variables, hence occurrences of such distribution are limited in the literature. Instead, Engle (2000) and Manganelli (2005) propose a recursive framework, which reduce the multivariate setting to a series of univariate problems, by making the following two assuming: a) weak exogeneity. b) the independence of innovation terms. Then each process can be estimated separately. The motivations and the main contributions of this thesis are based on the two assumptions.

First, following the recursive model of Engle (2000) and Manganelli (2005), the assumption of weak exogeneity of duration is often made. If this assumption is valid, then the marginal density and conditional densities can be estimated separately.

However, the consistent estimation of the parameters is based on the weak exogeneity assumptions, which is often left untested. Moreover, the consequences of the failure of the weak exogeneity are still open questions. This motivates our second Chapter, in which we analyse weak exogeneity problem in financial point processes. We consider the independence of innovation terms as a special case of weak exogeneity and propose three cases in which the weak exogeneity condition will break down. The simulation study suggests that a failure of the exogeneity assumption implies biased estimators. The biases are very large in the third case non-weakexogeneity, which makes the econometric inferences on the parameters unreliable or even misleading. In empirical analysis, we also derive an LM test for weak exogeneity and test the weak exogeneity of duration in a trivariate (duration, volume and volatility) system. The empirical results indicate that the weak exogeneity is often rejected for frequently traded stocks, but is less likely to be rejected for infrequently traded stocks. This suggests the trivariate should be modelled and estimated jointly.

Second, the recursive model assumes that the specific processes are independent. To incorporate the contemporaneous information, Engle (2000) and Manganelli (2005) specifies causality from duration to volume and from duration and volume to price volatility. However, modelling the distribution of price as being conditional on duration and volume is just one strategy to obtain their joint distribution. As pointed out by Engle and Sun (2007), it is also possible to go from the price process and model duration conditional on its contemporaneous return. Theoretically, variation in duration and variation in the price process would be related to the same news events or the underlying information process. Empirical studies by Grammig and Wellner (2002) and Hautsch (2008) also address the interdependence of the individual process.

In particular, Hautsch (2008) finds the existence of a common unobserved component that jointly drives the dynamics of the trade and price processes.

It is therefore necessary to extend the recursive model into a vector form, by allowing the three processes to be interdependent and relaxing weak exogeneity. This is the motivation of Chapter 3 and Chapter 4. Chapter 3 discusses the use of lognormal distribution for financial durations. We compare the performance of lognormal ACD with the alternative specifications. The empirical results show that Lognormal ACD model is superior to Exponential ACD model and Weibull ACD model. It performs similarly to Burr ACD model or generalized gamma ACD model. The study in chapter 3 opens a door to use lognormal distribution for nonnegative valued financial point process and provides further support for the methodology of chapter 4. In Chapter 4, we release the weak exogeneity assumption and propose general form of vector MEM. We further propose a multivariate lognormal for the distribution of the distribution of vector MEM, which allows the innovation terms to be interdependent. In this way, the two restrictions imposed by previous work are releases and the maximum likelihood is proved to be a suitable estimation strategy.

The vector MEM is then applied to the trade and quotes data from the New York Stock Exchange (NYSE) for the dynamics of trading duration, volume and price volatility. The empirical results show that the vector MEM captures the dynamics of the trivariate system successfully. We find that times of greater activity or trades with larger size coincide with a higher number of informed traders present in the market. But it is unexpected component of trading duration or trading volume that carry the information content. Moreover, the empirical results suggest a significant feedback effect from price process to trading intensity, in which the persistent quote changes and transient quote changes affect trading intensity in different direction.

Chapter 2 Weak Exogeneity in the Financial Point Processes

2.1 Introduction

The relationship between financial duration and market marks³ is critical for financial market microstructure studies. When modelling financial duration and other marks jointly, the multivariate extension of univariate ACD/MEM is required. However, the direct use of multivariate MEM model is restricted since joint probability distributions for nonnegative valued random variables are often not available in the literature.

To simplify the optimization procedure, a commonly used strategy in the literature is to decompose the joint distribution of duration and market marks into the product of the marginal density of duration and the conditional density of marks given duration. In estimation, if the weak exogeneity of duration is valid, then the marginal density of duration and the conditional density of marks can be estimated separately equation-by-equation. This approach simplifies the estimation procedure and is generally adopted in the empirical market microstructure. However, if the parameters in the conditional density depend on some of the parameters of the marginal density (for example, the weak exogeneity condition fails, the estimators would be inefficient or even inconsistent, leading to invalid inference (White (1981,1982))).

³ Engle (2000) use “marks” to denote trading duration, volatility and other variables associated with trading time. We adopt same idea here.

This chapter examines weak exogeneity problems in financial point processes. We consider three cases in which the weak exogeneity condition will break down and we use a Monte Carlo simulation to study the consequences of the failure of weak exogeneity. The simulation study suggested that a failure of the exogeneity assumption implied biased estimators. The bias is very large in third cases non-weak exogeneity. In empirical analysis, we derive an LM test, which is similar to Dolado, Rodriguez-Poo et al. (2004). We use a more fruitful specification of the conditional mean, which implies that the rejection of null is less likely due to the misspecification of conditional mean. Using two groups of high frequency data, we test both the weak exogeneity of duration and the joint weak exogeneity of duration and volume. The empirical results indicate the assumption of weak exogeneity is often rejected.

The remainder of this chapter is structured as follows. Section 2 reviews the related studies on weak exogeneity. Section 3 introduces the notion of weak exogeneity and methodology. Section 4 presents a simulation study to examine the consequences of incorrectly assuming weak exogeneity. Section 5 derives an LM test for weak exogeneity. Section 6 contains an empirical application.

2.2 Related Studies on Weak Exogeneity

Different definitions of exogeneity are clarified by Engle, Hendry et al. (1983), for example, weak exogeneity, strong exogeneity, super exogeneity and invariance. Weak exogeneity is proposed as an answer to the question of under what conditions can one estimate the parameters of conditional density without loss of information from neglecting the marginal process. The idea of weak exogeneity is expressed simply by saying that estimation and inference on the parameters of the marginal density and the conditional density can be undertaken separately, without loss of efficiency, if the endogenous variable in the marginal density is weakly exogenous for parameters in

the conditional density. Engle and Hendry (1993) develop the different classes of tests of weak exogeneity. In particular, if the marginal processes are constant, Wu-Hausman test is commonly used for test weak exogeneity.

The original Hausman test (Hausman 1978) contrasts two estimates obtained from different estimators (unconstrained and constrained parametric models). Under a null hypothesis, both of these estimators are consistent while only the second estimator is efficient. Under the alternative hypothesis of endogeneity, the first estimator is consistent while the second is not. This Hausman statistic has, under the null hypothesis, an asymptotically chi-squared distribution with the number of degrees of freedom equal to the number of endogenous regressors. An alternative to the Hausman contrast test is the two-stage Wald version test, originally derived by Wu (1973). In the first stage, by careful construction⁴, a reduced form model (marginal model) is specified for the endogenous variables which are estimated consistently. Then, the fitted values of the endogenous variables are computed and in the second stage, the conditional model is augmented by plugging in the fitted values as additional variables. If the fitted values are jointly significant in the conditional model, the null hypothesis of weak exogeneity is rejected. A simple Wald statistic can be used to test the joint significance. Effectively, this two-stage Wald version test leads to a test which is asymptotically equivalent to the Hausman contrast test [an algebraic derivation of this result can be found in the Davidson and MacKinnon (2004, Section 8.7)]. Using Monte Carlo simulation, Adkins, Campbell et al. (2012) shows, under a series of different conditions, that the Wald version of the Hausman test often has better properties than the contrast version.

⁴ See Terza, Basu and Rathouz (2008) for the conditions of choosing IV.

The Hausman test has been widely used in various areas, such as macroeconomics, health economics, and international trade. For example, Fischer (1993) and Boswijk and Urbain (1997) test the weak exogeneity of Swiss money Demand. Terza, Basu et al. (2008) address the endogeneity in a econometric model of health. Staub (2009) tests for the exogeneity of a binary explanatory variable in a count data regression model. Darrat, Hsu et al. (2000) test export exogeneity in Taiwan. However, Hausman tests suffer from three problems when applied to financial point process. Firstly, economic theory does not yield insights which guide the choice of instrumental variables. Secondly, the test is developed in a Gaussian/linear framework, whereas the market point process usually belongs to the exponential family. Thirdly, correct specification of the conditional mean is a fundamental assumption underlying the test, since the rejection of the null hypothesis could be due either to the absence of weak exogeneity or to the misspecification of the conditional mean.

We extend the Hausman test of weak exogeneity in a time series model and propose three cases in which weak exogeneity conditions will break down. A Monte Carlo simulation study is used to examine the consequences of the failing of weak exogeneity. In empirical analysis, we derive an LM test, which is similar to Dolado, Rodriguez-Poo et al. (2004). However, we use a more powerful specification of the conditional mean and test both weak exogeneity of duration and jointly weak exogeneity of duration and volume.

2.3 Methodology

2.3.1 Formal Definition of Weak Exogeneity

As in Engle, Hendry et al. (1983) and Engle and Hendry (1993), we start with a bivariate stochastic process $\{x_t, y_t\}$ and the joint density $f(x_t, y_t | \Omega_t; \theta)$, where Ω_t

denotes the information available up to period t , which includes lags and other important variables. Commonly, the joint density (x_t, y_t) can be factorized into the product of the marginal density x_t and conditional density of y_t given x_t

$$f(x_t, y_t | \Omega_t; \theta) = f_x(x_t | \psi_t; \theta^x) f_{y|x}(y_t | x_t, \Omega_t; \theta^y) \quad (2.1)$$

where $(\theta^x, \theta^y) \in \Theta$. Let $\mathcal{G} = f(\theta)$ be the parameters of interest, which are assumed to be present only in the conditional density. The key issue, addressed by Engle, Hendry et al. (1983), is to know under what conditions it is possible to estimate \mathcal{G} just as function of θ^y and without loss of information. In other words, that all the information needed for estimation of \mathcal{G} is $f_{y|x}$.

Engle, Hendry et al. (1983) define a variable of x_t as weakly exogenous for a set of parameters of interest \mathcal{G} if:

- i) $\mathcal{G} = f(\theta)$, \mathcal{G} is a function of parameters θ^y alone, and
- ii) θ^y and θ^x are variation free, i.e. $(\theta^x, \theta^y) \in \Theta^x \times \Theta^y$.

Consequently, if x_t is weakly exogenous for \mathcal{G} , there is no loss of information about \mathcal{G} from neglecting the process determining x_t . Otherwise, the estimation of θ^y would be inefficient or even inconsistent.

In econometric, the marginal density might also be interested. Weak exogeneity is also expressed simply by saying that estimation and inference on θ^x and θ^y can be undertaken separately without loss of information, if x_t is weak exogenous for θ^y . Engle, Hendry et al. (1983) further introduce the notation of “sequential cut” and “cross-restriction” to illustrate weak exogeneity, saying that x_t is weak exogenous for

θ^y , if $[f_x(x_t|\Omega_t;\theta^x), f_{y|x}(y_t|x_t, \Omega_t;\theta^y)]$ operates a sequential cut on $f(x_t, y_t|\Omega_t;\theta)$, or if θ^x and θ^y is not subject to “cross-restriction”.

2.3.2 Different Types of Weak Exogeneity in Financial Point Processes

Manganelli (2005) proposes a framework for the joint dynamics of trading duration, volume and price volatility. This model incorporates both causality and feedback effect among variables of interested and thereby can explain the various strategic models in the market microstructure literature. So we take Manganelli (2005)’s model for specification the dynamics of financial point processes. To simplify, we only consider the jointly distribution of duration and volume. Define $\{d_t, v_t\}, t=1, \dots, T$ as the two-dimensional time series associated with intraday trading duration and trading volume. In particular, duration is defined as the time elapsing between consecutive trades, volume is the trade size associated with each transaction. The bivariate trading process- duration, volume - can be modelled as follows:

$$\{d_t, v_t\} \sim f(d_t, v_t | \Omega_t; \theta) \quad (2.2)$$

where θ is a vector incorporating the parameters of interest.

In Manganelli (2005)’s framework, the joint distribution is decomposed into the product of marginal density of durations and the conditional density of volumes given durations:

$$\{d_t, v_t\} \sim g(d_t | \Omega_t; \theta_d) \cdot h(v_t | d_t, \Omega_t; \theta_v) . \quad (2.3)$$

Manganelli (2005) specifies the following univariate ACD/MEM model for duration and volume:

$$\begin{aligned} d_t &= \psi_t(\theta_d; \Omega_t) u_t, \quad u_t \sim i.i.d.(1, \sigma_u^2) \\ v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d.(1, \sigma_\eta^2) \end{aligned} \quad (2.4)$$

where (ψ_t, ϕ_t) are the conditional expectations of duration and volume, $\theta = (\theta_d, \theta_v)$ is a vector of s parameters of interest. The innovation terms are uncorrelated with each other by construction.

The log likelihood can be expressed as:

$$L(\theta^v, \theta^d) = \sum_{t=1}^T [\log g(d_t | \Omega_t; \theta_d) + \log h(v_t | d_t, \Omega_t; \theta_v)]. \quad (2.5)$$

Follows Manganelli (2005), the conditional mean of duration and the conditional mean of volume conditional on duration are expressed as:

$$\begin{aligned} g(d_t | \Omega_t; \theta^d) &\sim \begin{aligned} d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t \\ \psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1}, \end{aligned} \\ h(v_t | d_t, \Omega_t; \theta^v) &\sim \begin{aligned} v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t \\ \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t. \end{aligned} \end{aligned} \quad (2.6)$$

It is well know that estimation and inference on the parameters characterising each density can be undertaken separately, without loss of efficiency, if two of following condition hold: a) weak exogeneity, and b) the respective densities are correctly specified. Consequently, failing of weak exogeneity would result in inefficient or even inconsistent estimators, leading to unreliable inferences.

In the econometrics literature, the Hausman specification is usually used to test weak exogeneity. As explained by Engle, Hendry et al. (1983) , if none of the parameters in the marginal model appear in the conditional model, then weak exogeneity is valid. Therefore, testing weak exogeneity implies testing the significance of the predictor from the marginal model, in the conditional model. However, Hausman set is initially developed for a test of cross sectional model, whereas the ACD/MEM model is a time series model and the dynamics of

endogenous variables should also be considered. In this section, we extend the Hausman test of weak exogeneity in a time series model and propose three cases in which the weak exogeneity condition will break down. We use the so called “non-weak exogeneity” thereafter to express the notation that weak exogeneity condition breaks down.

Case 1: The first case of non-weak exogeneity is based on the Hausman specification. As explained by Engle, Hendry et al. (1983), Hausman test for weak exogeneity implies testing the significance of the predicted variable from the marginal model, in the conditional model. In the ACD/MEM models, it is natural to use the conditional expected value instead of predicted value, since the conditional expectation of duration is directly measured. This is also mentioned in Engle (2000). So to specify Hausman test for weak exogeneity in financial point process, we re-write equation (2.6) as:

$$\begin{aligned}
d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2), \\
\psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + a_3 \psi_{t-1}, \\
v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2), \\
\phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t + b_5 \psi_t.
\end{aligned} \tag{2.7}$$

Under assumption that the parameters of interest depend solely on the parameters of the conditional distribution, i.e. $\theta_v = f(b_0, b_1, b_2, b_3, b_4)$ and the expected duration ψ_t is estimated from the duration process (marginal density), then it suffices to test the significance of ψ_t in the volume process (conditional density) in order to test weak exogeneity of duration. If it is insignificant, the parameters of interest are not subject to cross equation restrictions and θ_v are variation free with respect to the parameters

of the duration process. Thus, the condition for weak exogeneity is that $b_5=0$ in equation (2.7).

If we look at in another way and assume

$$\begin{aligned} d_t - \psi_t &= \varepsilon_{1t} \sim i.i.d(0, \tilde{\sigma}_\varepsilon^2) \\ v_t - \phi_t &= \varepsilon_{2t} \sim i.i.d(0, \tilde{\sigma}_\eta^2). \end{aligned} \quad (2.8)$$

equation (2.7) then becomes⁵

$$\begin{aligned} d_t &= \alpha_0 + \alpha_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} + \varepsilon_{1t} \\ v_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + (b_4 + b_5) d_t + \varepsilon'_{2t} \end{aligned} \quad (2.9)$$

where $\varepsilon'_{2t} = -b_5 \varepsilon_{1t} + \varepsilon_{2t}$. Therefore, $Cov(\varepsilon_{1t}, \varepsilon'_{2t}) = -b_5 \text{var}(\varepsilon_{1t}) = -b_5 \tilde{\sigma}_\varepsilon^2$. The condition for weak exogeneity is $b_5=0$ or $Cov(\varepsilon_{1t}, \varepsilon'_{2t}) = 0$. In such a case, the parameters of interest θ_v are not subject to cross equation restrictions and are variation free with respect to the parameters from duration equation θ_d .

The Hausman specification test might also take another form, see for example Dolado, Rodriguez-Poo et al. (2004). They specify the following functional form for testing weak exogeneity:

$$\begin{aligned} v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t \\ \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + (b_4 + b_5 \psi_t) d_t. \end{aligned} \quad (2.10)$$

Generally speaking, if any linear or nonlinear forms of expected duration significantly enter the volume process (conditional density), the weak exogeneity of duration will break down⁶.

⁵ See Appendix 1 for proof.

⁶ Since the nonlinear term can be linearized as $f(\psi_t) \approx f(\bar{\psi}) + (\psi_t - \bar{\psi}) f'(\bar{\psi}_t)$. The proof thereafter is the same as is done in the Hausman test.

Case 2: The second case of non-weak exogeneity is motivated by Manganelli (2005)'s model. Manganelli (2005) initially considers the following specification for the duration and volume process.

$$\begin{aligned} d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2) \\ v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2) \\ \psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} + a_4 \phi_{t-1} \\ \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \psi_{t-1} + b_4 \phi_{t-1} + b_5 d_t. \end{aligned} \quad (2.11)$$

Writing it in matrix form:

$$\begin{pmatrix} \psi_t \\ \phi_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} a_3 & a_4 \\ b_3 & b_4 \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \phi_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix}. \quad (2.12)$$

Both the parameters in marginal density and conditional density are interested. In order to optimize the two processes separately, the assumption of weak exogeneity has to be imposed. In this case, the condition for weak exogeneity is $\alpha_4 = 0, b_3 = 0$, since only under this condition can $[g(d_t | \Omega_t; \theta^d), h(v_t | d_t, \Omega_t; \theta^v)]$ operate a sequential cut on $f(d_t, v_t | \Omega_t; \theta)$ whereupon there is no cross-section restrictions between marginal and conditional density (Engle, Hendry et al. 1983).

If we look at it in another way and assume:

$$\begin{aligned} d_t - \psi_t &= \varepsilon_{1t} \sim i.i.d(0, \tilde{\sigma}_\varepsilon^2) \\ v_t - \phi_t &= \varepsilon_{2t} \sim i.i.d(0, \tilde{\sigma}_\eta^2) \end{aligned} \quad (2.13)$$

then the above becomes⁷

$$\begin{pmatrix} d_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 + \alpha_3 & a_2 + \alpha_4 \\ b_1 + b_3 & b_2 + b_4 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} - \begin{pmatrix} a_3 & a_4 \\ b_3 & b_4 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix}. \quad (2.14)$$

⁷ See Appendix 2 for proof.

Again, the condition of weak exogeneity is that $\alpha_4 = 0$, $b_3 = 0$, since only under such condition can θ_d and θ_v be variation free and subject to no cross equation restrictions.

Generally speaking, if any lagged expected (or fitted) value from marginal model is present in the conditional model, or any lagged expected (or fitted) value from conditional model is present in marginal model, the weak exogeneity condition will break down.

Case 3: Motivated by the case 1 of non-weak exogeneity, we may consider a more restrictive case of non-weak exogeneity, where the innovations between marginal and conditional distributions are correlated. Let's look at the following model of duration and volume:

$$\begin{aligned} d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t, \\ v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \end{aligned} \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim i.i.d.(\mathbf{I}, \Omega) \quad (2.15)$$

$$\begin{aligned} \psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} \\ \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t \end{aligned}$$

where \mathbf{I} is the unit vector, $corr(\varepsilon_t, \eta_t) = \rho$, $\rho \neq 0$.

If the innovations from the marginal and conditional distributions are correlated, the weak exogeneity condition will break down, since the parameters of volume equation (θ_v) are subject to cross equation restrictions and are not variation free with respect to parameters from the duration equation (θ_d). In this case, the condition of weak exogeneity is that the innovation terms in marginal and conditional distribution are independent. Effectively, Hausman specification can also be viewed a special case where the innovations between the marginal and conditional distributions are

correlated. It is sufficient to test case 1 weak exogeneity in order to test case 3 weak exogeneity in empirical analysis.

Summary of conditions for weak exogeneity

- a) The expected (or fitted) value from marginal model does not enter in the specification of conditional density.
- b) The lagged expected (or fitted) value from marginal model is not present in the specification of conditional density, and the lagged expected (or fitted) value from conditional model is not present in the specification of marginal density.
- c) The innovation terms from the marginal and conditional distributions are uncorrelated.

The violation of either one of the above conditions will result in the breaking down of weak exogeneity.

2.4 Consequences of Incorrectly Assuming Weak Exogeneity- a Simulation Study

Based on the three cases of non-weak exogeneity above, we study the consequences of ignoring weak exogeneity in this section. We examine the consequences if one estimates the model under the assumption of weak exogeneity when there is none. To do so, we conduct a simulation study.

The experiments are designed as follows. The joint distribution of duration and volume is chosen as the benchmark model. The data is generated based on the fact that weak exogeneity condition breaks down. In particular, we generate the duration and volume data in accordance with each of the three cases of non-weak exogeneity discussed in section 2.3.2.

Case 1 The expected value from the marginal model is present in the conditional distribution. The data are generated according to equation (2.7).

Case 2 The lagged expected value from the marginal model is present in the conditional distribution and the lagged expected value from the conditional model is present in the marginal distribution. The data are generated according to equation (2.11)

Case 3 The innovations from the marginal and conditional distributions are correlated. The data are generated according to equation (2.15)

We chose the sample sizes at $N = 2000, 5000$, and 10000 respectively and the number of simulations equals 2000 . In the first two cases of non-weak exogeneity, we use an exponential distribution with mean value 1 to generate the random disturbances ε_t and η_t . In the third case, we use a bivariate exponential distribution with correlations $\rho = 0.1$ and $\rho = 0.5$ to generate the random disturbances ε_t and η_t jointly.

We then estimate each model by two approaches. In the first approach, we assume the weak exogeneity condition is valid. The marginal process of duration and conditional process of volume given duration are estimated separately. We denote this estimation method the conditional MLE. In the second approach, we estimate the model under the fact of non-weak exogeneity. The duration and volume processes are estimated jointly. We call this latter approach the full MLE. After estimation, we compare the estimation results of conditional MEL with those from the full MLE. In particular, we focus on a comparison of the bias/inconsistency and efficiency of the estimators. Table 2-1, Table 2-2 and Table 2-3 report the simulation results for the three cases.

Table 2-1: Case 1 simulation summary statistics. Estimated parameters

N=2000					N=5000				N=10000			
Conditional MLE		Full MLE			Conditional MLE		Full MLE		Conditional MLE		Full MLE	
Mean	SD	Mean	SD		Mean	SD	Mean	SD	Mean	SD	Mean	SD
a ₀	0.111	0.041	0.107	0.029	0.104	0.022	0.103	0.018	0.103	0.015	0.102	0.012
a ₁	0.049	0.014	0.049	0.013	0.049	0.009	0.049	0.008	0.050	0.006	0.049	0.005
a ₂	0.052	0.018	0.051	0.016	0.051	0.011	0.051	0.010	0.051	0.008	0.051	0.007
a ₃	0.842	0.038	0.846	0.026	0.847	0.022	0.848	0.017	0.848	0.015	0.849	0.012
b ₀	0.034	0.018	0.108	0.040	0.033	0.011	0.103	0.022	0.033	0.008	0.102	0.014
b ₁	0.062	0.022	0.051	0.024	0.061	0.014	0.051	0.014	0.061	0.010	0.050	0.010
b ₂	0.042	0.017	0.048	0.016	0.043	0.011	0.050	0.010	0.043	0.007	0.050	0.007
b ₃	0.715	0.033	0.802	0.049	0.716	0.020	0.799	0.028	0.716	0.014	0.800	0.019
b ₄	0.100	0.019	0.100	0.018	0.099	0.012	0.100	0.012	0.099	0.008	0.100	0.008
b ₅			-0.107	0.048			-0.101	0.027			-0.101	0.017

Data generation process (DGP):

$$\begin{aligned}
 d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2) & \psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} \\
 v_t &= \varphi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2) & \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t + b_5 \psi_t
 \end{aligned}$$

The random disturbances ε_t and η_t are generated from an exponential distribution with mean value 1.

The population parameter values:

$$a_0 = b_0 = 0.1, \quad a_1 = b_1 = 0.05, \quad a_2 = b_2 = 0.05$$

$$a_3 = 0.85, \quad b_3 = 0.80, \quad b_4 = 0.1, \quad b_5 = -0.1$$

The parameters values are chosen partly from the empirical work of Manganelli (2005).

Table 2-2: Case 2 simulation summary statistics. Estimated parameters

N=2000					N=5000				N=10000			
Conditional MLE		Full MLE			Conditional MLE		Full MLE		Conditional MLE		Full MLE	
Mean	SD	Mean	SD		Mean	SD	Mean	SD	Mean	SD	Mean	SD
a ₀	0.132	0.048	0.106	0.032	0.125	0.026	0.102	0.020	0.124	0.018	0.101	0.013
a ₁	0.038	0.018	0.048	0.017	0.038	0.011	0.049	0.011	0.039	0.007	0.050	0.007
a ₂	0.054	0.013	0.051	0.012	0.053	0.008	0.051	0.007	0.053	0.006	0.051	0.005
a ₃	0.759	0.067	0.846	0.062	0.767	0.036	0.849	0.038	0.767	0.026	0.849	0.026
a ₄			-0.050	0.035			-0.051	0.020			-0.050	0.014
b ₀	0.088	0.035	0.103	0.049	0.084	0.020	0.100	0.030	0.083	0.014	0.100	0.020
b ₁	0.057	0.039	0.052	0.036	0.056	0.024	0.051	0.022	0.055	0.017	0.050	0.015
b ₂	0.045	0.016	0.048	0.016	0.046	0.010	0.049	0.010	0.047	0.007	0.049	0.007
b ₃			-0.046	0.098			-0.047	0.061			-0.048	0.041
b ₄	0.771	0.046	0.795	0.053	0.774	0.027	0.798	0.032	0.775	0.019	0.799	0.021
b ₅	0.100	0.033	0.101	0.032	0.100	0.021	0.100	0.020	0.100	0.015	0.100	0.014

DGP:

$$\begin{aligned}
 d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2) & \psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + a_3 \psi_{t-1} + a_4 \phi_{t-1} \\
 v_t &= \varphi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2) & \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \psi_{t-1} + b_4 \phi_{t-1} + b_5 d_t
 \end{aligned}$$

The random disturbances ε_t and η_t are generated from an exponential distribution with mean value 1.

The population parameter values;

$$a_0 = b_0 = 0.1, \quad a_1 = b_1 = 0.05, \quad a_2 = b_2 = 0.05$$

$$a_3 = 0.85, \quad a_4 = -0.05, \quad b_3 = -0.05, \quad b_4 = 0.80, \quad b_5 = 0.1$$

The parameters values are chosen partly from the empirical work of Manganelli (2005).

Table 2-3: Case 3 simulation summary statistics. Estimated parameters
(Only conditional MLE is reported)

$\rho = 0.1$							$\rho = 0.5$					
N=2000		N=5000		N=10000			N=2000		N=5000		N=10000	
Mean	SD	Mean	SD	Mean	SD		Mean	SD	Mean	SD	Mean	SD
a0	0.108	0.033	0.104	0.021	0.102	0.014	0.110	0.030	0.103	0.018	0.102	0.012
a1	0.048	0.014	0.049	0.009	0.050	0.006	0.048	0.016	0.049	0.010	0.050	0.007
a2	0.051	0.012	0.050	0.007	0.050	0.005	0.051	0.012	0.051	0.008	0.050	0.005
a3	0.847	0.025	0.848	0.015	0.849	0.011	0.847	0.022	0.849	0.014	0.849	0.009
b0	0.090	0.041	0.086	0.025	0.084	0.018	0.050	0.067	0.041	0.044	0.036	0.018
b1	-0.098	0.042	-0.101	0.026	-0.102	0.019	-0.515	0.096	-0.528	0.059	-0.534	0.036
b2	0.040	0.015	0.041	0.009	0.041	0.007	0.021	0.018	0.019	0.012	0.019	0.008
b3	0.790	0.035	0.792	0.021	0.793	0.015	0.749	0.088	0.765	0.053	0.772	0.028
b4	0.279	0.037	0.279	0.023	0.280	0.017	0.796	0.049	0.794	0.031	0.794	0.022

DGP:

$$d_t = \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim i.i.d.(\mathbf{I}, \Omega), \quad \text{corr}(\varepsilon_t, \eta_t) = \rho, \rho \neq 0, \quad \begin{aligned} \psi_t &= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + a_3 \psi_{t-1} \\ \phi_t &= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t \end{aligned}$$

The disturbances ε_t and η_t are generated jointly from a bivariate exponential distribution with mean value \mathbf{I} and correlations $\rho = 0.1$ and $\rho = 0.5$.

