THREE ESSAYS ON MARKET MICROSTRUCTURE AND FINANCIAL ECONOMETRICS

by

Yi Xue

B.A., Wuhan University of China, 2002

M.A., Simon Fraser University, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Economics

SIMON FRASER UNIVERSITY
Spring 2009

APPROVAL

Name: **Yi Xue**

Degree: **Doctor of Philosophy**

Title of Project: **Three Essays on Market Microstructure and Financial Econometrics**

Examining Committee:

Chair:

Nicolas Schmitt
Professor, Department of Economics

Ramazan Gencay
Senior Supervisor
Professor, Department of Economics

Alexander Karaivanov
Supervisor
Assistant Professor, Department of Economics

Robert Jones
Supervisor
Professor, Department of Economics

Robert Gauer
Internal Examiner
Professor, Faculty of Business Administration, Finance

William A. Brock
External Examiner
Professor, Department of Economics University of Wisconsin - Madison

Date Defended/Approved: March 5, 2009

Abstract

This thesis consists of three essays that study three interdependent topics: microstructure foundation of volatility clustering, inefficiency of information diffusion and jump detection in high frequency financial time series data.

Volatility clustering, with autocorrelations of the hyperbolic decay rate, is unquestionably one of the most important stylized facts of financial time series. The first essay forms Chapter 1 which presents a market microstructure model that is able to generate volatility clustering with hyperbolic autocorrelations through traders with multiple trading frequencies using Bayesian information updating in an incomplete market. The model illustrates that signal extraction, which is induced by multiple trading frequency, can increase the persistence of the volatility of returns. Furthermore, it is shown that the local temporal memory of the underlying time series of returns and their volatility varies greatly with the number of traders in the market.

The second essay, Chapter 2, presents a market microstructure model showing that an increasing number of information hierarchies among informed competitive traders leads to a slower information diffusion rate and informational inefficiency. The model illustrates that informed traders may prefer trading with each other rather than with noise traders in the presence of the information hierarchies. Furthermore, it is shown that momentum can be generated from the trend following behavior pattern of noise traders.

I propose a new nonparametric test based on wavelets to detect jump arrivals in high frequency financial time series data, in the third essay, Chapter 3. It is demonstrated that the test is robust for different specifications of price processes and the presence of market microstructure noise and it has good size and power. Further, I examine the multi-scale jump dynamics in U.S. equity markets and the findings are as follows. First, the jump dynamics of equities are entirely different across different time scales. Second, although

arrival densities of positive jumps and negative jumps are symmetric across different time scales, the magnitude of jumps is distributed asymmetrically at high frequencies. Third, only twenty percent of jumps occur in the trading session from 9:30AM to 4:00PM, suggesting that jumps are largely determined by news rather than liquidity shocks.

Keywords: Trading frequency; Volatility clustering; Signal extraction; Hyperbolic decay; Information hierarchies; Information diffusion rate; Momentum: Jump detection; Wavelets; Directional jumps: Negative jumps; Positive jumps

Subject terms: Trading frequency; Volatility clustering; Hierarchical information; Momentum: Wavelets; Risk management

To my parents, for your life of love
To Shanshan, for your love of life

Acknowledgments

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. I owe my gratitude to all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

My deepest gratitude is to my senior advisor, Dr. Ramo Gençay. I have been amazingly fortunate to have such an advisor who gave me unwavering guidance during my graduate studies at Simon Fraser University. His mentorship was paramount in providing a well rounded experience to help me grow as a scholar but also as an independent thinker. His patience and support helped me overcome many crisis situations and finish this dissertation. I am grateful to him for holding me to a high research standard and enforcing strict validations for each research result, and thus teaching me how to do research.

My advisor, Dr. Alex Karaivanov, has been always there to listen and give advice. I am deeply grateful to him for the long discussions that helped me sort out the details of my work. I am also thankful to him for carefully reading and commenting on countless revisions of this manuscript. His insightful comments and constructive criticisms at different stages of my research were thought-provoking and they helped me focus my ideas.

I am grateful to Dr. Robbie Jones, Dr. Rob Grauer, and Dr. Buz Brock for their encouragement and insightful advice. I am also thankful to them for commenting on my views and helping me understand and enrich my ideas.

I would like to acknowledge Dr. Daneil Monte, Dr. Marie Rekkas, Dr. Simon Woodcock, and Dr. Stephen Easton for discussions on related topics that helped me improve my knowledge in the area.

I am also grateful to the following staff at the Department of Economics, for their various forms of snpport during my graduate study–Kathy Godson, Laura Nielson, Gwen

Wild, Tracey Sherwood, and Tim Coram.

Many friends have helped me stay clam and live free through these difficult years. I greatly value their friendship and I deeply appreciate their belief in me. Special thanks to Frank Han.

Finally and most importantly, none of this would have been possible without the love and patience of my family. My family, to whom this dissertation is dedicated to, has beeu a constant source of love, conceru, support aud strength all these years. I would like to thank my wife Shanshan. Her unyielding devotion, love, support, and encouragement were undeniably the ground upon which the past years of my life have been built. I thank my parents for their faith in me and allowing me to be whom I wanted. It was their quiet patience and support that gave me so much drive and an ability to tackle challenges head on.

Contents

List of Tables

List of Figures

Chapter 1

Trading Frequency and Volatility Clustering

1.1 Introduction

Over the last five decades, a broader picture of the time series features of asset prices has emerged. Among these features, return predictability at high trading frequencies and persistence of the variance of returns have received significant attention both theoretically and empirically. This latter feature is also known as volatility clustering and is unquestionably one of the most important stylized facts of financial time series. Engle (1982, 2000) and Bollerslev (1986) have proposed (G)ARCH-family models, which has been shown to be capable of capturing conditional volatility parsimoniously. In addition, as documented by Mandelbrot (1963), the autocorrelations of the variance of returns decay at a hyperbolic rate rather than exponentially. Several studies, however, have investigated the reasons and mechanisms behind such volatility persistence in market microstructure-type economic models, and have successfully generated volatility clustering.[1]

The microstructure model proposed in this chapter provides a framework for generating volatility clustering of the returns with autocorrelations of hyperbolic decay. In addition, the proposed mechanism is capable of generating a linearly trending price and a negative correlation at the first lag of returns. The formation of volatility clustering is due to the

[1]For example, Brock and LeBaron (1996), Cabrales and Hoshi (1996). Granger and Machina (2006) presents general mechanisms of how a time-invariant system can exhibit volatility clustering, although they do not provide microeconomic models that could lead to the system proposed in their paper.

combined effects of the presence of rational traders with multiple trading frequencies and their strategic interactions. It is natural to model traders with multiple trading frequencies, because not all traders trade at every possible opportunity.[2] The assumption that traders trade strategically is also plausible, since large investors are aware that their trades have an impact on the market price and take this effect into account. Note that no specific assumptions need to be made regarding the informational structure; information can be available either privately or publicly.

Specifically, I consider a discrete-time, multiperiod model in which traders trade a stock that has a limited risk absorption capacity (i.e., an upward sloping supply curve).[3] Traders are divided into two groups according to their trading frequency, while group size is random to prevent perfect signal extraction. Type A traders trade more frequently (trade every trading period) while type B traders trade less frequently (trade every second period).[4] In this model, traders may differ in the following respects: First, they may have different beliefs about fundamentals either due to different initial beliefs in the public signal environment or different realizations of signals in the private signal environment. Second, traders differ in trading strategies due to their different trading frequencies. Let the public signal environment with identical initial priors of fundamentals be the benchmark case. In such a benchmark case, the aggregate demand of type A traders depends on the presence of type B traders. Therefore, although there are no trades between groups, there will be an alternating pattern in price due to the multiple trading frequencies.

When traders of multiple trading frequency behave strategically, the volatility clustering is generated for two reasons. First, multiple trading frequencies lead to an alternating pattern in prices, which generates a serial correlation in the magnitude of returns. The alternating pattern in prices is partly due to the absence of infrequent traders, which lowers the size of the aggregate demand in every second period. The alternating pattern in prices is also partly due to the different strategies used by frequent traders depending on the presence of infrequent traders. Intuitively, frequent traders may behave like monopolists

[2]One example is the futures market, where typical traders are hedgers and speculators. The speculators in futures market generally have shorter trading horizons. Another example is the comparison of intraday traders and mutual fund managers, since mutual fund managers cannot conduct intraday trading due to regulatory restrictions.

[3]To allow for the strategic interaction between traders, I do not allow an infinite supply of the asset, such that large orders would have no impact on price.

[4]For simplicity, I do not model the arrivals of traders endogenously, although the main results will apply with endogenous arrivals.

in the absence of infrequent traders and like oligopolists in their presence. We label this source of volatility clustering *the alternating effect*. Secondly, in the private signal case, each group of traders has its own set of signals. Given different trading frequencies, it is natural for traders in one group to infer the other group's signals from the price. Infrequent traders can infer the signals from the price in the period when they are absent, because the prices are entirely determined by the demands of frequent traders. Therefore, past prices provide information that determines the current price. This feedback mechanism facilitates the formation of the volatility clustering (see, e.g., Brock and LeBaron (1996)). We label this source of volatility clustering *the signal extraction effect*.[5]

When there is strategic interaction between traders, then the group size of traders, i.e., the mean arrival of traders, also has an impact on the optimal strategy of traders. The strategic competition is more intense with larger group sizes. In the limit, the strategic interaction among large groups of traders converges to the competitive outcome. This decreases the persistence of the magnitude of returns significantly. These results show that when group sizes are large, the volatility clustering becomes negligible. Thus, the strategic behavior is necessary for the presence and the persistence of volatility clustering in this model.

There are two additional stylized facts that are generated by our model in addition to volatility clustering, namely, linearly trending prices and a negative correlation at the first lag of returns. The former is mainly due to the optimal trading strategy used by the traders in equilibrium. There are two ways for traders to make a profit in my model. First, traders who hold shares of the stock receive the payoff on the terminal trading date. Traders adjust their optimal holdings according to the realization of their own signals and to the other group's signals that they have extracted. Second, traders will harvest capital gains if they can correctly anticipate the price movement. Because traders are informationally large in this model, they can strategically adjust their holdings across the remaining trading dates in order to take the advantage of their own impact on prices, which leads to trending prices.[6] The negative first-order autocorrelation of returns is consistent with the concept of noisy

[5]This mechanism can generate volatility clustering even in the public signal case where signal extraction is absent. On the other hand, I find that signal extraction without multiple trading frequencies cannot generate volatility clustering.

[6]For example, if they believe that the stock is undervalued in the current period, they may adjust their holdings over several periods instead of just increasing their current period holding which may could drive the price up sharply and diminish the future capital gains.

rational expectation equilibrium (see e.g., Makarov and Rytchkov (2007)). The correlation between realized and expected returns can be shown to depend partly on the correlation between exogenous supply and the current price, which is negative.

This model generates a number of interesting and testable implications that are absent from standard models of asset pricing with uniform trading frequency. For instance, the traders with more precise signals have a marginal effect on the evolution of the equilibrium. This seemingly counterintuitive result can be explained by the fact that traders strategically adjust their optimal holdings over all trading dates. Perhaps the most novel feature of the model is that traders with different trading frequencies have different levels of impact on equilibrium prices and returns, with infrequent traders having a larger effect. This naturally results from the fact that infrequent traders have fewer trading dates to smooth their adjustment of optimal holdings.[7] Furthermore, I show that signal extraction not only helps traders to infer the fundamentals more precisely but also provides them with more accurate guesses as to the behavior of the other type of traders. This leads to greater persistence in the magnitude of returns. Naturally, this provides both a feedback mechanism and a forward mechanism, both of which contribute to the formation of volatility clustering.

Overall, the main contributions of this chapter are as follows. First, rational traders with multiple trading frequencies behaving strategically can generate volatility clustering, and this mechanism is robust with respect to different specifications of informational structure. The qualitative statistical properties of equilibrium including prices, returns and the magnitude of returns, are similar in the public and the private signal settings, with or without the same initial beliefs about fundamentals. Second, multiple trading frequency in the private signal environment can induce signal extraction, which contributes to the formation of volatility clustering and leads to hierarchical information. Hence, multiple trading frequencies within the private information environment provides theoretical justification for the existence of hierarchical information, where the infinite regress problem collapses (see, e.g., Townsend (1983), McNulty and Huffman (1996) and Bomfim (2001)). Third, return predictability is generated and is robust with respect to different informational structures.

Several papers have examined the role of multiple trading frequencies in different environments. For example, Christian and Jia (2005) try to determine the optimal trading

[7] As stated earlier, the proposed mechanism can generate similar stylized facts in various informational settings which may lead to identification problems. This issue can be easily solved by examining the impulse responses of traders with different trading frequencies.

frequency using a technical trading rule. Hauser *et al.* (2001) shows that the higher aggregate trading frequency, the more efficient the price discovery in a non-dealer market. To my knowledge, however, no paper which links the multiple trading frequencies to volatility clustering and hierarchical information. Unlike Hauser *et al.* (2001), who focus on the trading frequency determined by the institution at the aggregate level, this study examines trading frequency at the microstructure level.

There are a number of ways to generate volatility clustering. For instance, Brock and LeBaron (1996) studied asymmetric information, the adaptive beliefs model of stock price and volume, in which volatility clustering is generated from traders experimenting with different belief updating systems, where experimenting is based on the past profits and expected future profits. Cabrales and Hoshi (1996) built a heterogeneous beliefs asset pricing model in which the persistence of distribution of wealth can lead to volatility clustering. The approach in Haan and Spear (1998) was to develop a heterogeneous agent, incomplete market model of interest rates, in which persistence of financial frictions leads to volatility clustering. de Fontnouvelle (2000) investigated a costly information model of asset trading, in which agents need to pay to acquire information, which leads to volatility clustering in price. Timmermann (2001) studied an imperfect information model of asset pricing, in which Bayesian updating of parameter estimates leads to volatility clustering. Hommes (2006) has an excellent survey on generating volatility clustering through interacting agent models, see also Hommes (2008). In such models, these wide range of stylized facts are generated from the interaction between informed and noise traders through the adaptive belief updating and its evolution, as in Brock *et al.* (2005). One of our distinct contributions to this literature is the embedding of a Bayesian Nash equilibrium model which leads to the strategic interaction amongst informed traders. Such interaction in the presence of the multiple trading frequency is capable of generating a set of stylized facts widely observed in the finance literature.

Unlike Brock and LeBaron (1996), this model does not rely on the experimentation between different beliefs npdating systems. Although the feedback mechanism, i.e., the signal extraction, can contribute to the formation of volatility clustering of returns in this model, it is not essential. In this model, volatility clustering is generated even in the public signal environment where signal extraction is absent. Unlike Timmermann (2001), Bayesian information updating is not sufficient to generate volatility clustering in this model. My findings show that without strategic behavior, the volatility clustering is absent even with

Bayesian information updating.

This chapter is organized as follows. Section 1.1 describes the basic setting used in the chapter. Section 1.2 starts with the benchmark environment of the public signal case. In this case, although the setting is simplistic in that all signals are assumed to be publicly available, the model can generate the three stylized facts: First, prices display long memory and an upward sloping trend. Second, returns are stationary and display a negative first-order correlation. Third, the variance of returns (or the magnitude of returns) displays volatility clustering with hyperbolic decay rate. Section 1.3 considers the different information structure of the private signal case, which yields additional interesting results. The private signal case possesses all the features of the public signal case. In addition, multiple trading frequencies lead to signal extraction behavior, which adds to the temporal memory of the variance of returns and generates an information hierarchy. In Section 1.4, Monte Carlo simulations are carried out to illustrate the stylized facts, which are consistent with the theoretical results. Section 1.5 considers two extensions: First, in addition to improving the understanding of fundamentals, signal extraction can also help traders to predict the behaviors of the other group; we label this *sophisticated signal extractian.* Hence, signal extraction not only provides a feedback mechanism but also a forward-looking mechanism that links the prices to future prices. The other extension is to allow heterogeneous priors, which is shown to have negligible effects on the main findings. I conclude afterwards.

1.2 Basic Setting

I model a hypothetical financial market in which there is a single trading asset.[8] Assume that there are two groups of traders on the market, namely *Type A* and *Type B* traders. Type A traders come to the market every period (speculators), while Type B traders come every other period (fundamentalists). Figure 1.1 illustrates this multiple trading frequency market structure. During a trading period, traders receive a signal about the value of the underlying asset. For analytical tractability, I make two simplifying assumptions: trading dates are finite and traders maximize per period profit. The first assumption is innocuous. The myopic preference assumption may cause dissatisfaction, but it helps to avoid the large state variables problem which is endemic in the" forecasting the forecasts of others"

[8] I borrow the notation of Hong *et al.* (2006), who investigate the role of overconfidence in generating speculative bubbles.

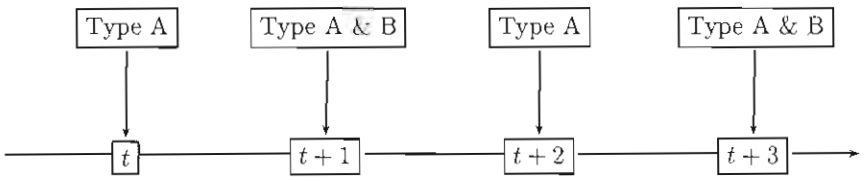


Figure 1.1: A FLOWCHART FOR MULTIPLE TRADING FREQUENCY. Type A traders (speculators) come to the market every trading date while type B traders (fundamentalists) come to the market every second trading date.

literature. However, it turns out that it is not a serious problem to ignore hedging demand when the agent's maximization problem has a quadratic form and the signals are normally distributed.[9]

There is a single traded asset in the economy, with $T+1$ trading dates: $t = 0, 1, \ldots, T$. The asset pays $\widetilde{f}$ at $t = T$, where $\widetilde{f}$ is a normally distributed random variable.[10] The supply function of outstanding shares is $Q_s = p - \alpha$, where Q_s is aggregate liquidity supply and α captures the fixed cost of providing outstanding shares.[11] The supply function takes on a simplest form, because the supply side is not the main focus of this chapter.

Two groups of traders, type A and type B, maximize a per-period objective function $E(W)$, taking into account their own influences on the equilibrium prices,[12] where W is the wealth of the trader. The type A traders come to the market at every date, but type B traders only come at $t = 0, t = 2, \ldots, t = T-1$ if T is odd, or $t = 1, t = 3, \ldots, t = T-1$ if T is even.

The arrivals of traders from both groups are assumed to be random. Random arrivals make it difficulty for traders to distinguish signals from the prices. The effective numbers of traders, n_t^A and n_t^B for type A and type B, are governed by an identical and independent

[9]However, myopia is not a main driver of equilibrium price and return dyuamics in this model. As indicated earlier, there are two channels for the proposed mechanism to generate volatility clustering, namely, "the alternating effect" and "the signal extraction effect". The first channel is a natural consequence of traders with multiple trading frequency. The second channel is naturally embedded in the private information environment. These two channels are independent of the myopic preference of traders.

[10]The payoff $\widetilde{f}$ can be interpreted as the liquidation value of the firm, and may be negative in the case of bankruptcy (considering the opportunity cost).

[11]The supply side is not modeled explicitly. The liquidity supply may come from noise traders or long term investors.

[12]Myopic preferences are adopted to avoid dynamic hedging problem and to obtain an analytically tractable solution.

(*iid*) normal distribution,[13] $N(n, \omega)$.

When $t = 0$, the two groups' prior beliefs about $\widetilde{f}$ are normally distributed, denoted by $N(\widehat{f}_0^A, \frac{1}{\tau_0})$ and $N(\widehat{f}_0^B, \frac{1}{\tau_0})$ where $\widehat{f}_0^A$ and $\widehat{f}_0^B$ can be different. Since I focus on the effect of trading frequency, the same prior beliefs are assumed, i.e., $\widehat{f}_0^A = \widehat{f}_0^B = \widehat{f}_0$. At $t = 1,2, \ldots, T-1$, each trader type receives one signal: $S_t^A = \widetilde{f} + \epsilon_t^A$ and $S_t^B = \widetilde{f} + \epsilon_t^B$ for $t = 1,2, \ldots, T-1$, where ϵ_t^A, ϵ_t^B are *iid* normally distributed $N(0, \frac{1}{\tau_\epsilon})$, where τ_ϵ is the precision of the signal.[14]

I start with the public signal case. Figure 1.2 illustrates the mechanism of price formation in the public signal case.

1.3 Public Signals

1.3.1 Equilibrium

I first solve for the beliefs of the two types of traders at time t. Using standard Bayesian updating formulas, these beliefs are easily characterized by the following proposition.

Proposition 1 *The beliefs of the two groups of traders at t are normally distributed as $N(f_t^A, \frac{1}{\tau_t})$ and $N(f_t^B, \frac{1}{\tau_t})$, where the precision is given by*

$$\tau_t = \tau_{t-1} + 2\tau_\epsilon \tag{1.1}$$

and the means are given by

$$\begin{aligned} \widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t}(S_t^A + S_t^B - 2\widehat{f}_{t-1}^A) \\ \widehat{f}_t^B &= \widehat{f}_{t-1}^B + \frac{\tau_\epsilon}{\tau_t}(S_t^B + S_t^A - 2\widehat{f}_{t-1}^B). \end{aligned} \tag{1.2}$$

If The traders are with identical priors, i.e. $\widehat{f_0^A} = \widehat{f_0^B}$, then their beliefs remain the same over all the trading period, i.e., $\widehat{f}_t^A = \widehat{f}_t^B = \widehat{f}_t$ for $t = 1,2, \ldots, T-1$. Without a loss of generality, I solve the model when T is even. Given beliefs as in Proposition 1, I can solve for the equilibrium asset holdings x_{T-1}^A for type A and x_{T-1}^B for type B traders, and the

[13] The arrival process of traders is independent of the economic variables. To avoid negative arrivals, I put a lower bound on the actual arrivals in the simulation study presented in Section 1.4.

[14] The inverse of the variance of the signal can be interpreted as the precision of the signal. If the variance of the signal increases, the precision of the signal decreases.

prices p_{T-1} at $T-1$. With risk neutral preferences, we can write the utility maximization problem faced by the ith type A trader as

$$\max_{x^A_{T-1,i}} E[\tilde{f} - (\alpha + \sum_{j=1}^{n^A_{T-1}} x^A_{T-1,j} + \sum_{j=1}^{n^B_{T-1}} x^B_{T-1,j})]x^A_{T-1,i}$$

where the first-order condition is

$$\widehat{f}_{T-1} - \alpha - (n-1)x^A_{T-1,j} - nx^B_{T-1,j} - 2x^A_{T-1,i} = 0.$$

Invoking symmetry, we have $x^A_{T-1,i} = x^B_{T-1,j}$ for all i,j in equilibrium, which leads to

$$x^A_{T-1} = \frac{\widehat{f}_{T-1} - \alpha}{2n+1}.$$

Hence, the Bayesian Nash equilibrium at $T-1$ can be characterized by

$$\begin{aligned} x^A_{T-1} &= x^B_{T-1} = x_{T-1} = \frac{\widehat{f}_{T-1} - \alpha}{2n+1} \\ p_{T-1} &= \alpha + (n^A_{T-1} + n^B_{T-1})x_{T-1} = \alpha + (n^A_{T-1} + n^B_{T-1})\frac{\widehat{f}_{T-1} - \alpha}{2n+1} \end{aligned} \tag{1.3}$$

Given the equilibrium price at $T-1$, we can use backward induction to derive the equilibrium holdings and price for $t = T-2$, $t = T-3, \ldots, t = 1$. If t is even, only type A traders arrive at market. Using an argument similar to the $T-1$ case, the optimal demand for the ith trader of type A is

$$\frac{E_t p_{t+1} - \alpha}{n+1}$$

where $E_t p_{t+1} = E[p_{t+1}|I_t]$, which is the conditional expectation of the next period price at time t given the information set I_t available at t, i.e., $I_t = \{p_{t-1}, p_{t-2}, p_{t-3}, \ldots, p_1, p_0\}$. Note that in period t, in order to determine the optimal holdings, traders need to forecast the next period price, p_{t+1}, and the optimal holdings of other traders.[15] If t is odd, both type A traders and type B traders come to the market. Using a similar argument as above, the optimal demand for the type A trader is

$$\frac{E_t p_{t+1} - \alpha}{2n+1}.$$

[15] p_t is not determined when traders make their decisions at time t. The joint decisions of all market participants lead to the equilibrium price p_t.

The equilibrium at $t = 0, 1, 2, \ldots, T-2$ differs from that at $t = T-1$ due to the myopic preference assumption. When $t = 0, 1, 2, \ldots, T-2$, traders only care about the price of the asset one period ahead because of their per period profit orientation. The optimal holding is determined by the forecast of the price in the next period which can be solved by backward induction and the expected holdings of other traders.

Proposition 2 *The Bayesian Nash equilibrium at t can be characterized by*
if t is odd

$$\begin{aligned} x_t^A &= x_t^B = x_t = \frac{2^{(T-t-1)/2} n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}_t - \alpha) \\ p_t &= \alpha + (n_t^A + n_t^B)\frac{2^{(T-t-1)/2} n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}_t - \alpha) \end{aligned}$$

if t is even

$$\begin{aligned} x_t^A &= \frac{2^{(T-t)/2} n^{(T-t-1)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t - \alpha) \\ p_t &= \alpha + (n_t^A)\frac{2^{(T-t)/2} n^{(T-t-1)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t - \alpha). \end{aligned}$$

Based on Proposition 1, Proposition 2 characterizes the Bayesian Nash equilibrium for all trading dates. The equilibrium price is a function of the remaining trading horizon and beliefs, which implies that the price dynamics are governed jointly by the trading horizon and the beliefs of traders.

1.3.2 Equilibrium Properties

The previous section characterized the Bayesian Nash equilibrium in the public signal case. To examine how the mechanism proposed in the chapter, namely, traders with the multiple trading frequencies behaving strategically, can generate the claimed stylized facts, including return predictability and volatility clustering, we need to study the properties of the equilibrium. The main properties of the equilibrium described are

1. Price series possess a linear trend and are linearly dependent.
2. Return series are stationary, and there is a negative autocorrelation of returns at first lag.
3. Return series display volatility clustering with hyperbolic decay autocorrelations.

Before elaborating on each of these properties, I will describe the equilibrium behavior of beliefs where the equilibrium price is a function of beliefs.

Beliefs

First, beliefs exhibit long memory as shown in Lemma 3.

Lemma 3

$$Cov(\widehat{f}_t, \widehat{f}_{t-j}) = \frac{1}{\tau_{t-j}} - \frac{1}{\tau_t}, \tag{1.4}$$

where τ_{t-j} is the precision of beliefs at $t-j$ for $j = 1, 2, 3, \ldots, t-1$.

Remarks:

1. Lemma 3 shows that beliefs have long memory and that the autocovariance function is a hyperbolic function of lags. To illustrate the nature of long memory and hyperbolic decay in beliefs, consider an impulse response experiment. For simplicity, assume that there is only one positive innovation at $t = 1$ and zero at other trading dates. Using Proposition 1, we know that the beliefs for traders at $t = 1, 2, \ldots, T-1$ are

$$\widehat{f}_t = \widehat{f}_0 + \frac{\tau_\epsilon}{\tau_t}(2t\epsilon_0 + \sum_{i=1}^{t}\epsilon_i^A + \sum_{i=1}^{i}\epsilon_i^B) \tag{1.5}$$

   where $\epsilon_0 = \widetilde{f} - \widehat{f}_0$. Given the innovation at $t = 1$, the belief $\widehat{f}_1$ is updated and the effect of the innovation on the belief at $t = 1$ is τ_ϵ/τ_1. Then, the effect of the innovation on the belief at $t+1$, $\widehat{f}_{t+1}$ is τ_ϵ/τ_{t+1} and the effect of the innovation on the belief at $t+j$ is τ_ϵ/τ_{t+j}. Note that $\tau_{t+j} = \tau_0 + 2(t+j)\tau_\epsilon$; this leads to the persistence of beliefs, and the decay rate is hyperbolic.

2. Lemma 3 also characterizes the limiting behavior of the beliefs, which converge to the true value of the underlying asset $\widetilde{f}$ as $t \to \infty$. As $t \to \infty$, $\tau_\epsilon/\tau_t \to 0$, and $2t\tau_\epsilon/\tau_t \to 1$. Equation 1.5 is reduced to $\widehat{f}_t = \widehat{f}_0 + \epsilon_0 = \widehat{f}_0 + \widetilde{f} - \widehat{f}_0 = \widetilde{f}$. This implies that the beliefs of traders converge asymptotically to the true value of the underlying asset.

Prices

For simplicity, I assume that $\alpha = 0$ so that I can work with the logarithm of price series,[16]

$$\begin{aligned} \log p_t &= A_t + \frac{t}{2}B + \log \widehat{f}_t \quad \text{for } t \text{ is odd} \\ \log p_t &= C_t + \frac{t}{2}B + \log \widehat{f}_t \quad \text{for } t \text{ is even} \end{aligned} \tag{1.6}$$

where $A_t = \log(n_t^A + n_t^B) + \frac{T-1}{2}\log 2 + (T-1)\log n - \frac{T+1}{2}\log(2n+1) - \frac{T-1}{2}\log(n+1)$, $B = [\log(2n+1) + \log(n+1) - \log 2 - 2\log n]$, $C_t = \log(n_t^A) + \frac{T}{2}\log 2 + (T-1)\log n - \frac{T}{2}\log(2n+1) - \frac{T}{2}\log(n+1)$.

Remarks:

1. From Equation 1.6, price series contain three components: A_t and C_t, $\frac{t}{2}B$ and $\log \widehat{f}_t$. A_t and C_t are exogenous random variables that are determined by the *iid* arrival process. B is a positive constant that acts as a drift parameter. Therefore, $\log p_t$ has a deterministic linear upward sloping trend. As shown in Equation 1.4, $\widehat{f}_t$ is a long memory process. Therefore the price process also has a long memory which arises from the hyperbolic decay of the beliefs.

2. Intuitively, prices have long memories because of the embedded belief process. Price is a linear function of beliefs and preserves the linear dependence structure of beliefs. As a result, the price series display long memory.

3. The deterministic trend originates from the strategic behaviors of traders. Traders will face a trade-off in deciding whether to increase their holdings. On the one hand, increasing their holdings today means that they can sell more at a higher price tomorrow. On the other hand, increasing their holdings will increase the cost of acquiring shares today. Without strategic behavior,[17] traders are not aware of their own impact on equilibrium price. They will therefore adjust their holdings until the price

[16] Monte Carlo simulations suggest that the results are not sensitive when α is nonzero.

[17] With the beliefs characterized in Proposition 1, we can solve for the equilibrium asset holdings x_{T-1}^A for type A and x_{T-1}^B for type B traders, and prices p_{T-1} at time $T-1$. With risk-neutral preferences, $E[\widetilde{f}] = p_{T-1}$ which implies that $p_{T-1} = \widehat{f}_{T-1}$. Using the market clear condition, we have $x_{T-1}^A = x_{T-1}^B = \frac{\widehat{f}_{T-1} - p_{T-1}}{n_{T-1}^A + n_{T-1}^B}$. This implies that at $T-2$, when only type A traders come to the market, $E[p_{T-1}] = p_{T-2}$, which implies $p_{T-2} = \widehat{f}_{T-2}$ and $p_t = \widehat{f}_t$ for $t = 1, 2, \ldots, T-1$. Hence, prices follow a martingale process and there is no upward sloping trend embedded in the price series.

difference is zero, which leads to zero expected profit for every trading date. With strategic behavior, traders will exploit their monopolistic powers to prevent the increase of prices in the current period to order to make profits. When the remaining trading horizon is longer, the strategic behavior of traders has a larger cumulative impact on price, which implies that the prices increase over time. Notice that this time trend can contribute to the time varying variance of price but not the volatility clustering of returns.

Returns

The gross returns[18] are defined by $r_t = p_t/p_{t-1}$ and the logarithm return $\log(r_t)$ at t is equal to .

$$\log r_t \quad = \quad D_t + \log Z_t$$

where $D_t = \log(n+1) - \log(n) + \log(n_t^A + n_t^B) - \log(2n_{t-1}^A)$ and $Z_t = \widehat{f}_t/\widehat{f}_{t-1}$. D_t is an exogenous random variable which is determined by the arrival process, and Z_t is the ratio of beliefs of two types of traders. In order to understand the properties of r_t, we need to study the time series properties of D_t and Z_t. Since it is difficult to get a closed form for the autocovariance function of Z_t and D_t, we rely on Monte Carlo simulations.

Remarks:

1. The negative first-order autocorrelation of returns is mainly due to the change in the mean intensity of arrivals due to multiple trading frequencies. As shown in Figure 1.3, D_t possesses a negative first-order autocorrelation with the magnitude -0.25 and no statistically significant autocorrelations at higher lags. In the meantime, Z_t possesses no statistically significant autocorrelations at any lag. With the joint effects of Z_t and D_t, the Monte Carlo study suggests that r_t has a negative first-order autocorrelation.

Volatility Clustering

Let $\log(r_t) = D_t + \log Z_t$, while $\log Z_t = log(\widehat{f}_t) - \log(\widehat{f}_{t-1})$. It is easy to show that D_t is a stationary process. Hence I focus on the time series properties of Z_t. Z_t can be recursively

[18] Alternatively, one can define returns by $r_t = \log(p_t) - \log(p_{t-1})$. These two specifications do not alter the core findings.

written as[19]

$$Z_t = 1 + \frac{\tau_{t-1}}{\tau_t}\left(\frac{S_t^A + S_t^B}{S_{t-1}^A + S_{t-1}^B}\left(1 - \frac{\tau_{t-2}}{\tau_{t-1}}\frac{1}{Z_{t-1}}\right) - 2\right), \tag{1.7}$$

which is a nonlinear function of Z_{t-1}.

Remarks:

1. For expositional purposes, I define $Z_t = f_t(Z_{t-1})$. Using Equation 1.7, I have $Var(\log(r_t) = Var[D_t] + Var[Z_t] = Var[D_t] + (\frac{\partial f_t}{\partial Z_{t-1}})^2 Var[Z_{t-1}]$. Remember that Z_{t-1} is a function of r_{t-1}, i.e., $Z_{t-1} = f_t^{-1}(r_{t-1})$. This suggests that $Var(\log(r_t) = Var[D_t] + (\frac{\partial f_t}{\partial Z_{t-1}})^2 Var(f_t^{-1}(r_{t-1}))$. As in Granger and Machina (2006), when Z_t is a nonlinear function of Z_{t-1}, it is evidence of volatility clustering.

2. Volatility clustering is mainly due to the time series properties of Z_t. As shown in Figure 1.3, Var(D_t) possesses no statistically significant autocorrelations at any lag. In contrast, the autocorrelations of Var(Z_t) at the first ten lags are all statistically significant with magnitudes ranging from 0.05 to 0.27. In addition, Var(Z_t) decay with a hyperbolic rate of 0.34.

Impact of Signal Precision

In addition to the stylized facts, this model generates interesting predictions that are absent from the standard asset pricing model. For example, this model predicts that the traders with more precise signals impose smaller effects on changing eqnilibrium prices. It is useful to consider an impulse response experiment. We start with an equilibrium where the beliefs of traders have already converged to the true value of the underlying asset, i.e., $\widehat{f}_0 = \widetilde{f}$. Suppose there is a large negative innovation in the signal at $t = 1$ with the magnitude $-\widetilde{f}$, i.e., $S_1^A = 0$ and $S_1^B = 0$. All other signals are equal to $\widetilde{f}$, i.e., $S_t^A = S_t^B = \widetilde{f}$. There is no noise in arrivals, i.e., $n_t^A = n_t^B = n$ and $\alpha = 0$. Then the beliefs, equilibrium

[19] See derivations in Appendix B.

prices, and returns can be computed according to Proposition 2

$$\begin{aligned}
\widehat{f}_t &= \widetilde{f}(1-\frac{2\tau_\epsilon}{\tau_t}) = \frac{\tau_{t-1}}{\tau_t}\widetilde{f} \\
p_t &= \frac{2^{(T-t+1)/2}n^{T-t}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}\widehat{f}_t \quad \text{for } t \text{ is odd} \\
p_t &= \frac{2^{(T-t)/2}n^{T-t}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}\widehat{f}_t \qquad \text{for } t \text{ is even} \\
r_t &= \frac{\tau_{t-2}}{\tau_t}.
\end{aligned}$$

When the precision of signal τ_ϵ increases, the innovation in the signal has a smaller effect on returns. To see this, note that $r_t = \frac{\tau_{t-2}}{\tau_t}$ can be rewritten as $1 - \frac{4}{\frac{\tau_0}{\tau_\epsilon}+2t}$. When τ_ϵ increases, r_t decreases. This implies that when traders have more precise signal, the change in price is smaller. This seemingly counterintuitive result is caused by the strategic adjustments of the optimal holdings of traders. When the signal is precise, traders tend to be reluctant to adjust their optimal holdings. This leads to traders with more precise signals imposing smaller effects on equilibrium prices and returns.

1.4 Private Signals

The model developed in the previous section is simple, yet capable of capturing the three stylized facts of financial data. Namely, prices display long memory and an upward sloping trend, returns are stationary and display a negative first-order correlation, and the variance of returns (magnitude of returns) displays volatility clustering with a hyperbolic decay rate. This section examines a variant of the model where the traders receive private signals instead of public signals. Signal extraction due to multiple trading frequencies in a private signal environment naturally contributes to the formation of volatility clustering. Intuitively, given different trading frequencies, it is natural for traders in one group to infer the other group's signals from the prices. Infrequent traders can infer the signals from the price in the period when they are absent, because the prices are entirely determined by the demands of the frequent traders. Therefore, the past prices provide information that determines the current price. This feedback mechanism facilitates the formation of the volatility clustering (see, e.g., Brock and LeBaron (1996)). Signal extraction generates other interesting findings as well, which provide insights into understanding private information trading in the market microstructure. For example, signal extraction in this model generates an

information hierarchy among traders in an *ex ante* symmetric information setting.[20] In this model, infrequent traders can infer signals received by frequent traders exactly. In addition, infrequent traders infer more precise signals than frequent traders. The information hierarchy created by multiple trading frequencies suggests that the asymmetry of information diffusion may be endogenously determined by trading frequencies, rather than exogenously given.

The importance of the private signal case is that it leads to signal extraction through the channel of multiple trading frequency. If the traders with several trading frequencies come to the market, we should expect to see "the alternating effect" lessen because of the smoothing effect of these traders. However, we should expect to see "the signal extraction effect" strengthen the impact on the formation of "volatility clustering". This is due to the fact that traders with several trading frequencies lead to hierarchical information through signal extraction behavior.

1.4.1 Informational Structure

Note that at $t = 0, 2, \ldots, T-4, T-2$, only type A traders are present in the market. This implies that the equilibrium prices in such periods are entirely determined by the behavior and beliefs of type A traders. Therefore, type B traders can extract the beliefs of the type A traders by inverting the equilibrium price function of the beliefs.

I assume that traders know the exact number of traders who came to the market in the last period, i.e., at time t, the information set for the traders is $I_t = \{P_{t-1}, N^A_{t-1}, N^B_{t-1}\}$, where $P_{t-1} = \{p_{t-1}, p_{t-2}, \ldots, p_1, p_0\}$, and $N^A_{t-1} = \{n^A_{t-1}, n^A_{t-2}, \ldots, n^A_0\}$ and $N^B_{t-1} = \{n^B_{t-1}, n^B_{t-2}, \ldots, n^B_0\}$.

1.4.2 Bayesian Nash Equilibrium

To describe signal extraction behavior and the evolution of beliefs, I first characterize the Bayesian Nash equilibrium. Given a sequence of beliefs after signal extraction, $\widehat{f}^A_t$ and $\widehat{f}^B_t$ for type A traders and type B traders respectively, we can solve for the equilibrium at time $T-1$.

[20] A symmetric information setting means that *ex ante*. all traders will receive the same number of signals per period, which are drawn from *iid* distribution.

With risk-neutral preferences, we can characterize the utility maximization problem faced by the *ith* type A trader as

$$\max_{x^A_{T-1,i}} E[\widetilde{f} - (\alpha + \sum_{j=1}^{n^A_{T-1}} x^A_{T-1,j} + \sum_{j=1}^{n^B_{T-1}} x^B_{T-1,j})]x^A_{T-1,i}.$$

The first-order condition is

$$\begin{aligned} \widehat{f}^A_{T-1} - \alpha - (n-1)x^A_{T-1,j} - nx^B_{T-1,j} - 2x^A_{T-1,i} &= 0 \\ \Rightarrow x^A_{T-1,i} &= \frac{\widehat{f}^A_{T-1} - \alpha - nx^B_{T-1,j}}{n+1}. \end{aligned}$$

Hence, the Bayesian Nash equilibrium at $T-1$ can be characterized by

$$\begin{aligned} x^A_{T-1} &= \frac{(n+1)\widehat{f}^A_{T-1} - n\widehat{f}^B_{T-1} - \alpha}{2n+1} \\ x^B_{T-1} &= \frac{(n+1)\widehat{f}^B_{T-1} - n\widehat{f}^A_{T-1} - \alpha}{2n+1} \\ p_{T-1} &= \alpha + n^A_{T-1}x^A_{T-1} + n^B_{T-1}x^B_{T-1}. \end{aligned}$$

Given the price in the $T-1$ period, we can derive the equilibrium for $t = 1, 2, 3, \ldots, T-2$:

Proposition 4 *The Bayesian Nash equilibrium at t is characterized by*
if t is odd:

$$\begin{aligned} x^A_t &= \frac{2^{(T-t-1)/2}n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}(\widehat{f}^A_t - \alpha) - \frac{2^{(T-t-1)/2}n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}^B_t - \alpha) \\ x^B_t &= \frac{2^{(T-t-1)/2}n^{(T-1-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}(\widehat{f}^B_t - \alpha) - \frac{2^{(T-t-1)/2}n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}^A_t - \alpha) \\ p_t &= \alpha + n^A_t x^A_t + n^B_t x^B_t \end{aligned}$$

if t is even:

$$\begin{aligned} x^A_t &= \frac{2^{(T-t)/2}n^{(T-1-t)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}^A_t - \alpha) \\ p_t &= \alpha + (n^A_t)\frac{2^{(T-t)/2}n^{(T-1-t)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}^A_t - \alpha). \end{aligned}$$

1.4.3 Signal Extraction

Using Proposition 4, we can describe the signal extraction behavior of traders. At $t = T-1$, type B traders know the actual arrivals n^A_{T-2} and n^B_{T-2} of both types of traders. and price p_{T-2} in the last period. From Proposition 8, $p_{T-2} = \alpha + (n^A_t)\frac{2n}{(2n+1)(n+1)}(\widehat{f}^A_{T-2} - \alpha)$. Therefore, the type B traders can invert the price formula to get $\widehat{f}^A_{T-2} = \alpha + \frac{(2n-1)(n+1)}{(2n)n^A_{T-2}}(p_{T-2} - \alpha)$. Hence, at $T-1$, type B traders know type A traders' belief at $T-2$, $\widehat{f}^A_{T-2}$. A similar analysis can be applied to type A traders and to other trading dates.

Proposition 5 *Type A traders know the exact beliefs of type B traders every other trading period, and type B traders know the exact beliefs of type A traders every trading period, i.e.,*

1. *When t is odd, type B traders know* $\widehat{f}^A_{t-1}, \widehat{f}^A_{t-2}, \ldots, \widehat{f}^A_0$ *while type A traders only know* $\widehat{f}^B_{t-2}, \widehat{f}^B_{t-4}, \ldots, \widehat{f}^B_0$.

2. *When t is even, type B traders know* $\widehat{f}^A_{t-1}, \widehat{f}^A_{t-2}, \ldots, \widehat{f}^A_0$ *while type A traders only know* $\widehat{f}^B_{t-1}, \widehat{f}^B_{t-3}, \ldots, \widehat{f}^B_0$.

Furthermore, traders can recover the private signals received by the other group by knowing the history of the beliefs. For instance, type B traders know the full history of $\widehat{f}^A_t$, and they understand that the difference in the beliefs is due to the signals. By inverting the Bayesian updating formula, they can even infer the private signals received by type A traders in addition to their beliefs.

