Duration, Attention, and Long Memory*

Fei Chen†

University of Pennsylvania

October 19, 2010

Abstract

This paper develops a class of models for the analysis of financial durations. We first establish a mixture of exponentials representation for general point processes. Based on the representation, we propose a new model, called the Markov switching multifractal duration (MSMD) model. We show the MSMD can explain most stylized facts of financial durations, especially the long memory feature. Extensive empirical study shows MSMD can predict long horizon durations better than the ACD model, which confirms that MSMD can explain long range dependence in financial durations.

*The author is deeply indebted to Francis X. Diebold, Frank Schorfheide, and Kyungchul Song for their invaluable inspiration and support. For helpful comments, the author thank Aureo de Paula. All the errors are the author's.

†Address: Fei Chen, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104; Email: chenf@sas.upenn.edu Tel: 610.742.7581

1 Introduction

The last two decades have witnessed a growing interest in theoretically and empirically modeling of the ultimate high-frequency financial data. A salient feature of these intra-day tick-by-tick data is that transactions are irregularly spaced in time.

Many empirical studies take these irregular durations as exogenous sampling scheme and aggregate the data to some fixed interval in accordance with the usual low-frequency data such as daily or weekly data. Such temporal aggregation facilitates empirical analysis, but also brings two issues. First, the aggregation will lose information and introduce unknown bias from a statistic view. Aït-Sahalia and Mykland (2003) discuss the effects of sampling randomness and sampling discreteness when estimating continuous time processes. Second, there is no theory guidance on how to choose length of the fixed interval. Bandi and Russell (2008) discuss how to choose the optimal sampling intervals when estimating realized volatilities.

More importantly, the duration is an endogenous economic variable with information content. It reflects the speed of the information flow on the financial market, see Hasbrouck (1999). Easley and O'Hara (1992) give an economic interpretation of the durations from a market microstructure view.

In this paper, we suggest a different, behavior-based interpretation and offer a new approach to model these irregular durations. We start with general point processes, and establish a mixture of exponentials representation for the durations from the martingale theory. Based on this representation, we develop a new duration model, the Markov switching multifractal duration (MSMD) model, which can be used for designing trading strategies and intra-day risk management.

The first econometric model to explore the information content of trading durations is the Autoregressive Conditional Duration (ACD) model proposed by Engle and Russell (1998). The ACD model assumes a multiplicative error form, where the duration is the product of the conditional mean and the error[1]. Such a specification has two components: the dynamics of the mean and the distribution of the error. Engle and Russell assume a GARCH-type dynamics with iid exponential or iid Weibull errors.

While GARCH-type dynamics explains the clustering effect in durations, i.e., short (long) durations tending to be followed by short (long) durations; the basic ACD model can not fully capture other stylized facts found in empirical studies. These include: overdispersion[2], the standard deviation being greater than the mean; long memory, autocorrelations decreasing hyperbolically; strong nonlinearities in the dynamics; heavy tail. See Gagliardini and Gourièroux (2008) for a summary.

Numerous extensions of the basic ACD models have been developed in the literature[3]. Various extensions have used different functional forms for the error distribution: generalized gamma distribution by Zhang et al. (2001),

[1]See Engle (2002) and Engle and Gallo (2006) for more details of multiplicative error model.

[2]Giot (2000) reports that some volume duration series (durations for volume to reach some threshhold) can exhibit underdispersion. This case will not be considered in this paper. All durations between transactions and price changes (quote changes) show overdispersion.

[3]For a partial list, see Lunde (1999), Jasiak (1999), Grammig and Maurer (2000), Bauwens and Giot (2000), Zhang et al. (2001), De Luca and Zuccolotto (2003), De Luca and Gallo (2004), Bauwens and Veredas (2004), Ghysels et al. (2004), Drost and Werker (2004), Fernandes and Grammig (2006), Meitz and Teräsvirta (2006), Hujer and Vuletic (2007), Sun et al. (2008), Deo et al. (2010).

Burr distribution by Fernandes and Grammig (2006), Weibull distribution by Deo et al. (2010), etc. Moreover, Drost and Werker (2004) even challenge the assumption of iid errors and use semiparametric alternatives.

On the one hand, these distributions are very flexible. On the other hand, the choice of a particular distribution is arbitrary, and is mostly based on convenience, familiarity and analytical tractability. Heckman and Singer (1984) give an example that two duration models with different error distributions can have the same statistical property. Though the example is for the single spell duration model, similar problem could exist for dynamic duration models. The arbitrarily chosen error distribution function is in fact very important. It not only has direct impacts on trading strategies and intra-day risk management, but also has serious implications on models that try to link durations and volatilities, e.g., Ghysels and Jasiak (1998), Engle (2000), Grammig and Wellner (2002).

This paper adopts a different approach. The trading process is a marked point process (PP) on the time line[4]. A PP can be represented by a series of durations, but the driving force underlying the durations is the continuous-time intensity process. Intensity-based modeling has already been applied to multivariate financial PPs, e.g., Russell (1999), Bauwens and Hautsch (2006), Bowsher (2007). We begin with the intensity process and use a time deformation method to build the link between intensities and durations. Our contribution is to establish a mixture of exponentials representation for durations. In this representation, durations can be written as iid exponential errors divided by mean intensities.

[4]See section 2 for the definition.

The MSMD model is built on the mixture of exponentials representation. The backbone of the MSMD model is the MSM intensity. We interpret trading intensity as a measure of the aggregate attention level from heterogeneous investors[5]. Different information events on the financial market will draw attention shocks that will last for different time scales. At any time, the total attention level is the aggregation of all the shocks from different time scales. The mathematical structure of MSM is particularly suitable to model such a situation.

We show that the MSMD model can explain most existing stylized features, especially the long memory feature. The long memory feature is an important property of financial time series. A lot of research interest is attracted to long memory in volatilities[6]. Recently, there is a view that long memory in volatilities is from long memory in durations, see Deo et al. (2009), Deo et al. (2010).

To validate the MSMD model, we implement extensive empirical study. Twenty stocks are randomly selected from the S&P 100 index. We estimate the MSMD model for all the twenty stocks and find that when the number of intensity components is seven, the MSMD model can describe the data well. We then compare the MSMD model with the ACD model. We do both in sample fitting and out of sample forecasting comparison. Both results are in favor of the MSMD model. A striking finding is that the long horizon prediction performance of the MSMD is much better than the ACD, which

[5] See section 3 for detail.

[6] See, among many others, Ding et al. (1993), Bollerslev and Mikkelsen (1996), Baillie et al. (1996), Comte and Renault (1996), Breidt et al. (1998), Andersen et al. (2001), Deo et al. (2006), Corsi (2009).

suggests the MSMD can be used to measure liquidity risk.

The rest parts of this paper is organized as follow. In section 2, we introduce notions of PPs and derive the mixture of exponentials representation. In section 3, we discuss economic interpretation of trading intensities. In section 4, we explain specifications of the MSMD model. In section 5, we show properties of the MSMD model. Section 6 is empirical work. Section 7 concludes.

