A MARKOV-SWITCHING MULTI-FRACTAL INTER-TRADE DURATION MODEL,
WITH APPLICATION TO U.S. EQUITIES

Fei Chen
Francis X. Diebold
Frank Schorfheide

A Markov-Switching Multi-Fractal Inter-Trade Duration Model, with Application to U.S. Equities
Fei Chen, Francis X. Diebold, and Frank Schorfheide
NBER Working Paper No. 18078
May 2012
JEL No. C22,C41

ABSTRACT

We propose and illustrate a Markov-switching multi-fractal duration (MSMD) model for analysis of inter-trade durations in financial markets. We establish several of its key properties with emphasis on high persistence (indeed long memory). Empirical exploration suggests MSMD's superiority relative to leading competitors.

Fei Chen
School of Economics, HUST
No 1037 Luoyu Road, Hongshan District
Wuhan, 430074 CHINA
chenf@econ.upenn.edu

Francis X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
and NBER
fdiebold@sas.upenn.edu

Frank Schorfheide
University of Pennsylvania
Department of Economics
3718 Locust Walk
McNeil 525
Philadelphia, PA 19104-6297
and NBER
schorf@ssc.upenn.edu

1 Introduction

The necessity of grappling with "big data," and the desirability of unlocking the information hidden within it, is a key modern development – arguably *the* key modern development – in all the sciences. Time-series econometrics, particularly time-series financial econometrics, is no exception. Big data in financial econometrics has both cross-sectional and time-series aspects. In the cross-sectional dimension it arises from the literally hundreds of thousands of assets that trade in global financial markets. In the time-series dimension it arises from the similarly huge number of trades, often many per second, that are now routinely executed for financial assets in liquid markets.

In this paper we are concerned in general with the time-series dimension, and in particular with the durations associated with inter-trade arrivals. Those trades are facilitated by modern hardware, software and algorithms, and they are continuously recorded electronically. The result are vast quantities of high-frequency (trade-by-trade) price data.

One might reasonably ask what, precisely, financial econometricians hope to learn from high-frequency trading data. Interestingly, such high-frequency data emerge as largely uninformative for some objects of interest (e.g., trend, or "drift," in log price), but highly informative for others (e.g., volatility), an insight traces in economics at least to Merton (1980). In particular, although precise estimation of trend benefits greatly from a long calendar span but not from high-frequency sampling, precise estimation of volatility benefits immensely from high-frequency sampling.[1] Accurate volatility estimation and forecasting, in turn, are crucial for financial risk management, asset pricing and portfolio allocation.

In this paper we are not directly interested in volatility; rather, as mentioned above, we are interested in inter-trade *durations*. However, high-frequency data is informative not only for the properties of volatility, but also for the properties of inter-trade durations. At one level that observation is trivial, as one obviously needs trade-by-trade data to infer properties of inter-trade durations. But at another level the observation is quite deep, linked to the insight that, via time-deformation arguments in the tradition of Clark (1973), properties of calendar-time volatility and transactions-time trade arrivals should be intimately-related.

In particular, the time-deformation perspective suggests that serial correlation in calendar-time volatility is driven by serial correlation in calendar-time trade counts. But serial cor-

---

[1]Indeed the mathematical foundation of the modern financial econometrics "realized volatility" literature initiated by Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002) is precisely the convergence of empirical quadratic variation to population quadratic variation as sampling frequency increases.

relation in calendar-time trade counts is driven by serial correlation in transactions-time trade-arrival intensity, which is ultimately driven by serial correlation in the information flow that drives trading. Hence the calendar-time "realized volatility" persistence revealed by high-frequency data should have parallels in calendar-time trade count persistence and transaction-time inter-trade duration persistence. Interestingly, the key and robust finding for realized volatility dynamics is not only high volatility persistence (of course...), but *tremendously* high volatility persistence, in the form of long memory.[2]

Against this background, in this paper we propose and evaluate a new model of inter-trade durations, closely-linked to the pioneering "multi-fractal" return volatility model of Mandelbrot, Fisher, and Calvet (1997), Calvet and Fisher (2001) and Calvet and Fisher (2004) as discussed and extended in Calvet and Fisher (2008). As we will discuss subsequently in detail, our model's construction and implications are quite different from existing dynamic duration models, most notably the prominent Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998) and variants thereof. Our model, which we call the Markov-switching multi-fractal duration (MSMD) model, captures high persistence in duration clustering; indeed it intrinsically captures long memory while nevertheless maintaining covariance stationarity. It also captures additional important features of observed durations, such as over-dispersion. Finally, as we will also demonstrate, it is highly successful empirically, largely dominating ACD.

We proceed as follows. In section 2, we review the important empirical regularities that routinely emerge in inter-trade durations in financial markets. In section 3 we develop our MSMD model, we characterize its properties relative to the earlier-documented empirical regularities. We also relate the MSMD model to other models that have appeared in the literature, such as the ACD model and the Stochastic Conditional Duration (SCD) model of Bauwens and Veredas (2004). In section 4 we apply the MSMD model to inter-trade duration data for a representative set of U.S. equities, evaluating its performance in both absolute terms as well as relative to the prominent ACD model. We conclude in section 5.

2   Empirical Regularities in Inter-Trade Duration Data

Here we highlight three important regularities in financial market inter-trade durations, illustrating them for a particular U.S. equity (Citigroup). We shall be interested subsequently

---

[2]Findings of long memory in realized volatility run consistently from the early work of Andersen, Bollerslev, Diebold, and Labys (2001) through scores of subsequent studies, as surveyed for example in Andersen, Bollerslev, Christoffersen, and Diebold (2013).
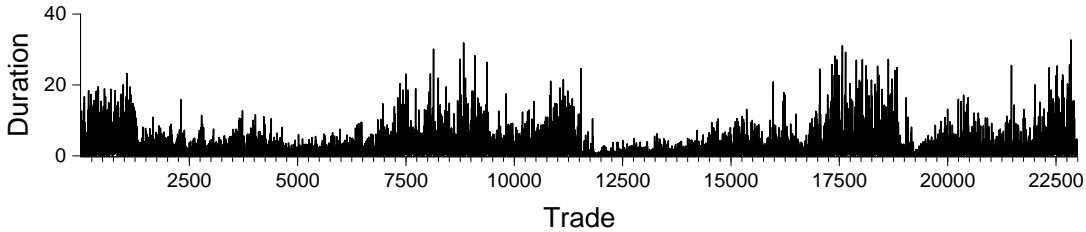
Figure 1: **Citigroup Duration Time Series.** We show a time-series plot of inter-trade durations between 10:00am and 4:00pm during February 1993, measured in minutes and adjusted for calendar effects.

in econometric duration models that capture those regularities.

2.1 Serially-Correlated Duration Dynamics and Self-Similar Clustering

Durations are highly persistent, as is clear visually from the time-series plot of Citigroup durations in Figure 1.[3] The transactions-time duration clustering is matched by a corresponding calendar time transactions-count clustering. Moreover, the calendar-time transactions count clustering tends to look similar regardless of observational frequency. In Figure 2, for example, we show Citigroup transaction counts in intervals of length 2, 5, 10 and 30 minutes during February 1993. Such invariance of qualitative features to observational frequency is called self-similarity and has long been emphasized in financial asset returns, particularly by Mandelbrot (e.g., Mandelbrot (1997)).

2.2 Over-Dispersion

Another property of inter-trade durations, related to their high persistence, is over-dispersion. Over-dispersion refers to the standard deviation exceeding the mean, in contrast to the equality that would obtain with exponential durations generated by *iid* trade arrivals. In Figure 3, for example, we show an exponential duration Q-Q plot for the Citigroup durations, which clearly indicates a non-exponential duration distribution characterized by a longer right tail than would occur under the exponential. Indeed the the sample standard deviation is 2.63

---

[3]We have adjusted the durations for calendar effects and will provide details subsequently in section 4.1.
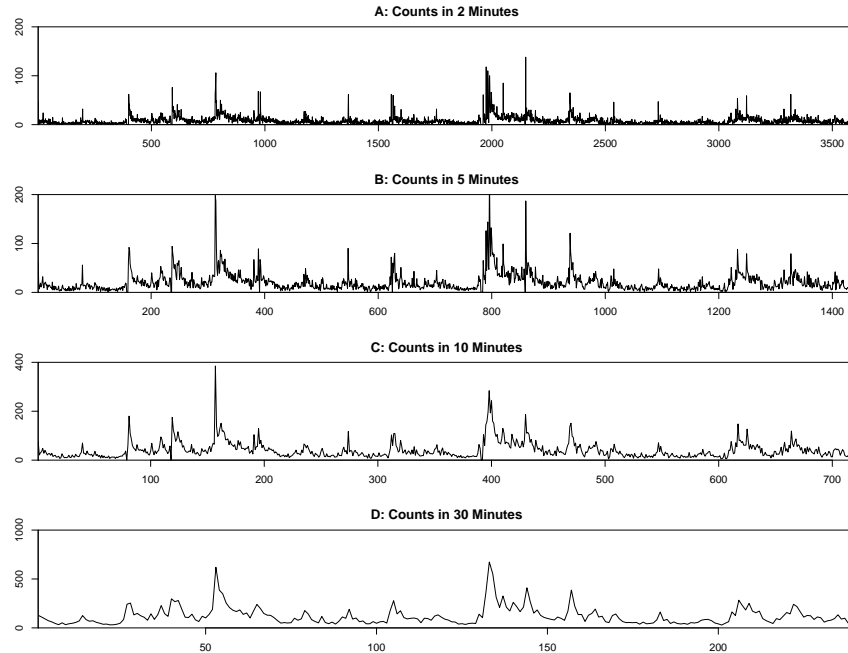
Figure 2: **Citigroup Trasactions Counts.** We show transaction counts in intervals of length 2, 5, 10 and 30 minutes between 10:00am and 4:00pm during February 1993.

and the sample mean is only 1.93. A Kolmogorov-Smirnov goodness-of-fit test rejects the exponential null hypothesis at a marginal significance level less than 10^{-6}.

2.3   Long Memory

We have already mentioned duration persistence and clustering, but we have saved for last the most important aspect of that persistence: long memory. Consider again the Citigroup duration data.[4] The time-series plot in Figure 1 clearly reveals high persistence, but the slow hyperbolic decay of the sample autocorrelation function in Figure 4 reveals significantly more. As is well-known, covariance stationary long memory dynamics are associated with very slow (hyperbolic) autocorrelation decay, in contrast to covariance-stationary short-memory (e.g., ARMA) dynamics, which are associated with very quick (exponential) autocorrelation decay.

It is interesting to note that long memory is self-similar. That is, it can be shown that if a flow variable like transactions counts is $I(d)$ at one observational frequency, then it is $I(d)$ at all other observational frequencies. Hence the observed long memory in duration

---

[4]For an extensive investigation of long memory in inter-trade durations from a model-free perspective, see Deo, Hsieh, and Hurvich (2010).
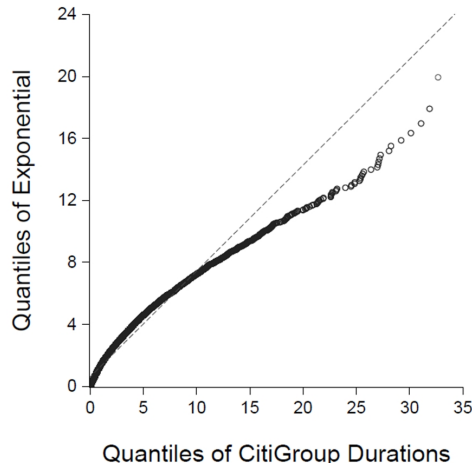
Figure 3: **Citigroup Duration Distribution.** We show an exponential QQ plot for Citigroup inter-trade durations between 10:00am and 4:00pm during February 1993, adjusted for calendar effects.

and transaction count dynamics coheres with the earlier-mentioned observed self-similarity in transaction-count clustering.

3   The Markov-Switching Multi-Fractal Duration Model

In this section we propose the Markov-Switching Multi-fractal Duration (MSMD) model and study its properties, proceeding as follows. In section 3.1 we sketch its basic structure, a mixture-of-exponentials representation with dynamics driven by a Markov-switching multi-fractal intensity process. In section 3.2 we establish some important properties of the MSMD model that match the earlier-documented features of inter-trade durations. In section 3.3 we provide an illustrative simulation. In section 3.4 we explore the relationship of the MSMD model to the general class of point processes. Finally, in section 3.5 we situate the MSMD model within some of the broader literature in inter-trade duration modeling, comparing and contrasting MSMD to existing approaches.

3.1   A New Model for Inter-Trade Durations

We specify our dynamic model of durations in terms of a mixture of exponential random variables in which the intensity evolves dynamically. Working in precisely parallel fashion to the multi-fractal stochastic volatility model of Calvet and Fisher (2008), we assume that
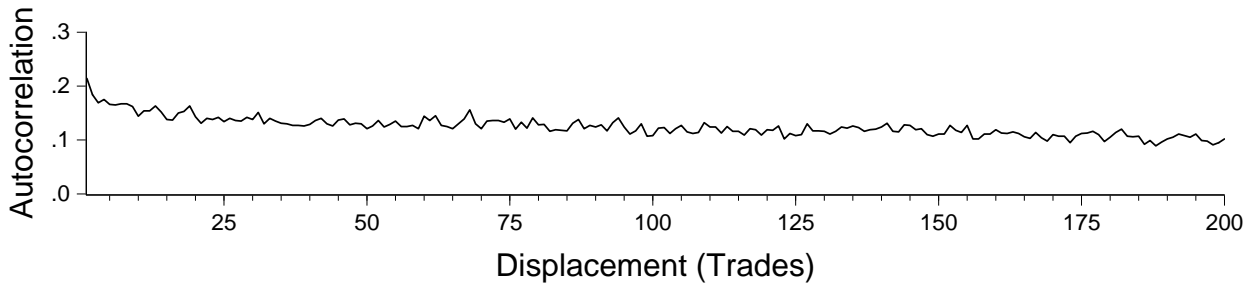
Figure 4: **Citigroup Duration Autocorrelations.** We show the sample autocorrelation function of Citigroup inter-trade durations between 10:00am and 4:00pm during February 1993, adjusted for calendar effects.

the intensities are composed multiplicatively of $\bar{k}$ dynamic components following two-state Markov-switching processes with different degrees of persistence, ranging from very low to very high. This produces the Markov-switching multi-fractal duration (MSMD) model.

Let us now introduce the model. We use $d_i = t_i - t_{i-1}$ to denote the duration between trades (events) occurring at times t_i and t_{i-1}. The durations are distributed according to

$$d_i \sim \frac{\epsilon_i}{\lambda_i}, \quad \epsilon_i \sim iidExp(1), \quad i = 1, \ldots, n, \tag{1}$$

where $Exp(1)$ refers to an exponential distribution with intensity parameter 1. Conditional on λ_i, the durations have a $Exp(\lambda_i)$ distribution, and we refer to the λ_i's as (mean) intensities. The intensities evolve according to the following Markov-switching multi-fractal process:

$$\lambda_i = \lambda \prod_{k=1}^{\bar{k}} M_{k,i}, \tag{2}$$

where λ is a positive constant and the $M_{k,i}$'s are positive intensity components, independent across k and following Markov renewal processes. That is, at time i $M_{k,i}$ is either renewed (drawn from a fixed distribution $f(M)$) or kept at its previous value, $k \in \{1, 2, 3, ..., \bar{k}\}$. We write

$$M_{k,i} = \begin{cases} \text{draw from } f(M) \text{ w.p. } \gamma_k \\ M_{k,i-1} \qquad \qquad \text{w.p. } 1 - \gamma_k, \end{cases} \tag{3}$$

where the renewal distribution $f(M)$ is identical for all $k = 1, ..., \bar{k}$, with $M > 0$ and

7

$E(M) = 1$. Renewal of $M_{k,i}$ corresponds to a new shock of type k hitting the system at time i, with the draw from $f(M)$ governing the magnitude of the shock. The value of γ_k determines the average lifetime and hence the persistence of an M shock. A large γ_k corresponds to M shocks with short expected lifetime and low persistence, and conversely.

