Jumps, Realized Densities, and News Premia *

Paul Sangrey
Amazon †

Current Version
This Version: July 11, 2020

Abstract

Announcements and other news continuously barrage financial markets, causing asset prices to jump hundreds of times each day. If price paths are continuous, the diffusion volatility nonparametrically summarizes the return distributions' dynamics. However, this is not true in the empirically-relevant case involving price jumps. To address this impasse, I derive both a tractable nonparametric continuous-time representation for the price jumps and an implied sufficient statistic for their dynamics. This statistic — jump volatility — is the instantaneous variance of the jump part and measures news risk. The realized density then depends, exclusively, on the diffusion volatility and the jump volatility. I develop estimators for both and show how to use them to nonparametrically identify continuous-time jump dynamics. I provide a detailed empirical application to the S&P 500 and show that the jump volatility premium is less than the diffusion volatility premium.

Keywords: Jumps, News Risk, Realized Volatility, High-Frequency Econometrics, Stochastic Volatility, Nonparametric Modeling, Semimartingales, Time Aggregation, Risk Premia

JEL Codes: C51, C55, C58, G12, G14, G17

*I am indebted to my advisors, Francis X. Diebold and Frank Schorfheide, as well as the rest of my committee: Xu Cheng and Amir Yaron. I have also benefited greatly from comments from Jaap Abbring, Hengjie Ai, Ben Connault, Frank DiTraglia, Winston Wei Dou, Jesús Fernández-Villaverde, A. Ronald Gallant, Robert J. Hodrick, Ilze Kalnina, Bryan Kelly, Yuan Liao, George Tauchen, Jessica Wachter, Bas J. M. Werker, and seminar participants at the University of Pennsylvania, the Wharton School at the University of Pennsylvania, the Rady School at the University of California San Diego, the Sauder School at the University of British Columbia, Tilburg University, the SoFiE Summer School at the University of Chicago, the Young Economics Symposium at Yale University, and the George Washington Student Research Conference. All remaining errors are my own.

†901 8th Ave, Apt 607, Seattle WA, 98104 Email: paul@sangrey.io. Web: sangrey.io

1. Introduction

The study of individuals' reactions to time-varying risk forms the core of modern finance and macroeconomics. Asset pricing, portfolio allocation, and performance evaluation all require investors to asses the risk they face in real time. Moreover, optimal financial regulation requires trading off risk and return at the societal level, and real-time risk measures form its core as well. The most general measure of this risk is the distribution of future returns as a function of the information available.

About twenty years ago, Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) substantially enhanced our understanding of the volatility by providing the nonparametric Realized Volatility estimator for the integrated diffusion volatility. Moreover, they showed that as long as price paths are continuous, the diffusion volatility entirely determines the price's continuous-time martingale dynamics. They also derived closed-form expressions for the discrete-time distributions as functions of integrated diffusion volatility by time-aggregating the continuous-time measures.

However, hundreds of quantitatively relevant news releases strike financial markets every day and cause the prices to jump. Aït-Sahalia and Jacod (2009a, 2009b, 2012a) even show that models with infinitely many jumps fit the data better than models with only finitely many jumps do. These jumps are also quite economically significant, (Andersen, Bollerslev, and Diebold 2007). For example, we need them to price derivatives well, (Pan 2002; Branger, Schlag, and Schneider 2008), and they play a fundamental role in driving the variance risk premium, (Todorov 2010; Drechsler 2013).

At present, however, no parsimonious representation with nonparametrically identified dynamics exists for jump processes. Gallant and Tauchen (2018) estimate a jump density after rescaling the returns using Barndorff-Nielsen and Shephard's (2004) bipower-variation estimator. However, they truncate away all the jumps above a shrinking threshold as they must because they rely on Todorov and Tauchen (2014). Consequently, this procedure provides a great deal of information about small jumps but cannot address the dynamics of large jumps. Conversely, Bollerslev, Todorov, and Li (2013) examine large jumps, but gain identification by exclusively focusing on the static problem. This paper studies both small and large jumps and allows for arbitrary dynamics.

To address this identification problem, I derive both a tractable nonparametric continuous-time representation for the price jumps and an implied sufficient statistic for their dynamics. This statistic — jump volatility — is the instantaneous variance of the jump part and measures news risk. The resulting realized density then depends, exclusively, on the diffusion and jump volatilities in continuous-time. In other words, volatilities control all of the distribution's short-horizon dynamics. I then time-aggregate this representation and derive closed-form expressions for the discrete-time densities and volatilities.

I nonparametrically identify the jump dynamics in the presence of stochastic diffusion volatility by deriving the first estimator for any instantaneous jump variation measure. I time-aggregate the instantaneous estimators to estimate the daily diffusion and jump volatilities.

I apply my estimators to the S&P 500 and first show that the diffusion volatility commands an economically and statistically significant premium, as in Brandt and Kang (2004) and Lettau and Ludvigson (2010). This risk premium literature has more recently focused on higher-order risk premia relationships, often analyzing them in terms of volatility of volatility or jumps, (Cheng, Renault, and Sangrey 2019; Dew-Becker, Giglio, and Kelly 2019). I show that the jump volatility premium is substantially less than the diffusion volatility premium.

The remainder of the introduction fixes ideas and explains the close connection between discontinuous information flows and jumps in asset prices. Section 2 lays out the data generating process, while Section 3 proves the representation theorems. Section 4 derives the estimators, and Section 5 characterizes their finite-sample performance in simulations. Section 6 describes my dataset, and Section 7 provides new stylized facts concerning the jump volatility dynamics. Section 8 shows that the jump volatility premium is less than the diffusion volatility premium. Section 9 concludes. The appendices contain proofs, robustness checks, and additional empirical results.

1.1. What Causes Jumps?

Aït-Sahalia and Jacod (2009a) find jumps drive $\approx 40\,\%$ of the squared variation in individual equities and $\approx 10\,\%$ of the variation in the market index using a ratio of bipower-type estimators. Additionally, almost every paper that explicitly tests for the degree of activity finds infinitely-active jumps, or at the very minimum a massive number, (Aït-Sahalia, Mykland, and Zhang 2005; Bakshi, Carr, and Wu 2008; Aït-Sahalia and Jacod 2009a).[1] From both a modeling and pricing perspective, a large number of jumps and infinitely-many are effectively equivalent in practice, as shown in Section 3.3.3. Even if the literature has not reached a consensus on the precise number and magnitude of the jumps, jumps are indisputably ubiquitous and crucial to understanding price dynamics.

There are two equivalent characterizations of jumps. First, a jump is a discontinuity in the price process — the price changes by such a large amount over such a small interval that we cannot draw a continuous line through it. However, this is a mathematical definition; we would like an economic characterization.

Various authors, such as Andersen, Bollerslev, Diebold, and Vega (2003) and Lahaye, Laurent, and Neely (2011), argue that jumps are responses of prices to news releases. These papers start with a series of news items that they a priori consider important, such as FOMC announcements, and show that prices react effectively instantaneously. However, many different sources cause discontinuities in investor's information sets. Other sources include Congressional decisions, a startup announcing a new product line on Twitter, effectively anything in a Bloomberg or Associated Press feed relevant for asset pricing, even private communications between financiers. The last example highlights the utter impossibility of listing all the potentially relevant events. We cannot construct investors' actual information sets. (Note, this paper uses *news* to refer to any headline news, i.e., it refers to any discontinuous change in information, not just traditional news sources such as

1. The single exception is Christensen, Oomen, and Podolskij (2014), whose procedure I show is biased downwards when jump variation is high in Appendix C.1.

newspapers. It does not refer to information that takes time to digest.)

As these examples illustrate, news often come at unpredictable times, and only a few investors may observe them, and so a priori choosing which news items are relevant necessarily excludes many relevant items. Besides, there is no reason to assume that the resultant price change is in any way substantial. Many news items cause a small, but measurable, impact on prices.

The connection between news and jumps is rather intuitive, and the empirics in these papers substantiate it. However, the connection is even more fundamental. Delbaen and Schachermayer (1994) show no-arbitrage implies prices are semimartingales.[2] Under some minor technical conditions, this implies that prices jump if and only if the information surprises us, (Huang 1985), i.e., the information available at τ cannot be predicted by the information we had before τ.

This result also explains why not all price changes are jumps. Not all information can be processed instantaneously. Some information takes time to process before the market participants can use it effectively. For example, after a firm announces its earnings, the headline reveals much of the information. However, many articles still analyze what each release implies about both the stock in question and other related assets. As various investors update their beliefs and buy or sell accordingly, other market participants see the information that is now revealed by the prices and also buy or sell. This process changes the asset's price, and it takes time. Gürkaynak, Kısacıkoğlu, and Wright (2018) distinguishes these two types of information flows and show that doing so substantially improves forecasting performance.

2. Data Generating Process

Models of prices differ along two different dimensions. They can be either continuous or discrete, and they can be either in continuous-time or in discrete-time. I write down a continuous-time DGP with jumps and derive the implied discrete-time representation. I also discuss the purely continuous special case that my DGP nests to provide a point of comparison.

2.1. Continuous-Time DGP

We know from Dambis (1965) and Dubins and Schwarz (1965) that continuous Itô semimartingales are stochastic volatility diffusions. That is, for some drift, $\mu(t)$, and diffusion volatility, $\sigma^2(t)$, we can represent the log-price process as $dp(t) = \mu(t)\, dt + \sigma(t)\, dW(t)$, where $W(t)$ is a Wiener process.

The standard nonparametric way to add jumps to these models is to assume that prices are Itô semimartingales. This representation is quite general because it only requires that prices are semimartingales and each of the components of the process have time-derivatives. The log-price being an Itô semimartingale implies that the jump part is an integral with respect to a Poisson random measure. Let n be a Poisson random measure with associated compensator, ν. The function $\delta(s, x)$ controls the magnitude of the process. In general, the triple (δ, n, ν) is not unique, which

2. Throughout this paper, I use no-arbitrage to refer to no-free-lunch with vanishing risk as is standard in continuous-time finance.

allows us to pick a particularly useful representation later.

Definition 1. Jump-Diffusion DGP (Grigelionis Form of an Itô Semimartingale)

$$p(t) = p(0) + \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \int_0^t \int_X \delta(s,x) 1\{\|\delta(x,s)\| \leq \mathbf{1}\}(n-\nu)(ds,dx)$$
$$+ \int_0^t \int_X \delta(s,x)\mathbf{1}\{\|\delta(x,s)\| > 1\} n(ds,dx)$$

I simplify Definition 1 by adding the following assumption.

Assumption Square-Integrable. The process, $p(t)$, is locally-square integrable.

Assumption Square-Integrable is relatively innocuous in practice because it holds whenever returns have conditional variances. Although many high-frequency papers initially allow for jumps that are so large no compensator exists, they almost always restrict themselves to processes that satisfy Assumption Square-Integrable when they derive estimators. Making Assumption Square-Integrable now simplifies notation because it implies that the jump measure has a predictable compensator. Assuming without loss of generality that $p(0) = 0$ gives $p(t) = \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \int_0^t \int_X \delta(s,x)(n-\nu)(ds,dx)$. I also assume without loss of generality that n is a standard Poisson random measure.

This representation is quite general and can handle a great variety of different price processes. In particular, it allows for both finite and infinite variation, and also places no restriction on the Blumenthal-Getoor index. In fact, Aït-Sahalia and Jacod (2012b) uses Definition 1 in their study of Blumenthal-Getoor indices.

However, the representation in Definition 1 is rather intractable and not identified. For each τ, $\delta(\tau,\cdot)$ is a function. For each set $A \subset X$, we have a Poisson process. It takes infinitely-many finite-sized open sets to partition $\mathbb{R}$. Each of these infinitely-many sets has a time-varying Poisson intensity. The δ function combines these intensities. To estimate this process, we would have to estimate these infinitely-many intensity parameters for each τ using only one realization.

2.2. Discrete-Time DGP

Before I relate the discrete- and continuous-time returns, we must know what a discrete-time return is. The discrete-time return is just the change in (an increment of) the price process over some length of time, say a day.[3] Throughout, I use subscripts to refer to daily objects, and functional notation to refer to stochastic processes. I index each variable by the time it first becomes known to the investor, i.e., becomes measurable in the filtration induced by the prices. For example, r_t is the daily return on date t, while $p(t)$ is the log-price at time t: $r_t \coloneqq \int_{t-1}^t dp(t)$.

This return has a density — h — in each period given the available information at the end of the day before — $\mathcal{F}_{t-1}$ — giving $r_t \,|\, \mathcal{F}_{t-1} \sim h\,(r_t \,|\, \mathcal{F}_{t-1})$. This predictive density fully characterizes the

3. Throughout, I focus on daily returns whose length I normalize to one, but there is nothing special about a day. We could perform the same analysis over any discrete length of time.

statistical risk that investors face. In particular, any statistical measure of risk, such as Expected Shortfall or Value-at-Risk, is a statistic of this density.

Daily returns are not very well-behaved objects in that they are unpredictable and their distributions vary substantially over time. Furthermore, we only observe one observation for each $h(r_t \,|\, \mathcal{F}_{t-1})$. Since $\mathcal{F}_{t-1}$ grows each day, $h(r_t \,|\, \mathcal{F}_{t-1})$ is a function-valued time-varying parameter. Modeling such parameters is quite difficult. Hence the literature (for example, the ARCH and GARCH models) focuses on representations for $h(r_t \,|\, \mathcal{F}_{t-1})$ in terms of a well-behaved sufficient statistic for the dynamics. The most common choice for x_t is some measure of volatility.

These models use x_t to separate $h(r_t \,|\, \mathcal{F}_{t-1})$ into three parts. The first — x_t — is well-behaved and predictable and hence easily forecastable. The second is noise as far as prediction is concerned and has an associated density — f. It affects the risk investors face but not the density's dynamics. The third part — G — is a process governing x_t's dynamics.

Both f and G are fixed across-time, and G is simple if we chose x_t well. This gives $r_t \,|\, \mathcal{F}_{t-1} \sim h(r_t \,|\, \mathcal{F}_{t-1}) = \int_{x_t} f(r_t \,|\, x_t)\, dG(x_t \,|\, \mathcal{F}_{t-1})$, replacing the question how should we model $h(r_t \,|\, \mathcal{F}_{t-1})$ with three related questions. What should we use for x_t? What should use for f? What should we use for G?

For example, consider the following simple stochastic volatility model: $r_t \sim \sigma_t \mathcal{N}(0,1)$ and $\log\left(\sigma_t^2\right) = \rho \log\left(\sigma_{t-1}^2\right) + \sigma_\sigma \mathcal{N}(0,1)$. This model uses volatility, σ_t^2, as x_t, f is a Gaussian distribution, and σ_t^2 follows an $AR(1)$ process in logs.

Now that we have a discrete-time DGP, we can define the *realized density*.

Definition 2 (Realized Density)**.**

$$RD_t \coloneqq f\left(r_t \,|\, x\right)\Big|_{x=x_t}$$

Just as the realized volatility, RV_t, is the particular value of the volatility that realizes in a given day, the realized density, RD_t, is the conditional density that realizes that day. For example, in the model given above, the realized density is $f(r_t \,|\, x_t) = f(r_t \,|\, \sigma_t^2)$.

The realized density is useful because it separates the dynamic and static parts of the process. Besides, it is precisely the part of the likelihood that high-frequency data identifies, as shown in Section 4. Once we have RD_t, we only need to model G. This is much simpler than modeling $h(r_t \,|\, \mathcal{F}_{t-1})$ directly as long as we chose a well-behaved x_t.

3. Modeling Jump Processes

The literature usually chooses a x_t that is some volatility measure. This section constructs a new volatility measure. This measure, unlike the jump part of the quadratic variation (i.e., the sum of squared jumps), is an ex-ante measure. This distinction is fundamental to the representation constructed below.

3.1. Jump Volatility

One way to define the instantaneous diffusion volatility, $\sigma^2(t)$ is as follows; $\sigma^2(t)$ is the appropriately standardized variance of the diffusion part of the process over a shrinking interval.[4]

Definition 3 (Instantaneous Diffusion Volatility)**.**

$$\sigma_t^2 \coloneqq \frac{1}{\Delta}\mathbb{E}\left[\left|p^D(t+\Delta) - p^D(t)\right|^2 \,\middle|\, \mathcal{F}_{t-}\right]$$

One key subtlety in Definition 3 is that we are only using the information available before t. Variances are forward-looking operators. This subtlety is not essential in the diffusion case. The ex-ante and ex-post measures coincide, and so the literature has not stressed it. In the jump case, however, it is fundamental.

Diffusion volatility's key advantage is that we can time-aggregate it easily. The daily volatility is just the integral of the high-frequency volatility: $\sigma_t^2 \coloneqq \int_{t-1}^t \sigma^2(t)\, ds$. The goal moving forward is to construct a sufficient statistic for the jump dynamics that also has this aggregation property. To do this, I define the *jump volatility* — $\gamma^2(t)$. Volatilities are variances, and so we can construct the jump analog to Definition 3. I substitute the diffusion part of the prices, $p^D(t)$, with the jump part — $p^J(t)$. In other words, the instantaneous jump volatility is the local variance of $p^J(t)$.

Definition 4 (Instantaneous Jump Volatility)**.**

$$\gamma^2(t) \coloneqq \frac{1}{\Delta}\mathbb{E}\left[\left|p^J(t+\Delta) - p^J(t)\right|^2 \,\middle|\, \mathcal{F}_{t-}\right].$$

The integrated jump volatility is defined in the obvious way: $\gamma_t^2 \coloneqq \int_{t-1}^t \gamma^2(s)\, ds$. We can also define $\gamma^2(t)$ in terms of Definition 1. The jump volatility is the time-derivative of the predictable quadratic variation of the jump part of the process.

