Statistical Analysis on Market Microstructure Models

by
Feng Liu

A Dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operation Research.

Chapel Hill
2010

Approved by:

Chuanshu Ji, Advisor

Günter Strobl, Committee Member

Douglas Kelly, Committee Member

Nilay Argon, Committee Member

Shankar Bhamidi, Committee Member

Abstract

FENG LIU: Statistical Analysis on Market Microstructure Models
(Under the direction of Chuanshu Ji.)

The field of market microstructure studies the trading mechanisms and costs of providing transaction services, with their impact on the short run behavior of security prices. Investors are involved in the market for securities and related information. The cost of a trade depends on the asymmetric information possessed by different participants in the trade. In this thesis, we perform empirical studies of stock microstructure data and infer on several market microstructure models. The generalized Roll model (1984) and Hasbrouck's (2009) approach, although considered as a good starting point, are too simple to be realistic. Kyle (1985), Glosten and Milgrom (1985) represent two most important models that involve asymmetric information and transaction costs. We derive a new characterization of Kyle's equilibrium model and develop new algorithms to solve recursive equations with computational efficiency. We also propose an extension of Kyle's model in which the private information is gradually revealed into the security price. Bayesian inference on model unknowns enables us to discover the trading strategies of the informed traders. These approaches facilitate real applications of market microstructure models.

Acknowledgments

I am largely indebted to Prof. Chuanshu Ji, who has been such a great advisor and mentor to me. His ingenious ideas and deep insight have constantly inspired me and guided me through the past five years. And his knowledge of statistics and financial mathematics has been a great source of enlightenment to me. I am especially gratefully for his great effort on my research topics and process.

My sincere gratitude also goes to other members of my committee. Prof. Günter Strobl from finance has been a great mentor to me and provides enormous support of my research. He has not only instructed me on microstructure topics but also interpreted the intricacy of microstructure issues with patience especially as I am coming from statistics major. His profound knowledge and valuable advices enlighten me throughout my research. Prof. Douglas Kelly, Prof. Nilay Argon and Prof. Shankar Bhamidi have read my dissertation carefully and provided numerous useful comments and feedbacks. I also learned a lot from the interesting courses they taught.

I cannot show my appreciation enough to my wife. She is always listening to me when I am frustrated, helping me out with her love at each and every moment. I also owe my deepest gratitude to my parents, who always love me, believe in me and support me no matter what. I hope that I will make them proud by my fulfilment.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The need for studies in financial economics becomes more urgent than ever after we have experienced the recent downturns in financial markets. Statistics play an increasing role in such studies. In the literature of mathematical finance and related statistical analysis, most of the elegant results hold under the assumption of a "perfect market" or "fully efficient market", i.e. no transaction costs, no bid/ask spread, same information shared by all market participants, no tax, no limit for short selling, etc. Conceivably that is far from what happens in real financial markets. The area of market microstructure studies what factors and mechanisms affect an asset price, how informed traders and uninformed ones differ in obtaining information and using it to optimize their trading strategies, etc. Research in this area ultimately will enhance our understanding of the real markets and have more practical impacts on various issues. That is what we plan to study.

Microstructure theory focuses on how specific trading mechanisms affect the price formation process. In a trading market, financial assets are not transformed but transferred from one investor to another. The field of market microstructure studies the costs for trading securities and the impact of trading costs on the short-run behavior of security prices. Costs are reflected in bid-ask spreads. Typical market participants

are the investors who demand or supply the ultimate immediacy and the dealers who facilitate the trading. An investor usually wishes to trade immediately and to buy low and sell high. In reality however, traders actually buy at an (higher) ask price and sell at a (lower) bid price. Those bid/ask prices are quoted by dealers (market makers, limit order traders), and the spread "ask price minus bid price" is the compensation that dealers receive for offering immediacy.

The literature on asset pricing often assumes that markets operate without costs and frictions whereas the essence of the market microstructure research is to analyze the impact of trading costs and various friction factors. The investors are generally involved in the market for securities and related information. The market for securities deals with the determinants of security prices such as earning per share etc. The market for information deals with the supply and demand of information. It incorporates the incentives of security analysis and related information transfer. The asymmetric information is closely related to transaction services since the cost of a trade depends on the information possessed by the participants in the trade.

The Efficient Market Hypothesis (EMH) asserts that financial markets are "informationally efficient", or that prices on traded assets already reflect all known information, and instantly change to reveal new information. Therefore, according to the theory, it is impossible to consistently outperform the market by using any information that the market already knows. Investors and researchers have disputed the EMH both empirically and theoretically. Dreman and Berry (1992) shows low P/E stocks have greater returns. Behavioral economists Daniel Kahneman, Amos Tversky, and Richard Thaler attribute the imperfections in financial markets to a combination of biases such as overconfidence, overreaction, information bias in reasoning and information processing. These errors in reasoning lead investors to avoid high-value stocks and buy growth stocks at expensive prices.

Market microstructure studies market friction and asymmetric information impact on security prices. When we look at the security price dynamics with respect to market microstructure, our focus has shifted from monthly or daily to minute or tick level with more features such as bid price, ask price, bid size, ask size, trade price and trade volume etc. The additional features of price and trading dynamics reflect the complexity of microstructure.

Let $\mathcal{F}_t$ be the *information set* available to the market at time t, the payoff of a security be a random variable, denote by v. Then the conditional expectation $p_t = E(v|\mathcal{F}_t)$ is referred to as the fundamental value or the *efficient price* of the security. The information set is the starting point for many microstructure models. One of the basic goals of microstructure analysis is a detailed study of how informational efficiency arises, and the process by which new information comes into play or is reflected in the price movement. In microstructure analysis, transaction prices are usually not martingales. By imposing economic or statistical structure, it is often possible to identify a martingale component of the price with respect to a particular information set.

Roll (1984) suggests a model of high frequency trade prices which incorporate trading dynamics. This model is fundamental to market microstructure models such that it illustrates the distinction between price movement due to fundamental security value and those attribute to the market organization and trading mechanism. To estimate the effective trading cost and returns formed from daily data, Hasbrouck (2009) advocates a Bayesian approach based on a generalized Roll model. This method accommodates a long time span by daily data, and the cost estimate is validated against microstructure data. Although the Roll model is too simple to capture many realistic features in the market, it articulates an important aspect of the bid-ask effect on trading price, and serves as a good starting point.

The public information set consists of some common knowledge concerning a proxy of probability structure of the economy, i.e. various possible scenarios of a terminal security value and associated types of agents. Most models make no provision for the updates of non-trade public information (e.g., news release). As trading unfolds, the most important updates to the public information set are market data, such as bid, ask, closing prices and volumes of trades. Private information may consist of signals about security value, i.e. more detailed knowledge of the terminal security value.

When all agents are *ex ante* identical, they are said to be *symmetric*. This does not exclude private values or private information. The symmetry means that all individual-specific variables (e.g., the coefficient of risk aversion, the signal) are identically distributed across all agents. The Roll model is still informational symmetric. In an *asymmetric* information model, a subset of agents has superior private information which leads to a trading advantage.

The majority of asymmetric information models in market microstructure examines market dynamics subject to a single source of uncertainty. At the end of the trading, the security payoff is known and realized. Thus, the trading process is an adjustment from one well-defined information set to another set. The dynamics are neither stationary nor ergodic, although path realizations could be stacked to disclose trading behaviors. There are two main sorts of asymmetric models, among others:

(a) **Sequential trade model**

Randomly selected traders arrive at the market one by one, *sequentially*, and *independently*. Reference include Copeland and Galai (1983), and Glosten and Milgrom (1985). In Glosten-Milgrom model, orders arrive and are executed by a market maker individually. The arrival rates of *informed* and *uninformed* traders are determined exogenously. Informed traders trade when chosen by this mechanism as if they have no future opportunities to trade. In other words, when trade

is profitable, they trade as much as possible.

(b) **Strategic trade model**

A seminal strategic model is studied in Kyle (1985). The Kyle model is a model of a batch-auction market, in which market makers see the order imbalance at each auction date. And market makers compete to fill the order imbalance, and matching orders are executed at market clearing prices. Unlike the sequential trade model, the strategic informed agent could trade at multiple times. Kyle develops the optimal trading behavior for the informed trader and shows that the agent will trade on his information only gradually, rather than exploit it to the maximum extent possible.

The essence of both models is that a trade from the informed trader will reveal his/her private information assuming traders are all rational. The "buy" order originates from a trader who has positive private information, but not from those who possess negative information (here we rule out "bluffing", i.e. the informed trader is "bluffing" if he knowingly sells an undervalued or buys an overvalued asset). And the competitive market makers will set their bid-ask quotes accordingly. In consequence, greater information asymmetry would lead to wider spreads in quotes. The spread and trade impact are major empirical implications of these models.

There is an extensive literature in market microstructure. Besides many research and survey papers, here we mention several good books: Brunnermeier (2001), de Jong, F. and Rindi, B. (2009), Harris, L. (2003), O'Hara, M. (1995), and Vives, X. (2008).

Our focus in this thesis is model-based empirical studies. Research in empirical market microstructure has two important aspects. On the economic side, certain trading mechanisms and market frictions, such as transaction costs and asymmetric information, are incorporated in a proposed microstructural model with a utility function. The

resulting constrained optimization problem is tackled, and its equilibrium solution will yield an optimal trading strategy from the perspective of each market participant, and a risk-neutral clearing price for every traded asset. In particular, the solution often enables us to interpret the economic impacts of certain parameters contained in the model. On the statistical side, inference on model parameters is performed based on real market data, usually represented by time series of asset prices and returns, trading volumes, orders and quotes, etc. whether they are daily or involving intra-day activities. More often than not, goodness-of-fit of the proposed model need not be satisfactory. Naturally, more sophisticated models can be considered. However, a purely statistical approach based on goodness-of-fit may not address the issue of economic interpretation. Financial economists always pay greater attentions to what we can learn from a model. To reconcile the economic and statistical aspects, a natural "spiral up" development can start with a basic economic model, fit it by market data; With observed deviations between the data and the proposed model, we proceed to modify the model and try to derive the corresponding equilibrium solution, then check it with data again, etc. The hope is to improve the goodness-of-fit and enhance the understanding of market behaviors in each new iteration along such a research path.

This approach will be illustrated in a framework of the celebrated strategic trade model in Kyle (1985). We aim at fitting several extensions of Kyle's models using intra-day data, and retaining its interpretability. Dynamic Bayesian modeling [see West and Harrison (1997)] appears to fit our need, because in many problems sequential updating between observed data and various unknowns follow a natural path. The unknowns include model parameters (e.g. market depth and noise trading volatility) and latent variables (e.g. an inside trader's order).

Our contributions in this thesis include a new characterization of Kyle's equilibrium

solutions. The algorithm we provide offers a computationally more efficient way to characterize the equilibrium solutions, which also enables us to develop similar equilibrium solutions in certain extensions to Kyle's model, such as the one with noisy signals of the asset value observed by the informed trader. We also propose an extended model to Kyle's in which the (reciprocal of) market depth $\{\lambda_n\}$ and the informed trading intensity $\{\beta_n\}$ form time series. A Bayesian inference procedure based on real intra-day market data (at Wharton's Database Services) is conducted for this dynamic model.

This dissertation is organized as follows. Several well-known market microstructure models are summarized in Chapter 2, including the generalized Roll model, sequential trade and strategic trade models. Chapter 3 describes our pre-modeling data analysis and key findings which pave the way for our research. In Chapter 4, we present related research on market liquidity. Chapter 5 outlines the new characterization of Kyle's model and inference on an extended model. We propose our dynamic Bayesian model in Chapter 6 and carry out simulation studies for the Bayesian inference. The empirical studies based on intra-day microstructure data are performed in Chapter 7. In Chapter 8, we discuss other research topics we have conducted.

Chapter 2

Basic Formulation and Major

Topics

2.1 Overview

The theoretical aspect of major microstructure research has concentrated on a number of important models. These models provide useful tools for microstructure research and results from the models are often directly referenced. Every model contains a number of parameters. A lot of efforts are devoted to interpret how those parameters impact market behaviors and investment decisions made by market participants. In contrast, systematic empirical studies are falling behind, i.e. little has been done in statistical inference on those proposed models and for model validation. Such limitation would hinder further development in applications of those models.

2.2 Market dynamics

2.2.1 Price movement and information set

When we look at the security price dynamics with respect to microstructure, our focus has shifted from monthly or daily to minute or tick level with more features at fine granularity. Such features include bid price, ask price, bid size, ask size, trade price and trade volume etc. The following figure illustrates ticker CSCO (Cisco System) traded on Jan, 3 2002 at second level, data source from TAQ (Trade and Quote) database. The trade price is augmented by bid/ask price quotes.
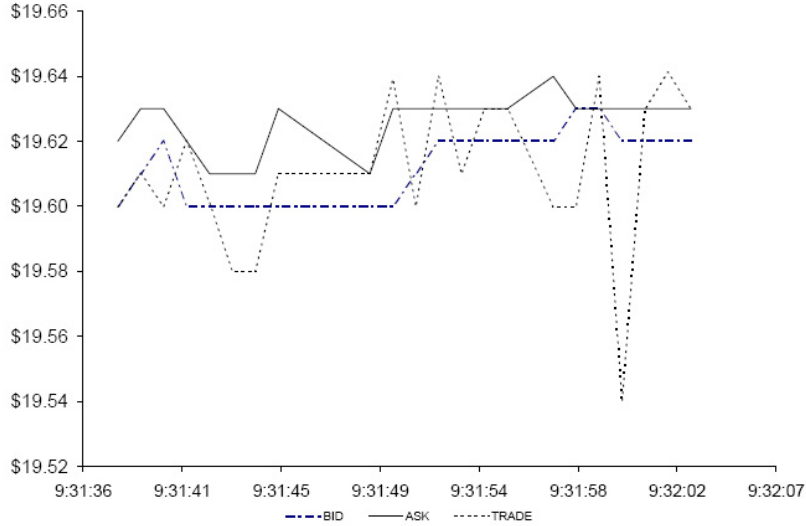
Figure 2.1: Sample Cisco trade price, bid, ask quotes

The three prices (bid, ask, and trade) differ. The ask (solid) is always higher than the bid (dot-dashed), and trades (dashed) usually occurs at posted bid and ask prices, but not always. They converge to each other. Those features reflect the complexity of market microstructure.

For simple illustration, consider the evolution of fundamental security value described by a random walk. Let p_t denote the transaction price at time t, where t index

regular point of time. The random-walk model with drift is: For $t = 1, 2, ...,$

$$p_t = p_{t-1} + \mu + u_t, \tag{2.1}$$

where u_t are iid $N(0, \sigma^2)$ random variables, and μ is the expected price change (the drift). In microstructure data samples the mean of μ is often small relative to its estimation variance, i.e, $E(\mu) < Var(\mu)$. It is often preferable to drop the mean return from the model in most microstructure analysis.

When $\mu = 0$, $E(p_t|p_{t-1}, p_{t-2}, ..) = p_{t-1}$, where $E(|p_t|) < \infty$ for all t. The p_t process follows a martingale. A more general definition involves conditioning on information sets.

Let $\mathcal{F}_t$ be the *information set* available to the market at the time t, the payoff of a security is a random variable, denote as v. Then the conditional expectation $p_t = E(v|\mathcal{F}_t)$, for all t is a martingale with respect to sequence of information sets $\mathcal{F}_t$. When the conditional information is all public information, the conditional expectation is referred to as the fundamental value of the security.

In microstructure analysis, transaction prices are usually not martingales. By imposing economic or statistical structure, it is often possible to identify a martingale component of the price with respect to a particular information set. In the random-walk equation (2.1), u_t are iid, the price process are time-homogeneous, that is, it exhibits same behavior whenever we sample it. This is plausible only if the economic environment underlying the security is also time-homogeneous. Securities like stocks often violate this condition and are not suitable to be approximated by a random walk. On the other hand, they usually have well-defined boundary conditions at maturity that affect their values during the process: securities have different valuation far from maturity compared to close to maturity. In market microstructure, the short-term behavior may still be approximated by a random-walk model, but this model is not a

valid choice to describe market behavior in the long-run.

2.2.2 Issues and interpretation

In equation (2.1), price change is $\Delta p_t = p_t - p_{t-1}$, which is iid with mean 0, variance $Var(u_t) = \sigma^2$, and μ set to 0. When we analyze the actual data samples, the short-run security price changes always exhibits extreme dispersion and auto-correlation between successive observations.

For financial security data samples, the price changes at time horizon often have sample distributions with fat tails. The standard assumption that price changes are normally distributed is violated. For a random variable X, the population moment of order α is defined as EX^α. If EX^α is finite, as $x \to \infty$, then the corresponding sample estimate $\sum X_t^\alpha / T$, T is the sample size, is the consistent estimate of of EX^α. To get a consistent estimate of the standard error of mean, we require a consistent estimate of the variance. Not all moments are finite if the normal assumption is violated. Recent studies suggest that finite moments for daily equity returns exist only up to order 3, and the trading volume only up to order 1.5 (Gabaix 2003). These findings impose substantial restrictions on the sort of microstructure models we could estimate. The existence of extreme values in finite samples may lead to many practical consequence. Increasing the sample size may not increase the precision as fast as we expected, and estimated parameters are very sensitive to model specification.

The price increments Δp_t in the random walk are iid and uncorrelated. But data samples show the first-order autocorrelations of price changes are usually negative and non-zero. For time series Δp_t, the autocovariance and autocorrelation is defined as $\gamma_k = Cov(\Delta p_t, \Delta p_{t-k})$ and $\rho_k = Corr(\Delta p_t, \Delta p_{t-k})$. When the mean is zero, γ_k could be estimated as the sample average $\hat{\gamma}_k = T^{-1} \sum_{t=1}^{T} \Delta p_t \Delta p_{t-k}$, and the autocorrelations as $\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0$.

We collect the data from WRDS TAQ database. The data samples we studied are MSFT (microsoft Inc) trade prices from Jan, 2 to Jan, 5 2002. There are 200,000+ trades in MSFT at tick level. The estimated first-order autocorrelation of price increments is $\hat{\rho}_1 = -0.4561$, with standard error of 0.004. The p-value of significance test is less than 10^{-5} which rejects zero autocorrelation hypothesis.

We would expect to find $\hat{\rho}_k = 0$ for $k = 1$ for a random walk model. But the empirical study verifies the contrary. The economic explanation about this contradiction motivates the Roll model, which explains autocorrelations of price increments by meaningful economic and statistic implication.

2.3 Generalized Roll model

2.3.1 Roll model

Roll (1984) suggests a model of high frequency trade prices which incorporate market dynamics. This model is fundamental to many market microstructure models such that it illustrates the distinction between price movement due to fundamental security value and those attribute to market organization and trading mechanism. The former arises from the earning capability and future cash flows of the underlying security, whereas the later are transient due to market behavior. The model provides meaningful economic interpretation, and in some cases, explains the market movement well.

For $t = 1, 2, ...,$

$$p_t = m_t + c\, q_t, \tag{2.2}$$

$$m_t = m_{t-1} + u_t, \tag{2.3}$$

which consists of an observation equation (2.2) and a state evolution equation (2.3),

where m_t denote the martingale efficient price at tth trade, p_t is the trade price. The q_t are direction indictors, which take on the value 1 (buy) or -1 (sell) with equal probability, the shocks $u_1, u_2, ...$ are iid $N(0, \sigma^2)$ random variables, the parameters $c > 0$ and $\sigma > 0$ represent the effective cost and the volatility respectively. The two sequences $\{q_t\}$ and $\{u_t\}$ are assumed to be independent. Note that only $\{p_t\}$ are observed, while $\{m_t\}$ and $\{q_t\}$ are treated as latent variables.

The model implies

$$\Delta p_t = c\Delta q_t + u_t, \tag{2.4}$$

from which it follows that $c = [-cov(\Delta p_t, \Delta p_{t-1})]^{1/2}$, if $cov(\Delta p_t, \Delta p_{t-1}) < 0$, and $c = 0$, otherwise. The first-order autocovariance is non-zero. Δp_t exhibits volatility and *negative* serial correlation as the result of effective cost. The intuition is: If m_t is fixed so that prices take on only two values, the bid and the ask, and if the current price is the ask, then the price change between the current price and the previous price must be either 0 or $-2c$, and the price change between the next price and the current price must be either 0 or $2c$. The moment estimate is feasible, however, only if the first-order sample autocovariance of the price change is negative.

If the dealers compete to the point where the costs are just covered, the bid and the ask are $m_t - c$ and $m_t + c$, with the spread 2c, a constant. We collect the data of 200,000 trades for MSFT on Jan, 2 to Jan, 5 2002 from TAQ, the first-order autocovariance is $\hat{\gamma}_1 = -0.00522$. This implies $c = \$0.035$, and bid-ask spread of $2c = \$0.070$; while the estimates from TAQ database shows the bid-ask spread is 0.0625 which is close.

2.3.2 Hasbrouck's approach

To estimate the effective trading cost and returns formed from daily data, Hasbrouck advocates a Bayesian approach based on the Roll model. This method accommodates a long time span by daily data, and the cost estimate is validated against microstructure data.

The unknowns comprise both the model parameters $\{c, \sigma^2\}$ and the latent data $\{q_t\}$. We could get posterior distribution $f(c|\sigma^2, p_1, p_2, ..., p_T)$ and $f(\sigma^2|c, p_1, p_2, ...p_T)$ via multivariate Bayesian methods. However, the posterior joint density $f(c, \sigma^2|p_1, p_2, ..., p_T)$ is not obtained in a closed-form. This motivates the Gibbs sampler. The Gibbs sampler constructs full posterior densities by iteratively simulating from full conditional distributions for c, σ and q_t.

The trading cost estimates from US stocks are formed from daily CRSP data. The CRSP/Gibbs estimates are very close to TAQ estimates (with correlation 0.96), which shows that the daily Gibbs estimates have strong validity. The estimation procedure tries to resolve the two components among the sample price path: the permanent innovations (due to the efficient price), and the transient effective cost (due to bid-ask effect). When $c >> \sigma$, the bid/ask bounce generates reversals that are easy to pick out which leads to clear resolution of the two components. When c is relative small, the bid/ask effect is swamped by innovations in the efficient price.

Empirical sample results:

- Ticker symbol NEWE (Jan 1990) $bid = 3.625$, $ask = 4.125$, $c \approx 0.25$, daily volatility $\sigma = 0.031$ clear resolution

- Ticker symbol MSFT (Jan 2005) $bid = 26.67$, $ask = 26.68$, $c \approx 0.005$, $\sigma = 0.073$ poor resolution

Although the Roll model captures important aspect of bid-ask effect on trading

price, it nevertheless lacks completeness in terms of determinants. For expected returns, it shows weak evidence of trading cost as a characteristic, and it shows no evidence that the trading cost variation is a risk factor. In fact, Glosten and Milgrom (1985) argues that c is determined endogenously and is unlikely to be independent of $\{m_t\}$, the permanent component.

Most microstructure models including Roll model are dynamic over time and have latent variables. Dynamic latent variable models can be formulated in state-space form and estimated by maximum likelihood. For Gaussian cases, it could be estimated using multivariate linear regression; For non-Gaussian latent variables(e.g., the buy/sell indicator), the estimation procedure often involves nonlinear smoothing or Bayesian MCMC methods. Hasbrouck's work sheds light on Bayesian type of analysis.

2.4   Sequential trade model

We begin with Glosten-Milgrom model. Consider one security valued at $V \in \{V_h, V_l\}$, with $\Pr(V_l) = \delta$. The value is revealed at the end of trade. There are two types of traders: the informed I and the uninformed U, the proportion of informed traders among the population is μ. The market maker posts bid and ask quotes, B and A. A trader is randomly drawn from the population. If the trader is informed, he buys if $V = V_h$, sells if $V = V_l$. If the trader is uninformed, she buys or sells randomly with equal probability. The market maker does not know the types of the trader. A buy is a purchase by the trader at the dealer's ask price, A; a sell is a trading at the bid, B. We assume that the competition among dealers drives the expected profit to zero. The market maker's inference given that the first trade is a buy or a sell can be summarized by his posterior belief about the low outcome.

Let $p_k(buy)$, (or $p_k(sell)$) $k = 1, 2...$, denote the probability of a low outcome given the kth trade is a buy (or a sell). p_0 is the unconditional probability being a low

outcome, which is δ. Let B_k denote kth order is buy, S_k denote kth order is sell. Then the market maker's posterior belief of a low outcome after the first trade is buy is,

$$p_1(buy) = \Pr(V_l|B_1) = \frac{\Pr(V_l, buy)}{\Pr(buy)} = \frac{\delta(1-\mu)}{1+\mu(1-2\delta)}, \tag{2.5}$$

and dealer's expectation of the value given first buy order is $E(V|B_1) = \Pr(V_l|buy)V_l + (1 - \Pr(V_l|buy))V_h$. If competition drives the expected profit to zero, then the posted "ask price" is the dealer's expected value.

$$A = E(V|B_1) = \frac{\delta(1-\mu)V_l + (1-\delta)(1+\mu)V_h}{1 - (1-2\delta)\mu}, \tag{2.6}$$

The bid price is similar, followed by a sell to the dealer. The dealer saw the first trader is a sell order and post the bid price.

$$p_1(sell) = \Pr(V_l|S_1) = \frac{\Pr(V_l, sell)}{\Pr(sell)} = \frac{\delta(1+\mu)}{1 - \mu(1-2\delta)}, \tag{2.7}$$

$$B = E(V|S_1) = \frac{\delta(1+\mu)V_l + (1-\delta)(1-\mu)V_h}{1 + (1-2\delta)\mu}, \tag{2.8}$$

The bid-ask spread is:

$$S = A - B = \frac{4(1-\delta)\delta(V_h - V_l)\mu}{1 - (1-2\delta)^2\mu^2}, \tag{2.9}$$

The dealer updates his belief and posts new quotes on each trades sequentially. This process repeats for k=1,2,... This updating procedure could be expressed in general forms since all probabilities in the event tree are constant except $p_k(.)$.

$$p_k(buy|p_{k-1}(.)) = \frac{p_{k-1}(.)\,(1-\mu)}{1 + \mu(1 - 2p_{k-1}(.))}, \tag{2.10}$$

$$p_k(sell|p_{k-1}(.)) = \frac{p_{k-1}(.)\,(1+\mu)}{1 - \mu(1 - 2p_{k-1}(.))}, \tag{2.11}$$

It can be shown that $p_k(buy|p_{k-1}(buy), p_{k-2}(sell)) = p_k(buy|p_{k-1}(sell), p_{k-2}(buy))$, for all k. The arrival sequence of the buy or sell orders does not matter. Therefore the proportion of buy or sell orders is deterministic to the outcome.

The conditional expectation of the ask can be decomposed as

$$A = E(V|buy) = E(V|U, buy) \Pr(U|buy) + E(V|I, buy) \Pr(I|buy), \qquad (2.12)$$

rearranging terms gives

$$(A - E(V|U, buy)) \Pr(U|buy) = -(A - E(V|I, buy)) \Pr(I|buy), \qquad (2.13)$$

In this model, the economic interpretation for equation (2.13) is that the gain from an uninformed trader on the left side is equal to the loss to the informed trader on the right side (subject to zero profit expectation for the market maker). There is net wealth transfer from the uninformed to the informed.

Although the trader is independently drawn from both population for order execution, one subset of the population (the informed) always trade in the same direction. The result is that orders are serially correlated. We will do empirical study on this topic in the next chapter.

One important economic justification of G-M model is trades update the price. For any security at kth given trade, a buy order on the $(k+1)th$ trade will make a upward revision in the conditional probability of a high outcome, and consequently increase both ask and bid quotes and drive trading price upward. In contrast, a sell order will drive price downward. The trade price impact is a particular useful empirical implication.

In the Roll model, we denote $\{q_t\}$ as the trade direction variable ($+1$ buy, -1 sell) with equal probability. In the G-M model, the order flow has no equal probability

attributes to asymmetric information processed by difference traders, the informed traders always trade in the direction of his knowledge.

The asymmetric information in the G-M model is μ, the proportion of the informed trader in the population. In equation (2.11) and (2.9), the asymmetric information parameter μ is positively related to $p_k(sell)$, and the bid-asked spread. The justification behind is when the market have more informed traders, a sell order will be more likely submitted by an informed trader instead of a uninformed, the probability of a low outcome after sell is high; similarly, the probability of a high outcome given buy order is also high. In consequence, the dealer will post wider bid-ask spread in response to the change of posterior beliefs. These results suggest use of the bid-ask spread or the impact of an order has on subsequent prices as proxies for the asymmetric information. We have more discussions in the empirical study.

The limitation of G-M model is the informed traders are drawn randomly by the market mechanism. When she is selected, she will trade once and the maximum (one unit of order). There are no trading strategies for the informed trader to maximize her profit. The order execution timing and order sizes are two important aspects to the informed in empirical work while remain unaddressed in G-M model.

2.5   Strategic trade model

We follow the basic framework in Kyle (1985) with modified notation.

- Fix an asset in what follows. Suppose N auctions take place sequentially over a trading period (e.g. day, month, year). For each $n = 0, 1, \ldots, N$, t_n denotes the time for the nth auction, with $0 = t_0 < t_1 < \cdots < t_N = 1$.

- There is a single informed trader who knows the liquidation value V of the asset, and let X_n denote the aggregate position of the insider trader after the nth auction,

18

so that $\Delta X_n = X_n - X_{n-1}$ denotes the quantity traded by the insider at the nth auction. However, $V \sim N(p_0, \Sigma_0)$ is considered as a random variable by a set of noise traders.

- The quantity traded by noise traders at the nth auction is denoted by $\Delta U_n \sim N(0, \sigma^2 \Delta t_n)$ with $\Delta t_n = t_n - t_{n-1}$. Assume $U_1, ..., U_N$ are independent, and V is independent of $\{U_1, ..., U_N\}$.