Consider now the renewal distribution $f(M)$. We take it as discrete, with two equally-likely points of support,

$$M = \begin{cases} m_0 & \text{w.p. } 1/2 \\ 2 - m_0 & \text{w.p. } 1/2, \end{cases} \tag{4}$$

where $m_0 \in (0, 2]$. We refer to this distribution as "binomial," although it differs slightly from a standard 0-1 Bernoulli trial. Note that the condition $m_0 \in (0, 2]$ implies $M > 0$ a.s. and $E(M) = 1$, as required. By combining (3) and (4) we obtain a two-state Markov process for $M_{k,i}$ that alternates between the states $s_1 = m_0$ and $s_2 = 2 - m_0$ with transition probabilities

$$\mathcal{P}(\gamma_k) = \begin{bmatrix} 1 - \gamma_k/2 & \gamma_k/2 \\ \gamma_k/2 & 1 - \gamma_k/2 \end{bmatrix}. \tag{5}$$

Here element jl of the matrix $\mathcal{P}(\gamma_k)$ specifies the transition from state s_j to s_l. The largest eigenvalue of $\mathcal{P}(\gamma_k)$ is equal to one, which implies that the Markov process has an equilibrium distribution. The equilibrium probabilities of states s_1 and s_2 are $1/2$ regardless of γ_k. The second eigenvalue is equal to $1 - \gamma_k$ and determines the persistence of the Markov chain.

Finally, to induce parsimony, we impose the following restriction on the sequence $\{\gamma_k\}_{k=1}^{\bar{k}}$:

$$\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k - \bar{k}}}, \tag{6}$$

where $\gamma_{\bar{k}} \in (0, 1)$ and $b \in (1, \infty)$. Hence, although the renewal distribution $f(M_k)$ is the same for all intensity components M_k, the renewal *probability*, γ_k, differs across k, creating a variety of intensity components ranging from low-frequency components that renew infrequently to high-frequency components that renew frequently, despite the fact that all renewals are of course stochastic. Small values of k relative to $\bar{k}$ lead to large values of $1 - \gamma_k$ and therefore produce "low-frequency" or "long-run" M_k shocks with low renewal probability and hence long expected lifetime and high persistence.

Equations (1), (2), (3), (4), and (6) define the MSMD model. Assembling everything, we can write it compactly as

$$d_i = \frac{\epsilon_i}{\lambda \prod_{k=1}^{\bar{k}} M_{k,i}} \tag{7a}$$

$$M_{k,i} = \begin{cases} M & \text{w.p. } 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}} \\ M_{k,i-1} & \text{w.p. } (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}} \end{cases} \tag{7b}$$

$$M = \begin{cases} m_0 & \text{w.p. } 1/2 \\ 2 - m_0 & \text{w.p. } 1/2, \end{cases} \tag{7c}$$

where $\epsilon_i \sim iid\ exp(1)$, $\bar{k} \in \mathbb{N}$, $\lambda > 0$, $\gamma_{\bar{k}} \in (0,1)$, $b \in (1, \infty)$ and $m_0 \in (0,2]$. Conditional on $\bar{k}$ we collect the remaining parameters in the vector $\theta_{\bar{k}} = (\lambda, \gamma_{\bar{k}}, b, m_0)'$. Note that the MSMD model has a $\bar{k}$-dimensional state vector $M_i = (M_{1,i}, M_{2,i}, \dots M_{\bar{k},i})$, and hence $2^{\bar{k}}$ states, despite its great parsimony (just four parameters, regardless of $\bar{k}$). Each latent intensity component takes the value m_0 in the high state and $2 - m_0$ in the low state. The parameter λ governs the overall intensity level; if all intensity components are in the high state, then $\lambda_i = m_0^{\bar{k}}$ (the upper bound for intensity) and if all intensity components are in the low state, then $\lambda_i = (2 - m_0)^{\bar{k}}$ (the lower bound for intensity). The parameter $\gamma_{\bar{k}}$ is the renewal probability of the most frequently-renewing intensity component $M_{\bar{k}}$. The other renewal probabilities $\gamma_1, ..., \gamma_{\bar{k}-1}$ increase monotonically (and at an increasing rate) toward $\gamma_{\bar{k}}$, with the precise pattern governed by the parameter b.

3.2   Implied Dynamics and Distributions

Here we study some properties of MSMD processes. We begin by showing stationarity, finite moments and ergodicity. Then we show that MSMD processes are consistent with the earlier-documented key stylized features of inter-trade financial market durations: serially-correlated duration dynamics and self-similar clustering, over-dispersion and long memory.
Stationarity, Ergodicity, and Finite Moments. Provided that $\gamma_k > 0$, the transition probability matrix (5) implies that the process $M_{k,i}$ is strictly stationary and ergodic. Moreover, because the $M_{k,i}$ processes are independent across k and independent of ϵ_i, and because the ϵ_i's are independent across i, the vector process $\xi_i = (\epsilon_i, M_{1,i}, \dots, M_{\bar{k},i})$ is strictly stationary and ergodic. Because the durations d_i are a measurable function of ξ_i, it follows that the sequence of durations is strictly stationary and ergodic. Because the moments of the $Exp(1)$ distribution are finite and there exists an upper bound $\bar{\lambda}$ and a lower bound $\underline{\lambda} > 0$ such that $\underline{\lambda} \le \lambda_i \le \bar{\lambda}$ for all i, we deduce that the moments of d_i are finite.
Serially-Correlated Duration Dynamics and Self-Similar Clustering. The MSMD model produces serially-correlated durations by design, due to the serial correlation in its conditional intensity, which in turn produces serial correlation in calendar-time transaction counts (clustering). Such clustering matches the predictions of many theoretical market

microstructure models. Those models emphasize the simultaneous presence of informed and uninformed traders, with informed trades occurring only when informational events occur, which produces transaction clustering in calendar time.[5]

Over-Dispersion. MSMD processes also display over-dispersion. The $Exp(1)$ distribution of ϵ_i implies that $\mathbb{E}[\epsilon_i] = 1$ and $\mathbb{E}[\epsilon_i^2] = 2$. Recall that $d_i = \epsilon_i/\lambda_i$. Because ϵ_i and λ_i are independent, we deduce from Jensen's Inequality that

$$
\begin{aligned}
Var(d_i) &= 2\mathbb{E}[\lambda_i^{-2}] - (\mathbb{E}[\lambda_i^{-1}])^2 \\
&\geq 2(\mathbb{E}[\lambda_i^{-1}])^2 - (\mathbb{E}[\lambda_i^{-1}])^2 \\
&= (\mathbb{E}[d_i])^2.
\end{aligned}
\tag{8}
$$

The inequality is strict if the distribution of λ_i has a non-zero variance, which leads to the over-dispersion property

$$
\sqrt{Var(d_i)} > \mathbb{E}[d_i].
\tag{9}
$$

It is interesting to note that, in the MSMD model and in reality, the over-dispersion of transactions-time duration distributions is induced by stochastic intensity. This is precisely analogous to the well-known fat-tails of calendar-time return distributions induced by stochastic volatility. Just as stochastic volatility "fattens" Gaussian conditional return distributions, so too does MSMD "over-disperse" exponential conditional duration distributions.

Long Memory. We now show that the MSMD process has long memory, by showing that the duration autocorrelation function, $\rho(h) = Corr(d_i, d_{i+h})$, decays hyperbolically. We begin by letting $\alpha_1 < \alpha_2$ denote two arbitrary numbers in the open interval $(0,1)$. Then the set of integers $\mathcal{H}_{\bar{k}} = \{h : \alpha_1 \log_b(b^{\bar{k}}) \leq \log_b h \leq \alpha_2 \log_b(b^{\bar{k}})\}$ contains a broad collection of lags. Notice that as $\bar{k} \to \infty$ the values of h included in the set $\mathcal{H}_{\bar{k}}$ also tend to infinity. The formal statement of the long-memory property is provided in the following proposition.

Proposition 1 *The MSMD autocorrelation function satisfies*

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

where $\delta = \log_b E(\tilde{M}^2) - \log_b\{[E(\tilde{M})]^2\}$ and $\tilde{M}$ has a binomial distribution that takes the values $2m_0^{-1}/(m_0^{-1} + (2 - m_0)^{-1})$ and $2(2 - m_0)^{-1}/(m_0^{-1} + (2 - m_0)^{-1})$ with probability 1/2.

---

[5]Early work includes Admati and Pfleiderer (1988) and Easley and O'Hara (1992). Surveys include O'Hara (1995), Hasbrouck (2007) and Vives (2008).

Proof The proposition is a direct consequence of results in Calvet and Fisher (2008), as follows. First note that we can define the mean-duration process

$$\phi_i = \frac{1}{\lambda_i} = \frac{1}{\lambda} \prod_{k=1}^{\bar{k}} M_{k,i}^{-1}. \tag{10}$$

Because $M_{k,i}$ is a two state Markov process that takes the values $s_1 = m_0$ and $s_2 = 2 - m_0$ it follows that $M_{k,i}^{-1}$ is also a two state Markov process that takes the values s_1^{-1} and s_2^{-1}. The mean of $M_{k,i}^{-1}$ is independent of k given by $\mu = (s_1^{-1} + s_2^{-1})/2$. Now define the mean-one process $\tilde{M}_{k,i} = M_{k,i}^{-1}/\mu$ such that

$$\phi_i = \frac{\mu^{\bar{k}}}{\lambda} \prod_{k=1}^{\bar{k}} \tilde{M}_{k,i}, \tag{11}$$

which highlights that ϕ_i is also a Markov-switching multi-fractal process. The autocovariances of ϕ_i are not affected by the factor $\mu^{\bar{k}}/\lambda$. Thus, with this transformation the MSMD duration process $d_i = \phi\epsilon_i$ has effectively the same probabilistic structure as the squared return process r_t^2 in Calvet and Fisher (2008), and the long-memory property of the durations follows from their Proposition 1.   **Q.E.D.**

As shown in Proposition 1, the overlaid Markov-switching processes in the MSMD model truly generate long memory in durations.[6] The MSMD long memory is reminiscent of well-known earlier results showing that long memory is obtained by aggregating short-memory processes of differing persistence. Those results include both aggregation of simple $AR(1)$ processes in discrete time as in Granger (1980) and aggregation of Ornstein-Uhlenbeck processes in continuous time as in Cox (1981).[7] Moreover, although the long memory arising in such environments is a limiting result as the number of component processes grows ($\bar{k} \to \infty$ in the case of MSMD), authors such as Alizadeh, Brandt, and Diebold (2002) have argued that that the limiting result holds quickly. Hence only a few components should be needed to approximate long memory accurately.
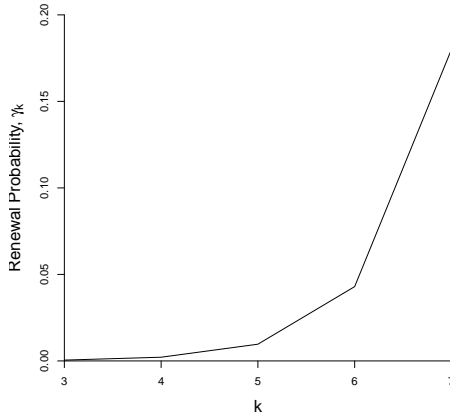
Figure 5: **Latent Intensity Component Renewal Probabilities.** We show the renewal probabilities ($\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}$) associated with the latent intensity components (M_k), $k = 3, ..., 7$. We calibrate the MSMD model with $\bar{k} = 7$, and with parameters that match our subsequently-reported estimates for Citigroup.

3.3 An Illustrative Simulation

To build insight we provide an illustrative simulation of the MSMD model. We calibrate the parameters to our later-obtained estimates for Citigroup ($m_0 = 1.23$, $\lambda = 1.17$, $\gamma_{\bar{k}} = 0.18$, $b = 4.52$), and we use $\bar{k} = 7$, which emerges as widely adequate in our subsequent empirical work.

In Figure 5 we show the renewal probabilities γ_k implied by our choices of $\gamma_{\bar{k}}$, $\bar{k}$ and b (recall that $\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}$), for $k = 3, ..., 7$. For smaller values of k, the renewal probabilities are very small and increasing only slowly, but they increase quickly starting around $k = 5$, reaching 0.18 when $k = \bar{k} = 7$.

In Figure 6 we show time series of simulated latent intensity components $M_{1,i}$, ..., $M_{7,i}$, overall latent intensities λ_i, and durations d_i. Despite the sample of length greater than 22,000 (again chosen to match our subsequent empirical work), the switching probabilities of M_1 and M_2 are so low that they never switch regime. Switching begins with M_3 and increases in frequency as we approach the highest-frequency component M_7, which switches very often. The resulting overall intensity and duration series are very highly persistent,

---

[6]In this regard the MSMD environment with $\bar{k} \to \infty$ is very different from the simple two-state Markov-switching environment studied by Diebold and Inoue (2001), who show in that environment how short-memory dynamics can be mis-diagnosed as long-memory.

[7]See also Corsi (2009).
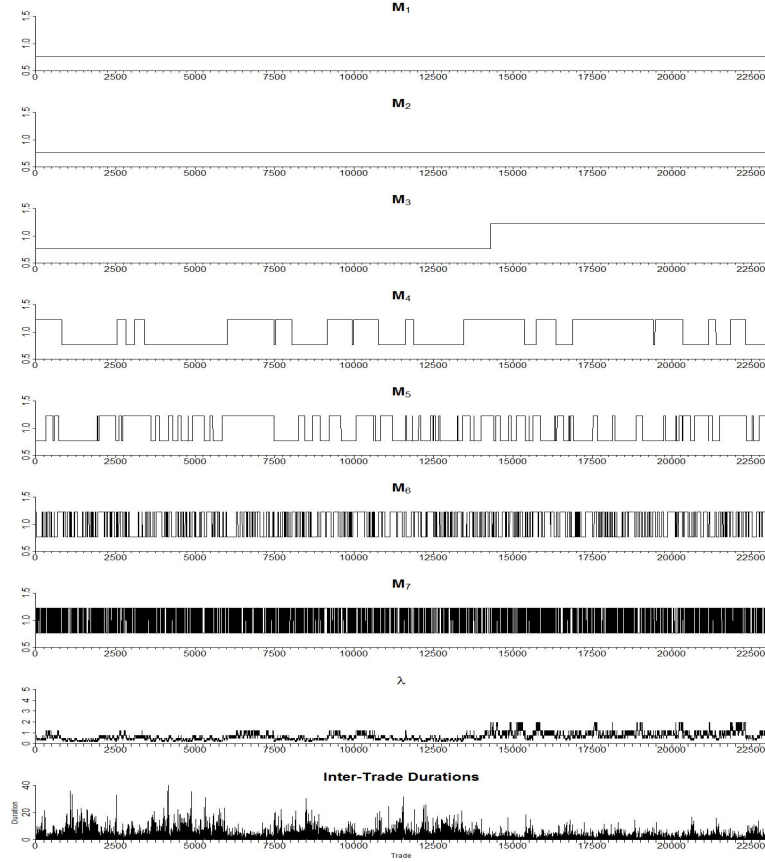
Figure 6: **Time-series plots of simulated** $M_{1,i}$**, ...,** $M_{7,i}$**,** λ_i**, and** d_i**.** We show simulated sample paths for an MSMD model with sample size ($N = 22,988$) and parameters calibrated to match our subsequently-reported estimates for Citigroup.

with episodes of clustering visually apparent in the time series plots at the bottom of Figure 6.

