

# A Compound Duration Model for High-Frequency Asset Returns

October 11, 2016

## Abstract

This paper builds a model of high-frequency equity returns by separately modeling the dynamics of trade-time returns and trade arrivals. Our main contributions are threefold. First, we characterize the distributional behavior of high-frequency asset returns both in ordinary clock time and in trade time. We show that when controlling for pre-scheduled market news events, trade-time returns of the near-month E-mini S&P 500 futures contract are well characterized by a Gaussian distribution at very fine time scales. Second, we develop a structured and parsimonious model of clock-time returns using a time-changed Brownian motion composed with a general, non-Lévy directing process. Particular cases of this model allow for leptokurtosis and volatility clustering in clock-time returns, even when trade-time returns are Gaussian. Finally, we highlight conditions for the directing process which are required in order to generate proper volatility dynamics while simultaneously matching the unconditional distribution of returns. In-sample fitting and out-of-sample realized volatility forecasting demonstrate the strength of our model relative to leading candidates.

**Keywords:** High-frequency trading, US Equities, News arrival.

**JEL Classification:** C22, C41, C58, G12, G14, G17.

# 1 Introduction

Modern electronic exchanges function in a manner that outwardly display the properties that are expected of an efficient, liquid market. Bid-offer spreads are narrow in comparison to the price of the underlying instrument being traded, volumes are high, high-frequency traders compete to make markets, and information regarding price discovery is disseminated at nearly the speed of light (see, e.g. Brogaard et al. (2014) and Hasbrouck and Saar (2013)). Within such an environment, the disparate, shifting spectrum of intentions of a wide range of market participants is continuously being aggregated, and so a naive, but nonetheless reasonable, expectation is that the Central Limit Theorem should play a fundamental role, and that short-period returns should adhere to a Gaussian distribution.

Indeed, from the pioneering work of Bachelier (1900) through the development of the Black-Scholes options pricing model (Black and Scholes (1973)), modern finance has traditionally held that market price movements can be approximated to a somewhat useful degree by a Gaussian random walk. In reality, observed distributions of market returns are markedly non-Gaussian. Regardless of venue and asset class, returns distributions invariably have fat tails and display the phenomenon of volatility clustering. A rich literature exists which describes both the characterization and the modeling of the observed departures from normality (for a review, see Bouchaud (2005)).

In this contribution, we carry out a ground-level re-examination of the process that generates short-period market returns within the context of high-frequency trading (over time scales ranging from milliseconds to minutes). We analyze a full year of recent, millisecond-resolution tick data from the extremely liquid near-month E-mini S&P 500 futures contract traded at the Chicago Mercantile Exchange.

Our analysis begins with a significant empirical insight: during our data sample period, Jan – Dec 2014, outside of pre-scheduled news announcement periods, high-frequency asset returns are well described by a Gaussian distribution when trade time is employed. Brada et al. (1966) introduced the notion of trade time to show that asset returns distributions are nearly Gaussian if the returns process is subordinated with successive transactions (trades) acting as the subordinator. Mandelbrot and Taylor (1967) showed that a Gaussian random walk composed with a subordinating trade-time process is fully consistent with a fat-tailed, Lévy-stable distribution, as suggested in Mandelbrot (1963). Clark (1973) used an alternative subordinator, time measured by volume of transactions, to obtain similar results. More recently, Ane and Geman (2000) shows that coarsely sampled intra-day returns also conform to a Gaussian distribution when measured in trade (transaction) time.

Our analysis demonstrates that the Gaussianity of trade-time returns does not immediately extend to high-frequency intra-day returns. That is, high-frequency, trade-time returns exhibit heavy tails and very weak volatility clustering when considered unconditionally throughout the day. However, when filtering price jumps by excluding the periods surrounding pre-scheduled news events, we confirm the existence of trade-time Gaussianity as well as a lack of volatility persistence. It is important to acknowledge that while removing jumps via news filtering is appropriate during our 2014 sample period, it may not be suitable for more volatile market periods outside of that time span.

Building on the empirical observations above, our paper makes two theoretical contributions. First, we develop a general time-changed Brownian motion model for high-frequency asset returns, similar to those of Press (1967), Mandelbrot and Taylor (1967) and Clark (1973), but allowing for more general (non-Lévy) directing processes. In particular, we allow the directing process to be characterized by inter-trade durations which follow the autoregressive conditional duration (ACD) model of Engle and Russell (1998) as well as the Markov-Switching Multifractal Duration (MSMD)

model of [Chen et al. \(2013\)](#) and [Zikes et al. \(2014\)](#), which builds on the work of [Mandelbrot et al. \(1997\)](#), [Calvet et al. \(1997\)](#) and [Fisher et al. \(1997\)](#), as well as subsequent work by [Calvet and Fisher \(2001\)](#), [Calvet and Fisher \(2002\)](#) and [Calvet \(2004\)](#). In this dimension our work differs substantially from that of [Ane and Geman \(2000\)](#): where they begin with a nonparametric estimate of the distribution of clock-time asset returns and work backwards to implicitly define the nonparametric density of trades that would be consistent with trade-time Gaussianity, we work forwards by first compounding a parametric distribution of trade-time returns with a parametric model of duration times (and hence, an associated trade arrival process) to characterize the distribution of clock-time returns. Our contribution is significant because it promotes a structured and parsimonious approach to approximating the observed evolution of asset returns.

Second, we develop conditions under which the directing process can generate realistic dynamics for asset returns. The early work on subordinated Brownian motion, cited above, focused entirely on the unconditional distribution of returns. Our work, however, shows that this same class of models is flexible enough to capture volatility persistence when the proper directing process is used. In a similar vein, [Carr and Wu \(2004\)](#) show how this class of models can account for the leverage effect.

The financial econometrics literature has utilized stochastic time changes to explain volatility dynamics, but this has typically been done by working with volatility directly. [Madan and Seneta \(1990\)](#) is an early example of this. More recently, [Andersen et al. \(2007\)](#), [Andersen et al. \(2010\)](#) and [Todorov and Tauchen \(2014\)](#) “devolatize” intra-day clock-time returns using short-term volatility measures in order to achieve conditional Gaussianity. Such devolatization devices are akin to stochastic time changes in that they control for the random flow of information and assume that the underlying returns process is Gaussian. The upshot is that they use latent volatility as a surrogate for information content while we use the observed transaction record. Our theoretical choice to focus on transactions is a result of the strong empirical evidence suggesting that this is the proper device for time deformation.

Using Monte Carlo simulations, we show that our compound duration model is a good characterization of observed clock-time returns and that the stochastic transformation between clock time and trade time for our data is most effectively explained using ACD and MSMD durations. In particular, we highlight the in-sample strengths of both versions of our model (ACD and MSMD durations) relative to a benchmark compound Poisson, as well as a more traditional GARCH model that has been adapted to high-frequency data ([Engle \(2000\)](#)). We also conduct an out-of-sample realized volatility forecasting exercise which demonstrates that despite its high degree of parameterization, the compound MSMD model significantly outperforms competing models.

Our paper proceeds as follows. We begin by describing our data in Section 2 and provide an analysis of the distributional characteristics of the data during news-affected and non-news-affected subperiods in Section 3. In Section 4, we describe the model and determine conditions under which it can produce volatility dynamics. Section 5 estimates the model and compares Monte Carlo simulations with observed data while Section 6 reports out-of-sample realized volatility forecasting results. Section 7 concludes.

## 2 Data

In this paper we focus our analysis exclusively on the Chicago Mercantile Exchange (CME) near-month E-mini S&P 500 Futures contract (commodity ticker symbol ES). Although the CME provides a variety of E-mini products, the E-mini S&P 500 futures contract is the most heavily traded, and for this reason it is commonly referred to as *the* E-mini. As indicated by its name, the E-mini

(ES) is a futures contract that trades at 1/5th the size of the standard S&P 500 futures contract. It has a notional value of 50 times the index. We obtained the full record of trades for the period 1 January 2014 to 31 December 2014 by parsing the CME historical files, encoded in FIX format, which we use to estimate and evaluate our model in Section 5.

Despite the fact that the E-mini is a futures contract that does not trade on equities exchanges, its statistical behavior characterizes the dynamics of equities markets as a whole. This is attributed to its liquidity and the relationship of price formation and information transmission between the futures and equities exchanges in Illinois and New Jersey, as studied in Laughlin et al. (2014).

E-mini futures trade Monday through Friday, starting at 5:00 p.m. Central Time on the previous day and ending at 4:15 p.m., with an additional daily maintenance trading halt from 3:15 p.m. to 3:30 p.m., Central Time. We aggregate multiple trades occurring within single milliseconds as unit transactions and assign to them the final, in-force price of the millisecond as the price of the trade. While not a perfect approximation, this assumption exploits the fact that multiple transactions with the same time stamp are nearly always attributable to a single aggressor order filling several resting orders at the same price level and also allows us to circumvent singularities associated with zero durations in our subsequent models. Our resulting data for the sample period contains a total of 25,095,624 such transactions.

The quoted price of the E-mini corresponds to the index value of the S&P 500. The minimum quoted price increment is  $\Delta P = 0.25$  index points, which corresponds to an actual increment of  $\$50 \times 0.25 = \$12.50$  for a single contract. For the remainder of the paper we will use *quoted* E-mini prices, measured in points, which correspond directly to the S&P 500 index value. Between January and December of 2014, the S&P 500 Index traded between  $P_{min} = 1732$  and  $P_{max} = 2088.75$ , indicating a typical minimum percentage increment in the index price of  $100 \times \frac{\Delta P}{(P_{min}+P_{max})/2} \sim 0.013\%$ .

Because we are interested in investigating the distributions of intra-day asset returns during news event periods and non-event periods, we consider subsamples of the data that sort according to news events. Since the E-mini is a futures contract on a market aggregate, it is almost exclusively affected by major macroeconomic announcements, and not by smaller scale, industry- or firm-specific news. For this reason, we classify news event periods with the EconoDay calendar ([econoday.com](http://econoday.com)), which lists major pre-scheduled news announcements in the U.S. and which powers calendars for outlets such as Bloomberg and the Wall Street Journal. In particular, we focus attention on the EconoDay ‘Market Moving Indicator’ and ‘Merit Extra Attention’ series, as past experience dictates that these are the events most relevant to markets. We build an event-driven dataset of all E-mini trades that occur during 1000-second (roughly 16-minute) windows following each news announcement during equities market hours, 9:30 a.m. – 4:00 p.m. ET, on all trading days in our data sample. A 1000-second window encompasses participants ranging from the fastest algorithmic traders to human traders who manually read the news, consider its implications, and trade on their resulting conclusions. We build a corresponding non-event-driven dataset that consists of all trades occurring outside of the 1000-second news windows.

As explained in Section 2, we also consider subsamples of the event-driven and non-event-driven data that confine attention to the 1000-second time window following 10:00 a.m. on each trading day, the most common news announcement time during regular market hours. That is, the event-driven subsample is comprised of trades from 10:00:00 – 10:16:40 a.m. for days on which there was a pre-scheduled 10:00 a.m. news announcement. The corresponding non-event-driven subsample constitutes all trades during the same time window on days when announcements were not scheduled. This latter subsample evades possible time-of-day effects and also results in greater balance in the number of observations in the event and non-event data.

For the remainder of the paper, we will refer to the event-driven subsample as the *active* data and the non-event subsample as the *passive* data. Our initial separation on news resulted in 179 days contributing news-affected returns and 252 days (all days in the sample period) contributing non-news-affected returns. However, due to complications with the timing of event windows<sup>1</sup>, we discarded the 24 days corresponding to the University of Michigan Consumer Sentiment reports. The resulting active dataset spans approximately 62 hours with 1,252,020 transaction records while the passive dataset spans roughly 1,412 hours with 15,556,531, or twelve times as many, records. In contrast, the reduced datasets (10:00 a.m. subsamples) contain 113 active periods and 121 passive periods (a little over 30 hours, each), corresponding to a total of 680,014 records and 650,582 records, respectively. Thus, our subsequent results that make use of the reduced samples are not an artifact of within-day cyclicalities (both data sets are drawn from the same intra-day time periods) or sample size.

### 3 Empirical Distributions of Intra-Day Asset Returns

In this section we emphasize some of the key features of observed intra-day asset returns distributions during our 2014 sample period. We define clock-time returns as

$$r_\delta(t) = p(t) - p(t - \delta) \quad (1)$$

where  $p(t)$  is the log price of an asset at time  $t$  and  $\delta$  is the clock-time duration under consideration (such as 1000 milliseconds). An alternative definition of returns can be provided in trade time, where time increments are measured by a fixed number of trades:

$$r_m(n) = p(n) - p(n - m), \quad (2)$$

where  $n$  represents the  $n$ -th trade and  $m$  represents the number of trades in a unit of time. Note that trade time is distinct from volume time: the former defines time by number of transactions (regardless of the size of the transactions) while the latter defines time by number of contracts traded. Although the two are closely related (volume is much less variable in trade time than clock time) they are distinct. As mentioned in the introduction, Clark (1973) subordinates returns with volume as a time increment. However, within the context of the historically well-documented volume/volatility relationship (Karpoff (1987), Gallant et al. (1992), Tauchen et al. (1996), Aldrich (2013)), Jones et al. (1994) shows that there is little additional information content in volume beyond the number of transactions. In each case, the objective is to use a measure of time that controls for latent information arrival. We choose to follow Jones et al. (1994) and adopt number of transactions as the best surrogate.

Empirical asset returns (measured in clock time) typically exhibit notable features such as leptokurtosis and conditional heteroskedasticity, regardless of time scale. Much effort has been expended over the course of decades to model the heavy tails of returns distributions (Mandelbrot (1963)) as well as the strong autocorrelation of volatility (Engle (1982) and Bollerslev (1986)). In the subsequent analysis we show that the observations of Brada et al. (1966), Mandelbrot and Taylor (1967), Clark (1973) and Ane and Geman (2000), that trade-time returns are nearly Gaussian, extend to high-frequency, intra-day returns during our 2014 sample period, when removing price jumps by filtering pre-scheduled news announcements.

---

<sup>1</sup>The University of Michigan Consumer Sentiment report is released twice monthly at 9:55 a.m., Eastern Time. Since some of these announcements also coincided with other 10:00 a.m. announcements during our sample period, a 1000-second time window would not have provided an appropriate filter relative to events on other days. For uniformity, we simply exclude the Consumer Sentiment days.

### 3.1 Intra-day Clock-Time and Trade-Time Distributions

As reported in Section 2, the active, event-driven data span 223,200 seconds (62.0 hours) with 1,252,020 transaction records, while the passive, non-event data span 5,081,400 seconds (1411.5 hours) with 15,556,531 records. This corresponds to an average of 5.61 transactions per second in the active sample and 3.06 transactions per second in the passive sample. To strike a balance, we assume in the remaining analysis that 1 transaction corresponds to 0.25 seconds, or 250 milliseconds.

Figures 1 and 2 display distributional features of clock-time returns and trade-time returns (respectively) for the full data sample, without separating news announcement periods. The clock-time intervals in Figure 1 are  $\delta = \{1000, 10000, 60000, 600000\}$  milliseconds (i.e. 1 second, 10 seconds, 1 minute and 10 minutes) and the trade-time intervals in Figure 2 are  $m = \{4, 40, 240, 2400\}$  trades. According to our approximation, these sets of intervals are roughly analogous. The first row of plots in each figure depict the empirical density functions of returns, superimposed upon Gaussian distributions that are estimated by maximum likelihood. These are shown with the vertical axis on a log scale in order to highlight discrepancies in the tails of the distributions. The second rows show corresponding quantile-quantile (Q-Q) plots of the distributions in the upper panels. Departures from Gaussianity are highlighted via nonlinearities in the Q-Q plots. The third and fourth rows depict sample autocorrelation functions (ACFs) of returns and squared returns for each time interval. Unlike the empirical densities and Q-Q plots in the upper rows of the figures, it is not correct to assume that returns are contiguous over the entire 6-month sample period when computing autocorrelations – the overnight gaps between market close and open introduce discontinuities in the data. For this reason, we compute ACFs on a day-by-day basis, and separately plot them in each panel on the lower rows of Figures 1 and 2. The 0.05, 0.5 and 0.95 quantiles are accentuated by bold-colored lines, and, for comparison, the full sample autocorrelations (treating the sample as contiguous) are shown as solid black lines. The black dotted lines depict 95% Bartlett bounds for the null hypothesis of no serial correlation.

