Bayesian econometric modelling of informed trading, bid-ask spread and volatility

Samuel Dua Oduro

Centre for Actuarial Science, Risk and Investment
School of Mathematics, Statistics and Actuarial Science

University of Kent

This dissertation is submitted in the subject of Actuarial Science for the degree of

Doctor of Philosophy

September 2016

I would like to dedicate this thesis to the memory of my deceased parents
Emmanuel Oduro Amankwaa and Beatrice Akua Agyapomaa

And to my wife Joyce and children Loretta, Rodney, Phoebe and Nathaniel whose sacrifices has seen me through this journey.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation is original and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Samuel Dua Oduro
September 2016

Acknowledgements

To God, I give thanks for his protection through this academic journey. May His name be praised. The thesis has been made possible through the contributions of numerous individuals whom I was privilege to meeting at Kent. Firstly I would like to express my profound gratitude to Professor Jim E. Griffin and Dr Jaideep S. Oberoi who constituted my joint-supervisory team. I am indeed very thankful for their immense guidance, support, constructive criticism and detailed scrutiny of the entire research work. It was a rare privilege to work with such intelligent, friendly, firm and accommodating gentlemen. Secondly, I would like to thank the School of Mathematics, Statistics and Actuarial Science for the generous funding which enabled me focus on my research. I would like to thank Professor David Veredas of Vlerick Business School in Belgium for agreeing to examine this thesis. I am most grateful to him. To Mrs Claire Carter, our wonderful and always cheerful postgraduate research officer, I say *Tapadh leibh*. You have always been there to support me not only on administrative procedures but personal life issues as well. Special thanks to Dr Dhirendra Kumar Sakaria, Dr Evangelia Mitrodima and Dr Abera Ayalew Muhamed for their friendship and support when the going went though in Canterbury. I appreciate the time I shared with you all. I am also very indebted to my mother-inlaw Madam Felicia Afia Asantewaa who took care of my children while I was away in Canterbury. I would like to appreciate Mr Samuel Mireku Agyapong and his wife Faustina Acheampong for the invaluable support to myself and my family before and during my relocation to Canterbury. Sincere thanks to my brother Mr Emmanuel Kwame Oduro and sister Mrs Giftina Appiah-Agyei Sarpong for their prayer support. Finally, I am very indebted to my wife Joyce and children Loretta, Rodney, Phoebe and Nathaniel who endured my absence while I was in Canterbury far away from home. No amount of financial reward will suffice the patience you had for me on occasions when I indirectly extended my frustration from this research to you. God bless you all.

Abstract

Recent developments in global financial markets have increased the need for research aimed at the measurement and possible reduction of liquidity risk. In particular, market crashes have been partly blamed on the sudden withdrawal of liquidity in markets and increases in liquidity risk. To this end, it is important to develop better approaches for inferring or quantifying liquidity risk. Liquidity risk caused by some investors trading on their information advantage (informed trading) has been a subject of market microstructure research in the last few decades. Researchers have employed information-based models that use observed or inferred order flow to investigate this problem. The Probability of Informed Trading (PIN) is a measure which uses inferred order flow to quantify the extent information asymmetry. However, a number of computational issues have been reported to effect the estimation of PIN. Using an alternative methodology, we address the numerical problem associated with the estimation of PIN. Varied evidence of a relationship between volume and bid-ask spread has been documented in the extant literature. In particular, theory suggests that bid-ask spread and volume are jointly driven by a common process as both variables measure an aspect of liquidity. The complex relationship between these variables is time-varying since the informed trading component of order flow changes as trading takes place. Thus, volume and bid-ask spread may provide insight on the time-varying composition of economic agents trading an asset. We exploit the nonlinear relationship between traded volume and bid-ask spread to develop a model that can be used to infer informed and uninformed trading components of volume. The structure of the model and estimation methodology enhances the sequential processing and incorporation of past volume and bid-ask spread as conditioning information. The model is applied to two equities that trade on the New York Stock Exchange. Finally, to increase our understanding on the effects of liquidity risk on volatility, we also examine whether separating volume into informed and uninformed components can provide further insight on the relationship between liquidity risk and volatility.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Financial markets provide platforms for diverse participants including private investors, institutional investors, brokers and designated market-makers to trade financial securities. In addition to the facilitation of trade between market participants, the markets collect and make available trade-related information such as prices, traded volume, time of trade, number of transactions and other relevant variables about assets. Market participants take account of this information when they make purchase and sale decisions.

Academic researchers also use the trade-related information from financial markets to develop theories and models that can be used to learn about the trading behaviour of market participants. A branch of the finance literature that studies investor trading behaviour in financial markets is called market microstructure. This research area has been defined differently by several academics. A common definition that can be drawn from papers including Madhavan (2000), O'Hara (1995) and Spulber (1996) is that market microstructure explores the evolution of asset prices while taking into consideration the following micro-level market issues:

1. The effect of market regulations and mechanisms on the evolution of asset prices.

2. How well the organisation of markets enhances the easy exchange of assets in large volumes with little or no impact on the price of the asset (liquidity).

3. How information generated from the demand and supply decisions of market participants, affects the price of an asset.

Market microstructure theory assumes that financial market participants have different information sets which influence their trading behaviour. Market participants, therefore, reveal the information they hold about the asset through their demand and supply decisions. Depending on the quality of information available, market participants alter their expectations on the stream of future cash-flow of an asset and hence the value they place on the asset.

At the core of the extant literature, economic agents in financial markets are categorised into informed and uninformed. The uninformed market participants are sometimes referred to as liquidity traders. This categorisation is based on the assumed motives behind the trading decisions of market participants. Informed market participants are considered to have superior knowledge or the sophistication to determine whether an asset is mis-priced. They, therefore, enter into trades hoping to gains from their information advantage. On the other hand, uninformed market participants do not have any information on the future price of the asset and hence trade on a multitude of reasons. These reasons may include portfolio re-balancing, the need for funds for other investment projects and consumption smoothing.

The likelihood of a market participant entering into a transaction with other market participants who may potentially have superior knowledge about the value of the asset creates what is referred to as *information asymmetry.* Bagehot (1971) made the argument that the differences between the prices at which investors are able and willing to buy or sell (bid-ask spread) an asset exist because some investors possess superior information. This implies that the size of the bid-ask spread is a function of information asymmetry. Information asymmetry is a fundamental source of uncertainty faced by market makers and liquidity providers. Investigating the presence or otherwise of informed trading in an asset and within the market is very important since information asymmetry affects the liquidity of an asset and the market in general.

In a seminal paper, Glosten and Milgrom (1985) formally presented a theoretical model for the idea of Bagehot (1971). According to Glosten and Milgrom (1985) traders arrive at the market sequentially to have their orders executed. In their model bid and ask prices are set based on the liquidity providers' belief of the proportion of informed traders in the market. Thus in the absence of exogenous transaction cost there exist a positive bid-ask spread. On the other hand, Kyle (1985) postulates that informed traders submit their orders strategically in a gradual manner. This strategic behaviour of informed traders ensures that the impact of their trades on asset price is minimal. Some academic papers have subsequently explored the information asymmetry problem by extending the work of Glosten and Milgrom (1985). Many papers based on initial work of Kyle (1985) also explore the impact of information asymmetry on trading cost.

Papers including Chordia et al. (2001), Acharya and Pedersen (2005), Brennan and Subrahmanyam (1996) and Easley and O'Hara (2003) argue that in equilibrium, high levels of information asymmetry create significant trading costs. This causes uninformed traders to demand higher returns resulting in assets being purchased at a discount. Chordia et al. (2001) also indicate that information asymmetry affects the volatility of assets. French and Roll (1986) found evidence of increased price volatility principally caused by private information of informed traders. Using a theoretical model, Wang (1993) also argues that in a market with information asymmetry, less informed traders demand an extra premium for the uncertainty of trading against better informed traders. Price volatility will, therefore, increase as less informed traders post quotes that widen the bid-ask spread.

1.1 Measuring Information Asymmetry

A number of approaches have been taken in the market microstructure literature to provide proxies for and measures of information asymmetry. In what follows we provide a brief review of some of the prominently used methods. This review however is not intended to be an exhaustive review of the numerous methods in the literature.

1.1.1 Spread decomposition models

A basic measure of illiquidity is the bid-ask spread. The bid-ask spread measures the rent market-makers charge for the provision of immediate liquidity. The adverse selection theory put forward by Glosten and Milgrom (1985) suggests that a trader offering to sell a large amount of his/her stock holdings unexpectedly will have to take a lower price for the asset if the counter-party to the trade believes that the seller of the stock has information which other traders do not know.

Various authors including Glosten and Harris (1988), foster1993variations, Madhavan et al. (1997), Huang and Stoll (1997) and Sadka (2006), have used a trade indicator regression model to decompose the bid-ask spread into *inventory-holding cost*, *adverse-selection cost* and *order-processing cost* components. Liquidity suppliers incorporate into the bid-ask spread the costs associated with the execution and processing of orders they receive. These may include costs such as brokerage fees and transaction tax. Apart from these costs, liquidity providers are exposed to the risk of trading with better informed traders. The adverse-selection component of the bid-ask spread is the compensation demanded by liquidity traders for trading with informed traders. Also, since market makers hold inventory to meet their obligation of providing immediate liquidity when demanded, they are exposed to price changes. Hence they demand compensation for this price risk in the form of inventory-holding cost.

Let changes in mid-quote that prevailed before the transaction at time t be denoted as $r_t = {}^{(p_t^{ask}+p_t^{bid})}/_2 - {}^{(p_{t-1}^{ask}+p_{t-1}^{bid})}/_2$. Buyer and seller initiated trades at time t are also denoted by $q_t = +1$ and $q_t = -1$ respectively. If S_t is the quoted spread prior to the transaction at time t, then the Huang and Stoll (1997) model which encompasses many of the trade indicator models is of the form

$$r_t = (\alpha + \beta)\frac{S_{t-1}}{2}q_{t-1} + \alpha(1 - 2\pi)\frac{S_{t-2}}{2}q_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{1.1}$$

where $E[q_{t-1}|q_{t-2}] = (1 - 2\pi)q_{t-2}$. The model parameters α and β are the adverse selection and inventory components of the quoted spread. The order processing component is calculated as $1-\alpha-\beta$. The probability that the trade at time t is opposite in sign to the trade at $t-1$ is π. The adverse selection component is used as a proxy for information asymmetry. A temporary increase in the information asymmetry between the informed and uninformed investors should cause a temporary positive deviation in the bid-ask spread from its normal level.

1.1.2 Price Impact Models

Informed market participants are likely to evaluate the impact of their trades on the price of the asset and hence would act strategically when trading. Kyle (1985) introduced one of the early strategic information models for a single asset market in which a monopolistic market maker operates. The market maker in this market sets break-even prices in such a way that the price sensitivity (referred to as *price impact*) to trades balances losses and gains resulting from transactions with informed and uninformed traders respectively.

In this model a trade from an informed trader should cause a permanent price impact because it partly reflects the traders' private information. The market subsequently incorporates this information into the price. Studies including Easley and O'Hara (1987), Glosten and Harris (1988), Glosten (1989) and Kyle (1985) argue that price impact of trade better captures the illiquidity effect of information asymmetry. The Kyle (1985) model is of the form

$$\Delta P_t = \gamma_0 + \gamma_1 X_t q_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{1.2}$$

where ΔP_t is price change, X_t is total volume of shares traded between time $t-1$ and t respectively. However, other researchers have used various transformations of volume such as square root of volume for X_t. For the same trading interval if the trade is inferred to be a buyer initiated trade we have $q_t = 1$ while a seller initiated trade is $q_t = -1$. In equation 1.2, γ_1 measures the effect of information asymmetry on prices while public information is captured in the error term ε_t. The model has been widely applied to different asset classes.

Cont et al. (2014), Foster and Viswanathan (1993), Glosten and Harris (1988) and Huh (2014) have extended and applied this model in various ways to answer the same research problem. The drawback of this model is that at very low frequencies such as daily level, aggregate trades will have to be classified as either buyer or seller initiated. This may render the estimates of the model parameters less accurate compared to estimating the model at high frequencies.

1.1.3 Vector Autoregressive (VAR) Models

Hasbrouck (1991) introduced the Vector Autoregressive (VAR) model to study the relationship between economic and financial variables. It is also used as a model to measure information asymmetry. Other models including the price impact model of Kyle (1985) have assumed that information asymmetry has instant impact on asset prices. However, the intuition behind the VAR is that the impact of information on asset prices takes effect gradually.

Defining signed volume at time t as $x_t = q_t X_t$, Hasbrouck (1991) proposed the following model,

$$r_t = \sum_{i=1}^{K} \alpha_i r_{t-i} + \sum_{i=0}^{K} \beta_i x_{t-i} + \varepsilon_{1,t}, \qquad \varepsilon_{1,t} \sim \mathcal{N}\Big(0, \sigma_1^2\Big) \tag{1.3a}$$

$$x_t^0 = \sum_{i=1}^{K} \gamma_i r_{t-i} + \sum_{i=1}^{K} \rho_i x_{t-i} + \varepsilon_{2,t}, \qquad \varepsilon_{2,t} \sim \mathcal{N}\Big(0, \sigma_2^2\Big). \tag{1.3b}$$

The model error terms $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are updates to public and private information sets respectively. Hasbrouck (1991) chose the value of K to be 5. The proxy for information in this model is x_t. It can be any trade related variable such as duration between trades. The estimation of the impulse response function $\sum_{i=0}^{K} E(r_{t+i})$ provides a measure of the private information of the trade.

1.1.4 Probability of Informed Trading (PIN)

The Probability of Informed Trading, introduced by Easley et al. (1996) is another measure of information asymmetry risk which is based on the asymmetric sequential trade model of Glosten and Milgrom (1985). Since the introduction of PIN, it has been extensively used as a proxy for liquidity risk in finance and in particular the market microstructure literature. Examples of the application of PIN as a risk measure are Borisova and Yadav (2015), and Chung and Li (2003).

PIN is used as a possible risk factor in the determination of expected asset returns. In the US market, Easley et al. (2002) extended Easley et al. (1996) to investigate the effect of information asymmetry on expected asset returns. They conclude that assets with higher PIN have correspondingly higher expected returns in comparison with assets that have lower PIN. In another study, Easley et al. (2010) established that PIN plays a significant role in providing explanatory power in a regression model where cross-sectional asset return is the response variable. Brennan et al. (2012) report the existence of a significant positive relationship between expected returns and price changes generated by sell orders. Motivated by the findings in Brennan et al. (2012), Subrahmanyam et al. (2013) studied the asymmetric relationship between the components of PIN. They found that the component of PIN attributable to trading based on unfavourable information (sell orders) is priced.

The PIN has traditionally been estimated on aggregate daily buyer and seller initiated trades. A number of studies including Boehmer et al. (2007), Easley et al. (1997*b*), Easley et al. (2010), Lei and Wu (2005), Vega (2006) and Lin and Ke (2011) have indicated that the PIN may be biased. The estimation of the underlying parameters of PIN is prone to numerical instability as a result of the nature of the likelihood function. This leads to corner solutions, especially for frequently traded assets.

In Chapter 2, we use a Bayesian approach to estimate the model parameters of the PIN measure. This alternative estimation methodology does not rely on any optimisation routine and hence avoids the numerical problems reported in the maximum likelihood estimation approach of calculating PIN. The methodology also provides a natural way of estimating the uncertainty about the model parameters and that of the PIN measure.

The Bayesian methodology Chapter 2 is implemented on high frequency buyer and seller initiated trades to aid the estimation of daily PIN. This is done in Chapter 3 where we compare the time series of daily PIN with the Volume Synchronized Probability of Informed Trading (VPIN) introduced by Easley et al. (2011) as an alternative measure of information asymmetry. The VPIN is widely used by many finance professionals to measure order toxicity.

Researchers are continually exploring the theoretical relationship between various market variables to build new information-based models which better estimate information asymmetry. One such relationship is that which exists between volume and bid-ask spread. In particular, theory suggests that bid-ask spread and volume are jointly determined. Copeland and Galai (1983), Glosten and Milgrom (1985) and Kyle (1985) posit that investors review their bid and ask quotes in response to their beliefs about the composition of market participants. In reviewing their quotes, market makers learn from the orders made by other investors who take the opposite side of the trade. A wider bid-ask spread may be an indication of a higher estimate of information asymmetry or other risks including inventory risk. A wider bid-ask spread will have a feedback effect on subsequent trading decisions. Informed traders experiencing a fall in their anticipated profits due to the increased cost of trading will reduce their order sizes subsequently.

Studies including Lesmond (2005) and Venkatesh and Chiang (1986) use the bid-ask spread to test for increased information asymmetry before the disclosure of events such as earnings or dividend announcements. Empirical predictions by Admati and Pfleiderer (1988), Copeland and Galai (1983), Easley and O'Hara (1987), Foster and Viswanathan (1990), Glosten (1987, 1989) and Kyle (1985) show that bid-ask spread is positively related to information asymmetry. Admati and Pfleiderer (1988) suggested that informed traders are attracted to the market when discretionary liquidity traders are present in the market. This way, informed traders can conceal the information content of their trades and hence minimise the possible impact of their trades on the cost of trading. Contrary to this intuition, Foster and Viswanathan (1993) show that volume and bid-ask spread exhibit similar characteristics during any typical trading day. Both volume and bid-ask spread decrease from a high level a few minutes after trading has begun in the morning and then rise again a few hours after lunchtime, peaking during the last hour of the trading day.

Hasbrouck (1991) also argued that bid-ask spreads respond continuously to trades. The dynamic changes in bid-ask spread suggest that market participants' perception of information asymmetry is not the same all the time. The bid-ask spread is, therefore, a natural measure of liquidity reflecting investors' expectations of market movements as they learn from the trading process. Thus, the temporal relationship between volume and bid-ask spread may provide insight on the time-varying composition of economic agents trading an asset.

None of the information-based models above has explored the relationship between volume and bid-ask spread in an attempt to infer the unobservable informed trading. Motivated by this, we propose an alternative approach of inferring informed trading in Chapter 4. We model the joint relationship between traded volume and bid-ask spread dynamically using a state space model while decomposing volume into two components with corresponding effects on bid-ask spread.

We depart from the use of derived variables such as the buyer or seller initiated trades or volume which has predominantly been used in the literature. Using our model, it is possible to account for the uncertainty about model parameters and unobserved processes. The structure of the model and estimation methodology enhances the incorporation of past volume and bid-ask spread as conditioning information. To the best of our knowledge, this is the first attempt at exploiting the predicted relationship between traded volume and bid-ask spread to extract unobserved informed and uninformed trading using the Kalman Filter in a Bayesian framework.

Other branches of the finance literature have extensively explored the relationship between asset returns and volume to learn about information in asset prices. This has led to many volatility forecasting models of varied complexity. In the market microstructure literature, the relationship between trade related data have been used to study the relationship between informed trading and volatility. These studies have resulted in mixed findings which are contingent on the underlying assumptions about the behaviour of market participants. In Chapter 5 we propose alternative models that can be used to explore the temporal relationship between volatility, informed trading and uninformed trading. The models exploit the predicted relationship between traded volume, bid-ask spread and volatility. We use the models to generate one-step-ahead volatility forecasts. The models investigated in Chapter 5 are compared with the Heterogeneous Autoregressive (HAR) model introduced by Corsi (2009).

The modelling approach we take in Chapters 4 and 5 are unique from what has been done in the literature in the sense that we do not rely on ordinary least squares estimation which assumes that the effect of information asymmetry is fixed over the entire sample period. We are also able to account for parameter uncertainty and fat tails in the observed market data.

We provide some conclusions in Chapter 6.

Chapter 2

A Bayesian Approach To Probability Of Informed Trading

2.1 Introduction

During the last three decades researchers and finance practitioners have been investigating how to quantify information asymmetry risk in financial markets. In recent times investigations about informed trading risk has increased partly in response to financial market crashes. A school of thought in the financial literature attributes the market crashes to the temporary withdrawal of liquidity by some investors. It is, therefore, appropriate that in times of market uncertainty and temporary liquidity dry-up, we revisit existing approaches used for quantifying information asymmetry risk.

The Probability of Informed Trading is a widely used measure of information asymmetry risk in the finance literature. The underlying parameters of PIN model are estimated using maximum likelihood estimation. In the estimation of the parameters underlying the PIN, a number of numerical computational issues have been documented in Boehmer et al. (2007), Easley et al. (1997*b*), Easley et al. (2010), Lei and Wu (2005), Vega (2006), Yan and Zhang (2012) and Lin and Ke (2011). The literature cited reports that due to the nature of the likelihood function of the PIN model sometimes MLE leads to floating-point exceptions. Secondly, the maximum likelihood estimates of some of the underlying parameters of PIN lie on the boundary of the parameter space. Also, one need to choose initial values of the MLE carefully to achieve stable results. Thus the estimates are likely to be dependent on the choice of the initial values used by the optimiser. Finally, in circumstances where the likelihood function has several maxima, the MLE optimiser may settle on a local maximum which may not necessarily be the global maximum we seek. These computational issues potentially effect the accuracy of the PIN estimate which in turn will impact any risk management decision drawn based on the PIN.

Boehmer et al. (2007), Easley et al. (2010, 1997*b*), Lei and Wu (2005), Vega (2006), Yan and Zhang (2012) and Lin and Ke (2011) have suggested alternative solutions to the computational problems of PIN estimation. However, there seems to be no concrete solution for the known problems. Motivated by the search for improvement in estimation of PIN as well as the search for alternative estimation methods to PIN, we employ a Bayesian approach to the estimation of the parameters of Easley et al. (1996) information asymmetry model. Using the Bayesian methodology, we can account for the uncertainty in the estimation of the model parameters. This approach also avoids the numerical problem of the MLE optimisers. Another motivation for using a Bayesian method is its ability to handle complex models where tractable analytical formulations are difficult to write down in closed-form and hence to estimate. Furthermore, we have a natural way of calculating the standard errors of the model parameters and the PIN from their respective posterior distributions.

In section 2.2, we provide a brief introduction to the theory underpinning the Easley et al. (1996) model. The estimation method is also discussed. We proceed with a description of our method of estimation in section 2.3. The theory and estimation method of Easley et al. (2002), which is an extension of the Easley et al. (1996) is described in section 2.4. In section 2.4, we give details of the Bayesian estimation method of the extended model. In section 2.6 and 2.7, we carry out empirical implementation of the methods detailed in sections 2.2, 2.3, 2.4 and 2.5 for a simulated data set and real data respectively. The results obtained from the empirical investigation are also discussed. Finally we provide some conclusions in section 2.8.

2.2 The Benchmark PIN Model

Glosten and Milgrom (1985) introduced the sequential information model for a market involving a risk-neutral market maker and two economic agents, namely the informed and uninformed traders. The informed and uninformed traders submit their orders to either buy or sell an asset. The market maker subsequently updates her information about the arrival of informed traders and then posts bid and ask prices that protect her against losses from trading with the informed traders. The market maker continues this Bayesian learning until all possible private information held by informed traders are incorporated into the price of the asset.

In the Glosten and Milgrom (1985) market setting, informed traders are assumed to be competitive and risk-neutral while liquidity traders buy or sell for reasons other than information on the fundamental value of the asset. Easley et al. (1996) proposed a structural model based on the sequential information model of Glosten and Milgrom (1985). Easley et al. (1996) model assume that within any trading day, the number of buyer and seller initiated trades from informed, and uninformed traders are realisations of independent Poisson distributions with mean μ and ϵ respectively. In the model a news event occur at the beginning of each trading day with probability α. With a probability δ, the news event on a "bad news day" will have a negative impact on the value of the asset. Otherwise on a "good news day", there will be a positive impact on the value of the asset.

On any given trading day liquidity traders are present in the market to either buy or sell the asset for reasons other than news. On a bad news day, informed traders expect an adverse effect on the value of the asset and are therefore likely to sell the asset. The total numbers of buy and sell orders on a bad news day are assumed to follow Poisson distributions with means ϵ and $\mu + \epsilon$ respectively. On a good news day, informed traders have an incentive to buy the asset if they judge that the current asset value is under-priced and therefore expect to make gains from their private information. The total number of buy and sell orders on a good news day are Poisson with means $\mu + \epsilon$ and ϵ respectively.

The assumption that informed traders trade on private information implies that informed traders are absent from the market on a day classified as a *no news* day. Thus on a no news day, the total number of buyer and seller initiated trades are each Poisson with mean ϵ. In this model, we do not observe the arrival of investors or the occurrence of a news event. However, we infer them from the observable market data. Figure 2.1 below is a representation of the information and order arrival process for the model

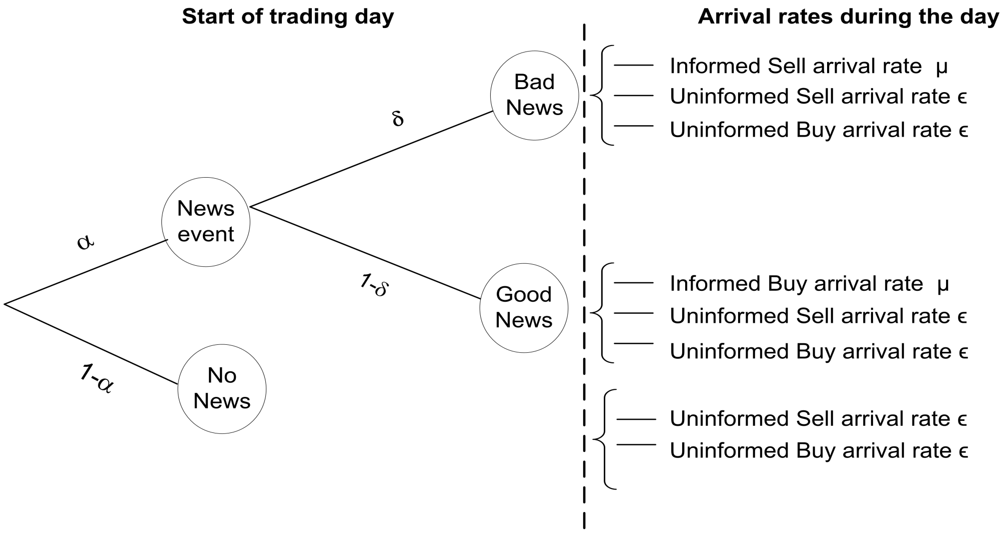


Fig. 2.1 Information and order arrival process in Easley et al. (1996) model

Let B_t and S_t denote the daily number of buyer and seller initiated trades inferred using the Lee and Ready (1991) trade classification algorithm. The density of a buy or sell order on any given trading day t is given as follows

$$P\left(B_t, S_t|\Theta\right) = \omega_1 \frac{e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{S_t}}{S_t!}\frac{\epsilon^{B_t}}{B_t!} + \omega_2 \frac{e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{B_t}}{B_t!}\frac{\epsilon^{S_t}}{S_t!} + \omega_3 \frac{e^{-2\epsilon}\epsilon^{B_t+S_t}}{B_t!\, S_t!}, \tag{2.1}$$

where $\Theta = (\alpha, \delta, \mu, \epsilon)$, $\omega_1 = \alpha\delta$, $\omega_2 = \alpha(1-\delta)$ and $\omega_3 = 1-\alpha$. The corresponding joint likelihood function over a number of trading days $t = 1, \ldots, T$ is

$$L\left(B_t, S_t|\Theta\right) = \prod_{t=1}^{T}\left[\omega_1 \frac{e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{S_t}}{S_t!}\frac{\epsilon^{B_t}}{B_t!} + \omega_2 \frac{e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{B_t}}{B_t!}\frac{\epsilon^{S_t}}{S_t!} + \omega_3 \frac{e^{-2\epsilon}\epsilon^{B_t+S_t}}{B_t!\, S_t!}\right]. \tag{2.2}$$

Easley et al. (1996) estimate the model parameters by maximising equation 2.2 and define PIN as

$$\text{PIN} = \frac{\alpha\mu}{\alpha\mu + 2\epsilon}, \tag{2.3}$$

which is interpreted as the ratio of the expected number of informed trades to the total number of trades. Easley et al. (1996) suggested a minimum of 60 days worth of data to achieve stable estimates of the parameters. Due to the floating-point exception, boundary solution problems of the MLE and other numerical issues of the maximum likelihood estimation reported in papers including Lei and Wu (2005), Vega (2006), Yan and Zhang (2012) and Lin and Ke (2011), we carry out the following factorisation of the log likelihood function

$$\log L\left(B_t, S_t|\Theta\right) = \sum_{t=1}^{T}\Bigg[-2\epsilon + (B_t + S_t)\ln\epsilon + \chi + \ln\left(e^{L_1-\chi} + e^{L_2-\chi} + e^{L_3-\chi}\right)\Bigg], \tag{2.4}$$

prior to maximisation. In equation 2.4, we have the following:

$L_1 = -\mu + S_t\ln(1+\frac{\mu}{\epsilon}) + \ln\omega_1$,

$L_2 = -\mu + B_t\ln(1+\frac{\mu}{\epsilon}) + \ln\omega_2$,

$L_3 = \ln\omega_3$, and

$\chi = \max(L_1, L_2, L_3)$.

Since the PIN is not a parameter estimate but rather a measure calculated based on the parameters any MLE optimser employed will not provide the standard error associated with the estimation of PIN. For this reason, we derive below the asymptotic variance of the PIN. Let $f(\Theta)$ be a multivariate function of the parameter set Θ. The delta method is a useful technique that can be used to derive the asymptotic variance of maximum likelihood estimators. According to the delta method (see Schervish (2012)), the asymptotic variance of $f(\Theta)$ is

$$Var[f(\Theta)] = \nabla f \times \sum\nolimits_{\Theta} \times (\nabla f)', \tag{2.5}$$

where $\nabla f = \frac{\partial f}{\partial\Theta}$ is a vector of the first derivates of the function $f(\Theta)$ with respect to the parameters. The term $\sum_{\Theta}$ is the variance-covariance matrix of the parameters Θ.

The PIN for the Easley et al. (1996) model can be re-written as $=\left(1+\frac{2\epsilon}{\alpha\mu}\right)^{-1}$.

$$PIN = \left(1+\frac{2\epsilon}{\alpha\mu}\right)^{-1}. \tag{2.6}$$

Since news arrival is a Bernoulli variable with parameter α, informed and uninformed trade arrival intensities are poisson random variables with parameters μ and ϵ respectively, we have $\frac{\alpha(1-\alpha)}{n}$, $\frac{\mu}{n}$ and $\frac{\epsilon}{n}$ as their respective asymptotic variances. The variance of PIN in the Easley et al. (1996) model can therefore be estimated via the delta method as follows. The gradient vector is

$$\begin{aligned}\nabla PIN =& \begin{pmatrix}\frac{\partial(PIN)}{\partial\alpha} & \frac{\partial(PIN)}{\partial\mu} & \frac{\partial(PIN)}{\partial\epsilon}\end{pmatrix}\\ =& \left[1+\frac{2\epsilon}{\alpha\mu}\right]^{-2}\begin{pmatrix}\frac{2\epsilon}{\alpha^2\mu} & \frac{2\epsilon}{\alpha\mu^2} & \frac{-2}{\alpha\mu}\end{pmatrix}\end{aligned} \tag{2.7}$$

and the variance-covariance matrix is $n^{-1}\begin{pmatrix}\alpha(1-\alpha) & 0 & 0\\ 0 & \mu & 0\\ 0 & 0 & \epsilon\end{pmatrix}$ assuming independence of parameters. Defining $\Omega = [1+\frac{2\epsilon}{\alpha\mu}]^{-2}$ and using equation 2.5 we have

$$Var(PIN) = n^{-1}\Omega\begin{pmatrix}\frac{2\epsilon}{\alpha^2\mu} & \frac{2\epsilon}{\alpha\mu^2} & \frac{-2}{\alpha\mu}\end{pmatrix}\begin{pmatrix}\alpha(1-\alpha) & 0 & 0\\ 0 & \mu & 0\\ 0 & 0 & \epsilon\end{pmatrix}\begin{pmatrix}\frac{2\Omega\epsilon}{\alpha^2\mu}\\ \frac{2\Omega\epsilon}{\alpha\mu^2}\\ \frac{-2\Omega}{\alpha\mu}\end{pmatrix}$$

and show that

$$Var(PIN) = \frac{4\Omega^2}{n}\left(\frac{\epsilon}{\alpha^2\mu^2}+\frac{\epsilon^2\mu}{\alpha^2\mu^4}+\frac{\alpha(1-\alpha)\epsilon^2}{\alpha^4\mu^2}\right) \tag{2.8}$$

We compare results of the maximisation of equation 2.4 with the estimates of the

Bayesian estimation methods which we detail in the next section.

2.3 Bayesian Inference Of Benchmark Model

Our goal is to learn about PIN and its underlying parameters from observed transaction data. The maximum likelihood estimation approach taken in the literature assumes that the model parameters are unknown but fixed. However, in a Bayesian setting, we assume that model parameters are random and unknown. The theory behind the PIN model relates to a market maker who sets quotes and updates her knowledge about the trading behaviour of other market participants. The methodology is well suited for the estimation of PIN and its parameters since it allows for the updating of knowledge about model parameters using data from the trading process. In Bayesian inference, we express the uncertainty about the unknown model parameters through the rules of probability. We achieve this through the Bayes' rule which states that the probability of the parameter set Θ given the observed data is

$$\begin{aligned} p(\Theta|B_t, S_t) =& \frac{p(\Theta)p(B_t, S_t|\Theta)}{p(B_t, S_t)} \\ &\propto p(\Theta)p(B_t, S_t|\Theta). \end{aligned} \tag{2.9}$$

The denominator in 2.9, $p(B_t, S_t) = \int p(B_t, S_t|\Theta)d\Theta$, is a normalising constant. The term $p(\Theta)$, referred to as the *prior density* is not dependent on the data. It is used to express the prior knowledge and uncertainty about the model parameters before observing the data. The term $p(B_t, S_t|\Theta)$, usually referred to as the *likelihood function* is the probability density function of the data conditional on the model parameters. In Bayesian inference, the primary object of interest is $p(\Theta|B_t, S_t)$ which is referred to as the *posterior density.* It summarises our updated knowledge of the model parameters having observed the data. It pools together information from the prior and likelihood to provide the updated information. From the posterior density, we can compute point estimates like the mean, mode and credible intervals for the model parameters. In this chapter, we employ two Bayesian Markov Chain Monte Carlo (MCMC) methods, namely the Gibbs Sampler and the Metropolis Hastings Algorithm to infer the parameters of the Easley et al. (1996) model. These methods are capable of exploring the entire support of the posterior distribution of the model parameters.

Given daily buyer and seller initiated trades B_t and S_t, and defining D_t as

$$D_t = \begin{cases} 1, & \text{bad news day with probability} \quad \omega_1 = \alpha\delta \\ 2, & \text{good news day with probability} \quad \omega_2 = \alpha(1-\delta) \\ 3, & \text{no news day with probability} \quad \omega_3 = 1-\alpha, \end{cases}$$

then we can write the following conditional buy and sell order distributions

$$\begin{array}{lll} S_t|D_t = 1 \sim Pn\,(\mu+\epsilon) & S_t|D_t = 2 \sim Pn\,(\epsilon) & S_t|D_t = 3 \sim Pn\,(\epsilon) \\ B_t|D_t = 1 \sim Pn\,(\epsilon) & B_t|D_t = 2 \sim Pn\,(\mu+\epsilon) & B_t|D_t = 3 \sim Pn\,(\epsilon), \end{array}$$

where $Pn\,(.)$ [1] is the probability mass function of a Poisson random variable.

2.3.1 *Method 1 : Gibbs Sampler*

The latent variable D_t is the process which determines the composition of traders in the market on a daily basis. This underlying process is unobservable and hence is inferred from transaction data, as a missing data problem within the Bayesian framework. Since we do not observe trader arrival rates, good, bad or no news days, we employ the data augmentation procedure to impute these missing observations. We do this by directly sampling from the posterior distribution of D_t conditional on the available data. In this section, we employ the theory of data augmentation to derive the density function of buy and sell trades.

[1] $Pn(x;\theta) = \frac{e^{-\theta}\theta^x}{x!}$

Density of buy and sell trades on a bad news day ($D_t = 1$)

Sell trades initiated by informed and liquidity traders on a bad news day are denoted as S_t^i and S_t^u respectively. These are assumed to follow Poisson distributions with means μ and ϵ respectively. Hence the total daily seller initiated trades, $S_t = S_t^i + S_t^u$ is also Poisson with mean $\mu + \epsilon$. Given the total number of sell orders S_t, the number of informed seller initiated trades S_t^i are binomial with S_t trials and probability ${}^{\mu}/_{\mu+\epsilon}$. Uninformed seller initiated trades are determined as $S_t^u = S_t - S_t^i$. All buyer initiated trades on a bad news day are made by liquidity traders. The distributions of buyer and seller initiated trades are given as follows:

$$S_t|D_t = 1 \sim Pn\,(\mu + \epsilon), \qquad S_t^i|S_t, D_t = 1 \sim Bin\Big(S_t,\ \tfrac{\mu}{\mu+\epsilon}\Big),$$
$$B_t|D_t = 1 \sim Pn\,(\epsilon),$$

where $Bin\,(.)$ [2] is the probability mass function of the Binomial random variable. The probability of a buy or sell initiated trade, on a bad news day is

$$\begin{aligned}
f_1\left(B_t, S_t, S_t^i, \Theta\right) =& P\left(B_t, S_t, S_t^i|D_t = 1, \Theta\right) \\
=& P\left(B_t|D_t = 1, \Theta\right) P\left(S_t, S_t^i|D_t = 1, \Theta\right) \\
=& P\left(B_t|D_t = 1, \Theta\right) P\left(S_t^i|S_t, D_t = 1, \Theta\right) P\left(S_t|D_t = 1, \Theta\right) \\
=& \frac{e^{-\epsilon}\epsilon^{B_t}}{B_t!}\binom{S_t}{S_t^i}\left(\frac{\mu}{\mu+\epsilon}\right)^{S_t^i}\left(\frac{\epsilon}{\mu+\epsilon}\right)^{S_t - S_t^i}\frac{(\mu+\epsilon)^{S_t}\,e^{-(\mu+\epsilon)}}{S_t!} \\
=& \binom{S_t}{S_t^i}\frac{e^{-(\mu+2\epsilon)}\epsilon^{B_t}}{B_t!\,S_t!}\mu^{S_t^i}\epsilon^{S_t - S_t^i} \\
=& \frac{e^{-\mu}\mu^{S_t^i}}{S_t^i!}\frac{e^{-\epsilon}\epsilon^{B_t}}{B_t!}\frac{e^{-\epsilon}\epsilon^{S_t - S_t^i}}{(S_t - S_t^i)!}.
\end{aligned} \tag{2.10}$$

Density of buy and sell trades on a good news day ($D_t = 2$)

The total number of buyer initiated trades B_t on a good news day comprises of buyer initiated trades B_t^i and B_t^u made by informed and uninformed traders respectively.