2 Point Processes and Mixture of Exponentials Representation

The purpose of this section is to derive the mixture of exponentials representation of PPs. To this end, we introduce basic concepts and tools of PPs. There are two fundamental approaches to characterize PPs, random measure and conditional intensity[7]. We only introduce the conditional intensity. The conditional intensity is a powerful tool for evolutionary PPs on the time line, because it introduces martingale-based methods to PPs.

2.1 Notation and Definition

A simple PP on $(0, \infty)$ is a sequence of nonnegative random variables $\{t_i\}_{i \in 1,2,\ldots}$ defined on some probability space (Ω, F, P), satisfying $0 < t_1 < t_2 < \cdots$, where t_i is the instant of the i-th occurrence of an event. Associated with each t_i, there could be some exogenous random variables. These variables are

[7]Textbook treatments of these two approaches can be found in Brémaud (1981), Karr (1991), Daley and Vere-Jones (2003), and Daley and Vere-Jones (2007).

called marks of the PP. In a trading process, the events are financial transactions. The marks could be volume, price, bid-ask quotes or other variables coming with each transaction[8].

A PP may also be represented via its associated counting process $N(t)$, where $N(t) = \sum_{i\geq 1} 1(t_i \leq t)$ is the number of events happened till time t. The internal history $\{F_t^N\}_{t\geq 0}$ of a PP is given by the σ-algebra generated by the observed past of the process, namely $F_t^N = \sigma(N(s) : 0 \leq s \leq t)$. A history F_t is a more general σ-algebra which could contain information about some exogenous variables, e.g., the marks. The internal history is the smallest history, $F_t^N \subseteq F_t$. Obviously $N(t)$ is F_t-adapted.

Let $\lambda(t)$ be a scalar, positive F_t-predictable process[9], then $\lambda(t)$ is called the F_t-conditional intensity of $N(t)$, if

$$E[N(s) - N(t)|F_t] = E[\int_t^s \lambda(u)du|F_t] \tag{1}$$

holds almost surely for all t, s with $0 \leq t \leq s$[10]. The definition of conditional intensity given by (1) is abstract. A more intuitive understanding of the

[8]By the definition, the trading process is usually not a simple PP. Multiple transactions at the same second are observed in TAQ database. It is believed that the simultaneous trades executed at the same second come from the same trader who has split a big order block in small blocks. We only keep one trade for each second. The thinned trading process is a simple PP

[9]See appendix A3 of Daley and Vere-Jones (2003) for definition of F_t-predictable. Sufficient conditions for $\lambda(t)$ to be F_t-predictable are $\lambda(t)$ is adapted to F_t, and the sample paths of $\lambda(t)$ are left continuous with right hand limits.

[10]For existence of $\lambda(t)$, see chapter 7 of Daley and Vere-Jones (2003) and chapter 14 of Daley and Vere-Jones (2007)

intensity can be got by letting $s \downarrow t$ in (1).

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E(N(t+\Delta t) - N(t)|F_{t-})$$
$$= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(N(t+\Delta t) - N(t) = 1|F_{t-}) \quad (2)$$

The above equation is not strict, but it shows the similarity between conditional intensity and hazard function.

The compensator of a PP is defined as $\Lambda(t) = \int_0^t \lambda(s)ds$. Let $M(t) = N(t) - \Lambda(t)$, then the process $M(t)$ is a martingale. One important result of the martingale-based PP theory is the random change of time theorem.

2.2 Random Change of Time

The random change of time theorem gives a method to transform non-Poisson processes to a homogeneous Poisson process. Though it is introduced as a pure mathematical method, it has an intuitive economic interpretation. In an ideal world without information flow, the trading process is a homogeneous Poisson process, i.e. the trading intensity is constant. In reality, the randomly arriving information flow distorts the trading intensity, and the trading process evolves on some operational or economic time scale that differs from the calendar or clock time. The random change of time method gives a functional mapping between the clock time and the economic time, which is so called time deformation.

Time deformation has been widely used in economic research, see, e.g. Clark (1973), Stock (1988), Carr and Wu (2004). The random change of time theorem gives a subordinator of a Poisson process.

Theorem 1 *Let $N(t)$ be a simple point process on $(0, \infty)$, adapted to filtration F_t. Suppose that $N(t)$ has the F_t-conditional intensity $\lambda(t)$ that satisfies:*

$$\int_0^\infty \lambda(t)dt = \infty.$$

For any $t \geq 0$, define the F_t-stopping time τ_t as the solution to:

$$\int_0^{\tau_t} \lambda(s)ds = t$$

then the point process $\tilde{N}(t) = N(\tau_t)$ is a homogenous Poisson process with intensity $\lambda = 1$.

Proof See Theorem T16, p.41, Brémaud (1981). ∎

The only condition for the theorem to hold is $\int_0^\infty \lambda(t)dt = \infty$. That is to say one can always expect more occurrences of the events in the future. This condition is satisfied by any trading process.

The theorem is well known in PP literature. But previous research has emphasis on using the theorem to construct goodness-of-fit test for the intensity process, e.g., chaper 7 of Daley and Vere-Jones (2003), Bowsher (2007). We first use it to establish the mixture of exponentials representation of PPs.

2.3 Mixture of Exponentials Representation

We will derive the mixture of exponentials representation by using the time deformation function, i.e. τ_t.

Let $\tilde{t}_i$ and t_i denote the time of ith event in the operational and clock time respectively. $\epsilon_i = \tilde{t}_i - \tilde{t}_{i-1}$ and $d_i = t_i - t_{i-1}$ are ith duration in different time scale. In the operational time scale, the trading process is a homogeneous Poisson process, so the distribution of the durations is iid exponential. That means $\epsilon_i \sim i.i.d.Exp(1)$. By the definition of τ_t, we have $\epsilon_i = \tilde{t}_i - \tilde{t}_{i-1} = \Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda(s)ds$. Let $\lambda_i = \Lambda(t_{i-1}, t_i)/d_i$ be the mean intensity, then we can write

$$d_i = \frac{\epsilon_i}{\lambda_i} \tag{3}$$

This is the mixture of exponentials representation, which is different from the multiplicative error form of the ACD models. Instead of modeling the conditional mean, we need to model the mean intensity λ_i.

3 Economic Interpretation of Trading Intensity

In last section, we establish the mixture of exponentials representation for durations. For a complete duration model, we need to specify the mean intensity λ_i. Before introducing the specification, we discuss some economic interpretation of trading intensity to motivate our specification.

The timing of certain type of trades and orders is a signal that can reveal the state of the market and plays an important role in the market microstructure theory, see, among others, Admati and Pfleiderer (1988), Easley and O'Hara (1992). Engle and Russell (1998) use the ACD model to test these microstructure models. Here we discuss a question in the other direction. What implications can these models offer on the trading intensities?

These structural models of the price discovery process usually assume there are two types of traders, informed and uninformed. The informed trader will arrive only when an information event happens. In other words, information flow affects intensities. But information doesn't trigger trading automatically. Investors have limited attention[11]. A investor has to allocate her attention to process all kinds of information then make trading decision. Both theoretical and empirical studies show that attention plays a significant role in investors' behavior, see, for example, Huberman (2001), Peng and Xiong (2006), Huang and Liu (2007), Barber and Odean (2008).