Remark 1 (Jump Volatility and the Predictable Quadratic Variation)**.** *Let $p(t)$ and $\int_X \delta^2(t,x)\, dx$ be Itô semimartingales satisfying Assumption Square-Integrable. Also assume that $\int_X \delta^2(t,x)\, dx$ is cádlág. Then the following holds where $\langle p^J \rangle(t)$ is the predictable quadratic variation of $p^J(t)$:*

$$\gamma^2(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} \int_t^{t+\Delta} \int_X \delta^2(s,x)\nu(dx,ds) = \lim_{\Delta \to 0} \frac{1}{\Delta}(\langle p^J \rangle(t+\Delta) - \langle p^J \rangle(t)).$$

The are three main advantages of γ_t^2 over the jump part of the quadratic variation. First, since jump processes are not absolutely continuous, there is no ex-post analog to $\gamma^2(t)$. We cannot take the quadratic variation's time derivative like we can take the predictable quadratic variation's time-derivative. Second, by conditioning on $\gamma^2(t)$, I construct a closed-form nonparametric continuous-time representation for $p(t)$ in Section 3.3. This representation avoids any truncation. Todorov and Tauchen (2014) must truncate all of the jumps above a shrinking threshold to derive their results

4. I use superscript D to refer to the diffusion part of the process.

while using an ex-post measure. As a final advantage, high-frequency data nonparametrically identify both $\gamma^2(t)$ and γ_t^2. Section 4 shows this by constructing consistent estimators for them.

3.2. Static Jump Processes (Variance-Gamma Process)

The next section constructs the static model that my model reduces to when there are no dynamics. Define $N(t)$ as the process that determines when $p^J(t)$ jumps:

$$N(t) \coloneqq \sum_{s \leq t} \mathbf{1}\{\left|p^J(t) - p^J(t-)\right| > 0\}. \tag{1}$$

Let $\kappa(t) \coloneqq \{p^D(t) \,|\, N(t) \neq N(t-)\}$ be the process that controls the jump magnitudes, i.e, $\kappa(t)$ is an ordered collection of $\mathcal{N}(0,1)$ random variables. Then the jump part of the prices process has the form: $p^J(t) = \sum_{s \leq t} \kappa(s) \Delta N(s)$.

This equation's variability comes from two places: the number of jumps and their magnitudes. Since we are in a time series context, the number of jumps and their locations carry the same information. Hence, we can rewrite the jump volatility as follows using the law of iterated expectations (where we assume $N(\tau) = 0$):

$$\gamma^2(\tau) = \lim_{\Delta \to 0} \mathbb{E}\left[\frac{\left|p_{\tau+\Delta} - p_\tau\right|^2}{\Delta} \,\middle|\, \mathcal{F}_\tau \right] = \lim_{\Delta \to 0} \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N(\tau+\Delta)} \frac{\mathbb{V}\mathrm{ar}\left(\kappa_i(\tau)\right)}{\Delta} \,\middle|\, \mathcal{F}_\tau, N\left(\tau+\Delta\right) \right] \,\middle|\, \mathcal{F}_\tau \right]. \tag{2}$$

We we can simplify this using the definition of $N(t)$: $\gamma^2(\tau) = \lim_{\Delta \to 0} \mathbb{E}[\frac{N(\tau+\Delta)}{\Delta}]\mathbb{E}[\kappa(\tau)^2] = \frac{\Delta}{\Delta} = 1$. That is the variance of the jump process is the intensity multiplied by the magnitude's variance. Changing the jump intensity or expected magnitude alters the variance of $p^J(t)$ in precisely the same way. This irrelevance is useful because the data do not identify the intensity and magnitude functions but do identify the volatility. This lack of identification no longer affects our results if we take $\mathbb{E}[N(t)] \to \infty$. In taking this limit, we must model the distribution of $\kappa(t)$ properly so that $p(t)$ remains square-integrable.[5] In particular, only finitely many jumps can exceed any fixed $\epsilon > 0$ in magnitude; otherwise, the price diverges. Consequently, we must shrink the size of the increments towards zero as we let $\mathbb{E}[N(t)] \to \infty$.

3.3. Jump Process Representation Theorem

3.3.1. Itô Semimartingales

Recall the simplified Grigelionis form of the semimartingale: $p(t) = \int_0^t \mu(s)\, ds + \int_0^t \sigma(s)\, dW(s) + \int_0^t \int_X \delta(s,x)(n-\nu)(ds,dx)$. I use the variance-gamma processes and the jump volatility discussed above to simplify the representation for the jump part of the process. To do this, I introduce some empirically-innocuous assumptions that are not entirely standard in the literature. First, $p(t)$ must have infinite-activity jumps. In other words, we need at least one jump in every finite interval. This

5. Just letting $p(t)$ be an ordered collection of $\mathcal{N}(0,1)$ variables does not work.

assumption implies two results. First, we do not need to track the probability that there are no jumps in a specific interval. Second, it identifies $\gamma^2(t)$. If we consider an interval without jumps, we obviously cannot estimate $\gamma^2(t)$ because there is no variation to identify it.

Assumption Infinite-Activity Jumps. The $p(t)$ process has infinite-activity jumps.

Assumption Infinite-Activity Jumps sounds very restrictive at first and contradicts the compound Poisson assumption often used in the literature. However, it is innocuous for two reasons. First, the literature is essentially unanimous in arguing that jumps are quite common in the data, as discussed in Section 1. Second, standard variance-gamma processes are limits of compound Poisson processes. As long as we have a sufficient number of jumps, the representation will work well in practice. Section 3.3.3 further discusses this.

The last assumption requires jump times to be unpredictable.

Assumption No Predictable Jumps. No stopping times τ exists such that the event $p(\tau) \neq p(\tau-)$ is contained in the information set $\mathcal{F}_{\tau-}$.

Having laid out the assumptions, I state the first main theorem. I later prove a more general proposition, Theorem 2. However, I have now described the environment sufficiently to make the result understandable. Stating the result now should make it easier to understand the structure of the next section. Throughout I will use $\mathcal{L}(t)$ to refer to the *standard* variance-gamma process, i.e., a variance-gamma process where all of the scale parameters equal one and all location parameters equal zero. This standard variance-gamma process is a Wiener process time-changed by an exponential process, making it the process generalization of the Laplace distribution.[6] Its increments are Laplace random variables.

Theorem 1 (Locally Square-Integrable Itô Semimartingales as Integrals)**.** *Let $p(t)$ be an Itô semimartingale with interval support satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Predictable Jumps. Then we can represent $p(t)$ as*

$$p(t) = \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \frac{1}{\sqrt{2}}\int_0^t \gamma(s)\,d\mathcal{L}(s).$$

This representation replaces the function $\delta(\tau,\cdot)$, with a single scalar $\gamma^2(\tau)$ for each τ. In addition the integrator is switched from a compensated Poisson random measure, $(n-\nu)$, to a standard variance-gamma process, $\mathcal{L}(t)$. As in the stochastic-volatility diffusion case, Theorem 1 does not imply that the innovations to $\gamma(s)$ and $p(t)$ are independent.[7] Hence, this representation can exhibit properties such as skewness and infinite variation that the variance-gamma process does not exhibit.

6. The Laplace distribution can be represented as a normal random variable with an exponentially-distributed variance.

7. Consider the geometric Brownian motion, $dX(t) = X(t-)\,dW(s)$, the innovations to the volatility and the process are the same.

3.3.2. Time-Change Representation

The proof of Theorem 1 relies on Theorem 2 and its corollaries, which I have not yet proven. In practice, Theorem 2 is the fundamental result. The other results are straightforward implications of it. Theorem 2 is a time-change representation for jump processes and is closely related to the time-change representations in the diffusion case. Consequently, let us recall those results.

The validity of the standard diffusion representation for general continuous martingales is implied by the Dambis-Dubins-Schwarz theorem, which shows that any continuous martingale time-changed by its predictable quadratic variation is a Wiener process. To put it in mathematical notation, Dambis (1965) and Dubins and Schwarz (1965) says that $p^D(t) \overset{\mathcal{L}}{=} W(\langle p^D \rangle(t))$, where the equals sign with an $\mathcal{L}$ above it refers to equality in law. This equation evaluates the Wiener process at the random-clock $\langle p^D(t) \rangle$.

The crucial difference between the jump part and the continuous part of a semimartingale is that the variation in the continuous part comes from variation in magnitudes, while the jump part has two sources of variation: the magnitudes and the locations. Intuitively, the Dambis-Dubins-Schwarz theorem separates the variation in any continuous martingale into a predictable part (the volatility) and i.i.d. innovations. By doing this, the martingale becomes a sum of appropriately scaled independent random variables. In other words, it is a "central limit theorem."[8] In fact, one method to prove static central limit theorems is deriving them from this result.

In the jump case, though, the dynamics are more complicated. Not only do the magnitudes vary, but the locations also vary. This is why the data do not identify existing time-change representations for jumps, such as those in (Monroe 1972; Geman, Madan, and Yor 2001; Barndorff-Nielsen and Shiryaev 2010). When we take the infill asymptotics, variation in both the jump sizes and locations still matters in the limit. In other words, a jump martingale is a sum of a random number of random summands. If the number of summands is geometrically-distributed, various geometric-stable central limit theorems tell us how the sum behaves as the expected number of summands approaches infinity, (Mittnik and Svetlozar 1993).

We can generalize the Itô semimartingale assumption in Theorem 1 by only requiring that $p^J(t)$ is an integral with respect to a Poisson random measure. In particular, $p(t)$'s characteristics do not need to have time-derivatives.

Theorem 2 (Time-Changing Jump Martingales)**.** *Let $p^J(t)$ be a purely discontinuous martingale with interval support satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Predictable Jumps that can be represented as $H * (n - \nu)$ where $H(t)$ is a predictable process, n a Poisson random measure, and ν its predictable compensator with Lebesgue base Lévy measure.*

Then $p^J(t)$ time-changed by its predictable quadratic variation is a standard variance-gamma process. In other words, $p^J(t) \overset{\mathcal{L}}{=} \mathcal{L}\left(\langle p^J \rangle(t)\right)$.[9]

8. Technically, this result is a law of large numbers in the diffusion case, not a central limit theorem because the convergence here is almost sure instead of in law.

9. Note, equality here only holds in law unlike in the Dambis-Dubins-Schwarz theorem, where it holds almost surely.

The proof of this theorem is in Appendix A. I present the intuition here. The first result we must establish is that the jump locations and magnitudes are conditionally independent. Thankfully, the Poisson random measure representation implies that the location and magnitude risk are independent given $\mathcal{F}_{t-}$.

Given some time τ I condition on the first jump after τ. I then show that the magnitudes are a continuous process in this new filtration after the appropriate time-change. Thus, I can apply the Dambis-Dubins-Schwarz theorem there, which results in a time-changed Wiener process. Since a Poisson random measure is generated by Poisson processes and the times between jumps for a Poisson process are exponential random variables, by keeping careful track of how the exponential time-changes aggregate, we get that the time-change coming from the locations is a standard gamma process. The quadratic variation of $p(t)$ combines the quadratic variations arising from each of two time-changes. Therefore, the initial process is a time-changed standard variance-gamma process.

Perhaps the most surprising part of Theorem 2 is that it implies we can "borrow-strength" from the small-jumps to learn about the large jumps without making any parametric assumptions. At first glance, this likely appears difficult to believe. This implies that local square-integrability is a stronger assumption than we might first expect. The Davis-Burkholder-Gundy inequality lets us bound the supremum of the deviations semimartingale by the 2nd-moment. That is, we can control all moments of the process if the 2nd moment is bounded. If all the moments between two representations, the distributions will as well.

Analogously, the standard central limit theorem and the Dambis-Dubins-Schwarz theorem are examples square-integrability gives us parametric results in the continuous case, and the geometric central limit theorems are examples where we get parametric results in the static analogue to a jump process. Perhaps, it should not be too surprising that we can generalize those results to the stochastic process case, just like we can generalize the standard central limit theorem.

Time-changed results are not particularly intuitive, and so we would like an integral representation as well. Corollary 2.1 is the jump analogue to stochastic volatility diffusion representations for continuous martingales.

Corollary 2.1 (Jumps Processes as Integrals)**.** *Let $p^J(t)$ be a purely discontinuous Itô martingale with interval support satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Predictable Jumps. Then $p^J(t) = \frac{1}{\sqrt{2}} \int_0^t \gamma(s)\, d\mathcal{L}(s)$, where $\mathcal{L}$ is a standard variance-gamma process.*

3.3.3. Processes with Finite-Activity Jumps

The most controversial assumption I make is Assumption Infinite-Activity Jumps. Various authors have claimed that the data have a large, but finite, number of jumps in each period. One might wonder what happens if the data violate this assumption. In that case, given an interval I, the price process is a point mass at zero if it does not jump. If the price does jump, we can use similar arguments to those above to construct a time-changed Wiener process. In other words, the ex-ante distribution over each interval is a mixture of a point mass at zero and a scale Gaussian mixture where the mixing weights between the two parts are the probability of the jump in that interval.

Theorem 3 (Time-Changing Finite-Activity Jump Martingales)**.** *Let $p^J(t)$ be a compound Poisson martingale with interval support and mean-zero jump magnitudes.*[10] *Let ν denote the intensity process, and $\rho(t)$ denote the time of the next jump. Then*

$$p^J(t) = \begin{cases} \mathcal{W}\left(\langle p^J \mid \rho \rangle(t)\right) & \text{with intensity } \nu \\ \delta_0(t) & \text{with intensity } 1 - \nu. \end{cases} \tag{3}$$

The critical difference between Theorem 2 and Theorem 3 is that we must keep track of the probability of no jump realizing. Unlike the infinite-activity case, the data do not identify time-change, and two states govern the process, not just one. The dependence of $\langle p^J \mid \rho \rangle$ on ρ disappears in the infinite-activity case because for any interval I the probability of a jump occurring equals one. Additionally, we need not directly assume that the magnitudes are mean-zero. We only require the initial process to be a martingale because we need not worry about the process's mean in intervals without jumps in the infinite activity case.

The main benefit of Theorem 3 is that it implies that Theorem 2 is the limiting case of a finite-activity process as the intensity approaches infinity (which is apparent in the proof of the infinite-activity case). This result is useful because it implies that the representation in Theorem 1 approximates the true DGP well if the intensity is relatively large.

3.4. Deriving the Realized Density

Having derived the continuous-time representation, we can solve the time-aggregation problem and derive the realized density. Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) concurrently derived the realized density in the diffusion case; although, they did not use that term. They showed that if volatility and prices are uncorrelated, σ_t^2 is a sufficient statistic for the dynamics under some technical conditions and that conditional on σ_t^2, the return density is Gaussian.

This conditional Gaussianity separates the daily return distribution into a well-behaved volatility component and a Gaussian noise component. To relate it to the previous discussion, we have the following decomposition for $h(r_t \mid \mathcal{F}_{t-1})$ if the price is a martingale:

$$f\left(r_t \mid \sigma_t^2\right) = f\Big|_{x_t = \left[\int_{t-1}^t \sigma^2(s)\, ds\right]} = \mathcal{N}\left(0, \int_{t-1}^t \sigma^2(s)\, ds\right). \tag{4}$$

I now discuss the realized density in the case with jumps. Recall that $dp(t) = \sigma(t)\, dW(t) + \int_X \delta(t,x)(n-\nu)(dx, dt)$. Conditional on $\sigma^2(t)$ and $\delta(t, \cdot)$, the jumps and diffusion parts are independent. Consequently, returns are the sum of two conditionally independent components. Densities of sums of independent components are convolutions of the summands' densities. We know, as discussed above, that the diffusion part is a Gaussian density whose variance equals the integrated

10. Compound Poisson processes do not have any predictable jumps and are locally-square integrable.

diffusion volatility. Hence, we only need to develop a parametric expression for the jump part.

Let $\mathcal{L}(0, x)$ refer to the Laplace density with mean zero and variance x and recall that $*$ is the standard convolution symbol. Then we have the following discrete-time representation.

Theorem 4 (Realized Density Representation)**.** *Let $p(t)$ be an Itô semimartingale with interval support satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Predictable Jumps. Let $\sigma^2(t)$ and $\gamma^2(t)$ be semimartingales whose martingale components are independent of the martingale component of $p(t)$. Then*

$$RD_t = \mathcal{N}\left(\int_{t-1}^{t} \mu(s)\, ds, \int_{t-1}^{t} \sigma^2(s)\, ds\right) * \mathcal{L}\left(0, \int_{t-1}^{t} \gamma^2(s)\, ds\right), \tag{5}$$

and the predictive density is

$$h\left(r_t \,\middle|\, \mathcal{F}_{t-1}\right) = \int_{\mu_t, \sigma_t^2, \gamma_t^2} RD_t\left(r_t \,\middle|\, \mu_t, \sigma_t^2, \gamma_t^2\right) dG\left(\mu_t, \sigma_t^2, \gamma_t^2 \,\middle|\, \mathcal{F}_{t-1}\right). \tag{6}$$

The intuition behind Theorem 4 is the following. If $\gamma^2(t)$ was constant, we could pull it out of the integral without affecting the distribution: $\int_{t-1}^{t} \gamma(t-1)\, d\mathcal{L}(s) \overset{\mathcal{L}}{=} \frac{\gamma_{t-1}}{\sqrt{2}} \int_{t-1}^{t} d\mathcal{L}(s)$. Since increments of the standard variance-gamma process are Laplace distributed, the second component is distributed $\mathcal{L}(0, 1)$. Consequently, conditionally on γ_{t-1}^2, we have a Laplace distribution with the specified variance. The $\sqrt{2}$ term arises because the scale of a Laplace distribution is the square root of one-half the variance. We can replace the constant assumption on the volatilities with independence conditions between the martingale components to recover the general case.