[2] $Bn(n; r; \theta) = \binom{n}{\theta}\theta^r(1-\theta)^{n-r}$

These buyer initiated trades are assumed to be Poisson with mean μ and ϵ, hence B_t is Poisson with $\mu{+}\epsilon$. Conditional on the total daily buyer initiated trades B_t, the number of buyer initiated trades from informed trades B_t^i are binomial distributions with B_t trials and probability ${}^{\mu}/_{\mu+\epsilon}$. The uninformed buyer initiated trade is calculated as $B_t^u = B_t - B_t^i$. All seller initiated trades on a good news day are made by liquidity traders. Thus we have $B_t|D_t = 2 \sim Pn\left(\mu+\epsilon\right)$, $B_t^i|B_t, D_t = 2 \sim Bin\Big(B_t, \frac{\mu}{\mu+\epsilon}\Big)$ and $S_t|D_t = 2 \sim Pn\left(\epsilon\right)$ as the distributions of the buyer and seller initiated trades on a good news day. The probability density of a buy and sell trades, on a good news day is given below

$$\begin{aligned}
f_2\left(B_t, S_t, B_t^i, \Theta\right) =& P\left(B_t, B_t, B_t^i|D_t = 2, \Theta\right) \\
=& P\left(S_t|D_t = 2, \Theta\right) P\left(B_t^i|B_t, D_t = 2, \Theta\right) P\left(B_t|D_t = 2, \Theta\right) \\
=& \frac{e^{-\epsilon}\epsilon^{S_t}}{S_t!}\binom{B_t}{B_t^i}\left(\frac{\mu}{\mu+\epsilon}\right)^{B_t^i}\left(\frac{\epsilon}{\mu+\epsilon}\right)^{B_t-B_t^i}\frac{\left(\mu+\epsilon\right)^{B_t}e^{-(\mu+\epsilon)}}{B_t!} \\
=& \frac{e^{-\mu}\mu^{B_t^i}}{B_t^i!}\frac{e^{-\epsilon}\epsilon^{S_t}}{S_t!}\frac{e^{-\epsilon}\epsilon^{B_t-B_t^i}}{(B_t-B_t^i)!}.
\end{aligned} \tag{2.11}$$

Density of buy and sell trades on a no news day ($D_t = 3$)

Easley et al. (1996) assume that informed traders do not trade on no news days. Hence the total number of buyer and seller initiated trades are made solely by liquidity traders. The arrivals are independent Poisson distributions $B_t|D_t = 3 \sim Pn\left(\epsilon\right)$ and $S_t|D_t = 3 \sim Pn\left(\epsilon\right)$ respectively. The probability density of a buyer or seller initiated trade on a no news day is

$$\begin{aligned}
f_3\left(B_t, S_t, \Theta\right) =& P\left(B_t, S_t|D_t = 3, \Theta\right) \\
=& P\left(S_t|D_t = 3, \Theta\right) P\left(B_t|\Theta\right) P\left(B_t|D_t = 3, \Theta\right) \\
=& \frac{e^{-2\epsilon}\epsilon^{B_t+S_t}}{B_t!\,S_t!}.
\end{aligned} \tag{2.12}$$

The Poisson mixture assumption underlying the model can be seen in equations 2.10, 2.11 and 2.12. We define the following indicator random variable $d_{t,j} = \mathbf{1}_{\{D_t=j\}}$, for $j = 1, 2, 3$. Putting equations 2.10, 2.11 and 2.12 together, the density function of buy and sell orders is

$$\begin{aligned} P\left(B_t, S_t|\Theta, D_t\right) &= \left[f_1\left(B_t, S_t, \Theta\right)\right]^{d_{t,1}} \left[f_2\left(B_t, S_t, \Theta\right)\right]^{d_{t,2}} \left[f_3\left(B_t, S_t, \Theta\right)\right]^{d_{t,3}} \\ &= \left[\frac{e^{-\mu}\mu^{S_t^i}}{S_t^i!}\frac{e^{-\epsilon}\epsilon^{B_t}}{B_t!}\frac{e^{-\epsilon}\epsilon^{S_t-S_t^i}}{(S_t-S_t^i)!}\right]^{d_{t,1}} \left[\frac{e^{-\mu}\mu^{B_t^i}}{B_t^i!}\frac{e^{-\epsilon}\epsilon^{S_t}}{S_t!}\frac{e^{-\epsilon}\epsilon^{B_t-B_t^i}}{(B_t-B_t^i)!}\right]^{d_{t,2}} \left[\frac{e^{-2\epsilon}\epsilon^{B_t+S_t}}{B_t!\,S_t!}\right]^{d_{t,3}}. \end{aligned} \tag{2.13}$$

Posterior Density And Full Conditional Distributions

Since the probability of news arrival α and its effect δ are both positive values that lie strictly in the interval (0, 1) we choose beta distributions for their prior distributions. Likewise we choose gamma prior distributions for the positive parameters μ and ϵ. Finally, we choose a Dirichlet prior for the type of day classifier D_t. The prior distributions for parameter set $\Theta = (\alpha, \delta, \mu, \epsilon, D_t)$ are

$P(\alpha|\rho, \phi) = \frac{1}{\mathcal{B}eta(\rho,\varepsilon)}\alpha^{\rho-1}(1-\alpha)^{\phi-1}$, $\quad P(\epsilon|\gamma_1, \beta_1) = \frac{\beta_1^{\gamma_1}}{\Gamma(\beta_1)}\lambda_s^{\gamma_1-1}e^{-\beta_1\lambda_s}$,

$P(\delta|\nu, \tau) = \frac{1}{\mathcal{B}eta(\nu,\tau)}\delta^{\nu-1}(1-\delta)^{\tau-1}$, $\quad D_t \sim \mathbf{D}irichlet(\pi_1, \pi_2, \pi_3)$ and

$P(\mu|\gamma_0, \beta_0) = \frac{\beta_0^{\gamma_0}}{\Gamma(\gamma_0)}\mu^{\gamma_0-1}e^{-\beta_0\mu}$.

With these conjugate prior distributions the resulting posterior distributions will have kernels which are proportional to standard probability distributions. The Gibbs Sampler can then be easily applied to sample from the posterior distributions. We set each of the hyper-parameters ρ, ϕ, ν, τ, γ_0, γ_1, β_0, β_1 to the value 1 and $\pi_1 = \pi_2 = \pi_3 = 1/3$. This means that α and δ can take on any number between zero and one with probability 1. The priors for ϵ and μ are informative since their respective means and variances are equal to 1. The hyper-parameter choices will have little influence on the parameter estimates since after a large enough iterations of the

Gibbs Sampler, the markov chain converges to the true parameter. From Bayes' theorem, the posterior density for the parameter set $\Theta = (\alpha, \delta, \mu, \epsilon)$ is proportional to the product of the likelihood and prior. This is given as

$$\begin{aligned}
&P\left(\Theta|B_t, S_t\right) \propto P\left(\Theta\right) \prod_{t=1}^{T} \left[P\left(B_t, S_t|D_t, \Theta\right) P\left(D_t|\Theta\right) \right] \\
&= \mu^{\gamma_0-1} e^{-\beta_0\mu} \epsilon^{\gamma_1-1} e^{-\beta_1\epsilon} \alpha^{\rho-1}(1-\alpha)^{\phi-1}\delta^{\nu-1}(1-\delta)^{\tau-1} \left[(\alpha\delta)^{T_1}(\alpha(1-\delta))^{T_2}(1-\alpha)^{T_3} \right] \\
&\times \prod_{t=1}^{T} \left[\frac{e^{-\mu}\mu^{S_t^i}}{S_t^i!} \frac{e^{-\epsilon}\epsilon^{B_t}}{B_t!} \frac{e^{-\epsilon}\epsilon^{S_t-S_t^i}}{(S_t-S_t^i)!} \right]^{d_{t,1}} \left[\frac{e^{-\mu}\mu^{B_t^i}}{B_t^i!} \frac{e^{-\epsilon}\epsilon^{S_t}}{S_t!} \frac{e^{-\epsilon}\epsilon^{B_t-B_t^i}}{(B_t-B_t^i)!} \right]^{d_{t,2}} \left[\frac{e^{-2\epsilon}\epsilon^{B_t+S_t}}{B_t!\, S_t!} \right]^{d_{t,3}}.
\end{aligned} \tag{2.14}$$

The full conditional distributions of the parameters $\Theta = (\alpha, \delta, \mu, \epsilon)$ are needed for the Gibbs Sampler. From equation 2.14 we have the following full conditional densities for the parameters

$$\mu \sim \mathcal{G}a\left(\gamma_0 + \sum_{t=1}^{T} [(S_t^i)^{d_{t,1}} + (B_t^i)^{d_{t,2}}],\ T_1 + T_2 + \beta_0 \right) \tag{2.15a}$$

$$\delta \sim \mathcal{B}e\left(\nu + T_1,\ \ T_2 + \tau \right) \tag{2.15b}$$

$$\alpha \sim \mathcal{B}e\left(\rho + T_1 + T_2,\ \ T_3 + \phi \right) \tag{2.15c}$$

$$\epsilon \sim \mathcal{G}a\left(\gamma_1 + \sum_{t=1}^{T} [B_t^{d_1} + S_t^{d_1} - (S_t^i)^{d_1} + S_t^{d_2} + B_t^{d_2} - (B_t^i)^{d_2} + B_t^{d_3} + S_t^{d_3}],\ 2T + \beta_1 \right), \tag{2.15d}$$

where T_1, T_2 and T_3 are the number of good, bad and no news days respectively such that $T = T_1 + T_2 + T_3$.

Gibbs Sampling Procedure

The algorithm recursively draw samples from the full conditional posterior distributions where the most recent values of the parameters are used in the simulation. The procedure is as follows:

- Choose an arbitrary initial type of day (good, bad and no news day) classification for the (B_t, S_t). Denote the initial classification as $\mathbf{D_t}^{(0)}$.
- Set initial values for the parameter set Θ. Denote it as $\Theta^{(0)}{=}\Big(\alpha^{(0)}, \delta^{(0)}, \epsilon^{(0)}, \mu^{(0)}\Big)$.
- Repeat for $\quad k = 1$ to G sweeps
  - Sample $B_t^{i(k)}|D_t^{(k-1)} \sim Bin\Big(B_t,\ \frac{\mu^{(k-1)}}{\mu^{(k-1)}+\epsilon^{(k-1)}}\Big), \quad t = 1, \ldots, T$
  - Sample $S_t^{i(k)}|D_t^{(k-1)} \sim Bin\Big(S_t,\ \frac{\mu^{(k-1)}}{\mu^{(k-1)}+\epsilon^{(k-1)}}\Big), \quad t = 1, \ldots, T$
  - Update $\mu^{(k)}|\alpha^{(k-1)}, \epsilon^{(k-1)}, \delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}$
  - Update $\alpha^{(k)}|\mu^{(k)}, \epsilon^{(k-1)}, \delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}$
  - Update $\epsilon^{(k)}|\mu^{(k)}, \alpha^{(k)}, \delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}$
  - Update $\delta^{(k)}|\mu^{(k)}, \alpha^{(k)}, \epsilon^{(k)}, B_t^{i(k)}, S_t^{i(k)}$
  - Compute $L_1 = \log \omega_1^{(k)} - \left(\mu^{(k)} + 2\epsilon^{(k)}\right) + B_t \log \epsilon^{(k)} + S_t \log\left(\mu^{(k)} + \epsilon^{(k)}\right)$
  - Compute $L_2 = \log \omega_2^{(k)} - \left(\mu^{(k)} + 2\epsilon^{(k)}\right) + S_t \log \epsilon^{(k)} + B_t \log\left(\epsilon^{(k)} + \mu^{(k)}\right)$
  - Compute $L_3 = \log \omega_3^{(k)} - 2\epsilon^{(k)} + (S_t + B_t) \log \epsilon^{(k)}$
  - compute $\chi = \max\left(L_1, L_2, L_3\right)$
  - Compute $p_1 = \frac{e^{L_1-\chi}}{\sum\limits_{j=1}^{3} e^{L_j-\chi}}$, $p_2 = \frac{e^{L_2-\chi}}{\sum\limits_{j=1}^{3} e^{L_j-\chi}}$ and $p_3 = \frac{e^{L_3-\chi}}{\sum\limits_{j=1}^{3} e^{L_j-\chi}}$
  - Update $\mathbf{D_t}^{(k)}$, the classification of (B_t, S_t) by sampling from the multinomial distribution with probability $\Big(p_1,\ p_2,\ p_3\Big)$,

where p_1, p_2 and p_3 are the probabilities that at the beginning of the trading day there will be bad news, good news and no news respectively.

2.3.2 *Method 2: Metropolis-Hastings Algorithm*

The Metropolis Hastings algorithm is a MCMC algorithm for drawing samples from the posterior distribution of high dimensional parameter and intractable complex model problems. The algorithm is used to draw samples of the parameter set $\Theta' = (\alpha', \delta', \mu', \epsilon')$ from an approximating distribution which has the same support as the posterior density. The approximating distribution which we denote as $q\Big(\Theta', \Theta^{(t-1)}\Big)$ is referred to as a *proposal density.*

The algorithm involves two basic steps. Firstly a draw from the proposal density is obtained. Secondly, the draw is either retained or rejected. Details of the algorithm are summarised as follows

1. Initialise the algorithm with values $\Theta^{(0)}$ from the parameter space of Θ.

2. At iteration t, a draw Θ' is taken from the proposal density $q\Big(\Theta', \Theta^{(t-1)}\Big)$ where $\Theta^{(t-1)}$ is the value of the parameter in the previous step.

3. The new draw is accepted with probability $\min\left\{1, \frac{\pi(\Theta')q(\Theta^{(t-1)},\Theta')}{\pi(\Theta^{(t-1)})q(\Theta',\Theta^{(t-1)})}\right\}$, where $\pi(\Theta)$ is the posterior density.

4. Steps 2 and 3 are repeated for a large number of iterations.

Now we proceed to derive the posterior density needed for the sampling. For easy comparison of results we use the same conjugate prior distributions and hyper-parameters for $\alpha, \delta, \mu, \epsilon$ and D_t which were chosen for the Gibbs Sampler. The choices as indicated earlier is to ensure that we have the draws not falling on the boundary of the respective parameter space. The joint posterior density of buyer and seller initiated trades is given as

$$
\begin{aligned}
&P\left(\Theta, D_t|B_t, S_t\right) \propto P(\alpha|\rho,\phi)P(\delta|\nu,\tau)P(\mu|\gamma_0,\beta_0)P(\epsilon|\gamma_1,\beta_1)\\
&\times \prod_{t=1}^{T}\left[\alpha\delta e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{S_t}\epsilon^{B_t} + \alpha(1-\delta)e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{B_t}\epsilon^{S_t} + (1-\alpha)e^{-2\epsilon}\epsilon^{B_t+S_t}\right]\\
&= \alpha^{\rho-1}(1-\alpha)^{\phi-1}\delta^{\nu-1}(1-\delta)^{\tau-1}\mu^{\gamma_0-1}e^{-\beta_0\mu}\epsilon^{\gamma_1-1}e^{-\gamma_1\epsilon}\\
&\times \prod_{t=1}^{T}\left[\alpha\delta e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{S_t}\epsilon^{B_t} + \alpha(1-\delta)e^{-(\mu+2\epsilon)}\left(\mu+\epsilon\right)^{B_t}\epsilon^{S_t} + (1-\alpha)e^{-2\epsilon}\epsilon^{B_t+S_t}\right],
\end{aligned}
\tag{2.16}
$$

Taking logarithm of the posterior density in equation 2.16 we have

$$
\begin{aligned}
\ln\left(P\left(\Theta|B_t, S_t\right)\right) = &\sum_{t=1}^{T}\ln\left[e^{L_1-\chi} + e^{L_2-\chi} + e^{L_3-\chi}\right]\\
&+\sum_{t=1}^{T}\left[-2\epsilon + (B_t+S_t)\ln(\epsilon) + \chi\right]\\
&+(\gamma_0-1)\ln\left(\mu\right) - \beta_0\mu + (\gamma_1-1)\ln\left(\epsilon\right) - \beta_1\epsilon + (\rho-1)\ln\left(\alpha\right)\\
&+(\phi-1)\ln\left(\alpha\right) + (\nu-1)\ln\left(\delta\right) + (\tau-1)\ln\left(\delta\right),
\end{aligned}
$$

where

$L_1 = -\mu + S_t\ln(1+\frac{\mu}{\epsilon}) + \ln(\alpha\delta)$,

$L_2 = -\mu + B_t\ln(1+\frac{\mu}{\epsilon}) + \ln[\alpha(1-\delta)]$,

$L_3 = \ln(1-\alpha)$ and

$\chi = \max(L_1, L_2, L_3)$.

We use a random walk Metropolis-Hastings algorithm with standard Gaussian innovations to draw samples from the posterior distribution of the parameters. The random walk proposal is $\Theta' = \Theta^{(t-1)} + \varepsilon$, where $\varepsilon_t \sim \mathcal{N}\left(0, \xi\right)$. Since the random walk proposal density is symmetric, the acceptance probability simplifies to

$$
\text{acceptance probability} = \min\left\{1, \frac{P(B_t, S_t|\Theta')P(\Theta')}{P(B_t, S_t|\Theta)P(\Theta)}\right\}. \tag{2.17}
$$

We start off the Bayesian estimation with arbitrary initial values for algorithm. The Markov chain initially explores regions of the parameter space around the initial values and finally converge to most probably parameter space. However, including samples around the initial values in the posterior mean calculation can produce substantial bias in the mean estimate. The practice of discarding an initial portion of a Markov chain sample so that the effect of initial values on the posterior inference is minimised is known as *burn-in* period. To improve the convergence of the Markov chain, we implemented the Adaptive Metropolis-Hastings algorithm (AMH) developed by Haario et al. (2001) which we briefly describe below.

The Haario et al. (2001) AMH algorithm

1. For each element of the parameter set $\Theta = (\alpha, \delta, \mu, \epsilon)$, set initial values $\Theta^{(0)}$, $\xi^{(0)}$, MCMC samples G, burnin n_0, and t_0
2. **For** $t = 1, 2, \ldots$ **do**
3. Sample $\Theta' = \Theta^{(t)} + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}\left(0, \xi^{(t)}\right)$
4. Accept the next iterate $\Theta^{(t+1)} = \Theta'$ with probability given in equation 2.17
5. Compute

$$\xi^{(t+1)} = \begin{cases} \xi^{(0)} & \text{if } t \leq t_0, \\ s_d \times \varphi \times I_d + \frac{s_d}{t-1} \times \left[\sum\limits_{j=1}^{t} \Theta^{(j)} \Theta^{(j)T} - \frac{\left(\sum\limits_{j=1}^{t} \Theta^{(j)}\right)\left(\sum\limits_{j=1}^{t} \Theta^{(j)}\right)^T}{t} \right], & \text{if } t > t_0 \end{cases}$$

   where φ is a small positive constant, I_d is a d-dimensional identity matrix and s_d is a scale parameter
6. **end for**
7. Collect samples $\Theta^{(n_0+1)}, \ldots, \Theta^{(n_0+G)}$

The scaling parameter $s_d = {}^{2.4^2}/_d$ where d is the dimension of Θ. This value was proposed in Gelman et al. (1996) to optimise the mixing properties of the Metropolis algorithm.

2.4 Extension Of The Benchmark PIN Model

Buyer and seller initiated trades made by liquidity traders in Easley et al. (1996) were assumed to be Poisson with the same arrival rates. Easley et al. (2002) relaxed this assumption since in reality, liquidity traders who want to buy or sell an asset do not arrive at the market at the same rate. In this model the means of the liquidity trader buyer and seller initiated trade distributions are λ_b and λ_s respectively. All other assumptions in the previous model remain unchanged.

Since order arrivals on a bad news day are assumed to follow independent Poisson distributions, the total number of buyer and seller initiated trades on a bad news day are Poisson with means λ_b and $\mu + \lambda_s$ respectively. On a good news day, the total number of buyer and seller initiated trades are also Poisson with parameters $\mu + \lambda_b$ and λ_s respectively. Likewise, on a no news day the total number of buyer and seller initiated trades are Poisson with parameters λ_b and λ_s respectively.

Figure 2.2 is a representation of the information and order arrival process for the model.

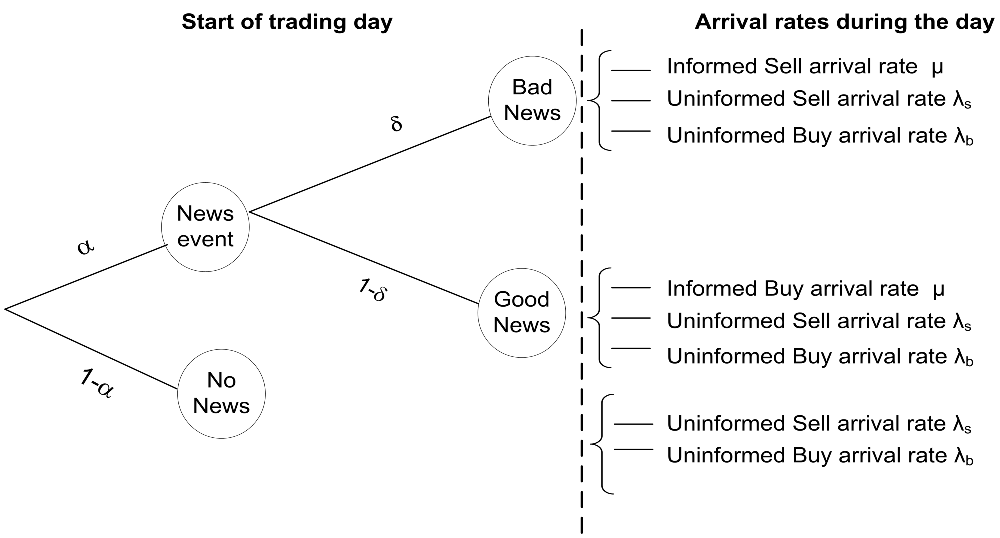


Fig. 2.2 Information and order arrival process in Easley et al. (2002) model

Using the definition of the variable D_t in section 2.3 the probability of a buyer or seller initiated trade is

$$\begin{aligned} P\left(B_t, S_t|\Theta\right) = {} & \omega_1 \frac{e^{-(\mu+\lambda_s)}\left(\mu+\lambda_s\right)^{S_t}}{S_t!}\frac{e^{-\lambda_b}\left(\lambda_b\right)^{B_t}}{B_t!} + \omega_3 \frac{e^{-\lambda_b}\left(\lambda_b\right)^{B_t}}{B_t!}\frac{e^{-\lambda_s}\left(\lambda_s\right)^{S_t}}{S_t!} \\ & + \omega_2 \frac{e^{-(\mu+\lambda_b)}\left(\mu+\lambda_b\right)^{B_t}}{B_t!}\frac{e^{-\lambda_s}\left(\lambda_s\right)^{S_t}}{S_t!}, \end{aligned} \tag{2.18}$$

where $\Theta = (\alpha, \delta, \mu, \lambda_b, \lambda_s)$. The corresponding joint likelihood function for buy or sell trades is

$$\begin{aligned} L\left(B_t, S_t|\Theta\right) = \prod_{t=1}^{T} \Bigg[& \omega_1 \frac{e^{-(\mu+\lambda_s)}\left(\mu+\lambda_s\right)^{S_t}}{S_t!}\frac{e^{-\lambda_b}\left(\lambda_b\right)^{B_t}}{B_t!} + \omega_3 \frac{e^{-\lambda_b}\left(\lambda_b\right)^{B_t}}{B_t!}\frac{e^{-\lambda_s}\left(\lambda_s\right)^{S_t}}{S_t!} \\ & + \omega_2 \frac{e^{-(\mu+\lambda_b)}\left(\mu+\lambda_b\right)^{B_t}}{B_t!}\frac{e^{-\lambda_s}\left(\lambda_s\right)^{S_t}}{S_t!} \Bigg]. \end{aligned} \tag{2.19}$$

Maximising equation 2.19, we estimate the parameter set Θ. In this model the probability of information based trading is calculated as

$$\text{PIN} = \frac{\alpha\mu}{\alpha\mu + \lambda_s + \lambda_b}. \tag{2.20}$$

As discussed in the preceding section, there are challenges with the optimisation of equation 2.19. Defining $M_t = 0.5[\min(B_t, S_t) + \max(B_t, S_t)]$, $x_s = \frac{\lambda_s}{\mu+\lambda_s}$, and $x_b = \frac{\lambda_b}{\mu+\lambda_b}$, Easley et al. (2002) used the following factorisation of the joint likelihood function

$$\begin{aligned} L(B_t, S_t|\Theta) = {} & \sum_{t=1}^{T} \Big[-\lambda_s - \lambda_b + M_t(\ln x_b + \ln x_s) + B_t \ln(\mu+\lambda_b) + S_t \ln(\mu+\lambda_s) \Big] \\ & + \sum_{t=1}^{T} \ln \Big[\alpha\delta e^{-\mu} x_b^{B_t - M_t} x_s^{-M_t} + \alpha(1-\delta) e^{-\mu} x_s^{S_t - M_t} x_b^{-M_t} + (1-\alpha) x_s^{S_t - M_t} x_b^{B_t - M_t} \Big], \end{aligned} \tag{2.21}$$

to enhance the maximisation. However, Lin and Ke (2011) argue that the above factorisation which we henceforth refer to as *EHO2002 factorisation* gives downward biased estimates of PIN due to the floating-point exceptions of optimisation routines in software packages. They also suggest that their factorisation below, which we also avoids the numerical problems in the EHO2002 factorisation. The factoristion which we refer to as *LinKe2011 factorisation* is given as

$$
\begin{aligned}
L(B_t, S_t) = \sum_{t=1}^{T} \Big[-\lambda_s - \lambda_b + B_t \ln(\mu + \lambda_b) + S_t \ln(\mu + \lambda_s) + e_{maxt} \Big] \\
+ \sum_{t=1}^{T} \ln \Big[\alpha\delta e^{e_1 - e_{maxt}} + \alpha(1-\delta) e^{e_2 - e_{maxt}} + (1-\alpha) e^{e_3 - e_{maxt}} \Big],
\end{aligned}
\tag{2.22}
$$

where

$$
\begin{aligned}
e_1 &= -\mu - S_t \ln(1 + \tfrac{\mu}{\lambda_s}), \\
e_2 &= -\mu - B_t \ln(1 + \tfrac{\mu}{\lambda_b}), \\
e_3 &= -B_t \ln(1 + \tfrac{\mu}{\lambda_b}) - S_t \ln(1 + \tfrac{\mu}{\lambda_s}) \quad \text{and} \\
e_{maxt} &= \max(e_1, e_2, e_3).
\end{aligned}
$$

According to Yan and Zhang (2012), the LinKe2011 factorisation is also not fully immune to the corner solutions problem. The authors suggest an adhoc approach of selecting initial values for the maximisation of the likelihood function. Using the following empirical first moments of buyer and seller initiated trades

$$E(B) = \alpha(1-\delta)\mu + \lambda_b \tag{2.23a}$$

$$E(S) = \alpha\delta\mu + \lambda_s, \tag{2.23b}$$

they propose the selection of the initial values of the model parameters as follows. Firstly they divide the interval $[0, 1]$ into equally spaced sub-intervals and choose equidistant values for α and δ from these sub-intervals as the initial values. They argue that since $\alpha(1-\delta)\mu$ in equation 2.23a is always positive, the empirical first moment $\bar{B}$ which is an estimate for $E(B)$ is always greater than λ_b. Hence initial values of

λ_b should be fractions γ of $E(B)$. Choosing $\gamma = (0.1, 0.3, 0.5, 0.7, 0.9)$ and solving equations 2.23a and 2.23b simultaneously they obtain $\mu = \frac{\bar{B}-\lambda_b}{\alpha(1-\delta)}$ and $\lambda_s = \bar{S} - \alpha\delta\mu$ as the initial values for μ and λ_s respectively. Combining α, δ and γ yields 125 sets of initial values. This large number of initial values did not completely solve the boundary solutions problem either.

For numerical implementation, we use the following factorised version of the log likelihood function

$$\begin{aligned}\ln L\left(B_t, S_t|\Theta\right) =& \sum_{t=1}^{T}\Big[-\lambda_b - \lambda_s + B_t \ln\lambda_b + S_t \ln\lambda_s + \chi\Big]\\ &+\sum_{t=1}^{T}\ln\Big[e^{L_1-\chi} + e^{L_2-\chi} + e^{L_3-\chi}\Big],\end{aligned} \tag{2.24}$$

where

$L_1 = -\mu + S_t \ln(1 + \frac{\mu}{\lambda_s}) + \ln(\alpha\delta)$,

$L_2 = -\mu + B_t \ln(1 + \frac{\mu}{\lambda_b}) + \ln[\alpha(1-\delta)]$,

$L_3 = \ln(1-\alpha)$ and

$\chi = \max(L_1, L_2, L_3)$.

The Bayesian estimation method does not require any adhoc selection of initial values. Any starting values of the parameters will yield feasible solutions which exclude boundary solutions for α and δ. In the maximum likelihood estimation of PIN from the parameters of equation 2.24, we would require the asymptotic variance of the PIN. Similar to what did in section 2.2, we derive the asymptotic variance of PIN for Easley et al. (2002) model can be estimated via the delta method as follows. The PIN for this model can be re-written as

$$PIN = \left[1 + \frac{(\lambda_s + \lambda_b)}{\alpha\mu}\right]^{-1} \tag{2.25}$$

Since news arrival is a Bernoulli variable with parameter α, informed and uninformed trade arrival intensities are poisson random variables with parameters μ and ϵ respectively, we have $\frac{\alpha(1-\alpha)}{n}$, $\frac{\mu}{n}$ $\frac{\lambda_s}{n}$ and $\frac{\lambda_b}{n}$ as their respective asymptotic variances. The variance of PIN in the Easley et al. (2002) model can therefore be estimated via the delta method as follows. The vector of derivatives of the PIN with respect to the parameters of the model

$$\begin{aligned}\nabla PIN =& \Big(\frac{\partial(PIN)}{\partial\alpha}\ \frac{\partial(PIN)}{\partial\mu}\ \frac{\partial(PIN)}{\partial\lambda_s}\ \frac{\partial(PIN)}{\partial\lambda_b}\Big)\\ =&\left(1+\frac{(\lambda_s+\lambda_b)}{\alpha\mu}\right)^{-2}\left[\frac{\lambda_s+\lambda_b}{\alpha^2\mu}\ \frac{\lambda_s+\lambda_b}{\alpha\mu^2}\ \frac{-1}{\alpha\mu}\ \frac{-1}{\alpha\mu}\right]\end{aligned} \tag{2.26}$$

and the variance-covariance matrix is $n^{-1}\begin{pmatrix}\alpha(1-\alpha) & 0 & 0 & 0\\ 0 & \mu & 0 & 0\\ 0 & 0 & \lambda_s & 0\\ 0 & 0 & 0 & \lambda_b\end{pmatrix}$ assuming independence of parameters. Defining $\Omega = \left[1+\frac{(\lambda_s+\lambda_b)}{\alpha\mu}\right]^{-2}$ and using equation 2.5, the asymptotic variance of PIN is

$$Var(PIN) = \frac{\Omega}{n}\Big(\frac{\lambda_s+\lambda_b}{\alpha^2\mu}\ \frac{\lambda_s+\lambda_b}{\alpha\mu^2}\ \frac{-1}{\alpha\mu}\ \frac{-1}{\alpha\mu}\Big)\begin{pmatrix}\alpha(1-\alpha) & 0 & 0 & 0\\ 0 & \mu & 0 & 0\\ 0 & 0 & \lambda_s & 0\\ 0 & 0 & 0 & \lambda_b\end{pmatrix}\begin{pmatrix}\Omega\frac{\lambda_s+\lambda_b}{\alpha^2\mu}\\ \Omega\frac{\lambda_s+\lambda_b}{\alpha\mu^2}\\ \Omega\frac{-1}{\alpha\mu}\\ \Omega\frac{-1}{\alpha\mu}\end{pmatrix}$$

and show that

$$Var(PIN) == \frac{\Omega^2}{n}\left[\frac{\alpha(1-\alpha)(\lambda_s+\lambda_b)^2}{\alpha^4\mu^2}+\frac{\mu(\lambda_s+\lambda_b)^2}{\alpha^2\mu^4}+\frac{(\lambda_s+\lambda_b)}{\alpha^2\mu^2}\right] \tag{2.27}$$

2.5 Bayesian Inference Of The Extended PIN Model

Similar to section 2.2, the Metropolis-Hastings and Gibbs sampling algorithms are used to estimate the model parameters. We choose the following prior probability distributions for the parameters $\alpha, \delta, \mu,$ λ_s,λ_b and D_t:

$$
\begin{aligned}
&P(\alpha|\rho,\phi) = \tfrac{1}{\mathcal{B}eta(\rho,\varepsilon)}\alpha^{\rho-1}(1-\alpha)^{\phi-1}, && P(\lambda_s|\gamma_1,\beta_1) = \tfrac{\beta_1^{\gamma_1}}{\Gamma(\beta_1)}\lambda_s^{\gamma_1-1}e^{-\beta_1\lambda_s},\\
&P(\delta|\nu,\tau) = \tfrac{1}{\mathcal{B}eta(\nu,\tau)}\delta^{\nu-1}(1-\delta)^{\tau-1}, && P(\lambda_b|\gamma_2,\beta_2) = \tfrac{\beta_2^{\gamma_2}}{\Gamma(\beta_2)}\lambda_b^{\gamma_2-1}e^{-\beta_2\lambda_b},\\
&P(\mu|\gamma_0,\beta_0) = \tfrac{\beta_0^{\gamma_0}}{\Gamma(\gamma_0)}\mu^{\gamma_0-1}e^{-\beta_0\mu}, && D_t \sim \mathbf{D}irichlet(\pi_1,\pi_2,\pi_3).
\end{aligned}
$$

2.5.1 *Method 1: Gibbs Sampler*

Density of buy and sell trades on a bad news day ($D_t = 1$)

On a bad news day, informed traders expect an adverse effect on the value of the asset and would sell the asset. Liquidity traders either buy or sell the asset for reasons other than information. The number of liquidity trader buy trades (B_t^u) follows a Poisson distribution with mean λ_b. Hence the total daily buy trades $B_t = B_t^u$, follows a Poisson distribution with mean λ_b. Similarly, informed trader sell trades (S_t^i) and liquidity trader sell trades (S_t^u) follow independent Poisson distributions with means μ and λ_s respectively. Thus the total sell orders ($S_t = S_t^i + S_t^u$) follow a Poisson distribution with mean $\mu + \lambda_s$. Conditioning on buy and sell trades, the probability of informed trader sell trades follows binomial distribution with S_t trials and parameter ${}^{\mu}/_{\mu+\lambda_s}$. The trade arrival distributions are given as

$$
\begin{aligned}
&S_t|D_t = 1 \sim Pn\left(\mu+\lambda_s\right) && B_t|D_t = 1 \sim Pn\left(\lambda_b\right),\\
&S_t^i|S_t, D_t = 1 \sim Bin\Big(S_t, \tfrac{\mu}{\mu+\lambda_s}\Big).
\end{aligned}
$$

The probability of a buy or sell trade on a bad news day is

$$
\begin{aligned}
f_1\left(B_t, S, S_t^i, \Theta\right) = P\left(B_t, S_t, S_t^i | D_t = 1, \Theta\right) \\
=& P\left(B_t | D_t = 1, \Theta\right) P\left(S_t, S_t^i | D_t = 1, \Theta\right) \\
=& P\left(B_t | D_t = 1, \Theta\right) P\left(S_t, S_t^i | D_t = 1, \Theta\right) \\
=& P\left(B_t | D_t = 1, \Theta\right) P\left(S_t^i | S_t, D_t = 1, \Theta\right) P\left(S_t | D_t = 1, \Theta\right) \\
=& \binom{S_t}{S_t^i} \frac{e^{-\left(\mu + \lambda_b + \lambda_{S_t}\right)}}{B_t!\, S_t!} {\lambda_b}^{B_t} {\lambda_{S_t}}^{S_t} \left(\frac{\mu}{\lambda_{S_t}}\right)^{S_t^i} \\
=& \frac{e^{-\lambda_b} \left(\lambda_b\right)^{B_t}}{B_t!} \binom{S_t}{S_t^i} \left(\frac{\mu}{\mu + \lambda_{S_t}}\right)^{S_t^i} \left(\frac{\lambda_{S_t}}{\mu + \lambda_{S_t}}\right)^{S_t - S_t^i} \frac{e^{-\mu + \lambda_{S_t}} (\mu + \lambda_{S_t})^{S_t}}{S_t!} \\
=& \binom{S_t}{S_t^i} \frac{e^{-\left(\mu + \lambda_b + \lambda_{S_t}\right)}}{B_t!\, S_t!} {\lambda_b}^{B_t} {\lambda_{S_t}}^{S_t - S_t^i} \mu^{S_t^i} \\
=& \frac{e^{-\mu} \mu^{S_t^i}}{S_t^i!} \frac{e^{-\lambda_b} \lambda_b^{B_t}}{B_t!} \frac{e^{-\lambda_{S_t}} \lambda_{S_t}^{S_t - S_t^i}}{(S_t - S_t^i)!}.
\end{aligned} \tag{2.28}
$$

Equation 2.28 is a product of Poisson processes of uninformed trader sell and buy trades, and informed trader sells. From the model assumptions and Poisson mixture structure, this is expected.

Density of buy and sell trades on a good news day($D_t = 2$)

Similarly, if the news content on a day is considered to be favourable, it is intuitive to expect informed traders to make purchases while liquidity traders either buy or sell the asset. Hence informed traders do not sell their assets during a period of anticipated good news. All sales on such a day come from liquidity traders. The trade arrivals distributions on such a day are as follows

$$
\begin{aligned}
&B_t | D_t = 2 \sim Pn\left(\mu + \lambda_b\right), \qquad S_t | D_t = 2 \sim Pn\left(\lambda_s\right), \\
&B_t^i | B_t, D_t = 2 \sim Bin\left(B_t, \tfrac{\mu}{\mu + \lambda_b}\right).
\end{aligned}
$$

The probability of a buyer or seller initiated trade on a good news day is

$$\begin{aligned} f_2\left(B_t, S_t, \Theta\right) =& P\left(B_t, S_t | D_t = 2, \Theta\right) \\ =& P\left(S_t | D_t = 2, \Theta\right) P\left(B_t^i | B_t, D_t = 2, \Theta\right) P\left(B_t | D_t = 2, \Theta\right) \\ =& \frac{e^{-\mu} \mu^{B_t^i}}{B_t^i!} \frac{e^{-\lambda_s} \lambda_s^{S_t}}{S!} \frac{e^{-\lambda_b} \lambda_b^{B_t - B_t^i}}{\left(B_t - B_t^i\right)!}. \end{aligned} \tag{2.29}$$

We recognise this as the product of three Poisson processes of informed trader buy trades, uninformed trader sell trades and uninformed buy trades.