At the aggregate level, the more attention is paid to a stock, the more likely a trade will happen. According to equation (2), this probability is the trading intensity. Thus the intensity is a measure for the aggregate attention level. This interpretation is very intuitive and is used by some researchers, e.g., Corwin and Coughenour (2008) use the number of transactions to approximate individual NYSE specialist's attention level.

Different information events can draw attention shocks with different persistence. For example, under current financial crisis, interest rate is low, so more risk-averse investors will withdraw their money from bank accounts and pay their attention to the stock market. This attention shock will last for months. During the crisis, there happens the oil spill in Mexico Gulf, which will attract investors' attention to oil industry. This shock will last for weeks. If there is some local news coverage about a gasoline company during the oil spill period, that is another attention shock which can last for days.

The above is just a simple example to illustrate that the attention or

[11]A large body of psychological literature shows that humans have limited attention. Limited attention is also related to the concept of "bounded rationality".

intensity shocks have a cascade structure. Numerous intensity shocks exist on the financial market. Some are observable, like public news. Some are unobservable, like private information. The shocks are spread through different social network, e.g., Hong et al. (2005) find that mutual fund managers in the same city are likely to trade the same stocks. At any time, the total intensity is the aggregation effect of all shocks. We use the MSM structure proposed by Calvet and Fisher (2004) to model such cascade shocks.

4 Markov Switching Multifractal Duration

One basic specification of the MSMD model is given in (3), where ϵ_i is specified as iid exponential distribution. We now give the specification of the intensity λ_i.

We assume the intensity has $\bar{k}$ components. Each component represents a shock at a particular frequency. All components contribute to the intensity through a *multiplicative* effect[12]. More precisely, we model λ_i as

$$\lambda_i = \lambda \prod_{k=1}^{\bar{k}} M_{k,i} \tag{4}$$

where λ is a positive constant. $M_{1,i}$, $M_{2,i}$, ..., $M_{\bar{k},i}$ are positive intensity components. The components are statistically independent with each other at any time. It is convenient to define the trading intensity state vector at time i as $M_i = (M_{1,i}, M_{2,i}, \dots M_{\bar{k},i})$.

[12]This multiplicative effect could become additive effect by taking logarithm. Then we can take the total intensity as a superposition of the $\bar{k}$ components with different frequencies. This is similar to the fourier series expansion, but we don't have the usual orthogonal condition here.

For each $k \in \{1, 2, 3, ..., \bar{k}\}$, the dynamics of the component $M_{k,i}$ is a Markov renew process. At time i, $M_{k,i}$ is either renewed, namely drawn from a fixed distribution M with probability γ_k, or remains its previous value $M_{k,i-1}$ with probability $1 - \gamma_k$.

The fixed distribution of M is the same for different components. A draw from M is the magnitude of a shock. Only positive shock is allowed, so M has a positive support $M > 0$. To prevent the shocks from exploding, we set $E(M) = 1$.

The renew probability γ_k is specified as

$$\gamma_k = 1 - (1 - \gamma_1)^{b^{(k-1)}} \tag{5}$$

where $\gamma_k \in (0, 1)$ and $b \in (1, \infty)$. This specification is introduced in Calvet and Fisher (2001) in connection with the discretization of a Poisson arrival process. The value of γ_k will determine the average lifetime or persistence of a $M_{k,i}$ shock. The larger the γ_k is, the shorter average lifetime the $M_{k,i}$ shock will have. Large k component stands for high frequency shock. Small k component stands for low frequency shock. An important feature of this specification is that all shocks, low frequency or high frequency, have stochastic lifetime.

Equations (3), (4) and (5) plus specification for M define a stochastic duration model, thus the MSM duration model.

5 Model Properties

We will establish some properties of the MSMD model, and show the model can explain most stylized features of financial durations.

5.1 Geometric Ergodicity

For duration models, the important properties are stationarity, ergodicity and finite higher-order moments. The strict stationarity of the MSM duration is obvious since each intensity component is independent and stationary. The existence of finite higher-order moments depends on the moment properties of M. For example, if we take M as binomial distribution, which we will do in our empirical study, then every finite moment of d_i exists. Here we give the ergodic property.

Proposition 2 *The MSM duration* $\{d_i\}$ *is geometrically ergodic.*

Proof From the definition, d_i is a hidden Markov model with the intensity vector M_i as the Markov chain. By Proposition 4 of Carrasco and Chen (2002), It is enough to show M_i is geometrically ergodic.

Let the support of M be S. M_i is a Markov chain on $S^{\bar{k}}$. Since each component $M_{k,i}$ is independent, we need to show $M_{k,i}$ is geometrically ergodic.

First we show $M_{k,i}$ is φ-irreducible T-chain. Take φ as Lebesgue measure μ^{Leb} on S. The μ^{Leb}-irreducibility is immediate from the assumption of positive densities for M. The transition kernel of $M_{k,i}$ is $P(x, A) = \gamma_k \int_A dF + (1-\gamma_k) 1_x(A)$. Let $T(x, A) = \gamma_k \int_A dF$, then $T(x, A)$ is a nontrivial continuous component of $P(x, A)$, by Proposition 6.2.4 of Meyn and Tweedie (1993), $M_{k,i}$ is a T-chain.

This implies that all compact sets in S are petite. We can choose any compact set C in S with positive probability measure as a test set. It is easy to check that $M_{k,i}$ satisfies conditional (ii) of Proposition 15.0.1 of Meyn and Tweedie (1993), then $M_{k,i}$ is geometrically ergodic. $\blacksquare$

5.2 Clustering Effect

The clustering effect is implied by some market microstructure models, e.g. Easley and O'Hara (1992). Those models suggest when information events happen, trades will cluster together. And the clustering effect is found in empirical studies.

The MSMD model can not only explain the clustering effect, i.e., when the highest frequency intensity component draws a large value, short durations will happen together. But also it predicts there is clustering effects at all time scales.

In Figure 1, we draw the counts or the number of transactions in 2 minutes, 5 minutes, 10 minutes and 30 minutes. We find clustering in all the time scales. This confirms the prediction of the MSMD model.

5.3 Nonlinearity

There is strong nonlinearity in the duration dynamics. Zhang et al. (2001) use their threshold ACD model to identify multiple structural breaks in the transaction duration data considered, and they find those break points matched nicely with real economic events. This is in agreement with our discussion in last section. Different events will draw different shocks, therefore cause regime switches. The MSMD is a Markov switching model. It has the non-linearity built in.

5.4 Overdispersion

The overdispersion property can be observed in all the samples we used for empirical study. Let $\mu_d = E(d_i)$, $\sigma_d^2 = \mathrm{Var}(d_i)$. We will have

Proposition 3

$$\sigma_d > \mu_d$$

Proof By definition, we have $\mu_d = E(d_i) = E(\frac{1}{\lambda_i})E(\epsilon_i) = E(\frac{1}{\lambda_i})$ and $\sigma_d^2 = \mathrm{Var}(d_i) = E(d_i^2) - [E(d_i)]^2 = E(\epsilon_i^2)E(\frac{1}{\lambda_i^2}) - [E(\frac{1}{\lambda_i})]^2$. it is easy to check $E(\epsilon_i^2) = 2$, so

$$\sigma_d^2 = 2E(\frac{1}{\lambda_i^2}) - [E(\frac{1}{\lambda_i})]^2$$

by Jensen's inequality $[E(\frac{1}{\lambda_i})]^2 < E(\frac{1}{\lambda_i^2})$, we get $\sigma_d > \mu_d$ ∎

5.5 Long Memory Feature

The duration autocorrelations decrease slowly with horizon. In Figure 2, we show four duration autocorrelations. A visual check will confirm the slowly decaying of autocorrelations.