Integrating RD_t out using its distribution G recovers (6) from (5). In practice, we likely want to model G directly. This model has the same form as the various stochastic volatility / GARCH type models in the diffusion case.

The primary assumption that Theorem 4 adds is the independence between the martingale components. As in Andersen, Bollerslev, Diebold, and Labys (2003) We need this assumption to time-aggregate the process because the marginal and conditional distributions given the volatilities of $p(t)$ must coincide. Importantly, this assumption restricts the leverage effect but does not assume away all dependence. The volatilities and drift can be arbitrarily related. As long as it takes a positive amount of time for the volatilities to affect the level of the prices or vice-versa, this assumption is satisfied. Besides, the observed correlation between the martingale parts is close to zero at high frequency as noticed by Aït-Sahalia, Fan, and Li (2013), who call it "the leverage effect puzzle."[11]

11. There is some evidence that this is an artifact of the estimation procedure, and so I leave to future work the optimal way to bring it into my framework. One way to do this is by keeping track of this correlation (as Bandi and Renò (2012) and Kalnina and Xiu (2017) do) and making Gaussian and Laplacian conditioning arguments.

4. Estimation

This section constructs estimators for $\sigma^2(t)$ and $\gamma^2(t)$ and their daily analogs. As is standard, the data do not identify the drift, and so we cannot estimate it. The proposed estimator for $\sigma^2(t)$ is adapted from Jacod and Rosenbaum (2013). I show that their estimator is still valid under my slightly more general assumptions. The estimator for $\gamma^2(t)$ is completely new. In particular, I develop a consistent estimator for $\gamma^2(\tau)$ for any fixed τ.[12] This estimator is the first to consistently estimate any instantaneous jump variation measure. This implies that high-frequency data nonparametrically identify instantaneous jump dynamics.

4.1. Assumptions

I start by fixing notation and stating some assumptions. The instantaneous volatility estimators take an appropriately defined average over an increasing number of increments over a shrinking interval. In other words, for a given index — n, we have a triangular array of increments. To make the notation even more complicated, we have both a true DGP with time-varying volatility and an approximate DGP, whose volatility is locally constant.

This setup implies we must keep track of both triangular arrays as we take limits. I adopt the notation used in Jacod and Protter (2012) for the most part. Specifically, I use $\Delta_i^n p$ to refer to a increment i in process $p(t)$ of length Δ^n, and I take limits with respect to n, that is $\{\Delta_i^n p\}$ is a triangular array of increments of $p(t)$. The assumptions used are very similar to the standard ones used in the literature. Assumption HL is essentially Jacod and Protter's (2012) Assumption H. The assumption on the jumps is slightly more general and more straightforward.

Assumption HL. 1. $\mu(t)$ is locally bounded.

2. $\sigma(t)$ is càdlàg (or càglàd).

3. $\gamma(t)$ is càdlàg (or càglàd).

To reduce notation, I adopt the convention from the literature that $i \in \mathbb{Z}, i \leq 0 \implies \Delta_i^n p = 0$ to make processes well-defined over the entire line, not just where we estimate them. This convention sets the processes equal to zero outside of the relevant window.

4.2. Instantaneous Volatility Estimators

In this section, I show that the diffusion volatility that I uses is consistent under the same assumptions I use elsewhere in the paper. The contribution here is admittedly small. Results under very similar assumptions already exist in the literature. I include it here for rhetorical consistency; having a single set of assumptions is useful.

The intuition behind the estimators' convergence is that we are averaging the volatilities over shrinking intervals that approach τ from the left so that we estimate the volatilizes' left limit. Let

12. In general, much of the theory that I develop can likely be extended to stopping times, but I leave that for future work.

k_n denote to the number of terms we averaging, $I(i,n) \coloneqq [(i-k_n-1)\Delta^n, (i-1)\Delta^n]$, and $\Delta^n_{i_n} p$ denote the change in p in $I(i,n)$. If we choose a sequence $i \to \tau$, the interval approaches τ from the left. Also, as $p(t)$ is one-dimensional, the driving Wiener and variance-gamma processes can be assumed to be one-dimensional without loss of generality.

Theorem 5 (Estimating the Instantaneous Diffusion Volatility)**.** *Let $p(t)$ be an Itô semimartingale with interval support satisfying Assumptions HL, Infinite-Activity Jumps, and Square-Integrable. Let k_n, Δ^n satisfy $k_n \to \infty$ and $k_n\sqrt{\Delta_n} \to 0$, and let $0 < \tau < \infty$ be a deterministic time. Define $i_n = i - k_n - 1$. Let $c_1(\Delta^n)^{1/4} < v_1^n < c_2\sqrt{\Delta^n}$ for some constants c_1, c_2 and $v_2^n \to 1$. Then*

$$\widehat{\sigma}^2_{i_n}(k_n, \tau-, p) \coloneqq \frac{1}{k_n \Delta^n} \sum_{m=0}^{k_n-1} v_2^n \left|\Delta^n_{i_n} p\right|^2 1\{\left|\Delta^n_{i_n} p\right| \leq v_1^n\} \xrightarrow{\mathbb{P}} \sigma^2(\tau-).$$

One might think we could use an analogous strategy to estimate $\gamma^2(t)$, i.e., form an estimator of $\left\langle p^J \right\rangle(t)$ by truncating away the small increments and take the time derivative of the resulting object. In fact, Jacod and Protter (2012, 256) show that this estimator converges to zero in their proof of the validity of their estimator for $\sigma^2(t)$. By considering a specific time τ, we implicitly condition on τ. Doing this reduces the variation in the locations, and shrinking the window eliminates variation from large jumps. If we also truncate away variation arising from the small jumps, we have no variation left to identify the jump volatility with.

Over a fixed interval, the quadratic variation of jump processes and diffusive processes are of the same asymptotic order as we shrink $\Delta^n_{i_n}$, (Jacod, Podolskij, and Vetter 2010). If we consider shrinking intervals, this is no longer the case. Instead, it is the absolute value of the stochastic volatility variance-gamma and diffusion processes that have similar asymptotic properties. The absolute value of a standard variance-gamma process, $|\mathcal{L}|(t)$, is a well-behaved object, just like the absolute value of a Wiener process, $|W|(t)$, and they vanish at the same asymptotic rate: $\sqrt{\Delta}$.[13] Consequently, the $\lim_{\Delta \to 0}\left|\Delta^n_{i_n} p(t)\right|$ contains both $\gamma^2(\tau)$ and $\sigma^2(\tau)$.[14]

Proposition 1 (Estimating the Instantaneous Absolute Volatility)**.** *Let $p(t)$ be an Itô semimartingale with interval support satisfying Assumptions HL, Infinite-Activity Jumps, and Square-Integrable. Let k_n, Δ^n satisfy $k_n \to \infty$ and $k_n\sqrt{\Delta^n} \to 0$, and let $0 < \tau < \infty$ be a deterministic time. Define $i_n \coloneqq i - k_n - 1$.*

Then the following holds, where $\operatorname{erfcx} \coloneqq \frac{2\exp(x^2)}{\sqrt{\pi}} \int_x^\infty \exp(-s^2)\, ds$*:*[15]

$$\frac{1}{k_n\sqrt{\Delta^n}} \sum_{m=0}^{k_n-1} |\Delta^n_{i_n+m} p| \xrightarrow{\mathbb{P}} \mathbb{E}|\mathcal{N}(0,1)|\sigma(\tau-) + \frac{\gamma(\tau-)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right).$$

As long as the volatilities are locally constant around τ, the implied parametric form gives a

13. Note, neither of the processes is a martingale. They are semimartingales.

14. It is worth noting that this estimator is for the instantaneous absolute value of the martingale part, it will not pick up non-martingale parts of the jump process.

15. This function, erfcx, is the scaled complementary error function. It is a reparameterization of Mill's ratio. Most scientific programming suites provide efficient, numerically-stable implementations.

limiting value for the absolute value as a function of $\sigma^2(\tau)$ and $\gamma^2(\tau)$. The expression on the right of the equation in Theorem 6 is the mean of the convolution of $|\mathcal{N}\left(0, \sigma^2(\tau-)\right)|$ and $|\mathcal{L}\left(0, \gamma^2(\tau-)\right)|$.

We combine this convolution and $\sigma^2(\tau)$ to estimate $\gamma^2(\tau)$. To do this, we must weight the difference between the absolute population moment as a function of $\gamma(\tau)$ and the absolute sample moment.

Theorem 6 (Estimating the Instantaneous Jump Volatility)**.** *Let $p(t)$ be an Itô semimartingale with interval support satisfying Assumptions HL, Infinite-Activity Jumps, and Square-Integrable. Let k_n, Δ^n satisfy $k_n \to \infty$ and $k_n\sqrt{\Delta^n} \to 0$, and let $0 < \tau < \infty$ be a deterministic time. Define $i_n = i - k_n - 1$. Let $\widehat{\sigma}_n(\tau-)$ converge in probability to $\sigma(\tau-)$. Let $\gamma(\tau) > 0$ and g be strictly-increasing, convex, and continuous, then the following holds:*

$$\widehat{\gamma}(k_n, \tau-, p) := \operatorname*{argmin}_{\gamma} g\left(\left|\frac{1}{k_n\sqrt{\Delta^n}}\sum_{m=0}^{k_n-1}\left|\Delta^n_{i_n+m}p\right| - \mathbb{E}|\mathcal{N}(0,1)|\widehat{\sigma}_n(\tau-) - \frac{\gamma\operatorname{erfcx}\left(\frac{\widehat{\sigma}_n(\tau-)}{\gamma}\right)}{2}\right|\right)$$
$$\xrightarrow{\mathbb{P}} \gamma(\tau-).$$

4.3. Implementation

We now have estimators for both the instantaneous and integrated volatilities. The difficult part is estimating the instantaneous volatilities. The integrated volatilities are their averages. In practice, two issues affect the analysis. First, we must remove market microstructure noise. To do this, I adopt the pre-averaging approach argued for in Podolskij and Vetter (2009, Eqn. (3.9)).[16] Define the function: $g(x) := (1-(2x-1)^2\mathbf{1}\{x >= 0\}\mathbf{1}\{x <= 1\}$. The pre-averaged data is the rolling average of the true data: $\bar{p}_{i_n} := \frac{1}{\kappa_n\sqrt{\int_0^1 g^2(s)\,ds}}\sum_{m=1}^{\kappa_n-1} g(\frac{m-1}{\kappa_n})\Delta^n_{i_n+m}p$. The g function corrects for the error introduced by the pre-averaging.

If $\kappa_n \propto 1/\sqrt{\Delta^n}$, we likely achieve the optimal rate in the presence of noise, but the noise leads to an asymptotic bias in most cases, (Jacod, Podolskij, and Vetter 2010). To avoid this, I set $\kappa_n = \lfloor\frac{\theta}{(\Delta^n)^{0.55}}\rfloor$. This rate is useful because we can apply the estimators directly to the pre-averaged data, and it is not obvious exactly what bias exists when estimating the instantaneous absolute variation.[17] I set $\theta = 0.5$, which both is recommend by Hautsch and Podolskij (2013) and works well in my simulations.

I apply Theorem 5 to estimate $\sigma^2(\tau-)$. To do this, we must choose v_2^n to converge to 1; I let $v_2^n = 1$. More importantly, I must choose the truncation threshold v_1^n. We need v_1^n to asymptotically upper bound the diffusion part. In the literature, papers usually set $v_1^n = c\widetilde{\sigma}(\tau-)\Delta_n^{0.49}$, where $\widetilde{\sigma}(\tau-)$ is a preliminary estimator for σ and c is a number of standard deviations chosen by the econometrician. Since the tails of Laplace and Gaussian variables both decline rapidly, solving this deconvolution problem is quite difficult, requiring a tight bound. The law of the iterated logarithm tightly bounds the deviations of a Gaussian variable, and so I use $v_1^n = \sqrt{2}\widetilde{\sigma}(\tau-)\sqrt{\Delta_n}\sqrt{\log(\log(1/\Delta_n))}$.

16. Proving my estimators are still consistent in the pre-averaging case is an interesting avenue for future research.
17. The transformation creating $\bar{p}_{i_n}$ does not affect the volatilities but does affect the mean.

The tails of the Laplace and Gaussian random variables both decline rapidly. The Gaussian density's tails are proportional to $\exp(-x^2/2)$, while the Laplace density's tails are proportional to $\exp(-x/\sqrt{2})$. Distinguishing these two is quite difficult in practice. Setting $v_1^n \propto \Delta^{.49}$, does not work particularly well in this scenario as I show in Section 5. This bound, although valid, is not particularly tight, and that matters when there are a lot of jumps.

To form a preliminary estimator, I start with the 1.25 times bipower variation and then iterate until convergence. We must start by overestimating the volatility to avoid incorrectly setting $\widehat{\sigma}(\tau-) = 0$ because that truncates away all the increments. I choose k_n, which controls the length of the interval over which the volatilities are treated as approximately constant, to equal $\bar{\kappa} + (\Delta^n)^{1/4}$ because that works well in simulations with market microstructure.

Now that we can estimate $\sigma^2(\tau-)$, we need an estimator for the local absolute value. I plug the pre-averaged data into Theorem 6. It is worth noting that the theory I develop is for the no-noise case; the particular implementation likely is not affected by the noise, but I have not proven that. An appealing avenue for future research would be extending these results to cover the noise case.

4.4. Integrated Volatilities

We want to estimate discrete increments of the volatilities. To do this, I average the instantaneous estimators each day. The diffusion estimator defined this way coincides with standard diffusion estimators in the literature up to edge effects.

Remark 2 (Consistency of the Integrated Estimators)**.** *Let $p(t)$ be an Itô semimartingale with interval support satisfying Assumptions HL, Infinite-Activity Jumps, and Square-Integrable. Let k_n, Δ^n satisfy $k_n \to \infty$ and $k_n\sqrt{\Delta_n} \to 0$. Define $i_n = i - k_n - 1$. Then*

$$\widehat{\sigma}_t^2 \coloneqq \frac{1}{\# t_n \in [t-1, t]} \sum_{t-1 < t_n \leq t} \widehat{\sigma}^2(k_n, t_n, p) \xrightarrow{\mathbb{P}} \int_{t-1}^{t} \sigma^2(s)\, ds, \tag{7}$$

and

$$\widehat{\gamma}_t^2 \coloneqq \frac{1}{\# t_n \in [t-1, t]} \sum_{t-1 < t_n \leq t} \widehat{\gamma}^2(k_n, t_n, p) \xrightarrow{\mathbb{P}} \int_{t-1}^{t} \gamma^2(s)\, ds. \tag{8}$$

Proof. The integrated estimators are averages of consistent estimators of $\sigma^2(t)$ and $\gamma^2(t)$. Averages of consistent estimators are consistent by the law of iterated expectations, Jensen's inequality applied to the square, and Chebyshev's theorem. □

Implementing the discrete volatility estimators is straightforward; we can take daily averages of the instantaneous volatilities. To estimate the realized density, plug the daily estimates into (5). Since this function is uniformly continuous given a lower bound on the volatilities, the resulting estimators should work well. Conversely, we cannot estimate the dynamics of the volatilities without time-series variation.

5. Simulations

One of my representation's key advantages is that we can simulate from it easily whenever we can simulate the instantaneous volatilities. Perhaps the most commonly used model for the diffusion volatility is the Cox-Ingersoll-Ross (CIR) process, also known as the square-root process. (A diffusion model whose volatility follows a CIR process is the Heston model.) It has the following form $dx(t) = \kappa(\theta - (x(t)) + \omega\sqrt{x(t)}\, dW(t)$, where θ is the asymptotic mean, κ is the mean-reversion rate, and ω is a scale parameter.

One nice feature of this model is that the volatility itself has volatility, but we only need to simulate one process. The qualitative features of the jump and diffusion volatilities are quite similar in practice, and so I adopt this model for the jump volatility.

5.1. Simulation Data Generating Process

I simulate a CIR process for both $\gamma^2(t)$ and $\sigma^2(t)$ using the full-truncation scheme of Lord, Koekkoek, and Van Dijk (2010). The parameters are given in Table 1. I chose the specific parameter values displayed below to match the discrete-time dynamics of the data.

Table 1: Volatility Parameters

	θ	κ	ω	Asymptotic Standard Deviation
$\sigma^2(t)$	5.00×10^{-5}	1	2.10×10^{-3}	4.60×10^{-4}
$\gamma^2(t)$	5.00×10^{-5}	1	2.10×10^{-3}	4.60×10^{-4}

Once I obtain $\sigma^2(t)$ and $\gamma^2(t)$, I plug them into $dp(t) = \sigma(t)\, dW(t) + \frac{\gamma(t)}{\sqrt{2}}\, d\mathcal{L}(t)$. This gives me a sequence of prices, which I use to estimate the volatilities.