The upper rows of Figure 1 demonstrate that clock-time returns are markedly non-Gaussian, suggesting that the underlying data generating process has higher probabilities of extreme events. This is true, even for the comparatively long intraday time interval of 10 minutes, considered by [Ane and Geman \(2000\)](#). In contrast, the upper rows of Figure 2 show that leptokurtosis, which is present for short trade time intervals, diminishes as the horizon expands, confirming the conclusions of [Ane and Geman \(2000\)](#) that were noted in the previous section. In all cases, departures from Gaussianity exhibit a slight left skewness, which is an empirical asymmetry often observed in financial returns. Both figures show that the autocorrelations of returns are not statistically different from zero at any lag, aside from the first, which corresponds to the reversion associated with bid/offer bounce. The pattern of autocorrelations across squared returns (typically used to diagnose persistence of volatility), however, is important. For  $\delta = 1$  second in the clock-time sample, the ACFs diminish very slowly, with the lowest quantiles hovering above the upper Bartlett bounds even after 20 lags. In contrast, for  $m = 4$  trades in the trade-time data, the median autocorrelation is encompassed by the Bartlett bounds by the fourth lag and the bundle of daily ACFs is only slightly shifted from the Bartlett bounds for all lags. The latter shift in the distribution of daily ACFs is indicative of volatility persistence, but much weaker than that of clock-time returns. Similar, but slightly weaker observations hold for  $\delta = 10$  seconds and  $m = 40$  trades: clock-time autocorrelations diminish slowly, with the median of daily ACFs only entering the Bartlett bounds at lag 20 (slightly more than 3 minutes) whereas the median of trade time autocorrelations is within the Bartlett bounds at all lags, and the full bundle of daily ACFs is only slightly offset (positively) from the Bartlett bounds at all lags. For the remaining time intervals, the relationship further weakens as there is less data to estimate the daily ACFs. It is important to note that the initial decay of autocorrelations

for squared returns in both samples is an artifact of bid/offer bounce, as reverisive behavior at the finest time scales results in positive correlation of returns magnitudes in adjacent periods. Since we are less interested in the dynamics of bid/offer reversion, we focus our attention on the longer horizon persistence of squared returns in the clock-time and (more weakly) trade-time data.

For the remainder of this section, we consider time aggregated returns for  $\delta = \{250, 500, 1000, 4000, 10000, 30000\}$  milliseconds (i.e. a maximum time scale of 30 seconds) and  $m = \{1, 2, 4, 40, 400, 4000\}$  trades. Figure 3 shows Q-Q plots of clock-time and trade-time returns during active and passive periods for these time intervals in the full-day dataset. To better observe departures from Gaussianity, the scale of the y-axes is half of those in Figures 1 and 2. The panels in the first and second rows of the figure compare the empirical quantiles of clock-time returns with the theoretical quantiles of the best-fit Gaussian distributions for active and passive periods, respectively. The panels in the third and fourth rows are the same for trade-time returns.

Not surprisingly, the upper rows of Figure 3 show that clock-time returns exhibit patterns of leptokurtosis, especially at fine time scales. The characteristic heavy tails of the empirical distributions are visible in the Q-Q plots as a convex-concave departure from linearity. The third row shows that trade-time returns in the non-jump-filtered, active subsample also exhibit leptokurtosis, although to a lesser extent than clock-time returns, and that the leptokurtosis diminishes with trade-time interval,  $m$ . We note, however, that the attenuation of tail heaviness in the active trade-time sample relative to the passive clock-time sample is largely due to sample size, as will be elaborated below. The final row of the figure highlights a distinct result: during passive, non-event time periods, trade-time returns conform quite closely to a Gaussian distribution for all time intervals. This is somewhat less true at the finest time scales, where a small fraction of the sample could be characterized as non-Gaussian jumps, however the evidence suggests that the bulk of the data is well characterized as Gaussian. This empirical result, which we develop in the remainder of the paper, is largely related to (1) the fact that pre-filtering news announcements results in a returns process that is well approximated by a continuous local martingale and (2) that trade-time (inter-trade durations) is a suitable time deformation device for rescaling the continuous local martingale so that it has Gaussian increments. As mentioned in the introduction and further explained in Section 5, inter-trade duration and stochastic volatility are similar devices for scaling continuous local martingales.

To address issues of sample size, which plays an important role in distributional behavior, Figure 4 depicts identical Q-Q plots for the reduced, 10:00 a.m. subsamples of active and passive returns. Recall that these subsamples result in far more balanced datasets. The broad features of tail heaviness are preserved in the first three rows of the figure, although the passive, clock-time returns do not exhibit the same propensity for extreme events as in the full-day sample. The result is that each of the passive subsamples (clock time and trade time) are not as leptokurtotic as their active-sample counterparts. This is exactly as we would anticipate, since periods following news announcements should have a higher frequency of large price movements and a more variable rate of information flow. In addition, the passive, trade-time returns in the final row of Figure 4, conform even more closely to the Gaussian distribution than their full-day counterparts: the small fraction of non-Gaussian jumps are no longer present. These latter two observations will be characterized in the model of Section 4, where non-Gaussianity will be generated through two channels: discrete jumps, and a stochastic time process. The observation that clock-time returns exhibit greater non-Gaussianity than trade-time returns is primarily related to variation in information flow as business time evolves stochastically with the wall clock. The additional observation that active-period returns have heavier tails than passive-period returns is related to greater heterogeneity in the stochastic time process (clock-time) as well as the arrival of discrete jumps (both clock-time and trade-time). Homogeneity of the time process and absence of large jumps in the passive-period

returns causes them to exhibit fewer departures from Gaussianity.

Gaussianity of high-frequency, trade-time returns is not immediately apparent when studying intra-day data because trading during news event periods has a large impact on the unconditional distribution of returns. That is, extreme price movements during periods of pre-scheduled news announcements cause substantial increases in tail probabilities of the unconditional distributions. Filtering those news announcements during our 2014 sample period, however, is an effective jump-removal device. This is most easily seen by comparing the Q-Q plots in Figure 2 with those in the lower rows of Figures 3 and 4. Recall that the y-axis scale in the latter figures is half that of the former. In contrast, as the bottom panels of Figure 3 and 4 indicate, if one only considers returns during non-event times, a Gaussian distribution provides an excellent fit. This is true for our very limited 10:00 a.m. subsample of passive period returns as well as for the full-day sample. For the remainder of this section, we will focus attention on the 10:00 a.m. subsample in order to ensure that sample sizes across active and passive periods are roughly equivalent.

Similar to the ACFs in Figures 1 and 2, Figures 5 and 6 depict day-by-day sample ACFs of returns and squared returns (respectively) for the clock-time and trade-time intervals considered in the Q-Q plots of Figure 4. As with the unconditional sample, Figure 5 shows that returns in each subsample exhibit no serial correlation at all time scales, except at the first lag, which demonstrates significant evidence of reversion. The strength of the reversion is greatest at the finest time scales, but is still present even for  $\delta = 30000$  ms and  $m = 120$ , where the ACF bundles for each dataset are slightly shifted towards negative values and eclipse the negative Bartlett bounds.

The panels in Figure 6 correspond to those of Figure 5, but depict ACFs of squared returns. The panels in the first two rows, which correspond to clock-time returns, are exemplary of the persistent nature of volatility: a high degree of autocorrelation at low lags and a subsequent low-level of persistence over a long horizon, which is exhibited by the vertical displacement of autocorrelation values for  $\delta \leq 1000$  ms. A similar, but weaker vertical asymmetry is also apparent for the active-period trade-time returns, where the median of daily ACFs hovers just below the upper Bartlett bound for all lags. The remaining panels show that passive-period trade-time returns exhibit volatility persistence at low lags, but that the persistence quickly disappears with no apparent long-horizon autocorrelation (i.e. no vertical displacement of the distributions of autocorrelations). While initial inspection of the first column of plots in Figure 6 would suggest a higher degree of autocorrelation in squared trade-time returns at low lags, this is attributed to bid/offer bounce (as previously mentioned). The weaker lag 1 autocorrelation in clock-time returns for  $\delta = 250$  ms is a result of smoothing over adjacent (reversive) transactions with a narrow clock-time window.

## 4 Model

Following [Mandelbrot and Taylor \(1967\)](#), we model the natural logarithm of asset prices,  $p(t)$ , as time-changed Brownian motion:

$$p(t) = \sigma B(\tau(t)), \quad (3)$$

where  $\{B(\tau)\}$  is Brownian motion and  $\{\tau(t)\}$  is an increasing process known as the *directing process*. It is common to require  $B(\tau)$  to be standard Brownian motion (as in [Mandelbrot and Taylor \(1967\)](#)), however in our empirical work we do not enforce the restriction of zero mean return despite the fact that average returns are not statistically different from zero over the horizons we consider. In the sequel, we will generally allow the Brownian process to follow

$$dB(\tau) = \mu\tau + dZ(\tau), \quad (4)$$

where  $dZ(\tau)$  represents the increments to standard Brownian motion. We restrict attention to pure-jump directing processes, which is a characteristic inherited by the price process. Intuitively,  $\{\tau(t)\}$  is a stochastic time change which represents business or transaction time. If  $\{\tau(t)\}$  is restricted to be a Lévy process, then its increments are, by definition, independent and it is referred to as a *subordinator*. Notable examples of such Lévy directing processes are the Gamma process, which results in the Variance-Gamma model of Madan and Seneta (1990) and Madan et al. (1998), and the Exponential process, which is related to the compound Poisson process. Our model, on the other hand, does not enforce the Lévy restriction of independent increments, but allows for more general directing processes. In fact, the dynamics of returns will depend importantly on the dynamics of the directing process.

Our empirical analysis in Section 3 suggests that the process for clock-time prices, Equation (3), should be augmented as

$$p(t) = \sigma B(\tau(t)) + J(t), \quad (5)$$

where  $J(t)$  is a purely discontinuous jump process. Equation (5) introduces two mechanisms that result in departures from Gaussianity for prices: (1) a stochastic time change (described in more detail below) and (2) discrete jumps. The first only introduces non-Gaussianity to clock-time returns, while the second does so for trade-time, and hence, clock-time returns. Since pre-filtering for news announcements effectively removes the jump mechanism in our 2014 sample period, we focus subsequent attention on Equation (3). This allows us to highlight the importance of stochastic time in the distributions of high-frequency asset returns.

For a given increasing, pure-jump directing process,  $\{\tau(t)\}$ , we define two auxiliary processes:

$$t_i = \min_t \tau(t) = i, \quad i = 1, 2, \dots \quad (6)$$

$$N(t) = \sum_{i=1}^{\infty} \mathbb{I}_{\{t_i \leq t\}}, \quad t > 0. \quad (7)$$

The point process  $\{t_i\}$  represents the jump or transaction times and  $\{N(t)\}$  is the counting process induced by the directing process. The point process increments,

$$d_i = t_i - t_{i-1}, \quad (8)$$

are the durations between transactions. The assumption of a Lévy directing process implies independence of the point process and inter-trade durations. On the other hand, non-independent increments to the directing process allows for time dependence in the processes defined in Equations (6) and (7) as well as the inter-trade durations.

The increments to the subordinated Brownian motion in Equation (3) can be expressed as

$$r_\delta(t) = \sum_{i=1}^{N_\delta(t)} r_i \quad (9)$$

where

$$r_\delta(t) = p(t) - p(t - \delta) \quad (10)$$

$$N_\delta(t) = N(t) - N(t - \delta) \quad (11)$$

$$r_i \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma), \forall i. \quad (12)$$

The interpretation is that increments to the subordinated Brownian motion are the summation of a random number of non-subordinated Brownian increments. The probability density function of  $r_\delta(t)$  is,

$$f(r_\delta(t)|\mu, \sigma) = \sum_{k=1}^{\infty} f\left(\sum_{i=1}^k r_i \middle| N_\delta(t) = k, \mu, \sigma\right) f(N_\delta(t) = k). \quad (13)$$

Equation (13) is a finite Gaussian mixture model with mixture weights that vary according to the probability distribution of  $N_\delta(t)$ . The model can also be characterized as a two-stage hierarchical model in which a number of trades is drawn from the distribution of  $N_\delta(t)$  in the first stage and a single  $\delta$ -period return is drawn from the Gaussian distribution of  $r_\delta(t) = \sum_{i=1}^{N_\delta(t)} r_i \sim \mathcal{N}(N_\delta(t)\mu, \sqrt{N_\delta(t)}\sigma)$  in the second stage.

Equation (13) is a form of the mixture of distributions hypothesis (MDH, Epps and Epps (1976)) for asset returns and has been utilized in various forms in the finance literature since the work of Press (1967). While certain cases of MDH models mix directly over time varying volatility, and hence allow for temporal dependence in the directing process, much of the focus of stochastic time change MDH models has been on Lévy directing processes, with the objective to match empirical moments of unconditional returns distributions. It is, in fact, the temporal dependence of the directing process which governs the dynamics of return volatility, a result that we develop at the end of this section. First, we highlight several specializations of the model in Equation (9), which we compare in Section 5. It is important to note that these specializations can either be characterized by their directing processes, associated point processes, counting processes or inter-trade durations. With the exception of the compound Poisson process, we will typically find it convenient to focus on their characterization in terms of inter-trade durations.

## 4.1 Compound Poisson Process

A starting point for modeling trade arrivals is to assume they follow a Poisson process:

$$N_\delta(t) \sim \text{Poisson}(\gamma\delta)$$

where  $\gamma$  is the trade arrival intensity parameter. In this case Equation (13) becomes

$$f(r_\delta(t)|\mu, \sigma) = \sum_{k=1}^{\infty} \frac{1}{\sigma\sqrt{2\pi k}} \exp\left\{-\frac{1}{2}\frac{(\sum_{i=1}^k r_i - k\mu)^2}{k\sigma^2}\right\} \times \exp\{-\gamma\delta\} \frac{(\gamma\delta)^k}{k!}. \quad (14)$$

This is the compound Poisson process developed by Press (1967) and Press (1968). Although, its density function cannot be obtained in closed form, it can be approximated by Monte Carlo simulation: first making independent draws from the Gaussian distribution and then accumulating random numbers of those Gaussians according to integer deviates drawn from the Poisson density. When  $\delta$  is large relative to  $\gamma$ , the  $\mathcal{N}(\gamma\delta, \sqrt{\gamma\delta})$  density serves as good approximation to the  $\text{Poisson}(\gamma\delta)$  density, which results in an approximate Gaussian density for  $r_\delta(t)$ . However, for small values of  $\delta$  (very short calendar time intervals) the compound Poisson process does exhibit leptokurtosis.

In this single case, we will often focus on the counting process as the representation of time deformation, although the assumption of Poisson trade arrivals implies that inter-trade durations are distributed as an Exponential random variable, with rate  $\gamma$ . This is also the only time deformation model that we will consider which has Lévy increments. In the remainder of this section, we will focus on models that represent time deformation with dependent, inter-trade durations.