Density of buy and sell trades on a no news day ($D_t = 3$)

Since informed traders do not trade on a no news day, the order arrivals would be wholly attributable to liquidity traders with distributions given as $B_t | D_t = 3 \sim Pn\left(\lambda_b\right)$ and $S_t | D_t = 3 \sim Pn\left(\lambda_s\right)$ respectively. The probability of a buyer or seller initiated trade, on a bad news day is

$$\begin{aligned} f_3\left(B_t, S_t, \Theta\right) =& P\left(B_t, S_t | D_t = 3, \Theta\right) \\ =& P\left(S_t | D_t = 3, \Theta\right) P\left(B_t | \Theta\right) P\left(B_t | D_t = 3, \Theta\right) \\ =& \frac{e^{-\lambda_s} \left(\lambda_s\right)^{S_t}}{S_t!} \frac{e^{-\lambda_b} \left(\lambda_b\right)^{B_t}}{B_t!}. \end{aligned} \tag{2.30}$$

Putting equations 2.28, 2.29 and 2.30 together, we obtain the following joint density function

$$\begin{aligned} P\left(B_t, S_t | D_t, \Theta\right) =& \left[f_1\left(B_t, S_t, \Theta\right)\right]^{d_{t,1}} \left[f_2\left(B_t, S_t, \Theta\right)\right]^{d_{t,2}} \left[f_3\left(B_t, S_t, \Theta\right)\right]^{d_{t,3}} \\ =& \left[\frac{e^{-\mu} \mu^{S_t^i}}{S_t^i!} \frac{e^{-\lambda_b} \lambda_b^{B_t}}{B_t!} \frac{e^{-\lambda_s} \lambda_s^{S_t - S_t^i}}{\left(S_t - S_t^i\right)!}\right]^{d_{t,1}} \left[\frac{e^{-\mu} \mu^{B_t^i}}{B_t^i!} \frac{e^{-\lambda_s} \lambda_s^{S_t}}{S_t!} \frac{e^{-\lambda_b} \lambda_b^{B_t - B_t^i}}{\left(B_t - B_t^i\right)!}\right]^{d_{t,2}} \\ & \times \left[\frac{e^{-\lambda_s} \left(\lambda_s\right)^{S_t}}{S_t!} \frac{e^{-\lambda_b} \left(\lambda_b\right)^{B_t}}{B_t!}\right]^{d_{t,3}}. \end{aligned} \tag{2.31}$$

Posterior Density And Full Conditional Distributions

Based on the prior distributions chosen, we write down the posterior density of the model parameters as

$$\begin{aligned}
P\left(\Theta|B_t,S_t\right) &\propto P\left(\Theta\right)\prod_{t=1}^{T}\left[P\left(B_t,S_t|D_t,\Theta\right)P\left(D_t|\Theta\right)\right] \\
&= \mu^{\gamma_0-1}e^{-\beta_0\mu}\lambda_s^{\gamma_1-1}e^{-\beta_1\lambda_s}\lambda_b^{\gamma_2-1}e^{-\beta_2\lambda_b}\alpha^{\rho-1}(1-\alpha)^{\phi-1}\delta^{\nu-1}(1-\delta)^{\tau-1} \\
&\times\left[(\alpha\delta)^{T_1}(\alpha(1-\delta))^{T_2}(1-\alpha)^{T_3}\right]\prod_{t=1}^{T}\left[\frac{e^{-\mu}\mu^{S_t^i}}{S_t^i!}\frac{e^{-\lambda_b}\lambda_b^{B_t}}{B_t!}\frac{e^{-\lambda_s}\lambda_s^{S_t-S_t^i}}{(S_t-S_t^i)!}\right]^{d_{t,1}} \\
&\times\left[\frac{e^{-\mu}\mu^{B_t^i}}{B_t^i!}\frac{e^{-\lambda_s}\lambda_s^{S_t}}{S_t!}\frac{e^{-\lambda_b}\lambda_b^{B_t-B_t^i}}{(B_t-B_t^i)!}\right]^{d_{t,2}}\left[\frac{e^{-\lambda_s}\left(\lambda_s\right)^{S_t}}{S_t!}\frac{e^{-\lambda_b}\left(\lambda_b\right)^{B_t}}{B_t!}\right]^{d_{t,3}}. \qquad (2.32)
\end{aligned}$$

To use the Gibbs sampler we need the full conditionals of the parameters of interest. From equation 2.32 the full conditional distributions for the parameter set Θ are given as follows

$$\mu \sim \mathcal{G}a\left(\gamma_0+\sum_{t=1}^{T}[(S_t^i)^{d_1}+(B_t^i)^{d_2}],\ T_1+T_2+\beta_0\right), \qquad (2.33a)$$

$$\delta \sim \mathcal{B}e\left(\nu+T_1+T_2,\ \ T_3+\tau\right), \qquad (2.33b)$$

$$\alpha \sim \mathcal{B}e\left(\rho+T_1+T_2,\ \ T_3+\phi\right), \qquad (2.33c)$$

$$\lambda_s \sim \mathcal{G}a\left(\gamma_1+\sum_{t=1}^{T}[S_t^{d_1}-(S_t^i)^{d_1}+S_t^{d_2}+S_t^{d_3}],\quad T_1+T_2+T_3+\beta_1\right), \qquad (2.33d)$$

$$\lambda_b \sim \mathcal{G}a\left(\gamma_2+\sum_{t=1}^{T}[B_t^{d_1}+B_t^{d_2}-(B_t^i)^{d_2}+B_t^{d_3}],\quad T_1+T_2+T_3+\beta_2\right). \qquad (2.33e)$$

where $d_{t,j}=\mathbf{1}_{\{D_t=j\}}$, for $j=1,2,3$ and $T=T_1+T_2+T_3$. It can be observed that the estimates of α and δ are highly dependent on the correct classification of news event periods.

Gibbs Sampling Procedure

The following algorithm is used to sample from the full conditional posterior distributions of the model parameters.

- Choose an arbitrary initial type of day (good, bad and no news day) classification for the (B_t, S_t). Denote the initial classification as $\mathbf{D_t}^{(0)}$.
- Set initial values for the parameter set Θ. Denote it as $\Theta^{(0)}=\left(\alpha^{(0)}, \delta^{(0)}, \lambda_s^{(0)}, \lambda_b^{(0)}, \mu^{(0)}\right)$
- Repeat for $\quad k = 1$ to G sweeps

  – Sample $B_t^{i(k)}|D_t^{(k-1)} \sim Bin\left(B_t, \frac{\mu^{(k-1)}}{\mu^{(k-1)}+\lambda_b^{(k-1)}}\right), \quad t = 1, \ldots, T$

  – Sample $S_t^{i(k)}|D_t^{(k-1)} \sim Bin\left(S_t, \frac{\mu^{(k-1)}}{\mu^{(k-1)}+\lambda_s^{(k-1)}}\right), \quad t = 1, \ldots, T$

  – Update $\mu^{(k)}|\lambda_s^{(k-1)}, \lambda_b^{(k-1)}, \alpha^{(k-1)}, \delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}$

  – Update $\lambda_s^{(k)}|\lambda_b^{(k-1)}, \alpha^{(k-1)}, \delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}, \mu^{(k)}$

  – Update $\lambda_b^{(k)}|\alpha^{(k-1)}, \delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}, \mu^{(k)}, \lambda_s^{(k)}$

  – Update $\alpha^{(k)}|\delta^{(k-1)}, B_t^{i(k)}, S_t^{i(k)}, \mu^{(k)}, \lambda_s^{(k)}, \lambda_b^{(k)}$

  – Update $\delta^{(k)}|B_t^{i(k)}, S_t^{i(k)}, \mu^{(k)}, \lambda_s^{(k)}, \lambda_b^{(k)}, \alpha^{(k)}$

  – Compute $L_1 = \log \omega_1^{(k)} - \left(\mu^{(k)} + \lambda_s^{(k)} + \lambda_b^{(k)}\right) + B_t \log \lambda_b^{(k)} + S_t \log\left(\lambda_s^{(k)} + \mu^{(k)}\right)$

  – Compute $L_2 = \log \omega_2^{(k)} - \left(\mu^{(k)} + \lambda_s^{(k)} + \lambda_b^{(k)}\right) + S_t \log \lambda_s^{(k)} + B_t \log\left(\lambda_b^{(k)} + \mu^{(k)}\right)$

  – Compute $L_3 = \log \omega_3^{(k)} - \left(\lambda_s^{(k)} + \lambda_b^{(k)}\right) + S_t \log \lambda_s^{(k)} + B_t \log \lambda_b^{(k)}$

  – compute $\chi = \max(L_1, L_2, L_3)$

  – Compute $p_1 = \frac{e^{L_1-\chi}}{\sum_{j=1}^{3} e^{L_j-\chi}}$, $p_2 = \frac{e^{L_2-\chi}}{\sum_{j=1}^{3} e^{L_j-\chi}}$ and $p_3 = \frac{e^{L_3-\chi}}{\sum_{j=1}^{3} e^{L_j-\chi}}$

  – Update $\mathbf{D_t}^{(k)}$, the classification of (B_t, S_t) by sampling from the multinomial distribution with probability (p_1, p_2, p_3),

where p_1, p_2 and p_3 are the probabilities that at the beginning of the trading there will be a bad news, good news and no news respectively.

2.5.2 *Method 2: Metropolis-Hastings Algorithm*

The posterior density for the parameter set $\Theta = (\alpha, \delta, \mu, \lambda_s, \lambda_b)$ which is proportional to the product of the likelihood and prior is given as

$$\begin{aligned}
P\left(\Theta|B_t, S_t\right) \propto & P(\alpha|\rho,\varepsilon)P(\delta|\nu,\tau)P(\mu|\gamma_0,\beta_0)P(\lambda_s|\gamma_1,\beta_1)P(\lambda_b|\gamma_2,\beta_2) \\
& \times \prod_{t=1}^{T}\Big[\alpha\delta e^{-(\mu+\lambda_s)}\left(\mu+\lambda_s\right)^{S_t}e^{-\lambda_b}\left(\lambda_b\right)^{B_t} + \left(1-\alpha\right)e^{-\lambda_b}\left(\lambda_b\right)^{B_t}e^{-\lambda_s}\left(\lambda_s\right)^{S_t} \\
& + \alpha\left(1-\delta\right)e^{-(\mu+\lambda_b)}\left(\mu+\lambda_b\right)^{B_t}e^{-\lambda_s}\left(\lambda_s\right)^{S_t}\Big] \\
\propto & \alpha^{\rho-1}(1-\alpha)^{\phi-1}\delta^{\nu-1}(1-\delta)^{\tau-1}\mu^{\gamma_0-1}e^{-\beta_0\mu}\lambda_s^{\gamma_1-1}e^{-\beta_1\lambda_s}\lambda_b^{\gamma_2-1}e^{-\beta_2\lambda_b} \\
& \times \prod_{t=1}^{T}\Big[\alpha\delta e^{-(\mu+\lambda_s)}\left(\mu+\lambda_s\right)^{S_t}e^{-\lambda_b}\left(\lambda_b\right)^{B_t} + \left(1-\alpha\right)e^{-\lambda_b}\left(\lambda_b\right)^{B_t}e^{-\lambda_s}\left(\lambda_s\right)^{S_t} \\
& + \alpha\left(1-\delta\right)e^{-(\mu+\lambda_b)}\left(\mu+\lambda_b\right)^{B_t}e^{-\lambda_s}\left(\lambda_s\right)^{S_t}\Big]. \qquad (2.34)
\end{aligned}$$

Similarly to the previous model we factorise the logarithm of the posterior density to avoid the floating point execution problems encountered in most packages. The logarithm of the posterior density is given as

$$\begin{aligned}
\ln\left(P\left(\Theta|B_t, S_t\right)\right) = & (\gamma_0-1)\ln\left(\mu\right) - \beta_0\mu + (\gamma_1-1)\ln\left(\lambda_s\right) - \beta_1\lambda_s \\
& + (\gamma_2-1)\ln\left(\lambda_b\right) - \beta_2\lambda_b + (\rho-1)\ln\left(\alpha\right) \\
& + (\phi-1)\ln\left(\alpha\right) + (\nu-1)\ln\left(\delta\right) + (\tau-1)\ln\left(\delta\right) \\
& + \sum_{t=1}^{T}\Big[-\lambda_b - \lambda_s B_t\ln\lambda_b + S_t\ln\lambda_s + \chi\Big] \\
& + \sum_{t=1}^{T}\ln\Big[e^{L_1-\chi} + e^{L_2-\chi} + e^{L_3-\chi}\Big],
\end{aligned}$$

where

$L_1 = -\mu + S_t\ln(1+\frac{\mu}{\lambda_s}) + \ln(\alpha\delta)$,

$L_2 = -\mu + B_t\ln(1+\frac{\mu}{\lambda_b}) + \ln[\alpha(1-\delta)]$,

$L_3 = \ln(1-\alpha)$ and

$\chi = \max(L_1, L_2, L_3)$.

We use the random walk Adaptive Metropolis Hastings algorithm with standard Gaussian innovations where we accept the proposals $\alpha', \delta', \mu', \lambda'_s$ and λ'_b of the parameters $\alpha, \delta, \mu, \lambda_s$ and λ_b from their respective posterior distributions with

$$\text{acceptance probability} = \min\left\{1, \frac{P\left(\alpha', \delta', \mu', \lambda'_s, \lambda'_b | B_t, S_t\right)}{P\left(\alpha, \delta, \mu, \lambda_s, \lambda_b | B_t, S_t\right)}\right\}. \tag{2.35}$$

To test the applicability of our estimation method we carry out simulation exercise on a hypothetical data set of aggregate buyer and seller initiated trades for Easley et al. (1996) and Easley et al. (2002) models. Subsequently the method is applied to real data for two assets. Henceforth we refer to Easley et al. (1996) and Easley et al. (2002) models as *Model I* and *Model II* respectively.

2.6 Hypothetical Data Implementation

We use the parameters in Table 2.1 below to simulate hypothetical data set. The corresponding PIN for Models I and II based on these parameters are 0.3846 and 0.2307 respectively. The simulated data set is of size $1,000$.

	α	δ	μ	ϵ	λ_s	λ_b
Model I	0.5	0.7	50	20		
Model II	0.5	0.7	30		20	30

Table 2.1 Parameter values for simulated data

The Gibbs Sampler is run for $G = 100,000$ sweeps for both Models I and II. The initial $n_0 = 30,000$ draws from the posterior distributions of the parameters are discarded. From the remaining $70,000$ draws we calculate the posterior mean and credible intervals for each parameter. Estimating Model I with the Adaptive Metropolis Hastings (AMH) algorithm, after iterations $t > t_0$, we set $\varphi = 0.01$ for the α' and δ' samples. Similarly, for the μ' and ϵ' we set $\varphi = 0.02$. We run the chain for $135,000$ iterations with $G = 100,000$, $n_0 = 35,000$, and $t_0 = 15,000$. The ratio of the total number of accepted draws to the number of iterations after the burnin period (*acceptance rate*) achieved is 0.1384 compared to the optimal figure of 0.234 suggested in Roberts et al. (1997).

Similarly, using $d = 5$, $s_d = 2.4^2$ and $\xi^{(0)} = 1$ we implement the Adaptive Metropolis Hastings Algorithm on the simulated data for Model II. For iterations $t > t_0$, we set $\varphi = 0.01$ for the α' and δ' samples. Similarly for the μ' , λ_b' and λ_s' sample we set $\varphi = 0.1$. We run the chain for $135,000$ iterations with $G = 100,000$, $n_0 = 35,000$, and $t_0 = 15,000$. The acceptance probability achieved is 0.1488 which is closer to the optimal figure of 0.234 of Roberts et al. (1997). For comparability, we carried out maximum likelihood estimation on the simulated data. The maximum likelihood estimates and PIN obtained by implementing the optimisation function *optim* in the statistical package R on equations 2.4 and 2.24 are presented in Tables 2.2 and 2.3 respectively. We carried out the MLE while varying the sample size (N) of the simulated data. For each sample size, we run the optimiser for $1,000$ runs each time changing the initial values. The runs for each sample size with the minimum negative log likelihood value are presented in the tables. The final columns of Tables 2.2 and 2.3 show the percentage of the $1,000$ runs for which the optimiser. Thus even with the factorisation of the log likelihood function one still needs to be careful about the choice of the initial values.

N		α	δ	μ	ϵ	PIN	-LogLike	Error Rate
60	Init. values	0.226	0.484	4	7			
	Est	0.450	0.555	49	21	0.350	-9986	5.8
	Std error	0.064	0.095	1.679	0.471	0.033		
200	Initl. values	0.270	0.969	3	6			
	Est	0.465	0.666	49	20	0.358	-33602	6.8
	Std error	0.035	0.048	0.904	0.258	0.017		
400	Initl. values	0.058	0.930	9	7			
	Est	0.515	0.747	5	20	0.385	-70683	10.5
	Std error	0.024	0.030	0.611	0.185	0.011		
600	Initl. values	0.406	0.507	6	7			
	Est	0.523	0.716	50	20	0.390	-106911	10.8
	Std error	0.020	0.025	0.495	0.151	0.009		
800	Initl. values	0.837	0.116	6	6			
	Est	0.510	0.713	50	20	0.388	-140539	8.6
	Std error	0.017	0.022	0.434	0.129	0.008		
1000	Initl. values	0.493	0.190	2	4			
	Est	0.520	0.709	50	20	0.393	-177195	8.1
	Std error	0.015	0.019	0.385	0.116	0.007		

Table 2.2 Maximum likelihood estimates of the simulated data for Model I

It can be observed that the parameter estimates in Tables 2.2 and 2.3 get closer to the actual values in Table 2.1 used in creating the hypothetical data set as the sample

N		α	δ	μ	λ_s	λ_b	PIN	-LogLike	Error Rate
60	Initl. values	0.634	0.037	2	2	6			
	Est	0.513	0.746	32	20	29	0.249	-9973	2.9
	Std error	0.066	0.081	1.513	0.732	0.802	0.023		
200	Initl. values	0.912	0.737	4	6	3			
	Est	0.413	0.707	32	20	30	0.206	-31654	3.7
	Std error	0.035	0.050	0.882	0.375	0.417	0.013		
400	Initl. values	0.088	0.580	5	0	10			
	Est	0.486	0.738	31	20	30	0.228	-66268	2.3
	Std error	0.025	0.031	0.576	0.276	0.293	0.009		
600	Initl. values	0.593	0.900	5	3	5			
	Est	0.504	0.724	31	20	30	0.236	-99894	2.3
	Std error	0.020	0.025	0.464	0.227	0.241	0.007		
800	Initl. values	0.898	0.514	2	4	5			
	Est	0.517	0.741	30.435	20	30	0.239	-134247	2.9
	Std error	0.017	0.021	0.398	0.201	0.208	0.006		
1000	Initl. values	0.611	0.136	3	6	5			
	Est	0.518	0.726	31	20	30	0.241	-167998	2.3
	Std error	0.016	0.019	0.356	0.178	0.188	0.005		

Table 2.3 Maximum likelihood estimates of the simulated data for Model II

size of the data increases. Likewise, the standard error associated with the estimation of the parameters and PIN decreases with increased sample size. From these results one can argue that even though Easley et al. (2002) proposes that 60 days of buyer and seller initiated trades are enough to give stable parameter estimates, the estimates based on 60 will yield large estimation errors.

Tables 2.4 and 2.5 are summaries of the marginal posterior distributions of the parameters obtained from the Bayesian estimation. Similarly to the MLE results the estimates from the Bayesian methods are also close to the original parameters used for simulating the data. With uninformative priors, we expect the Bayesian estimates to be close to the maximum likelihood estimates.

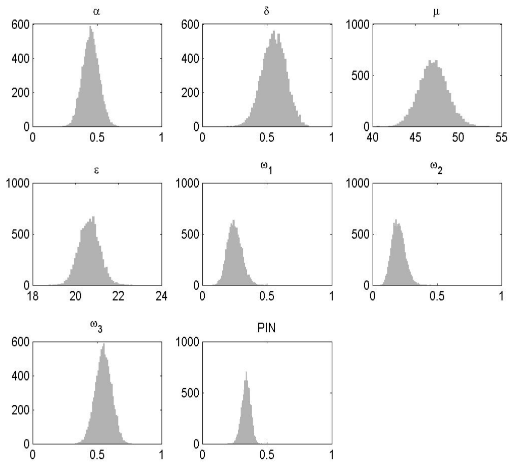
In Figure 2.3, we present the posterior distributions of the model parameters as well as the distribution of the PIN and type of trading day for a the hypothetical data of size 60.

	N		α	δ	μ	ϵ	PIN
Gibbs	60	Est	0.451	0.552	47	20.643	0.338
		Sdev	0.062	0.090	1.597	0.472	0.032
		LCI	0.331	0.370	44	20	0.272
		UCI	0.575	0.724	50	21	0.399
	200	Est	0.464	0.662	48	20	0.354
		Sdev	0.034	0.048	0.895	0.259	0.018
		LCI	0.396	0.563	47	20	0.318
		UCI	0.534	0.756	50	21	0.389
	400	Est	0.514	0.744	49	20	0.383
		Sdev	0.024	0.030	0.605	0.184	0.012
		LCI	0.465	0.681	48	20	0.358
		UCI	0.563	0.801	50	21	0.407
	600	Est	0.523	0.715	49	20	0.389
		Sdev	0.020	0.025	0.491	0.150	0.009
		LCI	0.483	0.664	48	20	0.370
		UCI	0.562	0.762	50	20	0.408
	800	Est	0.509	0.711	50	20	0.387
		Sdev	0.017	0.022	0.432	0.129	0.008
		LCI	0.474	0.667	49	20	0.370
		UCI	0.544	0.755	51	20	0.404
	1000	Est	0.520	0.708	51	20	0.392
		Sdev	0.015	0.019	0.385	0.116	0.007
		LCI	0.489	0.669	49	20	0.377
		UCI	0.551	0.746	51	20	0.407
AMH	60	Est	0.452	0.55	47	21	0.337
		Sdev	0.062	0.088	1.624	0.479	0.032
		LCI	0.332	0.374	44	20	0.269
		UCI	0.578	0.720	50	22	0.398
	200	Est	0.465	0.661	49	20.498	0.355
		Sdev	0.035	0.048	0.884	0.247	0.018
		LCI	0.397	0.561	47	20	0.318
		UCI	0.535	0.749	50	21	0.390
	400	Est	0.515	0.744	49	20	0.384
		Sdev	0.024	0.030	0.599	0.188	0.011
		LCI	0.466	0.685	48	20	0.360
		UCI	0.564	0.802	50	21	0.408
	600	Est	0.522	0.712	49	20	0.389
		Sdev	0.019	0.024	0.524	0.146	0.009
		LCI	0.484	0.662	48	20	0.370
		UCI	0.560	0.761	50	20	0.407
	800	Est	0.510	0.711	50	205	0.387
		Sdev	0.016	0.022	0.402	0.124	0.008
		LCI	0.475	0.664	49	20	0.370
		UCI	0.540	0.754	51	20	0.403
	1000	Est	0.521	0.707	50	20	0.393
		Sdev	0.015	0.019	0.408	0.12	0.007
		LCI	0.489	0.666	49	20	0.378
		UCI	0.550	0.745	51	20	0.407

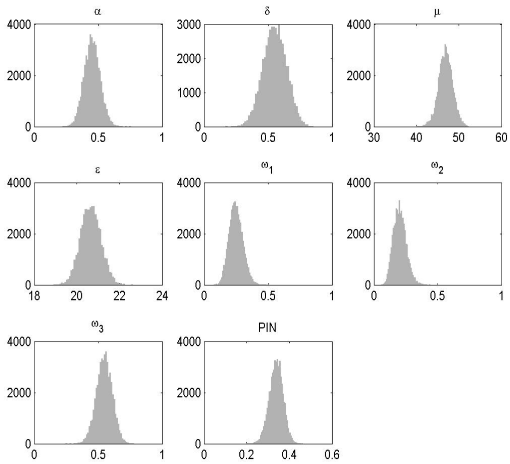
Table 2.4 Posterior estimates of the simulated data for Model I

	N		α	δ	μ	λ_s	λ_b	PIN
Gibbs	60	Est	0.523	0.720	30	20	29	0.280
		Sdev	0.064	0.078	1.464	0.731	0.784	0.028
		LCI	0.396	0.557	27	19	26.774	0.224
		UCI	0.648	0.860	33	21	30	0.335
	200	Est	0.416	0.701	31	20	30	0.244
		Sdev	0.035	0.050	0.882	0.378	0.416	0.016
		LCI	0.347	0.597	29	19	29	0.210
		UCI	0.485	0.794	32	20	31	0.277
	400	Est	0.487	0.734	30	20	30	0.270
		Sdev	0.025	0.031	0.575	0.277	0.294	0.011
		LCI	0.438	0.670	29	19	30	0.248
		UCI	0.536	0.793	31	20	31	0.292
	600	Est	0.505	0.722	30	20	30	0.278
		Sdev	0.020	0.025	0.464	0.227	0.239	0.011
		LCI	0.464	0.670	29	19	29	0.260
		UCI	0.545	0.771	31	20	30	0.296
	800	Est	0.517	0.740	30	20	30	0.281
		Sdev	0.017	0.021	0.399	0.203	0.208	0.007
		LCI	0.483	0.696	29	20	29	0.265
		UCI	0.552	0.781	31	20	30	0.296
	1000	Est	0.518	0.725	30	20	30	0.284
		Sdev	0.015	0.019	0.354	0.175	0.186	0.007
		LCI	0.486	0.686	30	19	29	0.270
		UCI	0.549	0.763	31	20	30	0.298
AMH	60	Est	0.524	0.713	30	20	28	0.244
		Sdev	0.063	0.077	1.471	0.732	0.757	0.024
		LCI	0.399	0.549	27	19	27	0.194
		UCI	0.648	0.851	33	21	30	0.290
	200	Est	0.416	0.699	31	20	30	0.204
		Sdev	0.034	0.049	0.907	0.38	0.418	0.014
		LCI	0.348	0.601	29	19	29	0.175
		UCI	0.485	0.791	33	20	31	0.232
	400	Est	0.487	0.733	30	20	30	0.227
		Sdev	0.024	0.031	0.533	0.274	0.295	0.009
		LCI	0.437	0.670	29	19	30	0.208
		UCI	0.535	0.792	31	20	31	0.247
	600	Est	0.506	0.722	30	20	30	0.236
		Sdev	0.019	0.024	0.459	0.206	0.233	0.007
		LCI	0.468	0.670	29	19	29	0.221
		UCI	0.545	0.770	31	20	30	0.252
	800	Est	0.516	0.739	30	20	30	0.238
		Sdev	0.018	0.020	0.396	0.192	0.200	0.007
		LCI	0.481	0.698	29	20	29	0.224
		UCI	0.550	0.777	31	20	30	0.251
	1000	Est	0.518	0.723	30	20	30	0.240
		Sdev	0.015	0.019	0.336	0.172	0.193	0.006
		LCI	0.488	0.686	30	20	30	0.228
		UCI	0.549	0.757	31	20	30	0.253

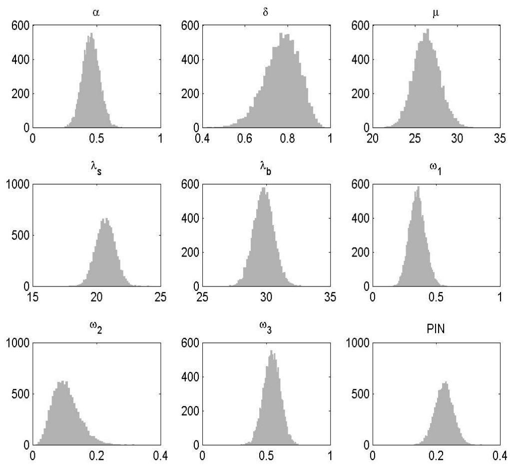
Table 2.5 Posterior estimates of the simulated data for Model II



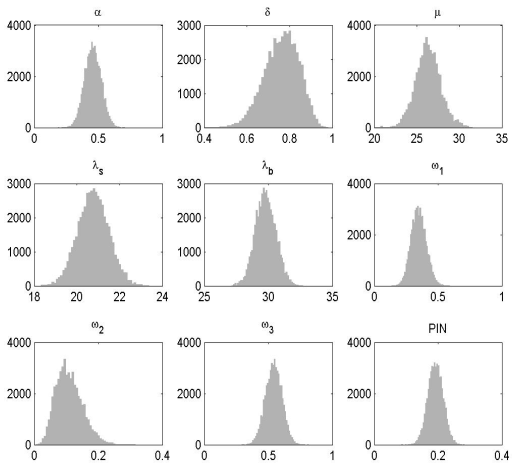
(a) Gibbs Model I



(b) AMH Model I



(c) Gibbs Model II



(d) AMH Model II

Fig. 2.3 Posterior distributions for the simulated data of size 60

The MLE computed from the optim function in *R* and the Bayesian estimates on the simulated data are compared with the maximum likelihood estimates from code written by Professor Noah Stoffman of Indiana University. Professor Stoffman's code uses the NLMIXED procedure in SAS to estimate the model parameters and PIN which are presented in Table 2.6 above. The estimates from the SAS code for Model I are also similar in magnitude to the estimates of the Bayesian methods and estimates of the fminsearchbnd function. Although slightly different, the confidence intervals of the estimates for Model II from Stoffman SAS code contains the maximum likelihood estimates of the optim function.

	Model I				**Model II**			
	Est	Sdev	LCI	UCI	Est	Sdev	LCI	UCI
α	0.450	0.064	0.321	0.578	0.439	0.066	0.305	0.573
δ	0.555	0.095	0.364	0.746	0.811	0.081	0.647	0.974
μ	49	1.679	46	53	28	1.601	21	31
ϵ	21	0.471	20	21				
λ_s					21	0.745	19	22
λ_b					30	0.807	29	32
PIN	0.351	0.034	0.282	0.418	0.195	0.025	0.144	0.246

Table 2.6 MLE for simulated data based on Noah Stoffman SAS [3]code

In the simulation exercise none of the buyer and seller initiated trades were large enough to cause overflow or underflow that arises from the terms like Θ^{B_t}, Θ^{S_t}, $e^{-\Theta}$ in the likelihood function of the two models. The value of e^{702} results in an overflow, therefore, causes MLE optimisers to become unstable. However the optimiser still failed in certain occasions. The failure of the optimisation function may be likely due to the fact that either of α and δ fell on the boundary of their parameter space. The results of the simulation exercise are encouraging since our estimation method can be applied to real data.

[3]Source of SAS code: http://kelley.iu.edu/nstoffma/

2.7 Real Data Estimation

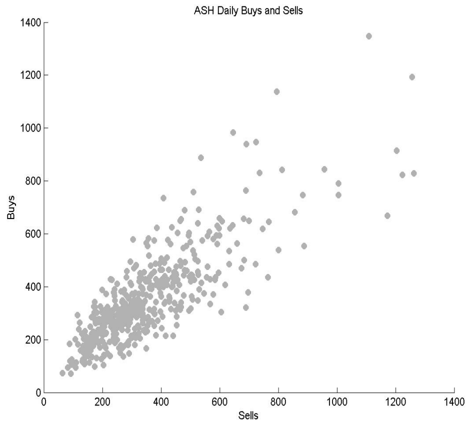
This section focuses on the empirical implementation of the Bayesian methodology on data that has been downloaded from Bloomberg terminals at the University of Kent. Comparison of the results are made with maximum likelihood estimation results. The data comprises of tick-by-tick transaction data covering the period 3^{rd} June 2013 to 15^{th} April 2015 for International Business Machines (IBM) and Ashland Oil (ASH) trading on the New York Stock Exchange (NYSE) only.

We excluded all transactions that occurred outside the normal trading hours of the Exchange. All transactions that had negative spreads were also removed from the sample. We further excluded all cases where the transaction price was higher (lower) than the ask (bid) price by more than 50 times the tick size (0.01). Finally, any transactions that occurred within the first and last 5 minutes of each trading day were also removed from the data set. After this data cleaning exercise, we obtained a total of $1,484,829$ and $332,799$ data points for IBM and ASH respectively.

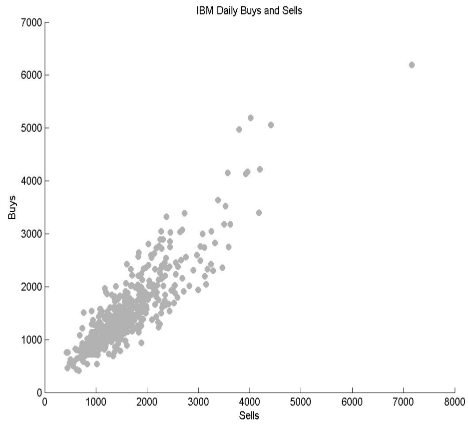
	ASH		**IBM**	
	Buys	Sells	Buys	Sells
No. days	468	468	466	466
Min	71	64	414	424
Median	316	315	1406	1523
Mean	356	353	1554	1630
Max	1348	1263	6192	7165

Table 2.7 Summary of daily buy and sell trades

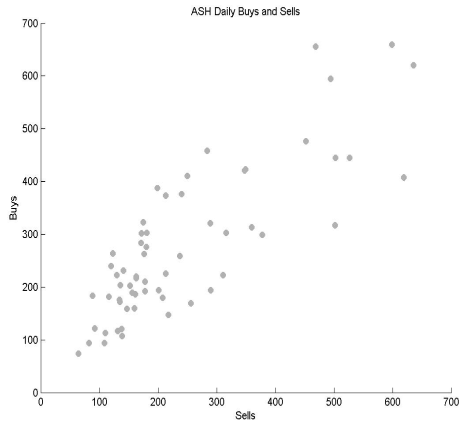
We use the Lee and Ready (1991) algorithm to classify the transaction data into buyer and seller initiated trades from which we compute the aggregate daily buyer and seller initiated trades. Table 2.7 is a summary of the daily buyer and seller initiated trades for both assets which we use in this chapter. It can be observed that the daily buyer and seller initiated trades for these assets are large enough to cause floating point exceptions. Corresponding scatter plots are shown in Figure 2.4. Noticeably, the buy and sell orders are very correlated. Other variables of this data set will be introduced in subsequent chapters where they are used.



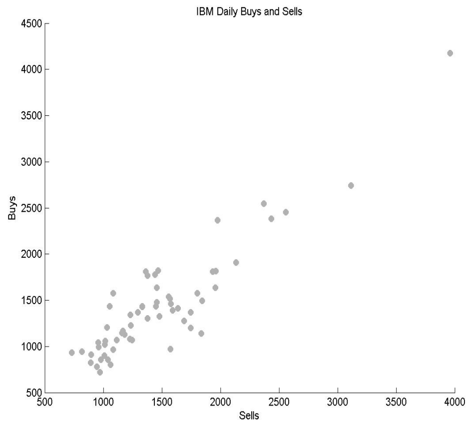
(a) ASH Full Sample



(b) IBM Full Sample



(c) ASH for last 60 days



(d) IBM for last 60 days

Fig. 2.4 Daily buy and sell orders

Results

Easley et al. (2002) indicated that buyer and seller initiated trades for 60 trading days is enough to provide stable estimates for the PIN model. With this in mind, we carried out MLE and Bayesian estimation of the PIN using buyer and seller initiated trades for the 60 trading days before 15^{th} April. The PIN estimate based on the last 60 is used as a proxy for the information asymmetry upon which the spread at the beginning of the 15^{th} April is set.

N		α	δ	μ	ϵ	PIN	% Error
ASH							
60	Initl. values	0.968	0.560	5	4		
	Est	0.295	0.388	2687	218	0.153	0
	Std error	0.058	0.114	5.397	1.449	0.009	
468	Initl. values	0.166	0.993	3	7		
	Est	0.272	0.513	330	310	0.126	0
	Std error	0.021	0.044	3.454	0.720	0.008	
IBM							
60	Initl. values	0.065	0.608	5	1		
	Est	0.311	0.631	857	1319	0.091	0.2
	Std error	0.059	0.110	11.326	3.579	0.005	
466	Initl. values	0.959	0.540	3	7		
	Est	0.251	0.601	1340	1424	0.105	1.2
	Std error	0.020	0.045	5.438	1.347	0.007	

Table 2.8 Maximum likelihood estimates of real data for Model I

N		α	δ	μ	λ_s	λ_b	PIN	% Error
ASH								
60	Initl. values	0.282	0.196	5	5	3		
	Est	0.360	0.727	245	174	254	0.171	0
	Std error	0.061	0.094	4.750	1.950	2.109	0.008	
468	Initl. values	0.329	0.465	3	4	7		
	Est	0.305	0.720	318	283	329	0.136	0
	Std error	0.021	0.037	2.374	0.938	0.895	0.008	
IBM								
60	Initl. values	0.591	0.566	8	5	4		
	Est	0.229	0.071	1016	1464	1207	0.080	0.2
	Std error	0.053	0.068	13.470	4.909	4.998	0.006	
466	Initl. values	0.427	0.660	4	6	3		
	Est	0.26	0.527	1312	1455	1388	0.107	12.4
	Std error	0.020	0.045	7.931	1.923	2.799	0.007	

Table 2.9 Maximum likelihood estimates of the real data for Model II

The optim function was run for a $1,000$ iterations each time changing the initial values. The run which resulted in the minimum negative log likelihood is presented in Tables 2.8 and 2.9 for the full sample and the last 60 trading days of Models I and II respectively. In Table 2.8 and 2.9 the optimiser did not fail for all the $1,000$ runs in the case of ASH. However in Table 2.9 the optimiser failed on 20 and 120 occasions for the IBM for the last 60 days and full sample respectively. Comparative estimates from Professor Stoffman's SAS code are also presented in Table 2.10.

		Full Sample				**Last 60 Days**			
		Model I		*Model II*		*Model I*		*Model II*	
		ASH	IBM	ASH	IBM	ASH	IBM	ASH	IBM
α	Est	0.302	0.072	0.271	0.310	0.300	0.174	0.333	0.113
	Sdev	0.021	0.012	0.023	0.021	0.059	0.063	0.060	0.045
	LCI	0.260	0.047	0.224	0.268	0.181	0.047	0.211	0.021
	UCI	0.344	0.097	0.318	0.353	0.418	0.301	0.455	0.204
δ	Est	0.527	0.777	0.515	0.721	0.388	0.540	0.550	0.563
	Sdev	0.042	0.066	0.058	0.037	0.114	0.214	0.111	0.216
	LCI	0.444	0.643	0.399	0.648	0.159	0.112	0.327	0.129
	UCI	0.611	0.903	0.631	0.795	0.618	0.968	0.772	0.996
μ	Est	304	1999	1956	306	269	200	253	200
	Sdev	2	13	7	2	5	18	5	19
	LCI	300	1974	1941	302	258	163	243	162
	UCI	309	2024	1970	311	280	237	263	238
ϵ	Est	305	**1982**			218	**2000**		
	Sdev	1	**0**			1	**0**		
	LCI	304	**1982**			215	**2000**		
	UCI	306	**1982**			221	**2000**		
λ_s	Est			**1422**	281			192	**2000**
	Sdev			**0**	1			2	**0**
	LCI			**1422**	279			188	**2000**
	UCI			**1422**	282			195	**2000**
λ_b	Est			1332	327			240	**2000**
	Sdev			5	1			2	**0**
	LCI			1323	323			236	**2000**
	UCI			1341	328			244	**2000**
PIN	Est	0.131	0.035	0.161	0.135	0.156	0.008	0.163	0.005
	Sdev	0.008	0.005	0.011	0.008	0.026	0.003	0.025	0.002
	LCI	0.115	0.024	0.138	0.119	0.104	0.002	0.113	0.001
	UCI	0.146	0.047	0.185	0.151	0.208	0.015	0.214	0.010

Table 2.10 MLE from Stoffman SAS code on real data

Using the SAS code of Stoffman we encountered the numerical computation challenges that have been reported in the literature. The estimates from the SAS code are dependent on the initial values used. Also, the routine behind the code in most cases got stuck in local maxima. We can see this in Table 2.10 where the optimiser failed in estimating λ_b, λ_s and ϵ which have been highlighted in red. For these parameters, the estimates returned by the NLMIXED procedure are exactly the initial values.