In Figure 7 we show an exponential QQ plot of the simulated durations. It clearly indicates a longer right tail than exponential, consistent with overdispersion. Indeed the exponential Kolmogorov-Smirnov statistic has a p-value less than 2.2e-16.

In Figure 8 we show the sample autocorrelation function of the simulated durations, which makes clear that they are more than just highly-persistent but short-memory; instead, their slow hyperbolic decay indicates long memory.
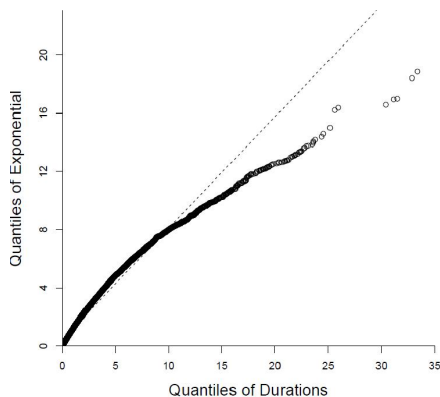
Figure 7: **QQ Plot, Simulated Durations.** We show a QQ plot for a simulated duration sample path for an MSMD model with sample size ($N = 22,988$) and parameters calibrated to match our subsequently-reported estimates for Citigroup.

3.4   MSMD, Point Process Theory, and Time Deformation

We now relate situate the MSMD approach within the broader framework of general point process theory.[8] We focus not on the Markov-switching multi-fractal process for λ_i, as we have discussed it previously and will return to it subsequently. Rather, we step back for a moment to consider the broader background associated with our parameterizing the the conditional intensity in a mixture-of-exponentials representation.

Formally, a point process (PP) on $(0, \infty)$ is a sequence of nonnegative random variables $\{T_i(\omega)\}_{i \in 1,2,...}$ defined on a probability space (Ω, F, P), satisfying $0 < T_1(\omega) < T_2(\omega) < \cdots$, where T_i is the instant of the i-th occurrence of an event. A PP is associated with a counting process, $N(t, \omega)$, where $N(t, \omega) = \sum_{i \geq 1} 1(T_i(\omega) \leq t)$ is the number of events up to and including time t. Under suitable regularity conditions the PP can be characterized by a stochastic intensity $\lambda(t, \omega)$. The heuristic interpretation of the process $\lambda(t, \omega)$ is that of a conditional hazard function:

$$\lambda(t, \omega) = \lim_{\Delta t \downarrow 0} \left(\frac{1}{\Delta t} P[N(t + \Delta t, \omega) - N(t, \omega) = 1 \mid F_{t-}] \right), \tag{12}$$

where F_{t-} is a sub-sigma field of F and can be intuitively interpreted as information available at time $t-$. Conditional on F_{t-} information, the intensity $\lambda(t, \omega)$ is non-stochastic, which is

---

[8]Throughout this subsection we only provide a brief heuristic discussion. For a more rigorous and complete treatment, see for example see Karr (1991) and Daley and Vere-Jones (2003, 2007).
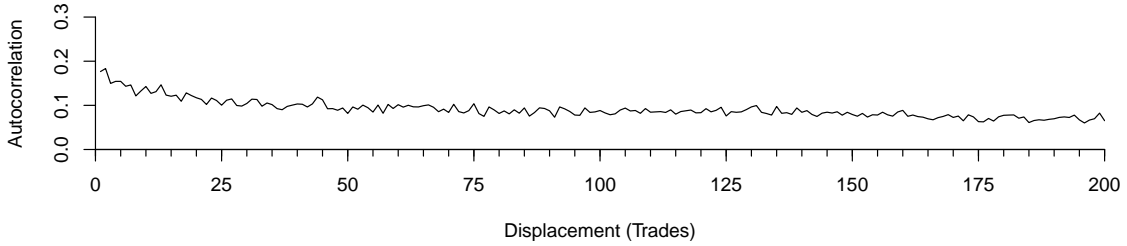
Figure 8: **Sample Autocorrelation Function, Simulated Durations.** We show the sample autocorrelation function for a simulated sample path from an MSMD model with sample size ($N = 22,988$) and parameters calibrated to match our subsequently-reported estimates for Citigroup.

why we will drop the ω argument subsequently.

On can interpret a PP as a dynamic, uncountable set of independent Bernoulli trials, one for each t. If we observe the process over the interval $(0, T]$ then the probability of "successes," i.e. observing trades in the context of our application, at times $t_1, \ldots, t_n$ is given by $\prod_{i=1}^{n} \lambda(t_i)$. Likewise, the probability of no successes in the interval (t_{i-1}, t_i) is given by $\exp[-\int_{t_{i-1}}^{t_i} \lambda(t)dt]$. Combining these two expressions, we deduce that the probability of events at and only at times $t_1, \ldots, t_n$ is

$$p(t_1, \ldots, t_n | \lambda(\cdot)) = \prod_{i=1}^{n} \left(\lambda(t_i) \exp \left[-\int_{t_{i-1}}^{t_i} \lambda(t)dt \right] \right). \tag{13}$$

Suppose that $\lambda(t)$ is constant on the interval $[t_{i-1}, t_i)$ and denote its value by λ_i. Moreover, let $d_i = t_i - t_{i-1}$. Then we can rewrite the probability as

$$p(t_1, \ldots, t_n | \lambda(\cdot)) = \prod_{i=1}^{n} \left(\lambda_i \exp \left[-\lambda_i d_i \right] \right), \tag{14}$$

which exactly corresponds to the mixture-of-exponentials specification (1).

Interestingly, the MSMD mixture-of-exponential representation can also be motivated using a time-deformation argument. Provided that the stochastic intensity $\lambda(t, \omega)$ exists and satisfies the condition $\int_0^\infty \lambda(t, \omega) = \infty$ almost surely, meaning that one always expects more trades in the context of our application, one can define the (stochastic) stopping time $\tau(t, \omega)$ as the (unique) solution to $\int_0^{\tau(t,\omega)} \lambda(s, \omega)ds = t$. It turns out that the process $\tilde{N}(t, \omega)$ defined by $\tilde{N}(t, \omega) = N(\tau(t, \omega), \omega)$ is a standard Poisson process with intensity one. As a

15

consequence the random variables $e_i(\omega) = \int_{T_{i-1}(\omega)}^{T_i(\omega)} \lambda(t, \omega) dt$ are iid $Exp(1)$. If the intensity is constant between events as assumed in the MSMD model, we again obtain $d_i \lambda_i \sim iid Exp(1)$ as in (1).

3.5   On the Comparative Structure of MSMD and Other Inter-Trade Duration Models

Duration dynamics and volatility dynamics are by necessity intimately related. In particular, we have noted the close link between the Markov-switching multi-fractal duration (MSMD) model and the pioneering Calvet-Fisher Markov-switching multi-fractal volatility (MSMV) model. But MSMD is not, of course, the first or only dynamic duration model, just as MSMV is not the first or only dynamic volatility model. Here, therefore, we contrast MSMD to some obvious competing dynamic duration models. We organize our discussion along four lines: mean intensity vs. mean duration models, parameter-driven vs. observation-driven models, long-memory vs. short-memory models, and "structural" vs. "reduced-form" long memory models. Throughout, we use the popular autoregressive conditional duration (ACD) model as an anchoring benchmark.[9]

Mean Intensity vs. Mean Duration. The ACD model of Engle and Russell (1998), which is parameterized in terms of the conditional mean duration, was the key breakthrough in modeling financial market inter-trade durations, and many varieties of ACD models now exist. The "multiplicative error model" (MEM) originally suggested by Engle (2002) is closely related; indeed ACD is a special case of MEM.[10] ACD/MEM assumes that

$$d_i = \varphi_i \epsilon_i, \tag{15}$$

where $\varphi_i = E(d_i | d_{i-1}, ..., d_1)$ and $\epsilon_i \sim iid(1, \sigma^2)$, and it focuses on parameterizing the conditional mean duration, φ_i. We have shown in section 3.2, however, that the MSMD mean-intensity model can be easily re-written as a mean-duration model where the mean duration follows a Markov-switching process. Hence the mean-intensity and mean-duration approaches are evidently more similar than might appear at first glance.

Parameter-Driven vs. Observation-Driven. A key difference between ACD/MEM and MSMD concerns "observation-driven" vs. "parameter-driven" structure. In the parlance of

---

[9]Hautsch (2012) provides a fine overview of much of the relevant literature on financial duration modeling.
[10]The MEM literature has also grown impressively; see, for example, Brownlees, Cipollini, and Gallo (2012) and Hautsch (2012).

Cox (1981), a model is observation-driven if conditional dynamics are driven by the history of observables (as for example in traditional ARCH volatility models), whereas it is parameter-driven if conditional dynamics are driven by the history of a latent variable (as for example in traditional stochastic volatility models).

The ACD model, as well as the important score-driven MEM extensions emphasized by Creal, Koopman, and Lucas (2010) and Harvey (2011), are clearly and intrinsically observation-driven. Their construction precisely parallels the ARCH model, which is also observation-driven. The motivating recognition is that the likelihood may always be factored into a product of conditional densities, where the conditioning is on observed historical data, so that if one specifies a model directly in terms of those conditional densities then likelihood evaluation is immediate and simple.

In contrast, the MSMD model is parameter-driven, as the evolution of the mean intensity follows a hidden Markov process. Rich MSMD duration dynamics arise from the large MSMD state space (recall that λ_i can take $2^{\bar{k}}$ distinct values). Simultaneously, however, MSMD is parsimonious. In addition to $\bar{k}$, which determines the total number of duration states, MSMD uses only two parameters, namely b and $\gamma_{\bar{k}}$, to determine the transition probabilities in this high-dimensional state-space. However, as we discuss below in section 4.3, likelihood evaluation is a bit more challenging (although still simple) for parameter-driven MSMD than for observation-driven ACD/MEM. In particular, MSMD likelihood evaluation requires use of a nonlinear filter as in Hamilton (1989).

The MSMD model is of course not the first parameter-driven duration model. For example, the Stochastic Conditional Duration (SCD) model of Bauwens and Veredas (2004), specifies the log expected duration ($\ln \varphi_i$ in the notation of (15)) as an autoregressive process.[11] Hence if ACD is the duration analogue of the GARCH model, then SCD is the duration analogue of the stochastic volatility model. Finally, Hautsch (2008) combines parameter-driven and observation-driven dynamics in a so-called stochastic ACD model.

Long-Memory vs. Short-Memory. Another key difference between ACD/MEM (as typically implemented) and MSMD concerns long memory. Before considering transactions-time durations, first consider calendar-time volatility. The volatility literature features long memory prominently. For example, long memory is routinely found and modeled in realized volatility calculated using intra-day data, as for example in Andersen, Bollerslev, Diebold, and Labys (2003). Indeed long memory features prominently even in the "conventional" stochastic volatility literature without intra-day data, as for example in the pioneering work

---

[11]Meddahi, Renault, and Werker (1998) also propose an interesting and closely-related model.

of Comte and Renault (1998) and Breidt, Crato, and De Lima (1998).

Now consider durations. The existing inter-trade dynamic duration modeling literature that began with Engle and Russell (1998) is of course aware of high persistence in inter-trade duration dynamics, and it works hard to capture that persistence. But given the close relationship between duration dynamics and volatility dynamics (and again, the routine finding of long memory in volatility dynamics), as well as the direct and model-free empirical evidence of long memory in inter-trade durations presented in section 2.3, long memory has been curiously absent from most of the duration literature. Two notable exceptions are the fractionally integrated ACD model of Jasiak (1999) (FI-ACD) and the long memory version of the SCD model proposed by Deo, Hsieh, and Hurvich (2010). MSMD is firmly in this long-memory camp.

Structural vs. Reduced-Form Long Memory.

The MSMD model's long-memory durations have an appealing structural flavor; that is, MSMD provides a natural long-memory duration generating mechanism by overlaying simple regime-switching processes with different degrees of persistence, potentially corresponding to market activity of different types of informed and uninformed traders. The regime-switching MSMD intensity components in turn parallel earlier two-state regime-switching models in the tradition of Hamilton (1989).

The structural long memory model of MSMD contrasts to "reduced-form" long memory model of Granger and Joyeux (1980). That approach works directly with representations of the form $(1 - L)^d y_t = x_t$, where x_t is a short-memory process, and $0 < d < 1/2$, with the fractional differencing operator $(1 - L)^d$ defined via the expansion

$$(1 - L)^d = 1 - dL + \frac{d(d - 1)}{2!}L^2 - \frac{d(d - 1)(d - 2)}{3!}L^3 + ...$$

Such a series is said to be (fractionally-) integrated of order d, and we write $y_t \sim I(d)$.

From the reduced-form perspective, all shocks have identical hyperbolically-decaying effects. That is the case, for example, with the reduced-form long-memory duration models of Jasiak (1999) and Deo, Hsieh, and Hurvich (2010). An attractive feature of MSMD, however, is its decomposition of reduced-form long memory into structural short-memory pieces coming from underlying cyclical variation at different frequencies, which again may correspond to market activity of different groups of informed and uninformed traders.

Symbol	Company Name	Symbol	Company Name
AA	ALCOA	ABT	Abbott Laboratories
AXP	American Express	BA	Boeing
BAC	Bank of America	C	Citigroup
CSCO	Cisco Systems	DELL	Dell
DOW	Dow Chemical	F	Ford Motor
GE	General Electric	HD	Home Depot
IBM	IBM	INTC	Intel
JNJ	Johnson & Johnson	KO	Coca-Cola
MCD	McDonald's	MRK	Merck
MSFT	Microsoft	QCOM	Qualcomm
T	AT&T	TXN	Texas Instruments
WFC	Wells Fargo	WMT	Wal-Mart
XRX	Xerox		

Table 1: **Twenty-Five U.S. Equities: Ticker symbols and company names.** We show twenty-five stocks selected randomly from the U.S. S&P 100 Index, ordered alphabetically.

4   Empirical Analysis of U.S. Equities

Here we apply the MSMD model to twenty-five U.S. equities selected randomly from those in the S&P 100 index. We study consolidated trade data extracted from the TAQ database. The sample period covers twenty trading days, from February 1, 1993 to February 26, 1993, 10:00 - 16:00.[12] Here in the main text we summarize graphically the results for all firms collectively, and we provide details for Citigroup, whose results are typical. In the appendices we provide details for all firms individually. We list the firms and their ticker symbols in Table 1.