Previous research doesn't pay much attention to the long memory feature. One reason is that the sum of parameters estimated in ACD models is nearly 1. This can explain some persistence of the durations. But the ACD models only have short memory. Recently, the long memory feature is in deed confirmed by the semiparametric analysis of Deo et al. (2010).

The autocorrelation function of durations is $\rho(n) = Corr(d_i, d_{i+n})$. Let $\alpha_1 < \alpha_2$ denote two arbitrary numbers in the open interval $(0, 1)$. The set of integers $I_{\bar{k}} = \{n : \alpha_1 \log_b(b^{\bar{k}}) \leq \log_b n \leq \alpha_2 \log_b(b^{\bar{k}})\}$ contains a broad collection of lags.

Proposition 4 *The autocorrelation of durations satisfies*

$$\sup_{n \in I_{\bar{k}}} \left| \frac{\ln \rho(n)}{\ln n^{-\delta}} - 1 \right| \to 0 \qquad as \quad \bar{k} \to +\infty$$

where $\delta = \log_b \mathbb{E}(M) - \log_b\{[\mathbb{E}(M^{1/2})]^2\}$

Proof By definition $Corr(d_i, d_{i+n}) = E(d_i d_{i+n}) - E(d_i)E(d_{i+n})$.

We calculate the first term:

$$E(d_i d_{i+n}) = E(\frac{\epsilon_i \epsilon_{i+n}}{\lambda_i \lambda_{i+n}}) = E(\lambda_i^{-1} \lambda_{i+n}^{-1}) = \prod_{k=1}^{\bar{k}} E(M_{k,i}^{-1} M_{k,i+n}^{-1}).$$

The last equality is valid by the independence of each component. We use iterated expectation to calculate the last term,

$$E(M_{k,i}^{-1} M_{k,i+n}^{-1}) = E[M_{k,i}^{-1} E(M_{k,i+n}^{-1} | M_{k,i}^{-1})],$$

where

$$E(M_{k,i+n}^{-1} | M_{k,i+n-1}^{-1}) = M_{k,i+n-1}^{-1}(1 - \gamma_k) + E(M^{-1})\gamma_k,$$

and

$$E(M_{k,i+n}^{-1} | M_{k,i+n-2}^{-1}) = M_{k,i+n-2}^{-1}(1 - \gamma_k)^2 + E(M^{-1})\gamma_k(1 - \gamma_k) + E(M^{-1})\gamma_k.$$

So we can get

$$E(M_{k,i+n}^{-1}|M_{k,i}^{-1}) = M_{k,i}^{-1}(1-\gamma_k)^n + E(M^{-1})[1-(1-\gamma_k)^n],$$

and

$$\begin{aligned} E(M_{k,i}^{-1}M_{k,i+n}^{-1}) &= E(M^{-2})(1-\gamma_k)^n + [E(M^{-1})]^2[1-(1-\gamma_k)^n] \\ &= [E(M^{-1})]^2[1+a(1-\gamma_k)^n] \end{aligned}$$

where $a = E(M^{-2})[E(M^{-1})]^{-2} - 1$.

Now we calculate the second term $E(d_i)E(d_{i+n}) = [E(d_i)]^2 = [E(\frac{1}{\lambda_i})]^2 = \prod_{k=1}^{\bar{k}}[E(M_{k,i}^{-1})]^2 = \prod_{k=1}^{\bar{k}}[E(M^{-1})]^2 = [E(M^{-1})]^{2\bar{k}}$. We already have $\sigma_d^2 = 2E(\frac{1}{\lambda_i^2}) - [E(\frac{1}{\lambda_i})]^2 = 2\prod_{k=1}^{\bar{k}} E(M^{-2}) - \prod_{k=1}^{\bar{k}}[E(M^{-1})]^2 = [E(M^{-1})]^{2\bar{k}}[2(1+a)^{\bar{k}} - 1]$, thus we get

$$\rho_n = Corr(d_i, d_{i+n}) = \frac{\prod_{k=1}^{\bar{k}}[1+a(1-\gamma_k)^n] - 1}{2(1+a)^{\bar{k}} - 1}$$

The rest of the proof just follows Proposition 1 of Calvet and Fisher (2004).

∎

5.6 Discussion

The traditional method to generate long memory is the fractional integration (FI) or $I(d)$ model, i.e., fractional difference operator acting on iid shocks. It is introduced to the econometrics literature by Granger and Joyeux (1980) as a parsimonious empirical method. In FI models, every shock has a long-lived effect. This is in contrast with $I(0)$ and $I(1)$ cases. In a stationary ARMA model, i.e., $I(0)$, every shock is transitory. In a non-stationary random walk model, i.e., $I(1)$, every shock is permanent. The FI process seems to provide a natural way to fill the gap between $I(0)$ and $I(1)$ processes. But in FI models, every shock still decay at the same rate. It introduces artificial mixing between long- and short-term dependence, which is illustrated by Comte and Renault (1998). It also blur the distinction between stationary and nonstationary processes.

Jasiak (1999) proposes the fractional integrated ACD model to capture long memory in durations. But this model suffers from the non-existence of moments problem. The second moment of the FIACD model doesn't exist. It is not a long memory model in the usual sense, autocorrelations decaying hyperbolically.

In the MSMD model, different shocks have different persistence, which is closer to our intuition. Another well-known mechanism to generate long memory is by occasional regime switches, see Diebold and Inoue (2001). This is why some researchers think the long memory in duration is spurious. It comes from some regime switches and parameter instabilities. The MSM structure can accommodate this occasional regime switches case. For example, suppose in our sample period, the lowest frequency intensity component $M_{1,i}$ gets renewed for only 2 times, i.e., $M_{1,i}$ take only three values x_1, x_2, x_3. Then the sample period can be split into 3 regimes according to the value of $M_{1,i}$. If we set

$$N = \lambda x \prod_{k=2}^{\bar{k}} M_{k,i}$$

where x is the value of $M_{1,i}$, we see in different regimes, N have different mean[13]. This is the occasional regime switches case discussed in Diebold and Inoue (2001). The MSMD model naturally combines long memory and nonlinearity.

6 Empirical Studies

We explain the MSMD model in previous sections. Now we do empirical studies to show the MSMD is not some statistical artifact. To this purpose, we have to specify M. As is discussed in section 4, M should satisfy $M > 0$ and $E(M) = 1$. Following Calvet and Fisher (2004), we specify M as a binomial variable taking value m_0 and $2 - m_0$ with equal probability. The binomial MSMD model has four parameters

$$\phi = (m_0, \lambda, b, \gamma_{\bar{k}}) \in \mathbb{R}_+^4.$$

Twenty stocks from the S&P 100 index are randomly selected and equally divided into two groups, high trading group and low trading group, according to the number of transactions in the sample period. We estimate the binomial MSMD model for all the twenty stocks. We find the MSMD mode with seven components thus MSMD(7) can give a good description of the data.