		Gibbs				**AMH**			
		Model I		*Model II*		*Model I*		*Model II*	
		ASH	IBM	ASH	IBM	ASH	IBM	ASH	IBM
α	Est	0.298	0.260	0.308	0.273	0.298	0.260	0.309	0.273
	Sdev	0.021	0.020	0.021	0.020	0.021	0.019	0.021	0.021
	LCI	0.256	0.221	0.267	0.234	0.257	0.222	0.268	0.232
	UCI	0.341	0.300	0.351	0.315	0.341	0.300	0.352	0.314
δ	Est	0.521	0.597	0.717	0.479	0.521	0.594	0.717	0.480
	Sdev	0.042	0.044	0.037	0.044	0.041	0.042	0.037	0.044
	LCI	0.437	0.508	0.639	0.392	0.442	0.508	0.639	0.392
	UCI	0.603	0.684	0.786	0.567	0.602	0.675	0.785	0.567
μ	Est	311	1297	312	1263	311	1296	313	1263
	Sdev	3	5	2	5	3	6	3	5
	LCI	306	1286	308	1251	304	1285	306	1252
	UCI	317	1308	317	1273	317	1308	318	1274
ϵ	Est	308	1422			308	1422		
	Sdev	1	1			1	1		
	LCI	307	1419			307	1419		
	UCI	309	1424			309	1424		
λ_s	Est			283	1463			283	1463
	Sdev			1	2			1	2
	LCI			281	1459			281	1459
	UCI			283	1467			283	1467
λ_b	Est			329	1369			329	1369
	Sdev			1	2			1	2
	LCI			327	1364			327	1365
	UCI			331	1373			330	1373
ω_1	Est	0.155	0.155	0.221	0.131	0.155	0.154	0.221	0.131
	Sdev	0.017	0.016	0.019	0.015	0.016	0.016	0.019	0.015
	LCI	0.123	0.123	0.185	0.102	0.124	0.125	0.184	0.101
	UCI	0.189	0.189	0.260	0.163	0.190	0.188	0.260	0.163
ω_2	Est	0.143	0.104	0.087	0.142	0.142	0.105	0.087	0.141
	Sdev	0.016	0.014	0.012	0.016	0.015	0.013	0.013	0.016
	LCI	0.112	0.078	0.063	0.112	0.112	0.079	0.064	0.111
	UCI	0.177	0.133	0.114	0.176	0.175	0.134	0.115	0.174
ω_3	Est	0.701	0.739	0.691	0.726	0.701	0.739	0.69	0.726
	Sdev	0.021	0.020	0.021	0.020	0.021	0.019	0.021	0.021
	LCI	0.658	0.699	0.648	0.684	0.658	0.699	0.647	0.685
	UCI	0.743	0.778	0.732	0.765	0.742	0.777	0.731	0.767
PIN	Est	0.131	0.106	0.145	0.105	0.131	0.105	0.136	0.108
	Sdev	0.008	0.007	0.008	0.007	0.008	0.007	0.008	0.007
	LCI	0.114	0.091	0.128	0.091	0.115	0.092	0.120	0.093
	UCI	0.147	0.120	0.162	0.119	0.147	0.120	0.152	0.122

Table 2.11 Posterior Estimates for the full sample

The posterior mean, standard error and credible intervals of the model parameters obtained from the Bayesian estimation for the full sample data are presented in Table 2.11. Table 2.12 is a corresponding summary of results for the last 60 trading days. The estimate of α shown in Table 2.11 for both assets is approximately 0.3 which is an indication that on average 3 out of 10 trading days in our sample are news driven. Given that news event occurs at the beginning, of a trading day there is approximately 72% chance that the news will have an effect on ASH in Model II. This result is in contrast with the 52% negative effect in Model I.

For IBM there is a 60% and 50% chance of a negative effect of news in Models I and Model II respectively. It can be observed that the Gibbs Sampler and AMH give similar results for the PIN estimates. Using the Gibbs Sampler the risk of trading with an informed trader in ASH is approximately 0.2 while that of IBM is also approximately 0.15. However, the risk of trading with an informed trader obtained from the AMH is approximately 0.13 and 0.1 respectively for ASH and IBM. On average there are about 300 and 1300 informed trader buyer and seller initiated trades for ASH and IBM respectively. Similarly there are about 300 and 1500 liquidity trader initiated buyer and seller initiated trades. Also on average in 70% of the days, order arrivals did not convey information that would have effect on the value of both assets.

The estimates for the last 60 days calculated from the Bayesian and MLE are also similar for both assets. It is worth noting that the MLE from optim function based on our factorisation in equations 2.4 and 2.24 are sensitive to the choice of initial values. In the code we allowed for randomly generated numbers to be used as initial values for the parameters in each model.

The parameters estimates for the full sample shown in Table 2.11 are lower in magnitude compared with the estimates of the last 60 days which are also shown in Table

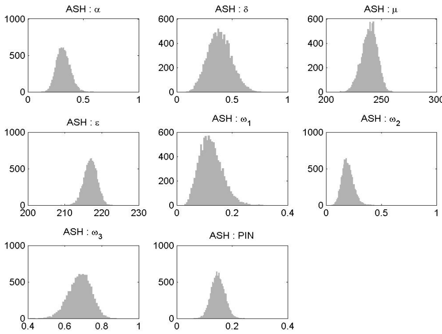
2.12. As expected the standard error of the estimates are lower for the full sample compared with the last 60 days. For instance the standard errors of α and δ for the last 60 days are about 2.5 and 3 times the standard errors of α and δ for the full sample.

Looking at the estimates of PIN, it can be observed that for IBM, the PIN for the last 60 is lower for the full sample. The standard errors are also lower than the corresponding figures of the last 60 days. However for ASH, the estimates of last 60 days are higher than that of the full sample. These results is an indication of less information that can be extracted from 60 days to provide a fair estimate of PIN, the probability of news arrival and its effect of news on the asset if it occurs.

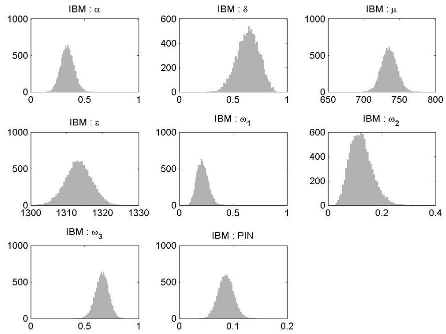
Finally, it can be observed that models I and II results in slightly different estimates. We argue that neither model is superior to the other based on the estimates. However model II provides extra insight on the behaviour of liquidity buy and sell traders rather than considring them to be homogeneous as in model I.

		Gibbs				AMH			
		Model I		*Model II*		*Model I*		*Model II*	
		ASH	IBM	ASH	IBM	ASH	IBM	ASH	IBM
α	Est	0.318	0.339	0.369	0.352	0.323	0.340	0.372	0.355
	Sdev	0.060	0.059	0.060	0.060	0.060	0.059	0.060	0.059
	LCI	0.207	0.225	0.255	0.240	0.210	0.230	0.258	0.242
	UCI	0.443	0.460	0.491	0.475	0.447	0.461	0.494	0.476
δ	Est	0.387	0.638	0.709	0.650	0.391	0.630	0.703	0.643
	Sdev	0.105	0.100	0.091	0.098	0.103	0.098	0.089	0.096
	LCI	0.191	0.431	0.517	0.446	0.199	0.428	0.516	0.444
	UCI	0.604	0.823	0.869	0.828	0.602	0.811	0.860	0.819
μ	Est	239	735	228	739	237	736	227	741
	Sdev	7	11	4	11	7	11	4	10
	LCI	225	715	219	719	220	712	219	720
	UCI	252	757	237	761	248	756	237	762
ϵ	Est	217	1313			217	1313		
	Sdev	2	4			2	4		
	LCI	213	1306			213	1306		
	UCI	220	1320			220	1320		
λ_s	Est			172	1281			172	1281
	Sdev			2	5			2	6
	LCI			168	1271			169	1271
	UCI			176	1293			176	1294
λ_b	Est			250	1311			250	1311
	Sdev			2	5			2	5
	LCI			246	1301			246	1302
	UCI			254	1320			254	1320
ω_1	Est	0.123	0.216	0.261	0.229	0.126	0.215	0.261	0.228
	Sdev	0.040	0.051	0.054	0.052	0.040	0.050	0.054	0.051
	LCI	0.053	0.124	0.162	0.136	0.057	0.124	0.164	0.135
	UCI	0.213	0.326	0.376	0.338	0.216	0.321	0.375	0.338
ω_2	Est	0.195	0.122	0.107	0.123	0.196	0.125	0.110	0.126
	Sdev	0.050	0.040	0.038	0.041	0.051	0.040	0.038	0.040
	LCI	0.106	0.054	0.044	0.054	0.107	0.058	0.048	0.057
	UCI	0.304	0.213	0.192	0.214	0.307	0.215	0.197	0.215
ω_3	Est	0.681	0.660	0.630	0.647	0.676	0.659	0.627	0.644
	Sdev	0.060	0.059	0.060	0.060	0.060	0.059	0.060	0.059
	LCI	0.556	0.539	0.508	0.524	0.552	0.538	0.505	0.523
	UCI	0.792	0.774	0.744	0.759	0.789	0.769	0.741	0.757
PIN	Est	0.148	0.086	0.196	0.092	0.149	0.086	0.166	0.092
	Sdev	0.023	0.013	0.026	0.014	0.023	0.013	0.022	0.014
	LCI	0.102	0.059	0.144	0.064	0.103	0.060	0.122	0.064
	UCI	0.195	0.114	0.247	0.120	0.195	0.114	0.210	0.119

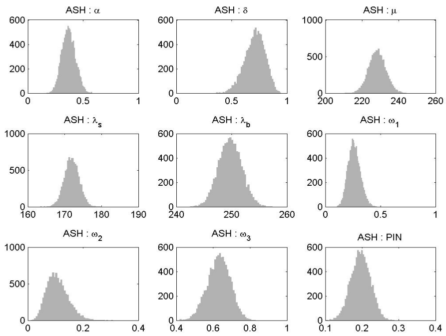
Table 2.12 Posterior estimates for last 60 days



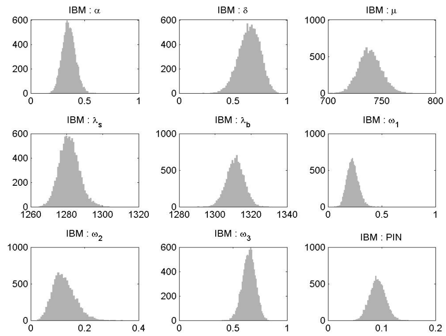
(a) ASH Gibbs Model I



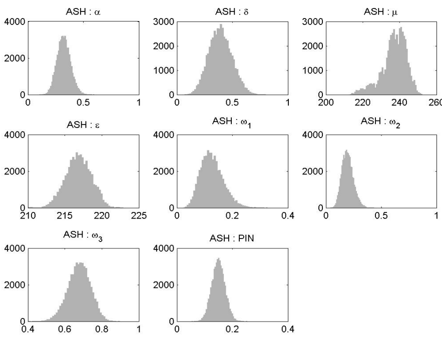
(b) IBM Gibbs Model I



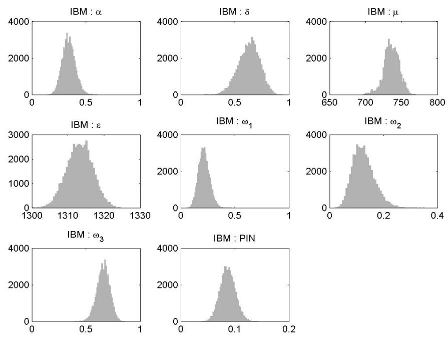
(c) ASH Gibbs Model II



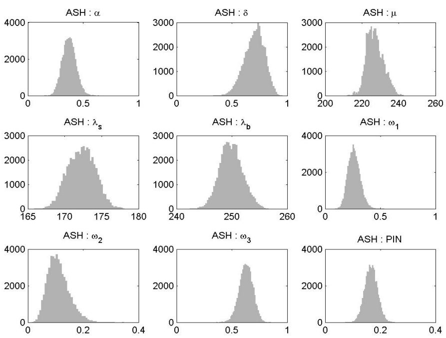
(d) IBM Gibbs Model II



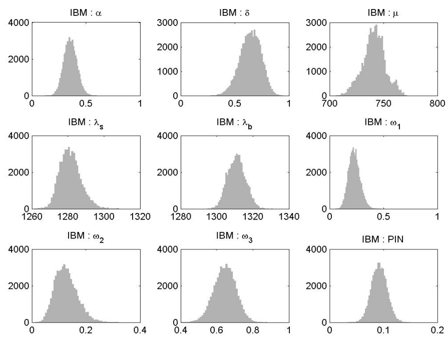
(e) ASH Metropolis Model I



(f) IBM Metropolis Model I

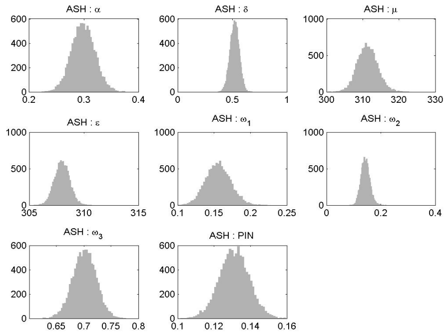


(g) ASH Metropolis Model II

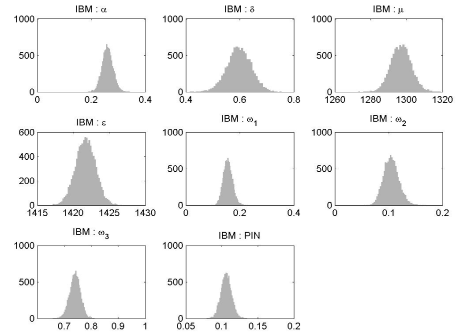


(h) IBM Metropolis Model II

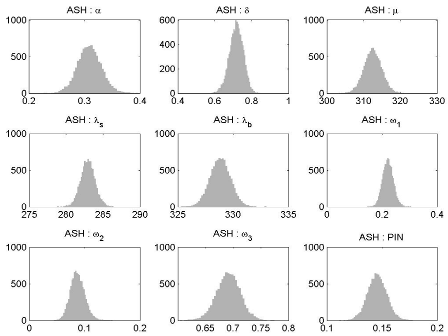
Fig. 2.5 Posterior estimates for recent 60 trading days



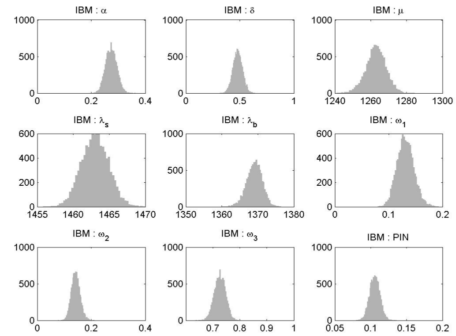
(a) ASH Gibbs Model I



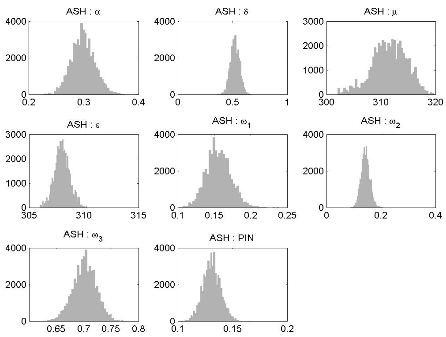
(b) IBM Gibbs Model I



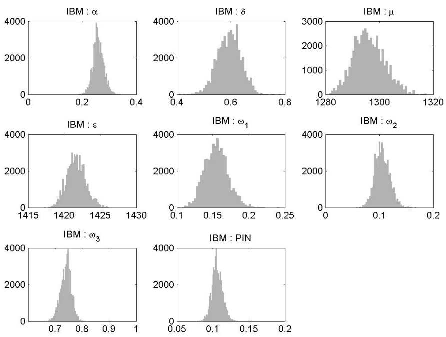
(c) ASH Gibbs Model II



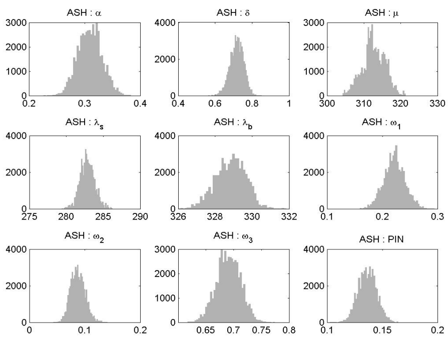
(d) IBM Gibbs Model II



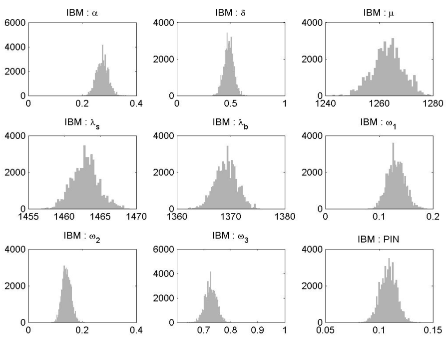
(e) ASH Metropolis Model I



(f) IBM Metropolis Model I



(g) ASH Metropolis Model II



(h) IBM Metropolis Model II

Fig. 2.6 Posterior estimates for entire sample

2.7.1 Discussion

In this chapter we have implemented the Metropolis-Hastings Algorithm and the Gibbs Sampler assuming relatively non-informative prior distributions for the model parameters. This, in essence, is equivalent to maximum likelihood estimation. However, the advantage of the MCMC methods over the classical MLE approach is that we have the full posterior distribution of the parameters from which other distributional properties can be derived. The Bayesian approach also avoids the problem of local maxima. Results from the estimation suggest that trading on private information in IBM is less in comparison to ASH. This result is consistent with findings in the literature suggesting that frequently traded assets have lower PIN compared to lower frequently traded assets.

Limitations of model assumptions

- The PIN models considered in this chapter assume that the buy and sell order arrival rates of informed traders are the same. This is to say that all informed traders have the same information set at all times. The assumption, however, is not realistic as investors may possess different information about an asset. Even in the absence of private information investors would process public information differently, and hence their trading behaviour would be non-homogeneous as assumed in these models.

- Easley et al. (1996) and Easley et al. (2002) models imply that the only source of information available to the investors is the direction of the trades. However, the dynamics of traded volume may carry information about the arrival of new information about the asset. Easley et al. (1997*a*) extended the Easley et al. (1996) model to a market where investors make small and large trades. The intuition is that large trades may be driven by the arrival of new information. However, the authors concluded that traded volume provided no further insights beyond what is contained in transaction prices. This conclusion may be attributed to the arbitrary classification of traded volume into small and large categories without recourse to the natural stochastic dynamics of the volume process.

- The Poisson arrival distribution assumption for daily arrival of orders may also not be appropriate due to large numbers of trades that occur during a normal trading day in financial markets. This is particularly the case for very liquid stocks. Sampling of trades at sub-intervals of the trading day may provide insight into the flow of information within the trading day. The Poisson assumption may be reasonable at such small intervals. We investigate the estimation of PIN using short time interval sampled trades in section 3.1.

- The assumption of days without any news is unrealistic as it will mean that there would be no trading on such days. However, we observe trades on every trading day indicating that there is some amount of news within the trading day.

2.8 Concluding Remarks

The empirical analysis carried out in this chapter indicate that the numerical instability problem in PIN estimation which has widely been reported can be avoided using Bayesian estimation methods. Bayesian methods provide flexible and efficient ways of estimating the model parameters while avoiding the non convergence problems of optimisation functions underlying maximum likelihood routines.

One challenging problem noted in papers that study PIN is the maximum likelihood estimation of the probability of news event α and the probability of a bad news event δ. In a considerable number of cases the MLE results in either a zero (0) or one (1) for these parameters which in turn yields biased PIN. However, these parameters need to be strictly between zero and one to make economic sense. The Bayesian methodology, on the other hand, does not suffer from this corner solutions problem. In the Bayesian approach, there is also no need for a careful selection and specification of initial values as is done in the MLE approach.

Chapter 3

Estimating Daily Information Asymmetry Risk From High Frequency Data

3.1 Introduction

In the preceding chapter we estimated the information asymmetry risk of an asset using daily aggregate buyer and seller initiated trades. However, buyer and seller initiated trades aggregated daily have the potential to conceal valuable information which otherwise could have been learnt if the aggregation were done at a relatively high frequency. In practice, investors are more interested in the evolution of information asymmetry in real time as trading of the asset progresses. This will enable them to better time their trades to minimise potential losses they may incur from trading with informed traders. In this chapter we utilise buyer and seller initiated trades sampled at relatively high frequency to explore daily estimated PIN. Comparisons are made with estimates of the Volume Synchronized Probability of Informed Trading (VPIN) of Easley et al. (2011).

Easley et al. (2008) were the first to investigate the time series properties of information asymmetry via their PIN measure. The authors used a bivariate Generalized Autoregressive Conditional Heteroscedastic (GARCH) to model the difference between buyer and seller initiated trades to infer the informed and liquidity arrival rates that underpin the PIN measure. They calculated PIN in the usual way as the ratio of expected number of informed trades to total trades. Their findings are that the arrival intensities of liquidity traders are negatively related to past arrival intensities of informed traders. Also they report that PIN is time-varying and that both informed and liquidity trader arrivals are persistent.

Easley et al. (2008) in their analysis did not account for the potential contributions of other available market variables such as volume, bid-ask spread and duration between trades. Tay et al. (2009) proposed a high frequency PIN based on Asymmetric Autoregressive Conditional model for trade direction and duration between high frequency trades. In their model, the authors used trade direction and duration between transactions to estimate the expected arrival rates of informed and uninformed traders. In predicting the probability of news arrival, Tay et al. (2009) used buyer and seller initiated volume as an explanatory variable in their model.

Easley et al. (2011) developed the Volume Synchronized Probability of Informed Trading as an extension of PIN. The extended model is intended to account for the information content of trading volume. The new measure is based on a predefined time interval or an arbitrarily chosen level of volume of shares traded which the authors define as *volume bucket.* The sequence of volume in a volume bucket is weighted using price changes over the sample period. According to the authors, the information content of volume in each volume bucket is assumed to be unchanged. Hence any unexpected increase or decrease in volume is an indication of the arrival of new information about the value of the asset. Using the cumulative distribution function of the normal distribution the price-weighted volume in each volume bucket is classified into buyer and seller initiated trades.

Since the introduction of VPIN, some papers including Wei et al. (2013), Andersen and Bondarenko (2014*a*,*b*) and Abad and Yagüe (2012) have raised concerns about the VPIN resulting in a considerable amount of debate and research that focus on the performance of VPIN in estimating information asymmetry risk.

In a recent paper, Kumar and Popescu (2013) extended the Copeland and Galai (1983) model to derive a new intra-day information asymmetry proxy called the **I**mplied **P**robability of **INF**ormed (PROBINF) trading, using dealer quoted bid and ask prices and market depth. The authors argue that dealer quotes may be considered as American Put and Call options since the dealer is obliged to trade at either the bid or ask price. Inverting the formula which equates the dealer expected gains and losses, they calculate PROBINF as the probability used for the calculation of the expected gains and losses. They find that PROBINF is highly correlated with PIN and hence an appropriate alternative to PIN.

In this chapter, our aim is to compute daily PIN using buyer and seller initiated trades that have been aggregated at a higher frequency. Sampling at high frequency has the benefit of reflecting fully the intra-day and inter-day dynamics of the information content of trades. A PIN estimated on trades sampled at relatively shorter time periods may be useful for learning about information asymmetry in a more dynamic way.

We infer daily PIN from buyer and seller initiated trades that has been sampled over 5 and 15-minute equally spaced time intervals of the trading day. There are 78 five (5) minute and 26 fifteen (15) minute time intervals in each trading day. We assume that the number of buyer and seller initiated trades within each time interval follow independent Poisson distributions. We focus only on the implementation of Easley et al. (2002) model.

As indicated earlier, Easley et al. (2011) introduced VPIN, a high frequency version of the probability of informed trading measure which has been adapted by finance professionals as a measure of order toxicity. We contrast the daily PIN estimate obtained from buyer and seller initiated trades sampled at 5 and 15 minute time intervals with VPIN. In what follows we give a brief description of the computational methodology of the VPIN.

Volume Synchronised Probability of Informed Trading

Easley et al. (2011) estimated VPIN using time-stamped transaction prices (P) and volume (V) as the basic input market microstructure variables. They sampled total volume and closing prices within 1 minute time intervals. The standard deviation $\sigma_{\Delta P}$, of changes in the 1 minute sample prices was also computed. They defined a variable called the volume bucket V^*, within which information is assumed to be homogeneous. Easley et al. (2011) proposed that the size of the volume bucket be calculated as $\frac{1}{L}$ $\times$ average daily volume over the sample period. They chose L to be 50 in their work. However, they argued that the VPIN is robust to any choice of L. From the 1 minute sampled data; volume is accumulated until the first volume bucket is reached. The corresponding price is picked as the closing price associated with the first volume bucket. Any excess volume is assigned to the next volume bucket. The procedure is continued until we have a series of pairs $\{(V_\tau, P_\tau)\}_{\tau=1}^{K}$ for K buckets.

Denoting the total traded volume in a volume bucket by V_i, Φ as the cumulative distribution function of the standard normal distribution and $t(\tau)$ the index of the last time-bar included in bucket τ, buy and sell volume in each bucket are determined as follows

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \Phi\Big(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\Big)$$

and

$$V_\tau^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \Big(1 - \Phi\Big(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\Big)\Big).$$

respectively. Denoting n as a rolling moving average window, the VPIN is calculated as

$$VPIN = \frac{\sum\limits_{\tau=1}^{n} |V_\tau^S - V_\tau^B|}{nV^*}. \tag{3.1}$$

Easley et al. (2011) calculated VPIN with $n = 50$ buckets where at each update point the first bucket is dropped, and a new bucket included. Thus starting with buckets 1 to 50 the first VPIN is calculated then buckets 2 to 51, 3 to 52 are used for calculating the second and third VPIN, and so on.

3.2 Empirical Analysis

Table 3.1 below is a summary of buyer and seller initiated trades sampled at 5 and 15-minute intervals. An observation from this table is that buyer and seller initiated trades at high frequency can also cause floating-point exceptions because in some intervals they are large.

	ASH		**IBM**	
	Buys	Sells	Buys	Sells
5 mins				
Min	0	0	0	0
Median	3	3	14	15
Mean	5	4	20	21
Max	179	241	647	574
15 mins				
Min	0	0	1	1
Median	9	9	45	49
Mean	14	14	62	65
Max	307	381	1480	964

Table 3.1 Summary of high frequency buy and sell trades

In Table 3.2 we provide the mean, median, first and third quartiles of the posterior distributions of the model parameters and PIN. The median of the probability of news event α, in ASH for the 5 and 15-minute intervals are 0.148 and 0.220 respectively. In the event of a negative or positive news arriving in a 5-minute interval, the ratio of the median informed trade arrivals μ, to the median liquidity trader arrivals (λ_s or λ_b) for ASH is $14/3 \approx 4$. In contrast the ratio is $24/9 \approx 2.5$ for the 15-minute sampled trades. Thus the order imbalance is higher in 5-minute intervals than in the 15-minute intervals. Although the negative effect of news at both sampling periods are similar in magnitude, the median probability 0.232 for an investor trading with an informed investor in the 5-minute interval is higher albeit marginal compared to 0.210 for the 15-minute interval sampled trades.

	5 min sampled trades						**15 min sampled trades**					
	Min	Q1	Median	Mean	Q3	Max	Min	Q1	Median	Mean	Q3	Max
ASH												
α	0.025	0.040	0.148	0.159	0.334	0.463	0.069	0.071	0.220	0.238	0.495	0.599
δ	0.059	0.108	0.508	0.507	0.882	0.940	0.068	0.099	0.515	0.509	0.898	0.932
μ	4	6	14	17	48	92	7	9	24	30	86	133
λ_s	0	1	3	4	8	13	1	3	9	10	26	43
λ_b	1	1	3	4	8	14	2	3	10	11	25	36
ω_1	0.008	0.011	0.060	0.082	0.253	0.424	0.022	0.025	0.089	0.122	0.380	0.504
ω_2	0.008	0.011	0.062	0.077	0.221	0.377	0.022	0.024	0.090	0.116	0.347	0.473
ω_3	0.536	0.665	0.851	0.840	0.959	0.974	0.400	0.504	0.779	0.761	0.928	0.930
PIN	0.073	0.130	0.232	0.233	0.356	0.466	0.077	0.117	0.210	0.216	0.360	0.474
IBM												
α	0.025	0.060	0.199	0.204	0.365	0.495	0.072	0.107	0.283	0.287	0.494	0.687
δ	0.038	0.095	0.498	0.501	0.910	0.959	0.047	0.091	0.498	0.500	0.903	0.939
μ	12	17	38	43	114	244	20	31	72	83	229	436
λ_s	3	6	16	17	36	89	9	16	48	52	114	273
λ_b	4	6	15	16	35	52	12	18	43	48	107	168
ω_1	0.008	0.012	0.089	0.103	0.279	0.374	0.026	0.028	0.122	0.142	0.365	0.529
ω_2	0.008	0.012	0.090	0.101	0.260	0.378	0.024	0.029	0.125	0.144	0.387	0.630
ω_3	0.504	0.634	0.800	0.795	0.939	0.974	0.312	0.505	0.716	0.712	0.892	0.927
PIN	0.078	0.105	0.190	0.193	0.298	0.386	0.061	0.099	0.174	0.178	0.288	0.352

Table 3.2 Summary of daily PIN estimates

Similar results are obtained for IBM where in the 5-minute interval, the ratio of informed to liquidity trader arrival rate is approximately 2.4 compared with 1.5 for the 15-minute interval trades. The results for IBM show that the PIN estimate is higher in the 5-minute interval sampled trades. In our sample, the average daily volume for ASH and IBM are approximately $533,000$ and $98,000$ respectively. The corresponding standard deviation of prices changes within 1, 5 and 15 minute time intervals for ASH are 0.0858, 0.1485 and 0.2442. That of IBM are also respectively 0.1087, 0.2298 and 0.3869. We used these figures as inputs for the VPIN calculation. A summary of VPIN over the entire sample period calculated from 1, 5 and 15 minute time bars are presented in Table 3.3. We chose $L = 50$ and $n = 50$ for each time bar.

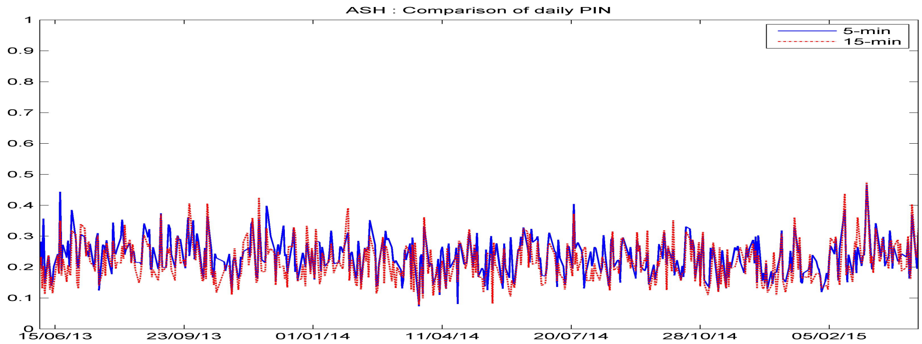
	ASH			**IBM**		
	1-min	5-min	15-min	1-min	5-min	15-min
Min	0.076	0.135	0.219	0.08	0.125	0.193
Q1	0.176	0.221	0.278	0.164	0.197	0.245
Median	0.350	0.363	0.373	0.316	0.336	0.346
Mean	0.358	0.371	0.385	0.331	0.343	0.354
Q3	0.577	0.569	0.555	0.568	0.525	0.504
Max	0.768	0.690	0.645	0.797	0.659	0.563

Table 3.3 Summary of VPIN

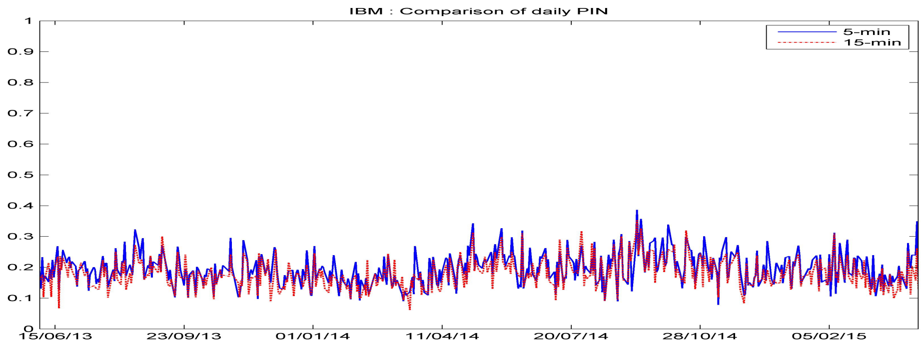
From Table 3.3, the VPIN of the 5 minute time bar for ASH over the sample period ranges between 0.135 and 0.690 with a corresponding average of 0.371. This is in contrast with the PIN for ASH which ranges between 0.073 and 0.466 obtained from the 5-minute buyer and seller initiated trades. Similarly the range (0.219 – 0.645) of the 15 minute time bar VPIN for ASH is higher than the PIN equivalent which is 0.077 to 0.474. The average VPIN of 15-minute time-bar for ASH is 0.385 compared with the posterior mean of 0.216 for the PIN. In the case of IBM, the PIN estimate from the 5-minute buyer and seller initiated trades is between 0.078 and 0.386. Part of this range overlaps with the lower part the range for the corresponding 5-minute time-bar VPIN which is 0.125 to 0.659.

Furthermore, it can be observed that the VPIN for ASH is higher than that of IBM. Comparing the VPIN with daily PIN computed from buyer and seller initiated trades, it can be seen that VPIN is consistently higher than the PIN for both assets. On the whole, we observe that the PIN estimate of ASH is greater than that of IBM. This is consistent with Easley et al. (1996) findings of infrequently traded assets having a higher risk of informed trading. The dynamic structure of VPIN is also quite different from the daily PIN.

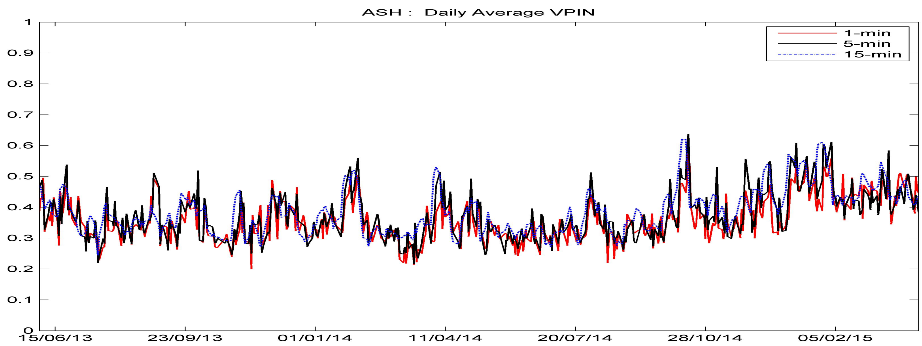
Figures 3.1a and 3.1b show daily PIN for ASH and IBM respectively. An equivalent version for the VPIN is also shown in Figures 3.1c and 3.1d. A series of 1, 5 and 15 time-bar VPIN within each trading day are averaged out to provide an estimate of a daily VPIN. These are shown in Figures 3.1c and 3.1d. Although both the PIN and VPIN are time-varying, the results show that the PIN is relatively more stable than the VPIN over the sample period. The daily PIN is calculated from 5 minute sampled buyer and seller initiated trades. Since the assumption is that over short time intervals order arrivals are constant, one would expect this stable behaviour of the daily PIN.



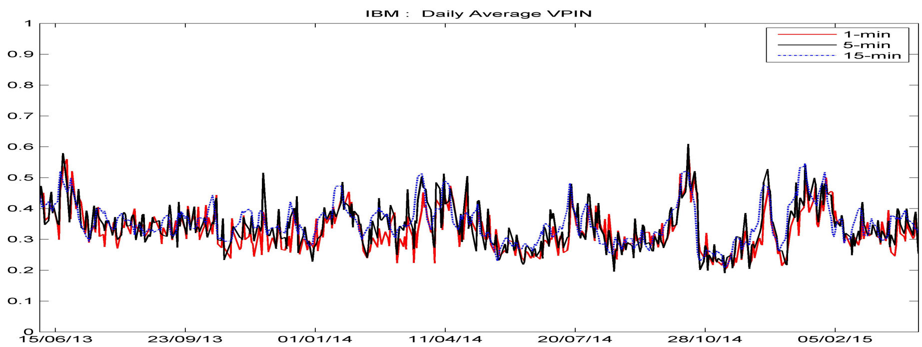
(a) ASH - daily PIN



(b) IBM - daily PIN



(c) ASH - daily average VPIN



(d) IBM - daily average VPIN

Fig. 3.1 Comparison of daily PIN and VPIN

3.3 Concluding Remarks

The empirical analysis carried out in this chapter indicates that the Bayesian methodology introduced in the previous chapter can be applied to high frequency sampled buyer and seller initiated trades. In addition the time series properties of daily PIN estimated from high frequency trades for our sample data are quite stable in comparison with VPIN estimates. Hence VPIN may provide more insight on the likelihood of Informed traders exploiting their information advantage.

Chapter 4

Learning About Informed Trading Via Volume - Spread Relationship

4.1 Introduction

It is well documented in the market microstructure literature that trade-related data such as the spread between bid and ask quotes, the number of transactions, traded volume, the duration between trades, trade direction and other derived variables contain valuable information that can be used to provide insights on the liquidity of an asset. To mention a few, research providing this theoretical and empirical finding include Kyle (1985), Manganelli (2005), Hasbrouck (1991), Easley and O'Hara (1992*b*), Easley and O'Hara (1987) and Dufour and Engle (2000). Bagehot's (1971) observations on the existence of information asymmetry in financial markets have spurred on extensive research that seeks to model and quantify information asymmetry risk. In chapter 2, we revisited the structural information model that uses order imbalance variables derived from trade-related data to infer the extent of informed trading in assets.

Traded volume of an asset has been used widely in the literature as a source of liquidity and information flow. The wide application of volume as a proxy for information flow is based on the mixture of distribution hypothesis (MDH) of Clark (1973). Clark posit that asset returns and volume are both driven by an unobserved information process. The unobserved information process generates trading decisions of investors. Based on this theory Lamoureux and Lastrapes (1994) use a mixture model for daily trading volume and returns to infer the process which jointly determines volume and asset returns. The joint relationship between volume and returns is widely studied in the finance literature. These include Bessembinder and Seguin (1993), Girard and Biswas (2007), Tauchen and Pitts (1983), Darrat et al. (2003) and Andersen (1996) to mention a few.

Another class of models that explores the information content of volume are based on the model of Kyle (1985). These models assume a linear relationship between price changes and order flow. To measure order flow, the direction of price change is used to sign volume. The order flow is used as a proxy for information in a regression model to estimate price impact.

We develop a joint model for volume and bid-ask spread. Volume is decomposed into informed and uninformed components. The informed and uninformed components of volume are associated with unobserved stochastic processes. The unobserved stochastic processes have corresponding effects on bid-ask spread. We use the unobserved stochastic processes as proxies for informed and uninformed trading. In the estimation of the model parameters we use a Bayesian methodology. In addition, we use the Bayesian method to infer the unobserved stochastic processes.

To learn about the composition of economic agents in the market and how the various agents reveal their private information through the trading process, Glosten and

Milgrom (1985) advanced the idea of sequential information hypothesis (SIAH). According to Glosten and Milgrom (1985), in a single asset market with a dedicated risk neutral market maker, investors arrive to the market randomly in a sequential manner. The market maker post prices at which she is willing to buy (bid price) and sell (ask price) an asset. Investors grouped into informed and uninformed are randomly chosen to trade. They can, however, choose to trade or not. As trading progress, the market maker revise the bid and ask prices. The bid and ask prices set by the market marker are conditional on her beliefs about the likelihood of the next purchase or sale coming from an informed or uninformed investor. Thus the sequential arrival and subsequent trading by the investors induce a Bayesian learning process for the market maker on the information held by investors. Once the private information held by the informed is incorporated into the price, there is an equilibrium. The learning process culminates in a series of alternating price discovery and equilibrium phases.

Drawing on the MDH and SIAH, Andersen (1996) as well as Mahieu and Bauer (1998) model the relationship between volume and asset returns in a joint mixture model. Conditional on the unobserved daily information that jointly determines volume and asset return volatility, they decompose volume into informed and uninformed components. Both components are assumed to be generated from a Poisson distribution. The informed component is driven by the unobserved daily information arrivals whereas the uninformed component is constant. The daily information arrival is modelled as a stochastic AR (1) process.

According to Mahieu and Bauer (1998) who also use the SIAH idea to decompose volume into informed and uninformed components, a small part of the daily trading volume is directly related to the unobserved private information. Other studies including Liesenfeld (2001) and Watanabe (2000) that extend the Andersen (1996) model assume that daily trading volume is normally distributed instead of the Poisson assumption in Andersen (1996). The work in chapter 2 is based on a model by Easley and O'Hara (1992*b*) who provided insight on the link between trading volume and unobserved private information. In their work, they proposed that volume has two components of which one can be forecasted. They found that bid-ask spread and the forecast-able component of volume are negatively related. In the model, private information signals cause volume to deviate from its normal level. Bessembinder (1994), Danielsson and Payne (2001), and Jorion (1996) among many others have employed the idea in Easley and O'Hara (1992*b*) to decompose volume into predictable and unpredictable components. The expected volume is predetermined while the unpredictable component of volume is more likely to be correlated with information asymmetry.