4.1   Intra-Day Calendar Effects

Just as with intra-day volatility in calendar time, durations display intra-day calendar effects. In particular, inter-trade durations tend to be shorter early and late in the day, and longer around lunch. Although the calendar effects are mild, they are nevertheless desirably removed prior to analysis. We do so using a simple deterministic dummy variable approach, as in Ghysels, Gourièroux, and Jasiak (2004). We divide the trading day into half-hour

---

[12]Trading also occurs 09:30-10:00, but following standard procedure (e.g., Engle and Russell (1998)), we delete transactions in 09:30-10:00 to remove effects of the opening auction.
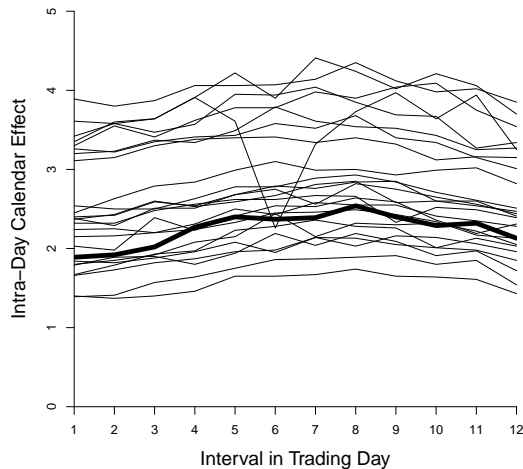
Figure 9: **Intra-Day Duration Calendar Effect Profile Bundle.** We plot estimated time-of-day effects for each firm for twelve half-hour intervals, 10:00am-4:00pm. The bold profile is Citigroup.

periods, and we form the associated indicator variables,

$$x_{ki} = \left\{ \begin{array}{l} 1, \text{ if } i \in k \\ 0, \text{ otherwise,} \end{array} \right.$$

$k = 1, ..., 12$. Then we estimate of the logarithmic dummy-variable regression,

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

The estimated coefficients are the estimated calendar effects. We graph them for all firms in Figure 9; that is, we provide a "profile bundle" in the terminology of Gallant, Rossi, and Tauchen (1993).[13] Finally, we form the adjusted duration series as

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

From this point onward, we use such adjusted durations exclusively, but in slight abuse of notation we use d_i instead of $\hat{d}_i$ to refer to our observations.

---

[13]In addition, we provide numerical details in Appendix A.

20
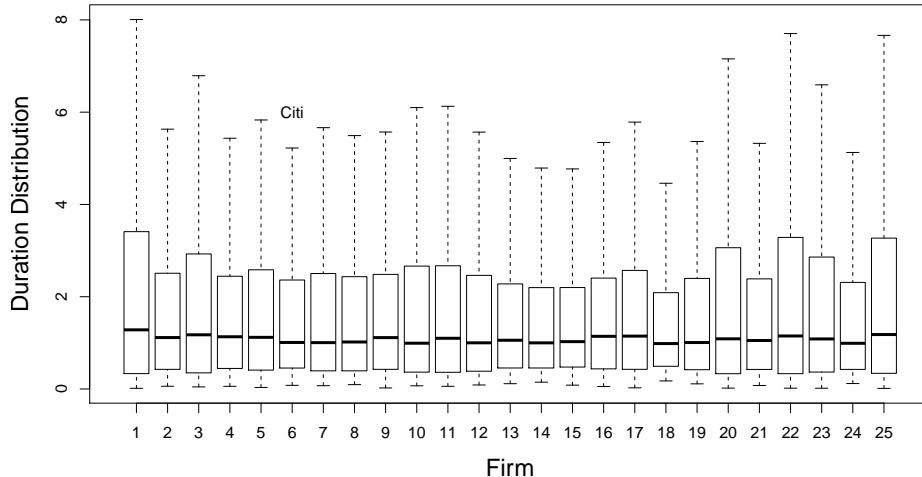
Figure 10: **Firm-by-Firm Duration Distributions (Box-and-Whisker Plots).** The box bottom is the 25th percentile, the box top is the 75th percentile, and the median is in between. The "whiskers" extend 1.5 interquartile ranges above and below the box, or to the edge of the data, whichever comes first.

4.2   Descriptive Statistics for Adjusted Durations

In Figure 10 we summarize the duration distributions for the twenty-five stocks using box-and-whisker plots, with Citigroup labeled. All show long right tails indicative of overdispersion. This is confirmed in Figure 11, in which we show a histogram of coefficients of variation (the standard deviation relative to the mean), for all twenty-five stocks. All are greater than one, and the distribution has a long right tail.[14] In Figure 12 we focus on dynamics rather than distributions, showing a profile bundle of sample autocorrelation functions (to displacement 200) for all twenty-five firms. The slow hyperbolic decay is apparent.

4.3   Maximum Likelihood Estimation

Prior to obtaining maximum-likelihood estimates, we first discuss here the likelihood function for a sequence of durations $d_{1:n} = \{d_1, \ldots, d_n\}$ governed by the MSMD model. By conditional factorization of the joint density, the likelihood is immediately

$$p(d_{1:n}|\theta_{\bar{k}}) = p(d_1|\theta_{\bar{k}}) \prod_{i=2}^{n} p(d_i|d_{1:i-1}, \theta_{\bar{k}}). \tag{16}$$

---

[14]For reference, the arrow indicates Citigroup in that histogram and all subsequent histograms.
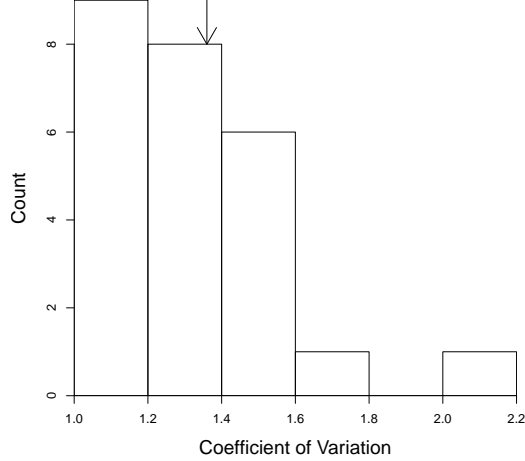
Figure 11: **Distribution of Duration Coefficients of Variation Across Firms.** We show a histogram of coefficients of variation (the standard deviation relative to the mean), as a measure of overdispersion relative to the exponential.

Conditional on λ_i, each duration d_i simply has an $Exp(\lambda_i)$ distribution with density $p(d_i|\lambda_i) = \lambda_i exp[-\lambda_i d_i]$, so evaluation of the likelihood would be trivial if the sequence of λ_i's were known. But the λ_i's are of course not known and must be replaced with estimates from an optimal filter as in Hamilton (1989), the only difference being that the MSMD model has $2^{\bar{k}}$ states rather than Hamilton's 2 state. To streamline the exposition we focus on the case of $\bar{k} = 2$, for a total of 4 states; generalization to $\bar{k} > 2$ is obvious but notationally tedious. Recall that we abbreviated the potential values of the states $M_{k,i}$ by $s_1 = m_0$ and $s_2 = 2 - m_0$. Using this notation, λ_i can take four values, $\lambda_i \in \{\lambda s_1 s_1, \lambda s_1 s_2, \lambda s_2 s_1, \lambda s_2 s_2\}$. Because the states $M_{k,i}$ evolve independently, the matrix of transition probabilities for the mean intensity is simply $\mathcal{P}_\lambda = \mathcal{P}(\gamma_1) \otimes \mathcal{P}(\gamma_2)$, where $\otimes$ denotes the Kronecker product and $\mathcal{P}(\gamma_k)$ was defined in (5). In slight abuse of notation, we characterize the distribution of λ_i through densities $p(\lambda_i|\cdot)$. Then we evaluate the likelihood recursively. First we initialize the hidden states with their equilibrium distribution in period $i = 0$. Then starting from $p(\lambda_{i-1}|d_{1:i-1}, \theta_{\bar{k}})$ we obtain the i'th point likelihood by evaluating conditional distribution $p(\lambda_i|d_{1:i-1}, \theta_{\bar{k}})$ (forecast of the mean intensity), then $p(d_i|d_{1:i-1}, \theta_{\bar{k}})$ (forecast of the duration), and then finally the desired $p(\lambda_i|d_{1:i}, \theta_{\bar{k}})$ (updating).

As discussed in Section 3.1, the MSMD model is a hidden Markov model. Its parameter vector is given by $\theta_{\bar{k}} = (\lambda, \gamma_{\bar{k}}, b, m_0)'$ and the number of states is $2^{\bar{k}}$. Hence $\bar{k}$ can be viewed as a model index, and we will use the notation $\mathcal{M}_j$ to denote a MSMD model with
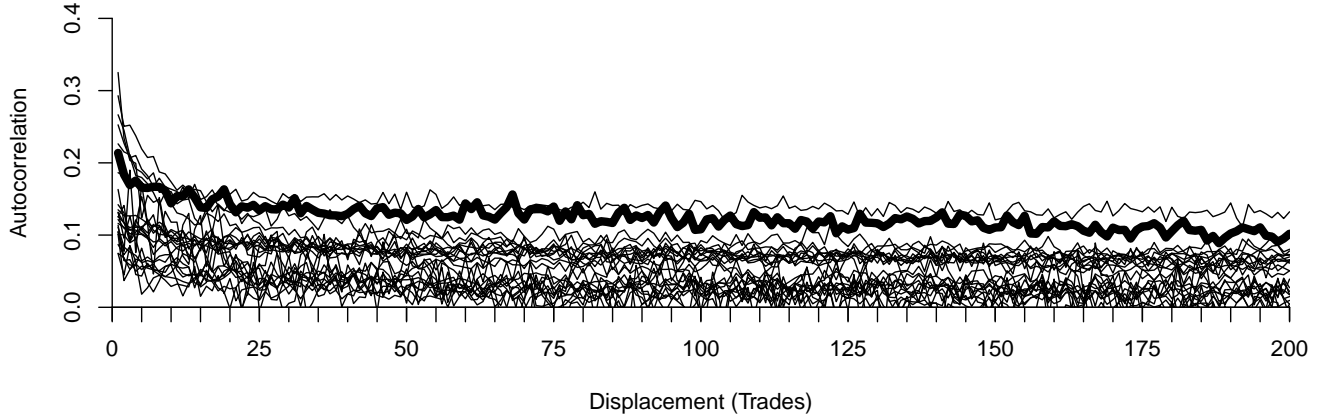
Figure 12: **Duration Autocorrelations Function Profile Bundle.** For each firm, we show the sample autocorrelation function of inter-trade durations between 10:00am and 4:00pm during February 1993, adjusted for calendar effects. The bold profile is Citigroup.

$\bar{k} = j$. While $\bar{k}$ affects the number of hidden states in the MSMD model, it does not alter the dimensionality of the parameter vector $\theta_{\bar{k}}$. In this sense the MSMD model remains parsimonious even if $\bar{k}$ is large.

Models $\mathcal{M}_j$ and $\mathcal{M}_l$, $j \le l$ are essentially non-nested. For instance, consider $\mathcal{M}_1$ versus $\mathcal{M}_2$. For model $\mathcal{M}_1$ the intensity can take the values $\lambda_i \in \{\lambda_1 m_{0,1}, \lambda_1(2 - m_{0,1})\}$, whereas the for $\mathcal{M}_2$ the intensity switches among $\lambda_i \in \{\lambda_2 m_{0,2}^2, \lambda_2 m_{0,2}(2 - m_{0,2}), \lambda_2 m_{0,2}^2\}$. Recall that the sequence $\{\gamma_k\}_{k=1}^{\bar{k}}$ that determines the transition probabilities $\mathcal{P}(\gamma_k)$ of the k'th intensity component $M_{k,i}$ is given by

$$\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}.$$

This restriction ensures that $\mathcal{M}_1$ and $\mathcal{M}_2$ are only nested if either $b = 1$ and $\gamma_{\bar{k}} = 1$ or $m_0 = 1$. The first case is explicitly excluded from the parameter space and the second case turned out to be empirically irrelevant.

We estimate the parameter vector $\theta_{\bar{k}}$ of model $\mathcal{M}_{\bar{k}}$ by maximizing the log likelihood function: $\hat{\theta}_{\bar{k}} = \mathrm{argmax}_{\theta_{\bar{k}} \in \Theta} \ln p(d_{1:n} | \theta_{\bar{k}}, \mathcal{M}_{\bar{k}})$ using the adjusted durations. The MSMD model potentially suffers from the following identification problems. For $\gamma_{\bar{k}} = 1$ the decay parameter b is non-identifiable. While we exclude $\gamma_{\bar{k}} = 1$ from the parameter space – recall that $\gamma_{\bar{k}} \in (0, 1)$ – the decay parameter becomes weakly identified near the upper bound for $\gamma_{\bar{k}}$. As b approaches its lower bound of one, the sequence $\{\gamma_k\}_{k=1}^{\bar{k}}$ approaches a sequence of constants. In the limit, the intensity components $M_{k,i}$ become exchangeable. Finally, if
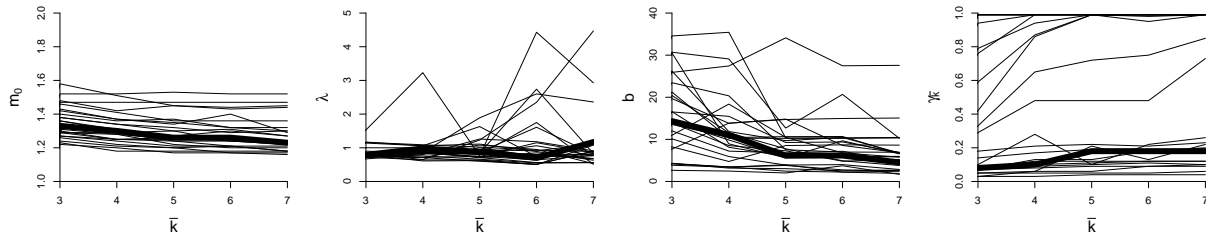
Figure 13: $\hat{m}_0$, $\hat{\lambda}$, $\hat{b}$ and $\hat{\gamma}_{\bar{k}}$ **Profile Bundles.** For each firm, we show maximum likelihood parameter estimates as a function of $\bar{k}$, for $\bar{k} = 3, ..., 7$. The bold profiles are Citigroup.

$m_0 \in (0, 2]$ equals one, then the intensity components are identical across states and neither b nor $\gamma_{\bar{k}}$ are identifiable.

We implement the maximum-likelihood estimation in MATLAB using the *fmincon* function to carry out a constrained maximization. After some experimentation we selected the interior-point option of the optimization algorithm. The empirically relevant bound on the parameters turns out to be $\gamma_{\bar{k}} < 1$, which we enforce by supplying the *fmincon* function with an upper bound of 0.999 for $\gamma_{\bar{k}}$. The identification problems associated with $b = 1$ and $m_0 = 1$ did not turn out to be empirically relevant. For each estimation, we explore a grid of starting values for the minimization routine. The computational time is increasing in $2^{\bar{k}}$. For instance, for $\bar{k} = 10$ the state transition matrix is a 1024×1024 matrix that needs to be re-calculated for every evaluation of the likelihood function. Moreover, the filter has to track probabilities for 1024 states. Fortunately, however, such high $\bar{k}$ values are empirically unnecessary, as we now show.

4.4 Parameter Estimates

We proceed as follows. First we discuss profile bundles for each estimated parameter (that is, parameter estimates as a function of $\bar{k}$, for each firm). Second, we discuss profile bundles for maximized likelihoods as a function of $\bar{k}$, which generally indicate that $\bar{k} = 7$ is adequate. Third, given the general adequacy of $\bar{k} = 7$, we discuss parameter distributions across firms conditional on $\bar{k} = 7$. In keeping with our running Citigroup example, we highlight Citigroup results throughout. We present detailed estimation results in Appendix C.