- Let p_n be the asset's market clearing price at the nth auction, $\Delta Y_n = \Delta X_n + \Delta U_n$ denote the total orders at the nth auction. The information set $\mathcal{F}_n^U$ available to uninformed traders (including a market maker and all noise traders) at t_n consists of the observations $\{p_1, ..., p_n; \Delta Y_1, ..., \Delta Y_n\}$.

- The informed trader (insider) has a richer information set available to him before making his move at the nth auction. Such a set $\mathcal{F}_{n-1}^I$ includes $\{X_1, ..., X_{n-1}; V\}$ in addition to $\mathcal{F}_{n-1}^U$. The insider chooses ΔX_n based on $\mathcal{F}_{n-1}^I$.

- After the move made by both insider and noise traders at the nth auction, the market maker determines the price p_n based on $\mathcal{F}_{n-1}^U$ and ΔY_n.

- Let

$$\pi_n = \sum_{i=n}^{N} (V - p_i) \, \Delta X_i \tag{2.14}$$

be the total profits of the insider to be made at auctions $n, n + 1, ..., N$, and $X = (X_1, ..., X_N)$, $P = (p_1, ..., p_N)$ denote the insider's trading strategy and the market maker's pricing rule respectively. Hence $\pi_n = \pi_n(X, P)$.

Definition 1. *A sequential auction equilibrium is defined as a pair (X, P) such that the following conditions hold:*

19

(*C1*) (*profit maximization*) *For $n = 1, ..., N$ and all $X^{'} = (X^{'}_1, ..., X^{'}_N)$ with $X^{'}_i = X_i$,*

  $i = 1, ..., n - 1$, *we have*

$$E[\pi_n(X, P)|\mathcal{F}^I_{n-1}] \geq E[\pi_n(X^{'}, P)|\mathcal{F}^I_{n-1}]. \tag{2.15}$$

(*C2*) (*market efficiency*) *For $n = 1, ..., N$ we have*

$$p_n = E(V|\mathcal{F}^U_{n-1}, \Delta Y_n). \tag{2.16}$$

Definition 2. *A sequential auction equilibrium (X, P) is called a linear equilibrium if the component functions of X and P are linear, and a recursive linear equilibrium in which there exist parameters $\lambda_1, ..., \lambda_N$ such that*

$$p_n = p_{n-1} + \lambda_n \; \Delta Y_n, \quad n = 1, ..., N. \tag{2.17}$$

The following theorem is the major result in Kyle (1985) which proves the existence and uniqueness of linear equilibrium, and characterizes those modeling parameters in it.

Theorem 1. *There exists a unique linear equilibrium (X, P), represented as a recursive linear equilibrium, characterized by (for $n = 1, ..., N$)*

$$\Delta X_n \;\; = \;\; \beta_n \; (V - p_{n-1}) \; \Delta t_n, \tag{2.18}$$

$$p_n \;\; = \;\; p_{n-1} + \lambda_n \Delta Y_n, \tag{2.19}$$

$$\Sigma_n \;\; = \;\; Var(V|\mathcal{F}^U_n), \tag{2.20}$$

$$E[\pi_n|\mathcal{F}^I_{n-1}] \;\; = \;\; \alpha_{n-1} \; (V - p_{n-1})^2 + \delta_{n-1}; \tag{2.21}$$

20

Given Σ_0, the parameters $\beta_n, \lambda_n, \Sigma_n, \alpha_n, \delta_n$ are the unique solutions to equations

$$\alpha_{n-1} = [4\lambda_n(1 - \alpha_n\lambda_n)]^{-1}, \tag{2.22}$$

$$\delta_{n-1} = \delta_n + \alpha_n \lambda_n^2 \sigma^2 \Delta t_n, \tag{2.23}$$

$$\beta_n \Delta t_n = (1 - 2\alpha_n\lambda_n) [2\lambda_n(1 - \alpha_n\lambda_n)]^{-1}, \tag{2.24}$$

$$\lambda_n = \beta_n \Sigma_n \sigma^{-2}, \tag{2.25}$$

$$\Sigma_n = (1 - \beta_n\lambda_n \Delta t_n) \Sigma_{n-1}, \tag{2.26}$$

subject to $\alpha_N = \delta_N = 0$ and $\lambda_n(1 - \alpha_n\lambda_n) > 0$.

In Kyle's model, nobody knows the market clearing price when they submit their orders. Because the liquidity trader order flow is exogenous, there are really only two players: the informed and the market maker. The informed trader wants to trade aggressively on her private information, i.e., buy a large quantity if her information is positive. But the market maker knows that if he sells into a large net buy customer, he himself is more likely to be on the wrong side of the trade. He protects himself by setting a price that is increasing in the *net order flow*. This acts as a brake on the informed trader's desires: if she wishes to buy a lot, she will have to pay a high price. This is the economic implication of Kyle's solution to the model.

Chapter 3

Data Structure and Empirical Study

We conduct our empirical studies on market microstructure models discussed previously. The results presented in this chapter motivate us both theoretical and empirical implications of those models.

3.1 Trade volume or order imbalance
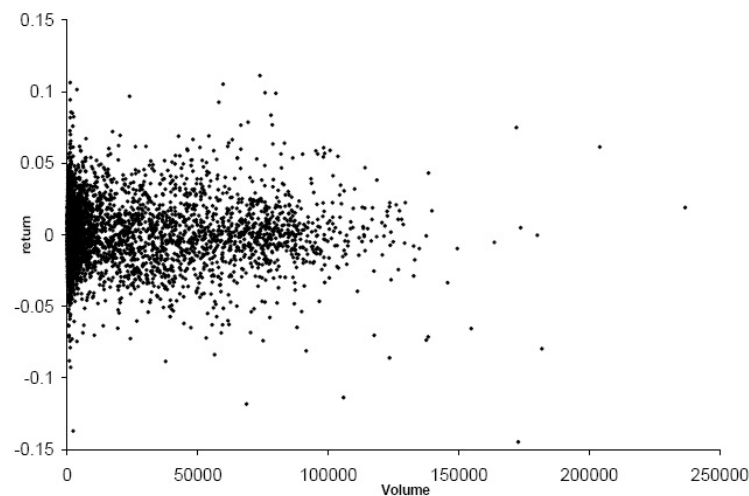
Figure 3.1: Relationship between trade volume and return

The *trade volume* is the total number of trade orders or trade size at specific time

frame, e.g. daily. The basic sequential trade model has one trade quantity in each trade. Trades in the real markets, of course, occur in varying quantities. The trade volume is an important market dynamics. To get a preliminary impression about trading volume and stock return, we obtain 10 randomly chosen firms, data from Jan 1988 to Dec 2004 from CRSP database, and plot the cross-sectional daily stock return over daily trade volume in Fig. 3.1. The summary statistics is shown in table 3.1.

Table 3.1: Summary statistics of daily returns vs. daily volume

Variables	Sample period	Observations	Mean	SD
Return	01/1988 - 12/2004	32890	0.00126	0.022964
Volume	01/1988 - 12/2004	32890	22173.54	30640.38

Variables	Max	Min	Skewness	Kurtosis
Return	0.195652	-0.15598	0.06162	3.839
Volume	236675	81	1.63798	2.60493

From Fig. 3.1, we know that volume are quite symmetric across zero return and high volume does not tend to be associated with high return. In Kyle's strategic trading model, the author conjectures a relationship between a firm's stock price change and its order flow. In Pasquariello and Vega (2009) empirical study, they address cross-trading effect with daily aggregated order imbalance. Chordia and Subrahmanyam (2004) show that the total number of transactions has greater explanatory power for stock-return fluctuation than trading volume. We will take similar approach with modified setting. The intuition is that total trade volumes can be decomposed into sell orders and buy orders, it is the order imbalance between sell and buy orders which drive the market movement.

We use intraday, transaction-level data from trade and quotes (TAQ) database during regular market hours (9:30am to 4:00pm). Corresponding daily price data comes from CRSP. We obtain MSFT (Microsoft) transaction level trading data on January

2001 as a starting point. First, we filter the TAQ data by deleting small number of traders and quotes representing possible data error (e.g., negative prices or quotes). We then assign the trades using the following procedure.

1. If a transaction occurs above (or below) the prevailing mid-point of bid-ask spread at that particular time, we assign buy (or sell) sign to that transaction.

2. If the transaction price is at the mid-point of the spread, we will label it a buy (or sell) if the sign of the last trade price change is positive (or negative).

We define the trade direction variable as +1 (buy) or -1 (sell) for each transaction, similar to the Roll model. Then we get the signed order flow by multiple trade direction and order size, denote as $\hat{\Delta y_t}$, where $\Delta y_t = \Delta x_t + \Delta u_t$ in Kyle's setting.

We denote *order imbalance* as the total number of signed order flows at given time period, e.g. daily. We would expect the signed order flows or order imbalance have greater explanatory power.

3.2 Trade direction

We use $\{q_t\}$, +1 (buy) or -1 (sell), $t = 1, 2...$ to denote intraday trade directions as we did in the previous chapter. In the Roll model, q_t has equal probability, which implies $E(q_t|\mathcal{F}_{t-1})$ to get zero in the empirical study. Each trading date has one series of high frequency trade directions. We got intraday estimates of the first-order autocorrelation of this high frequency series $\hat{\rho_k} = Corr(\hat{q_t}, \hat{q_{t-k}})$, with $k = 1$. Table 3.2 shows MSFT first-order autocorrelation for intraday trade directions in Jan 2001.

In table 3.2, the first column is the trading date. Within each trading date, we got positive correlation between $\{q_t\}$ and $\{q_{t-1}\}$ as shown in the second column. The intraday correlation estimates are all strongly nonzero by pearson's test. We are surprised to get very close intraday positive correlations among all trading dates, with

Table 3.2: Sample first-order autocorrelation of trade direction

Date	Autocorrelation	N	P-Value
2	0.3306	40237	<0.001
3	0.3049	58859	<0.001
4	0.3331	48718	<0.001
5	0.3341	42902	<0.001
8	0.3559	41755	<0.001
9	0.3543	55388	<0.001
10	0.3837	48945	<0.001
11	0.3232	41093	<0.001
12	0.3545	36273	<0.001
13	0.3653	33158	<0.001
17	0.3529	31529	<0.001
18	0.3786	42356	<0.001
19	0.3526	73470	<0.001
22	0.3871	37743	<0.001
23	0.4009	31299	<0.001
24	0.3995	45075	<0.001
25	0.3954	36869	<0.001
26	0.3934	34915	<0.001
29	0.3666	27078	<0.001
30	0.3769	25538	<0.001
31	0.3449	34231	<0.001
Mean	0.3614	41306.24	
Std	0.0270	11258.93	

mean 0.3614 and standard error 0.027.

These results have meaningful empirical implications. First, the assumption of the Roll model is not valid in practice. The sequence of the order types are more likely to pair with each other, buy after buy, sell after sell. Secondly, this may imply the asymmetric information processed by difference traders, since the informed traders always trade in the direction of his knowledge. Finally, this explains how day traders could make money by following the market. The daily order flows have high probability to be in the same trade directions sequentially. We would like to address this finding in our statistical inference.

The intraday signed order flows $\hat{\Delta y_t}$ on the other hand do not exhibit positive

autocorrelation, it is not significant against zero hypothesis. The order sizes are indeed exogenous along the time horizon (e.g. 500 size order could be followed by another 10 or 2000 size order).

3.3   Asymmetric information

In G-M model, the asymmetric information is μ, the proportion of the informed trader in the population. The asymmetric information parameter μ is positively correlated to $p_k(sell)$, and the bid-asked spread from the previous chapter. These results suggest use of the bid-ask spread or the impact of an order has on subsequent prices as proxies for the asymmetric information. We study the intraday bid-ask spread movement across one year.
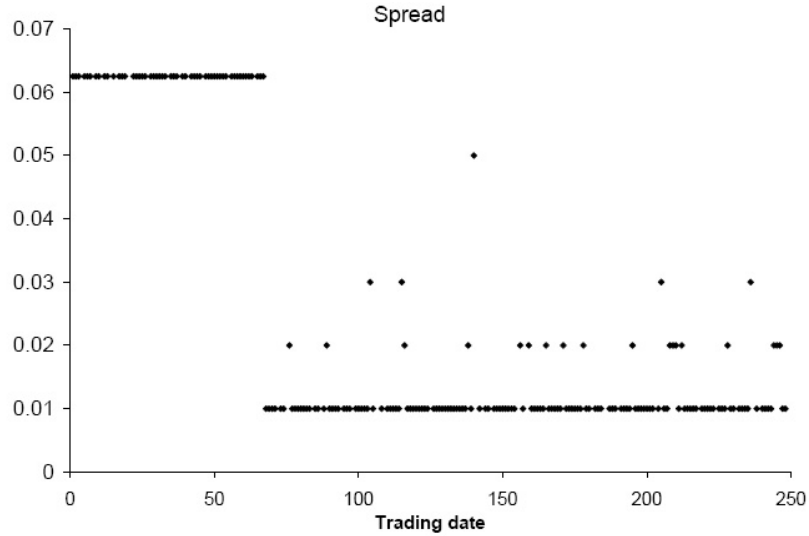
Figure 3.2: MSFT intraday bid-ask spread movement

Fig. 3.2 depicts the MSFT intraday bid-ask spread movement in year 2001. From the figure, we know that the intraday bid-ask spread are disjoint and stable. And the market maker does not post irrational bid-ask spreads based on trade price movement. The conclusion is that this spread remained constant at least for intraday or short time

period (i.e., one week).

On the other hand, we look at the trade price movement vs. the bid-ask spread. Fig. 3.3 illustrates the trade price movement in corresponding year. The bid-ask spread does

Figure 3.3: MSFT trade price movement

not have strong correlation with the trade price either positively or negatively. The spread contains dealer's posterior inference about the degree of informed traders, and this measurement is constant at short period. In Fig. 3.2 and 3.3, dealer's posterior beliefs about the degree of asymmetric information is decreasing, therefore we have seen higher spread in January-April than later the same year, while the trade price still move upward or downward in both directions.

The degree of informed trading among total market participants may not change in the short time period, at least from the dealer's viewpoint (dealers only see aggregated order flow, not distinguished one). This implies that the market maker makes no inference when he see the total order imbalance at tick level. He will shift the whole bid-ask band rather than change the spread itself.

3.4 Market depth

The *market depth* is an important characteristics of market dynamics. It refers to the size of an order flow innovation required to change the price in a given amount. In Kyle's framework, the market depth is λ_n^{-1}, with $p_n = p_{n-1} + \lambda_n \Delta Y_n$ for $n = 1, ..., N$. It deals with order imbalance with respect to the price increment. We do empirical studies at intraday transaction level.

First, we present the aggregated intraday transaction level order imbalance across trade price increments. Figure 3.4 illustrates MSFT aggregated order imbalance vs. price changes at each trading date using microstructure data. The correlation between
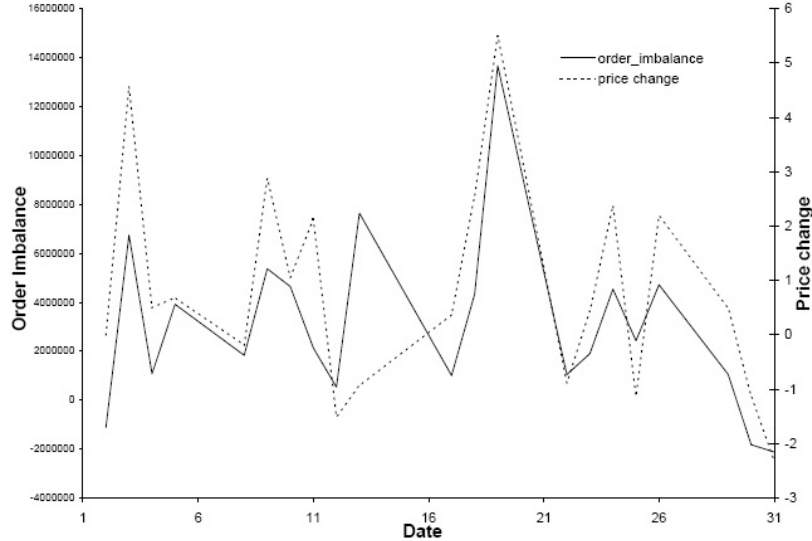
Figure 3.4: Aggregated intraday order imbalance vs. price change

the two series is 0.76. The results show strong explanatory power of order imbalance in the price change movement.

We conjecture the market depth (or λ_n) is constant in Kyle's model. We use regression to do our analysis. The t-statistics for λ coefficient is 5.09, with p-value less than 10^{-5} which rejects zero coefficient null hypothesis. The line fit chart is depicted in 3.5.
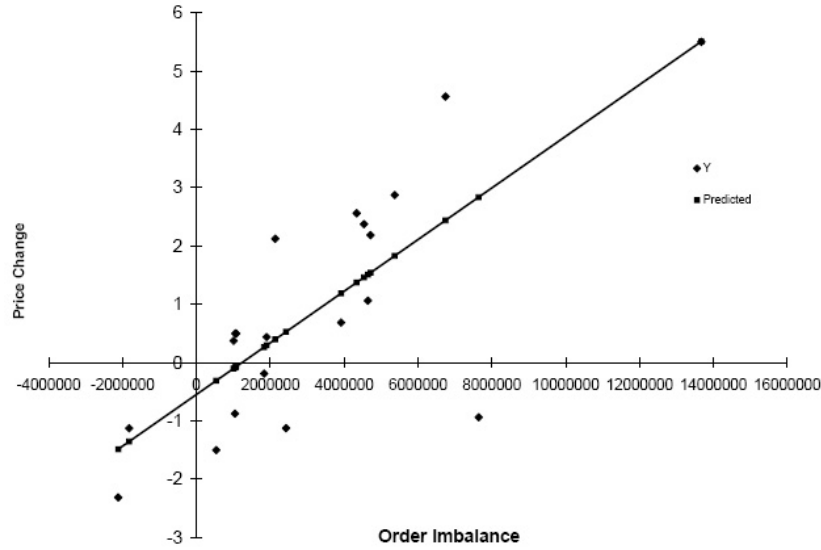
Figure 3.5: Line fit with market depth (or λ_n)

The empirical results show that the market depth could be modeled with the introduction of order imbalance or signed order flow. These findings have significant impact on the model inference.

In Kyle's model, the informed trader wants to trade aggressively on her private information, i.e., buy a large quantity if her information is positive. But the market maker knows that if he sells into a large net buy customer, he is more likely to be on the wrong side of the trade. He protects himself by setting a price that is increasing in the *net order flow*. This acts as a brake on the informed trader's desires: if she wishes to buy a lot, she will have to pay a high price. If there is an imbalance between buy and sell orders, the market maker makes up the difference in the Kyle's model. The results assert that it is the *order imbalance* that drives the price movement (price set by the market maker).

3.5 Market liquidity

In the previous section, we considered the liquidity parameter λ, and the results show that demanding liquidity has a cost. Intuitively, if you demand high liquidity, the price would be high. In Kyle's single period trading model, the λ takes following form:

$$\lambda = \frac{\sqrt{\Sigma_0}}{2\sigma}, \tag{3.1}$$

$\sqrt{\Sigma_0}/\sigma$ is ratio of volatilities, i.e., the value uncertainty vs. the noise order uncertainty. Therefore, the $\lambda\, \Delta Y$ is like a liquidity risk, where ΔY is the total order imbalance.

$$\lambda\, \Delta Y = \frac{\sqrt{\Sigma_0}}{2}\frac{\Delta Y}{\sigma}, \tag{3.2}$$

$\Delta Y/\sigma$ is proportional to the percentage of order imbalance. The higher σ, the lower the price impact. It is scaled by the value uncertainty $\sqrt{\Sigma_0}$. The higher the security value uncertainty, the higher the price impact.

The factor model starts with Fama and French (1992), which shows that factors related to company size and BtoM (Book to Market) ratio are able to explain a significant amount of the common stock variation in stock returns. They run the three-factor model of the form:

Fama and French 3-factor

$$R_j t - R_f t = \alpha_j + \beta_j\,(R_m t - R_f t) + \gamma_j\, SMB_t + \xi_j\, HML_t + \epsilon_j t, \tag{3.3}$$

where $R_j t$ is the return to portfolio j for time t, $R_f t$ is the risk-free return for time t. SMB_t is the small cap stock vs big cap stock, and HML_t is the high BtoM stocks vs low BtoM stocks. We would conjecture market liquidity as an additional factor given that demanding liquidity has a cost.

Chapter 4

Liquidity in market microstructure models

4.1 Liquidity and expect returns

Financial markets deviate from the perfect-market ideal in which there are no impediments to trade. A large and growing body of work has identified a variety of market imperfections, ranging from information asymmetries, participation costs, different forms of trading costs, inventory risk(i.e., the market maker, being exposed to the risk of price changes while he holds in inventory, requires compensation), to search frictions(i.e., a tradeoff between search and quick trading at a discount) etc. These cost of *illiquidity* should affect the securities prices if investors require compensation for bearing them. In addition, because liquidity varies over-time, risk-averse investors may require a compensation for being exposed to liquidity risk. These effects of liquidity on asset prices are important. Most papers focus on theoretical study of a specific imperfection that predicts how liquidity affects a security's expected return and/or the empirical connection between the two. The basic premise in these studies is that illiquidity is positively related to the expected returns.

We start the overview with different liquidity measures, and explore the effect of liquidity on assets expected returns by empirical evidence. The literature on liquidity is vast. Madhavan (2002), Bias, Glosten and Spatt (2005), Cochrane (2005), Vayanos and Wang (2009) have surveyed on liquidity and asset prices. While the effects of imperfections on market liquidity and further on expected returns have received much attention, their focuses are expected returns and mostly based on factor models, i.e, adjusted CAPM, adjusted Fama-French models etc. We then distinguish our work from those related literature such that we study the origins of illiquidity(e.g., in the form of bid-ask spreads or market impact) and fundamentals of the imperfections on the price movement with high frequency microstructure data.

4.2   Liquidity measures

One strength of a frictionless economy is that a security's cash flows and the pricing kernel are sufficient statistics for the pricing operation described as:

$$p_t = E_t\{ \ (p_{t+1} + d_{t+1}) \ \frac{m_{t+1}}{m_t}\}. \tag{4.1}$$

where m_t is the stochastic discount factor, d_t is the dividend process. Equation (4.1) is the main building block of standard asset pricing theory. The assumption of frictionless market is combined with *no arbitrage, agent optimality* and *equilibrium. No arbitrage* means that one can not make money in one state of nature without paying money in at least one other state of nature. *Agent optimality* derives investor's optimal portfolio choice only on a solution in the absence of arbitrage. If the investor's preferences are represented by an additively separable utility function $E_t \sum_s u_s(c_s)$ for a consumption process c, then $m_t = u'_t(c_t)$ is the marginal utility of consumption. In a complete market, agents $i = 1, ..., I$ with separable utility functions u^i compete to a competitive

equilibrium, (4.1) is satisfied with utility function $u_t = \sum_i \lambda^i u_t^i$ where λ_i is the Pareto Weights and depend on agents' endowments. In a frictionless market, the assumption of no arbitrage is essentially equivalent to the existence of a stochastic discount factor. That means the pricing kernel summarizes all the needed information contained in utility functions of agents, endowments, correlation with other securities etc.

In an economy with frictions, the price depends additionally on the security's liquidity and the liquidity of all other securities. In some liquidity models, there still exists a pricing kernel m such that (4.1) holds. In this case, illiquidity affects m_t, but the pricing of securities can still be summarized using a pricing kernel. The empirical analysis of Pastor and Stambaugh(2003) is based on an assumption that there exists an m that depends on a measure of aggregate illiquidity. In other models of illiquidity, however, there is no pricing kernel such that (4.1) applies. For instance, in transaction-cost-based models, securities with the same dividend cash flows have different prices if they have different transaction costs. Hence, a security's transaction cost not only affects the nature of market equilibrium, it is the fundamental attribute of the security.

If there does not exist a pricing kernel, the general equilibrium prices with illiquidity may depend on the fundamental parameters in a complicated way that does not have a closed-form expression. Nevertheless, we still can get important insight into the main principles how liquidity affects assets expected return under certain assumptions and with the assistance of empirical studies.

We consider the challenges of choosing a liquidity measure L. The problem of estimating liquidity on asset returns relies on how to measure liquidity since there is hardly a single measure that captures all of its aspects. Moreover, measures used in empirical studies are constrained by data availability. High-frequency data that enable the estimation of liquidity from the actual sequence of trade and quote become available in U.S only recently and are thus available for a relative short period of time.

Consequently, researchers requires a long time series to increase the power of the tests. In stock market outside U.S., high frequency data are hardly available, researchers need to find other measures of liquidity using low-frequency data, such as daily return data, and trading volume etc. The empirical studies from related work employ various measure of liquidities.

4.2.1   Bid-ask spread

Amihud and Mendelson(1986) studies the liquidity on stock's expected return using quoted bid-ask spreads. These predictions are tested using stock returns over the period 1961-1980 and data on quoted relative spreads. The spreads are the average of the beginning- and end-of-year end-of-day quotes, collected from Fitch quote sheets for NYSE and AMEX stocks. The estimation model is:

$$R_j = a + b\ \beta_j + c\ ln(S_j). \tag{4.2}$$

where R_j is the monthly stock portfolio return in excess of the 90-day Treasury Bill rate, β_j is the systematic risk, estimated from the preceding period, and S_j is the relative bid-ask spread. All coefficients are significant.

The model's estimations are: (1) the portfolio return *increases* with the bid-ask spread, which is the main result. (2) the return-spread slope *decreases* in the bid-ask spread, reflecting concavity which is due to clientele effect. In equilibrium, less liquid assets are allocated to investors with longer holding periods, which mitigates the compensation that they require for the costs of illiquidity. (3). the size effect reflects an aspect of liquidity - it is less costly to trade stocks of large companies - then the size effect should *weaken* if stock expected return equation includes the bid-ask spread, i.e., the bid-ask spread is known to be negatively related to firm size.
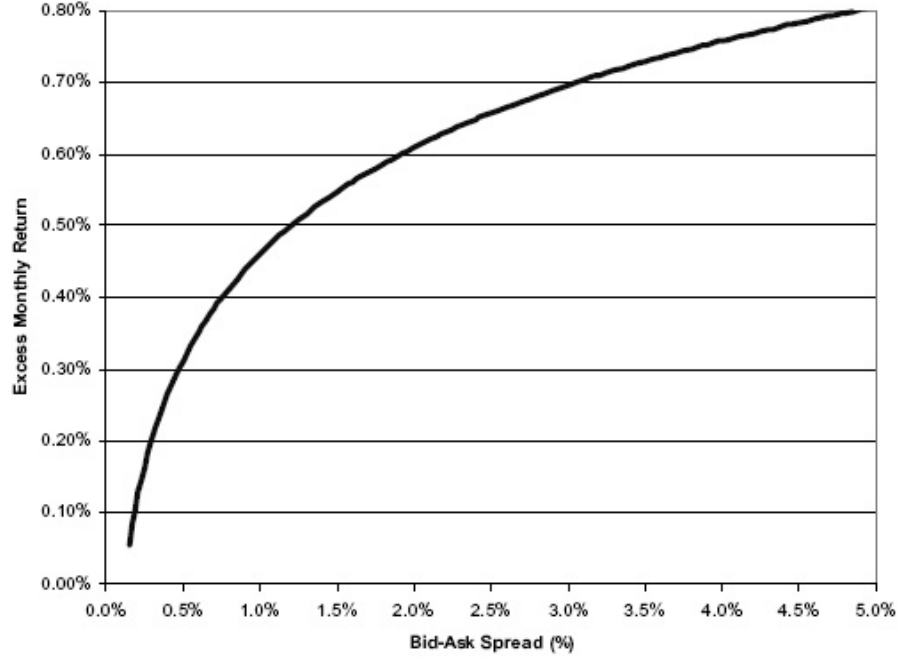
Figure 4.1: Relationship between stocks' excess monthly returns and bid-ask spreads

While on NYSE and AMEX, individual investors could trade through limit orders that had priority over the specialist's quotes and thus avoid the cost of spread although incurring the cost of risk and delay, on Nasdaq trading are done mostly through market makers, and investors have to endure the cost of spread. The estimated effect of the bid-ask spread is expected to be stronger when using Nasdaq stocks than NYSE and AMEX. This is shown in Eleswarapu(1997), who estimates a model where stock return is regressed on the stock's beta, relative spread and log(size). The estimation is performed for individual stocks employing Fama and MacBeth method. The consistent significant effect is the relative spread which has positive effect.

4.2.2 Kyle's λ

A finer measure of illiquidity is Kyle's λ, which is used by Brennan and Subrahmanyam(1996); Chordia, Huh and Subrahmanyam (2007) and Chordia et al (2009).

Chordia et al. named it "theory-based" illiquidity as it originates from Kyle's framework.

Brennan and Subrahmanyam(1996) estimate λ by regressing the price change, on the transaction size. The slope coefficient from the regression is Kyle's λ which measures the price impact of a unit of trade size, being larger for less liquid stock. The regression model also includes $\phi = D_t - D_{t-1}$, where $D_t = 1$ for a buy transaction and $D_t = -1$ for a sell transaction. The coefficient of this differential, ϕ, reflects the fixed cost of trading that is unrelated to the order size. The illiquidity variables that are used are: (1) $C_q = \lambda \, q/P$, the average of the marginal cost of trading, where q and P are monthly averages of trade size and price. (2) ϕ/P, the relative fixed cost of trading. These measures of illiquidity are then used in a cross-section regression of monthly NYSE stock returns for the years 1984-1991. The regression model employs Fama and French (1992) three factors model in addition to the illiquidity variables: The market return index, the small-minus-big firm return indexes and high-minus-low book-to-market return index.

The results show that C_q have a positive and significant effects on returns adjusted by Fama-French factors. In addition, C_q^2 has a negative and significant effect, consistent with Amihud and Mendelson (1986) clientele effect that generates an increasing and concave relationship between returns and illiquidity costs.