The assumption underlying the decomposition of volume in the papers cited above has been that daily volume is generated from either a normal or Poisson distribution. However, due to increased trading volume observed in markets in recent times, the normal and Poisson distributions are potentially inappropriate probability distributions for traded volume.

Volume and bid-ask spread are variables that measure different aspects of liquidity. Because of that, the relationship between the bid-ask spread and volume has also been studied extensively in the market microstructure literature. Exploiting the relationship between daily dollar volume and closing relative spread, Hallin et al. (2011) infer the common unobservable process that drive both volume and bid-ask spread. They found that the common unobserved process driving volume and bid-ask spread is effected by a unique shock which is a natural measure of market liquidity. According to Glosten and Milgrom (1985), the bid-ask spread is a function of informed trading, and that increased informed trading induce high ask prices and low bid prices. This leads to wider spreads that are intended to cover potential losses that might be incurred while an investor trades with another who may be informed. Studies like Bollen et al. (2004) explore the relationship between volume and spread with a linear regression model in which volume is used as a predictor of bid-ask spread. In this model an increase in daily volume is associated with a reduction in bid-ask spread. Copeland and Galai (1983) found that the bid-ask spread increases with market activity measures. Hence the bid-ask spread is positively related to volume. Li and Wu (2006) also studied the relationship between volume and bid-ask spread where they found a dynamic feedback relationship between informed trading volume and bid-ask spread. They found that informed trading is a significant predictor of bid-ask spread when compared with order imbalance variables that are used as measures of informed trading in spread regression models. A number of research papers including Glosten and Milgrom (1985), Glosten and Harris (1988), George et al. (1991) Madhavan et al. (1997) and Huang and Stoll (1997) decompose bid-ask spread into components relating to inventory carry cost, order processing cost and information asymmetry cost.

From the above, we infer that the dynamic relationship between volume and bid-ask spread may be useful for learning about the latent processes that drive informed and uninformed trading. Our aim is to exploit the theoretical relations between volume and bid-ask spread to estimate informed and uninformed trading. Andersen (1996) and the majority of the papers that extend it, assume that changes in daily volume are primarily due to fluctuations in informed trading while uninformed trading volume is time-invariant. According to Admati and Pfleiderer (1988) periods of high trading activity are associated with increased trading from both informed and uninformed traders. Thus uninformed trading may not necessarily be constant as has been modelled in Andersen (1996), Mahieu and Bauer (1998), Abanto-Valle et al. (2010) and others. The volume decomposition approach in our model relies on the predicted link between the components of volume and the bid-ask spread. The bid-ask spread is expected to be associated with the relative composition of informed and uniformed traders.

In the estimation of our model, we initially assume knowledge of the stochastic processes that drive unobserved informed and liquidity trading and then use a Gibbs Sampler to draw samples from the posterior distributions of the model parameters while conditioning on the observed data and the latent stochastic processes. Given the draws of the model parameters and the observed data, we employ a Kalman Filter to estimate optimal *mean* and *variance* of the latent stochastic processes for each trading day. We then use these moments to sample from the posterior distribution of the latent stochastic processes. This process of conditional sampling of model parameters and the latent stochastic processes is continued for a large number of times until the convergence of the Markov chain is achieved. The posterior means of the model parameters are then reported as estimates for the parameters.

The Bayesian estimation method allows for the incorporation of the history of volume and bid-ask spread as conditioning information. By conditioning on the history of volume and bid-ask spread we update our knowledge about the latent processes that drive informed and liquidity trading. Another advantage of our model structure is that we are able to extract the temporal information asymmetry through the joint relationship between volume and bid-ask with no recourse to trade classification. This is in contrast to information-based models such as the price impact of Kyle (1985), PIN and VPIN models of chapter 2 all of which rely on Lee and Ready (1991) trade classification algorithms. The Lee and Ready (1991) algorithm is used to infer the direction of trade. The direction of trade is then used to derive buyer and seller initiated trades used in the PIN model. Similarly, the trade direction obtained from the algorithm is used to create buyer and seller volumes for the price impact and VPIN models. However in our model we identify latent liquidity and informed trading effects through the theory which links volume and bid-ask spread. Also, we can take into account parameter uncertainty and the uncertainty about the latent stochastic processes through the Bayesian estimation approach.

4.2 The Model

We assume that in the absence of information which will effect the fundamental value of an asset, both informed and uniformed investors trade a certain amount of shares μ^i and μ^l respectively. Informed traders have private signals about the fundamental value of the asset. This private signal is driven by an unobservable stochastic process. Uninformed traders on the other hand are aware of the presence of informed traders and hence use the trading process to make inferences about their trading decisions. Finally, the informed and uniformed interpret public news differently and thus the innovations in their respective volume due to public news are random. In our model ε_t^i and ε_t^l represent the innovation in informed and uninformed processes. This assumption is consistent with Kandel and Pearson (1995) who argue that informed and uninformed investors interpret public news differently. Denoting informed and uninformed trading volume by V_t^i and V_t^l respectively, we decompose observed daily traded volume V_t as follows

$$\begin{aligned} V_t &= V_t^i + V_t^l \\ &= V_t^l\left(1 + \frac{V_t^i}{V_t^l}\right) \end{aligned} \tag{4.1}$$

For empirical estimation we model the logarithm of volume and hence have the following by taking logarithms of equation 4.1

$$\ln V_t = \ln V_t^l + \ln\left(1 + \frac{V_t^i}{V_t^l}\right).$$

We assume that $\ln V_t^l = \mu^l + \tau_t + \varepsilon_t^l$ and $\ln\left(1 + \frac{V_t^i}{V_t^l}\right) = \mu^i + h_t + \varepsilon_t^i$. Since the term $\ln\left(1+\frac{V_t^i}{V_t^l}\right)$ is a monotonic transformation of the ratio of informed to uninformed trading $\frac{V_t^i}{V_t^l}$, it is not effected by changes in overall volume. Based on these assumptions we treat volume as a stochastic process whose logarithmic transform can be modelled as

$$\begin{aligned}
\ln V_t &= \ln V_t^l + \ln\left(1 + \frac{V_t^i}{V_t^l}\right), \\
&= \mu^l + \mu^i + \tau_t + h_t + \varepsilon_t^i + \varepsilon_t^l \\
&= \mu + \tau_t + h_t + \varepsilon_t.
\end{aligned} \tag{4.2}$$

In the above, τ_t measures uninformed trading while h_t also measures the level of informed trading relative to uninformed trading. We assume that τ_t and h_t are independent stochastic processes. In addition τ_t is conditionally independent of its past given its most recent value. Likewise, given the most recent value h_{t-1}, h_t is conditionally independent of its history. Hence we model h_t and τ_t with the following AR (1) processes

$$\tau_t = \phi_\tau \tau_{t-1} + \varepsilon_{\tau,t}, \qquad \varepsilon_{\tau,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_\tau}}{\omega_t}\right) \tag{4.3}$$

$$h_t = \phi_h h_{t-1} + \varepsilon_{h,t}, \qquad \varepsilon_{h,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_h}}{\delta_t}\right), \tag{4.4}$$

where $\mathcal{N}(.)$ is the normal distribution. Our choice of model for the latent information processes τ_t and h_t is in line with Easley et al. (2008) and other authors who model the arrival intensities of informed and liquidity traders as first order auto-regressive processes. Persistence in the informed and uninformed trader information are given by the parameters ϕ_τ and ϕ_h respectively. The terms ω_t and δ_t in the error components of the latent processes are gamma distributions. They have been introduced to account for the possibility of fat tails. Additionally we assume that τ_t and h_t are uncorrelated with expectations equal to zero.

Let the ask and bid prices be denoted by P_t^a and P_t^a respectively. Then we define the relative spread for the ith time-stamped transaction as $RS_{t,i} = \frac{P_t^a - P_t^b}{0.5(P_t^a + P_t^b)}$. If M_0 is the number of time-stamped relative spreads in trading day t, we denote the average relative spread for trading day t as $S_t = \sum_{i=1}^{M_0} RS_{t,i}/M_0$. Let $y_t = \ln V_t$ and $x_t = \ln S_t$ be the logarithm of daily volume and average relative bid-ask spread of an asset respectively. Theoretical literature including Easley and O'Hara (1992a) as well as Glosten and Milgrom (1985) suggests that informed trading induces a wider bid-ask spread whereas as uninformed trading results in a narrower bid-ask spread. Thus the effect of informed trading on bid-ask spread will be positive. Likewise the effect of uniformed trading on bid-ask spread will be negative. We consider the following model

$$\mathbf{A1}: \quad y_t = \mu + \tau_t + h_t + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_y}}{\kappa_t}\right) \tag{4.5a}$$

$$x_t = \eta + \alpha_\tau \tau_t + \alpha_h h_t + \varepsilon_{x,t}, \qquad \varepsilon_{x,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_x}}{\varphi_t}\right) \tag{4.5b}$$

$$\tau_t = \phi_\tau \tau_{t-1} + \varepsilon_{\tau,t}, \qquad \varepsilon_{\tau,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_\tau}}{\omega_t}\right) \tag{4.5c}$$

$$h_t = \phi_h h_{t-1} + \varepsilon_{h,t}, \qquad \varepsilon_{h,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_h}}{\delta_t}\right), \tag{4.5d}$$

where $\alpha_\tau < 0$, $\alpha_h > 0$. The error terms $\varepsilon_{y,t}$, $\varepsilon_{x,t}$, $\varepsilon_{\tau,t}$ and $\varepsilon_{h,t}$ are assumed to be independent of each other. The term $\exp(\eta)$ is the minimum average relative bid-ask spread. The effect of informed and uninformed components of volume are modelled through α_h and α_τ respectively. We assume that, conditional on τ_t and h_t, volume and average bid-ask spread are independent.

The distribution of financial data are characterised by fat-tails. This mean that the data generating process is not normally distributed. To account for fat-tails in our models, we use a scale mixture of normals for the distribution of the error terms $\varepsilon_{y,t}$, $\varepsilon_{x,t}$, $\varepsilon_{\tau,t}$ and $\varepsilon_{h,t}$. The terms κ_t and φ_t which follow gamma distributions are used to achieve this scale mixture. Geweke (1993) proved that this formulation is equivalent to a specification that assume a Student-t distribution for the error terms. By introducing a t-distribution in the shock structure, we are able to capture extreme changes in the observed and unobserved processes. We therefore account effectively for potential extreme events. Empirical research suggests that the lags y_{t-1} and x_{t-1} of volume and bid-ask spread respectively convey information. Hence a natural direction of investigation is to find out whether τ_t and h_t will have significant impact on bid-ask spread if the lags are included in the model. We carry out this investigation by considering the following alternative model:

$$\mathbf{A2}: \quad y_t = \mu^* + \rho_y y_{t-1} + \tau_t + h_t + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_y}}{\kappa_t}\right) \tag{4.6a}$$

$$x_t = \eta^* + \rho_x x_{t-1} + \alpha_\tau \tau_t + \alpha_h h_t + \varepsilon_{x,t}, \qquad \varepsilon_{x,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_x}}{\varphi_t}\right) \tag{4.6b}$$

$$\tau_t = \phi_\tau \tau_{t-1} + \varepsilon_{\tau,t}, \qquad \varepsilon_{\tau,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_\tau}}{\omega_t}\right) \tag{4.6c}$$

$$h_t = \phi_h h_{t-1} + \varepsilon_{h,t}, \qquad \varepsilon_{h,t} \sim \mathcal{N}\left(0, \frac{\sigma^2_{\varepsilon_h}}{\delta_t}\right). \tag{4.6d}$$

The expectations of y_t and x_t in model A2 are $\mu = {}^{\mu^*}/_{(1-\rho_y)}$ and $\eta = {}^{\eta^*}/_{(1-\rho_x)}$ respectively. We our cast our models in state space form and then estimate them in a Bayesian setting. Models A1 and A2 can be represented in a state-space form as

$$Y_t = c + AX_t + Z\beta_t + u_t \qquad u_t \sim \mathcal{MVN}(0, W_t) \tag{4.7a}$$

$$\beta_t = d + T\beta_{t-1} + v_t \qquad v_t \sim \mathcal{MVN}(0, R_t) \tag{4.7b}$$

where $\mathcal{MVN}(.)$ is the density function of the multivariate normal distribution, $Y_t = (y_t, x_t)$ is the observed variables and the unobserved state variable(s) to be inferred is $\beta_t = (\tau_t, h_t)$. Other exogenous variables are collected in the vector X_t. Equation 4.7a, known as the observation equation links the observed data Y_t to the latent state β_t while the state transition represented by equation 4.7b defines how the latent state evolves over time. Harvey et al. (1992) proposes an alternative representation which also takes into account the fat tails in models A1 and A2. Their representation is not materially different from our representation and will lead to similar results.

In equation 4.7 we can only observe $Y_{1:N} = (Y_1, \ldots, Y_N)$. However we are interested in inferring the fixed parameter set $\Theta = (d, c, W_t, R_t, T, Z, A)$ as well as the latent state vector $\boldsymbol{\beta_{1:N}} = (\beta_1, \ldots, \beta_N)$. We have the following design matrices

$Z = \begin{pmatrix} 1 & 1 \\ \alpha_\tau & \alpha_h \end{pmatrix}$, $T = \begin{pmatrix} \phi_\tau & 0 \\ 0 & \phi_h \end{pmatrix}$, $W_t = \begin{pmatrix} \frac{\sigma_y^2}{\kappa_t} & 0 \\ 0 & \frac{\sigma_x^2}{\varphi_t} \end{pmatrix}$, $R_t = \begin{pmatrix} \frac{\sigma_{\varepsilon_\tau}^2}{\omega_t} & 0 \\ 0 & \frac{\sigma_{\varepsilon_h}^2}{\delta_t} \end{pmatrix}$, and $d = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ for both models. We also have $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $c = \begin{pmatrix} \mu \\ \eta \end{pmatrix}$, and $A = \begin{pmatrix} \rho_y & 0 \\ 0 & \rho_x \end{pmatrix}$, $c = \begin{pmatrix} \mu^* \\ \eta^* \end{pmatrix}$ for models A1 and A2 respectively.

The state-space representation enhances the modelling of the temporal relationship between observed variables and the unobservable processes in a more flexible manner. The estimation method also provides an approximation to the marginal posterior distributions of model parameters and the unobserved state. It enhances the measurement of the uncertainty associated with the estimate of the state variable and model parameters. To estimate the parameters Θ of the model in a Bayesian framework, we specify a prior distribution $P(\Theta)$. If $P(Y_{1:N}|\Theta, \boldsymbol{\beta_{1:N}})$ is the likelihood function of the observed data conditional on the latent state $\boldsymbol{\beta_{1:N}}$ and parameters Θ, then using Bayes' theorem, the joint posterior distribution of the model parameter set Θ and the latent state process conditional on the observed data is given as

$$P\left(\Theta, \beta_{1:N}|Y_{1:N}\right) \propto P(Y_{1:N}|\boldsymbol{\beta_{1:N}}, \Theta)P(\beta_{1:N}|\Theta)P(\Theta) \tag{4.8}$$

This joint posterior distribution is highly dimensional and most often analytically intractable and hence very complicated to work with. This makes direct simulation from the joint posterior distribution hard to perform. We use MCMC algorithms to explore the posterior density in equation (4.8). The idea behind the MCMC is to break the highly dimensional vectors of latent variables and parameters into pieces. Conditional on the latent process we use the Gibbs Sampler to infer the model parameters.

Gibbs Sampling Algorithm

We estimate the marginal posterior distributions using a Gibbs Sampler. This algorithm is used to draw samples from the pair $(\Theta, \boldsymbol{\beta})$, conditioning on one element of the pair and the observed series at a time. The steps below are taken to explore the posterior distribution of $p(\Theta, \beta_{1:N}|Y_{1:N})$.

1. We set the initial values of the parameter set Θ and $\boldsymbol{\beta_{1:N}}$.
2. Sample $\boldsymbol{\beta_{1:N}}$ using the Kalman Filter and the Carter and Kohn (1994) Forward Filtering Backward Sampling (FFBS) Algorithm.
3. Sample the Θ conditional on $Y_{1:N}$ and $\boldsymbol{\beta_{1:N}}$.
4. Steps 2 and 3 are repeated many times until we achieve convergence.

Full Conditional Posterior Distributions

We select the following conjugate prior distributions for the model parameters:

$$
\begin{array}{llll}
\eta \sim \mathcal{N}(\mu_\eta, \sigma^2_\eta) & \mu \sim \mathcal{N}(\mu_\mu, \sigma^2_\mu) & \omega_t \sim \mathcal{G}a\left(\frac{\nu_\tau}{2}, \frac{\nu_\tau}{2}\right) & \nu_\tau \sim \mathcal{G}a(a, b) \\
\rho_y \sim \mathcal{N}(\mu_{\rho_y}, \sigma^2_{\rho_y}) & \alpha_\tau \sim \mathcal{N}\left(\mu_{\alpha_\tau}, \sigma^2_{\alpha_\tau}\right) & \delta_t \sim \mathcal{G}a\left(\frac{\nu_h}{2}, \frac{\nu_h}{2}\right) & \sigma^2_{\varepsilon_\tau} \sim \mathcal{IG}(r, s) \\
\rho_x \sim \mathcal{N}(\mu_{\rho_x}, \sigma^2_{\rho_x}) & \alpha_h \sim \mathcal{N}\left(\mu_{\alpha_h}, \sigma^2_{\alpha_h}\right) & \nu_y \sim \mathcal{G}a(a, b) & \sigma^2_{\varepsilon h} \sim \mathcal{IG}(r, s) \\
\phi_\tau \sim \mathcal{N}(\mu_{\phi_\tau}, \sigma^2_{\phi_\tau}) & \kappa_t \sim \mathcal{G}a\left(\frac{\nu_y}{2}, \frac{\nu_y}{2}\right) & \nu_x \sim \mathcal{G}a(a, b) & \sigma^2_{\varepsilon_y} \sim \mathcal{IG}(r, s) \\
\phi_h \sim \mathcal{N}(\mu_{\phi_h}, \sigma^2_{\phi_h}) & \varphi_t \sim \mathcal{G}a\left(\frac{\nu_x}{2}, \frac{\nu_x}{2}\right) & \nu_h \sim \mathcal{G}a(a, b) & \sigma^2_{\varepsilon_x} \sim \mathcal{IG}(r, s),
\end{array}
$$

where $\mathcal{G}a(.)$ and $IG(.)$ are the distributions of the gamma and inverse gamma random variables respectively. With conjugate prior distributions, the kernel of the resulting posterior distributions are standard probability densities from which it is easy to draws samples using a Gibbs sampler. We set the hyper-parameters of the priors for ν_y, ν_x, ν_τ, and ν_h as follows $a = 0.16$ and $b = 0.04$. This sets the prior degrees of freedom for the distribution of the error terms in the models at 4 with a variance of 100, which reflects how uncertain we are about the true values of these parameters. Similarly we choose the following $\mu_{\alpha_h} = \mu_\eta = \mu_{\phi_\tau} = \mu_{\alpha_\tau} = \mu_{\phi_h} = \mu_\mu = 0$ and $\sigma^2_{\alpha_h} = \sigma^2_\eta = \sigma^2_{\phi_\tau} = \sigma^2_{\alpha_\tau} = \sigma^2_{\phi_h} = \sigma^2_\mu = \sigma^2_{\phi_h} = \sigma^2_\mu = 100$. These hyper-parameter choices mean that we are assuming that α_h, α_τ, ϕ_h, and ϕ_τ can take on any value. The uncertainty is expressed through the choice of a large prior variance 100. Finally, the hyper-parameters of $\sigma^2_{\varepsilon_\tau}$, $\sigma^2_{\varepsilon_h}$, $\sigma^2_{\varepsilon_y}$, and $\sigma^2_{\varepsilon_x}$ are set to be $r = s = 1$ which leads to relatively flat priors. Flat priors place little weight on any specific part of the parameter space. Thus we assume relatively no knowledge about the level and uncertainty of the model parameters. These choices imply that the estimation places more weight on the information held in the observed data. The respective full conditional posterior distributions of the parameters in model A1 are given as follows

$$\sigma^2_y \sim \mathcal{IG}\left(r_0 + \frac{N}{2} \quad, s_0 + 0.5\sum_{t=1}^{N} \kappa_t[y_t - \mu - \tau_t - h_t]^2\right)$$

$$\sigma^2_x \sim \mathcal{IG}\left(r_1 + \frac{N}{2} \quad, s_1 + 0.5\sum_{t=1}^{N} \varphi_t[x_t - \eta - \alpha_\tau\tau_t - \alpha_h h_t]^2\right)$$

$$\kappa_t \sim \mathcal{G}a\left(\frac{\nu_y+1}{2} \quad, \frac{\nu_y}{2} + \frac{1}{2\sigma^2_y}[y_t - \mu - \tau_t - h_t]^2\right)$$

$$\varphi_t \sim \mathcal{G}a\left(\frac{\nu_x+1}{2} \quad, \frac{\nu_x}{2} + \frac{1}{2\sigma^2_x}[x_t - \eta - \alpha_\tau\tau_t - \alpha_h h_t]^2\right)$$

$$\alpha_\tau \sim \mathcal{N}\left(\left(\frac{1}{\sigma^2_{\alpha_\tau}} + \frac{\sum_{t=1}^{N}\varphi_t\tau_t^2}{\sigma^2_x}\right)^{-1}\left(\frac{\mu_{\alpha_\tau}}{\sigma^2_{\alpha_\tau}} + \frac{\sum_{t=1}^{N}\varphi_t\tau_t(x_t-\eta-\alpha_h h_t)}{\sigma^2_x}\right), \left(\frac{1}{\sigma^2_{\alpha_\tau}} + \frac{\sum_{t=1}^{N}\varphi_t\tau_t^2}{\sigma^2_x}\right)^{-1}\right)$$

$$\alpha_h \sim \mathcal{N}\left(\left(\frac{1}{\sigma^2_{\alpha_h}} + \frac{\sum_{t=1}^{N}\varphi_t h_t^2}{\sigma^2_x}\right)^{-1}\left(\frac{\mu_{\alpha_h}}{\sigma^2_{\alpha_h}} + \frac{\sum_{t=1}^{N}\varphi_t h_t(x_t-\eta-\alpha_\tau\tau_t)}{\sigma^2_x}\right), \left(\frac{1}{\sigma^2_{\alpha_h}} + \frac{\sum_{t=1}^{N}\varphi_t h_t^2}{\sigma^2_x}\right)^{-1}\right)$$

$$\eta \sim \mathcal{N}\left(\left(\frac{1}{\sigma^2_\eta} + \frac{\sum_{t=1}^{N}\varphi_t}{\sigma^2_x}\right)^{-1}\left(\frac{\mu_\eta}{\sigma^2_\eta} + \frac{\sum_{t=1}^{N}\varphi_t(x_t-\alpha_\tau\tau_t-\alpha_h h_t)}{\sigma^2_x}\right), \left(\frac{1}{\sigma^2_\eta} + \frac{\sum_{t=1}^{N}\varphi_t}{\sigma^2_x}\right)^{-1}\right)$$

$$\mu \sim \mathcal{N}\left(\left(\frac{1}{\sigma_\mu^2}+\frac{\sum_{t=1}^{N}\kappa_t}{\sigma_y^2}\right)^{-1}\left(\frac{\mu_\mu}{\sigma_\mu^2}+\frac{\sum_{t=1}^{N}\kappa_t[y_t-\tau_t-h_t]}{\sigma_y^2}\right),\left(\frac{1}{\sigma_\mu^2}+\frac{\sum_{t=1}^{N}\kappa_t}{\sigma_y^2}\right)^{-1}\right).$$

$$\sigma_\tau^2 \sim \mathcal{IG}\left(r_2+\frac{N-1}{2}\quad, s_2+0.5\sum_{2=1}^{N}[\tau_t-\phi_\tau\tau_{t-1}]^2\right)$$

$$\sigma_h^2 \sim \mathcal{IG}\left(r_3+\frac{N-1}{2}\quad, s_3+0.5\sum_{2=1}^{N}[h_t-\phi_h h_{t-1}]^2\right)$$

$$\phi_h \sim \mathcal{N}\left(\left(\frac{1}{\sigma_{\phi_h}^2}+\frac{\sum_{t=2}^{N}h_{t-1}^2}{\sigma_h^2}\right)^{-1}\left(\frac{\mu_{\phi_h}}{\sigma_{\phi_h}^2}+\frac{\sum_{t=2}^{N}h_t h_{t-1}}{\sigma_h^2}\right),\left(\frac{1}{\sigma_{\phi_h}^2}+\frac{\sum_{t=2}^{N}h_{t-1}^2}{\sigma_h^2}\right)^{-1}\right)$$

$$\phi_\tau \sim \mathcal{N}\left(\left(\frac{1}{\sigma_{\phi_\tau}^2}+\frac{\sum_{t=2}^{N}\tau_{t-1}^2}{\sigma_\tau^2}\right)^{-1}\left(\frac{\mu_{\phi_\tau}}{\sigma_{\phi_\tau}^2}+\frac{\sum_{t=2}^{N}\tau_t\tau_{t-1}}{\sigma_\tau^2}\right),\left(\frac{1}{\sigma_{\phi_\tau}^2}+\frac{\sum_{t=2}^{N}\tau_{t-1}^2}{\sigma_\tau^2}\right)^{-1}\right)$$

$$\omega_t \sim \mathcal{G}a\left(\frac{\nu+1}{2}\quad, \frac{\nu_\tau}{2}+\frac{1}{2\sigma_\tau^2}[\tau_t-\phi_\tau\tau_{t-1}]^2\right)$$

$$\delta_t \sim \mathcal{G}a\left(\frac{\nu+1}{2}\quad, \frac{\nu_h}{2}+\frac{1}{2\sigma_h^2}[h_t-\phi_h h_{t-1}]^2\right).$$

However, since the full conditional posterior distributions of ν_y, ν_x, ν_τ and ν_h do not fall within any class of the standard probability distributions, we use the Adaptive Metropolis Hastings Algorithm which was introduced in the previous chapter to sample from the following posterior density functions

$$\nu_y|\kappa_t \propto \nu_y^{a-1}e^{-b\nu_y}e^{-\frac{\nu_y}{2}\sum_{t=1}^{N}\kappa_t}\prod_{t=1}^{N}\kappa_t^{\frac{\nu_y}{2}-1} \qquad \nu_\tau|\omega_t \propto \nu_\tau^{f-1}e^{-g\nu_\tau}e^{-\frac{\nu_\tau}{2}\sum_{t=1}^{N}\omega_t}\prod_{t=1}^{N}\omega_t^{\frac{\nu_\tau}{2}-1}$$

$$\nu_x|\psi_t \propto \nu_x^{c-1}e^{-d\nu_x}e^{-\frac{\nu_x}{2}\sum_{t=1}^{N}\psi_t}\prod_{t=1}^{N}\psi_t^{\frac{\nu_x}{2}-1} \qquad \nu_h|\delta_t \propto \nu_h^{k-1}e^{-l\nu_h}e^{-\frac{\nu_h}{2}\sum_{t=1}^{N}\delta_t}\prod_{t=1}^{N}\delta_t^{\frac{\nu_h}{2}-1}.$$

Sampling β

We utilise the Kalman Filter (Kalman, 1960), which is a recursive algorithm that provides an optimal estimate of the unobserved state variable β_t conditional on the observed data and other model parameters.

The Kalman Filter algorithm minimises the mean squared errors of the estimated state vector. The Kalman Filter and the Carter and Kohn (1994) FFBS algorithm are used to draw samples from the full conditional distribution of $\beta_1, \ldots, \beta_N$. There are several approaches to explaining and deriving the Kalman Filter. However, in this thesis, we focus on the Bayesian interpretation of the filter. Our goal is to obtain the posterior distribution of the state vector $\boldsymbol{\beta} = \beta_{1:N}$ conditional on knowledge of the model parameter set Θ and observed series. Conditioning on the observed series $y_1, \ldots, y_t$, the posterior distribution of the unobserved state at time t can be obtained recursively. We achieve this by noting that from from Bayes' theorem, the posterior distribution of the unobserved state β_t can be written as

$$\begin{aligned} p(\beta_t|y_t, Y_{1:t-1}, \Theta) &= \frac{p(y_t|\beta_t, Y_{1:t-1}, \Theta)p(\beta_t|Y_{1:t-1}, \Theta)}{p(y_t|Y_{1:t-1}, \Theta)} \\ &\propto p(y_t|\beta_t, Y_{1:t-1}, \Theta)p(\beta_t|Y_{1:t-1}, \Theta), \end{aligned} \tag{4.9}$$

where $p(y_t|\beta_t, Y_{1:t-1}, \Theta)$ is the likelihood function and $p(\beta_t|Y_{1:t-1}, \Theta)$, the prior distribution of β_t.

Kalman Filter Algorithm

At time $t-1$ we assume that given observed data up to time $t-1$, the state vector has a Gaussian distribution with mean $\beta_{t-1|t-1}$ and variance $P_{t-1|t-1}$. Thus $\beta_{t-1}|Y_{1:t-1} \sim \mathcal{N}(\beta_{t-1|t-1}, P_{t-1|t-1})$.

1. We initialise the first and second moments of the distribution of the state vector as $\beta_{0|0}$ and $P_{0|0}$. At time t before observing Y_t, the prior distribution of the state vector is $\beta_t|Y_{1:t-1} \sim \mathcal{N}(d + T\beta_{t-1|t-1}, TP_{t-t|t-1}T' + R_t)$, where $\beta_{t|t-1} = d + T\beta_{t-1|t-1}$ and $P_{t|t-1} = TP_{t-t|t-1}T' + R_t$.

2. After observing the Y_t, we update our knowledge about the state vector using the likelihood function. The posterior distribution of the state vector after observing Y_t then becomes $\beta_t|Y_t \sim \mathcal{N}(\beta_{t|t}, P_{t|t})$, where $\beta_{t|t}$ and $P_{t|t}$ are computed using the following recursions relations

   $\hat{Y}_t = c + AX_t + Z\beta_{t|t-1}$

   $F = ZP_{t|t-1}Z' + W_t$

   $K_t = P_{t|t-1}Z'F^{-1}$

   $\beta_{t|t} = \beta_{t|t-1} + K_t(Y_t - \hat{Y}_t)$

   $P_{t|t} = P_{t|t-1} - K_tZP_{t|t-1}.$

   K_t is known as the *Kalman gain* and the quantity $(Y_t - \hat{Y}_t)$ is the prediction error. At each time step, the previous *a posteriori* estimate of the state vector is used as the current *a priori* estimate.

3. Once we have the filtered state vector $\hat{\beta}_t$, we use the Carter and Kohn (1994) FFBS algorithm to sample the state vector from its full conditional distribution. To sample from the posterior distribution of the state vector β_t, we start with the filtered mean $\beta_{N|N}$ and variance $P_{N|N}$ at N as the posterior mean and variance.

   Thus conditional on the observed values of $Y_{1:N}$, β_N is normally distributed with mean $\beta_{N|N}$ and variance $P_{N|N}$. The posterior distribution of the state vector β_t for trading times $t = N-1, \ldots, 2, 1,$ is normal with mean and variance respectively given as

$$\mu = \Sigma \times \left(T'R_t^{-1}\beta_{t+1} - T'R_t^{-1}d + \Sigma^{-1}\hat{\beta}_t\right)$$

   and

$$\Sigma = \left(T'R_t^{-1}T + \Sigma^{-1}\right)^{-1}.$$

β_t is the sampled state vector at time $t+1$, $\hat{\beta}_t$ and Σ are the means and the variance-covariance matrix obtained from the Kalman filter at time t. One advantage of using the Kalman Filter is that it accumulates information about the observed series as it moves forward. This accumulated information is stored in mean and variance of the latent state. Details of how these quantities have been derived are provided in Carter and Kohn (1994).

Improving Sampling Efficiency

There are two ways of representing the state space model in equation 4.7. These are the *centered* and *non-centered* representations. The choice of representation in the estimation may have an effect on the simulation efficiency of the Gibbs Sampler. In our specific case, the centered representation is given as

$$Y_t = AX_t + Z\beta_t^* + u_t, \qquad u_t \sim \mathcal{MVN}(0, W_t) \tag{4.10a}$$

$$\beta_t^* = (I - T)Z^{-1}c + T\beta_{t-1}^* + v_t \qquad v_t \sim \mathcal{MVN}(0, R_t), \tag{4.10b}$$

while the non-centered representation is also given as

$$Y_t = c + AX_t + Z\beta_t + u_t, \qquad u_t \sim \mathcal{MVN}(0, W_t) \tag{4.11a}$$

$$\beta_t = T\beta_{t-1} + v_t \qquad v_t \sim \mathcal{MVN}(0, R_t), \tag{4.11b}$$

where $T = \begin{pmatrix} \phi_\tau & 0 \\ 0 & \phi_h \end{pmatrix}$, $Y_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$, $Z = \begin{pmatrix} 1 & 1 \\ \alpha_\tau & \alpha_h \end{pmatrix}$, $c = \begin{pmatrix} \mu \\ \eta \end{pmatrix}$ and $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Two main issues (see Strickland et al. (2008) and Kim et al. (1998)) arise when these representations are used for estimating model parameters. Firstly, the standard representation is affected by simulation inefficiency which leads to non-convergence of the MCMC for model parameters. Secondly, if the non-centered representation is the choice for estimation, the MCMC efficiency is hampered if the volatility of the latent process in equation 4.10b is very small or the latent process itself is highly persistent. Yu and Meng (2011) introduced a novel approach known as the Ancillarity-Sufficiency Interweaving Strategy (ASIS) to boost the simulation efficiency. The intuition behind ASIS is to estimate the model parameters by interweaving between the centered and non-centered representations through the transformation from β_t to β_t^* and vice-versa. Kastner and Frühwirth-Schnatter (2014) provides an example using the stochastic volatility model. In this chapter and the next, we have implemented the ASIS to improve the simulation efficiency for the sampling of μ and η.

4.3 Empirical Analysis

Details of the empirical estimation are provided in this section. We use volume and bid-ask spread of the cleaned data which we described in chapter 2. Table 4.1 is a summary of the logarithm of total daily volume and daily average relative bid-ask spread which we use in this chapter. The Gibbs Sampler was run for $100,000$ sweeps with a burn-in period of $30,000$ for both assets from 3^{rd} June 2013 to 15^{th} April 2015. In Tables 4.2 and 4.3 we present summaries of posterior means, lower and upper credible limits for parameters of model A1 and A2 respectively. The posterior densities of parameters in model A1 are shown in Figure 4.1. The estimates of μ for ASH and IBM in Table 4.2 are respectively 11.319 and 13.071. These estimates are very close to the respective

	y_t		x_t	
	ASH	IBM	ASH	IBM
Minimum	9.809	11.753	-8.103	-8.669
Maximum	12.907	14.765	-1.954	-2.118
Median	11.331	13.071	-2.757	-2.836
Mean	11.357	13.077	-2.755	-2.835
Skewness	0.097	0.128	0.009	-0.005
Kurtosis	3.334	3.237	3.086	2.850
Variance	0.248	0.208	0.067	0.054

Table 4.1 Descriptive statistics of data

empirical mean shown in Table 4.1. Similarly, estimates −7.385 and −8.032 of η for ASH and IBM respectively are of comparable magnitude to the empirical minimums shown in Table 4.1.

	ASH				**IBM**			
	Est	Stdev	Lci	Uci	Est	Stdev	Lci	Uci
μ	11.319	0.054	11.21	11.425	13.071	0.041	12.988	13.152
η	-7.385	0.022	-7.429	-7.341	-8.032	0.017	-8.066	-7.996
α_τ	-0.253	0.060	-0.379	-0.140	-0.258	0.054	-0.369	-0.155
α_h	0.212	0.062	0.098	0.343	0.216	0.049	0.123	0.318
ϕ_τ	0.861	0.049	0.750	0.941	0.849	0.038	0.768	0.918
ϕ_h	0.904	0.029	0.842	0.955	0.881	0.029	0.819	0.936
$\sigma^2_{\epsilon_y}$	0.069	0.010	0.051	0.090	0.046	0.005	0.035	0.058
$\sigma^2_{\epsilon_x}$	0.031	0.003	0.025	0.037	0.025	0.002	0.021	0.029
$\sigma^2_{\epsilon_\tau}$	0.058	0.010	0.040	0.082	0.044	0.006	0.032	0.058
$\sigma^2_{\epsilon_h}$	0.051	0.009	0.035	0.071	0.042	0.006	0.031	0.055
ν_y	22.520	17.910	5.778	75.044	39.529	24.922	11.379	106.764
ν_x	21.318	12.355	7.891	56.693	48.540	23.367	18.021	105.015
ν_τ	38.524	22.881	11.003	100.045	41.142	21.940	13.690	95.730
ν_h	40.877	20.663	13.685	94.900	40.135	22.402	13.157	95.676

Table 4.2 Model A1 - Posterior estimates

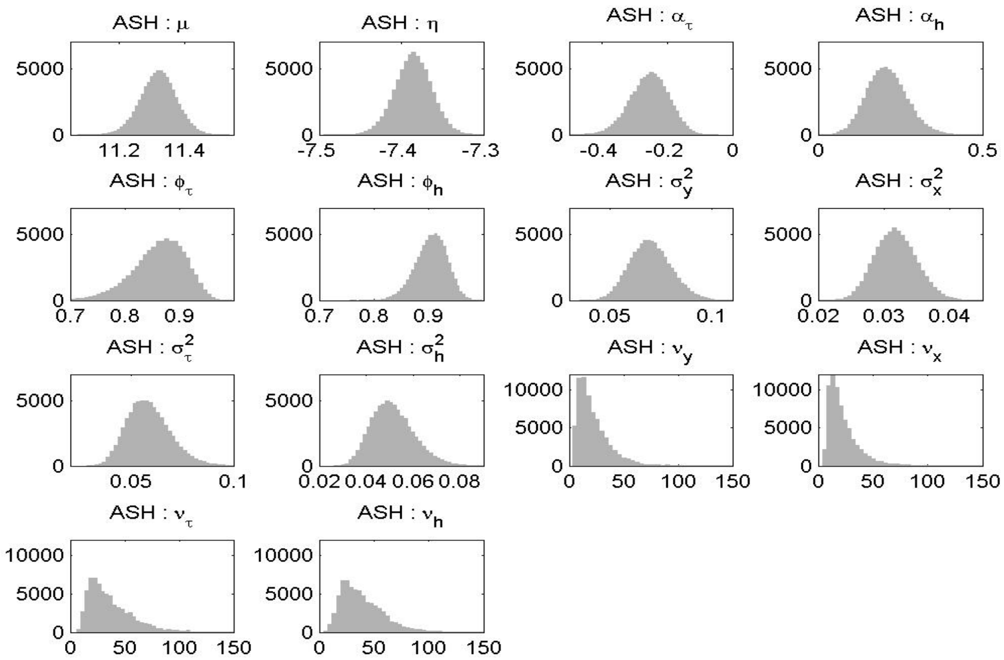
	ASH				**IBM**			
	Est	Stdev	Lci	Uci	Est	Stdev	Lci	Uci
μ^*	4.412	0.604	3.498	6.230	3.435	0.497	2.420	4.417
η^*	-2.970	0.350	-3.665	-2.303	-3.493	0.324	-3.665	-2.912
α_τ	-0.290	0.088	-0.469	-0.124	-0.176	0.074	-0.328	-0.038
α_h	0.099	0.071	0.004	0.263	0.064	0.051	0.002	0.192
ρ_y	0.611	0.053	0.451	0.691	0.737	0.038	0.662	0.814
ρ_x	0.597	0.047	0.503	0.687	0.565	0.040	0.487	0.637
ϕ_τ	-0.337	0.108	-0.528	-0.103	-0.311	0.107	-0.504	-0.081
ϕ_h	-0.152	0.162	-0.413	0.255	-0.232	0.120	-0.452	0.018
$\sigma^2_{\epsilon_y}$	0.067	0.011	0.047	0.091	0.048	0.007	0.035	0.063
$\sigma^2_{\epsilon_x}$	0.042	0.005	0.032	0.053	0.034	0.002	0.029	0.040
$\sigma^2_{\epsilon_\tau}$	0.065	0.011	0.045	0.089	0.048	0.007	0.036	0.064
$\sigma^2_{\epsilon_h}$	0.065	0.011	0.046	0.089	0.048	0.007	0.036	0.063
ν_y	33.031	20.487	8.460	86.799	36.143	20.819	10.291	89.289
ν_x	27.303	20.114	8.108	85.709	46.852	26.235	15.134	113.669
ν_τ	34.018	22.602	9.787	95.932	39.233	26.599	11.043	112.453
ν_h	31.913	22.026	8.085	93.430	37.057	23.518	10.097	101.201

Table 4.3 Model A2 - Posterior estimates

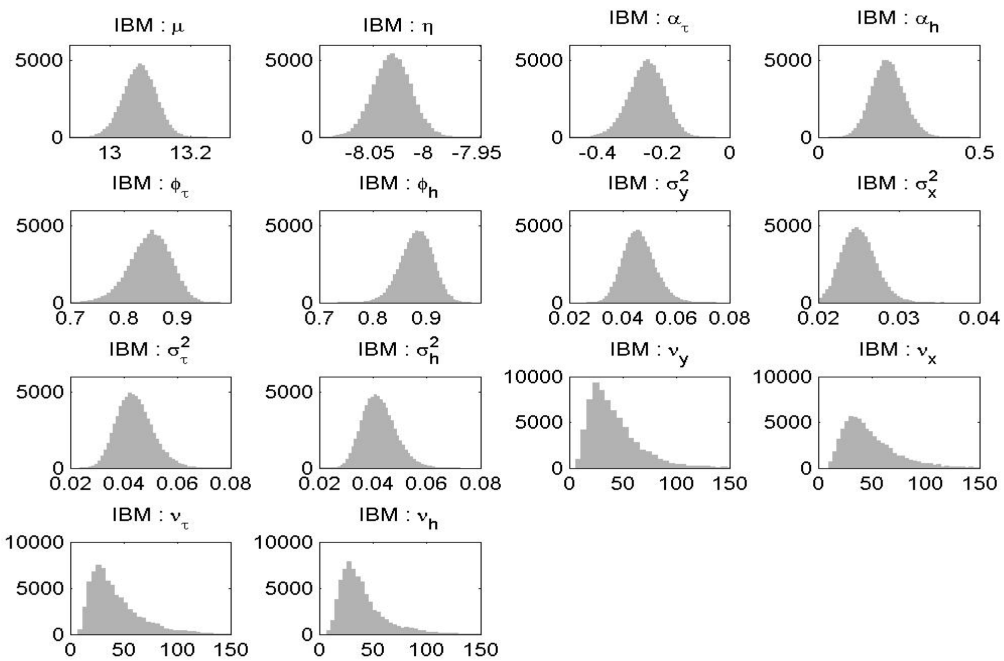
The effect of uninformed trading on bid-ask spread which is captured by the parameter α_τ for ASH and IBM are -0.253 and -0.258 respectively. The proxies of informed and uninformed trading are significantly persistent for both assets as shown by their respective estimates $\phi_\tau^{ASH} = 0.861$, $\phi_\tau^{IBM} = 0.849$, $\phi_h^{ASH} = 0.904$ and $\phi_h^{IBM} = 0.881$. The estimates of α_h, α_τ, ϕ_h and ϕ_τ are all significant as their respective credible intervals do not include the value zero.