## 4.2 Compound ACD Process

Engle and Russell (1998) introduce a dynamic model for inter-trade durations known as the autoregressive conditional duration (ACD) process. Widely considered the benchmark for dynamic models of inter-trade durations, the ACD process exploits a GARCH-style structure for expected waiting times:

$$d_i = \psi_i \varepsilon_i \quad (15)$$

$$\psi_i = \omega + \sum_{j=1}^m \alpha_j d_{i-j} + \sum_{j=1}^q \beta_j \psi_{i-j}, \quad (16)$$

where  $d_i$  represents inter-trade durations,  $\psi_i = \mathbb{E}[d_i]$  and  $\varepsilon_i$  represents an innovation at time  $i$ . Engle and Russell (1998) refer to Equations (15) and (16) as the ACD(m,q) model and separately consider both Exponential and Weibull innovations. As with the GARCH model, Engle and Russell (1998) demonstrate that a simple Exponential-ACD(1,1) model is very effective at capturing the dynamics of inter-trade durations.

While ACD models are typically combined with ARCH/GARCH models to generate high-frequency asset price dynamics (Engle (2000)), we propose a compound-duration model of the form in Equation (9), utilizing counting processes that are induced by ACD durations. We subsequently refer to this model as the compound ACD processes.

## 4.3 Compound Multifractal Process

Chen et al. (2013) and Zikes et al. (2014) develop the Markov-switching multifractal duration (MSMD) model, which builds on the multifractal volatility model of Calvet and Fisher (2001), Calvet and Fisher (2002) and Calvet (2004). While Zikes et al. (2014) develop a general characterization of the MSMD model, we focus on the specific case considered by Chen et al. (2013). Whereas the focus of this prior work was to explain the dynamics of inter-trade durations themselves, our innovation is to compose the MSMD model with a Brownian diffusion as a characterization of returns. The core components of the MSMD model are a set of  $\bar{k}$  latent state variables,  $M_{k,i}$ , that obey a two-state Markov-switching process with varying degrees of persistence,  $\gamma_k$ , for  $k = 1, 2, \dots, \bar{k}$ . That is, the distribution of trade durations,  $d_i$ , is governed by the equations,

$$d_i = \frac{\varepsilon_i}{\lambda_i} \quad (17a)$$

$$\varepsilon_i \sim Exp(1) \quad (17b)$$

$$\lambda_i = \lambda \prod_{k=1}^{\bar{k}} M_{k,i} \quad (17c)$$

$$M_{k,i} = \begin{cases} M & \text{with probability } \gamma_k \\ M_{k,i-1} & \text{otherwise} \end{cases} \quad (17d)$$

$$\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}} \quad (17e)$$

$$M = \begin{cases} m_0 & \text{with probability } 1/2 \\ 2 - m_0 & \text{otherwise.} \end{cases} \quad (17f)$$

Hence, the MSMD can be succinctly characterized by five parameters:  $\bar{k} \in \mathbb{N}$ ,  $\lambda > 0$ ,  $\gamma_{\bar{k}} \in (0, 1)$ ,  $b \in (1, \infty)$  and  $m_0 \in (0, 2]$ . The intuition is that conditional on knowing the values of the latent

state variables, inter-trade durations are Exponentially distributed with intensity parameter  $\lambda_i$ . However, as time evolves, the latent states,  $M_{k,i}$ , switch values with varying degrees of persistence,  $\gamma_k$ . This causes the unconditional distribution of trade durations to be a mixture of Exponentials, which is consistent with the over-dispersion property of the observed data, described in [Chen et al. \(2013\)](#). The latent states can be interpreted as shocks that have varying impacts over diverse timescales, some having short-horizon and others having long-horizon effects. Figure 7 depicts the times series of latent states from a simulation of a single trading day of data for the model we estimate in Section 5.1.2 and demonstrates heterogeneity of persistence across the states. The value  $b$  governs a tight relationship between the persistence parameters,  $\gamma_k$ , and is responsible for the parsimony of the model: even with a large number of latent states,  $\bar{k}$ , the model is always characterized by a total of five parameters. The choice of  $b$  dictates the degree of heterogeneity in values of persistence parameters. For more insight regarding the MSMD model, see [Chen et al. \(2013\)](#) and [Calvet and Fisher \(2008\)](#).

The distribution of  $r_\delta(t)$  retains its hierarchical structure under the MSMD model, with the number of trades per unit of time being drawn from a mixture of Poisson distributions in the first stage. We refer to  $r_\delta(t)$  as a *compound multifractal process* when the Gaussian mixture weights correspond to count probabilities associated with MSMD durations.

The variability of the MSMD stochastic intensity parameter,  $\lambda_i$ , causes MSMD durations to exhibit far greater heterogeneity than those of the constant-intensity Exponential. [Chen et al. \(2013\)](#) liken stochastic intensity in duration models to stochastic volatility in returns models: “Just as stochastic volatility ‘fattens’ Gaussian conditional returns distributions, so too does MSMD ‘over-disperse’ exponential conditional duration distributions.” (p. 9). In fact, we find that stochastic intensity plays the dual role of duration over-dispersion and returns tail fattening: the dispersion of probability over a greater variety of counts, relative to the simple Poisson model, induces greater heterogeneity in the Gaussian mixture, which generates fatter tails for  $r_\delta(t)$ . Intuitively, the random variable  $r_\delta(t)$  switches between a greater variety of differing sums of Gaussians with higher probability. In addition, the compound multifractal model explicitly generates volatility persistence by producing autocorrelation in the inter-trade duration distribution, via Markov-switching latent states. Similar mechanisms in the ACD process result in leptokurtosis and volatility persistence for clock-time returns.

#### 4.4 Volatility Persistence

The compound returns model of Equations (9) – (12) is able to simultaneously generate fat tails and volatility persistence. The latter feature, however, is dependent on the dynamics of the counting process, which we now show.

**THEOREM 1** Given a counting process  $N(t)$  and a Gaussian mixture process  $r_\delta(t)$  as specified in Equations (9) – (12), then

$$\gamma_{r_\delta^2,k} = \sigma^4 \gamma_{N_\delta,k}, \quad k = 0, 1, 2, \dots, \quad (18)$$

where  $\gamma_{X_\delta,k} = \text{Cov}(X_\delta(t), X_\delta(t - \delta k))$  is the  $k$ th  $\delta$ -period autocorrelation for  $X \in \{r^2, N\}$ .

## Proof

$$\gamma_{r_\delta^2, k} = \text{Cov}(r_\delta^2(t), r_\delta^2(t - k\delta)) \quad (19)$$

$$= \text{Cov}\left(\left[\sum_{i=1}^{N_\delta(t)} r_i\right]^2, \left[\sum_{j=1}^{N_\delta(t-k\delta)} r_j\right]^2\right) \quad (20)$$

$$= \text{Cov}\left(\underbrace{\sum_{i=1}^{N_\delta(t)} r_i^2 + 2 \sum_{i=1}^{N_\delta(t)-1} \sum_{n=i+1}^{N_\delta(t)} r_i r_n}_{A}, \underbrace{\sum_{j=1}^{N_\delta(t-k\delta)} r_j^2 + 2 \sum_{j=1}^{N_\delta(t-k\delta)-1} \sum_{m=j+1}^{N_\delta(t-k\delta)} r_j r_m}_{B}\right) \quad (21)$$

$$= \underbrace{\mathbb{E}[\text{Cov}(A, B | N_\delta(t), N_\delta(t - k\delta))]}_{=0} + \text{Cov}(E[A | N_\delta(t)], E[B | N_\delta(t - k\delta)]) \quad (22)$$

$$= \text{Cov}(N_\delta(t)\sigma^2, N_\delta(t - k\delta)\sigma^2) \quad (23)$$

$$= \sigma^4 \text{Cov}(N_\delta(t), N_\delta(t - k\delta)). \quad (24)$$

$$= \sigma^4 \gamma_{N_\delta, k}. \quad (25)$$

□

Theorem 1 shows that the persistence of squared returns is directly related to the persistence of the counting process. An immediate implication follows.

**LEMMA 1** Given a Lévy counting process,  $N(t)$ , i.e. a renewal process, a Gaussian mixture process  $r_\delta(t)$  as specified in Equations (9) – (12) does not exhibit serial correlation among its squared values.

It follows from Lemma 1 that the compound Poisson process will not exhibit volatility persistence. Further, any subordinated Brownian motion with a Lévy directing process, and its corresponding Gaussian mixture model, will not generate volatility dynamics. On the other hand, the ability of a non-Lévy directing processes to generate volatility dynamics will depend on the dynamics of the counting process. This suggests that the ability of the compound ACD and MSMD models to match volatility dynamics, will depend on the ability of their associated counting processes to match the dynamics of observed transaction counts. The ACD and MSMD models accomplish this by directly modeling serial dependence among inter-trade durations, which propagates to the associated counting processes. While the result stated above is an exact, finite-sample result, Deo et al. (2009) establish conditions under which long memory in durations asymptotically propagates to the associated counting process and clock-time returns within a pure-jump model. However, it is important to note that in our model, serial correlation of durations is sufficient but not necessary to generate the appropriate serial correlation in volatility via the counting process; it is possible for trade-count persistence to exist in the presence of independent durations.

## 5 Estimation and Results

Using the 10:00 a.m. subsample of jump-filtered, non-event E-mini returns data described in Sections 2 and 3, we estimate the parameters of the component distributions in Equation (13) and

simulate from the mixture model. In particular, we first estimate the duration models described in Section 4, which we then compound with an estimated Gaussian distribution for trade-time returns to synthesize a distribution for clock-time returns. Using Monte Carlo approximations for the distribution of clock-time returns, we evaluate the candidate models using several measures of goodness-of-fit and distributional distance.

## 5.1 Estimation

Since the observed inter-trade durations are not a contiguous sequence of data, we estimate each of the models on a daily basis, using the 1000 seconds of post 10:00 am data on the days without news announcements at that time. The result is 121 sets of estimates, whose means and standard deviations are reported in Table 1. As a matter of comparison, we also report the single point estimates for each model in Table 1, using the full sample of data as a single contiguous sequence of observations. Where units can be appropriately applied to the estimates, they correspond to millisecond time units and S&P 500 index point price increments. The following subsections provide details on estimation for each of the model components.

### 5.1.1 Poisson/Exponential

The assumption of Poisson-distributed trade arrivals with mean  $\gamma$  corresponds to duration times that are Exponentially distributed with mean  $\nu = \frac{1}{\gamma}$ . It is trivial to show that the maximum likelihood estimate of the Exponential mean is

$$\hat{\nu} = \frac{1}{\hat{\gamma}} = \frac{1}{n} \sum_{i=1}^n d_i, \quad (26)$$

where  $d_i, i = 1, 2, \dots, n$  are the observed inter-trade durations. The mean and standard deviation of daily estimates of  $\gamma$  are reported in Table 1. The interpretation of the estimate is that approximately  $\hat{\gamma} = 0.00535$  trades arrive each millisecond, or a single trade (buy or sell) arrives approximately every  $\hat{\nu} = 1/0.00535 = 187$  milliseconds.

### 5.1.2 MSMD

Following Chen et al. (2013) and Zikes et al. (2014), we evaluate the likelihood of the MSMD model, associated with Equations (17a) – (17f), using the nonlinear filtering method of Hamilton (1989) and maximize the likelihood with a standard hill-climbing algorithm. To estimate all parameters of the MSMD model, we iterate over candidate values of  $\bar{k}$  and estimate the remaining four parameters,  $\lambda$ ,  $\gamma_{\bar{k}}$ ,  $b$ , and  $m_0$ . Although we do not report the full set of results here, we find that the log likelihood (both the mean of daily log likelihood values and the single value for the full sample) is maximized for  $\bar{k} = 8$ , with a plateau that begins at  $\bar{k} = 7$ . For purposes of concision, we adopt  $\bar{k} = 7$  for the remainder of the paper.

### 5.1.3 ACD

As outlined in Engle and Russell (1998) the likelihood function for the ACD model can be obtained in closed form. We obtain daily and full-sample estimates of the model via maximum likelihood.

### 5.1.4 Gaussian

Building on the empirical observations of Section 3 and the assumptions of the compound duration model of Equation (9), we treat non-news trade-time returns as i.i.d. Gaussian random variables and estimate their moments via maximum likelihood. In this case, the maximum likelihood estimates of the parameters of the distribution are simply the sample average and standard deviation of trade-time returns. Daily and full sample estimates are reported in Table 1.

## 5.2 Simulation and Evaluation

We obtain Monte Carlo approximations of clock-time returns distributions using the Gaussian mixture model, expressed in Equation (13). We do this in a hierarchical fashion, first simulating inter-trade durations from the Exponential, MSMD and ACD models, pairing the durations with independent draws of trade-time returns from the estimated Gaussian density, and finally aggregating individual returns within a fixed clock time interval. For each model, we generate a dataset equivalent to the observed data: 121 days, each comprising 1,000,000 milliseconds worth of data.

Following the procedure outlined above, we aggregate returns for clock-time intervals  $\delta = \{250, 500, 1000, 5000, 10000, 30000\}$  milliseconds until we obtain  $n = \{484000, 242000, 121000, 24200, 12100, 4033\}$  clock-time returns, respectively, which are almost identical to the number of observations in the data for those time intervals<sup>2</sup>. The individual simulations under each of the duration models use the same trade-time returns; they only differ in the elapsed time between observations. Since empirical inter-trade durations are discrete and observed at integral millisecond values, we discretize the simulated inter-trade durations (which are continuous valued) by rounding to the nearest millisecond. As zero durations are not allowed in our framework, all simulated durations below one millisecond are rounded upward.

As a matter of comparison, we also estimate and simulate clock-time returns under the ACD-GARCH(1,1), suggested by Engle (2000), using ACD durations in the GARCH volatility equation. In particular, following Engle (2000),

$$\frac{r_i}{d_i} = \sigma_i \varepsilon_i \quad (27)$$

$$\sigma_i^2 = \omega + \alpha \varepsilon_{i-1}^2 + \beta \sigma_{i-1}^2 + \gamma d_i^{-1} \quad (28)$$

$$\varepsilon_i \stackrel{i.i.d.}{\sim} WN(0, 1). \quad (29)$$

Since the compound Poisson process has no ability to generate serial correlation in clock-time return volatility, the ACD-GARCH model is a more reasonable benchmark against which to compare the compound duration model.

Table 2 reports summary statistics for observed inter-trade durations, as well as the daily simulations under each of the duration models. The table reports the averages of daily estimates of the first four moments of durations, in addition to the single point estimates obtained when treating each data set as a single contiguous sample. The table shows that the Exponential and ACD durations do a much better job at approximating the mean and standard deviation of daily durations in the data than the MSMD model, which has values that are an order of magnitude larger than all of the others. This is also true of the first two moments when treating the data sets as a single sample. It is readily apparent, however, that the dispersion of daily mean and standard deviations for the Exponential and ACD estimates are extremely low – a little more than

---

<sup>2</sup>600 seconds of data, corresponding to 10:00:00 – 10:01:40 a.m. on 3/24, 3/31, 4/30, 7/31, 10/23, and 12/16/2014, were excluded from the sample due to 1000-second news intervals starting at 9:45:00 a.m. on those days.

two orders of magnitude lower than those of the data. The MSMD durations, despite being too high, come closer to capturing the right relative dispersion of daily estimates. For daily skewness, the Exponential and ACD durations are roughly half as skewed as the data, while the MSMD durations exhibit roughly twice the skewness. Likewise, the dispersion of Exponential and ACD daily skewness is too low and the that of the MSMD is too high. Similar to skewness, daily kurtosis values of the ACD model are about three times lower than those of the data, and the daily kurtosis values of the MSMD model are almost five times larger. In this latter case, the dispersion of the daily estimates, relative to the average of kurtosis values, is much more in line with the data for the ACD model than for the MSMD model. The Exponential model kurtosis values are even lower than those of the ACD model. Interestingly, when treating the daily data sets as single contiguous samples, the skewness and kurtosis of the E-mini inter-trade durations are much higher and the MSMD model does a much better job (an order of magnitude in each case) of matching their empirical counterparts. One final note regards over-dispersion, or the ratio of mean and standard deviation of durations: over-dispersion of E-mini inter-trade durations in the data is between 0.35 and 0.55 (depending on whether one uses daily averages or single point estimates). Relative to the data, ACD durations are very under-dispersed, with a ratio of mean to standard of roughly 0.9, while MSMD durations are even more over-dispersed, with values between 0.17 and 0.23 (again, daily vs. full sample values).