Chordia, Huh and Subrahmanyam (2009) consider the illiquidity λ in Kyle-type framework with extension to N informed traders and each informed trader i observes a signal with an error $\epsilon_i, i = 1, 2, ...N$, where $\epsilon_i \sim N(0, v_\epsilon)$. The asset payoff is $W = \tilde{W} + \delta$, where $\tilde{W}$ is expected payoff, and $\delta \sim N(0, v_\delta)$. The informed traders maximize their expected profit, while the uninformed traders who trades randomly and submit order

size, $z \sim N(0, v_z)$. The author shows the Kyle's measure λ is given by:

$$\lambda = \frac{v_\delta}{(N+1)v_\delta + 2v_\epsilon} \sqrt{\frac{N(v_\delta + v_\epsilon)}{v_z}}. \tag{4.3}$$

where N is the number of informed traders. v_δ is the variance of the asset payoff, v_z is the variance of uninformed trades, v_ϵ is the variance of signal innovation. Note that this measure requires proxies, for instance, a proxy for the variance of the signal innovation, as well as that of the signal itself. Each of those variance is represented by different proxies. v_δ is proxied through the earnings volatility from the most recent eight quarters. v_ϵ is proxied by the earnings surprise defined as the absolute value of the current EPS minus the EPS forecast four quarters ago. v_z is proxied by the average daily dollar volume (in million dollars) within the previous month.

The main model is still multi-factor model. The key contribution is that it uses Kyle's λ to derive a liquidity measure and to establish the connection between liquidity and expected returns.

4.2.3 Daily returns, trading volume and ILLIQ

Researchers often use alternative measures based on daily data on volume, shares outstanding, and prices, which are available for most markets.

Brennan, Chordia, and Subrahmanyam(1998) uses stock's dollar trading volume as a measure of liquidity in a multi-factor asset pricing model, version of the APT. Datar, Naik and Radcliffe (1998) use stock turnover (the ratio of stock volume to the number of shares outstanding) as a measure of liquidity. The logic behind the scene is that if in equilibrium less liquid stocks are allocated to investors with longer holding periods (Amihud and Mendelson, 1986), or investors reduced their trading frequency of illiquid stocks (Constantinides, 1986), then even though liquidity is not directly observed, it

37

can be inferred from the average holding period of the stock, which is the reciprocal of the stock turnover. Datar et al. estimate the cross-section of NYSE stock returns on stock returns in years 1963-1991, controlling for size, book-to-market ratio and beta, employing the Fama and MacBeth method. The result is that the longer the average holding time which implies lower liquidity, or low turnover, the high the expected return.

Amihud (2002) examines the effect of illiquidity on the cross-section of stock returns using an illiquidity measure called $ILLIQ$, where $ILLIQ = |R|/(P * VOL)$, where R is daily return, P is the closing daily price and VOL is the number of shares traded during the day. Intuitively, $ILLIQ$ reflects the relative price change that is induced by a given dollar volume, which is related to Kyle's pricing impact λ, but on a daily basis.

4.3 Liquidity risk

Liquidity varies over time which means the investors are uncertain what transaction cost they will incur in the future. Secondly, since liquidity affects the level of prices, liquidity fluctuations can affect the asset volatility itself. For both reasons, liquidity fluctuations constitute a new level of risk that impact the fundamental risk. This section explores liquidity models of the effect of a security's liquidity risk on its expected returns.

Acharya and Pedersen (2005) presents a model which gives rise to a "liquidity-adjusted CAPM" model that shows how liquidity risks are captured by three liquidity betas, and how shocks affect future expected returns. Re-writing the one-beta CAPM in net returns in terms of gross returns, we get a liquidity-adjusted CAPM for gross returns. Acharya and Pedersen introduce three liquidity betas, $\beta^{L1}, \beta^{L2}, \beta^{L3}$.

$$E_t(r_{t+1}^i) = r^f + E_t(c_{t+1}^i) + \lambda_t(\beta_t + \beta_t^{L1} - \beta_t^{L2} - \beta_t^{L3}). \tag{4.4}$$

where $\lambda_t = E_t(r_{t+1}^M - c_{t+1}^M - r^f)$ is the risk premium. $E_t(c_{t+1}^i)$ is the expected relative illiquidity cost. The models states that the required excess return is the expected relative cost plus four betas times the risk premium.

The first liquidity beta β_t^{L1} measures the covariance between the asset's illiquidity and the market illiquidity. The model implies the expected return increases with this covariance, because investors want to be compensated for holding a security that becomes illiquid when the market in general becomes illiquid. The second liquidity beta β_t^{L2} measures the exposure of asset i to marketwide illiquidity, which is usually negative. This beta affects return negatively because investors are willing to accept a lower return in times of market illiquidity. The more negative the exposure to the market illiquidity, the greater is the expected return. The third liquidity beta β_t^{L3} measures the sensitivity of the asset's illiquidity to the market conditions. This beta is also negative for most stocks.

Chapter 5

Inference on an Extended Kyle's Model

Research in empirical market microstructure has two important aspects. On the economic side, certain trading mechanisms and market frictions, such as transaction costs and asymmetric information, are incorporated in a proposed microstructural model with a utility function. The resulting constrained optimization problem is tackled, and its equilibrium solution will yield an optimal trading strategy from the perspective of each market participant, and a risk-neutral clearing price for every traded asset. In particular, the solution often enables us to interpret the economic impacts of certain model parameters contained in the model. On the statistical side, inference on model parameters is performed based on real market data, usually represented by time series of asset prices and returns, trading volumes, orders and quotes, etc. whether they are daily or involving intra-day activities. More often than not, goodness-of-fit of the proposed model need not be satisfactory. Naturally, more sophisticated models can be considered. However, such purely statistical approach may deviate us from an interpretable economic framework, and it may not be clear what we can learn from an improved model even it fits data better.

This work begins with the celebrated strategic trade model in Kyle (1985). It aims at fitting a modified version of Kyle's model using some intra-day data, and retaining its interpretability. Our study consists of three parts:

[1] an alternative characterization of the equilibrium solution to Kyle's model;

[2] derivation of the equilibrium solution to an extended Kyle's model in which the informed trader observes a noisy signal of the asset value;

[3] A case study of simulated equilibrium solutions.

[4] MCMC dynamic Bayesian inference on a proposed extension of Kyle's model in the next chapter.

Kyle (1985) proves the existence and uniqueness of a linear equilibrium solution in which the parameters are derived via a set of recursive formulas. In Part [1], we provide a new method to reproduce those parameters. Our method is computationally more convenient and direct. It also paves a road for deriving equilibrium solutions to certain extended models. One extension is analyzed in Part [2], in which the informed trader observes a noisy signal of the asset value v instead of v itself. In Part [4], we perform Bayesian inference on an extended model based on real microstructure data.

5.1 New derivation of Kyle's equilibrium solution

Recall the basic framework and major result in Kyle (1985).

- Fix an asset in what follows. Suppose N auctions take place sequentially over a trading period (e.g. day, month, year). For each $n = 0, 1, \ldots, N$, t_n denotes the time for the nth auction, with $0 = t_0 < t_1 < \cdots < t_N = 1$.

- There is a single informed trader who knows the liquidation value v of the asset, and trades a quantity Δx_n at the nth auction. However, $v \sim N(p_0, \Sigma_0)$ is considered as a random variable by (uninformed) noise traders.

- The quantity traded by noise traders at the nth auction is denoted by $\Delta u_n \sim N(0, \sigma_u^2 \Delta t_n)$ with $\Delta t_n = t_n - t_{n-1}$. Assume $\Delta u_1, ..., \Delta u_N$ are independent, and they are also independent of v.

- Let $\Delta y_n = \Delta x_n + \Delta u_n$ be the batch order at the nth auction, and $\mathcal{F}_{n-1}^U$ denote the information set available to uninformed traders (including a market maker and all noise traders) at the beginning of the nth auction, consisting of the observations $\{p_1, ..., p_{n-1}; \Delta y_1, ..., \Delta y_{n-1}\}$, where p_i represents the asset's market clearing price determined at the ith auction.

- The informed trader (insider) has a richer information set available to him before making his move at the nth auction. Such a set $\mathcal{F}_{n-1}^I$ includes $\{\Delta x_1, ..., \Delta x_{n-1}; v\}$ in addition to $\mathcal{F}_{n-1}^U$. The insider chooses Δx_n based on $\mathcal{F}_{n-1}^I$.

- After the move made by the insider at the nth auction, the market maker determines the price p_n based on $\mathcal{F}_{n-1}^U$ and Δy_n.

- Let

$$\pi_n = \sum_{i=n}^{N} (v - p_i) \Delta x_i \tag{5.1}$$

be the total (future) profits of the insider to be made at auctions $n, n+1, ..., N$, and $X = (\Delta x_1, ..., \Delta x_N)$, $P = (p_1, ..., p_N)$ denote the insider's trading strategy and the market maker's pricing rule respectively. Hence $\pi_n = \pi_n(X, P)$.

Definition 3. *A sequential auction equilibrium is defined as a pair (X, P) such that the following conditions hold:*

(*C1*) (*profit maximization*)   *For* $n = 1, ..., N$ *and all* $X^{'} = (\Delta x^{'}_1, ..., \Delta x^{'}_N)$ *with* $\Delta x^{'}_i = \Delta x_i$, $i = 1, ..., n - 1$, *we have*

$$E[\pi_n(X, P)|\mathcal{F}^I_{n-1}] \geq E[\pi_n(X^{'}, P)|\mathcal{F}^I_{n-1}]. \tag{5.2}$$

(*C2*) (*market efficiency*)   *For* $n = 1, ..., N$ *we have*

$$p_n = E(v|\mathcal{F}^U_{n-1}, \Delta y_n). \tag{5.3}$$

Definition 4. *A sequential auction equilibrium* (X, P) *is called a linear equilibrium if the component functions of* X *and* P *are linear, and a recursive linear equilibrium if there exist parameters* $\lambda_1, ..., \lambda_N$ *such that*

$$p_n = p_{n-1} + \lambda_n \Delta y_n, \quad n = 1, ..., N. \tag{5.4}$$

The following theorem is the major result in Kyle (1985) which proves the existence and uniqueness of linear equilibrium, and characterizes those modeling parameters in it.

Theorem 2. *There exists a unique linear equilibrium* (X, P), *represented as a recursive linear equilibrium, characterized by (for* $n = 1, ..., N$)

$$\Delta x_n = \beta_n \ (v - p_{n-1}) \ \Delta t_n, \tag{5.5}$$

$$p_n = p_{n-1} + \lambda_n \Delta y_n, \tag{5.6}$$

$$\Sigma_n = Var(v|\mathcal{F}^U_n), \tag{5.7}$$

$$E[\pi_n|\mathcal{F}^I_{n-1}] = \alpha_{n-1} \ (v - p_{n-1})^2 + \delta_{n-1}; \tag{5.8}$$

Given Σ_0 and σ_u^2, the parameters $\beta_n, \lambda_n, \Sigma_n, \alpha_n, \delta_n$ are the unique solutions to equations

$$\alpha_{n-1} = [4\lambda_n(1 - \alpha_n\lambda_n)]^{-1}, \tag{5.9}$$

$$\delta_{n-1} = \delta_n + \alpha_n \lambda_n^2 \sigma_u^2 \Delta t_n, \tag{5.10}$$

$$\beta_n \Delta t_n = (1 - 2\alpha_n\lambda_n) [2\lambda_n(1 - \alpha_n\lambda_n)]^{-1}, \tag{5.11}$$

$$\lambda_n = \beta_n \Sigma_n \sigma_u^{-2}, \tag{5.12}$$

$$\Sigma_n = (1 - \beta_n\lambda_n \Delta t_n) \Sigma_{n-1}, \tag{5.13}$$

subject to $\alpha_N = \delta_N = 0$ and the second order condition

$$\lambda_n(1 - \alpha_n\lambda_n) > 0. \tag{5.14}$$

5.1.1 The original derivation of Kyle's solution

As is suggested on page 1325 in Kyle (1985), combining (5.11) and (5.12) yields

$$(1 - \lambda_n^2\sigma_u^2\Delta t_n/\Sigma_n)(1 - \alpha_n\lambda_n) = \frac{1}{2}, \tag{5.15}$$

which is a cubic equation in λ_n, given nonnegative values of α_n, Σ_n and σ_u^2. (5.15) has three real roots. The middle one is the only solution that satisfies the second order condition. Overall, the sequences $\{\lambda_n\}, \{\beta_n\}, \{\alpha_n\}, \{\delta_n\}$ and $\{\Sigma_n\}$ can be determined by iterating $n = N, N - 1, ..., 1$ backwards, given a pair (Σ_0, σ_u^2) and the boundary condition $\alpha_N = \delta_N = 0$. Since Σ_N is also unknown, we have to set an initial value arbitrarily and run a search until it converges. The detail is given as follows.

Given Σ_0, σ_u^2 and the boundary condition $\alpha_N = \delta_N = 0$, an iterative algorithm consists of the following steps:

S1: Make an initial guess Σ_N^* for Σ_N;

S2: Get $\lambda_N = \frac{\sqrt{\Sigma_N^*}}{\sigma_u \sqrt{2\Delta t_n}}$ using $\alpha_N = 0$ and (5.15);

S3: Set $n = N$;

S4: Get β_n and Σ_{n-1}^* from (5.12) and (5.13);

S5: Get α_{n-1} from (5.9);

S6: Solve the cubic equation (5.15) and use its middle root for λ_{n-1};

S7: Replace n by $n-1$ and go to S4 if $n > 0$;

S8: If $|\Sigma_0^* - \Sigma_0| > \epsilon$ where ϵ is a prescribed error bound, go to S2 with a different initial value Σ_N^*, and repeat ...

This backward induction search algorithm contains an outside loop and an inside loop: the outside loop, as shown in S1 — S8, determines Σ_N^* up to an acceptable error, while the inside loop solves a cubic equation for each n in S6. Even for a fixed pair (Σ_0, σ_u^2), the computational complexity for a desirable target result is $O(N^2)$. However, we can only fix (Σ_0, σ_u^2) in a simulation study. In an empirical study using real market data, Σ_0 and σ_u^2 have to be treated as unknowns and estimated. Conceivably, the required computational complexity for that task will increase rapidly and make the algorithm impractical. That is why we propose the following alternative algorithm, which is more efficient and has not been explored, to the best of our knowledge.

5.1.2 An alternative characterization of Kyle's solution

Proposition 1. *Assume the same conditions in Theorem 2 with $\Delta t_n = \frac{1}{N} \ \forall \ n$, and let $d_n = \alpha_n \lambda_n$. Then for every $n = N, N-1, ..., 1$ (running backwards and in particular, $d_N = 0$ follows from $\alpha_N = 0$), there exists a unique real root $d_{n-1} \in (0, 1/2)$ for the cubic equation*

$$8 \, d_{n-1}^3 - 8 \, d_{n-1}^2 - 2 \, K_n \, d_{n-1} + K_n = 0, \tag{5.16}$$

45

where

$$K_n = \frac{1}{1 - 2d_n}. \tag{5.17}$$

Other parameters in Theorem 2 are determined iteratively by

$$\Sigma_n = \frac{1}{2(1 - d_n)} \Sigma_{n-1}, \tag{5.18}$$

$$\lambda_n = \left[\frac{(1 - 2d_n) \Sigma_n}{2(1 - d_n)\Delta t_n \sigma_u^2} \right]^{1/2}, \tag{5.19}$$

$$\beta_n = \frac{1 - 2d_n}{2\lambda_n (1 - d_n)\Delta t_n}. \tag{5.20}$$

Furthermore, the sequence $\{d_n\}$ satisfies the property

$$\frac{1}{2} > d_1 > d_2 >, ..., > d_{N-1} > d_N = 0. \tag{5.21}$$

The sequence $\{d_n\}$ plays a central role in obtaining other parameters in Kyle's model. d_n has two factors: α_n as the coefficient for a quadratic utility function representing the expected future (at auctions $n, n+1, ..., N$) profit from the informed trader; and λ_n as a measure for the market depth (a smaller value of λ_n corresponds to a deeper market). There is another important parameter β_n, which models the informed trading intensity. The following proposition, derived from Proposition 1, depicts how the sequences $\{\beta_n\}$ and $\{\lambda_n\}$ will evolve as more auctions take place.

Proposition 2.

$$\frac{\beta_n}{\beta_{n-1}} \triangleq h_n = \sqrt{2(1 - d_{n-1}) \frac{1 - 2d_n}{1 - 2d_{n-1}}} > 1 \tag{5.22}$$

$$\frac{\lambda_n}{\lambda_{n-1}} \triangleq k_n = \sqrt{\frac{1 - 2d_n}{1 - 2d_{n-1}} \frac{1 - d_{n-1}}{2(1 - d_n)^2}} > \frac{1}{2}. \tag{5.23}$$

See Appendix for detailed proofs of Propositions 1 and 2.

46

Remark: There are several advantages for the proposed algorithm given in Proposition 1.

- Computational efficiency: With this algorithm, the cubic equation (5.16) only need to be solved for each n *once for all*, i.e. it does not depend on the values of Σ_0 and σ_u^2. Therefore, this part of computation is purely off-line. Having solved for the entire sequence $d_n, n = N - 1, ..., 1$, we can run a *forward* algorithm, with $n = 1, ..., N$ and an assigned pair (Σ_0, σ_u^2) to obtain other sequences $\{\lambda_n\}, \{\beta_n\}, \{\alpha_n\}, \{\delta_n\}$ and $\{\Sigma_n\}$. Suppose we have done the calculation for a given N, and decide to run it again for a larger $N' > N$. Then we can reuse the result of $d_N, d_{N-1}, ..., d_1$ for $d_{N'}, d_{N'-1}, ..., d_{N'-N+1}$, and continue to calculate only new values for $d_{N'-N}, ..., d_1$. Moreover, the only computational errors involved in the new algorithm come from numerical solutions for (5.16). No "trial-and-error" with different values for Σ_N^* in the previous numerical search is required. The greater value of N, the more efficient the new algorithm will be.

- From Proposition 2, we learn that the informed trader increases his orders as more auctions take place. As trading unfolds and more information is released to him, the insider has no incentive to hide his private information hence trades more aggressively. Following our derivation, neither ratio β_n/β_{n-1} nor ratio λ_n/λ_{n-1} depend on any other parameters in the model, except for the auction horizon N. However, the initial values β_0 and λ_0 do depend on the inputs Σ_0 and σ_u^2 [see (5.12)], and such dependence will carry on in subsequent values β_n and λ_n. Once the sequence $\{d_n\}$ is solved, ratios for both sequences will be uniquely determined. Moreover, the sequence $\{\beta_n\}$ exhibits a consistent growth, while λ_n does not reveal this property.

- As we mentioned, our main objective is to improve goodness-of-fit for an extended

47

Kyle's model in parameter estimation with real data. The new algorithm turns out to offer a useful clue for what extensions we may consider. We will elaborate on that part in the next section.

Figure 5.1 demonstrates the numerical results of $\{d_n\}$ sequence given number of periods N. $\{d_n\}$ is a decreasing sequence as we expect. It also shows that the beginning portion of $\{d_n\}$ sequence are concentrated within the range of $0.45 - 0.5$ when the total number of periods $N > 10$. When N is large, $\{d_n\}$ would be decreasing slowly for the majority of time periods, and drop sharply at the end of trading.
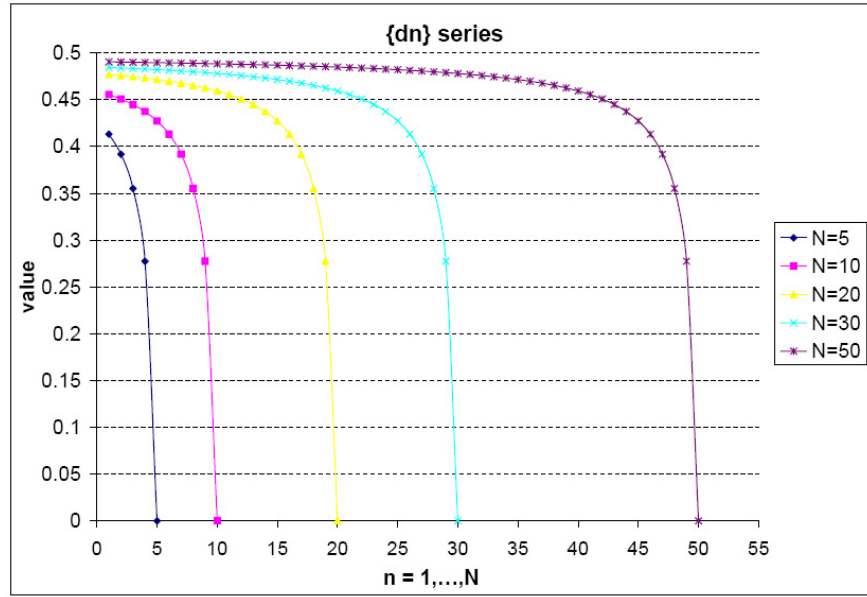
Figure 5.1: $\{d_n\}$ series with various sample periods N

We illustrate the patterns of λ_n and β_n in figure 5.2 given same initial condition with different sample periods N. β_n represents the insider's strategy while λ_n is the reciprocal of market depth. β_n is an increasing sequence under Kyle's equilibrium model, and it follows a pattern of flat at the beginning and gradually more steep toward the end of trading periods. If we compare β_n across different N, the results are favorable to the insider if he is trading in a longer time horizon, in which he could exercise his strategy

48

over longer periods. The λ_n sequence is at no surprise. It flattens out through the entire time horizon, and drop at the end of trading periods.
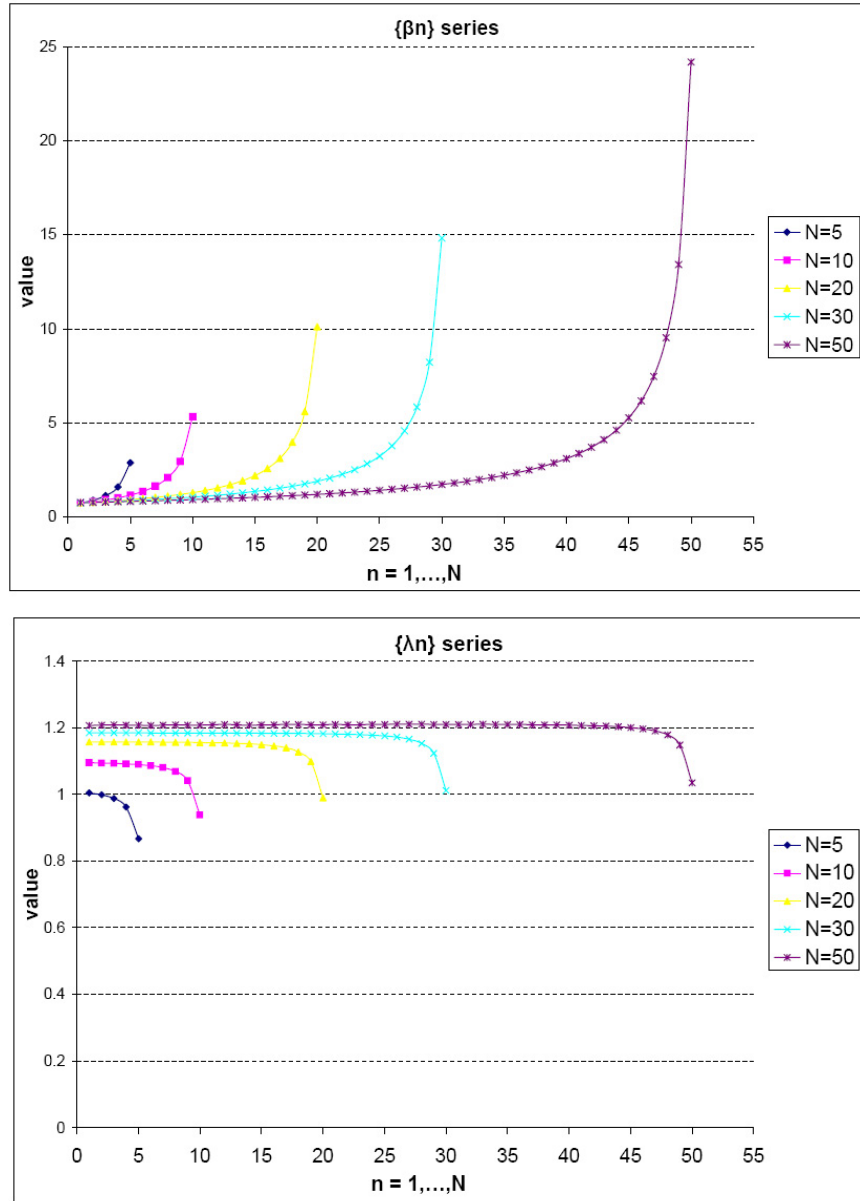

Figure 5.2: $\{\beta_n\}$ and $\{\lambda_n\}$ series with various sample periods N

49

at each auction

5.2 An extension of Kyle's model

In this section, we consider an extended Kyle's model in which the informed trader observes a noisy signal about true value v at each auction, but not v itself. We will focus on the case of sequential (multiple) auctions after skimming over the single period case.

5.2.1 The single period case

Consider an asset with payoff $v \sim N(p_0, \Sigma_0)$. The quantity traded by noise traders is denoted by $u \sim N(0, \sigma_u^2)$. Different from the original Kyle's model, here we assume the informed trader observes a signal $s = v + \epsilon$ at the beginning of the period where $\epsilon \sim N(0, \sigma_\epsilon^2)$. Conditioning on s, the informed trader maximizes his expected profit by choosing his trading strategy x. Assume that v, u, and ϵ are independent of each other. There is a competitive risk-neutral market maker, who sets the asset price as $p(y) = E(v|y)$ based on the batch order $y = x + u$.

Lemma 1. *There exists a unique equilibrium (X, P), in which the insider's trading strategy X and the market maker's pricing rule P are linear functions of s and y respectively:*

$$x(s) = \beta \ (s - p_0), \tag{5.24}$$

$$p(y) = p_0 + \lambda \ y, \tag{5.25}$$

where

$$\lambda = \frac{\Sigma_0}{2\sqrt{\sigma_u^2(\Sigma_0 + \sigma_\epsilon^2)}}, \tag{5.26}$$

$$\beta = \sqrt{\frac{\sigma_u^2}{\Sigma_0 + \sigma_\epsilon^2}}. \tag{5.27}$$

Proof: Let $\pi = [v - p(y)]\ x$. Following the linearity assumptions (5.24), (5.25) and conditioning on the signal s, the informed trader will choose $x = x(s)$ to maximize his expected profit

$$
\begin{aligned}
E(\pi|s) &= E[(v - p_0 - \lambda y)\ x|s] \\
&= x\ E(v - p_0|s) - x\ \lambda\ E(x + u|s) \\
&= x\ E(v - p_0|s) - \lambda\ x^2, \tag{5.28}
\end{aligned}
$$

where the projection theorem implies

$$
\begin{aligned}
E(v - p_0|s) &= \frac{Cov(v - p_0, s)}{Var(s)}\ (s - Es) \\
&= \frac{\Sigma_0}{\Sigma_0 + \sigma_\epsilon^2}\ (s - p_0) \\
&= \gamma\ (s - p_0), \tag{5.29}
\end{aligned}
$$

with

$$\gamma = \frac{\Sigma_0}{\Sigma_0 + \sigma_\epsilon^2}. \tag{5.30}$$

Maximizing $E(\pi|s)$ with respect to x leads to $-2\lambda x + E(v - p_0|s) = 0$, hence

$$x = \beta\ (s - p_0) \tag{5.31}$$

with

$$\beta = \frac{\gamma}{2\lambda}. \tag{5.32}$$

Furthermore, the projection theorem and (5.25) imply

$$
\begin{aligned}
p(y) &= E(v|y) \\
&= p_0 + \frac{Cov(v,y)}{Var(y)} (y - Ey) \\
&\triangleq p_0 + \lambda\, y,
\end{aligned} \tag{5.33}
$$

and

$$
\begin{aligned}
\lambda &= \frac{Cov(v, \beta\, (v - p_0 + \epsilon) + u)}{Var(y)} \\
&= \frac{\beta \Sigma_0}{\beta^2\, (\Sigma_0 + \sigma_\epsilon^2) + \sigma_u^2}.
\end{aligned} \tag{5.34}
$$

Therefore, (5.26) and (5.27) follow from (5.30) and (5.32). Moreover,

$$E(\pi|s) = \frac{\Sigma_0 \sqrt{\sigma_u^2}}{2(\Sigma_0 + \sigma_\epsilon^2)^{3/2}}\, (s - p_0)^2.$$

The ex-ante profit for the insider is given by

$$E(\pi) = \frac{\Sigma_0\, \sqrt{\sigma_u^2}}{2\, (\Sigma_0 + \sigma_\epsilon^2)^{1/2}}. \tag{5.35}$$

The special case with $\epsilon = 0$ will return to the original Kyle's solution, i.e., $\lambda = \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}}$ and $\beta = \sqrt{\frac{\sigma_u^2}{\Sigma_0}}$. Note that the noisy signal reduces the insider's profit compared to the case with $\epsilon = 0$ as we expected. The higher the noise signal (larger σ_ϵ^2), the lower the profit, also the lower the insider's trading intensity β, since the insider would trade more cautiously due to his imperfect knowledge about the asset value v.