The population parameter values:

$$a_0 = b_0 = 0.1, \quad a_1 = b_1 = 0.05, \quad a_2 = b_2 = 0.05$$

$$a_3 = 0.85, \quad b_3 = 0.80, \quad b_4 = 0.1$$

The parameters values are chosen partly from the empirical work of Manganelli (2005).

From Table 2-1 (the first case of non-weak exogeneity), the means of the full MLE are all close to the population means. As the number of observations increases, the standard deviation of the full MLE gets smaller and the performance generally improves. The full MLEs work well as a whole. On the other hand, the performance of the conditional MLE is somewhat different to that of the full MLE. For the marginal distribution, the means of the conditional MLEs are unbiased and consistent in general. And the standard deviations of conditional MLEs are slightly larger than that of full MLEs, suggesting an efficient gain when duration and volume are estimated jointly. For the conditional distribution, both b_2 and b_3 are downward biased. In particular, the sum of b_2 and b_3 is downward biased towards smaller estimated persistence for volume. The conditional MLE of b_1 is greater than its population means, suggesting the impact of duration on volume is over estimated. As the number of observations increases, the performance of the conditional MLEs generally improves. However, the same characteristics of conditional MLE continue to hold. It seems that when the first weak exogeneity condition breaks down, the conditional MLEs for marginal distribution work fine, where the conditional MLEs for conditional distribution are biased. The poor performance of the conditional MLE is due to the fact that the information from the marginal distribution contains some of the information of the conditional distribution.

From Table 2-2 (the second case of non-weak exogeneity), we get similar results for the full MLE approach. The means of the full MLE are all close to the population values and the full MLE works well as a whole. The performance of the conditional MLE is different to that of the full MLE. In the duration process, both a_1 and a_3 are smaller than the population values. And the sum of a_1 and a_3 is downward biased

towards smaller estimated persistence for duration. The conditional MLE of a_2 is larger than those from the full MLEs, suggesting that the impact of volume on duration is over estimated. The same result is also hold for the volume process, where the sum of b_2 and b_4 is downward biased towards smaller estimated persistence and b_1 is upward biased suggesting a larger duration impact on volume. Besides, the conditional MELs are less efficient than those from the full MLEs in general, but the efficient loss is not significant in most of cases. It can be seen that the same characteristics of the conditional MLEs continue to hold when $N= 2000, 5000$ and 10000 .

Table 2-3 (the third case of non-weak exogeneity) only reports the results from conditional MLE. The full MLE and joint estimation method will be discussed in chapter 4. The conditional MLEs of the marginal distribution are unbiased and consistent in this case, even if the correlation between the marginal and conditional distributions is high. In the conditional distribution, the conditional MLE of b_3 is unbiased and consistent when the correlation of errors between the marginal distribution and the conditional distribution is relatively small ($\rho = 0.1$), and it gets slightly biased and inconsistent when the correlation is relatively high ($\rho = 0.5$). The greater differences are observed for the conditional MLEs of b_1 and b_4 , which evaluate the impact of duration on volume. It can be seen that the conditional MLE of b_1 is negative in this case, and the negative size increases drastically as the correlation of the errors increases. The conditional MLE of b_4 is much larger than its population mean. As the correlation of the errors increases, the conditional MLE of b_4 gets larger. Thus, the incorrectly assuming case 3 weak exogeneity has severe consequences on the estimation results, which makes the inferences on the parameters unreliable or

even misleading. Again, the same characteristics of the conditional MLEs continue to hold when sample size increases.

To summarize, the simulation study suggests that a failure of the exogeneity assumption implies biased estimators. The biases are very large in the third case non-weak exogeneity. In particular, the failure of weak exogeneity assumptions have severe effects on the conditional distribution, where the persistence of volume will be downward biased and the impact of duration on volume will be over estimated. Besides, the failure of the weak exogeneity also implies inefficient estimators, but the efficiency loss is relatively small. The simulation results are partially consistent with White (1981, 1982). The results indicate that the econometric inferences on the parameters are unreliable or even misleading if weak exogeneity condition fails.

It is therefore necessary to conduct a test for weak exogeneity before estimation in empirical analysis.

2.5 An LM Test for Weak Exogeneity in Financial Point Processes

In this section, we will derive a Lagrange-multiplier (LM) or efficient score test for weak exogeneity. It proves to be particularly useful since it only requires estimation of the restricted model.

In the literature, the Hausman test is often used to test weak exogeneity. However, correct specification of the conditional mean is a fundamental assumption for the validity of the test since the rejection of the null hypothesis could be due to either the rejection of weak exogeneity or the result of misspecification of the conditional mean. To take this into consideration, we introduce an Augmented ACD (AACD) model (Fernandes and Grammig 2006) for the specification of the conditional mean of duration. The AACD model of Fernandes and Grammig (2006) is given by

$$d_t = \psi_t(\theta_d; \Omega_t) \varepsilon_t$$

where ε_t is i.i.d with mean value 1, and

$$\frac{\psi_t^\lambda - 1}{\lambda} = \omega^* + \alpha^* \psi_{t-1}^\lambda [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^v + \beta \frac{\psi_{t-1}^\lambda - 1}{\lambda}.$$

The AACD model is then obtained by rewriting it as

$$\psi_t^\lambda = \omega + \alpha \psi_{t-1}^\lambda [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^v + \beta \psi_{t-1}^\lambda \quad (2.16)$$

where $\omega = \lambda \omega^* - \beta + 1$ and $\alpha = \lambda \alpha^*$.

The AACD model provides a flexible functional form and permits the conditional duration process $\{\psi_t\}$ to respond in distinct manners to small and large shocks. The shock impact curve $g(\varepsilon_t) = [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^v$ incorporates such asymmetric responses through the shift and rotation parameters b and c , respectively. The shape parameter v plays a similar role to λ , which determines whether the Box-Cox transformation is concave ($\lambda \leq 1$) or convex ($\lambda \geq 1$)

Appendix 3 summarizes the typology of ACD models which can be nested by the AACD model. The AACD model provides a flexible functional form and encompasses most of the current ACD models. The rejection of the null is less likely to be due to misspecification of the conditional mean.

To simplify, only the case one weak exogeneity is discussed in the LM test and in empirical analysis. It is necessary but not sufficient for weak exogeneity. The case two weak exogeneity can be derived in the same way. As it is mentioned, the first case of non-weak exogeneity is a special case of the third case of non-weak exogeneity. The test procedure also serves as a test of the case three weak exogeneity.

2.5.1 Testing Weak Exogeneity of Duration

Let us specify the duration and volume as represented by the AACD and augmented autoregressive conditional volume (AACV) models respectively with the errors belonging to the exponential distribution family (exponential, Weibull or Burr distribution); for example

$$\begin{aligned}
 d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t \\
 \psi_t^\lambda &= \omega_1 + \alpha_1 \psi_{t-1}^\lambda [|\varepsilon_{t-1} - b_1| + c_1(\varepsilon_{t-1} - b_1)]^v + \beta_1 \psi_{t-1}^\lambda \\
 v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t \\
 \phi_t^\lambda &= \omega_2 + \alpha_2 \phi_{t-1}^\lambda [|\eta_{t-1} - b_2| + c_2(\eta_{t-1} - b_2)]^v + \beta_2 \phi_{t-1}^\lambda + a_0 d_t + a_1 \psi_t.
 \end{aligned} \tag{2.17}$$

As explained in section 2.3, it suffices to test $H_0 : a_1 = 0$ in order to test for weak exogeneity of duration. In such a case, the parameters of interest are not subject to cross equation restrictions and are variation free with parameters from marginal model.

Due to the inherent complexity of the AACD model, the LM or efficient score testing principle is proved to be particularly useful for this purpose, since it requires estimation under the null hypothesis only.

Under H_0

$$\begin{aligned}
 v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t \\
 \phi_t^\lambda &= \omega_2 + \alpha_2 \phi_{t-1}^\lambda [|\eta_{t-1} - b_2| + c_2(\eta_{t-1} - b_2)]^v + \beta_2 \phi_{t-1}^\lambda + a_0 d_t.
 \end{aligned} \tag{2.18}$$

Under H_1

$$\begin{aligned}
 v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t \\
 \phi_t^\lambda &= \omega_2 + \alpha_2 \phi_{t-1}^\lambda [|\eta_{t-1} - b_2| + c_2(\eta_{t-1} - b_2)]^v + \beta_2 \phi_{t-1}^\lambda + a_0 d_t + a_1 \psi_t.
 \end{aligned} \tag{2.19}$$

Assuming that the densities are correct, the general theory of ML leads to a simple score test for $a_1 = 0$. Given correctly specified duration and volume models, the quasi log-likelihood function is

$$L = -\sum_{t=1}^T (\log \phi_t + v_t / \phi_t + \log \psi_t + d_t / \psi_t). \quad (2.20)$$

The quasi log-likelihood MLE approach is most suitable since it allows for a wide range of different distributions capturing all possible supports of the point process.

Moreover, under H_0 of weak exogeneity, the AACD and AACV model can be estimated separately. Then, the score/LM test has the familiar form

$$S = -\hat{i}^{a_1}(\hat{\theta}_c)' \hat{I}^{a_1}(\hat{\theta}_c)^{-1} \hat{i}^{a_1}(\hat{\theta}_c) \quad (2.21)$$

where $\hat{i}^{a_1}(\hat{\theta}_c) = \frac{\partial L}{\partial \theta}$ and $\hat{I}^{a_1}(\hat{\theta}_c) = \frac{\partial^2 L}{\partial \theta \partial \theta'}$ are the components corresponding to a_1

in the empirical score and Hessian from constrained model. Under mild regularity conditions it is well known that, the score test has an asymptotically $\chi^2(1)$ distribution under H_0 .

2.5.2 Testing Joint Weak Exogeneity of Duration and Volume

The above testing approach enables a test of weak exogeneity of duration for one market mark (volume or volatility). In market microstructure, sometimes more than two variables are interested and modelled jointly. For example, Manganelli (2005) models duration, volume and volatility jointly. The joint distribution is decomposed into the product of the three components: the marginal distribution of duration, the conditional distribution of volume given duration, the conditional distribution of volatility given duration and volume. In such case, testing joint weak exogeneity of duration and volume is needed. In this section, we propose the joint weak exogeneity test principle. We start with the joint distribution of duration, volume and volatility and test the joint weak exogeneity of duration and volume for volatility equation.

However, this approach can be extended to test the joint weak exogeneity of more than two variables.

The joint distribution is decomposed into the product of the three components, specially,

$$(d_t, v_t, r_t) \sim f(d_t, v_t, r_t | \Omega_t; \theta) = g(d_t | \Omega_t; \theta^d) \cdot h(v_t | d_t, \Omega_t; \theta^v) \cdot k(r_t | d_t, v_t, \Omega_t; \theta^r). \quad (2.22)$$

The log likelihood can be expressed as:

$$L(\theta^d, \theta^v, \theta^r) = \sum_{t=1}^T [\log g(d_t | \Omega_t; \theta^d) + \log h(v_t | d_t, \Omega_t; \theta^v) + \log k(r_t | d_t, v_t, \Omega_t; \theta^r)]. \quad (2.23)$$

As illustrated in the previous section, we allow for a more flexible functional form for duration, volume and volatility process. This results in a LM score test. The conditional expectation of duration and volume are assumed to follow an AACD and AACV process and the conditional expectation of volatility are assumed to follow an Asymmetric Power GARCH (APGARCH) (Ding, Granger et al. 1993) process, which is similar to the AACD specification. Then the volatility model has the following form:

$$\begin{aligned} \hat{r}_t &= \sigma_t(\theta_v; d_t, v_t, \Omega_t) \xi_t, \quad \xi_t \sim i.i.d(0,1), \\ \sigma_t^\lambda &= \omega_3 + \alpha_3 \sigma_{t-1}^\lambda [|\xi_{t-1} - b_3| + c_3(\xi_{t-1} - b_3)]^v + \beta_3 \sigma_{t-1}^\lambda + a_{01} d_t + a_{11} \psi_t + a_{02} v_t + a_{12} \phi_t. \end{aligned} \quad (2.24)$$

For the same reason, it suffices to test that $\alpha_{11} = \alpha_{12} = 0$ in order to test jointly the weak exogeneity of duration and volume.

Under H_0

$$\begin{aligned} \hat{r}_t &= \sigma_t(\theta_v; d_t, v_t, \Omega_t) \xi_t, \quad \xi_t \sim i.i.d(0,1), \\ \sigma_t^\lambda &= \omega_3 + \alpha_3 \sigma_{t-1}^\lambda [|\xi_{t-1} - b_3| + c_3(\xi_{t-1} - b_3)]^v + \beta_3 \sigma_{t-1}^\lambda + a_{01} d_t + a_{02} v_t. \end{aligned} \quad (2.25)$$

The density for both the AACD and AACV models is the exponential while for the APGARCH model it is a standard normal distribution. Assuming that the densities are correct, the general theory of ML leads to a simple Score test for $\alpha_{11} = \alpha_{12} = 0$.

Given that the Augmented GARCH, AACV and AACD models are correctly specified, the quasi log-likelihood function is

$$L = -\sum_{t=1}^T (\log \phi_t + v_t / \phi_t + \log \psi_t + d_t / \psi_t + \log \sigma_t^2 + \mu_t^2 / \sigma_t^2). \quad (2.26)$$

Moreover, under H_0 of weak exogeneity, the APGARCH and the AACD and AACV models can be estimated separately. The score/LM test has the familiar form

$$S = -\hat{i}^a(\hat{\theta}_c)' \hat{I}^\alpha(\hat{\theta}_c)^{-1} \hat{i}^a(\hat{\theta}_c) \quad (2.27)$$

where $\hat{i}^a(\hat{\theta}_c) = \frac{\partial L}{\partial \theta}$ and $\hat{I}^\alpha(\hat{\theta}_c) = \frac{\partial^2 L}{\partial \theta \partial \theta'}$ are, respectively, the components of the empirical score and Hessian from unconstrained model corresponding to α_{11} and α_{12} . Under mild regularity conditions it is well known that, the score test has an asymptotic $\chi^2(2)$ distribution under H_0 .

2.5.3 Power of the Test

How powerful is the LM test for weak exogeneity in this study? How many observations do we need to have? To answer these questions, we need to conduct an investigation of the statistical power of the LM test. We begin with the ACD/MEM model below.

$$\begin{aligned} d_t &= \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2) \\ v_t &= \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2) \\ \psi_t &= a_0 + \alpha_1 d_{t-1} + a_2 \psi_{t-1} \\ \phi_t &= b_0 + b_1 v_{t-1} + b_2 \phi_{t-1} + b_3 d_t + b_4 \psi_t \end{aligned} \quad (2.28)$$

$$H_0: b_4 = 0$$

$$H_1: b_4 \neq 0$$

Choosing a 5% significance level, the simulation results indicate that the empirical test size is 6.7% for sample size $n=10000$, 7.2% for sample size $n=5000$ and 8.9% for sample size $n=2000$.

To explore the power of the test, we generate data under the alternative hypothesis and estimate the model under null hypothesis⁸. Under the alternative hypothesis the parameter b_4 varies between -0.2 to 0.2 with step 0.025. Given the sample size and empirical test size, the power of the test is the probability of rejecting a hypothesis when it is false. The results of the LM test for different sample sizes and effective test size are listed in Table 2-4.

Table 2-4: Power of the test

b_4	Power of test		
	N=2000	N=5000	N=10000
-0.200	0.825	0.989	1.000
-0.175	0.736	0.977	1.000
-0.150	0.614	0.933	0.999
-0.125	0.457	0.803	0.972
-0.100	0.354	0.610	0.878
-0.075	0.232	0.403	0.600
-0.050	0.149	0.210	0.347
-0.025	0.084	0.093	0.141
0.000	0.052	0.051	0.051
0.025	0.041	0.057	0.098
0.050	0.054	0.120	0.236
0.075	0.105	0.277	0.502
0.100	0.164	0.458	0.781
0.125	0.260	0.684	0.938
0.150	0.414	0.848	0.987
0.175	0.547	0.924	0.997
0.200	0.642	0.970	1.000

Note: Power of the test is the percentage rejections of the LM tests at empirical significant level for testing $b_4 = 0$ against $b_4 \neq 0$

⁸ To avoid present of negative value of volume, we use logarithmic version of ACD model for GDP process and estimation.

As the sample size increases, the power of LM test increases. It can also be seen when b_4 decreases to 0, the power tends to be 5%. The test power grows quickly to 1 as b_4 move away from zero. The results are also plotted in Figure 2-1. They are appropriately symmetric. The simulation shows that the LM test has good power to test for weak exogeneity in a financial market point process.

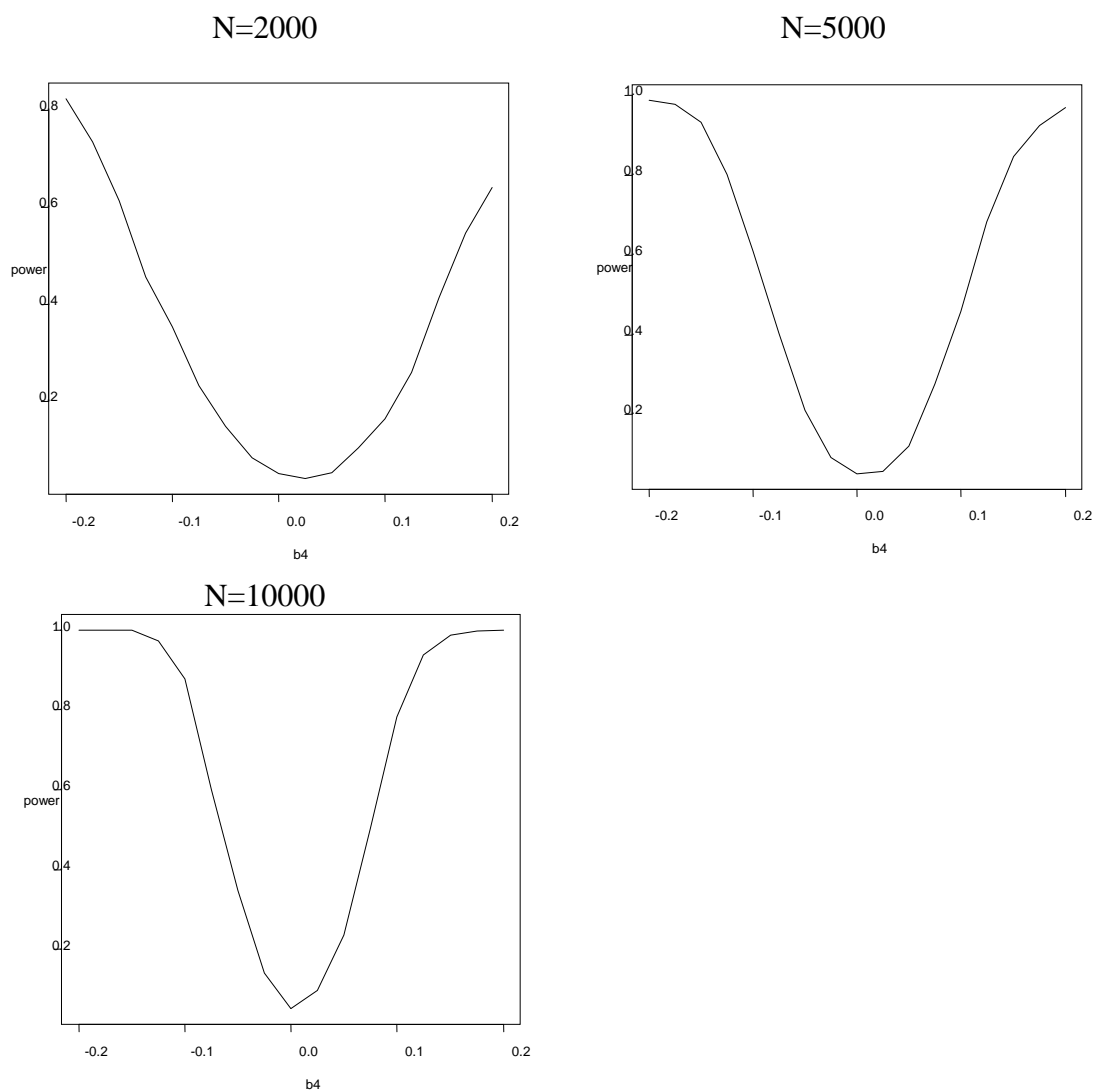


Figure 2-1: Power of the test

2.6 Empirical Analysis

In this section, we use the method discussed in section 5 to test weak exogeneity of duration for two groups of high frequency data. The empirical analysis starts with the joint distribution of the three variables: duration, volume and price volatility. These three variables are key factors in analysing market microstructure. Specifically, we will test weak exogeneity of duration for volume process and weak exogeneity of duration for volatility process respectively. We also test the joint weak exogeneity of duration and volume for volatility process.

2.6.1 Data

We use the data from the Trades and Quotes (TAQ) dataset at NYSE. The TAQ data consists of two parts: the first reports the trade data while the second lists the quote data (bid and ask data). The data were kindly provided by Manganelli (2005). He constructed 10 deciles of stocks covering the period from Jan 1,1998 to June 30, 1999, on the basis of the 1997 total number of trades of all stocks quoted on the NYSE. We randomly selected 5 stocks from the eighth decile (frequently traded stocks) and 5 from the second decile (infrequently traded stocks) covering the period from Jan 1,1998 to June 30, 1999. The tickers and names of the ten stocks are reported in Table 2-5.

Table 2-5: Stock used in this analysis

A. Frequently traded		B. Infrequently traded	
TRN	Trinity Industries	ABG	Group ABG
R	Ryder System Inc.	OFG	Oriental Finl Grp Hold Co.
ARG	Airgas Inc.	LSB	LSB Industries Inc.
GAS	Nicor Incorporated	FEP	Franklin Electronic Publisher
TCB	TCF Financial Corp.	HTD	Huntingdon Life S.G.

Before the analysis began, we adopted Manganelli (2005)'s strategy to prepare the data. First, all trades before 9:30 am or after 4:00 pm were discarded. Second, durations over night were computed as if the overnight periods did not exist. For example, the time elapsing between 15:59:50 and 9:30:05 of the following day is only 15 seconds. We keep overnight duration because our samples for infrequently traded stocks are very small. Eliminating this duration would cause the loss of important data for these stocks. Third, all transaction data with zero duration are eliminated. These transactions are treated as one single transaction, and the related volumes are summed. Fourth, to deal with the impact of dividend payments and trading halts, we simply deleted the first observation whose price incorporated the dividend payment or a trading halt. Fifth, to adjust the data for stock splits, we simply multiplied the price and volume by the stock split ratio. Sixth, the price of each transaction is calculated as the average of the prevailing bid and ask quote. To obtain the prevailing quotes, we use the 5 second rule used by Lee and Ready (1991) which links each trade to the quote posted at least 5 seconds before, since the quotes can be posted more quickly than trades are recorded. This procedure is standard in microstructure studies. Seventh, the returns were computed as the difference of the log of the prices. To obtain a return sequence that is free of the bid-ask bounce that affects prices (see Campbell et al., 1997, chapter 3), we follow Ghysels, Gouriéroux et al. (2004) in using the residuals of an ARMA(1,1) model estimated on the return data.

The second issue to be addressed prior to the analysis concerns the intraday pattern in the data. It is well known that duration, volume and volatility exhibit strong intraday periodic components, with a high trading activity at the beginning and end of the day. To adjust for this, we make use of a method used by Engle (2000). We regress the durations, volumes and returns squares on a piecewise cubic spline with

knots at 9:30, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 15:30 and 16:00. The original series are then divided by the spline forecast to obtain the adjusted series. Figure 2-2 depicts the nonparametric estimate of daily pattern of duration and return square for one typical stock ARG. Generally, less frequently traded stocks do not exhibit any regular intraday pattern. More frequently traded stocks typically show the inverted U pattern for duration, the L pattern for return squares, and no regular pattern for volume.

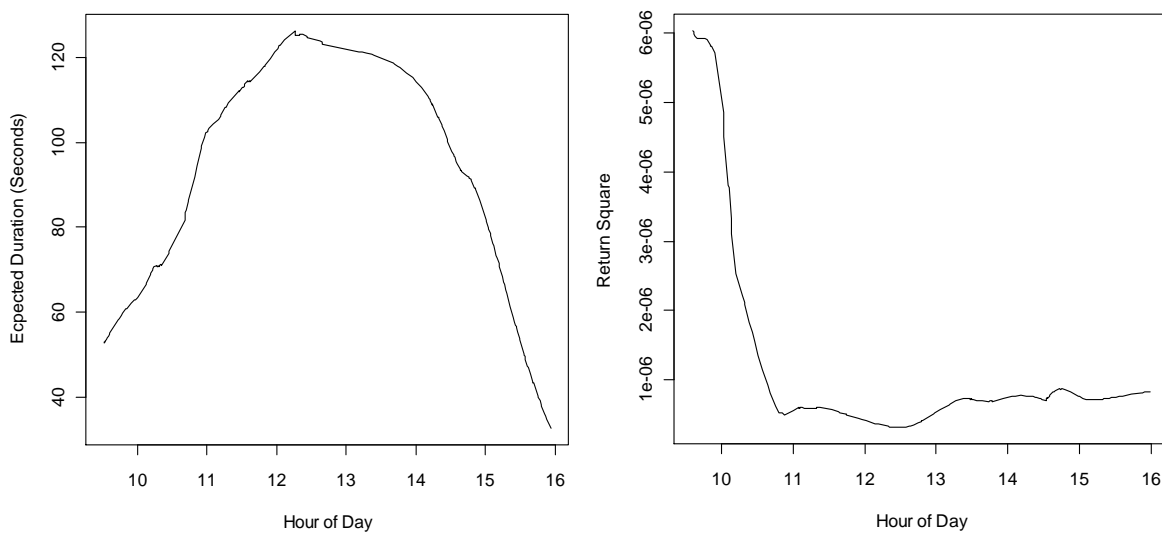


Figure 2-2: Nonparametric estimate of daily pattern of transaction duration and return square.

Some summary statistics for the cleaned data are reported in Table 2-6. For the frequently traded stocks, the number of observations range from 33,850 to 63,862 in the sample period, and the average trading duration ranges from 137 seconds to 259 seconds. For the infrequently traded stocks, the number of observation ranges from 2,074 to 7,212 in the sample period, with the average trading duration ranging from

1,215 seconds to 4,215 seconds. The trading volume does not show any difference between frequently traded stocks and infrequently traded stocks. The number of trading volumes ranges from 833 to 5,295. Ljung–Box statistics suggests that duration, volume and volatility show strong serial correlations. And this is particularly true for high frequently traded stocks, which motivates the ACD models.

Table 2-6: Summary statistics for the 10 stocks

	Obs	Mean		LB(20)		
		Duration	Volume	Duration	Volume	Variance
TRN	55582	157.86	1369.43	3780.09	1383.35	3769.80
GAS	41999	212.93	827.77	5951.85	2338.08	4073.09
TCB	55208	158.94	1855.20	4171.36	2644.11	2925.82
R	63862	137.41	1800.74	14072.3	7276.91	23685.7
ARG	33850	259.2	1280.70	3780.09	1383.35	3769.80
ABG	2074	4214.88	5259.05	120.28	225.07	146.00
OFG	7212	1214.58	833.86	523.16	1343.43	738.09
LSB	2962	2962.19	1971.61	481.41	435.69	523.58
HTD	2505	3422.28	3943.59	268.52	682.92	297.01
FEP	4405	1989.58	1565.89	2431.00	660.60	788.81

Notes: LB(20) denotes Ljung–Box statistics for order 20. The mean statistics report the average valued for the raw data. LB (20) statistics report serial correlation for the data after adjusting the intraday pattern.

2.6.2 Testing for Weak Exogeneity - Empirical Results

Table 2-7 reports the LM test statistics. The first row is the LM statistics for weak exogeneity of duration for volume process and the second row is the LM statistics for weak exogeneity of duration for volatility process. The third row is the LM statistics for jointly weak exogeneity of duration and volume for volatility process.