Figure 8 depicts daily sample autocorrelation functions for the observed and simulated inter-trade durations. The faint lines represent ACFs for each day and the darker lines represent 0.05, 0.5 and 0.95 quantiles of the distributions of autocorrelations at each lag. By construction, the Exponential durations are independent of each other, and so do not exhibit any serial correlation. The ACD and MSMD models, on the other hand, do a much better job of capturing observed dynamics: the median ACD autocorrelation remains elevated until lag 7 or 8, while that of the MSMD does not fall to zero until roughly lag 15. The empirical counterpart falls rapidly until lag 10-15, at which point it remains elevated at a low level and with very slow decay. Further, the quantiles of the empirical autocorrelations are heavily skewed toward positive values for the full range of lags that are depicted in the plot. This low-level persistence of observed durations is not captured by any of the models, although the MSMD performs slightly better than the ACD in this dimension: its quantiles are skewed towards positive autocorrelation until about lag 40. It is important to note that while the MSMD model appears to have some directional advantage in matching the autocorrelation structure of observed durations, it is certainly much noisier than the ACD model.

Figure 9 shows Q-Q plots of the clock-time returns simulations for each value of  $\delta$  that we consider. The panels in the first (red) row correspond to the compound Poisson (Exponential) model, the second (yellow) row corresponds to the compound ACD model, the third (green) row corresponds to the compound MSMD model and the fourth (purple) row corresponds to the ACD-GARCH(1,1) model. The panels in the final (blue) row of the figure are a reproduction of the E-mini passive-period clock-time Q-Q plots shown in Figure 4. We note that while the limits of the axes are constant within row, they differ across rows: the vertical axis for the ACD-GARCH model is two orders of magnitude larger than the others, in order to capture the full range of simulated quantiles. It is immediately apparent from the figure that the compound Poisson model does a very poor job of capturing leptokurtosis, except at the lowest values of  $\delta$ , where the tails of the distributions of the simulated returns are still lighter than those of the data. The compound ACD model only provides a slight improvement. On the other hand, the compound MSMD and ACD-GARCH models exhibit much heavier tails for all time scales – tails that are heavier than those of the data. Interestingly, as  $\delta$  increases, the central quantiles of the ACD-GARCH conform more closely to a Gaussian, with only a very small proportion of extreme quantiles. These extreme tails are most

likely the additive result of persistence in both the volatility and duration equations: the simple GARCH(1,1) is known to exhibit leptokurtosis, which now appears to be accentuated in the presence of duration autocorrelation. The compound MSMD, however, exhibits more appropriate (although slightly extreme) tail heaviness, although the leptokurtosis comes at a cost: a preponderance of zero returns. The zero returns are caused by an excess of long inter-trade durations, which are long enough to span the respective clock-time intervals. The preponderance of long inter-trade durations is generated by the same mechanism which generates tail heaviness: serial correlation. Intuitively, strong serial correlation in durations causes both a preponderance of short waiting times (which results in a higher frequency of tail events) as well as long waiting times (which results in zero returns).

Panels (a) and (b) of Table 3 report summary statistics that formalize the degree of leptokurtosis in the simulated returns depicted in Figure 9. Under the null hypothesis of Gaussian clock-time returns, the variance of the sample excess kurtosis,  $\hat{\kappa}$ , is:

$$\sigma_{\hat{\kappa}}^2 = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}, \quad (30)$$

where  $n$  is the number of clock-time returns during a single day (i.e.  $n = 4000$  for  $\delta = 250$  ms). An implication of this result is that  $\sqrt{n}\hat{\kappa} \xrightarrow{d} \mathcal{N}(0, 24)$  under the null hypothesis of Gaussianity. Panel (a) of Table 3 reports the fraction of times  $\hat{\kappa} \in (-z_{1-\alpha/2}\sigma_{\hat{\kappa}}, z_{1-\alpha/2}\sigma_{\hat{\kappa}})$ , where  $z_{1-\alpha/2}$  is the  $(1-\alpha/2)$  quantile of the standard Normal density. We set  $\alpha = 0.05$ . The values in the table demonstrate that for low values of  $\delta$ , all models and the data never accept the hypothesis of Gaussianity on all days. However, as  $\delta$  increases, the simulated compound Poisson/Exponential returns suddenly switch from near 0% acceptance to very high acceptance rates, while the compound MSMD and ACD-GARCH models increase their acceptance rates by only a small amount. The only model that reasonably tracks the data in this case is the compound ACD, which moderately increases acceptance rates for increasing  $\delta$ .

Panel (b) of Table 3 reports the Kullback-Leibler divergences of the empirical distributions of daily Kurtosis values for each model and each clock-time interval  $\delta$ , relative to the theoretical limiting Gaussian density under the null hypothesis for clock-time returns. In order to compute the distance measures, we used a Gaussian kernel smoother to approximate the empirical density of kurtoses. Once again, the compound Poisson/Exponential performs worst, with distances from the limiting distribution that are much too high or too low relative to those observed in the data, except for  $\delta = 30000$ . In a less extreme manner, the same is true of the compound ACD model. This latter result is interesting in light of the fact that the compound ACD has the best acceptance rates in panel (a). The remaining models do a better job of tracking the data, with the ACD-GARCH model exhibiting distances that are slightly more congruent with the data than those of the MSMD.

Although we do not show them here, the daily sample autocorrelation functions of returns under each of the models exhibit no autocorrelation and look almost identical to those shown in the bottom panels of Figure 5. One exception is the sharp negative autocorrelation at the first lag: while none of the models capture the negative autocorrelation attributed to bid/offer bounce and mean reversion in finely sampled data (low  $\delta$ ), this dynamic is not explicitly modeled in our framework and is not expected to be present.

Figure 10 shows daily sample autocorrelation functions for squared returns simulated under each of the duration models. The panels of Figure 10 correspond to those of Figure 9. Persistence among the autocorrelations is present in all but the compound Poisson/Exponential model. The variability of the compound MSMD, and ACD-GARCH(1,1) models appears to be much larger than that of

the data, and the shape of the autocorrelations for the ACD-GARCH models does not appear to conform as closely to the data for low values of  $\delta$ : in the case of the ACD-GARCH(1,1), the low-lag autocorrelations begin at values that are too extreme and subsequently drop too quickly. The compound ACD model, on the other hand, appears to exhibit both the right variability and shape, relative to what is observed in the data, but for  $\delta \leq 1000$  ms the distribution of autocorrelations is symmetrically displaced around zero by lag 10, while that of the data remains positively displaced for many more lags. This latter feature is captured by the compound MSMD model which, despite excessive variation, has a distribution of autocorrelations with an appropriate displacement for low values of  $\delta$ .

Figure 10 shows that there is an apparent trade-off between matching the long-horizon persistence in autocorrelations of squared returns and the lag-by-lag sampling variability of those autocorrelations across days. To summarize the trade-off, panel (e) of Table 3 reports a mean squared error (MSE) criterion for the autocorrelations in each panel of the figure. To compute the MSE, we integrate (sum across lags) the squared deviations of the median autocorrelations from zero and add them to the integral (sum across lags) of the variance of daily estimates. The resulting values for the compound Exponential and ACD models are almost uniformly lower than those of the data, especially for  $\delta \leq 1000$ . This is primarily attributed to these models' inability to capture long-horizon persistence, as the autocorrelations diminish to zero too quickly. On the other hand, the compound MSMD reports MSE values that are much more elevated than those of the data, especially for  $\delta \leq 1000$ , owing to the excessive variability of autocorrelations across days for each lag. While the ACD-GARCH model also exhibits elevated MSE values for the lowest time scales, they fall faster than those of the compound MSMD and do a better job matching the MSE values of the data for a variety of time scales than the other models.

Panels (c) and (d) of Table 3 report hypothesis acceptance rates and Kullback-Leibler divergences for daily Ljung-Box statistics computed using squared clock-time returns. The Ljung-Box statistic is defined as

$$Q_l = n(n+2) \sum_{i=1}^l \frac{\hat{\rho}_i^2}{n-i},$$

where  $\hat{\rho}_i$  is the sample autocorrelation for lag  $i$ ,  $n$  is the number of observations in the data (for each day), and  $l$  is the number of lags over which the statistic is computed. Under the null hypothesis that all autocorrelations are jointly zero,  $Q_l \sim \chi^2(l)$ . For the summary statistics reported in Table 3, we set  $l = 30$ , but the results are robust to a variety of other choices. Similar to panel (a), panel (c) reports the fraction of times the daily Ljung-Box statistic falls within a one-sided  $(1 - \alpha)$   $\chi^2$  confidence interval:  $\hat{Q}_l \in (0, \chi^2_{1-\alpha}(l))$ , where  $\alpha = 0.05$ . The values in the table show that the compound Poisson/Exponential and compound ACD models almost always accept the null hypothesis, unlike the data which frequently rejects for low values of  $\delta$ . The remaining models display acceptance rates that are similar to the data, with the ACD-GARCH(1,1) performing slightly better for  $\delta \leq 5000$  ms.

Panel (d) of Table 3 reports Kullback-Leibler divergences for the empirical densities of daily Ljung-Box statistics relative to the theoretical  $\chi^2$  distribution under the null hypothesis of no serial correlation in squared clock-time returns. The table shows that the compound Poisson/Exponential and compound ACD models exhibit divergences that are almost uniformly lower than those of the data, although the compound ACD model is reasonably close for  $\delta \geq 1000$  ms. The compound MSMD model, however, displays distances that are uniformly too high relative to the data and the ACD-GARCH model is characterized by large distances for  $\delta \leq 1000$  ms and low distances for  $\delta \geq 5000$  ms.

### 5.3 Discussion

The foregoing results highlight the strengths of compound duration models in the presence of a time deformation process that exhibits serial correlation. While no model performs uniformly better than the rest, the compound MSMD and compound ACD models demonstrate a number of strengths (typically for differing ranges of  $\delta$ ) and achieve a reasonable balance in terms of matching both the tail heaviness and volatility clustering that is observed in the jump-filtered data. This emphasizes the importance of viewing asset prices as subordinated Brownian motion with a non-Lévy directing process. Several observations are in order.

First, the empirical results corroborate the finding of Theorem 1, which shows that the lack of dependence in the compound Poisson trade arrival process renders it unable to deliver serial correlation in volatility. On the other hand, the ability of the remaining models to capture volatility persistence is directly related to the persistence of its inter-trade durations: the autocorrelation structure observed in Figure 8 immediately propagates to Figure 10. Notably, the manner in which the compound MSMD model exhibits a low level of volatility persistence for many lags when  $\delta \leq 1000$  ms is directly related to the low-level persistence of MSMD durations shown in Figure 8.

Second, we see that the compound Poisson model does not exhibit an appropriate degree of leptokurtosis. While a more flexible distributional assumption for the trade arrival process might rectify this deficiency, volatility dynamics will not be present when the directing process of Equation (3) is Lévy. In particular, although the variance gamma models of Madan and Seneta (1990), Madan et al. (1998) and Carr et al. (2002) are much better at capturing extreme events, the nature of their directing processes precludes serial correlation in volatility.

Third, despite empirical advances in measuring the leverage effect in high-frequency data (Aït-Sahalia et al. (2013)) and theoretical advances in explaining the leverage effect via time deformation (Carr and Wu (2004)), we do not attempt to explain the leverage effect in this paper. Our model treats the distributions of trade-time returns and trade arrival as independent, which, as suggested by Renault and Werker (2011), is an oversimplification. This is a simplification we adopt in order to focus on other influences that the trading process has on the distribution of asset returns and their dynamics. Future research could generalize our model to jointly describe trade-time returns and durations and to incorporate the leverage effect. One method would be to require the baseline MSMD intensity parameter,  $\lambda$ , to depend on past or concurrent returns, or other joint factors.

Fourth, a significant portion of the financial econometrics literature works directly with volatility as a time deformation device for continuous local martingales. Madan and Seneta (1990), Andersen et al. (2007), Andersen et al. (2010) and Todorov and Tauchen (2014) are examples. In our work, we arrive at an approximation to a continuous local martingale by pre-filtering news announcement periods and view both time-varying inter-trade duration and time-varying volatility as outgrowths of a latent time deforming information process for the continuous local martingale. This latter point is emphasized both by Theorem 1 and by the fact that inter-trade duration is inversely related to the distribution of trade arrival,  $f(N_\delta(t) = k)$ , in Equation (13), which is directly related to the volatility of clock-time returns,  $\text{Std}(r_\delta(t)) = \sqrt{N_\delta(t)}\sigma$ . We work directly with durations because they are more primitive to the trading process and because our empirical work suggests that this might be the most direct way to model time deformation. Further, while the effects of microstructure noise render volatility estimation infeasible for time scales less than a minute, we encounter no difficulty in estimating the distribution of trade arrivals at all time scales (including less than a second). We thus allow volatility to be a direct result of inter-trade duration, which is our surrogate for a time deformation device.

Following the points above, our model has important economic implications for understanding the dynamics and market prices of volatility. As will be shown in Section 6, we can use simulations

from our estimated models to compute measures of realized volatility in clock-time. In addition, expressing trading rate over time horizon  $\delta$  as the median inter-trade duration,  $d_{\text{med}}(\delta)$ , we can use the estimated Gaussian volatility parameter  $\hat{\sigma}$  to calculate the corresponding implied annual volatility of the S&P 500 index that stems from Brownian motion,

$$\hat{\sigma}_{\text{VIX}} = 100 \times \sqrt{N_\delta M_\delta} \hat{\sigma}, \quad (31)$$

where  $N_\delta$  is the number of trade arrivals generated by the model simulation over horizon  $\delta$  and where  $M_\delta$  is the number of  $\delta$  unit time intervals in a year (assuming 6.5-hour trading days, and 250 trading days per year). Expressed as a one-sigma expectation, this “diffusion VIX” can be directly compared to the concurrent, observed VIX index value.

Figure 11 depicts scatter plots of daily volatility estimates against  $1/\sqrt{d_{\text{med}}(\delta)}$  for the 121 days in our passive-period data set. We have transformed the x-axis values by the function  $g(x) = 1/\sqrt{x}$ , since this function roughly linearizes the relationships and since it is an approximation for the number of trade arrivals per unit of time (a millisecond, in this application). We set  $\delta = 1$  hour. Panel (a) of the figure shows estimated 1-minute realized volatility for each day (blue points) as well as 1-minute realized volatility obtained by simulating 1 hour of data from each of the candidate models using the estimated parameters for corresponding days. The median inter-trade durations used for the x-axis values are separately measured for the empirical data and for each of the simulations. The ACD-GARCH RV estimates are so extreme that very few fall on the plot itself – if we had adjusted the scale of the vertical axis to include these points, the estimates of the other models would not have been discernible. Panel (b) similarly compares the daily closing value of the VIX (blue points) with the same daily estimates of trade rate,  $1/\sqrt{d_{\text{med}}(\delta)}$ . This latter panel also includes model-implied VIX estimates (as described above) and median inter-trade durations from the same simulations used to compute RV in panel (a). These computations make explicit use of the assumption of Brownian diffusion under trade time.