5.2.2 The multiple period case

We now turn our attention to the multiple period case. The set up in Kyle's model (Section 1) will be followed with a distinction: We assume the informed trader observes a noisy signal $s_n = v + a_n \epsilon_n$ at the nth auction, where the errors $\epsilon_1, ..., \epsilon_N$ are iid $N(0, \sigma_\epsilon^2)$ random variables, the sequence $\{\epsilon_n\}$ is independent of v, and $a_1, ..., a_N$ are positive constants that decrease as $n \to N$. The information set $\mathcal{F}_{n-1}^I$, as a knowledge base for the informed trader to determine his strategy Δx_n at the nth action, now contains $\{\Delta x_1, ..., \Delta x_{n-1}; s_1, ..., s_n\}$, in addition to the uninformed traders' information set $\mathcal{F}_{n-1}^U$, which includes $\{p_1, ..., p_{n-1}; \Delta y_1, ..., \Delta y_{n-1}\}$. It makes sense for the informed trader to base his decision at the nth auction on the difference $E(v|s_1, ..., s_n) - p_{n-1}$ since he does not know v. For instance, the strategy $\Delta x_n = \beta_n(\bar{s}_n - p_{n-1}) \, \Delta t_n$ may be chosen with $\bar{s}_n = \frac{1}{n} \sum_{i=1}^{n} s_i$. However, that would present some technical difficulty in the derivation of recursive equilibrium solution to this problem. In what follows, the trading rule $\Delta x_n = \beta_n(s_n - p_{n-1}) \, \Delta t_n$ will be adopted which makes it easier to extend Kyle's original solution to the current setting.

Theorem 3. *There exists a unique linear equilibrium (X, P), represented as a recursive linear equilibrium, characterized by (for $n = 1, ..., N$)*

$$\Delta x_n = \beta_n \, (s_n - p_{n-1}) \, \Delta t_n, \tag{5.36}$$

$$\Delta p_n = p_n - p_{n-1} = \lambda_n \Delta y_n = \lambda_n(\Delta x_n + \Delta u_n), \tag{5.37}$$

$$E[\pi_n | \mathcal{F}_{n-1}^I] = \alpha_{n-1} \, (s_n - p_{n-1})^2 + \delta_{n-1}. \tag{5.38}$$

Given Σ_0, σ_u^2, σ_ϵ^2, the sequences β_n, λ_n, α_n, Σ_n, δ_n are the unique solutions to difference

equations

$$\alpha_{n-1} = \frac{b_n^2}{4\lambda_n \ (1 - \alpha_n\lambda_n)}, \quad where \ b_n = \frac{\Sigma_{n-1}}{\Sigma_{n-1} + a_n^2\sigma_\epsilon^2}, \tag{5.39}$$

$$\beta_n \ \Delta t_n = \frac{b_n \ (1 - 2\alpha_n\lambda_n)}{2\lambda_n \ (1 - \alpha_n\lambda_n)}, \tag{5.40}$$

$$\lambda_n = \frac{1}{\frac{\beta_n\Delta t_n}{b_n} + \frac{\sigma_u^2}{\beta_n\Sigma_{n-1}}} = \frac{\beta_n\Sigma_n}{\beta_n^2\Delta t_n \ (b_n^{-1} - 1) \ \Sigma_{n-1} + \sigma_u^2}, \tag{5.41}$$

$$\Sigma_n = (1 - \beta_n\lambda_n\Delta t_n) \ \Sigma_{n-1}, \tag{5.42}$$

$$\delta_{n-1} = \delta_n + \alpha_n\lambda_n^2 \ \sigma_u^2\Delta t_n + \ \alpha_n(1 - b_n) \ \Sigma_{n-1} + \alpha_n \ a_{n+1}^2\sigma_\epsilon^2, \tag{5.43}$$

subject to the boundary conditions $\alpha_N = 0$, $\delta_N = 0$ *and the second order condition*

$$\lambda_n \ (1 - \alpha_n\lambda_n) > 0. \tag{5.44}$$

See Apendix for the proof of Theorem 3.

Note: In the special case with $\epsilon = 0$, (5.39) — (5.43) go back to the solutions for the original Kyle's difference equations in Theorem 2.

Now we make some additional assumptions and present the following proposition (also to be proved in Appendix), which is an extension of Proposition 1, and plays an important role in characterization of the equilibrium solution provided in Theorem 3.

Proposition 3. *Assume* $\frac{a_n^2\sigma_\epsilon^2}{\Sigma_{n-1}} \equiv c \in (0, 1)$ *(constant), hence* $b_n = \frac{1}{1+c} \triangleq \gamma \in \left(\frac{1}{2}, 1\right)$ *for all* $n = 1, ..., N$. *We also assume* $\Delta t_n = 1/N \triangleq \Delta$ *for all* n. *Given* Σ_0, σ_ϵ^2, σ_u^2 *and* Δ, *let* $q_n = \alpha_n\lambda_n$ *and consider the cubic equation*

$$8q_{n-1}^3 - 8q_{n-1}^2 - 2K_nq_{n-1} + K_n = 0 \tag{5.45}$$

with $n = N, N - 1, ..., 1$ moving backwards, where

$$K_n = \frac{\gamma^4}{(1 - 2q_n) \left[2(1 - q_n) - \gamma(1 - 2q_n)\right]}. \tag{5.46}$$

For every n, (5.45) has a unique root $q_n \in \left(0, \frac{1}{2}\right)$ that is economically meaningful. Having obtained the sequence $\{q_n\}$, other parameters can be derived recursively (moving forwards):

$$\Sigma_n = \left[1 - \frac{\gamma(1 - 2q_n)}{2(1 - q_n)}\right] \Sigma_{n-1}, \tag{5.47}$$

$$\lambda_n = \left[\frac{(1 - 2q_n) \ \gamma \ \Sigma_{n-1}}{4(1 - q_n)^2 \ \sigma_u^2 \Delta}\right]^{1/2} = \left\{\frac{(1 - 2q_n) \ \gamma \ \Sigma_n}{2(1 - q_n) \left[2(1 - q_n) - \gamma(1 - 2q_n)\right] \sigma_u^2 \Delta}\right\}^{1/2} \tag{5.48}$$

$$\beta_n = \frac{\gamma(1 - 2q_n)}{2\lambda_n \ (1 - q_n)\Delta}, \tag{5.49}$$

$$\alpha_{n-1} = \frac{\gamma^2}{4\lambda_n \ (1 - q_n)}. \tag{5.50}$$

See Apendix for the proof of Proposition 3.

5.3   Simulation study of Kyle's equilibrium solution

In this section, we carry out simulation studies of equilibrium solution with the help of new algorithm developed from part [1].

We use $\sigma_u = 0.5$, $\Sigma_0 = 0.4$, $p_0 = 2.0$ and $N = 50$. Figure 5.3 demonstrates the equilibrium solution to model parameters. With all model parameters, we do simulation on order flows from both insider as well as from the noise traders, trade prices and profit. The true value of underlying security is assumed to be $V = 2.5$.
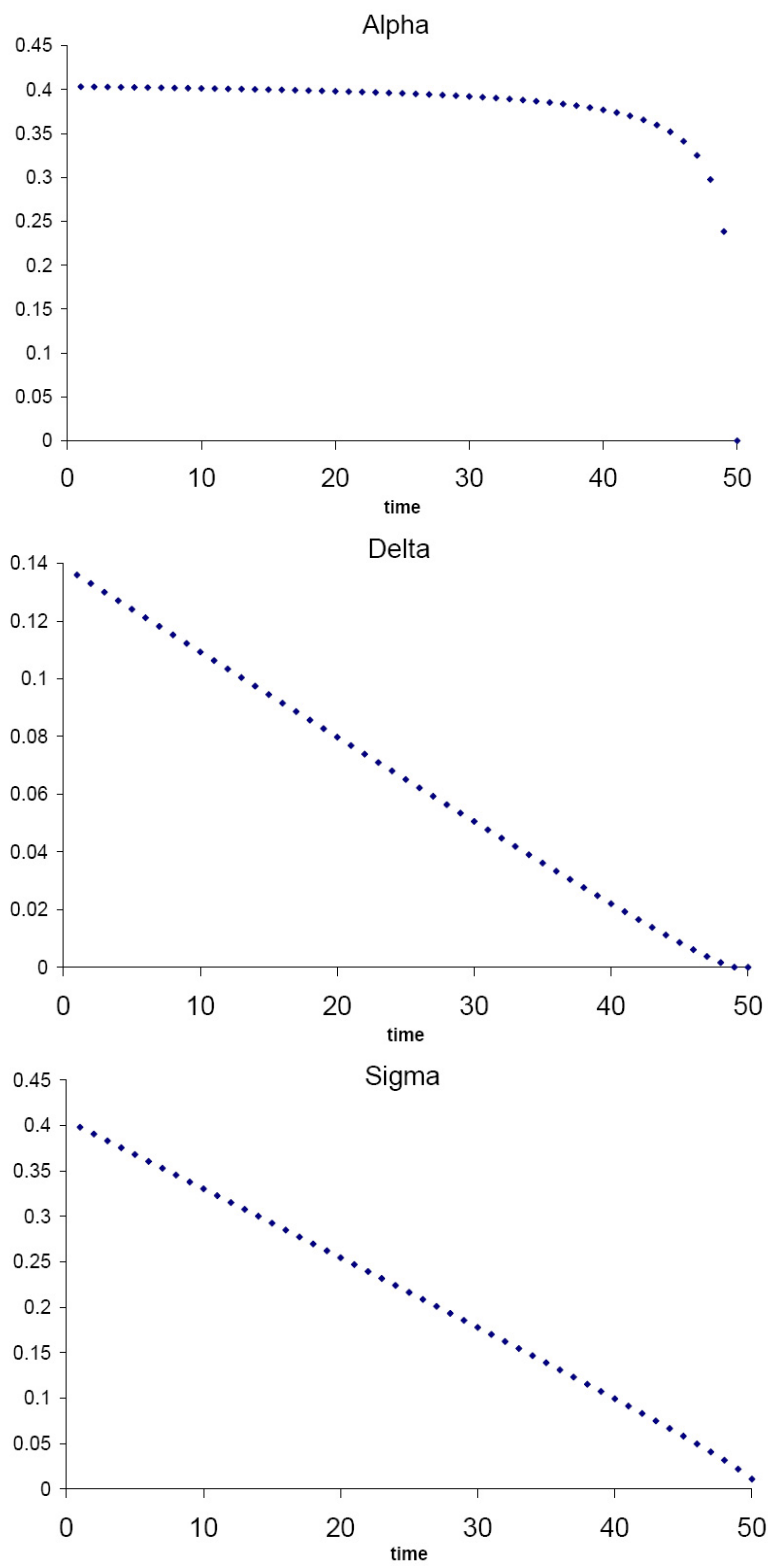
Beta

Lambda
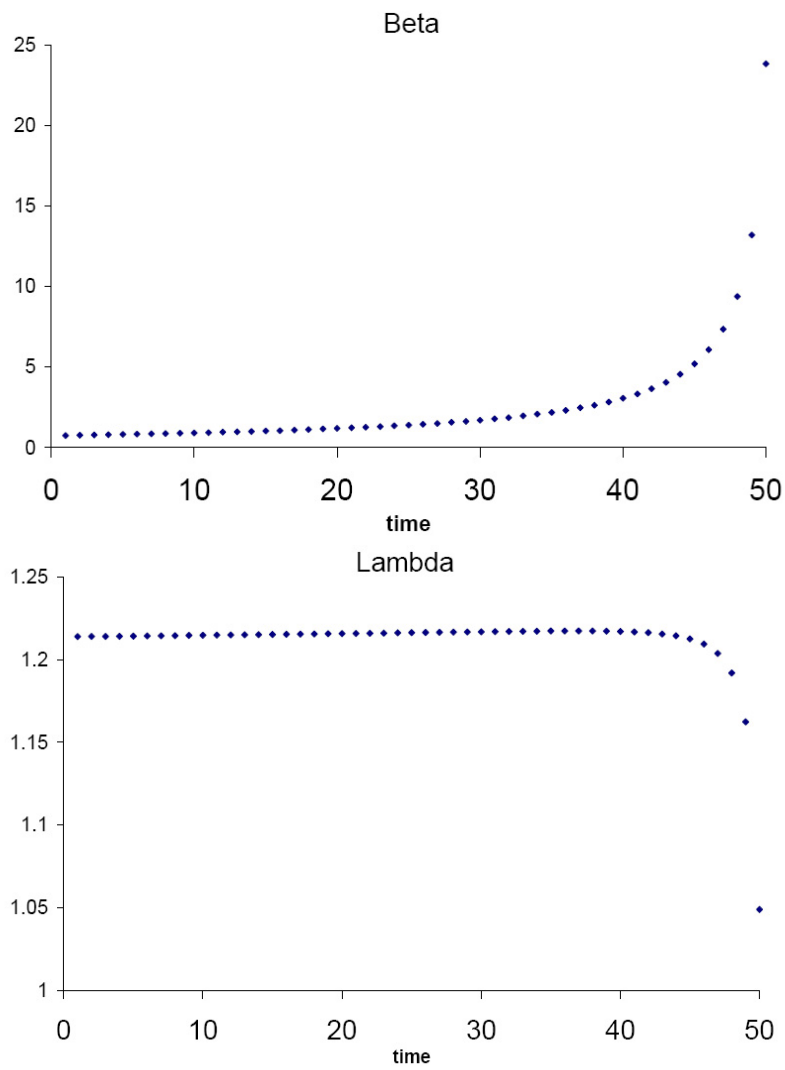
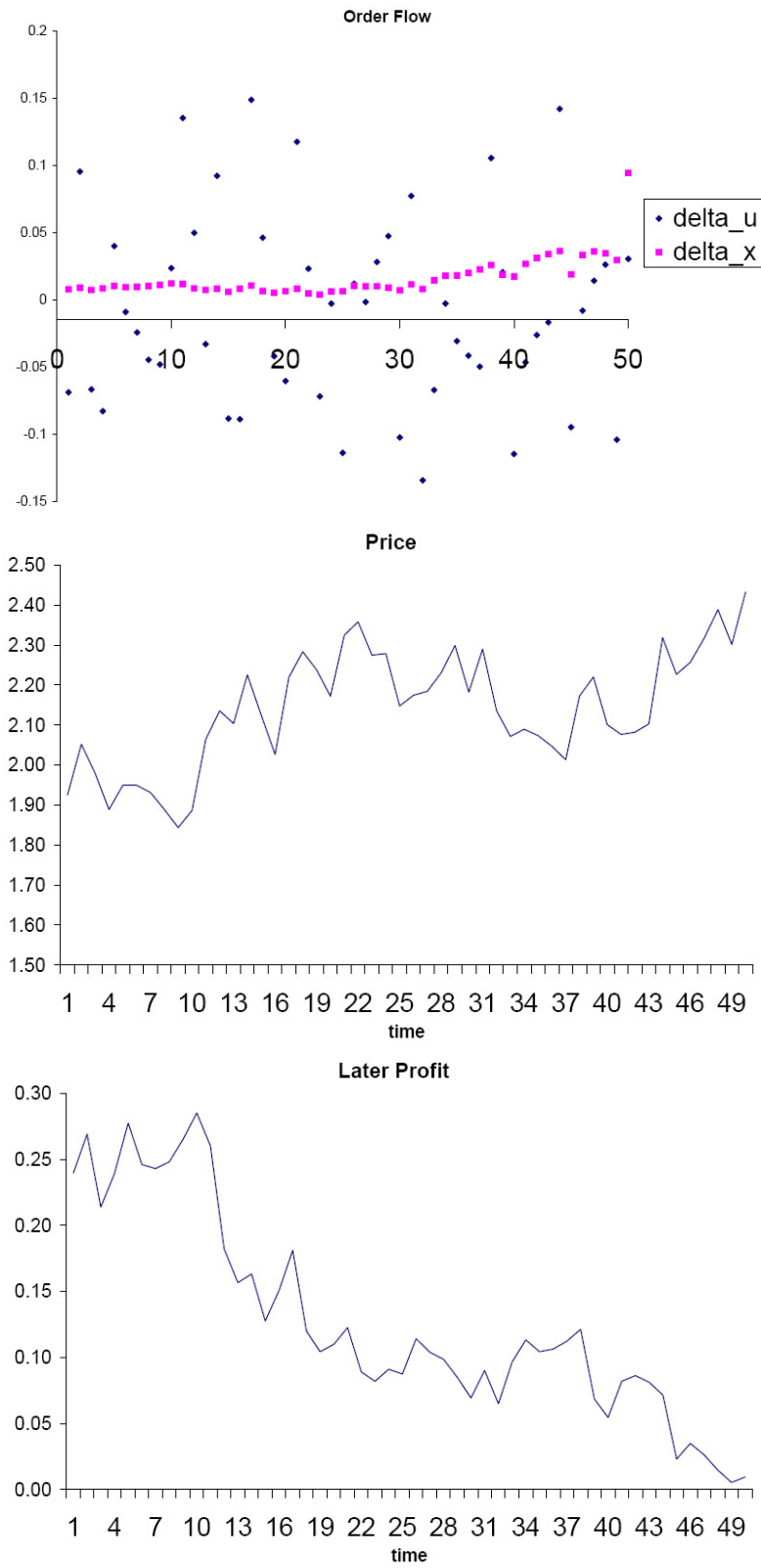Figure 5.3: Kyle model parameters, $\beta, \lambda, \alpha, \delta, \Sigma$

Figure 5.4: Order flows, trade prices and profit

There are some findings from the solutions to Kyle's model. For market depth(inverse of the λ), it is almost constant. Neither increasing nor decreasing market depth is consistent with behavior by the informed trader which is "stable" enough to sustain an equilibrium. For the informed trader, the optimal trading strategy β is through *gradual* submission, rather than abruptly. When true value revelation becomes close, the informed has no incentive to hide his private information and trade aggressively to benefit from the last-minute opportunity. The α and δ define the quadratic profit function. The leftover information Σ diminishes as trading unfolds which is not surprising. The end of period information left Σ_N (N=50 in our case) is non-zero which is determined dynamically from equilibrium solution.

The order flows and trade price evolution from figure 5.4 are illuminating with following properties: (1).The informed trader is actively disguise his informed orders(Δx) among the uninformed orders(Δu). (2). Informed orders are larger after negative uninformed trades. i.e.,Assume informed trader has positive projection, V is greater than current price p. After a large uninformed buy, the informed sends a small buy order; whereas, after a large uninformed sell, the informed submits a large buy order.(3). Informed order size increases over time. As trading unfolds, informed orders become more aggressive. (4) The volatility of trade prices is determined by the noise traders but not by the insider. There is a sense in which the "trading volume" of the insider is small. Despite his small trading volume, however, the insider ultimately *determines* what price is established at the end of trading. He does this because his trades, unlike the trades of noise traders, are positively correlated from period to period. (5). The end of trade price would not converge to the true value while trade price moves toward its value. $p_{50} = 2.43$ against true value of 2.5 in this simulation.

Chapter 6

Dynamic Bayesian Inference

In this chapter, we propose a Bayesian time series model, and carry out simulation studies. The model is originated from Kyle's model, and in particular, motivated by the derivation of Kyle's equilibrium solution as shown in Proposition 3 from last chapter.

We establish the theoretical framework for Bayesian time series. First we look at Bayesian approach, since they constitute the foundation of the field; we move on to Bayesian time series with important properties but also their limitations. We developed a Bayesian time series model based on ideas originally put forth in market microstructure models. We then discuss Markov Chain Monto Carlo (MCMC) algorithms with proposal of different methods. The empirical studies and Bayesian model assessment will be presented in the next chapter.

Statistical time series analysis using state space models were developed in the 1970-80's. Although the model was originally introduced as a method primarily for use in aerospace-related research, it has been applied to modeling data from economics, Harvey and Pierse (1984), West and Harrison (1997), Durbin and Koopman (2001). The impressive growth of applications is largely due to the possibility of solving complicated problems using Monte Carlo methods in a Bayesian framework. A Bayesian approach has several advantages, both methodological and computational. The time

series models like ARMA can be regarded in terms of state space models. But Bayesian framework offer more flexibility in treating non-stationary time series, or modeling hierarchical data with structural changes, and are often more easily interpretable. The more general class of Bayesian models extend the analysis to non-Gaussian and non-linear dynamic systems. We start with basic notions.

6.1 Basic notions

The Bayesian estimation assumes a T x 1 vector $y \doteq (y_1, ..., y_T)'$ of observations described through a probability density $p(y|\theta)$. The unknown parameter θ is a d-dimensional vector, where $\theta \in \Theta \subseteq R^d$. The difference between the Bayesian and the classical approach lies in the mathematical nature of θ. In the classical framework, it is assumed that there exist a true and fixed value for parameter θ. Conversely, the Bayesian approach consider θ as a random variable which is characterized by a *prior* distribution denoted by $\pi(\theta|\eta)$, where η is a vector of *hyerparameters*. Inference concerning θ is then based on its *posterior* distributions, given by

$$p(\theta|y, \eta) = \frac{p(y, \theta|\eta)}{p(y|\eta)} = \frac{f(y|\theta)\pi(\theta|\eta)}{\int_\Theta f(y|u)\pi(u|\eta)du} \tag{6.1}$$

The result of the integral in the denominator is sometimes written as $m(y|\eta)$, the *marginal* distribution of the data y given the value of the hyperparameter η. If η is known, we often suppress it in the notation and the posterior distributions would be written as $p(\theta|y)$.

If we are unsure as to the proper value for η, the proper Bayesian solution would be to quantify this uncertainty in a second-stage prior distribution called *hyperprior*. Denoting this distribution by $h(\eta)$, the posterior for θ is now obtained by marginalizing

over η,

$$
\begin{aligned}
p(\theta|y) &= \frac{p(y,\theta)}{p(y)} = \frac{\int p(y,\theta,\eta)d\eta}{\int \int p(y,u,\eta)d\eta du} \\
&= \frac{\int f(y|\theta)\pi(\theta|\eta)h(\eta)d\eta}{\int \int f(y|u)\pi(u|\eta)h(\eta)d\eta du}
\end{aligned} \tag{6.2}
$$

In principle, this is no reason why the hyperprior for η can not itself depend on a collection of unknown parameters λ, resulting in a generalization of (6.2) featuring a second-stage prior $h(\eta|\lambda)$ and third-stage prior $g(\lambda)$. This type of specification of a model over several levels is called *Bayesian hierarchical modeling*, with each new distribution forming a new level in the hierarchy. This constitutes an important framework to model structural changes.

The main objective of a statistical analysis is *forecasting*; the event of interest is the value of a future observation y^*. The prediction of a future value y^* given the data y is observed by computing the conditional distribution of y^* given y is called *predictive distribution*. It can be written as,

$$
\pi(y^*|y) = \int f(y^*,\theta|y)d\theta = \int f(y^*|y,\theta)p(\theta|y)d\theta \tag{6.3}
$$

where $p(\theta|y)$ is posterior distribution of θ.

6.2 Time series in Bayesian framework

A univariate or multivariate time series is described probabilistically by a sequence of random variables or vectors $(Y_n : n = 1,2,...)$, where the index n denotes time. One basic problem is to make forecasts about the value of the next observation, Y_{n+1} say, having observed data up to time n, $Y_1 = y_1, ..., Y_n = y_n$ or $Y_{1:n} = y_{1:n}$ for short. If we are able to specify the joint density $\pi(y_1, ..., y_n)$ for any $n \geq 1$, and Bayesian forecasting

would be solving by computing the predictive density,

$$\pi(y_{n+1}|y_{1:n}) = \frac{\pi(y_{1:n+1})}{\pi(y_{1:n})} \tag{6.4}$$

Suppose θ is the unknown parameter, and a prior $\pi(\theta)$ on θ, we obtain the joint density $\pi(y_{1:n}) = \int \pi(y_{1:n}|\theta)\pi(\theta)d\theta$. More importantly, Bayesian time series characterize some favorable mechanism which are not present from other approaches.

The dependence structure in Bayesian time series is normally assumed to be conditional independence, i.e., $Y_1, ..., Y_2$ are conditionally independent and identically distributed (i.i.d) given parameter θ, θ could have complicated structures, or even time dependent. $\pi(y_{1:n}|\theta) = \prod_{i=1}^{n} \pi(y_i|\theta)$.

Note that $Y_1, ..., Y_n$ are only conditionally independent: the observations $y_1, ..., y_n$ provides information about the unknown value of θ and, through θ, on the value of the next observation Y_{n+1}. Thus, Y_{n+1} depends, in a probabilistic sense, on the past observations $Y_1, ..., Y_n$. The predictive density can be computed as

$$
\begin{aligned}
\pi(y_{n+1}|y_{1:n}) &= \int \pi(y_{n+1}, \theta|y_{1:n})d\theta \\
&= \int \pi(y_{n+1}|\theta, y_{1:n})\pi(\theta|y_{1:n})d\theta \\
&= \int \pi(y_{n+1}|\theta)\pi(\theta|y_{1:n})d\theta
\end{aligned} \tag{6.5}
$$

where $\pi(\theta|y_{1:n})$ is the posterior density of θ, conditioning on the data $y_1, ..., y_n$. As we can see, the posterior density can be computed via Bayes formula,

$$\pi(\theta|y_{1:n}) = \frac{\pi(y_{1:n}|\theta)\pi(\theta)}{\pi(y_{1:n})} \propto \prod_{t=1}^{n} \pi(y_t|\theta)\pi(\theta)$$

Note that the denominator $\pi(y_{1:n})$ is a normalized constant which does not depend on θ. Therefore, the posterior density could be computed *recursively*. At time $(n-1)$, the

information available about θ is described by conditional density

$$\pi(\theta|y_{1:n-1}) \propto \prod_{t=1}^{n-1} \pi(y_t|\theta)\pi(\theta)$$

This density plays the role of *prior* when we move to time n. Once the new information about y_n becomes available, we just have to compute the likelihood function, which is $\pi(y_n|\theta, y_{1:n-1}) = \pi(y_n|\theta)$ by the assumption of conditional independence, and update the *prior* $\pi(y_{1:n-1})$ by Bayes rule, obtaining

$$\pi(\theta|y_{1:n-1}, y_n) \propto \pi(\theta|y_{1:n-1})\pi(y_n|\theta) \propto \prod_{t=1}^{n-1} \pi(y_t|\theta)\pi(\theta)\pi(y_n|\theta)$$

This recursive structure is critical in time series analysis. In contrast to non-Bayesian time series approaches, Bayesian analysis offers more flexibility and requires no model fitting with all past and current information due to the fact the past information has all been incorporated in its conditional densities. We "update" density functions with newly available information.

6.3 Dynamic Bayesian Model

We develop a Dynamic Bayesian Factor Model (DBFM) which incorporates multivariate dynamic factors, and time-varying patterns that arise from the microstructure settings. The DBFM, in basic form, models dynamic multivariate factors such as economic conditions, hierarchical structures, common time-varying component in multiple time series etc; in its advanced form, it models time-varying volatilities like stochastic volatility (SV) or ARCH/GARCH.

The dynamic factor model is an extension to dynamic linear model. There has been an increasing interest in the application of dynamic linear model (henceforth DLM),

or state space models in time series analysis. They allow a natural interpretation of a time series as the combination of several components, such as trend, regressive or volatility components. At the same time, the computation can be implemented by recursive algorithms. The problems of estimation and forecasting are solved by recursively computing the conditional distribution of the quantities of interest, given the available information. Such desirable features are quite naturally treated in a Bayesian framework.

Dynamic factor models can be used to model univariate or multivariate time series, also in the presence of non-stationary, structural changes, and irregular patterns.

We present the DLM in its generalized matrix notation.

Definition 5. *The dynamic linear model, in its basic form, employs the following relationship,*

$$Y_t = F_t'\theta_t + \nu_t \quad \nu_t \sim N[0, V_t] \tag{6.6}$$

$$\theta_t = G_t\theta_{t-1} + \omega_t \quad \omega_t \sim N[0, W_t] \tag{6.7}$$

$$(\theta_0|D_0) \sim N[m_0, c_0] \tag{6.8}$$

where ν_t and ω_t are independent. m_0 and c_0 are some prior moments.

Denote Y_t a (r x 1) vector of observations on the time series over time $t = 1, 2, ...n$. F_t is a known (n x r) matrix for regression. G_t is a known (n x n) matrix for state space evolution. V_t is a (r x r) matrix, and W_t is a (n x n) matrix for variance-covariance.

Equation (6.6) is called observation equation, which defines sampling distribution for Y_t conditional on θ_t. Given θ_t, Y_t is independent of all other observations and states. $F_t'\theta_t$ is the mean response, and ν_t is the observation error.

Equation (6.7) is called system, or evolution equation, which defines the time evolution of the state vector. The conditional independence structure of state space shows

that, given θ_{t-1} and the known value of G_t and W_t, the conditional distribution of θ_t is fully independent of Y_{t-1}. The system error is ω_t.

Consider a univariate DLM model that is closed to external information at time $t \geq 1$, such that given initial prior information D_0 at $t = 0$, at any future time t, the available information set is $D_t = \{Y_t, D_{t-1}\}$, where Y_t is the observed value at time t.

At any time t, for all $k \geq 0$, define,

$$
\begin{aligned}
a_t(k) &= E(\theta_{t+k}|y_{1:t}) & (6.9)\\
R_t(k) &= Var(\theta_{t+k}|y_{1:t}) & (6.10)\\
f_t(k) &= E(Y_{t+k}|y_{1:t}) = E[F'_{t+k}\theta_{t+k}|y_{1:t}], & (6.11)\\
Q_t(k) &= Var(Y_{t+k}|y_{1:t}) & (6.12)
\end{aligned}
$$

The Forecast Function is $f_t(k)$. For k strictly greater than 0, the forecast function provides the expected values of future observations given current information.

$$
f_t(k) = E[Y_{t+k}|D_t] \quad \text{for all } k \geq 1 \tag{6.13}
$$

Our model is a fully dynamic multivariate factors representation of market microstructure models by the integration of microstructure settings from Kyle's framework.

6.4 The Model

We continue the derivation from previous chapter, and present formal definitions of our time series model.