Assuming that the identical initial prior beliefs $\widehat{f}^A_0 = \widehat{f}^B_0$ are common knowledge, then the beliefs of each type of traders can be determined as follows:

Proposition 6 *The beliefs of two groups of traders at t are normally distributed, denoted by* $N(f^A_t, \frac{1}{\tau_t})$ *and* $N(f^B_t, \frac{1}{\tau_t})$,*where the precision for type A traders is given by*

$$\begin{aligned}
\tau^A_1 &= \tau^A_0 + \tau_\epsilon \\
\tau^A_2 &= \tau^A_1 + 2\tau_\epsilon \\
\tau^A_t &= \tau^A_{t-1} + \tau_\epsilon, && \text{for } t \text{ is } 2k-1,\ k \geq 2 \\
\tau^A_t &= \tau^A_{t-1} + \tau_\epsilon + \widehat{\tau}^A_t, && \text{for } t \text{ is } 2k,\ k \geq 2
\end{aligned}$$

where $\widehat{\tau}^A_t = \frac{1+(\tau^B_{t-1})^2}{(1+\tau^B_{t-1})^2}\tau_\epsilon$.

The mean for type A traders is given by

$$\begin{aligned}
\widehat{f}_1^A &= \widehat{f}_0^A + \frac{\tau_\epsilon}{\tau_1}(S_1^A - \widehat{f}_0^A) \\
\widehat{f}_2^A &= \widehat{f}_1^A + \frac{\tau_\epsilon}{\tau_2^A}(S_2^A + S_1^B - 2\widehat{f}_1^A) \\
\widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t^A}(S_t^A - \widehat{f}_{t-1}^A) \qquad \text{for } t = 2k-1,\ k \geq 2 \\
\widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t^A}(S_t^A - \widehat{f}_{t-1}^A) + \frac{\widehat{\tau}_t^A}{\tau_t^A}(\widehat{S}_t^A - \widehat{f}_{t-1}^A) \qquad \text{for } t = 2k,\ k \geq 2
\end{aligned}$$

where $\widehat{S}_t^A = \frac{\tau_{t-1}^B S_{t-2}^B}{1+\tau_t^B{}_1} + \frac{S_{t-1}^B}{1+\tau_t^B{}_1}$.

The precision for type B traders is given by

$$\begin{aligned}
\tau_1^B &= \tau_0^B + \tau_\epsilon \\
\tau_t^B &= \tau_{t-1}^B + 2\tau_\epsilon \quad \text{for } t \geq 2
\end{aligned}$$

and the mean for type B traders is given by

$$\begin{aligned}
\widehat{f}_1^B &= \widehat{f}_0^B + \frac{\tau_\epsilon}{\tau_1^B}(S_1^B - \widehat{f}_0^B) \\
\widehat{f}_t^B &= \widehat{f}_{t-1}^B + \frac{\tau_\epsilon}{\tau_t^B}(S_t^B + S_{t-1}^A - 2\widehat{f}_{t-1}^B) \quad \text{for } t \geq 2
\end{aligned}$$

Remarks:

1. Type B traders know the exact private signals received by type A traders, where type A traders only have estimates of the type B traders' private signals ($\widehat{S}_t^A$). This implies that there is an informational hierarchy among traders. Type B traders know all the signals that type A traders know. Notice that we start from an *ex ante* symmetric setting.

2. Type B traders extract higher precision signals: the signals extracted by type B traders are of precision τ_ϵ, while the signals extracted by type A traders are of precision $\widehat{\tau}_t^A = \frac{1+(\tau_{t-1}^B)^2}{(1+\tau_{t-1}^B)^2}\tau_\epsilon$. It is easy to see that $(1+\tau_{t-1}^B)^2 > 1+(\tau_{t-1}^B)^2$. Therefore, $\widehat{\tau}_t^A = \frac{1+(\tau_{t-1}^B)^2}{(1+\tau_{t-1}^B)^2}\tau_\epsilon < \tau_\epsilon$.[21]

[21] An exception occurs at $t = 2$. This is the only trading date where type A traders can extract an exact signal received by type B traders. This exception originates from the assumption that initial beliefs are identical. When I relax this assumption in Section 1.5, type A traders are no longer able to extract exact signals.

3. Signal extraction facilitates the formation of volatility clustering by changing the formation of the beliefs. It is easy to see that due to signal extraction, there is a correlation between the traders' beliefs which is absent in the public signal case. The price is linear in beliefs and preserves the correlation between the traders' beliefs in its own correlation across trading dates.[22]

1.4.4 Equilibrium Properties

The previous section, I characterized the Bayesian Nash equilibrium in the public signal case. In order to examine how the mechanism proposed in the chapter, namely, traders with the multiple trading frequencies behaving strategically can generate return predictability and volatility clustering, we need to study the properties of the equilibrium. In addition, I are going to examine the difference between the equilibria in the private signal case and the public signal case. The main properties of the equilibrium in the private signal case are

1. Price series possess a linear trend and are linearly dependent.
2. Return series are stationary and there is a negative autocorrelation of returns at first lag.
3. Return series display volatility clustering with hyperbolic decay autocorrelations.

Before elaborating on each of these properties, I will describe the equilibrium behavior of beliefs where the equilibrium price is a function of beliefs.

Beliefs

First, beliefs exhibit long memory as shown in Lemma 7.

[22] In addition to the beliefs channel, signal extraction can facilitate the formation of volatility clustering by providing a forward looking mechanism. As discussed in Section 1.5, when traders use signal extraction not only to improve the understanding of fundamentals (via the beliefs channel) but also to predict the behaviors of the other group of traders, the effect of signal extraction on the formation of volatility clustering is stronger.

Lemma 7 *The autocovariance functions of beliefs $\widehat{f}_t^A$ and $\widehat{f}_t^B$ can be characterized by*

$$\begin{aligned} Cov(\widehat{f}_t^A, \widehat{f}_{t-j}^A) &= \frac{1}{\tau_{t-j}^A} - \frac{1}{\tau_t^A} \\ Cov(\widehat{f}_t^B, \widehat{f}_{t-j}^B) &= \frac{1}{\tau_{t-j}^B} - \frac{1}{\tau_t^B} \\ Cov(\widehat{f}_t^A, \widehat{f}_{t-j}^B) &= \frac{\sum_{k=1}^{[t/2]} \widehat{\tau}_k^A}{\tau_t^A \tau_{t-j}^B}, \end{aligned}$$

where $[x]$ is the integer part of x.

Remarks:

1. As in the public signal case, the signal innovation will have an impact on the beliefs even at long lags and the decay rate is hyperbolic. To see this, rewrite the equations for the beliefs in the form

$$\begin{aligned} \widehat{f}_t^A &= \widehat{f}_0 + \frac{\tau_\epsilon}{\tau_t^A}(t\epsilon_0 + \sum_{i=1}^{t} \epsilon_i^A) + \frac{\sum_{k=1}^{t/2} \widehat{\tau}_t^A \epsilon_0}{\tau_t^A} + \frac{\sum_{k=1}^{t/2} \widehat{\tau}_t^A \widehat{\epsilon}_{2k}^A}{\tau_t^A} \\ \widehat{f}_t^B &= \widehat{f}_0 + \frac{\tau_\epsilon}{\tau_t^B}((2t-1)\epsilon_0 + \sum_{i=1}^{t-1}(\epsilon_i^A + \epsilon_i^B) + \epsilon_t^B) \end{aligned} \tag{1.8}$$

   where $\epsilon_0 = \widetilde{f} - \widehat{f}_0$ and $\widehat{\epsilon}_t^A = \widehat{S}_t^A - \widetilde{f}$. Using a similar impulse response experiment as in the public signal case, one can show that the beliefs have long memory and that the decay rate is hyperbolic.

2. It then follows that when $t \to \infty$, the beliefs of each group converge asymptotically to the true value of the underlying asset.

3. The correlation between beliefs is nonzero. It originates from the signal extraction behavior. For instance, as shown in Proposition 6, type B traders know the private signals received by type A traders in the last period. Hence, there is a common signal incorporated in the beliefs of type B traders in the current trading period and type A traders in the last trading period. This imposes a correlation structure on the beliefs of the traders. It turns out that the correlation structure of the beliefs is preserved in the prices, since the price is linear in beliefs.

Prices

Next, I examine the properties of prices. The prices characterized in Proposition 4 can be rewritten as

$$\begin{aligned} p_t &= \alpha + (n_t^A - \frac{n}{n+1}n_t^B)K_t(\widehat{f}_t^A - \alpha) + (n_t^B - \frac{n}{n+1}n_t^A)K_t(\widehat{f}_t^B - \alpha) \quad \text{for } t \text{ is odd} \\ p_t &= \alpha + \frac{2^{1/2}(2n+1)^{1/2}}{(n+1)^{3/2}}n_t^A K_t(\widehat{f}_t^A - \alpha) \quad \text{for } t \text{ is even} \end{aligned}$$

where $K_t = \frac{2^{(T-t-1)/2}n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}$.

For simplicity, I assume that $\alpha = 0$ and that there is no noise in the arrivals, i.e, $n_t^A = n_t^B = n$. Therefore, the prices can be characterized by:

$$\begin{aligned} p_t &= \frac{n}{n+1}K_t(\widehat{f}_t^A + \widehat{f}_t^B) \quad \text{for } t \text{ is odd} \\ p_t &= \frac{2^{1/2}(2n+1)^{1/2}n}{(n+1)^{3/2}}K_t(\widehat{f}_t^A) \quad \text{for } t \text{ is even} \end{aligned} \tag{1.9}$$

where $K_t = \frac{2^{(T-t-1)/2}n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}$.

Remarks:

1. As in the public signal case, (log) price has a linear trend over time: $\log K_t$ is linear in t. Therefore K_t increases exponentially.[23] Again, the deterministic trend arises from strategic behaviors of traders. Due to differences in beliefs, they may adjust their optimal holdings at different rates. This further coutributes to the alternating pattern in prices.

2. The prices display long memories because of the embedded belief process. Likewise, the price is linear in beliefs and preserves the dependence structure of the beliefs. Formally, the autocovariance function of prices can be characterized by

$$\begin{aligned} \text{Cov}(p_t, p_{t-2j-1}) &= \frac{n}{n+1}\frac{(n+1/2)^{1/2}n}{(n+1)^{3/2}}K_t K_{t-2j-1}\text{Cov}(\widehat{f}_t^A + \widehat{f}_t^B, \widehat{f}_{t-2j-1}^A) \\ \text{Cov}(p_t, p_{t-2j}) &= \frac{n^2}{n+1^2}K_t K_{t-2j}\text{Cov}(\widehat{f}_t^A + \widehat{f}_t^B, \widehat{f}_{t-2j}^A + \widehat{f}_{t-2j}^B) \end{aligned}$$

[23] $\log K_t = (T-1)/2\log 2 + (T-1)/2\log n - (T-1)/2\log(2n+1) - (T-3)\log(n+1) + t/2[\log(2n+1) + \log(n+1) - \log 2 - 2\log n]$.

where

$$\begin{aligned}
\mathrm{Cov}(\widehat{f}_t^A + \widehat{f}_t^B, \widehat{f}_{t-j}^A + \widehat{f}_{t-j}^B) &= \frac{1}{\tau_{t-j}^A} - \frac{1}{\tau_t^A} + \frac{1}{\tau_{t-j}^B} - \frac{1}{\tau_t^B} + \frac{\sum_{k=1}^{[t/2]} \widehat{\tau}_k^A}{\tau_t^A \tau_{t-j}^B} + \frac{\sum_{k=1}^{[t/2]} \widehat{\tau}_k^A}{\tau_{t-j}^A \tau_t^B} \\
\mathrm{Cov}(\widehat{f}_t^A + \widehat{f}_t^B, \widehat{f}_{t-j}^A) &= \frac{1}{\tau_{t-j}^A} - \frac{1}{\tau_t^A} + \frac{\sum_{k=1}^{[t/2]} \widehat{\tau}_k^A}{\tau_t^A \tau_{t-j}^B}.
\end{aligned}$$

Returns

Next, I examine the return series. Define $r_t = p_t/p_{t-1}$. Since it is difficult to get a closed form for the autocovariance function, I rely on Monte Carlo simulations. Monte Carlo simulations show that the autocorrelation of r_t at the first lag is -0.25 and is statistically significant, while the autocorrelations at other lags are statistically indistinguishable from zero.

Volatility Clustering

Let $r_t = X_t + Y_t$, where $X_t = \frac{\widehat{f}_t^A}{2\widehat{f}_{t-1}^A}$ and $Y_t = \frac{\widehat{f}_t^B}{2\widehat{f}_{t-1}^A}$. X_t can be rewritten as

$$X_t = 1 + \frac{\tau_{t-1}}{\tau_t}\left(\frac{S_t^A}{S_{t-1}^A}\left(1 - \frac{\tau_{t-2}}{\tau_{t-1}}\frac{1}{X_{t-1}}\right) - 1\right)$$

Remarks:

1. For expositional purposes, I define $X_t = m_t(X_{t-1})$. Using the same methods of analysis as in the public signal case, we obtain

$$\begin{aligned}
\mathrm{Var}(r_t) &= \mathrm{Var}(Y_t) + \mathrm{Var}(X_t) + 2\mathrm{Cov}(X_t, Y_t) \\
&= \mathrm{Var}(Y_t) + \left(\frac{\partial m_t}{\partial X_{t-1}}\right)^2 \mathrm{Var}(X_{t-1}) + 2\mathrm{Cov}(X_t, Y_t)
\end{aligned}$$

Remember that X_{t-1} is a function of r_{t-1}, i.e., $X_{t-1} = m_t^{-1}(r_{t-1})$. This suggests that $Var(r_t) = \mathrm{Var}(Y_t) + (\frac{\partial m_t}{\partial X_{t-1}})^2 \mathrm{Var}(m_t^{-1}(r_{t-1})) + 2\mathrm{Cov}(X_t, Y_t)$. As in Granger and Machina (2006), when X_t is a nonlinear function of X_{t-1}, it is evidence of volatility clustering.

2. Signal extraction adds terms to the expression for volatility clustering. $\mathrm{Cov}(X_t, Y_t)$ is the covariance between the beliefs ratios of type A traders, $\frac{\widehat{f}_t^A}{2\widehat{f}_{t-1}^A}$ and $\frac{\widehat{f}_t^B}{2\widehat{f}_{t-1}^A}$. As shown in Lemma 7, the covariance between beliefs is positive. This shows that signal extraction facilitates the formation of volatility clustering.

Role of Multiple Trading Frequency

Through a multiple trading frequency mechanism, this model generates interesting predictions that are absent from the standard asset pricing model. I already discussed one example in the public signal case, namely, that signal precision has an impact on the equilibrium prices and returns. With signal extraction, this model is capable of generating other interesting predictions.

1. Traders with different trading frequencies have different levels of impact on equilibrium prices and returns. To see this, it is useful to consider a heuristic argument. Let us compare the cumulative impact of one innovation in the signal for type A traders or type B traders. As shown in Equation 1.8, the cumulative impulse response of signal for type A traders is $\sum_{i=1}^{t} \frac{\tau_\epsilon}{\tau_t^A} + \sum_{i=1}^{t} \frac{\tau_\epsilon}{\tau_t^B}$. In contrast, the cumulative impulse response of signal for type B traders is $\sum_{i=1}^{t} \frac{\widehat{\tau}_t^A}{\tau_t^A} \frac{1}{1+\tau_{t-1}^B} + \sum_{i=1}^{t} \frac{\tau_\epsilon}{\tau_t^B}$. Note that $\widehat{\tau}_t^A < \tau_\epsilon$ and $\frac{1}{1+\tau_{t-1}^B} < 1$. Given that $\widehat{\tau}_t^A$ is positive every other period, the cumulative impulse response of the infrequent traders to a signal is smaller. Notice that the smaller cumulative impulse response is due to the fact that type A traders are not capable of doing signal extraction every trading period and their signal precision is lower.

2. The timing schedule of signals has an impact on equilibrium prices and returns. It arises from the fact that type A traders are only able to extract exactly the signal of type B traders on $t = 2$. If the bad news arrives at a later time, type A traders incorporate it into their beliefs differently.

1.5 Simulations

This section illustrates how the model is able to generate the claimed stylized facts, namely, linearly trending prices, negative first-order autocorrelation of returns and volatility

clustering. In the simulations, the parameters are set at $\alpha = 100$, $T = 150$, $n = 50$, $\tilde{f} = 150$, $\tau_0 = 10$, $\tau_\epsilon = 3$ and $\omega = 400$. I also report the robustness of the findings with different sets of parameter values.

1.5.1 Public Signals

The simulatiou results illustrate that (log) prices are linearly trending, returns are stationary and possess a negative first lag autocorrelation, which is consistent with the literature (see, e.g., Dacorogna *et al.* (2001)). In addition, as shown in Figure 1.4, the autocorrelation function of the variauce of returns decays hyperbolically. Table 1.3 reports the average autocorrelations of r_t and $\mathrm{Var}(r_t)$ across 100 simulations in the public sigual case. It shows that for r_t, there is a statistically significant negative autocorrelation at the first lag with a magnitude of -0.48.[24] There are no statistically siguificant autocorrelations at other lags. For $\mathrm{Var}(r_t)$, all autocorrelations are statistically siguificant, with magnitudes rangiug from 0.07 to 0.347.

The fixed cost of providing liquidity, α, mainly affects the position of the prices. Varyiug α does not alter the pattern of the prices and returns. Iu addition, changing α has negligible effects on the dependence structure of returns and the variance of returns. Chauges in α mainly change the position of prices but not the slope of price series. Because the fixed cost is constant over time, the return series will be independent of α. Figure 1.5 shows the simulation results for $\alpha = 80$, $\alpha = 150$ and $\alpha = 200$. The simulation results are consistent with our intuition. The dependence structure of the returns and the variance of returns are almost the same for different α values.

Next, I examine the meau arrivals of traders of each group, n. As a common property of the Cournot game, as n increases, the prices become more volatile. In the mean time, the returns display less dependence structure, and less dependence structure at the second moment (the variance of returns). Intuitively, as n increases, the strategic outcome converges to the competitive outcome, which involves less strategic behavior and rednces the dependence structure of the returns and the variance of returns. Fignre 1.6 shows the autocorrelation functiou of returns and the variance of returns when $n = 50$, $n = 100$, $n = 200$.

[24] The magnitude of the first order autocorrelation of returns is larger than is observed empirically. However, if traders are price takers instead of strategic traders, the first order autocorrelation coefficient of returns is not statistically significant from zero. In practice, the market participants should be a mixture of small traders who are price takers and institution traders who have influence on price. The interaction of the mixture of traders may lead to a lower first order autocorrelation coefficient in magnitude.

The simulations show that as n increases, the variance of returns displays less dependence structure. When $n = 200$, the autocorrelation coefficients of the variance of returns at all but the first lag are very small. In contrast, when $n = 50$, the variance of returns displays evidence of volatility clustering.

I continue to study the effects of the variance of arrivals, ω. Figure 1.7 suggests that ω has a negligible effect on the dependence structure of return and the variance of returns. It mainly affects the volatility of prices. If ω is lower, then the prices become smoother (less volatile). This is consistent with our intuition that ω is the parameter characterizing the arrival process only, which is independent of the updating procedure. This means that ω should not have any explanatory power in the dependence structure as opposed to n, which can affect the dependence structure of the variance of returns by affecting the interaction between traders.

Finally, I investigate the effects of τ_ϵ, the precision of the signal. When the precision of the signal is high, it tends to put more weight on the signals rather than the beliefs of the last period. Therefore, the potential high precision signal τ_ϵ can reduce the dependence structure. The simulation results suggest that the magnitude of this effect is small.

1.5.2 Private Signals

The simulation results are quite similar to the public signal case. Price series have an upward sloping trend and there is a dependence structure in returns, while the returns are stationary. The first-order autocorrelation coefficient of the returns is negative. As in Figure 1.8, the autocorrelation function of the variance of returns exhibits hyperbolic decay, which is evidence of volatility clustering. Table 1.4 reports the average autocorrelations of r_t and $\mathrm{Var}(r_t)$ in 100 simulations. It shows that for r_t, the autocorrelation at the first lag is -0.25, which is statistically significant, while the autocorrelations at other lags are statistically indistinguishable from zero. For $\mathrm{Var}(r_t)$, all the autocorrelations are statistically significant. The magnitude ranges from 0.1 to 0.29. Compared to the public signal case, the persistent structure of $\mathrm{Var}(r_t)$ is of a similar magnitude.

I also present the results of the experiments in terms of parameters in Figure 1.9, Figures 1.10 and 1.11. The model behavior resembles the public signal case. The dependence structures of returns and the variance of returns, which are captured by the autocorrelation function, are similar when changing the carrying cost, α. Changing the mean intensity of arrival n has negligible effects on dependence structure of returns. As n increases, however,

the variance of returns becomes less persistent, which is reflected in smaller magnitude of the autocorrelation coefficients. Different valnes of the standard deviation of arrivals ω have negligible effects on thc dependence structure of rcturns and the variance of rcturns.

To summarize, this simulation provides an illustration of the robustness of the proposed mechanism. In both the public and the private signal cases, the simnlation demonstrates the three stylized facts. First, the prices display long memory and an upward sloping trend. Second, the returns are stationary and display a negative first-order correlation. Third, the variance of returns (the magnitude of returns) display volatility clustering and the decay rate is hyperbolic. Furthermore, the simulations suggest that there is an inverse relationship between the mean intensity of the arrivals and the persistence structure of the variance of returns. In both cases, as the mean intensity of the arrivals n increases, the magnitudes of the autocorrelations become smaller, indicating that the variance of returns is less persistent. As guided by the theoretical framework, it is reasouable to believe that this diminishing effect of mcan intcnsity of the arrivals is due to less strategic bchaviors of traders, as the Bayesian Nash equilibrium converges to the competitive equilibrium. This framework suggests that strategic behavior contributes to the persistent structure in the variance of returns.

1.6 Extensions

1.6.1 Sophisticated Guess

The previous analysis shows that signal extraction facilitates the formation of volatility clustering through a beliefs channel. It is interesting to consider additional mechanisms led by signal extraction. In additional to the beliefs channel, traders seek the short-run profit opportunities by predicting the other party's behavior using the extracted beliefs. Intuitively, this forward looking mechanism will bring even more persistence to the prices and the magnitude of returns.

For expositional purposes, first consider the Bayesian Nash equilibrium at time $T-1$. Time $T-1$ is the last trading date. Therefore, traders only care about the true value of the underlying asset. The profit maximization problem faced by the *ith* type A trader can

be characterized by

$$\max_{x^A_{T-1,i}} E[\tilde{f} - (\alpha + \sum_{j=1}^{n^A_{T-1}} x^A_{T-1,j} + \sum_{j=1}^{n^B_{T-1}} x^B_{T-1,j})]x^A_{T-1,i}.$$

Therefore the equilibrium at time $T-1$ is the same as in the signal extraction case we studied in Section 1.3

$$\begin{aligned} x^A_{T-1} &= \frac{(n+1)\widehat{f}^A_{T-1} - n\widehat{f}^B_{T-1} - \alpha}{2n+1} \\ x^B_{T-1} &= \frac{(n+1)\widehat{f}^B_{T-1} - n\widehat{f}^A_{T-1} - \alpha}{2n+1} \\ p_{T-1} &= \alpha + n^A_{T-1}x^A_{T-1} + n^B_{T-1}x^B_{T-1}. \end{aligned}$$

At time $T-2$, only type A traders come to the market, and they care about the per period profit instead, i.e., they care about the price in the next period p_{T-1} instead of $\tilde{f}$. Therefore, the profit maximization problem faced by the *ith* type A trader at time $T-2$ is

$$\max_{x^A_{T-2,i}} E^A[p_{T-1} - (\alpha + \sum_{j=1}^{n^A_{T-2}} x^A_{T-2,j} + \sum_{j=1}^{n^B_{T-2}} x^B_{T-2,j})]x^A_{T-2,i}$$

which implies that

$$x^A_{T-2} = \frac{E^A[p_{T-1}] - \alpha}{n+1}.$$

By backward induction,

$$\begin{aligned} E^A[p_{T-1}] - \alpha &= E^A[n^A_{T-1}x^A_{T-1} + n^B_{T-1}x^B_{T-1}] \\ &= nE^A[\frac{n}{2n+1}(\widehat{f}^A_{T-1} + \widehat{f}^B_{T-1} - 2\alpha)] \\ &= \frac{n^2}{2n+1}(E^A[\widehat{f}^A_{T-1}] + E^A[\widehat{f}^B_{T-1}]). \end{aligned}$$

At $t = T-2$, the best guess for $\widehat{f}^A_{T-1}$ is $\widehat{f}^A_{T-2}$, the current mean of the beliefs. Without signal extraction, the best guess for $\widehat{f}^B_{T-1}$ is also $\widehat{f}^A_{T-2}$, because there is no further information on the beliefs of type B traders. With signal extraction, type A traders may be able to find a better guess instead of their own belief, for instance, $\widehat{f}^B_{T-3}$. One direct impact of this new guess is that signal extraction behavior can alter the optimal holdings of traders and the equilibrium price.

Given a sequence of beliefs after signal extraction, $\widehat{f}^A_t$ and $\widehat{f}^B_t$ for type A and type B traders, I can characterize the Bayesian Nash equilibrium as:

Proposition 8 *The Bayesian Nash equilibrium at t can be characterized by*

$$\begin{aligned}
x_t^A &= \frac{2^{(T-t-3)/2}n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}(\widehat{f}_t^A + \widehat{f}_{t-2}^B - 2\alpha) \\
&\quad - \frac{2^{(T-t-3)/2}n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}_{t-1}^A + \widehat{f}_t^B - 2\alpha) \\
x_t^B &= \frac{2^T n^{(T-t-1)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}(\widehat{f}_t^B + \widehat{f}_{t-1}^A - 2\alpha) \\
&\quad - \frac{n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}_t^A + \widehat{f}_{t-2}^B - 2\alpha) \\
p_t &= \alpha + n_t^A x_t^A + n_t^B x_t^B
\end{aligned}$$

if t is odd;

$$\begin{aligned}
x_t^A &= \frac{2^{(T-t-2)/2}n^{(T-t-1)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t^A + \widehat{f}_{t-1}^B - 2\alpha) \\
p_t &= \alpha + (n_t^A)x_t^A
\end{aligned}$$

if t is even;

Fignre 1.13 demonstrates the simulation result and Table 1.5 demonstrates the average autocorrelations of r_t and $\mathrm{Var}(r_t)$ across 100 simulations. We can see that the variance of r_t displays a more persistent structure which is reflected by larger magnitude of autocorrelations ranging from 0.168 to 0.383. This suggests that signal extraction can affect the formation of prices and induce more persistence in the variance of returns.

1.6.2 Heterogeneous Priors

Previous analysis assumed that the prior beliefs are the same for both types of traders. In this section, I extend the models studied in the previous sections to allow for the heterogeneous priors, i.e. $\widehat{f}_0^A \neq \widehat{f}_0^B$.

Public Signals

I start with the public signal case, in which both traders receive the same signal but they will hold different beliefs. I will solve the Bayesian Nash equilibrium in a similar way to the private signal case. Formally, the beliefs of both groups can be characterized by

Proposition 9 *The beliefs of two groups of traders at t are normally distributed, denoted by* $N(f_t^A, \frac{1}{\tau_t})$ *and* $N(f_t^B, \frac{1}{\tau_t})$*,where the precision is given by*

$$\tau_t = \tau_{t-1} + 2\tau_\epsilon$$

and the means are given by

$$\begin{aligned} \widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t}(S_t^A + S_t^B - 2\widehat{f}_{t-1}^A) \\ \widehat{f}_t^B &= \widehat{f}_{t-1}^B + \frac{\tau_\epsilon}{\tau_t}(S_t^B + S_t^A - 2\widehat{f}_{t-1}^B). \end{aligned}$$

We can characterize the Bayesian Nash equilibrium at t as:

Proposition 10 *The Bayesian Nash equilibrium at t can be characterized by:*

$$\begin{aligned} x_t^A &= \frac{2^{(T-t-1)/2}n^{(T-1-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}(\widehat{f}_t^A - \alpha) - \frac{2^{(T-t-1)/2}n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}_t^B - \alpha) \\ x_t^B &= \frac{2^{(T-t-1)/2}n^{(T-1-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-3)/2}}(\widehat{f}_t^B - \alpha) - \frac{2^{(T-t-1)/2}n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t-1)/2}}(\widehat{f}_t^A - \alpha) \\ p_t &= \alpha + n_t^A x_t^A + n_t^B x_t^B \end{aligned}$$

if t is odd;

$$\begin{aligned} x_t^A &= \frac{2^{(T-t)/2}n^{(T-1-t)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t^A - \alpha) \\ p_t &= \alpha + (n_t^A)\frac{2^{(T-t)/2}n^{(T-1-t)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t^A - \alpha). \end{aligned}$$

if t is even.

Remarks:

1. Monte Carlo simulations confirm that heterogeneous beliefs in the public signal case will not change qualitative features of the main results. From Proposition 10, we can see that the Bayesian Nash equilibrium is quite similar to the private signal case. This makes sense because with heterogeneous priors, the beliefs are different at each trading date even with public signals.

Private Signals

Note that with heterogeneous beliefs, type A traders cannot extract exact signals at any trading date as opposed to the homogeneous priors case, in which they could at $t = 2$. This is the only modification of the equilibrium in the private signals case with heterogeneous priors. Instead, type A traders can only get a composite signal combined with the initial belief of type B traders $\widehat{f}_0^B$ and their belief $\widehat{f}_1^B$ at $t = 1$. For $t = 3, 4, 5, \ldots, T-1$, the signal extraction behavior of both types is identical to the homogeneous priors case. Formally,

Proposition 11 *The beliefs of the two groups of traders at t are normally distributed, denoted by $N(f_t^A, \frac{1}{\tau_t})$ and $N(f_t^B, \frac{1}{\tau_t})$, where the precision for type A traders is given by*

$$\begin{aligned}
\tau_1^A &= \tau_0^A + \tau_\epsilon \\
\tau_t^A &= \tau_{t-1}^A + \tau_\epsilon, \qquad \text{for } t \text{ is } 2k-1,\ k \geq 2 \\
\tau_t^A &= \tau_{t-1}^A + \tau_\epsilon + \widehat{\tau}_t^A, \qquad \text{for } t \text{ is } 2k,\ k \geq 1
\end{aligned}$$

where $\widehat{\tau}_t^A = \frac{1+(\tau_{t-1}^B)^2}{(1+\tau_{t-1}^B)^2}\tau_\epsilon$. The mean for type A is given by

$$\begin{aligned}
\widehat{f}_1^A &= \widehat{f}_0^A + \frac{\tau_\epsilon}{\tau_1}(S_1^A - \widehat{f}_0^A) \\
\widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t^A}(S_t^A - \widehat{f}_{t-1}^A), \qquad \text{for } t \text{ is } 2k-1,\ k \geq 2 \\
\widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t^A}(S_t^A - \widehat{f}_{t-1}^A) + \frac{\widehat{\tau}_t^A}{\tau_t^A}(\widehat{S}_t^A - \widehat{f}_{t-1}^A), \qquad \text{for } t \text{ is } 2k,\ k \geq 1
\end{aligned}$$

where $\widehat{S}_t^A = \frac{\tau_{t-1}^B S_{t-2}^B}{1+\tau_t^B{}_1} + \frac{S_{t-1}^B}{1+\tau_{t-1}^B}$. The precision for type B traders is given by

$$\begin{aligned}
\tau_1^B &= \tau_0^B + \tau_\epsilon \\
\tau_t^B &= \tau_{t-1}^B + 2\tau_\epsilon \quad \text{for } t \geq 2
\end{aligned}$$

and the mean for type B traders is given by

$$\begin{aligned}
\widehat{f}_1^B &= \widehat{f}_0^B + \frac{\tau_\epsilon}{\tau_1^B}(S_1^B - \widehat{f}_0^B) \\
\widehat{f}_t^B &= \widehat{f}_{t-1}^B + \frac{\tau_\epsilon}{\tau_t^B}(S_t^B + S_{t-1}^A - 2\widehat{f}_{t-1}^B) \quad \text{for } t \geq 2
\end{aligned}$$

Figure 1.12 depicts the average autocorrelations of r_t and $\mathrm{Var}(r_t)$ across 100 simulations in the public signal case and in the private signal case with heterogeneous priors. It shows

that the magnitudes of the average autocorrelations are similar to those in the cases with homogeneous beliefs. This implies that the changing from the homogeneous priors to the heterogeneous priors have negligible effects on the persistence structure of the returns and the variance of returns.

In summary, heterogeneous beliefs affect the equilibrinm in a minor way. In the public signal case, traders have different beliefs. The differences in beliefs are constant over time and equal to the difference in priors. In the private signal case, it changes the belief updating of type A traders at only one trading date, $t = 2$. As a result, the qualitative properties of prices, returns, and volatility of returns do not change.

1.7 Conclusions

This chapter has developed a discrete-time multiperiod model of volatility clustering due to the combined effects of rational traders with multiple trading frequencies and their strategic interactions. First, multiple trading frequencies lead to an alternating pattern in prices which generates a serial correlation in the magnitude of the returns. Secondly, signal extraction provides a feedback mechanism, which induces a correlation between the past prices and the current price. This facilitates the formation of the volatility clustering. In addition, the proposed mechanism is capable of generating linearly trending prices and a negative correlation at the first lag of returns.

I also find that the number of traders has an impact on the formation of volatility clustering. This is a consequence of the fact that when the mean intensity of arrivals increases, the strategic competition outcome will converge to the competitive outcome. Hence, the effect of strategic interaction diminishes. Monte Carlo simulations show that in all settings, as the mean arrivals of traders increase, the variance of returns becomes less persistent. In the extreme, when the nnmber of traders is sufficiently large, the model prcdicts that there is only one statistically significant first-order autocorrelation of variance of retnrns, while other autocorrelation coefficients are statistically indistinguishable from zero.

This model yields several interesting predictions. First, traders with more precise signals have a smaller impact on the evolution of equilibrium prices and returns. Secondly, traders with different trading frequencies impose different levels of impact on the equilibrium prices and returns. Frequent traders respond to the signals in smaller magnitudes and this is

transformed into a smaller cumulative impact on the evolution of the equilibrium.

I show that an information hierarchy can be generated in an *ex ante* symmetric setting through signal extraction. Thus when trading frequencies are different, signal extraction can endogenously determine the information diffusion. The informational advantages of the traders may be due to their trading frequency. One potentially interesting avenue for future research is to endogenize the trading frequencies in a more general model where trading frequencies and an information hierarchy are simultaneously determined. Doing so would allow us to address additional issues, such as the microstructure impact of trading from information diffusion. For instance, in the context of this model, trading frequency is exogenously determined and leads to signal extraction which generates volatility clustering and information hierarchy. It is not obvious which factors make traders choose to trade less frequently. I leave the clarification of these issues for the future work.

Lags	D_t			$\mathrm{Var}(D_t)$		
	Mean	Std	p-value	Mean	Std	p-value
1	-0.255	0.073	0	0.018	0.041	0
2	0.006	0.082	0.444	0.007	0.060	0.252
3	-0.008	0.060	0.203	-0.004	0.046	0.338
5	-0.002	0.064	0.770	-0.003	0.052	0.560
6	-0.004	0.070	0.608	-0.007	0.047	0.174
7	-0.009	0.076	0.233	-0.003	0.050	0.549
8	0.003	0.068	0.693	-0.006	0.0430	0.198
9	-0.005	0.070	0.507	0.006	0.067	0.356
10	-0.006	0.069	0.419	-0.005	0.044	0.239

Table 1.1: MONTE CARLO STUDY OF D_t (ARRIVAL COMPONENT). First column reports the mean of autocorrelations. Second column reports the variance of autocorrelations. Third column reports the p-value of t test for null hypothesis that the mean equals zero. The columns four to six are the corresponding results for variance of D_t.

Lags	Z_t			$\mathrm{Var}(Z_t)$		
	Mean	Std	p-value	Mean	Std	p-value
1	-0.024	0.172	0.290	0.271	0.152	0
2	0.009	0.157	0.624	0.220	0.179	0
3	-0.036	0.161	0.043	0.200	0.141	0
4	0.001	0.171	0.938	0.144	0.160	0
5	0.006	0.160	0.641	0.130	0.155	0
6	0.001	0.173	0.924	0.108	0.164	0
7	-0.002	0.157	0.865	0.097	0.177	0
8	0.008	0.164	0.486	0.079	0.167	0
9	0.015	0.172	0.126	0.071	0.166	0
10	0.002	0.163	0.871	0.054	0.174	0

Table 1.2: MONTE CARLO STUDY OF Z_t (BELIEF COMPONENT). First column reports the mean of autocorrelations. Second column reports the variance of autocorrelations. Third column reports the p-value of t test for null hypothesis that the mean equals zero. The columns four to six are the corresponding results for variance of Z_t.

Lags	r_t			$\text{Var}(r_t)$		
	Mean	Std	p-value	Mean	Std	p-value
1	-0.489	0.058	0	0.347	0.165	0
2	-0.005	0.097	0.579	0.094	0.153	0
3	0.001	0.105	0.138	0.099	0.155	0
4	-0.002	0.109	0.893	0.091	0.152	0
5	-0.006	0.088	0.414	0.083	0.164	0
6	0.013	0.098	0.534	0.083	0.154	0
7	-0.011	0.093	0.187	0.082	0.162	0
8	0.008	0.110	0.308	0.086	0.152	0
9	-0.004	0.099	0.291	0.079	0.155	0
10	-0.003	0.105	0.958	0.071	0.160	0

Table 1.3: AUTOCORRELATIONS OF r_t AND VAR(r_t) IN PUBLIC SIGNAL CASE. First column reports the average of the autocorrelations across 100 simulations. Second column reports the standard deviation of the autocorrelations across 100 simulations. Third column reports the p-value of t test for the null hypothesis that the mean eqnals zero. The columns four to six are the corresponding results for $\text{Var}(r_t)$.

Lags	r_t			$\text{Var}(r_t)$		
	Mean	Std	p-value	Mean	Std	p-valne
1	-0.461	0.077	0	0.293	0.151	0
2	-0.018	0.131	0.173	0.125	0.171	0
3	0.017	0.114	0.134	0.121	0.159	0
4	-0.001	0.119	0.905	0.125	0.157	0
5	0.010	0.127	0.413	0.128	0.160	0
6	-0.009	0.131	0.519	0.119	0.160	0
7	0.016	0.115	0.173	0.114	0.172	0
8	-0.014	0.125	0.280	0.102	0.162	0
9	0.013	0.122	0.291	0.114	0.158	0
10	0.001	0.110	0.902	0.109	0.170	0

Table 1.4: ACFS OF r_t AND VAR(r_t) IN SIGNAL EXTRACTION CASE. First colnmn reports the mean of autocorrelations. Second column reports the variance of autocorrelations. Third column reports the p-value of t test for nnll hypothesis that the mean equals zero. The colnmns four to six are the corresponding resnlts for $\text{Var}(r_t)$.

Lags	r_t			$\text{Var}(r_t)$		
	Mean	Std	p-value	Mean	Std	p-value
1	-0.460	0.099	0	0.383	0.154	0
2	-0.023	0.166	0.169	0.214	0.160	0
3	0.024	0.142	0.090	0.204	0.165	0
4	-0.003	0.144	0.860	0.202	0.163	0
5	0.011	0.150	0.479	0.198	0.163	0
6	-0.006	0.151	0.698	0.188	0.176	0
7	-0.019	0.135	0.221	0.179	0.187	0
8	0.019	0.144	0.188	0.166	0.165	0
9	0.002	0.140	0.179	0.175	0.153	0
10	0.001	0.136	0.913	0.168	0.178	0

Table 1.5: ACFs OF r_t AND $\text{VAR}(r_t)$ WHEN TRADERS USE SOPHISTICATED GUESS. First column reports the mean of autocorrelations. Second column reports the variance of autocorrelations. Third column reports the p-value of t test for null hypothesis that the mean equals zero. The column four to six are the corresponding results for $\text{Var}(r_t)$.

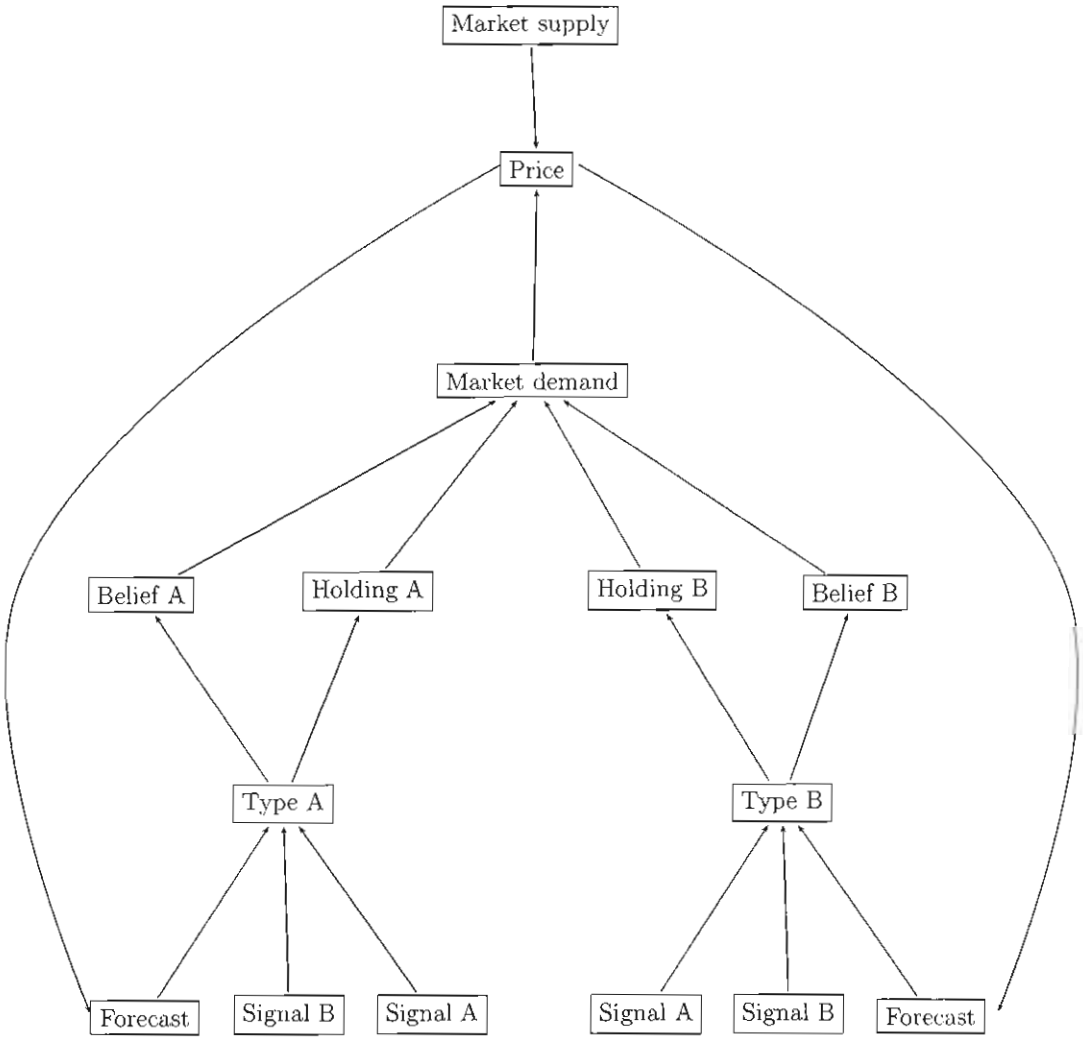
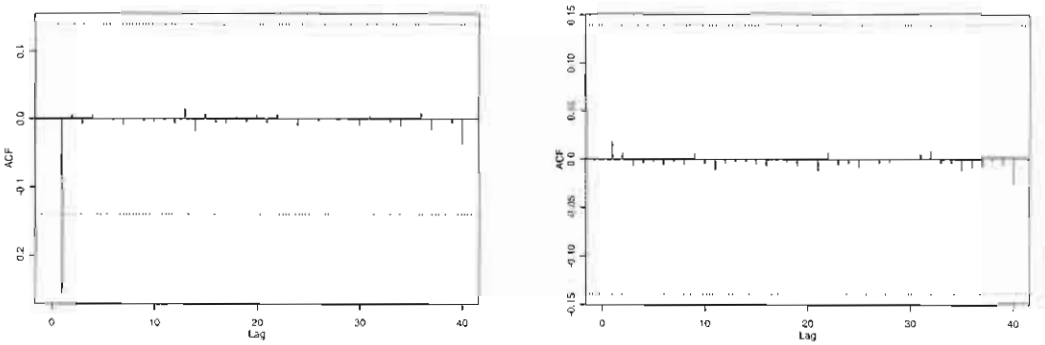
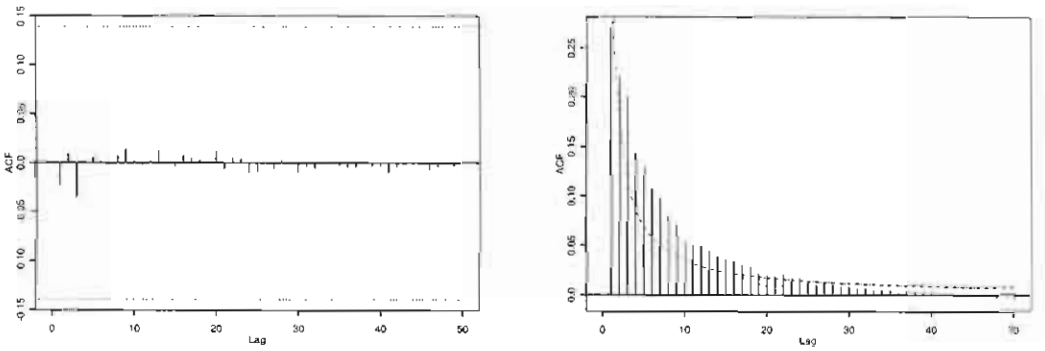


Figure 1.2: A FLOWCHART FOR THE PUBLIC SIGNAL CASE: HOW THE PRICE IS GENERATED? In the public signal case, each trader receives a signal, observes the signal of other traders, and forecasts the next period's price at the beginning of the current trading period. Based on this information, each trader updates his belief about the value of the underlying asset and adjusts his optimal holdings which are aggregated into the market demand. Combined with the market supply, the price is determined.