Figure 11 identifies a very important fact: empirical volatility and trade rate exhibit a strong relationship. Although the source of correlation between these quantities is not determined through simple scatters, the relationship supports the conclusion of our work, that latent information flow drives time-varying volatility through trade arrival. What is most remarkable is the ability of the compound MSMD model to closely replicate daily volatility estimates, while the compound ACD is almost equivalent to the compound Poisson/Exponential. Although this is an admittedly in-sample exercise, it shadows the out-of-sample results reported in the next section.

In summary, as our work in Sections 3 and 4 suggest that asset prices follow a Brownian diffusion in trade time, and as the volatility of a Brownian diffusion scales linearly with the square root of the number of increments, we view clock-time volatility as directly related to the trade rate of the driving process. The implication is that understanding volatility dynamics and prices can largely be accomplished by understanding the process driving trade arrival. Further, our framework allows for the estimation of volatility at very fine time scales, without discarding or smoothing over large quantities of observations.

## 6 Realized Volatility Forecast Performance

Each of the models proposed in Section 4 and evaluated in Section 5 has a differing number of parameters: the compound Poisson has 2, the compound ACD as 5, the compound MSMD has 7 and the ACD-GARCH(1,1) has 4. To deal with the potential issue of over-fitting, and as an alternate method of comparison, we evaluate each model’s performance in out-of-sample realized volatility (RV) forecasting. We propose the following forecasting procedure:

1. For each day that the E-mini traded and for each model under consideration, use all transactions from 2:15 - 3:00 p.m. to estimate the parameters of the model, which we will denote  $\hat{\theta}_M$ .
2. Simulate one hour of transactions/returns from the estimated model:

$$r_M(t + j\delta) = N\left(N_{\hat{\theta}_M}(t + j\delta)\hat{\mu}, \sqrt{N_{\hat{\theta}_M}(t + j\delta)}\hat{\sigma}\right), \quad \text{for } j = 1, \dots, T/\delta, \quad (32)$$

where  $t$  represents the start of the forecasting period (3:00 p.m.),  $T$  is the total number minutes in the time period (in our case  $T = 60$  minutes),  $T/\delta$  represents the number of 1- or 5-minute time intervals in the hour,  $M$  denotes the model under consideration (with corresponding estimates  $\hat{\theta}_M$ ) and where  $N_{\hat{\theta}_M}(t + j\delta)$  represents the number of trade arrivals in interval  $j$  under the model simulation.

3. Compute 1-minute and 5-minute realized volatility for the simulated data, where  $\delta$ -minute realized volatility is defined as

$$RV_M(\delta) = \left( \sum_{j=1}^{T/\delta} r_M^2(t + j\delta) \right)^{1/2}, \quad (33)$$

and compare with observed 1-minute and 5-minute realized volatilities for E-mini returns

$$RV_{ES}(\delta) = \left( \sum_{j=1}^{T/\delta} r_{ES}^2(t + j\delta) \right)^{1/2}, \quad (34)$$

between 3:00 and 4:00 p.m.

The overall objective is to utilize a contiguous segment of transactions each day that are not affected by news announcements in order to estimate and forecast. Although a number of news releases are regularly scheduled for 2:00 p.m. Eastern Time, the time span we use for fitting and forecasting is free of macroeconomic news announcements during the test period.

Panel (a) of Table 6 reports summary quantiles for the empirical distributions of the realized volatilities. There are a total of 244 days in the test period. Since  $T = 60$  minutes in our exercise, all of the estimates in the table correspond to hourly volatility computed under two values of  $\delta$ . As expected, panel (a) shows that hourly volatility is a very noisy quantity. This is both a result of inherent noisiness, as well as estimation error. Although we have a wealth of data (thousands of transactions for each day, during the time window under consideration), we cannot simply improve our RV estimates by reducing  $\delta$  and sampling the data more finely: microstructure effects introduce a bias that render this infeasible ([Andersen et al. \(1999\)](#)). Much of the empirical RV literature has found 5-minute RV to be a suitable, bias-free (but noisy) quantity ([Patton and Sheppard \(2009\)](#)); we also report 1-minute RV since it allows the use of more, albeit potentially bias-affected, data.

Panel (b) of the table reports root-mean-square errors (RMSE) for both 1-minute and 5-minute RV forecasts under each model ( $M$ ), relative to the estimated 1-minute and 5-minute E-mini RV values ( $ES$ ):

$$RMSE_{M,ES} = \left( \sum_{i=1}^D (RV_{ES,i}(\delta) - RV_{M,i}(\delta))^2 \right)^{1/2}, \quad (35)$$

where  $D = 244$  represents the number of days in our forecasting period. The results show that the compound MSMD model readily beats each of the others in terms of out-of-sample RV forecasting, despite the fact that it has more parameters. Not surprisingly, the compound ACD model performs much better than the compound Poisson/Exponential, and the ACD-GARCH performs worst. The latter result is due to the fact that the ACD-GARCH model has a high probability of generating returns with large magnitudes (recall the scale of the axes in Figure 9) during relatively short intervals of time. Consequentially, the RMSE of the ACD-GARCH model is two orders of magnitude worse than all of the others.

Panel (c) of the table reports two sets of Diebold and Mariano (1995) statistics. The statistics proposed by Diebold and Mariano (1995) test the null hypothesis of equal predictive accuracy for competing forecasts, even in the presence of contemporaneous and serial correlation. Since our forecasts are separated by periods of time that are large relative to the forecast time scale, it is not expected that serial correlation is an issue. We corroborate this fact with the empirical autocorrelation functions, although we do not report them here for considerations for space. As such, we compute Diebold-Mariano statistics both for the case of no serial correlation among forecasts, as well as a rather long truncation lag of  $k = 30$  for the ACF. The former statistics are reported above the diagonal in panel (c) and the latter are reported below the diagonal. In all cases, the test statistics are significant at the 5% level or better and demonstrate that the differences in forecast performance reported in panel (b) are not a result of sampling error.

## 7 Conclusion

The empirical and theoretical work of this paper is intimately linked to the work of Mandelbrot (1963), Clark (1973), Brada et al. (1966), Mandelbrot and Taylor (1967) and Ane and Geman (2000), which show that fat-tailed returns distributions are consistent with a Gaussian random walk directed by an appropriate stochastic process. Our empirical insight is that during our 2014 sample period, filtering market-wide news announcements effectively reduces the price process for a highly liquid market asset (the E-Mini S&P 500 near-month futures contract) to a continuous local martingale, and that the subordinating process for those continuous prices is simply characterized by a model of high-frequency trade arrival. Our theoretical contribution is to compose a parsimonious inter-trade duration model that serves as an appropriate directing process with a Gaussian random walk to arrive at a hierarchical model of returns in clock time and to show when this model can simultaneously explain the dynamics and unconditional nature of observed returns. Returns in the tail of the distribution arise as a consequence of faster random walks generated by periods where the trading rate is high and volatility persistence is generated by serial correlation within the trade arrival process. The upshot is that outside of pre-scheduled news-affected periods, the observed non-Gaussianity in the E-mini returns distribution can be fully attributed to the temporal clustering of trades. It is important to recognize that these results are relevant to our particular sample period and may not extend to more volatile market states.

Further work appears warranted. While the compound ACD and compound MSMD models provide the best in-sample fits and the compound MSMD model provides the best out-of-sample fit compared to other models we consider, they both fall far short of explaining the low-level of persistence exhibited in observed inter-trade durations. To the extent that trade-time returns abide by a random walk (as suggested by our empirical work), improvements in modeling the directing process would yield excellent dividends in explaining returns dynamics. Further, while we have treated durations and returns as mutually independent processes, characterizing their joint distribution could enhance the model.

Finally, it is possible that the Gaussian spectrum of trade-time returns is a feature that is largely specific to the heavily traded E-mini. Expanding the work of this paper to a broader set of assets could lead to important innovations to the model. Additionally, it would be useful to adapt the model to explain the evolution of returns under conditions of market stress.

Model	Parameter	Mean	Std. Dev.	Full Sample
Exponential	$\gamma$	0.00535	0.00232	0.00541
MSMD	$\lambda$	0.107	0.0206	0.0963
MSMD	$\gamma_{\bar{k}}$	0.637	0.0945	0.679
MSMD	$b$	5.37	0.838	5.17
MSMD	$m_0$	0.124	0.0243	0.134
ACD	$\omega$	2.17	0.483	1.90
ACD	$\alpha$	0.226	0.0265	0.225
ACD	$\beta$	0.423	0.0874	0.469
Gaussian	$\mu$	0.000	0.000000217	0.000
Gaussian	$\sigma$	0.0000606	0.00000441	0.0000616

**Table 1:** Parameter estimates for duration models and Gaussian trade-time returns distributions.

Statistic	Model			
	Exponential	MSMD	ACD	Data
Mean	197	4350	304	392
SD(Mean)	2.78	3720	8.13	1150
Full Sample Mean	197	2700	304	185
SD	197	18,900	339	709
SD(SD)	3.90	11,700	12.7	1770
Full Sample SD	197	15,800	339	529
Skew	1.98	9.77	2.55	4.12
SD(Skew)	0.106	4.06	0.270	0.632
Full Sample Skew	1.98	13.9	2.57	36.7
Kurt	5.78	128	10.6	27.1
SD(Kurt)	1.09	111	3.93	10.4
Full Sample Kurt	5.82	252	10.9	3520

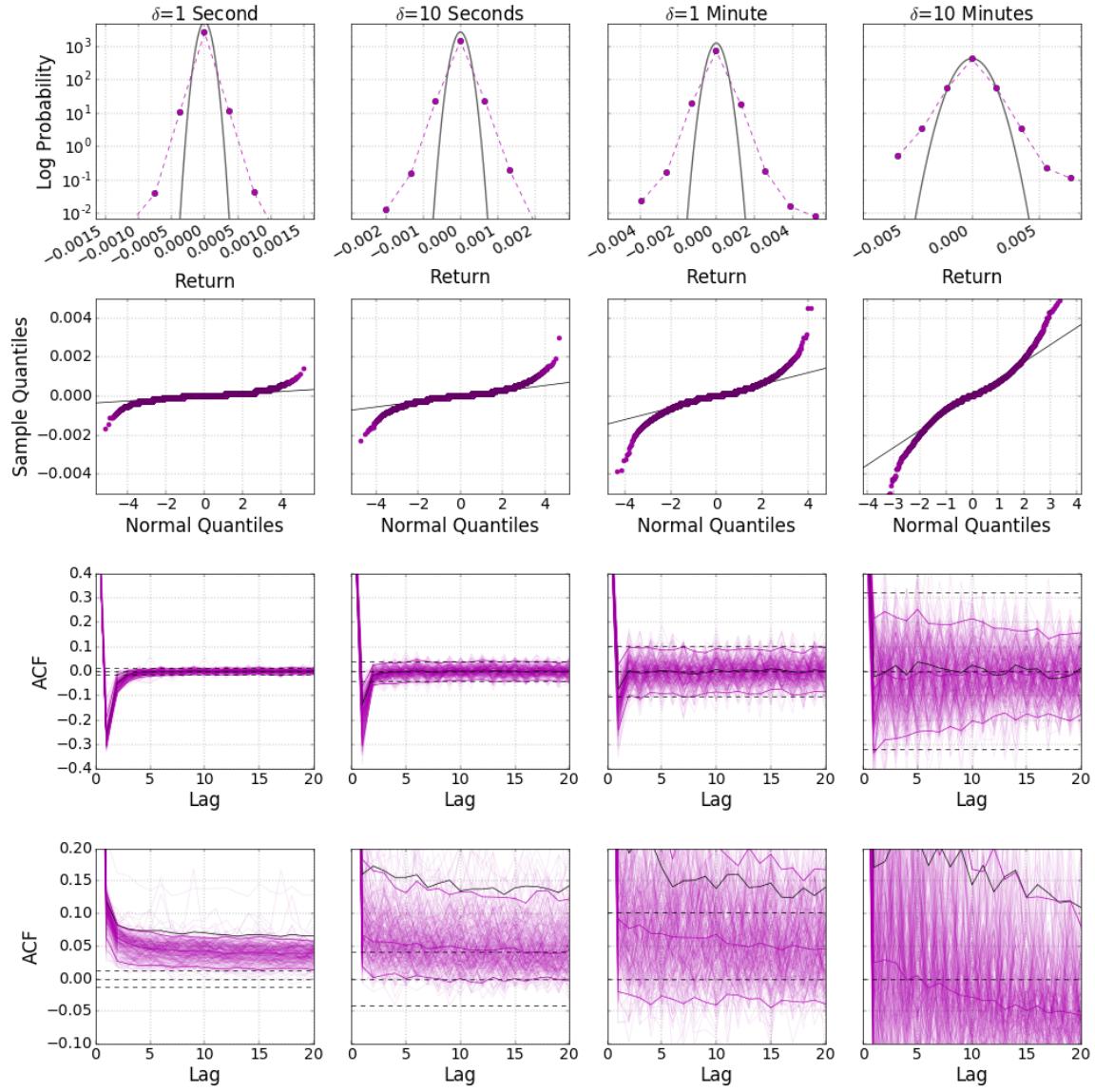
**Table 2:** Summary statistics for observed inter-trade durations, as well as the daily simulations under each of the models. The first four moments are computed day by day, and averages across days are reported. Single point estimates obtained by treating each data set (comprised of 58 days) as one contiguous sequence of observations are also reported.