We then run a horse race between the binomial MSMD(7) model and the ACD(1,1) model. We compare both in sample fitting and out of sample forecasting of the two competing models. For in sample fitting, we compare the likelihoods. For out of sample forecasting, we compare the mean square

[13] In regime 1, $E[N] = \lambda x_1$. In regime 2, $E[N] = \lambda x_2$. In regime 3, $E[N] = \lambda x_3$.

prediction errors. Forecasting at three horizons, 1-step, 5-step, and 20-step is implemented. A fixed scheme is chosen to compare forecasting, see Pagan and Schwert (1990) and West and McCracken (1998). We choose this scheme mainly because the estimation process involves computationally intensive numerical maximizations. We take 11000 observations as the maximum sample size for forecasting comparison and split the sample into two sets, a fitting set and a testing set. The fitting set has 10000 observations, and the later has 1000 observations. If the total observations are less than 11000 for some low trading stocks, we take roughly the last 1000 observation as testing set, previous observations as fitting set. Competing models are estimated only once on the fitting set and then the estimated parameters are used in forming predictions for observations in the testing set.

6.1 Data Description

The data for empirical study are the consolidated trades data extracted from TAQ database. The time period is from February 1, 1993 to February 26, 1993, which has 20 trading days. We only keep transactions during the open time, from 9:30 a.m. EST to 4:00 p.m. EST. All over night durations are omitted. Following Zhang et al. (2001), transactions in the opening period from 9:30 am to 10:00 a.m. are also deleted to remove the opening auction effect. There are still simultaneous trades and zero durations. Following Engle and Russell (1998), we delete the zero durations. Table 1 gives the symbol and company name of the twenty stocks.

6.2 Daily Seasonality

The raw durations have strong diurnal daily pattern, i.e. the average duration is short at morning opening time and afternoon close time, and long at noon time. This daily seasonality is documented by many empirical studies. There are several methods to remove the seasonality. We adopt the method used by Ghysels et al. (2004). The main step is to regress the logarithm of the raw duration on the indicator variables that indicate the time of day. A day is divided into 12 subperiods. Each subperiod is 30 minutes. We consider the regression

$$\log d_i = \sum_{k=1}^{12} a_k x_{ki} + \epsilon_t = a' x_i + \epsilon_i$$

where $x_{ki} = 1$, if time i belongs to the intraday subperiod k, and 0 otherwise. Then the seasonally adjusted series is defined by

$$\hat{d}_i = d_i \exp(-\hat{a}' x_i)$$

where $\hat{a}$ denotes the OLS estimator of a. The data from now on are all seasonally adjusted data.

6.3 Data Statistics

Table 2 and Table 3 show the summary statistics for both low and high trading groups. The stock of Merck & Co has the most, 54242 durations during the sampling period. While the stock of ALCOA has the least, 2989 durations. The longest duration is 76.83. The shortest duration is 0.02. All twenty stocks show overdispersion even after the seasonality adjusting.

For all stocks, duration mean is greater than duration median. Each stock's kurtosis is much bigger than 3, implying heavy tail distributions.

6.4 Estimation Method

Each intensity component can take only two values in the binomial MSMD model. For $\bar{k}$ components, there are $2^{\bar{k}}$ states of the intensity state vector M_i. For the finite number of states, we can easily get the likelihood of the model by standard filtering procedures.

We initiate the distribution of M_i with the ergodic distribution, then use Bayes' law to update the distribution of M_i, and compute the likelihood for each observation. MLE is used to estimate the parameters. Like other hidden Markov models, local maximums exist. Multiple initial conditions are tried to find the MLE estimation.

6.5 Model diagnostics

Several types of diagnostic tests have been proposed to evaluate the fast growing ACD models, see Li and Yu (2003), Fernandes and Grammig (2005), Meitz and Teräsvirta (2006), Chen and Hsieh (2010). Unfortunately, the MSMD model is a latent variable model. These tests can not be used. Instead, we use the information matrix test developed by White (1982). The test is based upon the asymptotic equivalence of the Hessian and outer product forms of Fisher's information matrix, when the model is correctly specified. The up-right elements of the information matrix (total 10 elements) are selected to form the test statistic S_{IM}. Note $S_{IM} \sim \chi^2_{10}$

6.6 Estimation Results

We present 5 estimation results: table 4 for Bank of America Corp., table 5 for Ford Motor, table 6 for IBM, table 7 for Coca-Cola, table 8 for Microsoft.

From these tables, a clear trend is that when the number of intensity components increases, the likelihood increases too. High components MSMDs usually have higher likelihood. When the number of components increases to seven, the MSMD can give a good description to the data. This can be seen from the p-value of the information matrix test.

6.7 Comparison with ACD

From last section, we see that MSMD model with seven intensity components usually can give a good description of the data. From now on, we fix the number of components at seven and use the MSMD(7) model to compare with the ACD(1,1) model. We choose the ACD(1,1) model because it is the leading example.

In sample fitting results for the low trading group are in table 9. Out of sample forecasting results for low trading group are in table 10. In sample fitting results for the high trading group are in table 11. Out of sample forecasting results for high trading group are in table 12.

It is obvious that the MSMD model can fit the data better. The log likelihoods of MSMD(7) are higher than ACD(1,1) for all twenty stocks.

For 1-step forecasting, the results of both MSMD and ACD model are comparable. MSMD(7) does better forecasting for the high trading group, while ACD(1,1) can give more precise forecasting for the low trading group.

For 5-step and 20-step forecasting, the MSMD model dominates the ACD model for all 20 stocks.

This is clear evidence that the ACD model can only capture short run dynamics, while the MSMD can capture longer horizon dynamics. An interesting observation is that the mean square prediction error of the MSMD model doesn't change much when the forecasting horizon changes.

Maybe it is more fare to compare the MSMD models with a long memory ACD model. But as we explain in last section, the fractional integrated ACD model doesn't have second order moment. It is not a long memory model in the common sense. So we ignore it. Another possible candidate is the long memory stochastic duration (LMSD) model of Deo et al. (2010). One problem with LMSD is that it does not allow iid exponential errors[14]. For now, we don't consider the LMSD model.

7 Concluding Remarks

In this paper, we propose a new model, the MSMD model to analyze financial durations. Compared to the conditional mean modeling of ACD models, our method focus on intensity modeling.

We first establish a mixture of exponentials representation for general point processes, then model the intensity process as a MSM process. We show the MSMD model has good properties. It can explain most of the stylized facts of durations.

Extensive empirical study shows the MSMD model can do good long

[14]As Deo et al. (2010) reports the LMSD model doesn't converge when using iid exponential errors.

horizon forecasting. The intertrade duration is a natural measure of market liquidity and their variability is related to liquidity risk. Our model could be used for the analysis of liquidity on financial markets. For example, we can use bayesian method to update the probabilities of different intensity components. The high frequency component shocks can be regarded as liquidity shocks.