(a) ACF of simulated D_t

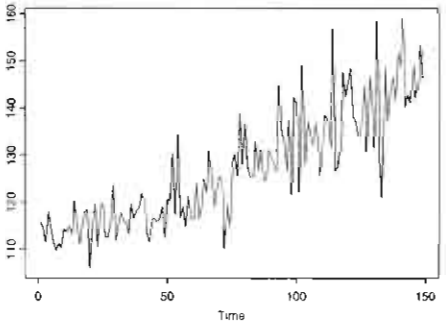
(b) ACF of simulated $\mathrm{Var}(D_t)$



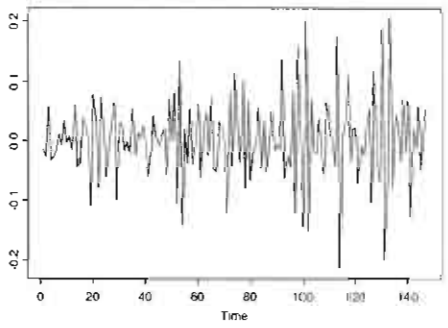
(c) ACF of simulated Z_t

(d) ACF of simulated $\mathrm{Var}(Z_t)$

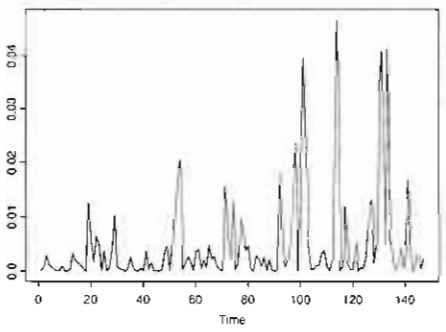
Figure 1.3: MONTE CARLO STUDY OF AUTOCORRELATION FUNCTION (ACF) OF D_t AND Z_t. (a) Average autocorrelations of D_t across 100 simulations. (b) Average autocorrelations of $\mathrm{Var}(D_t)$. (c) Average autocorrelations of Z_t. (d) Average autocorrelations of $\mathrm{Var}(Z_t)$. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



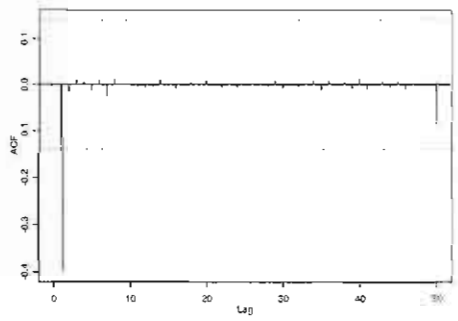
(a) Simulated prices



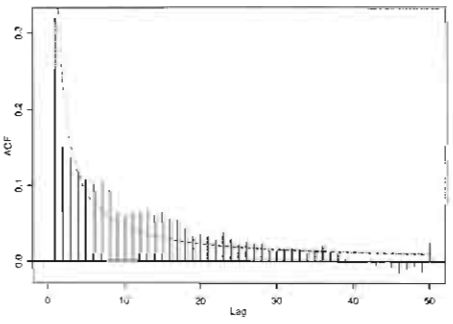
(b) Simulated returns



(c) Simulated variance of return

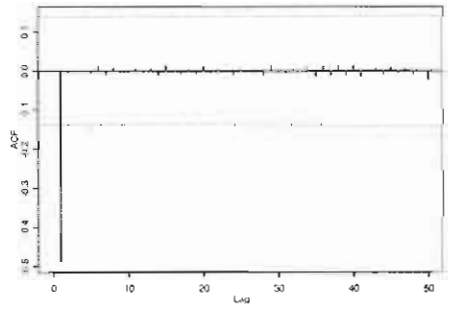


(d) ACF of simulated returns

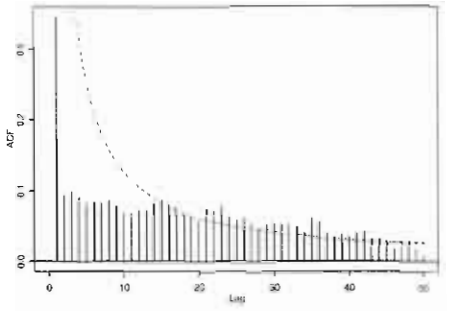


(e) ACF of simulated the variance of returns

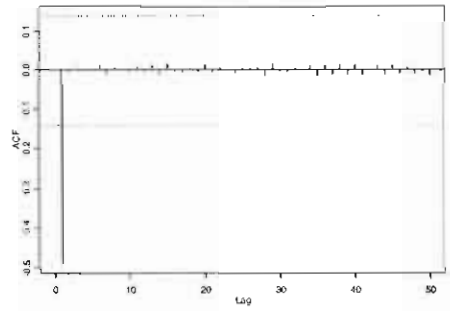
Figure 1.4: SIMULATION RESULTS IN THE PUBLIC SIGNAL CASE. (a) Time series of simulated prices. (b) Time series of simulated returns. (c) Time series of simulated variance of returns. (d) Average autocorrelations of returns across 100 simulations (e) Average autocorrelations of variance of returns. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



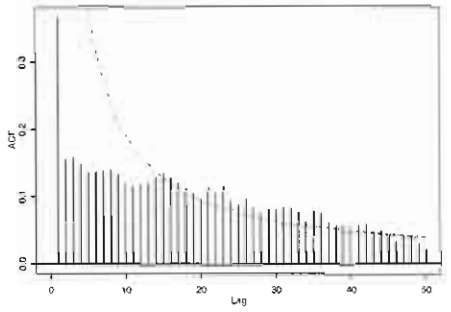
(a) ACF of returns for $\alpha = 80$



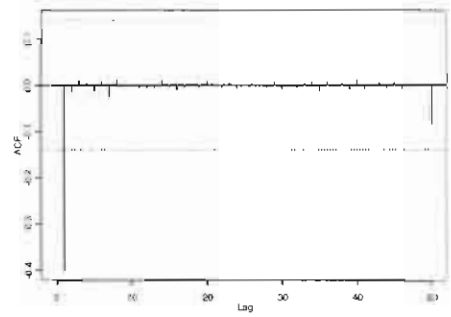
(b) ACF of variance of returns for $\alpha = 80$



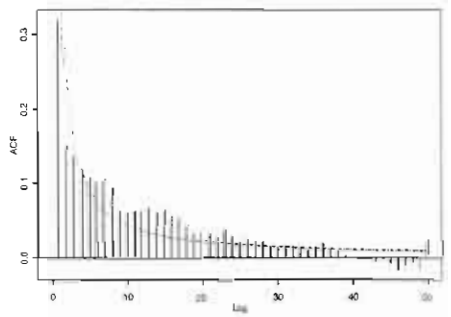
(c) ACF of returns when $\alpha = 150$



(d) ACF of variance of returns for $\alpha = 150$

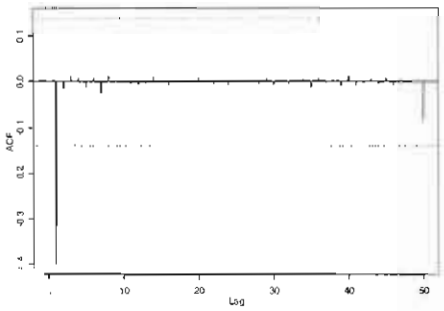


(e) ACF of returns for $\alpha = 200$

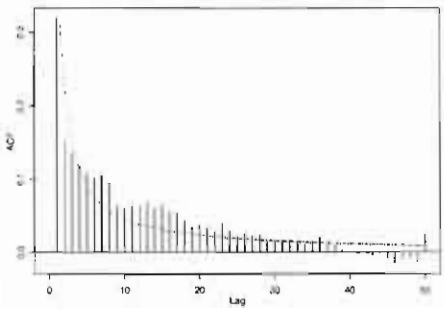


(f) ACF of variance of returns for $\alpha = 200$

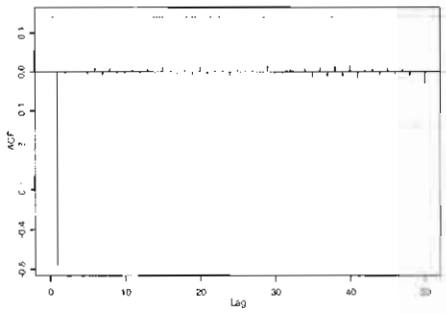
Figure 1.5: DEPENDENCE STRUCTURE (AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t)) AND CARRYING COST IN THE PUBLIC SIGNAL CASE. (a) Average autocorrelations of returns across 100 simulations for $\alpha = 80$. (b) Average autocorrelations of variance of returns for $\alpha = 80$. (c) Average autocorrelations of returns for $\alpha = 150$. (d) Average autocorrelations of variance of returns for when $\alpha = 150$. (e) Average autocorrelations of returns for $\alpha = 200$. (f) Average autocorrelations of variance of returns for $\alpha = 200$. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



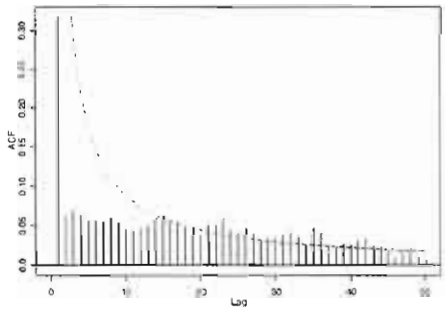
(a) ACF of returns for $n = 50$



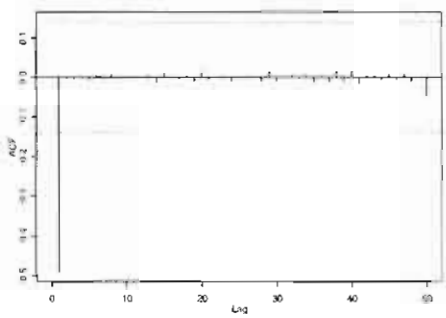
(b) ACF of variance of returns for $n = 50$



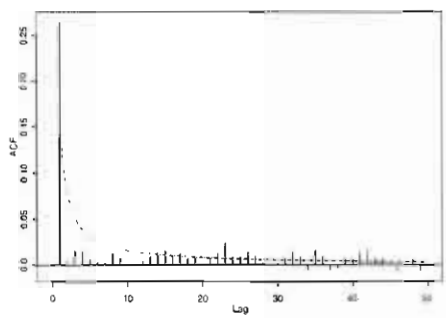
(c) ACF of returns for $n = 100$



(d) ACF of variance of returns for $n = 100$

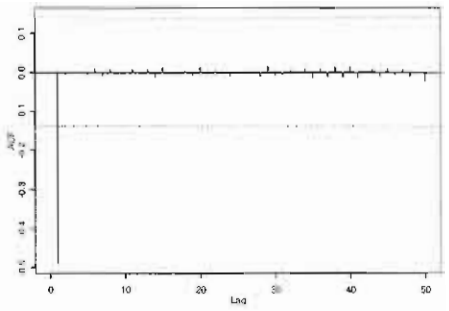


(e) ACF of returns for $n = 200$

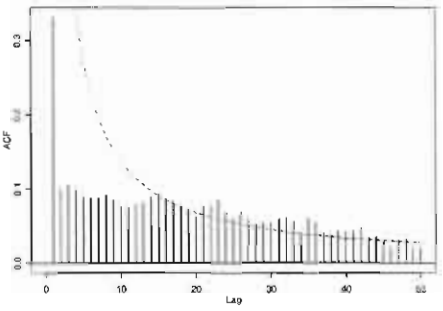


(f) ACF of variance of returns for $n = 200$

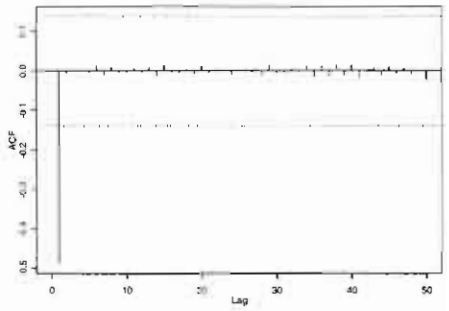
Figure 1.6: DEPENDENCE STRUCTURE (AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t)) AND MEAN ARRIVALS IN THE PUBLIC SIGNAL CASE. (a) Average autocorrelations of returns across 100 simulations for $n = 50$. (b) Average autocorrelations of variance of returns for $n = 50$. (c) Average autocorrelations of returns for $n = 100$. (d) Average ACF of variance of returns for $n = 100$. (e) Average autocorrelations of returns when $n = 200$. (f) Average autocorrelations of variance of returns for $n = 200$. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



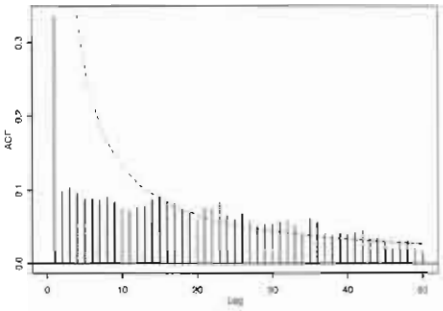
(a) ACF of returns for $\sqrt{\omega} = 10$



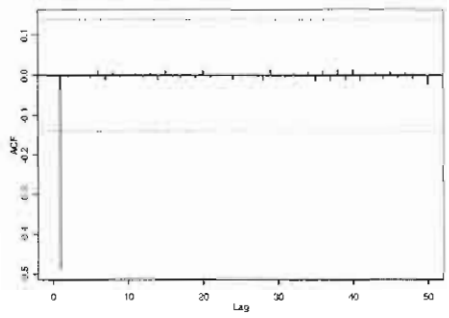
(b) ACF of variance of returns for $\sqrt{\omega} = 10$



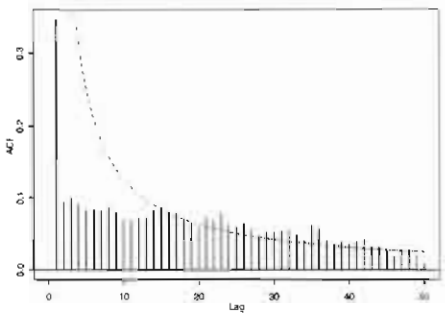
(c) ACF of returns for $\sqrt{\omega} = 20$



(d) ACF of variance of returns for $\sqrt{\omega} = 20$

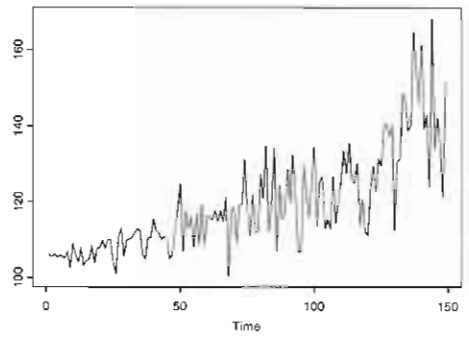


(e) ACF of returns for $\sqrt{\omega} = 30$

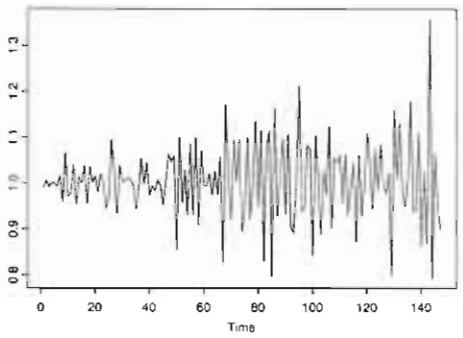


(f) ACF of variance of returns for $\sqrt{\omega} = 30$

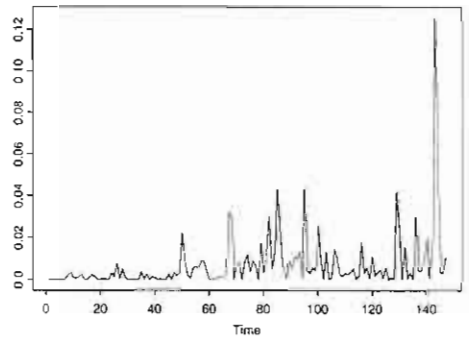
Figure 1.7: DEPENDENCE STRUCTURE (AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t)) AND STANDARD DEVIATION OF ARRIVALS IN THE PUBLIC SIGNAL CASE. (a) Average autocorrelations of returns across 100 simulations for $\sqrt{\omega} = 10$. (b) Average autocorrelations of variance of returns for $\sqrt{\omega} = 10$. (c) Average autocorrelations of returns for $\sqrt{\omega} = 20$. (d) Average autocorrelations of variance of returns for $\sqrt{\omega} = 20$. (e) Average antocorrelations of returns for $\sqrt{\omega} = 30$. (f) Average autocorrelations of variance of returns for $\sqrt{\omega} = 30$. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



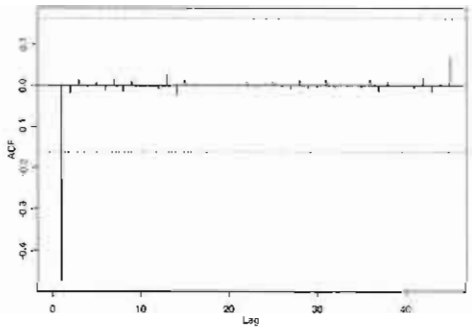
(a) Simulated prices



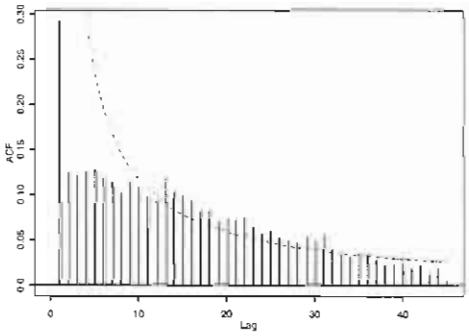
(b) Simulated returns



(c) Simulated variance of returns

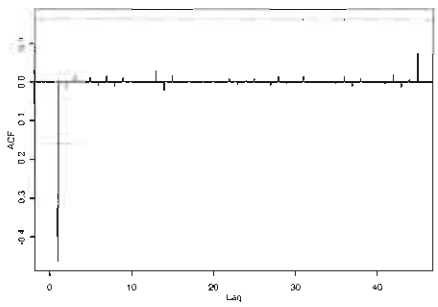


(d) ACF of returns

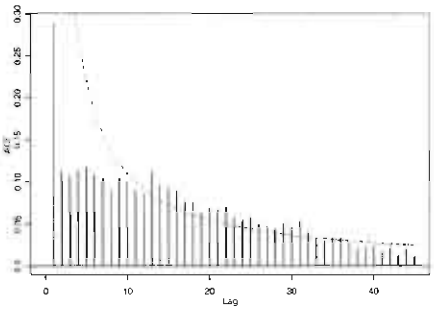


(e) ACF of variance returns

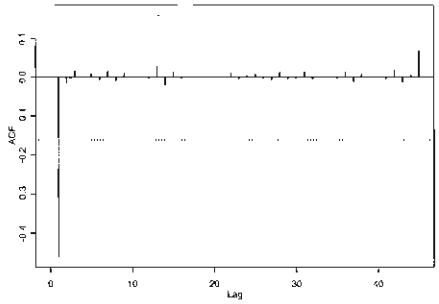
Figure 1.8: SIMULATION RESULTS IN THE PRIVATE SIGNAL CASE. (a) Time series of simulated prices. (b) Time series of simulated returns. (c) Time series of simulated variance of returns. (d) Average autocorrelations of returns across 100 simulations (e) Average autocorrelations of variance of returns. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



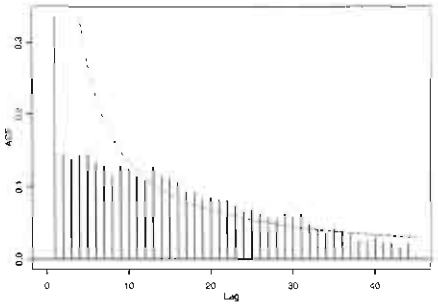
(a) ACF of returns for $\alpha = 80$



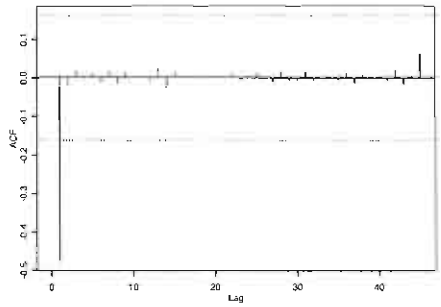
(b) ACF of variance returns for $\alpha = 80$



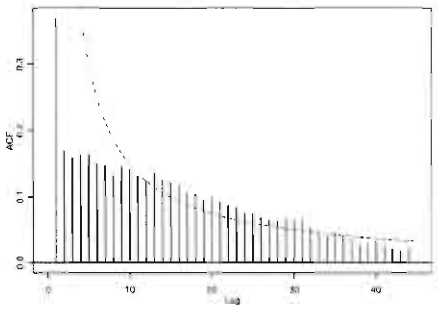
(c) ACF of returns for $\alpha = 100$



(d) ACF of variance of returns for $\alpha = 100$

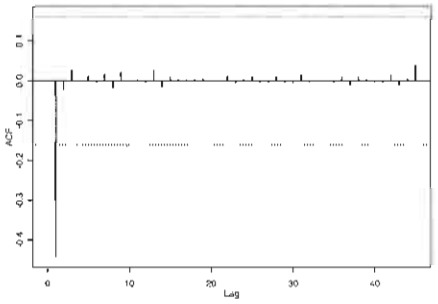


(e) ACF of returns for $\alpha = 120$

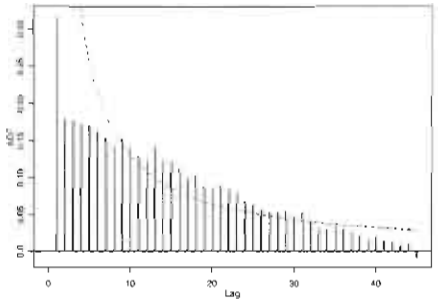


(f) ACF of variance of returns for $\alpha = 120$

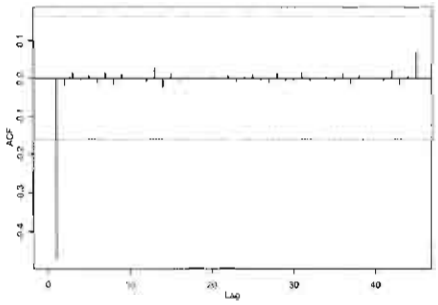
Figure 1.9: DEPENDENCE STRUCTURE (AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t)) AND CARRYING COST IN THE PRIVATE SIGNAL CASE. (a) Average autocorrelations of returns for $\alpha = 80$ across 100 simulations. (b) Average autocorrelations of variance of retnrns for $\alpha = 80$. (c) Average antocorrelations of returns for $\alpha = 100$. (d) Average autocorrelations of variance of returns for $\alpha = 100$. (e) Average antocorrelations of returns for $\alpha = 120$. (f) Average antocorrelations of variance of returns for $\alpha = 120$. For ACF plots. a hyperbolic decay function of antocorrelations is imposed.



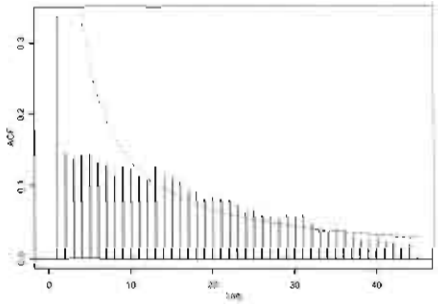
(a) ACF of returns for $n = 30$



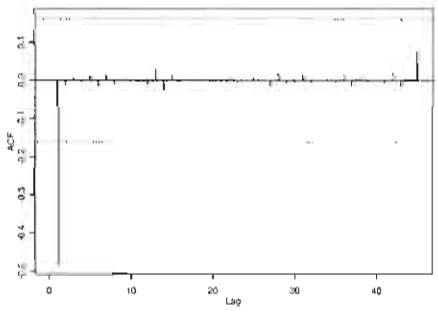
(b) ACF of variance of returns for $n = 30$



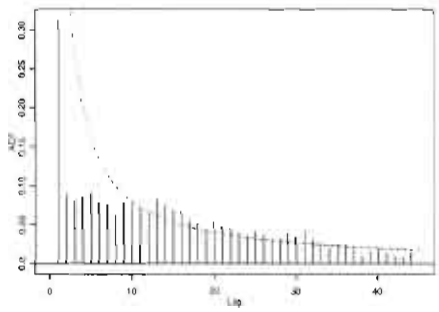
(c) ACF of returns for $n = 50$



(d) ACF of variance of returns for $n = 50$

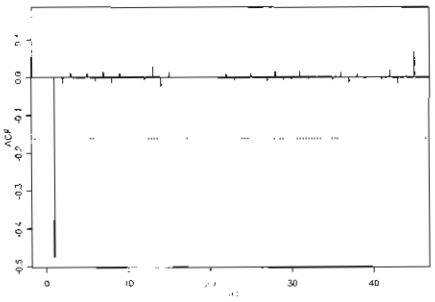


(e) ACF of returns for $n = 80$

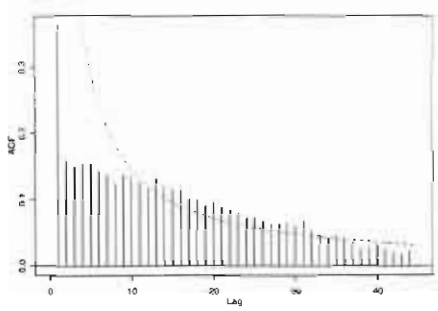


(f) ACF of variance of returns for $n = 80$

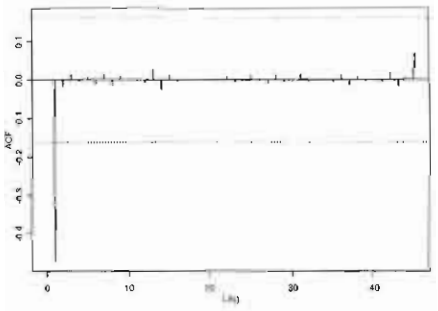
Figure 1.10: DEPENDENCE STRUCTURE (AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t)) AND MEAN ARRIVALS IN THE PRIVATE SIGNAL CASE. (a) Average autocorrelations of returns across 100 simulations for $n = 30$. (b) Average autocorrelations of variance of returns for $= 50$. (c) Average autocorrelations of returns for $n = 50$. (d) Average autocorrelations of variance of returns for $n = 80$. (e) Average autocorrelations of returns for $n = 80$. (f) Average autocorrelations of variance of returns for $n = 80$. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



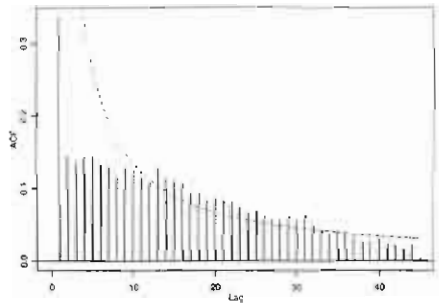
(a) ACF of returns for $\sqrt{\omega} = 10$



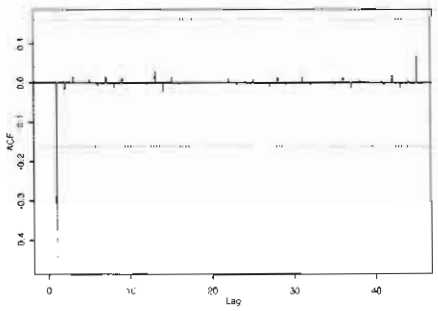
(b) ACF of variance of returns for $\sqrt{\omega} = 10$



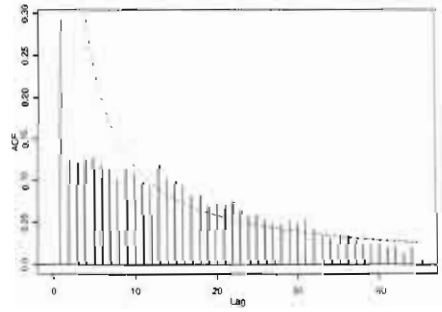
(c) ACF of retnrns for $\sqrt{\omega} = 20$



(d) ACF of variance of returus for $\sqrt{\omega} = 20$

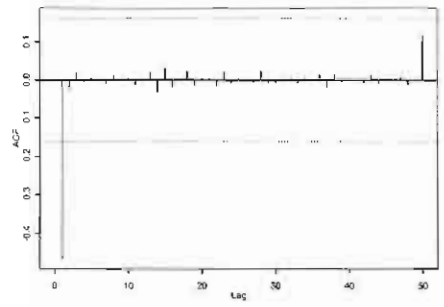


(e) ACF of returns for $\sqrt{\omega} = 30$

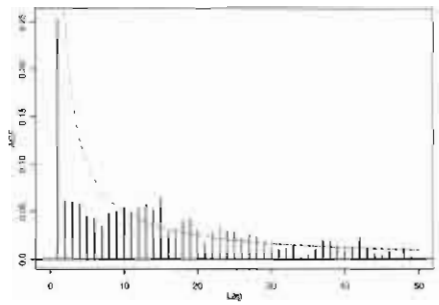


(f) ACF of variance of returns for $\sqrt{\omega} = 30$

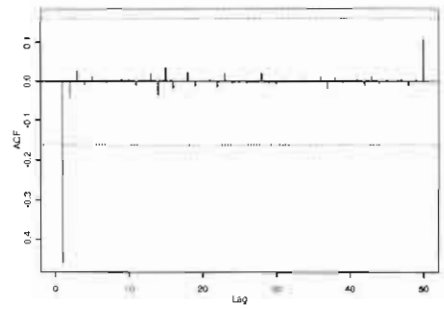
Figure 1.11: DEPENDENCE STRUCTURE (AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t)) AND VARIANCE OF ARRIVALS IN THE PRIVATE SIGNAL CASE. (a) Average autocorrelations of returns across 100 simulations for $\sqrt{\omega} = 10$. (b) Average autocorrelations of variance of retnrns for $\sqrt{\omega} = 10$. (c) Average autocorrelations of returns for $\sqrt{\omega} = 20$. (d) Average autocorrelations of variance of returns for $\sqrt{\omega} = 20$. (e) Average autocorrelations of returns for $\sqrt{\omega} = 30$. (f) Average autocorrelations of variance of returns for $\sqrt{\omega} = 30$. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



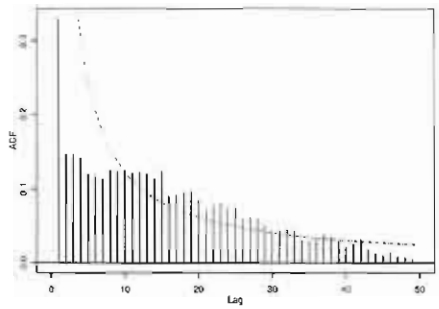
(a) ACF of r_t with public signal



(b) ACF of $\mathrm{Var}(r_t)$ with public signal

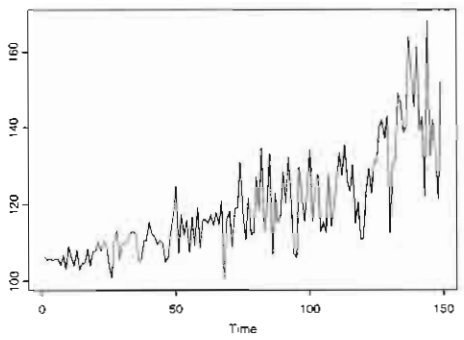


(c) ACF of r_t with private signal

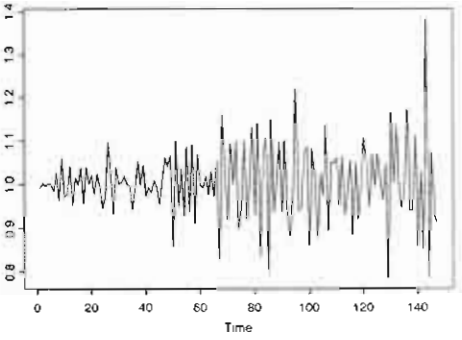


(d) ACF of $\mathrm{Var}(r_t)$ with private signal

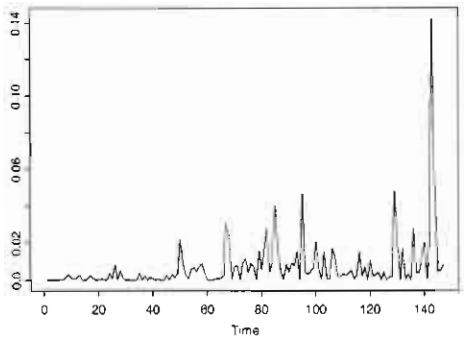
Figure 1.12: THE EFFECTS OF HETEROGENEOUS PRIORS ON AUTOCORRELATION FUNCTION (ACF) OF r_t AND VAR(r_t). (a) Average autocorrelations of r_t across 100 simulations with public signal. (b) Average autocorrelations of $\mathrm{Var}(r_t)$ with public signal. (c) Average autocorrelations of r_t with private signal. (d) Average autocorrelations of $\mathrm{Var}(r_t)$ with private signal. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.



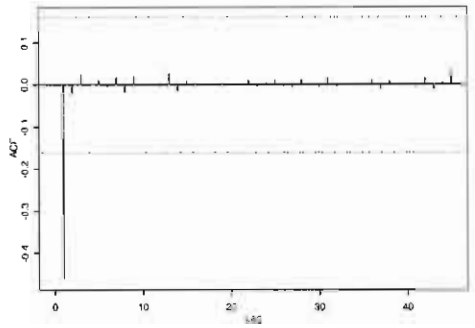
(a) Simulated prices



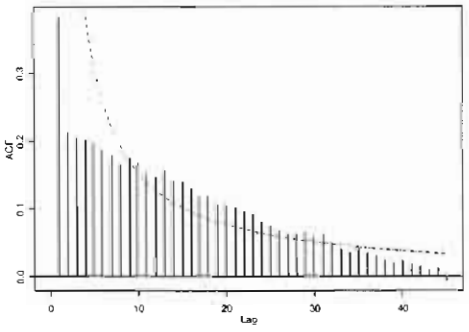
(b) Simulated returns



(c) Simulated variance of returns



(d) ACF of returns



(e) ACF of variance of returns

Figure 1.13: SIMULATION RESULTS WHEN TRADERS USE SOPHISTICATED GUESS. (a) Time series of simulated prices. (b) Time series of simulated returns. (c) Time series of simulated variance of returns. (d) Average autocorrelations of returns (e) Average autocorrelations of variance of returns. For ACF plots, a hyperbolic decay function of autocorrelations is imposed.

Chapter 2

Hierarchical Information and Price Discovery

2.1 Introduction

The rate of iuformatiou diffusion, and, consequently, price discovery, is conditional on not only the design of the market microstructure, but also the iuformational structure. The existing financc literature demonstrates that by polarizing traders into informed traders and noise traders, price discovery can be very slow (see Kyle (1985)) or very fast (see Foster and Viswanathan (1993)) depending on the market microstructure.[1] The goal of this chapter is to understand the impact of the informational structure on the degree of information diffusion inefficiency. I show for instance, that even in a market where there are many informed traders, the rate of information diffusion can be very slow if the information is distributed hierarchically.

Specifically, I consider a discrete time, infinite trading horizon model in which traders trade a single asset with a probability of liquidation in every period. The traders differ in two ways. First, traders are in different information hierarchy levels such that traders in a higher information hierarchy level embed the information of those in lower hierarchy levels. Such a hierarchy leads to a mechanism where informed traders in a higher hierarchy

[1] Kyle (1985) shows that when a monopolistically informed trader strategically trades with noise traders, the monopolist will prevent the information from being released, which in turn leads to a slower rate of information diffusion. On the other hand, information is released almost instantly when informed traders possessing identieal information compete very aggressively. as analyzed in Foster and Viswanathan (1993).

level may prefer trading against informed traders in lower hierarchies. In order to trade against informed traders in lower hierarchies, a trader needs to maintain the informational advantage by preventing the spread of the information incorporated into the price (at a fast rate). Thus, this informational mechanism can lead to inefficiency in information diffusion. This is more likely to happen when the number of information hierarchies increases. Second, traders in the lowest information hierarchy, i.e., noise traders, do not receive signals about fundamentals and trade on price direction. Thus, this setting includes the traditional stylized setting of the dichotomy of informed traders and noise traders as a special case. When the informed traders trade with the noise traders, they take into account the fact that noise traders trade on price direction. This can lead to the generation of the momentum, i.e., positive autocorrelation in returns.

The informativeness of price, or the rate of the information diffusion, is a function of the layers of hierarchies among informed traders for the following reasons. First, the expectation error of the value of the underlying asset originating from the signal extraction of the partially informed traders makes the price less informative. The accuracy of the expectation formed by the partially informed traders decreases as the number of the layers of the hierarchies increases. This decrease in accuracy is partly due to the fact that it is increasingly hard for the partially informed traders to infer whether the movement in price is due to a change in fundamental value of the underlying asset or the liquidity shock brought by noise traders.[2] Second, the prevention of information disclosure by informed traders in higher information hierarchies makes the price less informative. When the riskiness of trading with other informed traders relative to the riskiness of trading with noise traders decreases, the informed traders will prevent information disclosure in order to make a profit from other informed traders. I show that informed traders in higher information hierarchies are more likely to be profitable if they trade against other informed traders when the price deviates sufficiently far away from the fundamental value of the underlying asset and the number of information hierarchy layers increases. As a result, the information diffusion is slower when the number of information hierarchy layers among informed traders increases.

In addition, rather than assuming that the behavior of noise traders is independently

[2]For instance, a positive innovation in the value of the underlying asset is the observational equivalent to a shortage of liquidity supply provided by noise traders from the perspective of the partially informed traders.

and identically distributed across time, as in the standard literature, I argue that the behavioral pattern of noise traders consists of predictable (based on price direction) and an unpredictable (idiosyncratic liquidity shock) components. In turn, this model generates a number of interesting and testable implications that are absent from existing models of rational noisy equilibrium. For instance, the predictable pattern of noise traders' behavior resulting in persistence structure in prices, regardless of the distributional properties of the fundamentals. In addition, interaction between informed traders and noise traders can lead to various market liquidity levels. Perhaps the most novel feature of this model is that the trend-following behavior pattern of noise traders can generate momentum in returns, i.e., a positive autocorrelation in returns. One explanation comes from the self-fulfilling type argument that when noise traders believe there is momentum, they will follow the trend; while when noise traders follow the trend, momentum can be generated.

Overall, the main contributions of this chapter are as follows. First, this chapter proposes a framework to study the impact of hierarchical information and the layers of hierarchy on the speed of price discovery. I show that the speed of price discovery and the informativeness of price decrease when the number of the information hierarchy layers increases. Second, the predictable behavior pattern of noise traders has a significant impact on the persistence structure of prices and returns. The trend-following behavior pattern facilitates the formation of momentum in returns. Third, bounded rationality of noise traders (the unpredictable componeut) may generate profits for them when informed traders believe it is too risky to trade with them and choose to trade with other informed traders, thus providing justification for the existence of noise traders.

The rest of the chapter is organized as follows. In Section 2.2, a case of two information hierarchies is discussed where only one informed trader and noise traders are presented in the market. This model serves as a benchmark to motivate the extensions that follow and demonstrates that the predictable pattern of noise traders results in the persistence in prices and facilitates the formation of persistence in returns. In addition, the trend-following behavior of noise traders can geuerate momentum in returns. In section 2.3, an information hierarchy with three levels, two types of informed traders and noise traders, is studied. The information diffusion is slower compared to the benchmark case. In section 2.4, a general case of N information hierarchies is investigated. The information diffusion speed decreases as the number of information hierarchy levels increases. In section 2.5, Monte Carlo simulations are carried out to demonstrate the impact of the predictable pattern of

noise traders and of the number of hierarchical information levels. Finally, I conclude.

2.2 The Case of Two Information Hierarchies

I begin by providing a simple version of this model without modeling the interactions between informed traders at different hierarchical information levels. This special case helps develop the intuition for how the interaction between a competitive informed trader who has no influence on the formation of market price and noise traders affects the equilibrium properties and the formation of return predictability, in particular, the formation of momentum. The benchmark case is also interesting in its own right as it provides a framework to study how the behavior of noise traders affects the equilibrium properties that can not be captured by existing noisy rational expectation literatures. In Section 2.3, I extend this version to allow for interactions between informed traders by explicitly modeling the information hierarchy among informed traders.[3]

2.2.1 Financial Assets

Consider two traded assets. One is a riskless asset with a fixed rate of return of $1+r$. The other asset is composed of shares or claims on a hypothetical firm, which pays no dividends but has a chance of being liquidated every period. The probability of being liquidated in period $t+1$, conditional on the firm's surviving until period t, is assumed to be constant λ. When liquidation happens, the firm pays the shareholders a liquidation value V_t, which is assumed to follow a stochastic process. Throughout the chapter, I assume that V_t is Gaussian. V_t can be independently and identically distributed (iid) or serially correlated. The normality assumption is for the purpose of simplicity as it permits the existence of a linear equilibrium. The market price for the risky asset is P_t.

2.2.2 Game Structure

Two groups of investors, one informed trader and continuum of noise traders, trade the asset on every trading date. At the beginning of each trading date, noise traders supply a certain amount of the shares of the underlying asset to the market. The supply of noise traders is stochastic, which captures the bounded rationality of noise traders. The informed

[3]I borrow the basic setting from Makarov and Rytchkov (2007).

trader observes the current market price and submits his demand. The price can be adjusted upward or downward, and the informed trader will adjust his optimal demand for the number of shares of the underlying asset accordingly. The market clears when the demand from the informed trader equals to the supply from noise traders.

2.2.3 Noise Traders

Traditionally, the behavior of liquidity suppliers, i.e., noise traders has been assumed in the literature to be independently and identically distributed (*iid*) over time. With the purpose of investigating the effects of the interactions between noise and informed traders on the equilibrium outcomes, I model the behavior of noise traders in a way consistent with Harris (2003). Harris (2003) describes the noise traders as types of traders who trade based on their beliefs concerning the price change direction rather than the fundamentals of the underlying asset. According to this description, a typical example of noise trader is a technical trader. Technical traders trade based on "pattern recognition" type techniques which aid in the formation of the beliefs concerning the direction of prices.[4] Therefore, I assume that there is a predictable components of the aggregate supply of noise traders (technical traders) that should be a function of past prices. In additiou, there is an unpredictable component which captures the difference in opinions among noise traders or technical traders. Hence, I model the aggregate supply of the shares of the underlying asset from noise traders throughout the chapter Q_t^s as $\beta P_{t-1} + \Theta_t$, where P_t is the price of the trading asset. βP_{t-1} captures the predictable component of the aggregate supply of shares from noise traders. Θ_t is the *iid* Gaussian with mean 0 and variance σ_Θ^2.[5] The random component is for the purpose of preventing prices from being fully revealing, in the spirit of Grossman and Stiglitz (1980).[6]

β is an aggregate measure of noise traders' supply of shares of the underlying asset. The sign of β indicates the aggregate response of the noise traders to the price change direction.

[4]There are two different types of technical analysis: subjective and objective analysis. Subjective analysis captures the fact that different traders may come up with different conclusions based on the same information set due to subjective judgments or priors. Even in objective analysis, different traders may have different confidence levels or tolerance levels which lead to different trading decisions

[5]The mean is not necessarily 0. It could be a positive number with a large magnitude, which would guarantee that noise traders always supply a positive number of shares. The zero mean assumption could be interpreted as demeaned supply.