Model	$\delta$					
	250	500	1000	5000	10000	30000
<b>(a) Fraction of Accepted Hypotheses (Daily Kurtosis)</b>						
Compound Exp.	0	0	0.141	0.934	0.959	0.967
Compound ACD	0	0	0	0.711	0.909	0.967
Compound MSMD	0	0	0	0	0	0.0909
ACD-GARCH	0	0	0	0.0536	0.170	0.625
Data	0	0	0.157	0.636	0.769	0.975
<b>(b) Kullback-Leibler Divergence (Daily Kurtosis)</b>						
Compound Exp.	67.1	12.8	2.68	0.0402	0.0422	0.0694
Compound ACD	57.1	17.4	6.66	0.418	0.0564	0.0443
Compound MSMD	7.56	6.74	5.98	4.21	3.46	1.98
ACD-GARCH	6.44	5.24	4.56	2.56	2.05	0.873
Data	4.25	3.37	2.66	1.33	0.741	0.0379
<b>(c) Fraction of Accepted Hypotheses (Ljung-Box)</b>						
Compound Exp.	0.942	0.967	0.967	0.983	0.967	0.983
Compound ACD	0.851	0.917	0.950	0.975	0.950	0.992
Compound MSMD	0.0579	0.157	0.372	0.744	0.950	0.992
ACD-GARCH	0.0268	0.259	0.705	0.982	0.982	0.964
Data	0.248	0.653	0.884	0.950	0.967	0.965
<b>(d) Kullback-Leibler Divergence (Ljung-Box)</b>						
Compound Exp.	0.0950	0.0711	0.0845	0.146	0.205	0.398
Compound ACD	0.103	0.0807	0.138	0.143	0.209	0.383
Compound MSMD	3.58	2.63	1.84	0.880	0.838	1.45
ACD-GARCH	3.23	2.07	0.823	0.583	0.668	0.739
Data	1.61	0.455	0.201	0.205	0.360	0.329
<b>(e) MSE of Autocorrelations</b>						
Compound Exp.	1.02	0.989	0.994	0.895	0.783	0.627
Compound ACD	1.33	1.19	1.12	0.949	0.885	0.682
Compound MSMD	4.10	4.05	3.94	3.05	2.54	0.838
ACD-GARCH	3.57	3.21	2.16	0.976	0.886	0.585
Data	2.31	1.89	1.59	1.06	0.910	0.655

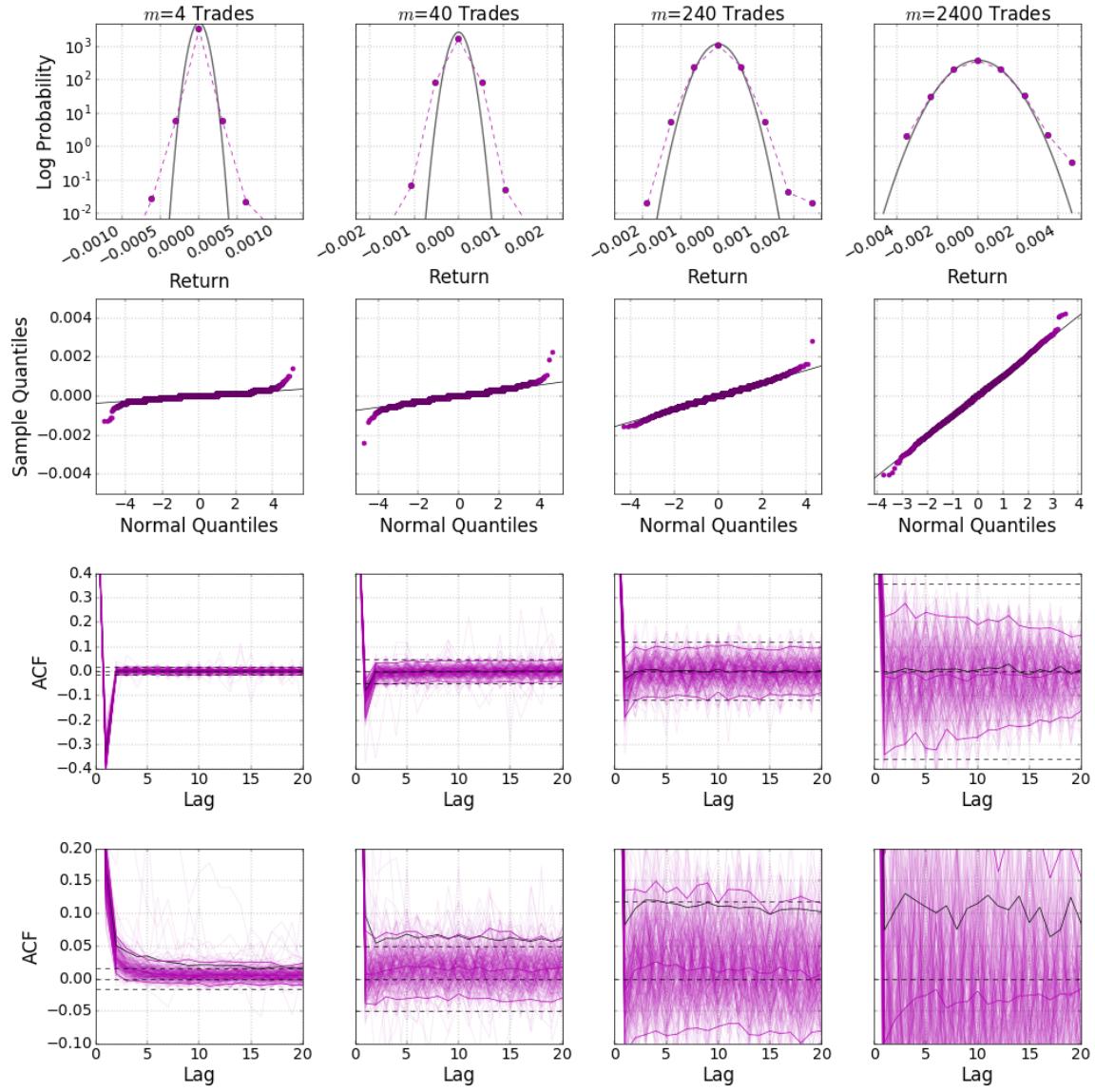
**Table 3:** Sample statistics under null hypotheses related to the distributions of daily kurtosis and autocorrelation estimates of squared returns. Panels (a) and (c) report fractions of accepted null hypotheses and panels (b) and (d) report Kullback-Leibler distances of empirical distributions of daily estimates from limiting distributions under the null hypotheses. Panel (e) reports the integral of lag-by-lag mean squared error (MSE) of daily autocorrelations. All statistics are computed for each of the compound duration models considered in this paper as well as the ACD-GARCH(1,1) and the passive-period clock-time E-mini data, using returns computed across clock-time intervals  $\delta = \{250, 500, 1000, 5000, 10000, 30000\}$  milliseconds.

(a) Realized Volatility Empirical Quantiles (%)					
Time Interval	Min	0.05	Median	0.95	Max
1 min	0.06556	0.1140	0.1844	0.4328	0.7187
5 min	0.04886	0.07883	0.1651	0.4366	0.9176
(b) RMSE (%)					
Time Interval	Compound Exp.	Compound ACD	Compound MSMD	ACD-GARCH	
1 min	0.4290	0.2596	0.1245	11.90	
5 min	0.4432	0.2907	0.1368	11.83	
(c) Diebold-Mariano Statistics					
	Comp. Exp.	Comp. ACD	Comp. MSMD	ACD-GARCH	
Comp. Exp.	NA	7.87	12.3	-28.4	
Comp. ACD	11.0	NA	8.74	-28.4	
Comp. MSMD	13.6	6.95	NA	-28.4	
ACD-GARCH	-25.2	-25.2	-25.2	NA	

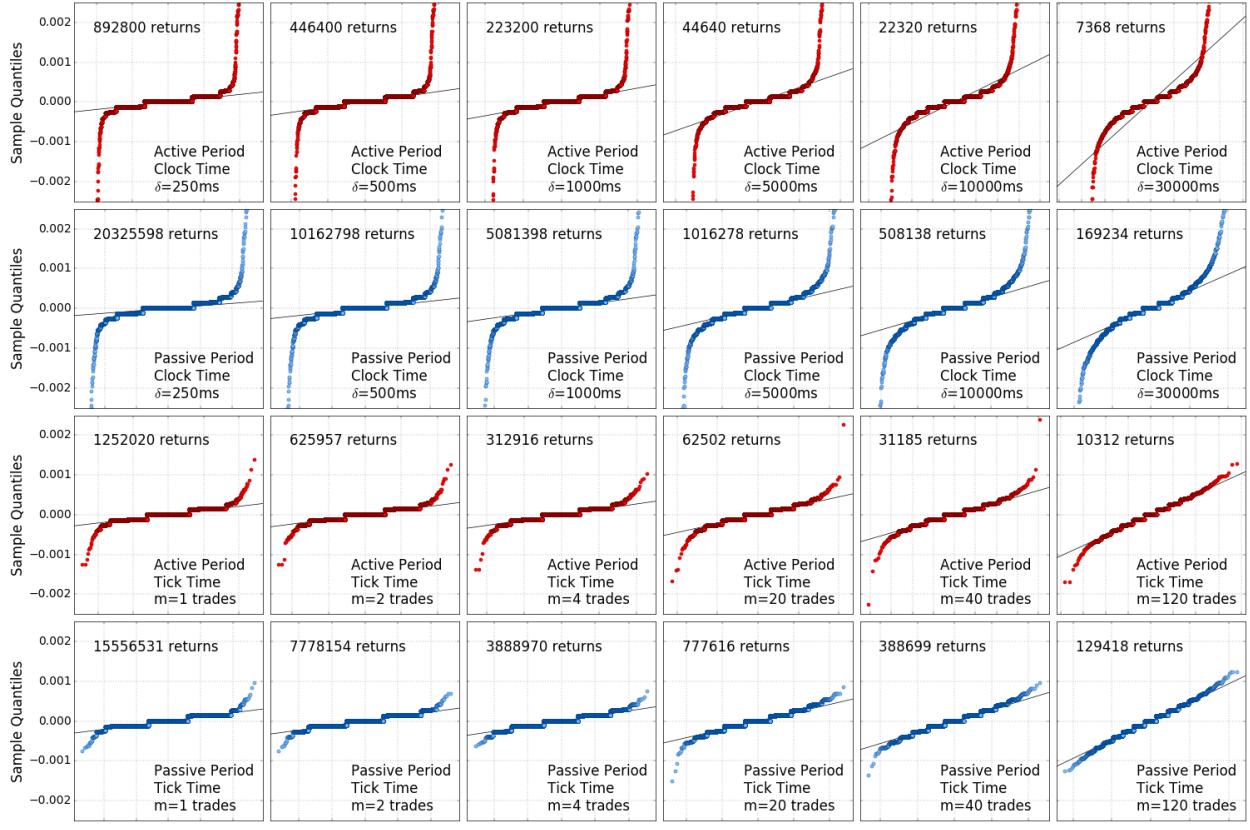
**Table 4:** Out of sample realized volatility (RV) forecasting results. Panel (a) reports quantiles of the empirical distributions of 1-minute and 5-minute RV. Panel (b) reports root-mean-square errors of 1-minute and 5-minute RV under each model. Panel (c) reports two sets of pairwise Diebold-Mariano statistics under the null hypothesis of equal prediction accuracy among models – the values above the diagonal assume no serial correlation in forecast errors and those below the diagonal truncate the error ACFs at lag  $k = 30$ .



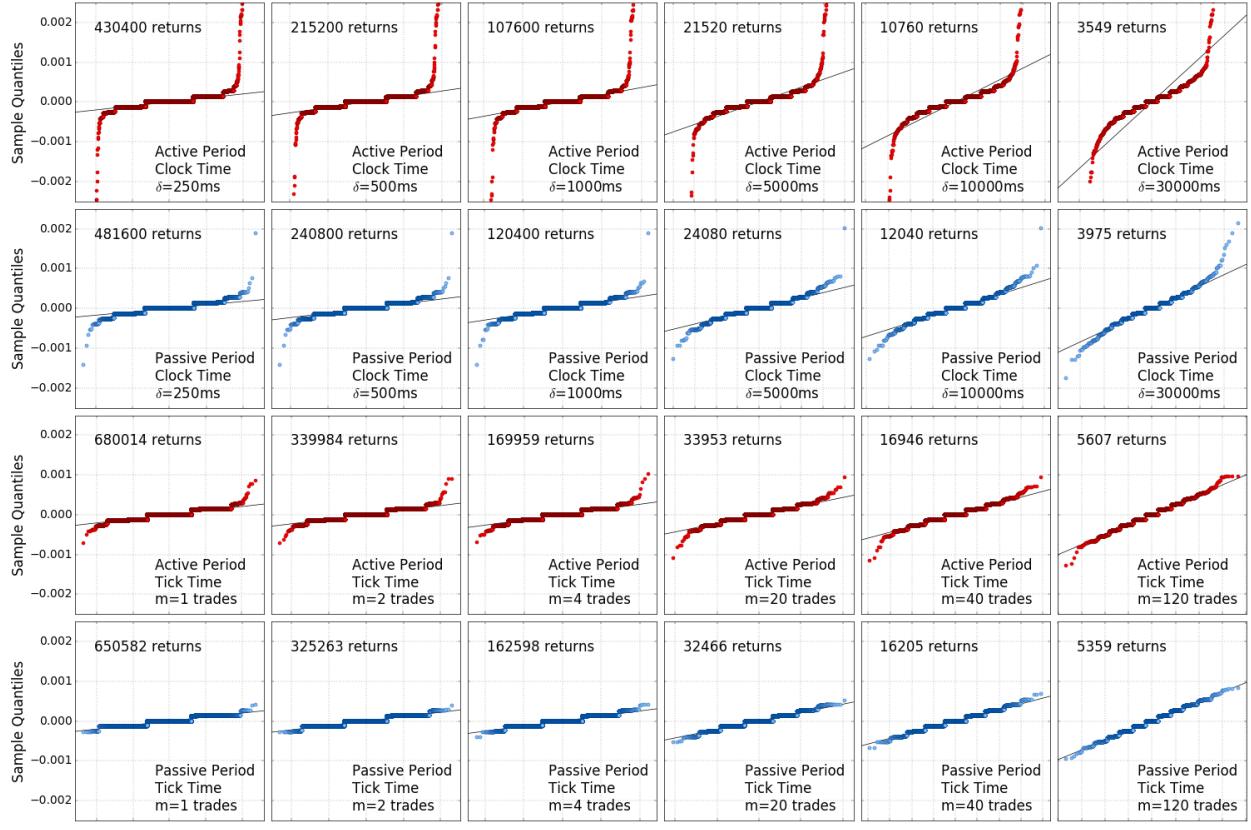
**Figure 1:** Empirical densities, Q-Q plots and autocorrelations of clock-time returns returns and squared clock-time returns. Clock-time intervals correspond to 1 second, 10 seconds, 1 minute and 10 minutes.



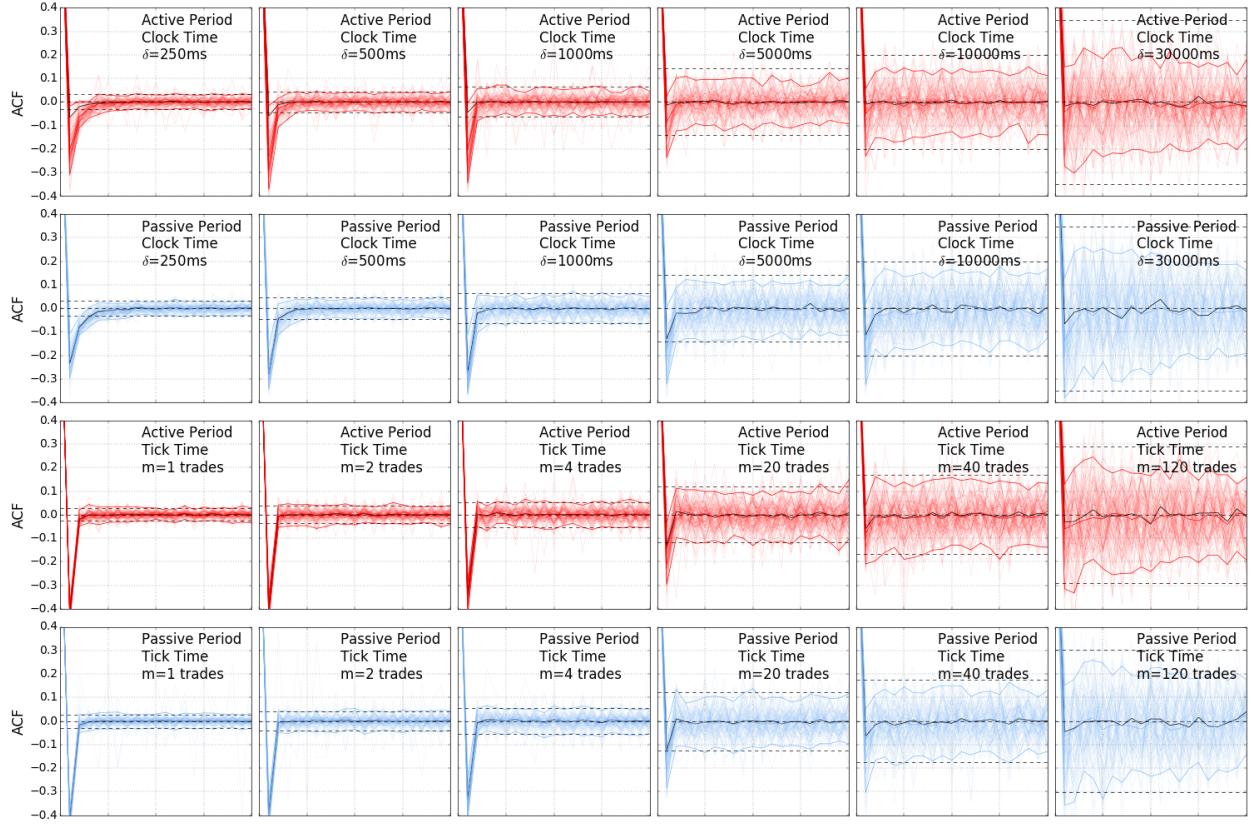
**Figure 2:** Empirical densities, Q-Q plots and autocorrelations of trade-time returns returns and squared trade-time returns. Trade-time intervals correspond to 4 trades, 40 trades, 240 trades and 2400 trades.