Another interesting direction for future work is to link durations, or intensities to volatilities. The MSM volatility model of Calvet and Fisher (2004) has a lot of similarities with the MSM duration model. The driving force of the MSM volatility and the MSM intensity could be the same.

References

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

Aït-Sahalia, Y. and Mykland, P. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* 71(2), 483–549.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96(453), 42–55.

Baillie, R., Bollerslev, T., and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 74(1), 3–30.

Bandi, F. and Russell, J. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75(2), 339–369.

Barber, B. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2), 785.

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique* 117–149.

Bauwens, L. and Hautsch, N. (2006). Stochastic Conditional Intensity Processes. *Journal of Financial Econometrics* 4(3), 450–493.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations. *Journal of Econometrics* 119(2), 381–412.

Bollerslev, T. and Mikkelsen, H. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73(1), 151–184.

Bowsher, C. (2007). Modelling security market events in continuous time: intensity based, multivariate point process models. *Journal of Econometrics* 141(2), 876–912.

Breidt, F., Crato, N., and De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83(1-2), 325–348.

Brémaud, P. (1981). *Point Processes and Queues, Martingale Dynamics.* Springer, New York.

Calvet, L. and Fisher, A. (2001). Forecasting multifractal volatility. *Journal of econometrics* 105(1), 27–58.

Calvet, L. and Fisher, A. (2004). How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics* 2(1), 49.

Carr, P. and Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71(1), 113–142.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18(01), 17–39.

Chen, Y. and Hsieh, C. (2010). Generalized Moment Tests for Autoregressive Conditional Duration Models. *Journal of Financial Econometrics* 8(3), 345–391.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 135–155.

Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics* 73(1), 101–149.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8(4), 291–323.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7(2), 174–196.

Corwin, S. and Coughenour, J. (2008). Limited attention and the allocation of effort in securities trading. *The Journal of Finance* 63(6), 3031–3067.

Daley, D. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, New York.

Daley, D. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Verlag, New York.

De Luca, G. and Gallo, G. (2004). Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics and Econometrics* 8(2).

De Luca, G. and Zuccolotto, P. (2003). Finite and infinite mixtures for financial durations. *Metron* 61, 431–455.

Deo, R., Hsieh, M., and Hurvich, C. (2010). Long memory in intertrade durations, counts and realized volatility of nyse stocks. *Journal of Statistical Planning and Inference* 140, 3715–3733.

Deo, R., Hurvich, C., and Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131(1-2), 29–58.

Deo, R., Hurvich, C., Soulier, P., and Wang, Y. (2009). Conditions for the Propagation of Memory Parameter from Durations to Counts and Realized Volatility. *Econometric Theory* 25(03), 764–792.

Diebold, F. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics* 105(1), 131–159.

Ding, Z., Granger, C., and Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance* 1(1), 83–106.

Drost, F. and Werker, B. (2004). Semiparametric duration models. *Journal of Business and Economic Statistics* 22(1), 40–50.

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

Engle, R. (2000). The econometrics of ultra-high-frequency data. *Econometrica* 68(1), 1–22.

Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17(5), 425–446.

Engle, R. and Gallo, G. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131(1-2), 3–27.

Engle, R. and Russell, J. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.

Fernandes, M. and Grammig, J. (2005). Nonparametric specification tests for conditional duration models. *Journal of Econometrics* 127(1), 35–68.

Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics* 130(1), 1–23.

Gagliardini, P. and Gouriéroux, C. (2008). Duration time-series models with proportional hazard. *Journal of Time Series Analysis* 29(1), 74–124.

Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics* 119(2), 413–433.

Ghysels, E. and Jasiak, J. (1998). GARCH for irregularly spaced financial data: the ACD-GARCH model. *Studies in Nonlinear Dynamics and Econometrics* 2(4), 133–149.

Giot, P. (2000). Time transformations, intraday data, and volatility models. *Journal of Computational Finance* 4(2), 31–62.

Grammig, J. and Maurer, K. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal* 3(1), 16–38.

Grammig, J. and Wellner, M. (2002). Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics* 106(2), 369–400.

Granger, C. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis* 1(1), 15–29.

Hasbrouck, J. (1999). Trading fast and slow: Security market events in real time. *Working paper*, New York University.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2), 271–320.

Hong, H., Kubik, J., and Stein, J. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance* 60(6), 2801–2824.

Huang, L. and Liu, H. (2007). Rational inattention and portfolio selection. *Journal of Finance* 62(4), 1999–2040.

Huberman, G. (2001). Familiarity breeds investment. *Review of Financial Studies* 14(3), 659.

Hujer, R. and Vuletic, S. (2007). Econometric analysis of financial trade processes by discrete mixture duration models. *Journal of Economic Dynamics and Control* 31(2), 635–667.

Jasiak, J. (1999). Persistence in intertrade durations. *Working paper*, Department of Economics, York University.

Karr, A. (1991). *Point processes and their statistical inference.* Dekker, New York.

Li, W. and Yu, P. (2003). On the residual autocorrelation of the autoregressive conditional duration model. *Economics Letters* 79(2), 169–175.

Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. *Working paper*, Aalborg University.

Meitz, M. and Teräsvirta, T. (2006). Evaluating models of autoregressive conditional duration. *Journal of Business and Economic statistics* 24(1), 104–124.

Meyn, S. and Tweedie, R. (1993). *Markov chains and stochastic stability.* Springer-Verlag London.

Pagan, A. and Schwert, G. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics* 45, 267–290.

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80(3), 563–602.

Russell, J. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. *Mimeo*, University of Chicago, Graduate School of Business.

Stock, J. (1988). Estimating continuous-time processes subject to time deformation: an application to postwar US GNP. *Journal of the American Statistical Association* 83, 77–85.

Sun, W., Rachev, S., Fabozzi, F., and Kalev, P. (2008). Fractals in trade duration: capturing long-range dependence and heavy tailedness in modeling trade duration. *Annals of Finance* 4(2), 217–241.

West, K. and McCracken, M. (1998). Regression-based tests of predictive ability. *International Economic Review* 39(4), 817–840.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society* 50(1), 1–25.

Zhang, M., Russell, J., and Tsay, R. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* 104(1), 179–207.

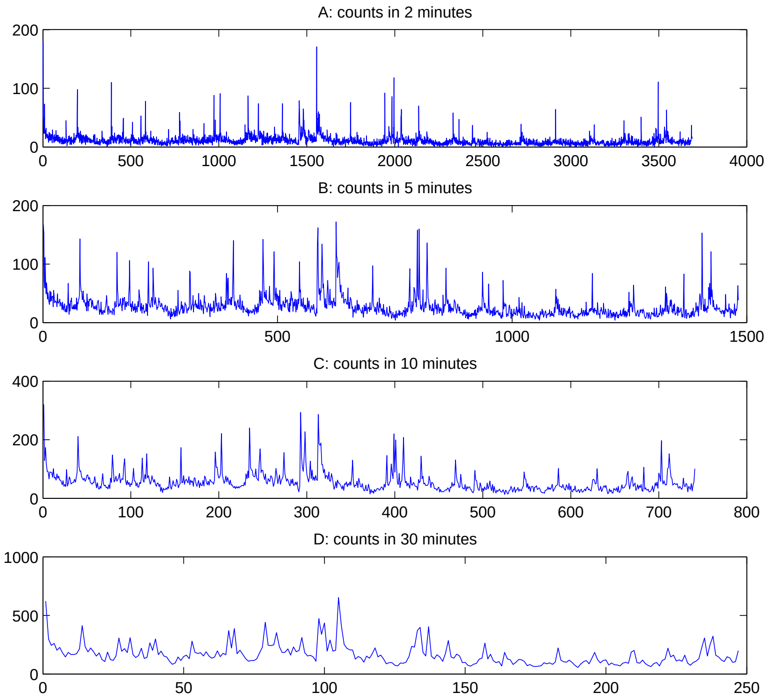

Figure 1: **Clustering Effect at Different Time Scale**
Panel A B C D are counting data for the same period. The data is IBM transaction data from February 1 1993 to December 31 1993. Vertical axis is number of counts. Horizontal axis is time index.