[6]The predictable component is specified for simplicity, although it is easy to extend this analysis to allow for many price lags without chauging the main results.

For instance, if $\beta < 0$, noise traders will sell the traded asset when the unit price is negative (price is decreasing). It seems that the noise traders behave as if they follow the trend of the price. We label noise traders with negative β *"trend followers"*. In contrast, if $\beta > 0$, noise traders will sell the trading asset when the unit change of price of the traded asset is positive (price is increasing), i.e., noise traders act against the trend of the price, and we label the noise traders with positive β *"contrarians"*. The magnitude of β measures the sensitivity of noise traders to price direction changes. The larger the magnitude of β, the more sensitive noise traders are to the changes in price direction. That is, if the magnitude of β is large, with a small change of price, noise traders adjust their holdings of the underlying asset to a large extent. The magnitude of β can be interpreted as the aggregate number of noise traders present in the market as well. β could be time-varying. Throughout this chapter, however, I do not intend to model the evolution of noise traders and assume β to be a time invariant parameter. I will elaborate more on the relationship between β and equilibrium properties later.

2.2.4 Informed Traders

In this benchmark model, I only consider a single informed investor. In each trading period t, the investor receives a signal S_t about the fundamental value of the underlying asset V_t:

$$S_t = V_t + b_s \eta_t \tag{2.1}$$

where $\eta_t \sim iidN(0,1)$. b_s is the standard deviation of the signal and $1/b_s^2$ is the precision of the signal. The signal is more informative when b_s decreases. The investor is assumed to have a mean-variance preference, i.e., $E(W_t) - \frac{1}{2}\delta Var(W_t)$, where W_t is his wealth level at t and δ is the preference parameter. When $\delta = 0$, the traders are risk neutral, and when δ increases, the traders become more risk averse. In order to obtain a closed-form solution, I use myopic preference to abstract away from dynamic hedging considerations.[7] Let Q_{t+1} be excess return, i.e.,

$$Q_{t+1} = \lambda V_{t+1} + (1-\lambda)P_{t+1} - (1+r)P_t \tag{2.2}$$

[7]Myopic preference is a simplifying assumption, and its main purpose is to obtain an analytically tractable solution.

where $\lambda V_{t+1} + (1-\lambda)P_{t+1}$ is the expected return from the holding of one share of the trading asset and $(1+r)P_t$ is the opportunity cost of holding. Therefore, the per period utility maximization problem for the informed trader is

$$\max_{X_t} E[Q_{t+1}|F_t]X_t - \frac{1}{2}\delta Var[Q_{t+1}|F_t]X_t^2$$

where F_t is the information set available for informed traders that contains the current price, the history of past prices, and all the received signals, that is, $F_t = \{P_t, P_{t-1}, \ldots, P_0, S_t, S_{t-1}, \ldots, S_0\}$, and X_t is the holding of the informed investor at trading period t. For the expositional purposes, I use $E_t[.] = E[.|F_t]$ throughout the chapter. Hence, the optimal demand of the shares of the underlying asset from informed traders at t is

$$X_t = \frac{E_t[Q_{t+1}]}{\delta Var_t[Q_{t+1}]} \tag{2.3}$$

Let $\omega_t = \frac{1}{\delta Var_t[Q_{t+1}]}$ so that $X_t = \omega_t E_t[Q_{t+1}]$. Thus, the informed trader adjnsts his holding of the risky asset proportional to the expected return from the holdings. The risk averse coefficient affects the proportion of the informed trader's investment in the risky asset. When $\delta = 0$, the optimal holding X_t is not well-defined for $E_t[Q_{t+1}] \neq 0$, and X_t could be any amonnt for $E_t[Q_{t+1}] = 0$. Namely, if the informed trader is risk-neutral, he would like to borrow an infinite amount of money to invest in the risky asset if its expected return is greater than zero. He will spend any portion of his wealth on the risky asset if its expected return is equal to zero because the risky asset is indifferent from the riskless asset in this case. Similarly, when δ increases, the trader becomes more risk averse; he will invest less and less of his wealth into the risky asset. In particular, if $\delta = \infty$, he will not invest any of his wealth into risky asset.

2.2.5 Equilibrium

I focus on stationary and linear expectation equilibrium, where $\text{Var}_t[Q_{t+1}]$ is constant such that $\omega_t = \frac{1}{\delta \text{Var}_t[Q_{t+1}]} = \omega$ is time invariant and there is no bubble in prices. Hence $X_t = \omega E_t[Q_{t+1}]$. Market clearing implies $X_t = \beta P_{t-1} + \Theta_t$. Hence,

$$P_t = -\frac{\beta P_{t-1}}{\omega(1+r)} - \frac{\Theta_t}{\omega(1+r)} + \frac{1}{1+r}E_t[\lambda V_{t+1} + (1-\lambda)P_{t+1}] \tag{2.4}$$

Defining $P_t^* = P_t - \rho P_{t-1}$, we can rewrite Equation 2.4 as

$$P_t^* = -\frac{\Theta_t}{\omega(1+r)} + \frac{1}{1+r}E_t[\lambda V_{t+1} + (1-\lambda)P_{t+1}^*]$$

where ρ solves:

$$(1-\lambda)\rho^2 + (1+r)\rho + \beta = 0. \tag{2.5}$$

The detailed derivation is given in Appendix. Iterating Equation 2.5 forward and invoking the no-bubble constraint, i.e., $lim_{j\to\infty}(\frac{1-\lambda}{1+r})^j P_{t+j} = 0$, we have:

$$\begin{aligned} P_t^* &= -\frac{\Theta_t}{\omega(1+r)} + \sum_{s=0}^{\infty}(\frac{1-\lambda}{1+r})^s E_t[\lambda V_{t+s+1} - \frac{(1-\lambda)\Theta_{t+s+1}}{\omega(1+r)}] \\ &= -\frac{\Theta_t}{\omega(1+r)} + \sum_{s=0}^{\infty}(\frac{1-\lambda}{1+r})^s E_t[\lambda V_{t+s+1}] \end{aligned}$$

Hence,

$$P_t = \rho P_{t-1} - \frac{\Theta_t}{\omega(1+r)} + \sum_{s=0}^{\infty}(\frac{1-\lambda}{1+r})^s E_t[\lambda V_{t+s+1}] \tag{2.6}$$

Equatiou 2.6 shows that the equilibrium price at time t is the sum of three terms. The first term incorporates the predictable pattern of noise traders' behavior. As shown in Equation 2.5, the serial correlation structure of prices (ρ) is a function of the predictable pattern of noise traders' behavior (β). Intuitively, because the informed trader understands that the predictable pattern of the aggregate supply from the noise traders is a function of past prices, he can forecast the mean of the aggregate supply from the noise traders. Thus, the trader can form a better forecast of the price in the next period by taking into account this information, which will help increase his profit. As a result, the price in the next period is correlated with the past prices, which generates the serial correlation. As demonstrated in a later section, this serial correlation in prices brought by the predictable pattern in noise traders' behavior generates a serial correlation in returns as well. Therefore, the momentum could be generated. Notice that this result is independent of the assumption of the fundamental value of the underlying asset. I will elaborate more on this in later sections. The second term compensates for the risks for informed traders, which originate from the non-predictable component of aggregate noise traders' behavior. The third term is the expected payoff of asset holdings.

2.2.6 Equilibrium Properties

Stationary Equilibrium

This section studies the conditions for the equilibrium price to be stationary. I rewrite Equation 2.5 as

$$P_t - \rho P_{t-1} = -\frac{\Theta_t}{\omega(1+r)} + \sum_{s=0}^{\infty}(\frac{1-\lambda}{1+r})^s E_t[\lambda V_{t+s+1}] \tag{2.7}$$

Notice that in Equation 2.7, the right hand side (RHS) is assumed to be covariance stationary and has finite variance given the stationary assumption of V_t. Therefore, the price is covariance stationary when $|\rho| < 1$. Formally,

Proposition 12 *The relationship between the stationarity of equilibrium and β can be summarized as follows:*

1. *When noise traders are trend followers ($\beta < 0$), the maximum number of stationary equilibria is 1. Formally, there exist a stationary equilibrium if and only if the probability of liquidation is not sufficiently large ($\lambda < \beta + r + 2$).*

2. *When noise traders are contrarians and not sensitive to price direction change ($\frac{1+r}{2} > \beta > 0$) and the probability of liquidation is sufficiently large ($\lambda > \beta - r$), there is a unique stationary equilibrium. However, if the probability of liquidation is not sufficiently large, ($\lambda < \beta - r$), there exists two stationary equilibria.*

3. *When noise traders are contrarians and they are sensitive to price direction change ($\beta > \frac{1+r}{2}$) and if the probability of liquidation is sufficiently small ($\lambda < \frac{1+r}{2}$), there exist two stationary equilibria. Additionally, if $\lambda > \beta - r$, there is a unique equilibrium. Otherwise, if $\frac{1+r}{2} < \lambda < \beta - r$, there is no stationary equilibrium.*

Proposition 12 suggests that the predictable component β and the probability of liquidation λ play important roles in determining of the stationarity of the equilibrium price process. As shown in Proposition 12, the equilibrium can be stationary only if the liquidation probability is not large. This result is consistent with the observation that when the probability of being liquidated in the following period is very large, an informed trader is reluctant to adjust the optimal holding in response to the short run profitable opportunity originating from the trading pattern of noise traders. When the probability of liquidation is

small, the riskiness of exploiting the short run profitable opportunity from noise traders is relatively low. In that case, the informed trader may be more willing to adjust his holding accordingly. Therefore, the market will be cleared because the demand from the informed trader will match the supply from noise traders. Therefore, a stationary equilibrium can exist only if the probability of liquidation is sufficiently low.

Momentum

Momentum is defined as the rate of acceleration of a security's price or volume in technical analysis terms. Once a momentum trader sees an acceleration in a stock's price, earnings, or revenues, the trader will often take a long or short position in the stock with the hope that its momentum will continue in either an upward or downward direction. This strategy relies more on short-term movements in price rather than on the fundamental value of companies. Jegadeesh and Titman (1993) show that the momentum trading strategy can generate abnormal profit. Since its discovery, momentum has been one of the most resilient anomalies challenging the market efficiency hypothesis. It is well known that any theory seeking to explain momentum should be able to generate positive serial correlations in returns. In this section, I show the ability of my model to generate momentum, and particularly the positive autocorrelation in returns.

I consider two specifications of V_t. To start, I assume V_t are iid with $E_t[V_{t+s}] = \mu$. Hence, Equation 2.6 becomes

$$P_t = \rho P_{t-1} - \frac{\Theta_t}{\omega(1+r)} + \frac{\lambda\mu}{r+\lambda} \tag{2.8}$$

Equation 2.8 demonstrates that if $|\rho| < 1$, P_t follows an autoregressive (AR) (1) process and $\text{Corr}[P_t, P_{t-s}] = \rho^s$ follows directly. Defining return as the difference between the price levels, i.e., $r_t = P_t - P_{t-1}$, we have

$$r_t = \rho r_{t-1} - \frac{\Theta_t - \Theta_{t-1}}{\omega(1+r)} \tag{2.9}$$

Equation 2.9 shows that given $|\rho| < 1$, returns follow an autoregressive moving average model (ARMA)(1,1) process with mean 0. It can further be shown that $\text{Corr}[r_t, r_{t-1}] = \rho^2 + \rho - 1$. It follows directly that there can be a positive serial correlation of returns at the first lag. Formally, if $\rho < \frac{-\sqrt{5}-1}{2}$ or $\rho > \frac{\sqrt{5}-1}{2}$, the returns display a positive autocorrelation at first

lag, i.e., momentum is generated.[8]

The result of positive serial correlations in returns provides a rational explanation of momentum. From Proposition 12, if we restrict ourselves to stationary equilibrium, the only possible scenario resulting in positive serial correlations in returns is when ρ is positive which corresponds to $\beta < 0$ and $\lambda < \beta + r + 2$ case. This suggests that it is only possible in my model to generate a momentum anomaly when noise traders behave like trend followers. This is consistent with a self-fulfilling explanation. The reason why noise traders behaves like trend-followers is because the momentum anomaly exists, and when noise traders behaves like trend-followers, the momentum anomaly can be generated.

Next, I consider V_t to be a stationary AR(1) process. That is, $V_t = aV_{t-1} + b_v\epsilon_t$, where ϵ_t are *iid* standard normals, b_v is the inverse of the square root of the precision of the innovations, and $a < 1$. When b_v increases, the precision of the signal decreases; that is, the V_t becomes noisier. Then $E_t[V_{t+s}] = a^s V_t$. Hence

$$P_t = \rho P_{t-1} - \frac{\Theta_t}{\omega(1+r)} + \frac{\lambda a V_t}{1+r-(1-\lambda)a} \tag{2.10}$$

Defining return as the difference in price, i.e., $r_t = P_t - P_{t-1}$, we have:

$$r_t = \rho r_{t-1} - \frac{\Delta\Theta_t}{\omega(1+r)} + \frac{\lambda a}{1+r-a(1-\lambda)}\Delta V_t, \tag{2.11}$$

where $\Delta\Theta_t = \Theta_t - \Theta_{t-1}$ and $\Delta V_t = V_t - V_{t-1}$. It can be shown that the unconditional covariance of returns is

$$\mathrm{Cov}[r_t, r_{t-1}] = \rho\mathrm{Var}[r_{t-1}] + \frac{\sigma_\Theta^2}{\omega^2(1+r)^2} + \frac{\lambda^2 a^2}{(1+r-a(1-\lambda))^2}\mathrm{Var}[V_{t-1}] \tag{2.12}$$

We are still able to generate positive serial correlations in returns when $\rho > 0$. From Proposition 12, when noise traders are trend-followers, the ρ can be positive and momentum can exist. In addition, even if ρ is negative, it is still possible for momentum to exist as long as the right hand side (RHS) of Equation 2.12 is positive. In other words, if the fundamental value of the underlying asset is serially correlated, the momentum could exist regardless of the behavior pattern of the noise traders.

[8] In order to have a positive ρ which is consistent with the observed price autocorrelation coefficient, I choose $\rho > \frac{\sqrt{5}-1}{2}$. Combined with the annual risk free rate $r = 0.03$ and liquidation probability $\lambda = 10\%$, we need to have $\beta < -1$.

What if There is no Pattern in Noise Traders' Behavior?

To illustrate of the benefits of the assumption that there is a predictable component in noise traders' behavior, i.e. βP_{t-1}, I investigate the case where there is no predictable pattern in noise traders' behavior, i.e., $\beta = 0$. That is, the liquidity supply from noise traders is Θ_t, where Θ_t is *iid* normal. This is consistent with the standard assumption about noise traders' behavior in the literature (for instance, Kyle (1985)). The equilibrium price can be shown to be

$$P_t = -\frac{\Theta_t}{\omega(1+r)} + \sum_{s=0}^{\infty}(\frac{1-\lambda}{1+r})^s E_t[\lambda V_{t+s+1}] \tag{2.13}$$

This shows that the statistical properties of prices are fully determined by the statistical properties of the fundamental value of the underlying asset in the absence of the predictable pattern in noise traders' behavior. Hence, the extra gains from the assumption of a predictable pattern in noise traders' behavior can be summarized as follows:

1. The assumption of a predictable pattern, i.e., $\beta \neq 0$, results in persistence in price regardless of the statistical assumption of fundamentals. On the contrary, if there is no predictable pattern, i.e., $\beta = 0$, the persistence structure in price depends on the statistical assumption of fundamentals. With the presence of the predictable pattern in noise traders' behavior, price is persistent even when V_t is *iid*. The persistence in prices is partly due to the fact that the informed trader adjusts his optimal holding based on the expectation of the predictable pattern in the noise traders' behavior. This predictable pattern in noise traders' behavior preserves the correlation structure of prices across time. The persistence in prices is also due to the fact that the noise traders utilize the information concerning past prices to adjust their position, which determines the price in the current period. This feedback mechanism can also help the formation of the persistence structure in prices. On the contrary, if there is no predictable pattern in noise traders' behavior, the equilibrium price is entirely determined by the distributional assumption of fundamentals, i.e., V_t. To see this, from Equation 2.13 I can demonstrate that if V_t is *iid*, P_t is *iid*. If V_t is an AR(1) process, then P_t is an AR(1) process.

2. The assumption of a predictable pattern, i.e., $\beta \neq 0$, generates momentum (the positive serial correlation in returns), while if there is no predictable pattern, i.e., $\beta = 0$,

momentum cannot be generated in my model. When there is no predictable pattern in noise traders' behavior, generally, asymmetric information cannot generate momentum alone. Intuitively, rational traders require higher compensation for holding a larger amount of the risky asset. This leads to a positive relationship between returns in the next period and the supply from noise traders. In the meantime, there is a negative relationship between price and the supply from noise traders. Recall that return in this period is the difference between prices. If supply is assumed to be *iid*, realized returns are negatively correlated.[9] On the contrary, when there is a predictable pattern in noise traders' behavior, as indicated earlier, momentum or positive serial correlation in returns can be generated in my model. Notice that there are two opposing effects that generate momentum. One is the negative slope in the demand curve of informed traders and the other one is trend-following behavior of noise traders. I show that under some parameters values, the noise traders' trend following behavior may dominate the downward demand curve effects, which generates a positive serial correlation in returns and provides an explanation for the existence of momentum.

2.3 The Case of Three Information Hierarchies

I now extend the simple model of the previous section to allow for the interaction of informed traders at different hierarchical information levels. I consider two types of informed traders, corresponding to two informational hierarchical levels. To obtain a closed form solution, I further simplify the setting. Assume there are two factors that jointly determine the fundamental values of the underlying asset. That is, V_t is a function of the two factors V_t^1 and V_t^2. For simplicity, I assume that the function is linear, i.e., $V_t = V_t^1 + V_t^2$. In addition, I assume that there is no noise in the signals. Namely, a type 1 trader (fully informed trader) receives two signals per period, $S_{1,t} = V_t^1$ and $S_{2,t} = V_t^2$. A Type 2 trader (partially informed trader) receives only one signal $S_{2,t} = V_t^2$. Furthermore, I assume that V_t^1 and V_t^2 are AR(1) processes, i.e., $V_t^1 = aV_{t-1}^1 + b_V\epsilon_t^1$ and $V_t^2 = aV_{t-1}^2 + b_V\epsilon_t^2$.

The information set for the type 2 trader is $F_{2,t} = \{P_t, P_{t-1}, \ldots, V_t^2, V_{t-1}^2, \ldots\}$, and the information set for the type 1 trader is $F_{1,t} = \{P_t, P_{t-1}, \ldots, V_t^1, V_{t-1}^2 1, \ldots, V_t^2, V_{t-1}^2, \ldots\}$. By construction, $F_{2,t} \subset F_{1,t}$, and it captures the idea of a hierarchical information structure. It

[9] $\mathrm{Cov}[Q_t, r_t] \to \mathrm{Cov}[\Theta_t, P_t - P_{t-1}] = \mathrm{Cov}|\Theta_t, P_t] < 0.$

is well known from the existing literature on forecasting the forecasts of others, the infinite regress problem can be avoided with a hierarchical information structure. Intuitively, a fully informed trader knows everything a partially informed trader knows. Therefore, a fully informed trader knows exactly the expectation formed by a partially informed trader on the signal received by the fnlly informed trader. Then the infinite regress problem collapses.[10] Formally, the three information hierarchies' equilibrium is characterized as:

Proposition 13 *If a type 1 trade observes V_t^1 and V_t^2 and a type 2 trader only observes V_t^2, the equilibrium price is*

$$P_t = \rho P_{t-1} + P_V V_t + P_\Theta \Theta_t + P_\Delta(\widehat{V}_t^1 - V_t^1),$$

where

$$\begin{aligned} P_\Theta &= \frac{\rho}{\beta} \\ P_V &= -\frac{a\lambda}{\frac{\beta}{\rho\Omega} + a(1-\lambda)} \\ P_\Delta &= -\frac{a\omega_2(\lambda + (1-\lambda)\frac{a\lambda}{-\frac{\beta}{\rho\Omega} - a(1+\lambda)})}{\frac{\beta}{\rho} + a\omega_1(1-\lambda)c} \end{aligned}$$

where $\widehat{V}_t^1 = E[V_t^1|F_{2,t}]$, $\Omega = \omega_1 + \omega_2$, $\omega_1 = Var[Q_{t+1}|F_{1,t}]$, $\omega_2 = Var[Q_{t+1}|F_{2,t}]$, and where ρ solves

$$(1-\lambda)\Omega\rho^2 - (1+r)\Omega\rho - \beta = 0$$

and c solves

$$\frac{\rho^2}{\beta^2}\sigma_\Theta^2(1-c)(1-a^2c) - cb_V^2\left(\frac{a\lambda}{\frac{\beta}{\rho\Omega} + a(1+\lambda)} + P_\Delta\right)^2 = 0$$

The expectation errors follow an AR(1) process:

$$\widehat{V}_t^1 - V_t^1 = ac(\widehat{V}_{t-1}^1 - V_{t-1}^1) - b_V c\epsilon_t^1 + kP_\Theta\Theta_t$$

[10] It is well known that when a hierarchical information structure exists, the fully informed traders can infer the exact expectations of the partially informed traders. Then signal extraction problem between these traders can be characterized by a finite number of the state variables that include the expectations of the partially informed traders. If there is no hierarchical information structnre, the signal extraction problem needs to be characterized by an iufinite number of the states variables that include the infinite iteration of expectation among traders, for instance, trader A's expectation of trader B's expectation, trader B's expectation of trader A's expectation of trader B's expectation, and so on.

where $k = -\frac{1-c}{\frac{a\lambda}{\frac{\beta}{\rho\Omega}+a(1+\lambda)}+P_\Delta}$.

Proposition 13 shows that the equilibrium price in the case of three information hierarchies consists of four terms, instead of three terms as in the case of two information hierarchies . The extra term is $P_\Delta(\widehat{V}_t^1 - V_t^1)$, which captures the expectation error of type 2 traders in guessing the signal received by type 1 traders. That is, the forecasting behavior of type 2 traders adds noise to the equilibrium price. Further, as shown in Proposition 13, the expectation error follows an AR(1) process and is thus persistent.

Intuitively, the persistent structure in prices leads to a slower information diffusion rate. If there is no persistent structure in prices, the innovation in fundamentals can be incorporated immediately. But when prices are persistent, the innovation in fundamentals can bave long lasting effects so that the prices can not immediately be adjusted to reflect the fundamental value of the underlying asset. The persistent structure is due to the combined effects of persistent structure through the predictable pattern in noise traders and the persistent expectation errors formed by partially informed traders. Thus, I label the former effect the "beta effect" and the latter the "hierarchical effect". I will continue to investigate the roles of these two effects on the information diffusion rate in the next section.

2.3.1 Information Diffusion Rate

As mentioned earlier, the persistent structure in prices leads to a slower information diffusion rate. The slower information diffusion rate comes from two sources: the "beta effect" and the "hierarchical effect".

Beta effect

To study the role of the predictable pattern of noise traders' behavior in generating a slower information diffusion rate, I abstract from the hierarchical information setting. That is, I consider the two information hierarchical levels case. Without loss of generality, I restrict myself to the AR specification of fundamental evolution, i.e., $V_t = aV_{t-1} + b_v\epsilon_t$, where ϵ_t are *iid* standard normals, b_v is the inverse of the square root of the precision of the innovations, and $a < 1$. As b_v increases, the V_t becomes noisier. Then $E_t[V_{t+s}] = a^s V_t$. Hence

$$P_t = \rho P_{t-t} - \frac{\Theta_t}{\omega(1+r)} + \frac{\lambda a V_t}{1+r-(1-\lambda)a} \tag{2.14}$$

When there is no predictable pattern in noise traders' behavior, i.e., $\beta = 0$, the equilibrium price is

$$P_t = -\frac{\Theta_t}{\omega(1+r)} + \frac{\lambda a V_t}{1+r-(1-\lambda)a} \tag{2.15}$$

To demonstrate a slower information diffusion rate, I show that the impulse response in prices to an innovation in liquidity shocks brought by noise traders, i.e., in Θ_t, is larger. Intuitively, when there is an innovation in fundamentals, if the impulse response of prices is larger, it will take a longer time for prices to "settle down". That is, the information diffusion rate is slower.

Lemma 14 *The information diffusion rate is slower if there is a predictable pattern in noise traders' behavior.*

Proof: If there is no predictable pattern in noise traders' behavior, i.e., $\beta = 0$, the impulse response of prices to an innovation in liquidity shocks brought by noise traders, i.e., Θ_t, is IR_t^1 such that

$$\begin{aligned} IR_t^1 &= -\frac{1}{(1+r)\Omega} \\ IR_t^1 &= 0 \end{aligned} \tag{2.16}$$

for $t = 0, 1, \ldots$. Notice that $IR_t \geq 0$ for all t. Meanwhile, if there is a predictable pattern in noise traders' behavior, i.e., $\beta \neq 0$, the impulse response of prices to an innovation in the fundamental value of the underlying asset, i.e., V_t, is IR_t^2 such that

$$\begin{aligned} IR_0^2 &= -\frac{1}{(1+r)\Omega} \\ IR_t^2 &= \rho IR_{t-1}^2 \end{aligned} \tag{2.17}$$

for $t = 0, 1, \ldots$. It is easy to see that $IR_t^2 - IR_t^1 = \rho IR_{t-1}^2 > 0$ for all $t \geq 0$. That is, the impulse response in prices to an innovation in liquidity shocks is greater when $\beta \neq 0$ than when $\beta = 0$. $IR_t^2 > IR_t^1$ for all t implies that the difference between the current price and equilibrium price is smaller when $\beta = 0$. Therefore, it takes less time for price to converge in the case where noise traders do not have predictable pattern, i.e., $\beta = 0$. In other words, the information diffusion rate is slower if there is a predictable pattern in noise traders' behavior.

Q.E.D

Lemma 14 shows that the existence of a predictable pattern in noise traders' behavior leads to a slower information diffusion rate. The slower information diffusion rate is due to the fact that informed traders make trading decisions taking into account the predictable pattern in noise traders' behavior that generates the persistent structure in prices.

In addition, I investigate the deviation of prices from the fundamental value of the underlying asset due to the existence of the predictable pattern in noise traders' behavior. If the deviation is different than zero persistently, then the information is incorporated into prices at a slower rate.

Formally, if there is no predictable pattern in noise traders' behavior, i.e., $\beta = 0$, there is no deviation of prices from the fundamental value of the underlying asset because of the complete market. That is, when $\beta = 0$, the price in every period fully reflects the fundamental valne of the underlying asset. Meanwhile, if there is a predictable pattern in noise traders' behavior, i.e., $\beta \neq 0$, as shown in Equation 2.14, the price in every period deviates from the fundamental value of the underlying asset. The deviation is

$$D_t = \rho P_{t-1} \tag{2.18}$$

for $t = 0, 1, \ldots$. It is easy to see that $D_t > 0$ for all t. That is, the predictable pattern in noise traders' behavior leads to deviation of prices from the fundamental value of the underlying asset. The information about the fundamental value of the underlying asset, i.e., innovation in V_t, is incorporated into the price at a longer horizon. It also takes prices a longer time to adjust to account for the innovation in fundamental value of the underlying asset.

Hierarchical Effect

This section study the role of hierarchical information levels in generating a slower information diffusion rate. To isolate the beta effect brought by the predictable pattern in noise traders' behavior, I impose $\beta = 0$ in this part.

Proposition 15 *The information diffusion rate is slower in the case of three information hierarchies than in the case of two information hierarchies.*

Proof: Formally, the three information hierarchies' equilibrium when $\beta = 0$ is characterized as: If a type 1 trader observes V_t^1 and V_t^2 and a type 2 trader only observes V_t^2, the

equilibrium price is

$$P_t = P_V V_t + P_\Theta \Theta_t + P_\Delta(\widehat{V}_t^1 - V_t^1),$$

where

$$\begin{aligned} P'_\Theta &= -\frac{1}{\Omega(1+r)} \\ P'_V &= \frac{a\lambda}{1+r-(1-\lambda)a} \\ P'_\Delta &= \frac{\omega_2(1+r)P'_V}{\Omega(1+r)-\omega_1 ac} \end{aligned}$$

where $\widehat{V}_t^1 = E[V_t^1|F_{2,t}]$, $\Omega = \omega_1 + \omega_2$, $\omega_1 = \mathrm{Var}[Q_{t+1}|F_{1,t}]$, $\omega_2 = \mathrm{Var}[Q_{t+1}|F_{2,t}]$, where c solves

$$\frac{\sigma_\Theta^2}{b_V^2}\frac{(1-a\frac{1-\lambda}{1+r})^2}{\Omega^2 a^2\lambda^2}(1-c)(1-a^2c)(1+\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r}))^2 - c(\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r}))^2 = 0$$

The expectation errors follow an AR(1) process:

$$\widehat{V}_t^1 - V_t^1 = ac(\widehat{V}_{t-1}^1 - V_{t-1}^1) - b_V c\epsilon_t^1 + kP'_\Theta\Theta_t$$

where $k = (1-c)/(\frac{a\lambda}{1+r-a(1-\lambda)}\frac{\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r})}{1+\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r})})$. It is easy to see that $P'_V > 0$. In addition, $P'_V > P'_\Delta > 0$. To see that,

$$\begin{aligned} P'_\Delta &= \frac{\omega_2(1+r)P'_V}{\Omega(1+r)-\omega_1 ac} \\ &= \frac{1}{\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r})+1}P'_V \end{aligned} \tag{2.19}$$

Notice that $\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r})+1 > 1$, so that $P'_V > P'_\Delta > 0$. The impulse response of prices to an innovation in liquidity shocks of the underlying asset IR_t^3 is

$$\begin{aligned} IR_0^3 &= (kP_\Delta + 1)P'_\Theta \\ IR_t^3 &= a^t c^t k P'_\Theta \quad for \quad t > 0 \end{aligned} \tag{2.20}$$

Notice that because shocks in the supply by noise traders, i.e., the innovations in Θ_t, affect the price persistently because that they are incorporated into the persistent expectation errors formed by the partially informed trader. It takes time for the partially informed traders to learn that liquidity shocks are irrelevant to the fundamental value of the underlying asset. Thus, it takes prices a longer time to adjust to the shocks in supply by noise traders and the information diffusion rate is slower.

Q.E.D

I further investigate the deviation of prices from the fundamental value of the underlying asset due to the existence of the predictable pattern in noise traders' behavior. If the deviation is consistently non-zero, then the information is incorporated into prices at a slower rate.

Formally, if there is no predictable pattern in noise traders' behavior, i.e., $\beta = 0$, there is no deviation of prices from the fnndamental value of the underlying asset because of the complete market. That is, when $\beta = 0$, the price in every period fully reflects the fundamental value of the underlying asset. Meanwhile, with three hierarchical information levels, the price in every period deviates from the fundamental value of the underlying asset. The deviation is

$$D_t = P'_{\Delta}(\widehat{V}_t^1 - V_t^1) \tag{2.21}$$

for $t = 0, 1, \ldots$. It is easy to see that $D_t > 0$ for all t. That is, the case of three information hierarchies leads to the deviation of prices from the fundamental value of the underlying asset. The information about the fundamental value of the underlying asset, i.e., innovation in V_t, is incorporated into the price at a longer horizon. Therefore, it also takes prices a longer time to adjust to the innovation in fnndamental value of the underlying asset.

2.3.2 Stationary Equilibrium

Similar to the case of two information hierarchies, there are two possible values of ρ for the equilibrium. Solving the roots explicitly, we have

$$\rho_{1,2} = \frac{(1+r) \pm \sqrt{(1+r)^2 + 4(1-\lambda)\frac{\beta}{\Omega}}}{2(1-\lambda)} \tag{2.22}$$

From Proposition 13,we have $P_t - \rho P_{t-1} = +P_V V_t + \frac{\rho}{\beta}\Theta_t + P_{\Delta}(\widehat{V}_t^1 - V_t^1)$. The right hand side (RHS) of the equation is assumed to be stationary and has finite variance given the stationary assumption of V_t. Therefore, the price is stationary when $|\rho| < 1$. Formally,

Proposition 16 *The relationship between the stationarity of equilibrium and β can be summarized as follows:*

1. *When noise traders are contrarians, the maximum number of stationary equilibria is 1. Formally, there exist a unique stationary equilibrium if and only if the probability of liquidation is not sufficiently large ($\lambda < 2 + r - \frac{\beta}{\Omega}$).*

2. *When noise traders are trend followers and they are not too sensitive to price direction changes* ($-\frac{\Omega(1+r)}{2} < \beta < 0$)*, and if the probability of liquidation is not sufficiently large* ($\lambda < -r - \frac{\beta}{\Omega}$)*, two stationary equilibria exist. Otherwise, if liquidation is sufficiently large* ($\lambda > -r - \frac{\beta}{\Omega}$)*, there exists a unique stationary equilibrium.*

3. *When noise traders are contrarians and they are sensitive to price direction changes* ($\beta < -\frac{\Omega(1+r)}{2}$)*, and the probability of liquidation is sufficiently large* ($-r - \frac{\beta}{\Omega} < \lambda$*, there is a unique stationary equilibrium. If the probability of liquidation is sufficiently small* ($\lambda < \frac{1-r}{2}$)*, there are two stationary equilibria. Otherwise, i.e.,* $\frac{1-r}{2} < \lambda < -r - \frac{\beta}{\Omega}$*, there is no stationary equilibrium.*

Proposition 16 suggests that in addition to the number of equilibria, the predictable pattern of noise traders' behavior can affect the stationarity of the equilibrium price process as well. It is interesting to note that the impact of the predictable pattern of noise traders' behavior affects the stationarity of the equilibrium price process in a different way. For instance, in the case of two information hierarchies, when noise traders are trend-followers, there can be two stationary equilibria, while at most one equilibrium can exist in the case of three information hierarchies when noise traders are trend-followers.

2.3.3 Persistent Prices

Proposition 13 allows us to calculate the correlation structure of prices explicitly. It can be shown that in stationary equilibrium, the variance of price is

$$\mathrm{Var}[P_t] = \frac{1}{1-\rho^2}\frac{a^2\lambda^2}{(\frac{\beta}{\rho\Omega}+a(1+\lambda))^2}\mathrm{Var}[V_t] + \frac{\rho^2}{\beta^2}\sigma_\Theta^2 + P_\Delta^2\frac{1}{1-\rho^2 c^2}(b_V^2 c^2 + \frac{k^2\sigma_\Theta^2}{\omega^2(1+r)^2}).$$

This demonstrates that there are two sources of noise in the price. One is the random (unpredictable) behavior of noise traders, and the other is from the expectation errors of partially informed trader. Notice that if the noise traders' behavior becomes noisier, i.e., σ_Θ increases, the variance of expectation errors of the partially informed trader also increases. Intuitively, partially informed trader will try to distinguish the effects of noise traders' behavior and of the signals of the fully iuformed trader on prices. Because noise traders' behavior does not reveal any information on fundamental value of the underlying asset, it is not helpful in forming expectations of the fundamental value of the underlying asset. If the behavior of noise traders becomes noisier, it is harder for the partially informed traders

to extract useful information on fundamentals. The partially informed trader will make greater errors in forming expectations of fundamental value of the underlying asset, which also leads to a noisier market price.

Furthermore, the correlation structure of prices can be recursively represented as:

$$\begin{aligned}
\mathrm{Cov}[P_t, P_{t-1}] &= \rho \mathrm{Var}[P_{t-1}] + \frac{1}{1-\rho^2}\left(\frac{a^2\lambda^2}{(\frac{\beta}{\rho\Omega}+a(1+\lambda))^2}\right)\mathrm{Cov}[V_t, V_{t-1}] \\
&\quad + P_\Delta^2 \mathrm{Cov}[\widehat{V}_t^1 - V_t^1, \widehat{V}_{t-1}^1 - V_{t-1}^1] \\
&= \rho \mathrm{Var}[P_t] + \frac{a}{1-\rho^2}\left(\frac{a^2\lambda^2}{(\frac{\beta}{\rho\Omega}+a(1+\lambda))^2}\right) + P_\Delta^2 ac \\
\mathrm{Cov}[P_t, P_{t-s}] &= \rho \mathrm{Cov}[P_t, P_{t-s+1}] + \frac{a^s}{1-\rho^2}\left(\frac{a^2\lambda^2}{(\frac{\beta}{\rho\Omega}+a(1+\lambda))^2}\right) + P_\Delta^2 (ac)^s
\end{aligned}$$

This shows that the prices are more persistent compared to the case of two information hierarchies because of the additional correlation structure brought about by the expectation errors, i.e., $P_\Delta^2(ac)^2$.

2.3.4 Return Predictability

The return is defined by the difference in prices, i.e., $r_t = P_t - P_{t-1}$. From Proposition 13, return is

$$r_t = \rho r_{t-1} - \frac{a\lambda}{\frac{\beta}{\rho\Omega}+a(1+\lambda)}\Delta V_t + \frac{\rho}{\beta}\Delta\Theta_t + P_\Delta \Delta(\widehat{V}_t^1 - V_t^1), \tag{2.23}$$

where $\Delta V_t = V_t - V_{t-1}$, $\Delta\Theta_t = \Theta_t - \Theta_{t-1}$ and $\Delta(\widehat{V}_t^1 - V_t) = (\widehat{V}_t^1 - V_t^1) - (\widehat{V}_{t-1}^1 - V_{t-1}^1)$.

Notice that the statistical structure of returns is determined by the statistical properties of ΔV_t, $\Delta\Theta_t$, and $\Delta(\widehat{V}_t^1 - V_t)$, which correspond to evolution of the fundamental value of underlying asset, noise traders' behavior, and expectation errors. Notice that ΔV_t follows an ARMA(1,1) process, $\Delta(\widehat{V}_t^1 - V_t)$ follows an ARMA(1,1) process and $\Delta\Theta_t$ follows a MA(1) process. Formally,

$$\begin{aligned}
\Delta V_t = V_t - V_{t-1} &= a(V_{t-1} - V_{t-2}) + b_V(\epsilon_t - \epsilon_{t-1}) = a\Delta V_{t-1} + b_V(\epsilon_t - \epsilon_{t-1}) \\
\Delta(\widehat{V}_t^1 - V_t) &= ac(\Delta(\widehat{V}_t^1 - V_t)) - (b_V c\epsilon_t^1 - k\frac{\rho}{\beta}\Theta_t - b_V c\epsilon_{t-1}^1 + k\frac{\rho}{\beta}\Theta_{t-1})
\end{aligned} \tag{2.24}$$

Similar to the analysis of prices, the correlation structure of returns is

$$\begin{aligned}
\mathrm{Cov}[r_t, r_{t-1}] &= \rho \mathrm{Var}[r_{t-1}] + \frac{a^2\lambda^2}{(\frac{\beta}{\rho\Omega} + a(1+\lambda))^2}\mathrm{Cov}[\Delta V_t, \Delta V_{t-1}] - \frac{\rho^2}{\beta^2}\sigma_\Theta^2 \\
&\quad + P_\Delta^2 \mathrm{Cov}[\Delta(\widehat{V}_t^1 - V_t), \Delta(\widehat{V}_{t-1}^1 - V_{t-1})] \\
\mathrm{Cov}[r_t, r_{t-s}] &= \rho \mathrm{Cov}[r_t, r_{t-s+1}] + \frac{a^2\lambda^2}{(\frac{\beta}{\rho\Omega} + a(1+\lambda))^2}\mathrm{Cov}[\Delta V_t, \Delta V_{t-s}] - \frac{\rho^2}{\beta^2}\sigma_\Theta^2 \\
&\quad + P_\Delta^2 \mathrm{Cov}[\Delta(\widehat{V}_t^1 - V_t), \Delta(\widehat{V}_{t-s}^1 - V_{t-s})] \qquad (2.25)
\end{aligned}$$

This shows that there is a serial correlation structure in returns. Moreover, it is possible to generate momentum, i.e., a positive correlation in returns, under certain parameter values.

In this section, I investigate the impact of interaction between informed traders at different information hierarchies, namely, the fully informed trader and partially informed trader, on eqnilibrium properties. I show that compared to the case of two information hierarchies, the market price contains an extra term, which is the expectation error originating from signal extraction by the partially informed trader and adding noise to the market price. This leads to a slower information diffnsion rate. The slower information diffusion rate is due to the combined effects of the "beta" effect (the persistence structure brought by the predictable pattern in noise traders' behavior) and the "hierarchical effect" (the extra noise brought by the signal extraction behavior of the partially informed trader). Therefore, the increased number of iuformation hierarchies decreases the information diffusion rate. That is, when there is an innovation in the fundamental value of the underlying asset, it takes a longer time for the market price to incorporate the innovation and reflects the fundamental value of the underlying asset. Furthermore, prices become more persistent. The extra persistence is also due to the persistence in the expectation errors made by the partially informed trader, which are due to the persistence of the belief updating. In addition, momentum in returns can be generated as well. Two sources contribute to the formation of momentum. First, as in the case of two information hierarchies, the predictable pattern of noise traders' behavior can aid in the formation of momentum via the self-fulfilling mechanism. Second, the autocorrelations in expectation errors can contribute to the momentum as well.

2.4 Case of N Information Hierarchies

I extend the Case of three information hierarchies to allow for N information hierarchies to study the impact of increasiug the number of information hierarchies. Intuitively, with

increasing number of information hierarchies, the fully informed traders may want to trade with the traders in lower information hierarchies. This is because it is possible that the riskiness of trading with the partially informed traders is lower than trading with noise traders because the riskiness originating from the unpredictable pattern of their behavior may be relatively larger. Formally, I assume there are N factors determining the fundamental value of the underlying asset in a linear fashion, i.e., $V_t = \sum_{n=1}^{N-1} V_t^n$. In addition, I assume that there is no noise in signals for all types of traders, and the type i informed trader observes $V_t^i, \ldots, V_t^{N-1}$. Assume V_t is an AR(1) process, i.e., $V_t^n = aV_{t-1}^n + b_V \epsilon_t^n$ for $n = 1, 2, \ldots, N-1$. The information set for a type i trader is $F_{i,t} = \{P_t, P_{t-1}, \ldots, V_t^i, V_{t-1}^i, \ldots, V_t^{N-1}, \ldots\}$. By construction, $F_{i,t} \subset F_{j,t}$ if $i > j$.

Proposition 17 *If a type i informed trader observes $V_t^i, \ldots, V_t^{N-1}$, for $i = 1, 2, \ldots, N-1$. The equilibrium price is*

$$P_t = \rho P_{t-1} + P_V V_t + P_\Theta \Theta_t + P_{\Delta^1}(\widehat{V}_{2,t}^1 - V_t^1) + \sum_{s=3}^{N-1} P_{\Delta^{s-1}}(\sum_{n=1}^{s-1} \widehat{V}_{s,t}^n - \sum_{n=1}^{s-1} V_t^n),$$

where

$$\begin{aligned} P_\Theta &= \frac{\rho}{\beta} \\ P_V &= \frac{a\lambda}{-\frac{\beta}{\rho\Omega} - a(1-\lambda)} \\ P_{\Delta^{N-2}} &= \frac{a\omega_{N-1}(\lambda + (1-\lambda)P_V)}{-\frac{\beta}{\rho} - a(1-\lambda)\sum_{i=1}^{N-2}\omega_i} \end{aligned} .$$

for $1 < s < N-2$,

$$P_{\Delta^s} = \frac{a\sum_{i=s}^{N-1}\omega_i(\lambda + (1-\lambda)P_V) - a\omega_{s+1}\sum_{i=s+1}^{N-2} P_{\Delta^i}}{-\frac{\beta}{\rho} - a(1-\lambda)\sum_{i=1}^{s-1}\omega_i}$$

for $s = 1$,

$$P_{\Delta^1} = \frac{a\omega_2(\lambda + (1-\lambda)P_V) - ac\omega_2\sum_{i=2}^{N-2} P_{\Delta^i}}{-\frac{\beta}{\rho} - a\omega_1(1-\lambda)c}$$

where $\widehat{V}^j_{s,t} = E[V^j_t|F_{s,t}]$ *for* $j = 1, 2, \ldots, s-1$; $\Omega = \sum_{n=1}^{N-1} \omega_n$; $\omega_s = Var[Q_{t+1}|F_{s,t}]$, ρ *solves*

$$(1-\lambda)\Omega\rho^2 - (1+r)\Omega\rho - \beta = 0$$

and c *solves*

$$\frac{\rho^2}{\beta^2}\sigma^2_\Theta(1-c)(1-a^2c) - cb^2_V(P_V - \sum_{s=1}^{N-2} P_{\Delta^s})^2 = 0$$

The expectation errors from type 2 traders follow an AR(1) process:

$$\widehat{V}^1_{2,t} - V^1_t = ac(\widehat{V}^1_{2,t-1} - V^1_{t-1}) - b_V c\epsilon^1_t + kP_\Theta\Theta_t$$

where $k = \frac{1-c}{\frac{a\lambda}{\frac{\beta}{\rho\Omega}-a(1+\lambda)} - \sum_{s=1}^{N-2} P_{\Delta^s}}$.