First, consider the parameter estimate profile bundles in Figure 13. Citigroup's m_0 estimates are robust and slightly decreasing across $\bar{k}$, with $\hat{m}_0 = 1.33, ..., 1.23$ for $\bar{k} = 3, ..., 7$, respectively. The entire profile bundle of m_0 estimates behaves similarly. Citigroup's λ

24
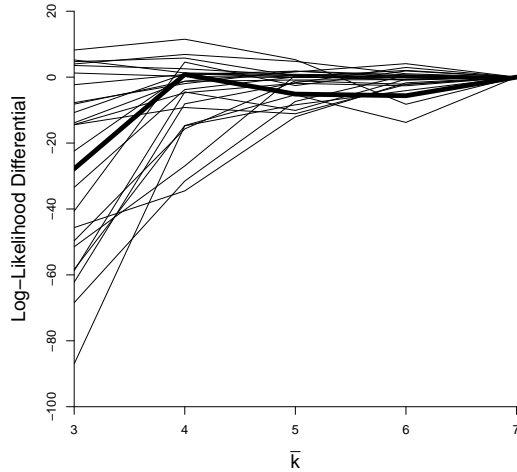
Figure 14: **Log-Likelihood-Differential Profile Bundle.** We show log-likelihood differentials for all firms as a function of $\bar{k}$, with respect to the value for $\bar{k} = 7$, which is therefore identically equal to 0. The bold profile is Citigroup.

estimates are rather stable in $\bar{k}$, but there is a noticeable increase when moving from $\bar{k} = 6$ to $\bar{k} = 7$, with $\hat{\lambda}_0 = 0.79, 0.98, 0.88, 0.70, 1.17$ for $\bar{k} = 3, ..., 7$, respectively. Most firms' $\hat{\lambda}$ profiles are quite stable, but a few are noticeably "bumpy." Citigroup's b estimates are monotonically decreasing in $\bar{k}$, with $\hat{b} = 14.22, ..., 4.52$ for $\bar{k} = 3, ..., 7$, respectively. Most firms' $\hat{b}$ profiles are stable or slightly decreasing in $\bar{k}$. Citigroup's $\gamma_{\bar{k}}$ are quite stable in $\bar{k}$ with, for example, $\hat{\gamma}_{\bar{k}} = 0.18, 0.18, 0.18$ for $\bar{k} = 5, 6, 7$, respectively. As for the profile bundle of $\gamma_{\bar{k}}$ estimates, recall that $\gamma_{\bar{k}}$ is the renewal probability of the most frequently-renewing latent intensity component. Interestingly, the profile bundles separate rather neatly into "high" and "low" groups.

Now we focus on profile bundles of maximized log likelihood functions $\ln p(d_{1:n}|\hat{\theta}_{\bar{k}}, \mathcal{M}_{\bar{k}})$, which we show in Figure 14. In particular, we show log likelihood differentials with respect to $\bar{k} = 7$, which is therefore identically equal to 0, and moreover, a profile increasing in $\bar{k}$ corresponds to a likelihood increasing in $\bar{k}$. Most profiles increase in $\bar{k}$ and level off quickly, around $\bar{k} = 4$; a few are basically flat in $\bar{k}$. While the largest log-likelihood differential is -85, most of the differentials are less than 10 in absolute value. Since the model complexity, i.e. number of parameters, does not increase with $\bar{k}$ the log-likelihood differentials may be interpreted as log-posterior odds. It is rare that the log-likelihood differential is positive and if it is the evidence in favor of $\bar{k} < 7$ tends to be small. The Citigroup profile fits the general pattern, increasing from $\bar{k} = 3$ to $\bar{k} = 4$, and then staying roughly constant.

25
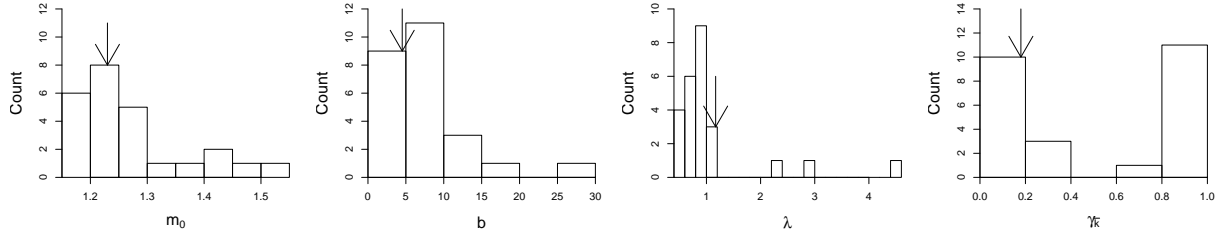
Figure 15: **Distributions of MSMD Parameter Estimates Across Firms, $\bar{k} = 7$.** We show histograms of maximum likelihood parameter estimates across firms, obtained using $\bar{k} = 7$. The arrow denotes Citigroup.

Given the log likelihood results, we now focus exclusively on $\bar{k} = 7$ and consider cross-firm distributions of parameter estimates, as summarized in Figure 15. The distribution of m_0 estimates is skewed, with most estimates around roughly 1.25, but some as high as 1.4 or even 1.5. The distribution of λ estimates is quite dispersed, with some estimates greater than 1 and some less. Most firms' $\hat{\lambda}$ are less than or near 1, but there is again a long right tail. The distribution of b estimates is yet again right-skewed, with a mode in the neighborhood of 7. The distribution of $\gamma_{\bar{k}}$ estimates, in contrast, is strongly bimodal, with a low "low mode" (near 0) and a high "high mode" (near 1), corresponding to the earlier-discussed separation of $\gamma_{\bar{k}}$ profile bundles into "high" and "low" groups.

In Figure 16 we assemble the information in the various parameter estimates in an informative way. We show latent intensity component (M_k) renewal probability profile bundles, estimated using $\bar{k} = 7$. That is, for each firm we show $\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}$ as a function of k, $k = 3, ..., 7$, where $\bar{k} = 7$.[15] Because the estimated b tends to be substantially larger than 1 for all firms, $b^{k-\bar{k}}$ is close to zero for small values of k and all profiles have similarly small γ_k's for $k \leq 3$. But as k grows, a clear bifurcation emerges, with roughly half the profiles rising only slowly, and half rising much more quickly toward 1. Citigroup is one of the "slowly-rising" firms, with γ_k rising from approximately 0 to .18 for as k goes from 3 to 7, meaning that its inter-trade durations are quite persistent.

---

[15]Recall that $1 - \gamma_k$ corresponds to the eigenvalue of the transition probability matrix $\mathcal{P}(\gamma_k)$ that governs the persistence of the intensity component M_k. Thus, firms for which γ_k stays close to zero have very persistent inter-trade durations.
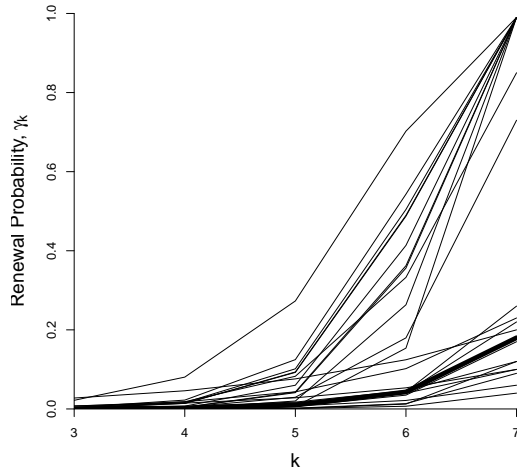
Figure 16: **Estimated Intensity Component Renewal Probability Profile Bundle.**
The bold profile is Citigroup.

4.5   Absolute and Relative MSMD Performance

We now examine the performance of the estimated MSMD model, considering absolute fit
as well as measures of relative fit and forecasting performance. We assess absolute per-
formance of the MSMD model using White's information matrix test, and we assess per-
formance relative to the leading ACD(1,1) competitor using Bayesian information criterion
(BIC) differentials as well as pseudo-out-of-sample root-mean-squared forecast error (RMSE)
differentials.[16]

White's Information-Matrix Test for MSMD(7).

Given the mixture-of-exponentials structure of the MSMD model, we could presumably
construct specification tests based on the residual $\hat{\epsilon}_i = d_i \hat{\lambda}_i$, assessing whether the residuals
are iid exponential after accounting for the fact that the residual is based on the estimate
$\hat{\lambda}_i$ rather than the true λ_i.[17] Instead we take a different and general approach, performing
the information matrix (IM) test of White (1982), which is a generic specification test for
models estimated by MLE.

White (1982)'s IM test was originally developed for models with *iid* observations. Ex-
tensions to time series models are discussed, for instance, in White (1994). In the context of

---

[16]We consider the ACD(1,1) model because it is a widely-used benchmark in the literature. We chose the
ubiquitous (1,1) specification based on preliminary empirical investigation.

[17]In fact, such tests have a long tradition in the specification analysis of point process models that rely on
the time-deformation result discussed in Section 3.4 (e.g., Ogata (1988)).
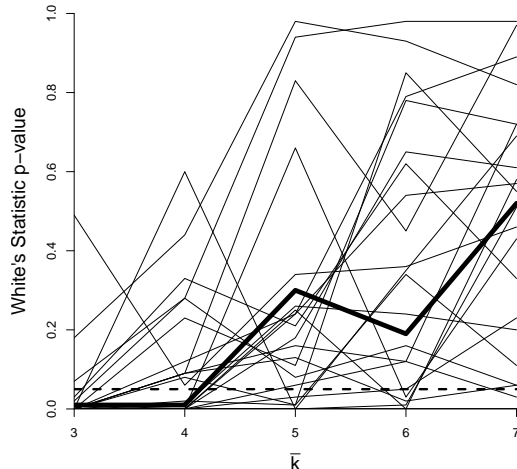
Figure 17: **White Statistic p-Value Profile Bundle.** The bold profile is Citigroup. For visual reference we include a horizontal line at p=0.05.

a time series model, IM tests focus on two properties of a correctly specified model. The so-called dynamic first-order IM tests assess whether the score process is serially uncorrelated. Under correct specification, the score process should be a martingale difference sequence. The so-called second-order IM tests mimic White (1982)'s test for *iid* data and assess the large-sample equivalence of the Hessian and the outer product of scores.

We report p-values for a second-order IM test (see White (1994, Section 11.1). We use all 10 non-redundant elements of the information matrix to form the test statistic $White$. Under the null hypothesis of correct specification the limit distribution takes the form $White \sim \chi_{10}^2$. In Figure 17 we show the profile bundle of White statistic p-values. There is rather wide variation, but there is also a clear pattern of increase in $White$ statistics with $\bar{k}$. Indeed for the preferred value of $\bar{k} = 7$, almost none of the twenty-five p-values are less than 0.05. Thus the second-order IM test generally reveals little or no evidence in the data against our preferred MSMD specification.

Bayesian Information Criterion for MSMD(7) vs. ACD(1,1). We now consider the performance of MSMD(7) relative to the canonical ACD(1,1) model. The ACD(1,1) model is

$$d_i = \varphi_i \epsilon_i \tag{17}$$

$$\varphi_i = \omega + \alpha d_{i-1} + \beta \varphi_{i-1}$$

$$\epsilon_i \sim iid Exp(1)$$
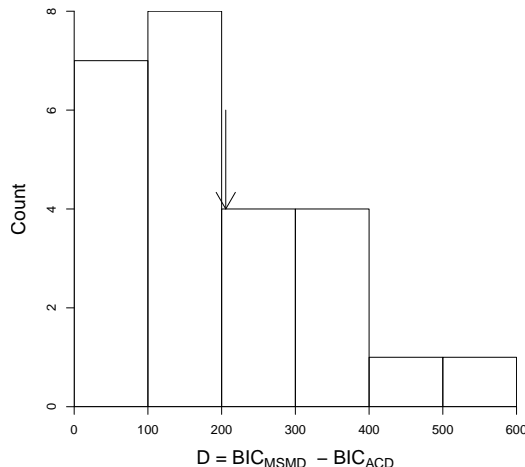
Figure 18: **Distribution of Differences in BIC Values Across Firms, $\gamma_{\bar{k}} = 7$.** We compute differences as MSMD - ACD. We show a histogram.

The ACD(1,1) has three parameters whereas our MSMD(7) model has four parameters. We estimate both models by maximum likelihood and use the Bayesian information criterion, $BIC = -2\ln L + k\ln(n)$, to adjust the maximized likelihood function for model complexity, that is, the number of estimated parameters.[18] In Figure 18 we show the distribution of BIC differences ($BIC_{MSMD} - BIC_{ACD}$). The differences are typically large, and the distribution is centered around 600. For Citigroup the BIC differential is 200. To put the number into perspective consider the following back-of-the-envelope calculation. The sample consists of 22,988 observations. Dividing the differential by the number of observations yields 0.013. Thus, from a Bayesian perspective, on average each duration observation shifts the odds in favor of the MSMD model by a factor of 1.013. Thus, while each individual observation only provides weak evidence in favor of our specification, the nearly 23,000 observations in combination provide overwhelming evidence in favor of the MSMD model.

Out-of-Sample Forecasting for MSMD(7) vs. ACD(1,1). BIC is an asymptotic approximation of a Bayesian marginal data density, which in turn can be interpreted as a measure of one-step-ahead predictive performance.[19] We now conduct a pseudo-out-of-

---

[18]For ease of interpretation and comparability to the maximized log likelihoods reported in Appendix C and Figure 18, we multiply the BIC by -1/2, which converts it to $BIC = \ln L - k\ln(n)/2$; that is, maximized log likelihood less a degrees-of-freedom penalty. Hence our BIC is in "log likelihood units," so that "bigger is better."

[19]The marginal data density is defined as $p(d_{1:n}) = \int p(d_{1:n}|\theta)d\theta$ and its logarithm can be decomposed as $p(d_{1:n}) = \sum \ln_{i=1}^{n} p(d_i|d_{1:i-1}) = \sum \ln_{i=1}^{n} \int p(d_i|d_{1:i-1},\theta)p(\theta|d_{1:i-1})d\theta$. Here $\int p(d_i|d_{1:i-1},\theta)p(\theta|d_{1:i-1})d\theta$ can be viewed as score for a one-step-ahead forecast of d_i based on information about $d_{1:i-1}$.

29

sample forecasting experiment and directly compute RMSEs for one-step-ahead as well as multi-step-ahead forecasts. Because our sample is large, we do not re-estimate the parameters of the MSMD and the ACD model for each forecast origin. Instead we simply estimate the models based on an initial sample $d_{1:n_0}$ and then keep the parameter estimates fixed for the subsequent forecast origins $i = n_0 + 1, \ldots, n$. For more than half of the stocks we have more than 11,000 inter-trade durations. In this case we choose $n_0 = 10,0000$. For all other stocks we choose n_0 such that $n - n_0 > 1,000$. Thus, the forecast evaluations are based on more than 1,000 forecasts for all stocks. The shortest estimation sample consists of $n_0 = 1,000$ observations. Detailed information in n and n_0 for each stock is provided in the Appendix.

Under the quadratic loss function implicit in the RMSE criterion, the best predictor is the conditional mean. For the ACD model we simply iterate the expression for φ_i, which represents the expected duration, forward using (17). The forecasting procedure for the MSMD model is slightly more complicated. Using our filter, we construct a posterior (based on the information available at the forecast origin n_*) for the intensity components: $p(M_{1,n_0}, \ldots, M_{\bar{k},n_0}|\hat{\theta}, d_{1:n_*})$. We then use the transition matrices $\mathcal{P}(\hat{\gamma}_k)$ to forecast the intensity components and compute the expected duration.