Table 2-7: Weak Exogeneity Test -- LM Test Statistics

	<u>TRN</u>	<u>R</u>	<u>ARG</u>	<u>TCB</u>	<u>GAS</u>	<u>ABG</u>	<u>OFG</u>	<u>LSB</u>	<u>FEP</u>	<u>HTD</u>
Volume	2.78	20.2	15.8	74.5	>100	0.72	2.00	4.20	8.91	3.02
Volatility	>100	>100	0.51	9.81	>100	>100	>100	2.14	41.0	3.78
Volatility- J	31.9	>100	>100	>100	>100	>100	>100	>100	>100	>100

Note: Critical values $\chi^2(1)_{0.05}=3.84$, $\chi^2(2)_{0.05}=5.99$

$$\chi^2(1)_{0.01}=6.63, \chi^2(2)_{0.01}=9.21$$

First, let's look at the frequently traded stocks. In volume equation, the null hypothesis that duration is weak exogenous is rejected in 4 out of 5 cases. In volatility equation, the same result is found. This suggests that the weak exogeneity of duration in financial point processes is not supported by the frequently traded data. The testing results for joint weak exogeneity of duration and volume in volatility equation are similar. We can see that the null hypothesis is rejected in all the 5 stocks, which suggests the duration and volume are not jointly weak exogenous in volatility equation.

A different picture emerges from infrequently traded stocks. In volume equation, the null of weak exogeneity of duration is not rejected in 4 out of 5 cases (under 1% level). And in volatility equation, the null is not rejected for 2 out of 5 stocks. However, the joint weak exogeneity of duration and volume is rejected in all the 5 cases. The different results found for frequently traded stocks and infrequently traded stocks are striking.

In general, the weak exogeneity of duration is rejected for frequently traded stocks, while it is less likely to be rejected for infrequently traded stocks. But the jointly weak exogeneity of duration and volume is rejected in all of the cases. This indicate that that the empirical model of Engle (2000) and Manganelli (2005) on market

microstructure analysis, in which duration and marks are estimated separately, may only be suitable for infrequently traded stocks. It is more efficient to estimate duration, volume, and price volatility jointly for frequently traded stocks.

2.7 Conclusion

A common practice when modelling several financial point processes jointly is to factor the joint density into the product of the marginal density of duration and conditional density of marks given duration. In estimation, the assumption of weak exogeneity of duration is made in order to estimate the marginal density and conditional density separately. This chapter analyses the issues related to weak exogeneity in financial point processes. We propose three cases of in which the weak exogeneity condition will break down, which extends the application of the Hausman test of weak exogeneity to a time series model. We then do a simulation to study the consequences of incorrectly assuming weak exogeneity in estimation. We find that incorrectly assuming weak exogeneity implied biased estimators. The biases are very large in the third case non-weak exogeneity. In particular, the failure of weak exogeneity assumptions have severe effects on the conditional distribution, where the persistence of volume will be downward biased and the impact of duration on volume will be overestimated. This makes econometric inferences on the parameters are unreliable or even misleading.

In empirical analysis, we derive a test for weak exogeneity based on LM test principles. The LM test is attractive because it only requires estimation of the restricted model. A simulation study suggests that the LM test has good power. We apply the method to two groups of high frequency data. The empirical results indicate that weak exogeneity of duration is often rejected for frequently traded stocks, but is less likely to be rejected for infrequently traded stocks.

Chapter 3 The Lognormal Autoregressive Conditional Duration (LNACD) Model and a Comparison with Alternative ACD Models

3.1 Introduction

In microstructure, the timing of transactions is a key factor in understanding economic theory. For example, the time duration between market events has been found to have a deep impact on the behaviour of market agents and on the intraday characteristics of the price process. Recent models in market microstructure literature based on asymmetric information argue that time may convey information and should be modelled as well.

Econometric modelling the dynamics of transaction time was discussed, for example, in the context of an Autoregressive Conditional Duration (ACD) model by Engle and Russell (1998). They model the arrival times of trades as random variables that follow a point process. The reference ACD model is extended in several ways.

The motivation of this chapter derives from an idea that the ACD model can be formulated as an ARMA specification but the innovation of the ARMA process is highly non-Gaussian. We ask what is the distribution of the ARMA innovation and under what conditions is the innovation is Gaussian distributed.

To answer this question, we begin with a logarithmic version of the ACD model (Bauwens and Giot 2000). The duration x_t is defined as the time elapsed between events occurring at time t_i and t_{i-1} , so that $x_t = t_i - t_{i-1}$. Then the log-ACD model has a standard form; $x_t = \psi_t \varepsilon_t$, $\log \psi_t = \omega + \alpha \log x_{t-1} + \beta \log \psi_{t-1}$. After transformation, the log-ACD model is equivalent to $\log(x_t) = \bar{c} + (\alpha + \beta) \log(x_{t-1}) + e_t - \beta e_{t-1}$, where $e_i = \log(\varepsilon_i)$ and $\bar{c} = \omega - \beta c$. This is an ARMA specification. It is interesting to observe that as long as the innovation of the ACD model follows a lognormal distribution, the innovation of ARMA form will be normal distributed.

The commonly used specifications of the duration distribution in the ACD literature are the exponential distribution (Engle and Russell 1998), Weibull distribution (Engle and Russell 1998; Bauwens and Giot 2000), Burr distribution (Grammig and Maurer 2000; Fernandes and Grammig 2006), and generalized gamma (GG) distribution (Lunde 1999). The use of the lognormal distribution in duration modelling has attracted less interest in the literature. Additionally, empirical studies have found that the hazard function of several types of financial durations may be increasing for small durations and decreasing for long duration (Grammig and Maurer 2000; Bauwens and Veredas 2004). Specifically, the financial duration usually exhibits an inverted U-shaped hazard function. The lognormal distribution seems to capture this stylized factor very well. This motivates us to develop a lognormal ACD (LNACD) model and evaluate its performance.

There are several advantages of using the lognormal distribution to specify the ACD model: (a) the LNACD model permits a hump-shaped hazard function for financial duration (compared with the Exponential ACD and Weibull ACD models), (b) it only has one shape parameter, which implies a simpler computation and estimation procedure (compared to the Burr ACD and generalized gamma ACD

models). (c) it opens the door to use lognormal distribution for other financial point process.

The remainder of this chapter is as follow. Section 2 briefly reviews the ACD models. Section 3 specifies the LNACD model and its likelihood function. Section 4 introduces two latest developed specification tests for the financial duration model. Section 5 reports the empirical results.

3.2 ACD Models

3.2.1 Model Specification

The duration x_t is defined as the time elapsed between events occurring at time t_i and t_{i-1} , such that $x_t = t_i - t_{i-1}$. The basic reference ACD model is proposed by Engle and Russell(1998). They model the conditional expected duration $\psi_t = E(x_t|\Omega_t)$, where Ω_t is the conditioning information set generated by the durations proceeding x_t , as a linear function of past duration and past expected duration:

$$\psi_t = \omega + \alpha x_{t-1} + \beta \psi_{t-1} \quad (3.1)$$

The ACD model is further characterized by the assumption that the standard durations $\varepsilon_t = \frac{x_t}{\psi_t}$, where $f(\psi_t) = \psi_t / E(\varepsilon_t)$, are independently and identically distributed.

To ensure positivity of the conditional duration, common restrictions on the coefficients are that $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. However, if additional explanatory variables implied by market microstructure are included in equation (3.1), a negative value of conditional duration may arise. To avoid this situation, Bauwens and Giot (2000)

propose a logarithmic version of the ACD model. Two specifications are considered, referred to as Log-ACD1 and Log-ACD2, respectively:

$$\log \psi_t = \omega + \alpha \log x_{t-1} + \beta \log \psi_{t-1} \quad (3.2)$$

$$\log \psi_t = \omega + \alpha \varepsilon_{t-1} + \beta \log \psi_{t-1} \quad (3.3)$$

Since the ACD model is very similar to the GARCH model, it is not surprising that the linear ACD model can be extended in several ways. A flexible specification is the augmented ACD (AACD) model by Fernandes and Grammig (2006). The AACD model is obtained by a Box-Cox transformation of the conditional duration and permits an asymmetric response to small and large shocks. The first-order parameterization is given by:

$$\psi_t^\lambda = \omega + \alpha \phi_{t-1}^\lambda [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^v + \beta \psi_{t-1}^\lambda \quad (3.4)$$

Other specifications of the ACD model can be found in a summary paper Pacurar (2008); for instance, the Bauwens and Veredas (2004) stochastic conditional duration model and the Zhang, Russell et al. (2001) threshold ACD model. One can also incorporate additional regressors in the ACD model to model the microstructure effect.

3.2.2 Density Assumption

Engle and Russell (1998) initially consider the exponential distribution for the error ε_i with density

$$f(\varepsilon_t) = \exp(-\varepsilon_t), \quad \varepsilon_t \geq 0 \quad (3.5)$$

The exponential distribution has a flat hazard function, which is too restrictive. Engle and Russell (1998) also consider the Weibull distribution for the error, which nests the exponential distribution as special case ($\gamma = 1$). The standard Weibull density function is :

$$f(\varepsilon_t | \gamma) = \gamma(\varepsilon_t)^{\gamma-1} \exp(-\varepsilon_t^\gamma), \quad \varepsilon_t \geq 0 \quad (3.6)$$

The hazard function is derived analytically as:

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \gamma t^{\gamma-1} \quad (3.7)$$

The Weibull distribution is flexible and permits both an increasing hazard function if $\gamma > 1$ or a decreasing hazard function if $\gamma < 1$

However, Bauwens and Veredas (2004) and Grammig and Maurer (2000) questioned the assumption of a monotonic hazard. They find the hazard function of several types of financial durations may be increasing for small durations and decreasing for long durations. To account for this stylized factor, Grammig and Maurer (2000) propose the use of the Burr distribution and Lunde (1999) propose the use of GG distribution for financial durations. Both the Burr and the GG distributions have two shape parameters which allow for hump-shaped hazard functions. Moreover, the Weibull and exponential distributions can be nested in the Burr and GG distributions. The density and hazard functions for the standard Burr distribution are:

$$f(\varepsilon_t | \kappa, \sigma^2) = \frac{\kappa(\varepsilon_t)^{\kappa-1}}{(1 + \sigma^2 \varepsilon_t^\kappa)^{1+1/\sigma^2}}, \quad \varepsilon_t \geq 0 \quad (3.8)$$

$$\lambda(t) = \frac{\kappa(t)^{\kappa-1}}{1 + \sigma^2 t^\kappa} \cdot \quad (3.9)$$

The density for GG distribution is:

$$f(\varepsilon_t | \alpha, \delta) = \frac{\delta(\varepsilon_t)^{\alpha-1}}{\Gamma(\alpha)} \exp(-\varepsilon_t^\delta), \quad \varepsilon_t \geq 0 \quad (3.10)$$

There is no closed form Hazard function for the GG distribution since it involves the incomplete gamma integral. However, with numerical analysis, Tony (1990) and Lunde (1999) show the conditions under which the hazard functions is increasing, decreasing, U-shaped and inverted U-shaped.

The GG and Burr distributions both allow for flexibility of the hazard function; however, the Burr distribution seems more popular in ACD literature. The main reason lies in the fact that it has closed form survival and hazard functions. However, not all moments for Burr distribution necessarily exist unless some restrictions are imposed on parameters. This may result in poor modelling of the higher (unconditional) moments of duration (Bauwens, Galli et al. 2003); Bauwens, Galli et al. (2003). On the other hand GG distributions involve the incomplete gamma integral and hence none of these functions can be written in closed form. This has led to the GG distribution being less frequently used in ACD literature to some extent. But several specification tests (Fernandes and Grammig 2005; Allen, Lazarov et al. 2009) find that generalized gamma ACD model usually performs better and it is the only model pass the specification test in some cases.

3.3 Methodology

3.3.1 Lognormal ACD Model

In probability theory, a lognormal distribution is a probability distribution of a random variable whose logarithm is normally distributed. See Appendix 7 for introduction of lognormal distribution.

The lognormal distribution is very commonly used in reliability analysis, but less so in financial duration models. To our knowledge, the only paper which mentioned the estimation of ACD models with log-normally distributed errors is Allen, Lazarov et al. (2009). The reasons for this might be (a) it cannot produce a closed form CDF and Hazard function, and (b) the shape of the hazard function is less clear (Telang and Mariappan 2008). However, by comparing the density and hazard function produced by the Burr and lognormal distributions (Figure 3-1), we find that they have a very similar shape. Since it only requires one shape parameter to generate a hump shaped hazard function, it makes the lognormal distribution quite attractive for modelling a duration distribution. This motivates us to investigate the performance of lognormal ACD models.

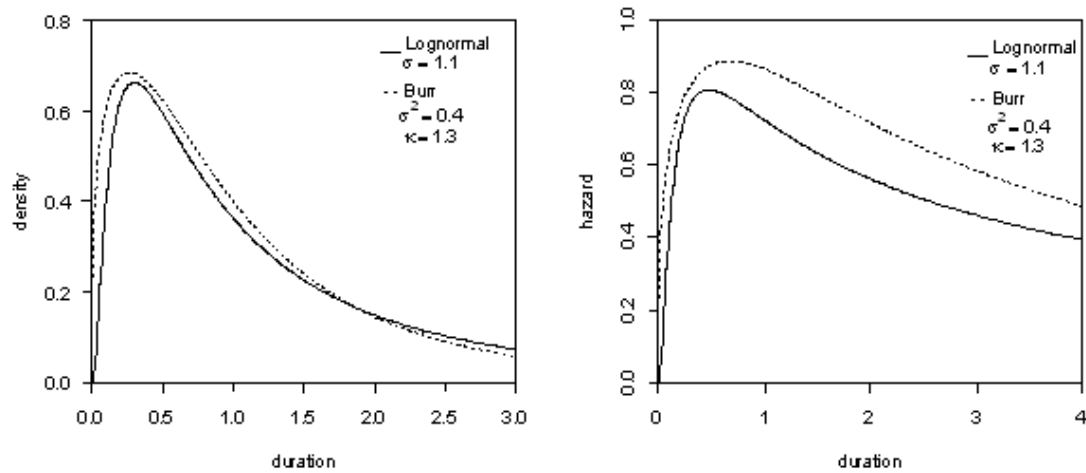


Figure 3-1: Density and Hazard of lognormal and Burr distribution

Defining the dynamics of duration x_t as ACD (1,1) specification of Engle and Russell (1998)

$$\psi_t = \omega + \alpha x_{t-1} + \beta \psi_{t-1} \quad (3.11)$$

and $\varepsilon_t = x_t / f(\psi_t)$, where $f(\psi_t) = \psi_t / E(\varepsilon_t)$, are independently and identically distributed. We consider the properties of the standard lognormal distribution that

arises when the mean of its normal counterpart is zero and the standard deviation is σ . Then the lognormal distribution only has one shape parameter σ . The density function for a standard lognormal distribution is,

$$f(\varepsilon_t | \sigma) = \frac{1}{\sqrt{2\pi\sigma^2} \varepsilon_t} \exp\left[-\frac{\log \varepsilon_t^2}{2\sigma^2}\right], \quad \varepsilon_t \geq 0 \quad (3.12)$$

If the innovations of the ACD model follows a lognormal distribution, then

$$f\left(\frac{x_t}{f(\psi_t)} | \Omega_t, \theta\right) = \frac{f(\psi_t)}{\sqrt{2\pi\sigma^2} x_t} \exp\left[-\frac{(\log x_t - \log(f(\psi_t)))^2}{2\sigma^2}\right] \quad (3.13)$$

where $f(\psi_t) = \psi_t / E(\varepsilon_t) = \psi_t / \exp(\frac{1}{2}\sigma^2)$

The conditional density of x_t is then:

$$f(x_t | \Omega_t, \theta) = \frac{1}{\sqrt{2\pi\sigma^2} x_t} \exp\left[-\frac{(\log x_t - \log(\psi_t) + 0.5\sigma^2)^2}{2\sigma^2}\right] \quad (3.14)$$

The log-likelihood function can be expressed as:

$$L = \sum_{t=1}^T \left\{ -\frac{1}{2} \log(2\pi) - \log(\sigma) - \log(x_t) - \frac{(\log x_t - \log \psi_t + 0.5\sigma^2)^2}{2\sigma^2} \right\} \quad (3.15)$$

3.3.2 ARMA Specification of Lognormal ACD Model

One of advantage in using the lognormal ACD model is that it has an equivalent ARMA specification with an innovation that follows a Gaussian distribution. To explain this, let's change the lognormal ACD model slightly and consider the Log-ACD specification of Bauwens and Giot (2000).

$$\begin{aligned} x_t &= \psi_t \varepsilon_t \\ \log \psi_t &= \omega + \alpha \log x_{t-1} + \beta \log \psi_{t-1} \end{aligned} \quad (3.16)$$

Note that in this specification, we assume that mean of its normal counterpart is $-0.5\sigma^2$ to guarantee that the expectation of the lognormal error term is one. Then taking logs on the first part of equation (3.16)

$$\begin{aligned}\log(x_t) &= c + \log(\psi_t) + e_t \\ \log \psi_t &= \omega + \alpha \log x_{t-1} + \beta \log \psi_{t-1}\end{aligned}\tag{3.17}$$

where $c = -0.5\sigma^2$ and e_t is iid $N(0, \sigma^2)$.

Rewriting equation (3.17), we get

$$\log(x_t) = \bar{c} + (\alpha + \beta) \log(x_{t-1}) + e_t - \beta e_{t-1}\tag{3.18}$$

where $\bar{c} = c + \omega - \beta c$

This is an ARMA (1,1) process where the long dependence of duration is measured by the coefficient of lagged duration, which has the same size as what is measured by the Log-ACD model. The stationarity of the ARMA model is guaranteed by the condition that $\alpha + \beta < 1$ and the invertibility of the ARMA model is guaranteed by the condition that $|\beta| < 1$.

3.3.3 The Shape of Hazard Function of Lognormal Distribution – Numerical Analysis

The hazard function involves the integral of the normal distribution, which does not have any closed form. The behaviour of the hazard rate of the lognormal random variable, as has been reported in some recent publications, is quite misleading. Wadsworth, Stephens and Godfrey (1986) state “for standard deviation σ approximately equal to 0.5, the failure rate is constant; for σ less than 0.2, the failure rate is increasing, while for σ greater than 0.8, it is decreasing”. Sweet (1990) criticises this judgement analytically. He expresses the hazard rate of the lognormal

distribution in terms of the probability density function of a standard normal and plots the curves of the lognormal hazard rate for σ equal to 0.3, 0.5 and 0.7 respectively and it shows a hump-shaped hazard rate. Telang and Mariappan (2008) carry out analytical and numerical investigations of the behaviour of the hazard rate for the lognormal distribution. It was shown that the hazard rate is a unimodal function with convexity upwards. Hence we do a numerical simulation to clarify the shape of hazard function in this section.

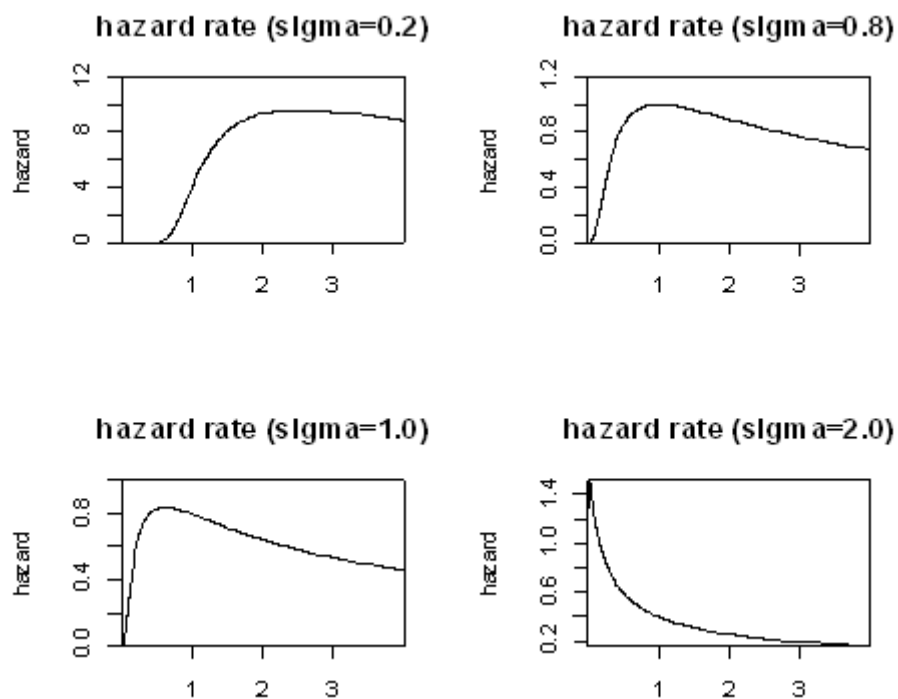


Figure 3-2: Hazard of a lognormal distribution with different σ

Following Sweet (1990), we simply choose a sequence of σ to generate the hazard rate. As we can see in Figure 3-2, the hazard function seems to be either an increasing ($\sigma = 0.2$), hump-shaped ($\sigma = 0.8$ and $\sigma = 1$) or decreasing ($\sigma = 2.0$) function. We further investigate the conditions under which the shape of hazard function is determined. For σ between 0.2 and 1.5, it has been confirmed that hazard

function is hump shaped. The numerical study focus on hazard functions when σ is below 0.2 and above 1.5. We do a simulation study over the interval 0.1 to 0.2 with a step of 0.01 and over the interval 1.5 to 2.5 with a step of 0.1. We use gradient analysis. If its gradient is positive, the hazard function is increasing. If it is zero, the hazard function is constant. And if it is negative, the hazard function is decreasing. We set the criteria that when the 99% numbers of the gradients are positive (negative), the hazard function is increasing (decreasing). The replicate number is 1000. Appendix 5 reports the numerical output. The results are summarized as follow:

1. $0.17 < \sigma < 1.8$ hump-shaped
2. $\sigma > 1.8$ decreasing
3. $\sigma < 0.17$ unstable⁹

From the above analysis, it can be seen that the hazard rates of the lognormal distribution takes on two desirable shapes: decreasing and skewed hump-shaped. The results are consistent with Arnold L. Sweet (1990) and slightly different to Wadsworth, Stephens and Godfrey(1986).

Apparently, the hazard function is not as flexible as in the case of the GG or Burr distribution. For example, the GG distribution can show a decreased, increased U-shaped and inverted U-shaped hazard function, while the lognormal distribution only shows a decreased and U-shaped hazard function. But considering there is only one free parameter, this performance is quite impressive. Furthermore, empirical studies on financial duration work find that the hazard function is either decreasing or has a skewed inverse U-shape (Grammig and Maurer 2000).This makes the lognormal a promising distribution to use when modelling and estimating the ACD model.

⁹ When the standard deviation is smaller than 0.17, the hazard rate is infinite for large values. The smaller the standard deviation, the more will be the numbers of infinite hazard rates. This is not acceptable in reliability analysis. We call this case “unstable”.

3.4 Specification Tests

In this section, we introduce the method to test the adequacy of Lognormal ACD model. Despite the variety of ACD specification in the literature, the question of testing the specification of a particular model has so far attracted less interest. A common way of evaluating ACD models consists of examining the dynamic and distributional properties of the estimated standardized duration. If the estimated model is adequate, the standardized durations are independent and identically distributed (*i.i.d.*). The approach used by Engle and Russell (1998) and followed by subsequent authors, consists of applying the Ljung-Box Q-statistic to test for serial correlation.

However, the question of whether the distribution of the duration is correctly specified is not addressed by this test. In the following we review two recently proposed specification test methods.

3.4.1 Testing Financial Duration Model via Density Forecasts

The first way to test specification of duration model is on the basis of evaluation of density forecasts, which was originally developed by Shephard (1994), Diebold, Gunther et al. (1998) and Kim, Shephard et al. (1998). The motivation behind these procedures is rather intuitive and easily understood. The sequence of probability integral transforms of the one-step-ahead density forecast has a distribution iid uniform $U(0,1)$ under the null hypothesis that the one-step-ahead prediction of the conditional density of duration is the correct density forecast for the data-generating process. Bauwens, Giot et al. (2004) introduce this method for testing financial duration, which is adopted by Allen, Lazarov et al. (2009).

Let us denote by $\{p_t(x_t|\Omega_t)\}_{t=1}^m$ a sequence of one-step-ahead density forecasts and by $\{f_t(x_t|\Omega_t)\}_{t=1}^m$ the sequence of densities defining the data-generating process.

Diebold et al show that the forecast user would always prefer a model which produces the correct density function, regardless of their loss function. This suggests that forecasts should be evaluated by assessing whether the forecasting densities are correct, i.e. where

$$\{p_t(x_t|\Omega_t)\}_{t=1}^m = \{f_t(x_t|\Omega_t)\}_{t=1}^m \quad (3.19)$$

The true density $\{f_t(x_t|\Omega_t)\}_{t=1}^m$ is never observed but Rosenblatt (1952) provides the solution by evaluating its probability integral transform. Under the null hypothesis that the model is correctly specified, the sequence of probability integral transforms

$$z = \int_{-\infty}^{x_t} p_t(u)du \text{ is iid } U(0,1).$$

A straightforward χ^2 goodness-of-fit test can be computed by exploiting the statistical property of the uniform distribution. In addition, Diebold et al. recommend graphical tools that complement statistical tests for *i.i.d.* uniformity. For example, by plotting a histogram based on the empirical z sequence, to detect departures from uniformity, and by plotting the autocorrelogram for z -sequence, to identify potential deficiencies of a model to account for the dynamics of the duration model.

3.4.2 Non-parametric Testing of Conditional Duration Model

One drawback of the density evaluation method is that the effect of parameter estimation is not considered. Fernandes and Grammig (2005) introduce tests for the distribution of the error term based on a comparison between parametric and non-parametric estimates of the density function of the standard durations.

The first step consists in estimating the conditional duration process by QML. The second step then gauges the closeness between parametric and nonparametric

estimates of the baseline density function of the residuals. To be more specific, Fernandes and Grammig (2005) test the null

$$H_0: \exists \theta_g \in \Theta \quad \text{such that} \quad g(x, \theta_g) = g(x) \quad (3.20)$$

where $g(x)$ is the true density of the standardized duration and $g(x, \theta_g)$ the density implied by the parametric model. The alternative hypothesis is that there is no such $\theta_g \in \Theta$. The true density $g(\cdot)$ is of course unknown. The authors advise to estimate the density function using a non-parametric kernel method, which produces consistent estimates irrespective of the parametric specification of the distribution. Since the parametric density estimator is consistent only under the null, the natural test is to gauge the closeness between these two density estimates (Fernandes and Grammig) therefore measure the following distance:

$$\Psi_g = \int_0^\infty I(x \in S) \{g(x, \theta) - g(x)\}^2 g(x) dx \quad (3.21)$$

where $I(\cdot)$ is the indicator function. The compact subset S is introduced to avoid regions in which density estimation is unstable. This test is referred as D-test. The sample analogy of (3.21) reads

$$\Psi_{\hat{g}} = \frac{1}{T} \sum_{t=1}^T I(x \in S) \{g(x, \hat{\theta}) - \hat{g}(x)\}^2 \quad (3.22)$$

where $\hat{\theta}$ and $\hat{g}(\cdot)$ denote consistent estimates of the true parameter θ_g and density $g(\cdot)$, respectively. The null hypothesis is rejected if the D-test statistic $\Psi_{\hat{g}}$ is large enough. Under the null and with a set of regularity assumptions, Fernandes and Grammig show that the limiting distribution of the D statistic is

$$h_n^{1/2}(\Psi_{\hat{g}} - \hat{\sigma}_D^2) \xrightarrow{d} N(0, \hat{\sigma}_D^2) \quad (3.23)$$

h_n denotes the bandwidth used for the density estimation and $\hat{\delta}_D$ and $\hat{\sigma}_D^2$ are consistent estimates of $\delta_D = e_K E[I(x \in S)g_x]$ and $\sigma_D^2 = v_K E[I(x \in S)(g_x)^3]$, respectively, where $e_K \equiv \int_u K^2(u)du$ and $v_K \equiv \int_v \left\{ \int_u K(u)K(u+v)du \right\}^2 dv$

The test inspects the whole distribution of the residuals, not only a limited number of moment restrictions, and is shown to be nuisance parameter free. All results are derived under mixing conditions; hence there is no need to perform a previous test for serial independence of the standardized durations.

Since the standardized duration is bounded and strictly positive, D-test statistic may perform poorly due to the boundary bias that haunts non-parametric estimation using fixed kernels. One solution is to work with log-duration whose support is unbounded, using the result that for $Y = \log(X)$ $Y = \log(X)$ we have

$$f_Y(y) = f_X[\exp(y)]\exp(y).$$

3.5 Empirical Application

3.5.1 Data

In this section, we use real data to assess the significance of using the lognormal as the distribution of duration in ACD model. The main purpose of this empirical analysis is to compare the performance of lognormal ACD and alternative specification of ACD models. Grammig and Maurer (2000) has developed a Burr ACD model and intensively compares the Burr ACD with other specification of ACD models. For comparison reason, we use same data used by Grammig and Maurer (2000). A full description of the data can also be found in Bauwens and Giot (2000,2003), who have constructed a database from the NYSE Trade and Quote (TAQ) raw data. They choose the data for the months of September, October, and

November 1996 and construct the price duration of five actively traded stocks: Boeing, Coca-Cola, Disney and Exxon and IBM. The price duration is defined by thinning the quote process with respect to a minimum change in the mid-price of the quotes. More specifically, price duration is defined as the time interval needed to observe a cumulative change in the mid-quote of at least \$0.125. Durations between events recorded outside the regular opening hours of the NYSE (9:30am to 4:00 pm) are discarded.