Consider an asset that pays off, $v \sim N(p_0, \Sigma_0)$, p_0 is the expected value, and Σ_0 is the value uncertainty at the beginning of the trade. There are multiple trading

periods, with $n = 1, ..., N$, and $t = 0, ..., 1$. The quantity traded by noise traders is denoted by Δu_n, where $\Delta u_n \sim N(0, \sigma_u^2 \Delta t_n)$. The informed trader observes true value v. The informed trader maximizes his expected profit by choosing optimal trading strategy Δx_n. It is assumed that v, Δu_n are all independent. There are competitive risk-neutral market-makers, who set the prices of asset equal to the expected value of the liquidation values conditional on the information about the quantities traded by the strategic insider trader and noise traders, $\Delta y_n = \Delta x_n + \Delta u_n$. We also assume $\Delta t_n = 1/N$ for all n. The model is specified as follows,

$$\text{order flows}: \quad \Delta y_n = \beta_n (v - p_{n-1}) \Delta t_n + \Delta u_n, \quad (6.14)$$

$$\beta_n = \beta_{n-1} + \mu_{n-1} + \partial \beta_n \quad (6.15)$$

$$\mu_n = \mu_{n-1} + \partial \mu_n \quad (6.16)$$

$$\Delta u_n \sim N(0, \sigma_u^2 \Delta t_n) \quad (6.17)$$

$$\partial \beta_n \sim N(0, \sigma_\beta^2) \quad (6.18)$$

$$\partial \mu_n \sim N(0, \sigma_\mu^2) \quad (6.19)$$

$$\text{price}: \quad \Delta p_n = p_n - p_{n-1} = \lambda_n \Delta y_n + \epsilon_n \quad (6.20)$$

$$\lambda_n = \lambda_{n-1} + \partial \lambda_n \quad (6.21)$$

$$\epsilon_n \sim N(0, \sigma_\epsilon^2) \quad (6.22)$$

$$\partial \lambda_n \sim N(0, \sigma_\phi^2) \quad (6.23)$$

where β_n, μ_n and λ_n are dynamic components, $\partial \beta_n$, $\partial \mu_n$ and $\partial \lambda_n$ are stochastic terms which are assumed to be uncorrelated. In our model, μ_n has the interpretation of incremental growth in the level of the series over the time interval $n - 1$ to n, evolving during that interval according to the addition of the stochastic element $\partial \mu_n$. The level β_n at time n evolves systematically via the addition of the growth μ_n and undergoes a

further stochastic shift via the addition of $\partial \beta_t$.

The state vector is defined as $\theta_n = (\beta_n, \mu_n)'$, $n = 1, ..., N$. Denoting $m_{n-1} = (\hat{\beta}_{n-1}, \hat{\mu}_{n-1})'$, where $\hat{\beta}_{n-1}$ and $\hat{\mu}_{n-1}$ are estimates at time $n - 1$. The *one-step ahead* state estimates and point forecasts are given by,

$$a_n = G_n m_{n-1} = \begin{bmatrix} \hat{\beta}_{n-1} + \hat{\mu}_{n-1} \\ \hat{\mu}_{n-1} \end{bmatrix} \tag{6.24}$$

$$f_n = F_n a_n = (\hat{\beta}_{n-1} + \hat{\mu}_{n-1})(v - p_{n-1})\Delta t_n \tag{6.25}$$

The *kth*-period ahead state forecast is:

$$a_n(k) = (\hat{\beta}_n + k \; \hat{\mu}_n) \tag{6.26}$$

which is a linear function of k.

The model is always observable, since the observability matrix is always full-rank, with Ω:

$$\Omega = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{6.27}$$

Therefore, given $\sigma_\mu > 0$, there exist limiting values for R_n, C_n, and A_n (see West and Harrison; 1997). In particular, the gain A_n converges to a constant matrix $A = (A_1 \; A_2)'$ (Theorem 7.2).

Given the existence of limiting values, we obtain the *asymptotic properties* of updating equations. Writing $e_n = Y_n - f_n$, the updating equations in terms of limiting

values:

$$Y_n = (\hat{\beta}_{n-1} + \hat{\mu}_{n-1})(v - p_{n-1})\Delta t + e_n \tag{6.28}$$

$$\hat{\beta}_n = \hat{\beta}_{n-1} + \hat{\mu}_{n-1} + A_1 \, e_n \tag{6.29}$$

$$\hat{\mu}_n = \hat{\mu}_{n-1} + A_2 \, e_n \tag{6.30}$$

With a limiting second difference equation

$$Y_n - 2Y_{n-1} + Y_{n-2} = e_n - \phi_1 e_{n-1} + \phi_2 e_{n-2} \tag{6.31}$$

where

$$\phi_1 = 2 - A_1 - A_2 \tag{6.32}$$

$$\phi_2 = 1 - A1 \tag{6.33}$$

This can be written in terms of backshift operator B as

$$(1 - B)^2 Y_n = (1 - \phi_1 B + \phi_2 B^2)e_n \tag{6.34}$$

There are a number of point prediction methods which take the form of equation (6.31). Such methods include Box and Jenkins' predictor, Exponential weighted regression, and Holt's liner methods. Our inference employs Bayesian MCMC methods.

6.5 Markov Chain Monte Carlo (MCMC)

In Bayesian analysis, two types of strategies are used in the summarization of the posterior distributions. If the sampling density has a familiar functional form, such as a member of an exponential family, and a *conjugate prior* is chosen for the parameter,

then the posterior distribution is often expressed in terms of conjugate probability distributions. In this case, conjugate priors permit posterior densities to emerge without numerical integration. However, the easy calculations of this specification comes with a price due to the restrictions they impose on the form of the prior. A second type of computation strategy is called *simulation-based* methods. In many cases, it is unlikely that the conjugate prior is an adequate representation of the prior state of knowledge, and the posterior distribution is not a familiar functional form. In such cases, the asymptotic approximations or Monte Carlo methods are required.

In terms of simulation-based methods, rejection sampling with a suitable choice of proposal density is a general method for simulating from an arbitrary posterior distribution. Importance sampling and sampling importance re-sampling (SIR) are alternative methods for computing integrals and simulating from a general posterior distribution. The SIR algorithm is especially useful when one wishes to integrate the sensitivity of a posterior distribution with respect to changes in the prior and likelihood functions.

Monte Carlo integration and posterior approximation via rejection sampling or importance sampling involve direct simulation from an sampling distribution, usually viewed as an approximation to the true density $p(X)$. However, when the dimension of the model becomes large, both rejection sampling and importance sampling can be difficult to setup because they require the construction of a suitable proposal density. Markov Chain Monte Carlo (MCMC) algorithm becomes the only way to approximate the posterior density in high-dimensional problems. The idea of MCMC sampling was first introduced by Metropolis, Rosenbluth, Teller (1953) and was subsequently generalized by Hasting (1970). A general and detailed statistical theory of MCMC methods can be found in Tierney (1994).

The MCMC sampling strategy relies on the construction of a Markov chain with

70

realizations $\theta^{[0]}, \theta^{[1]}, ..., \theta^{[j]}, ...$ in the parameter space Θ. Under appropriate regularity conditions (see Tierney 1994), asymptotic results guarantee that as j tends to infinity, then $\theta^{[j]}$ converge in distribution to $p(\theta|y)$. Hence, the realized values of the chain can be used to make inference about the joint posterior. All we required are algorithms to construct the chains.

We first introduce Gibbs sampler algorithm, and then present the MCMC analysis on two Bayesian models: the original Kyle's model, and our dynamic model. Our simulation studies are based on the new algorithms we developed from the previous chapter.

6.5.1   The Gibbs sampler

The Gibbs sampler is an algorithm based on successful generations from the full conditional densities. An elementary exploration can be found in Casella and George (1992). See also Gelfand and Smith (1990), Tanner and Wong (1987). The full condition density $p(\theta_i|\theta_{\neq i}, y)$ is the posterior density of the ith element of parameter $\theta \doteq (\theta_1, ..., \theta_d)'$, d is the total dimension of parameter θ, given all other elements, where elements of θ can be scalars or sub-vectors. The Gibbs algorithm works as follows:

1. Initialize the iteration counter of the chain to $j = 1$ and
   set the initial value $\theta^{[0]} \doteq (\theta_1^{[0]}, ... \theta_d^{[0]})'$

2. Generate a new value $\theta^{[j]}$ from $\theta^{[j-1]}$ through successive
   generation values:
   $$\theta_1^{[j]} \sim p(\theta_1|\theta_{\neq 1}^{[j-1]}, y)$$
   $$\theta_2^{[j]} \sim p(\theta_2|\theta_1^{[j]}, \theta_3^{[j-1]}, ..., \theta_d^{[j-1]}, y)$$
   $$\vdots$$
   $$\theta_d^{[j]} \sim p(\theta_d|\theta_{\neq d}^{[j]}, y)$$

3. Change counter j to $j+1$ and go back to step 2 until convergence is reached.

As the number of iterations increases, the chain approaches its stationary distribution and convergence is then assumed to hold approximately (Tierney 1994). Sufficient conditions for the convergence of the Gibbs sampler are given in Roberts and Smith (1994). As noted in Chib and Greenberg (1996), these conditions ensure that each full conditional density is well defined and that the support of the joint posterior is not separated into disjoint regions since this would prevent exploration of the full parameter space.

The Gibbs sampler is most frequently used MCMC sampling strategy when it is easy to write down full conditional densities from which we could generate draws. When the expression $p(\theta_i|\theta_{\neq i}, y)$ has no functional form, we might consider reject methods (Ripley 1987), the Giddy-Gibbs sampler when θ_i is univariate (Ritter and Tanner 1982), adaptive rejection sampling (Gilks and Wild 1992) or Metroplis-Hasting Algorithm.

6.5.2 MCMC for original Kyle's model

The original Kyle's model has exact solutions to a set of difference equations. We derived new methods of solving these difference equations which boils down to solving just one cubic equation on $\{d_n\}$ sequence, where $d_n = \alpha_n \lambda_n$.

The parameter sequence $\{d_n\}$ depend on the choice of N, the number of periods, and independent on all other parameters. Given $\{d_n\}$ values, we could solve all other parameters exactly *once for all* instead of recursive approximation. Consequently, we only requires two priors for original Kyle's model.

Original Kyle's model can be written as:

$$\Delta u_n \quad \sim \quad N(0, \sigma_u^2 \Delta t_n), \tag{6.35}$$

$$\Delta y_n \quad = \quad \beta_n \; (v - p_{n-1}) \; \Delta t_n + \Delta u_n, \tag{6.36}$$

$$p_n \quad = \quad p_{n-1} + \lambda_n (\Delta x_n + \Delta u_n), \tag{6.37}$$

$$\Sigma_n \quad = \quad Var(v | \mathcal{F}_n^U), \tag{6.38}$$

where $\Delta t_n = \frac{1}{N}, \forall \; n = 1, ..., N$.

In order to write the posterior likelihood function, we define $\psi \doteq (\Sigma_0, \sigma_u^2)$. And vectors $Y_n \doteq (\Delta y_1, \Delta y_2, ..., \Delta y_n)'$, and $X_n \doteq \{(v - p_0)\Delta t_1, (v - p_1)\Delta t_2, ..., (v - p_{n-1})\Delta t_n\}'$. X_n and Y_n are observed information at time n. Given the initial prior information tion D_0 at time $t = 0$, at any future time, the available information set is thus $D_n = \{D_{n-1}, X_n, Y_n\}$.

Define $\Theta_n = (\theta_0, \theta_1, ...\theta_n)$, which includes all latent variables, in our case, $\theta \doteq \{\lambda, \beta, \alpha, \delta\}$. The conditional posterior function of ψ can be expressed as follows:

$$p(\psi | \Theta_n, D_n) = p(\Sigma_0 | \Theta_n, D_n) p(\sigma_u^2 | \Theta_n, D_n) \tag{6.39}$$

this implies the independence between two priors.

The likelihood function,

$$\ell(\sigma_u^2 | \Theta_n, D_n) \propto p(\sigma_u^2 | D_0)(\sigma_u)^{-n} \exp\left[-\frac{n}{2\sigma_u^2}(Y - X\beta)'(Y - X\beta) \right] \tag{6.40}$$

We propose the following *priors* on the parameters Σ_0 and σ_u^2 of the preceding

model:

$$\sigma_u^2 \quad \sim \quad IG(\alpha, \beta) \tag{6.41}$$

$$\Sigma_0 \quad \sim \quad N(\mu, \sigma^2) I_{\{\Sigma_0 > 0\}} \tag{6.42}$$

where $\alpha, \beta, \mu, \sigma^2$ are hyperparameters, $I_{\{.\}}$ is the indicate function which equals unity if the constraints holds and zero otherwise. The prior σ_u^2 has inverted gamma distribution, while Σ_0 has improper prior distribution since Σ_0 follows exact equilibrium solutions given the available information set. And $p(\psi) = p(\sigma_u^2)p(\Sigma_0)$.

Then we construct the joint posterior densities:

$$p(\psi|\Theta_n, D_n) \propto \ell(\psi|\Theta_n, D_n)p(\psi) \tag{6.43}$$

6.5.3   MCMC for our dynamic Bayesian model

In this subsection, we provide derivations of MCMC via our dynamic model. Our proposed model provides more insights about market behaviors when we carry out empirical studies in the next chapter.

The state vector for period n is defined as

$$\theta_n = (\beta_n, \mu_n)' \tag{6.44}$$

where $n = 1, ..., N$. Write $\Theta_n = \{\theta_0, \theta_1, ...\theta_n\}$ for all n state vectors. This is sometimes called *latent variable.*

We obtain,

$$\Delta u_n \quad \sim \quad N(0, V), \tag{6.45}$$

$$\omega_n \doteq (\partial \beta_n, \partial \mu_n)' \quad \sim \quad N(0, W) \tag{6.46}$$

Where $V = \sigma_u^2 \Delta t_n$, and W is the variance-covariance matrix of vector $(\partial\beta_n, \partial\mu_n)'$. Both V and W are constant and independent over time n. Define model *parameters*,

$$\psi = \{V, W\} \tag{6.47}$$

We also define the information set at time n as D_n. Let vectors $Y_n \doteq (\Delta y_1, \Delta y_2, ..., \Delta y_n)'$, and $X_n \doteq \{(v - p_0)\Delta t_1, (v - p_1)\Delta t_2, ..., (v - p_{n-1})\Delta t_n\}'$. X_n and Y_n are observed data at time n. Given the initial prior information $D_0 = \theta_0$ at time $t = 0$, at any future time, the available information set is thus $D_n = \{D_{n-1}, X_n, Y_n\}$.

This structure is evident in considering MCMC based on Gibbs sampling. Gibbs sampling suggests the full posterior distribution:

$$p(\Theta_n, \psi | D_n) \tag{6.48}$$

may be simulated by iterating between two conditional posteriors

$$p(\Theta_n | \psi, D_n) \leftrightarrow p(\psi | \Theta_n, D_n) \tag{6.49}$$

Note that this procedure is different from original Gibbs sampling. This involves sampling Θ_n, and separately ψ, from their full, multivariate conditional posteriors, rather than sequencing through individual scalar elements as in the original definition of Gibbs sampling.

Gibbs sampling procedure:

1. Simulating $p(\Theta_n | \psi, D_n)$

   This procedure is standard. Based on the observation that the analysis is conditioning on ψ, simulation of LHS of (6.49) will be accessible as a multivariate normal distribution.

2. Simulating $p(\psi|\Theta_n, D_n)$

   This procedure depends on the model form, we will illustrate in details.

3. Repeat these steps until convergence

In our proposed model, the conditional posterior for ψ is

$$p(\psi|\Theta_n, D_n) = p(V|\Theta_n, D_n)p(W|\Theta_n, D_n) \tag{6.50}$$

We assume prior independence between parameter V and W.

It can be shown that, the posterior likelihood function for V is:

$$\ell(V|\Theta_n, D_n) \propto p(V|D_0)(V)^{-n/2} \exp\left[-\frac{n}{2V}(Y - X\beta)'(Y - X\beta)\right] \tag{6.51}$$

and the posterior likelihood function for W,

$$\ell(W|\Theta_n, D_n) \propto p(W|D_0)|W|^{-n/2} \exp\left[-\frac{n}{2} \; trace(\hat{W}_n W^{-1})\right] \tag{6.52}$$

where $\omega_n = \theta_n - G_t\theta_{n-1} = (\omega_{n,1}, \omega_{n,2})'$ for each n, and $\hat{W}_n = n^{-1}\sum_{t=1}^{n} \omega_t\omega_t'$.

We propose the following *priors* on the parameters $\psi = \{V, W\}$ and $W = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}$:

$$V \;\sim\; IG(a_1, b_1) \tag{6.53}$$

$$W_1 \;\sim\; IG(a_2, b_2) \tag{6.54}$$

$$W_2 \;\sim\; IG(a_3, b_3) \tag{6.55}$$

where $a_{\{.\}}, b_{\{.\}}$ are hyperparameters. V, W_1 and W_2 have inverted gamma distributions. And $p(\psi) = p(V)p(W)$.

Then we obtain the joint posterior density:

$$p(\psi|\Theta_n, D_n) \propto \ell(\psi|\Theta_n, D_n)p(\psi) \tag{6.56}$$

6.6 Simulation

It is instructive to compare our dynamic models through simulation studies. The target data set is generated from original Kyle's model. Two simulations are carried out on the target data set: one is Bayesian MCMC for our dynamic model, another is MCMC for original Kyle's model.

Recall that the original Kyle's model only depends on two parameters, we first assume $\sigma_u = 0.5$, $\Sigma_0 = 0.4$, $N = 50$, and obtain sequences of all parameters $\{\lambda_n, \alpha_n, \beta_n, \Sigma_n, \delta_n\}$ for $n = 1, ..., N$. Then we take market initial condition at p_0, and simulate the market prices sequence and order imbalance. The order flow sequence consists of aggregated order flows from both the informed trader as well as the uninformed.

Suppose we know the market information i.e., the market price evolution and the aggregated order imbalance, that represent what may be observed in a real market situation. And all model parameters are unknown, we then perform MCMC simulations and infer on model unknowns.

Given the price sequence $\{p_n\}$, and net order flow sequence $\{\Delta y_n\}$, with $n = 1, ..., N$, we run Bayesian MCMC for our proposed model.

In table 6.1, our model has the estimate of posterior mean of $\sigma_u^2 = 0.27315$ which is close to the true value $\sigma_u^2 = 0.25$. The deviance information (DIC = -115.6) suggests good fit (details about Bayesian model selection will be discussed in the next chapter). Figure 6.1 shows MCMC method for our dynamic modeling picks up the trend well; the turning point is near $n = 42$ which is consistent with actual data. Figure 6.2 presents Bayesian results for λ_n. The posterior mean of λ_n is flat throughout and
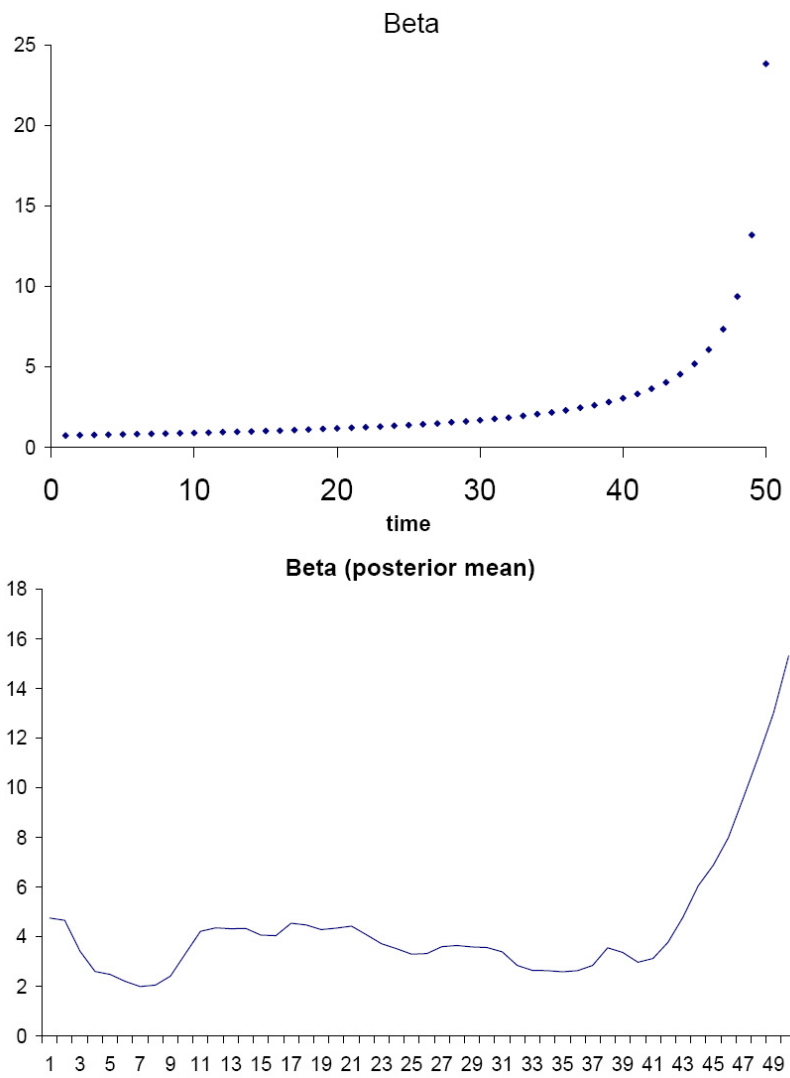
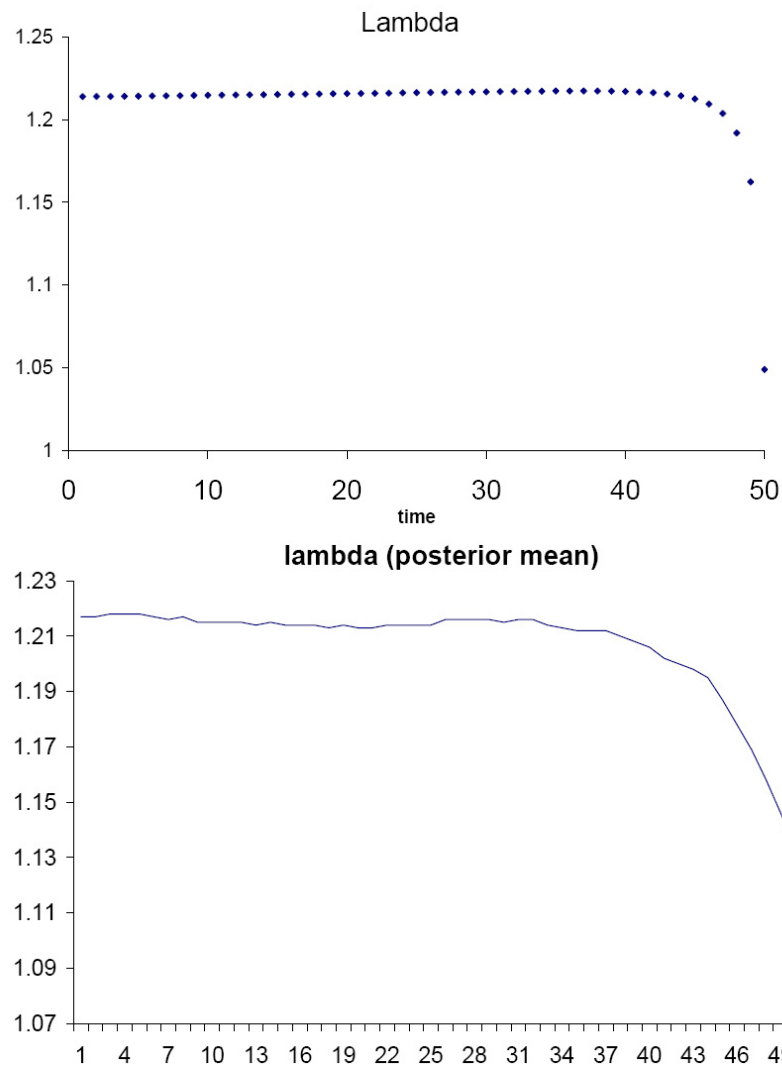Figure 6.1: Insiders' strategy, original vs. MCMC for our model

Figure 6.2: Reciprocal of market depth, original vs. MCMC for our model

Table 6.1: Parameters and Deviance Bayesian results

Parameters	Mean	SD
σ_u^2	0.27315	0.095075
σ_β^2	0.06745	0.122625
σ_μ^2	0.03597	0.053575
σ_ϵ^2	0.00005	3.461E-7
σ_ϕ^2	0.00050	0.000019
deviance	-115.50	0.12940

drop at the end of trading periods as we expect. We notice that the variance term σ_ϵ^2 has fairly small mean 0.00005 and small variance which represents the stochastic term of price equation. This is true since this is how we construct the dataset initially. The data generated from original Kyle's model does not carry stochastic term, the price evolution is proportion to total order imbalance with the price set by the market market. The Bayesian results reflect the relationship well with minor stochastic terms. This assumption may be further investigated when we do empirical studies using real market data set. Figure 6.3 shows results of order flows from our dynamic model. The insider is actively hide his position among noise order flows while he is trading in the direction of his own knowledge.

We then illustrate details of our Bayesian MCMC analysis.

Figure 6.4 displays trace plots of the simulated draws of variance component. Note that we have 6000 iterations with first half burn-in. The simulated draws appear to have reached the main support of the posterior of variance component.

Figure 6.5 shows density plots of parameters. They all have single modal shape which suggest good support for the posterior mean.

One can observe the autocorrelation structure of the sequences by using autocorrelation plots produced in Figure 6.6. The autocorrelations are very close to one for lag one and reduce quickly as a function of the lag.

In the second simulation, we perform MCMC method for original Kyle's model, i.e.,
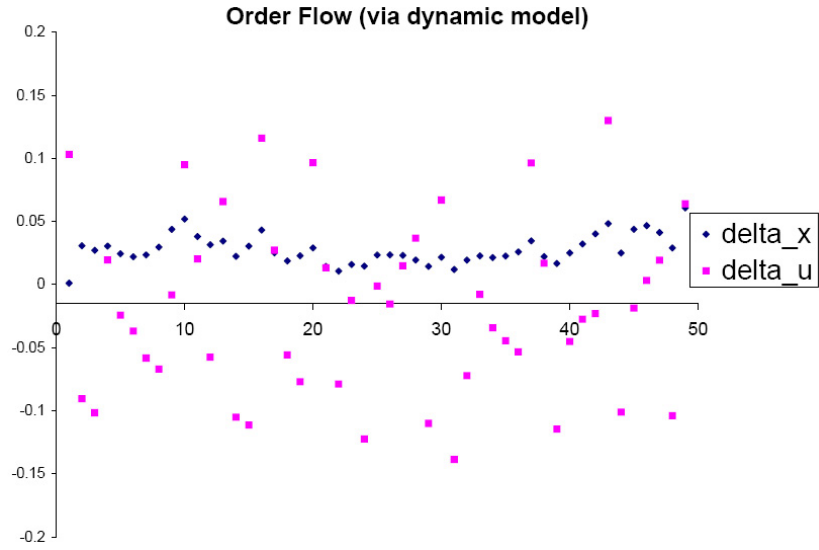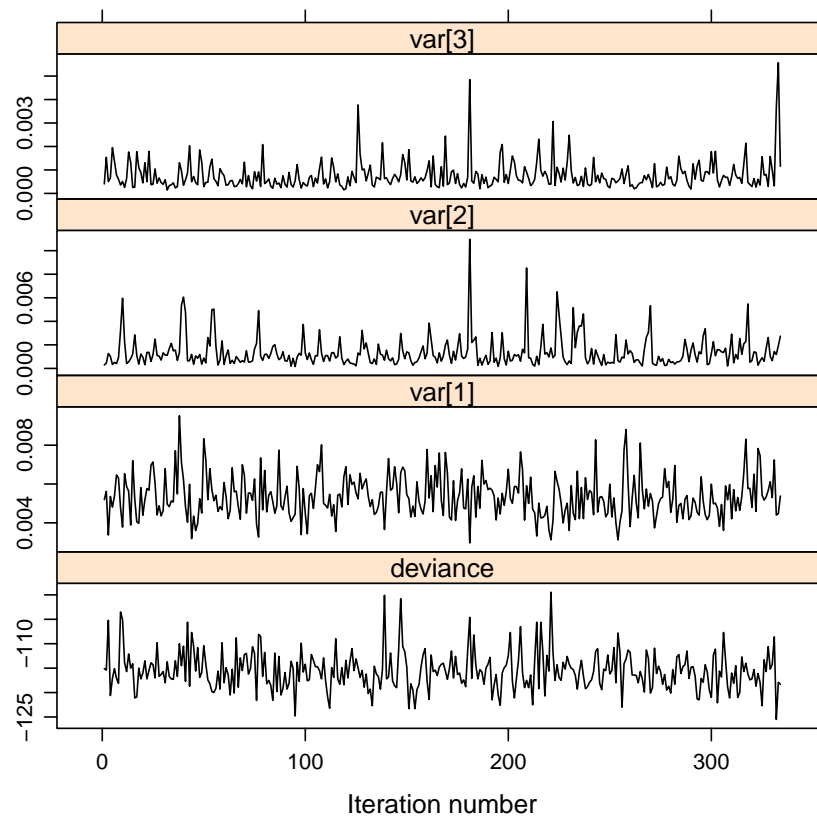
Figure 6.3: Bayesian results of order flows

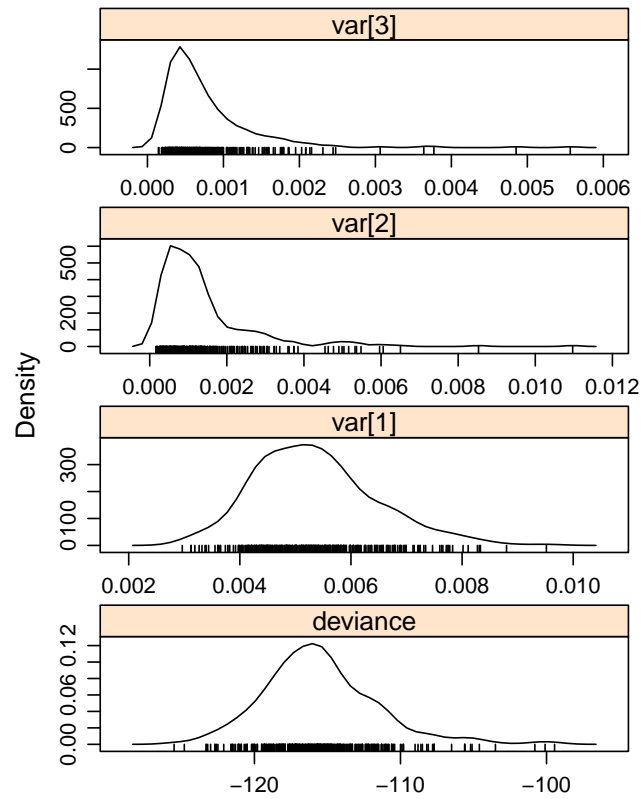Figure 6.4: Parameter and deviance trace plots

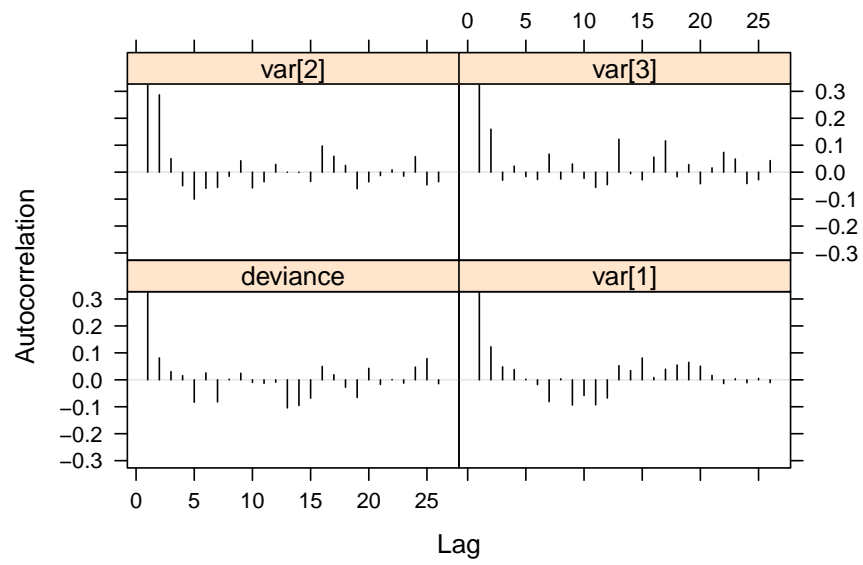Figure 6.5: Parameter and deviance density estimates

Figure 6.6: Parameter and deviance autocorrelation plots

MCMC using two-variable Kyle's model on the same data set we created. The posterior mean for σ_u is 0.5002 vs. actual value of 0.5, and the posterior mean for Σ_0 is 0.3402 vs. true value 0.4. The Bayesian methods provide robust and consistent results in our simulation studies. Since there are only two unknown variables, the model performs well if we can infer on those two variables successfully, and other parameters are derived accordingly. The β_n series shown in Figure 6.7 is at no surprise.