The expectation errors from type s *traders follow an AR(1) process, for* $s > 2$*:*

$$\sum_{n=1}^{s-1} \widehat{V}^n_{s,t} - \sum_{n=1}^{s-1} V^n_t = a(\sum_{n=1}^{s-1} \widehat{V}^n_{s,t-1} - \sum_{n=1}^{s-1} V^n_{t-1}) - b_V \sum_{i=1}^{s-1} \epsilon^i_t.$$

Remarks:

1. *Observational equivalence of informed traders in higher information hierarchies.* Notice that, the noise traders' behavior only affects the type 2 trader (the partially informed trader in second highest information hierarchy) directly. Partially informed traders in other information hierarchies need to distinguish more than two sources of randomness, and they only have the market price as an identifying instrument. One interesting result in this setting is that the forecasting of the signals owned by the trader in a higher hierarchy collapses. Formally, $E[E[V^s_t|F_{j,t}]|F_{i,t}] = E[V^s_t|F_{i,t}]$, which follows directly from the law of iterated expectation. It simply states that the traders' best guess in lower hierarchies concerning the signal received by the traders in higher hierarchies is their own expectation of the fundamental value of the underlying asset. This suggests that traders in lower information hierarchies simply cannot distinguish the traders in higher information hierarchies, implying that all traders in higher information hierarchies are observationally equivalent in the perspective of a trader in a lower hierarchy.

2. *Traders in higher information hierarchies may want to trade with informed traders in lower information hierarchies.* Intuitively, the fully informed trader knows the optimal trading decisions of any partially informed traders in lower hierarchies. Therefore, there is only one type of risk if the fully informed trader wants to trade with the partially informed traders, which originates from the probability of being liquidated. Meanwhile, in addition to the liquidation risk, the fully informed trader faces extra risk when he wants to trade with noise traders, which is due to unpredictable part of noise traders' behavior. In other words, it may be more profitable for the fully informed trader to trade with the partially informed traders rather noise traders. It can be shown that under certain parameter values, the fully informed trader will trade against the partially informed traders rather than noise traders. I will elaborate more on it in a later section.

3. *The rationale for the presence of noise traders.* Following the analysis above, when traders with an informational advantage choose not to trade with noise traders, noise traders could make a profit. The conventional wisdom that noise traders could not make a profit in the long run dictates that informed traders have the informational advantages and do at least well as noise traders. However, in my model, informed traders may prefer trading among themselves rather than trade against noise traders. This may lead to profits for noise traders even in the long run. Thus, the presence of noise traders on the market is justified.

2.4.1 Information Diffusion Rate

Similar to the case of three information hierarchies, the persistent structure in prices leads to a slower information diffusion rate. The slower information diffusion rate is due to two sources: the "beta effect" and the "hierarchical effect", as in the case of three information hierarchies. The "beta effect" works similar to that in the case of three informational hierarchies and we restrict ourselves to the study of the "hierarchical effect". In Particular, we want to study the relationship between information diffusion rate and the number of information hierarchies.

Hierarchical Effect

This section studies the roles of the number of information hierarchies in generating a slower information diffusion rate. To isolate the beta effect brought by the predictable pattern in noise traders' behavior, I impose $\beta = 0$ in this section.

Proposition 18 *The information diffusion rate slows as the number of information hierarchies increases.*

Proof: Formally, the N information hierarchies equilibrium when $\beta = 0$ is characterized as: If a type i informed trader observes $V_t^1, \ldots, V_t^{N-1}$, for $i = 1, 2, \ldots, N-1$. The equilibrium price is

$$P_t = P_V'' V_t + P_\Theta'' \Theta_t + P_{\Delta^1}''(\widehat{V}_{2,t}^1 - V_t^1) + \sum_{s=3}^{N-1} P_{\Delta^{s-1}}''(\sum_{n=1}^{s-1} \widehat{V}_{s,t}^n - \sum_{n=1}^{s-1} V_t^n),$$

where

$$\begin{aligned} P_\Theta'' &= -\frac{1}{\Omega(1+r)} \\ P_V'' &= \frac{a\lambda}{1+r-(1-\lambda)a} \\ P_{\Delta^{N-2}}'' &= \frac{a\omega_{N-1}(\lambda + (1-\lambda)P_V'')}{(1+r)\Omega - a(1-\lambda)\sum_{i=1}^{N-2}\omega_i} \end{aligned}$$

for $1 < s < N-2$,

$$P_{\Delta^s}'' = \frac{a\sum_{i=s}^{N-1}\omega_i(\lambda + (1-\lambda)P_V) - a\omega_{s+1}\sum_{i=s+1}^{N-2} P_{\Delta^i}}{(1+r)\Omega - a(1-\lambda)\sum_{i=1}^{s-1}\omega_i}$$

for $s = 1$,

$$P_{\Delta^1}'' = \frac{a\omega_2(\lambda + (1-\lambda)P_V'') - ac\omega_2\sum_{i=2}^{N-2} P_{\Delta^i}''}{(1+r)\Omega - a\omega_1(1-\lambda)c}$$

where $\widehat{V}_{s,t}^j = E[V_t^j | F_{s,t}]$ for $j = 1, 2, \ldots, s-1$; $\Omega = \sum_{n=1}^{N-1}\omega_n$; $\omega_s = \mathrm{Var}[Q_{t+1}|F_{s,t}]$, where c solves

$$\frac{\sigma_\Theta^2}{b_V^2}\frac{(1-a\frac{1-\lambda}{1+r})^2}{\Omega^2 a^2\lambda^2}(1-c)(1-a^2c)(1+\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r}))^2 - c(\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r}))^2 = 0$$

The expectation errors from type 2 traders follow AR(1) process:

$$\widehat{V}^1_{2,t} - V^1_t = ac(\widehat{V}^1_{2,t-1} - V^1_{t-1}) - b_V c\epsilon^1_t + kP''_\Theta \Theta_t$$

where $k = (1-c)/(\frac{a\lambda}{1+r-a(1-\lambda)} \frac{\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r})}{1+\frac{\omega_1}{\omega_2}(1-ac\frac{1-\lambda}{1+r})})$.

The expectation errors from type s traders follow AR(1) process, for $s > 2$:

$$\sum_{n=1}^{s-1} \widehat{V}^n_{s,t} - \sum_{n=1}^{s-1} V^n_t = a(\sum_{n=1}^{s-1} \widehat{V}^n_{s,t-1} - \sum_{n=1}^{s-1} V^n_{t-1}) - b_V \sum_{i=1}^{s-1} \epsilon^i_t.$$

It is easy to see that $P''_V > 0$.

The impulse response of prices to an innovation in liquidity shocks of the underlying asset IR^4_t, is

$$\begin{aligned} IR^4_0 &= (kP''_{\Delta^1} + 1)P''_\Theta \\ IR^4_t &= a^t c^t k P''_\Theta \quad for \quad t > 0 \end{aligned} \tag{2.26}$$

Notice that liquidity shocks, i.e., the innovations in Θ_t, affect the price persistently because that the shocks in supply by noise traders are incorporated into persistent expectation errors formed by the partially informed traders. It takes time for the partially iuformed traders to learn that the changes in prices are due to shocks in supply and are irrelevant to the fundamental value of the underlying asset. Thus, it takes prices a longer time to adjust to account for the shocks in supply of noise traders and the information diffusion rate is slower.

Q.E.D

Further, I investigate the deviation of prices from the fundamental value of the underlying asset due to the existence of the predictable pattern in noise traders' behavior. If the deviation is consistently non-zero, the information is incorporated into prices at a slower rate.

Formally, if there is no predictable pattern in noise traders' behavior, i.e., $\beta = 0$, there is no deviation of prices from the fundamental value of the underlying asset because of the complete market. That is, when $\beta = 0$, the price in every period fully reflects the fundamental value of the underlying asset. Meanwhile, with N hierarchical information levels, the price in every period deviates from the fundamental value of the underlying asset. The deviation is

$$D_t = P_{\Delta^1}(\widehat{V}^1_{2,t} - V^1_t) + \sum_{s=3}^{N-1} P_{\Delta^{s-1}}(\sum_{n=1}^{s-1} \widehat{V}^n_{s,t} - \sum_{n=1}^{s-1} V^n_t) \tag{2.27}$$

for $t = 0, 1, \ldots$. It is easy to see that $D_t > 0$ for all t. That is, the case of N information hierarchies leads to deviation of prices from the fundamental value of the underlying asset. That is, the information about the fundamental value of the underlying asset, i.e., innovation in V_t is incorporated into the price at a longer horizon. In addition, as the number of information hierarchies increases, the deviation could be even larger because of the existence of the expectation errors formed by all partially informed traders.

2.4.2 Trading Among Informed Traders

As mentioned earlier, the traders in higher information hierarchies may want to trade against the traders in lower information hierarchies. To illustrate, let us consider the case of three information hierarchies. As shown in Appendix A, the demand from the fully informed trader is

$$\omega_1(((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)aV_t + (1-\lambda)P_\Delta E[(\widehat{V}^1_{t+1} - V^1_{t+1})|F_{1,t}] \quad (2.28)$$

while the demand from the partially informed trader is

$$\omega_2(((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)a(\widehat{V}^1_t + V^2_t)) \quad (2.29)$$

where $\widehat{V}^1_t = a(1 - k(P_V - P_\Delta))\widehat{V}^1_{t-1} + k(P_V - P_\Delta)V^1_t + kP_\Theta\Theta_t$.

Consider a scenario where there is a supply shock and no change in the fundamental value of the underlying asset, such that noise traders choose to supply more, i.e., $\Delta\Theta_t > 0$. Although there is no change in the fundamental value of the underlying asset, both types of informed traders will adjust their optimal demands accordingly. Notice that partially informed traders will adjust their expectations of the fundamental value of the asset downward, due to the fact that they cannot distinguish the decrease in fundamental value of the underlying asset from a positive supply shock. Hence, the quantity of adjustment for partially informed traders is $(\lambda + (1-\lambda)P_V)akP_\Theta\Delta\Theta_t$ and for fully informed traders is $(1-\lambda)P_\Delta E[(\widehat{V}^1_{t+1} - V^1_{t+1})|F_{1,t}]$.

As indicated earlier in Proposition 13, the expectation error is an AR(1) process and thus persistent. Therefore, from the perspective of the fully informed trader, the expectation error should have the same signs for time t and time $t+1$.

If $P_\Theta < 0$ and thus $(\lambda + (1-\lambda)P_V)akP_\Theta < 0$, partially informed traders will choose to decrease their optimal demand by mistakenly believing that the fundamental value of the asset decreases. This is simply because they cannot distinguish the decrease in fundamental

value in the underlying asset from the a positive supply shock. On the other hand, when $P_\Delta < 0$ and thus $(1-\lambda)P_\Delta < 0$, fully informed traders will trade based on the expectation errors from partially informed traders. Notice that partially informed traders will adjust their expectation of the fundamental value of the nnderlying asset downward, i.e., $\widehat{V}_t^1 - V_t^1 < 0$. From the persistence of the expectation error, fully informed traders will believe $E[(\widehat{V}_{t+1}^1 - V_{t+1}^1)|F_{1,t}] < 0$. Thus, they will increase the holding of the underlying asset in order to make profit from the mistakes of the partially informed traders. Therefore, fully informed traders trade against partially informed traders by trading in opposite directions. In summary, I show that when $P_\Theta < 0$ aud $P_\Delta < 0$, fully informed traders trade against partially informed traders in order to profit from the mistakes of the partially informed traders.

Proposition 19 *If $\beta < 0$, $\rho > 0$ and $0 < k < 1$, trading amongst informed traders can happen. In addition, in this case, $P_V < 0$.*

$\beta < 0$ implies that noise traders are trend followers and $\rho > 0$ implies that the price has a positive autocorrelation coefficient which matches the empirical observation of the financial time scries of price. According to the Bucy-Kalman filter formula, k is the weight used in the belief updating. A positive k suggests that new information and current belief both receive attention. As a result, the partially informed traders adjust their belief in a wrong way when there is a positive shock in supply.

A negative P_V suggests that when there is a positive innovation in fundamental value of the underlying asset, the price may decrease in response. This is inefficient because such trading amongst informed traders can prolong the proccdure of information diffusion and thus the rate of information diffusion is slower.

Proposition 19 characterizes one set of possible parameter values which can generate trading amongst informed traders. That is, partially informed traders mistakenly adjust their beliefs in a wrong way when they cannot distinguish the positive shock in supply from a decrease in fundamental value of the underlying asset. If fully informed traders know that noise traders are trend followers and partially informed traders make a mistake in adjusting their beliefs, they will trade against the partially informed traders to make a profit. This profitablc opportunity originates from the mistakes made by thc partially informed tradcrs. This is inefficient, because if that is the case, it leads to a negative response of current market price to a positive innovation in the fundamental value of the underlying asset. This

distortion of price response is due to the informational arbitrage amongst informed traders.

2.5 Numerical Analysis

In this section, I present a nnmerical study of the model. Using this numerical study, I can show the statistical properties of equilibrium, namely, the price structure, and the retnrn predictability. Second, I want to investigate the "beta effect" and the "hierarchical effect" and their implications.

The parameters used in the simulation study are as follows. The probability of liquidation λ is 5%, which corresponds to an expected lifespan of 20 periods. The risk-free rate r is 1%, which corresponds to the annual yield on a treasury bill. The parameters that describe the fundamental value of the underlying asset are chosen to match the volatility of returns at 12% and I set $a = 0.85$ and $b_V = 0.3$. I vary thc above parameters and find that the different sets of parameters do not change the main results qualitatively.

2.5.1 Impact of Noise Traders

First, I study the effects of noise traders' behavior. I mainly study the behavior of noise traders in the case of two information hierarchies and the case of three information hierarchies. To match the prices' empirical behavior, we want to have a positive AR(1) coefficient that is close to 1. I choose $\beta = -2$ in the case of two information hierarchies and $\beta = -20$ in the case of three information hierarchies. The Figure 2.1 shows a typical set of plots of price series, return series, and persistent structure which is captured by autocorrelations of prices and returns. It shows that the prices are persistent, and there is a first order positive serial correlation with the magnitude 0.8 in returns. There is a positive serial correlation in returns which suggests the existence of momentum. Similar patterns can also be observed in the plot of the case of three information hierarchies, although the first order serial correlation in returns drops and returns become much less persistent, which is a better approximation of the empirical behavior of returns. This decrease in persistence strncture in returns may be due to the existence of interactions between the informed traders which diminishes the effect of the persistent structure brought by the predictable pattern in noise traders' behavior. Thus, when there are more information hierarchies, the magnitude of the momentum effect is smaller, as shown in Table 2.2. The magnitude of the first order autocorrelation of returns decreases when the number of information hierarchies increases.

Furthermore, Table 2.2 indicates that the persistence structure in returns decreases when the number of information hierarchies increases.

Table 2.1 shows that the simulation results for "beta effect" in the case of three information hierarchies. First column reports AR coefficient of price series, ρ. It demonstrates that when the magnitude of β increases, the AR coefficient ρ decreases. Intuitively, when the size of noise trading on a financial market increases, the behavior of the market price is driven by the noise traders' behavior. The forecasting behavior of informed traders which contribute to the persistence of price in a diminishing manner. The second column reports the first lag serial correlation in returns which also suggests that the magnitude of "momentum effect" decreases as the magnitude of β increases.

There is an ARCH effect in returns which suggests the volatility clustering of returns. Intuitively, because there are potentially two equilibrium with two ρs, the switching from equilibrium can induce the volatility clusteriug of returns. The persistence structure of squared returns decreases as the magnitude of β increases as evidenced by the decrease in the reported sum of the autocorrelation coefficients of first 10 lags of squared returns.

Define $\Delta_t = \widehat{V}_t^1 - V_t^1$. Table 2.1 reports the effects of the predictable pattern of the noise traders behavior. β become more negative implyiug that the noise traders trade more and more aggressively as trend-followers.

The negative correlation between Δ_t and V_t is interesting. It implies that when there is a positive shock in the fundamental value of the underlying asset, the expectation is biased downward. Therefore, the partially informed trader takes a fundamental risk. When noise traders trade more and more aggressively, the partially informed trader takes more and more fundamental risk. The seventh column shows that given the noise traders are trend-followers, they are right about price direction. It is may be due to that the fully informed trader may find out that it is not that profitable to trade with noise traders and he chooses to trade with the partially informed traders using his own informational advantage. Notice that in this setting, it may be less risky to trade with the partially informed traders than with noise traders. Although there is predictable pattern in noise traders' behavior, it becomes increasingly risky when noise traders trade aggressively. Hence, the fully informed trader instead chooses to exploit his informational advantage over the partially informed traders.

The negative correlation between expectation errors and supply shocks is also consistent with our intuition. Intuitively, if there is a positive shock in liquidity supply, the price will

go down. The partially informed trader cannot distinguish whether the drop in price is due to a negative shock iu fundaiuentals or a positive shock in supply. Hence, he adjusts his expectation downward.

The deviation of price from the fundamentals is increasing in β. When the magnitude of β decreases, i.e, traders trade less aggressively, ρ increases. When ρ approaches 1, there is a sharp increase in the deviation corresponding to a 98% increase in mean of price. It seems that as noise traders trade less aggressively, the price becomes more persistent and the a bubble, i.e., the deviation of market price from the fundamental value of the underlying asset, is generated.

2.5.2 Impacts of the Number of Hierarchies

I choose $\beta = -20$ for all values of N. Table 2.2 reports the effects of the number of information hicrarchies. First thing to note is that the corresponding AR coefficients of price monotonically increase when N increases, implying that price becomes more persistent. Intuitively, if there are more information hierarchies, the expectation errors from the signal extraction behavior of all partially informed traders make the price more persistent. The weight of the past price in determining the current price increases as more partially informed traders try to do signal extraction. Thus, the current price is more correlated with the past price, i.e., the AR coefficient for price increases.

The autocorrelation coefficient of return decreases as the number of the information hicrarchies increases as I indicated earlier. To be consistent with the empirical finding, the number of information hierarchies should be greater than 9, which generates a 0.1 to 0.2 autocorrelation coefficient. ARCH effect exists in the returns as well. And the persistence structure of squared returns increases as the number of information hierarchies increase.

As shown in the last column of Table 2.2, the slower information diffusion rate is captured by the number of time periods it takes for the price to converge back to the long run equilibrium price. To do this, I start with a positive supply shock with a magnitude 0.1 and set all V_ts constant. Then I examine the difference between the current price and the fundamental value of the underlying asset. If the difference is sufficiently small, the price converges to the fundamental value.[11] Then I report the uumber of periods for the price to converge. We can see clearly that as the number of information hierarchies increases, it takes

[11] In the simulation. I use 10^{-4} as a threshold level.

a longer time for price to converge back to the long run fundamental value of the underlying asset. This provides simulation evidence of one of the main results of the chapter: As the number of information hierarchies increases, the information diffusion is slower. A similar experiment is studied for the shock in the fundamental value of the underlying asset, and a similar pattern is found.

2.6 Conclusion

In this chapter, I develop a discrete time, infinite time period model to understand the relationship between hierarchical information and price discovery. The partially informed traders trade a stock based on their signals and expectation of the signals received by fully informed traders. The expectation errors from the signal extraction behavior of the partially informed traders is preserved in the market price, which prevents the information from being released at the same rate as if there was no hierarchical information among traders and the information was released instantaneously. As the number of informational hierarchies increases, it becomes harder for the partially informed traders to distinguish between the shock in fundamental value of the underlying asset and liquidity shock brought about by noise traders. The slower price discovery process of the financial market can also be partly due to the fact that fully informed traders may prefer to trade with partially informed traders. In order to make a profit from the partially informed traders, the fully informed trader should prevent the information disclosure and takes the opposite position of the partially informed traders if they make large expectation errors. Therefore, this informational arbitrage leads to a slower information diffusion rate.

In addition, noise traders' behavior is modeled with a predictable behavior pattern. This framework yields a number of interesting findings. For instance, the predictable pattern of noise traders brings the persistence structure in prices regardless of the statistical properties of the fundamentals of the underlying asset. In addition, the predictable pattern of noise traders' behavior can aid in the formation of the momentum, i.e., the positive autocorrelation in returns.

One potentially interesting avenue for future work is to endogenize the evolution of the predictable pattern of noise traders' behavior. Doing so would allow us to determine the relationship between the information diffusion and liquidity of the market. For instance, this extra uncertainty of the predictable pattern in noise traders' behavior may affect the

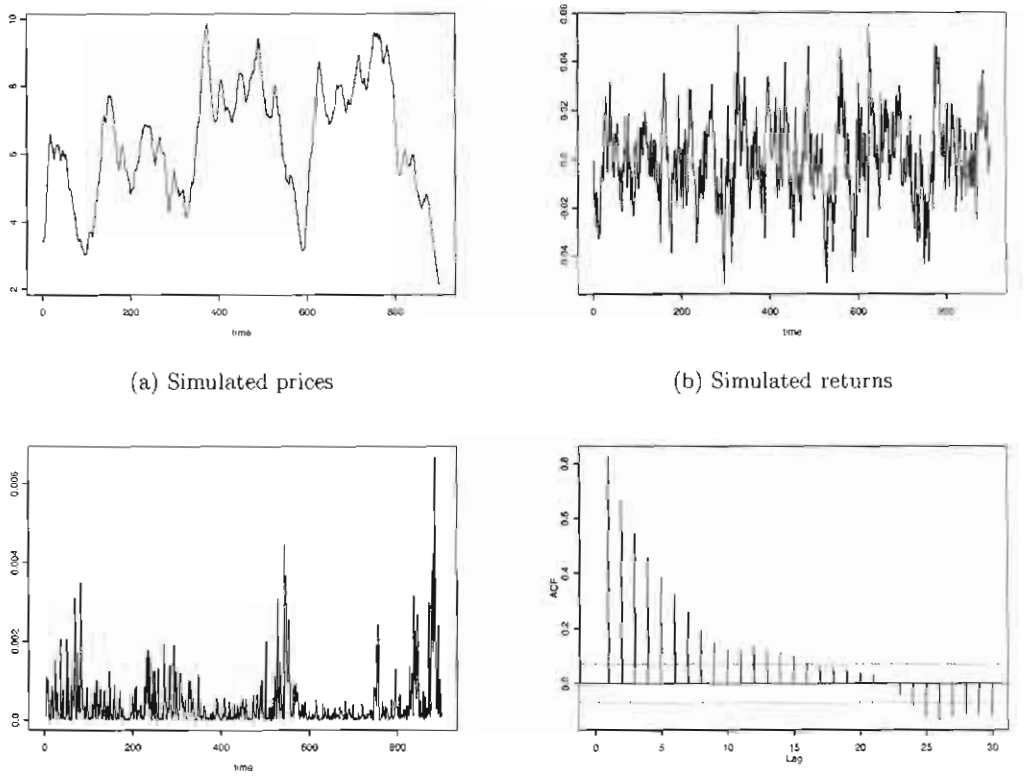
interactions among informed traders and further affect the information diffusion speed. Meanwhile, the evolution of the liquidity supply of noise traders will affect the market depth. This study may provide insights into the correlation of market depth and information diffusion speed.

β	ρ	Corr$[r_t, r_{t-1}]$	ARCH	$\sum \lvert ACF(r_t^2) \rvert$	Corr$[V_t, \Delta_t]$	Corr$[\Theta_t, \Delta_t]$	Bubble
-11	0.995	0.753	0	1.879	-0.460	-0.508	0.670
-15	0.960	0.736	0	1.092	-0.541	-0.331	0.025
-20	0.922	0.723	0	0.964	-0.586	-0.213	0.010
-25	0.880	0.697	0	0.867	-0.607	-0.130	0.006
-30	0.833	0.657	0	0.775	-0.618	-0.068	0.003
-35	0.776	0.585	0	0.587	-0.624	-0.018	-0.001
-40	0.701	0.536	0	0.640	-0.626	-0.028	0.013
-45	0.594	0.488	0	0.506	-0.630	-0.069	0.003

Table 2.1: EFFECTS OF THE PREDICTABLE PATTERN OF THE NOISE TRADERS' BEHAVIOR. The first column reports the corresponding AR coefficient of the price series, ρ. The second column reports the first lag serial correlation in returns. The third column reports the P-value of the ARCH test with the H_0: No ARCH effects in returns for the first 10 lags. The fourth column reports the sum of the magnitudes of the autocorrelation function (ACF) coefficients of the squared returns for the first 10 lags. The fifth column reports the correlation between fundamental value and the expectation errors of the partially informed trader. The sixth column reports the correlation between liquidity shocks and expectation errors from the partial informed trader. The seventh column reports the deviation of price from the fundamental value discounted using a risk free rate and corrected for liquidity risk. The fundamental value of the asset is defined as $\frac{\lambda a V_t}{1+r-(1-\lambda)a}$. The equilibrium price without liquidity shocks is used for comparison with the purpose of isolating the effects of β. Formally, $P_t^{comparison} = \rho P_{t-1} - \frac{a\lambda}{\frac{\beta}{\rho\Omega} + a(1-\lambda)} V_t + P_\Delta(\widehat{V}_t^1 - V_t^1)$.

N	ρ	Corr$[r_t, r_{t-1}]$	ARCH	$\sum \lvert ACF(r_t^2) \rvert$	Adjustment Periods
5	0.859	0.692	0	0.797	7
6	0.895	0.490	0	0.855	11
7	0.920	0.301	0	1.282	18
9	0.939	0.208	0	1.578	29
10	0.952	0.154	0	1.848	32
14	0.987	0.113	0	2.978	36
15	0.992	0.109	0	2.156	39

Table 2.2: EFFECTS OF THE INCREASING NUMBER OF HIERARCHICAL INFORMATION LEVELS. $\beta = -20$ for all the cases. The first column reports the corresponding AR coefficient of the price series, ρ. The second column reports the first lag serial correlation in returns. The third column reports the P-value of the ARCH test with the H_0: No ARCH effects in returns for the first 10 lags. The fourth column reports the sum of the magnitudes of the autocorrelation function (ACF) coefficients of the squared returns for the first 10 lags. The fifth column reports the number of time periods for the price to converge back to the stationary equilibrium price level given a positive shock in supply of noise traders.



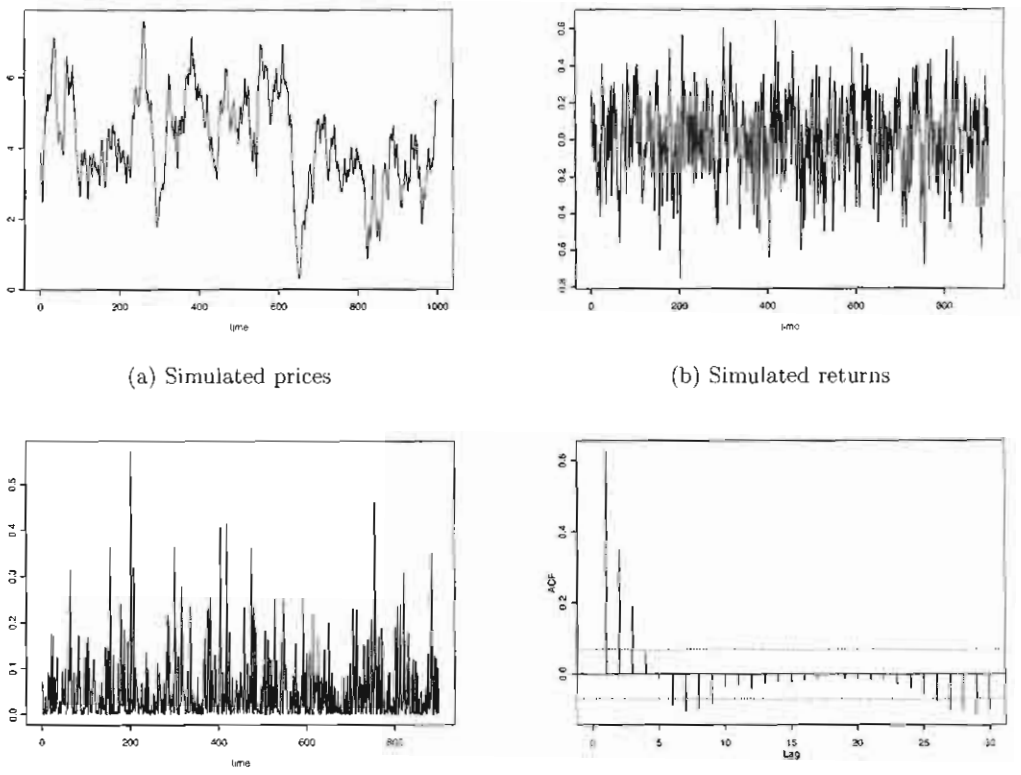
(a) Simulated prices

(b) Simulated returns

(c) Simulated variance of returns

(d) ACF of returns

Figure 2.1: SIMULATION RESULTS IN TWO INFORMATION HIERARCHIES CASE. (a) Time series of simulated prices. (b) Time series of simulated returns. (c) Time series of simulated variance of returns. (d) Average autocorrelations of returns across 100 simulations.



(a) Simulated prices

(b) Simulated returns

(c) Simulated variance of returns

(d) ACF of returns

Figure 2.2: SIMULATION RESULTS IN THREE INFORMATION HIERARCHIES CASE. (a) Time series of simulated prices. (b) Time series of simulated returns. (c) Time series of simulated variance of returns. (d) Average autocorrelations of returns across 100 simulations.

Chapter 3

Jump Detection by Wavelets

In the last two decades, statistics and finance literature have produced substantial empirical evidence that many financial time series contain surprise elements or jumps. It is well understood that compared to continuous price changes, jumps have distinctly different modeling, inference, and testing requirements for the valuation of derivative securities, require a certain distributional framework for inferring extreme risks, and require specialized statistical measures of optimal portfolio allocation. Thus, understanding what drives jumps in underlying securities, how to characterize jumps both theoretically and empirically, and having efficient tests available for jumps that are sufficiently robust to withstand mis-specification and small sample bias is imperative.

This chapter proposes a method based on maximum overlap wavelet transformation (MODWT) to detect and estimate the exact jump location, jump size, and the number of jumps in a given time interval from high frequency data. A useful property of MODWT is that the number of scaling and wavelet coefficients is equal to the number of data points. With this feature of MODWT and a zero phase distortion wavelet function, the location of jumps can be detected precisely from noisy time series processes. In addition to practical jump detection criterion as provided by Fan and Wang (2007), this chapter provides the asymptotic distribution of the test statistic, and the test demonstrates good power and size. I show that the test using MODWT with *Haar* filter is equivalent to Lee and Mykland (2008). Additionally, when a wavelet filter with less leakage, e.g., a least asymmetric wavelet filter, is used, the performance of the test improves.[1] The improvement originates from the

[1] A filter with less leakage is one that is closer to being an ideal band-pass filter.

fact that the wavelet coefficients of a wavelet filter with less leakage contain the weighted average of microstructure noise and continuous changes in price processes and jumps. At locations with no jumps, the weighted average is closer to zero than a simple difference of price process. Thus, identifying spurious jumps becomes easier while preserving power in detecting true jumps.[2]

Substantial research effort has been dedicated to jump detection in asset prices. Among previous studies, Andersen *et al.* (2003) proposed a method using a jump-robust estimator of realized volatility, Barndorff-Nielsen and Shephard (2004, 2006) proposed a bi-power variation (BPV) measure to separate the jump variance and diffusive variance. Lee and Mykland (2008) developed a rolling-based nonparametric test for jumps and estimates of jump size and jump arrival time. Jiang and Oomen (2008) proposed a jump test based on the idea of "variance swap" and explicitly accounted for market microstructure noise. Johannes (2004) and Dungey *et al.* (2007) found significant evidence for jumps in U.S. treasury bond prices and rates. Piazzesi (2003) demonstrated that jump modeling leads to improved bond pricing in the U.S. treasury market. Andersen *et al.* (2007) showed that incorporation of jump components could improve the forecasting of return volatility. As indicated by Fan and Wang (2007), when market returns contain jumps, separating the variation in returns into jump and diffusion components is important for efficient estimation of realized volatility.

Although several tests can be used to identify the existence of jumps, particularly with high frequency data, these tests cannot provide an exact location of jumps even ex post. The estimation of the exact jump location is important for understanding the jump density and its distribution (see Dungey *et al.* (2007) and Jiang *et al.* (2008)). Moreover, the exact detection of the jump location is of great importance for improving derivative hedging. In the presence of jumps, the market is incomplete, and hedging based on the continuous price process thus leads to hedging error. The exact detection of jump location allows the development of hedging strategies and an efficient dynamic rebalancing of hedging portfolios.

The use of wavelets is of importance in jump detection for the following reasons. First, the ability of wavelets to decompose noisy time series data into different time scales is essential for distinguishing jumps from continuous movement in underlying asset and microstructure noise. Intuitively, if there is no noise in the data, a jump can be defined as

[2]Spurious jumps are false jumps due to large movements originates from the presence microstructure noise or large spot volatility.

the change in the mean of the time series. If there is no noise, locating the jump by examining the location where the mean price changes abruptly would be sufficient. Wavelet coefficients contain this information. However, financial and economic time series are far from noise-free and the degree of noise becomes substantial at higher frequencies, i.e., market microstructure noise. Market microstructure noise is due to imperfections of trading processes that comprise a vast array of issues collectively known as market microstructure and include price discreteness, infrequent trading, and bid-ask bounce effects. The wavelet coefficients at jump locations are larger than other wavelet coefficients due to the fact that wavelet coefficients decay at a different rate for continuous processes and jump processes. In a given small time interval, changes in continuous price processes should be very close to zero while on the contrary, jumps are never close to zero. Such information is contained in wavelet coefficients at the jump locations (see Wang (1995) and Fan and Wang (2007)).[3] Second, estimation of jump size is highly correlated with the estimation of integrated volatility. As shown in Fan and Wang (2007), wavelets shows superior estimation of the integrated volatility, which can be used to improve the efficiency of the estimation of jump size.

This chapter implements the wavelet jump test to examine the jump dynamics of three individual equities in U.S. equity markets and finds that the jump dynamics of equities are entirely different across different time scales. This suggests that choosing a proper sampling frequency is important for extracting full jump dynamics. Based on the data set used in the chapter, this suggests that the popular choice of five-minute sampling frequency may neglect a large proportion of jump dynamics embedded in transaction prices. Additionally, although jump arrival densities of positive jumps and negative jumps are symmetric across time scales, the magnitude of jumps is asymmetrically distributed at high frequencies. This suggests that a skewed distribution for the magnitude of jumps should be employed in risk management or asset pricing practices concerning high frequency trading. Finally, only twenty percent of jumps occur in the trading session from 9:30 AM to 4:00 PM, which suggests that the jumps are largely determined by news and not by liquidity shocks. This also suggests that the mutual fund traders or other institutional traders who are prohibited by regulation from intra-day trading should pay closer attention to news outside of trading

[3]Wang (1995) and Fan and Wang (2007) proposed a wavelets-based procedure using the different convergence rates of wavelet coefficients with or without jumps. In addition to jump size, Fan and Wang (2007) could estimate the number of jumps and an estimated interval of jump location, although an exact point estimate of jump location was still absent.

sessions.

Overall, the main contributions of the chapter are as follows. First, a nonparametric jump detection test based on wavelets is proposed and the asymptotic distribution of the test statistic is established. This test is showu to improve jump detection by identifying spurious jumps more accurately. In addition, it provides exact jump location due to the combined effects of MODWT and a zero phase distortion filter. Second, I show that the jump detection test is robust in the presence of microstructure noise. Third, the empirical implementation of the wavelet test in U.S. equity markets demonstrates a dramatic change of jump dynamics across time scales, the asymmetric distribution of jump magnitudes at high frequencies and the occurences of the majority of jumps outside the trading session.

The rest of the chapter is organized as follows. Section II provides the theoretical framework for jump detection in the absence of microstructure noise and introduces the wavelet-based jump test statistic. The asymptotic distribution of the test statistic follows a scaled normal distribution and the scalar is determined by the properties of the wavelet filter. Section III extends the framework to investigate the performance of the test in the presence of the microstructure noise. I show that the asymptotic distribution of the test statistic under the null hypothesis remains the same. Section IV presents the Monte Carlo simulation results, which demonstrate the finite sample behavior of the proposed test. I show that the test has desirable size and power in small samples. Section V discusses the empirical examination of jump dynamics in equity markets. I conclude afterwards.

3.1 No Microstructure Noise

I employ a one-dimensional asset return process. Let the logarithm of the market price of underlying asset be written as $P_t = \log S_t$ where S_t is the asset price at t. For expositional purposes, I restrict myself to finding the jumps in asset prices as follows. When there are no jumps in the market prices, P_t is represented as

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \tag{3.1}$$

where the two terms correspond to the drift and diffusion parts of X_t. In the diffusion term, W_t is a standard Brownian motion, and the diffusion variance σ_t^2 is called spot volatility. Equivalently, P_t can be characterized as

$$dP_t = \mu_t dt + \sigma_t dW_t. \tag{3.2}$$

When there are jumps, P_t is given by

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \sum_{l=1}^{N_t} L_l \tag{3.3}$$

where N_t represents the number of jumps in P_t up to time t and L_l denotes the jump size. Or equivalently, P_t can be modeled as

$$dP_t = \mu_t dt + \sigma_t dW_t + L_t dN_t \tag{3.4}$$

where N_t is a counting process that is left un-modeled. I assume jump sizes L_l are independently and identically distributed. They are also independent of other random components W_t and N_t.

Observations of P_t, the log price, are only available at discrete times $0 = t_0 < t_1 < t_2 < \ldots < t_n = T$. For simplicity, I assume observations are equally spaced: $\Delta t = t_i - t_{i-1}$. Following Lee and Mykland (2008), I impose the following necessary assumptions on price processes throughout this chapter: For any small $\epsilon > 0$,

$$\text{A1.1} \quad \sup_i \sup_{t_i \le u \le t_{i+1}} |\mu_u - \mu_{t_i}| = O_p(\Delta t^{\frac{1}{2}-\epsilon})$$

$$\text{A1.2} \quad \sup_i \sup_{t_i \le u \le t_{i+1}} |\sigma_u - \sigma_{t_i}| = O_p(\Delta t^{\frac{1}{2}-\epsilon})$$

The assumption $A1.1$ and $A1.2$ can be interpreted as the drift and diffusion coefficients not changing dramatically over a short time interval. Formally, this states that the maximum change in mean and spot volatility in a given time interval is bounded above. The assumptions $A1.1$ and $A1.2$ guarantee that the available discrete data are reasonably well-behaved such that the data are a good approximation of the continuous process of the underlying asset. The availability of high frequency financial data allows us to improve the approximation of the continuous underlying asset process using discrete data.

3.1.1 Intuition and Definition of the Test

Wavelets

A wavelet is a small wave that grows and decays in a limited time period.[4] To formalize the notion of a wavelet, let $h = (h_0, h_1, \ldots, h_{L-1})$ be a finite length discrete wavelet (or high

[4]This section follows Gençay *et al.* (2001) closely.

pass) filter coefficient for a wavelet such that it integrates (sums) to zero

$$\sum_{0}^{L-1} h_l = 0$$

and has unit energy

$$\sum_{0}^{L-1} h_l^2 = 1$$

where L is the length of the wavelet filter. Using wavelet filter coefficients, let scaling filter coefficients be

$$g_l = (-1)^{l+1} h_{L-1-l} \text{ for } l = 0, \ldots, L-1.$$

With both wavelet filter coefficients and scaling filter coefficients, I can decompose the data using the (discrete) wavelet transformation (DWT). Formally, let me introduce the DWT throngh a simple matrix operation. Let x to be the dyadic length vector ($N = 2^J$) of observations. The length N vector of discrete wavelet coefficients w is obtained via

$$w = W\mathrm{x}$$

where W is an $N \times N$ orthonormal matrix defining the DWT. The vector of wavelet coefficients can be organized into $J+1$ vectors, $w = [w_1, w_2, \ldots, w_J, v_J]^T$, where w_j is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$, and v_J is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$.

The matrix W is composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. Let

$$h_1 = [h_{1,N-1}, h_{1,N-2}, \ldots, h_{1,1}, h_{1.0}]^T$$

be the vector of zero-padded unit scale wavelet filter coefficients in reverse order, where T is the matrix transpose operation. Thus, the coefficients $h_{1,0}, \ldots, h_{1,L-1}$ are taken from an appropriate ortho-normal wavelet family of length L, and all values $L < t < N$ are defined to be zero. Now circularly shift h_1 by factors of two so that

$$\begin{aligned} h_1^{(2)} &= [h_{1,1}, h_{1,0}, h_{1,N-1}, h_{1,N-2} \ldots, h_{1,3}, h_{1,2}]^T \\ h_1^{(4)} &= [h_{1,3}, h_{1,2}, h_{1,1}, h_{1,0} \ldots, h_{1,5}, h_{1,4}]^T \end{aligned}$$

and so on. Define the $N/2 \times N$ dimensional matrix W_1 to be the collection of $N/2$ circularly shifted versions of h_1. Hence,

$$W_1 = [h_1^{(2)}, h_1^{(4)}, \ldots, h_1^{(N/2-1)}, h_1]$$

Let h_2 be the vector of zero-padded scale 2 wavelet filter coefficients defined similarly to h_1. W_2 is constructed by circularly shifting the vector h_2 by factor of four. Repeat this to construct W_j by circularly shifting the vector h_j (the vector of zero-padded scale j wavelet filter coefficients) by 2^j. The matrix V_J is simply a column vector whose elements are all equal to $1/\sqrt{N}$. Then, the $N \times N$ dimensional matrix W is $W = [W_1, W_2, \ldots, W_J, V_J]^T$.

To complete the construction of the ortho-normal matrix W, we must be able to explicitly compute the wavelet filter coefficients for scales $1, \ldots, J$. Define the wavelet filter h_j, l for scale $\lambda_j = 2^{j-1}$ as

$$H_{j,k} = H_{1,2^{j-1}k \bmod N} \prod_{l=0}^{j-2} G_{1,2^l \bmod N}, \; k = 0, \ldots, N-1.$$

The modulus operator is required in order to account for the boundary of a finite length vector of observations. Thus, we are implicitly assuming that x is periodic. Also, let us define the scaling filter g_J for scale λ_J as

$$G_{j,k} = \prod_{l=0}^{J-1} G_{1,2^l k \bmod N}, \; k = 0, \ldots, N-1.$$

Note that if the data do not represent a dyadic length vector, then we must account for boundary issues. This is one of the reasons why maximum overlap discrete wavelet transformation (MODWT) is attractive when dealing with empirical time series problems.

The following properties are important for distinguishing the MODWT from the DWT:

1. The MODWT can accommodate any sample size N, while DWT restricts the sample size to multiples of 2.

2. The wavelet and scaling coefficients of a MODWT are associated with zero phase filters. Thus, events that feature in the original time series can be properly aligned with features in the MODWT coefficients.

3. The MODWT is invariant to circular shifts in the original time series. This property does not hold for the DWT.

4. The MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT.

For both MODWT and DWT, wavelet coefficients contain information about the high frequency movements in the data series, while scaling coefficients contain information about the low frequency movements. In the next section, I investigate the application of the wavelet method to jump detection for high-frequency time series data.

Intuition of the Test

Wavelets has been shown to be useful for jump detection when the underlying process is a diffusion process as shown in Fan and Wang (2007). As mentioned earlier, Fan and Wang (2007) used a special property of the wavelet expansion, i.e., the localization property, if a function is Holder-continuous with a jump at a point,[5] then the wavelet coefficients of the high pass filter close to that point decay at order $2^{j(\frac{1}{2})}$, where j is a scale for wavelet decomposition. This special feature was used to separate jumps from the continuous parts and microstructure noise (see Wang (1995)).

Although Fan and Wang (2007) showed the effectiveness of the wavelets method for jump detection with a diffusion process convincingly, the statistic of the jump detection test was not formally defined. Thus, the distributional properties of the jump detection procedure were not discussed. Meanwhile, because discrete wavelet transformation (DWT) was applied to the data, the jump location could only be estimated with an informational loss in time domain.