As documented by a previous empirical study(Giot 2000; Bauwens, Giot et al. 2004),the duration process features a strong time-of-day effect, which stems from predetermined market characteristics such as opening/closing of trading or lunch time for traders. To adjust for this, we employ a method used by Engle (2000). The raw duration is first regressed on a cubic spline using knots of every half hour. Separate splines are used for each day of the week. The adjusted durations are then obtained via dividing the raw duration by the spline forecast. For brevity of notation, we will henceforth refer to the time-of-day adjusted durations simply as durations. Table 3-1reports the descriptive statistics for the price duration. The price durations for the group of stocks exhibit two features: high serial correlation and some degree of overdispersion, which motivates ACD modelling.

Table 3-1: Descriptive statistics for the price durations.

Stock	Sample size	Mean	Overdispersion	Q(10)
Boeing (BA)	2620	1.001	1.338	322.3
Coca-Cola(KO)	1690	1.002	1.171	69.7
Disney (DIS)	2160	1.004	1.209	137.3
IBM	6728	1.015	1.427	1932.6

Note: Overdispersion stands for the ratio between standard deviation and mean. $Q(10)$ denotes the Ljung-Box statistics for order 10.

3.5.2 Estimation Results

In estimation, we use first two thirds of the sample for each stock. The remainder of the dataset is reserved for out-of-sample analysis. To simplify, we restrict the model for estimation to the ACD (1, 1) specification and exclude any pre-determined explanatory variables. For purpose of comparison, alternative ACD models, specified in section 2, are estimated in addition to the Lognormal ACD model. The alternative ACD models are, henceforth, referred to as the EACD (Exponential ACD model), the WACD (Weibull ACD model), the BACD (Burr ACD model) and the GGACD (Generalized gamma ACD model). In addition, we estimate both ACD (1,1) and Log-ACD(1,1) specifications for the mean, and find that the results do not differ qualitatively. The estimation results for the Log-ACD models are not reported here for the sake of brevity. Maximum Likelihood estimation results and robust standard errors of the ACD models are reported in Table 3-2.

For LNACD estimates, all coefficients are significant. The sum of a and β is always greater than 0.9, implying high persistence of duration. The estimated shape parameters σ are all significant and appropriately equal 1.2, suggesting an estimated inverted U-shaped hazard function from LNACD model.

For EACD and WACD models, the estimates of α and β are similar and the shape parameter of the WACD model (γ) is small but very close to 1. This suggests that nothing much is gained by replacing the exponential distribution with the Weibull distribution for the specification of the error. Lunde (1999) argues that the sum of a and β of the EACD and WACD models is upward biased towards greater persistence, while the BACD and GGACD models produces an unbiased estimate of persistence, which is also confirmed from our empirical results. The spurious persistence of the

conditional mean implied by EACD and WACD model is due to the lack of flexibility of the distribution.

It is notable that the lognormal ACD estimates of a and β are significantly different to the EACD and WACD estimates. Indeed, they are close to the BACD and GGACD estimates. In particular, the LNACD seems to give quite similar coefficient estimates for a and β to the GGACD estimates, although the lognormal distribution is not nested by the generalized gamma distribution and the former has one parameter less. Given the fact that empirical studies support GGACD and BACD model, the LNACD model performs pretty well in terms of the coefficient estimates.

The shape parameters estimated from BACD model are significantly different from zero and the shape parameters estimated from GGACD model are significantly different from one. This empirical evidence does not support a reduction from the Burr or GG specification to a simpler distribution (i.e Exponential or Weibull distribution). The lognormal distribution, which is restricted to one free shape parameter, is not nested in the Burr or GG distribution. Thus, justifying the flexibility brought about by the use of the Burr and GG formulations does not imply a rejection of the lognormal formulation of the ACD model.

With respected to the hazard function, both BACD and GGACD estimates support an inverted U-shaped conditional hazard function, which is also consistent with the LNACD model.

Table 3-2: ACD model estimates results

		ω	a	β	κ	σ^2
BA	EACD	0.031 (0.023)	0.114 (0.041)	0.861 (0.059)		
	WACD	0.034 (0.025)	0.121 (0.042)	0.851 (0.061)	0.895 (0.016)	
	BACD	0.057 (0.033)	0.169 (0.046)	0.789 (0.067)	1.093 (0.036)	0.339 (0.061)
	LNACD	0.089 (0.047)	0.199 (0.056)	0.759 (0.081)	1.259 (0.022)	
	GGACD	0.081 (0.040)	0.188 (0.040)	0.744 (0.069)	0.550 (0.043)	2.422 (0.334)
KO	EACD	0.159 (0.042)	0.109 (0.026)	0.727 (0.051)		
	WACD	0.159 (0.026)	0.109 (0.042)	0.727 (0.051)	0.959 (0.019)	
	BACD	0.161 (0.042)	0.124 (0.030)	0.715 (0.051)	1.124 (0.050)	0.286 (0.079)
	LNACD	0.181 (0.048)	0.136 (0.036)	0.711 (0.055)	1.182 (0.025)	
	GGACD	0.178 (0.043)	0.122 (0.027)	0.692 (0.050)	0.568 (0.048)	2.564 (0.389)
DIS	EACD	0.074 (0.030)	0.046 (0.015)	0.889 (0.033)		
	WACD	0.074 (0.031)	0.046 (0.015)	0.888 (0.034)	0.969 (0.018)	
	BACD	0.099 (0.044)	0.048 (0.018)	0.867 (0.049)	1.219 (0.045)	0.396 (0.067)
	LNACD	0.153 (0.088)	0.050 (0.023)	0.830 (0.085)	1.148 (0.022)	
	GGACD	0.137 (0.078)	0.051 (0.017)	0.825 (0.078)	0.567 (0.045)	2.684 (0.389)
XON	EACD	0.065 (0.037)	0.046 (0.016)	0.890 (0.048)		
	WACD	0.066 (0.038)	0.045 (0.016)	0.889 (0.049)	0.962 (0.016)	
	BACD	0.102 (0.055)	0.039 (0.015)	0.863 (0.061)	1.250 (0.044)	0.464 (0.068)
	LNACD	0.156 (0.323)	0.031 (0.020)	0.827 (0.323)	1.144 (0.020)	
	GGACD	0.945 (0.037)	0.053 (0.029)	0.000 (.)	0.411 (0.065)	4.840 (1.475)
IBM	EACD	0.010 (0.005)	0.090 (0.019)	0.905 (0.021)		
	WACD	0.010 (0.005)	0.090 (0.019)	0.904 (0.021)	0.985 (0.011)	
	BACD	0.017 (0.009)	0.112 (0.029)	0.880 (0.033)	1.263 (0.025)	0.420 (0.038)
	LNACD	0.032 (0.017)	0.136 (0.039)	0.853 (0.050)	1.121 (0.012)	
	GGACD	0.023 (0.011)	0.125 (0.026)	0.859 (0.033)	0.536 (0.028)	3.092 (0.298)

Note: If it is the WACD model or the LNACD model, κ stands for the shape parameter of (γ) and (σ^2) , respectively. If it is GGACD model, κ and σ^2 stand for the corresponding shape parameter of δ and a

3.5.3 Specification Test Results

To further evaluate the performance of the Lognormal ACD model, we employ the test procedures introduced in section 3.4. We conduct both in-sample and out-of-sample tests. The alternative specification of the ACD model is tested at the same time for comparative purposes. One advantage of our tests is that they both allow for visual diagnostic checks, which are helpful for interpreting the numerical outputs. For sake of brevity, we take the Boeing and EXXON results to illustrate the visual diagnostic checks. The graphs for the other stocks are more or less the same. They are attached in Appendix 6.

Figure 3-3 and Figure 3-4 depict the non-parametric density function and their parametric counterparts implied by EACD, WACD, LNACD, BACD and GGACD for the two stocks. The non-parametric density is computed based on a Gaussian kernel and log-durations. For the Boeing stock, the non-parametric density appears to fluctuate tightly around the parametric densities implied by LNACD, BACD and GGACD, while the performance of LNACD seems the best. Densities implied by EACD and WACD are more or less further away from the non-parametric density. For EXXON stock, the similar results hold. The parametric densities implied by LNACD, BACD and GGACD are close to the non-parametric density, while the GGACD seems to match the non-parametric density most closely. Densities implied by EACD, WACD are deviate considerably away from non-parametric density.

Figure 3-5 and Figure 3-6 plot a histogram of z . Z-sequences is the distribution of the empirical probability integral transform produced by the conditional duration

forecast. Under the null hypothesis, the z -sequence is *iid* $U(0,1)$. For EXXON, the histograms for EACD and WACD have a distinct non-uniform distribution. It can be seen that far too few realizations fall into the very low tails of the forecast densities. One would expect more observations under the null hypothesis (data were really generated by the assumed data generating process). On the other hand, small (but not very small) durations are over-represented: the frequencies associated with the third to seventh histogram bins are above the confidence interval. The reason for this result can be found as follows. Because the estimates of the γ parameter for the Weibull distribution are smaller than one, the density tends to infinity as x tends to zero and then drops dramatically as x increases. As a result, there are not enough very small durations and fewer small (but not too small) durations. A similar explanation holds for the exponential distribution. It can be seen that the z -histogram for GGACD matches the z -histogram of uniform distribution most closely. The LNACD also has similar z -histogram to the uniform distribution and the performance of BACD is slightly worse than LNACD. For Boeing, the z -histograms of all the 5 models seem not to match the z -histogram of the uniform distribution, but still the ones implied by LNACD BACD and GGACD are the closest.

Figure 3-7 and Figure 3-8 depict the autocorrelations for z that are significant at the 5% level. For the two stocks, all ACD specifications capture the duration dynamics in more or less the same way. The LNACD, BACD and GGACD show slight advantages. The last two columns in Table 3-3 contain the number of autocorrelations (out of 50) for z that is significant. It can be seen that GGACD model performs best in most of the cases, followed by BACD and LNACD models.

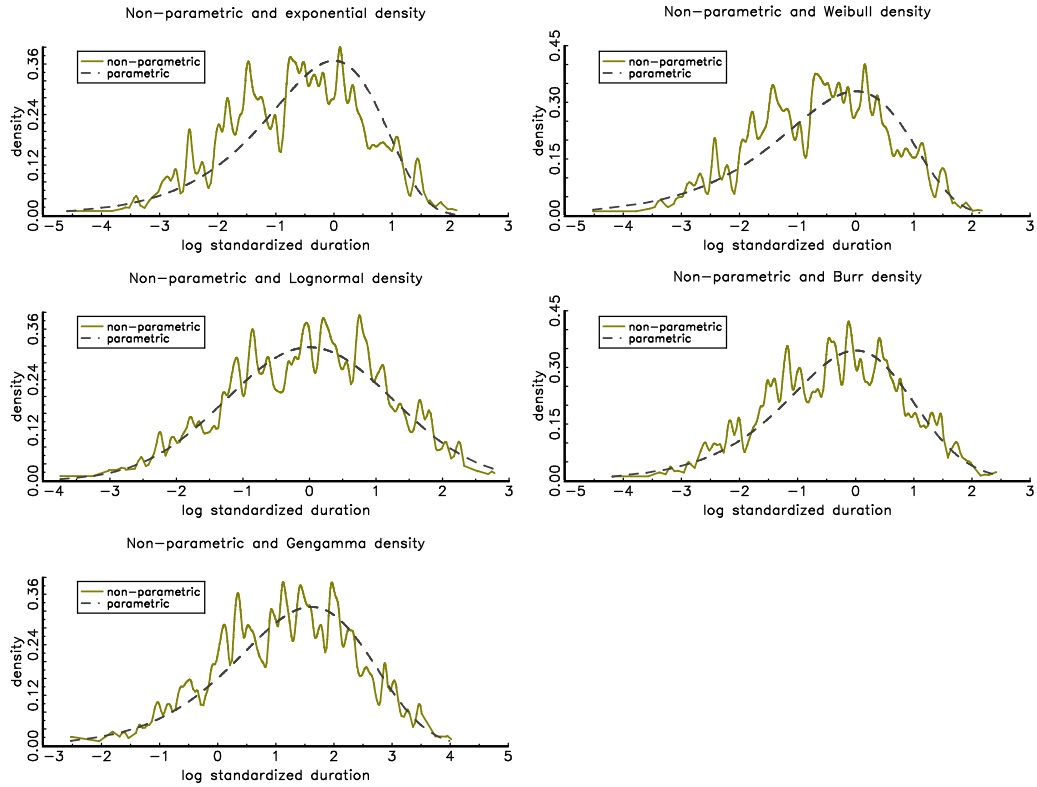


Figure 3-3: Non-parametric and parametric densities: BA Out sample results

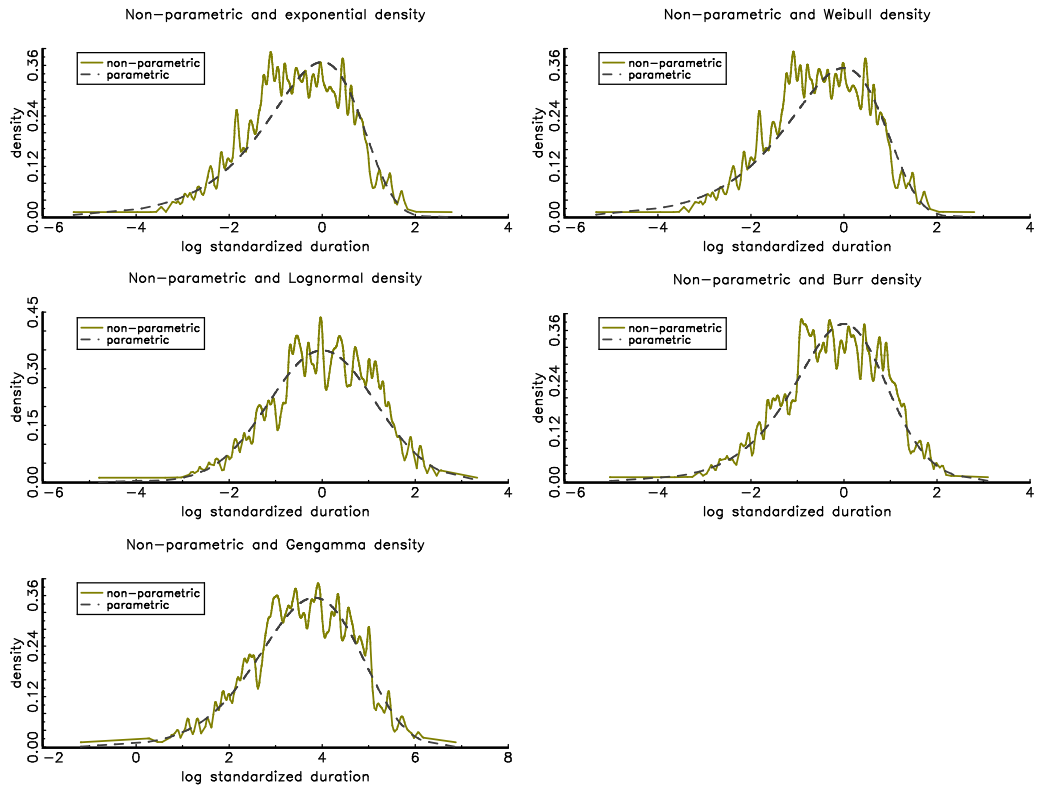
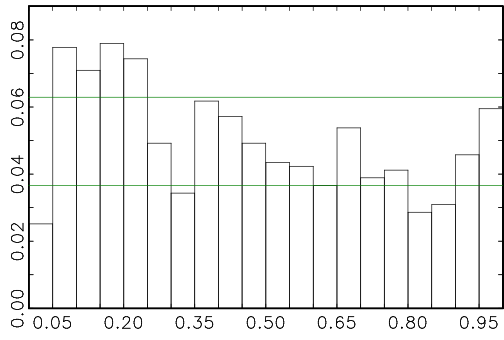
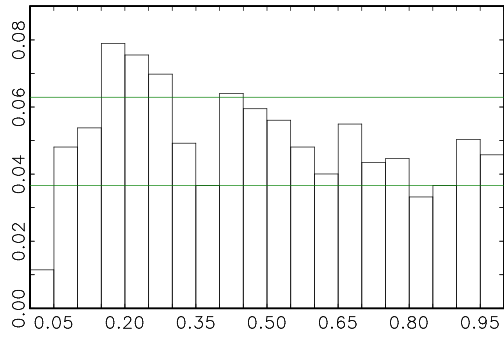


Figure 3-4: Non-parametric and parametric densities: XON Out sample results

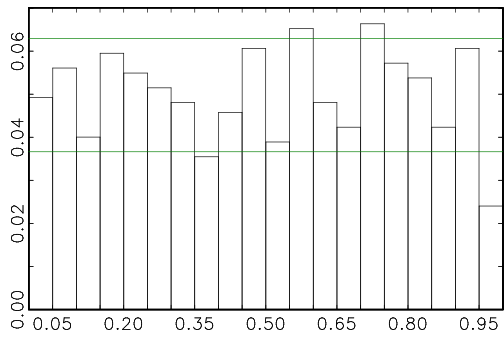
Histogram of Z – Exponential ACD model



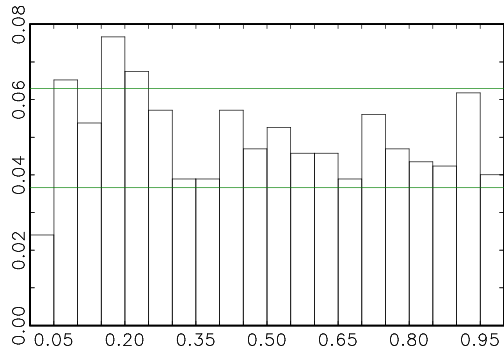
Histogram of Z – Weibull ACD model



Histogram of Z – Lognormal ACD model



Histogram of Z – Burr ACD model



Histogram of Z – Gengamma ACD model

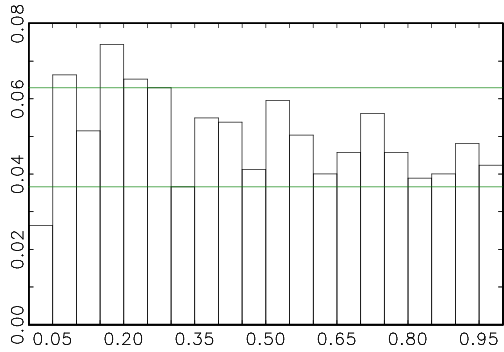
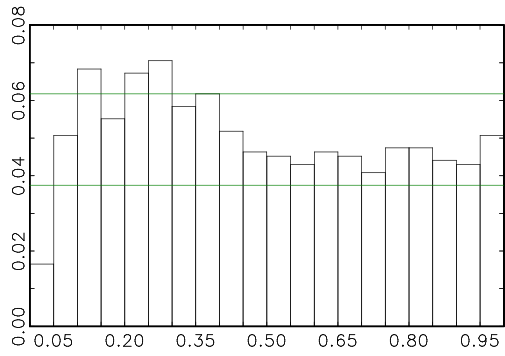
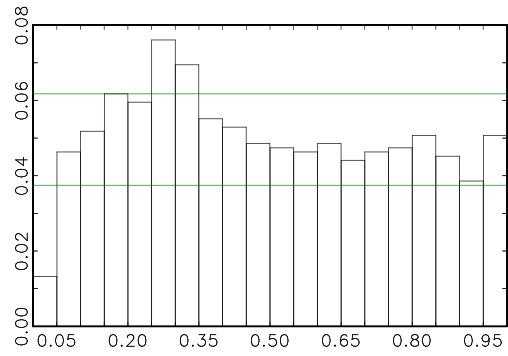


Figure 3-5: Histogram of Z: BA Out sample results

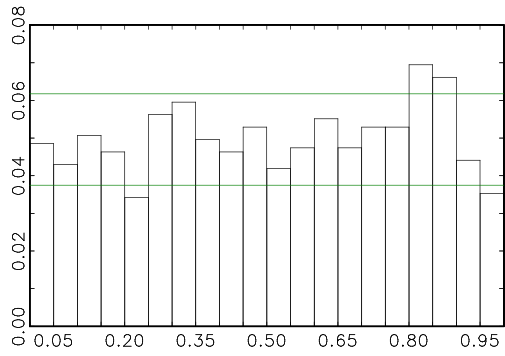
Histogram of Z – Exponential ACD model



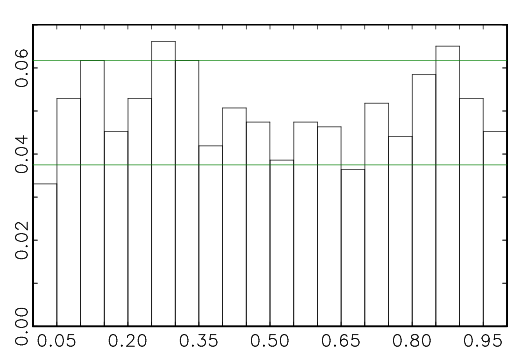
Histogram of Z – Weibull ACD model



Histogram of Z – Lognormal ACD model



Histogram of Z – Burr ACD model



Histogram of Z – Gengamma ACD model

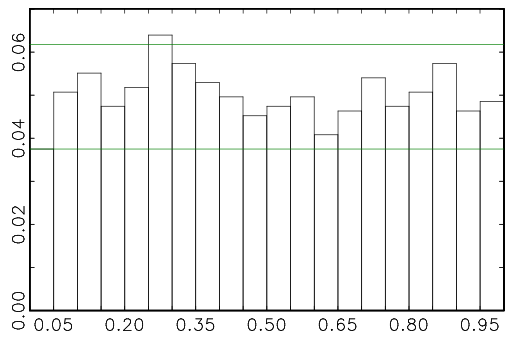


Figure 3-6: Histogram of Z, XON out of sample results.

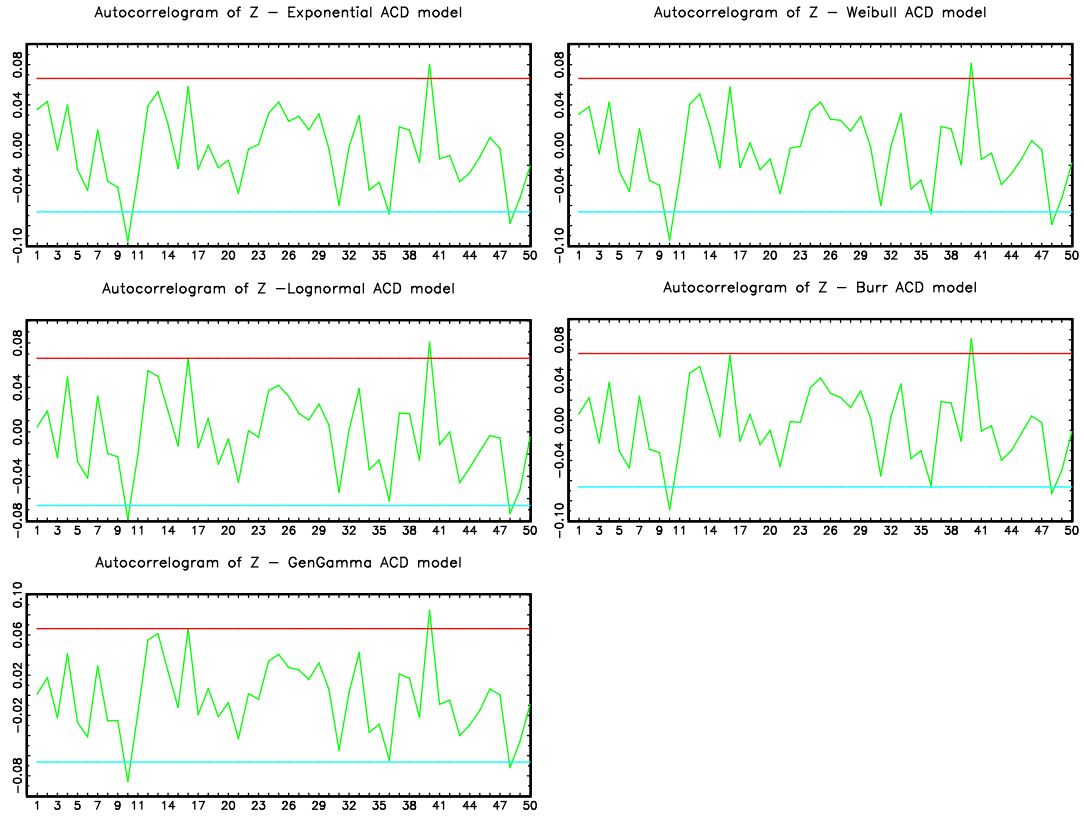


Figure 3-7: Autocorrelation of z for BA out sample

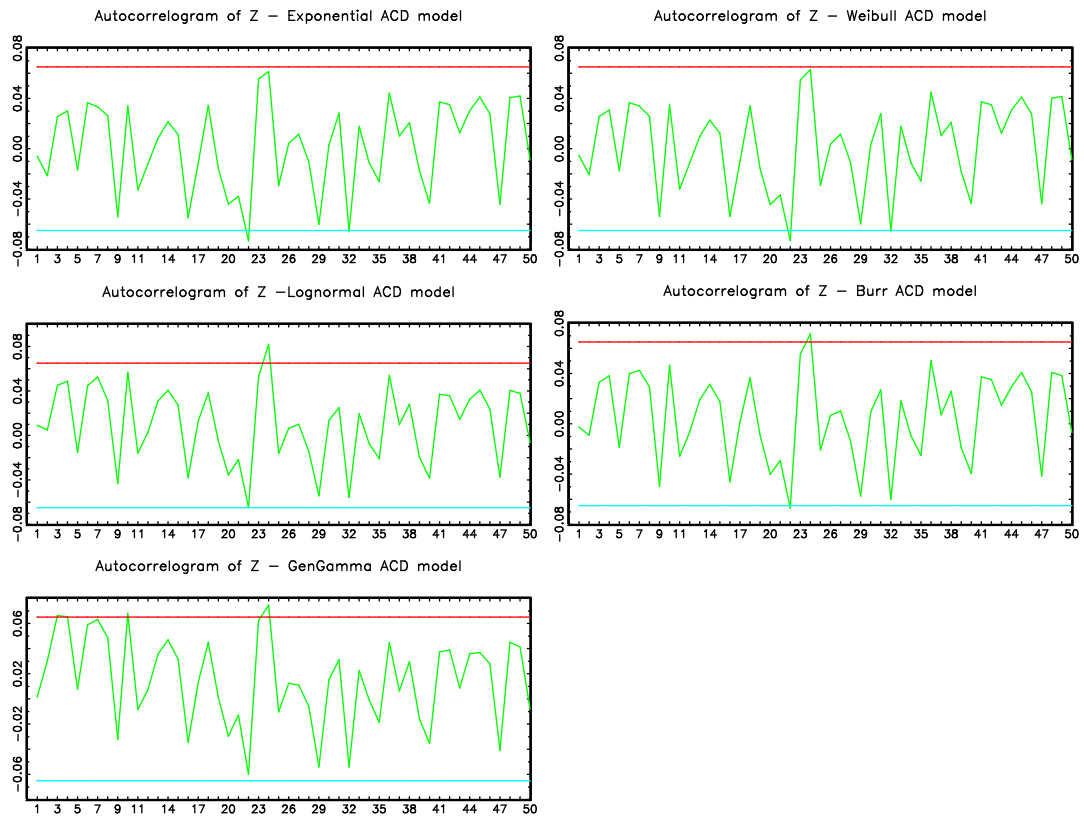


Figure 3-8 : Autocorrelation of z for XON out of sample

Table 3-3: Specification test results.

		D-test		$\chi^2 - GOF$		AC(z)	
Stock		In sample	Out of sample	In sample	Out of sample	In sample	Out of sample
BA	EACD	0.000	0.000	0.000	0.000	8	4
	WACD	0.000	0.000	0.000	0.000	9	4
	LNACD	0.000	17.24	0.003	0.53	6	3
	BACD	13.77	0.94	20.10	0.01	7	3
	GGACD	43.84	0.47	32.00	0.04	6	3
KO	EACD	2.91	82.07	0.000	1.77	2	0
	WACD	31.62	87.68	0.02	11.40	2	0
	LNACD	54.68	20.84	11.4	8.29	2	0
	BACD	66.57	96.89	19.72	6.15	2	0
	GGACD	94.39	97.70	69.76	46.07	2	0
DIS	EACD	0.00	0.00	0.00	0.00	5	20
	WACD	0.00	0.00	0.00	0.00	5	20
	LNACD	0.14	26.22	0.40	0.00	4	29
	BACD	15.98	0.00	5.10	0.00	4	24
	GGACD	10.78	0.00	46.23	0.00	3	23
XON	EACD	0.00	0.67	0.00	0.00	4	2
	WACD	0.00	2.80	0.00	0.01	4	2
	LNACD	46.78	32.30	14.63	10.77	3	1
	BACD	13.72	26.06	2.56	5.52	3	2
	GGACD	53.14	91.31	12.65	87.76	3	4
IBM	EACD	0.00	0.00	0.00	0.00	7	4
	WACD	0.00	0.00	0.00	0.00	7	4
	LNACD	0.00	3.47	0.00	0.00	5	4
	BACD	0.28	0.00	0.03	0.00	6	6
	GGACD	34.49	0.00	1.37	0.00	5	5

The numerical statistical output is summarized in Table 3-3. The D-test statistics are computed based on a Gaussian kernel and log-durations. The corresponding p-values of are reported in the table. A large p-value means that the model passes the test and fits data well. The tests indicate a clear rejection of EACD and WACD model, with the sole exception of Coca-Cola. For the other four stocks, the in sample test results are in favour of GGACD model, while the out of sample results support

LNACD model. IBM is an exception, since all models are rejected, but the GGACD and LNACD still produce the best results.