Table 6.2: Parameters and Deviance Bayesian results via original Kyle's model

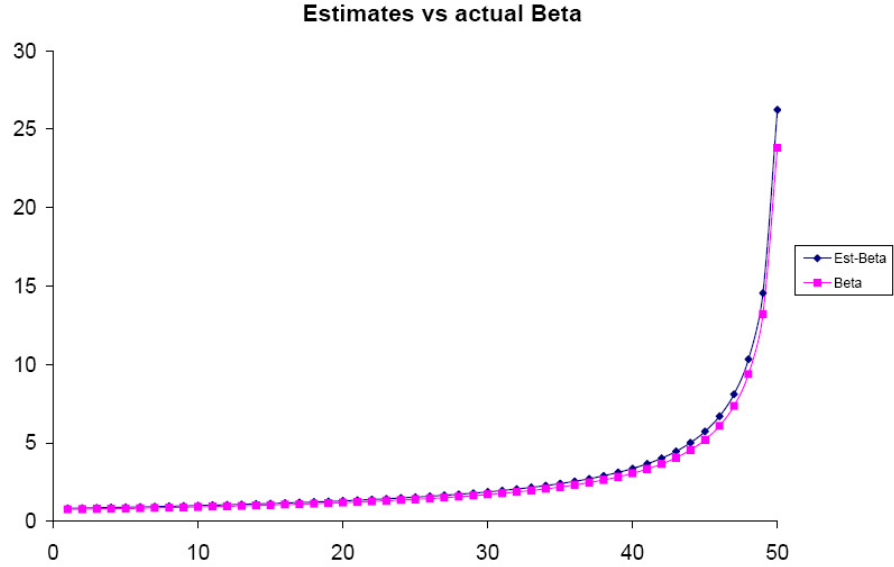| Parameters	Mean	SD	True value
Σ_0	0.3402	0.02768	0.4
σ_u	0.5002	0.0021	0.5
deviance	-114.510	0.07764	

Figure 6.7: MCMC Estimates of beta via original Kyle's model vs. actual beta

Chapter 7

Empirical studies and Bayesian model selection

In previous chapters, we derived our dynamic model and validated through simulation studies, we turn to applications of the model in real market condition. First, we set up our target data. Secondly, we compare dynamic model vs. original kyle's model. We do model inference on different length of real market data in the third section. The empirical results on different stocks and various periods will be discussed in the fourth section. Finally, we summarize Bayesian model diagnosis on different models.

7.1   Data

We study the informed trading prior to the corporate earning release which took place every quarter on NYSE stocks. The intraday, transaction-level data are collected from trade and quote (TAQ) database. The size of intraday data set is enormous because it contains all trades and quotes per second. The daily time frame is regular trading hours from 9:30AM to 4:00PM. Corresponding daily data come from CRSP and quarterly earning data are from COMPUSTAT database. Since every stock has its own earning announcement date and varies by calendar year, we have separate treatment to adjust

different quarterly earning periods (start, end, and number of trading periods) even for the same stock.

All trade data are matched against quote database to find the corresponding quote information. The intra-day data are aggregated to small intervals, in our case, 30 minutes. We use the algorithm mentioned in chapter 3. All trades are translated to signed order imbalance.

The data set are aggregated to different number of periods with observed market information. To study the model performance on the size of data set, we have different length of periods on the same data set. For example, the whole period should span a quarter of high frequency data, which is roughly three-month data depending on two consecutive earning release dates. We also take a subset of the whole data set, for instance, the last one-month data to study the behavior of the model since the original model would expect more intensive activity at the end of trading periods, this is the spirit of original model. We would like to explore the robustness of our dynamic model under various trading periods.

We randomly choose 5 NYSE stocks on four different earning announcement periods. The four sampled periods are: 2003Q4, 2005Q2, 2006Q3 and 2008Q3, that cover sample periods from both boom periods like from year 2003, 2005 and stressed economic environment like from year 2008, that is what we have experienced most recently.

7.2   Dynamic Model vs. original Kyle's Model

We take one NYSE stock, ticker IBM, and study one sample period at 3rd quarter of calendar year 2008. The quarterly data for IBM 2008Q3 consists of roughly three-month intraday data.

Figure 7.1 illustrates dynamic model results for the insider's strategy. It is relatively smooth at the beginning of that period, and starts to trade aggressively at period

$n = 565$ which is around Oct,2 2008. We further look at the trade price evolution during that period and check if there is any irregular pattern. From Figure 7.2, we do not see disconnectivity during the whole 3-month period. The price data exhibits rather smooth pattern while the insiders' strategy has steep slope toward the end of trading periods. The result shows that, the dynamic model produces some information about insiders' strategy which may not be observed or derived by price evolution alone.

We observe our dynamic model performs well under both *positive* and *negative* true value position, i.e., the true value (which is known to the insider) is higher (or lower) than the prevailing market prices. In this example, the insider has negative true value projection, i.e., his knowledge about the true value is lower than market prices, and he should take the short position. The model inference confirms that the insider's strategy is still positive and grow aggressively (short more) over the time.
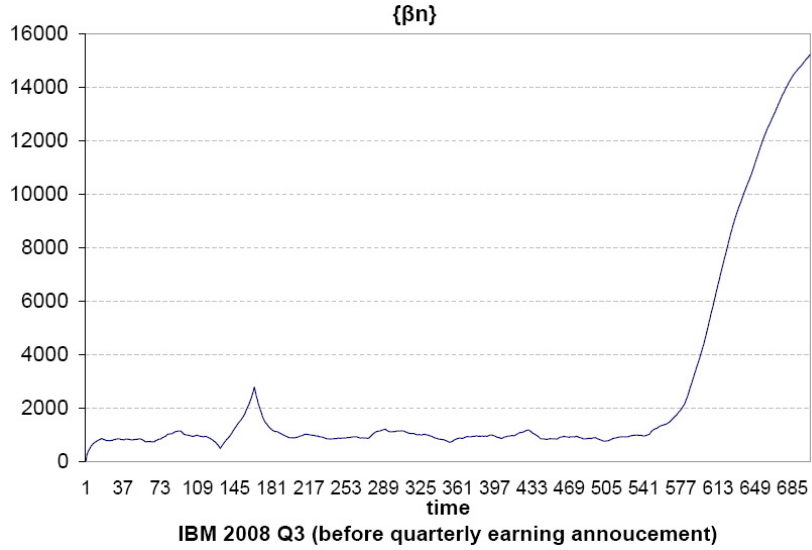

Figure 7.1: Insiders' strategy, MCMC for dynamic model on IBM 2008Q3

We also calculate order flows from our dynamic model. Figure 7.3 illustrates the insider is actively hiding his position among the order flows from noise traders, and he trades more aggressively toward the end of trading periods. Notice that order flows are negative for the insider which means "sell" positions since he is trading in the direction
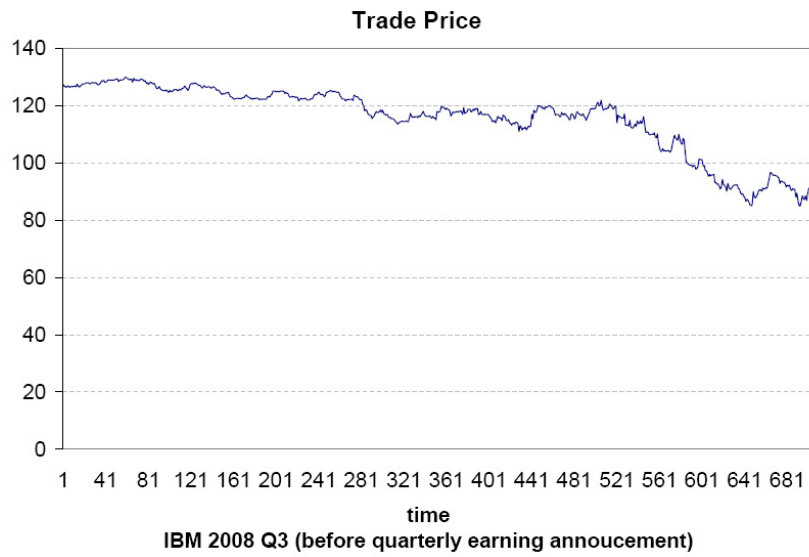
87

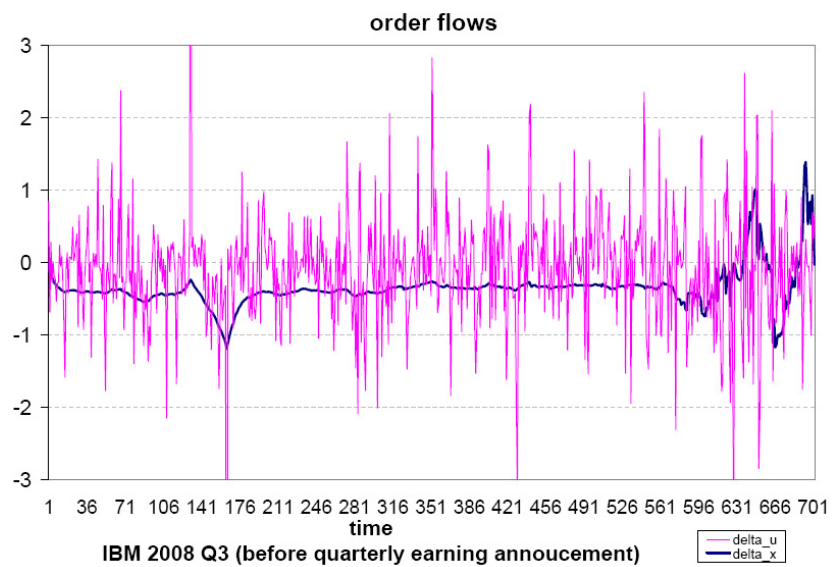Figure 7.2: Trade prices on IBM prior to 2008Q3 earning announcement

Figure 7.3: Order flows, MCMC for our model on IBM 2008Q3

of his knowledge, in this case an unfavorable true value.

Figure 7.4 shows MCMC results of the reciprocal of market depth, i.e., λ_n sequence. The λ_n sequence is little jumpy from the model inference which means the market depth could be noisy on the high frequency data. This parameter reflects the actual market condition that may vary rapidly across the time, such market conditions include macro economic condition, correlation with other industry stocks etc. If we take a further look at the data, we may still discover that the λ_n sequence drops at the end of trading periods.
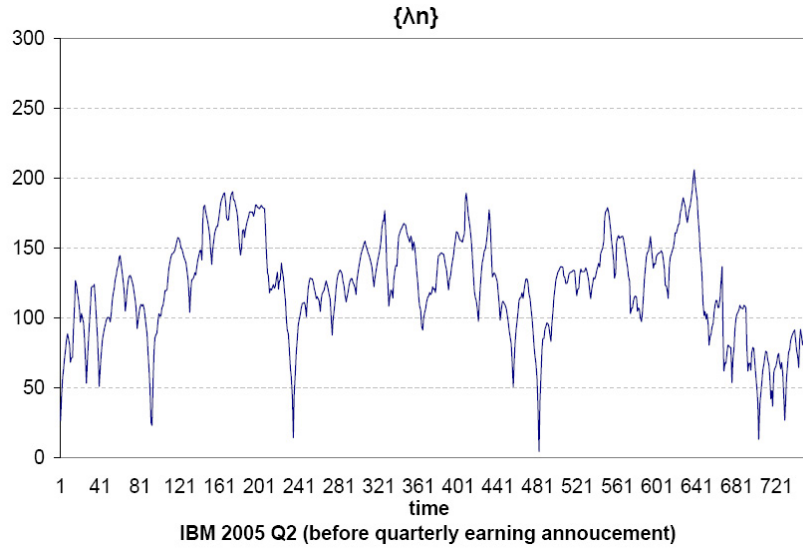

Figure 7.4: Reciprocal of market depth, MCMC for our model on IBM 2005Q2

We then run Bayesian MCMC for original Kyle's model. The parameter results are shown in Table 7.1. The original Kyle's model has deviance of 2324.019.

We present the Bayesian MCMC results for β_n and λ_n using original Kyle's model in Figure 7.5 and 7.6. Since the original model is fully determined by two parameters and number of periods, the upward slope are close to the end of trading periods, in this case, takes place in the last day of trading periods (Oct, 16 2008) before earning announcement. We notice the same pattern as simulated data of Kyle's model.
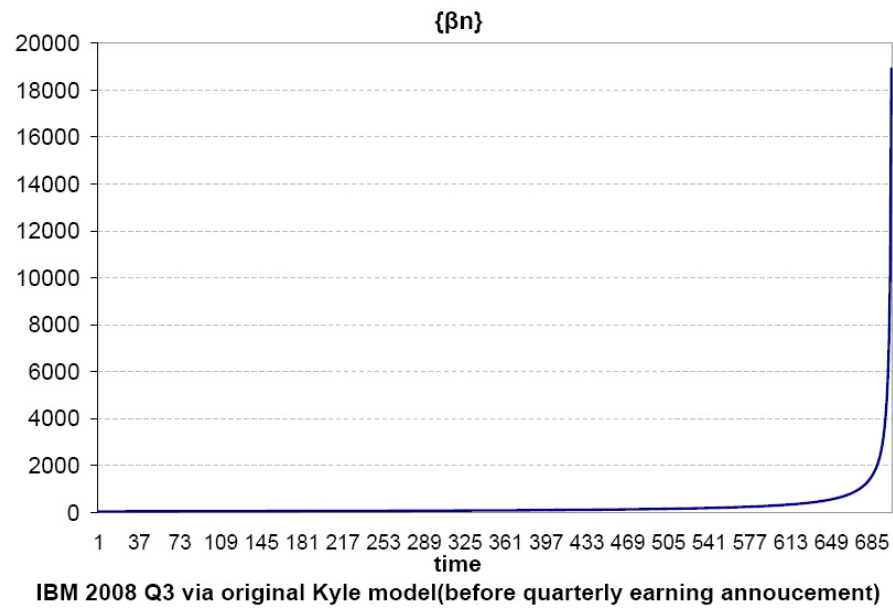
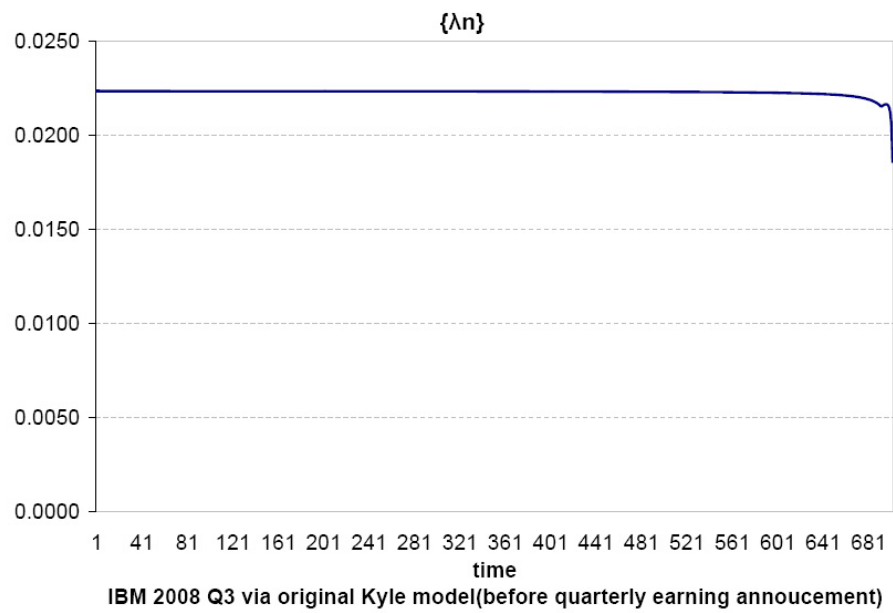Figure 7.5: Beta, MCMC for original Kyle's model on IBM 2008Q3

Figure 7.6: Lambda, MCMC for original Kyle's model on IBM 2008Q3

Table 7.1: Parameters and Deviance results via original Kyle's model on IBM 2008Q3

Parameters	Mean	SD
Σ_0	0.567	0.07693
σ_u	33.636	0.02558
deviance	2324.019	0.06774

7.3 Model fitting with different sample periods

In this section, we compare model performance on various length of periods. In real market condition, we may observe portion of the data set, like one-month of the data close to the earning release, or we may want to explore using subset of the data. The spirit of the original kyle's model is the insider would trade more intensively at the end of the trading periods.

In the first analysis, we observe only last month data for IBM 2008Q3 without knowing what happened in the first two months. And we run Bayesian MCMC for our dynamic model. We observe exactly same pattern as with three-month data, see figure 7.7. The left figure shows results for complete quarter data while the right figure shows only last-month data are applied via Bayesian methods. We then take derivative of estimates and get $m = 1$ at $n = 591$ for quarterly and $n = 151$ for monthly, the first one happens on Oct 2, 2008, 11:30AM, the second one is on Oct 2, 2008, 12:00PM, they are right on with each other.

Table 7.2: MCMC for original Kyle's model using different length of data

Parameters	Mean1Qtr	SD1Qtr	Mean1Mo	SD1Mo
Σ_0	0.567	0.07693	0.474	0.06802
σ_u	33.636	0.02558	14.711	0.01929
deviance	2324.019	0.06774	730.34	0.06901

We then perform Bayesian MCMC for original Kyle's model on the same data sets. Table 7.2 shows the result of MCMC for original Kyle's model using one quarter and one month separately. In Figure 7.8, the corresponding insider strategies are plotted
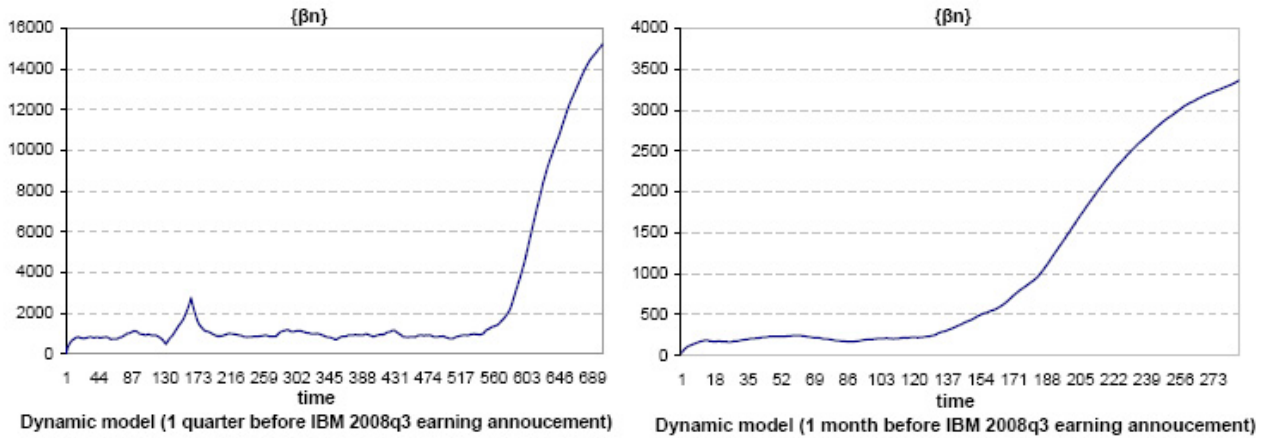
Figure 7.7: MCMC for dynamic model using different length of data

over different periods. We get derivative $m = 1$ at $n = 682$ and $n = 277$ respectively, which is 11:30AM on Oct, 15 2008 and 10:30AM on Oct, 16 2008. They have one-day gap.
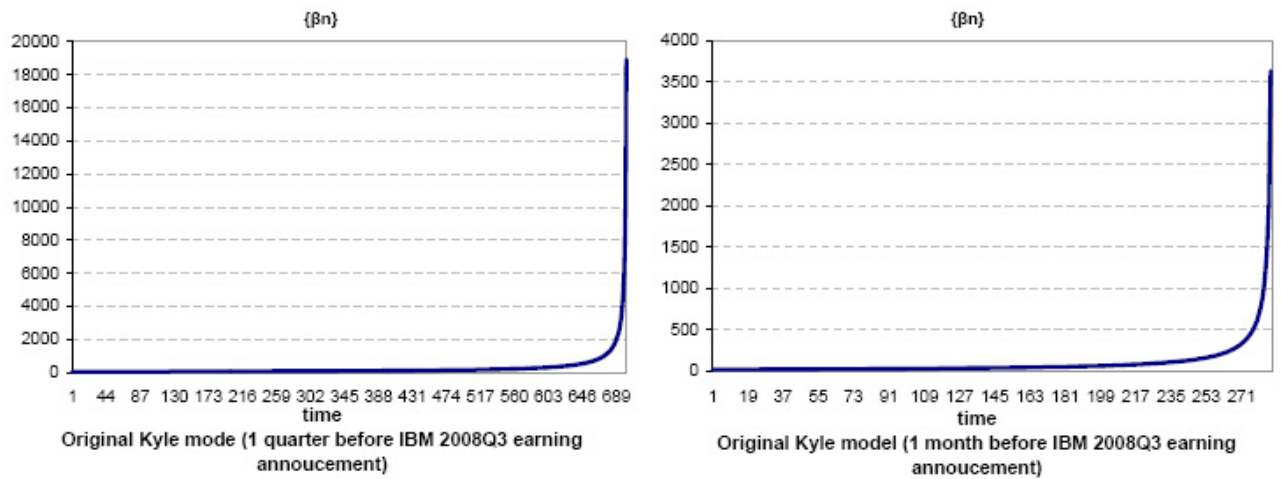
Figure 7.8: MCMC for original Kyle's model using different length of data

We did exercises on other sample periods, and on other stocks, the conclusion remains unchanged. The Bayesian dynamic models provide robust results in terms of various length of periods.

7.4   Sample stocks

We choose 5 randomly sampled stocks on four different periods. The four sampled periods are: 2003Q4, 2005Q2, 2006Q3 and 2008Q3. Each period span the whole period of time between two consecutive earning announcement dates, which is approximately three months. The sampled periods are chosen from favorable economic environment as well as stress environment.

Figure 7.9-7.13 illustrate 5 stocks on four different periods. The overall trend of β_n is increasing as we expected, while the insiders take different approaches to execute their strategies.