This chapter proposes a framework that treats jump detection using wavelets formally. Before I mathematically define the jump detection statistic J_W, I address the basic intuition behind the proposed detection technique as follows. Imagine that asset prices evolve continuously over time. Suppose that due to an announcement or other informational shock, a jump in (log) prices occurs at time t_i. Given the additive nature of the jump, we expect to see the mean level of the price process to shift. Thus, if we examine high frequency movement, we will find a large movement in prices or a return of large magnitude. To illustrate this, let return be $r_{t_i} = P_{t_i} - P_{t_{i-1}}$. At time t_i, r_{t_i} can be approximated by $(\mu_{t_i} - \mu_{t_{i-1}})\Delta t + (\sigma_{t_i} - \sigma_{t_{i-1}})W_{t_{i-1}}\Delta t + L_{t_i}$. For a small time interval, assumptions A1.1 and

[5]Holder-continuous is an attribute of a function $g : R^d \to R$. g is said to be Holder-continuous if there exist constants C and $0 \leq E \leq 1$ such that for all u and v in R^d:$|g(u) - g(v)| \leq C||u - v||^E$.

A1.2 imply that the first two terms in the return expression should be close to zero while the third term, which is equal to the jump size, should be different from zero. Thus, we should expect to see an absolute large value for the return r_{t_i} at the jump location. Note that the return r_t is the difference of log price P_t. Thus, the information about the return should be stored in the wavelet filter coefficients if we apply the wavelet transformation to P_t. Therefore, we should see a wavelet coefficient of a large magnitude at jump location once the wavelet transformation is applied to the data.

However, due to price discreteness, we can observe a large magnitude of change at the jump location if there is also a large spot volatility. Due to price discreteness, the sampling frequency of the data is bounded below. Increased volatility leads to a large movement in returns. Thus, it is difficult to distinguish whether the observed large movement in prices is due to a jump in price process or a volatility of large magnitude. This is not an issue if the time unit of the observed price process is infinitely small. Thus, this problem should be alleviated if the data is examined at a higher frequency. However, we wish to consider this problem explicitly for the purpose of determining test efficiency. This can be accomplished by normalizing the absolute value of wavelet coefficients via division by the estimated spot volatility. In this chapter, I apply the bi-power variation estimator suggested by Barndorff-Nielsen and Shephard (2004) as a consistent estimate of the spot volatility in the presence of jumps.

Furthermore, I use the maximum overlap discrete wavelet transformation (MODWT) instead of discrete wavelet transformation. The reasons are as follows. First, MODWT generates an equal number of wavelet coefficients (high pass filter) as the original data series. Combined with zero phase correction, the locations of the wavelet coefficients naturally reveal information concerning the original data in the time domain. Thus, the jump location detection is reduced to a jump detection problem. Second, MODWT shows superiority over DWT in decomposing the movements in the data series into high and low frequency. The performance of the test relies on its ability to decompose the data into different scales by wavelet transformation. The MODWT is a suitable choice for the main purpose, which is to detect the exact jump location of the data series in this chapter.

Definition of the Test

Formally, let $P_{j,t}$ be the wavelet coefficients of P_t, where the scale of the wavelet transformation is $j = 1, \ldots, \log_2(T)$. Intuitively, the first scale level of wavelet decomposition

should be of major concerns for jump detection. I define the test statistic as J_W, which tests at time t_i whether a jump occurs for $i = 0, \ldots, T$ as

$$\widehat{J}_W(i) = \frac{P_{1,t_i} - \widehat{m}_i}{\widehat{\sigma}_{t_{i-1}}} \tag{3.5}$$

where $\widehat{\sigma}^2_{t_{i-1}} = \frac{1}{i-3}\sum_{k=2}^{i-1} |P_{1,t_k}||P_{1,t_{k-1}}|$, and $\widehat{m}_i = \frac{1}{i-1}\sum_{k=1}^{i-1} P_{1,t_k}$.

Note that I employ a bi-power variation method to estimate the integrated volatility of the underlying process. There are alternative methods for estimating the integrated volatility; these include two scale realized volatility estimators (TSRV) (Zhang *et al.* (2005) and multi-scale realized volatility estimators (MSRV) (Zhang (2006)). However, Barndorff-Nielsen and Shephard (2004) demonstrated that the presence of jumps in the underlying asset will change the asymptotic behavior of the tests. Additionally, it was shown that a bi-power variation estimator is robust in the presence of the jumps.[6]

3.1.2 Null Distribution: No Jumps

Haar Filter Case

Under the null hypothesis that there is no jump occuring at time t_i, as Δt goes to zero, $\widehat{J}_W(i)$ should go to zero under assumptions $A1.1$ and $A1.2$. Theorem 20 relates the proposed statistic in the case of MODWT with *Haar* filter to the test statistic proposed in Lee and Mykland (2008) (Lee and Mykland test, LM test).

Theorem 20 *Suppose assumptions A1.1 and A1.2 are satisfied. If there is no jump in* (t_{i-1}, t_i), *as* $\Delta t \to 0$

$$sup_i \ \ |\widehat{J}_W(i) - J_W(i)| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \quad \text{where} \quad J_W(i) = \frac{U_i - \overline{U}_{i-1}}{c}. \tag{3.6}$$

Here $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable,* $\overline{U}_{i-1} = \frac{1}{i-1}\sum_{j=1}^{i-1} U_j$, *and* $c = E[|U_i|] = \frac{\sqrt{2}}{\sqrt{\pi}}$ *is a constant, where* W_t *follows a Brownian motion process.*

Proof: Using MODWT with *Haar* filter to transform the data, the wavelet coefficients at scale level 1 are

$$P_{1,t_i} = \frac{1}{2}(P_{t_i} - P_{t_{i-1}}) = \frac{1}{2}(\log S_{t_i} - \log S_{t_{i-1}}). \tag{3.7}$$

[6]Fan and Wang (2007) found that there are, on average, four jumps in foreign exchange market.

Hence, we have

$$\log S_{t_i} - \log S_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \mu_u du + \int_{t_{i-1}}^{t_i} \sigma_u dW(u). \tag{3.8}$$

Imposing assumption A1.1, we have

$$\int_{t_{i-1}}^{t_i} \mu_u du - \mu_{t_{i-1}} \Delta t = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.9}$$

This implies

$$\sup| \int_{t_{i-1}}^{t_i} [\mu_u - \mu_{u-1}] du| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.10}$$

Similarly, we have

$$\sup| \int_{t_{i-1}}^{t_i} [\sigma_u - \sigma_{u-1}] du| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.11}$$

It can be shown that[7]

$$\log S_{t_i} - \log S_{t_{i-1}} - \widehat{m}_i = \sigma_{t_{i-1}} \sqrt{\Delta t}(U_i - \overline{U}_{i-1}) + O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.12}$$

where $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, which is an independently identically distributed (iid) normal and $\overline{U}_{i-1} = \frac{1}{i-1} \sum_{j=0}^{i-1} U_j$.

For the denominator, following Barndorff-Nielsen and Shephard (2004), we have

$$\begin{aligned} plim_{\Delta t \to 0} c^2 \widehat{\sigma}^2_{t_{i-1}} &= plim_{\Delta t \to 0} \frac{1}{(i-3)\Delta t} \sum_{k=2}^{i-1} |\log S_{t_k} - \log S_{t_{k-1}}| |\log S_{t_{k-1}} - \log S_{t_{k-2}}| \\ &= c^2 \sigma^2_{t_{i-1}} \end{aligned} \tag{3.13}$$

where $c = E[|U_i|] = \frac{\sqrt{2}}{\sqrt{\pi}}$. Hence

$$\widehat{J}_W(i) = \frac{U_i - \overline{U}_{i-1}}{c} + O_p \Delta^{\frac{3}{2}-\epsilon}. \tag{3.14}$$

Q.E.D

[7] See Lee and Mykland (2008) for a detailed derivation.

D(4) Filter Case

Theoretically, the *Haar* filter localizes relatively well for jump detection in a continuous price process, but in reality, data are only discretely available. The discretization of data might be accompanied by discretization error, a type of microstructure noise whose effects of microstructure noise on volatility are not negligible. Thus, large returns generated from a discrete data set might be due to noise with large instantaneous volatility. Distinguishing a jump from a large return due to a "volatility effect" is difficult. I propose to the use of an alternative wavelet filter that is closer to being an ideal band-pass filter. Such a filter assigns different weights to the returns in a moving window where the wavelet coefficients capture the weighted average of differenced data (the returns). This enhances the efficiency of distinguishing the jumps from returns due to "volatility effect". However, this simultaneously increases the difficulty of detecting the exact location of the jumps. This trade-off may require an optimal design of wavelet filters that can achieve the overall efficiency in both jump detection and jump location. The optimal design of the wavelet is beyond the scope of this chapter and I only show the distribution of the test statistic with a D(4) filter in this section. A Monte Carlo study is conducted in a later section which demonstrates that nsing a least asymmetric filter rather than a *Haar* filter improves the efficiency of the test.

Under the null hypothesis that no jump occurs at time t_i, as Δt goes to zero, $\widehat{J}_W(i)$ should go to zero under the assumptions of $A1.1$ and $A1.2$. Theorem 21 demonstrates the distribution of the proposed statistic in the case of MODWT with $D(4)$ filter.

Theorem 21 *Assuming zero drift in the underlying price process and supposing assumptions A1.1 and A1.2 are satisfied, if there is no jump in* (t_{i-1}, t_i), *as* $\Delta t \to 0$

$$sup_i \ \ |\widehat{J}_W(i) - J_W(i)| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \quad \text{where} \quad J_W(i) = \frac{U_i}{c}. \tag{3.15}$$

Here $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable, and a constant* $c = \frac{\sqrt{2}}{\sqrt{3}}$, *where* W_t *follows a Brownian motion process.*

Proof: Assuming zero drift, using MODWT with $D(4)$ filter to transform the data, the

wavelet coefficients at scale 1 are

$$\begin{aligned} P_{1,t_i} &= \frac{1}{4\sqrt{2}}((1-\sqrt{3})P_{t_i} + (\sqrt{3}-3)P_{t_{i-1}} + (3+\sqrt{3})P_{t_{i-2}} + (-1-\sqrt{3})P_{t_{i-3}}) \\ &= \frac{1}{4\sqrt{2}}((1-\sqrt{3})(P_{t_i} - P_{t_{i-1}}) + (1+\sqrt{3})(P_{t_{i-2}} - P_{t_{i-3}}) - 2(P_{t_{i-1}} - P_{t_{i-2}})) \\ &= \frac{1}{4\sqrt{2}}((1-\sqrt{3})(\log S_{t_i} - \log S_{t_{i-1}}) + (1+\sqrt{3})(\log S_{t_{i-2}} - \log S_{t_{i-3}}) \\ &\quad -2(\log S_{t_{i-1}} - \log S_{t_{i-2}})). \end{aligned} \tag{3.16}$$

Hence, we have

$$\begin{aligned} \log S_{t_i} - \log S_{t_{i-1}} &= \int_{t_{i-1}}^{t_i} \mu_u du + \int_{t_{i-1}}^{t_i} \sigma_u dW(u) \\ \log S_{t_{i-1}} - \log S_{t_{i-2}} &= \int_{t_{i-2}}^{t_{i-1}} \mu_u du + \int_{t_{i-2}}^{t_{i-1}} \sigma_u dW(u) \\ \log S_{t_{i-2}} - \log S_{t_{i-3}} &= \int_{t_{i-3}}^{t_{i-2}} \mu_u du + \int_{t_{i-3}}^{t_{i-2}} \sigma_u dW(u) \end{aligned} \tag{3.17}$$

Imposing assumption A1.1, we have

$$\int_{t_{i-1}}^{t_i} \mu_u du - \mu_{t_{i-1}} \Delta t = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.18}$$

This implies

$$\sup| \int_{t_{i-1}}^{t_i} [\mu_u - \mu_{u-1}] du| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.19}$$

Similarly, we have

$$\sup| \int_{t_{i-1}}^{t_i} [\sigma_u - \sigma_{u-1}] du| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.20}$$

It can be shown that

$$\begin{aligned} \log S_{t_i} - \log S_{t_{i-1}} &= \sigma_{t_{i-1}} \sqrt{\Delta t}(U_i) + O_p(\Delta t^{\frac{3}{2}-\epsilon}) \\ \log S_{t_{i-1}} - \log S_{t_{i-2}} &= \sigma_{t_{i-2}} \sqrt{\Delta t}(U_{i-1}) + O_p(\Delta t^{\frac{3}{2}-\epsilon}) \\ \log S_{t_{i-2}} - \log S_{t_{i-3}} &= \sigma_{t_{i-3}} \sqrt{\Delta t}(U_{i-2}) + O_p(\Delta t^{\frac{3}{2}-\epsilon}) \end{aligned} \tag{3.21}$$

where $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, which is an independently identically distributed (iid) normal. Hence, P_{1,t_i} is a linear combination of iid normal variables U_is. Therefore, P_{1,t_i} is normally distributed with mean 0 and variance

$$\sigma_{P,t_i} = \frac{1}{4\sqrt{2}} \sqrt{(4-2\sqrt{3})\sigma^2_{t_{i-1}} + (4+2\sqrt{3})\sigma^2_{t_{i-3}} + 4\sigma^2_{t_{i-2}}}.$$

For the denominator, in order to apply the resnlts from Barndorff-Nielsen and Shephard (2004), we redefine P_{1,t_i} to be a new diffnsion process

$$P_{1,t_i} = \sigma_{P,t_i} dW_X \tag{3.22}$$

σ_{P,t_i} is continnous, so we can apply the result from Barndorff-Nielsen and Shephard (2004). Hence:

$$\sigma^2_{P,t_{i-1}} = \sum_{k=2}^{i-1} |P_{1,t_i}||P_{1,t_{i-1}}| \tag{3.23}$$

$$\begin{aligned} plim_{\Delta t \to 0} c^2 \widehat{\sigma}^2_{X,t_{i-1}} &= plim_{\Delta t \to 0} \frac{1}{(i-3)\Delta t} \sum_{k=2}^{i-1} |P_{1,t_i}||P_{1.t_{i-1}}| \\ &= c^2 \sigma^2_{P,t_{i-1}} \end{aligned} \tag{3.24}$$

where $c = \frac{\sqrt{2}}{\sqrt{3}}$. Hence

$$J_W(i) = \frac{U_i}{c} + O_p \Delta^{\frac{3}{2}-\epsilon}. \tag{3.25}$$

This can be extended to allow for nonzero drift case.

Q.E.D

Therefore, the distribution of the test statistic is unchanged under a $D(4)$ filter.

3.2 With Microstructure Noise

Due to market microstructure, high frequency data are noisy. A common modeling approach is to treat microstructure noise as ordinary "observational error", and then assume that the observed high-frequency data P_t^* are equal to the latent, true log-price process P_t of a security plus market microstructnre noise ϵ_t, thus:

$$P_t^* = P_t + \epsilon_t \tag{3.26}$$

where P_t^* is the logarithm of the observable transaction price of the security observed at time t and increment in ϵ_t is mean 0 *iid* noise with variance η^2 and independent of P_t. When there are no jumps in the market price, P_t^* is represented as

$$P_t^* = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_{1,s} + \epsilon_t \tag{3.27}$$

where the three terms correspond to the drift and diffusion parts of P_t and the *iid* market microstructure noise. In the diffusion term, W_t is a standard Brownian motion, and the diffusion variance σ_t^2 is called spot volatility. Equivalently, P_t^* can be characterized as

$$dP_t^* = \mu_t dt + \sigma_t dW_{1,t} + \eta dW_{2,t}. \tag{3.28}$$

When there are jumps, P_t^* is given by

$$P_t^* = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_{1,s} + \epsilon_t + \sum_{l=1}^{N_t} L_l \tag{3.29}$$

where N_t represents the number of jumps in P_t^* up to time t and L_l denotes the jump size. Equivalently, P_t^* can be modeled as

$$dP_t^* = \mu_t dt + \sigma_t dW_{1,t} + \eta dW_{2,t} + L_t dN_t \tag{3.30}$$

where N_t is a counting process that is left un-modeled. I assume jump sizes L_l are independently and identically distributed and also independent of other random components W_t and N_t. Additionally, microstructure noise is perfectly correlated with price.

3.2.1 Under the Null: No Jumps

Theorem 22 characterizes the asymptotic behavior of the proposed test statistic when market microstructure noise is presented.

Theorem 22 *Assuming zero drift in the underlying process and supposing assumptions A1.1 and A1.2 are satisfied, if there is no jump in* (t_{i-1}, t_i), *as* $\Delta t \to 0$

$$sup_i \ \ |\widehat{J}_W(i) - J_W(i)| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \quad \text{where} \quad J_W(i) = \frac{U_i}{c}. \tag{3.31}$$

Here $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable and a constant* $c = E[|U_i|] = \frac{\sqrt{2}}{\sqrt{\pi}}$.

Proof: Let us investigate the denominator and numerator of $J_W(i)$. As a simple demonstration, let us assume a zero drift case. For the numerator: using MODWT with a *Haar* filter to transform the data, the wavelet coefficients at scale level 1 are

$$P_{1,t}^* = \frac{1}{2}(P_t^* - P_{t-1}^*) = \frac{1}{2}(\log S_t - \log S_{t-1} + \epsilon_t - \epsilon_{t-1}). \tag{3.32}$$

Recall that

$$(\log S_{t_i} - \log S_{t_{i-1}}) = \sigma_{t_{i-1}}\sqrt{\Delta t}U_i + O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.33}$$

and

$$\epsilon_{t_i} - \epsilon_{t_{i-1}} = \eta\sqrt{\Delta t}U_i + O_P(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.34}$$

where $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, which is an independently and identically distributed normal. Therefore,

$$2P^*_{1,t} = (\eta + \sigma_{t_{i-1}})\sqrt{\Delta t}U_i + O_p(\Delta t^{\frac{3}{2}-\epsilon}) \tag{3.35}$$

For the denominator, following Barndorff-Nielsen and Shephard (2004), we have

$$\begin{aligned}
& \frac{1}{i-3}\sum_{k=2}^{i-1}|P^*_{1,t_k}||P^*_{1,t_{k-1}}| \\
= \; & \frac{1}{i-3}\sum_{k=2}^{i-1}|(\log S_{t_k} - \log S_{t_{k-1}}) + (\epsilon_{t_k} - \epsilon_{t_{k-1}})||(\log S_{t_{k-1}} - \log S_{t_{k-2}}) + (\epsilon_{t_{k-1}} - \epsilon_{t_{k-2}})| \\
= \; & \frac{1}{i-3}\sum_{k=2}^{i-1}|\log S_{t_k} - \log S_{t_{k-1}}||\log S_{t_{k-1}} - \log S_{t_{k-2}}| \\
& + \frac{2}{i-3}\sum_{k=2}^{i-1}|\log S_{t_k} - \log S_{t_{k-1}}||(\epsilon_{t_{i-1}} - \epsilon_{t_{k-2}})| \\
& + \frac{1}{i-3}\sum_{k=2}^{i-1}|(\epsilon_{t_k} - \epsilon_{t_{k-1}})||(\epsilon_{t_{k-1}} - \epsilon_{t_{k-2}})|
\end{aligned} \tag{3.36}$$

Recall that

$$plim_{\Delta t\to 0}\frac{1}{i-3}\sum_{k=2}^{i-1}|\log S_{t_k} - \log S_{t_{k-1}}||\log S_{t_{k-1}} - \log S_{t_{k-2}}| \;=\; c^2\sigma^2_{t_{i-1}} \tag{3.37}$$

where $c = E[|U_i|] = \frac{\sqrt{2}}{\sqrt{\pi}}$. The second term in Equation 3.36 behaves as in

$$\begin{aligned}
& plim_{\Delta t\to 0}\frac{2}{i-3}\sum_{k=2}^{i-1}|\log S_{t_k} - \log S_{t_{k-1}}||(\epsilon_{t_{i-1}} - \epsilon_{t_{k-2}})| \\
= \; & plim \;\; 2E[|\log S_k - \log S_{k-1}|]E[|\epsilon_{l-1} - \epsilon_{l-2}|] = 2c^2\sigma_{t_{i-1}}\eta
\end{aligned} \tag{3.38}$$

The third term in Equation 3.36 behaves as in

$$plim_{\Delta t\to 0}\frac{1}{i-3}\sum_{k=2}^{i-1}|(\epsilon_{t_k} - \epsilon_{t_{k-1}})||(\epsilon_{t_{k-1}} - \epsilon_{t_{k-2}})| \;= plim \;\; c^2\eta^2 \tag{3.39}$$

Hence

$$2\frac{1}{i-3}\sum_{k=2}^{i-1}|P^*_{1,t_k}||P^*_{1,t_{k-1}}| \ \ plim \ \ (\sigma_{t_{i-1}}+\eta)^2c^2 \tag{3.40}$$

$$\widehat{J}_W(i) = \frac{U_i}{c} + O_p\Delta^{\frac{3}{2}-\epsilon}. \tag{3.41}$$

Q.E.D

Theorem 22 indicates that the statistic $\widehat{J}_W(i)$ is robust in the presence of microstructure noise. Intuitively, in the presence of microstructure noise, underlying process as $dP^*_t = dP^*_t = \mu_t dt + \sigma^*_t dW_t$ can be rewritten as long as the noise increment is iid, where σ^*_t is a function of spot volatility σ_t and the volatility of microstructure noise η. Thus, it is equivalent to have a larger spot of volatility. If we want to estimate the integrated volatility, distinguishing the spot volatility of the underlying process (σ_t) from the microstructure noise (η) is a challenging task when σ^*_t is the only information available (See Zhang *et al.* (2005) and Zhang (2006)). However, the numerator of the test statistic is also composed of microstructure noise. Hence, we only need to estimate the "new spot volatility" of the underlying process, σ^*_t. Note that the new underlying process is well-defined so that the bi-power variation estimator is a consistent estimator of σ^*_t (or $\sigma_t + \eta$). Therefore, the asymptotic distribution of the test is not changed in the presence of microstructure noise.

Issues related to microstructure noise are important in finance literature, especially for estimation of integrated volatility. Microstructure noise might be due to the imperfections in trading processes, including price discreteness, infrequent trading, and bid-ask bounce effects. It is well-known that higher price sampling frequencies are linked to a larger the impact of microstructure noise. Zhang *et al.* (2005) demonstrated that estimation of integrated volatility via a realized volatility method is severely contaminated by microstructure noise. Fan and Wang (2007) assumes a very small noise ratio in detecting jumps. The distribution of the test proposed in Ait-Sahalia and Jacod (2009) is different in the presence of microstructnre noise. Lee and Mykland (2008) chose a rather low sampling frequency (fifteen minutes) in an empirical study in order to avoid the impact of microstructure noise. Because the asymptotic distribution of this statistic is robust in the presence of microstructure noise, I do not need to decrease the sampling frequency. Thus, I claim that this technique of jump detection is more efficient.

3.3 Monte Carlo Simulations

In this section, I examine the effectiveness of the wavelet-based jump test using Monte Carlo simulations. The performance of the test statistic is examined at different sampling frequencies. An Euler method is used to generate continuous diffusion processes and the burn-in period observations is discarded to avoid the effects of the initial value.

3.3.1 Under the Null

This subsection illustrates the simulated test statistic under the null hypothesis of no jump in a given period of time. The asymptotic distribution of the statistic is a scaled standard normal and the scaling factor is $E[U] = \frac{\sqrt{2}}{\sqrt{\pi}}$. Formally, I consider

$$dP_t = \mu_t dt + \sigma_t dW_t. \tag{3.42}$$

where μ_t is the drift in price process and σ_t is the diffusion or spot volatility in the price process. I consider four scenarios: $\mu_t = 0$ and $\sigma_t = \sigma$ (zero drift and constant volatility as the benchmark), $\mu_t \neq 0$ and $\sigma_t = \sigma$ (non-zero drift and constant volatility), $\mu_t = 0$ and $\sigma_t \neq \sigma$ (zero drift and stochastic volatility), and $\mu_t \neq 0$ and $\sigma_t \neq \sigma$ (non-zero drift and stochastic volatility). Specifically, I assume $\mu_t = 1$ or $\mu_t = 0$ for non-zero and zero drift cases, respectively.[8] I employ an Ornstein-Uhlenbeck process as the volatility model for the stochastic volatility case:

$$\begin{aligned} dP_t &= \mu_t dt + \sigma_t dW_{1,t} \\ d\log\sigma_t^2 &= k(\log\overline{\sigma}^2 - \log\sigma_t^2) + \delta dW_{2.t} \end{aligned} \tag{3.43}$$

where k measures the recovery rate of volatility to the mean and $\log\overline{\sigma}^2$ can be interpreted as the long run mean of volatility, δ is the diffusion parameter for volatility process.[9] Following Fan and Wang (2007), I assume that the correlation between $W_{1,t}$ and $W_{2,t}$ is ρ, which is negative and captures the asymmetric impact of the innovation in price process. In this section, following Fan and Wang (2007), I assume $k = -0.1$, $\log\overline{\sigma}^2 = -6.802$, $\delta = 0.25$.

[8] I also investigated various specifications for drift part as well, for instance the mean reverting process as Ornstein-Uhlenbeck process. I found that the specification of drift part has negligible impact on the main result.

[9] I also investigated other specifications of the process and found that different specifications of volatility processes did not qualitatively changed the main results.

Figure 3.1 shows the density plot of the statistic for 1 million observations when a *Haar* filter is used. The top left panel shows the zero mean and constant volatility case; top right panel depicts the non-zero mean and constant volatility case; bottom left panel shows the zero mean and stochastic volatility case; and bottom right panel depicts the non-zero mean and stochastic volatility case.

Figure 3.2 shows the density plot of the statistic for one million observations when a $S8$ filter is used. The top left panel shows the zero mean and constant volatility case; top right panel depicts the non-zero mean and constant volatility case; bottom left panel shows the zero mean and stochastic volatility case; and bottom right panel depicts the non-zero mean and stochastic volatility case. The reason for choosing $S8$ rather $D4$ filter is that the $S8$ belongs to the least asymmetric filter class which supposedly permits nearly zero phase distortion; this is helpful in conveying information about the jump location in the time domain. Simulations show that jump detection using a $D4$ filter is similar to that using a $S8$ filter.

Additionally, Figure 3.1 and Figure 3.2 show that the test statistic follows a standard normal distribution when volatility is constant. When volatility is stochastic, the test statistic has fat tails. I found that when the frequency increases, i.e., $\Delta t \to 0$, the fat tail of the test statistic diminishes. The fat-tailness of the test statistics originates from two sources. First, the estimation of spot volatility using bi-power variation is unfavorable if there are insufficient data close to the boundary of the wavelet transformation. Second, the *Haar* filter has excessive leakage in decomposing the data into noise component and continuous components. It is notable that $S8$ filter outperforms the *Haar* filter while the distribution of the test statistics under the null hypothesis using the $S8$ filter is closer to a normal distribution even in stochastic volatility cases.[10]

[10]Simulations show that the fat-tailness of null distribution originates from two sources: the poor performance of bi-power variation estimator for small samples and the choice of wavelet filter. Poor performance of the bi-power variation estimator for small samples occurs when there are only a few data points available for estimation, e.g., when the filter location is close to the boundary of the data set. Additionally, simulations demonstrate that the wavelet filter with less leakage can decrease the fat-tailness of the null empirical distribution. This is mainly due to the fact that wavelet filters with less leakage will more sophisticatedly utilize information both at the estimated location and in the neighborhood of the estimated location. This mitigates the small sample estimation problem brought about by bi-power variation.

3.3.2 Size and Power

First, I present the size of the test statistic in Table 3.1.[11] As shown in the table, the test statistic has a good size. The probability of false jump detection is quite close to the theoretical value, i.e., 1% when the 99% quantile from the simulated distribution of the test statistic is used and 5% when the 95% quantile is used.

Subsequently, I study the power of the test statistic. For the purposes of illustration, I allow only one jump per simulation. In particular, I study a sample of 1,024 observations and a jump occurring at the $819th$ observation. The jump size could be large, e.g., 3σ (three standard deviation of return volatility) or could small, e.g, 0.1σ (10% standard deviation of return volatility).[12] I also examine the performance of the test statistic at the different time scales. In the simulations, the number of repetitions is 1,000.

I employ two measures of performance to characterize the test. One is the power of the test, which is the probability of detecting the actual jump at the time i when the jump occurs. Note that when I apply the test to the simulated data, I let it detect the location of jump itself without using the information about the actual jump location. The power of test is calculated by the number of cases where the test detects a jump at $819th$ observation divided by the number of repetitions. I also consider another measure, a success rate measure, which is consistent with the probability of spurious detection of jumps (GSD) as in Lee and Mykland (2008). Specifically, if the test detects the true jump without detecting any other spurious jumps, I call it a success. Recall that there is only one jump in the true data, i.e., the test should detect one and only one jump at the $819th$ observation. Therefore, the second measure is the success rate, i.e., the number of successes divided by the number of repetitions.

When the test statistic exceeds a threshold, the null hypothesis of no jumps is rejected. There are two threshold levels used in this chapter: 95% and 99% quantiles of the null distribution. I also investigate the performance of the test statistic with different time scales; Thus, I let the time step used to generate the continuous process of the underlying asset be 252 times the number of observations per day. For instance, I choose the time step between observations to be $252 * 1$ for daily observations. In the simulations, I choose the

[11] I use 95% and 99% critical values from the standard normal distribution multiplied by a scalar $c = \frac{\sqrt{\pi}}{\sqrt{2}}$.

[12] Fan and Wang (2007) found that the ratio of jump variation to the total return variation is 1 to 1.5 in the foreign exchange market.

observations per day to be 24 or 12. Thus, the sample size of 1024 corresponds to two weeks of hourly data or one month of bi-hourly data. The findings suggest that the power of the test is improved when sampling frequency is higher. This is consistent with the findings of Lee and Mykland (2008).

Table 3.2 shows the power comparisons for the $S8$ filter case, $Haar$ filter case (LM), linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and difference test of Jiang and Oomen (2008) (JO). The wavelet-based test has good power and outperforms other tests at both the 95% and 99% quantiles. For a wide range of jump sizes (0.5σ to 3σ), both $S8$ and $Haar$ filters can detect the actual jump without failure. For jump size 0.25σ when the $S8$ filter is used, the test can detect the actual jump with 83% probability if the 95% quantile is used and 62% probability if 99% quantile is used. Similarly, for the 0.1σ case, the $S8$ filter detects jumps with 21% and 8% probability, respectively. The $Haar$ filter detects the actual jumps with a probability 97% and 91% for 95% and 99% quantiles, respectively, for jump size 0.25σ. For jump size 0.1σ, the $Haar$ filter detects the actual jump with the probabilities 34% and 16% accordingly. Recall that wavelet coefficients of the $Haar$ filter are simple differences of prices, whereas wavelet coefficients of $S8$ filter are weighted averages of returns in the neighborhood of a given location. Thus, it is easier for a $Haar$ filter to detect large movement in returns and, not surprisingly, the $Haar$ filter marginally outperforms the $S8$ filter when the magnitude of jump is very small. Both the $Haar$ filter and the $S8$ filter outperform BNS and JO tests using the 95% or 99% quantile. This suggests that wavelet-based methods can be more effective for detecting large movements in the presence of jumps. Note that a large variation in microstructure noise is equivalent to decrease the relative magnitude of jump size to the spot volatility of underlying process. From the simulations, wavelet-based methods indeed demonstrate superior performance for detecting jumps even when the jump size relative to spot volatility is small. This is important for detection of jumps in high frequency financial time series data, where microstructure noise inevitably exists.

As a performance measure, power only considers the ability of a test to detect the true jumps while ignoring the possibility of spurious jumps detection. Intuitively, if we set a threshold low enough to reject the null hypothesis at every point of time, the actual jumps can be detected without failure if they exists. However, too many spurious jumps are detected. Thus success rate is employed as another measure. Recall that success rate measures the ability of a test to detect the actual jumps only. That is, if and only if

the test only detects the true jumps, it is regarded as successful. This measure is more appropriate for describing the performance of jump detection tests. Table 3.3 presents success rate comparisons for the $S8$ filter, $Haar$ filter (LM), linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and difference test of Jiang and Oomen (2008) (JO). When the 95% quantile is used as threshold level, the success rate is zeros for both the $S8$ and $Haar$ filter. In contrast, BNS provides a 15% to 25% success rate. The JO test provides a 5% success rate. Combined with the power results, the zero success rate suggests that both the $Haar$ and $S8$ filters always detect at least one spurious jump in addition to the actual jump. The BNS and JO tests show a superior performance for not detecting spurious jumps compared to $Haar$ filter and $S8$ filter using 95% quantile as the threshold level. However, when 99% quantile is used as threshold level, the $S8$ filter outperforms BNS, LM, and JO tests at all jump size levels. This shows that when a more stricter threshold level is imposed, i.e., a 99% quantile rather a 95% quantile, the $S8$ filter outperforms the BNS and JO tests in both power and success rate.

Additionally, the $S8$ filter is recommended rather than the $Haar$ filter for the following reasons. First, when a 99% quantile is used as a threshold, the success rate of $S8$ filter is satisfactory compared to the $Haar$ filter and other alternative candidate tests. Second, the $S8$ filter can provide comparable power to the $Haar$ filter when the magnitude of jump is in a range from 0.25σ to 3σ. Thus, when a filter with less leakage is used, the test performance improves.

Overall, the proposed test demonstrates satisfactory power for detecting jumps. Based on simulations, I recommend the $S8$ filter or the other filters with less leakage and the use of a stringent threshold level (at least 99% quantile) for empirical implementation of jump detection. The filter with less leakage possesses similar power to the $Haar$ filter (Lee and Mykland test), but the success rate is higher. This improvement occurs because the wavelet filter with less leakage averages the noise embedded in the underlying process to facilitate the to differentiation of the jump from the noise (volatility effect).

3.4 Empirical Analysis for U.S. Equity Markets

In this section, I apply the jump detection test to three major U.S. individual equities transaction prices to determine their jump dynamics. To characterize jump dynamics, the frequency of arrivals and the magnitude of jumps need to be characterized. Using a jump

detection method, the jump locations can be estimated in a given time interval.[13] Thus, the arrival density of jumps per trading day can be subsequently calculated. To characterize the magnitude of the jumps, I employ a method that follows Fan and Wang (2007). Formally, for each estimated jump location t_i, I choose a small neighborhood, $t_i \mp \gamma$, where small $\gamma > 0$. Then, I calculate the average of the prices over $[t_i - \gamma, t_i]$ and $[t_i, t_i + \gamma]$. Let P_{t_i-} and P_{t_i-} to denote the averages accordingly. The jump size is estimated by $L_{t_i} = P_{t_i+} - P_{t_i-}$. Fan and Wang (2007) shows that this estimator of jump size is consistent when the neighborhood γ is chosen such that $\gamma \sim T^{-1/2}$, where T is the sample size.

3.4.1 Multi-Scale Jump Dynamics

I use ultra-high frequency tick data from transactions on the New York Stock Exchange (NYSE) collected from the Trade and Quote (TAQ) database. I apply the jump detection test to the log transaction prices. The time span is three months from January 1st 2008 to March 31st 2008, which represents latest data available and which has never been investigated in the literature. I choose three equities (Wal-Mart (WMT), IBM (IBM) and General Electric (GE)) to compare their jump dynamics. Recall that the test is robust in the presence of microstructure noise, I employ one minute data to improve the efficiency of estimation.[14]

For the comparison purposes, I also report jump dynamics at different time scales: specifically, jump dynamics are reported using one-minute data, five-minute data and fifteen-minute data.[15] This comparison offers useful insights for both the theoretical framework for jump detection and the implementation of empirical jump detection. Due to the robustness of this test in the presence of microstructure noise, the difference in jump dynamics detected at different time scales should account for the difference in jump arrival densities at different time scales. Intuitively, if two jumps of similar magnitudes occur in opposite directions, i.e., a positive jump and negative jump, a large difference in price level should not be observed; thus we will claim that there were no jumps in this time interval. Specifically, low frequency data might ignore jump dynamics that could not be captured due to the sampling

[13] In this chapter, the time interval is chosen to be one day.

[14] Tick level transaction data are irregularly spaced time series data and need to be converted into regularly spaced time series data for the implementation of the test. See Dacorogna *et al.* (2001) for a discussion about transforming irregularly spaced time series data into regularly spaced time series data.

[15] Lee and Mykland (2008) used fifteen-minute transaction data.

frequency. In order to examine the jump dynamics at a pre-chosen sampling frequency, we implicitly assume that the smallest duration between two jumps should be larger than the time interval implied by the sampling frequency. If this is true, significant differences should not be observed in jump dynamics across different time scales. Otherwise, a higher sampling frequency should be used to extract full jump dynamics.

Table 3.4 and Figure 3.3, Figure 3.6 and Figure 3.9 report the dynamics of jump arrivals for these three equities. Jump dynamics for the same equity are entirely different at different time scales; specifically, the jump arrival densities for GE are 11.7, 3.7, and 1.5 jumps per day at one-minute, five-minutes, and fifteen-minutes sampling frequency, respectively. Similar patterns are found for IBM and WMT. The jump arrival densities for IBM are 11.3, 3.6, and 1.2 jumps per day at one-minute, five-minutes, fifteen-minutes sampling frequency, respectively. The jump arrival densities for WMT are 11, 3.6, and 1.3 at one-minute, five-minutes, fifteen-minutes, respectively. Thus, the jump arrival densities at one-minute sampling frequency per day for all three equities are higher than jump arrival densities at five-minutes and fifteen-minutes. As stated earlier, if the sampling frequency is high enough to observe full jump dynamics, significant differences should not be observed in jump arrival densities across time scales. Therefore, sampling data at fifteen-minutes ignores a significant portion of jump dynamics that occurs at higher frequencies than the sampling frequency. Thus, sampling data at fifteen-minutes even five-minutes might not be appropriate for the purpose of risk management or dynamic hedging which requires continuous adjustments of positions. If the jump arrival density is estimated at an incorrect frequency, the impact of jumps will be underestimated. Another suggestion is that the averaging of jumps with opposite directions should be considered for dynamic hedging. It should also be noted that the jump arrival densities are similar for all three equities at the same scale. This suggests that macroeconomic news might play an important role in the formation of jumps that accounts for "common trend".

3.4.2 Positive Jumps Versus Negative Jumps

I further investigated the jump dynamics for jumps of different directions, i.e., positive jumps and negative jumps. Table 3.4, Figure 3.4, Figure 3.7, and Figure 3.10 report the arrival densities for jumps of different directions. Jump arrivals are symmetric for all equities. Thus, the arrival densities of positive jumps and negative jumps are similar across different time scales. For instance, the arrival densities of positive and negative jumps for GE at

one-minute sampling frequency are 6.1 and 5.6 per day, respectively. The arrival densities of positive and negative jumps for GE at five-minutes are 1.8 and 1.9 per day, respectively. The arrival densities of positive and negative jumps for GE at fifteen-minutes are 0.75 and 0.75 per day, respectively. This pattern also holds for IBM and WMT.

Furthermore, the magnitudes of jumps are asymmetric at high frequencies for all equities. At the one-minute sampling frequency, the mean size of negative jumps for GE is -0.04% of the return while the mean size of the positive jumps is 0.01% of the return. In the case of IBM, the mean size of jumps is -0.02% and 0.10% respectively. This suggests that the magnitudes of jumps are not symmetric at high sampling frequencies. Symmetric distribution of jump sizes for derivative security pricing should not be assumed in practice. This asymmetry in jump magnitude is decreased at lower sampling frequencies.

3.4.3 Trading Session Versus Off-Trading Session

I also decompose the jump dynamics of all these equities into two sessions: the day trading session (9:30 AM to 4:00 PM for NYSE) and the off-trading session. Notice that for hedging purposes, the jumps in the day trading session are more relevant because dynamic hedging needs to continuously account for the impact of jumps on the prices of underlying assets. Mutual fund traders or other institutional investors cannot engage in intra-day trading due to regulatory constraints. Thus, the jumps in the off-trading sessions might be more relevant. Table 3.4, Figure 3.5, Figure 3.8, and Figure 3.11 report the jump dynamics in the day trading session. It demonstrates that the majority of jumps occurs in off-trading sessions. Only twenty percent of jumps occurs in the day trading session. The average number of jumps occur in day trading time session at one-minute frequency per day is 2 to 3 which is comparable to the findings of Fan and Wang (2007). This suggests that mutual fund or other institutional traders should pay closer attention to the macroeconomic news or other prescheduled news that contribute to the majority of jump arrivals. This also suggests that jumps are mostly determined by informational factors other than liquidity shocks brought about by noise traders.

Additionally, the average sizes of positive and negative jumps in the trading session are different across time scales. It seems that at high frequencies, the magnitudes of the jumps of different directions are quite asymmetric. This asymmetry diminishes as the sampling frequency decreases. This suggests that normality assumption of jump sizes might not be a good approximation for high frequency data; a skewed distribution might be required for

modeling of jump size when conducting high frequency trading.

3.5 Conclusions

This chapter introduces a new nonparametric test based on the wavelets method to detect jump arrival times in high frequency financial time series data. This newly proposed test is motivated by the ability of the wavelets method to decompose the data into different time scales. This localization property of wavelets is shown to be superior for jump detection. I show that the distribution of the test under the null hypothesis of no jumps is asymptotically normal. I demonstrate that the test is robust for different price processes and the presence of market microstructure noise. A Monte Carlo simulation is conducted to demonstrate the test has good power and size. I also demonstrate that the use of wavelet filters with less leakage improves the success rate of the test, i.e., the ability of the test to only detect the true jumps.

An empirical implementation is then conducted for U.S. equity markets, and jump dynamics changes dramatically across time scales. This suggests that choosing a proper sampling frequency is very important for investigating the full jump dynamics. Additionally, the arrival densities of positive jumps and negative jumps are similar, but the magnitudes of the jumps are asymmetrically distributed (at high frequencies). Finally, the majority of jumps occur outside of the day trading session and only twenty percent of jumps occur in the day trading session.

One potentially interesting avenue for future research is to relate the jumps to macroeconomic news and liquidity shocks. Recall that jumps can be irregular (due to macroeconomic news) or regular (due to liquidity shocks brought about by noise traders), but this distinction has not received much attention in the literature. Intuitively, irregular jumps cannot be assumed to occur regularly and should be modeled separately. Thus, it is important to separate the detected jumps into two categories: irregular and regular jumps. Doing so helps to conditionally predict jumps and improve asset pricing and hedging activities in practice.[16]

[16] I am currently working on this.

Haar filter	*S*8 filter	*Haar* filter	*S*8 filter
95% confidence level		99% confidence level	
	$\mu = 0$ and $\sigma_t = \sigma$		
0.0510	0.0490	0.0110	0.0100
	$\mu \neq 0$ and $\sigma_t = \sigma$		
0.0490	0.0500	0.0110	0.0100
	$\mu = 0$ and $\sigma_t \neq \sigma$		
0.0490	0.0490	0.0100	0.0100
	$\mu \neq 0$ and $\sigma_t \neq \sigma$		
0.0510	0.0500	0.0100	0.0100

Table 3.1: SIZE OF TEST STATISTIC AT 95% AND 99% QUANTILES. Size is defined as the detection of spurious jumps under the null hypothesis of no jumps. The number of Monte Carlo simulations is 1000. Four scenarios are investigated; specifically, the case of zero mean and constant variance, the case of non-zero mean and constant variance, the case of zero mean and non-constant variance. and the case of non-zero mean and non-constant variance.