The χ^2 goodness-of-fit test is designed to evaluate the models' density forecasts. The results are in line with D-test results. EACD and WACD are clearly rejected for all stocks, including Coca-Cola. The GGACD produces the largest p-value in most of cases. BACD and LNACD have a similar performance and perform slightly worse than GGACD model. However, the superiority performance compared to EACD and WACD is obvious. The p-values for IBM, once again, are smaller than 5% for all specifications.

In all, the performance of the LNACD specification is quite impressive, considering that it has one fewer free parameter than the BACD and GGACD specifications. The GGACD specification performs best in most cases while LNACD is superior to specifications other than GGACD.

3.6 Conclusion

In this chapter, we extend the Engle and Russell (1998) ACD model to a Lognormal ACD model. The Lognormal ACD model permits a humped-shaped hazard function with one free shape parameter which demonstrates computational advantages compared to the ACD specification. We compare the performance of the Lognormal ACD model with an alternative specification of the ACD model. The empirical results show that the Lognormal ACD model is superior to the Exponential and Weibull ACD models and its performance is similar to the Burr ACD and Generalized Gamma ACD models. The significance of this study is that it opens a door to use lognormal distribution for financial point processes.

Chapter 4 The Dynamics of Trading Duration, Volume and Price Volatility – a Vector MEM Model

4.1 Introduction

Microstructure theory generally indicates that trading duration and trading volume convey information with respect to fundamental asset prices, and reflect the behaviour of financial market participants.¹⁰ Since French and Roll (1986) have found evidence that price volatility is caused by private information that affects prices when informed investors trade, the empirical studies on trade and price processes have been based on increasingly on the analysis of the dynamics of trading duration, volume and price volatility. However, prior research on this issue is based on a recursive framework, in which the trade and price processes are independent of each other.

In this chapter, we extend the recently developed recursive framework of Engle (2000) and Manganello (2005) for high frequency data to a vector MEM model in which the trading duration, volume and price volatility are involved simultaneously and are interdependent. Based on the results from Chapter 3, we further propose a multivariate lognormal for the distribution of the vector model, which allows the innovation terms to be correlated contemporaneously. In addition, maximum

¹⁰ In general, duration is considered to reflect the trading strategy of informed traders or is an indicator of liquidity (Easley and O'Hara 1992), while volume is viewed as an important determinant of the strength of a market move and reflects information about changes in investors' expectations (Harris and Raviv, 1993).

likelihood is proposed as a suitable estimation strategy. The vector MEM release two restrictions often imposed by previous empirical work and incorporates various causal and feedback effects among these variables. We also construct impulse response functions that show how the price reacts to a perturbation of its long-run equilibrium. The method is applied to a trade and quote dataset of the NYSE, and the model is estimated using a sample of ten stocks.

Our empirical results are generally consistent with the previous findings in the empirical microstructure literature (see, for example, Dufour and Engle (2000), Engle (2000) and Manganelli (2005)). But our work is novel in two ways. First, we find that duration and duration shocks have a significant impact on price volatility, while only the unexpected components of volume are considered to carry information content with respect to price. This generally suggests that it is the unexpected components of trading characteristics rather than the trading variables themselves that carry information content with respect to fundamental asset prices. In addition, impulse response analysis shows that shocks to duration or volume are incorporated appropriately into the price within one trading day for frequently traded stocks, but this takes up to one week for infrequently traded stocks. Second, our empirical results suggest that volatility has a negative impact on trading intensity, while volatility shock has a positive impact on trading intensity. We explain this by considering the persistent quote change (volatility) to be motivated by information based reason, and transient quote change (volatility shock) to be motivated by inventory based reason. The results confirm Hasbrouck (1988,1991)'s prediction that persistent quote changes (volatility) reduce trading intensity and transient quote changes increase trading intensity.

The remainder of this chapter is organized as follows. Section 2 reviews the relevant literature; the theoretical and empirical work on the relationship of duration, volume and volatility are reviewed in this section. Section 3 outlines the empirical motivation and describes the model and methodology used in the analysis. Section 4 introduces the high frequency data and empirical results. Section 5 concludes the chapter.

4.2 Relevant Theoretical and Empirical studies

Theoretically, the market microstructure literature explains trading activity using two types of model: information based and inventory based models. Accordingly, predictions of the relations between duration, volume and price volatility differ. In empirical analysis, the operation of the market is customarily undertaken by using time-series, high-frequency data. The dynamics of such positive-valued variables is generally modelled by a type of ACD model. In this section, we first review the market microstructure prediction of the relations between duration, volume and volatility, and then the ACD model of the relevant empirical findings on these relationships.

4.2.1 Market Microstructure Predictions

The market microstructure literature explains trading activity using two types of model: information based and inventory based models. Specifically, trading occurs either for information motivated or liquidity motivated reasons. A compact overview of the market microstructure was discussed in Chapter 2.2. This section gives reviews on relevant predictions with respect to the relations between trading duration, volume and price volatility.

The market indicators, commonly of most interest, are time stamps of trades, the best bid/ask quote updates, the traded volume, and the best bid-ask price. Among the key variables considered, the timing of the trade plays an important role. It is ignored initially, and incorporated explicitly into market microstructure models by Diamond and Verrecchia (1987) and Easley and O'Hara (1992). Diamond and Verrecchia (1987) develop a rational expectations model with short-sale constraints. They summarize the time effect of trade as “No trade means bad news”. In their model, the informed traders' actions are driven by the arrival of private information, while uninformed traders are assumed to trade for reasons unrelated to the arrival of such information. If the news is bad, informed traders will wish to sell (or, alternatively, to short-sell if they do not own the stock). Given short-sale constraints, there may be no trade. Therefore, long durations are associated with bad news and should lead an adjustment of the prices and hence to increase the return volatility. This is summarized as “No trade means bad news”. The implicit prediction from their model is that long durations increase the price volatility of the next trade.

Easley and O'Hara (1992) provide a different explanation for the role of time. Informed traders only trade when there is new information (whether good or bad) arriving in the market. So variations in trading intensity are closely related to the change in the participation rate of informed traders. It follows that short trade duration is a signal that informed traders are participating in the market. Consequently, the market maker adjusts his/her prices to reflect the increased risk of trading with informed traders, which reveals a higher volatility and wider bid–ask spreads in the market. To summarize, ‘No trade means no news’. In the strategic trading assumption, the informed trader may choose to segment large volume trades into a greater number of smaller-volume, information-based trades, and hence conceal their type and make

full use of private information. It follows that both trading intensity and trading volume may provide information concerning the behaviour of market participants. A consequence of this is that short durations and high volumes should increase the price volatility of the next trade.

A relationship between time duration and price volatility is also explained by the model of Admati and Pfleiderer (1988). It is assumed that frequent trading is associated with liquidity traders, and therefore low trading means that liquidity (discretionary) traders are inactive, which leaves a high proportion of informed traders in the market. This again translates into quick price adjustment and hence high volatility.

Goodhart and O'Hara (1997) examine the price effect of trade. Traders may learn over time from the information-based model, and adjust their speed of trading in reaction to this. For example, a large change in a market maker's mid-quote price may be a signal to the informed traders that their private information has been revealed to the market makers, assuming that no new signal has been released subsequently. This means that private information is no longer superior, and therefore the incentive to trade disappears, which decreases trading intensity. However, from the inventory model perspective, large quote changes would immediately attract opposite-side traders, thus increasing trading intensity. In addition, when uninformed traders behave strategically (O'Hara 1995), it becomes more complex, since the uninformed will increase the probability they attach to the risk of informed trading when they observe large absolute returns or large trading volume. Consequently, they will reduce the overall trading intensity. Hasbrouck (1988,1991) explains the two effects using the short-run and long-run characteristics of trading behaviour. The private information is persistent and long-lived; the persistent quote change is related to private information,

and should have a negative impact on trading intensity. The inventory level in stationary and inventory control is inherently a transient concern, the transient quote change is related to inventory control, and has a positive impact on trading intensity.

Table 4-1 summarizes the related market microstructure literature and its predictions.

Table 4-1: Summary of the related market microstructure literature

Model		Authors and year	Main feature	Predictions
Information-based model	Sequential trade model	Glosten and Milgrom (1985)	All agents act competitively	Volume is positive correlated with price volatility
		Diamond and Verrecchia (1987)	Short sale constraints Incorporating time	No trade means bad news (long durations increase the price volatility of the next trade.)
	Strategic trade model	Kyle (1985)	Informed traders act strategically Long-lived information	
		Easley and O'Hara (1992)	Incorporating time	No trade means no news (short durations and high volumes increase the price volatility of the next trade.)
		Admati and Pfleiderer (1988) Parlour (1998)	Uninformed traders also act strategically Short-lived information Rational expectations	Trade intensity increases, the informativeness of trades decreases. Large quote change is a risk of informed trading; liquidity traders may leave or slow down trading activity
Inventory-based model	Ho and Stoll (1981) O'Hara and Oldfield (1986) Hasbrouck (1991)		Market makers use price to balance their inventory	Large quote changes attract opposite-side traders, thus increasing trading intensity

4.2.2 Empirical Studies

Empirical investigation of market microstructure predictions is subject to the availability of high-frequency transaction data. Statistically speaking, high-frequency data are realizations of point processes; that is, the arrival of the observations is random. This, jointly with other unique features of financial data (long memory; strong skewness; and kurtosis) implies that new methods and new econometric models are needed. It was first addressed, by Engle and Russell (1998) in the context of an ACD model for the dynamics of transaction time. It represents the time duration as product of a (autoregressive) scale factor and non-negative valued random process.

In the ACD framework, the trade characteristics associated with time are incorporated and modelled simultaneously, so that the market microstructure predictions can be evaluated at the transaction level. Zhang, Russell et al. (2001) develop an threshold ACD model and find that the fast trading regime is characterized by wider spread, larger volume and high volatility, all of which proxy for informed trading. Taylor (2004) models future market trading duration using various augmentations of the basic ACD model, and confirms that bid–ask spread and transaction volume have a significant impact on the subsequent trading intensity.

The most significance work is done by Engle (2000). He proposes a recursive framework to represent the dynamics of duration and other trading characteristics jointly, so that various market microstructure predictions can be tested empirically. In Engle (2000), the joint density of duration and volatility is expressed as the product of the marginal density of the duration times and the conditional density of the volatility, given the duration. The result provides evidence of the bad-news effect of long durations, which is the reverse of the Diamond and Verrecchia (1987) result. The recursive framework of Engle (2000) reduces the complexity of the model, since each

process is estimated separately, and used widely by later empirical works. For example, Engle and Sun (2007) model the joint density of the duration and the tick-by-tick returns within a recursive framework. They build an econometric model for estimating the volatility of the unobserved efficient price change. Using this model, it is easy to forecast the volatility of returns over an arbitrary time interval through simulation using all the observations available.

Manganelli (2005) notes that other high-frequency data (trading volume, bid–ask spread) share similar characteristics to duration (for example, they are positive-valued and persistently clustered over time), so that their dynamics can be represented using the same autoregressive process. He further extends Engle (2000)’s model by incorporating the trading volume and develops a framework to model jointly duration, volume and price volatility. Following Engle (2000), the joint distribution of duration, volume and volatility is decomposed into the product of the marginal distribution of duration; the marginal distribution of volume, given duration; and the conditional distribution of volatility, given duration and volume. Further assumptions of weak exogeneity are made, such as that the three processes are independent so they can be estimated separately. Manganelli (2005) studies the causal and feedback effects among the three variables and found that times of greater activity coincided with a larger fraction of informed traders being present in the market. However, his empirical results suggest that lagged volatility increases trading intensity, which is in contrast to Easley and O’Hara (1992), but confirms the inventory based model predictions that large returns attract opposite side traders and increase trading intensity.

Grammig and Wellner (2002) extend Engle (2000)’s model in another way. One of the key assumptions of Engle (2000)’s recursive model is that the duration and

volatility processes are independent so they can be estimated separately. Grammig and Wellner (2002) notice that duration and volatility might be interdependent. They formulate an interdependent intraday duration and volatility model. In this model, conditional volatility and intraday duration evolve simultaneously. The conditional volatility is formulated as a GARCH process, with time-varying parameters that are functions of the expected intraday duration. Their empirical results show that lagged volatility significantly reduces transaction intensity, which is consistent with Easley and O'Hara (1992).

The interdependence of these trading processes are also addressed by Hautsch (2008). He analyses the return volatility, trade size and trading duration under the Multivariate Error Model (MEM) framework. Rather than using transaction data, Hautsch (2008) uses the cumulated five-minute data and focuses on the study of the underlying common factors that jointly drive the trading processes. He finds that the common factor captures most causal relations and cross-dependencies between the individual variables. The existence of common factors is an indicator of the interdependence of the three processes.

In addition to the ACD framework, the vector autoregressive (VAR) model is used in the study of high frequency data. For example, Bowe, Hyde et al. (2009) used a trivariate VAR model to analyse the interrelationship between trading volume, duration and price volatility, which is similar to Dufour and Engle (2000). But it is also similar to the recursive model and assumes that trade and price processes are cross-independent. Using the data from an emerging futures market, they find that duration is affected positively by volatility, which is consistent with Diamond and Verrecchia (1987).

To summarize the empirical studies, the recursive frameworks are generally adopted for the analysis of high frequency data, but this is challenged by some empirical evidence. The empirical results with respect to the relations of trade and price process as are partially contradictory and there is no uniform conclusion at present.

4.3 Methodology

In this section, we first specify the dynamics of duration volume and price volatility according to the Engle (2000) and Manganelli (2005) recursive framework and discuss the statistic and economic concerns with this framework. We then extend the recursive framework of Engle (2000) and Manganelli (2005) to a vector specification in which trading duration, volume and price volatility evolve simultaneously and are interdependent.

4.3.1 Duration, Volume and Price Volatility --- a Recursive Framework

Define $\{d_t, v_t, r_t\}, t=1, \dots, T$ as the three-dimensional time series associated with intraday trading duration, trading volume and the return process, respectively. In particular, duration is defined as the time elapsing between consecutive trades, volume is the trade size associated with each transaction and return is measured as the mid-quote change. The trivariate trading process - duration, volume and return volatility - can be modelled as follows:

$$\{d_t, v_t, r_t\} \sim f(d_t, v_t, r_t | \mathcal{F}_{t-1}; \theta) \quad (4.1)$$

where \mathcal{F}_{t-1} denotes the information available up to period $t-1$, and θ is a vector incorporating the parameters of interest.

In the recursive model (Manganelli 2005), the joint distribution is decomposed into the product of three components: marginal density of durations, the conditional density of volumes given durations and the conditional density of the return volatility given durations and volumes. Specially,

$$\{d_t, v_t, r_t\} \sim g(d_t | \mathcal{F}_{t-1}; \theta_d) \cdot h(v_t | d_t, \mathcal{F}_{t-1}; \theta_v) \cdot k(r_t | d_t, v_t, \mathcal{F}_{t-1}; \theta_r). \quad (4.2)$$

For the dynamics of such a nonnegative valued financial point process, Engle and Russell (1998) first propose an ACD specification for financial duration. They model duration as the product of its conditional expectation and the non-negative supported innovation term,

$$d_t = \psi_t(\theta_d; \mathcal{F}_{t-1}) u_t, \quad u_t \sim i.i.d.(1, \sigma_u^2). \quad (4.3)$$

The ACD model is further characterized by the assumptions that the conditional duration ψ_t follows a GARCH-type process and the innovations are independently and identically distributed. The base (1,1) specification of ψ_t is:

$$\psi_t = \omega + \alpha d_{t-1} + \beta \psi_{t-1}. \quad (4.4)$$

The logarithmic version is also specified (Bauwens and Giot 2000) to ensure positivity of the conditional duration,

$$\log \psi_t = \omega + \alpha \log d_{t-1} + \beta \log \psi_{t-1}. \quad (4.5)$$

To close the model, the parametric density function for the innovations is needed. Engle and Russell (1998) initially consider the exponential and Weibull distribution, which is extended later by Grammig and Maurer (2000), Allen, Lazarov et al. (2009) and Xu (2011), offering more flexible density and hazard functions.

Following the ACD model, Manganelli (2005) considers similarly specifications for volume and volatility. Then the trivariate system has the following specifications:

$$\begin{aligned}
d_t &= \psi_t(\theta_d; \mathcal{F}_{t-1})u_t, \quad u_t \sim i.i.d.(1, \sigma_u^2) \\
v_t &= \phi_t(\theta_v; d_t, \mathcal{F}_{t-1})\eta_t, \quad \eta_t \sim i.i.d.(1, \sigma_\eta^2) \\
\hat{r}_t &= \sqrt{h_t(\theta_r; d_t, v_t, \mathcal{F}_{t-1})}\xi_t, \quad \xi_t \sim i.i.d.(0, 1) \\
\text{or } \hat{r}_t^2 &= h_t(\theta_r; d_t, v_t, \mathcal{F}_{t-1})\xi_t, \quad \xi_t \sim i.i.d.(1, \sigma_\xi^2)
\end{aligned} \tag{4.6}$$

where \hat{r}_t^2 is the proxy for volatility¹¹, (ψ_t, ϕ_t, h_t) are the conditional expectations of duration, volume and volatility, respectively, and $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ is a vector of s parameters of interest. Manganelli (2005) considers the univariate exponential distribution for the innovations in this specification.

To capture the causal and feedback effect among these variables, he specifies the following first order autoregressive conditional model:

$$\begin{aligned}
\psi_t &= w_1 + (a_{11}d_{t-1} + a_{12}v_{t-1} + a_{13}\hat{r}_{t-1}^2) + (b_{11}\psi_{t-1} + b_{12}\phi_{t-1} + b_{13}h_{t-1}), \\
\phi_t &= w_2 + (a_{21}d_{t-1} + a_{22}v_{t-1} + a_{23}\hat{r}_{t-1}^2) + (b_{21}\psi_{t-1} + b_{22}\phi_{t-1} + b_{23}h_{t-1}) + a_0^{12}d_t, \\
h_t &= w_3 + (a_{31}d_{t-1} + a_{32}v_{t-1} + a_{33}\hat{r}_{t-1}^2) + (b_{31}\psi_{t-1} + b_{32}\phi_{t-1} + b_{33}h_{t-1}) + a_0^{13}d_t + a_0^{23}v_t.
\end{aligned} \tag{4.7}$$

Under the restrictions of weak exogeneity¹² ($b_{ij} = 0$ for $i = j$) and independence of the innovations terms, the three components are estimated separately. This approach is generally adopted in the existing empirical literature (see, for example, Engle (2000), Dufour and Engle (2000), Manganelli (2005) and Engle and Sun (2007)).

4.3.2 Econometric Concerns

Following Manganelli (2005), there are two concerns regarding the recursive model. First, it assumes that the specific processes are independent. To incorporate the contemporaneous information, Manganelli (2005) specifies causality from duration to volume and from duration and volume to price volatility. However, modelling the

¹¹ In order to obtain a price change sequence which is free of the bid-ask bounce that affects price, we follow Ghysels, et al. (1998) and \hat{r}_t is obtained as the residuals of an ARMA(1,1) process of return series. See also in Hautsch (2008). One advantage of using \hat{r}_t is that it avoids the problem of exact zero values in r_t .

¹² This corresponds the second case weak exogeneity proposed in Chapter 2.

distribution of price as being conditional on duration and volume is just one strategy to obtain their joint distribution. As pointed out by Engle and Sun (2007), it is also possible to go from the price process and model duration conditional on its contemporaneous return. Theoretically, variation in duration and variation in the price process would be related to the same news events or the underlying information process. Empirical studies also address this issue. For example, Hautsch (2008) finds the existence of a common unobserved component that jointly drives the dynamics of the trade and price processes. This common component explains most of the causality between the trade and the price processes, even if the contemporaneous effect of the trade variable on the price variable is controlled. We test first case of weak exogeneity in chapter 2 and the empirical result show the existence of cross-restriction between the duration and price process. Therefore, the advisable approach is to allow the innovation terms to be contemporaneous correlated, and specify a vector form for the dynamics of the trivariate system.

Second, Manganelli (2005) assumes weak exogeneity¹³, which means the conditional expectation of one variable is a function only of its own past conditional expectation, while the past conditional expectations of other variables are not taken into consideration. This strategy has been adopted by most empirical microstructure papers (see, for example, Dufour and Engle (2000)). However, we argue that this assumption is too restrictive. When studying the price impact of trade, various specifications of duration and volume should be considered. For example, trade innovation is an exclusive a manifestation of the private information of the informed trader. Engle (2000) and Wuensche, Grammig et al. (2007) argue that it is the unexpected components of the trade process that carry informational content with

¹³ This corresponds the second case of weak exogeneity discussed in Chapter 2.

respect to the fundamental asset price, since price change is unpredictable. And the same happens for the feedback effects from price to trading intensity. For example, Grammig and Wellner (2002) find that expected volatility and volatility shocks have significant effects on trading intensity. Manganelli (2005) conducts a robustness test on this restriction. Specifically, he regresses the residuals of the three equations against past conditional expectations of other variables. The results indicate that the coefficients of past expected variables are almost never significant, and thus the recursive model is correctly specified. However, the robustness check might be misleading, since the dynamics of expected variables have been distorted when estimating and predicting the expected variables using recursive models. It is also shown by Grammig and Maurer (2000) in a simulation study that the misspecification of the conditional mean has severe consequences for the expectation of conditional duration.

We therefore extend the recursive model into a vector form, by allowing the three processes to be interdependent and relaxing weak exogeneity.

4.3.3 Vector MEM

Let $x_t = (d_t, v_t, r_t)'$, $\mu_t = (\psi_t, \phi_t, h_t)'$ and $\varepsilon_t = (u_t, \eta_t, \xi_t)'$. Following Engle (2002) and Cipollini, Engle et al. (2007) we write this system of equations as a trivariate vector multiplicative error model (MEM). The three-dimensional vector MEM for duration, volume and volatility is:

$$x_t = \mu_t \Theta \varepsilon_t = \text{diag}(\mu_t) \varepsilon_t \quad (4.8)$$

where Θ denotes the Hadmard (element by element) product and $\text{diag}(\cdot)$ denotes a diagonal matrix with the vector in the augment as main diagonal. The innovation vector is a 3 dimensional random variable defined over a $[0, +\infty)^3$ support. ε_t has a

mean vector I with all components unity and general variance-covariance matrix Σ , i.e., $\varepsilon_t | \mathcal{F}_{t-1} \sim D(I, \Sigma)$. μ_t is defined as before, except that now we are dealing with 3-dimensional vector. The multivariate specification for μ_t is:

$$\mu_t = \omega + \sum_{l=1}^p A_l x_{t-l} + \sum_{l=1}^q B_l \mu_{t-l} + A_0 z_t \quad (4.9)$$

where z_t is a vector of predetermined variables

The first two moment conditions of the vector MEM are given by:

$$\begin{aligned} E(x_t | \mathcal{F}_{t-1}) &= \mu_t \\ \text{Var}(x_t | \mathcal{F}_{t-1}) &= \mu_t \mu_t' \Theta \Sigma = \text{diag}(\mu_t) \Sigma \text{diag}(\mu_t) \end{aligned} \quad (4.10)$$

which is a positive defined matrix by construction, as emphasized by Engle (2002).

We do not specify recursively the contemporaneous relationship from duration to volume and from duration and volume to volatility (Manganelli 2005). However, we allow the innovation terms to be contemporaneously correlated. By this specification, the conditional expectation of one variable is a function not only of its own past conditional expectation, but also of past conditional expectations of other variables. The two restrictions imposed by the recursive model are released.

The mean equation is further extended to be a logarithmic version to ensure the positivity of the individual processes without imposing additional parameter restrictions.

$$\ln(\mu_t) = \omega + \sum_{l=1}^p A_l \ln(x_{t-l}) + \sum_{l=1}^q B_l \ln(\mu_{t-l}) + A_0 \ln(z_t). \quad (4.11)$$

4.3.4 Specification of ε_t

A completely parametric formulation of the vector MEM requires a full specification of the conditional distribution of ε_t . In the ACD literature, Engle and

Russell (1998) initially consider the exponential and Weibull distribution for the error ε_t , which is extended later by Grammig and Maurer (2000) to be a Burr distribution, by Lunde (1999) to be a generalized gamma distribution, and recently by Allen, Lazarov et al. (2009) and also in our third Chapter to be a lognormal distribution. Figure 4-1 plots the comparison of density functions implied by these parametric distributions. It can be seen that only the exponential distribution implies a monotonically decreasing density function, while the others imply hump shaped density functions. In our third Chapter, we also tests the specification of the duration distributions, and finds that the lognormal ACD model is superior to the Exponential ACD and Weibull ACD models, while its performance is similar to the Burr or Generalized Gamma ACD models. It is well known that price volatility is lognormally distributed, while Andersen, Bollerslev et al. (2001) and Cizeau, Liu et al. (1997), among others, also showed that the lognormal distribution fitted the realized volatility distribution very well.

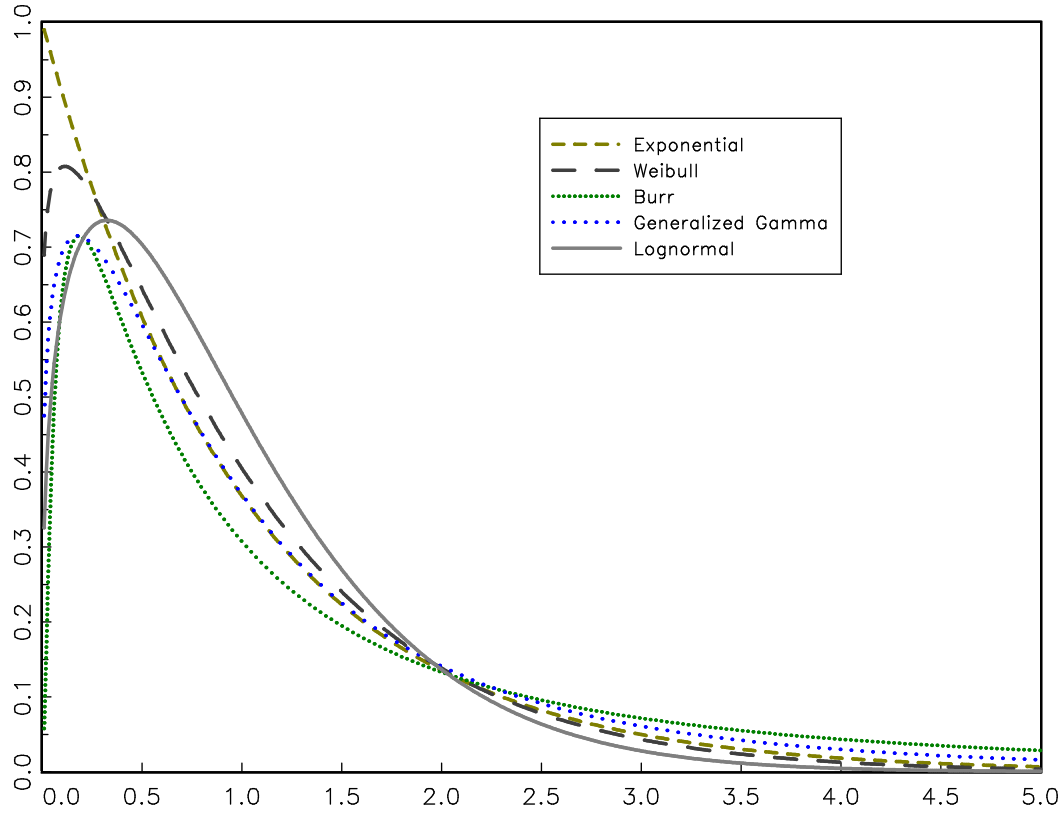


Figure 4-1: A comparison of parametric densities

So we propose to use the multivariate lognormal distribution for the MEM. Indeed, the multivariate lognormal distribution seems to be the only feasible choice in the specification of vector MEM. It has a closed form conditional density function, so that ML estimation can be conducted. Cipollini, Engle et al. (2007) consider appropriate multivariate gamma versions but find that the only useful version admits only positive correlation, which is too restrictive. The multivariate lognormal distribution admits both positive and negative correlations. Moreover, Allen, Chan et al. (2008) prove that the lognormal distribution is sufficiently flexible to provide a good approximation to a wide range of non-negative distributions, and is also sufficiently accurate so as not to induce unnecessary numerical difficulties.