In Figures 4.2 and 4.3 we present the relative bid-ask spread, volume and the latent components of volume for ASH and IBM respectively in model A1. It can be observed that both latent components of informed and uninformed trading in both assets change over time. A closer look at the uninformed (τ_t) and informed (h_t) trading processes show that there are certain trading days when both are high, both are low or one of them is high and the other is low. The persistence and dynamic behaviour of the informed and uninformed trading processes provide support for the decomposition of volume and the modelling approach we employ. The latent components exhibit a cyclical pattern which coincides with quarterly earnings announcement days. Our model structure is thus flexible and is able to accommodate all possible behaviours of informed and uninformed trading.

The empirical results of model A2 indicate that the effect of informed trading (α_h) on spread is significant but weak. The informed trading is noise. Even though the effect of the uninformed (α_τ) is significantly negative, the process is noisy. The dynamics of the informed and uninformed in model A2 imply that all relevant information about informed and uninformed trading are already captured in the lag volume. In contrast in model A1, τ_t and h_t have accumulated information about previous informed and uninformed trading.



(a) ASH



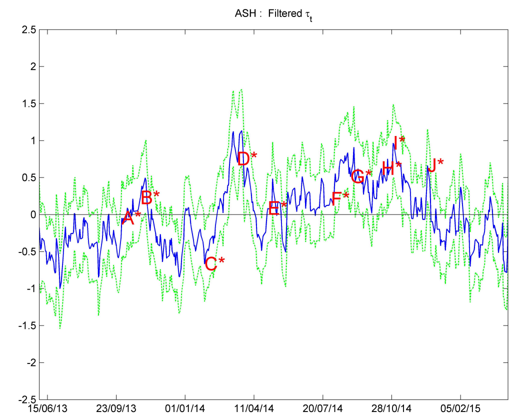
(b) IBM

Fig. 4.1 Model A1 - Posterior distributions of parameters

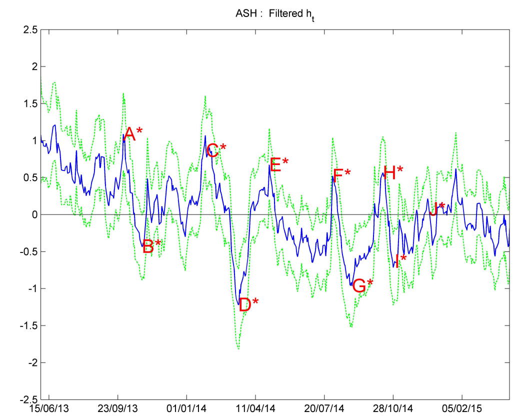
In a situation where τ_t and h_t are both high, bid-ask spread is expected to be wide. Volume on the other hand may also be high because a sudden increase in informed trading relative to uninformed trading may lead to increased volume. Notice that trading days labelled **E***, **F*** and **H*** in Figure 4.2 as well as **A** and **E** in Figure 4.3 both the informed and uninformed trading components were high. Figures 4.2e, 4.2f, 4.3e and 4.3f show that as expected, the bid-ask spread increased on these days when both components of volume were high. It can also be observed that the volume was high on these days.

According to theory, it is expected that bid-ask spread will be narrow when uniformed trading is high relative to informed trading. The corresponding volume will also see an increase due to the increased trading from uninformed traders. Trading days marked **B***, **D*** and **I*** for ASH show that τ_t is high and h_t is low. The bid-ask spread on these days are narrow as expected. The respective volumes on those days are also high. We observe similar patterns in IBM one 18^{th} March 2014 and 16^{th} March 2015 which are marked **C** and **F** respectively on Figure 4.3.

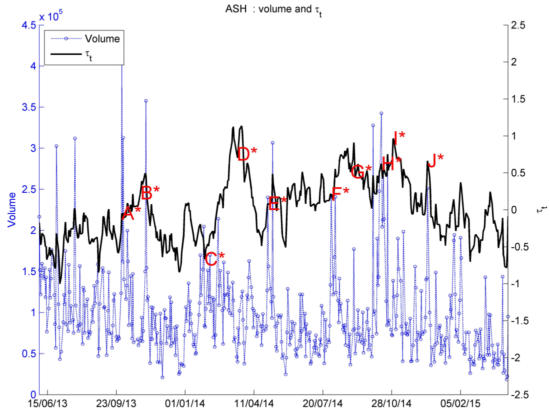
It is expected that the bid-ask spread will be wide when informed trading is high and uninformed trading low. The traded volume in such a scenario is also expected to be high. Points **B** (22^{nd} January 2014) and **D** are two trading days where we observe such dynamics of the information proxies for IBM. The 22^{nd} January 2014 is a day after IBM had reported weak performance claimed to be driven by fallen revenues in the so-called BRIC countries (Brazil, Russia, India and China). The dynamics of τ_t and h_t on this day provides evidence in support of the argument that investors interpret public information differently. For ASH we observe the same dynamics of the informed and uniformed proxies on the days labelled **C*** and **G*** in Figure 4.2.



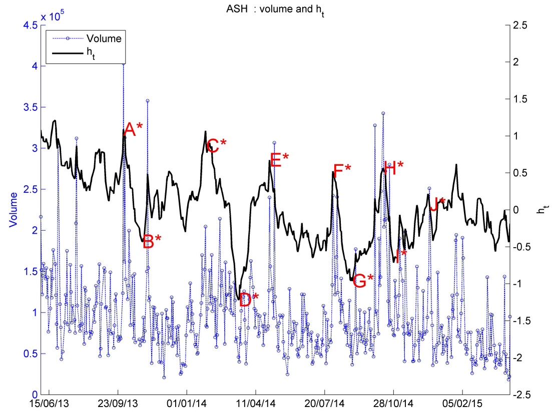
(a) ASH τ_t



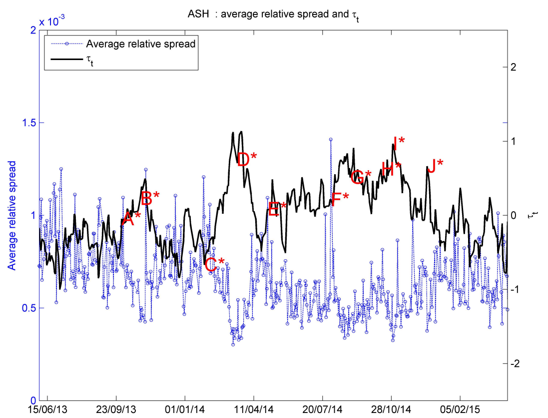
(b) ASH h_t



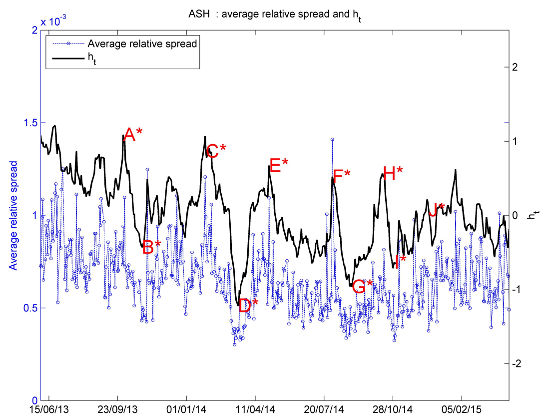
(c) ASH - Volume and τ_t



(d) ASH - Volume and h_t

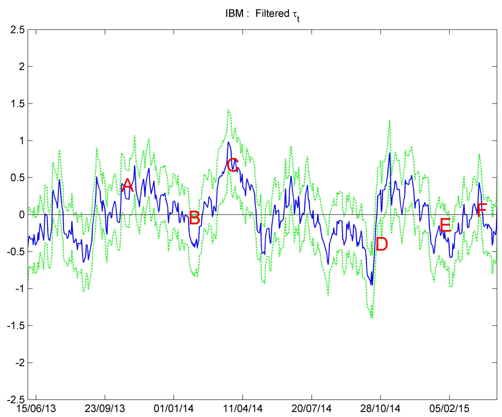


(e) ASH - Spread and τ_t

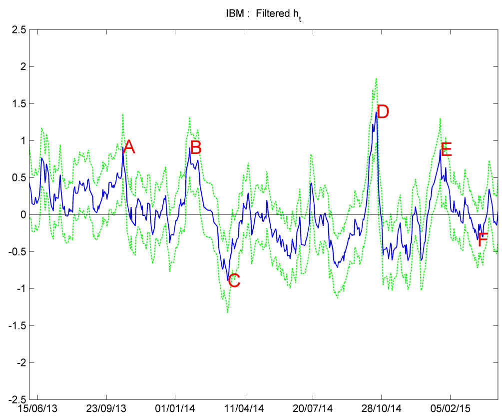


(f) ASH - Spread and h_t

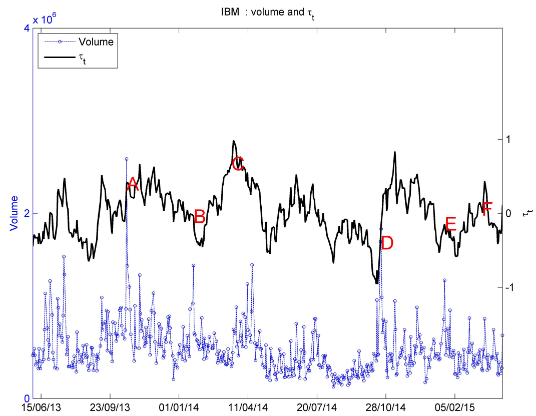
Fig. 4.2 ASH model A1 - latent processes, volume and relative spread



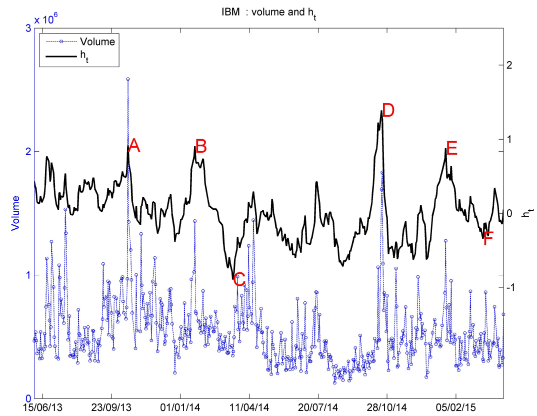
(a) IBM τ_t



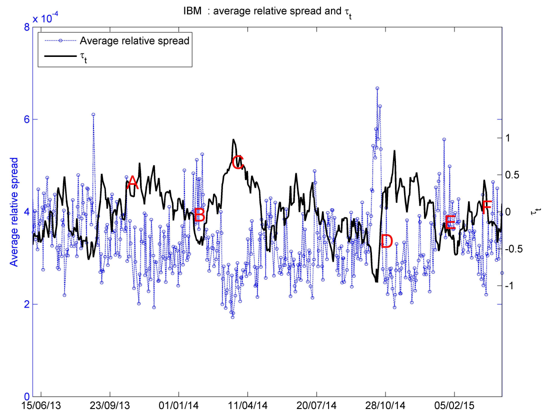
(b) IBM h_t



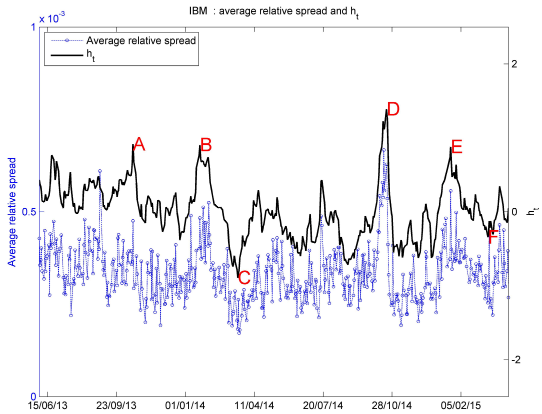
(c) IBM - Volume and τ_t



(d) IBM - Volume and h_t

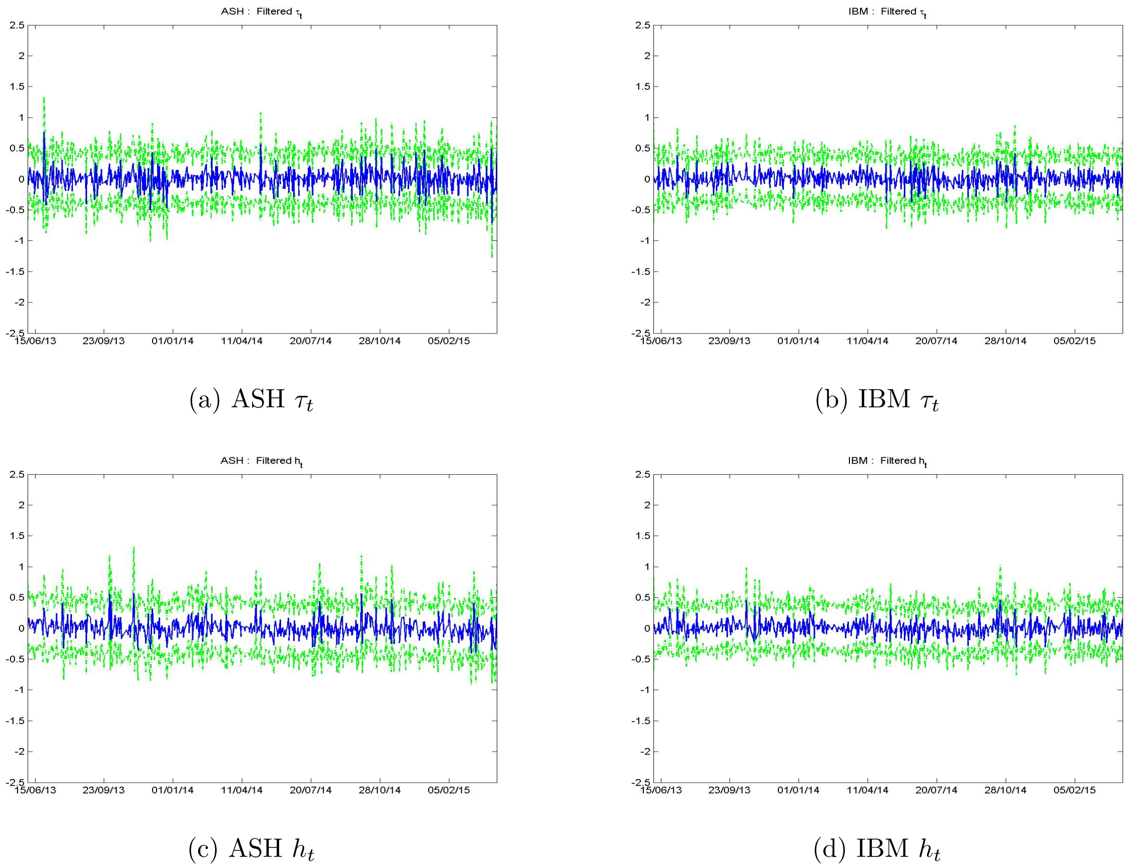


(e) IBM - Spread and τ_t



(f) IBM - Spread and h_t

Fig. 4.3 IBM model A1 - latent processes, volume and relative spread



(a) ASH τ_t

(b) IBM τ_t

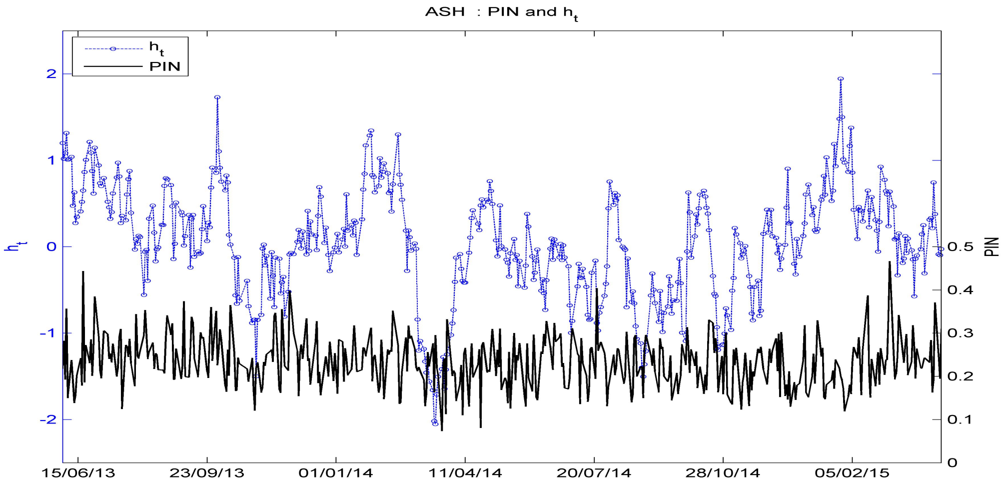
(c) ASH h_t

(d) IBM h_t

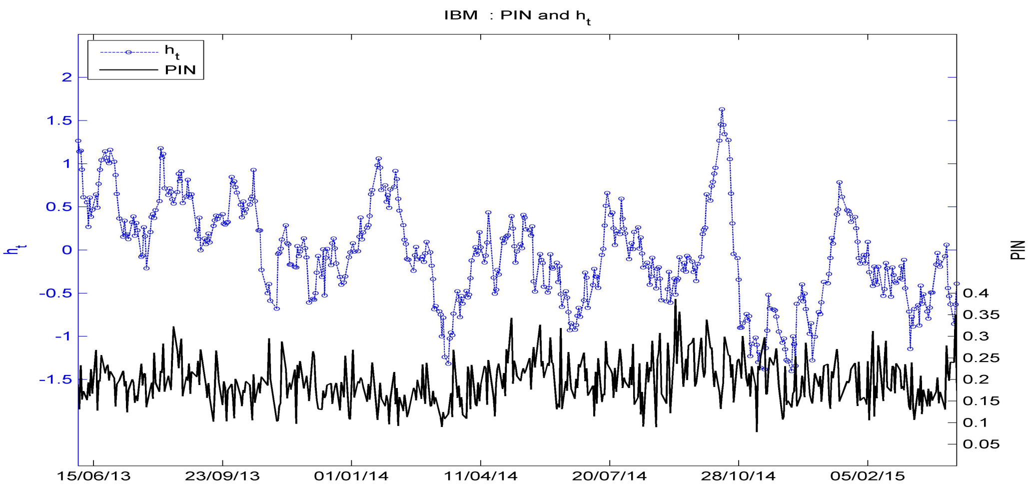
Fig. 4.4 Model A2 - latent processes

On the 20^{th} October 2014 (point labelled **D**), the CEO of IBM confirmed that IBM is to pay GlobalFoundries \$1.5 billion to take on its ailing semiconductor technologies business. IBM also abandoned its promise of delivering a \$20 earnings per share by the year 2015 as the company also announced another round of layoffs. The news on the deal between IBM and GlobalFoundries was revealed on Sunday 19^{th} via Bloomberg. From Figures 4.3a and 4.3b, the informed component of volume started increasing about a week and a half before the earnings announcement while the uniformed component (τ_t) decreased. The uninformed component only started increasing once the news about the deal had been revealed via Bloomberg on the 19^{th}. It is reasonable to conjecture that before the 19^{th} October 2014 some investors traded on their superior information which drove uninformed trading down.

In Figure 4.5, we present a comparison of the informed trading process in our model with the daily probability of informed trading of chapter 2.



(a) ASH - 5min based PIN and h_t



(b) IBM -5min based daily PIN and h_t

Fig. 4.5 Comparison of daily PIN and h_t

It is clear from the graph that while both h_t and PIN are time varying, the PIN is more stable than the h_t. The daily PIN is calculated from 5 minute sampled buyer and seller initiated trades. Since the PIN model assume that over short intervals volume is constant, one would expect that PIN estimated from buy and sell trades sampled over such short intervals will not change very much. Intuitively one would expect a positive correlation between the 2 series and simultaneous jumps. However this is not clear from the plots possibly due to the approaches taken.

4.4 Concluding Remarks

In this chapter, we have exploited existing market microstructure theory about the relationship between volume and bid-ask spread to develop an alternative and complementary approach of inferring the components of trading volume. We depart from the use of derived order flow imbalance variables such as buyer or seller initiated trades but rather we use observed market data to infer information asymmetry. An additional advantage of our modelling approach is that we can account for the uncertainty about model parameters and the unobserved processes. Prior beliefs about effects of various components of volume on bid-ask spread are also easily incorporated in to the estimation.

The empirical analysis carried out shows that uninformed and informed trading components of traded volume are both persistent. The effect of the uninformed component of volume on bid-ask spread is relatively higher in ASH compared with IBM. The components which have been used as proxies for informed and liquidity trading seem to capture the effects of news events of the assets considered in this thesis. Our model also shows that the informed trading component of volume increases in anticipation either to favourable or unfavourable news about the asset.

A limitation of our model is that the effect of informed trading on bid-ask spread is unidirectional. It can be argued that there is a feedback relationship between informed trading and bid-ask spread. This feature has not been addressed in our model. In future work, it is our intention to incorporate this feedback effect into our model. We believe this will enhance the performance of the model in capturing the latent process as well as to infer the direction of the next periods' liquidity. Model parameters in our model are invariant to changes in the market and general economic conditions. This is left for future work where we consider a time-varying parameter state space model structure.

Chapter 5

Investigating The Link Between Volatility, Informed And Uninformed Trading

5.1 Introduction

Volatility is an important unobservable characteristic of assets associated with information flow in financial markets. It is vital in financial markets because it is used as an input for risk management, asset allocation and derivative pricing. Due to its importance, research on volatility has seen the development of numerous models of varying complexity.

The information held by investors is revealed to the market through the volume of the orders they submit. Traded volume has therefore been used in the literature as a proxy for volatility. Clark (1973) used the mixture of distributions hypothesis (MDH) to initiate the idea and discussion on the link between volume and volatility. The MDH postulates that changes in the price of an asset and traded volume are jointly dependent on a common unobserved information process. This unobserved information process is normally interpreted as volatility. The MDH has spurred on a vast literature of volatility models including Andersen (1996) who decompose traded volume into two components. The decomposition of volume into two components in Andersen (1996) is based on market microstructure theory of informational heterogeneity among investors.

The effect of informed trading on volatility has been studied extensively in the literature (see for example Foster and Viswanathan (1994), Admati and Pfleiderer (1988), Madhavan et al. (1997), Amihud and Mendelson (1980), Harris (1987) and Kyle (1984)). These studies and others have reported mixed findings on the relationship between informed trading and volatility. For example Foster and Viswanathan (1994) find that informed trading and volatility are negatively related. They argue that informed traders trade competitively among themselves. The competitive trading between informed traders leads to the revelation of more private information and hence a reduction in the uncertainty about the value of the asset. The uncertainty about the value of the asset, therefore, leads to a reduced return volatility. Admati and Pfleiderer (1988) looks at a market in which some uninformed traders are strategic and can choose the timing of their trades. Since the uninformed traders do not want their trades to impact the price of the asset, they tend to trade at the same time. Informed traders also have the incentive to trade when uninformed traders cluster in the market so that they can also conceal their information. Admati and Pfleiderer (1988) also argue that if informed traders have the same piece of information, then an increase in informed trading will have no effect on volatility. However, if the informed traders have diverse pieces of information about the asset trading, then more private information will be generated. The excess diverse private information generated will increase the uncertainty about the fundamental value of the asset hence volatility.

In chapter 3, we employed the idea of a market comprising of informed and uniformed investors to separate volume into two components. The decomposition of volume in chapter 3 is in the spirit of Andersen (1996) which is based on the joint dependence of volume and returns on volatility as postulated in the MDH. There is also evidence of a relationship between volatility and bid-ask spread. Bollerslev and Melvin (1994) found a positive relationship between bid-ask spread and volatility in the foreign exchange market. In their paper, the author's treated volatility and bid-ask spread exogenous to each other. They used a GARCH model to estimate volatility from daily returns and then used the estimated volatility as an explanatory variable in a probit regression model in which the bid-ask spread is the response variable. Wang and Yau (2000) also find evidence of a joint dependence of bid-ask spread and volatility in the futures market. In their work, they found that there is a negative relationship between volume and bid-ask spread while volume and price volatility exhibited a positive relationship. Also, they found that the bid-ask spread and volatility have a positive relationship. The previous days' trading volume also had a negative relationship with volatility. Harris (2002) suggested that volatility, information asymmetry are the most important determinants of the bid-ask spread and that volatility has a strong indirect effect on bid-ask spread. Therefore a joint model of volume, volatility and bid-ask spread can offer an avenue to learn more about the relationship between volatility, informed and uniformed trading.

Our objective in this chapter is to examine the relationship between informed trading, uninformed trading and volatility. We exploit the volume decomposition idea and model developed in chapter 3 for this investigation. The resulting model has the potential to be used to forecast volatility. Four volatility models are considered and compared with the Heterogeneous Autoregressive (HAR) and Autoregressive (AR) models of realized variance. In the empirical implementation of our models, we divide our data into in-sample and out-of-sample periods. We employ the Bayesian methodology in discussed in chapter 3 on the in-sample data to estimate the model parameters and also to filter the unobserved informed and uninformed components of volume. We use the samples drawn from the marginal posterior distributions of model parameters in the in-sample-period as discrete approximations for the model parameters in the out-of-sample period. Similarly, we use the unobserved informed and uninformed trading processes obtained from the Kalman filter as prior distributions for the unobserved informed and uninformed trading processes out-of-sample period. Using a Sequential Monte Carlo (SMC) method we update our knowledge of the unobserved informed and uninformed trading as new observations of volume, bid-ask spread and realized variance become available. Simultaneously, we generate one day ahead volatility forecasts. Another advantage of using the SMC approach is that the entire history of the observed data will not be needed for making forecasts. This is because all relevant information needed for forecasting are accumulated in the current state of unobserved informed trading components.

5.2 The Models

Let $y_t = \ln V_t$, $x_t = \ln S_t$ and $z_t = \ln RV_t$ be the logarithm of daily volume (V_t), average relative bid-ask spread (S_t) and realized variance (RV_t) of an asset respectively. Since proxies for volatility are subject to jumps and other microstructure noise, we used the *realized kernel* matlab function of the *Oxford MFE Toolbox* that is robust to microstructure noise to calculate the realized variance of our data set.

In what follows we describe the models considered in this chapter. The rationale for the first model we consider is to revisit the volume-volatility relationship. In this model $\tilde{\tau}_t$ is the underlying volume which is assumed to evolve smoothly over time and changes with the arrival of information about the asset. The model is as follows:

$$\mathbf{C1}: \quad y_t = \tilde{\tau}_t + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_y}}{\kappa_t}\Big) \tag{5.1a}$$

$$z_t = \beta_0 + \beta_{\tilde{\tau}}\tilde{\tau}_t + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_z}}{\xi_t}\Big) \tag{5.1b}$$

$$\tilde{\tau}_t = \phi_{\tilde{\tau}}\tilde{\tau}_{t-1} + \varepsilon_{\tilde{\tau},t}, \qquad \varepsilon_{\tilde{\tau},t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_{\tilde{\tau}}}}{\omega_t}\Big). \tag{5.1c}$$

Fluctuations in volume due to transitory information arrival are modelled through $\varepsilon_{y,t}$. We expect the parameter $\beta_{\tilde{\tau}}$ to be positive in line with the mixture of distributions hypothesis. Model C1 does not provide us with a way to discern the behaviour of the informed and uninformed components of volume and their respective effects on volatility. We therefore consider an alternative model which allows for the learning of the effect of informed and uninformed trading on volatility and bid-ask spread. The model is an extension of model A1 which we considered in chapter 3. Our interest is to know whether bid-ask spread provides additional information beyond what is already contained in the volume-volatility relationship in model C1. The alternative model is given as

$$\mathbf{C2}: \quad y_t = \mu + \tau_t + h_t + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_y}}{\kappa_t}\Big) \tag{5.2a}$$

$$z_t = \beta_0 + \beta_\tau\tau_t + \beta_h h_t + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_z}}{\xi_t}\Big) \tag{5.2b}$$

$$x_t = \eta + \alpha_\tau\tau_t + \alpha_h h_t + \varepsilon_{x,t}, \qquad \varepsilon_{x,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_x}}{\varphi_t}\Big) \tag{5.2c}$$

$$\tau_t = \phi_\tau\tau_{t-1} + \varepsilon_{\tau,t}, \qquad \varepsilon_{\tau,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_\tau}}{\omega_t}\Big) \tag{5.2d}$$

$$h_t = \phi_h h_{t-1} + \varepsilon_{h,t}, \qquad \varepsilon_{h,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_h}}{\delta_t}\Big). \tag{5.2e}$$

The previous days' volatility (z_{t-1}) has been commonly used in AR (1) type models as a predictor of present day volatility (z_t). Model C2 makes no provision for lag volatility as a predictor of present day volatility. In the next model we incorporate lag volatility (z_{t-1}) and exploit only the effect of h_t on volatility. To investigate the extent to which h_t reduce or increase the impact of lag volatility on present day volatility, we consider the following model:

$$\textbf{C3}: \quad y_t = \mu + \tau_t + h_t + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_y}}{\kappa_t}\Big) \tag{5.3a}$$

$$z_t = \beta_0 + \beta_1(z_{t-1} - \beta_0) + \beta_h h_t + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_z}}{\xi_t}\Big) \tag{5.3b}$$

$$x_t = \eta + \alpha_\tau \tau_t + \alpha_h h_t + \varepsilon_{x,t}, \qquad \varepsilon_{x,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_x}}{\varphi_t}\Big) \tag{5.3c}$$

$$\tau_t = \phi_\tau \tau_{t-1} + \varepsilon_{\tau,t}, \qquad \varepsilon_{\tau,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_\tau}}{\omega_t}\Big) \tag{5.3d}$$

$$h_t = \phi_h h_{t-1} + \varepsilon_{h,t}, \qquad \varepsilon_{h,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_h}}{\delta_t}\Big). \tag{5.3e}$$

Model C3 assumes that the proxy for uninformed trading effect the bid-ask spread only but not volatility. It is however likely that both informed and uninformed trading effect volatility in different ways. For example Li and Wu (2006) found that uninformed trading reduces volatility. In the next model, both τ_t and h_t are included in the volatility model to confirm or otherwise the effect of τ_t on volatility. If all information is reflected in z_{t-1} then we expect the estimates of β_h and β_τ to be zero. The proxy for uninformed and informed trading in this model affects both volatility and bid-ask spread. The extended model we investigate is

$$\textbf{C4}: \quad y_t = \mu + \tau_t + h_t + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_y}}{\kappa_t}\Big) \tag{5.4a}$$

$$z_t = \beta_0 + \beta_1(z_{t-1} - \beta_0) + \beta_\tau \tau_t + \beta_h h_t + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_z}}{\xi_t}\Big) \tag{5.4b}$$

$$x_t = \eta + \alpha_\tau \tau_t + \alpha_h h_t + \varepsilon_{x,t}, \qquad \varepsilon_{x,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_x}}{\varphi_t}\Big) \tag{5.4c}$$

$$\tau_t = \phi_\tau \tau_{t-1} + \varepsilon_{\tau,t}, \qquad \varepsilon_{\tau,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_\tau}}{\omega_t}\Big) \tag{5.4d}$$

$$h_t = \phi_h h_{t-1} + \varepsilon_{h,t}, \qquad \varepsilon_{h,t} \sim \mathcal{N}\Big(0, \frac{\sigma^2_{\varepsilon_h}}{\delta_t}\Big). \tag{5.4e}$$

In models C2, C3 and C4, we expect the estimate of β_h to be positive as the literature posit simultaneous increases in volatility and bid-ask spread on the arrival of private information. Persistence in uninformed and informed trading are given by the parameters ϕ_τ and ϕ_h respectively. The term $\exp(\eta)$ is the minimum average relative bid-ask spread. The effect of informed and uninformed components of volume on bid-ask spread are modelled through α_h and α_τ respectively. The other assumption we make in all the models is that conditional on τ_t and h_t, bid-ask spread, volume and volatility are independent. To account for fat-tails in our models, we use a scale mixture of normals for the distribution of the error terms $\varepsilon_{z,t}$, $\varepsilon_{y,t}$, $\varepsilon_{x,t}$, $\varepsilon_{\tau,t}$ and $\varepsilon_{h,t}$. Geweke (1993) proved that this formulation is equivalent to a specification that assumes a Student-t distribution for the error terms. By introducing a t-distribution in the shock structure, we can capture extreme changes in the observed and unobserved processes. Thus we have an effective way of dealing with outliers and extreme events. Since we estimate the parameters of the models above in a Bayesian setting, we choose $\eta \sim \mathcal{N}(\mu_\eta, \sigma^2_\eta)$,

$$\begin{array}{llll}
\beta_0 \sim \mathcal{N}(\mu_{\beta_0}, \sigma^2_{\beta_0}), & \mu \sim \mathcal{N}(\mu_\mu, \sigma^2_\mu), & \xi_t \sim \mathcal{G}a\left(\frac{\nu_z}{2}, \frac{\nu_z}{2}\right), & \nu_\tau \sim \mathcal{G}a(a,b), \\
\beta_1 \sim \mathcal{N}(\mu_{\beta_1}, \sigma^2_{\beta_1}), & \alpha_\tau \sim \mathcal{N}\left(\mu_{\alpha_\tau}, \sigma^2_{\alpha_\tau}\right), & \delta_t \sim \mathcal{G}a\left(\frac{\nu_h}{2}, \frac{\nu_h}{2}\right), & \sigma^2_{\varepsilon_\tau} \sim \mathcal{IG}(r,s), \\
\beta_h \sim \mathcal{N}(\mu_{\beta_h}, \sigma^2_{\beta_h}), & \alpha_h \sim \mathcal{N}\left(\mu_{\alpha_h}, \sigma^2_{\alpha_h}\right), & \nu_y \sim \mathcal{G}a(a,b), & \sigma^2_{\varepsilon h} \sim \mathcal{IG}(r,s), \\
\beta_\tau \sim \mathcal{N}(\mu_{\beta_\tau}, \sigma^2_{\beta_\tau}), & \kappa_t \sim \mathcal{G}a\left(\frac{\nu_y}{2}, \frac{\nu_y}{2}\right), & \nu_x \sim \mathcal{G}a(a,b), & \sigma^2_{\varepsilon_y} \sim \mathcal{IG}(r,s), \\
\phi_\tau \sim \mathcal{N}(\mu_{\phi_\tau}, \sigma^2_{\phi_\tau}), & \varphi_t \sim \mathcal{G}a\left(\frac{\nu_x}{2}, \frac{\nu_x}{2}\right), & \nu_z \sim \mathcal{G}a(a,b), & \sigma^2_{\varepsilon_z} \sim \mathcal{IG}(r,s), \\
\phi_h \sim \mathcal{N}(\mu_{\phi_h}, \sigma^2_{\phi_h}) & \omega_t \sim \mathcal{G}a\left(\frac{\nu_\tau}{2}, \frac{\nu_\tau}{2}\right), & \nu_h \sim \mathcal{G}a(a,b), \text{ and} & \sigma^2_{\varepsilon_x} \sim \mathcal{IG}(r,s)
\end{array}$$

as the prior distributions for the model parameters; where $\mathcal{G}a(.)$, $\mathcal{N}(.)$ and $IG(.)$ are the distributions of the gamma, normal and inverse gamma random variables. We choose conjugate priors so that the posterior distributions will have kernels that can easily be sampled using the Gibbs sampler. We set the hyper-parameters of the priors for ν_z, ν_y, ν_x, ν_τ, and ν_h as follows $a = 0.16$ and $b = 0.04$. This sets the prior degrees of freedom of the distribution of the error terms at 4. The variance of the prior degrees of freedom is 100, which reflects how uncertain we are about the true values of these parameters. Similarly we choose the following $\mu_{\beta_0} = \mu_{\beta_1} = \mu_{\beta_\tau} = \mu_{\beta_h} = \mu_{\alpha_h} = \mu_\eta = \mu_{\phi_\tau} = \mu_{\alpha_\tau} = \mu_{\phi_h} = \mu_\mu = 0$ and $\sigma^2_{\beta_0} = \sigma^2_{\beta_1} = \sigma^2_{\beta_\tau} = \sigma^2_{\beta_h} = \sigma^2_{\alpha_h} = \sigma^2_\eta = \sigma^2_{\phi_\tau} = \sigma^2_{\alpha_\tau} = \sigma^2_{\phi_h} = \sigma^2_\mu = \sigma^2_{\phi_h} = \sigma^2_\mu = 100$. These hyper-parameter choices mean that before observing the data, β_0, β_h, β_τ, α_h, α_τ, ϕ_h, and ϕ_τ have no effect and can take on any value. The uncertainty is expressed through the large prior variance 100. Finally, the hyper-parameters of $\sigma^2_{\varepsilon_\tau}$, $\sigma^2_{\varepsilon_h}$, $\sigma^2_{\varepsilon_y}$, $\sigma^2_{\varepsilon_z}$, and $\sigma^2_{\varepsilon_x}$ are set to be $r = s = 1$ resulting in relatively flat priors which place little weight on any part of the parameter space. Thus we assume relatively no knowledge about the level and uncertainty of the model parameters. These choices imply that in the estimation of the model parameters, more weight on the information held in the observed data. In addition, the estimation of the model parameters are not sensitive to the choice of the hyper-parameters since the MCMC algorithm converges to the most probable space of the parameter after the burnin period. The corresponding full conditional posterior distributions similar to the ones in chapter 3 were derived and used for the MCMC sampling.