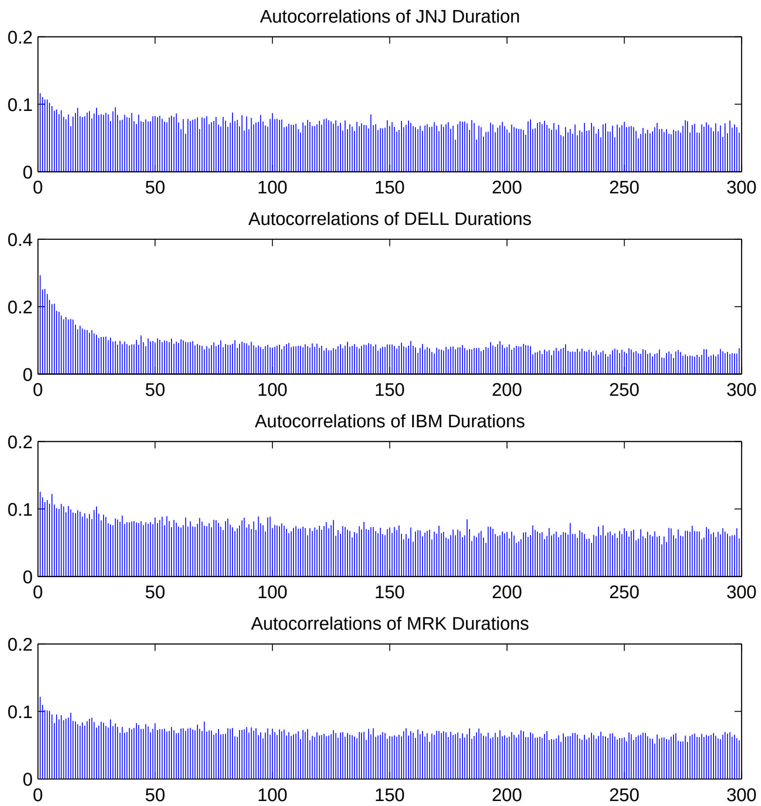

Figure 2: **Sample Autocorrelation Functions for 4 Stocks**
The sampling period is from February 1 1993 to February 26 1993.

Symbol	Company Name
AA	ALCOA
ABT	Abbott Laboratories
AXP	American Express Inc
BAC	Bank of America Corp
CSCO	Cisco Systems
DELL	Dell
DOW	Dow Chemical
F	Ford Motor
GE	General Electric Co.
IBM	International Business Machines
INTC	Intel Corporation
JNJ	Johnson & Johnson Inc
KO	The Coca-Cola Company
MCD	McDonald's Corp
MRK	Merck & Co.
MSFT	Microsoft
TXN	Texas Instruments
WFC	Wells Fargo
WMT	Wal-Mart
XRX	Xerox Corp

Table 1: **Twenty Stocks: symbol and company name**

Stock	Mean	Median	Max	Min	STD	Skew	Kurt	OD	n
AA	2.66	1.28	39.58	0.01	3.77	3.07	17.09	1.42	2989
AXP	2.22	1.17	42.33	0.05	2.93	3.13	19.35	1.32	10531
BAC	1.97	1.12	27.14	0.03	2.44	2.9	15.69	1.24	7939
DOW	1.96	1.11	37.59	0.02	2.49	3.6	28.03	1.27	6902
GE	2.03	1.1	27.13	0.06	2.56	2.72	13.85	1.26	14798
KO	1.82	1.14	26.31	0.05	2.06	2.49	12.47	1.13	15542
MCD	1.93	1.15	22.17	0.03	2.26	2.58	12.77	1.17	7441
TXN	2.56	1.15	55.41	0.02	3.7	3.39	23.54	1.44	4235
WFC	2.47	1.08	78.65	0.02	4.05	5.18	54.37	1.64	4047
XRX	2.56	1.15	55.41	0.02	3.7	3.39	23.54	1.44	4235

Table 2: **Basic Statistics: Low Trading Group**
Skew is Skewness. Kurt is Kurtosis. OD is overdispersion which is equal to std/mean. n is the number of observations in the sampling period.

Stock	Mean	Median	Max	Min	STD	Skew	Kurt	OD	n
ABT	1.86	1.11	26.3	0.06	2.21	2.84	15.89	1.19	16929
CSCO	2.22	1.01	56.3	0.07	3.52	4.53	36.29	1.59	17963
DELL	2.13	1.02	76.83	0.09	3.4	5.15	48.3	1.6	24160
F	2.18	0.99	49.25	0.07	3.13	3.5	22.48	1.44	15562
IBM	1.75	1.06	35.87	0.12	2.03	3.01	19.16	1.16	31895
INTC	1.81	1.0	50.38	0.15	2.41	4.17	34.57	1.33	41957
JNJ	1.72	1.03	29.56	0.08	2.01	3.1	19.1	1.17	24208
MRK	1.61	0.98	24.66	0.18	1.78	2.95	17.31	1.11	54242
MSFT	2.01	1.01	53.68	0.11	2.94	4.43	37.43	1.46	29191
WMT	1.77	0.99	31.92	0.12	2.11	2.88	16.23	1.19	33899

Table 3: **Basic Statistics: High Trading Group**
Skew is Skewness. Kurt is Kurtosis. OD is overdispersion which is equal to std/mean. n is the number of observations.