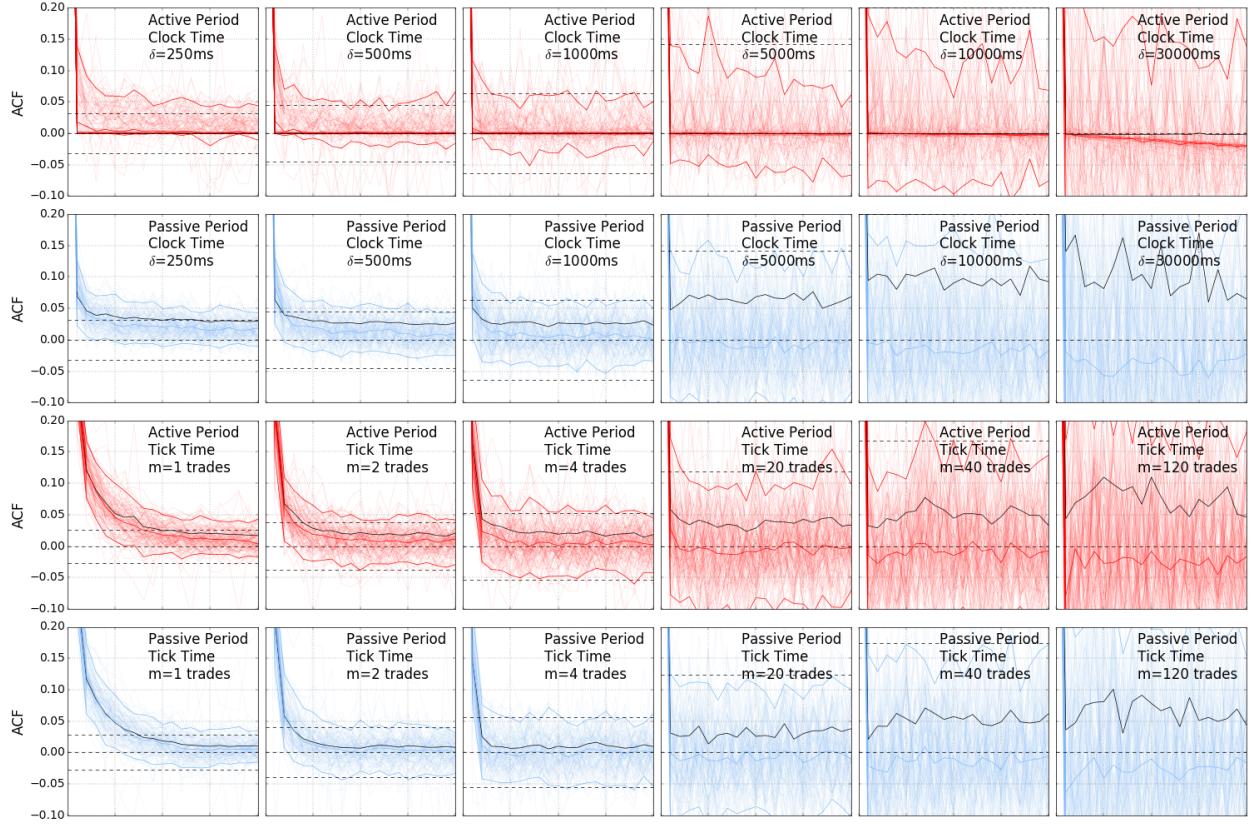
**Figure 3:** Sample Q-Q plots for returns in both clock time and trade time and in both active and passive samples of the full-day dataset. Each panel corresponds to a clock-time or trade-time interval during either the active or passive subsample of data. The vertical axes depict sample quantiles of the data and the horizontal axes depict theoretical quantiles of best-fit Gaussian distributions.



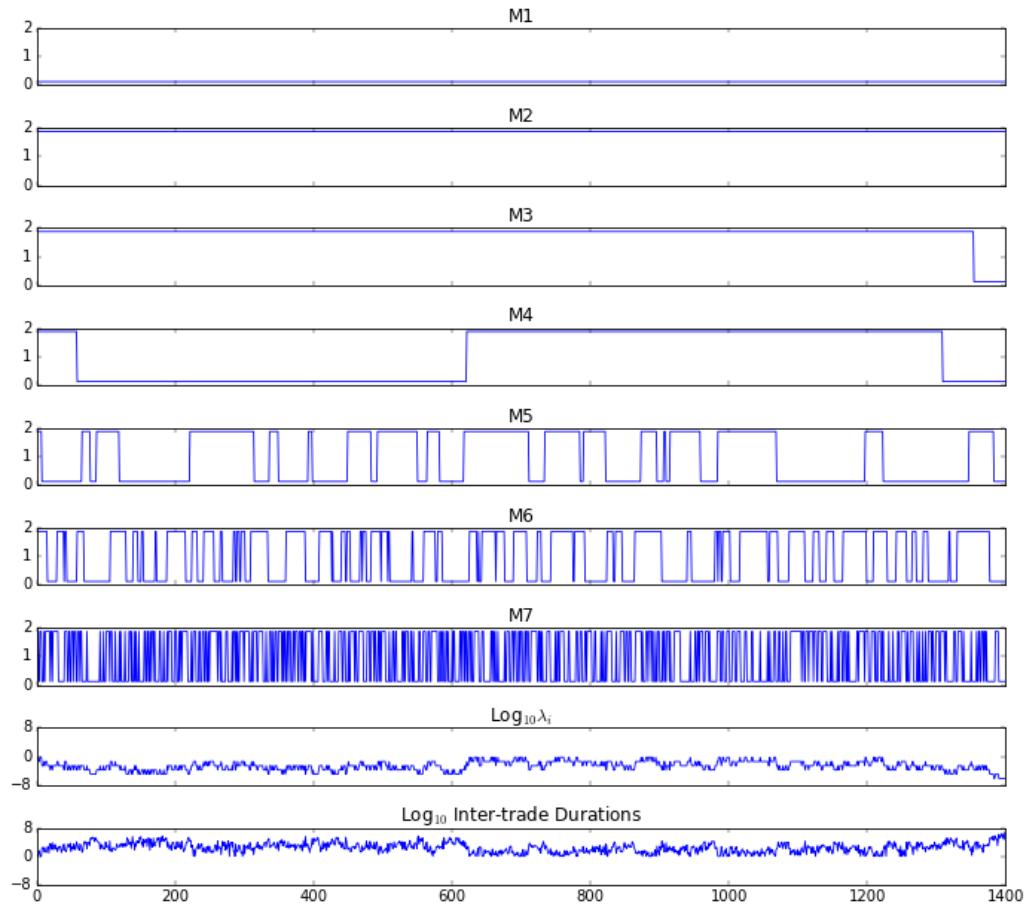
**Figure 4:** Sample Q-Q plots for returns in both clock time and trade time and in both active and passive samples of the reduced, 10:00 a.m. dataset. Each panel corresponds to a clock-time or trade-time interval during either the active or passive subsample of data. The vertical axes depict sample quantiles of the data and the horizontal axes depict theoretical quantiles of best-fit Gaussian distributions.



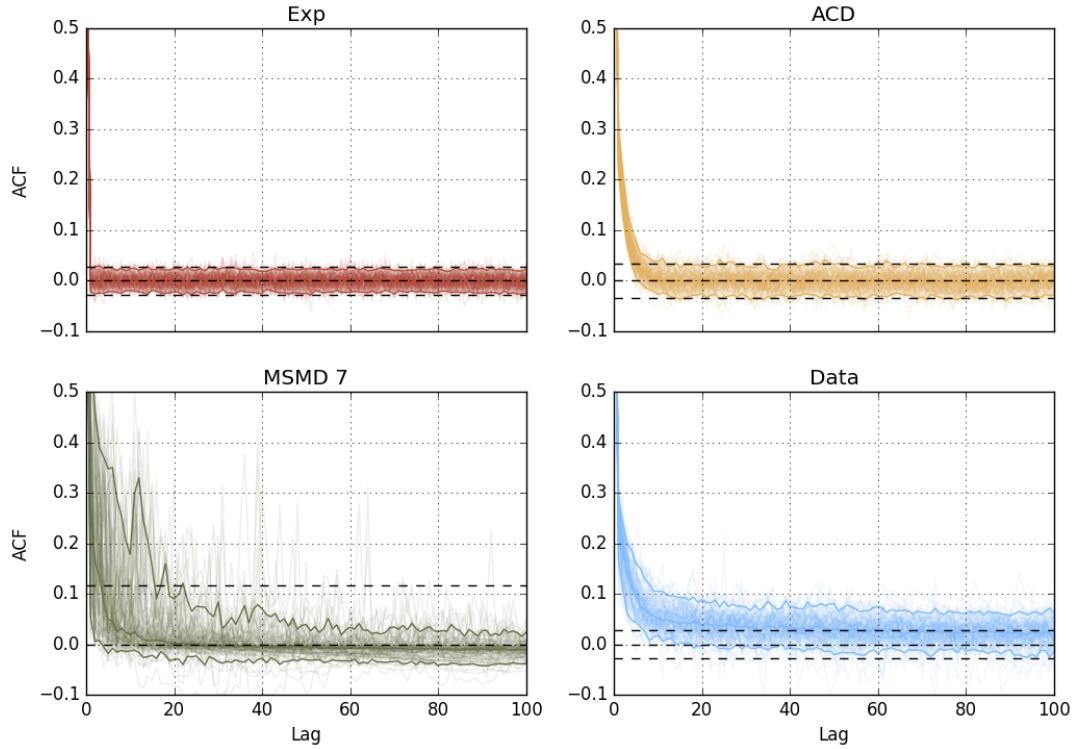
**Figure 5:** Sample autocorrelation functions of returns in both clock time and trade time and in both active and passive subsamples. Each panel corresponds to a clock-time or trade-time interval during either the active or passive subsample of data.



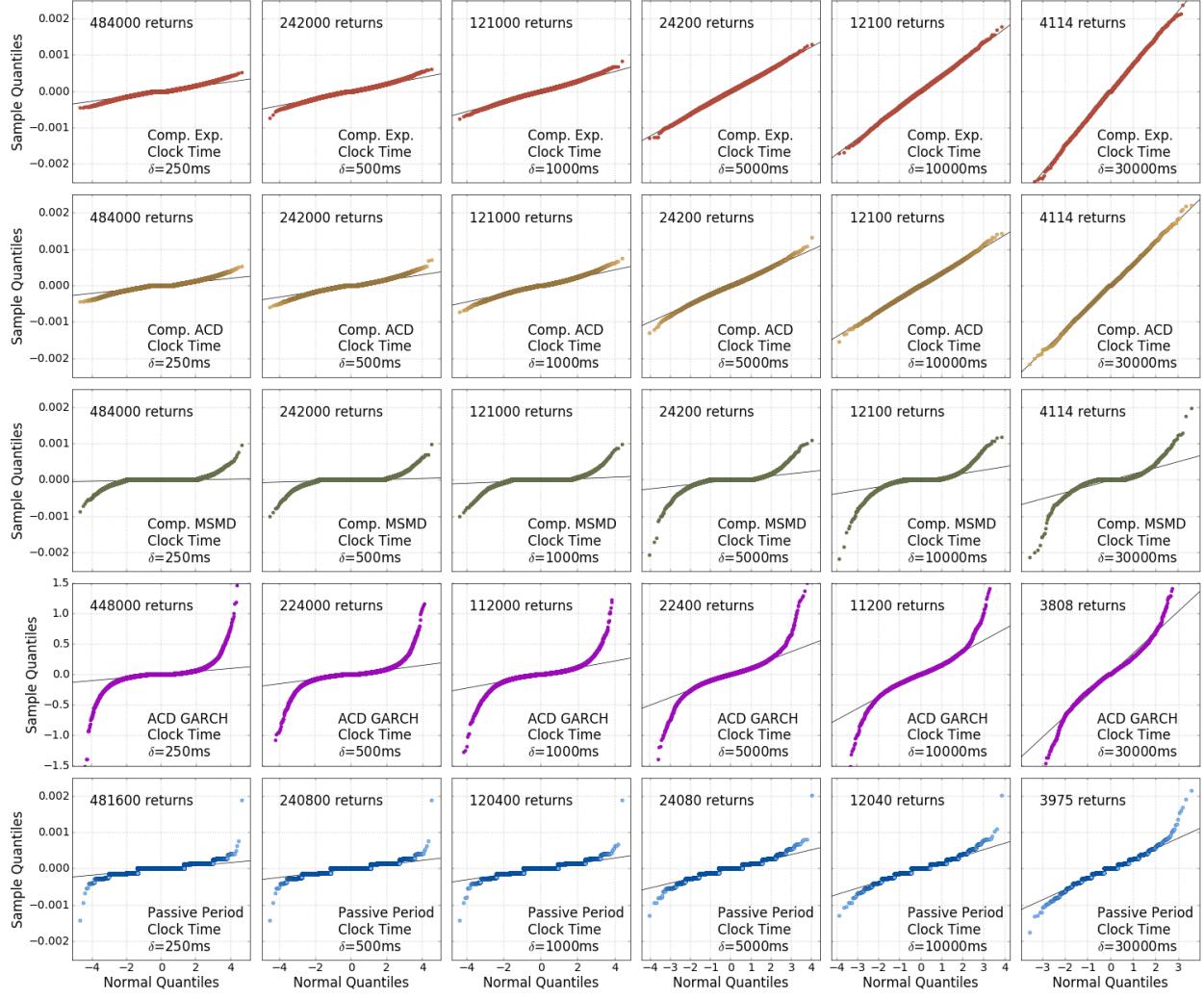
**Figure 6:** Sample autocorrelation functions of squared returns in both clock time and trade time and in both active and passive subsamples. Each panel corresponds to a clock-time or trade-time interval during either the active or passive subsample of data.