Figure 7.9: Results of IBM through four different periods

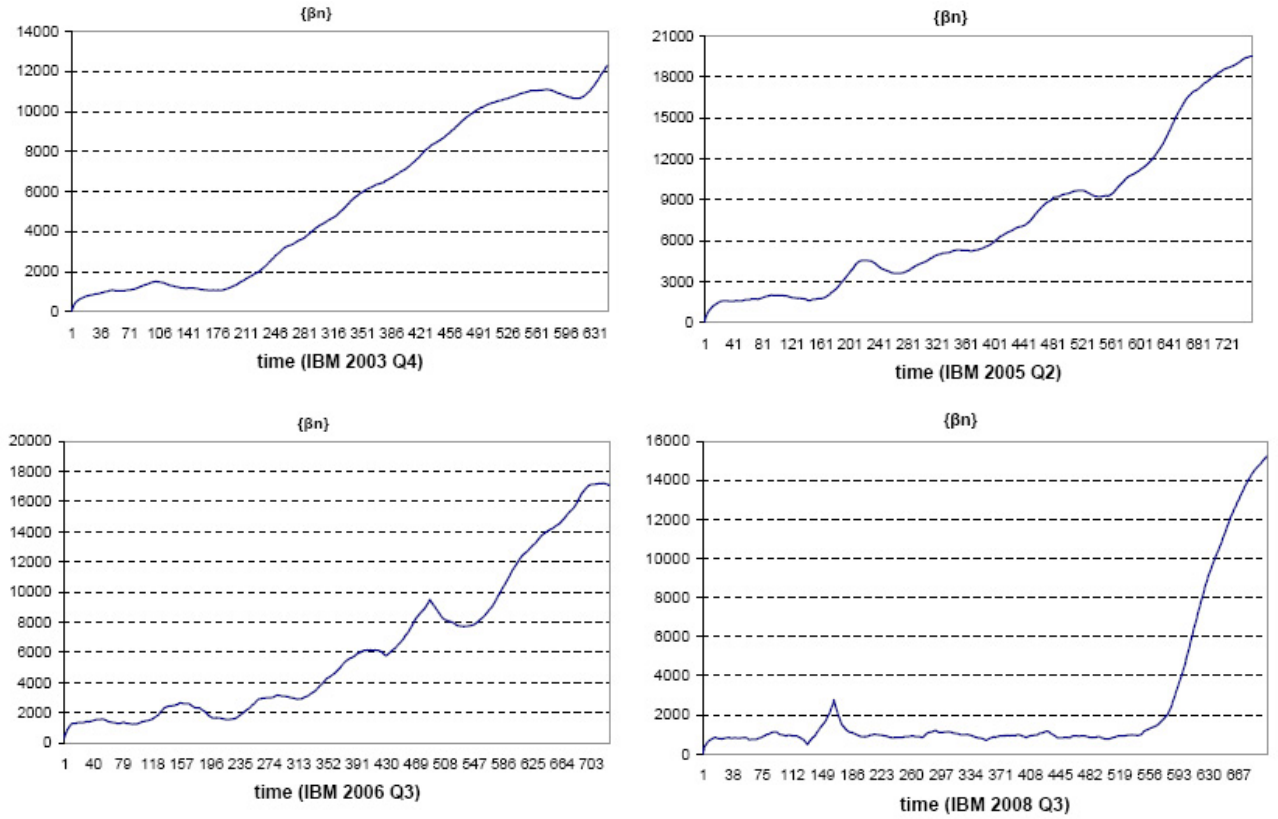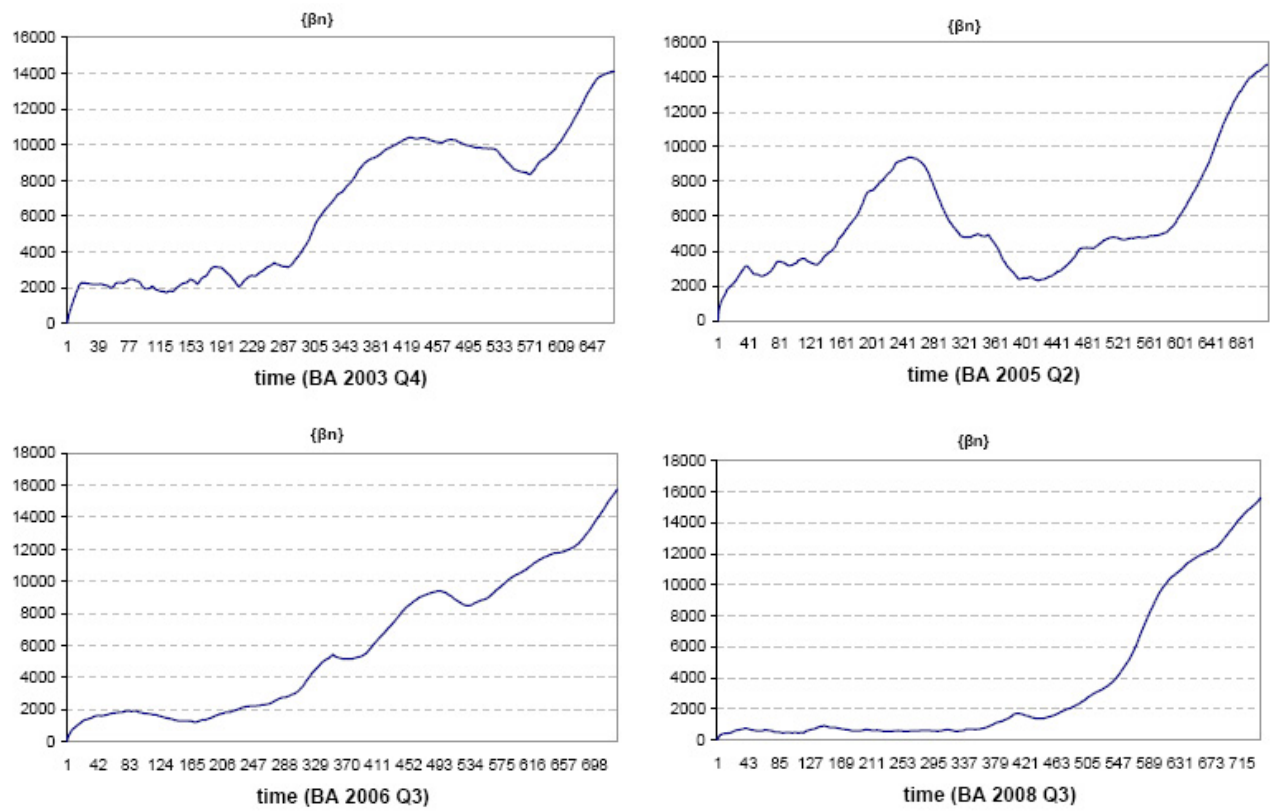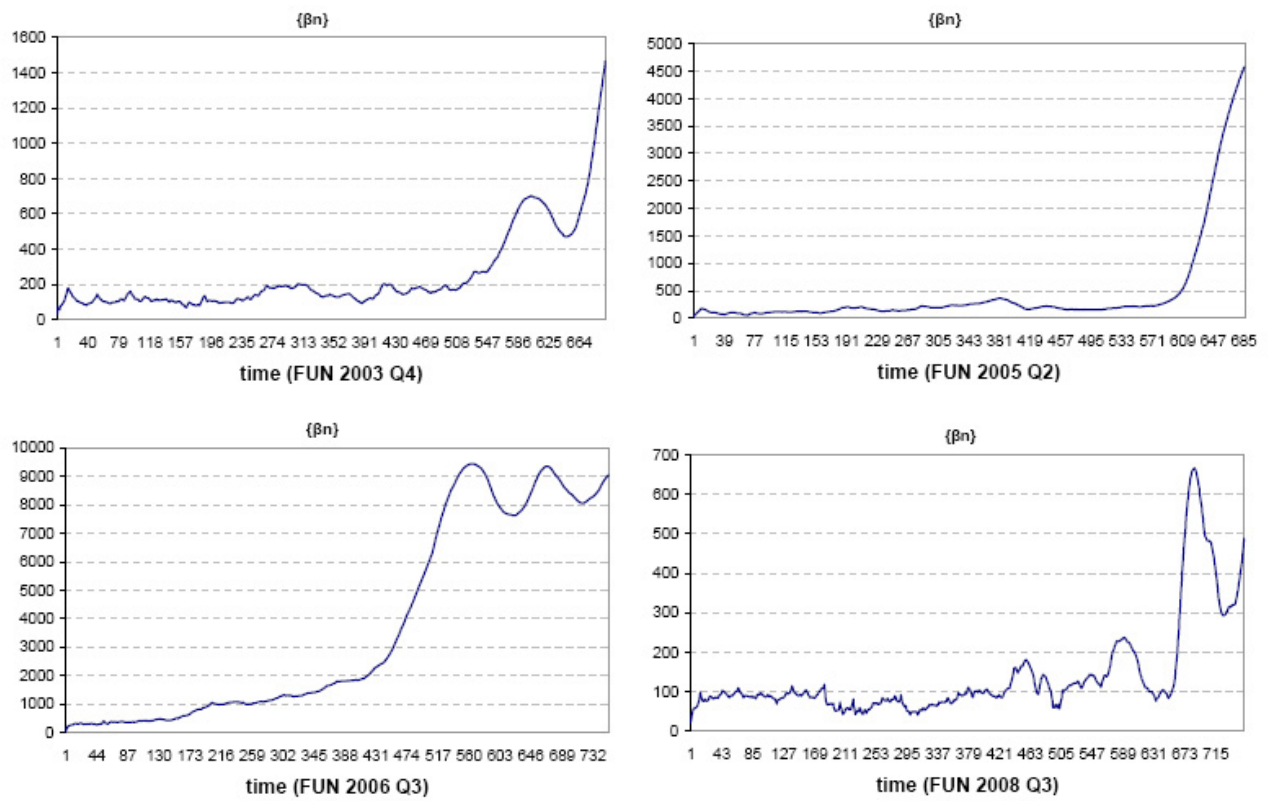Figure 7.10: Results of BA through four different periods

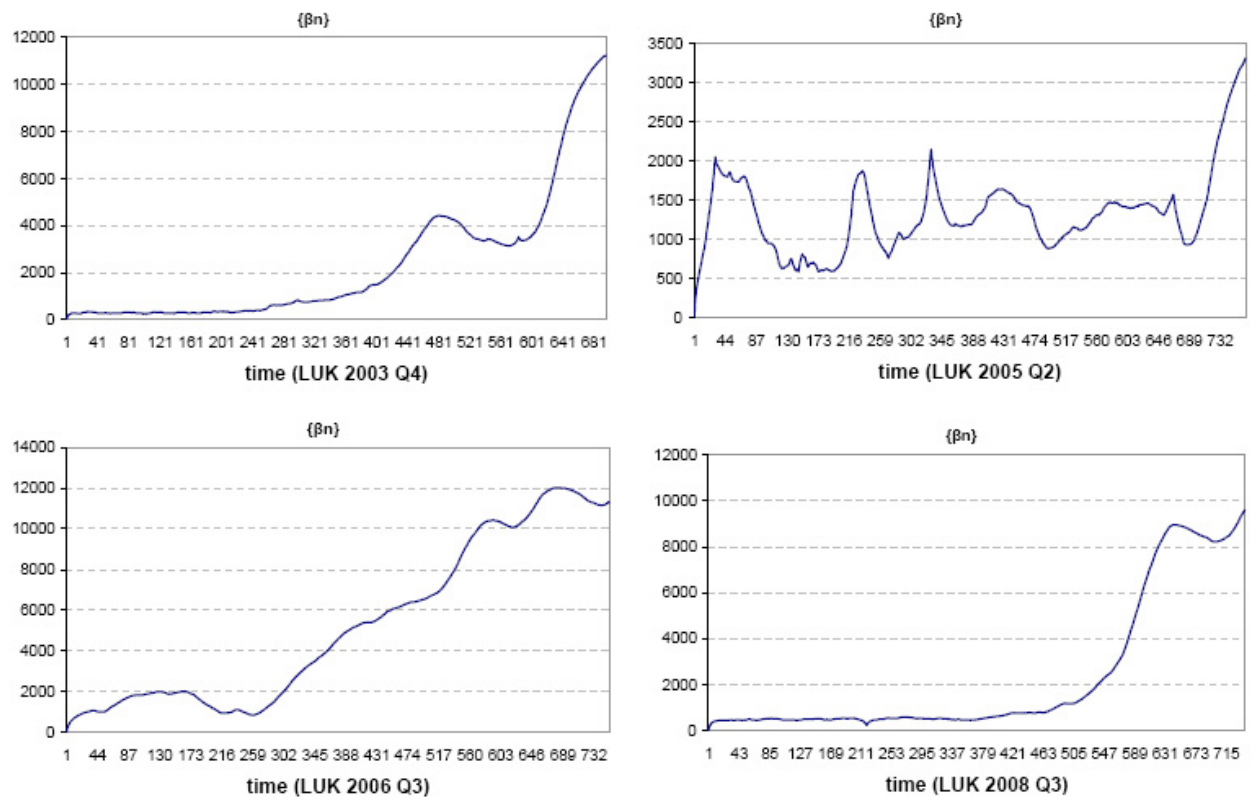Figure 7.11: Results of FUN through four different periods

95

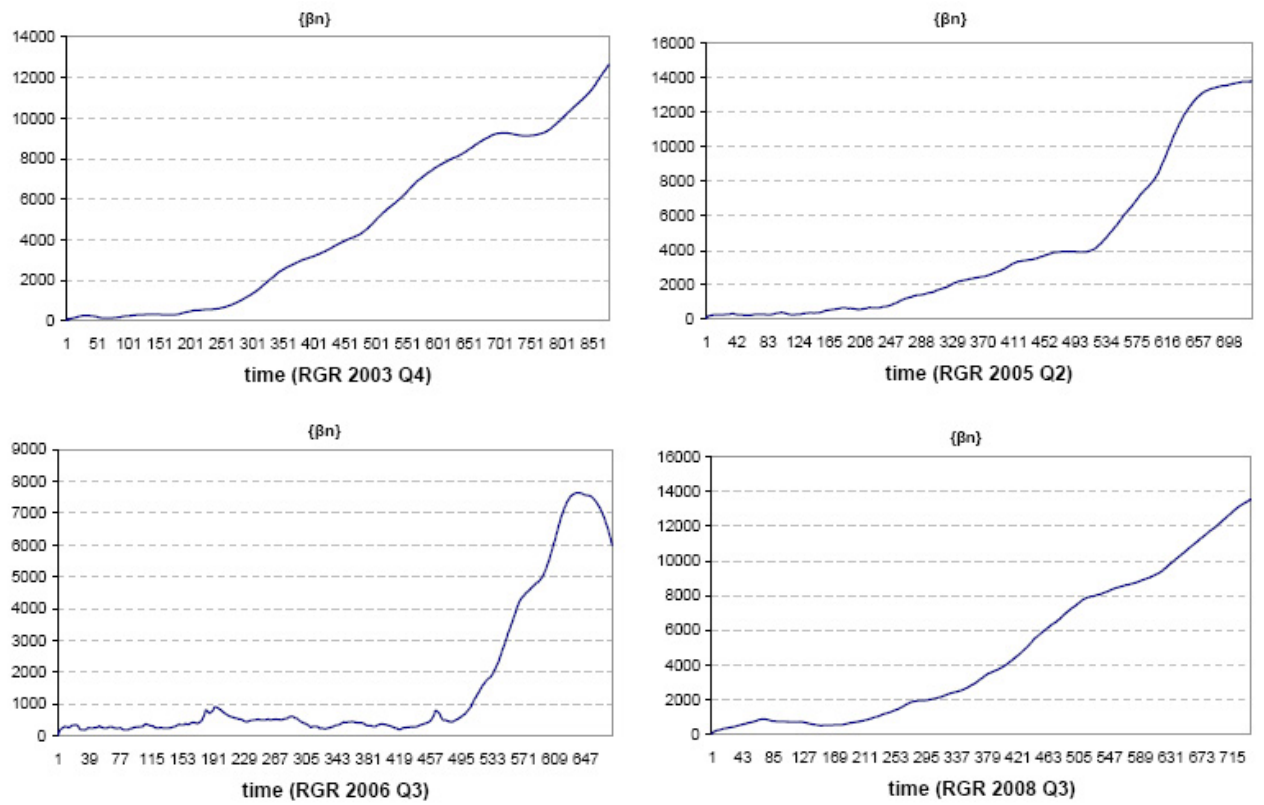Figure 7.12: Results of LUK through four different periods

Figure 7.13: Results of RGR through four different periods

7.5 Bayesian model selection

We provide model selection on three models: Kyle's two-variable model, dynamic model and GARCH(1,1) Bayesian model. The selection method is based on DIC and PLC and we developed methods to calculate those metrics. The modified GARCH (1,1) Bayesian model has advantage over the other two models in terms of model selection.

1 The DIC criteria

Spiegelhalter et al. (2002) propose a generalization of the AIC, whose asymptotic justification is not appropriate for hierarchical (3 or more level) models. The generalization is based on the posterior distribution of the deviance statistic,

$$D(\theta) = -2 \log f(y|\theta) + 2 \log h(y), \tag{7.1}$$

where $f(y|\theta)$ is the likelihood function and $h(y)$ is some standardizing function of the data. These authors suggest summarizing the fit of a model by the posterior expectation of the deviance, $\overline{D} = E_{\theta|y}[D]$ and the complexity of a model by the effective number of parameters p_D. In the case of Gaussian models, one can show that a reasonable definition of p_D is the expected deviance minus the deviance evaluated at the posterior expectations.

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \overline{D} - D(\overline{\theta}), \tag{7.2}$$

The *Deviance Information Criterion*(DIC) is then defined as

$$DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\theta}), \tag{7.3}$$

with smaller values of DIC indicating a better-fitting model. Both building blocks of DIC and p_D, $E_{\theta|y}[D]$, $D(E_{\theta|y}[\theta])$ are estimated via MCMC methods.

DIC is remarkably general, and computed as part of an MCMC run without any need

for extra sampling, or complicated loss function determination. Moreover, experience with DIC to date suggests it works remarkably well, despite the fact that no formal justification for it is yet available outside of posteriors that can be well approximated by a Gaussian distribution. Still, DIC is by no means universally accepted by Bayesians as a suitable model choice tool.

Since model comparison using DIC is not invariant to parametrization, the most sensible parametrization must be carefully chosen beforehand. Unknown scale parameters and other restructuring of the model can also lead to subtle changes in the DIC value.

Finally, DIC will depend on what part of the model specification is considered to be part of the likelihood, and what is not. Spiegelhalter et al. (2002) refer to this as the *focus* issue, i.e., determining which parameters are of primary interest, and which should count in p_D.

Our bayesian MCMC kyle inference has hierarchical parametrization, and DIC are sensitive to the scale and choice of parameter space.

2 Posterior predictive loss criteria

Another approach is also implemented from the output from posterior simulation is the *posterior predictive loss* approach of Gelfand and Ghosh (1998). The PLC is based on the posterior predictive distributions, and hence, irrespective of the dimensions of the parameter space, it can be used to compare different models.

We define $y_{l,obs}, l = 1, ..., n$ as observed data, and $y_{l,pred}, l = 1, ..., n$ as prediction. The selected model are those perform well under *balanced* loss function: this loss function penalizes actions both for departure from the corresponding observed value "fit" as well as for departure from what we expect the prediction to be "smoothness". The loss puts weights k and 1 on these two components, respectively, to allow for adjustment of relative regret for the two types of departure.

For squared loss, the resulting criterion becomes:

$$D_k = \frac{k}{k+1} \, G + P, \tag{7.4}$$

where $G = \sum_{l=1}^{n} (\mu_l - y_{l,obs})^2$, and $P = \sum_{l=1}^{n} \sigma_l^2$. μ_l, and σ_l^2 are the mean and variance of the predictive distribution of $y_{l,pred}$ given the observed data $y_{l,obs}$.

The components of D_k have natural interpretations. G is a goodness-of-fit term, while P is a penalty term. To clarify, we are seeking to penalize complexity and reward parsimony, just as DIC and other penalized likelihood criteria do. For a poor model we expect large predictive variance and poor fit. As the model improves, we expect to do better on both terms. But as we start to overfit, we will continue to do better with regard to goodness of fit, but also begin to inflate the variance.

If under model m we have parameters $\theta^{(m)}$, then,

$$p(y_{l,pred}|y) = \int p(y_{l,pred}|\theta^{(m)})p(\theta^{(m)}|y)d\theta^{(m)}, \tag{7.5}$$

Hence each posterior realization (say, θ^*) can be used to draw a corresponding $y_{l,pred}$ from $p(y_{l,pred}|\theta^{(m)} = \theta^*)$. The resulting $y_{l,pred}^*$ has marginal distribution $p(y_{l,pred}|y)$. With samples from this distribution we can obtain μ_l and σ_l^2.

3 Model assessment

We fit three different models and calculate DIC and PLC on different sample periods. The DIC numbers have marginal difference since DIC is sensitive to parameter spaces while PLC provides better assessment. The GARCH(1,1) Bayesian model has the lowest DIC and PLC, our dynamic model ranks the second. The original Kyle's model has the worst goodness-of-fit in terms of DIC and PLC. Table 7.3 illustrates the model assessment for sample period of IBM 2001Q2.

Table 7.3: Model assessment by different criteria

Model	DIC	PLC
Original Kyle's mode	839.60	4062.955
Dynamic model	821.00	3485.356
GARCH(1,1) dynamic model	820.20	2094.378

Chapter 8

Other Topics

Two models with asymmetric information, among others, have occupied arguably the most significant stage in the development of market microstructure research. They complement each other in a couple of aspects. The sequential trade model represented by Glosten and Milgrom (1985), referred to as G-M model in what follows, incorporates the trade orders (buy and sell) and the related bid-ask spread explicitly, but omits the aspects of trading time and volume. On the other hand, the strategic trade model introduced by Kyle (1985) articulates the informed trader's trading strategy in terms of trading time and volume, but does not separate buy and sell trading orders, hence there is no bid-ask spread included in the model.

A major task in our research is to perform model-based empirical studies in the area of market microstructure, especially using high frequency intra-day data. The G-M model appears to be oversimplified, i.e., the assumptions of several Bernouli variables in the G-M settings makes it inconvenient to be related to TAQ data. Although we have extended G-M model in two ways, the modified G-M settings are still not appropriate for empirical studies. We illustrate the major research we conducted.

Kyle's model (and the extensions we study) incorporates informational asymmetry in the informed trader's strategy and it also quantifies the market liquidity by using

$\{\lambda_n\}$. Since it uses batch orders instead of limit orders, it does not model transaction costs directly which is another important factor in market microstructure. Glosten and Milgrom (1985) proposes a limit order model (G-M model), and Back and Baruch (2004) provides a unified view on Kyle's model and the G-M model. A further research topic for us down the road will concern empirical studies related to the G-M model and the work of Back and Baruch (2004).

1: The first extension of G-M model

G-M studies a discrete time series. At each time epoch, not only the order buy or sell is a binary variable, but also the possible value for an asset is assumed to be either V_l (low) or V_h (high). Such simplification makes it inconvenient to connect G-M with Kyle's model in which both the security value and trading volume follow normal distributions. More importantly, making the security value a continuous variable will pave a road for empirical studies.

In what follows, we propose a modified G-M setting using interchangeable (or exchangeable) sequences in probability. In fact, G-M model is a special application of interchangeability. Recall that random variables $X_1, ..., X_n$ are said to be interchangeable if their joint distribution is invariant under permutations. A sequence of random variables $X = \{X_n, \ n = 1, 2, ...\}$ is said to be interchangeable if every finite subset of X is interchangeable. A useful example of X is a mixture of iid random variable that we are to define. See Chow and Teicher (2003) or other probability books for more detailed discussion on interchangeability.

- Let $\mu \in (0, 1)$ represent the proportion of informed traders. Assume one trading order is processed per unit time.

- At time 0, let the security value be a normal random variable $V_0 \sim N(p_0, \Sigma_0)$ whose density is denoted by $f_0(\cdot)$.

- Generate a sample $V_0 = v$ from $f_0(\cdot)$. Define the conditional probability for an

informed trader to place a buy order when observing v by an inverse logit transformation,

$$P_{I1}(B|v) = 1 - P_{I1}(S|v) = \frac{e^{\beta v}}{1 + e^{\beta v}} \quad \text{with the symbols } B = \text{``buy'' and } S = \text{``sell''},$$

, β is the coefficient. While for an uninformed (noise) trader,

$$P_{U1}(B|v) = 1 - P_{U1}(S|v) = \frac{1}{2}, \quad \forall v$$

i.e. overall, noise traders are insensitive to a particular asset value when they decide whether to buy or sell. Overall, we have

$$P_1(B|v) = 1 - P_1(S|v) = \mu P_{I1}(B|v) + (1 - \mu)P_{U1}(B|v). \tag{8.1}$$

- The unconditional probability of the trading order is given by

$$P_1(B) = 1 - P_1(S) = \mu \int_{-\infty}^{\infty} \frac{e^{\beta v}}{1 + e^{\beta v}} f_0(v) \, dv + \frac{1 - \mu}{2} \doteq K_1. \tag{8.2}$$

- The posterior density of the security value V_1 after one trade resulted from a buy order will follow from Bayes' formula:

$$f_1(v|B) = f_0(v) \, P_1(B|v)/P_1(B) = \frac{1}{K_1} f_0(v) \, [\mu \frac{e^{\beta v}}{1 + e^{\beta v}} + \frac{1 - \mu}{2}], \tag{8.3}$$

and similarly,

$$f_1(v|S) = f_0(v) \, P_1(S|v)/P_1(S) = \frac{1}{(1 - K_1)} f_0(v) \, [\mu \frac{1}{1 + e^{\beta v}} + \frac{1 - \mu}{2}], \tag{8.4}$$

Note that these two conditional densities are different from the unconditional

density of V_1. Moreover, the bid price, ask price and bid-ask spread can be defined respectively by

$$B_1 = E(V_1|S) \quad \text{with respect to the density } f_1(v|S)$$

$$A_1 = E(V_1|B) \quad \text{with respect to the density } f_1(v|B)$$

and $A_1 - B_1$.

- To move forward with the evolution of more orders, let O_n denote the order variable associated with the nth trade taking a value "B" or "S", and V_n be the asset value after the nth trade. Assume O_n, $n = 1, 2, ...$ are *conditionally* iid Bernoulli random variables given $v \in \mathbb{R}$, with the common conditional distribution $\{P_1(B|v),\ P_1(S|v)\}$ given in (8.1). Therefore, the sequence $\{O_n,\ n = 1, 2, ...\}$ is interchangeable. Note that (unconditionally) $O_1, O_2, ...$ are dependent with a positive correlation. Symbolically, Bayes' formula yields the posterior density for the asset value V_n given the trading flow $\{O_1, ..., O_n\}$:

$$f_n(v|O_1, ..., O_n) = \frac{f_0(v)\ P(O_1, ..., O_n|v)}{P(O_1, ...O_n)} = \frac{f_0(v)\ \prod_{i=1}^{n} P(O_i|v)}{P(O_1, ..., O_n)}, \qquad (8.5)$$

Theorem 4. *The trading flow $\{O_1, ..., O_n\}$ is invariant with respect to any particular order of $O_1, ..., O_n$. If we let b_n and s_n denote the total numbers of buy and sell orders respectively among $O_1, ..., O_n$, then*

$$f_n(v|O_1, ..., O_n) = f_n(v|b_n) = \frac{f_0(v)\ [P_1(B|v)]^{b_n}[P_1(S|v)]^{s_n}}{\int_{-\infty}^{\infty} f_0(v')\ [P_1(B|v')]^{b_n}[P_1(S|v')]^{s_n}\ dv'}. \qquad (8.6)$$

Proof. Given two order sequences, B_1, S_2 or S_1, B_2, we have,

$$
\begin{aligned}
f_2(v|B_1, S_2) &= \frac{f_1(v|B_1)\ \Pr(S_2|B_1, v)}{\Pr(B_1, S_2)} = \frac{f_1(v|B_1)\ \Pr(S_2|v)}{\int_{-\infty}^{\infty} f_1(v|B_1)\Pr(S_2|v)dv} \\
&= \frac{f_0(v)[\mu\frac{e^{\beta v}}{1+e^{\beta v}} + \frac{1-\mu}{2}][\mu\frac{1}{1+e^{\beta v}} + \frac{1-\mu}{2}]}{\int_{-\infty}^{\infty} f_0(v)[\mu\frac{e^{\beta v}}{1+e^{\beta v}} + \frac{1-\mu}{2}][\mu\frac{1}{1+e^{\beta v}} + \frac{1-\mu}{2}]dv} \\
&= \frac{f_1(v|S_1)\ \Pr(B_2|v)}{\int_{-\infty}^{\infty} f_1(v|S_1)\Pr(B_2|v)dv} \\
&= \frac{f_1(v|S_1)\ \Pr(B_2|S_1, v)}{\Pr(B_2, S_1)} \\
&= f(v|S_1, B_2),
\end{aligned}
$$

Replace $f_0(v)$ with $f_{n-1}(v)$ gives the proof by induction. $\qquad\square$

Because the posterior densities depend on the trading record path $O_1, ..., O_n$ only through a 1D summary statistic b_n (or s_n). Now we have a general formula for the posterior density:

$$
f_n(v|b_n) = \frac{f_0(v)\ [P_1(B|v)]^{b_n}[P_1(S|v)]^{n-b_n}}{\int_{-\infty}^{\infty} f_0(v')\ [P_1(B|v')]^{b_n}[P_1(S|v')]^{n-b_n}\ dv'}. \tag{8.7}
$$

The following theorem states that the bid and ask price quotes converge over time to the true value.

Theorem 5. *Suppose* $v = v^*$, *let* ask_n *and* bid_n *denote bid and ask price after nth trade, for* $n = 1, 2...$, *then they converge exponentially to* v^*.

$$
\lim_{n\to\infty} ask_n = \lim_{n\to\infty} bid_n = v^* \quad a.s. \tag{8.8}
$$

Proof. Given $\epsilon > 0$, let $P(v), \forall v \in (-\infty, \infty)$ denote the probability of v at interval $[v - \epsilon, v + \epsilon]$. If we let b_n and s_n denote the total numbers of buy and sell orders

respectively among $O_1, ..., O_n$. For $v = v'$, we have,

$$P_n(v') = \int_{v'-\epsilon}^{v'+\epsilon} f_n(v|b_n)dv,$$

$$\frac{P_n(v')}{P_n(v^*)} = \frac{f_0(v') \, [P_1(B|v')]^{b_n} [P_1(S|v')]^{s_n}}{f_0(v^*) \, [P_1(B|v^*)]^{b_n} [P_1(S|v^*)]^{s_n}},$$

Take logarithm,

$$\frac{1}{n} \log \frac{P_n(v')}{P_n(v^*)} = \frac{1}{n} \log \frac{f_0(v')}{f_0(v^*)} + \frac{b_n}{n} \log \frac{P_1(B|v')}{P_1(B|v^*)} + (1 - \frac{b_n}{n}) \log \frac{P_1(S|v')}{P_1(S|v^*)}, \quad (8.9)$$

The first term on the right hand side of equation (8.9) is 0 as $n \to \infty$.

Since $v = v^*$, by Strong Law of Large Numbers,

$$\frac{b_n}{n} \to P_1(B|v^*) \quad as \quad n \to \infty$$

Similarly, $s_n/n \to P_1(S|v^*)$ as $n \to \infty$, Let $P_1(B|v^*) = q$, $q \in [0,1]$ which is fixed for $v = v^*$. Also let $P_1(B|v') = y$, $y \in [0,1]$ which is random $\forall v'$, equation (8.9) could be written as,

$$\frac{1}{n} \log \frac{P_n(v')}{P_n(v^*)} \to f(y) = q \log \frac{y}{q} + (1-q) \log \frac{1-y}{1-q} \quad as \quad n \to \infty,$$

$$f'(y) = \frac{q}{y} - \frac{1-q}{1-y} = 0,$$

$$y^* = q,$$

$$f"(y) = \frac{-q}{y^2} - \frac{1-q}{(1-y)^2} < 0,$$

indicates y^* is maximum.

$$f(y) < f(q) = 0, \quad (y \neq q)$$

$$\frac{P_n(v')}{P_n(v^*)} \to 0 \quad as \quad n \to \infty \quad \forall v' \neq v^*,$$

107

It follows that, $\lim_{n \to \infty} ask_n = \lim_{n \to \infty} bid_n = v^*$ exponentially a.s. $\square$

We can summarize the results so far in

Theorem 6. *Let* $Y = \{(b_n, \ V_n) : \ n = 0, 1, 2, ...\}$ *and* $b_0 = 0$.

[1] *Y is a Markov chain over the state space* $\mathbb{N}_+ \times \mathbb{R}$ *where* $\mathbb{N}_+ = \{0, 1, 2, ...\}$, *whose transition probability function is given by* (8.7) *and*

$$P(b_{n+1}|v, b_n) = [P_1(B|v)]^{b_{n+1}-b_n}[P_1(S|v)]^{1-(b_{n+1}-b_n)}. \tag{8.10}$$

[2] *$V_0, V_1, ...$ are iid $N(p_0, \Sigma_0)$ random variables.*

[3] *$O_1, O_2, ...$ are interchangeable random variables with a constant correlation* $corr(O_i, O_j) = \rho \geq 0 \ \forall \ i \neq j$. *More specifically, $O_1, O_2, ...$ are a mixture of iid random variables, i.e. they are iid conditioning on $V_0 = v$.*

[4] *The (predicted) ask price and bid price after n trades are given by*

$$A_n \ = \ E(V_n|b_{n-1}, O_n = B), \tag{8.11}$$

$$B_n \ = \ E(V_n|b_{n-1}, O_n = S). \tag{8.12}$$

Given $v = v^*$ (insider knows v^*), they converge exponentially to v^* as $n \to \infty$.

2: The second extension of G-M model

In the first extended G-M model, an over-simplified assumption is every order (buy or see) has to be of size one. Apparently, variable trading volumes affect the fluctuation of asset prices, and should not be ignored in market microstructure studies. Now we propose another extension of G-M model that incorporates trading volumes.

Let the same notation O_n represent the (signed) volume (# of shares) in the nth trade, where the sign indicates the direction of order imbalance: $O_n > 0$ (resp. $O_n < 0$)

corresponds to a net "buy order" (resp. "sell order") effect. We basically follow the formulation in the first extended G-M model except for defining O_n's as continuous random variables and making related adjustments accordingly.