Jump size	$S8$ filter	*Haar* filter (LM)	Linear test (BNS)	Difference Test (JO)
		Frequency = 2 *hours*, 95% confidence level		
3.00σ	1.0000	1.0000	0.8550	0.8430
1.00σ	1.0000	1.0000	0.8110	0.7690
0.50σ	1.0000	1.0000	0.7110	0.6470
0.25σ	0.8500	0.9720	0.5330	0.4400
0.10σ	0.2160	0.3360	0.1110	0.0990
		Frequency = 1 *hour*, 95% confidence level		
3.00σ	1.0000	1.0000	0.8740	0.8610
1.00σ	1.0000	1.0000	0.8460	0.8210
0.50σ	1.0000	1.0000	0.8030	0.7550
0.25σ	0.9300	0.9760	0.6380	0.5540
0.10σ	0.2180	0.3800	0.1090	0.0980
		Frequency = 2 *hours*, 99% confidence level		
3.00σ	1.0000	1.0000	0.8550	0.8430
1.00σ	1.0000	1.0000	0.8110	0.7690
0.50σ	1.0000	1.0000	0.7110	0.6470
0.25σ	0.6260	0.9080	0.5330	0.4400
0.10σ	0.0820	0.1380	0.1110	0.0990
		Frequency = 1 *hour*, 99% confidence level		
3.00σ	1.0000	1.0000	0.8740	0.8610
1.00σ	1.0000	1.0000	0.8460	0.8210
0.50σ	1.0000	1.0000	0.8030	0.7550
0.25σ	0.7000	0.9120	0.6380	0.5540
0.10σ	0.1020	0.1860	0.1090	0.0980

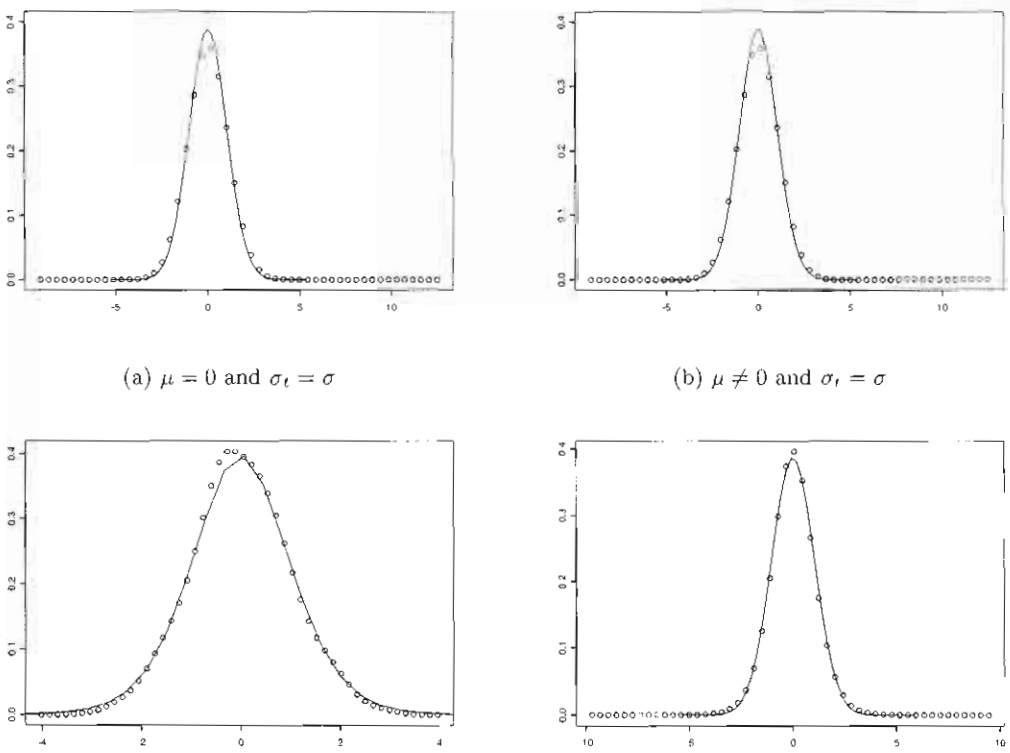
Table 3.2: POWER COMPARISON WITH OTHER JUMP TESTS. The simulation only allows one jump and assnmes constant volatility and non-zero drift part in price process. The number of repetitions is 1,000. The power of the test is defined as the probability that the test will detect the actual jump (even when the test also detects spurious jumps). The table shows the power of my test ($S8$ filter, *Haar* filter corresponding to Lee and Mykland (2008) (LM)), linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and difference test of Jiang and Oomen (2008) (JO). The time interval for integration of the linear (BNS) and difference tests (JO) is one day The jump sizes are 10%, 25%, 50%, 100%, 300% of the spot volatility.

Jump size	$S8$ filter	*Haar* filter (LM)	Linear test (BNS)	Difference Test (JO)
		$Frequency = 2\ hours$, 95% confidence level		
3.00σ	0.0000	0.0000	0.2511	0.0650
1.00σ	0.0000	0.0000	0.2426	0.0484
0.50σ	0.0000	0.0000	0.1190	0.0010
0.25σ	0.0000	0.0000	0.0840	0.0000
0.10σ	0.0000	0.0000	0.0510	0.0000
		$Frequency = 1\ hour$, 95% confidence level		
3.00σ	0.0000	0.0000	0.1884	0.0090
1.00σ	0.0000	0.0000	0.1686	0.0076
0.50σ	0.0000	0.0000	0.0910	0.0009
0.25σ	0.0000	0.0000	0.0000	0.0000
0.10σ	0.0000	0.0000	0.0000	0.0000
		$Frequency = 2\ hours$, 99% confidence level		
3.00σ	0.8020	0.0400	0.2511	0.0650
1.00σ	0.1980	0.0100	0.2426	0.0484
0.50σ	0.0502	0.0080	0.1190	0.0010
0.25σ	0.0240	0.0040	0.0840	0.0000
0.10σ	0.0020	0.0000	0.0510	0.0000
		$Frequency = 1\ hour$, 99% confidence level		
3.00σ	0.2400	0.1120	0.1884	0.0090
1.00σ	0.2120	0.1000	0.1686	0.0076
0.50σ	0.1560	0.0190	0.0910	0.0009
0.25σ	0.1120	0.0120	0.0000	0.0000
0.10σ	0.0040	0.0040	0.0000	0.0000

Table 3.3: SUCCESS RATE COMPARISON WITH OTHER JUMP TESTS. The simulation only allows one jump and assumes constant volatility and non-zero drift part in the price process. The number of repetitions is 1,000. Success rate is defined as the probability that the test will detect only true jumps. The table shows the success rate of my test ($S8$ filter, *Haar* filter corresponding to Lee and Mykland (2008) (LM)), Linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and Difference test of Jiang and Oomen (2008) (JO). The time interval for integration of the linear (BNS) and difference tests (JO) is one day. The jump sizes are 10%, 25%, 50%, 100%, 300% of the spot volatility.

Sampling Frequency	N_T	$N_{T,+}$	$N_{T,-}$	N_t	$Size_{N_t}$	$Size_{N_{t,+}}$	$Size_{N_{t,-}}$
			GE				
1-min	11.7	6.10	5.60	2.70	−0.04%	0.01%	−0.04%
5-mins	3.70	1.80	1.90	0.90	−0.09%	0.01%	−0.09%
15-mins	1.50	0.75	0.75	0.70	−0.03%	0.04%	−0.04%
			IBM				
1-min	11.3	5.90	5.40	2.70	−0.001%	0.10%	−0.02%
5-mins	3.60	1.80	1.80	0.70	−0.05%	0.02%	−0.05%
15-mins	1.20	0.60	0.60	0.40	−0.05%	0.03%	−0.07%
			WMT				
1-min	11.0	6.00	5.00	3.50	−0.01%	0.08%	−0.01%
5-mins	3.60	1.90	1.70	1.00	−0.20%	0.05%	−0.05%
15-mins	1.30	0.80	0.50	0.40	−0.02%	0.01%	−0.02%

Table 3.4: JUMP DYNAMICS OF INDIVIDUAL EQUITIES. This table contains the jump dynamics of three U.S. individual equities: GE, IBM, and WMT based on transaction prices from the New York Stock Exchange (NYSE) duriug three months from January 1st to March 31st, 2008. N_T is the average number of total jumps estimated in each day. $N_{T,+}$ is the average of the number of positive jumps estimated for each day. $N_{T,-}$ is the average number of negative jumps estimated for each day. N_t is the average number of trading session jnmps estimated for each day. $Size_{N_t}$ is the average magnitude of the trading-session jumps estimated for each day. $Size_{N_{t,+}}$ is the average magnitude of positive trading-session jumps estimated for each day. $Size_{N_{t,-}}$ is the average magnitude of the negative trading-session jumps estimated for each day.



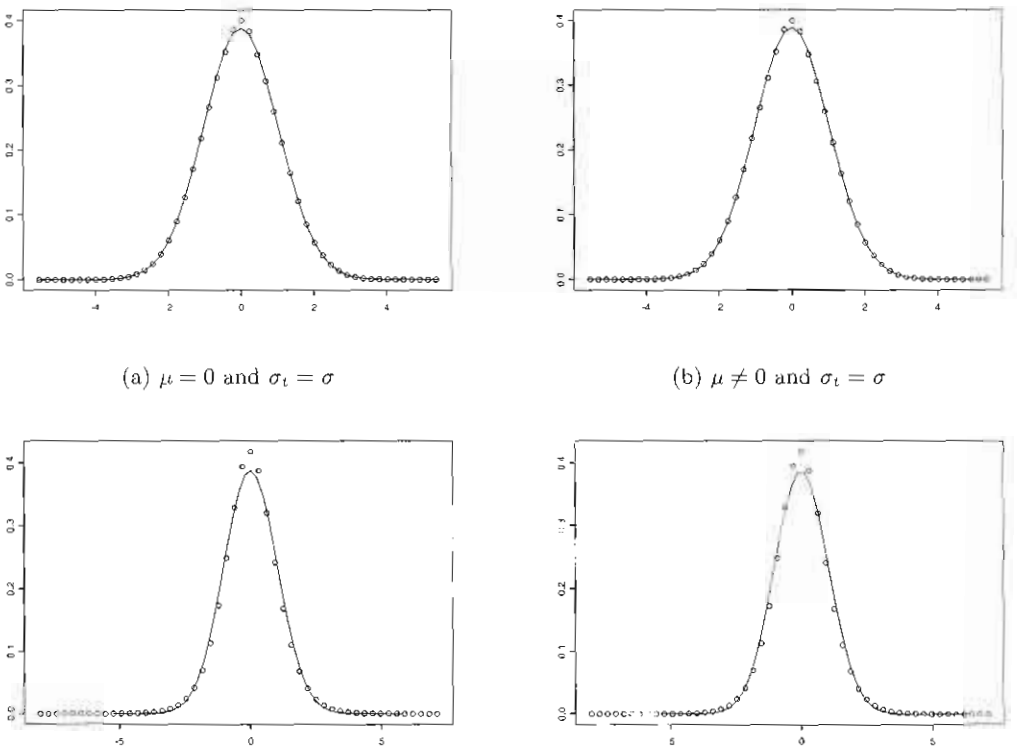
(a) $\mu = 0$ and $\sigma_t = \sigma$

(b) $\mu \neq 0$ and $\sigma_t = \sigma$

(c) $\mu = 0$ and $\sigma_t \neq \sigma$

(d) $\mu \neq 0$ and $\sigma_t \neq \sigma$

Figure 3.1: DENSITY PLOT OF THE SIMULATED STATISTIC UNDER THE NULL HYPOTHESIS WITH THE *Haar* FILTER. (a) Density plot of the simulated statistic with zero mean and constant volatility. (b) Density plot of the simulated statistic with non-zero mean and constant volatility. (c) Density plot of the simulated statistic with zero mean and stochastic volatility. (d) Density plot of the simulated statistic with non-zero mean and stochastic volatility. For each plot, a standard normal density function is imposed. The solid line is a standard normal density function and the line with circles is an empirical null distribution of the jump statistic from one million simulations.



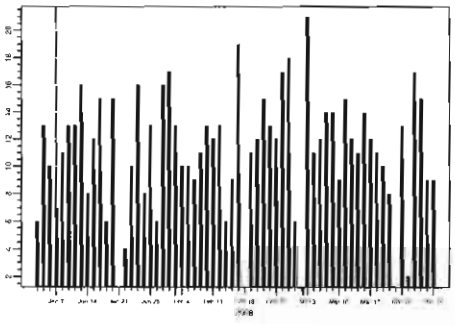
(a) $\mu = 0$ and $\sigma_t = \sigma$

(b) $\mu \neq 0$ and $\sigma_t = \sigma$

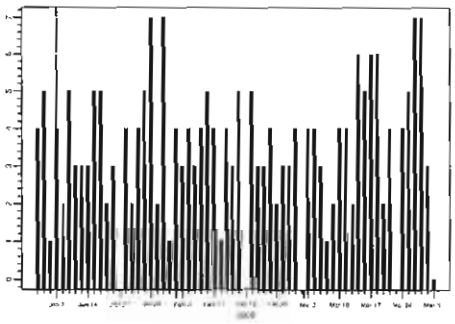
(c) $\mu = 0$ and $\sigma_t \neq \sigma$

(d) $\mu \neq 0$ and $\sigma_t \neq \sigma$

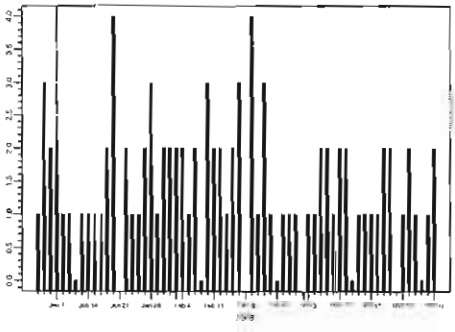
Figure 3.2: DENSITY PLOT OF THE SIMULATED STATISTIC UNDER THE NULL HYPOTHESIS WITH THE $S8$ FILTER. (a) Density plot of the simulated statistic with zero mean and constant volatility. (b) Density plot of the simnlated statistic with non-zero mean and constant volatility. (c) Density plot of the simulated statistic with zero mean and stochastic volatility. (d) Density plot of the simulated statistic with nou-zero mean and stochastic volatility. For each plot, a standard normal density function is imposed. The solid line is a standard normal density function and the line with circles is an empirical null distribution of the jump statistic from one million simulations.



(a) Total jumps at 1-min frequency

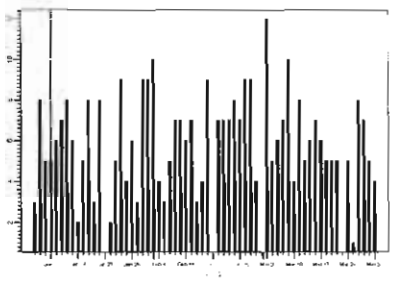


(b) Total jumps at 5-mins frequency

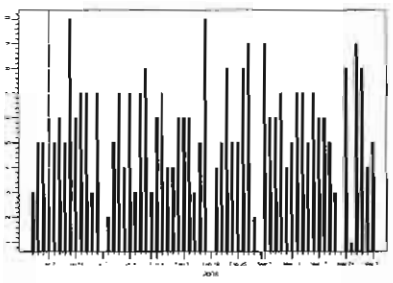


(c) Total jumps at 15-mins frequency

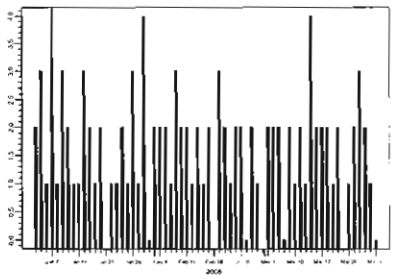
Figure 3.3: MULTI-SCALE JUMP DYNAMICS OF GENERAL ELECTRONIC (GE) FROM JANUARY 1 TO MARCH 31, 2008. (a) The total number of jumps estimated for each day using one-minute data. (b) The total number of jumps estimated in each day using five-minutes data. (c) The total number of jumps estimated in each day using fifteen-minutes data.



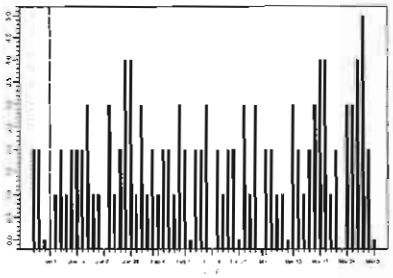
(a) Positive jumps at 1-min frequency



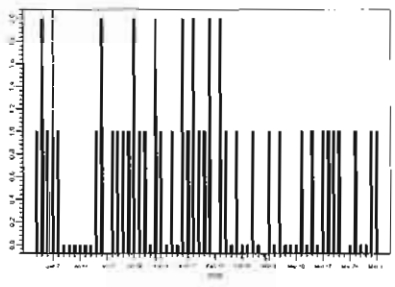
(b) Negative jumps at 1-min frequency



(c) Positive jumps at 5-mins frequency



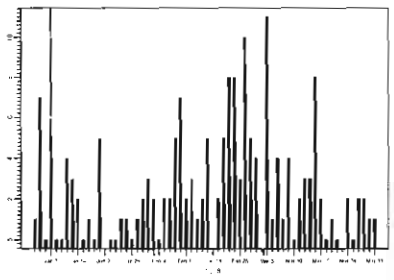
(d) Negative jumps at 5-mins frequency



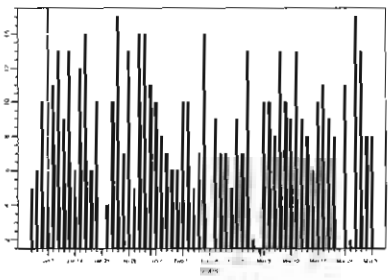
(e) Positive jumps at 15-mins frequency

(f) Negative jumps at 15-mins frequency

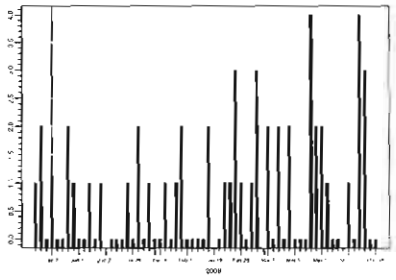
Figure 3.4: MULTI-SCALE DIRECTIONAL JUMP DYNAMICS OF GENERAL ELECTRONIC (GE) FROM JANUARY 1 TO MARCH 31, 2008. (a) The number of positive jumps estimated for each day using one-minute data. (b) The number of negative jumps estimated for each day using one-minute data. (c) The number of positive jumps estimated for each day using five-minutes data. (d) The number of negative jumps estimated for each day using five-minutes data. (e) The number of positive jumps estimated for each day using fifteen-minutes data. (f)The number of negative jumps estimated for each day using fifteen-minutes data.



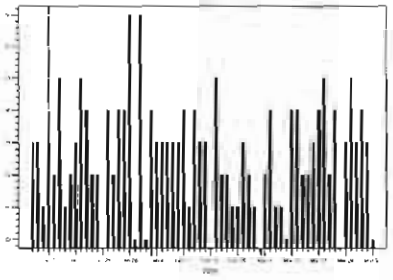
(a) Trading session jumps at 1-min frequency



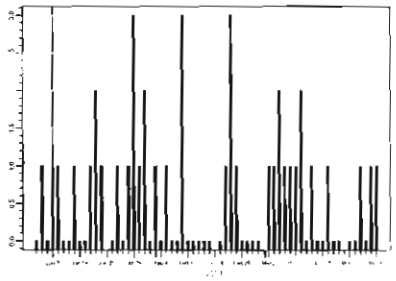
(b) Off-trading session jumps at 1-min frequency



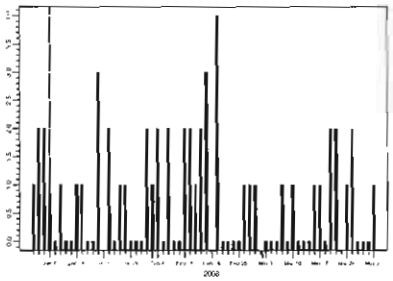
(c) Trading session jumps at 5-mins frequency



(d) Off-trading session jumps at 5-mins frequency

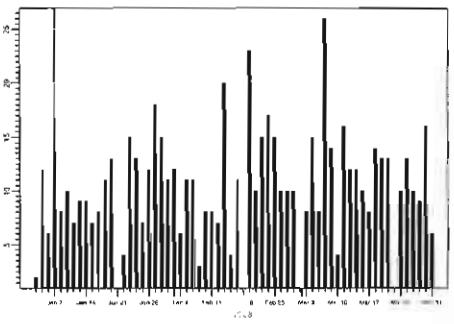


(e) Trading session jumps at 15-mins frequency

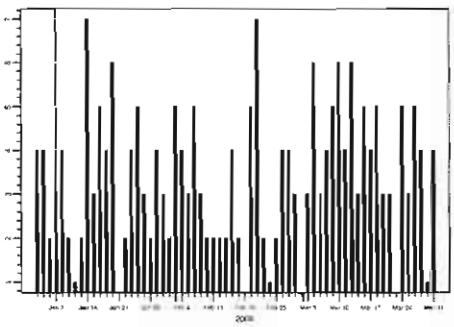


(f) Off-trading session jumps at 15-mins frequency

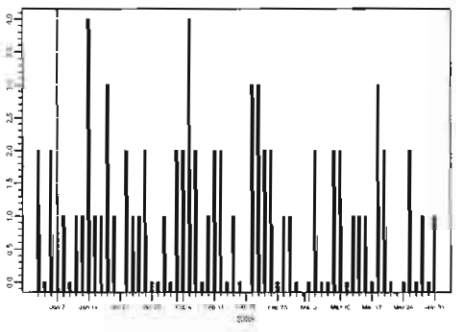
Figure 3.5: MULTI-SCALE JUMP DYNAMICS OF GENERAL ELECTRONIC (GE) FROM JANUARY 1 TO MARCH 31, 2008. (a) The number of trading session jumps estimated for each day using one-minute data. (b) The nuruber of off-trading session jumps estimated for each day using one-minute data. (c) The number of trading sessiou jumps estimated for each day using five-minutes data. (d) The number of off-trading session jumps estimated for each day using five-minutes data. (e) The number of trading session jumps estimated for each day using fifteen-minutes data. (f) The number of off-trading session jumps estimated in each day using fifteeu-minutes data.



(a) Total jumps at 1-min frequency

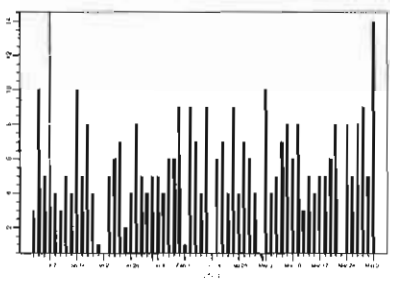


(b) Total jumps at 5-mins frequency

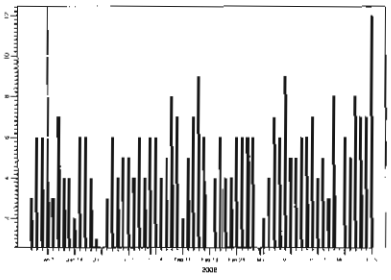


(c) Total jumps at 15-mins frequency

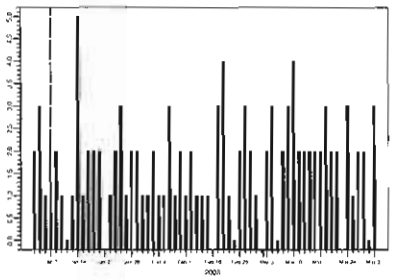
Figure 3.6: MULTI-SCALE JUMP DYNAMICS OF INTERNATIONAL BUSINESS MACHINE (IBM) FROM JANUARY 1 TO MARCH 31, 2008. (a) The total number of jumps estimated for each day using one-minute data. (b) The total number of jumps estimated in each day using five-minutes data. (c) The total number of jumps estimated in each day using fifteen-minutes data



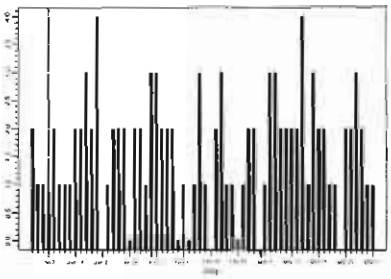
(a) Positive jumps at 1-min frequency



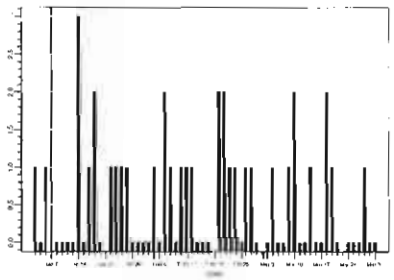
(b) Negative jumps at 1-min frequency



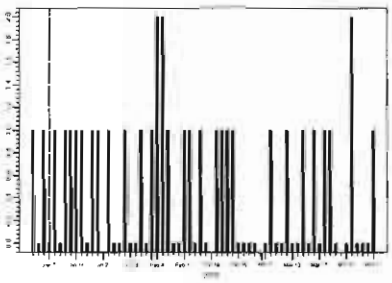
(c) Positive jumps at 5-mins frequency



(d) Negative jumps at 5-mins frequency

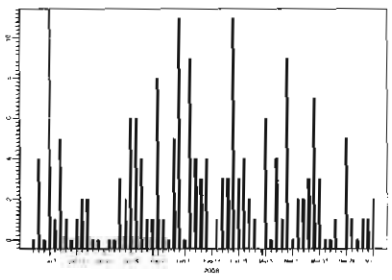


(e) Positive jumps at 15-mins frequency

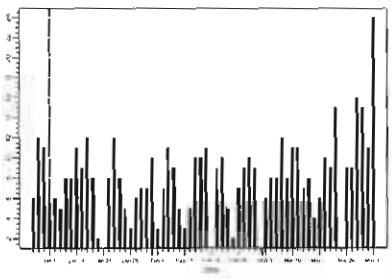


(f) Negative jumps at 15-mins frequency

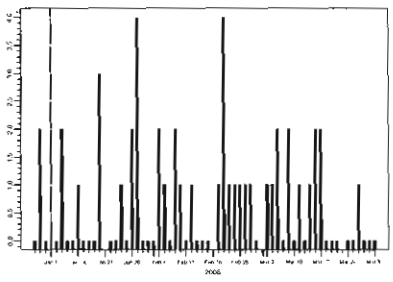
Figure 3.7: MULTI-SCALE DIRECTIONAL JUMP DYNAMICS OF INTERNATIONAL BUSINESS MACHINE (IBM) FROM JANUARY 1 TO MARCH 31, 2008. (a) The number of positive jumps estimated for each day using one-minute data. (b) The number of negative jumps estimated for each day using one-minute data. (c) The number of positive jumps estimated for each day using five-minutes data. (d) The number of negative jumps estimated for each day using five-minutes data. (e) The number of positive jumps estimated for each day using fifteen-minutes data. (f)The number of negative jumps estimated for each day using fifteen-minutes data.



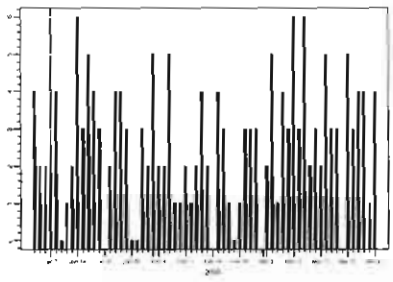
(a) Trading session jumps at 1-min frequency



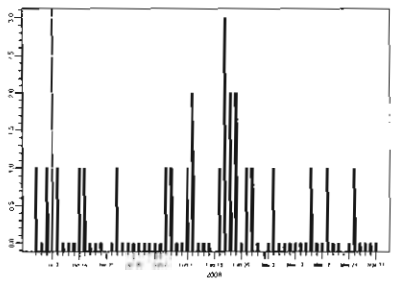
(b) Off-trading session jumps at 1-min frequency



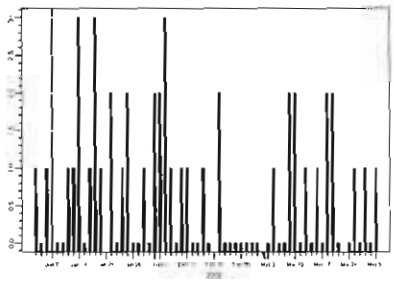
(c) Trading session jumps at 5-min frequency



(d) Off-trading session jumps at 5-min frequency

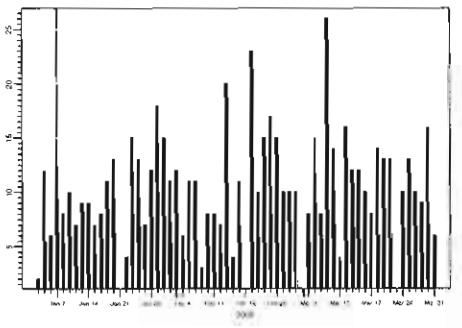


(e) Trading session jumps at 15-min frequency

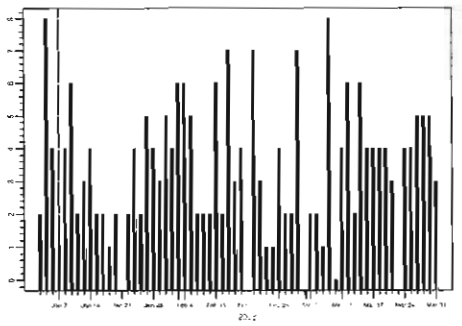


(f) Off-trading session jumps at 15-min frequency

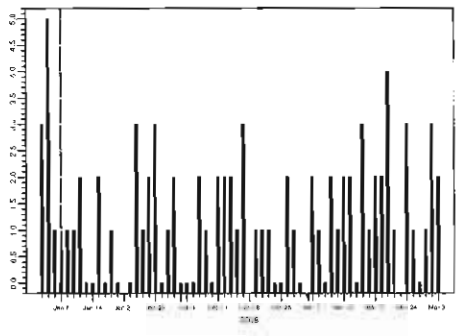
Figure 3.8: MULTI-SCALE JUMP DYNAMICS OF INTERNATIONAL BUSINESS MACHINE (IBM) FROM JANUARY 1 TO MARCH 31, 2008. a) The number of trading session jumps estimated for each day using one-minute data. (b) The number of off-trading session jumps estimated for each day using one-minute data. (c) The number of trading session jumps estimated for each day using five-minutes data. (d) The number of off-trading session jumps estimated for each day using five-minutes data. (e) The number of trading session jumps estimated for each day using fifteen-minutes data. (f) The number of off-trading session jumps estimated in each day using fifteen-minutes data.



(a) Total jumps at 1-min frequency

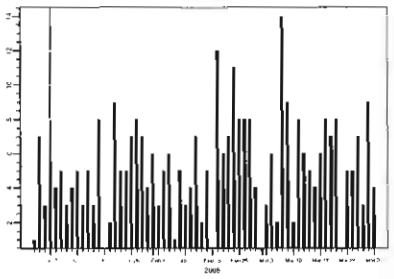


(b) Total jumps at 5-mins frequency

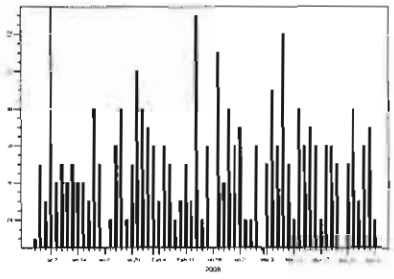


(c) Total jumps at 15-mins frequency

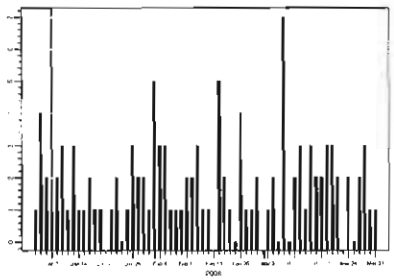
Figure 3.9: MULTI-SCALE JUMP DYNAMICS OF WAL-MART FROM JANUARY 1 TO MARCH 31, 2008. (a) The total number of jumps estimated for each day using one-minute data. (b) The total number of jumps estimated in each day using five-minutes data. (c) The total number of jumps estimated in each day using fifteen-minutes data.



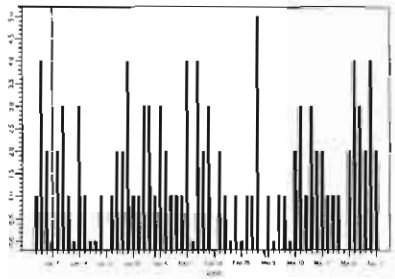
(a) Positive jumps at 1-min frequeucy



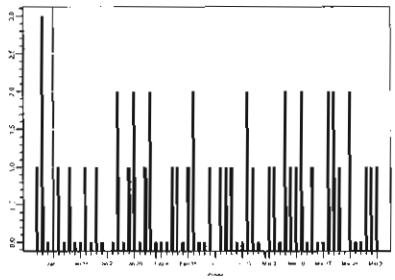
(b) Negative jumps at 1-min frequency



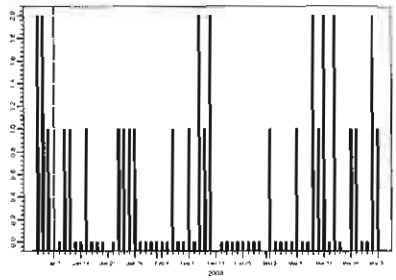
(c) Positive jumps at 5-mius frequency



(d) Negative jumps at 5-mins frequency

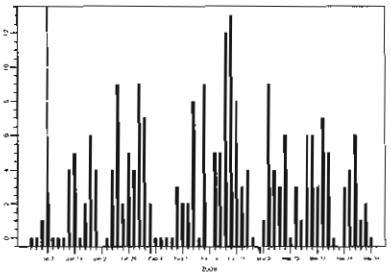


(e) Positive jumps at 15-mins frequency

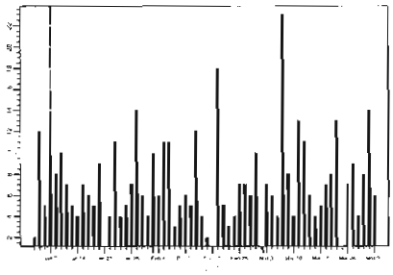


(f) Negative jumps at 15-mins frequency

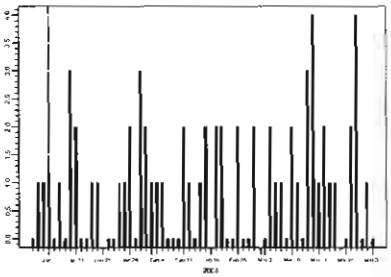
Figure 3.10: MULTI-SCALE DIRECTIONAL JUMP DYNAMICS OF WAL-MART FROM JANUARY 1 TO MARCH 31, 2008. (a) The number of positive jumps estimated for each day using one-minute data. (b) The number of negative jumps estimated for each day using one-minute data. (c) The number of positive jumps estimated for each day using five-minutes data. (d) The number of negative jumps estimated for each day using five-minutes data. (e) The number of positive jumps estimated for each day using fifteen-minutes data. (f) The number of negative jumps estimated for each day using fifteen-minutes data.



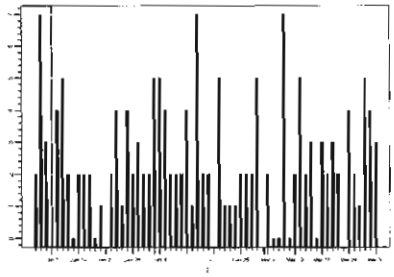
(a) Trading session jumps at 1-min frequency



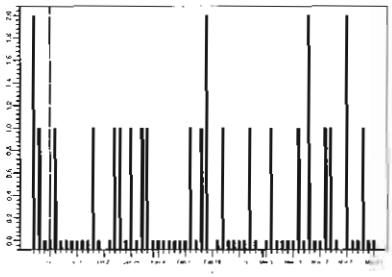
(b) Off-trading session jumps at 1-min frequency



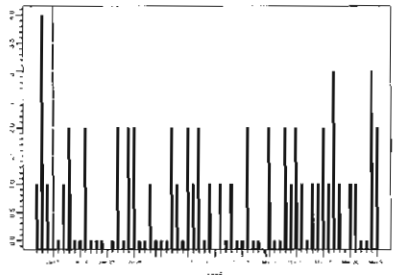
(c) Trading session jumps at 5-mins frequency



(d) Off-trading session jumps at 5-mins frequency



(e) Trading session jumps at 15-mins frequency



(f) Off-trading session jumps at 15-mins frequency

Figure 3.11: MULTI-SCALE JUMP DYNAMICS OF WAL-MART FROM JANUARY 1 TO MARCH 31, 2008. a) The number of trading session jumps estimated for each day using one-minute data. (b) The number of off-trading session jumps estimated for each day using one-minute data. (c) The number of trading session jumps estimated for each day using five-minutes data. (d) The number of off-trading session jumps estimated for each day using five-minntes data. (e) The number of trading session jumps estimated for each day using fifteen-minutes data. (f) The number of off-trading session jumps estimated in each day using fifteen-minutes data.

Appendix A

Appendices

A.1 Appendix A: Proofs

Proof to Proposition 1

Proof: With identical initial beliefs $\widehat{f}_0^A = \widehat{f}_0^B$, and invoking DeGroot (1970), this leads to $\widehat{f}_t^A = \widehat{f}_t^B$ for $t = 1, 2, \ldots, T-1$:

$$\begin{aligned}
\tau_t &= \tau_{t-1} + 2\tau_\epsilon \\
\widehat{f}_t^A &= \widehat{f}_{t-1}^A + \frac{\tau_\epsilon}{\tau_t}(S_t^A + S_t^B - 2\widehat{f}_{t-1}^A) \\
\widehat{f}_t^B &= \widehat{f}_1^B + \frac{\tau_\epsilon}{\tau_t}(S_t^B + S_t^A - 2\widehat{f}_{t-1}^B).
\end{aligned}$$

Q.E.D

Proof to Proposition 2

Proof: The proof is by induction. The Bayesian Nash equilibrium at $T-1$ is characterized by

$$\begin{aligned}
x_{T-1}^A &= x_{T-1}^B = x_{T-1} = \frac{\widehat{f}_{T-1} - \alpha}{2n+1} \\
p_{T-1} &= \alpha + (n_{T-1}^A + n_{T-1}^B)x_{T-1} \\
&= \alpha + (n_{T-1}^A + n_{T-1}^B)\frac{\widehat{f}_{T-1} - \alpha}{2n+1}.
\end{aligned}$$

Therefore, the equilibrium characterization holds when $t = T-1$. Suppose the Proposition

holds at $t = 2k$, where k is a integer such that $2k \leq T-1$, i.e,

$$\begin{aligned} x^A_{2k} &= \frac{2^{(T-2k)/2} n^{(T-2k-1)}}{(2n+1)^{(T-2k)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k} - \alpha) \\ p_{2k} &= \alpha + (n^A_{2k})\frac{2^{(T-2k)/2} n^{(T-2k-1)}}{(2n+1)^{(T-2k)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k} - \alpha). \end{aligned}$$

Therefore at $t = 2k-1$, the ith type A traders face the maximization problem which can be characterized by

$$max_{x^A_{2k\ \ 1,i}} E[p_{2k} - (\alpha + \sum_{j=1}^{n^A_{2k-1}} x^A_{2k-1,j} + \sum_{j=1}^{n^B_{2k-1}} x^B_{2k-1,j})]x^A_{2k-1,i}.$$

First-order condition to the above utility maximization problem is:

$$E[p_{2k}] - \alpha - (n-1)x^A_{2k-1,j} - nx^B_{2k-1,j} - 2x^A_{2k-1,i} = 0.$$

Invoking the symmetry result, we will have $x^A_{2k-1} = x^B_{2k-1}$ in equilibrium, then Equation A.1 can be rewritten as:

$$x^A_{2k-1,i} = \frac{E[p_{2k}] - \alpha}{2n+1}.$$

At $t = 2k-1$,

$$\begin{aligned} E[p_{2k}] &= E[\alpha + (n^A_{2k})\frac{2^{(T-2k)/2} n^{(T-2k)}}{(2n+1)^{(T-2k)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k} - \alpha)] \\ &= \alpha + n\frac{2^{(T-2k)/2} n^{(T-2k)}}{(2n+1)^{(T-2k)/2}(n+1)^{(T-2k)/2}}E(\widehat{f}_{2k} - \alpha) \\ &= \alpha + \frac{2^{(T-2k)/2} n^{(T-2k+1)}}{(2n+1)^{(T-2k)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k-1} - \alpha). \end{aligned}$$

Therefore, the optimal holdings for type A traders and type B traders at $t = 2k-1$ can be characterized by

$$x^A_{2k-1,i} = x^B_{2k-1,i} = \frac{E[p_{2k}] - \alpha}{2n+1} = \frac{2^{(T-2k)/2} n^{(T-2k+1)}}{(2n+1)^{(T-2k+2)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k-1} - \alpha).$$

Further, the price at $t = 2k-1$ is

$$p_{2k-1} = \alpha + (n^A_{2k-1} + n^B_{2k-1})\frac{2^{(T-2k)/2} n^{(T-2k)}}{(2n+1)^{(T+2k+2)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k-1} - \alpha).$$

At $t = 2k-2$, the type A traders face the maximization problem which can be characterized by

$$max_{x^A_{2k-2,i}} \quad E[p_{2k-1} - (\alpha + \sum_{j=1}^{n^A_{2k-2}} x^A_{2k-2,j}]x^A_{2k-2,i}.$$

First-order condition to the above utility maximization problem is:

$$E[p_{2k-1}] - \alpha - (n-1)x^A_{2k-1,j} - 2x^A_{2k-1,i} = 0.$$

Invoking the symmetry result, we will have $x^A_{2k-1} = x^B_{2k-1}$ in equilibrium, then Equation A.1 can be rewritten as:

$$x^A_{2k-2,i} = \frac{E[p_{2k-1}] - \alpha}{n+1}.$$

Similarly, at $t = 2k-2$

$$\begin{aligned} E[p_{2k-1}] &= E[\alpha + (n^A_{2k-1} + n^B_{2k-1})\frac{2^{(T-2k)/2}n^{(T-2k)}}{(2n+1)^{(T-2k+2)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k-1} - \alpha)] \\ &= \alpha + 2n\frac{2^{(T-2k)/2}n^{(T-2k)}}{(2n+1)^{(T-2k+2)/2}(n+1)^{(T-2k)/2}}E(\widehat{f}_{2k-1} - \alpha) \\ &= \alpha + \frac{2^{(T-2k+2)/2}n^{(T-2k+1)}}{(2n+1)^{(T-2k+2)/2}(n+1)^{(T-2k)/2}}(\widehat{f}_{2k-2} - \alpha). \end{aligned}$$

Therefore, the optimal holding for type A traders and type B traders at $t = 2k-2$ can be characterized by

$$\begin{aligned} x^A_{2k-2,i} &= \frac{E[p_{2k-1}] - \alpha}{n+1} \\ &= \frac{2^{(T-2k+2)/2}n^{(T-2k+1)}}{(2n+1)^{(T-2k+2)/2}(n+1)^{(T-2k+2)/2}}(\widehat{f}_{2k-2} - \alpha) \\ &= \frac{2^{(T-t)/2}n^{(T-t-1)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t - \alpha). \end{aligned}$$

Furthermore, the price at $t = 2k-2$ is

$$p_t = \alpha + (n^A_t)\frac{2^{(T-t)/2}n^{(T-t-1)}}{(2n+1)^{(T-t)/2}(n+1)^{(T-t)/2}}(\widehat{f}_t - \alpha).$$

Q.E.D

Proof to Proposition 5

Proof : At t, both types of traders know actual arrivals of both types of traders' last period, N^A_{t-1} and N^B_{t-1}, and price in the last period, p_{t-1}. When t is odd, $t-1$ is even. From Proposition 4,

$$p_{t-1} = \alpha + (n^A_{t-1})\frac{2^{(T-t+1)/2}n^{(T-t)}}{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t+1)/2}}(\widehat{f}^A_{t-1} - \alpha). \tag{A.1}$$

Type B can infer $\widehat{f}^A_{t-1}$ according to

$$\widehat{f}^A_{t-1} = \alpha + \frac{(2n+1)^{(T-t+1)/2}(n+1)^{(T-t+1)/2}}{2^{(T-t+1)/2}n^{(T-t)}(n^A_{t-1})}(p_{t-1} - \alpha), \tag{A.2}$$

while type A cannot extract any information about $\widehat{f}^B_{t-1}$. When t is even, $t-1$ is odd. From Proposition 4,

$$p_{t-1} = \alpha + (n^A_{t-1} - \frac{n}{n+1}n^B_{t-1})K_{t-1}(\widehat{f}^A_{t-1} - \alpha) + (n^B_{t-1} - \frac{n}{n+1}n^A_{t-1})K_{t-1}(\widehat{f}^B_{t-1} - \alpha) \tag{A.3}$$

where $K_{t-1} = \frac{2^{(T-t)/2}n^{(T-t)}}{(2n+1)^{(T-t+2)/2}(n+1)^{(T-t-2)/2}}$. Type B can infer $\widehat{f}^A_{t-1}$ according to

$$\widehat{f}^A_{t-1} = \alpha + \frac{(p_{t-1} - \alpha)}{(n^A_{t-1} - \frac{n}{n+1}n^B_{t-1})K_{t-1}} - \frac{(n^B_{t-1} - \frac{n}{n+1}n^A_{t-1})K_{t-1}(\widehat{f}^B_{t-1} - \alpha)}{(n^A_{t-1} - \frac{n}{n+1}n^B_{t-1})K_{t-1}}. \tag{A.4}$$

And type A can also infer $\widehat{f}^B_{t-1}$ in a similar way:

$$\widehat{f}^B_{t-1} = \alpha + \frac{(p_{t-1} - \alpha)}{(n^B_{t-1} - \frac{n}{n+1}n^A_{t-1})K_{t-1}} - \frac{(n^A_{t-1} - \frac{n}{n+1}n^B_{t-1})K_{t-1}(\widehat{f}^A_{t-1} - \alpha)}{(n^B_{t-1} - \frac{n}{n+1}n^A_{t-1})K_{t-1}}. \tag{A.5}$$

Hence,

1. When t is odd, type B traders know $\widehat{f}^A_{t-1}$, $\widehat{f}^A_{t-2}$, ..., $\widehat{f}^A_0$ while type A only know $\widehat{f}^B_{t-2}$, $\widehat{f}^B_{t-4}$, ..., $\widehat{f}^B_0$.