Assume ε_t follows a multivariate lognormal distribution such that $\varepsilon_t | \Omega_t \sim \ln N(\nu, D)$ ¹⁴. The density function is¹⁵:

$$f(\varepsilon_t | \mathcal{F}_{t-1}, D) = (2\pi)^{-K/2} |D|^{-1/2} \left(\prod_{i=1}^K \varepsilon_{i,t}^{-1} \exp \left(-\frac{1}{2} (\ln \varepsilon_t - \nu)' D^{-1} (\ln \varepsilon_t - \nu) \right) \right) \quad (4.12)$$

where $\varepsilon_t > 0$. The conditional density of x_t is then:

$$f(x_t | \mathcal{F}_{t-1}, \theta) = (2\pi)^{-K/2} |D|^{-1/2} \prod_{i=1}^K x_{i,t}^{-1} \exp \left(-\frac{1}{2} (\ln x_t - \ln \mu_t - \nu)' D^{-1} (\ln x_t - \ln \mu_t - \nu) \right). \quad (4.13)$$

The log likelihood of the model is then:

$$l = \sum_{t=1}^T l_t = \sum_{t=1}^T \ln f(x_t | \mathcal{F}_{t-1}, \theta) \quad (4.14)$$

where

$$\begin{aligned} l_t = \ln f(x_t | \mathcal{F}_{t-1}, \theta) = & -\frac{K}{2} \ln(2\pi) - \frac{1}{2} |D| - \sum_{i=1}^K \ln(x_{i,t}) \\ & - \frac{1}{2} (\ln x_t - \ln \mu_t - \nu)' D^{-1} (\ln x_t - \ln \mu_t - \nu) \end{aligned} \quad (4.15)$$

The first and second moments of the multivariate lognormal distribution are given by:

$$\begin{aligned} \tau = E(\varepsilon) = (\tau_1, \tau_2, \dots, \tau_k)' \quad \tau_i &= e^{v_i + \frac{1}{2} d_{ii}} = 1 \\ \Sigma = E(\varepsilon - \tau)(\varepsilon - \tau)' = \sigma_{ij} \quad \sigma_{ij} &= e^{(v_i + v_j + \frac{d_{ii} + d_{jj}}{2})} (e^{d_{ij}} - 1) = e^{d_{ij}} - 1 \\ \rho_{ij} &= \frac{e^{d_{ij}} - 1}{\sqrt{(e^{d_{ii}} - 1)(e^{d_{jj}} - 1)}} \end{aligned}$$

where $v_i = -\frac{1}{2} d_{ii}$ and d_{ij} is the ij th element of D . It is clear that if

$(\ln \varepsilon_1, \ln \varepsilon_2, \dots, \ln \varepsilon_k)$ are independent, then $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)$ are also independent and vice versa. The multivariate lognormal distribution allows both positive and negative

¹⁴ where $v_i = -\frac{1}{2} d_{ii}$ to guarantee that $E(\varepsilon_t | \mathcal{F}_{t-1}) = I$

¹⁵ See Appendix 7 for the derivation of density function for multivariate lognormal distribution.

correlation, which is much more flexible than the multivariate gamma distribution (Cipollini, Engle et al. 2007).

The lognormal belongs to the exponential family. The parameters are still consistently estimated, even if the chosen density is wrong. However, the asymptotic distribution of the QML estimator differs from that of the ML estimator. The variance-covariance matrix is not the inverse of the Fisher information. It has the so-called ‘sandwich’ form.

$$\sqrt{N}(\hat{\theta}_{QML} - \theta) \rightarrow N(0, I^{-1}(\hat{\theta})J(\hat{\theta})I^{-1}(\hat{\theta})) \quad (4.16)$$

where $I(\hat{\theta}) = -E \left[\frac{\partial^2 \ln L(\hat{\theta}; x)}{\partial \hat{\theta} \partial \hat{\theta}'} \right]$, $J(\hat{\theta}) = E \left[\frac{\partial \ln L(\hat{\theta}; x)}{\partial \hat{\theta}} \left(\frac{\partial \ln L(\hat{\theta}; x)}{\partial \hat{\theta}} \right)' \right]$ are,

respectively, the components of the empirical average Hessian and the empirical average outer product of the gradients evaluated at the estimates $\hat{\theta}$.

4.3.5 Impulse Response Function

Following the vector MEM, we can derive the impulse response functions. We concentrate on the first order model and exclude the predetermined variables.

$$\begin{aligned} x_t &= \mu_t \Theta \varepsilon_t, \\ \ln \mu_t &= \omega + A \ln x_{t-1} + B \ln \mu_{t-1}. \end{aligned} \quad (4.17)$$

In the impulse response, we work on the impulse of $\nu_0 = \ln \varepsilon_0$ on the natural logarithmic of the interested variable $\ln x_t$. The impulse responses function of the model (4.17) for $t > 0$ is¹⁶:

$$\frac{\partial \ln x_t}{\partial \nu_0} = \Phi_t \quad (4.18)$$

¹⁶ See Appendix 8 for proof.

where $\Phi_t = (A + B)^{t-1}(A - B)$, $\Phi_0 = I$.

This process can be rewritten in such a way that the residuals of different equations are uncorrelated. For this purpose, we choose a decomposition of the white noise covariance matrix $\Sigma_v = W\Sigma_\tau W'$, where Σ_τ is a diagonal matrix with positive diagonal elements and W is a lower triangular matrix with unit diagonal. Thus,

$$\ln x_t = \bar{\omega} + \sum_{i=0}^{\infty} \Psi_i \tau_{t-i}, \quad \Psi_i = \Phi_i W^{-1}. \quad (4.19)$$

Then the impulse response function of the model (4.17) for $t > 0$ is:

$$\frac{\partial \ln x_t}{\partial \tau_0} = \Psi_t \quad (4.20)$$

The standard errors for the impulse response are computed as followings. Let

$$\theta_{(p \times 1)} = [\theta_d', \theta_v', \theta_r']' \quad \text{and} \quad \kappa_t \equiv \text{vec}(\Phi_t(\theta)) \quad . \quad \text{If} \quad \sqrt{T}(\hat{\theta} - \theta) \xrightarrow{a} N(0, Q) \quad , \quad \text{then}$$

$$\sqrt{T}(\hat{\kappa}_t - \kappa_t) \xrightarrow{a} N(0, G_t Q G_t') \quad , \quad \text{where} \quad G_t = \frac{\partial \kappa_t}{\partial \theta'} .$$

4.3.6 Vector ARMA Representation

One of the advantages of using the lognormal distribution for the vector MEM model is that it has an equivalent Vector ARMA specification with an innovation that follows a multivariate Gaussian distribution.

From the following log vector MEM model,

$$x_t = \mu_t \Theta \varepsilon_t, \quad (4.21)$$

$$\ln(\mu_t) = \omega + \sum_{l=1}^p A_l \ln(x_{t-l}) + \sum_{l=1}^q B_l \ln(\mu_{t-l}) + A_0 \ln(z_t). \quad (4.22)$$

If we take logs of equation (4.21), we obtain

$$\ln(x_t) = \ln(\mu_t) + \ln(\varepsilon_t) = c + \ln(\mu_t) + e_t \quad (4.23)$$

where $e_t | \mathcal{F}_{t-1} \sim iid N(0, \Pi)$.

Then,

$$\ln(\mu_t) = \ln(x_t) - c - e_t, \quad (4.24)$$

$$\sum_{l=1}^q B_l \ln(\mu_{t-l}) = \sum_{l=1}^q B_l \ln(x_{t-l}) + \sum_{l=1}^q B_l c - \sum_{l=1}^q B_l e_{t-l}. \quad (4.25)$$

Substituting $\ln(\mu_t)$ and $\sum_{l=1}^q B_l \ln(\mu_{t-l})$ into equation (4.25), it follows that

$$\ln(x_t) = \bar{c} + \left(\sum_{l=1}^p A_l + \sum_{l=1}^q B_l \right) \ln(x_{t-l}) + e_t - \sum_{l=1}^q B_l e_{t-l} + A_0 \ln(z_t) \quad (4.26)$$

where $\bar{c} = c + \omega - \sum_{l=1}^q B_l c$.

It is interesting that the vector MEM model is equivalent to a VARMA specification. In particular, it provides a good way to adopt the VARMA inference to make inferences in the vector MEM model¹⁷.

4.4 Empirical Analysis

4.4.1 Data

We use the data from the Trades and Quotes (TAQ) dataset at NYSE. The TAQ data consists of two parts: the first reports the trade data, while the second lists the quote data (bid and ask data) posted by the market maker. The data were kindly provided by Manganelli (2005). He constructed 10 deciles of stocks covering the period from Jan 1, 1998 to June 30, 1999, on the basis of the 1997 total number of trades of all stocks quoted on the NYSE. We randomly selected 5 stocks from the eighth decile (frequently traded stocks) and 5 from the second decile (infrequently

¹⁷ See Lütkepohl (2005) for the inference of VARMA model

traded stocks) covering the period from Jan 1,1998 to June 30, 1999. The tickers and names of the ten stocks are reported in Table 4-2.

Table 4-2: Stocks used in the analysis

A. Frequently traded		B. Infrequently traded	
TRN	Trinity Industries	ABG	Group ABG
R	Ryder System Inc.	OFG	Oriental Finl Grp Hold Co.
ARG	Airgas Inc.	LSB	LSB Industries Inc.
GAS	Nicor Incorporated	FEP	Franklin Electronic Publisher
TCB	TCF Financial Corp.	HTD	Huntingdon Life S.G.

We adopt the same strategy used in Chapter 2 to prepare the data and adjust the time of day effect. Please see Chapter 2 of the detailed specification. Some summary statistics for the cleaned data are reported in Table 4-3. For the frequently traded stocks, the number of observations range from 33,850 to 63,862 in the sample period, and the average trading duration ranges from 137 seconds to 259 seconds. For the infrequently traded stocks, the number of observation ranges from 2,074 to 7,212 in the sample period, with the average trading duration ranging from 1,215 seconds to 4,215 seconds. The trading volume does not show any difference between frequently traded stocks and infrequently traded stocks. The number of trading volumes ranges from 833 to 5,295. The multivariate Ljung–Box statistics, computed according to Hosking (1980) and is given by

$$MLB(s) := n(n+2) \sum_{j=1}^s \frac{1}{n-j} \text{trace}(\hat{C}_j' \hat{C}_0^{-1} \hat{C}_j' \hat{C}_0^{-1}) \sim \chi^2(k^2 s) \quad (4.27)$$

where k denotes the dimension of the process (in this case $k=3$), s is the number of lags taken into account, and \hat{C}_j is the j th residual autocovariance matrix. It is apparent that duration, volume and volatility show strong serial autocorrelations, and this is particularly true for high frequency traded stocks. The large multivariate Ljung-

Box statistics in the table indicate that the trivariate system reveals strong dynamic and contemporaneous dependencies. These indicators suggest the use of vector form MEM.

Table 4-3: Summary statistics for the 10 stocks

	Obs	Mean		LB(20)			MLB(20)
		Duration	Volume	Duration	Volume	Variance	
TRN	55582	157.86	1369.43	3780.09	1383.35	3769.80	12744.02
GAS	41999	212.93	827.77	5951.85	2338.08	4073.09	19049.05
TCB	55208	158.94	1855.20	4171.36	2644.11	2925.82	14716.45
R	63862	137.41	1800.74	14072.3	7276.91	23685.7	58049.96
ARG	33850	259.2	1280.70	3780.09	1383.35	3769.80	12744.02
ABG	2074	4214.88	5259.05	120.28	225.07	146.00	760.08
OFG	7212	1214.58	833.86	523.16	1343.43	738.09	3557.98
LSB	2962	2962.19	1971.61	481.41	435.69	523.58	2110.88
HTD	2505	3422.28	3943.59	268.52	682.92	297.01	1571.99
FEP	4405	1989.58	1565.89	2431.00	660.60	788.81	4564.73

Notes: LB(20) denotes Ljung–Box statistics for order 20. MLB(20) denotes multivariate Ljung–Box statistics. The mean statistics report the average valued for the raw data. LB and MLB statistics report serial correlation for the data after adjusting the time of day effect. Critical values for LB statistics $\chi^2(20)_{0.05}=31.41$, $\chi^2(20)_{0.01}=37.57$. $\chi^2(180)_{0.05}=212.30$, $\chi^2(180)_{0.01}=227.06$.

We also depict the non-parametric density and parametric densities implied by the exponential and lognormal distributions.¹⁸ Figure 4-2 reports the comparison of parametric and non-parametric densities for one typical stock LSB. It can be seen that the lognormal distribution fits with the true density very well for the duration data. This result is consistent with Xu (2011). For volume data, we are surprised to find the lognormal distribution has the best performance. And the raw data fluctuates closely around the lognormal distribution. Even for volatility, the lognormal distribution also

¹⁸ See Xu (2011a) and Grammig and Maurer (2000) for the discussion of parametric and non-parametric density.

performs well. For brevity, the other 9 stocks have are not been reported for brevity, but these findings are robust across the stocks.

The data we use in this chapter strongly support the multivariate lognormal MEM model for the dynamics of duration, volume and price volatility.

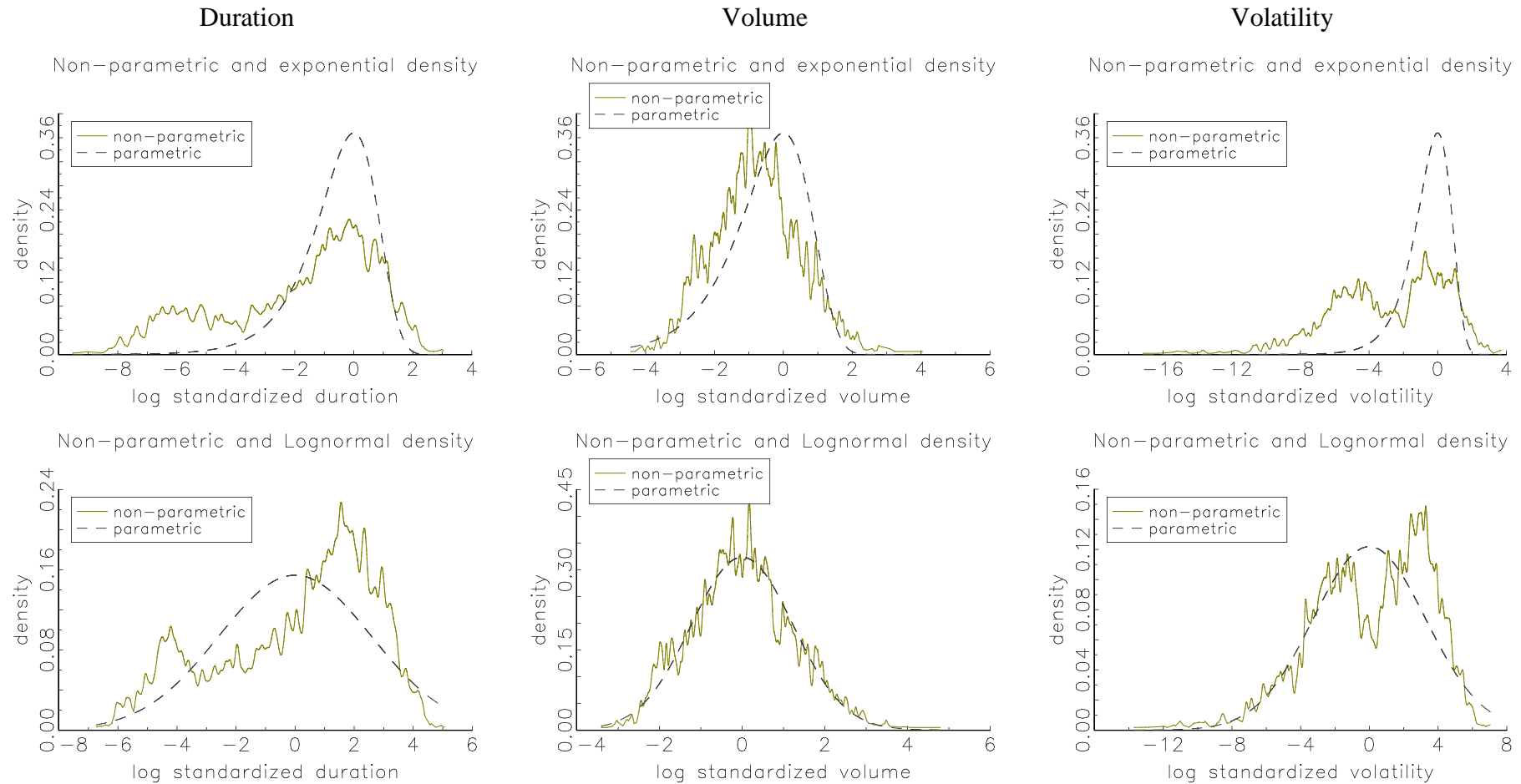


Figure 4-2: A comparison of parametric density and non-parametric densities--LSB

4.4.2 Empirical Model

In the empirical analysis, we are interested in the causal and feedback effects among the variables. In contrast to the previous recursive model, we allow trade duration, volume and innovations of these variables to affect price volatility and vice versa: the volatility and volatility shocks are allowed to affect trading intensity. So we specify and estimate the following vector MEM:

$$x_t = \mu_t \Theta \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim D(I, \Sigma) \quad (4.28)$$

$$\ln \mu_t = \omega + A \ln x_{t-1} + B \ln \mu_{t-1} + C \ln \frac{x_{t-1}}{\mu_{t-1}}$$

where B is a diagonal matrix and C is a matrix where the diagonal elements are zero. Then, a_{31} (a_{32}) measures the impact of duration (volume) on price volatility, c_{31} (c_{32}) measures the impact of duration (volume) shocks on price volatility, a_{13} measures the impact of volatility on trading intensity and c_{13} measures the impact of volatility shocks on trading intensity. The estimation results and various diagnostics for the five frequently traded stocks are reported in Table 4-4 and results for the five infrequently traded stocks are reported in Table 4-5.

The first purpose of empirical analysis is to examine the performance of the vector MEM. Considering the diagnostic statistics of the model, these suggest that the vector MEM improves the dynamic properties of the model significantly, as we can see from the sharp drop in the Ljung-Box statistics. This is particularly true for the volatility process. Moreover, the vector MEM reduces the multivariate Ljung-Box statistic significantly, indicating that the vector MEM does a good job in capturing the multivariate dynamics and interdependencies between the individual processes. For frequently traded stocks, the dynamics of the system are still not captured completely

by the model. But this is commonly the case with such large time series (see, for example, Engle (2000)). For infrequently traded stocks, the dynamics of the system are captured completely by the vector MEM.

In Manganelli(2005)'s recursive model, the assumption of weak exogeneity is made in the specification of the conditional mean. The past expected variables are assumed not to carry any information ($c_{ij} = 0$). Manganelli (2005) and Dufour and Engle (2000) also conduct robustness tests of this restriction, in which the residuals of the three models are regressed against lagged expected variables. They find that the lagged expected variables are insignificant. However, we find that the most lagged expected variables are significant (c_{ij}) in our vector MEMs. It is particularly true for infrequently traded stocks. The LR tests also suggest that the lagged expected variables are jointly significant in almost all cases. We argue that the robustness checks conducted by Manganelli (2005) and Dufour and Engle (2000) are misleading, since the dynamics of expected variables has been distorted by the marginal model. Therefore, the weak exogeneity assumption is not supported by the empirical data. The lagged expected variables should be incorporated in this trivariate system.

Table 4-4: Estimation results and diagnostics: frequently traded stocks.

	ARG	TRN	TCB	GAS	R
α_{11}	0.060***	0.089	0.055***	0.062	0.064
α_{12}	0.107**	0.228***	0.122***	0.216***	0.168
α_{13}	0.012***	0.007***	0.025***	0.009	0.018
α_{21}	-0.067**	0.113**	-0.003**	0.121	0.018
α_{22}	0.125	0.124	0.098	0.118***	0.125
α_{23}	-0.009	-0.004	-0.009***	-0.007***	-0.011**
α_{31}	-0.337***	-0.204***	-0.387***	-0.065	0.071***
α_{32}	-0.109	0.219***	0.371***	-0.019	-0.434
α_{33}	0.389***	0.241***	0.278***	0.316***	0.195
b_{11}	0.939***	0.730***	0.942***	0.724***	0.912***
b_{22}	0.508***	0.638***	0.695***	0.706***	0.606**
b_{33}	0.239***	0.301***	0.075***	0.246***	0.629***
c_{12}	-0.171***	-0.331***	-0.202***	-0.338***	-0.265
c_{13}	-0.023***	-0.017***	-0.035***	-0.014***	-0.032
c_{21}	0.064	-0.127**	-0.013***	-0.132	-0.031
c_{23}	0.004	0.005**	0.008***	0.007***	0.009***
c_{31}	0.015	-0.084***	0.015***	-0.170	-0.414***
c_{32}	0.629***	0.353***	0.260***	0.474***	0.882**
LR test ¹⁹					
$H_0: c_{ij} = 0, i \neq j$	240	519	345	408	1587
Diagnostics					
LL	-221998	-358758	-365788	-260314	-407773
BIC	444247	717780	731838	520883	815812
MLB	565.8***	991.6**	1018***	684.3***	1086
LB_d	101.4***	36.23**	104.0***	52.82***	60.37***
LB_v	95.97***	184.1***	182.3***	83.37***	355.9***
LB_r^2	174.3***	308.7***	457.0***	219.9***	70.30***

Note: *** denotes significance at 1% level. ** denotes significance at 5% level. LL denotes Log likelihood function. BIC denotes Bayes Information Criterion. LB denotes Ljung-Box statistics of flitted residuals and MLB denotes multivariate Ljung-Box statistic. The Ljung-Box statistics are computed based on 20 lags. Critical values of LR statistics $\chi^2(6)_{0.05}=12.59$, $\chi^2(6)_{0.01}=16.81$

¹⁹ We estimate five different vector MEMs for comparison. The results have not reported for brevity. LR test is based on the likelihood values of restricted and unrestricted models.

Table 4-5: Estimation results and diagnostics: infrequently traded stocks.

	ABG	HTD	LSB	HUN	FEP
α_{11}	0.042***	0.019**	0.032	0.056***	0.056***
α_{12}	-0.079***	0.015***	0.049	-0.002	0.016
α_{13}	-0.021***	0.018***	-0.004	0.013**	0.085
α_{21}	0.019***	-0.564***	-0.083	-0.005	-0.051
α_{22}	0.231***	0.198	0.166	0.133**	0.185***
α_{23}	-0.079***	-0.090**	-0.009	-0.025***	-0.155
α_{31}	-0.023***	-0.446	-0.145***	0.032	-0.062
α_{32}	-0.599***	-0.408***	-0.140	-0.374	-0.134
α_{33}	-0.079***	-0.090**	-0.009	-0.025***	-0.155
b_{11}	0.912***	0.980***	0.967***	0.932***	0.910***
b_{22}	0.366***	0.290***	0.569	0.708***	0.643***
b_{33}	0.682***	0.665***	0.522***	0.67***	0.318**
c_{12}	0.108***	-0.061***	-0.118	-0.013***	-0.103***
c_{13}	0.024***	-0.038***	0.006	-0.028**	-0.094
c_{21}	-0.034***	0.549***	0.087	0.000	0.042
c_{23}	0.079***	0.095**	0.013	0.028***	0.149***
c_{31}	-0.024***	0.370	0.028	-0.160***	-0.006***
c_{32}	0.919***	0.663***	0.389	0.721**	0.580**
LR test					
$H_0: c_{ij} = 0, i \neq j$	35.1	54.7	10.4	110	34.0
Diagnostics					
LL	-14392	-17116	-19370	-37243	-28574
BIC	28967	34420	38932	74695	57310
MLB	221.3**	191.3	249.3***	167.2	195.7
LB_d	36.47**	32.81**	48.37**	25.04	29.12
LB_v	22.17	17.11	37.52***	19.98	10.38
LB_r^2	27.78	21.57	35.27**	12.73	65.36***

Note: *** denotes significance at 1% level. ** denotes significance at 5% level. LL denotes Log likelihood function. BIC denotes Bayes Information Criterion. LB denotes Ljung-Box statistics of flitted residuals and MLB denotes multivariate Ljung-Box statistic. The Ljung-Box statistics are computed based on 20 lags. Critical values of LR statistics $\chi^2(6)_{0.05}=12.59$, $\chi^2(6)_{0.01}=16.81$

4.4.3 Empirical Results

The second purpose of empirical analysis is to examine the dynamics relationship of the trivariate system. Looking first at the price volatility (h_t) process. The coefficient of duration (a_{31}) and coefficient of duration shocks (c_{31}) in the volatility equation are negative and significant in most cases. This is consistent with Easley and O'Hara (1992), indicating that trades with short duration or the shocks of trading intensity are related to the arriving of new information, which reveals a higher volatility impact. The implicit application is that market makers will associate the higher trading activity or trading activity that is higher than its expected level as a signal of informed trading.

The volume coefficient (a_{32}) is only significant for 4 out of 10 stocks and the sign is unclear. However, the volume shocks coefficient (c_{32}) are all significant and positive. This implies that the unexpected component of volume rather than the raw volume carry information. Implicitly, market makers will only consider trade size that is larger than its expected level as a signal of private information, and adjust bid-ask price accordingly. The expected large trade size is simply for liquidity reason. The results partly support the prediction from Easley and O'Hara (1987,1992).

This exercise of the price impact of trade is novel in two aspects. First, most empirical market microstructure literature (see, for example, Dufour and Engle (2000) and Manganelli (2005)) uses raw duration (volume) to determine the presence of informed traders in the market. We highlight that it is the unexpected components of trade that carry information with respect to asset prices. Second, in contrast to Manganelli (2005), our findings are generally robust for less frequently stocks. There

is no reason why the informed traders should avoid taking advantage of their private information if it is related to infrequently traded stocks.

The strikingly different results, with respect to the feedback effects from the price process to trading intensity, are found in the duration equation. For the frequently traded stocks, the coefficient of volatility (a_{13}) is always positive but significant for 3 out of 5 stocks and the coefficients on volatility innovation (c_{13}) is always negative but significant for 4 out of 5 stocks. Following Hasbrouck (1988,1991), we explain this by considering the persistent quote change (volatility) to be information motivated and transient quote change (volatility shock) to be inventory motivated. Then our results are consistent with microstructure predictions. For example, information motivated large absolute quote changes indicate a risk of informed trading and the liquidity traders may leave or slow down the trading activity to avoid adverse selection (Admati and Pfleiderer 1988; Easley and O'Hara 1992), while inventory motivated large quote changes may attract opposite side traders and increase trading intensity. Similar results can be found for infrequently traded stocks, but the effects are less significant.

In the existing empirical microstructure literature, Dufour and Engle (2000) and Manganello (2005) find that short durations follow large returns, while Grammig and Wellner (2002) find that lagged volatility significantly reduces trade intensity. Our findings enhance the existing literature by incorporating both of these effects in one model.

4.4.4 Impulse Response Analysis

From the estimates of the MEM in equation (4.28), we generate the impulse responses which trace the effect of a one-time shock to one of the innovations on the future values of the endogenous variables. The impulse response function for two representative stocks TRN and ABG are plotted in Figure 4-3 and Figure 4-4. This gives the effects of a variation on the forecast up to the 10th trade. Since the impulse-response functions are plotted in transaction time, they are not directly comparable among different stocks. We use the Manganelli (2005) method to approximate the calendar time the system takes to return to its long-run equilibrium, by multiplying the number of transactions by their average duration. The average duration per trade of the two representative stocks is 158 seconds for TRN and 4215 seconds for ABG. This implies, for example, that a shock to the duration of TRN is absorbed by the expected duration after about 27 trades, or, on average, after 1.2 hours. In the case of ABG, the same shock is absorbed after 54 transactions, which corresponds, on average, to a period of 63.3 hours. Similar results hold for the other impulse-responses, indicating that the more traded the stock, the faster the market returns to its full information equilibrium after an initial perturbation. In particular, this is consistent with the (plausible) assumption that the more frequently traded the stock the higher the number of informed traders.