We summarize the results in Figure 19. The three panels show histograms of RMSE differences (MSMD(7) - ACD(1,1)) at forecast horizons $h = 1$, $h = 5$, and $h = 20$. For each horizon we are forecasting the duration d_{n_*+h}, which is the duration between trade n_*+h-1 and $n_* + h$ (as opposed to the duration between trade n_* and $n_* + h$). For Citigroup the RMSE differentials are -0.02, -2.50, and -2.00, respectively. Across stocks, the one-step-ahead forecasting performances of the MSMD and ACD models are comparable, and most of the differentials are close to zero. However, for five-step-ahead and 20-step-ahead forecasting the MSMD model dominates the ACD model for all 25 stocks. The forecasting gain from using MSMD is huge. For instance, for Citigroup the RMSE reduction is about 50%, clearly illustrating the benefits of MSMD's more-accurate approximation to longer-run dynamics.

5   Summary and Concluding Remarks

The inter-trade duration is an important and natural measure of market liquidity, and its variability is related to liquidity risk. In this paper, we proposed a new Markov-switching multi-fractal (MSMD) model of inter-trade durations. MSMD is a parameter-driven long-memory model of conditional intensity dynamics, with the long memory driven by structural markov-switching components. The popular standard ACD duration model neglects al+l of
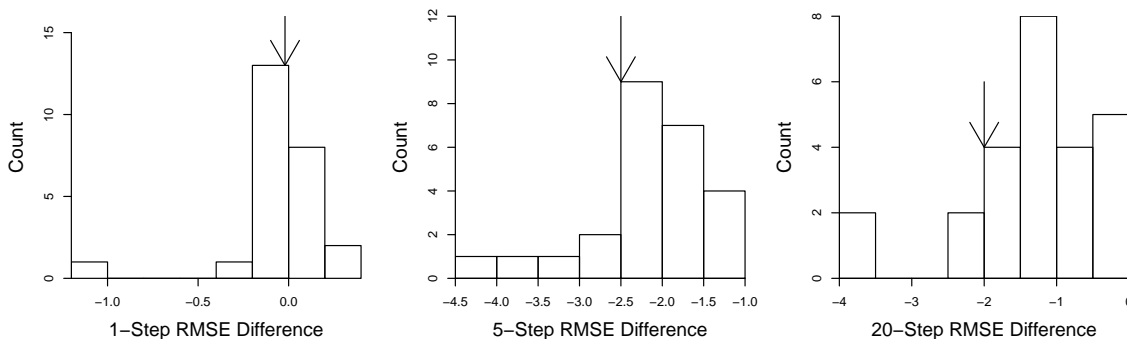
Figure 19: **Distribution of Differences in RMSE Across Firms.** Differences are computed as MSMD - ACD.

those features. A few other notable duration models such as SCD have featured them in isolation or in smaller assemblies, but none have featured them all. MSMD does so in a simple and parsimonious fashion, successfully capturing the key features of financial market inter-trade durations: long-memory dynamics and over-dispersed distributions. Both model selection criteria and out-of-sample forecast comparisons strongly favor MSMD relative to ACD. We leave it for future research to explore multivariate extensions and specifications tailored to confront the massive trade counts observed in recent years.

References

ADMATI, A., AND P. PFLEIDERER (1988): "A Theory of Intraday Trading Patterns: Volume and Price Variability," *Review of Financial Studies*, 1, 3–40.

ALIZADEH, S., M. BRANDT, AND F. DIEBOLD (2002): "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047–1091.

ANDERSEN, T., T. BOLLERSLEV, P. CHRISTOFFERSEN, AND F. DIEBOLD (2013): "Risk Measurement for Market Risk Management," in M. Harris, G. Constantinedes and R. Stulz (eds.), *Handbook of the Economics of Finance*, Elsevier.

ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society B*, 64, 253–280.

BAUWENS, L., AND D. VEREDAS (2004): "The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations," *Journal of Econometrics*, 119, 381–412.

BREIDT, F., N. CRATO, AND P. DE LIMA (1998): "The Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325–348.

BROWNLEES, C., F. CIPOLLINI, AND G. GALLO (2012): "Multiplicative Error Models," in L. Bauwens, C. Hafner and S. Laurent (eds.), *Volatility Models and Their Applications*, Wiley, in press.

CALVET, L., AND A. FISHER (2001): "Forecasting Multifractal Volatility," *Journal of Econometrics*, 105, 27–58.

——— (2004): "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes," *Journal of Financial Econometrics*, 2, 49–83.

——— (2008): *Multifractal Volatility: Theory, Forecasting, and Pricing*, Elsevier.

CLARK, P. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

COMTE, F., AND E. RENAULT (1998): "Long Memory in Continuous-Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323.

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

COX, D. (1981): "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93–115.

CREAL, D., S. KOOPMAN, AND A. LUCAS (2010): "Generalize Autoregressive Score Models with Applications," Manuscript, University of Chicago and Free University of Amsterdam.

DALEY, D., AND D. VERE-JONES (2003): *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, Springer.

——— (2007): *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, Springer.

DEO, R., M. HSIEH, AND C. HURVICH (2010): "Long Memory in Intertrade Durations, Counts and Realized Volatility of NYSE Stocks," *Journal of Statistical Planning and Inference*, 140, 3715–3733.

DIEBOLD, F., AND A. INOUE (2001): "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159.

EASLEY, D., AND M. O'HARA (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 576–605.

ENGLE, R. (2002): "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425–446.

ENGLE, R., AND J. RUSSELL (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

GALLANT, A. R., P. ROSSI, AND G. TAUCHEN (1993): "Nonlinear Dynamic Structures," *Econometrica*, 61, 871–907.

GHYSELS, E., C. GOURIÈROUX, AND J. JASIAK (2004): "Stochastic Volatility Duration Models," *Journal of Econometrics*, 119, 413–433.

GRANGER, C., AND R. JOYEUX (1980): "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of time series analysis*, 1, 15–29.

GRANGER, C. W. J. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

HAMILTON, J. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

HARVEY, A. (2011): "Exponential Conditional Volatility Models," Manuscript, University of Cambridge.

HASBROUCK, J. (2007): *Empirical Market Microstructure*, Oxford University Press.

HAUTSCH, N. (2008): "Capturing Common Components in High-Frequency Financial Time Series: A Multivariate Stochastic Multiplicative Error Model," *Journal of Economic Dynamics and Control*, 32, 3978–4015.

HAUTSCH, N. (2012): *Econometrics of Financial High-Frequency Data*. Springer Verlag, New York, *Econometrics of Financial High-Frequency Data*, Springer.

JASIAK, J. (1999): "Persistence in Intertrade Durations," Manuscript, Department of Economics, York University.

KARR, A. (1991): *Point Processes and their Statistical Inference*, Dekker.

MANDELBROT, B. (1997): *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer.

MANDELBROT, B., A. FISHER, AND L. CALVET (1997): "A Multifractal Model of Asset Returns," Cowles Foundation Discussion Paper 1164.

MEDDAHI, N., E. RENAULT, AND B. WERKER (1998): "Modeling High-Frequency Data in Continuous Time," Manuscript, University of Montreal.

MERTON, R. (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

OGATA, Y. (1988): "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9–27.

O'HARA, M. (1995): *Market Microstructure Theory*, Cambridge University Press.

VIVES, X. (2008): *Information and Learning in Markets: The Impact of Market Microstructure*, Princeton University Press.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

——— (1994): *Estimation Inference and Specification Analysis*. New York: Cambridge University Press.

Appendices

(For Web Publication Only)

A   Firm-by-Firm Intraday Calendar Effects

Ticker	Name	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	$\hat{a}_7$	$\hat{a}_8$	$\hat{a}_9$	$\hat{a}_{10}$	$\hat{a}_{11}$	$\hat{a}_{12}$
AA	Alcoa	3.61	3.58	3.64	3.91	4.22	3.90	4.41	4.24	4.02	4.21	4.06	3.70
ABT	Abbott Labs	2.38	2.29	2.51	2.51	2.68	2.75	2.56	2.82	2.75	2.65	2.56	2.44
AXP	Amex	2.45	2.63	2.79	2.84	2.99	3.10	2.99	3.00	2.93	2.99	3.02	2.82
BA	Boeing	2.37	2.43	2.60	2.53	2.59	2.67	2.81	2.85	2.58	2.60	2.60	2.41
BAC	B of A	3.11	3.15	3.30	3.38	3.40	3.41	3.34	3.40	3.32	3.12	3.16	3.15
C	Citigroup	1.89	1.92	2.02	2.27	2.40	2.37	2.39	2.54	2.40	2.29	2.32	2.13
CSCO	Cisco	2.03	1.98	2.39	2.23	2.41	2.54	2.53	2.64	2.59	2.36	2.19	2.31
DELL	Dell	1.67	1.79	1.93	1.97	2.23	2.28	2.36	2.28	2.26	2.01	2.13	2.03
DOW	Dow	3.20	3.23	3.37	3.34	3.49	3.79	3.61	3.54	3.52	3.43	3.25	3.27
F	Ford	2.32	2.32	2.48	2.56	2.62	2.62	2.67	2.66	2.33	2.52	2.49	2.39
GE	GE	2.40	2.42	2.59	2.54	2.68	2.79	2.76	2.84	2.85	2.60	2.54	2.47
HD	Home Depot	1.80	1.85	1.92	2.08	2.15	2.44	2.15	2.32	2.30	2.32	2.22	2.05
IBM	IBM	1.84	1.82	1.88	1.96	2.08	1.95	2.14	2.03	2.16	2.14	2.06	1.96
INTC	Intel	1.39	1.41	1.57	1.64	1.75	1.86	1.87	1.89	1.91	1.80	1.85	1.54
JNJ	J & J	2.15	2.16	2.20	2.24	2.33	2.45	2.46	2.49	2.43	2.34	2.17	2.13
KO	Coca-Cola	2.54	2.50	2.50	2.63	2.78	2.78	2.89	2.93	2.84	2.71	2.61	2.51
MCD	McDonald's	3.26	3.22	3.35	3.41	3.43	3.58	3.52	3.68	3.40	3.34	3.15	3.01
MRK	Merck	1.40	1.37	1.40	1.46	1.65	1.65	1.67	1.74	1.65	1.64	1.61	1.43
MSFT	Microsoft	1.79	1.89	1.90	1.80	1.94	2.19	2.04	2.19	2.09	1.91	1.97	1.72
QCOM	Qualcomm	3.35	3.60	3.64	3.91	3.61	2.26	3.32	3.73	3.97	3.64	3.94	3.25
T	AT&T	2.24	2.25	2.20	2.31	2.50	2.43	2.58	2.55	2.55	2.41	2.35	2.28
TXN	Texas Inst.	3.42	3.58	3.47	3.57	3.95	3.94	4.04	3.85	3.69	3.67	3.27	3.34
WFC	Wells Fargo	3.30	3.55	3.41	3.62	3.78	3.78	3.98	3.90	4.04	4.09	3.75	3.54
WMT	Wal-Mart	1.66	1.73	1.82	1.87	1.96	1.98	2.14	2.13	2.05	2.01	1.98	1.85
XRX	Xerox	3.89	3.80	3.87	4.06	4.06	4.07	4.14	4.35	4.12	3.98	4.02	3.85

Table 2: **Estimated Intraday Calendar Effects, Twenty-Five Firms Randomly Se-lected From S&P 100.** We show estimated coefficients on time-of-day dummies, for half-hour intervals from 10 AM to 4 PM, for February 1993. We order firms across rows alphabetically.

B Firm-by-Firm Descriptive Statistics

Ticker	Name	Mean	Med	Max	Min	Std	Skew	Kurt	OD	n	n_0
AA	Alcoa	2.66	1.28	39.58	0.01	3.77	3.07	17.09	1.42	2,989	1,000
ABT	Abbott Labs	1.86	1.11	26.30	0.06	2.21	2.84	15.89	1.19	16,929	10,000
AXP	Amex	2.22	1.17	42.33	0.05	2.93	3.13	19.35	1.32	10,531	9,000
BA	Boeing	1.82	1.13	25.00	0.06	2.03	2.50	12.94	1.12	17,111	10,000
BAC	B of A	1.97	1.12	27.14	0.03	2.44	2.90	15.69	1.24	7,939	6,000
C	Citigroup	1.93	1.01	32.67	0.08	2.63	3.59	22.46	1.36	22,988	1,0000
CSCO	Cisco	2.22	1.01	56.30	0.07	3.52	4.53	36.29	1.59	17,963	10,000
DELL	Dell	2.13	1.02	76.83	0.09	3.40	5.15	48.30	1.60	24,610	10,000
DOW	Dow Chemical	1.96	1.11	37.59	0.02	2.49	3.60	28.03	1.27	6,902	5,000
F	Ford	2.18	0.99	49.25	0.07	3.13	3.50	22.48	1.44	15,562	10,000
GE	GE	2.03	1.10	27.13	0.06	2.56	2.72	13.85	1.26	14,798	10,000
HD	Home Depot	1.97	1.00	38.62	0.09	2.65	3.47	22.28	1.35	25,113	10,000
IBM	IBM	1.75	1.06	35.87	0.12	2.03	3.01	19.16	1.16	31,895	10,000
INTC	Intel	1.81	1.00	50.38	0.15	2.41	4.17	34.57	1.33	41,957	10,000
JNJ	J & J	1.72	1.03	29.56	0.08	2.01	3.10	19.10	1.17	24,208	10,000
KO	Coca-Cola	1.82	1.14	26.31	0.05	2.06	2.49	12.47	1.13	15,542	10,000
MCD	McDonald's	1.93	1.15	22.17	0.03	2.26	2.58	12.77	1.17	7,441	6,000
MRK	Merck	1.61	0.98	24.66	0.18	1.78	2.95	17.31	1.11	54,242	10,000
MSFT	Microsoft	2.01	1.01	53.68	0.11	2.94	4.43	37.43	1.46	29,191	10,000
QCOM	Qualcomm	3.09	1.09	201.25	0.02	6.81	10.00	202.57	2.20	4,416	3,000
T	AT&T	1.79	1.05	22.78	0.08	2.09	2.62	13.26	1.16	21145	10,000
TXN	Texas Inst.	2.56	1.15	55.41	0.02	3.70	3.39	23.54	1.44	4,235	3,000
WFC	Wells Fargo	2.47	1.08	78.65	0.02	4.05	5.18	54.37	1.64	4,047	3,000
WMT	Wal-Mart	1.77	0.99	31.92	0.12	2.11	2.88	16.23	1.19	33,899	10,000
XRX	Xerox	2.50	1.18	31.80	0.01	3.42	2.83	14.62	1.37	2,933	1,000

Table 3: **Descriptive Statistics, Twenty-Five Firms Randomly Selected From S&P 100.** We show sample mean, median, standard deviation, skewness, kurtosis, over-dispersion (Std/Mean), sample size n (number of inter-trade durations), and size of estimation sample for pseudo-out-of-sample forecasting exercise (n_0) for inter-trade durations between 10 AM and 4 PM during February 1993. We order firms across rows alphabetically.