5.2. Simulation Results

This section focuses on the daily volatility results as they form sufficient statistics for all of the daily objects, which the applications study. This section also reports results for the truncation-based estimator used by Li, Todorov, and Tauchen (2017), (LTT); the bipower estimator of Barndorff-Nielsen and Shephard (2004) and Podolskij and Vetter (2009), (Bipower); and bipower estimators computed on 5-minute data (5 Minute) to provide a point of comparison. In the jump case, these estimators do not converge to γ_t^2 but rather to the jumps' quadratic variation. However, since γ_t^2 is the jumps' predictable quadratic variation, these estimators should still be asymptotically unbiased for γ_t^2.

I first estimate the model using the estimation procedure described in Section 4 without the microstructure correction described there. I report averages over 250 days. It is worth noting that the daily estimates are independent, i.e., I do not smooth across days.

The proposed estimators, however, perform quite well at this frequency outperforming the other estimators by approximately an order of magnitude. Table 2 reports the average root mean square

Table 2: Relative Simulation Error without Microstructure

Obs. per Min.	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$				$\frac{\mathbb{E}[(\widehat{\gamma}_t-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$			
	BNS	LTT	5 Min.	Proposed	BNS	LTT	5 Min.	Proposed
≈ 2	0.37	0.40	0.40	0.46	0.72	1.01	0.80	0.72
≈ 12	0.38	0.40	0.42	0.16	0.70	1.01	0.83	0.21
≈ 60	0.40	0.43	0.45	0.05	0.68	1.01	0.87	0.07
≈ 180	0.39	0.41	0.43	0.07	0.69	1.01	0.85	0.07

errors of 250 days worth of various estimators.

Appendix C shows that the results are robust. Table 7 reports a simulation with microstructure noise, which is adapted from Christensen, Oomen, and Podolskij (2014). The results are similar. The previous methods perform better when estimating σ_t^2, but the proposed method still substantially outperforms in estimating γ_t^2. I also simulate a process where we have Poisson jumps and an average of one jump per day. Even though this dramatically violates the infinite-activity assumption, the proposed method still works well (Table 8).

6. Data

The methods developed in this paper require high-frequency data. For the analysis to be interesting, we need a dataset that faces a dense stream of relevant news. I chose SPY (SPDR S&P 500 ETF Trust) which is an exchange-traded fund that mimics the S&P 500. I obtain it from the Trade and Quotes (TAQ) database at Wharton Research Data Services (WRDS).

Since this paper only use one asset, and SPY is one of the most liquid assets traded, we can essentially choose the frequency at which we want to observe the underlying price. In order to balance market-microstructure noise, computational cost, and efficiency of the resulting estimators, I sample at the 1-second frequency. The data used starts in 2003 and ends in September 2017. Since the asset is only traded during business hours, this leads to 3713 days of data with an average of $\approx 24\,000$ observations per day. The dataset takes up about 4.4 GiB of memory. It is also worth noting that SPY is by far the most liquid exchange-traded fund, especially in recent years, reducing the effect of market microstructure, such as bid-ask spreads, bounces, and rounding error.

7. Volatility: Empirics

I separate this empirical part into two subsections. The first section characterizes the static properties of the volatilities, and the second characterizes their dynamic properties. In particular, it shows that both volatilities are highly persistent, displaying long-memory.

7.1. Statics

The results concerning σ_t^2 are broadly consistent with previous work on the topic. Since this paper introduces γ_t^2, the stylized facts regarding its features are new. Thankfully, in practice, the volatilities have very similar dynamics, and so much of the intuition regarding σ_t^2 can be directly translated to γ_t^2.

As can be seen in Figure 1a, the volatilities are very closely related; their correlation equals 0.93. They both significantly increase during crises. Interestingly, σ_t^2 spiked more than γ_t^2 did during the Financial Crisis and seems to spike more during other recessions as well.

Figure 1: Volatilities

(a) Time Series (b) Log Densities

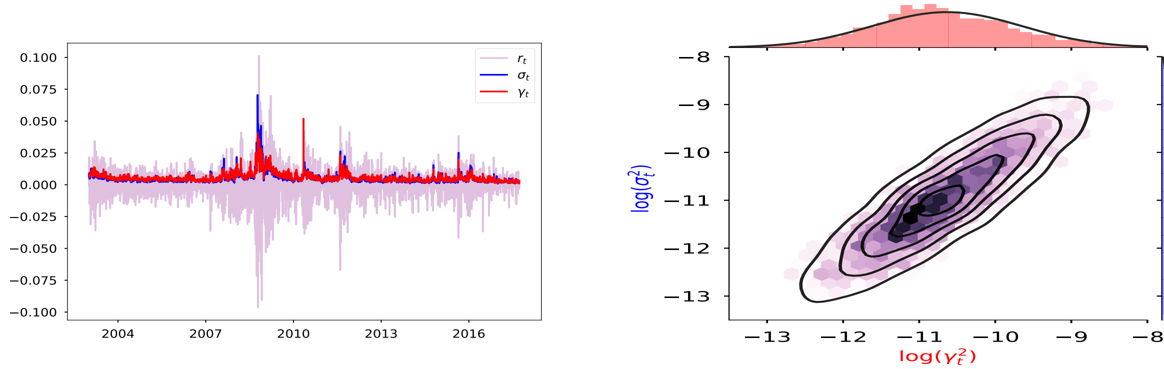


Figure 1b plots the two marginal log-volatility distributions along with their joint distribution. As can be seen from the graph, both marginal distributions are skewed right.[18] This skewness implies that the volatilities are more likely to take on abnormally large values than take on abnormally small ones. This fact is particularly noteworthy as these are distributions of log-volatilities, and taking the logarithm already removes a large amount of the skewness.

Andersen et al. (2001) argue that realized volatilities are approximately log-Gaussian. One might expect this to continue to hold in this case. The black lines in Figure 1b are Gaussian densities fit to the data for comparison purposes. At a qualitative level, the log-volatilities are roughly log-Gaussian. They are slightly skewed and slightly kurtotic, even after taking logs, which we can also see in Table 3.

Table 3: Volatility Summary Statistics

	σ_t^2	γ_t^2	$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$	$\log(\sigma_t^2)$	$\log(\gamma_t^2)$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$
Mean	4.47×10^{-5}	3.68×10^{-5}	0.56	−10.91	−10.64	−13.15	−2.17
Std. Dev.	1.52×10^{-4}	9.12×10^{-5}	0.12	1.13	0.98	1.03	0.22
Skew.	15.65	11.81	−0.18	0.71	0.55	0.72	−0.95
Kurt.	376.55	250.23	2.92	4.12	3.81	4.10	4.88

18. The only reason that the diffusion density might appear to be skewed left is that it is plotted sideways.

We are interested not just in the volatilities' univariate dynamics, but also in their interrelationships. We know from Figure 1b that the volatilities move together. To investigate this further, Table 4 reports the correlations between the various volatility measures and daily excess returns.

Table 4: Volatility Correlations

	σ_t^2	γ_t^2	$\sigma_t^2 + \gamma_t^2$	$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$	rx_t
σ_t^2	1.00	0.74	0.96	−0.29	−0.11
γ_t^2	0.74	1.00	0.89	−0.10	−0.13
$\sigma_t^2 + \gamma_t^2$	0.96	0.89	1.00	−0.23	−0.13
$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$	−0.11	−0.13	−0.13	1.00	0.12

Table 4 replicates standard results regarding the relationship between volatility and returns. We also see that jump and diffusion volatility and highly positively correlated.

7.2. Dynamics

This section studies the dynamics of log-volatilities because they are closer to Gaussian, and so the true conditional expectations are likely closer to approximately linear. Table 5 reports independent $AR(1)$ regressions for each volatility and a joint $VAR(1)$ regression to gain some high-level understanding of the dynamics. Both series are quite persistent and predictable. However, we still have economically significant innovations.[19]

Table 5: Autoregressive Models

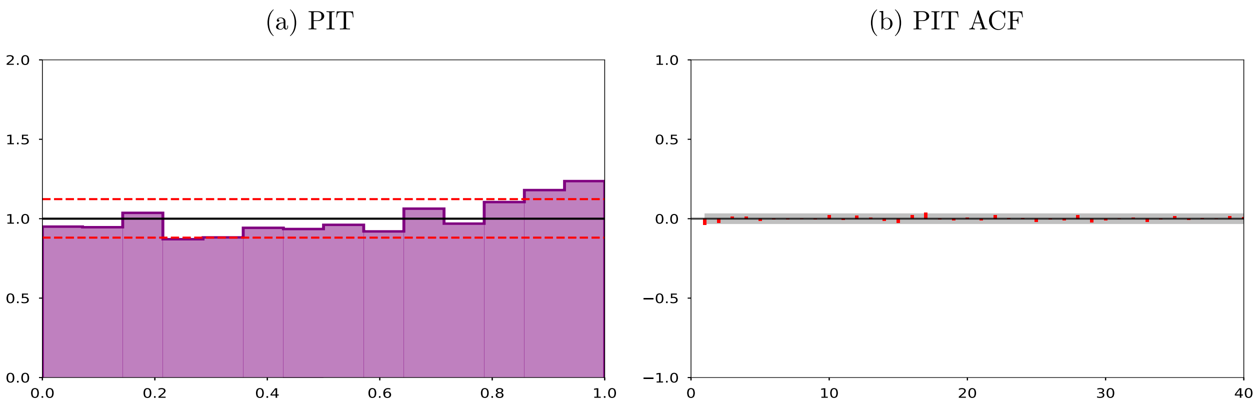
	Intercept	$\log(\sigma_{t-1}^2)$	$\log(\gamma_{t-1}^2)$	Innovation Variance	$\mathbb{R}^2$
	AR(1)				
$\log(\sigma_t^2)$	−1.63	0.85		0.31	72 %
	(−1.82, −1.45)	(0.83, 0.87)			
$\log(\gamma_t^2)$	−1.78	0.83		0.25	69 %
	(−1.97, −1.59)	(0.82, 0.85)			
	VAR(1)				
$\log(\sigma_t^2)$	−0.84	0.56	0.38	0.33	
	(−1.04, −0.64)	(0.52, 0.59)	(0.33, 0.42)		
$\log(\gamma_t^2)$	−1.80	0.34	0.48	0.27	72 %
	(−1.98, −1.62)	(0.31, 0.38)	(0.44, 0.52)		

19. This section's results come with the significant caveat that I am using estimated regressors and do not correct for this in my statistical results. For the most part, the evidence is so overwhelming the conclusions should not be affected, but, in some of the more borderline cases, it may be an issue.

7.3. Realized Density Evaluation

Section 5 showed that the estimators work well in simulations. It would be useful to know if they worked well in the data as well. Besides, perhaps the assumptions justifying the integrated-Laplace representation do not hold in practice. Thankfully, Theorem 4 is a valid conditional density, and we can consistently estimate the conditioning variables. Consequently, techniques developed to analyze conditional densities work well here.

Figure 2: Realized Density Evaluation



Each day, I take the $\widehat{\sigma}_t^2$ and $\widehat{\gamma}_t^2$ and compute $\widehat{RD_t}$. I can draw from $\widehat{RD_t}$ easily, and so I compute the probability integral transform (PIT) of the demeaned daily return using simulation. This procedure jointly evaluates the density representation, the time-aggregation procedure, and the estimation of the parameters.

As can be seen in Figure 2, the PIT is close to uniform. The only deviation is in the far right tail. I did not correct for the skewness in the data when I computed RD_t, i.e., the volatility and returns are not independent at the daily level. We can see this in the graph. However, the deviation is not large, and for most risk-measures, we are far more concerned about the left-tail. This procedure estimates that tail almost perfectly. It is also worth noting that I needed to assume this symmetry in the discrete-time representation, but not the continuous-time one. The deviations here do not invalidate that representation at all. Furthermore, Figure 2b shows the deviations are most perfectly uncorrelated across time. This lack of correlation implies that the densities' dynamics are estimated quite well.

8. News Premia: Empirics

8.1. News Premia

In practice, σ_t^2 and γ_t^2 are very heavily correlated (74 %) and so regressing on them does not lead to robust results. Moreover, interaction terms in those regressions are often significant. To isolate the effect of the jumps, I reparameterize the process in terms of $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ and $\sigma_t^2 + \gamma_t^2$. To make the innovations closer to Gaussian-distributed and avoid the need for interaction terms, I report elasticities, i.e., I apply a log transformation. Hence, the preferred specification is

$$rx_t = \beta_0 + \beta_1 \log\left(\sigma_t^2 + \gamma_t^2\right) + \beta_2 \log\left(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}\right) + \epsilon_t. \tag{9}$$

I report the OLS and robustness results in Appendix E. The identified results in the other specifications either agree with the main specification or are insignificant. The OLS results are consistent with the literature.

The analysis below uses the daily excess return, rx_t, to make the results more easily comparable with those in the literature. I construct rx_t by taking r_t and subtracting the log yield on the 10 year treasury bill, which I obtain from FRED. I annualize rx_t (multiplied it by 252) to make the results more interpretable. I use Newey-West heteroskedasticity and autocorrelation (HAC) robust standard errors and report t-statistics in square brackets. I use Bartlett's kernel with the optimal bandwidth, per Newey and West (1994).

Risk premia are forward-looking by definition, and so we must isolate the predictable variation in the regressors. Intuitively, we want to regress returns on expected volatilities. If we interpret coefficients from a contemporaneous regression, like those reported in Table 12 (which is in the appendix), as risk premia, we have the classic endogenous regressors problem because the leverage effect is the correlation between the regressors and error terms.

The most common way to handle endogenous regressors is using instrumental variables, which is what I do. In particular, I use the lagged regressors as instruments. The lagged volatilities are valid instruments. First, they explain a large amount of the variation in the regressors. I adopt an approximate heterogeneous autoregressive (HAR) specification to choose lags used as instruments, (Corsi 2009). To be precise, I use $\sigma_{t-l}^2 + \gamma_{t-l}^2$ and $\frac{\gamma_{t-l}^2}{\sigma_{t-l}^2 + \gamma_{t-l}^2}$ for $l \in \{1, 2, 5, 25\}$. I report the results from the first-stage regressions in Table 16. The F-statics reported there are far outside the week-instrument region. Second, they are predetermined, and hence by definition independent of the date-t innovation.

I consider two specifications. The leading specification uses $\log(\sigma_t^2 + \gamma_t^2)$ as my first regressor and $\log(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2})$ as my second regressor. I also consider a specification with $\log(\sigma_t^2)$ as the first regressor and $\log(\sigma_t^2 + \gamma_t^2)$ as the second regressor.

The results are highly statistically significant. Contrary to much of the previous literature, I find returns are highly predictable at the daily level. As theory predicts, I find that the predictable parts of $\log(\sigma_t^2 + \gamma_t^2)$, $\log(\sigma_t^2)$, and $\log(\gamma_t^2)$ have strong positive relationships with returns in univariate regression. Conversely, when we run a bivariate regression the coefficient associated with jumps changes sign. Table 11 shows that $\log(\sigma_t^2)$ and $\log(\gamma_t^2)$ are highly positively correlated. Consequently, it should not be surprising that $\log(\gamma_t^2)$ has different signs in the univariate and bivariate regressions. The regression on the jump proportion has similar results because it also adjusts for movements in $\sigma_t^2 + \gamma_t^2$.

We want to interpret the magnitude of the coefficients, not just the sign. Since I regress annualized excess log-return on the $\log(\sigma_t^2 + \gamma_t^2)$ and $\log(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2})$, the estimates are elasticities.

Table 6: News Premia Estimates

Regressors	Specifications					
Intercept	2.95 [6.61]	−2.45 [−5.12]	−5.04 [−0.58]	3.27 [7.25]	2.95 [6.07]	2.32 [4.15]
$\log\left(\sigma_t^2+\gamma_t^2\right)$	0.24 [6.61]		0.14 [2.68]			
$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$		−5.01 [−5.86]	−4.15 [−4.93]			
$\log(\sigma_t^2)$				0.25 [6.53]		1.86 [5.18]
$\log\left(\gamma_t^2\right)$					0.23 [5.40]	−1.74 [−4.53]

These elasticities are highly statistically and economically significant. Consider the first column. The elasticity of rx_t with respect to $\sigma_t^2+\gamma_t^2$ equals 0.24. In other words, a 1 % increase in $\sigma_t^2+\gamma_t^2$ for an entire year increases the expected yearly return by 0.24 %.[20] For comparison, the average year-to-year difference in average $\sigma_t^2+\gamma_t^2$ in my sample is $\approx 50\,\%$. It increased by $\approx 150\,\%$ between 2007 and 2008.

The average annual absolute difference in $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ is lower, equaling 6.13 %, but the regression coefficient is substantially larger. A 1 % change in $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ for an entire year changes expected returns by −5.01 %. In both cases, the implied movements in risk premia from year to year are substantial. I am not the first researcher to find movements in risk premia that are this large. Martin (2017) reports changes of similar magnitude.

As we would expect, when risk as measured by σ_t^2, γ_t^2, or $\sigma_t^2+\gamma_t^2$ increases, expected returns increase. Surprisingly, once we control for σ_t^2 the jump (news) premium is negative. It is always smaller than the diffusion volatility premium. This stylized fact is rather difficult to explain using standard finance models for two reasons. First, it means that two risk factors move at the daily frequency or faster. Second, it means that the shocks that are large relative to the amount of time over which they occur command a smaller premium than shocks which are small relative to the amount of time over which they occur. I am not the first person to find counter-intuitive relationships between risk premia and the size of shocks. See, for example, Dew-Becker, Giglio, and Kelly (2019). I also consider several other specification and setups to this problem in the appendix. The results either agree or are statically insignificant.