$\bar{k}$	3	4	5	6	7
m_0	1.3278	1.2857	1.3104	1.311	1.3108
	(0.0)	(0.01)	(0.01)	(0.01)	(0.01)
λ	0.8494	0.8518	0.836	1.2135	0.9256
	(0.02)	(0.06)	(0.03)	(0.07)	(0.05)
$\gamma_{\bar{k}}$	0.7052	0.9194	0.8595	0.8696	0.8663
	(0.01)	(0.13)	(0.01)	(0.14)	(0.14)
b	11.37	7.14	11.58	11.93	11.81
	(0.48)	(1.8)	(0.28)	(3.2)	(3.3)
$-\ln L$	12917	12916	12912	12919	12912
S_{IM}	13.8137	11.5906	2.803	3.1441	1.7233
p	0.1817	0.3134	0.9857	0.9778	0.9981

Table 4: **MSMD Estimation: BAC**
S_{IM} is the statistic for White's Information Matrix Test, $S_{IM} \sim \chi^2_{10}$. p is the p-value of S_{IM}

$\bar{k}$	3	4	5	6	7
m_0	1.431 (0.0)	1.4145 (0.01)	1.3367 (0.01)	1.3191 (0.01)	1.2982 (0)
λ	1.0045 (0.04)	0.7475 (0.02)	1.0393 (0.06)	1.2841 (0.13)	1.1107 (0.04)
$\gamma_{\bar{k}}$	0.7584 (0.05)	0.8052 (0.06)	0.999 (0)	0.999 (0)	0.999 (0)
b	19.6802 (2.5)	18.2764 (2.4)	9.9257 (1.8)	8.0253 (1.4)	6.1077 (1.8)
$-\ln L$	25684	25674	25649	25647	25639
S_{IM}	34.7749	24.7862	24.9896	19.1013	3.6787
p	0.0001	0.0058	0.0054	0.039	0.9607

Table 5: **MSMD Estimation: F**
S_{IM} is the statistic for White's Information Matrix Test, $S_{IM} \sim \chi^2_{10}$. p is the p-value of S_{IM}

$\bar{k}$	3	4	5	6	7
m_0	1.2739	1.2361	1.2098	1.2098	1.1791
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
λ	0.8173	0.8725	0.8361	0.6925	0.7686
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
$\gamma_{\bar{k}}$	0.0448	0.0471	0.0533	0.0538	0.0643
	(0.0)	(0.0)	(0.01)	(0.01)	(0.01)
b	10.12	5.93	3.90	3.97	2.85
	(1.7)	(0.91)	(0.45)	(0.47)	(0.31)
$-\ln L$	47745	47711	47696	47697	47695
S_{IM}	39.6988	20.3271	9.0755	7.3585	3.5432
p	0	0.0263	0.5250	0.6912	0.9656

Table 6: **MSMD Estimation: IBM**
S_{IM} is the statistic for White's Information Matrix Test, $S_{IM} \sim \chi^2_{10}$. p is the p-value of S_{IM}

$\bar{k}$	3	4	5	6	7
m_0	1.3214	1.2449	1.2049	1.2409	1.2449
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
λ	1.271	1.0381	0.6612	0.6811	2.4128
	(0.06)	(0.04)	(0.02)	(0.02)	(0.17)
$\gamma_{\bar{k}}$	0.1378	0.3459	0.7213	0.3597	0.3481
	(0.02)	(0.08)	(0.15)	(0.08)	(0.21)
b	74.08	13.26	6.78	12.03	13.41
	(27.02)	(2.43)	(1.38)	(2.41)	(4.17)
$-\ln L$	24441	24429	24423	24427	24421
S_{IM}	25.0972	9.7062	12.8718	16.8928	5.0148
p	0.0052	0.4666	0.2309	0.0768	0.8902

Table 7: **MSMD Estimation: KO**
S_{IM} is the statistic for White's Information Matrix Test, $S_{IM} \sim \chi^2_{10}$. p is the p-value of S_{IM}

$\bar{k}$	3	4	5	6	7
m_0	1.379	1.339	1.3199	1.2935	1.2984
	(0.0)	(0.0)	(0.01)	(0.01)	(0.01)
λ	0.7844	0.852	0.7173	0.8548	1.2205
	(0.02)	(0.03)	(0.03)	(0.05)	(0.1)
$\gamma_{\bar{k}}$	0.0931	0.102	0.1115	0.1184	0.1229
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
b	4.2922	3.1773	3.0434	2.5351	2.7579
	(0.48)	(0.28)	(0.28)	(0.21)	(0.22)
$-\ln L$	44696	44642	44635	44631	44632
S_{IM}	86.0721	12.3657	16.7955	10.8251	10.6235
p	0	0.2613	0.079	0.3713	0.3876

Table 8: **MSMD Estimation: MSFT**
S_{IM} is the statistic for White's Information Matrix Test, $S_{IM} \sim \chi^2_{10}$. p is the p-value of S_{IM}

Stock	In Sample Fitting $-\ln L$	
	MSMD(7)	ACD(1,1)
AA	5598.4	5791.3
AXP	17828	18010
BAC	12919	13078
DOW	11148	11246
GE	24403	24783
KO	24441	24547
MCD	12106	12207
TXN	7690.3	7978.9
WFC	7050.7	7315.3
XRX	7690.3	7978.9

Table 9: **Model Comparison: in Sample Fit**
This is low trading group.

Stock	Out of Sample Forecasting					
	1-step MSPE		5-step MSPE		20-step MSPE	
	MSMD(7)	ACD(1,1)	MSMD(7)	ACD(1,1)	MSMD(7)	ACD(1,1)
AA	18.1833	16.4306	18.7463	41.3716	19.8247	25.3063
AXP	17.4886	16.5861	17.8916	64.0274	18.61	61.2588
BAC	8.0538	7.9268	8.1647	24.6852	8.2475	16.329
DOW	10.1717	9.9576	10.4884	31.0425	10.7088	24.3516
GE	5.7333	5.6145	5.898	20.4824	6.1323	15.2521
KO	3.0342	3.0314	3.044	11.3846	3.1105	6.6252
MCD	6.16	6.0905	6.3188	20.5726	6.2952	12.8945
TXN	12.2909	11.1517	12.7247	28.581	13.4088	15.351
WFC	3.1323	8.5206	3.1509	23.8728	3.1785	13.2856
XRX	12.2909	11.1517	12.7247	28.581	13.4088	15.351

Table 10: **Model Comparison: out of Sample Forecast**
This is low trading group. MSPE is mean square prediction error.

Stock	In Sample Fitting $-\ln L$	
	MSMD(7)	ACD(1,1)
ABT	26324	26460
CSCO	28610	29110
DEll	37903	38254
F	25652	26253
IBM	47695	47810
INTC	61895	62032
JNJ	35893	36011
MRK	77378	77486
MSFT	44635	44984
WMT	50644	50841

Table 11: **Model Comparison: in Sample Fit**
This is high trading group.

Stock	Out of Sample Forecasting					
	1-step MSPE		5-step MSPE		20-step MSPE	
	MSMD(7)	ACD(1,1)	MSMD(7)	ACD(1,1)	MSMD(7)	ACD(1,1)
ABT	2.1301	2.132	2.1416	9.6389	2.203	2.8122
CSCO	3.1284	3.2781	3.2912	10.4741	3.4586	3.7093
DEll	6.678	7.0334	7.0476	22.2225	7.3236	15.0193
F	6.9681	6.6235	7.2714	25.4917	7.5698	21.5794
IBM	2.819	2.8114	2.8035	12.7333	2.8708	8.0736
INTC	7.4043	7.5486	7.7788	25.825	8.5547	14.7445
JNJ	3.2169	3.315	3.2208	14.035	3.183	9.0553
MRK	0.9111	0.915	0.9153	4.5424	0.9008	2.8697
MSFT	9.7861	10.2364	10.3726	36.8851	10.8266	30.3269
WMT	8.9516	9.0933	9.0395	28.4147	9.1606	17.329

Table 12: **Model Comparison: out of Sample Forecast**
This is high trading group. MSPE is mean square prediction error.