C Firm-by-Firm Estimation and Forecasting

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.47	1.47	1.47	1.47	1.47
	(0.02)	(0.08)	(0.02)	(0.02)	(0.02)
$\widehat{\lambda}$	0.83	0.65	1.23	2.35	4.47
	(0.06)	(0.43)	(0.24)	(0.40)	(0.81)
$\widehat{\gamma_{\bar{k}}}$	0.99	0.99	0.99	0.99	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	7.68	13.85	14.72	14.96	15.07
	(2.09)	(16.68)	(4.70)	(2.85)	(1.34)
$\ln L$	-5595.32	-5599.27	-5599.68	-5600.08	-5600.54
$White$	117.71	25.66	13.76	7.80	8.24
	(p=0.00)	(p= 0.00)	(p= 0.18)	(p= 0.65)	(p= 0.61)
BIC	–	–	–	–	11233.09
1-step RMSE	–	–	–	–	4.26
5-step RMSE	–	–	–	–	4.33
20-step RMSE	–	–	–	–	4.45
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-5783.64
BIC	–	–	–	–	11591.29
1-step RMSE	–	–	–	–	4.05
5-step RMSE	–	–	–	–	6.43
20-step RMSE	–	–	–	–	5.03

Table 4: **Model Estimation and Forecasting: AA**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. *White* is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.29	1.25	1.24	1.21	1.21
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.80	0.85	1.08	0.95	0.53
	(0.14)	(0.07)	(0.04)	(0.04)	(0.02)
$\widehat{\gamma_{\bar{k}}}$	0.07	0.13	0.13	0.18	0.17
	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)
$\widehat{b}$	13.89	8.53	7.28	4.81	5.24
	(1.09)	(1.56)	(0.72)	(0.53)	(0.22)
$\ln L$	-26342.34	-26329.71	-26326.88	-26324.39	-26328.48
$White$	39.00	16.39	15.01	21.18	17.70
	(p= 0.00)	(p= 0.09)	(p= 0.13)	(p= 0.02)	(p= 0.06)
BIC	–	–	–	–	52695.91
1-step RMSE	–	–	–	–	1.46
5-step RMSE	–	–	–	–	1.46
20-step RMSE	–	–	–	–	1.48
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-26434.75
BIC	–	–	–	–	52898.71
1-step RMSE	–	–	–	–	1.46
5-step RMSE	–	–	–	–	3.10
20-step RMSE	–	–	–	–	1.68

Table 5: **Model Estimation and Forecasting: ABT**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.38	1.34	1.30	1.31	1.27
	(0.48)	(0.02)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.76	0.95	0.87	1.61	0.81
	(1.12)	(0.26)	(0.07)	(0.15)	(0.04)
$\widehat{\gamma_{\bar{k}}}$	0.79	0.94	0.99	0.95	0.99
	(0.10)	(0.28)	(0.00)	(0.01)	(0.00)
$\widehat{b}$	30.70	29.10	12.69	20.68	10.24
	(0.59)	(4.69)	(1.62)	(0.18)	(1.08)
$\ln L$	-17888.32	-17844.68	-17835.35	-17843.72	-17829.99
$White$	29.73	28.11	7.64	19.56	10.16
	(p= 0.00)	(p= 0.00)	(p= 0.66)	(p= 0.03)	(p= 0.43)
BIC	–	–	–	–	35697.03
1-step RMSE	–	–	–	–	4.18
5-step RMSE	–	–	–	–	4.23
20-step RMSE	–	–	–	–	4.31
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-18042.24
BIC	–	–	–	–	36112.27
1-step RMSE	–	–	–	–	4.07
5-step RMSE	–	–	–	–	8.00
20-step RMSE	–	–	–	–	7.83

Table 6: **Model Estimation and Forecasting: AXP**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. *White* is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.30	1.25	1.26	1.24	1.22
	(0.05)	(0.01)	(0.02)	(0.01)	(0.02)
$\widehat{\lambda}$	1.17	1.06	1.63	0.74	0.82
	(0.25)	(0.05)	(0.17)	(0.03)	(0.09)
$\widehat{\gamma_{\bar{k}}}$	0.10	0.28	0.10	0.22	0.26
	(0.04)	(0.01)	(0.03)	(0.07)	(0.06)
$\widehat{b}$	26.21	11.86	9.70	10.44	6.77
	(26.69)	(2.62)	(1.47)	(1.35)	(1.31)
$\ln L$	-26717.54	-26712.29	-26714.19	-26705.30	-26703.08
$White$	26.55	26.13	20.41	18.28	12.83
	(p= 0.00)	(p= 0.00)	(p= 0.03)	(p= 0.05)	(p= 0.23)
BIC	–	–	–	–	53445.15
1-step RMSE	–	–	–	–	1.79
5-step RMSE	–	–	–	–	1.80
20-step RMSE	–	–	–	–	1.83
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-26802.11
BIC	–	–	–	–	53633.4
1-step RMSE	–	–	–	–	1.78
5-step RMSE	–	–	–	–	3.69
20-step RMSE	–	–	–	–	3.34

Table 7: **Model Estimation and Forecasting: BA**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.36	1.31	1.27	1.25	1.23
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	1.16	1.09	1.02	0.93	0.89
	(0.06)	(0.05)	(0.03)	(0.05)	(0.11)
$\widehat{\gamma}_{\bar{k}}$	0.42	0.86	0.99	0.98	0.99
	(0.06)	(0.25)	(0.00)	(0.02)	(0.00)
$\widehat{b}$	20.43	11.58	7.76	4.91	3.80
	(0.24)	(2.07)	(0.43)	(0.68)	(0.33)
$\ln L$	-12917.76	-12910.83	-12910.44	-12910.30	-12909.57
$White$	20.09	11.32	13.26	8.07	11.33
	(p=0.03)	(p=0.33)	(p=0.21)	(p=0.62)	(p=0.33)
BIC	–	–	–	–	25856.52
1-step RMSE	–	–	–	–	2.84
5-step RMSE	–	–	–	–	2.86
20-step RMSE	–	–	–	–	2.87
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-13073.60
BIC	–	–	–	–	26174.14
1-step RMSE	–	–	–	–	2.82
5-step RMSE	–	–	–	–	4.97
20-step RMSE	–	–	–	–	4.04

Table 8: **Model Estimation and Forecasting: BAC**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.33	1.30	1.26	1.26	1.23
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.79	0.98	0.88	0.70	1.17
	(0.08)	(0.03)	(0.05)	(0.27)	(0.10)
$\widehat{\gamma_{\bar{k}}}$	0.08	0.10	0.18	0.18	0.18
	(0.03)	(0.01)	(0.02)	(0.01)	(0.06)
$\widehat{b}$	14.33	11.06	6.14	6.16	4.52
	(11.23)	(0.39)	(0.55)	(2.29)	(1.61)
$\ln L$	-34872.83	-34844.23	-34850.18	-34850.67	-34845.03
White	25.01	22.89	11.81	13.63	9.12
	(p= 0.01)	(p= 0.01)	(p= 0.30)	(p= 0.19)	(p= 0.52)
BIC	–	–	–	–	69730.23
1-step RMSE	–	–	–	–	2.41
5-step RMSE	–	–	–	–	2.44
20-step RMSE	–	–	–	–	2.50
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-35055.65
BIC	–	–	–	–	70141.43
1-step RMSE	–	–	–	–	2.43
5-step RMSE	–	–	–	–	4.94
20-step RMSE	–	–	–	–	4.50

Table 9: **MSMD Model Estimation and Forecasting: C**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. *White* is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.42	1.37	1.34	1.40	1.29
	(0.00)	(0.01)	(0.08)	(0.01)	(0.02)
$\widehat{\lambda}$	0.73	0.82	0.84	2.74	0.83
	(0.03)	(0.04)	(2.77)	(0.17)	(0.04)
$\widehat{\gamma_{\bar{k}}}$	0.14	0.17	0.19	0.18	0.20
	(0.01)	(0.02)	(0.50)	(0.00)	(0.02)
$\widehat{b}$	2.65	2.47	2.02	3.73	1.68
	(0.80)	(0.34)	(1.88)	(0.32)	(0.08)
$\ln L$	-28592.08	-28588.80	-28595.00	-28608.54	-28600.31
$White$	45.80	43.11	47.90	33.29	36.75
	(p=0.00)	(p=0.00)	(p=0.00)	(p=0.00)	(p=0.00)
BIC	–	–	–	–	57239.8
1-step RMSE	–	–	–	–	1.77
5-step RMSE	–	–	–	–	1.81
20-step RMSE	–	–	–	–	1.86
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-29098.61
BIC	–	–	–	–	58226.61
1-step RMSE	–	–	–	–	1.81
5-step RMSE	–	–	–	–	3.24
20-step RMSE	–	–	–	–	1.93

Table 10: **Model Estimation and Forecasting: CSCO**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.40	1.36	1.37	1.32	1.32
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
$\widehat{\lambda}$	0.67	0.80	1.26	1.18	0.91
	(0.02)	(0.04)	(0.09)	(0.04)	(0.07)
$\widehat{\gamma_{\bar{k}}}$	0.08	0.09	0.10	0.11	0.10
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{b}$	4.18	3.45	3.80	2.75	2.59
	(0.98)	(0.66)	(0.55)	(0.15)	(0.30)
$\ln L$	-37957.67	-37903.42	-37909.02	-37899.56	-37898.92
$White$	63.21	35.57	35.94	11.09	7.37
	(p= 0.00)	(p= 0.00)	(p=0.00)	(p=0.35)	(p=0.69)
BIC	–	–	–	–	75838.28
1-step RMSE	–	–	–	–	2.58
5-step RMSE	–	–	–	–	2.65
20-step RMSE	–	–	–	–	2.71
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-38240.47
BIC	–	–	–	–	76511.27
1-step RMSE	–	–	–	–	2.65
5-step RMSE	–	–	–	–	4.71
20-step RMSE	–	–	–	–	3.88

Table 11: **Model Estimation and Forecasting: DELL**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.29	1.28	1.25	1.25	1.22
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.72	0.64	0.64	0.51	0.90
	(0.03)	(0.02)	(0.02)	(0.02)	(0.07)
$\widehat{\gamma_{\bar{k}}}$	0.59	0.87	0.99	0.99	0.99
	(0.10)	(0.05)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	16.52	15.47	9.26	9.38	6.89
	(0.41)	(0.44)	(1.37)	(1.18)	(0.27)
$\ln L$	-11159.40	-11149.76	-11146.89	-11147.64	-11145.02
$White$	13.80	10.04	3.01	4.40	5.92
	(p= 0.18)	(p= 0.44)	(p= 0.98)	(p= 0.93)	(p= 0.82)
BIC	–	–	–	–	22325.4
1-step RMSE	–	–	–	–	3.19
5-step RMSE	–	–	–	–	3.24
20-step RMSE	–	–	–	–	3.27
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-11238.01
BIC	–	–	–	–	22502.54
1-step RMSE	–	–	–	–	3.16
5-step RMSE	–	–	–	–	5.57
20-step RMSE	–	–	–	–	4.93

Table 12: **Model Estimation and Forecasting: DOW**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.43	1.37	1.35	1.32	1.30
	(0.01)	(0.01)	(0.04)	(0.01)	(0.01)
$\widehat{\lambda}$	1.00	1.00	0.94	0.91	1.12
	(0.09)	(0.32)	(1.16)	(0.06)	(0.02)
$\widehat{\gamma_{\bar{k}}}$	0.76	0.99	0.99	0.99	0.99
	(0.03)	(0.02)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	19.68	13.82	14.82	8.55	6.85
	(2.07)	(2.68)	(2.69)	(1.00)	(0.33)
$\ln L$	-25684.19	-25673.06	-25650.57	-25640.26	-25638.54
$White$	55.32	16.75	22.23	11.29	15.59
	(p= 0.00)	(p= 0.08)	(p= 0.01)	(p= 0.34)	(p= 0.11)
BIC	–	–	–	–	51315.69
1-step RMSE	–	–	–	–	2.64
5-step RMSE	–	–	–	–	2.70
20-step RMSE	–	–	–	–	2.75
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-26227.00
BIC	–	–	–	–	52482.96
1-step RMSE	–	–	–	–	2.57
5-step RMSE	–	–	–	–	5.05
20-step RMSE	–	–	–	–	4.65

Table 13: **Model Estimation and Forecasting: F**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.36	1.30	1.27	1.30	1.27
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.90	0.85	0.80	1.75	0.51
	(0.09)	(0.03)	(0.50)	(0.10)	(0.03)
$\widehat{\gamma_{\bar{k}}}$	0.99	0.99	0.99	0.99	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	30.54	8.94	5.61	9.66	6.56
	(3.12)	(0.54)	(0.41)	(0.65)	(0.46)
$\ln L$	-24406.93	-24404.04	-24406.14	-24409.90	-24410.93
$White$	56.46	41.13	26.16	22.58	9.20
	(p= 0.00)	(p= 0.00)	(p= 0.00)	(p= 0.01)	(p= 0.51)
BIC	–	–	–	–	48860.27
1-step RMSE	–	–	–	–	2.39
5-step RMSE	–	–	–	–	2.43
20-step RMSE	–	–	–	–	2.48
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-24766.13
BIC	–	–	–	–	49561.07
1-step RMSE	–	–	–	–	2.37
5-step RMSE	–	–	–	–	4.53
20-step RMSE	–	–	–	–	3.91

Table 14: **Model Estimation and Forecasting: GE**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.31	1.28	1.22	1.20	1.20
	(0.01)	(0.10)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.76	0.90	0.71	0.80	0.67
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
$\widehat{\gamma_{\bar{k}}}$	0.05	0.06	0.06	0.09	0.09
	(0.01)	(0.01)	(0.05)	(0.01)	(0.02)
$\widehat{b}$	23.46	20.34	7.53	6.99	6.94
	(4.73)	(3.10)	(6.60)	(0.85)	(1.28)
$\ln L$	-38439.49	-38402.59	-38378.43	-38370.54	-38371.06
$White$	51.38	30.45	12.69	8.96	8.61
	(p= 0.00)	(p= 0.00)	(p= 0.24)	(p= 0.54)	(p= 0.57)
BIC	–	–	–	–	76782.64
1-step RMSE	–	–	–	–	0.95
5-step RMSE	–	–	–	–	0.95
20-step RMSE	–	–	–	–	0.96
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-38474.64
BIC	–	–	–	–	76979.67
1-step RMSE	–	–	–	–	0.95
5-step RMSE	–	–	–	–	2.15
20-step RMSE	–	–	–	–	2.03

Table 15: **Model Estimation and Forecasting: HD**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.27	1.24	1.21	1.21	1.18
	(0.01)	(0.01)	(0.03)	(0.12)	(0.01)
$\widehat{\lambda}$	0.82	0.87	0.84	1.06	0.77
	(0.02)	(0.03)	(0.29)	(0.33)	(0.03)
$\widehat{\gamma_{\bar{k}}}$	0.05	0.05	0.05	0.05	0.06
	(0.00)	(0.00)	(0.01)	(0.04)	(0.02)
$\widehat{b}$	10.12	5.93	3.90	4.00	2.85
	(0.99)	(0.56)	(1.91)	(3.64)	(0.31)
$\ln L$	-47744.73	-47710.78	-47696.20	-47697.22	-47695.07
$White$	40.16	16.37	5.86	9.93	3.42
	(p=0.00)	(p= 0.09)	(p= 0.83)	(p= 0.45)	(p= 0.97)
BIC	–	–	–	–	95431.62
1-step RMSE	–	–	–	–	1.68
5-step RMSE	–	–	–	–	1.67
20-step RMSE	–	–	–	–	1.69
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-47794.69
BIC	–	–	–	–	95620.49
1-step RMSE	–	–	–	–	1.68
5-step RMSE	–	–	–	–	3.57
20-step RMSE	–	–	–	–	2.84