9. Conclusion

This paper investigates how jumps affect investors' risk. When a news shock causes the representative investor's information set to jump, she reprices the assets. I introduce jump volatility — γ_t^2 —

20. The reason that I only considered a 1 % change is that the approximation of log-differences as percent differences only holds for small changes.

which is a sufficient statistic for the jump part of price dynamics. I introduce the realized density, RD_t, to reduce tracking the returns' predictive density — $h(r_t \,|\, \mathcal{F}_{t-1})$ — to volatility forecasting. I do this by representing infinite-activity jump processes as stochastic volatility variance-gamma processes. I develop estimators for the instantaneous and integrated jump and diffusion volatilities, thereby nonparametrically identifying instantaneous jump dynamics.

These volatility estimators outperform those in the literature when the data have lots of jumps. I apply these estimators to the S&P 500 using high-frequency data from SPY. I show that the jump volatility is relatively well-behaved and is roughly log-Gaussian and that γ_t^2 has long-memory and co-moves a great deal with σ_t^2.

I next analyze how jumps affect expected returns. I show that σ_t^2 commands a positive daily risk premium as the theory predicts. I then show that $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ commands an economically and statistically significant negative premium. The magnitude of this premium implies there are substantial benefits to adding jumps to our models relative to just using volatility. This magnitude also poses interesting questions for finance theory. In particular, the data require at least two risk factors to explain daily risk premia dynamics in the S&P 500.

References

Aït-Sahalia, Yacine, Jianqing Fan, and Yingying Li. 2013. "The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency." *Journal of Financial Economics* 109 (1): 224–249.

Aït-Sahalia, Yacine, and Jean Jacod. 2009a. "Estimating the Degree of Activity of Jumps in High Frequency Data." *The Annals of Statistics* 37 (5A): 2202–2244.

———. 2009b. "Testing for Jumps in a Discretely Observed Process." *The Annals of Statistics* 37 (1): 184–222.

———. 2012a. "Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data." *Journal of Economic Literature* 50 (4): 1007–1050.

———. 2012b. "Identifying the Successive Blumenthal-Getoor Indices of a Discretely Observed Process." *The Annals of Statistics* 40 (3): 1430–1464.

Aït-Sahalia, Yacine, Per A. Mykland, and Lan Zhang. 2005. "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise." *Review of Financial Studies* 18 (2): 351–416.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. 2007. "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility." *The Review of Economics and Statistics* 89 (4): 701–720.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics* 61 (1): 43–76.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. "Modeling and Forecasting Realized Volatility." *Econometrica* 71 (2): 579–625.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega. 2003. "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange." *The American Economic Review* 93 (1): 38–62.

Bakshi, Gurdip, Peter Carr, and Liuren Wu. 2008. "Stochastic Risk Premiums, Stochastic Skewness in Currency Options, and Stochastic Discount Factors in International Economies." *Journal of Financial Economics* 87 (1): 132–156.

Bandi, Federico M., and Roberto Renò. 2012. "Time-varying Leverage Effects." *Journal of Econometrics* 169 (1): 94–113.

Barlow, Martin T. 1978. "Study of a Filtration Expanded to Include an Honest Time." *Probability Theory and Related Fields* 44 (4): 307–323.

Barndorff-Nielsen, Ole E., and Neil Shephard. 2002. "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64 (2): 253–280.

———. 2004. "Power and Bipower Variation with Stochastic Volatility and Jumps." *Journal of Financial Econometrics* 2 (1): 1–37.

Barndorff-Nielsen, Ole E., and Albert Nikolaevich Shiryaev. 2010. "Change of Time and Change of Measure." In *Advanced Series on Statistical Science & Applied Probability,* edited by Ole E. Barndorff-Nielsen, vol. 13. Toh Tuck Link, Singapore: World Scientific.

Bollerslev, Tim, Viktor Todorov, and Sophia Zhengzi Li. 2013. "Jump Tails, Extreme Dependencies, and the Distribution of Stock Returns." *Journal of Econometrics* 172 (2): 307–324.

Brandt, Michael W., and Qiang Kang. 2004. "On the Relationship Between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach." *Journal of Financial Economics* 72 (2): 217–257.

Branger, Nicole, Christian Schlag, and Eva Schneider. 2008. "Optimal Portfolios when Volatility Can Jump." *Journal of Banking & Finance* 32 (6): 1087–1097.

Cheng, Xu, Eric Renault, and Paul Sangrey. 2019. *Identification Robust Inference for Structural Stochastic Volatility Models.* Working Paper. University of Pennsylvania, April.

Christensen, Kim, Roel C.A. Oomen, and Mark Podolskij. 2014. "Fact or Friction: Jumps at Ultra High Frequency." *Journal of Financial Economics* 114 (3): 576–599.

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

Dambis, K. E. 1965. "On the Decomposition of Continuous Submartingales." *Theory of Probability and its Applications* 10 (3): 401–10.

Delbaen, Freddy, and Walter Schachermayer. 1994. "A General Version of the Fundamental Theorem of Asset Pricing." *Mathematische Annalen* 300 (1): 463–520.

Dew-Becker, Ian, Stefano Giglio, and Bryan Kelly. 2019. *Hedging Macroeconomic and Financial Uncertainty and Volatility.* Working Paper 3284790. Social Science Research Network, April.

Drechsler, Itamar. 2013. "Uncertainty, Time-Varying Fear, and Asset Prices." *The Journal of Finance* 68 (5): 1843–1889.

Dubins, Lester E., and Gideon Schwarz. 1965. "On Continuous Martingales." *Proceedings of the National Academy of Sciences of the United States of America* 53 (5): 913–916.

Gallant, A. Ronald, and George Tauchen. 2018. "Exact Bayesian Moment Based Inference for the Distribution of the Small-time Movements of an Itô Semimartingale." Indirect Estimation Methods in Finance and Economics, *Journal of Econometrics* 205 (1): 140–155.

Geman, Hélyette, Dilip B. Madan, and Marc Yor. 2001. "Time Changes for Lévy Processes." *Mathematical Finance* 11 (1): 79–96.

Gürkaynak, Refet S., Burçin Kısacıkoğlu, and Jonathan H. Wright. 2018. *Missing Events in Event Studies: Identifying the Effects of Partially-Measured News Surprises.* Working Paper, Working Paper Series 25016. National Bureau of Economic Research, September.

Hautsch, Nikolaus, and Mark Podolskij. 2013. "Preaveraging-based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence." *Journal of Business & Economic Statistics* 31 (2): 165–183.

Huang, Chi-Fu. 1985. "Information Structure and Equilibrium Asset Prices." *Journal of Economic Theory* 35 (1): 33–71.

Huang, Xin, and George Tauchen. 2005. "The Relative Contribution of Jumps to Total Price Variance." *Journal of Financial Econometrics* 3 (4): 456–499.

Jacod, Jean, Mark Podolskij, and Mathias Vetter. 2010. "Limit Theorems for Moving Averages of Discretized Processes Plus Noise." *The Annals of Statistics* 38 (3): 1478–1545.

Jacod, Jean, and Phillip Protter. 2012. "Discretization of Processes." In *Stochastic Modelling and Appplied Probability,* edited by Boris Rozovksiĭ and Peter W. Glynn, vol. 67. Berlin, Germany: Springer-Verlag.

Jacod, Jean, and Mathieu Rosenbaum. 2013. "Quarticity and Other Functionals of Volatility: Efficient Estimation." *Annals of Statistics* 41 (4): 1462–1484.

Kalnina, Ilze, and Dacheng Xiu. 2017. “Nonparametric Estimation of the Leverage Effect: A Trade-Off Between Robustness and Efficiency.” *Journal of the American Statistical Association* 112 (517): 384–396.

Lahaye, Jérôme, Sébastien Laurent, and Christopher J. Neely. 2011. “Jumps, Cojumps and Macro Announcements.” *Journal of Applied Econometrics* 26 (6): 893–921.

Lettau, Martin, and Sydney C. Ludvigson. 2010. “Measuring and Modeling Variation in the Risk-Return Tradeoff.” Chap. 11, edited by Yacine Aït-Sahalia and Lars Peter Hansen, 1:617–690. Handbook of Financial Econometrics. Elsevier.

Li, Jia, Viktor Todorov, and George Tauchen. 2017. “Jump Regressions.” *Econometrica* 85 (1): 173–195.

Lord, Roger, Remmert Koekkoek, and Dick Van Dijk. 2010. “A Comparison of Biased Simulation Schemes for Stochastic Volatility Models.” *Quantitative Finance* 10 (2): 177–194.

Martin, Ian. 2017. “What is the Expected Return on the Market?” *The Quarterly Journal of Economics* 132 (1): 367–433.

Medvegyev, Peter. 2007. *Stochastic Integration Theory.* Edited by R. Cohen, S.K. Donaldson, S. Hildebrant, T.J. Lyons, and M.J. Taylor. Oxford Graduate Texts in Mathematics. New York: Oxford University Press.

Mittnik, Stefan, and Rachev T. Svetlozar. 1993. “Modeling Asset Returns with Alternative Stable Distributions.” *Econometric Reviews* 12 (3): 261–330.

Monroe, Itrel. 1972. “On Embedding Right Continuous Martingales in Brownian Motion.” *The Annals of Mathematical Statistics:* 1293–1311.

Newey, Whitney K., and Daniel McFadden. 1994. “Large Sample Estimation and Hypothesis Testing.” Chap. 36, edited by Robert Engle and Daniel McFadden, 4:2111–2245. Handbook of Econometrics. Elsevier.

Newey, Whitney K., and Kenneth D. West. 1994. “Automatic Lag Selection in Covariance Matrix Estimation.” *The Review of Economic Studies* 61 (4): 631–653.

Oomen, Roel C. A. 2006. “Comment.” *Journal of Business & Economic Statistics* 24 (2): 195–202.

Pan, Jun. 2002. “The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study.” *Journal of Financial Economics* 63 (1): 3–50.

Podolskij, Mark, and Mathias Vetter. 2009. “Bipower-type Estimation in a Noisy Diffusion Setting.” *Stochastic Processes and Their Applications* 119 (9): 2803–2831.

Stambaugh, Robert F. 1999. "Predictive Regressions." *Journal of Financial Economics* 54 (3): 375–421.

Todorov, Viktor. 2010. "Variance Risk-Premium Dynamics: The Role of Jumps." *Review of Financial Studies* 23 (1): 345–383.

Todorov, Viktor, and George Tauchen. 2014. "Limit Theorems for the Empirical Distribution Function of Scaled Increments of Itô Semimartingales at High Frequencies." *The Annals of Applied Probability* 24, no. 5 (October): 1850–1888.

Appendix A Representation Theorems

A.1 Theorem 1 Square-Integrable Semimartingales as Integrals

Proof. We can replace the jump part of the Grigelionis representation with an integral with respect to the standard-variance gamma process where the root jump volatility is the integrator using Corollary 2.1. □

A.2 Remark 1 Jump Volatility and Predictable Quadratic Variation

Proof. The jumps are integrals with respect to Poisson random measures and the jumps are unpredictable, and so $[p]^J(t) = \sum_{s \leq t} \Delta p(s)^2 = \int_0^t \int_X \delta^2(s,x) n(ds,dx)$. This implies that $\langle p \rangle^J(t) = \mathbb{E}[[p]^j(t) \,|\, \mathcal{F}_{t-}] = \mathbb{E}[\int_0^t \int_X \delta^2(s,x) n(ds,dx) \,|\, \mathcal{F}_{t-}] = \int_0^t \int_X \delta^2(s,x) \nu(ds,dx)$, where the last inequality holds by the definition of locally-square integrable.

We also must show that the limit of the expectation approaches $\gamma^2(t)$:

$$\lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}\left[\left| p^J(t+\Delta) - p^J(t) \right|^2 \,\middle|\, \mathcal{F}_{t-} \right] = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}\left[\left| \int_t^{t+\Delta} \delta(s,x)(n-\nu)(ds\, dx) \right|^2 \,\middle|\, \mathcal{F}_{t-} \right]. \tag{10}$$

By the Itô Isometry, we can rewrite (10) as: $\lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}[\int_t^{t+\Delta} \delta^2(s,x) n(ds\, dx) \,|\, \mathcal{F}_{t-}]$. Define $\gamma^2(t) \coloneqq \int_X \delta^2(t,x) \nu(dx)$. Choosing δ so that dx, ds are independent and ν is proportional to the Lebesgue measure, we have $\lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}[\int_t^{t+\Delta} \int_X \delta^2(s,x)(ds,dx) \,|\, \mathcal{F}_{t-}] = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}[\Delta \gamma^2(t) + \int_t^{t+\Delta} (\gamma^2(t) - \gamma^2(s))\, ds \,|\, \mathcal{F}_{t-}]$. We can split this into the value of the jump volatility at t and deviations from it: $= \lim_{\Delta \to 0} \gamma^2(t) + \frac{1}{\Delta} \Delta O(\mathbb{E}[\sup_{t \leq s \leq t+\Delta} \left| \gamma^2(t) - \gamma^2(s) \right| \,|\, \mathcal{F}_{t-}]) = \gamma^2(t)$, where the last inequality holds by the Davis-Burkholder-Gundy inequality. □

A.3 Converting Poisson Processes to Continuous Martingales

Lemma 7. *Converting Poisson Processes to Continuous Martingales*

Let $Y(t)$ be a compound Poisson process with compensator $\nu(t)$. Define $N(t) \coloneqq \sum_{s \leq t} \mathbf{1}\{Y(t) \neq Y(t-)\}$. Let $\nu^*(t) \coloneqq \nu(N(t))$. Define $\rho(t) \coloneqq \min\{\tau \,|\, \Delta N(\tau) = 1, \tau > t\}$. Define the expanded *filtration*

$$\mathcal{F}^{\rho}_t \coloneqq \cup_{\epsilon>0}\mathcal{F}^{Y}_{t-\epsilon} \cap \sigma\left(\{\rho < t\}\right).^{21} \tag{11}$$

Then $Y^*(t) \coloneqq Y(t) - \nu^*(t)$ is a martingale in $\mathcal{F}^Y_t$, and there exists a continuous martingale in $\mathcal{F}^\rho_t$ that equals $Z(t) \coloneqq Y^*(N^{-1}(t))$ in probability.

Proof. We can assume that ν has Lebesgue base Lévy measure without loss of generality; otherwise, we just rescale appropriately. We have that Y^* is a martingale in $\mathcal{F}_t$ because all that time-changing ν is doing is making the compensator be a jump process with the same jump times as $Y(t)$, $\rho(t)$, with jump magnitudes $\Delta\nu(\rho) = \nu(1) - \nu(0)$.

Since $N(t)$ is a deterministic process in $\mathcal{F}^\rho_t$ and we are expanding by an honest time, $Y^*(t)$ is still a semimartingale, (Barlow 1978). Hence, we only need to show that the martingale property holds. Consider $\mathbb{E}\left[Z(1) \,|\, \mathcal{F}^\rho_\tau\right]$. Time-changing $Y(t)$ by $N^{-1}(t)$ turns 1 into ρ because there is 1 jump whose time satisfies $t \leq \rho$. That is, we pick up the one jump in Y which occurs at ρ. Meanwhile, $\nu^*(N^{-1}(1)) = \nu(1)$: $\mathbb{E}[\Delta Y(\rho) - (\nu(1) - \nu(0)) \,|\, \mathcal{F}^\rho_\tau] = 0 \implies \mathbb{E}[Z(t) - Z(t-1) \,|\, \mathcal{F}_{t-1}] = 0$ for all t.

Consider $\tau < 1$, then the change in $N(t)$ equals zero, and so the increment has mean zero. If we consider $\tau > 1$, we can combine these two arguments. Consequently, $Z(t)$ is a martingale in the $\mathcal{F}^\rho_t$ filtration.

Now, $\mathcal{F}^\rho$ and $\mathcal{F}^\rho_{t-}$ coincide by (11), which implies that $\mathcal{F}^\rho_t$ is generated by the predictable σ-algebra or, equivalently, the continuous processes. As a result, for any process adapted to this filtration, there exists a continuous martingale that equals it in probability.

□

A.4 Theorem 2 Time-Changing Jump Martingales

Proof. Let $Y(t) \coloneqq p^J(t)$. Fix a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}))$, where the space is a very good filtered extension such that $Y(t) = H \star (n - \nu)$, and n is a Poisson random measure with associated compensator ν. This compensator has associated Lebesgue base Lévy measure λ. (Since H must be left-continuous for this process to be well-defined, we can assume without loss of generality that H is predictable.) This H is a two-dimensional process that controls all of the dynamics of $Y(t)$.

To prove this result, first, I switch the base Lévy measure to one that has this property. Second, I time-change the process in each strip to handle the dynamics. Third, I switch from an integral with respect to infinitely-many Poisson processes to one with respect to a single Laplace process by appropriately combining the Poisson processes.

The process $Y(t)$ jumps only finitely-many times in any $A \subset \mathbb{R}$, with $0 \notin A$. Let $A \subset \mathbb{R}\{0\}^c$ be an open interval. Define $\widetilde{n}$ such that $\widetilde{n}([0,t] \times A) = \frac{1}{x}\exp(-x)\mathbf{1}_A \star n$, with associated compensator $\widetilde{\nu}$. This implies the base Lévy measure equals $1/x\exp(-x)$. This measure is absolutely-continuous to n for any A because it is a compensated compound Poisson processes without atoms. Jumps of size zero, on the other hand, do not affect $Y(t)$'s paths. Hence, we can represent $Y(t)$ as an integral with respect to $\widetilde{n} - \widetilde{\nu}$, (Medvegyev 2007, Example 5.51).