As indicated earlier, the models specified above may be used to forecast volatility. We, therefore, compare their performance with the models listed below which we refer to as **benchmark models**. Defining $z_t^{(w)}$ and $z_t^{(m)}$ as the logarithm of weekly and monthly realized variance, the benchmark models are

$$\textbf{AR (1)}: \quad z_t = \beta_0 + \beta_1(z_{t-1} - \beta_0) + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \sigma^2_{\varepsilon_z}\Big) \tag{5.5a}$$

$$\textbf{HAR}: \quad z_t = \beta_0 + \beta_1 z_{t-1} + \beta_w z_t^{(w)} + \beta_m z_t^{(m)} + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \sigma^2_{\varepsilon_z}\Big) \tag{5.5b}$$

$$\textbf{C5}: \quad z_t = \beta_0 + \beta_1 z_{t-1} + \lambda y_{t-1} + \varepsilon_{z,t}, \qquad \varepsilon_{z,t} \sim \mathcal{N}\Big(0, \sigma^2_{\varepsilon_z}\Big). \tag{5.5c}$$

Corsi (2009) used the idea of heterogeneity of investors with different investment time horizons to develop the Heterogeneous Autoregressive (HAR) model of realized variance. The model has become a common benchmark model due to its simplicity and forecasting performance. Because of this, we compare our models against it. Similarly, AR (1) models are also popular in the volatility forecasting literature hence its inclusion in the benchmark models. Lag volume is reported to contain information that can be used to forecast volatility. For this reason model C5 is used as an additional comparator model. Model C5, HAR and AR (1) models are estimated in the ordinary least square regression model framework.

5.3 Model Estimation

Models C1, C2, C3, and C4, can be represented in a state space form. We combine the Gibbs Sampling algorithm in chapter 3 with a Sequential Monte Carlo (SMC) method to estimate our models. In what follows we give a brief description of the SMC method.

Sequential Monte Carlo Methods

Sequential Monte Carlo methods, often referred to as *particle filters*, are simulation-based algorithms developed to aid the approximation of intractable integrals. Due to their flexibility and efficiency in filtering complex models, SMC methods have been widely applied in finance and econometric research. Among the numerous research applying SMC are Chib et al. (2002) and Lopes and Tsay (2011). The SMC is flexible in the sense that there is no need to store the entire history of data but rather only the most recent observation. SMC methods use discrete probability distributions consisting of weighted draws from posterior distributions known as *particles* to approximate continuous probability distributions. Prado and West (2010) and Sekerke (2015) provide‘ a detailed exposition to the theory and application of SMC.

As a new observation of the variable(s) being modelled become available the particles are updated based on their weights. Let Θ be the collection of model parameters, a general representation of our state space model can be stated as

$$Y_t \sim p(Y_t|\beta_t, \Theta) \tag{5.6a}$$

$$\beta_t \sim p(\beta_t|\beta_{t-1}, \Theta) \tag{5.6b}$$

$$p_0 \sim p(\beta_0|\Theta), \quad t = 1, \ldots, N, \tag{5.6c}$$

where $p(\beta_0)$ is the prior distribution of the latent state β_t. Defining $\psi_{0:t} = \Big(\psi_0, \psi_1, \ldots, \psi_t\Big)$ and $z_{1:t} = \Big(z_1, z_2, \ldots, z_t\Big)$, suppose $\Big\{\psi_{0:t}^{(i)}, \omega_t^{(i)}\Big\}_{i=1}^{M}$ can be used to characterise a posterior probability distribution $p(\psi_{0:t}|z_{1:t})$ of a latent state vector at some point t conditional on some observed variable $z_{1:t}$. If $\delta(.)$ is the *dirac* delta function, then the posterior distribution at time t may be approximated as

$$p(\psi_{0:t}|z_{1:t}, \theta) \approx \sum_{i=1}^{M} \omega_t^i \delta(\psi_{0:t} - \psi_{0:t}^{(i)}), \tag{5.7}$$

which is a discrete weighted approximation to the true posterior. Using *importance sampling*, the weights can be calculated. The importance sampling procedure relies on the following insight. Suppose that it is difficult to draw samples from the probability density $p(\psi)$. However it is possible to draw samples from the probability density $\pi(\psi)$ and that $p(\psi) \propto \pi(\psi)$. Assuming that $\psi^i \sim q(\psi), i = 1, \ldots M$ are samples easily drawn from $q(.)$, the *importance density*, then a weighted approximation to the density $p(\psi)$ is given as follows

$$p(\psi) \approx \sum_{i=1}^{M} \omega_t^{(i)} \delta(\psi - \psi^{(i)}), \tag{5.8}$$

where the normalised weights are defined as $w_t^i \propto \frac{\pi(\psi^i)}{q(\psi^i)}$. This means that samples $\psi_{0:t}$

drawn from an importance density $q(\psi_{0:t}|z_{1:t})$ have weights in equation (5.7) given as

$$w_t^i \propto \frac{p(\psi_{0:t}^i|z_{1:t}, \Theta)}{q(\psi_{0:t}^i|z_{1:t}, \Theta)}. \tag{5.9}$$

Now returning to our case and defining $\beta_{0:t} = (\beta_0, \beta_1, \ldots, \beta_t)$ and $Y_{1:t} = (Y_1, Y_2, \ldots, Y_t)$, if at every point we have samples approximating the density $p(\beta_{0:t-1}|Y_{1:t-1}, \Theta)$, we are interested in updating this density to $p(\beta_{0:t}|Y_{1:t}, \Theta)$ as new observations become available at t. If the importance density can be factorised as follows

$$q(\beta_{0:t}|Y_{1:t}, \Theta) = q(\beta_t|\beta_{0:t-1}, Y_{1:t}, \Theta)q(\beta_{0:t-1}|Y_{1:t-1}, \Theta), \tag{5.10}$$

then it is possible to draw samples $\beta_{0:t}^i \sim q(\beta_{0:t-1}|Y_{1:t}, \Theta)$. In order to calculate the weights we first express the posterior density $p(\beta_{0:t}|Y_{1:t})$ in terms of the prior distribution $p(\beta_{0:t-1}|Y_{1:t-}, \Theta)$, likelihood $p(Y_t|\beta_t, \Theta)$ and transition $p(\beta_t|\beta_{t-1}, \Theta)$. The posterior density is given as

$$\begin{aligned} p(\beta_{0:t}|Y_{1:t}, \Theta) &= \frac{p(Y_t|\beta_{0:t}, Y_{1:t-1}, \Theta)p(\beta_t|\beta_{t-1}, \Theta)}{p(Y_t|Y_{1:t-1}, \Theta)} \\ &= \frac{p(Y_t|\beta_t, \Theta)p(\beta_t|\beta_{t-1}, \Theta)}{p(Y_t|Y_{t-1}, \Theta)} p(\beta_{0:t-1}|Y_{1:t-1}, \Theta) \\ &\propto p(Y_t|\beta_t, \Theta)p(\beta_t|\beta_{t-1}, \Theta)p(\beta_{0:t-1}|Y_{1:t-1}, \Theta). \end{aligned} \tag{5.11}$$

Substituting equation 5.10 and 5.11 into equation 5.9 we have the following updated weights

$$\begin{aligned} w_t^i &\propto \frac{p(Y_t|\beta_t^i, \Theta)p(\beta_t^i|\beta_{t-1}^i, \Theta)p(\beta_{0:t-1}^i|Y_{1:t-1}, \Theta)}{q(\beta_t^i|\beta_{0:t-1}^i, Y_{1:t}, \Theta)q(\beta_{0:t-1}^i|Y_{1:t-1}, \Theta)} \\ &= \omega_{t-1}^i \frac{p(Y_t|\beta_t^i, \Theta)p(\beta_t^i|\beta_{t-1}^i, \Theta)}{q(\beta_t^i|\beta_{0:t-1}^i, Y_{1:t}, \Theta)}. \end{aligned} \tag{5.12}$$

With these weights the posterior density of the state vector can be approximated as follows

$$\hat{p}(\beta_{0:t}|Y_{1:t},\Theta) \approx \sum_{i=1}^{M} \omega_t^{(i)} \delta(\beta_{0:t} - \beta_{0:t}^{(i)}). \tag{5.13}$$

It can be shown that as the number of particles M becomes large the approximation in 5.13 approaches the true posterior density $p(\beta_{0:t}|Y_{1:t})$. Various algorithms based on the importance sampling have been proposed in the literature to extract the latent states. In this thesis we employ the algorithm of Liu and West (2001) described below. The Liu and West (2001) allows the estimation of model parameters as well as the filtering of the unobserved state vector.

Liu and West (2001) Algorithm

1. Start with a set of initial particles $\left\{\beta_0^{(i)}, \Theta_0^{(i)}\right\}_{i=1}^{M}$ with weights $w_0^{(i)} = \frac{1}{M}$.
2. For each trading day $t = 1, \ldots, N$.

   (a) For each particle $i = 1, \ldots, M$, we calculate point estimates

   $$\boldsymbol{m}_t^{(i)} = \gamma\Theta^{(i)} + (1-\gamma)\bar{\Theta}$$
   $$\boldsymbol{\mu}_t^{(i)} = E(\beta_t|\beta_{t-1}^{(i)}, \Theta^{(i)})$$

   of the pair (β_t, Θ) where γ and $\bar{\Theta} = E[\Theta^{(i)}]$ are a shrinkage parameter and the mean model parameters.

   (b) For $j = 1, \ldots, M$:

      i. Draw an auxiliary integer variable k, with probability $p(k) \propto w_{t-1}^{(j)} p\Big(Y_t|\boldsymbol{\mu}_t^{(j)}, \boldsymbol{m}_t^{(j)}\Big)$

      ii. Sample a new parameter vector $\Theta^{(j)} \sim \mathcal{N}\Big(\boldsymbol{m}_t^{(k)}, (1-\gamma)^2\Sigma\Big)$, where $\Sigma = Var[\Theta]$ is the variance of model parameters.

      iii. Sample a value of current state vector $\beta_t^{(k)}$ from $p\Big(\beta_t|\beta_{t-1}^{(k)}, \Theta^{(k_j)}\Big)$

iv. Assign each particle $\beta_t^{(k)}$ with a corresponding importance weight

$$\tilde{w}_t^{(k)} \propto \frac{p\Big(Y_t|\beta_t^{(k)}, \Theta^{(k)}\Big)}{p\Big(Y_t|\boldsymbol{\mu}_t^{(k)}, \boldsymbol{m}_t^{(k)}\Big)}$$

(c) Normalize the weights

$$w_t^{(j)} = \frac{\tilde{w}_t^{(j)}}{\sum\limits_{s=1}^{M} \tilde{w}_t^{(s)}}$$

3. Repeating Step 2 a large number of times produces the triplet $(\Theta^{(k)}, \beta_t^{(k)}, \omega_t^{(k)})$ for $k = 1, \ldots, M$ as the posterior approximation

5.4 Empirical Analysis

We use 1 minute sampled returns to calculate daily realized variance. The Gibbs Sampling algorithm described in chapter 3 is implemented on observed data from 3^{rd} of June 2013 to 4^{th} November 2014 for each asset. The sampler is run for $100,000$ sweeps with a burn-in period $30,000$. We use the remaining $70,000$ draws from the posterior distributions of the model parameters and the Kalman filter as initial particles for the Sequential Monte Carlo algorithm. The SMC is applied on the observed data for the period 15^{th} November 2014 to 15^{th} April 2015. The out of sample period, therefore, has 103 and 102 trading days for ASH and IBM respectively. The Liu and West (2001) particle filtering and parameter smoothing algorithm is then employed to extract the latent information processes and also to compute one day ahead forecasts. Ordinary least square estimates of AR (1), HAR and model C5 are shown in Table 5.1.

		AR (1)		**HAR**		**Model C5**	
		ASH	IBM	ASH	IBM	ASH	IBM
β_0	Est	-14.651	-15.643	-5.275	-7.228	-10.251	-10.829
	Sdev	0.024	0.021	1.227	1.338	1.099	1.274
	LCI	-14.655	-15.646	-5.449	-7.418	-10.423	-11.029
	UCI	-14.648	-15.640	-5.100	-7.037	-10.078	-10.628
β_1	Est	0.450	0.505	0.200	0.275	0.408	0.433
	Sdev	0.046	0.045	0.065	0.064	0.049	0.051
	LCI	0.443	0.499	0.191	0.266	0.401	0.425
	UCI	0.456	0.511	0.209	0.284	0.416	0.441
λ	Est					0.138	0.149
	Sdev					0.054	0.051
	LCI					0.130	0.141
	UCI					0.147	0.158
β_w	Est			0.449	0.445		
	Sdev			0.099	0.092		
	LCI			0.435	0.432		
	UCI			0.463	0.458		
β_m	Est			-0.007	-0.182		
	Sdev			0.105	0.094		
	LCI			-0.022	-0.195		
	UCI			0.007	-0.168		
$\sigma^2_{\epsilon_z}$	Est	0.215	0.174	0.201	0.161	0.212	0.171

Table 5.1 Estimates from benchmark models

The mean, standard error and credible intervals of the posterior distributions of the model parameters are summarised in Tables 5.2 and 5.3 for models C2, C3 and C4.

		C1	**C2**	**C3**	**C4**			**C1**	**C2**	**C3**	**C4**
μ	Est		11.360	11.366	11.362	$\sigma^2_{\epsilon_y}$	Est	0.105	0.065	0.068	0.064
	Sdev		0.054	0.055	0.052		Sdev	0.016	0.010	0.010	0.009
	LCI		11.243	11.248	11.250		LCI	0.075	0.047	0.049	0.046
	UCI		11.458	11.468	11.456		UCI	0.140	0.086	0.091	0.085
β_0	Est	-24.239	-14.725	-14.731	-14.726	$\sigma^2_{\epsilon_z}$	Est	0.154	0.092	0.097	0.089
	Sdev	1.321	0.046	0.050	0.045		Sdev	0.022	0.013	0.014	0.013
	LCI	-26.934	-14.825	-14.836	-14.823		LCI	0.111	0.068	0.071	0.064
	UCI	-21.755	-14.639	-14.638	-14.642		UCI	0.199	0.119	0.126	0.117
β_1	Est			-0.029	-0.052	$\sigma^2_{\epsilon_x}$	Est		0.026	0.028	0.027
	Sdev			0.063	0.062		Sdev		0.003	0.003	0.002
	LCI			-0.153	-0.173		LCI		0.021	0.022	0.021
	UCI			0.097	0.072		UCI		0.033	0.034	0.033
$\beta_{\tilde{\tau}}/\beta_\tau$	Est	0.837	-0.340		-0.368	$\sigma^2_{\epsilon_{\tilde{\tau}}}/\sigma^2_{\epsilon_\tau}$	Est	0.045	0.059	0.067	0.059
	Sdev	0.115	0.160		0.169		Sdev	0.007	0.010	0.012	0.010
	LCI	0.620	-0.664		-0.709		LCI	0.032	0.041	0.046	0.041
	UCI	1.073	-0.037		-0.042		UCI	0.062	0.082	0.093	0.082
β_h	Est		0.763	0.936	0.821	$\sigma^2_{\epsilon_h}$	Est		0.057	0.051	0.057
	Sdev		0.137	0.185	0.163		Sdev		0.009	0.008	0.009
	LCI		0.503	0.603	0.517		LCI		0.040	0.036	0.041
	UCI		1.042	1.323	1.158		UCI		0.079	0.069	0.078
η	Est		-7.404	-7.408	-7.405	ν_y	Est	14.59	32.345	25.533	25.522
	Sdev		0.023	0.024	0.023		Sdev	9.470	23.327	19.458	16.637
	LCI		-7.453	-7.460	-7.453		LCI	4.803	7.555	6.465	6.729
	UCI		-7.361	-7.362	-7.363		UCI	41.713	100.697	80.998	69.395
α_τ	Est		-0.388	-0.224	-0.392	ν_z	Est	16.758	13.716	16.192	14.506
	Sdev		0.093	0.044	0.093		Sdev	13.156	9.901	12.970	10.739
	LCI		-0.576	-0.314	-0.578		LCI	5.572	4.886	5.013	5.038
	UCI		-0.209	-0.138	-0.213		UCI	53.434	42.597	54.020	45.789
α_h	Est		0.348	0.421	0.355	ν_x	Est		43.580	39.648	44.564
	Sdev		0.080	0.095	0.081		Sdev		26.353	23.702	24.423
	LCI		0.197	0.244	0.204		LCI		12.738	11.952	13.426
	UCI		0.512	0.615	0.523		UCI		113.230	102.242	106.647
$\phi_{\tilde{\tau}}/\phi_\tau$	Est	0.999	0.757	0.801	0.751	$\nu_{\tilde{\tau}}/\nu_\tau$	Est	39.857	34.518	36.319	36.363
	Sdev	0.001	0.077	0.066	0.076		Sdev	23.247	21.764	21.002	24.423
	LCI	0.997	0.600	0.660	0.599		LCI	11.306	10.340	10.885	10.753
	UCI	1.001	0.901	0.919	0.896		UCI	100.979	90.996	89.810	105.933
ϕ_h	Est		0.866	0.881	0.862	ν_h	Est		26.248	31.921	26.091
	Sdev		0.046	0.042	0.046		Sdev		17.730	20.751	16.915
	LCI		0.771	0.790	0.767		LCI		7.396	9.168	7.596
	UCI		0.950	0.957	0.949		UCI		71.256	88.117	72.427

Table 5.2 ASH - Posterior estimates

		C1	**C2**	**C3**	**C4**			**C1**	**C2**	**C3**	**C4**
μ	Est		13.066	13.069	13.062	$\sigma^2_{\epsilon_y}$	Est	0.061	0.043	0.044	0.044
	Sdev		0.040	0.039	0.041		Sdev	0.007	0.006	0.006	0.006
	LCI		12.984	12.990	12.978		LCI	0.046	0.032	0.034	0.033
	UCI		13.145	13.146	13.141		UCI	0.077	0.057	0.058	0.057
β_0	Est	-22.909	-15.684	-15.683	-15.687	$\sigma^2_{\epsilon_z}$	Est	0.160	0.068	0.065	0.063
	Sdev	0.903	0.042	0.041	0.042		Sdev	0.015	0.009	0.011	0.010
	LCI	-24.674	-15.772	-15.767	-15.774		LCI	0.131	0.050	0.046	0.046
	UCI	-21.132	-15.603	-15.607	-15.607		UCI	0.191	0.088	0.088	0.086
β_1	Est			-0.057	-0.065	$\sigma^2_{\epsilon_x}$	Est		0.022	0.023	0.022
	Sdev			0.056	0.057		Sdev		0.002	0.002	0.002
	LCI			-0.166	-0.177		LCI		0.018	0.019	0.018
	UCI			0.054	0.048		UCI		0.027	0.028	0.027
$\beta_{\tilde{\tau}}/\beta_\tau$	Est	0.554	-0.101		-0.145	$\sigma^2_{\epsilon_{\tilde{\tau}}}/\sigma^2_{\epsilon_\tau}$	Est	0.051	0.049	0.052	0.049
	Sdev	0.069	0.157		0.168		Sdev	0.007	0.007	0.007	0.007
	LCI	0.417	-0.423		-0.493		LCI	0.038	0.036	0.038	0.036
	UCI	0.689	0.190		0.160		UCI	0.068	0.066	0.069	0.066
β_h	Est		1.123	1.234	1.177	$\sigma^2_{\epsilon_h}$	Est		0.052	0.050	0.052
	Sdev		0.153	0.162	0.165		Sdev		0.008	0.007	0.008
	LCI		0.848	0.916	0.866		LCI		0.038	0.037	0.038
	UCI		1.445	1.554	1.516		UCI		0.070	0.066	0.070
η	Est		-8.042	-8.042	-8.042	ν_y	Est	34.001	39.927	41.272	42.413
	Sdev		0.018	0.018	0.018		Sdev	22.840	23.012	27.244	23.615
	LCI		-8.078	-8.079	-8.079		LCI	9.010	11.591	10.807	11.884
	UCI		-8.007	-8.007	-8.007		UCI	92.999	98.691	112.424	102.130
α_τ	Est		-0.324	-0.272	-0.337	ν_z	Est	31.129	39.384	41.730	35.196
	Sdev		0.083	0.033	0.083		Sdev	20.776	25.677	27.171	21.080
	LCI		-0.495	-0.339	-0.509		LCI	8.400	10.399	11.133	9.552
	UCI		-0.169	-0.208	-0.183		UCI	88.706	105.937	115.984	89.901
α_h	Est		0.344	0.369	0.336	ν_x	Est		51.648	50.326	50.875
	Sdev		0.070	0.069	0.076		Sdev		26.161	25.558	26.852
	LCI		0.205	0.235	0.195		LCI		16.663	17.651	16.678
	UCI		0.507	0.511	0.492		UCI		117.716	116.8702	118.608
$\phi_{\tilde{\tau}}/\phi_\tau$	Est	0.999	0.807	0.822	0.799	$\nu_{\tilde{\tau}}/\nu_\tau$	Est	38.908	39.741	37.135	41.981
	Sdev	0.000	0.052	0.044	0.052		Sdev	24.041	23.829	21.294	24.384
	LCI	0.997	0.697	0.732	0.690		LCI	10.991	11.562	11.518	13.358
	UCI	1.001	0.904	0.904	0.895		UCI	100.429	105.321	93.264	106.017
ϕ_h	Est		0.806	0.804	0.810	ν_h	Est		28.651	30.445	28.432
	Sdev		0.052	0.053	0.051		Sdev		19.618	20.029	18.598
	LCI		0.699	0.697	0.706		LCI		9.154	8.749	8.431
	UCI		0.905	0.905	0.907		UCI		77.799	83.472	76.932

Table 5.3 IBM - Posterior estimates

The posterior estimate $\beta_{\tilde{\tau}}^{ASH} = 0.837$ of model C1 shown in Table 5.2 indicates a positive relationship between volume and volatility. The underlying volume $\tilde{\tau}$ in model C1 for ASH is significantly persistent as given by the posterior estimate $\phi_{\tilde{\tau}}^{ASH} = 0.999$.

We obtain similar results for IBM which can be found in Table 5.3. Intuitively we expect the previous period's volatility (z_{t-1}) to have a positive and significant effect on the current period (z_{t-1}) volatility. It can be observed from Table 5.1 that for the benchmark models the effect of z_{t-1} modelled via β_1 is positive and significant. On the contrary, the effect of z_{t-1} in the comparable proposed model namely C3 and C4 is negative and insignificant as can be seen in Tables 5.2 and 5.3. This negative insignificant effect is likely to be as a result of the β_h capturing the expected positive effect. Hence a possible remnant of the negative effect of uniformed trading being captured by β_1

Impact Of Informed Trading On Volatility - β_h

From Table 5.2, the posterior estimate of informed trading on volatility for ASH in model C2 is $\beta_h^{ASH} = 0.763$. Its confidence interval $(0.503,\ 1.042)$ is an indication of a significant positive effect of informed trading on volatility. Similarly $\beta_h^{ASH} = 0.936$ for model C3 is also positive and significant. The effect of h_t on volatility in model C3 is larger than that in model C2. The value $\beta_1^{ASH} = -0.029$ shows the insignificant effect of z_{t-1} on z_t. The estimate in model C4, $\beta_h^{ASH} = 0.821$, also shows that the effect is significantly positive. In comparison this estimate in model C4 is larger than the estimate in model C2.

The results for models C2, C3 and C4 for IBM shown in Table 5.3 are very similar to that obtained for ASH. The posterior estimates of the effect of informed trading on volatility are $\beta_h^{IBM} = 1.123$, $\beta_h^{IBM} = 1.234$ and $\beta_h^{IBM} = 1.177$ for models C2, C3 and C4 respectively. Their corresponding credible intervals also tells us that the effect is positive and significant. From these results, there is evidence that z_{t-1} provides no additional information in the models.

Impact Of Uninformed Trading On Volatility - β_τ

Posterior estimates of uninformed trading on volatility for ASH in models C2 and C4 are respectively $\beta_\tau^{ASH} = -0.340$ and $\beta_\tau^{ASH} = -0.368$. These are also negative and significant. The effect is marginally high in model C4 compared with model C2. The estimates $\beta_\tau^{IBM} = -0.101$ and $\beta_\tau^{IBM} = -0.145$, also show that the effect of uninformed trading on volatility is negative. However the credible intervals $(-0.423,\ 0.190)$ for model C2 and $(-0.493,\ 0.160)$ for model C4 indicate that the effect of uninformed trading on volatility in IBM is insignificant. From these results, there is evidence that z_{t-1} provides no additional information.

Additional Learning Of α_h And α_τ From Realized Variance

From Table 5.2 and 5.3 it can be observed that α_h and α_τ for both assets are significant across all models. For ASH, the posterior estimate for model C2 is $\alpha_h^{ASH} = 0.348$ with a credible interval $(0.197,\ 0.512)$. This estimate is higher than $\alpha_h^{ASH} = 0.212$ obtained from model A1 in chapter 3. Similarly the estimate $\alpha_h^{IBM} = 0.344$ for IBM is also higher than that obtained in the previous chapter.

Likewise, the effect of the uninformed trading on bid-ask spread $\alpha_\tau^{ASH} = -0.253$ and $\alpha_\tau^{IBM} = -0.258$ in the previous chapter are higher than the respective effects $\alpha_\tau^{ASH} = -0.388$ and $\alpha_\tau^{IBM} = -0.324$ obtained in model C2. These results is an indication that volatility provides additional information that can be used to infer the effect of informed and uninformed trading on bid-ask spread.

Persistence Of τ_t And h_t

The estimates of ϕ in the unobserved informed and uninformed trading for both assets across all models are considerably high. The estimates of persistence in the informed trading for models C2, C3 and C4 for ASH are $\phi_h^{ASH} = 0.866$, $\phi_h^{ASH} = 0.881$ and $\phi_h^{ASH} = 0.862$ respectively. Their respective credible intervals show that these are highly significant. We obtain similar significant and high persistence in the informed trading for IBM. The estimates of persistence in uniformed trading are similarly significant for both assets and across all models.

In all the models considered, the posterior estimates of the parameters $\nu_y, \nu_z, \nu_x, \nu_h$ and ν_τ show that the distribution of log of the observed series and latent processes are approximately Gaussian in both assets. The parameter estimates are stable across all models for both assets.

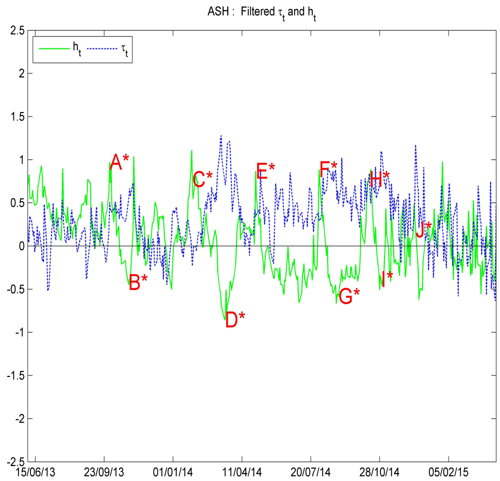
Temporal Behaviour Of Volatility, Informed And Uninformed Trading

We compare the components of volume inferred from models C2, C3 and C4 with the realized variance for ASH and IBM in Figures 5.1 and 5.2 respectively. On the days marked **E***, **F*** and **H*** in Figure 5.1, both the informed and uninformed trading components were high. It can be observed that volatility of ASH on these days are high as expected. Likewise, the volatility of IBM was high on the days where both informed and uninformed trading were high. This relationship between volatility, informed and uninformed trading can be seen in Fig 5.2.

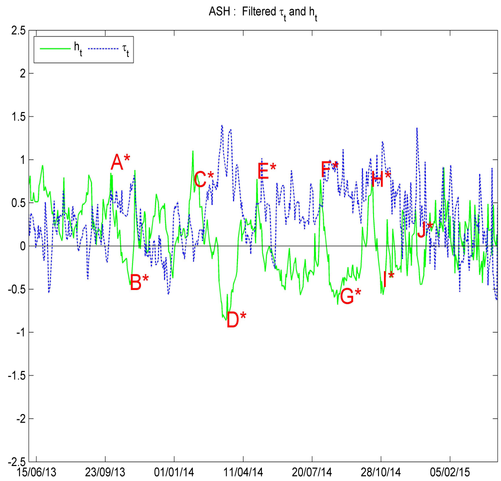
A trading day with low informed trading and a high uninformed trading is predicted to result in low volatility. We notice this relationship on the days marked **B*** and **D*** in Figure 5.1 as well as points marked **C** and **F** in Figure 5.2.

On the other hand, a trading day with high informed trading and a low uninformed trading is predicted to have high volatility. We notice this relationship on the days marked **C*** in Figure 5.1 as well as points marked **B**, and **D** in Figure 5.2.

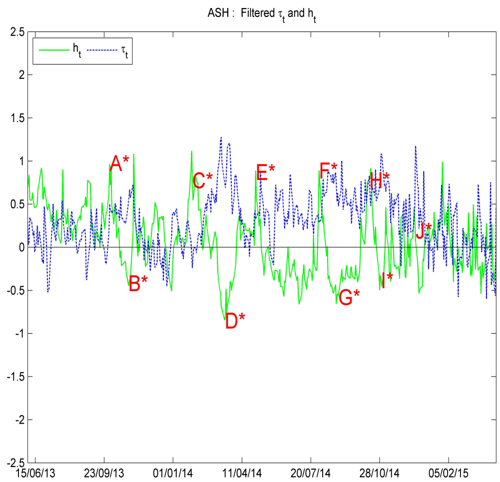
The findings above is an indication that we can use the joint relationships between volume, realized variance and bid-ask spread to understand the effects of latent liquidity risk on volatility.



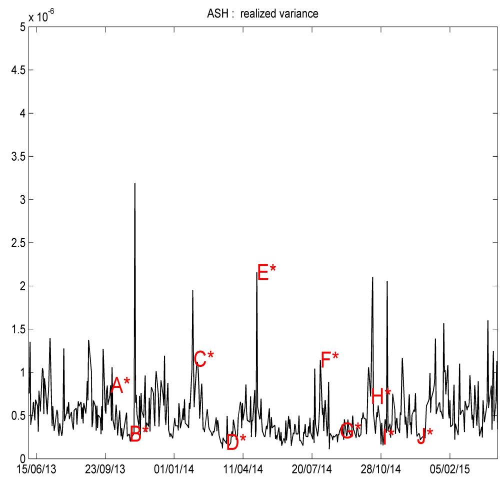
(a) Model C2



(b) Model C3

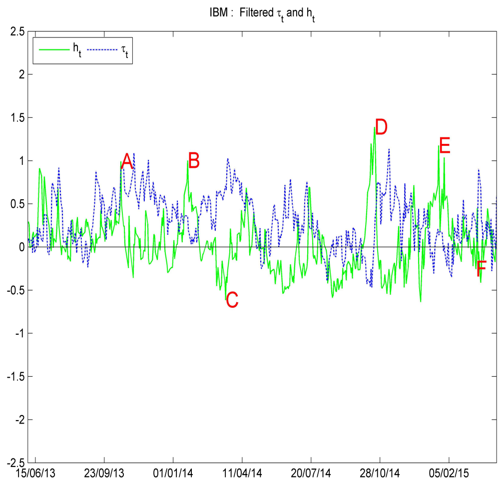


(c) Model C4

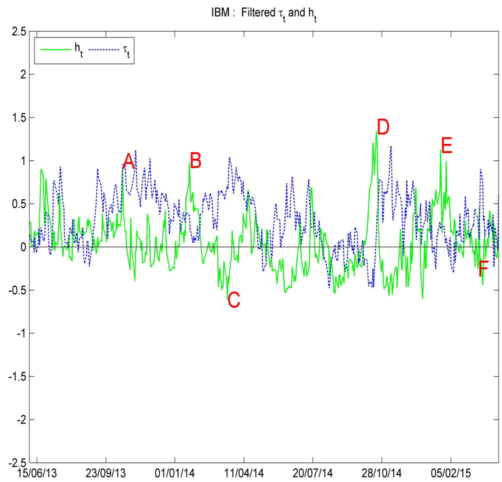


(d) Realized variance

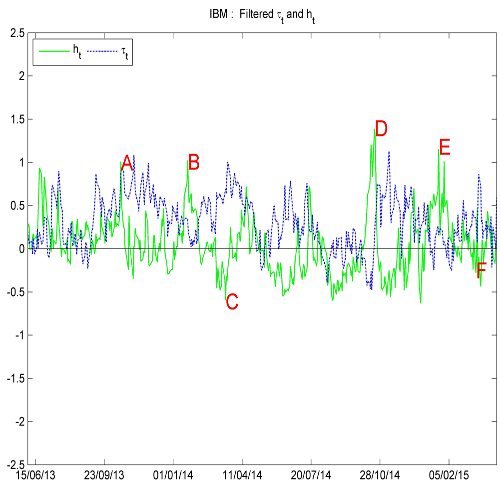
Fig. 5.1 ASH - realized variance compared with h_t and τ_t



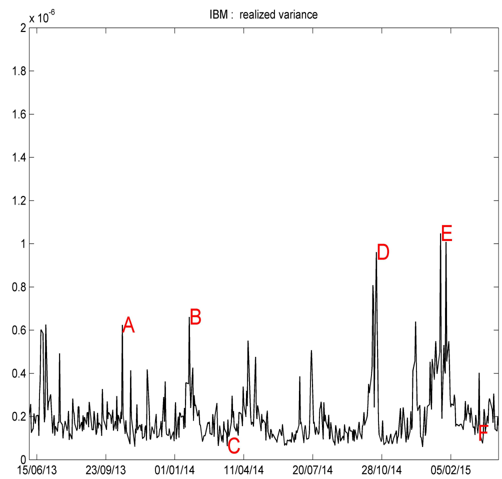
(a) Model C2



(b) Model C3



(c) Model C4



(d) Realized variance

Fig. 5.2 IBM - realized variance compared with h_t and τ_t

5.5 One Day Ahead Volatility Forecasting

The models developed in the preceding sections may be utilised to produce volatility forecasts. In what follows we provide a brief summary of Bayesian forecasting approach we take. The forecasts from the state space model are based on the posterior predictive distribution, the draws of the parameters and the latent state. Conditional on the information available $Y_{1:t}$, the one step ahead predictive density of latent state process β_t can be computed as

$$p(\beta_{t+1}|Y_{1:t},\Theta) = \int p(\beta_{t+1}|\beta_t, Y_{1:t},\Theta)p(\beta_t|Y_{1:t},\Theta)d\beta_t. \tag{5.15}$$

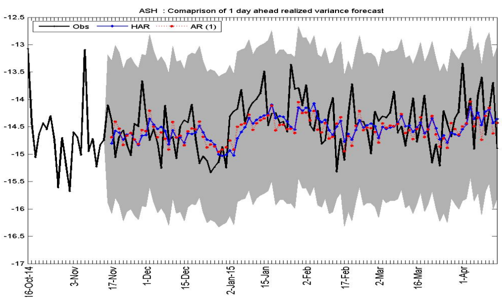
The one step ahead predictive density of the observable Y_{t+1}, conditional on information available at time t is also computed as

$$p(Y_{t+1}|Y_{1:t}) = \int p(Y_{t+1}|\beta_{t+1},\Theta)p(\beta_{t+1}|Y_{1:t},\Theta)p(\Theta|Y_{1:t})d\beta_{t+1}d\Theta. \tag{5.16}$$

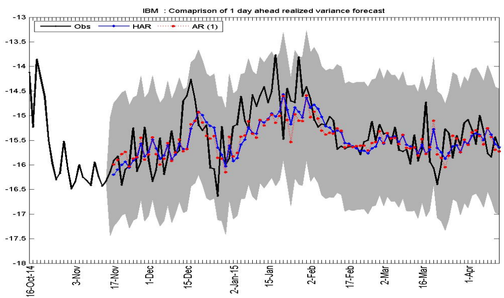
There are three uncertainties which are in the predictive density shown in equation 5.16. Firstly the uncertainty about the model parameters are captured by the posterior distribution $p(\Theta|Y_{1:t})$. Secondly there is uncertainty about the future evolution of the latent state vector which is represented by $p(\beta_{t+1}|\beta_t,\Theta,Y_{1:t})$. Finally the uncertainty about the future realizations of observed data is captured in $p(Y_{t+1}|\beta_{t+1},\Theta)$.

For model C5, AR (1) and the HAR model, we use the parameter estimates obtained from the entire history of the observed series up to time $t-1$ to compute the forecast for the period t. The one day ahead forecast is then computed using the lag 1 value of the observed series. On the other hand, the forecasts from the SMC are just based on the current observed data point and the most recent values of the unobserved state β_t which contains all the needed knowledge accumulated from the history of the observed series.

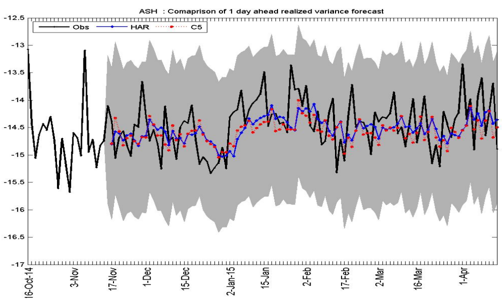
In Figure 5.3 below, we present the one day ahead forecasts and 95% confidence intervals of the AR (1) and model C5 compared with the HAR model. The daily forecast, which is the mean of the predictive distribution with corresponding 95% credible intervals for models C1, C1, C3, and C4 compared with the HAR model forecasts are shown in Figure 5.4. For a well calibrated interval forecast, it is expected that observed data points fall inside the 95% credible intervals. It can be noted that the HAR, AR (1) and model C5 are unable to capture peaks and troughs in the volatility dynamics. In contrast, the high and low periods of the volatility dynamics are better modelled in models C1, C2, C3 and C4. Comparisons of forecasts for all models are also shown in Figures 5.5 and 5.6



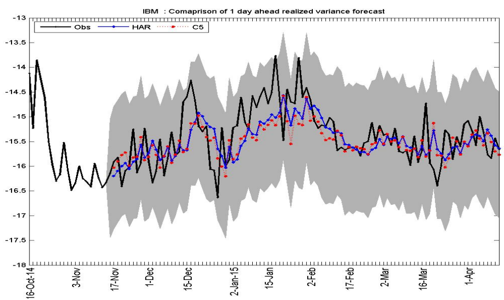
(a) ASH Model AR (1)



(b) IBM Model AR (1)

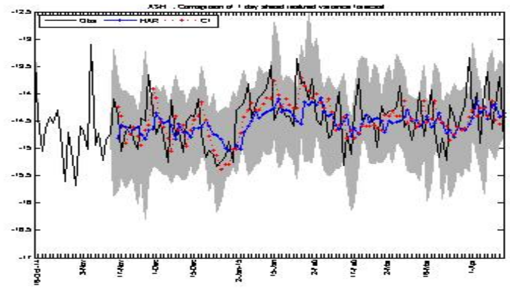


(c) ASH Model C5

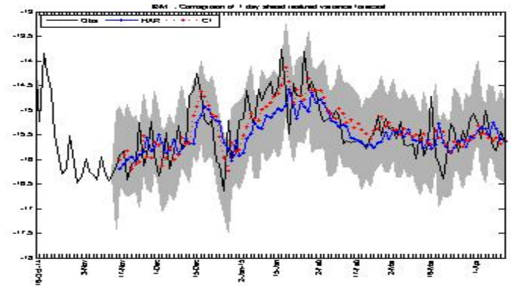


(d) IBM Model C5

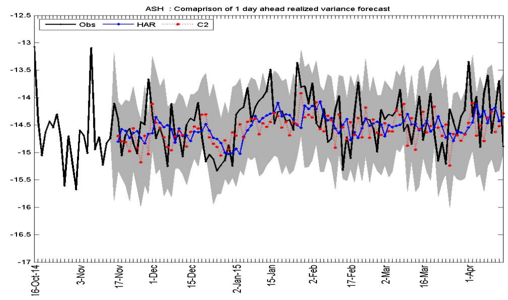
Fig. 5.3 Benchmark models - comparison of one day ahead volatility forecasts.