Table 16: **Model Estimation and Forecasting: IBM**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.32	1.29	1.30	1.24	1.22
	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
$\widehat{\lambda}$	0.76	0.71	1.00	0.74	0.76
	(0.02)	(0.01)	(0.03)	(0.04)	(0.04)
$\widehat{\gamma_{\bar{k}}}$	0.07	0.08	0.09	0.09	0.10
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
$\widehat{b}$	3.86	3.42	3.97	2.21	1.92
	(0.48)	(0.27)	(0.37)	(0.19)	(0.06)
$\ln L$	-61934.00	-61888.79	-61895.91	-61891.47	-61893.35
$White$	76.93	83.01	83.32	64.88	67.82
	(p=0.00)	(p= 0.00)	(p= 0.00)	(p= 0.00)	(p= 0.00)
BIC	–	–	–	–	123829.3
1-step RMSE	–	–	–	–	2.72
5-step RMSE	–	–	–	–	2.79
20-step RMSE	–	–	–	–	2.92
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-62030.70
BIC	–	–	–	–	124093.3
1-step RMSE	–	–	–	–	2.75
5-step RMSE	–	–	–	–	5.08
20-step RMSE	–	–	–	–	3.84

Table 17: **Model Estimation and Forecasting: INTC**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.23	1.18	1.18	1.18	1.18
	(0.00)	(0.01)	(0.01)	(0.01)	(0.03)
$\widehat{\lambda}$	0.75	0.71	0.60	0.51	0.90
	(0.01)	(0.02)	(0.01)	(0.01)	(0.04)
$\widehat{\gamma_{\bar{k}}}$	0.08	0.12	0.12	0.12	0.12
	(0.02)	(0.03)	(0.01)	(0.04)	(0.00)
$\widehat{b}$	21.23	10.25	10.25	10.24	10.33
	(2.34)	(1.01)	(0.40)	(1.64)	(0.83)
$\ln L$	-35902.36	-35890.60	-35891.17	-35891.85	-35891.62
$White$	9.45	17.73	11.24	11.02	9.75
	(p= 0.49)	(p= 0.06)	(p= 0.34)	(p= 0.36)	(p= 0.46)
BIC	–	–	–	–	71823.62
1-step RMSE	–	–	–	–	1.79
5-step RMSE	–	–	–	–	1.79
20-step RMSE	–	–	–	–	1.78
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-35961.84
BIC	–	–	–	–	71953.96
1-step RMSE	–	–	–	–	1.82
5-step RMSE	–	–	–	–	3.75
20-step RMSE	–	–	–	–	3.01

Table 18: **Model Estimation and Forecasting: JNJ**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.24	1.21	1.20	1.21	1.18
	(0.01)	(0.52)	(0.01)	(0.34)	(0.01)
$\widehat{\lambda}$	0.78	0.76	0.66	0.55	0.56
	(0.02)	(3.80)	(0.02)	(0.35)	(0.09)
$\widehat{\gamma_{\bar{k}}}$	0.33	0.65	0.72	0.75	0.85
	(0.05)	(0.49)	(0.09)	(0.57)	(0.02)
$\widehat{b}$	12.02	7.21	6.78	7.17	4.69
	(2.21)	(2.89)	(1.11)	(8.03)	(0.64)
$\ln L$	-24426.66	-24423.53	-24422.53	-24423.63	-24424.36
$White$	25.44	23.20	12.45	12.75	13.51
	(p=0.00)	(p= 0.01)	(p= 0.26)	(p= 0.24)	(p= 0.20)
BIC	–	–	–	–	48887.33
1-step RMSE	–	–	–	–	1.74
5-step RMSE	–	–	–	–	1.74
20-step RMSE	–	–	–	–	1.76
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-24534.83
BIC	–	–	–	–	49098.61
1-step RMSE	–	–	–	–	1.74
5-step RMSE	–	–	–	–	3.37
20-step RMSE	–	–	–	–	2.57

Table 19: **Model Estimation and Forecasting: KO**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.29	1.29	1.24	1.24	1.24
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.79	0.61	0.62	0.50	1.08
	(0.03)	(0.02)	(0.01)	(0.01)	(0.03)
$\widehat{\gamma_{\bar{k}}}$	0.99	0.99	0.99	0.99	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	34.56	35.43	10.32	10.37	10.46
	(5.50)	(5.54)	(0.53)	(0.45)	(0.91)
$\ln L$	-12094.83	-12095.83	-12097.68	-12098.48	-12098.37
$White$	17.27	12.02	16.77	14.35	17.55
	(p= 0.07)	(p= 0.28)	(p= 0.08)	(p= 0.16)	(p= 0.06)
BIC	–	–	–	–	24232.4
1-step RMSE	–	–	–	–	2.48
5-step RMSE	–	–	–	–	2.51
20-step RMSE	–	–	–	–	2.51
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-12201.87
BIC	–	–	–	–	24430.48
1-step RMSE	–	–	–	–	2.47
5-step RMSE	–	–	–	–	4.54
20-step RMSE	–	–	–	–	3.59

Table 20: **Model Estimation and Forecasting: MCD**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. *White* is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.22	1.20	1.17	1.17	1.17
	(0.04)	(0.01)	(0.00)	(0.00)	(0.00)
$\widehat{\lambda}$	0.71	0.83	0.75	0.64	0.78
	(0.10)	(0.05)	(0.01)	(0.01)	(0.01)
$\widehat{\gamma_{\bar{k}}}$	0.03	0.03	0.04	0.04	0.04
	(0.04)	(0.00)	(0.01)	(0.00)	(0.01)
$\widehat{b}$	16.59	8.12	5.97	5.94	5.96
	(1.32)	(0.21)	(0.83)	(0.67)	(0.66)
$\ln L$	-77426.73	-77402.21	-77374.72	-77375.22	-77375.26
$White$	29.02	16.47	14.41	15.28	7.02
	(p=0.00)	(p= 0.09)	(p= 0.16)	(p= 0.12)	(p= 0.72)
BIC	–	–	–	–	154794.1
1-step RMSE	–	–	–	–	0.95
5-step RMSE	–	–	–	–	0.96
20-step RMSE	–	–	–	–	0.95
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-77449.99
BIC	–	–	–	–	154932.7
1-step RMSE	–	–	–	–	0.96
5-step RMSE	–	–	–	–	2.13
20-step RMSE	–	–	–	–	1.69

Table 21: **Model Estimation and Forecasting: MRK**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.38	1.34	1.32	1.29	1.30
	(0.03)	(0.01)	(0.03)	(0.00)	(0.01)
$\widehat{\lambda}$	0.78	0.85	0.72	0.85	0.68
	(0.16)	(0.02)	(0.02)	(0.07)	(0.03)
$\widehat{\gamma}_{\bar{k}}$	0.09	0.10	0.11	0.12	0.12
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
$\widehat{b}$	4.29	3.18	3.04	2.54	2.70
	(0.20)	(0.23)	(0.28)	(0.17)	(0.16)
$\ln L$	-44695.91	-44641.79	-44635.38	-44630.75	-44633.69
$White$	60.03	12.87	15.57	6.39	7.04
	(p=0.00)	(p=0.23)	(p=0.11)	(p=0.78)	(p= 0.72)
BIC	–	–	–	–	89308.51
1-step RMSE	–	–	–	–	3.13
5-step RMSE	–	–	–	–	3.22
20-step RMSE	–	–	–	–	3.29
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-44978.82
BIC	–	–	–	–	89988.48
1-step RMSE	–	–	–	–	3.20
5-step RMSE	–	–	–	–	6.07
20-step RMSE	–	–	–	–	5.51

Table 22: **Model Estimation and Forecasting: MSFT**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.58	1.51	1.45	1.43	1.44
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
$\widehat{\lambda}$	1.14	1.06	1.11	0.89	0.63
	(0.10)	(0.07)	(0.10)	(0.07)	(0.13)
$\widehat{\gamma_{\bar{k}}}$	0.18	0.21	0.22	0.21	0.23
	(0.03)	(0.02)	(0.03)	(0.03)	(0.16)
$\widehat{b}$	4.30	3.44	2.45	2.24	2.43
	(1.01)	(0.38)	(0.28)	(0.29)	(0.73)
$\ln L$	-7597.95	-7568.28	-7564.41	-7562.45	-7564.50
$White$	41.27	15.36	12.69	6.32	5.08
	(p= 0.00)	(p= 0.12)	(p= 0.24)	(p= 0.79)	(p= 0.89)
BIC	–	–	–	–	15162.57
1-step RMSE	–	–	–	–	5.58
5-step RMSE	–	–	–	–	6.36
20-step RMSE	–	–	–	–	6.91
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-7871.92
BIC	–	–	–	–	15769.02
1-step RMSE	–	–	–	–	5.92
5-step RMSE	–	–	–	–	10.63
20-step RMSE	–	–	–	–	10.80

Table 23: **Model Estimation and Forecasting: QCOM**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.27	1.25	1.25	1.25	1.22
	(0.01)	(0.04)	(0.07)	(0.54)	(0.01)
$\widehat{\lambda}$	0.80	0.88	0.71	0.56	0.56
	(0.04)	(0.04)	(0.35)	(1.65)	(0.03)
$\widehat{\gamma_{\bar{k}}}$	0.29	0.48	0.48	0.48	0.73
	(0.15)	(0.05)	(0.76)	(0.10)	(0.04)
$\widehat{b}$	13.55	10.42	10.65	10.68	6.61
	(4.05)	(0.97)	(1.73)	(1.68)	(1.18)
$\ln L$	-32707.60	-32684.19	-32685.15	-32685.87	-32685.19
$White$	49.12	29.23	17.68	15.33	19.86
	(p=0.00)	(p= 0.00)	(p= 0.06)	(p= 0.12)	(p= 0.03)
BIC	–	–	–	–	65410.22
1-step RMSE	–	–	–	–	1.23
5-step RMSE	–	–	–	–	1.24
20-step RMSE	–	–	–	–	1.24
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-32886.98
BIC	–	–	–	–	65803.84
1-step RMSE	–	–	–	–	1.28
5-step RMSE	–	–	–	–	2.49
20-step RMSE	–	–	–	–	1.52

Table 24: **Model Estimation and Forecasting: T**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.48	1.42	1.45	1.44	1.45
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
$\widehat{\lambda}$	0.83	0.86	1.89	2.60	2.36
	(0.05)	(0.04)	(0.27)	(0.51)	(0.32)
$\widehat{\gamma_{\bar{k}}}$	0.99	0.99	0.99	0.99	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	8.23	4.75	8.53	8.70	8.63
	(0.79)	(0.55)	(1.04)	(1.22)	(1.77)
$\ln L$	-7689.50	-7688.41	-7693.84	-7695.12	-7694.20
$White$	22.17	98.37	12.57	38.89	8.54
	(p=0.01)	(p= 0.00)	(p= 0.25)	(p= 0.00)	(p= 0.58)
BIC	–	–	–	–	15421.8
1-step RMSE	–	–	–	–	3.51
5-step RMSE	–	–	–	–	3.57
20-step RMSE	–	–	–	–	3.66
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-7974.67
BIC	–	–	–	–	15974.39
1-step RMSE	–	–	–	–	3.34
5-step RMSE	–	–	–	–	5.35
20-step RMSE	–	–	–	–	3.92

Table 25: **Model Estimation and Forecasting: TXN**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.46	1.41	1.36	1.36	1.36
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\widehat{\lambda}$	0.81	0.72	0.74	1.15	0.86
	(0.07)	(0.06)	(0.06)	(0.10)	(0.02)
$\widehat{\gamma_{\bar{k}}}$	0.94	0.99	0.99	0.99	0.99
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	14.25	9.86	5.64	5.77	5.88
	(0.96)	(0.91)	(0.51)	(0.31)	(0.69)
$\ln L$	-7059.32	-7053.53	-7049.78	-7051.41	-7051.57
$White$	21.49	12.04	4.17	2.99	2.86
	(p= 0.02)	(p= 0.28)	(p= 0.94)	(p= 0.98)	(p= 0.98)
BIC	–	–	–	–	14136.36
1-step RMSE	–	–	–	–	1.77
5-step RMSE	–	–	–	–	1.78
20-step RMSE	–	–	–	–	1.78
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-7291.16
BIC	–	–	–	–	14607.24
1-step RMSE	–	–	–	–	2.92
5-step RMSE	–	–	–	–	4.89
20-step RMSE	–	–	–	–	3.64

Table 26: **Model Estimation and Forecasting: WFC**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$
$\widehat{m_0}$	1.26	1.22	1.20	1.17	1.16
	(0.03)	(0.00)	(0.00)	(0.00)	(0.01)
$\widehat{\lambda}$	0.75	0.83	0.83	0.80	0.81
	(0.37)	(0.02)	(0.01)	(0.02)	(0.03)
$\widehat{\gamma_{\bar{k}}}$	0.03	0.06	0.21	0.13	0.22
	(0.01)	(0.04)	(0.05)	(0.03)	(0.11)
$\widehat{b}$	10.97	18.38	10.48	6.80	5.76
	(1.33)	(11.31)	(1.52)	(0.93)	(1.07)
$\ln L$	-50726.70	-50654.46	-50648.08	-50643.79	-50639.69
$White$	94.54	20.87	22.78	5.54	8.80
	(p= 0.00)	(p= 0.02)	(p= 0.01)	(p= 0.85)	(p= 0.55)
BIC	–	–	–	–	101321.1
1-step RMSE	–	–	–	–	2.99
5-step RMSE	–	–	–	–	3.01
20-step RMSE	–	–	–	–	3.03
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-50774.41
BIC	–	–	–	–	101580.1
1-step RMSE	–	–	–	–	3.02
5-step RMSE	–	–	–	–	5.33
20-step RMSE	–	–	–	–	4.16

Table 27: **Model Estimation and Forecasting: WMT**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. $White$ is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.

	$\bar{k} = 3$	$\bar{k} = 4$	$\bar{k} = 5$	$\bar{k} = 6$	$\bar{k} = 7$
$\widehat{m_0}$	1.52	1.52	1.53	1.52	1.52
	(0.02)	(0.02)	(0.10)	(0.02)	(0.03)
$\widehat{\lambda}$	1.53	3.23	0.76	4.43	2.93
	(0.19)	(0.47)	(0.35)	(0.94)	(1.12)
$\widehat{\gamma_{\bar{k}}}$	0.99	0.99	0.99	0.99	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\widehat{b}$	25.91	27.43	34.11	27.48	27.58
	(6.18)	(2.91)	(37.63)	(6.93)	(1.58)
$\ln L$	-5360.69	-5361.73	-5362.07	-5361.97	-5361.94
White	30.32	8.34	31.07	28.50	29.34
	(p= 0.00)	(p= 0.60)	(p= 0.00)	(p= 0.00)	(p= 0.00)
BIC	–	–	–	–	10755.82
1-step RMSE	–	–	–	–	3.65
5-step RMSE	–	–	–	–	3.69
20-step RMSE	–	–	–	–	3.71
ACD Comparison:	–	–	–	–	–
$\ln L$	–	–	–	–	-5552.51
BIC	–	–	–	–	11128.97
1-step RMSE	–	–	–	–	3.43
5-step RMSE	–	–	–	–	5.37
20-step RMSE	–	–	–	–	4.04

Table 28: **Model Estimation and Forecasting: XRX**. We report estimation results for the MSMD model with $\bar{k}$ intensity components. Standard errors appear in parentheses beneath estimated parameters. *White* is White's omnibus information matrix test of model specification adequacy, with marginal significance levels in parentheses. See text for details.