21. Note, we expanding $\mathcal{F}^A_{t-}$ not $\mathcal{F}^A_t$.

The benefit of switching the base Lévy measure is that it implies the associated predictable integrator, $\widetilde{H}$, is $O_p(1/x \exp(-x))$. For a stopping time τ, by the Davis-Burkholder-Gundy inequality: $C_1 \mathbb{E}\left[[Y\mathbf{1}_A](\tau)\right] \leq \mathbb{E}\left[|(Y\mathbf{1}_A)(\tau)|\right] \leq C_2 \mathbb{E}[[Y\mathbf{1}_A](\tau)]$. The middle term is just expectation of the absolute value of a Poisson random variable. Hence, $\infty > \int_A \widetilde{H}(s)\, d\widetilde{\lambda}(s) \implies \infty > \int_A \frac{\widetilde{H}(s)}{s} \exp(-s)\, ds$. Since A has arbitrary finite measure, the integrand must be $O_p(1)$. Note, we have only bounded it from above, $\widetilde{H}(s)$ may equal zero. However, since we assumed $H(x)$ has interval support, we can assume it is bounded from below over any A in the support.[22]

Define $Y^A(t) \coloneqq Y(t)\mathbf{1}\{Y(t) \in A\} - \int_0^t \int_A \widetilde{H}(x,s)\, d\widetilde{\lambda}(x)\, ds$. This process is a martingale with respect to $\mathcal{F}_t$. To see this note, first $Y(t)\mathbf{1}\{Y(t) \in A\} = \mathbf{1}\{Y(t) \in A\} \star n$; second, $\mathbf{1}\{Y(t) \in A\} \star \nu = \int_0^t \int_A \widetilde{H}(x,s)\, d\widetilde{\lambda}(x)\, ds$ by the definition of base Lévy measure; and, third, ν is the predicable compensator of n. Consequently, increments of $Y^A(t)$ are a martingale difference sequence, and so $Y^A(t)$ is a martingale with respect to the filtration it generates — $\mathcal{F}_t^A$.

I now use a time-change argument to convert $Y^A(t)$ to a continuous martingale in an expanded filtration. Doing this will let me use the Dambis-Dubins-Schwarz theorem on a continuous martingale. Define the expanded filtration $\mathcal{F}_t^\rho \coloneqq \cup_{\epsilon > 0} \mathcal{F}_{t-\epsilon}^A \cap \sigma\left(\{\rho < t\}\right)$ and $N^A(t) \coloneqq \sum_{s \leq t} \mathbf{1}\{Y^A(t) \neq Y^A(t-)\}$. Define $\nu^* = \widetilde{\nu}(N(t))$ and $Z^A(t) \coloneqq (Y^A - \nu^*)(N^{-1}(t))$.

By Lemma 7, we can assume without loss of generality that $Z^A(t)$ is a martingale in $\mathcal{F}_t^A$ and a continuous martingale in $\mathcal{F}_t^\rho$. By the Dambis, Dubins & Schwarz theorem, a continuous martingale is a Wiener process when time-changed by its quadratic variation. Therefore, $Z^A(t) \overset{a.s.}{=} \mathcal{W}(\langle Z^A \rangle(t))$, where $\mathcal{W}$ is the standard Wiener process.

There are two main limitations to this result. First, we want an expression for Y, not just for each of the Z^A. Second, $\mathcal{F}_t^\rho$ is not the filtration generated by the data, and so we must consider what this representation implies about the process in that filtration.

To resolve the first problem, we must aggregate the strips. Consider aggregating all of the $N^A(t)$. From the definition of $\widetilde{\nu}$, aggregating the $N^A(t)$ creates a Poisson random measure with intensity measure $\widetilde{\nu}(x) = x^{-1}\exp(-x)dx$. This is just the intensity measure of a gamma process.

For a countable partition of $Z(t)$'s support, $A_1, A_2 \ldots,\ Z = \sum_{A_i} Z^{A_i}$. Furthermore, Wiener processes are stable under countable sums as long as the variance remains finite, which it does in this context because the initial process is locally-square integrable. Because the Poisson processes in the different A are conditionally independent, the Weiner processes are as well. Consequently, we can drop the A superscript in the previous argument, and $\langle Z \rangle = \sum_{A_i} \langle Z_i^A \rangle$.

Since we are conditioning on additional variation over any interval I,

$$\langle Y \,|\, \mathcal{F}_t^Y \rangle \leq \langle Z \,|\, \mathcal{F}_t^Y \rangle \leq \langle Y \,|\, \mathcal{F}_t^Y, \text{at least one jump in } I \rangle = \langle Y \,|\, \mathcal{F}_t^Y \rangle, \tag{12}$$

where the last equality follows because we are conditioning on a probability one event by the infinitely-activity assumption. Because intervals generate the Borel σ-algebra over $\mathbb{R}_+$, this implies that $\langle Y \rangle = \langle Z \rangle$. Then since $\langle Y \rangle$ is measurable with respect to $\mathcal{F}_t$ and the jumps in $N(t)$ are totally

22. Any A with $H(x) = 0$, for $x \in A$ has $\mathbb{E}[\mathbf{1}\{Y(t) \in A\}] = 0$.

inaccessible this implies that $\langle Z \rangle$ and $N(t)$ are independent.

This implies that

$$(Y-\nu^*)(N^{-1}(t)) \overset{\mathcal{L}}{=} Z(t) \overset{\mathcal{L}}{=} \mathcal{W}(\langle Z \rangle(t)) \implies (Y-\nu^*)(t) \overset{\mathcal{L}}{=} \mathcal{W}(\langle Z \rangle(N(t))) \overset{\mathcal{L}}{=} \mathcal{W}(N(\langle Z \rangle(t))), \quad (13)$$

where the last inequality follows by noting we can interchange two different time-changes if they are independent and at least one of them is a Lévy process. We can combine $\mathcal{W}$ and N, and get a Laplace process: $(Y-\nu^*)(t) \overset{\mathcal{L}}{=} \mathcal{L}(\langle Z \rangle)(t)$.

All we have left to show is that $Y-\nu^*$ and $Y-\widetilde{\nu}$ induce the same distributions. It is sufficient to show that $\widetilde{\nu}$ and ν^* induce the same distributions over short-intervals because that implies that they act the same as integrators for predictable processes (which are automatically independent of $N(t)$). All that the time-change does is convert $\widetilde{\nu}$ into a step function with the steps at the jumps of $N(t)$. Consider an interval I of length ξ. Now, $|N(\xi)-\xi|$ declines exponentially as the tails of a gamma random variable does. In addition, because $N(t)$ is infinite-active the jump locations form a dense subset of $\mathbb{R}_+$. Since $\widetilde{\nu}$ is an finite-variation predictable process, it is a semimartingale. Consequently, the Davis-Burkholder-Gundy applies, and so the maximal deviation of a p-polynomial is less than ξ^p. Consequently, the maximal deviation is bounded above by $C\xi\xi^p|N(\xi)-\xi|$ where C is a constant. This is proportional to $\xi^{p+1} \cdot \exp(-\xi)\text{polynomial}(\xi) \to 0$ as $\xi \to 0$. This holds for any p, and the polynomials are sufficient to generate the Borel σ-algebra. □

A.5 Theorem 2 Time-Changing Finite-Activity Jump Martingales

Proof. Let $Y(t) \coloneqq p^J(t)$ be a finite-activity jump martingale (a compensated compound Poisson process). Fix a stochastic basis $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. Note, we can assume without loss of generality that there exists a $\delta(s,x)$ so that the base Lévy measure λ is independent of time (dt). Define the expanded filtration $\mathcal{F}_t^\rho \coloneqq \cup_{\epsilon>0}\mathcal{F}_{t-\epsilon}^Y \cap \sigma(\{\rho < t\})$ and $N(t) \coloneqq \sum_{s \leq t} \mathbf{1}\{Y(t) \neq Y(t-)\}$.

Define $Z(t) \coloneqq Y(N^{-1}(t))$. As in the proof of Lemma 7, if $\mathbb{E}[Z(t) \,|\, \mathcal{F}_t^\rho] = Z(t)$ there exists a continuous martingale in $\mathcal{F}_t^\rho$ that equals $Z(t)$ in probability. Consider $\mathbb{E}[Z(1)-Z(0) \,|\, \mathcal{F}_t^\rho] = \mathbb{E}[\Delta Y(\rho) \,|\, \mathcal{F}_t^\rho] = 0$ because it is just the expected jump magnitude, which is zero by assumption. For some $\tau < 1$, $N(t)$ is constant, and so the difference is zero. For $\tau > 1$, we can combine these arguments. By the Dambis, Dubins, and Schwarz theorem, $Z(t) \overset{a.s.}{=} \mathcal{W}(\langle Z \rangle)$, where we extend the space as necessary. We have a sequence of two time-changes: $N(t)$ and $\langle Z \rangle(t)$.

The question is what is the relationship between $\langle Z \,|\, \mathcal{F}_t^\rho \rangle$ and $\langle Y \rangle$. Unlike in the proof of Theorem 2, we cannot assume that they are independent. For an interval I, we must keep track of the probability of no jump, this event is not independent of the number of jumps, and so we cannot interchange filtrations like we do in that proof. Note, $Z(t)$ and $Y(t)$ have the same jump magnitudes. Consider an interval I, by subdividing I as necessary we can assume without loss of generality that I has at most one jump. If I, has zero jumps, both $N(t)$ and $Y(t)$ have zero variation, and so $\langle Z \,|\, \mathcal{F}_t^\rho \rangle$ and $\langle Y \,|\, \mathcal{F}_t^\rho \rangle$ equal zero. Conversely, if I contains one jump, then both equal $\mathbb{E}[(\Delta Y)^2 \,|\, \mathcal{F}_{t-}^Y, \rho]$. Hence, the $\langle Z \,|\, \mathcal{F}_t^\rho \rangle = \langle Y \,|\, \mathcal{F}_t^\rho \rangle$. Since ρ and $\mathcal{F}_{t-}^Y$ jointly generate $\mathcal{F}_t^\rho$ and

$\langle Y \mid \rho \rangle = \langle Y \mid \mathcal{F}_t^{\rho} \rangle$, we have a compound Poisson mixture of time-changed Weiner processes. That is

$$Y(t) = \begin{cases} \mathcal{W}\left(\langle Y \mid \rho \rangle (t)\right) & \text{with intensity } \nu \\ \delta_0(t) & \text{with intensity } 1-\nu. \end{cases} \tag{14}$$

□

A.6 Theorem 1 Jump Processes as Integrals

Proof. Since $Y(t)$ is an Itô semimartingale, $Y(t) = \int_0^t \int_{\mathbb{R}} \delta(s,x)(ds,dx)$. This implies that its predicable quadratic variation, $K(t)$, equals $\int_0^t \int_{\mathbb{R}} \delta^2(s,x)(dx,ds)$ with time-derivative $k(t)$ equal to $\int_{\mathbb{R}} \delta^2(s,x)\,dx$.

Let $J(t)$ be the purely-discontinuous martingale part of $Y(t)$, then Theorem 2 implies that $J(\langle Y \rangle^{-1})(t) \overset{\mathcal{L}}{=} \mathcal{L}(t)$, or equivalently, $J(t) \overset{\mathcal{L}}{=} \int_0^{\langle Y \rangle^{-1}(t)} d\mathcal{L}(s)$. Then since $\frac{k}{\sqrt{2}}\mathcal{L}(1) = \mathcal{L}(k^2)$, where $\mathcal{L}(1)$ is a standard Laplace random variable, and $k(t)$ is a predicable process (and so independent of $\mathcal{L}$), $J(t) = \int_0^t k(s)\,d\mathcal{L}(s)$.

□

A.7 Theorem 4 Realized Density Representation

Proof. Consider the diffusion part of the process:

$$h\left(p^D(t) - p^D(t-1) \,\middle|\, \mathcal{F}_{t-1}\right) = h\left(\sum_{n \in \frac{1}{\Delta},\ldots,0} \int_{t-n\Delta}^{t-(n+1)\Delta} \sigma(s)\,dW(s) \,\middle|\, \mathcal{F}_t\right). \tag{15}$$

If Δ is small enough, we can pull $\sigma^2(t)$ out of the integral because requiring the integrand to be predictable does not affect the value of the process. This equals

$$h\left(\lim_{\Delta \to 0} \sum_{n \in 1, \frac{1}{\Delta}} \sigma(t-n\Delta) \int_{t-n\Delta}^{t-(n+1)\Delta} dW(s) \,\middle|\, \mathcal{F}_{t-1}\right). \tag{16}$$

Since the martingale components of $\sigma^2(t)$ are independent of W, we can condition on the entire path of $\sigma^2(t)$ without affecting the distribution of the increments of W:

$$h\left(\sqrt{\int_{t-1}^{t} \sigma^2(s)\,ds} \int_{t-1}^{t} dW(s) \,\middle|\, \mathcal{F}_{t-1}\right) \overset{\mathcal{L}}{=} \int_{\sigma_t^2} N\left(0, \int_t^{t+\Delta} \sigma^2(s)\,ds\right) dG\left(\sigma_t^2 \,\middle|\, \mathcal{F}_{t-1}\right) \tag{17}$$

The argument for the jump volatility follows, mutatis mutandis. The only real difference is that the scale (the expectation of the absolute deviation) of the Laplace distribution equals $\sqrt{2}$ times the standard deviation.

You can just carry the drift through the analysis and then add it back at the end. To combine the jump and diffusion realized densities, note that the density of independent variables is the convolution of the densities. The integrators are pure-jump and diffusive martingales, and so they are orthogonal. Consequently, the jump and diffuse parts are independent conditional on the drift and the volatilities. Also, I derived RD_t in the argument above; it is merely the function inside the integral.

□

APPENDIX B VOLATILITY ESTIMATION

B.1 Assumption HL implies Assumption SHL

Similar to Jacod and Protter (2012), I prove the global convergence by proving local convergence. I do this by slightly modifying Jacod and Protter's (2012) Assumption SH in the same way that I modified their Assumption H in Section 4.1. I also slightly modify the literature's Assumption SH. Let ω index the underlying probability space Ω.

Assumption SHL. We have Assumption HL and there is a constant K such that the following hold for all t and all ω:

$$\|b(t,\omega)\| < K, \ \|\sigma(t,\omega)\| < K, \ \|\gamma(t,\omega)\| < K.$$

These two assumptions are closely related. Assumption HL is the local version of Assumption SHL. Assumption HL only restricts the local behavior of the function, while Assumption SHL make the equivalent conditions globally. Since convergence in the Skorokhod topology only depends upon local behavior, convergence under Assumption SHL implies convergence under Assumption HL. The arrow with $\mathcal{L}$-s above it denotes stable convergence in law.

Lemma 8 (HL implies SHL)**.** *If an Itô semimartingale $p(t)^n \xrightarrow{\mathcal{L}\text{-s}} p(t)$ under Assumption SHL, then $p(t)^n \xrightarrow{\mathcal{L}\text{-s}} p(t)$ under Assumption HL, and the equivalent statement holds for convergence in probability.*

Proof. Let $U^n(p)(t)$ and $U(p)(t)$ refer to two processes that are defined in terms of $p(t)$. In the first step, I define a process in terms of $p(t)$ that satisfies Assumptions SHL and Infinite-Activity Jumps and characterize its relationship to $p(t)$. In the second step, I show that if that $p(t)$ satisfies Assumptions HL and Infinite-Activity Jumps, then $U^n(p)(t) \xrightarrow{\mathcal{L}\text{-s}} U(p)(t)$ under Assumption SHL implies $U^n(p)(t) \xrightarrow{\mathcal{L}\text{-s}} U(p)(t)$ under Assumption HL. I then show that Assumption Infinite-Activity Jumps is unnecessary, and similar statements hold for convergence in probability and convergence of stopped processes.

Step 1

We can assume without loss of generality that $\mu(0) = 0$, and so there is a localizing sequence τ_j such that $\|\mu(t)\| \leq j$ if $0 \leq t \leq \tau_j$. Define the stopping times $R_j = \inf\left(t : \|p(t)\| + \|\sigma(t)\| \geq \xi\right)$ and the stopping times $Q_j = \inf(t : \|p(t)\| + \|\gamma(t)\| \geq \xi)$. These increase to ∞ as well. Therefore, we can set $S_j = \tau_j \wedge R_j \wedge Q_j$.

Define the following processes:

$$\mu^{(j)}(t) = \mu(t \wedge S_j),\ \sigma^{(j)}(t) = \sigma(t \wedge S_j),\ \gamma^{(j)}(t) = \gamma(t \wedge S_j) \tag{18}$$

and

$$p^{(j)}(t) = \begin{cases} 0 & \text{if } S_j = 0 \\ p(0) + \int_0^t \mu^{(j)}(s)\,ds + \int_0^t \sigma^{(j)}(s)\,dW(s) + \int_0^t \gamma^{(j)}\,d\mathcal{L}(s) & \text{if } S_j > 0. \end{cases} \tag{19}$$

The local characteristics of $p^{(j)}(t)$ and $p(t)$ agree when $t < S_j$ because they are defined to be the same. If $S_j = 0$, then $\|p(t)\| = 0$, and so we are equal there as well. Furthermore, if we use the same driving measures $W(t)$ and $\mathcal{L}(t)$ to represent both processes, the equality is not just in distribution, but ω by ω, where the original processes are defined relative to an event space Ω. In addition, $p^{(j)}(t)$ satisfies Assumption SHL, since $\|p^{(j)}(t)\| \leq 3\xi$.