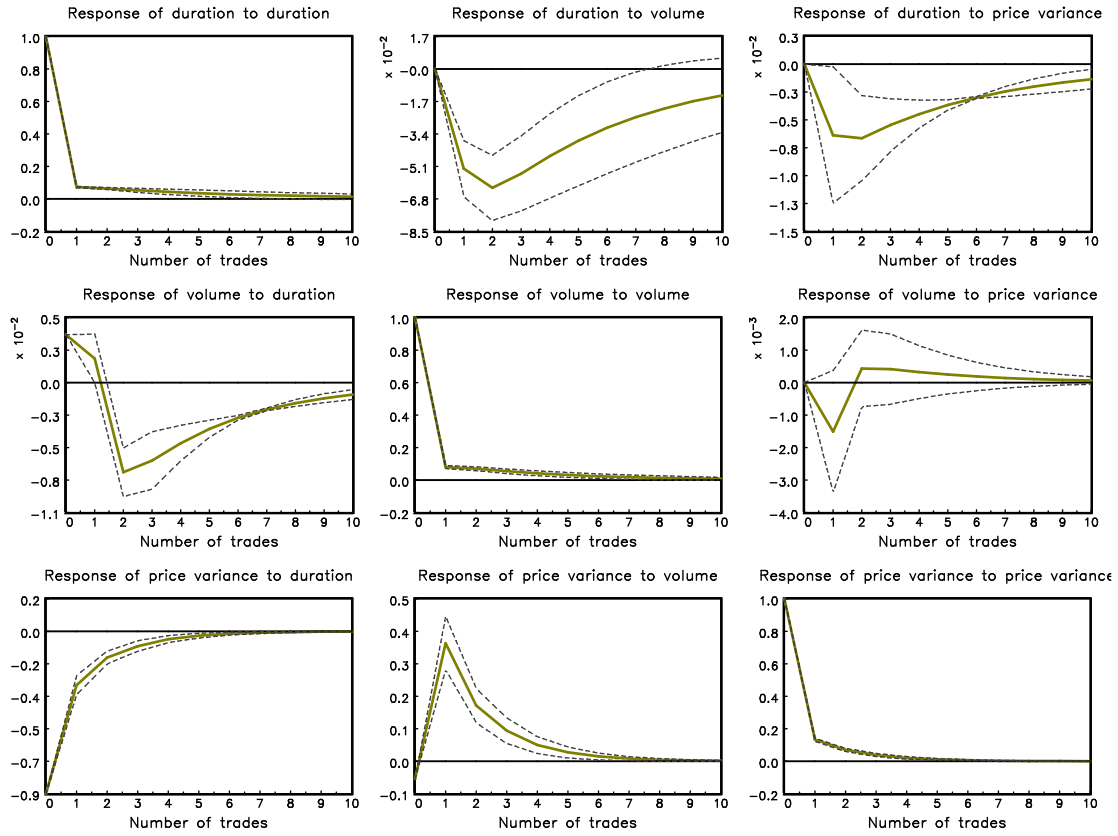


Figure 4-3: Impulse response function for TRN

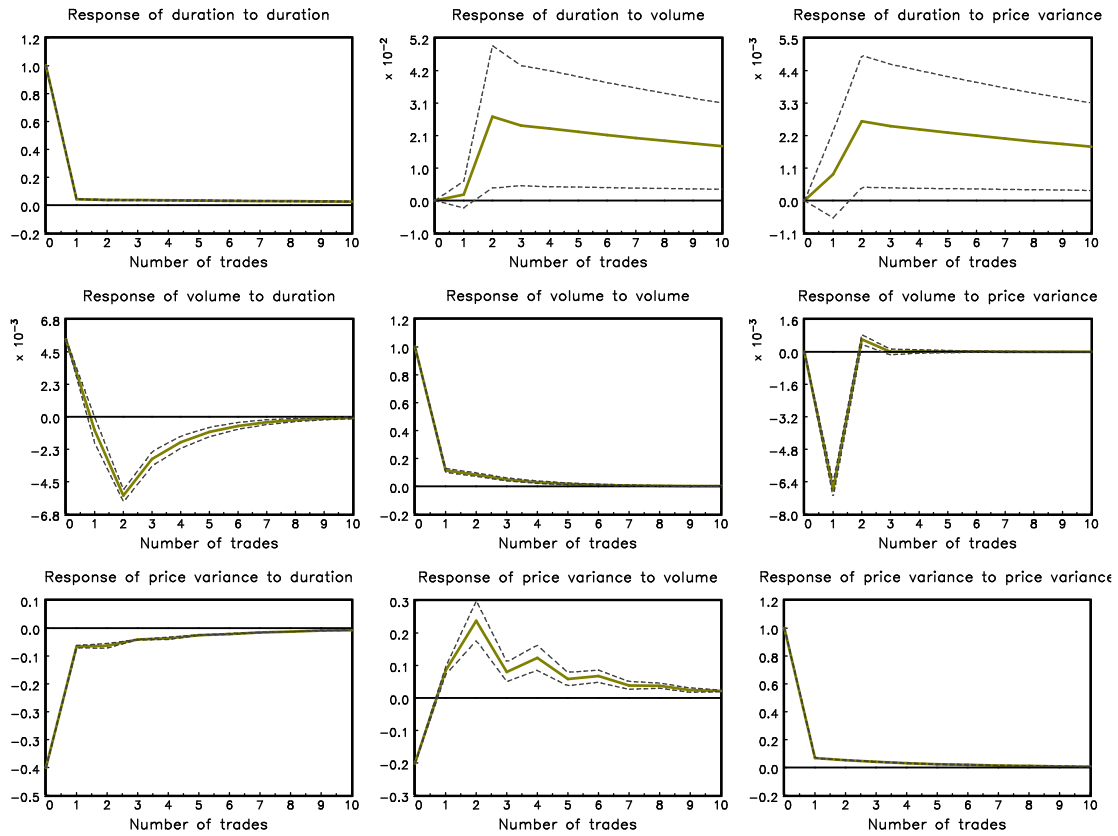


Figure 4-4: Impulse response function for ABG

Table 4-6 summarizes the results for the other stocks, confirming that the price volatility of frequently traded stocks converges much faster to its long-run equilibrium²⁰ after an initial perturbation. In general, for frequently traded stocks, the new information is implicitly incorporated in the price within one trading day, while it takes up to a week for the new information to be included into the price for infrequently traded stocks. Overall, the effect is to suggest that the market is reasonably efficient. This result, in contrast to Kyle (1985), confirms Admati and Pfleiderer (1988) and Holden and Subrahmanyam (1992)'s finding that information is short lived. For example, Holden and Subrahmanyam (1992) show that with multiple informed traders there will be more aggressive trading in the early periods, causing more information to be revealed earlier in the process.

Table 4-6: Time (in hours) it takes to absorb shocks to the long term equilibrium variances

	ARG	TRN	TCB	GAS	R
Shock to duration	2.5	1.2	0.8	1.7	3.1
Shock to volume	2.5	1.2	0.8	1.7	3.1
Shock to price volatility	2.4	1.1	0.7	1.7	3.0
	ABG	HTD	LSB	HUN	FEP
Shock to duration	63.3	59.0	37.8	29.7	7.7
Shock to volume	69.1	61.9	38.7	31.3	8.9
Shock to price volatility	63.3	59.0	38.7	29.3	7.2

²⁰ The threshold at which the shock producing the impulse–response is assumed to be absorbed is at $1e-7$ for shocks. That is, Table 7 reports the time it takes for the impulse–response of the variance to fall below $1e-7$.

4.5 Conclusion

In this chapter, we extend the recursive framework of Engle (2000) and Manganelli (2005) for the transaction data to a vector MEM, in which trading duration, volume and price volatility are interdependent. We further propose a multivariate lognormal for the distribution of the vector MEM, which allows the innovations terms to be contemporaneously correlated. In this way, we can build a system that incorporates various causal and feedback effects among these variables. The method is applied to the trade and quote dataset of the NYSE and the model is estimated using a sample of 10 stocks. The empirical findings are summarized as follows:

- (1) The diagnostic statistics show that the vector MEM improves the dynamic properties of the model significantly. Moreover, the lagged (un)expected variables are widely significant in the MEM model, challenging the weak exogeneity assumptions used in the empirical market microstructure literature.
- (2) We find a significant price impact of trade. However, we highlight the effect of unexpected components of trading characteristics. Both duration and duration shocks carry price information, while only unexpected volume carries most of the volume related information content.
- (3) We also find significant feedback effects, with volatility and volatility shocks affecting duration in different directions. This finding confirms Hasbrouck (1988,1991)'s prediction that persistent quote changes are driven by private information, which decreases trading intensity, while the transient quote changes are motivated by inventory control, which would attract opposite side traders and

increase trading intensity. However, this effect is only robust for frequently traded stocks.

- (4) With the impulse response, we find that the new information is implicitly incorporated in to the price within one trading day for frequently traded stocks, and it takes up to one week for infrequently traded stocks.

Chapter 5 General Conclusion

The study of financial market behaviour is increasingly based on the econometrics of HFD. The intrinsic feature of HFD is represented by the transaction or tick-by-tick data in which events are recorded one by one as they arise. Consequently, these data are naturally irregularly spaced in time and are realized as point processes. This, jointly with other unique features (such as nonnegative valued; long memory; strong skewness and kurtosis) implies that new methods and new econometric models are needed. Econometric modelling of HFD was first addressed, by Engle and Russell (1998) in the context of an ACD model whose explicit object is the modelling of time between events, and then extended by Engle (2002) and Manganelli (2005) in the context of an Multiplicative Error Model for the modelling of other nonnegative valued financial point processes. The basic idea is to model the nonnegative valued process in terms of the product of a (conditional autoregressive) scale factor and an innovation process with nonnegative support.

Extending the ACD/MEM model into a multivariate setting is frustrated by the limitation of a multivariate nonnegative random process. The full specification of the model requires the joint probability distribution of nonnegative random variables: hence occurrences of such specifications are limited in the literature. Thereby, a common strategy adopted in this study is to reduce the multivariate setting to a series of univariate problems, by making the following two assumptions: a) Weak exogeneity. b) The independence of innovation terms. Then, the multivariate estimation can be done separately equation by equation, as univariate MEM. The object of this thesis is to examine issues related to this strategy and to propose a way to model several nonnegative valued financial point processes jointly. In particular,

we are interested in modelling the dynamics of trading duration, volume and price volatility. Three main Chapters have been developed for this purpose.

We begin with the analysis of weak exogeneity in the second Chapter. The independence of innovation terms is considered as a special case of weak exogeneity in this Chapter. We propose three cases in which the weak exogeneity condition will break down. The simulation study suggests that a failure of the exogeneity assumption implies biased estimators. The biases are very large in the third case nonweak exogeneity, which makes the econometric inferences on the parameters unreliable or even misleading. In empirical analysis, we also derive an LM test for weak exogeneity and test the weak exogeneity of duration in a trivariate (duration, volume and volatility) system. The empirical results indicate that the weak exogeneity is often rejected for frequently traded stocks, but is less likely to be rejected for infrequently traded stocks.

In the analysis of weak exogeneity, we find that as long as the innovation term of ACD model follows a lognormal distribution, the equivalent ARMA model will be a Gaussian distributed. To our knowledge, the lognormal distribution, which is nonnegatively supported, is less interested in the ACD literatures. This motivates us to develop a lognormal ACD model and empirically evaluate its performance in the third Chapter. The Lognormal ACD model permits a humped-shaped hazard function with one free shape parameter, which shows a computation advantage comparing with the existence ACD specification in the Literature. The empirical results show that Lognormal ACD model is superior to Exponential ACD model and Weibull ACD model. It performs similarly to Burr ACD model or generalized gamma ACD model. Moreover, it provides a door of using lognormal distribution for other nonnegative valued financial point processes.

In the fourth Chapter, we propose a general form of vector MEM for the dynamics of several nonnegative valued financial point processes jointly. The vector MEM relaxes these two restrictions imposed by previous work, by allowing interdependence among the variables and releasing weak exogeneity restrictions. Based on results from Chapter three, we further propose to use the multivariate lognormal distribution for the vector MEM. And the maximum likelihood is proposed as a suitable estimation strategy. The model is then applied to the trade and quotes data from the New York Stock Exchange (NYSE) for the dynamics of trading duration, volume and price volatility. The empirical results show that the vector MEM captures the dynamics of the trivariate system successfully. We find that times of greater activity or trades with larger size coincide with a higher number of informed traders present in the market. We highlight that it is unexpected component of trading duration or trading volume that carry the information content. Moreover, the empirical results suggest a significant feedback effect from price process to trading intensity, in which the persistent quote changes and transient quote changes affect trading intensity in different direction.

With respect to further research, the methodology developed in Chapter 4 can easily be extended to model any nonnegative valued variables. An interesting extension is to model financial volatilities. For example, there are different measures of volatility, but no individual one appears to be a sufficient measure on its own. One possibility is to consider absolute daily returns, daily high-low range and daily realized volatility in the vector MEM for forecasting volatility (see Engle and Gallo (2006)). A second example, the multivariate GARCH model is usually used in modelling dynamics interactions among volatilities in different markets. But it is hindered by parametric limitations. However, one can model directly the volatility

proxy (i.e. daily range) for each market and insert other markets' volatility in the expression of its conditional expectations in the vector MEM. This is a very promising possibility, since there is no parametric limitation.

References

- Adkins, L. C., R. C. Campbell, et al. (2012). "The Hausman Test, and Some Alternatives, with Heteroskedastic Data." Advances in Econometrics **29**: 515-546.
- Admati, A. R. and P. Pfleiderer (1988). "A theory of intraday patterns: Volume and price variability." Review of Financial Studies **1**(1): 3-40.
- Allen, D., F. Chan, et al. (2008). "Finite sample properties of the QMLE for the Log-ACD model: Application to Australian stocks." Journal of Econometrics **147**(1): 163-185.
- Allen, D., Z. Lazarov, et al. (2009). "Comparison of alternative ACD models via density and interval forecasts: Evidence from the Australian stock market☆." Mathematics and Computers in Simulation **79**(8): 2535-2555.
- Andersen, T. G., T. Bollerslev, et al. (2001). "The distribution of realized stock return volatility." Journal of Financial Economics **61**(1): 43-76.
- Bauwens, L., F. Galli, et al. (2003). "The moments of Log-ACD models." Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
- Bauwens, L., F. Galli, et al. (2003). The moments of Log-ACD models, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
- Bauwens, L. and P. Giot (2000). "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks." Annales d'Economie et de Statistique **60**: 117-149.

- Bauwens, L. and P. Giot (2003). "Asymmetric ACD models: introducing price information in ACD models." Empirical Economics **28**(4): 709-731.
- Bauwens, L., P. Giot, et al. (2004). "A comparison of financial duration models via density forecasts." International Journal of Forecasting **20**(4): 589-609.
- Bauwens, L., W. Pohlmeier, et al. (2008). High frequency financial econometrics: recent developments, Physica Verlag.
- Bauwens, L. and D. Veredas (2004). "The stochastic conditional duration model: a latent variable model for the analysis of financial durations." Journal of Econometrics **119**(2): 381-412.
- Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity." Journal of Econometrics **31**(3): 307-327.
- Boswijk, H. P. and J. P. Urbain (1997). "Lagrange-multiplier tests for weak exogeneity: a synthesis." Econometric Reviews **16**(1): 21-38.
- Bowe, M., S. Hyde, et al. (2009). "The interrelationship between trading volume, duration and futures contract price volatility." Manchester Business School working paper.
- Cipollini, F., R. F. Engle, et al. (2007). "A model for multivariate non-negative valued processes in financial econometrics." Unpublished paper: Stern School of Business, New York University.
- Cizeau, P., Y. Liu, et al. (1997). "Volatility distribution in the S&P500 stock index." Physica A: Statistical Mechanics and its Applications **245**(3): 441-445.
- Dacorogna, M. M. (2001). An introduction to high-frequency finance, Academic Pr.
- Darrat, A. F., M. K. Hsu, et al. (2000). "Testing export exogeneity in Taiwan: further evidence." Applied Economics Letters **7**(9): 563-567.

- Davidson, R. and J. G. MacKinnon (2004). "Econometric Theory and Methods." Oxford University Press, New York.
- Diamond, D. W. and R. E. Verrecchia (1987). "Constraints on short-selling and asset price adjustment to private information." Journal of Financial Economics **18**(2): 277-311.
- Diebold, F. X., T. A. Gunther, et al. (1998). "Evaluating Density Forecasts with Applications to Financial Risk Management." International Economic Review **39**: 863-883.
- Ding, Z., C. W. J. Granger, et al. (1993). "A long memory property of stock market returns and a new model." Journal of Empirical Finance **1**(1): 83-106.
- Dolado, J., J. M. Rodriguez-Poo, et al. (2004). "Testing Weak Exogeneity in the Exponential Family: An Application to Financial Marked-Point Processes." CORE Discussion Paper No. 2004/49.
- Dufour, A. and R. F. Engle (2000). "Time and the price impact of a trade." Journal of Finance **55**(6): 2467-2498.
- Easley, D., N. M. Kiefer, et al. (1997). "The information content of the trading process." Journal of Empirical Finance **4**(2): 159-186.
- Easley, D. and M. O'Hara (1987). "Price, trade size, and information in securities markets." Journal of Financial Economics **19**(1): 69-90.
- Easley, D. and M. O'Hara (1992). "Time and the process of security price adjustment." Journal of Finance **47**(2): 577-605.
- Engle, R. (2002). "New frontiers for ARCH models." Journal of Applied Econometrics **17**(5): 425-446.
- Engle, R. F. (2000). "The Econometrics of Ultra - high - frequency Data." Econometrica **68**(1): 1-22.

- Engle, R. F. and G. M. Gallo (2006). "A multiple indicators model for volatility using intra-daily data." Journal of Econometrics **131**(1): 3-27.
- Engle, R. F. and D. F. Hendry (1993). "Testing superexogeneity and invariance in regression models." Journal of Econometrics **56**(1): 119-139.
- Engle, R. F., D. F. Hendry, et al. (1983). "Exogeneity." Econometrica **51**(2): 277-304.
- Engle, R. F. and J. R. Russell (1998). "Autoregressive conditional duration: A new model for irregularly spaced transaction data." Econometrica **66**(5): 1127-1162.
- Engle, R. F. and Z. Sun (2007). "When is noise not noise—a microstructure estimate of realized volatility." Manuscript, Stern School, New York University **38**.
- Fernandes, M. and J. Grammig (2005). "Nonparametric specification tests for conditional duration models." Journal of Econometrics **127**(1): 35-68.
- Fernandes, M. and J. Grammig (2006). "A family of autoregressive conditional duration models." Journal of Econometrics **130**(1): 1-23.
- Fischer, A. M. (1993). "Is money really exogenous? Testing for weak exogeneity in Swiss money demand." Journal of Money, Credit and Banking **25**(2): 248-258.
- French, K. R. and R. Roll (1986). "Stock return variances: The arrival of information and the reaction of traders." Journal of Financial Economics **17**(1): 5-26.
- Ghysels, E., C. Gouriéroux, et al. (2004). "Stochastic volatility duration models." Journal of Econometrics **119**(2): 413-433.
- Giot, P. (2000). "Time transformations, intraday data, and volatility models." Journal of Computational Finance **4**(2): 31-62.
- Glosten, L. R. and P. R. Milgrom (1985). "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." Journal of Financial Economics **14**(1): 71-100.

- Goodhart, C. A. E. and M. O'Hara (1997). "High frequency data in financial markets: Issues and applications." Journal of Empirical Finance **4**(2): 73-114.
- Grammig, J. and K. O. Maurer (2000). "Nonmonotonic hazard functions and the autoregressive conditional duration model." The Econometrics Journal **3**(1): 16-38.
- Grammig, J. and M. Wellner (2002). "Modeling the interdependence of volatility and inter-transaction duration processes." Journal of Econometrics **106**(2): 369-400.
- Harris, M. and A. Raviv (1993). "Differences of opinion make a horse race." Review of Financial Studies **6**(3): 473-506.
- Hasbrouck, J. (1988). "Trades, quotes, inventories, and information." Journal of Financial Economics **22**(2): 229-252.
- Hasbrouck, J. (1991). "Measuring the information content of stock trades." Journal of Finance **46**(1991): 179-207.
- Hausman, J. A. (1978). "Specification tests in econometrics." Econometrica **46**(6): 1251-1271.
- Hautsch, N. (2008). "Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model." Journal of Economic Dynamics and Control **32**(12): 3978-4015.
- Ho, T. and H. R. Stoll (1981). "Optimal dealer pricing under transactions and return uncertainty." Journal of Financial Economics **9**(1): 47-73.
- Holden, C. W. and A. Subrahmanyam (1992). "Long-Lived Private Information and Imperfect Competition." Journal of Finance **47**(1): 247-270.
- Hosking, J. R. M. (1980). "The multivariate portmanteau statistic." Journal of the American Statistical Association **75**(371): 602-608.

- Kim, S., N. Shephard, et al. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models." The Review of Economic Studies **65**(3): 361-393.
- Kyle, A. S. (1985). "Continuous auctions and insider trading." Econometrica **53**(6): 1315-1335.
- Lütkepohl, H. (2005). "New introduction to multiple time series analysis." Cambridge Univ Press.
- Lee, C. M. C. and M. J. Ready (1991). "Inferring trade direction from intraday data." Journal of Finance **46**(2): 733-746.
- Lunde, A. (1999). "A generalized gamma autoregressive conditional duration model." Department of Economics, Politics and Public Administration, Aalborg University, Denmark.
- Madhavan, A. (2000). "Market microstructure: A survey." Journal of Financial Markets **3**(3): 205-258.
- Manganelli, S. (2005). "Duration, volume and volatility impact of trades." Journal of Financial Markets **8**(4): 377-399.
- O'Hara, M. (1995). Market microstructure theory, Blackwell.
- O'Hara, M. and G. S. Oldfield (1986). "The Microeconomics of Market Making." Journal of Financial and Quantitative Analysis **21**(4): 361-376.
- Pacurar, M. (2008). "Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature." Journal of Economic Surveys **22**(4): 711-751.
- Rosenblatt, M. (1952). "Remarks on a multivariate transformation." The Annals of Mathematical Statistics **23**(3): 470-472.

- Russell, J. R. and R. F. Engle (2005). "A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times." Journal of Business and Economic Statistics **23**(2): 166-180.
- Shephard, N. (1994). "Partial non-Gaussian state space." Biometrika **81**(1): 115-131.
- Staub, K. E. (2009). "Simple tests for exogeneity of a binary explanatory variable in count data regression models." Communications in Statistics—Simulation and Computation® **38**(9): 1834-1855.
- Sweet, A. L. (1990). "On the hazard rate of the lognormal distribution." Reliability, IEEE Transactions on **39**(3): 325-328.
- Taylor, N. (2004). "Trading intensity, volatility, and arbitrage activity." Journal of Banking & Finance **28**(5): 1137-1162.
- Telang, A. and V. Mariappan (2008). "Hazard Rate of Lognormal Distribution: An Investigation." International Journal of Performability Engineering **4**(2): 103.
- Terza, J. V., A. Basu, et al. (2008). "Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling." Journal of Health Economics **27**(3): 531-543.
- Tony, Lancaster. (1990). The econometric analysis of transition data, Cambridge: University Press.
- White, H. (1981). "Consequences and detection of misspecified nonlinear regression models." Journal of the American Statistical Association **76**(374): 419-433.
- White, H. (1982). "Maximum likelihood estimation of misspecified models." Econometrica **50**(1): 1-25.
- Wu, D. M. (1973). "Alternative tests of independence between stochastic regressors and disturbances." Econometrica **41**(4): 733-750.

Wuensche, O., J. Grammig, et al. (2007). "Time and the Price Impact of a Trade: A Structural Approach." CFR Working Paper.

Xu, Y. (2011). "Lognormal ACD model and comparison with alternative specification" " Working paper, Cardiff Business School, UK.

Zhang, M. Y., J. R. Russell, et al. (2001). "A nonlinear autoregressive conditional duration model with applications to financial transaction data." Journal of Econometrics **104**(1): 179-207.

Appendices

Appendix 1 Proof of Case 1 Weak Exogeneity

Suppose we have the following model for analysis of case 1 weak exogeneity:

$$d_t = \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2).$$

$$v_t = \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2).$$

$$\psi_t = a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1}$$

$$\phi_t = b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t + b_5 \psi_t.$$

If we assume

$$d_t - \psi_t = \varepsilon_{1t} \sim i.i.d(0, \tilde{\sigma}_\varepsilon^2)$$

$$v_t - \phi_t = \varepsilon_{2t} \sim i.i.d(0, \tilde{\sigma}_\eta^2) \quad ,$$

it implies that,

$$d_t = a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} + (d_t - \psi_t)$$

$$= a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} + \varepsilon_{1t} \quad ,$$

$$v_t = b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t + b_5 \psi_t + (v_t - \phi_t)$$

$$= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + b_4 d_t + b_5 (d_t - \varepsilon_{1t}) + \varepsilon_{2t}$$

$$= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + (b_4 + b_5) d_t - b_5 \varepsilon_{1t} + \varepsilon_{2t} \quad .$$

$$= b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \phi_{t-1} + (b_4 + b_5) d_t + \varepsilon'_{2t}$$

Using matrix form for the two equation,

$$\begin{pmatrix} d_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_3 & 0 \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \phi_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_4 + b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -b_5 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Appendix 2 Proof of Case 2 Weak Exogeneity

Support we have the following model for analysis of case 2 weak exogeneity:

$$d_t = \psi_t(\theta_d; \Omega_t) \varepsilon_t, \quad \varepsilon_t \sim i.i.d(1, \sigma_\varepsilon^2).$$

$$v_t = \phi_t(\theta_v; d_t, \Omega_t) \eta_t, \quad \eta_t \sim i.i.d(1, \sigma_\eta^2).$$

$$\psi_t = a_0 + a_1 d_{t-1} + a_2 v_{t-1} + \alpha_3 \psi_{t-1} + a_4 \phi_{t-1},$$

$$\phi_t = b_0 + b_1 d_{t-1} + b_2 v_{t-1} + b_3 \psi_{t-1} + b_4 \phi_{t-1} + b_5 d_t.$$

Using matrix form,

$$\begin{pmatrix} \psi_t \\ \phi_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \phi_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix} + \begin{pmatrix} a_4 & a_5 \\ b_4 & b_5 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix}$$

Using the same method, the above model can be transformed into:

$$\begin{pmatrix} d_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \phi_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix} + \begin{pmatrix} a_4 & a_5 \\ b_4 & b_5 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} d_t - \psi_t \\ v_t - \phi_t \end{pmatrix}$$

If we assume $d_t - \psi_t = \varepsilon_{1t} \sim i.i.d(0, \tilde{\sigma}_\varepsilon^2)$, it becomes

$$\begin{pmatrix} d_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} d_{t-1} - \varepsilon_{1t-1} \\ v_{t-1} - \varepsilon_{2t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix} + \begin{pmatrix} a_4 & a_5 \\ b_4 & b_5 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\begin{pmatrix} d_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 + \alpha_4 & a_2 + \alpha_5 \\ b_1 + b_4 & b_2 + b_5 \end{pmatrix} \begin{pmatrix} d_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_5 & 0 \end{pmatrix} \begin{pmatrix} d_t \\ v_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} - \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix}$$

Appendix 3 Typology of ACD Models

Augmented ACD

$$\psi_t^\lambda = \omega + \alpha \phi_{t-1}^\lambda [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^\nu + \beta \psi_{t-1}^\lambda$$

Asymmetric power ACD ($\lambda = \nu$)

$$\psi_t^\lambda = \omega + \alpha \phi_{t-1}^\lambda [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^\lambda + \beta \psi_{t-1}^\lambda$$

Asymmetric logarithmic ACD ($\lambda \rightarrow 0$ and $\nu = 1$)

$$\log \psi_t = \omega + \alpha [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^\lambda + \beta \log \psi_{t-1}$$

Asymmetric ACD ($\nu = \lambda = 1$)

$$\psi_t = \omega + \alpha [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^\lambda + \beta \psi_{t-1}$$

Power ACD ($\lambda = \nu$ and $b=c=0$)

$$\psi_t^\lambda = \omega + \alpha x_{t-1}^\lambda + \beta \psi_{t-1}^\lambda$$

Box-Cox ACD ($\lambda \rightarrow 0$ and $b=c=0$)

Dufour and Engle (2000)

$$\log \psi_t = \omega + \alpha \varepsilon_{t-1}^\lambda + \beta \log \psi_{t-1}$$

Logarithmic ACD type I ($\lambda, \nu \rightarrow 0$ and $b=c=0$)

Bauwens and Giot's (2000)

$$\log \psi_t = \omega + \alpha \log x_{t-1} + \beta \log \psi_{t-1}$$

Logarithmic ACD type II ($\lambda \rightarrow 0, \nu = 1$ and $b=c=0$)

Bauwens and Giot's (2000)

$$\log \psi_t = \omega + \alpha \varepsilon_{t-1} + \beta \log \psi_{t-1}$$

Linear ACD ($\lambda = \nu = 1$ and $b=c=0$)

Engle and Russell (1998)

$$\psi_t = \omega + \alpha x_{t-1} + \beta \psi_{t-1}$$

Appendix 4 Survival and Hazard Rate

Suppose there is a random variable T with the continuous cumulative distribution function (CDF) $F(t) = \int_0^t f(s)ds = \text{Prob}(T \leq t)$ (where t is a realization of T). Then, its **survival function** is expressed as: $S(t) = 1 - F(t) = \text{Prob}(T \geq t)$. That is, the survival function is the probability that the length of the spell is at least t .