- With $\mu \in (0,1)$ denoting the proportion of informed traders, we still assume one trading order is processed per unit time. Suppose $V_0 \sim N(p_0, \Sigma_0)$ with the density $f_0(\cdot)$.

- Given $V_0 = v$, suppose O_n, $n = 1, 2, \ldots$ are conditionally iid with a common conditional density (as a mixture of two normal densities) defined by

$$g_1(u|v) = \mu \ g_{I1}(u|v) + (1 - \mu) \ g_{U1}(u|v), \tag{8.13}$$

  where

$$g_{I1}(\cdot \ |v) \sim N(\mu_I(v), \sigma_I^2(v)) \quad \text{and} \quad g_{U1}(\cdot \ |v) \sim N(0, \sigma_U^2) \tag{8.14}$$

  for an informed trader and a noise trader respectively.

- Once again, given the trading volume history $O_i = u_i$, $i = 1, \ldots, n$, Bayes' formula yields the posterior density for the asset value V_n:

$$f_n(v|u_1, \ldots, u_n) = \frac{f_0(v) \ \prod_{i=1}^{n} g_1(u_i|v)}{\int_{-\infty}^{\infty} f_0(v') \ \prod_{i=1}^{n} g_1(u_i|v') \ dv'}. \tag{8.15}$$

  However, (8.15) is not as practically useful as (8.7) due to the lack of summary statistics such as b_n in (8.7). Consequently, the posterior density becomes genuinely path-dependent. As more and more trades are made, the required memory grows exponentially, which makes the updating intractable.

Chapter 9

Conclusions

This thesis studies market microstructure models at high frequency tick level. We propose a novel dynamic Bayesian model derived from Kyle's microstructure framework in which informational asymmetries and liquidity needs motivate trade. Bayesian inference allows us to interpret the economic impacts of certain parameters contained in the model. A new algorithm is developed to characterize Kyle's solution with a significant reduction in computational complexity when the model is expanded to encompass the real data. The empirical studies on NYSE TAQ data illustrate the applications of Kyle's model and our dynamic model. The approach developed in this thesis can be extended to characterize variation of microstructure models and other dynamic determinants.

The dynamic model is estimated and tested for intraday data on five randomly chosen stocks over four different sample periods. Our study focuses on informed trading between two consecutive quarterly earning announcement. The test of our dynamic model indicates that the specification is broadly consistent with the original Kyle's model through simulation, while our dynamic model performs vastly better than the equilibrium solution in real data. This work is subject to future validation with new measures and compared against other microstructure models.

The magnitude of trading intensity β_n from Bayesian inference provides interesting

implications of insider's strategies, while we do not find persuasive influence of market depth λ_n^{-1} through our dynamic model. In the meantime, we observe strong persistency in the order imbalance volatility. The GARCH(1,1) modification is a prominent candidate to embrace such property. We develop our GARCH dynamic model and implement it with empirical data analysis. The results are promising. The impact could be substantial to explain high volatility persistency measures founded in both price evolution and order flows. Thus, future research in this area is likely to stem from theoretical work that bridge equilibrium solution and GARCH dynamic model, and the interplay between price and order imbalance volatility processes.

Appendix

A: Proof of Proposition 1:

First of all, we show by backward induction that $\lambda_n, \alpha_n, \beta_n$ are nonnegative for all n. Starting from the second order condition (5.14) and $\alpha_N = 0$ we have $\lambda_N > 0$, which plus (5.9) implies $\alpha_{N-1} > 0$. Then $\lambda_{N-1} > 0$ and $\alpha_{N-1}\lambda_{N-1} < 1$ would follow from (5.14), etc. This argument shows $\alpha_n > 0$ for $n = N-1, N-2, ..., 1$ and $\lambda_n > 0$ for $n = N, N-1, ..., 1$. Furthermore, (5.12) implies $\beta_n > 0 \ \forall \ n$.

Next, with $d_n = \alpha_n\lambda_n$, it follows from (5.9) that

$$\alpha_{n-1} = \frac{1}{4\lambda_n(1-d_n)}, \quad \text{and}$$

$$d_{n-1} = \alpha_{n-1}\lambda_{n-1} = \frac{\lambda_{n-1}}{4\lambda_n(1-d_n)}, \tag{A1}$$

which implies

$$\frac{\lambda_n}{\lambda_{n-1}} = \frac{1}{4\ d_{n-1}(1-d_n)}. \tag{A2}$$

From (5.12), we have

$$\frac{\lambda_n}{\lambda_{n-1}} = \frac{\beta_n}{\beta_{n-1}} \frac{\Sigma_n}{\Sigma_{n-1}}. \tag{A3}$$

Plugging (5.13) in (A3) yields

$$\frac{\lambda_n}{\lambda_{n-1}} = \frac{\beta_n}{\beta_{n-1}} (1 - \beta_n\lambda_n\Delta t_n). \tag{A4}$$

Note that (5.11) implies

$$\beta_n \ \Delta t_n = \frac{1 - 2d_n}{2\lambda_n \ (1-d_n)}, \tag{A5}$$

from which substituting β_n and β_{n-1} into (A4) leads to

$$\frac{\lambda_n}{\lambda_{n-1}} = \frac{1 - 2d_n}{2\lambda_n (1 - d_n)} \frac{2\lambda_{n-1} (1 - d_{n-1})}{1 - 2d_{n-1}} \frac{1}{2(1 - d_n)}, \tag{A6}$$

and

$$\frac{\lambda_n^2}{\lambda_{n-1}^2} = \frac{1 - 2d_n}{2(1 - d_n)^2} \frac{1 - d_{n-1}}{1 - 2d_{n-1}}. \tag{A7}$$

Squaring both sides of (A2) and matching (A7), we have

$$\frac{8 \ d_{n-1}^2 (1 - d_{n-1})}{1 - 2d_{n-1}} = \frac{1}{1 - 2d_n} \triangleq K_n, \tag{A8}$$

which can be expressed as the cubic equation (5.16):

$$8 \ d_{n-1}^3 - 8 \ d_{n-1}^2 - 2K_n \ d_{n-1} + K_n = 0,$$

with the boundary condition $d_N = 0$, $K_N = 1$ following $\alpha_N = 0$. Furthermore, (5.18) – (5.20) follow from (5.11) – (5.13) straightforwardly.

To show the existence and uniqueness of the solution in the interval $(0, 1/2)$ to equation (5.16) for each n, note that the definition $\Sigma_n = Var(v|\mathcal{F}_n^U)$ appears to imply $\Sigma_n < \Sigma_{n-1}$, which along with (5.18) shows $[2(1 - d_n)]^{-1} < 1$, hence $K_n > 0$ and $0 < d_n < 1/2$ for all $n = N - 1, ..., 1$. However, these assertions need to be justified more rigorously via backward induction.

Rewrite (5.16) as $L(d_{n-1}) = R(d_{n-1})$ where

$$L(d_{n-1}) \ = \ 8d_{n-1}^2 \ (d_{n-1} - 1), \tag{A9}$$

$$R(d_{n-1}) \ = \ K_n \ (2d_{n-1} - 1). \tag{A10}$$

Since $0 < d_{n-1} < 1 \; \forall \; n$ by the second order condition, $L(d_{n-1}) < 0 \; \forall \; n$. Starting from $K_N = 1$, we have $2d_{N-1} - 1 < 0$, hence $d_{N-1} < 1/2$ and $K_{N-1} > 0$. Successive iterations between (A10) and (5.17) lead to $d_n \in (0, 1/2)$ for all $n = N - 1, ..., 1$. Note for every $n = N - 1, ..., 1$, the first-order derivatives

$$
\begin{aligned}
L'(d_{n-1}) &= 8d_{n-1}(3d_{n-1} - 2) < 0, \\
R'(d_{n-1}) &= 2K_n > 0
\end{aligned}
$$

imply that $L(d_{n-1})$ decreases in $d_{n-1} \in (0, 1/2)$ and $R(d_{n-1})$ increases. Observing $L(0) = 0$, $L(1/2) = -1$, $R(0) = -K_n < 0$ and $R(1/2) = 0$, there is a unique intersection between the two curves $L(d_{n-1})$ and $R(d_{n-1})$ at $d_{n-1} \in (0, 1/2)$. Also, K_n is increasing in d_n and satisfies $K_n > 1$.

Finally, as for the monotonicity of sequence $\{d_n\}$, notice that the graph of $L(\cdot)$ remains unchanged in the interval $[0, 1/2]$ but that of $R(\cdot)$ will change due to its dependence on K_n for different n. More specifically, $K_{N-1} > 1 = K_N$ pushes the (vertical axis) intercept for $R(d_{N-2})$ downwards relative to the intercept for $R(d_{N-1})$, which implies $d_{N-2} > d_{N-1}$. Following the inductive argument, we obtain (5.21):

$$
\frac{1}{2} > d_1 > d_2 >, ..., > d_{N-1} > d_N = 0.
$$

This completes the proof of Proposition 1.

B: Proof of Proposition 2:

First,

$$
h_n = \sqrt{2(1 - d_{n-1}) \frac{1 - 2d_n}{1 - 2d_{n-1}}}
$$

easily follows from (5.20) and (A7). Since $\frac{1 - 2d_n}{1 - 2d_{n-1}} > 1$ and $d_{n-1} < 1/2$, we have $h_n > 1$.

Next, (A7) implies

$$k_n = \sqrt{\frac{1 - 2d_n}{1 - 2d_{n-1}} \frac{1 - d_{n-1}}{2(1 - d_n)^2}}.$$

Furthermore, $k_n > 1/2$ follows from $\frac{1-2d_n}{1-2d_{n-1}} > 1$ and $\frac{1-d_{n-1}}{2(1-d_n)^2} > 1/4$.

This completes the proof of Proposition 2.

C: Proof of Proposition 3:

To characterize the equilibrium solution for the case with noisy signals observed by the informed trader, we rely on the same method applied to Proposition 1, instead of Kyle's original approach.

Define $q_n = \alpha_n \lambda_n$, from equation (5.39)

$$\begin{aligned}
\alpha_{n-1} &= \frac{\gamma^2}{4\lambda_n (1 - q_n)}, \\
q_{n-1} &= \alpha_{n-1}\lambda_{n-1} = \frac{\lambda_{n-1}}{\lambda_n} \frac{\gamma^2}{4(1 - q_n)},
\end{aligned} \tag{A11}$$

Which implies

$$\frac{\lambda_n}{\lambda_{n-1}} = \frac{\gamma^2}{4q_{n-1}(1 - q_n)}. \tag{A12}$$

From equation (5.41), we have,

$$\begin{aligned}
\frac{\lambda_n}{\lambda_{n-1}} &= \frac{\beta_n}{\beta_{n-1}} \frac{\Sigma_n}{\Sigma_{n-1}} \frac{\beta_{n-1}^2\left(\frac{1}{\gamma} - 1\right)\Sigma_{n-2}\Delta + \sigma_u^2}{\beta_n^2\left(\frac{1}{\gamma} - 1\right)\Sigma_{n-1}\Delta + \sigma_u^2} \\
&\triangleq \frac{\beta_n}{\beta_{n-1}} \frac{\Sigma_n}{\Sigma_{n-1}} W.
\end{aligned} \tag{A13}$$

Let $E_{n-1}(\cdot)$, $Var_{n-1}(\cdot)$, $Cov_{n-1}(\cdot, \cdot)$ and $corr_{n-1}(\cdot, \cdot)$ denote the conditional mean, variance, covariance and correlation operators respectively given $\mathcal{F}_{n-1}^U$. Since order flows from the insider are much less volatile than the trades from the uninformed and the

insider is actively hiding his trading behavior, $Var_{n-1}(\Delta x_n) << Var_{n-1}(\Delta u_n)$. Therefore,

$$
\begin{aligned}
Var_{n-1}(\Delta x_n) &= Var_{n-1}[\beta_n \Delta \ (v - p_{n-1} + \epsilon_n)] = \beta_n^2 \ (\Sigma_{n-1} + a_n^2 \sigma_\epsilon^2) \ \Delta^2 \\
&= \beta_n^2 \ \Sigma_{n-1} \Delta^2 / \gamma << Var_{n-1}(\Delta u_n) = \sigma_u^2 \Delta,
\end{aligned}
$$

which leads to $W \approx 1$. (5.42) turns (A13) to

$$
\frac{\lambda_n}{\lambda_{n-1}} = \frac{\beta_n}{\beta_{n-1}} \ (1 - \beta_n \lambda_n \Delta). \tag{A14}
$$

(5.40) now becomes

$$
\beta_n \ \Delta = \frac{\gamma(1 - 2q_n)}{2\lambda_n(1 - q_n)}, \tag{A15}
$$

from which substituting β_n and β_{n-1} into (A14) we have

$$
\frac{\lambda_n}{\lambda_{n-1}} = \frac{1 - 2q_n}{2\lambda_n \ (1 - q_n)} \frac{2\lambda_{n-1} \ (1 - q_{n-1})}{1 - 2q_{n-1}} \frac{2(1 - q_n) - \gamma(1 - 2q_n)}{2(1 - q_n)}. \tag{A16}
$$

Therefore,

$$
\frac{\lambda_n^2}{\lambda_{n-1}^2} = \frac{1 - 2q_n}{1 - 2q_{n-1}} \frac{1 - q_{n-1}}{1 - q_n} \frac{2(1 - q_n) - \gamma(1 - 2q_n)}{2(1 - q_n)}. \tag{A17}
$$

Squaring both sides of equation (A12) and equating with equation (A17),

$$
\frac{8q_{n-1}^2(1 - q_{n-1})}{1 - 2q_{n-1}} = \frac{\gamma^4}{(1 - 2q_n) \ [2(1 - q_n) - \gamma(1 - 2q_n)]} = K_n, \tag{A18}
$$

116

which is equivalent to the cubic equation

$$8q_{n-1}^3 - 8q_{n-1}^2 - 2K_n q_{n-1} + K_n = 0. \tag{A19}$$

Note that the boundary condition $\alpha_N = 0$ implies $q_N = \alpha_N \lambda_N = 0$. Moreover, $K_N = \frac{\gamma^4}{2-\gamma} \in (1/24, 1/2)$.

We then show there exists one unique root which makes economic sense. From the second order condition from (5.44), we have $\lambda_n(1 - q_n) > 0$, then $q_n < 1$ for all n.

Follow from equation (5.47), we have $0 < \frac{\gamma(1-2q_n)}{2(1-q_n)} < 1$, which leads to $q_n < 1/2$, for all n. We could rewrite (A19) as

$$8q_{n-1}^2(q_{n-1} - 1) = 2K_n q_{n-1} - K_n. \tag{A20}$$

It is straightforward to show that there exists a unique intersection $q_{n-1} \in (0, 1/2)$ between the RHS function and the RHS function in (A20).

This completes the proof of Proposition 3.

D: Proof of Theorem 3:

We follow a backward induction procedure. Denote the conditional mean and variance given $\mathcal{F}_n^I$ by $E_n^I(\cdot)$ and $Var_n^I(\cdot)$ respectively. Suppose at the $(n+1)$th auction, the insider's expected profit satisfies

$$E_n^I(\pi_{n+1}) = \alpha_n(s_{n+1} - p_n)^2 + \delta_n. \tag{A21}$$

Since $\pi_n = (v - p_n)\,\Delta x_n + \pi_{n+1}$, we have

$$E_{n-1}^I(\pi_n) = \max_{\Delta x} E_{n-1}^I[(v - p_n)\Delta x + \alpha_n(s_{n+1} - p_n)^2 + \delta_n]. \tag{A22}$$

A linear equilibrium requires

$$p_n = p_{n-1} + \lambda_n(\Delta x_n + \Delta u_n) + h, \tag{A23}$$

where h is a function of $\Delta y_1, ..., \Delta y_{n-1}$. Plugging (A23) into (A22) and evaluating the conditional expectation,

$$
\begin{aligned}
& E_{n-1}^{I}(\pi_n) \\
= & \max_{\Delta x} E_{n-1}^{I}[(v - p_{n-1} - \lambda_n \Delta x - h)\,\Delta x + \alpha_n(s_{n+1} - p_{n-1} - \lambda_n \Delta x - h - \lambda_n \Delta u_n)^2 + \delta_n] \\
= & \max_{\Delta x} \{[E_{n-1}^{I}(v - p_{n-1}) - \lambda_n \Delta x - h]\,\Delta x + \alpha_n\, E_{n-1}^{I}(s_{n+1} - p_{n-1})^2 \\
& \quad + \alpha_n(\lambda_n \Delta x + h)^2 - 2\alpha_n\,(\lambda_n \Delta x + h)\, E_{n-1}^{I}(s_{n+1} - p_{n-1}) + \alpha_n\,\lambda_n^2 \sigma_u^2 \Delta t_n + \delta_n\}. \tag{A24}
\end{aligned}
$$

Recall $b_n = \frac{\Sigma_{n-1}}{\Sigma_{n-1} + a_n^2 \sigma_\epsilon^2}$. Note that for each n, the information set $\mathcal{F}_{n-1}^{U}$ does not include s_n. We have

$$
\begin{aligned}
E_{n-1}^{I}(v - p_{n-1}) &= \frac{Cov_{n-1}(v - p_{n-1}, s_n)}{Var_{n-1}(s_n)}\,(s_n - p_{n-1}) = b_n\,(s_n - p_{n-1}); \tag{A25} \\
E_{n-1}^{I}(s_{n+1} - p_{n-1}) &= E_{n-1}^{I}[(v - p_{n-1}) + a_{n+1}\epsilon_{n+1}] = b_n\,(s_n - p_{n-1}); \tag{A26} \\
E_{n-1}^{I}(s_{n+1} - p_{n-1})^2 &= E_{n-1}^{I}[(v - p_{n-1}) + a_{n+1}\epsilon_{n+1}]^2 \\
&= E_{n-1}^{I}(v - p_{n-1})^2 + a_{n+1}^2 \sigma_\epsilon^2 \\
&= Var_{n-1}^{I}(v - p_{n-1}) + [E_{n-1}^{I}(v - p_{n-1})]^2 + a_{n+1}^2 \sigma_\epsilon^2 \\
&= Var_{n-1}(v)\,\{1 - [corr_{n-1}(v, s_n)]^2\} + b_n^2(s_n - p_{n-1})^2 + a_{n+1}^2 \sigma_\epsilon^2 \\
&= \Sigma_{n-1}\,(1 - b_n) + b_n^2(s_n - p_{n-1})^2 + a_{n+1}^2 \sigma_\epsilon^2. \tag{A27}
\end{aligned}
$$

Note: Compared to $E_{n-1}(v - p_{n-1}) = 0$, (A25) shows a distinction between the informed trader and uninformed ones.

The first order condition in maximizing the future profit yields

$$b_n(s_n - p_{n-1}) - 2\lambda_n \Delta x - h + 2\alpha_n \lambda_n (\lambda_n \Delta x + h) - 2\alpha_n \lambda_n \ b_n(s_n - p_{n-1}) = 0.$$

Hence

$$\Delta x = \frac{b_n(1 - 2\alpha_n \lambda_n)}{2\lambda_n \ (1 - \alpha_n \lambda_n)} \ (s_n - p_{n-1}) + h \ f_n, \tag{A28}$$

with $f_n = \frac{2\alpha_n \lambda_n - 1}{2\lambda_n(1 - \alpha_n \lambda_n)}$ and the second order condition

$$\lambda_n \ (1 - \alpha_n \lambda_n) > 0. \tag{A29}$$

From (A24), we can verify (5.39):

$$\alpha_{n-1} = \frac{b_n^2}{4\lambda_n(1 - \alpha_n \lambda_n)}.$$

We claim that $h = 0$. In fact, the market efficient condition implies $E(\Delta p_n | \mathcal{F}_{n-1}^U) = 0$. It follows from (A23) and (A28) that

$$E\{\Delta p_n | \mathcal{F}_{n-1}^U\} = h \ (1 + \lambda_n f_n) = 0, \tag{A30}$$

which implies $h = 0$, and also (5.40):

$$\beta_n \ \Delta t_n = \frac{b_n \ (1 - 2\alpha_n \lambda_n)}{2\lambda_n \ (1 - \alpha_n \lambda_n)}.$$

It follows from the pricing equation $p_n = E_{n-1}(v| \ \mathcal{F}_{n-1}^U; \Delta y_n)$ and the projection

theorem that

$$
\begin{aligned}
p_n - p_{n-1} &= E[v - p_{n-1}|\; \mathcal{F}^U_{n-1}; \Delta y_n] \\
&= \frac{Cov_{n-1}(v - p_{n-1}, \Delta y_n)}{Var_{n-1}(\Delta y_n)}\; (\Delta y_n) \\
&= \lambda_n\; \Delta y_n, && \text{(A31)}
\end{aligned}
$$

hence

$$
\begin{aligned}
\lambda_n &= \frac{Cov_{n-1}(v - p_{n-1}, \Delta x_n + \Delta u_n)}{Var_{n-1}(\Delta x_n + \Delta u_n)} \\
&= \frac{\beta_n \Delta t_n\; \Sigma_{n-1}}{\beta_n^2 (\Delta t_n)^2 (\Sigma_{n-1} + a_n^2 \sigma_\epsilon^2) + \sigma_u^2 \Delta t_n} \\
&= \frac{\beta_n \Sigma_{n-1}}{\beta_n^2 \Delta t_n\; (\Sigma_{n-1} + a_n^2 \sigma_\epsilon^2)\; + \sigma_u^2} \\
&= \frac{1}{\frac{\beta_n \Delta t_n}{b_n} + \frac{\sigma_u^2}{\beta_n \Sigma_{n-1}}}. && \text{(A32)}
\end{aligned}
$$

Since the (joint) conditional distribution of the vector $(v - p_{n-1},\; \Delta y_n)$ given $\mathcal{F}^U_{n-1}$ is bivariate Gaussian, (5.42) follows from

$$
\begin{aligned}
\Sigma_n &= Var\{v - p_{n-1}|\; \mathcal{F}^U_{n-1}; \Delta y_n\} \\
&= \Sigma_{n-1}\; \{1 - [corr_{n-1}(v - p_{n-1}, \Delta y_n)]^2\} \\
&= \Sigma_{n-1}\; \left\{1 - \frac{[Cov_{n-1}(v - p_{n-1}, \Delta x_n + \Delta u_n)]^2}{Var_{n-1}(v)\; Var_{n-1}(\Delta x_n + \Delta u_n)}\right\} \\
&= \Sigma_{n-1}\; \left\{1 - \frac{\lambda_n Cov_{n-1}(v - p_{n-1}, \Delta x_n + \Delta u_n)}{Var_{n-1}(v)}\right\} \\
&= \Sigma_{n-1}\; \left(1 - \frac{\lambda_n \beta_n \Delta t_n\; \Sigma_{n-1}}{\Sigma_{n-1}}\right) \\
&= \Sigma_{n-1}\; (1 - \lambda_n \beta_n \Delta t_n).
\end{aligned}
$$

Now (A32) and (5.42) imply

$$\frac{\lambda_n \beta_n \Delta t_n}{b_n} + \frac{\lambda_n \sigma_u^2}{\beta_n \Sigma_{n-1}} = 1,$$

and

$$\lambda_n \beta_n^2 \Delta t_n \ \left(b_n^{-1} - 1\right) \ \Sigma_{n-1} + \lambda_n \sigma_u^2 = \beta_n \Sigma_n,$$

which turns to (5.41) easily.

Finally, collecting relevant terms in (A24) entails

$$\delta_{n-1} = \delta_n + \alpha_n \lambda_n^2 \ \sigma_u^2 \Delta t_n + \alpha_n (1 - b_n) \ \Sigma_{n-1} + \alpha_n \ a_{n+1}^2 \sigma_\epsilon^2. \tag{A33}$$

Bibliography

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

Amihud, Y., and Mendelson, H. (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics* **17**, 223-249.

Amihud, Y., (2002). Illiquidity and stock returns: Cross-section and time series effects. *Journal of Financial Markets* **5**, 31-56.

Amihud, Y., Mendelson, H. and Pedersen, L.H. (2005). Liquidity and asset prices. *Foundations and Trends in Finance* **4**, 269-364.

Back, K. (1992). Insider trading in continuous time. *Review of Financial Studies* **5**, 387-409.

Back, K. and Baruch, S. (2004). Information in securities markets: Kyle meets Glosten and Milgrom. *Econometrica* **72**, 433-465.

Biais, B., Glosten, L. and Spatt, C. (2005), Market Microstructure: a Survey of Microfoundations, Empirical Results and Political Implications. *Journal of Financial Markets* **8**, 217-264

Brenna, M.J., Chordia, T. and Subrahmanyam, A. (1998). *Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics* **49**, 345-373

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

Brunnermeier, M.K. (2001). *Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis, and Herding.* Oxford Univ. Press.

Casella, G and George, E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician* Vol. 46, No. 3. pp. 167-174.

Chib, S. and Greenberg, E. (1996). Markov Chain Monte Carlo Simulation Methods in Econometrics. *Journal of Econometrics* **75**, 79-97.

Chordia, T., Huh, S., and Subrahmanyam, A. (2007). The Cross-Section of Expected Trading Activity. *Review of Financial Studies* **20**, 709-740.

Chordia, T., Huh, S-W. and Subrahmanyam, A. (2009). Theory-based illiquidity and asset pricing. *The Review of Financial Studies* **22**, 3629-3668.

Chow, Y., and Teicher, H. (2003). *Probability Theory, Independence, Interchangeability, Martingales.* Springer Texts.

Cochrane, J. (2005). Liquidity, Trading and Asset Prices. *NBER Reporter*, 1-12.

Constantinides, G.M. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy* **94**, 842-862.

Copeland, Thomas C., and Daniel Galai. (1983). Information effects of the bid-ask spread *Journal of Finance* **38**, 1457-1469.

Datar, V., Naik, N. and Radcliffe, R. (1998) Liquidity and stock returns: An alternative test. *Journal of Financial Markets* **1**, 205-219.

de Jong, F. and Rindi, B. (2009). *The Microstructure of Financial Markets.* Cambridge Univ. Press.

Durbin, J. and Koopman, S.J. (2001) Time Series Analysis by State Space Methods. Oxford University Press.

Dreman David N. and Berry Michael A. (1992). Overreaction, Underreaction, and the Low-P/E Effect. *Financial Analysts Journal* **51 (4)**, 2130.

Easley, David and Maureen O'Hara. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics* **19**, 69-90.

Eleswarapu, V.R. (1997). Cost of transacting and expected returns in the Nasdaq market. *Journal of Finance* **52**, 2113-2127.

Fama, E. F. and French, Kenneth R. (1992). The Cross-Section of Expected Stock. Returns. *Journal of Finance* **47**, 427-465.

Gelfand, A. E. and Ghosh, S. K. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika* **Vol. 85, No. 1**, 1-11.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of American Statistics Association* **85**, 398-409.

Gilks, W.R., and Wild, P. (1992) Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society.* series C **41**. 337-348.

Glosten, L. and Milgrom, P. (1985). Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* **13**, 71-100.

Grossman, S.J. and Stiglitz, J.E. (1980). On the impossibility of informationally efficient markets. *American Economic Review* **70**, 393-408.

Harris, L. (2003). *Trading and Exchanges: Market Microstructure for Practitioners.* Oxford Univ. Press.

Harvey, A.C. and Pierse, R.G. (1984). Estimating Missing Observations in Economic Time Series. *Journal of the American Statistical Association* **79**, 125-131.

Hasbrouck, J. (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford Univ. Press.

Hasbrouck, J. (2009). Trading costs and returns for US equities: estimating effective costs from daily data. *Journal of Finance* (forthcoming).

Hastings, W.K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications *Biometrika* **57**, 97109.

Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica* **53**, 1315-1335.

Kyle, A.S. (1989). Information speculation with imperfect competition. *Review of Economic Studies* **56**, 317-356.

Luttmer, E.G. (1996). Asset pricing in economies with frictions. *Econometrica* **64**, 1439-1467.

Madhavan, A. (2000). Market Microstruture: A survey. *Journal of Financial Markets* **3**, 205-258.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, **21**, 10871092.

O'Hara, M. (1995). *Market Microstructure Theory.* Blackwell Publishing.

Pasquariello, P. and Vega, C. (2009). Strategic cross-trading in the U.S. stock market. Working paper.

Pástor, Ľ. and Stambaugh, R.F. (2003). Liquidity risk and expected stock returns. *The Journal of Political Economy* **111**, 642-685.

Ripley, B. D. (1987) Stochastic Simulation. Wiley, 237pp.

Ritter, C., and Tanner, M.A. (1992) The Gibbs Stopper and the Griddy Gibbs Sampler. *Journal of American Statistical Association* **87** 861-868.

Roberts, G. O. and Smith, A. F. M. (1994) Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* Volume 49, Issue 2, Pages 207-216

Roll, Richard. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance.* **39**, 1127-1139.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society* series B **55**, No. 1, pp. 3-23

Spiegel, Matthew, and Avanidhar Subrahmanyam. (1992). Informed Speculation and Hedging in a Noncompetitive Securities Market. *Review of Financial Studies.* **5**, 307-329.

Spiegelhalter, D.J., Best, N.G, Carlin, B.P. and Van der Linde, A., (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society.* Series B, **64(4)**, 583-616.

Stoll, Hans R. (1978). The supply of dealer services in securities markets. *Journal of Finance.* **33**, 1133-1151.

Tanner, M.A. and Wong, W.H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* Vol. 82, **No.398** pp. 528-540.

Tierney, L (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics* Vol. 22, **No. 4** pp. 1701-1728.

Vayanos, D. and Wang, J. (2009). Liquidity and asset prices: a unified framework. Working paper.

Vayanos, D. and Wang, J. (2009). Theories of Liquidity. *Foundations and Trends in Finance*, forthcoming.

Vives, X. (2008). *Information and Learning in Markets.* Princeton Univ. Press.

West, M. and Harrison, P.J. (1997) Bayesian Forecasting and Dynamic Models. Springer-Verlag, New York, (2nd Edn)