Step 2

By the proof of Lemma 4.4.9 in Jacod and Protter (2012), the above statement is sufficient to show that the estimators defined above imply convergence stably-in-law. Then this holds for any process, and so it holds for the stopped versions above. Convergence stably-in-law implies convergence in probability if the two processes are defined on the same probability space, which we do not change above. So if the initial result implies convergence in probability, the new one does as well.

If $p(t)$ does not satisfy Assumption Infinite-Activity Jumps, then it is locally a convolution of a Laplacian mixture and the zero process by Theorem 3. Replacing part of the sample path with 0 does not violate any boundedness conditions. Therefore, we can replace $p^{(j)}(t)$ with the 0 process when necessary, and so the result even holds if Assumption Infinite-Activity Jumps does not hold. □

B.2 Proposition 1 Instantaneous Absolute Volatility

Proof. I start this proof by deriving the mean of the absolute volatility under an assumption that $\sigma(t)$ and $\gamma(t)$ are locally constant. I then show that the estimator in that situation converges to its mean. I then relax the assumption of locally-constant volatility.

Step 1

In this section, I start by applying Itô's Formula for convex functions to $|p|(t)$ to separate its variation into its jump and continuous components. Recall the left-derivative of the absolute value function is $f'_{-} = \operatorname{sign}(x)$. Using Medvegyev (2007, Theorem 6.65), where $A(t)$ is a finite-valued increasing process, we can rewrite $|p(t)|$ as

$$|p(t)| = \int_0^t \operatorname{sign}(p(s-))\,dp(s) + A(t) = \int_0^t \operatorname{sign}(p(s-))\,dW(s) + \int_0^t \operatorname{sign}(p(s-))\,d\mathcal{L}(s) + A(t). \quad (20)$$

This $A(t)$ can be absorbed into the drift term of $p(t)$ and vanishes as $\Delta \to 0$.

If the Laplace part and the diffusion parts have the same sign, $|p|(t) - A(t)$ is the sum of the absolute values of the two processes. Since the innovation processes are independent and symmetric, this occurs one-half of the time. If they have different signs, the situation is more complicated. In that case, $\operatorname{sign}(p(s-))$ is the same as the sign of the larger, in magnitude, of the two processes. Let $\Omega^{\mathcal{L}}$ denote the set where the Laplace part has a larger magnitude and Ω^W denote the part where the diffusion part does. Then

$$\begin{aligned} &|p(t)| - A(t) \,|\, \text{the signs differ} \quad (21)\\ &= \int_0^t \operatorname{sign}(W(s-))1_{\Omega^W}(s-)\sigma(s)\,dW(s) - \int_0^t \operatorname{sign}(\mathcal{L}(s-))1_{\Omega^W}(s-)\gamma(s)\,d\mathcal{L}(s)\\ &+ \int_0^t \operatorname{sign}(\mathcal{L}(s-))1_{\Omega^{\mathcal{L}}}(s-)\gamma(s)\,d\mathcal{L}(s) - \int_0^t \operatorname{sign}(W(s-))1_{\Omega^{\mathcal{L}}}(s-)\sigma(s)\,dW(s). \end{aligned}$$

Let Δ be the length of an interval over which $\gamma(t)$ and $\sigma(t)$ are constant, and let $|\psi|$ and $|\phi|$ denote the densities of the absolute values of a Laplace and Gaussian variables, respectively. Then we can rewrite an increment of (21) as follows conditional on the signs differing as follows:

$$\int_0^\infty \int_x^\infty (y-x)\psi_{\gamma,\Delta}(x)|\phi|_{\sigma,\Delta}(y)\,dx\,dy + \int_0^\infty \int_y^\infty (x-y)\psi_{\gamma,\Delta}(x)|\phi|_{\sigma,\Delta}\,dy\,dx. \quad (22)$$

Plugging in the parametric forms of ψ and ϕ gives:[23]

$$\frac{\sqrt{\Delta}}{\sqrt{2}}\left(-\gamma + \frac{2}{\sqrt{\pi}}\sigma + \gamma \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right)\right) + \frac{\gamma\sqrt{\Delta}}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right) \sqrt{\Delta}\left(m_1\sigma + \frac{\gamma}{\sqrt{2}}\left(2 \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right) - 1\right)\right). \quad (23)$$

When both parts have have the same sign, the absolute value is just the sum of the absolute values and so we can rewrite (21) given that the signs equal as $m_1\sigma\sqrt{\Delta} + \frac{\gamma}{\sqrt{2}}\sqrt{\Delta}$. By taking the average of this and (23), we can solve for (21):

$$\mathbb{E}|p(t)| - A(t) = m_1\sigma\sqrt{\Delta} + \frac{\gamma\sqrt{\Delta}}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right). \quad (24)$$

Step 3

This section considers the asymptotic behavior of the estimator. It proves convergence in mean-square, which implies convergence in probability. Let Ω_n be the set where the two increments have the same sign and let λ_n be its accompanying Lebesgue measure.

23. A standard computer algebra system can be used to perform the requisite integration.

Assume that $\sigma(t)$ and $\gamma(t)$ are step functions, there exists a sequence $\{\tau_j\}$ such that $\sigma(t)$ and $\gamma(t)$ are constant in (τ_j, τ_{j+1}). Hence,

$$p(t) = \sum_j \int_{\tau_j}^{\tau_{j+1}} \sigma(\tau_j)\, dW(s) + \int_{\tau_j}^{\tau_{j+1}} \gamma(\tau_j)\, d\mathcal{L}(s) \tag{25}$$

Consider the squared norm of the difference between the estimator and its expectation. It is worth noting that as k_n gets large, we are averaging over times earlier and earlier with respect to τ, which is why the bottom part of the integral is growing with k_n, not the top part. We can assume without loss of generality $\sigma(t)$ and $\gamma(t)$ are constant over $\tau - k_n\Delta^n, \tau$ by taking $k_n\Delta^n$ to 0 faster than the mesh of the τ_j goes to zero, (which it may not at all). Consequently, we let τ depend upon n in our notation.

We can rewrite the sample and population difference as

$$\mathbb{E}\left[\left\| \frac{1}{k_n\Delta^n} \sum_{m=0}^{n} \left|\Delta^n_{i_n+m} p\right| - \left| m_1\sigma(\tau_n) + \frac{\gamma(\tau_n)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau_n)}{\gamma(\tau_n)}\right)\right| \right\|^2\right] \tag{26}$$

$$= \frac{1}{k_n^2\Delta^n}\mathbb{E}\left[\left\| \sum_{m=0}^{k_n} \left| \int_{\tau(n,m+1)}^{\tau(n,m)} \sigma(\tau(n,m+1))\, dW(s) + \int_{\tau(n,m+1)}^{\tau(n,m)} \gamma(\tau(n,m)\, d\mathcal{L}(s) \right| \right.\right. \tag{27}$$

$$\left.\left. - k_n\sqrt{\Delta^n}\left| m_1\sigma(\tau_n) + \frac{\gamma(\tau_n)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau_n)}{\gamma(\tau_n)}\right)\right| \right\|^2\right].$$

By applying (24), we have

$$= \frac{1}{k_n^2\Delta^n}\mathbb{E}\left[\left\| k_n\sqrt{\Delta^n}\left| m_1\sigma(\tau(n,m)) + \frac{\gamma(\tau(n,m))}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau(n))}{\gamma(\tau(n,m))}\right)\right| + A(t) \right.\right. \tag{28}$$

$$\left.\left. - k_n\sqrt{\Delta^n}\left| m_1\sigma(\tau_n) + \frac{\gamma(\tau_n)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau_n)}{\gamma(\tau_n)}\right)\right| \right\|^2\right].$$

Simplifying implies this equals

$$O_p\left(\frac{k_n^2(\Delta^n)^2}{k_n^2\Delta^n}\right) + O_p(\Delta^n), \tag{29}$$

since $A(t)$ has finite-variation.

Step 5

I now show the step function approximation is innocuous. Consider a sequence $\tau_n \to \tau$, and define $\tilde{\sigma}(t) = \sigma(\max \tau_n : \tau_n \leq t)$, and similarly for $\tilde{\gamma}(t)$. Define $\xi_x^2(t) = \sup_{s_1,s_2<t\wedge\tau} |x(s_1) - \tilde{x}(s_2)|^2$ for x equal to σ and γ, and let $\xi_b^2(t) = \sum_{s_1,s_2<t\wedge\tau} |b(s_1) - b(s_2)|$. These functions exist and are almost surely finite by localization since σ, γ, and b are locally-bounded. Now, consider the squared distance between any semimartingale satisfying our assumptions and the one used in (25). Let

$t_1, t_2 < \tau$. Consider:

$$\mathbb{E}\left[\left\| \int_{t_1}^{t_2} \mu(s)\, ds + \int_{t_1}^{t_2} \sigma(s)\, dW(s) + \frac{1}{2}\int_{t_1}^{t_2} \gamma(s)\, d\mathcal{L}(s) - \left(\int_{t_1}^{t_2} \tilde{\sigma}(s)\, dW(s) + \frac{1}{2}\int_{t_1}^{t_2} \tilde{\gamma}(s)\, d\mathcal{L}(s)\right) \right\|^2\right]. \tag{30}$$

Increasing the range is valid because all of the integrands are positive:

$$\leq \mathbb{E}\left[\int_{t_1}^{\tau} \mu(s)^2(s) + \int_{t_1}^{\tau} |\tilde{\sigma}(s) - \sigma(s)|^2\, ds + \frac{1}{2}\int_{t_1}^{\tau} |\tilde{\gamma}(s) - \gamma(s)|^2\, ds\right]. \tag{31}$$

Then we can bound each of the terms:

$$= (O(1)\gamma_b^2(\tau) + O(1)\gamma_\sigma^2(\tau) + O(1)\gamma_\gamma^2(\tau))(\tau - t_2) = O(1)(\tau - t_2). \tag{32}$$

In other words, if we choose a sequence of meshes so that the supremum of their magnitudes $\Delta^n \to 0$ and the minimal $\tau - k_n\Delta^n \to 0$, the entire square converges. As one might expect from the definition of integration, approximating the integrands by step functions is innocuous.

Finally, we combine the previous steps to bound the entire process. Note, since variances of sums can be written in terms of variance of the original parts and their covariance, the asymptotic rate at which the quadratic variation decreases towards zero equals the larger of the asymptotic rates at which its constituent components do. Let $Y(t)$ be the absolute value of the process derived in (25). Consider the mean-square deviation of the estimator from its limiting value:

$$\frac{1}{k_n^2\Delta_n}\mathbb{E}\left[\left\| \sum_{m=0}^{k_n-1} |\Delta_{i_n+m}^n p| - Y(t) + Y(t) - \left(m_1\sigma(\tau-)k_n\sqrt{\Delta^n} + \frac{\gamma(\tau-)}{\sqrt{2}}\operatorname{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right)k_n\sqrt{\Delta^n}\right)\right\|^2\right].$$

Splitting the term into two parts and applying (29) and (32) gives: $\frac{1}{k_n^2\Delta_n}\left(O(\Delta_n^2 k_n^2) + O(\Delta_n k_n)\right) \to 0$.

∎

B.3 Theorem 5 Instantaneous Diffusion Volatility

Proof. By localization, we can replace Assumption HL with Assumption SHL without loss of generality. Also, the jump martingale part of the process is a sum of an integral with respect to $\mathcal{L}(t)$ and $\delta_0(t)$, where the weights depend upon the jumps' intensity by Theorem 3. The jump increments of that part are almost surely zero, and so if we separate the space into parts where $\mathcal{L}(t)$ is active and where $\delta_0(t)$ is active, we only have to deal with the first section. Consequently, we can assume without loss of generality that the jump part is an integral of with respect to $\mathcal{L}(t)$. The part of the proof regarding the process's continuous part does change.

Step 1

I proceed by showing convergence in mean-square, which implies convergence in probability. Note, $\left|\Delta^n_{i_n+m} p\right|^2 = O_p(\Delta^n)$ for all i, since $p(t)$ is an integral with bounded integrands and integrators whose quadratic variation is proportional to Δ^n. Consider the jump part of the variation. To prove initial estimator's consistency, I must show that the jump part converges to zero.

Following Jacod and Protter (2012, 258), for all $w, x, y, z \in \mathbb{R}$, $\epsilon \in (0,1]$, and $v \geq 1$,

$$\left|(x+y+z+w)1\{|x+y+z+w| < v\} - x^2\right| \leq K\frac{|x|^4}{v^2} + \epsilon x^2 + \frac{K}{\epsilon}((v^2 \wedge y^2) + z^2 + w^2). \tag{33}$$

Define the following four processes. The continuous variation is split into two parts, one with locally constant volatility and the other with additional deviation coming from changes in the volatility: $Y^n(t) := \sigma(\tau_n)(W_t - W_{\tau_n})1\{\tau_n \leq t\}$, $Y'^n(t) := \int_{\tau_n \wedge t}^t (\sigma(s) - \sigma(\tau_n)\, dW(s)$, $Z^n(t) := \int_{\tau_n \wedge t}^t \gamma(s)\, d\mathcal{L}(s)$, and $B^n(t) := \int_{\tau_n \wedge t}^t \mu(s)\, ds$.

Note, $p(\tau_n \wedge t) = Y^n(t) + Y'^n(t) + Z^n(t) + B^n(t)$. Now, we can use (33) with $x = \frac{\Delta^n_{i_n+m} Y^n}{\sqrt{\Delta^n}}$, $y = \frac{\Delta^n_{i_n+m} Z^n_{i_n+m}}{\sqrt{\Delta^n}}$, and $w = \frac{\Delta^n_{i_n+m} B^n_{i_n+m}}{\sqrt{\Delta^n}}$. The main issue here is showing that all of the parts except for $Y^n(t)$ converge to zero because it has correct variance. Take $v = \frac{v_n}{\sqrt{\Delta^n}} = \omega_n$, where $\omega_n = o_p(1/\Delta^n)$ and $1/\omega_n$ is $o_p(\sqrt{\Delta})$. Then $\frac{1}{k_n \Delta^n}\sum_{m=0}^{k_n-1}\left|(Y^n_t)^2 - (p^n_t)^2\right|$ is bounded by

$$\frac{1}{k_n}\sum_{m=0}^{k_n-1}\left(\frac{K}{\omega_n^2}\left|\frac{\Delta^n_{i_n+m}Y^n}{\sqrt{\Delta^n}}\right|^4 + \epsilon\left|\frac{\Delta^n_{i_n+m}Y^n}{\sqrt{\Delta^n}}\right|^2 + \frac{K}{\omega_n^2\epsilon}\left|\frac{\Delta^n_{i_n+m}Z}{\sqrt{\Delta^n}\omega_n}\right|^2 + \frac{K}{\epsilon}\left|\frac{\Delta^n_{i_n+m}Y'^n}{\sqrt{\Delta^n}}\right|^2 + \frac{K}{\epsilon}\left|\frac{\Delta^n_{i_n+m}B}{\sqrt{\Delta^n}}\right|^2\right). \tag{34}$$

Set $\xi_n = \sum_{s \in |\tau_n, \tau_n + (k_n+2)\Delta^n|}|\sigma(s) - \sigma(\tau_n)|^2$, which is bounded and converges to zero, and $\phi_n = \sum_{s \in |\tau_n, \tau_n + (k_n+2)\Delta^n|}|\gamma(s)|^2$ The key hard part is bounding $\Delta^n_{i_n+m}Z$. Clearly, $E\left|\Delta^n_{i_n+m}\right| \leq \phi_n\sqrt{\Delta^n}$. Consider the part of the variation in $Z(t)$ that comes from jumps smaller than 1 in magnitude, where 1 is an arbitrary constant picked for the sake of simplicity:

$$\mathbb{E}|\mathcal{L}(0,\phi_n) \wedge 1| = \phi_n\sqrt{\Delta^n} - \exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right)(\phi_n\sqrt{\Delta^n} + 1) \leq O\left(\frac{1}{\sqrt{\Delta^n}}\right)\exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right). \tag{35}$$

In addition, since τ_n is a stopping time, the probability that a jump exceeds 1 in the previous k_n periods declines to 0 almost surely with Δ^n. Consequently, $\frac{1}{\omega_n^2}\frac{\Delta^n_{i_n+m}Z}{\sqrt{\Delta^n}\omega_n} \overset{a.s.}{\in} O_p\left(\frac{1}{\Delta^n\omega_n^3}\right)\exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right) = o_p(1)$ as exponential functions decay faster than polynomials increase.

I use K to refer to an arbitrary constant, which may change. The drift term is $\Delta^n_{i_n+m}B$, and so $\left|\Delta^n_{i_n+m}B\right| \leq K\Delta^n$. Meanwhile, $\mathbb{E}[\left|\Delta^n_{i_n+m}Y^n\right|^4 | \mathcal{F}_{(i_n+m-1)\Delta^n}] \leq K(\Delta^n)^2$, and $\mathbb{E}[\left|\Delta^n_{i_n+m}Y'^n\right|^n | \mathcal{F}_{(i_n+m-1)\Delta^n}] \leq K\Delta^2\mathbb{E}[\xi_n|\mathcal{F}_{(i_n+m-1)\Delta^n} \leq K\Delta^n$. As a consequence, we have the following where ξ_n is some sequence converging to zero:

$$\mathbb{E}[\left|(Y^n_t)^2 - (p^n_t)^2\right|] \leq K\epsilon + \frac{K}{\epsilon}(o_p(1) + o_p(1) + \mathbb{E}[\xi_n]). \tag{36}$$

Taking $n \to \infty$ and then $\epsilon \to 0$, makes the left-hand side (36) converge to zero.