Given that the spell has lasted until time t , the probability that it will end in the next short interval of time, say Δt , is $l(t, \Delta t) = \text{Prob}(t \leq t + \Delta t | T \geq t)$.

This characterization is represented as the **hazard rate**:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t \leq t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t S(t)} = \frac{f(t)}{S(t)}.$$

The hazard rate is the probability of an event occurring in the time interval $[t, t + \Delta t]$, given that it did not occur before time t . The integrated hazard function is expressed as $\Lambda(t) = \int_0^t \lambda(s)ds$ for which $S(t) = e^{-\Lambda(t)}$. So the integrated hazard function is related to the survival rate, such that $\Lambda(t) = -\ln S(t)$.

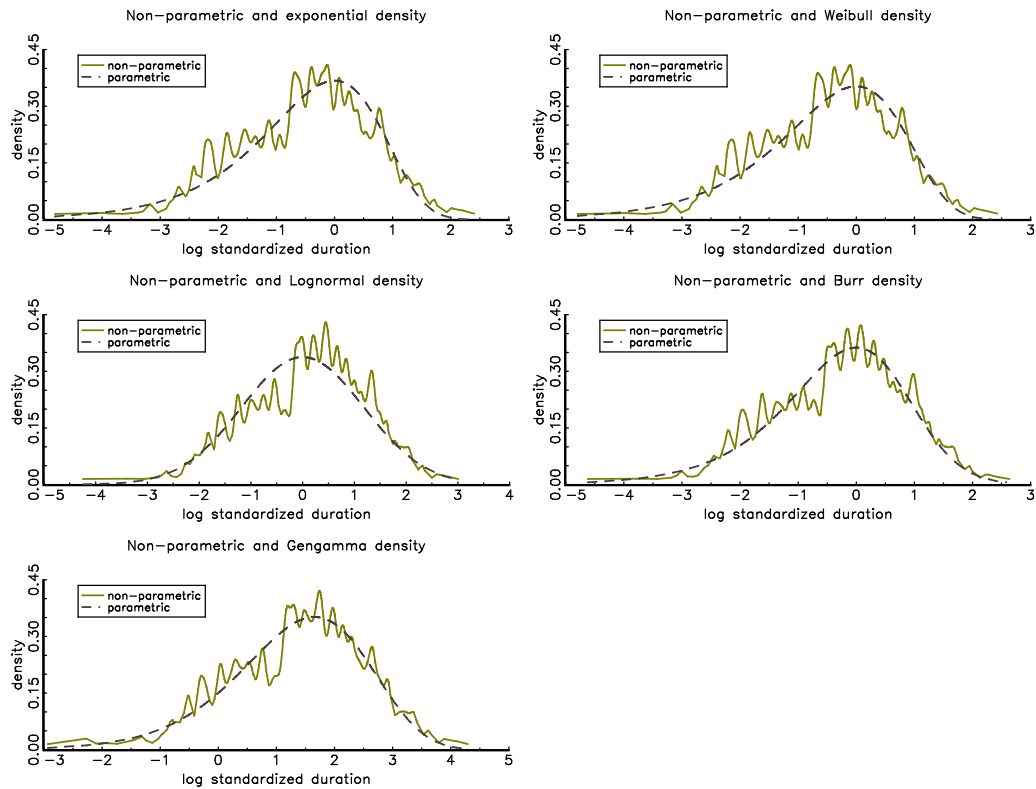
Appendix 5 Hazard Rate of Lognormal Distribution

Simulation study of hazard rate for different

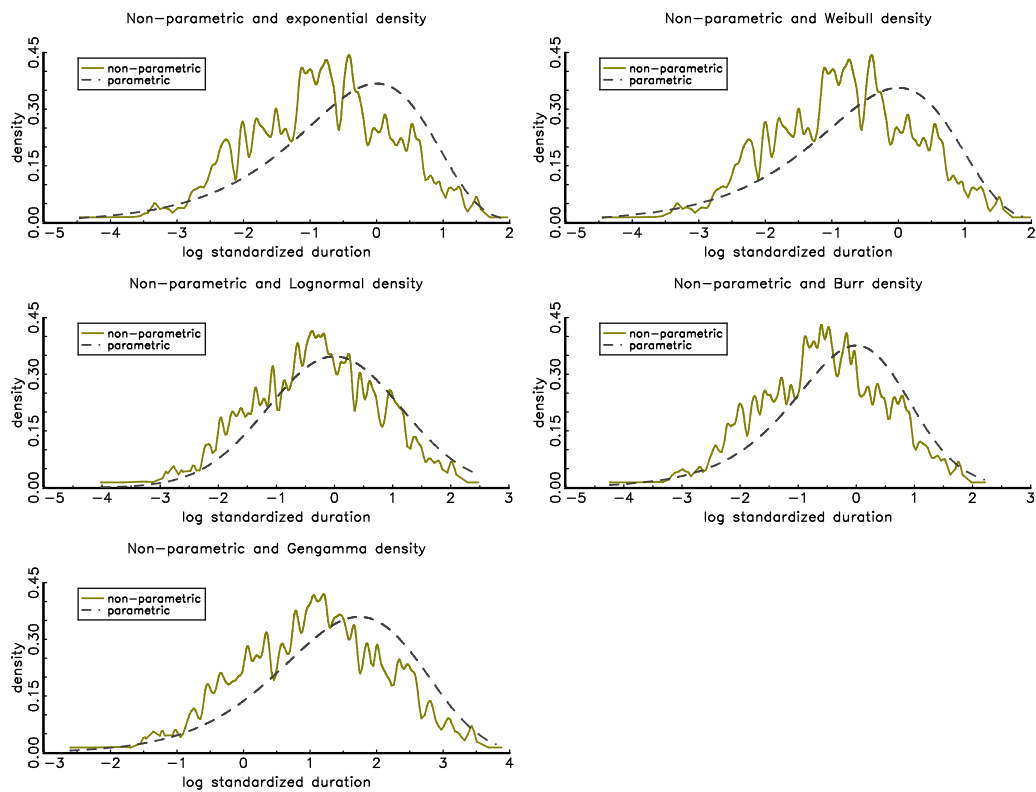
σ	Hazard rate	Percept of positive grads	Percept of negative grads
2.4		0.0008	0.9992
2.3		0.001	0.999
2.2		0.002	0.998
2.1		0.003	0.997
2.0		0.005	0.995
1.9		0.008	0.992
1.8		0.012	0.988
1.7		0.017	0.983
1.6		0.024	0.976
1.5		0.035	0.965

σ	Percept of Infinite Hazard rate	Percept of Positive grad	Percept of Negative grad
0.20	0.000	0.643	0.356
0.19	0.000	0.679	0.321
0.18	0.000	0.703	0.297
0.17	0.000	0.704	0.296
0.16	0.058	0.705	0.237
0.15	0.133	0.707	0.161
0.14	0.202	0.674	0.124
0.13	0.266	0.633	0.102
0.12	0.323	0.592	0.085
0.11	0.378	0.551	0.072
0.10	0.427	0.512	0.061

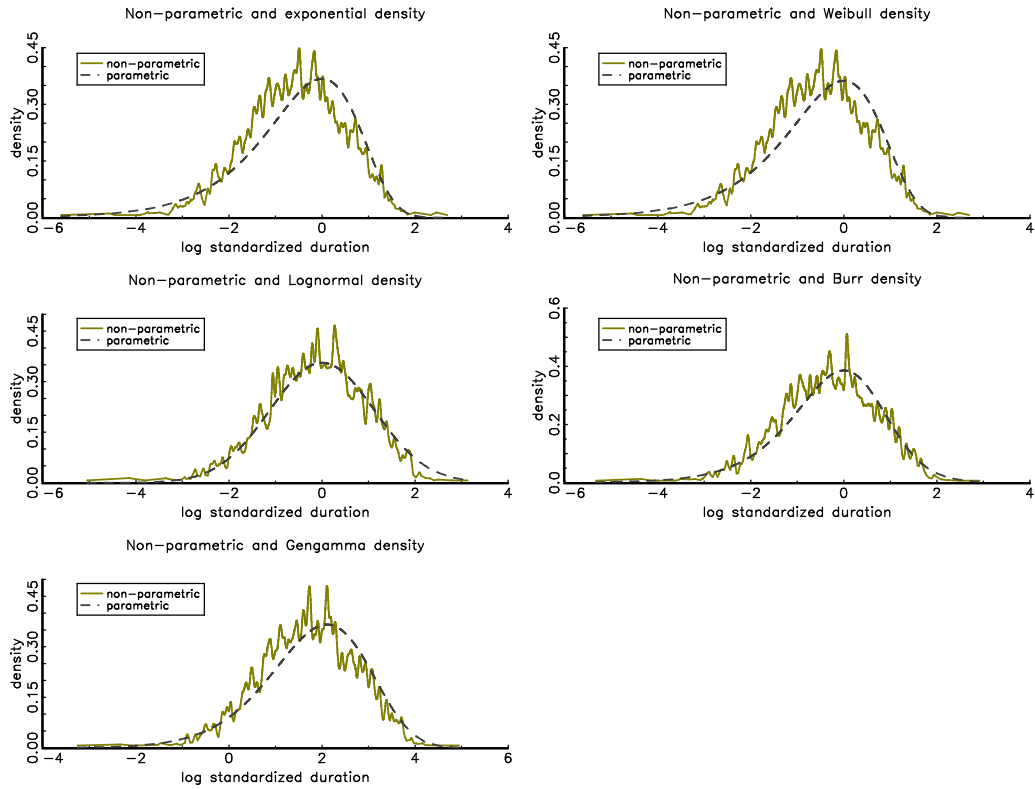
Appendix 6 The Visual Diagnostic Checks for other Three Stocks



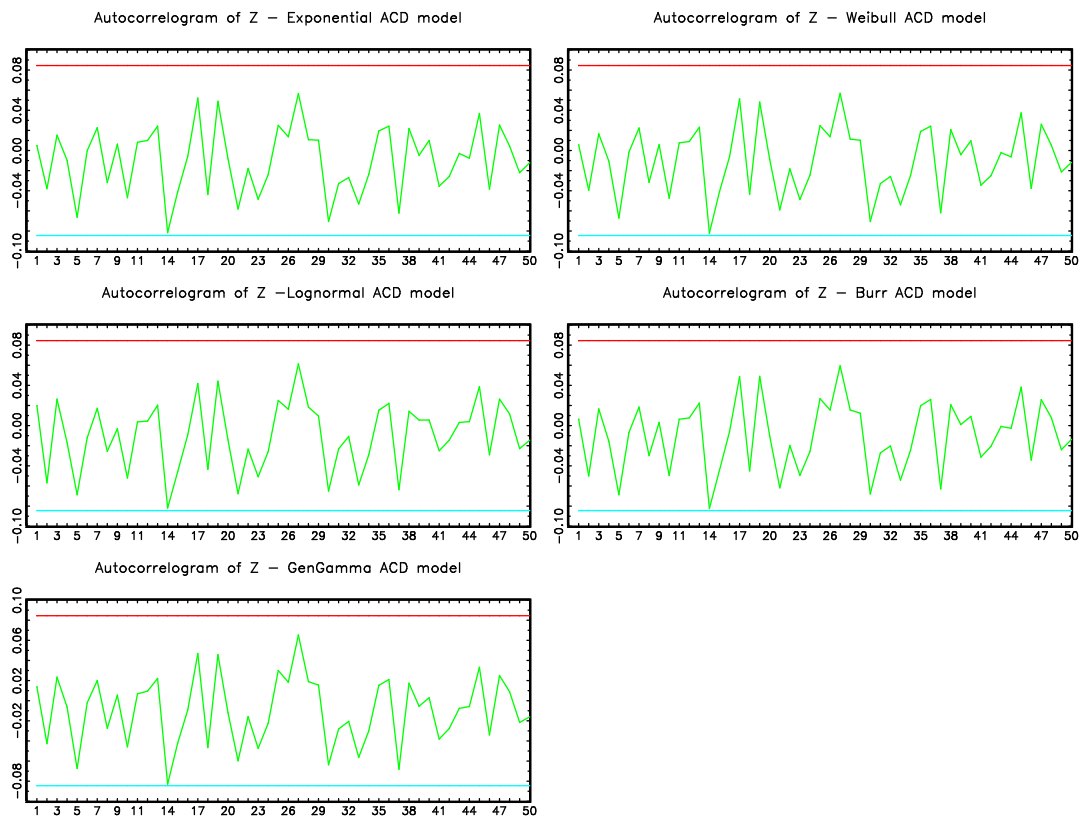
Non-parametric and parametric densities: KO out of sample result



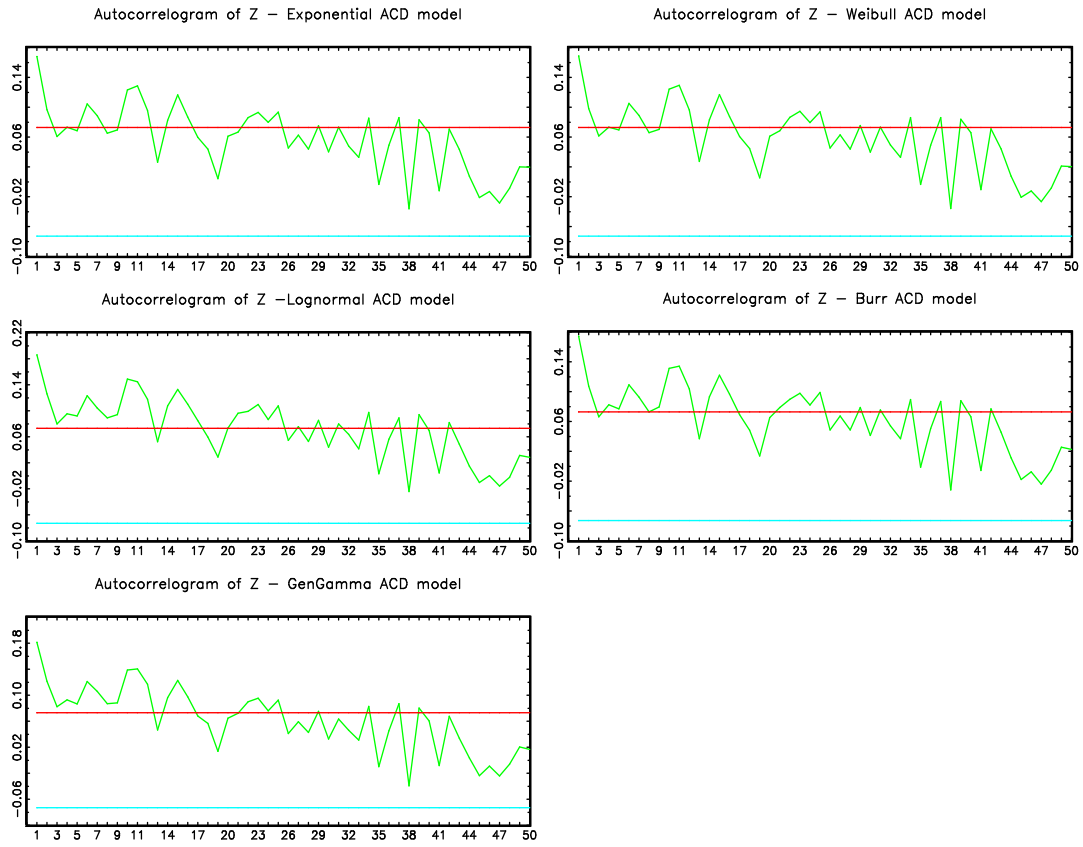
Non-parametric and parametric densities: DIS out of sample result



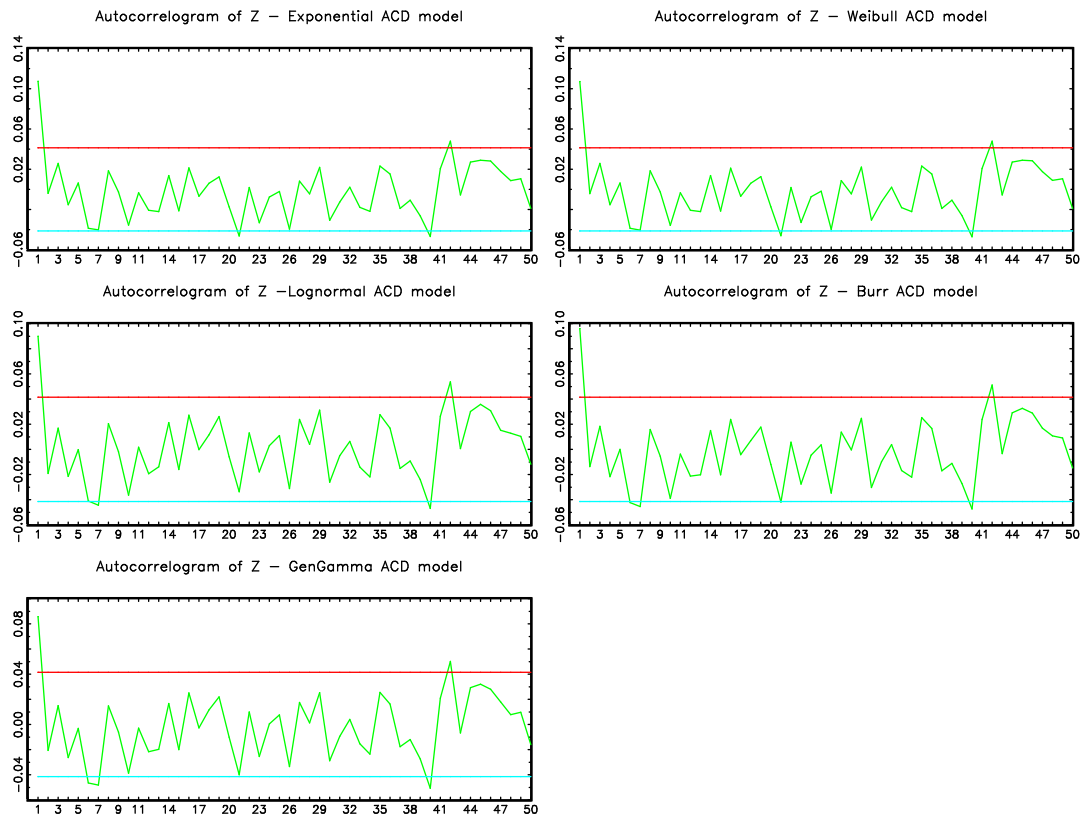
Non-parametric and parametric densities: IBM out of sample result



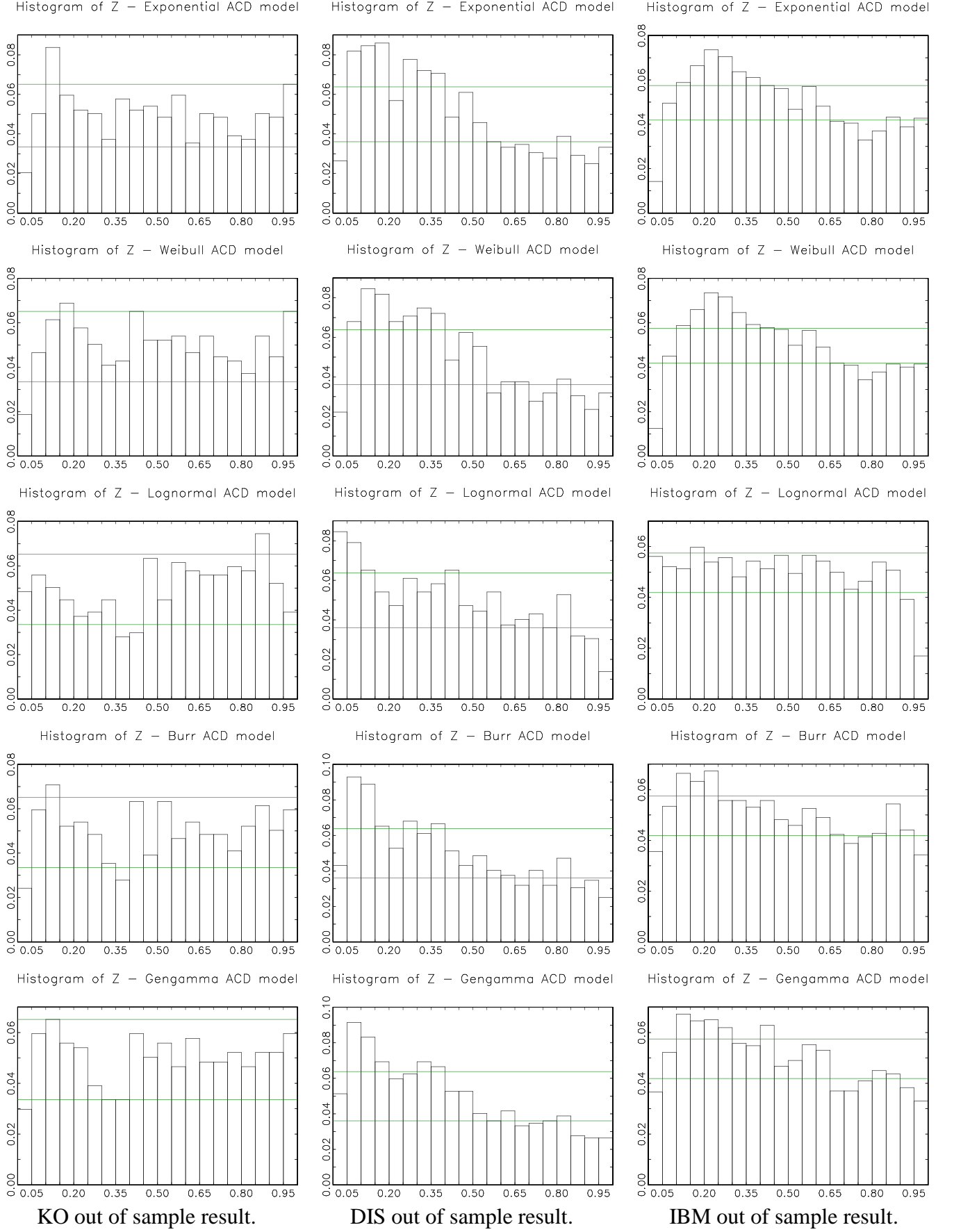
Autocorrelation of z : KO out of sample result



Autocorrelation of z: DIS out of sample result



Autocorrelation of z : IBM out of sample result



Appendix 7 Lognormal Distribution

Univariate lognormal distribution

A lognormally-distributed random variable is a random variable whose logarithm is normally-distributed. Consider a lognormally-distributed x , whose logarithmic transformation $y = \log(x)$ is normally-distributed with mean μ and standard deviation σ . The probability density function for a lognormal distribution is given by,

$$f(x | \sigma) = \frac{1}{\sqrt{2\pi\sigma^2} x} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right], \quad x \geq 0$$

As noted, for example, in Hines and Montgomery (1990). This distribution is skewed with a longer tail to the right of the mean. When μ and σ are known for y , the corresponding mean and variance for x can be found from the following:

$$E(x) = e^{\mu + \frac{1}{2}\sigma^2}$$

$$Var(x) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

Multivariate lognormal distribution

Let $y = (y_1, y_2, \dots, y_k)$ is a k -dimensional random variable having multivariate normal distribution with mean v and covariance matrix $D = (d_{ij})$. The probability density function of y is defined as:

$$f_y(y | D) = (2\pi)^{-k/2} |D|^{-1/2} \exp\left(-\frac{1}{2}(y - v)' D^{-1}(y - v)\right)$$

Then, the variable, $x = \exp(y)$, have a multivariate lognormal distribution. It is defined as $x \sim \ln N(v, D)$. Use Jacobian transformation, and $y = h(x) = \ln(x)$ the probability density for multivariate lognormal distribution has the following form:

$$f_x(x|D) = f_y(h(x)|D) \left| \frac{dh}{dx} \right|$$

$$= (2\pi)^{-k/2} |D|^{-1/2} * (x_1 * x_2 * \dots * x_k)^{-1} * \exp\left(-\frac{1}{2}(\ln x - v)' D^{-1}(\ln x - v)\right) \quad y_i > 0$$

Law and Kelton (2000) show the covariance and correlation of the bivariate

lognormal variables $x = (x_1, x_2, \dots, x_k)$ as follows:

$$\mu = E(x) = (\mu_1, \mu_2, \dots, \mu_k)'$$

$$\mu_i = e^{v_i + \frac{1}{2}d_{ii}}$$

$$\Sigma = E(x - \mu)(x - \mu)' = \sigma_{ij}$$

$$\sigma_{ij} = e^{(v_i + v_j + \frac{d_{ii} + d_{jj}}{2})} (e^{d_{ij}} - 1)$$

$$\rho_{ij} = \frac{e^{d_{ij}} - 1}{\sqrt{(e^{d_{ii}} - 1)(e^{d_{jj}} - 1)}}$$

where d_{ij} is the ij th element of D . It is clear that if y_1, y_2, \dots, y_k are independent, then

x_1, x_2, \dots, x_k are also independent and vice verse.

Jacobian transformation

Let $y = (y_1, \dots, y_k)$ be a k -dimensional random variable with probability density function (pdf) $f_y(y): f_y(y): R^k \rightarrow R$. Define some 1:1 differentiable transformation of y into x using $g: R^k \rightarrow R^k$,

$$g(y) = \begin{bmatrix} g_1(y) \\ \vdots \\ g_k(y) \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = x$$

with inverse

$$h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_k(x) \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} = y$$

The pdf of y , the transformed random variable, is

$$f_x(x) = f_y(h(x)) \left| \frac{dh}{dx} \right|$$

where

$$\left| \frac{dh}{dx} \right| = \left| \frac{\partial(h_1, \dots, h_k)}{\partial(x_1, \dots, x_k)} \right| = \begin{vmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \dots & \frac{\partial h_1}{\partial x_k} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \dots & \frac{\partial h_2}{\partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_k}{\partial x_1} & \frac{\partial h_k}{\partial x_2} & \dots & \frac{\partial h_k}{\partial x_k} \end{vmatrix}$$

Appendix 8 Proofs of Impulse Response Function

Suppose we have the following vector MEM model:

$$x_t = \mu_t \ominus \varepsilon_t,$$

$$\ln \mu_t = \omega + A \ln x_{t-1} + B \ln \mu_{t-1}$$

Firstly, we transform the vector MEM into a VARMA model, by substituting

$\ln \mu_t$ with $\ln x_t - \ln \frac{x_t}{\mu_t}$. Then,

$$\begin{aligned} \ln x_t &= \omega + A \ln x_{t-1} + \ln \frac{x_t}{\mu_t} + B \left(\ln x_{t-1} - \ln \frac{x_{t-1}}{\mu_{t-1}} \right) \\ &= \omega + A' \ln x_{t-1} + \ln \varepsilon_t + B' \ln \varepsilon_{t-1} \end{aligned}$$

where $A' = A + B$, and $B' = -B$.

The causal and feedback effect are not affected by this transformation. Therefore, it is feasible to assume that $\ln \varepsilon_t$ follows a multivariable Gaussian distribution. Then, (quasi) maximum likelihood estimation can be used to estimate the parameters of VARMA model. Suppose v_t is a multivariable Gaussian distributed random variables, then

$$\ln x_t = \bar{\omega} + A' \ln x_{t-1} + v_t + B' v_{t-1}$$

where $v_t \sim N(0, \Sigma_v)$, $\bar{\omega} = \omega + D + BD$, $D = \text{dia}(\Sigma_v)$.

In the impulse response, we work on the impulse of v_t on the $\ln x_t$ in a standard way. Writing the VARMA (1,1) equation as an infinite VAR model:

$$\ln x_t = \bar{\omega} + \sum_{i=0}^{\infty} \Phi_i v_{t-i} = \bar{\omega} + \Phi(L)v_t$$

where $\Phi(L) = (I - A'L)^{-1}(I + B'L) = I + (A' + B')L + A'(A' + B')L^2 + A'^2(A' + B')L^3 + \dots$

and $\Phi_i = A'^{i-1}(A' + B') = (A + B)^{i-1}(A - B)$, $\Phi_0 = I$.

The impulse response function is :

$$\frac{\partial \ln x_t}{\partial v_{t-s}} = \Phi_s$$

Or the impulse response function for $t > 0$ is :

$$\frac{\partial \ln x_t}{\partial v_0} = \Phi_t$$

This process can be rewritten in such a way that the residuals of different equations are uncorrelated. For this purpose, we choose a decomposition of the white noise covariance matrix $\Sigma_v = W\Sigma_\tau W'$, where Σ_τ is a diagonal matrix with positive diagonal elements and W is a lower triangular matrix with unit diagonal. This decomposition is obtained from the Choleski decomposition $\Sigma_v = PP'$ by defining a diagonal matrix D which has the same main diagonal as P and by specifying $W = PD^{-1}$ and $\Sigma_\tau = DD'$.

$$\ln x_t = \bar{\omega} + \sum_{i=0}^{\infty} \Psi_i \tau_{t-i}, \quad \Psi_i = \Phi_i W^{-1}$$

Then the impulse response function for $t > 0$ is :

$$\frac{\partial \ln x_t}{\partial \tau_0} = \Psi_t$$