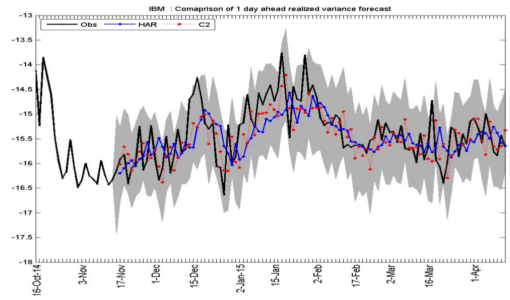
(a) ASH Model C1



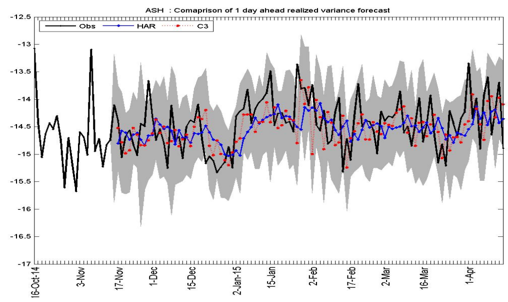
(b) IBM Model C1



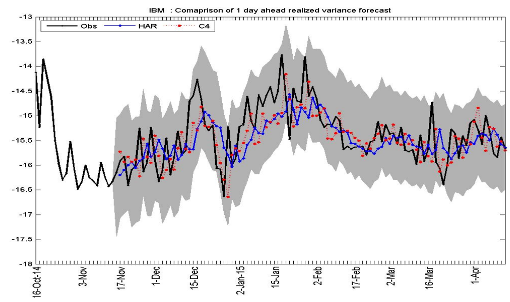
(c) ASH Model C2



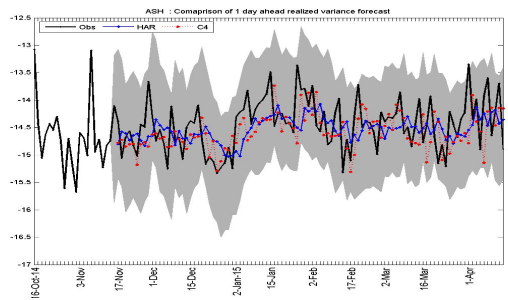
(d) IBM Model C2



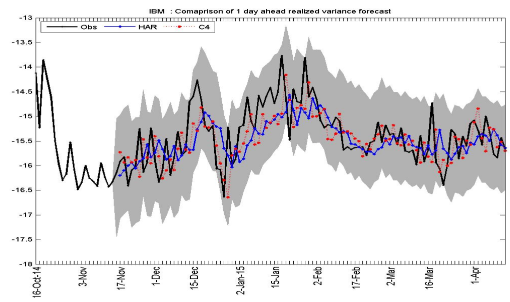
(e) ASH Model C3



(f) IBM Model C3

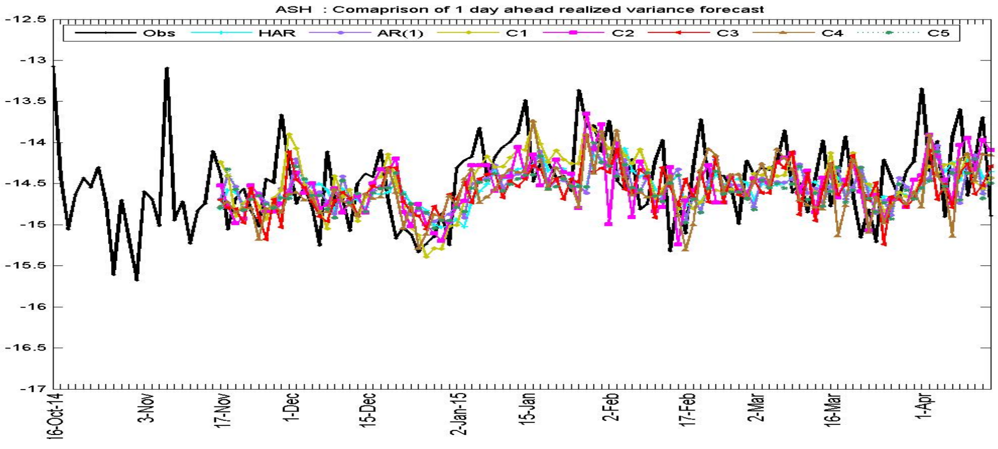


(g) ASH Model C4

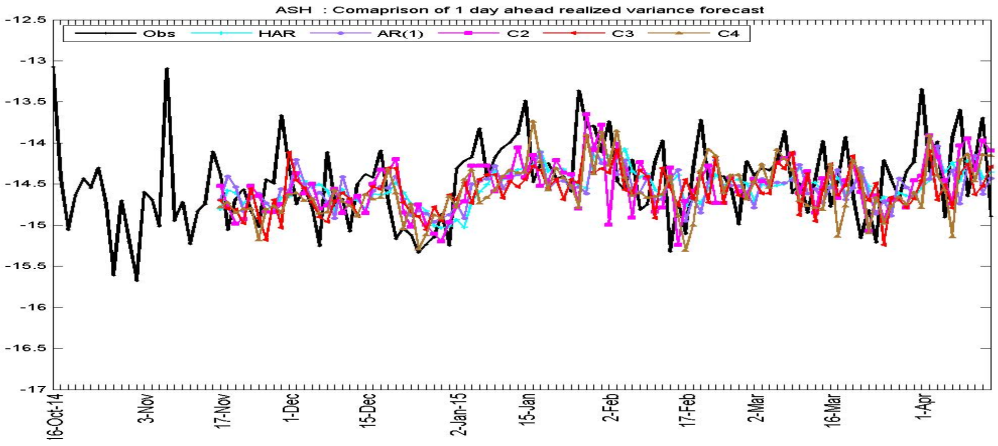


(h) IBM Model C4

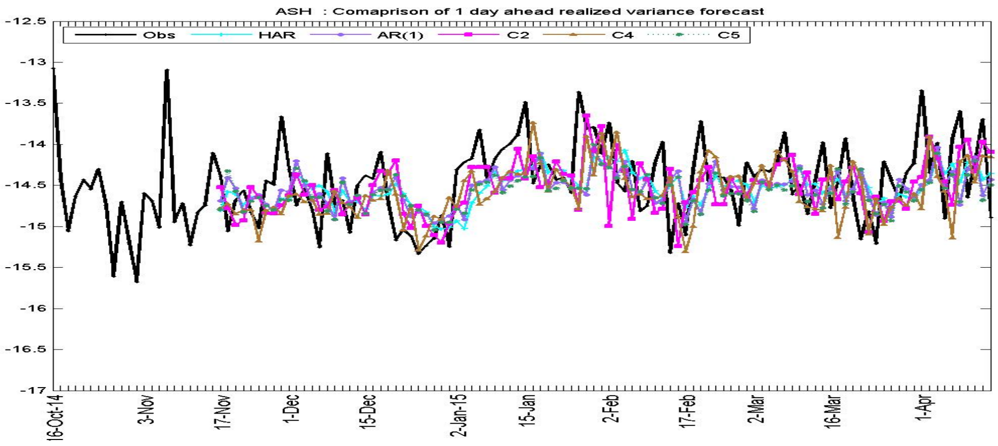
Fig. 5.4 HAR versus other models - comparison of one day ahead volatility forecasts.



(a)

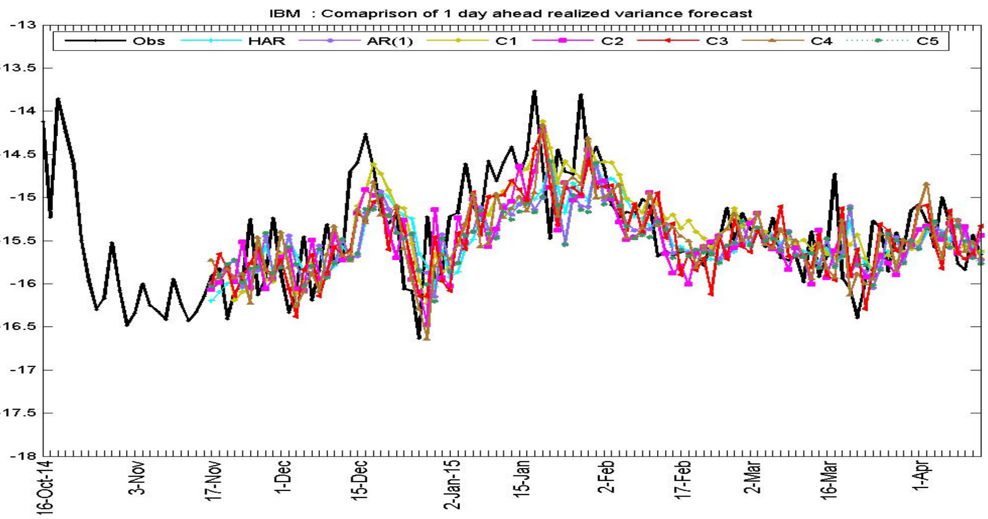


(b)

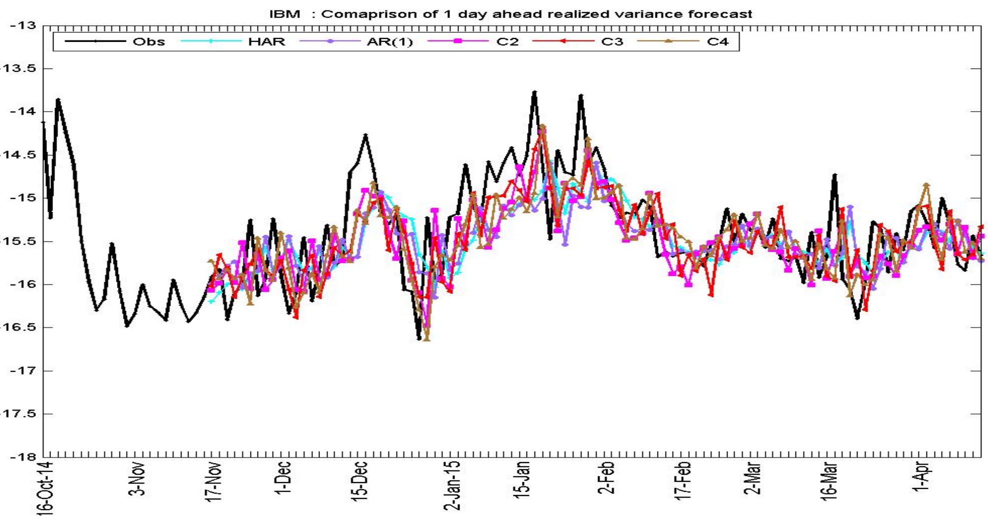


(c)

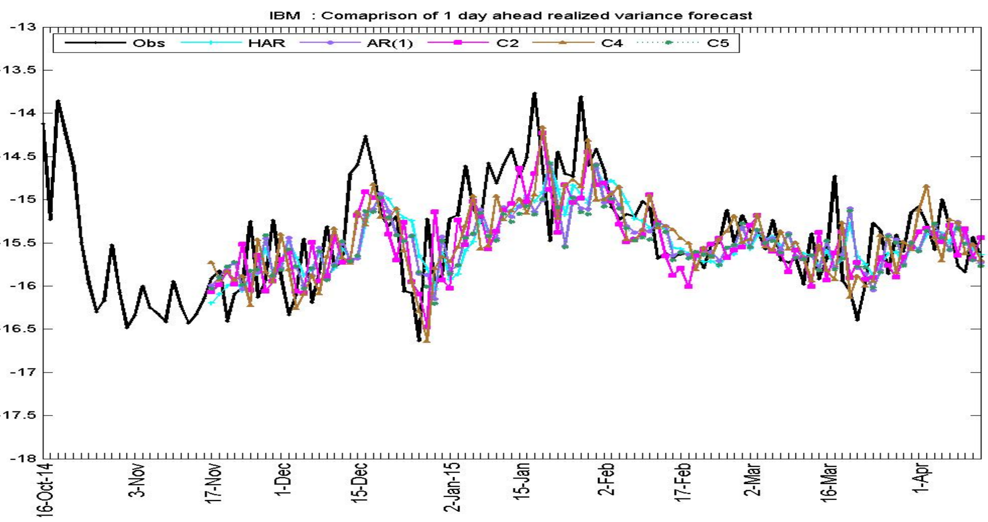
Fig. 5.5 ASH - Comparison of one day ahead volatility forecasts



(a)



(b)



(c)

Fig. 5.6 IBM - Comparison of one day ahead volatility forecasts

In a Bayesian estimation setting, the predictive likelihood is of particular relevance when the objective is to choose between models based on forecast comparisons. We employ the log predictive score to assess the fit of the models. To assess the forecasting accuracy of the models, the root mean square of the forecast errors (RMSE) are computed for each model. Let $\hat{Y}_t$ denote the one day ahead forecast of the observed Y_t, the RMSE and average log predictive score for a one-day ahead forecast are respectively given as

$$RMSE = \sqrt{\frac{1}{(N-T)} \sum_{t=T+1}^{N} (\hat{Y}_t - Y_t)^2} \tag{5.17}$$

and

$$LPS = -\frac{1}{(N-T)} \sum_{t=T+1}^{N} \log p(Y_t|Y_{t-1}, \hat{\Theta}), \tag{5.18}$$

where M is the number of particles used in the SMC, $\hat{Y}_t = \frac{1}{M}\sum_{i=1}^{M} Y_t^i$ and $\log p(Y_t|Y_{t-1}, \hat{\Theta}) = \frac{1}{M}\sum_{i=1}^{M} \log p(Y_t|\Theta^{(i)}, Y_{1:t}, \beta_{t-1}^{(i)})$. The number of days in the training sample and the number of days for which one-step ahead forecasts were generated are T and $N-T$ respectively.

The LPS provides guidance on the overall fit of the model. A smaller value of LPS is an indication of a better model fit. In making decisions about the comparative analysis of model fit, we base our evaluation of the models solely on the LPS. We do this because model evaluation based on point forecasts typically disregards the uncertainty surrounding predictions. The LPS and RMSE for all models are presented in Table 5.4. The RMSE of the HAR and AR(1) models for IBM are respectively given as 0.47992 and 0.49216. These estimates show that the HAR model does a better job at forecasting the one day ahead volatility of IBM than the AR (1). In addition, for IBM, the HAR and AR(1) models have less forecasting accuracy in comparison with models C1, C2, C3 and C4 since the RMSE of these models are smaller albeit marginal than that of the HAR and AR(1) models. Thus, the bid-ask spread in models C2, C3 and

		RMSE		LPS	
	Model	**ASH**	**IBM**	**ASH**	**IBM**
Full Model	C1			12.56415	12.07233
	C2			12.21805	11.65621
	C3			12.31803	11.57342
	C4			12.14872	11.54800
Realized	AR (1)	0.46512	0.49216		
	HAR	0.45517	0.47992		
	C1	0.44867	0.43728		
	C2	0.47347	0.47133		
	C3	0.48708	0.45031		
	C4	0.48883	0.47350		
	C5	0.49991	0.49486		

Table 5.4 RMSE and LPS of one day ahead volatility forecasts

C4 provide additional information that can be used in forecasting short term volatility of IBM. For IBM, model C5 has the largest RMSE amongst all the models. We obtain mixed results for ASH. The RMSE of ASH for model C1 is the smallest as in the case of IBM. However the RMSE of ASH for the HAR and AR (1) models are smaller in comparison with C2, C3, C4 and C5 which is the opposite results we found for IBM.

Amongst the models we propose, the RMSE of model C1 for IBM is the smallest in comparison with models C2, C3 and C4 all of which include bid-ask spread as an additional variable. Similar results for ASH were obtained. These results seem to suggest that additional sources of information beyond that contained in volume is needed for the prediction of short term volatility of the asset. For both assets, model C5 is the worst performing models it has the largest RMSE. From the results, it seems model one performs better than other models. In model C1 we don not separate out the informed and liquidity components. It is likely that without the separation of the components some information is masked hence through some smoothing out effect. This might account for its seemingly relative good performance.

Using the results from the RMSE and LPS it is not clear which of the models is superior in forecasting the volatility. Hansen et al. (2011) introduced a method that can be used to rank several models in terms of their superiority. In what follows we utilise the Model Confidence Set of Hansen et al. (2011) to provide a ranking of the models considered in this chapter. We first describe briefly the methodology and the proceed to utilise the *mcs* matlab function of the *Oxford MFE Toolbox* for the empirical implementation.

Given that we have and observed quantity Y_t and a corresponding forecast from model i as $\hat{Y}_{i,t}$. Let $L_{i,t} = L(Y_t, \hat{Y}_{i,t})$ be a loss function for $i = 1, \ldots, M$ models. Hansen et al. (2011) defines a relative performance measure $d_{ij,t} = L_{i,t} - L_{j,t},\ i, j \in M$ between models i and j respectively. The following sample loss statistics are computed

$$\bar{d}_{ij} \equiv \frac{1}{(N-T)} \sum_{t=T+1}^{N} d_{ij,t}, \quad t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{Var(\bar{d}_{ij})}}, \quad \forall i, j \in M$$

$$\bar{d}_{i.} \equiv -\frac{1}{M} \sum_{j \in M} \hat{d}_{ij}, \qquad t_{i.} = \frac{\bar{d}_{i.}}{\sqrt{Var(\bar{d}_{i.})}} \qquad \forall i, j \in M,$$

where $\hat{d}_{ij}$ measures the relative sample loss between the i^{th} and j^{th} models. The sample loss of the i^{th} model relative to the average across all models in the set M is given by $\bar{d}_{i.}$. The quantities $Var(\bar{d}_{ij})$ and $Var(\bar{d}_{i.})$ are the bootstrap variances of $\bar{d}_{ij}$ and $\bar{d}_{i.}$ respectively. Under this method the null hypothesis is that all models are of equal predictive power. By sequentially comparing the competing models and retaining the superior ones, we arrive at the Model Confidence Set. In our case we chose the loss function to be the difference between the observed value an its forecast, that is $L_{i,t} = Y_t - \hat{Y}_{i,t}$. The results of the empirical implementation is given in Table 5.5.

The results in Table 5.5 gives mixed findings. In respect of ASH, only three models were retained with C4 being ranked as superior. However in the case of IBM, two

	Model	p-value	Rank
IBM	C2	0.688	2
	C5	1	1
ASH	C3	0.954	3
	C5	0.978	2
	C4	1	1

Table 5.5 Model Confidence Set for one day ahead volatility forecasts

models were retained. Model C5 is ranked as superior to model C2. Model C2 was ranked to be the next superior model. The results above suggests a further look at the alternative models deeper.

5.6 Concluding Remarks

In this chapter, we have exploited the relationship between volume, bid-ask spread and volatility. The relationship provided insight on the temporal relationship between volatility, informed and uninformed trading. The empirical findings are that informed trading has a significant positive effect on volatility for the assets considered. On the other hand, uninformed trading seems to significantly effect relatively illiquid assets while it has no effect on liquid assets. Since we had only two assets in this study, it would be interesting to implement this model on a large number of assets. Implementing our model on a large number of data sets will provide us with a broader picture of the link between volatility, informed and uninformed trading.

Secondly, we have undertaken a comparison of some alternative volatility forecasting models. The empirical analysis carried out indicates that our models have the potential to be used in forecasting volatility. Using the log predictive score as a measure of model fit, we find that bid-ask spread contributes additional information to the volume-volatility relationship for the forecasting of short-term volatility. From the results, one would be inclined to say that models C2, C3 and C4 compare equally with the

AR (1) and HAR models. However, in the estimation process, the parameters of the AR (1) and HAR models were recalculated with the entire history of the realized variance every time a new forecast is generated. This is in contrast to the assumption of invariant model parameters in models C1, C2, C3 and C4. Using the SMC we did not have the need for the entire history of the volume, bid-ask spread and realized variance yet we obtained results that are comparable and in certain cases superior to the AR (1) and HAR. Thus the relationship between volume, volatility and bid-ask spread is very insightful for the forecasting of volatility.

Chapter 6

Conclusions And Further Research

The contribution of this thesis is threefold. First, we provide an alternative method for the empirical estimation of the Probability of Informed Trading. Second, we propose a new method of inferring liquidity and informed trading. Finally, we show that the information contained in bid-ask spread can be used to forecast volatility. We summarise the main conclusions of this thesis and then discuss some areas for possible future research.

In Chapter 2 we developed a Bayesian estimation methodology for the PIN measure. The main conclusions of this work were:

- The Bayesian estimation method provides a way to circumvent the numerical instability problem of the MLE which has been reported in some papers that studied PIN.
- PIN is higher when computed on high frequency data than on low frequency data. Thus PIN estimated from daily buy and sell trades may underestimate information asymmetry.
- High frequency data collected over the most recent few days is sufficient to compute the risk of informed trading.

- Maximum likelihood estimation of PIN is dependent on initial values and the factorised form of the likelihood function.

The methodology developed in Chapter 2 was then applied to high frequency buy and sell trades in Chapter 3. The findings of a comparison between PIN and VPIN indicates that the VPIN gives better insight on times series behaiour information asymmetry risk.

Chapter 4 utilised the theoretical relationship between traded volume and the bid-ask spread to extract and learn about the dynamics of informed trading. This is the first attempt to infer informed trading from the joint relationship between volume and bid –ask spread using Bayesian methods. Our findings are as follows

- The dynamics of inferred informed trading component of volume mimics patterns in corporate events of the assets considered.
- The informed and liquidity driven components of volume are time-varying and persistent.

Finally in Chapter 5, we investigated the relationship between volatility, informed and uninformed trading using the model developed in Chapter 3. These models are based on the intuition behind the link between volume, bid – ask spread and volatility. The models were also bench-marked against the AR(1) and HAR volatility models. The main conclusions from the chapter are:

- Informed trading effect positively on volatility in all assets studied.
- Uninformed trading does not effect volatility of liquid assets but has a significant negative effect on less liquid assets.
- The out-of–sample forecasting performance of the models considered in this thesis are comparable to the HAR and AR (1) models.

- The bid-ask spread provides additional information that can be used for the forecasting of short-term (i.e., daily) volatility.

In Chapters 4 and 5, model parameters were considered to be fixed. However as market conditions change over time, it is reasonable to conjecture that the parameters would change with the economic conditions. Secondly, we also assumed that the relationship between informed trading and bid-ask spread is unidirectional from informed trading to bid-ask spread. There is evidence to suggest that there is a feedback relationship between these variables. The models considered can be re-stated to account for the feedback relationship. In the thesis, we have focused on daily information and volatility. It is desirable to explore the informed trading and volatility in almost real time. This will require the implementation of the models in this thesis on high frequency data. Future research is intended to be carried out using long time series for a large number of assets.

Lastly, the impact of trade on asset price which is a measure if the liquidity of an asset changes over time. The Kyle (1985) parameter (lambda) is a measure of the price impact of trade. However, the Kyle (1985) parameter has been treated as fixed over sample periods. Using SMC techniques, it is possible to estimate the time-varying liquidity in Kyle (1985) model as a stochastic process. We intend to consider this problem in future.

References

Abad, D. and Yagüe, J. (2012), 'From pin to vpin: An introduction to order flow toxicity', *The Spanish Review of Financial Economics* **10**(2), 74–83.

Abanto-Valle, C. A., Migon, H. S. and Lopes, H. F. (2010), 'Bayesian modeling of financial returns: a relationship between volatility and trading volume', *Applied Stochastic Models in Business and Industry* **26**(2), 172–193.

Acharya, V. V. and Pedersen, L. H. (2005), 'Asset pricing with liquidity risk', *Journal of Financial Economics* **77**(2), 375–410.

Admati, A. R. and Pfleiderer, P. (1988), 'A theory of intraday patterns: Volume and price variability', *Review of Financial Studies* **1**(1), 3–40.

Amihud, Y. and Mendelson, H. (1980), 'Dealership market: Market-making with inventory', *Journal of Financial Economics* **8**(1), 31–53.

Andersen, T. G. (1996), 'Return volatility and trading volume: An information flow interpretation of stochastic volatility', *The Journal of Finance* **51**(1), 169–204.

Andersen, T. G. and Bondarenko, O. (2014*a*), 'Assessing measures of order flow toxicity and early warning signals for market turbulence*', *Review of Finance* p. rfu041.

Andersen, T. G. and Bondarenko, O. (2014*b*), 'Vpin and the flash crash', *Journal of Financial Markets* **17**, 1–46.

Bagehot, W. (1971), 'The only game in town', *Financial Analysts Journal* **27**(2), 12–14.

Bessembinder, H. (1994), 'Bid-ask spreads in the interbank foreign exchange markets', *Journal of Financial Economics* **35**(3), 317–348.

Bessembinder, H. and Seguin, P. J. (1993), 'Price volatility, trading volume, and market depth: Evidence from futures markets', *Journal of Financial and Quantitative Analysis* **28**(01), 21–39.

Boehmer, E., Grammig, J. and Theissen, E. (2007), 'Estimating the probability of informed trading-does trade misclassification matter?', *Journal of Financial Markets* **10**(1), 26–47.

Bollen, N. P., Smith, T. and Whaley, R. E. (2004), 'Modeling the bid/ask spread: measuring the inventory-holding premium', *Journal of Financial Economics* **72**(1), 97–141.

Bollerslev, T. and Melvin, M. (1994), 'Bid—ask spreads and volatility in the foreign exchange market: An empirical analysis', *Journal of International Economics* **36**(3), 355–372.

Borisova, G. and Yadav, P. K. (2015), 'Government ownership, informed trading, and private information', *Journal of Corporate Finance* **33**, 196–211.

Brennan, M. J., Chordia, T., Subrahmanyam, A. and Tong, Q. (2012), 'Sell-order liquidity and the cross-section of expected stock returns', *Journal of Financial Economics* **105**(3), 523–541.

Brennan, M. J. and Subrahmanyam, A. (1996), 'Market microstructure and asset pricing: On the compensation for illiquidity in stock returns', *Journal of Financial Economics* **41**(3), 441–464.

Carter, C. K. and Kohn, R. (1994), 'On gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Chib, S., Nardari, F. and Shephard, N. (2002), 'Markov chain monte carlo methods for stochastic volatility models', *Journal of Econometrics* **108**(2), 281–316.

Chordia, T., Subrahmanyam, A. and Anshuman, V. R. (2001), 'Trading activity and expected stock returns', *Journal of Financial Economics* **59**(1), 3–32.

Chung, K. H. and Li, M. (2003), 'Adverse-selection costs and the probability of information-based trading', *Financial Review* **38**(2), 257–272.

Clark, P. K. (1973), 'A subordinated stochastic process model with finite variance for speculative prices', *Econometrica: Journal of the Econometric Society* pp. 135–155.

Cont, R., Kukanov, A. and Stoikov, S. (2014), 'The price impact of order book events', *Journal of Financial Econometrics* **12**(1), 47–88.

Copeland, T. E. and Galai, D. (1983), 'Information effects on the bid-ask spread', *The Journal of Finance* **38**(5), 1457–1469.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Danielsson, J. and Payne, R. (2001), Measuring and explaining liquidity on an electronic limit order book: evidence from reuters d2000-2, *in* 'EFA 2001 Barcelona Meetings'.

Darrat, A. F., Rahman, S. and Zhong, M. (2003), ‘Intraday trading volume and return volatility of the djia stocks: A note’, *Journal of Banking & Finance* **27**(10), 2035–2043.

Dufour, A. and Engle, R. F. (2000), ‘Time and the price impact of a trade’, *The Journal of Finance* **55**(6), 2467–2498.

Easley, D., Engle, R. F., O’Hara, M. and Wu, L. (2008), ‘Time-varying arrival rates of informed and uninformed trades’, *Journal of Financial Econometrics* **6**(2), 171–207.

Easley, D., Hvidkjaer, S. and O’Hara, M. (2002), ‘Is information risk a determinant of asset returns?’, *The Journal of Finance* **57**(5), 2185–2221.

Easley, D., Hvidkjaer, S. and O’Hara, M. (2010), ‘Factoring information into returns.’, *Journal of Financial & Quantitative Analysis* **45**(2), 293 – 309.

Easley, D., Kiefer, N. M. and O’Hara, M. (1997*a*), ‘The information content of the trading process’, *Journal of Empirical Finance* **4**(2), 159–186.

Easley, D., Kiefer, N. M. and O’Hara, M. (1997*b*), ‘One day in the life of a very common stock’, *Review of Financial Studies* **10**(3), 805–835.

Easley, D., Kiefer, N. M., O’Hara, M. and Paperman, J. B. (1996), ‘Liquidity, information, and infrequently traded stocks’, *The Journal of Finance* **51**(4), 1405–1436.

Easley, D., López de Prado, M. and O’Hara, M. (2011), ‘The microstructure of the ‘flash crash’: Flow toxicity, liquidity crashes and the probability of informed trading’, *The Journal of Portfolio Management* **37**(2), 118–128.

Easley, D. and O’Hara, M. (1987), ‘Price, trade size, and information in securities markets’, *Journal of Financial Economics* **19**(1), 69–90.

Easley, D. and O’Hara, M. (1992*a*), ‘Adverse selection and large trade volume: The implications for market efficiency’, *Journal of Financial and Quantitative Analysis* **27**(02), 185–208.

Easley, D. and O’Hara, M. (1992*b*), ‘Time and the process of security price adjustment’, *The Journal of Finance* **47**(2), 577–605.

Easley, D. and O’Hara, M. (2003), ‘Microstructure and asset pricing’, *Handbook of the Economics of Finance* **1**, 1021–1051.

Foster, F. D. and Viswanathan, S. (1990), ‘A theory of the interday variations in volume, variance, and trading costs in securities markets’, *Review of Financial Studies* **3**(4), 593–624.

Foster, F. D. and Viswanathan, S. (1993), ‘Variations in trading volume, return volatility, and trading costs; evidence on recent price formation models’, *The Journal of Finance* **48**(1), 187–211.

Foster, F. D. and Viswanathan, S. (1994), 'Strategic trading with asymmetrically informed traders and long-lived information', *Journal of Financial and Quantitative Analysis* **29**(04), 499–518.

French, K. R. and Roll, R. (1986), 'Stock return variances: The arrival of information and the reaction of traders', *Journal of Financial Economics* **17**(1), 5–26.

Gelman, A., Roberts, G. and Gilks, W. (1996), 'Efficient metropolis jumping hules', *Bayesian Statistics* **5**(599-608), 42.

George, T. J., Kaul, G. and Nimalendran, M. (1991), 'Estimation of the bid–ask spread and its components: A new approach', *Review of Financial Studies* **4**(4), 623–656.

Geweke, J. (1993), 'Bayesian treatment of the independent student-t linear model', *Journal of Applied Econometrics* **8**(S1), S19–S40.

Girard, E. and Biswas, R. (2007), 'Trading volume and market volatility: Developed versus emerging stock markets', *Financial Review* **42**(3), 429–459.

Glosten, L. R. (1987), 'Components of the bid-ask spread and the statistical properties of transaction prices', *The Journal of Finance* **42**(5), 1293–1307.

Glosten, L. R. (1989), 'Insider trading, liquidity, and the role of the monopolist specialist', *Journal of Business* pp. 211–235.

Glosten, L. R. and Harris, L. E. (1988), 'Estimating the components of the bid/ask spread', *Journal of Financial Economics* **21**(1), 123–142.

Glosten, L. R. and Milgrom, P. R. (1985), 'Bid, ask and transaction prices in a specialist market with heterogeneously informed traders', *Journal of Financial Economics* **14**(1), 71–100.

Haario, H., Saksman, E. and Tamminen, J. (2001), 'An adaptive metropolis algorithm', *Bernoulli* pp. 223–242.

Hallin, M., Mathias, C., Pirotte, H. and Veredas, D. (2011), 'Market liquidity as dynamic factors', *Journal of Econometrics* **163**(1), 42–50.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011), 'The model confidence set', *Econometrica* **79**(2), 453–497.

Harris, L. (1987), 'Transaction data tests of the mixture of distributions hypothesis', *Journal of Financial and Quantitative Analysis* **22**(02), 127–141.

Harris, L. (2002), *Trading and exchanges: Market microstructure for practitioners*, Oxford university press.

Harvey, A., Ruiz, E. and Sentana, E. (1992), 'Unobserved component time series models with arch disturbances', *Journal of Econometrics* **52**(1-2), 129–157.

Hasbrouck, J. (1991), 'Measuring the information content of stock trades', *The Journal of Finance* **46**(1), 179–207.

Huang, R. D. and Stoll, H. R. (1997), 'The components of the bid-ask spread: A general approach', *Review of Financial Studies* **10**(4), 995–1034.

Huh, S.-W. (2014), 'Price impact and asset pricing', *Journal of Financial Markets* **19**, 1–38.

Jorion, P. (1996), Risk and turnover in the foreign exchange market, *in* 'The microstructure of foreign exchange markets', University of Chicago Press, pp. 19–40.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering* **82**(1), 35–45.

Kandel, E. and Pearson, N. D. (1995), 'Differential interpretation of public signals and trade in speculative markets', *Journal of Political Economy* pp. 831–872.

Kastner, G. and Frühwirth-Schnatter, S. (2014), 'Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models', *Computational Statistics & Data Analysis* **76**, 408–423.

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: likelihood inference and comparison with arch models', *The Review of Economic Studies* **65**(3), 361–393.

Kumar, R. and Popescu, M. (2013), 'The implied intra-day probability of informed trading', *Review of Quantitative Finance and Accounting* pp. 1–15.

Kyle, A. (1985), 'Continuous auctions and insider trading', *Econometrica: Journal of the Econometric Society* pp. 1315–1335.

Kyle, A. S. (1984), 'Market structure, information, futures markets, and price formation.'.

Lamoureux, C. G. and Lastrapes, W. D. (1994), 'Endogenous trading volume and momentum in stock-return volatility', *Journal of Business & Economic Statistics* **12**(2), 253–260.

Lee, C. M. C. and Ready, M. J. (1991), 'Inferring trade direction from intraday data', *The Journal of Finance* **46**(2), 733–46.

Lei, Q. and Wu, G. (2005), 'Time-varying informed and uninformed trading activities', *Journal of Financial Markets* **8**(2), 153–181.

Lesmond, D. A. (2005), 'Liquidity of emerging markets', *Journal of Financial Economics* **77**(2), 411–452.

Li, J. and Wu, C. (2006), 'Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume*', *The Journal of Business* **79**(5), 2697–2739.

Liesenfeld, R. (2001), 'A generalized bivariate mixture model for stock price volatility and trading volume', *Journal of Econometrics* **104**(1), 141–178.

Lin, W. H.-W. and Ke, W.-C. (2011), 'A computing bias in estimating the probability of informed trading', *Journal of Financial Markets* **14**(4), 625–640.

Liu, J. and West, M. (2001), Combined parameter and state estimation in simulation-based filtering, *in* 'Sequential Monte Carlo methods in practice', Springer, pp. 197–223.

Lopes, H. F. and Tsay, R. S. (2011), 'Particle filters and bayesian inference in financial econometrics', *Journal of Forecasting* **30**(1), 168–209.

Madhavan, A. (2000), 'Market microstructure: A survey', *Journal of Financial Markets* **3**(3), 205–258.

Madhavan, A., Richardson, M. and Roomans, M. (1997), 'Why do security prices change? a transaction-level analysis of nyse stocks', *Review of Financial Studies* **10**(4), 1035–1064.

Mahieu, R. and Bauer, R. (1998), 'A bayesian analysis of stock return volatility and trading volume', *Applied Financial Economics* **8**(6), 671–687.

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*, Vol. 108, Blackwell Cambridge.

Prado, R. and West, M. (2010), *Time series: modeling, computation, and inference*, CRC Press.

Roberts, G. O., Gelman, A., Gilks, W. R. et al. (1997), 'Weak convergence and optimal scaling of random walk metropolis algorithms', *The annals of applied probability* **7**(1), 110–120.

Sadka, R. (2006), 'Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk', *Journal of Financial Economics* **80**(2), 309–349.

Schervish, M. J. (2012), *Theory of statistics*, Springer Science & Business Media.

Sekerke, M. (2015), *Bayesian Risk Management: A Guide to Model Risk and Sequential Learning in Financial Markets*, John Wiley & Sons.

Spulber, D. F. (1996), 'Market microstructure and intermediation', *The Journal of Economic Perspectives* **10**(3), 135–152.

Strickland, C. M., Martin, G. M. and Forbes, C. S. (2008), 'Parameterisation and efficient mcmc estimation of non-gaussian state space models', *Computational Statistics & Data Analysis* **52**(6), 2911–2930.

Subrahmanyam, A., Huh, S. and Brennan, M. (2013), 'The pricing of good and bad private information', *Available at SSRN 2199723* .

Tauchen, G. E. and Pitts, M. (1983), 'The price variability-volume relationship on speculative markets', *Econometrica: Journal of the Econometric Society* pp. 485–505.

Tay, A., Ting, C., Tse, Y. K. and Warachka, M. (2009), 'Using high-frequency transaction data to estimate the probability of informed trading', *Journal of Financial Econometrics* **7**(3), 288–311.

Vega, C. (2006), 'Stock price reaction to public and private information', *Journal of Financial Economics* **82**(1), 103–133.

Venkatesh, P. and Chiang, R. (1986), 'Information asymmetry and the dealer's bid-ask spread: A case study of earnings and dividend announcements', *The Journal of Finance* **41**(5), 1089–1102.

Wang, G. H. and Yau, J. (2000), 'Trading volume, bid–ask spread, and price volatility in futures markets', *Journal of Futures Markets* **20**(10), 943–970.

Wang, J. (1993), 'A model of intertemporal asset prices under asymmetric information', *The Review of Economic Studies* **60**(2), 249–282.

Watanabe, T. (2000), 'Bayesian analysis of dynamic bivariate mixture models: Can they explain the behavior of returns and trading volume?', *Journal of Business & Economic Statistics* **18**(2), 199–210.

Wei, W. C., Gerace, D. and Frino, A. (2013), 'Informed trading, flow toxicity and the impact on intraday trading factors', *Australasian Accounting, Business and Finance Journal* **7**(2), 3–24.

Yan, Y. and Zhang, S. (2012), 'An improved estimation method and empirical properties of the probability of informed trading', *Journal of Banking & Finance* **36**(2), 454–467.

Yu, Y. and Meng, X.-L. (2011), 'To center or not to center: That is not the question—an ancillarity–sufficiency interweaving strategy (asis) for boosting mcmc efficiency', *Journal of Computational and Graphical Statistics* **20**(3), 531–570.