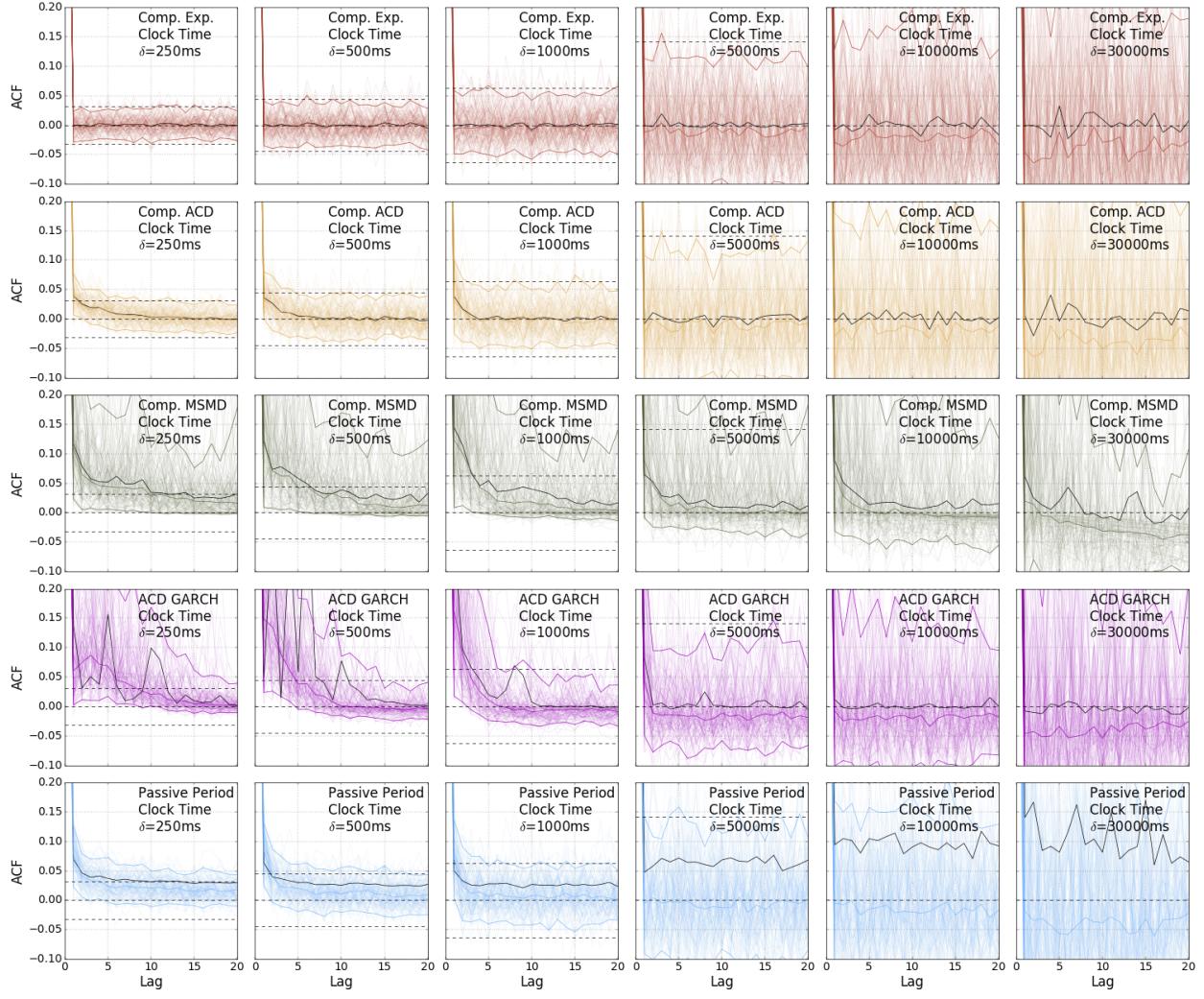
**Figure 7:** Time series of latent states for a simulation of 6.5 hours of data (one trading day) from the estimated MSMD model.



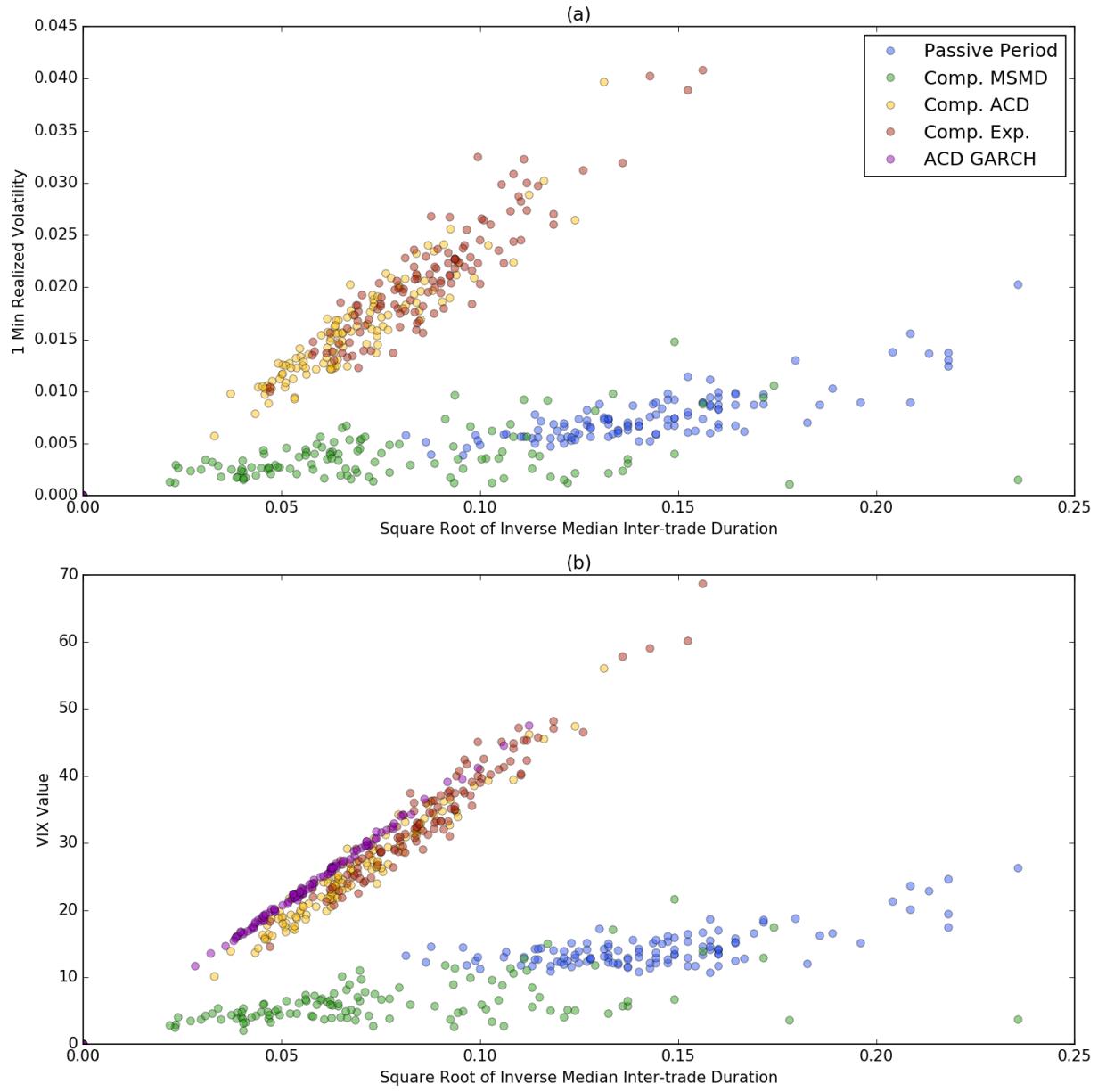
**Figure 8:** Daily sample autocorrelation functions for observed durations (blue) and durations simulated under the estimated Exponential (red), ACD (yellow) and MSMD (green) models. Faint lines represent daily ACFs and darker lines represent 0.05, 0.5 and 0.95 quantiles of the distributions of autocorrelations at each lag. The dotted black lines are 95% Bartlett bounds for the null hypothesis of no serial correlation.



**Figure 9:** Sample Q-Q plots for simulated clock-time returns under each of the compound duration models considered in this paper as well as the ACD-GARCH(1,1). The bottom row of panels is a reproduction of the E-mini passive-period clock-time Q-Q plots shown in Figure 4. Q-Q plots are shown for returns computed across clock-time intervals  $\delta = \{250, 500, 1000, 5000, 10000, 30000\}$  milliseconds.



**Figure 10:** Sample autocorrelation functions for simulated clock-time squared returns under each of the compound duration models considered in this paper as well as the ACD-GARCH(1,1). The bottom row of panels is a reproduction of the E-mini passive-period clock-time ACFs shown in Figure 6. ACFs are shown for squared returns computed across clock-time intervals  $\delta = \{250, 500, 1000, 5000, 10000, 30000\}$  milliseconds.



**Figure 11:** Scatter plots of volatility estimates against square root of inverse median inter-trade duration. Panel (a) shows 1-minute realized volatility for E-mini data and simulations from each of the candidate models, on the 121 days in our passive-period sample. Panel (b) shows closing values of the VIX index and VIX values implied by the simulations from the candidate models.

## References

- Aït-Sahalia, Y., Fan, J., and Li, Y. (2013), “The leverage effect puzzle: Disentangling sources of bias at high frequency,” *Journal of Financial Economics*, 109, 224–249.
- Aldrich, E. M. (2013), “Trading Volume in General Equilibrium with Complete Markets,” *Working Paper*.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (1999), “Realized Volatility and Correlation,” *Working Paper*.
- Andersen, T. G., Bollerslev, T., and Dobrev, D. (2007), “No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications,” *Journal of Econometrics*, 138, 125–180.
- Andersen, T. G., Bollerslev, T. I. M., and Frederiksen, P. E. R. (2010), “Continuous-time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns,” *Journal of Applied Econometrics*, 261, 233–261.
- Ane, T. and Geman, H. (2000), “Order Flow, Transaction Clock, and Normality of Asset Returns,” *The Journal of Finance*, LV, 2259–2284.
- Bachelier, L. (1900), *Théorie de la Spéculation*, Paris: Gauthier-Villars.
- Black, F. and Scholes, M. (1973), “The Pricing of Options and Corporate Liabilities,” *Journal of Political Economy*, 81, 637–654.
- Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.
- Bouchaud, J.-P. (2005), “The subtle nature of financial random walks.” *Chaos*, 15, 26104.
- Brada, J., Ernst, H., and van Tassel, J. (1966), “The Distribution of Stock Price Differences: Gaussian after All?” *Operations Research*, 14, 334–340.
- Brogaard, J., Hendershott, T., and Riordan, R. (2014), “High Frequency Trading and Price Discovery,” *Review of Financial Studies*, 27, 2267–2306.
- Calvet, L. and Fisher, A. (2001), “Forecasting multifractal volatility,” *Journal of Econometrics*, 105, 27–58.
- (2002), “Multifractality in Asset Returns: Theory and Evidence,” *The Review of Economics and Statistics*, 84, 381–406.
- Calvet, L., Fisher, A., and Mandelbrot, B. (1997), “Large Deviations and the Distribution of Price Changes,” *Cowles Foundation Discussion Paper No. 1165*.
- Calvet, L. E. (2004), “How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes,” *Journal of Financial Econometrics*, 2, 49–83.
- Calvet, L. E. and Fisher, A. J. (2008), *Multifractal Volatility: Theory, Forecasting, and Pricing*, Burlington, MA: Academic Press.
- (2012), “Extreme Risk and Fractal Regularity in Finance,” *SSRN Electronic Journal*.

- Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002), "The Fine Structure of Asset Returns: An Empirical Investigation," *The Journal of Business*, 75, 305–333.
- Carr, P. and Wu, L. (2004), "Time-changed Lévy processes and option pricing," *Journal of Financial Economics*, 71, 113–141.
- Chen, F., Diebold, F. X., and Schorfheide, F. (2013), "A Markov-switching multifractal inter-trade duration model, with application to US equities," *Journal of Econometrics*, 177, 320–342.
- Clark, P. K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.
- Deo, R., Hurvich, C. M., Soulier, P., and Wang, Y. (2009), "Conditions for the Propagation of Memory Parameter From Durations To Counts and Realized Volatility," *Econometric Theory*, 25, 764.
- Diebold, F. X. and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 134–144.
- Easley, D., de Prado, M. M. L., and O'Hara, M. (2012), "The Volume Clock: Insights into the High Frequency Paradigm," *Journal of Portfolio Management*.
- Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1008.
- (2000), "The Econometrics of Ultra-High-Frequency Data," *Econometrica*, 68.
- Engle, R. F. and Russell, J. R. (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.
- Epps, T. W. and Epps, M. L. (1976), "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis," *Econometrica*, 44, 305–321.
- Fama, E. F. (1965), "The Behavior of Stock-Market Prices," *The Journal of Business*, 38, 34–105.
- Fisher, A., Calvet, L., and Mandelbrot, B. (1997), "Multifractality of Deutschemark / US Dollar Exchange Rates," *Cowles Foundation Discussion Paper No. 1165*.
- Gallant, A. R., Rossi, P. E., and Tauchen, G. (1992), "Stock Prices and Volume," *The Review of Financial Studies*, 5, 199–242.
- Ghysels, E., Gouriéroux, C., and Jasiak, J. (1995), "Market Time and Asset Price Movements: Theory and Estimation," *CIRANO Working Paper No. 95s-32*.
- Grammig, J. and Maurer, K.-O. (2000), "Non-monotonic hazard functions and the autoregressive conditional duration model," *Econometrics Journal*, 3, 16–38.
- Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.
- Hasbrouck, J. (2003), "Intraday Price Formation in U.S. Equity Index Markets," *The Journal of Finance*, LVIII, 2375–2399.

- Hasbrouck, J. and Saar, G. (2013), “Low-latency trading,” *Journal of Financial Markets*, 16, 646–679.
- Jones, C. M., Kaul, G., and Lipson, M. L. (1994), “Transactions, Volume, and Volatility,” *The Review of Financial Studies*, 7, 631–651.
- Karpoff, J. M. (1987), “The Relation Between Price Changes and Trading Volume: A Survey,” *The Journal of Financial and Quantitative Analysis*, 22, 109–126.
- Kullback, S. and Leibler, R. A. (1951), “On Information and Sufficiency,” *The Annals of Mathematical Statistics*, 22, 79–86.
- Laughlin, G., Aguirre, A., and Grundfest, J. (2014), “Information Transmission between Financial Markets in Chicago and New York,” *The Financial Review*, 49, 283–312.
- Lunde, A. (1999), “A Generalized Gamma Autoregressive Conditional Duration Model,” *Working Paper*, 1–30.
- Madan, D. B., Carr, P. P., and Chang, E. C. (1998), “The Variance Gamma Process and Option Pricing,” *European Finance Review*, 2, 79–105.
- Madan, D. B. and Seneta, E. (1990), “The Variance Gamma (V.G.) Model for Share Market Returns,” *The Journal of Business*, 63, 511–524.
- Mandelbrot, B. (1963), “The Variation of Certain Speculative Prices,” *The Journal of Business*, 36, 394–419.
- Mandelbrot, B., Fisher, A., and Calvet, L. (1997), “A Multifractal Model of Asset Returns,” *Cowles Foundation Discussion Paper No. 1164*.
- Mandelbrot, B. and Taylor, H. M. (1967), “On the Distribution of Stock Price Differences,” *Operations Research*, 15, 1057–1062.
- Mandelbrot, B. B. (1973), “Comments on: ”A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices,” by Peter K . Clark,” *Econometrica*, 41, 157–159.
- Patton, A. J. and Sheppard, K. (2009), “Optimal combinations of realised volatility estimators,” *International Journal of Forecasting*, 25, 218–238.
- Press, S. J. (1967), “A Compound Events Model for Security Prices,” *The Journal of Business*, 40, 317–335.
- (1968), “A Modified Compound Poisson Process with Normal Compounding,” *Journal of the American Statistical Association*, 63, 607–613.
- Renault, E. and Werker, B. J. (2011), “Causality effects in return volatility measures with random times,” *Journal of Econometrics*, 160, 272–279.
- Rydberg, T. H. and Shephard, N. (2003), “Dynamics of Trade-by-Trade Price Movements: Decomposition and Models,” *Journal of Financial Econometrics*, 1, 2.
- Tauchen, G., Zhang, H., and Liu, M. (1996), “Volume, volatility, and leverage: A dynamic analysis,” *Journal of Econometrics*, 74, 177–208.

- Tauchen, G. E. and Pitts, M. (1983), “The Price Variability-Volume Relationship on Speculative Markets,” *Econometrica*, 51, 485.
- Todorov, V. and Tauchen, G. (2014), “Limit Theorems for the Empirical Distribution Function of Scaled Increments of Ito Semimartingales at High Frequencies,” *The Annals of Applied Probability*, 24, 1850–1888.
- Zikes, F., Barunik, J., and Shenai, N. (2014), “Modeling and Forecasting Persistent Financial Durations,” *Working Paper*.