2. When t is even, type B traders know $\widehat{f}^A_{t-1}$, $\widehat{f}^A_{t-2}$, ..., $\widehat{f}^A_0$ while type A only know $\widehat{f}^B_{t-1}$, $\widehat{f}^B_{t-3}$, ..., $\widehat{f}^B_0$.

Q.E.D

Proof to Proposition 6

Proof: This is done by induction. Notice that at $t = 1$, there is no need to do signal extraction because there are no signals for traders at $t = 0$. Therefore at $t = 1$, both types of traders will update their belief according to their own signals, i.e:

$$\begin{aligned}
\tau_1^A &= \tau_0^A + \tau_\epsilon \\
\tau_1^B &= \tau_0^B + \tau_\epsilon \\
\widehat{f}_1^A &= \widehat{f}_0^A + \frac{\tau_\epsilon}{\tau_1}(S_1^A - \widehat{f}_0^A) \\
\widehat{f}_1^B &= \widehat{f}_0^B + \frac{\tau_\epsilon}{\tau_1}(S_1^B - \widehat{f}_0^B).
\end{aligned}$$

At $t = 2$, type B traders know $\widehat{f}_1^A$, and type A traders know $\widehat{f}_1^B$. For type B traders, they can infer the signal type A traders received at $t = 1$, $S_1^A = (\widehat{f}_1^A - \widehat{f}_0^A)\frac{\tau_1^A}{\tau_\epsilon} + \widehat{f}_0^A$. In the meantime, type A traders are also able to infer the exact signal type B received at $t = 1$, $S_1^B = (\widehat{f}_1^B - \widehat{f}_0^B)\frac{\tau_1^B}{\tau_\epsilon} + \widehat{f}_0^B$ when $t = 2$. Therefore type A traders update their beliefs incorporating their own signal S_2^A and the signal extracted S_1^B, while type B traders update their beliefs incorporating their own signal S_2^B and the signal extracted S_1^A, i.e:

$$\begin{aligned}
\tau_2^A &= \tau_1^A + 2\tau_\epsilon \\
\tau_2^B &= \tau_1^B + 2\tau_\epsilon \\
\widehat{f}_2^A &= \widehat{f}_1^A + \frac{\tau_\epsilon}{\tau_2^A}(S_2^A + S_1^B - 2\widehat{f}_1^A) \\
\widehat{f}_2^B &= \widehat{f}_1^B + \frac{\tau_\epsilon}{\tau_2^B}(S_2^B + S_1^A - 2\widehat{f}_1^B).
\end{aligned}$$

At $t = 3$, type B traders know $\widehat{f}_2^A$ and $\widehat{f}_1^A$, while type A traders have no further information. Similarly, type B traders can infer exact signal $S_2^A = (\widehat{f}_2^A - \widehat{f}_1^A)\frac{\tau_2^A}{\tau_\epsilon} + 2\widehat{f}_1^A - S_1^B$ while type A traders cannot, i.e:

$$\begin{aligned}
\tau_3^A &= \tau_2^A + \tau_\epsilon \\
\tau_3^B &= \tau_2^B + 2\tau_\epsilon \\
\widehat{f}_3^A &= \widehat{f}_2^A + \frac{\tau_\epsilon}{\tau_3^A}(S_3^A - \widehat{f}_2^A) \\
\widehat{f}_3^B &= \widehat{f}_2^B + \frac{\tau_\epsilon}{\tau_3^B}(S_3^B + S_2^A - 2\widehat{f}_2^B).
\end{aligned}$$

At $t = 4$, type B can infer $S_3^A = (\widehat{f}_3^A - \widehat{f}_1^A)\frac{\tau_3^A}{\tau_\epsilon} + \widehat{f}_3^A$, while type A knows $\widehat{f}_3^B$ and $\widehat{f}_1^B$.

Type A traders understand that

$$\begin{aligned}
\widehat{f}_3^B &= \widehat{f}_2^B + \frac{\tau_\epsilon}{\tau_3^B}(S_3^B + S_2^A - 2\widehat{f}_2^B) \\
&= \widehat{f}_1^B + \frac{\tau_\epsilon}{\tau_2^B}(S_2^B + S_1^A - 2\widehat{f}_1^B) + \frac{\tau_\epsilon}{\tau_3^B}(S_3^B + S_2^A - 2\widehat{f}_2^B) \\
&= \widehat{f}_1^B + \frac{\tau_\epsilon}{\tau_2^B}(S_2^B + S_1^A - 2\widehat{f}_1^B) + \frac{\tau_\epsilon}{\tau_3^B}(S_3^B + S_2^A - 2\widehat{f}_1^B - \frac{2\tau_\epsilon}{\tau_2^B}(S_2^B + S_1^A - 2\widehat{f}_1^B)).
\end{aligned}$$

Therefore, type A traders cannot exact the exact signal anymore. They can only know $\tau_\epsilon S_2^B + \frac{\tau_\epsilon}{\tau_3^B} S_3^B$, which is normally distributed. They can always normalize this combined signal to update their belief. Let $\widehat{S}_4^A = \frac{\tau_3^B S_2^B}{1+\tau_3^B} + \frac{S_3^B}{1+\tau_3^B}$ be the normalized signal, where $\widehat{S}_4^A$ is normally distributed with mean $\widetilde{f}$, precision $\widehat{\tau}_4^A = \frac{(1+\tau_3^B)^2}{(1+\tau_3^B)^2}\tau_\epsilon$. Therefore,

$$\begin{aligned}
\tau_4^A &= \tau_3^A + \tau_\epsilon + \widehat{\tau}_4^A \\
\tau_4^B &= \tau_3^B + 2\tau_\epsilon \\
\widehat{f}_4^A &= \widehat{f}_3^A + \frac{\tau_\epsilon}{\tau_4^A}(S_4^A - \widehat{f}_3^A) + \frac{\widehat{\tau}_4^A}{\tau_4^A}(\widehat{S}_4^A - \widehat{f}_3^A) \\
\widehat{f}_2^B &= \widehat{f}_1^B + \frac{\tau_\epsilon}{\tau_2^B}(S_2^B + S_1^A - 2\widehat{f}_1^B).
\end{aligned}$$

Suppose the proposition holds at $t = 2k$.[1] This implies

$$\begin{aligned}
\widehat{f}_{2k}^A &= \widehat{f}_{2k-1}^A + \frac{\tau_\epsilon}{\tau_{2k}^A}(S_{2k}^A - \widehat{f}_{2k-1}^A) + \frac{\widehat{\tau}_{2k}^A}{\tau_{2k}}(\widehat{S}_{2k}^A - \widehat{f}_{2k-1}^A) \\
\widehat{f}_{2k}^B &= \widehat{f}_{2k-1}^B + \frac{\tau_\epsilon}{\tau_{2k}^B}(S_{2k}^B + S_{2k-1}^A - 2\widehat{f}_{2k-1}^B)
\end{aligned}$$

where $\widehat{S}_{2k}^A = \frac{\tau_{2k-1}^B S_{2k-2}^B}{1+\tau_{2k-1}^B} + \frac{S_{2k-1}^B}{1+\tau_{2k-1}^B}$.

At $t = 2k+1$, type B traders know $\widehat{f}_{2k}^A$, $\widehat{f}_{2k-1}^A$, while type A traders have no further information. Notice that $\widehat{S}_{2k}^A$ is a function of S_{2k-2}^B and S_{2k-1}^B. This implies type B traders know $\widehat{S}_{2k}^A$. Therefore, type B traders can infer exact signal

$$S_{2k}^A = \frac{\tau_{2k}^A}{\tau_\epsilon}(\widehat{f}_{2k}^A - \widehat{f}_{2k-1}^A - \frac{\widehat{\tau}_{2k}^A}{\tau_{2k}^A}(\widehat{S}_{2k}^A - \widehat{f}_{2k-1}^A)) + \widehat{f}_{2k-1}^A \tag{A.6}$$

[1] We have already shown that it holds at $t = 4$.

while type A traders cannot infer anything new. Therefore, at $t = 2k + 1$,

$$\begin{aligned}
\tau^A_{2k+1} &= \tau^A_{2k} + \tau_\epsilon \\
\tau^B_{2k+1} &= \tau^B_{2k} + 2\tau_\epsilon \\
\widehat{f}^A_{2k+1} &= \widehat{f}^A_{2k} + \frac{\tau_\epsilon}{\tau^A_{2k+1}}(S^A_{2k+1} - \widehat{f}^A_{2k}) \\
\widehat{f}^B_{2k+1} &= \widehat{f}^B_{2k} + \frac{\tau_\epsilon}{\tau^B_{2k+1}}(S^B_{2k+1} + S^A_{2k} - 2\widehat{f}^B_{2k}).
\end{aligned}$$

In other words, the proposition holds at $t = 2k + 1$. To complete the proof, we also need to examine the beliefs updating at $t = 2k + 2$. At $t = 2k + 2$, type B traders know $\widehat{f}_{2k+1}$ and $\widehat{f}_{2k}$. Therefore, they can infer $S^A_{2k+1} = \frac{\tau^A_{2k+1}}{\tau_\epsilon}(\widehat{f}^A_{2k+1} - \widehat{f}^A_{2k}) + \widehat{f}^A_{2k}$. Therefore, for type B traders, the belief updating at $t = 2k + 2$ is characterized as

$$\begin{aligned}
\tau^B_{2k+2} &= \tau^B_{2k+1} + 2\tau_\epsilon \\
\widehat{f}^B_{2k+2} &= \widehat{f}^B_{2k+1} + \frac{\tau_\epsilon}{\tau^B_{2k+2}}(S^B_{2k+2} + S^A_{2k+1} - 2\widehat{f}^B_{2k+1})
\end{aligned}$$

while type A traders only know $\widehat{f}_{2k+1}$ and $\widehat{f}_{2k-1}$. They understand the belief updating of type B at $t = 2k + 1$ is

$$\begin{aligned}
\widehat{f}^B_{2k+1} &= \widehat{f}^B_{2k} + \frac{\tau_\epsilon}{\tau^B_{2K+1}}(S^B_{2k+1} + S^A_{2k} - 2\widehat{f}^B_{2k}) \\
&= \widehat{f}^B_{2k-1} + \frac{\tau_\epsilon}{\tau^B_{2k}}(S^B_{2k} + S^A_{2k-1} - 2\widehat{f}^B_{2k-1}) + \frac{\tau_\epsilon}{\tau^B_{2k+1}}(S^B_{2k+1} + S^A_{2k} - 2\widehat{f}^B_{2k}) \\
&= \widehat{f}^B_{2k-1} + \frac{\tau_\epsilon}{\tau^B_{2k}}(S^B_{2k} + S^A_{2k-1} - 2\widehat{f}^B_{2k-1}) \\
&\quad + \frac{\tau_\epsilon}{\tau^B_{2k+1}}(S^B_{2k+1} + S^A_{2k} - 2(\widehat{f}^B_{2k-1} + \frac{\tau_\epsilon}{\tau^B_{2k}}(S^B_{2k} + S^A_{2k-1} - 2\widehat{f}^B_{2k-1})).
\end{aligned}$$

Therefore, type A can infer a composite signal $\widehat{S}^A_{2k+2} = \frac{\tau^B_{2k+1}S^B_{2k}}{1+\tau^B_{2k+1}} + \frac{S^B_{2k+1}}{1+\tau^B_{2k+1}}$, with the precision $\widehat{\tau}^A_{2k+2} = \frac{1+(\tau^B_{2k+1})^2}{(1+\tau^B_{2k+1})^2}\tau_\epsilon$. In summary, the belief updating of type A traders at $t = 2k + 2$ is

$$\begin{aligned}
\tau^A_{2k+2} &= \tau^A_{2k+1} + \tau_\epsilon + \widehat{\tau}^A_{2k+2} \\
\tau^B_{2k+2} &= \tau^B_{2k+1} + 2\tau_\epsilon \\
\widehat{f}^A_{2k+2} &= \widehat{f}^A_{2k+1} + \frac{\tau_\epsilon}{\tau^A_{2k+2}}(S^A_{2k+2} - \widehat{f}^A_{2k+1}) + \frac{\widehat{\tau}^A_{2k+2}}{\tau^A_{2k+2}}(\widehat{S}^A_{2k+2} - \widehat{f}^A_{2k+1}) \\
\widehat{f}^B_{2k+2} &= \widehat{f}^B_{2k+1} + \frac{\tau_\epsilon}{\tau^B_{2k+2}}(S^B_{2k+2} + S^A_{2k+1} - 2\widehat{f}^B_{2k+1}).
\end{aligned}$$

Hence, the proposition holds at $t = 2k + 2$ and this completes the proof.

Q.E.D

Proof to Proposition 12

It is straight forward to check that the proposition holds. Then we leave out for readers' exercise.

Q.E.D

Proof to Proposition 13

If hierarchical information structure is assumed, the infinite regress problem collapses. The iterated expectations is reduced to $\widehat{V}_t^1 = E[V_t^1|F_{2,t}]$. We restrict ourself to the linear rational expectation equilibrium, where the price is a linear function takes the form of

$$P_t = \rho P_{t-1} + P_V V_t + P_\Theta \Theta_t + P_\Delta(\widehat{V}_t^1 - V_t^1) \tag{A.7}$$

Given Equation A.7, the demand for type 1 trader, the fully informed trader and the demand of type 2 trader, the partially informed trader are

$$\begin{aligned} X_t^1 &= \omega_1 E[Q_{t+1}|F_{1,t}] \\ X_t^2 &= \omega_2 E[Q_{t+1}|F_{2,t}] \end{aligned} \tag{A.8}$$

Using Equation A.7, we have

$$\begin{aligned} Q_{t+1} &= \lambda V_{t+1} + (1-\lambda)P_{t+1} - (1+r)P_t \\ &= \lambda V_{t+1} + (1-\lambda)(\rho P_t + P_V V_{t+1} + P_\Theta \Theta_{t+1} + P_\Delta(\widehat{V}_{t+1}^1 - V_{t+1}^1)) - (1+r)P_t \\ &= (1-\lambda)\rho P_t + (\lambda + (1-\lambda)P_V)(aV_t + b_V \epsilon_t) + (1-\lambda)P_\Theta \Theta_{t+1} \\ &\quad + (1-\lambda)P_\Delta(\widehat{V}_{t+1}^1 - V_{t+1}^1) - (1+r)P_t \end{aligned} \tag{A.9}$$

Hence, we have

$$\begin{aligned} E[Q_{t+1}|F_{2,t}] &= ((1-\lambda)\rho - (1+r))P_t + [\lambda + (1-\lambda)P_V]a[\widehat{V}_t^1 + V_t^2] + (1-\lambda)P_\Delta(\widehat{V}_{t+1}^1 - \widehat{V}_{t+1}^1) \\ &= ((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)a(\widehat{V}_t^1 + V_t^2) \\ &= ((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)aV_t + (\lambda + (1-\lambda)P_V)a(\widehat{V}_t^1 - V_t^1) \\ E[Q_{t+1}|F_{1,t}] &= ((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)aV_t + (1-\lambda)P_\Delta E[(\widehat{V}_{t+1}^1 - V_{t+1}^1)|F_{1,t}] \end{aligned} \tag{A.10}$$

Therefore, aggregate market demand is

$$\begin{aligned} X_t^1 + X_t^2 &= \Omega((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)aV_t) \\ &\quad + (\omega_1(1-\lambda)P_\Delta E[\widehat{V}_{t+1}^1 - V_{t+1}^1|F_{1,t}] + \omega_2(\lambda + (1-\lambda)P_V)a(\widehat{V}_t^1 - V_t^1) \end{aligned} \tag{A.11}$$

Market clearing condition implies that

$$\begin{aligned}\beta P_{t-1} + \Theta_t &= \Omega((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)aV_t) \\ &+(\omega_1(1-\lambda)I'_\Delta E[\widehat{V}^1_{t+1} - V^1_{t+1}|F_{1,t}] + \omega_2(\lambda + (1-\lambda)P_V)a(\widehat{V}^1_t - V^1_t)\end{aligned} \tag{A.12}$$

Matching coefficients with Equation A.7, we have

$$\begin{aligned}\rho &= \frac{\beta}{\Omega((1-\lambda)\rho - (1+r))} \\ P_V &= -\frac{a\Omega(\lambda + (1-\lambda)P_V}{\Omega((1-\lambda)\rho - (1+r))} \\ P_\Theta &= \frac{1}{\Omega((1-\lambda)\rho - (1+r))}\end{aligned} \tag{A.13}$$

Hence, we can solve all parameters except P_Δ and they are

$$\begin{aligned}0 &= (1-\lambda)\Omega\rho^2 - (1+r)\Omega\rho - \beta \\ P_V &= \frac{a\lambda}{-\frac{\beta}{\rho\Omega} - a(1-\lambda)} \\ P_\Theta &= \frac{\rho}{\beta}\end{aligned} \tag{A.14}$$

Now, we need to calculate P_Δ. In order to do that, we need to model the filtering expectation problem for type 2 trader explicitly. What type 2 trader can effectively observe is $(P_V - P_\Delta)V^1_t + P_\Theta\Theta_t$. He need to forecast V^1_t. It is essentially a filter problem. We set up the system as

$$\begin{aligned}V^1_t &= aV^1_{t-1} + b_V\epsilon^1_t \\ y_t &= (P_V - P_\Delta)V^1_t + P_\Theta\Theta_t\end{aligned}$$

Where ϵ^1_t is iid normal with mean 0 and variance 1 and Θ_t is iid normal with mean 0 and variance σ^2_Θ. Applying Kalman-Bucy filter[2], we have

$$\widehat{V}^1_t = a(1 - k(P_V - P_\Delta))\widehat{V}^1_{t-1} + k(P_V - P_\Delta)V^1_t + kP_\Theta\Theta_t \tag{A.15}$$

where k solves

$$P^2_\Theta\sigma^2_\Theta a^2(P_V - P\Delta)k^2 + (P^2_\Theta\sigma^2_\Theta(1-a^2) + b^2_V(P_V - P_\Delta)^2)k - b^2_V(P_V - P_\Delta) = 0$$

Let $c = 1 - k(P_V - P_\Delta)$, we can rewrite Equation A.15 into

$$\widehat{V}^1_t - V^1_t = ac(\widehat{V}^1_{t-1} - V^1_{t-1}) - b_V c\epsilon^1_t + kP_\Theta\Theta_t \tag{A.16}$$

[2]See Jazwinski (1970) for a description.

If we impose Equation A.16 on the market clearing condition, we have

$$P_\Delta = \frac{\omega_1 ac(1-\lambda)P_\Delta + \omega_2 a(\lambda + (1-\lambda)P_V)}{\Omega((1-\lambda)\rho - (1+r)} \tag{A.17}$$

Hence, we can solve for P_Δ and c simultaneously

$$P_\Delta = \frac{a\omega_2(\lambda + (1-\lambda)\frac{a\lambda}{\frac{\beta}{\rho\Omega} - a(1-\lambda)})}{-\frac{\beta}{\rho} - a\omega_1(1-\lambda)c}. \tag{A.18}$$

where c solves

$$\frac{\rho^2}{\beta^2}\sigma_\Theta^2(1-c)(1-a^2c) - cb_V^2(P_V - P_\Delta)^2 = 0$$

Q.E.D

Proof to Proposition 17

We only prove it when $N = 4$. The proof for $N > 4$ is similar. First, we have two partially informed traders, type 2 trader who knows V_t^2, V_t^3, and type 3 trader who only knows V_t^3. Notice that there is only one instrnment for them to filter the nseful information for them, i.e., the market price of the underlying asset. We conjecture the price to take form of:

$$P_t = \rho P_{t-1} + P_V V_t + \frac{\rho}{\beta}\Theta_t + P_{\Delta^1}(\widehat{V}_{2,t}^1 - V_t^1) + P_{\Delta^2}(\widehat{V}_{3,t}^1 + \widehat{V}_{3,t}^2 - V_t^1 - V_t^2), \tag{A.19}$$

where $\widehat{V}_{2,t}^1 = E[V_t^1|F_{2,t}]$, $\widehat{V}_{3,t}^1 = E[V_t^1|F_{3,t}]$, $\widehat{V}_{3,t}^1 = E[V_t^2|F_{3,t}]$.

Given Equation A.19, the demands from all traders are:

$$\begin{aligned} X_t^1 &= \omega_1 E[Q_{t+1}|F_{1,t}] \\ X_t^2 &= \omega_2 E[Q_{t+1}|F_{2,t}] \\ X_t^3 &= \omega_3 E[Q_{t+1}|F_{3,t}] \end{aligned}$$

Using Equation A.19, we have

$$\begin{aligned} Q_{t+1} &= \lambda V_{t+1} + (1-\lambda)P_{t+1} - (1+r)P_t \\ &= \lambda V_{t+1} + (1-\lambda)(\rho P_t + P_V V_{t+1} + P_\Theta \Theta_{t+1} + P_{\Delta^1}(\widehat{V}_{2,t+1}^1 - V_{t+1}^1) \\ &\quad + P_{\Delta^2}(\widehat{V}_{3,t+1}^1 + \widehat{V}_{3,t+1}^2 - V_{t+1}^1 - V_{t+1}^2)) - (1+r)P_t \\ &= ((1-\lambda)\rho - (1+r))P_t + (\lambda + (1-\lambda)P_V)(aV_t + b_V\epsilon_t) + (1-\lambda)P_\Theta\Theta_{t+1} \\ &\quad + (1-\lambda)(P_{\Delta^1}(\widehat{V}_{2,t+1}^1 - V_{t+1}^1) + P_{\Delta^2}(\widehat{V}_{3,t+1}^1 + \widehat{V}_{3,t+1}^2 - V_{t+1}^1 - V_{t+1}^2) \end{aligned} \tag{A.20}$$

Hence, we have:

$$\begin{aligned}
E[Q_{t+1}|F_{3,t}] &= ((1-\lambda)\rho-(1+r))P_t+(\lambda+(1-\lambda)P_V)a(\widehat{V}^1_{3,t}+\widehat{V}^2_{3,t}+V^3_t)\\
&\quad +(1-\lambda)P_{\Delta^1}(E[\widehat{V}^1_{2,t+1}|F_{3,t}]-\widehat{V}^1_{3,t+1})\\
&= ((1-\lambda)\rho-(1+r))P_t+(\lambda+(1-\lambda)P_V)a(\widehat{V}^1_{3,t}+\widehat{V}^2_{3,t}+V^3_t)\\
&= ((1-\lambda)\rho-(1+r))P_t+(\lambda+(1-\lambda)P_V)aV_t+\\
&\quad (\lambda+(1-\lambda)P_V)a(\widehat{V}^1_{3,t+1}+\widehat{V}^2_{3,t+1}-V^1_{t+1}-V^2_{t+1})\\
E[Q_{t+1}|F_{2,t}] &= ((1-\lambda)\rho-(1+r))P_t+(\lambda+(1-\lambda)P_V)a(\widehat{V}^1_t+V^2_t+V^3_t)\\
&\quad +(1-\lambda)P_{\Delta^2}E[(\widehat{V}^1_{3,t+1}+\widehat{V}^2_{3,t+1}-V^1_{t+1}-V^2_{t+1}|F_{2,t})]\\
&= ((1-\lambda)\rho-(1+r))P_t+(\lambda+(1-\lambda)P_V)aV_t+(\lambda+(1-\lambda)P_V)a(\widehat{V}^1_{2,t}-V^1_{2,t})\\
&\quad +(1-\lambda)P_{\Delta^2}E[(\widehat{V}^1_{3,t+1}+\widehat{V}^2_{3,t+1}-V^1_{t+1}-V^2_{t+1}|F_{2,t}]\\
&\quad -(1-\lambda)P_{\Delta^2}E[\widehat{V}_{2,t+1}-V^1_{t+1}|F_{2,t}]\\
E[Q_{t+1}|F_{1,t}] &= ((1-\lambda)\rho-(1+r))P_t+(\lambda+(1-\lambda)P_V)aV_t+(1-\lambda)P_{\Delta^1}E[(\widehat{V}^1_{2,t+1}-V^1_{t+1})|F_{1,t}]\\
&\quad +(1-\lambda)P_{\Delta^2}E[(\widehat{V}^1_{3,t+1}+\widehat{V}^2_{3,t+1}-V^1_{t+1}-V^2_{t+1})|F_{1,t}] \qquad \text{(A.21)}
\end{aligned}$$

Imposing market clearing condition and match the coefficients with Equation A.19, we have:

$$\begin{aligned}
\rho &= \frac{\beta}{\Omega((1-\lambda)\rho-(1+r))}\\
P_V &= -\frac{a\Omega(\lambda+(1-\lambda)P_V}{\Omega((1-\lambda)\rho-(1+r))}\\
P_\Theta &= \frac{1}{\Omega((1-\lambda)\rho-(1+r))}
\end{aligned} \qquad \text{(A.22)}$$

Hence, we can solve all parameters except P_{Δ^1} and P_{Δ^2} and they are:

$$\begin{aligned}
0 &= (1-\lambda)\Omega\rho^2-(1+r)\Omega\rho-\beta\\
P_V &= \frac{a\lambda}{-\frac{\beta}{\rho\Omega}-a(1-\lambda)}\\
P_\Theta &= \frac{\rho}{\beta}
\end{aligned} \qquad \text{(A.23)}$$

Now, we need to calculate P_{Δ^1} and P_{Δ^2}. In order to do that, we need to model the filtering expectation problem for type 2 trader and type 3 trader explicitly. What type 2 trader can effectively observe is $(P_V-P_{\Delta^1}-P_{\Delta^2})V^1_t+P_\Theta\Theta_t$. And type 3 trader can effectively observe

$(P_V - P_{\Delta^2})(V_t^1 + V_t^2) + P_{\Delta^1}(\widehat{V}_{2,t}^1) + P_\Theta \Theta_t$. Type 2 trader need to forecast V_t^1 while type 3 trader need to forecast V_t^1, V_t^2 and $\widehat{V}_{2,t}^1$. Notice that in hierarchical information structure, $F_{3,t} \subset F_{2,t}$. Therefore we have $E[E[V_t^1|F_{2,t}]|F_{3,t}] = E[V_t^1|F_{3,t}] = \widehat{V}_{3,t}^1$ which follows directly from law of iterated expectation.

Next, let us write the filter problem for type 2 trader first:

$$\begin{aligned} V_t^1 &= aV_{t-1}^1 + b_V \epsilon_t^1 \\ y_t &= (P_V - P_{\Delta^1} - P_{\Delta^2})V_t^1 + P_\Theta \Theta_t \end{aligned}$$

where ϵ_t^1 is *iid* normal with mean 0 and variance 1 and Θ_t is *iid* normal with mean 0 and variance σ_Θ^2. Applying Kalman-Bucy filter, we have:

$$\widehat{V}_t^1 = a(1 - k(P_V - P_\Delta))\widehat{V}_{t-1}^1 + k(P_V - P_\Delta)V_t^1 + kP_\Theta \Theta_t \tag{A.24}$$

where k solves

$$P_\Theta^2 \sigma_\Theta^2 a^2 (P_V - P_\Delta)k^2 + (P_\Theta^2 \sigma_\Theta^2 (1 - a^2) + b_V^2 (P_V - P_\Delta)^2)k - b_V^2 (P_V - P_\Delta) = 0$$

where $P_\Delta = P_{\Delta^1} + P_{\Delta^2}$.

Let $c = 1 - k(P_V - P_\Delta)$, we can rewrite Equation A.24 into

$$\widehat{V}_t^1 - V_t^1 = ac(\widehat{V}_{t-1}^1 - V_{t-1}^1) - b_V c \epsilon_t^1 + kP_\Theta \Theta_t \tag{A.25}$$

We continue to investigate the filter problem faced by type 3 trader:

$$\begin{bmatrix} V_t^1 + V_t^2 \\ \widehat{V}_{2,t}^1 - V_t^1 \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & ac \end{bmatrix} \begin{bmatrix} V_{t-1}^1 + V_{t-1}^2 \\ \widehat{V}_{2,t-1}^1 - V_{t-1}^1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_V \epsilon_t \\ -b_V c \epsilon_t^1 + kP_\Theta \Theta_t \end{bmatrix} \tag{A.26}$$

The second line of Equation A.26 means that type 3 trader do not make systematic expectation errors. It implies that in equilibrium, type 3 trader can forecast the right parameter values which determine the evolution process of the expectation errors for type 2 trader but not the errors themselves. If we apply Kalman-Bucy filter again, we have:

$$\begin{aligned} \begin{bmatrix} \widehat{V}_{3,t}^1 + \widehat{V}_{3.t}^2 \\ \widehat{V}_{3,t}^1 - \widehat{V}_{3,t}^1 \end{bmatrix} &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} k_1(P_V - P_{\Delta^2}) & k_1 P_{\Delta^1} \\ k_2(P_V - P_{\Delta^2}) & k_2 P_{\Delta^1} \end{bmatrix} \right) \begin{bmatrix} a & 0 \\ 0 & ac \end{bmatrix} \begin{bmatrix} \widehat{V}_{3,t-1}^1 + \widehat{V}_{3,t-1}^2 \\ \widehat{V}_{3,t-1}^1 - \widehat{V}_{3,t-1}^1 \end{bmatrix} \\ &\quad + \begin{bmatrix} k_1((P_V - P_{\Delta^2})(V_t^1 + V_t^2) + P_{\Delta^1}(\widehat{V}_{2,t}^1) + P_\Theta \Theta_t) \\ k_2((P_V - P_{\Delta^2})(V_t^1 + V_t^2) + P_{\Delta^1}(\widehat{V}_{2,t}^1) + P_\Theta \Theta_t) \end{bmatrix}, \end{aligned} \tag{A.27}$$

where k_1 and k_2 are parameters describe the weights used in filtering. Notice that it is really complicated to solve the parameters matrix directly. But we do not need to solve for the parameters value for filtering. We only need to describe the evolution of expectation errors of type 3 trader. Notice that $\widehat{V}^1_{3,t} - \widehat{V}^1_{3,t} = 0$ by law of iterated expectation. Equation A.27 can be reduced to:

$$\begin{aligned}\begin{bmatrix}\widehat{V}^1_{3,t} + \widehat{V}^2_{3,t} \\ 0\end{bmatrix} &= \left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}k_1(P_V - P_{\Delta^2}) & k_1 P_{\Delta^1}\\ k_2(P_V - P_{\Delta^2}) & k_2 P_{\Delta^1}\end{bmatrix}\right)\begin{bmatrix}a & 0\\ 0 & ac\end{bmatrix}\begin{bmatrix}\widehat{V}^1_{3,t-1} + \widehat{V}^2_{3,t-1}\\ 0\end{bmatrix} \\ &+ \begin{bmatrix}k_1((P_V - P_{\Delta^2})(V^1_t + V^2_t) + P_{\Delta^1}(\widehat{V}^1_{2,t}) + P_\Theta \Theta_t)\\ k_2((P_V - P_{\Delta^2})(V^1_t + V^2_t) + P_{\Delta^1}(\widehat{V}^1_{2,t}) + P_\Theta \Theta_t)\end{bmatrix},\end{aligned} \tag{A.28}$$

Explicit calculation shows:

$$\begin{aligned}\widehat{V}^1_{3,t} + \widehat{V}^2_{3,t} - V^1_t - V^2_t &= a(1 - k_1(P_V - P_{\Delta^2}))(\widehat{V}^1_{3,t-1} + \widehat{V}^2_{3,t-1} - V^1_{t-1} - V^2_{t-1}) + k_1 P_\Theta \Theta_t \\ &\quad + k_1 P_{\Delta^1}(\widehat{V}^1_{2,t} - V^1_t) + (1 - k_1(P_V - P_{\Delta^2}))b_V(\epsilon^1_t + \epsilon^2_t) \\ P_{\Delta^1}(V^1_{2,t} - V^1_t) &= a(P_V - P_{\Delta^2})(\widehat{V}^1_{3,t-1} + \widehat{V}^2_{3,t-1} - V^1_{t-1} - V^2_{t-1}) \\ &\quad - b_V(P_V - P_{\Delta^2}(\epsilon^1_t + \epsilon^2_t) - P_\Theta \Theta_t)\end{aligned} \tag{A.29}$$

Substitute the second line of Equation A.29 into the first line of Equation A.29, we have

$$(\widehat{V}^1_{3,t} + \widehat{V}^2_{3,t} - V^1_t - V^2_t) = a(\widehat{V}^1_{3,t-1} + \widehat{V}^2_{3,t-1} - V^1_{t-1} - V^2_{t-1}) - b_V(\epsilon^1_t + \epsilon^2_t) \tag{A.30}$$

Equation A.25 and Eqnation A.30 describe the evolution of expectation errors of type 2 trader and type 3 trader respectively. Using them with market clearing conditions, we have:

$$\begin{aligned}P_{\Delta^2} &= -\frac{a\omega_3(\lambda + (1-\lambda)P_V)}{\frac{\beta}{\rho} - a(\omega_1 + \omega_2)(1-\lambda)} \\ P_{\Delta^1} &= \frac{a\omega_2(\lambda + (1-\lambda)P_V)) - ac(1-\lambda)\omega_2 P_{\Delta^2}}{-\frac{\beta}{\rho} - ac(1-\lambda)\omega_1}\end{aligned}$$

where c solves

$$\frac{\rho^2}{\beta^2}\sigma^2_\Theta(1-c)(1-a^2c) - cb^2_V\left(\frac{a\lambda}{-\frac{\beta}{\rho\Omega} - a(1+\lambda)} - P_{\Delta^1} - P_{\Delta^2}\right)^2 = 0$$

Q.E.D

Proof to Proposition 19

From Proposition 13, $P_\Theta = \frac{\rho}{\beta}$. Hence, if $\beta < 0$ and $\rho > 0$, $P_\Theta < 0$. $P_\Theta < 0$suggests that if there is a positive shock in noise traders' supply, price should decrease in response.

Then we only need to check the sign of P_Δ.

First, we prove that if $P_\Delta < 0$, then $P_V < 0$. Suppose not, $P_V > 0$.

From Proposition 13, $P_V = \frac{a\lambda}{-\frac{\beta}{\rho\Omega} - a(1-\lambda)}$, the sign of P_V is determined by the sign of $-\frac{\beta}{\rho\Omega} - a(1-\lambda)$. To see this, if $P_V > 0$, that is, $-\frac{\beta}{\rho\Omega} - a(1-\lambda) > 0$. We must have $\Omega(-\frac{\beta}{\rho\Omega} - a\frac{\omega_1}{\Omega}(1-\lambda)c) > 0$ and $a\omega_2(\lambda + (1-\lambda)P_V) > 0$. From Proposition 13, P_Δ is

$$\begin{aligned} P_\Delta &= \frac{a\omega_2(\lambda + (1-\lambda)P_V)}{-\frac{\beta}{\rho} - a\omega_1(1-\lambda)c} \\ &= \frac{a\omega_2(\lambda + (1-\lambda)P_V)}{\Omega[-\frac{\beta}{\rho\Omega} - a\frac{\omega_1}{\Omega}(1-\lambda)c]} \end{aligned} \tag{A.31}$$

Recall $\Omega = \omega_1 + \omega_2$ and both ω_1 and ω_2 are positive. Hence, $\frac{\omega_1}{\Omega} < 1$. Therefore, both numerator and denominator are positive. Hence, $P_\Delta > 0$. That is, to have a $P_\Delta < 0$, P_V should be negative.

If P_V is negative, i.e., $-\frac{\beta}{\rho\Omega} - a(1-\lambda) < 0$.

There are two possible scenarios: first, the numerator in Equation A.31 is negative, i.e., if $P_V < -\frac{\lambda}{1-\lambda}$. The denominator of Equation A.31, $(-\frac{\beta}{\rho\Omega} - a\frac{\omega_1}{\Omega}(1-\lambda)c)$, should be positive in order for $P_\Delta < 0$. That is, $a\frac{\omega_1}{\Omega}(1-\lambda)c < -\frac{\beta}{\rho\Omega} < a(1-\lambda)$. Notice that this requires that $\frac{\omega_1}{\Omega}c < 1$. Because $\frac{\omega_1}{\Omega} < 1$, it is sufficient to have $c < 1$. From the proof to Proposition 13, we know that $0 < k < 1$ can guarantee $0 < c < 1$.

Second, the numerator in Equation A.31 is positive, i.e., if $0 > P_V > -\frac{\lambda}{1-\lambda}$. However, it is not possible. To see that,

$$\begin{aligned} P_V + \frac{\lambda}{1-\lambda} &= \frac{a\lambda(1-\lambda) + \lambda[-\frac{\beta}{\rho\Omega} - a(1-\lambda)]}{[-\frac{\beta}{\rho\Omega} - a(1-\lambda)][1-\lambda]} \\ &= \frac{-\lambda\frac{\beta}{\rho\Omega}}{[-\frac{\beta}{\rho\Omega} - a(1-\lambda)][1-\lambda]} \end{aligned} \tag{A.32}$$

We know that $\beta < 0$, $\rho > 0$, $\Omega > 0$ and $\lambda > 0$. Hence, the numerator of Equation A.32 is positive. For the denominator, $[-\frac{\beta}{\rho\Omega} - a(1-\lambda)] < 0$ because of $P_V < 0$ and $[1-\lambda] > 0$. Thus, the denominator of Equation A.32 is negative. That is, $P_V + \frac{\lambda}{1-\lambda} < 0$. In other word, $(\lambda + (1-\lambda)P_V) < 0$. That is, only the first scenario is possible.

Q.E.D

A.2 Appendix B: Derivations

Derivation of Equation 1.7

We start with Z_t. By definition, $Z_t = \frac{\widehat{f}_t}{\widehat{f}_{t-1}}$, which can be written as

$$Z_t = 1 + \frac{\tau_\epsilon}{\tau_t}\left(\frac{S_t^A + S_t^B}{\widehat{f}_{t-1}} - 2\right) \tag{A.33}$$

Note that $\frac{1}{Z_{t-1}} = \frac{\widehat{f}_{t-2}}{\widehat{f}_{t-1}}$, which can be written as

$$\frac{1}{Z_{t-1}} = \frac{\tau_{t-1}}{\tau_{t-2}} - \frac{\tau_\epsilon}{\tau_{t-2}\widehat{f}_{t-1}}(S_{t-1}^A + S_{t-1}^B)$$

which can be used to solve $\frac{1}{\widehat{f}_{t-1}}$

$$\frac{1}{\widehat{f}_{t-1}} = \frac{\left(\frac{\tau_{t-1}}{\tau_\epsilon} - \frac{\tau_{t-2}}{\tau_\epsilon}\frac{1}{Z_{t-1}}\right)}{(S_{t-1}^A + S_{t-1}^B)}.$$

Substitute Equation A.34 into Equation A.33, we get the expression.

Derivation of Equation 2.5

In order to get Equation 2.5, we need to rewrite Equation 2.4 into

$$P_t - \rho P_{t-1} = -\frac{\Theta_t}{\omega(1+r)}\Delta + \frac{\Delta}{1+r}E[\lambda V_{t+1} + (1-\lambda)(P_{t+1} - \rho P_t)]$$

where Δ is a constant. Hence,

$$\left(1 + \frac{\Delta\rho(1-\lambda)}{1+r}\right)P_t = \rho P_{t-1} - \frac{\Theta_t}{\omega(1+r)}\Delta + \frac{\Delta}{1+r}E[\lambda V_{t+1} + (1-\lambda)P_{t+1}]$$

In order to match the coefficients of Equation 2.4, we need to have:

$$\begin{aligned} \frac{\rho}{\Delta} &= -\frac{\beta}{\omega(1+r)} \\ \Delta &= 1 + \frac{\Delta\rho(1-\lambda)}{1+r} \end{aligned}$$

The first line implies that $\Delta = -\frac{\beta}{\rho\omega(1+r)}$ and substitute it into second line. Then we have:

$$(1-\lambda)\rho^2 + (1+r)\omega\rho + \beta = 0.$$

Bibliography

Ait-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process. *Annals of Statistics*, **37**, 184–222.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 579–625.

Andersen, T., Bollerslev, T., and Diebold, F. (2007). Roughing it up: including jump components in the measurement modeling and forecasting of return volatility. *Review of Economices and Statistics*. forthcoming.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, **2**, 1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, **4**, 1–30.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*.

Bomfim, A. N. (2001). Heterogeneous forecasts and aggregate dynamics. *Journal of Monetary Economics*, **47**, 145–161.

Brock, W. A. and LeBaron, B. (1996). A dynamic structural model for stock return volatility and trading volume. *Review of Economics and Statistics*, **78**, 94–110.

Brock, W. A., Hommes, C., and Wagener, F. (2005). Evolutionary dynamics in markets with many trader types. *Journal of Mathematical Economics*, **41**, 7–42.

Cabrales, A. and Hoshi, T. (1996). Heterogeneous beliefs, wealth accumulation, and asset price dynamics. *Journal of Economic Dynamics and Control*, **20**, 1073–1100.

Christian, D. and Jia, M. (2005). Optimal trading frequency for active asset management: Evidence from technical trading rules. *Journal of Asset Management*, **5**, 305–326.

Dacorogna, M., Gençay, R., Muller. U. A., Pictet, O., and Olsen, R. (2001). *An Introduction to High-Frequency Finance*. Academic Press.

de Fontnouvelle, P. (2000). Information dynamics in financial markets. *Macroeconomic Dynamics*, **4**, 139–169.

DeGroot, M. (1970). *Optimal statistic decisions*. McGraw Hill Press, New York.

Dungey, M., McKenzie, M., and Smith, V. (2007). Empirical evidence on jumps in the term structure of the U.S. treasury market. manuscript.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Engle, R. F. (2000). The econometrics of ultra high frequency data. *Econometrica*, **68**, 1–22.

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, **102**, 1349–1362.

Foster, F. and Viswanathan, S. (1993). The effect of public information and competition on trading volume and price volatility. *The Review of Financial Studies*, **6**, 23–56.

Gençay, R., Selçuk, F., and Whitcher, B. (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press, New York.

Granger, C. W. J. and Machina, M. J. (2006). Structural attribution of observed volatility clustering. *Journal of Econometrics*, **135**, 15–29.

Grossman, S. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, **70**, 1393–408.

Haan, W. D. and Spear, S. (1998). Volatility clustering in real interest rates: theory and evidence. *Journal of Monetary Economics*, **41**, 431–453.

Harris, L. (2003). *Trading and exchange*. Oxford University Press.

Hauser, S., Levy, A., and Yaari, U. (2001). Trading frequency and the efficiency of price discovery in a non-dealer market. *European Journal of Finance*, **7**, 187–197.

Hommes, C. H. (2006). Heterogeneous agent models in economics and finance. In L. Tesfatsion and K. Judd, editors, *Hand-book of Computational Economics, Volume 2: Agent-Based Computational Economics*, pages 1109–1186. Elsevier Science B.V.

Hommes, C. H. (2008). Interacting agents in finance. In L. Blume and S. Durlauf, editors, *the New Palgrave Dictionary of Economics*. Palgrave Macmillan.

Hong, H., Scheinkman, J., and Xiong, W. (2006). Asset float and speculative bubbles. *Journal of Finance*, **31**, 307–327.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implicatious for stock market efficiency. *Journal of Finance*, **48**, 65–91.

Jiang, G. J. and Oomen, R. C. (2008). Testing for jumps when asset prices are observed with noise - a swap variance approach. *Journal of Econometrics*, **144**, 352–370.

Jiang, G. J., Lo., I., and Verdelhan, A. (2008). Information shocks, jumps and price discovery: evidence from the U.S treasury market. Working Paper.

Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance*, **59**, 227–260.

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

Lee, S. S. and Mykland, P. A. (2008). Jumps in finacial markets: a new nonparametric test and jump dynamics. *Review of Financial Studies*, **21**, 2536–2563.

Makarov, I. and Rytchkov, O. (2007). Forecasting the forecasts of others: implication for asset pricing. Working Paper.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, **36**, 307–332.

McNulty, M. S. and Huffman, W. E. (1996). Market equilibria with endogenous, hierarchical information. *Journal of Economic Dynamics and Control*, **20**, 607–626.

Piazzesi, M. (2003). Bond yields and the federal reserve. *Journal of Political Economy*, **113**, 311–344.

Timmermann, A. (2001). Structural breaks, incomplete information and stock prices. *Journal of Business and Economics Statistics*, **19**, 299–314.

Townsend, R. M. (1983). Forecasting the forecasts of others. *Journal of Political Economy*, **91**, 546–587.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika*, **82**, 385–397.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Working Paper.

Zhang, L., Ait-Sahalia, Y., and Mykland, A. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, **100**, 1394–1411.