Step 2

Consider $\lim_{n\to\infty} \frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}|Y_t^n|^2$ is. Its definition implies it converges to the variance of the increment:

$$\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}|\sigma_{\tau_n}(W_t - W_{\tau_n})1\{\tau_n \le t\}|^2 = \sigma(\tau_n)^2\frac{1}{k_n}\sum_{m=0}^{k_n-1}\left|\frac{\Delta^n_{i_n+m}W}{\sqrt{\Delta^n}}\right|^2 \to \sigma(\tau_n)^2 \tag{37}$$

Since the square is a convex function, we can combine these two previous limits, given that the original expression converges to $\sigma(\tau_n)^2$. However, this is the local integrated volatility evaluated at τ_n, which was the object of interest. Multiplying the expression by a value converging to 1 does not change the results.

□

B.4 Theorem 6 Instantaneous Jump Volatility

Proof. Note, 0-subscripts denote population objects. Consider

$$\widehat{Q}_n(\gamma) \coloneqq g\left(\left|\frac{1}{k_n\sqrt{\Delta}}\sum_{m=0}^{k_n-1}\left|\Delta^n_{i_n+m}p\right| - m_1\hat{\sigma}(\tau-) - \gamma\,\mathrm{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right)\right|\right). \tag{38}$$

Note that $\widehat{Q}_n(\gamma)$ is implicitly a continuous function of $\hat{\sigma}_n(\tau-)$. By assumption, $\hat{\sigma}_n(\tau-)\xrightarrow{\mathbb{P}}\sigma_0$, and so we can suppress that dependence in our notation and plug in σ_0. In addition, g is an increasing function and both g and the absolute value are convex, continuous functions, we can use the continuous mapping theorem to derive the limiting value of $\widehat{Q}_n(\gamma)$.

$$Q_0(\gamma) \coloneqq g\left(\left|\gamma_0\,\mathrm{erfcx}\left(\frac{\sigma_0}{\gamma_0\sqrt{2}}\right) - \gamma\,\mathrm{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right)\right|\right). \tag{39}$$

Clearly, this equals zero when $\gamma = \gamma_0$. If both $\widehat{Q}_n(\gamma)$ and $Q_0(\gamma)$ are strictly convex, the minimum is unique. Define $A(\sigma,\gamma) \coloneqq \gamma\,\mathrm{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right)$. Showing $A(\sigma,\gamma)$ is strictly increasing for all σ is sufficient to show this convexity because of properties assumed about g and the absolute-value function. Consider

$$\frac{\partial}{\partial\gamma}\gamma\,\mathrm{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right) = \mathrm{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right) - \frac{\sigma}{\gamma^2\sqrt{2}}\frac{\partial}{\partial x}\mathrm{erfcx}(x)\bigg|_{x=\frac{\sigma}{\gamma\sqrt{2}}}. \tag{40}$$

Since erfcx is a positive, decreasing function, the last term is negative, and so the entire equation is strictly positive. This result implies that $\widehat{Q}_n(\gamma)$ and $Q_0(\gamma)$ are both strictly convex as functions of γ, which then implies the minimum given above is strict.

Since $\gamma_0 > 0$, it lies in the interior of a convex set. By Newey and McFadden (1994, Theorem 2.7), $\hat{\gamma}_n$ is a unique minimizer and $\hat{\gamma}_n\xrightarrow{\mathbb{P}}\gamma_0$.

□

Appendix C Simulation Results (For online publication only)

In this section, I report several other simulation results. They are similar to those reported in the paper, i.e., paper's results are quite robust.

C.1 Market Microstructure

The data have substantial market microstructure noise. To mimic its effect, I follow Christensen, Oomen, and Podolskij (2014). They assume we observe $r_{i_n} + u_{i_n}$, here u_{i_n} follows

$$u_{i_n} = \beta u_{i_n - 1} + \epsilon_{i_n}, \quad \epsilon_{i_n} \overset{i.i.d.}{\sim} \mathcal{N}(0, \omega^2(1-\beta^2)). \tag{41}$$

I set $\omega^2 = 1.00 \times 10^{-10}$ because that is the value obtained from the data using the jump robust noise variance bipower-type estimator of Oomen (2006):

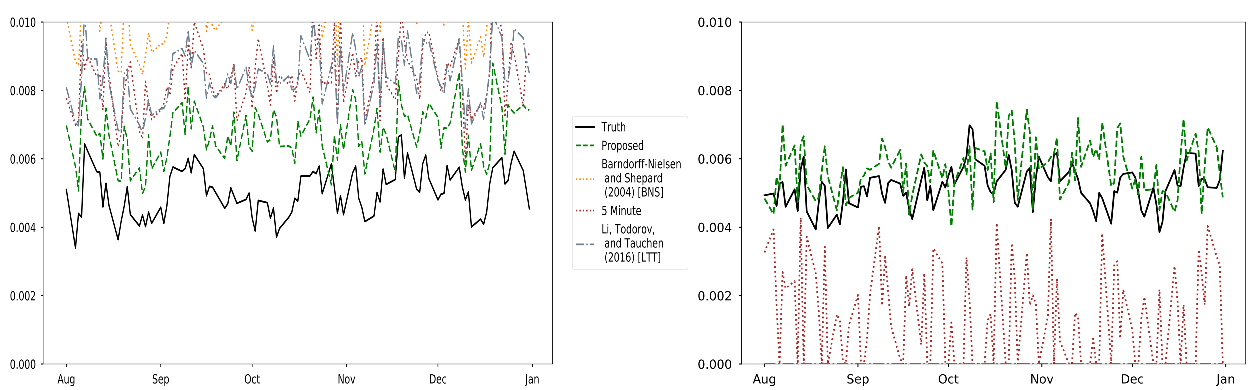
$$\frac{1}{T}\sum_{t=1}^{T-1} \frac{\Delta^n}{2} \sum_{t-1 < i_n, i_n - 1 < t} \left|\Delta_{i_n}^n p\right|\left|\Delta_{i_n - 1}^n p\right|. \tag{42}$$

I set $\beta = 0.77$, which is the value used in Christensen, Oomen, and Podolskij (2014). They set it to match the sign of the S&P 500 futures contract on the day of the 2010 Flash Crash.

I now add the market-microstructure correction described in Section 4. I also set $\theta = 0.5$ (the constant for the pre-averaging correction) and $\bar{\kappa} = 1000$ (the constant for the instantaneous estimator), which are the values used in the actual estimation. I chose these values because they appeared to work well in the simulated data. As we can see in Figure 3, the estimators are slightly biased upwards in this scenario, especially the estimators for σ_t^2.

Figure 3: Discrete-Time Simulation Results with Microstructure

(a) σ_t (b) γ_t



Even though they are slightly biased upwards, the proposed estimators perform reasonably well in practice. This claim does not hold for the other estimators in the literature. In Table 7, I report the mean-square error of the previous estimates averaged over a year's worth of simulations. Here I have approximately 1/2 the average error in estimating σ_t^2 and 1/5 the error in estimating γ_t^2.

Although the jump variation estimators in the literature are not consistent for γ_t^2, they should be asymptotically unbiased. In large finite-samples, they appear both biased and inconsistent.

Table 7: Relative Simulation Error with Microstructure

Obs. per Min.	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$				$\frac{\mathbb{E}[(\widehat{\gamma_t}-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$			
	BNS	LTT	5 Min.	Proposed	BNS	LTT	5 Min.	Proposed
≈ 2	0.74	0.41	0.42	1.00	1.01	1.01	0.83	0.65
≈ 12	0.82	0.46	0.46	0.36	1.01	1.01	0.82	0.41
≈ 60	1.11	0.69	0.69	0.36	1.01	1.01	0.84	0.21
≈ 180	1.58	1.06	1.06	0.85	1.01	1.09	0.81	0.18

C.2 Simulation Error with only a few Jumps

I simulate the volatilities using the DGP described in Table 1. Then instead of simulating the prices using Section 5.1, I follow Huang and Tauchen (2005) and assume the jump locations follow a time-invariant Poisson distribution and the magnitudes are Gaussian distributed. I set the Poisson's intensity to result in an average of one jump per day. I set the variance of the magnitude so that the jump process has the volatility given by γ_t^2. This DGP should be quite difficult for my procedure because there are very few jumps. It drastically violates the infinite-activity assumption. I also add the microstructure noise as described in (41).

Table 8: Relative Simulation Error with Microstructure and Poisson Jumps

Obs. per Min.	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$				$\frac{\mathbb{E}[(\widehat{\gamma_t}-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$			
	BNS	LTT	5 Min.	Proposed	BNS	LTT	5 Min.	Proposed
≈ 2	0.88	0.12	0.20	0.88	1.01	1.01	0.78	0.34
≈ 12	0.95	0.13	0.21	0.51	1.01	1.01	0.79	0.32
≈ 60	1.17	0.32	0.41	0.09	1.01	1.01	0.80	0.39
≈ 180	1.55	0.77	0.85	0.58	1.01	1.01	0.75	0.36

APPENDIX D VOLATILITY: EMPIRICAL RESULTS (FOR ONLINE PUBLICATION ONLY)

Table 9: Univariate Autoregressive Models: AR(BIC)

	$\log(\sigma_t^2)$		$\log(\gamma_t^2)$	
Intercept	−0.68	(−0.88, −0.48)	−0.62	(−0.81, −0.42)
Lag 1	0.54	(0.51, 0.58)	0.46	(0.43, 0.49)
Lag 2	0.15	(0.11, 0.18)	0.17	(0.13, 0.21)
Lag 3	0.06	(0.02, 0.09)	0.05	(0.02, 0.09)
Lag 4	0.07	(0.04, 0.11)	0.08	(0.04, 0.11)
Lag 5	0.04	(0.00, 0.08)	0.09	(0.05, 0.13)
Lag 6	0.00	(−0.03, 0.04)	0.01	(−0.02, 0.05)
Lag 7	−0.00	(−0.04, 0.04)	0.01	(−0.03, 0.05)
Lag 8	−0.00	(−0.04, 0.04)	−0.02	(−0.05, 0.02)
Lag 9	0.08	(0.04, 0.11)	0.08	(0.05, 0.12)
$\mathbb{R}^2$	76 %		74 %	
Innovation Variance	0.31		0.25	

Table 10: Vector Autoregression Models: VAR(BIC)

	$\log(\sigma_t^2)$		$\log(\gamma_t^2)$	
Intercept	−0.33	(−0.54, −0.11)	−0.88	(−0.73, −0.69)
$\log(\sigma_{t-1}^2)$	0.40	(0.36, 0.44)	0.24	(0.24, 0.27)
$\log(\gamma_{t-1}^2)$	0.25	(0.20, 0.29)	0.30	(0.19, 0.34)
$\log\sigma_{t-2}^2$	0.11	(0.07, 0.16)	0.01	(−0.06, 0.04)
$\log\gamma_{t-2}^2$	0.01	(−0.03, 0.06)	0.13	(0.09, 0.17)
$\log(\sigma_{t-3}^2)$	0.05	(0.01, 0.09)	−0.01	(−0.12, 0.03)
$\log(\gamma_{t-3}^2)$	−0.00	(−0.05, 0.04)	0.06	(0.02, 0.10)
$\log\sigma_{t-4}^2$	0.07	(0.02, 0.11)	−0.03	(−0.08, 0.01)
$\log\gamma_{t-4}^2$	0.01	(−0.03, 0.06)	0.11	(0.05, 0.15)
$\log(\sigma_{t-5}^2)$	0.03	(−0.01, 0.07)	−0.02	(−0.04, 0.02)
$\log(\gamma_{t-5}^2)$	0.04	(−0.00, 0.09)	0.14	(−0.01, 0.18)
$\mathbb{R}^2$	76 %		75 %	
Innovation Covariance	$\begin{pmatrix} 0.31 & 0.17 \\ 0.17 & 0.24 \end{pmatrix}$			

Table 11: Log-Volatility Correlations

	$\log(\sigma_t^2)$	$\log\left(\gamma_t^2\right)$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	rx_t	$\mathbf{1}\{\text{FOMC}\}_t$
$\log\left(\sigma_t^2\right)$	1.00	0.90	0.97	−0.50	−0.18	0.06
$\log\left(\gamma_t^2\right)$	0.90	1.00	0.98	−0.08	−0.14	0.09
$\log\left(\sigma_t^2+\gamma_t^2\right)$	0.97	0.98	1.00	−0.29	−0.16	0.08
$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	−0.29	−0.08	−0.29	1.00	0.13	0.04

Appendix E Volatility and Excess Returns (For online publication only)

E.1 Contemporaneous Regression

This section replicates the standard result that volatility and returns are contemporaneously negatively correlated, (Lettau and Ludvigson 2010) and splits it into relationships with $\sigma_t^2 + \gamma_t^2$ and $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$.

Table 12: $\mathbb{E}\left[rx_t \,\middle|\, \sigma_t^2 + \gamma_t^2, \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right]$ (OLS)

Regressors	Specifications			
Intercept	−4.55 [5.81]	1.02 [6.48]	−3.17 [−3.94]	−1.27 [−0.54]
$\log\left(\sigma_t^2 + \gamma_t^2\right)$	−0.46 [−5.58]		−0.39 [−4.12]	−0.19 [−0.85]
$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$		1.65 [5.81]	1.13 [4.06]	3.91 [1.07]
$\log\left(\sigma_t^2 + \gamma_t^2\right)\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$				0.29 [0.80]
$\bar{\mathbb{R}}^2$	2.67 %	1.61 %	3.35 %	3.42 %

E.2 Risk Premia Estimates: Additional Results

Estimating risk premia is difficult because the signal-to-noise ratio is quite low. However, because the regressions are run at the daily level and there are 3700 datapoints, most of the potential biases are not relevant. For example, the Stambaugh (1999) bias declines at $\frac{1}{\#\text{ datapoints}}$ rate and so should not noticeably affect my estimates.

I also consider several other specification below. The volatility coefficients are robust to the particular instruments chosen (Table 14). Results from running the regression over a subsamples agree with the main results, but are not always statistically significant (Table 15). Estimates in levels and unweighted estimates either agree with the main results or are not statistically significant (Table 15 and Table 13).

I am using estimated regressors, and so we might worry about the generated regressors problem. However, because I have a lot of intra-day data, the regressors should be estimated relatively precisely. Furthermore, the error that still exists mostly arises from the difficulty in disentangling $p^J t$) and $p^D(t)$. The deconvolution procedure depends upon the magnitude of the increments relative to their standard deviation, not their sign. Consequently, it should be approximately independent across time, and so the instruments used above are an asymptotically valid way of dealing with this type of measurement error.

Table 13: News Premia Estimates in Levels

(Volatility is measured in yearly terms. (252 * daily)).
$l \in \{1, 2, 5, 25\}$.

Regressors				Instruments			
Intercept	σ_t^2	γ_t^2	$(\sigma_t^2)(\gamma_t^2)$	$\mathbf{1}\{\text{FOMC}\}_t$	$\sigma_{t-l}^2 \dots$	$\gamma_{t-l}^2 \dots$	$(\sigma_{t-l}^2)(\gamma_{t-l}^2)$
0.28	0.08				✓		
[8.85]	[3.11]						
0.28	0.08			✓	✓	✓	✓
[8.80]	[2.74]						
0.27		0.07		✓			
[7.52]		[2.53]					
0.24		0.10		✓	✓	✓	✓
[6.63]		[3.33]					
0.31	0.16	−0.09			✓	✓	
[7.25]	[1.35]	[−0.77]					
0.24	0.02	0.09		✓	✓	✓	✓
[5.57]	[0.22]	[1.05]					
0.23	0.36	−0.50	−0.00	✓	✓	✓	✓
[5.75]	[1.39]	[−1.14]	[1.53]				

Table 14: News Premia Estimates: Other Instruments

$$\psi_t := \log\left(\sigma_t^2 + \gamma_t^2\right),\ \phi_t := \log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$$

Regressors				Instruments			
Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	ψ_{t-1}	ϕ_{t-1}	$\mathbf{1}\{\text{FOMC}\}_t$	$\phi_{t-l}\ldots$	$\psi_{t-l}\ldots$	$\psi_{t-l}\phi_{t-l}\ldots$
				$l \in \{1, 2, 5, 25\}$			
3.05		0.25			✓		
[6.83]		[6.07]					
3.01		0.25		✓	✓	✓	✓
[6.03]		[6.77]					
−2.02			−4.21			✓	
[−3.99]			[−4.63]				
−2.05			−4.28	✓	✓	✓	✓
[−5.28]			[−6.20]				
0.25		0.17	−3.47	✓	✓	✓	✓
[0.36]		[3.83]	[−5.01]				
0.11	0.25	0.16	−3.57	✓	✓	✓	✓
[0.14]	[1.24]	[3.59]	[−5.02]				
				$l = 1$			
3.10		0.25		✓			
[6.85]		[6.10]					
−2.71			−5.44			✓	
[−2.86]			[−3.20]				
−1.43		0.11	−5.22	✓	✓		
[−0.84]		[1.35]	[−2.94]				
−1.03	0.29	0.12	−4.76	✓	✓	✓	✓
[−0.76]	[1.36]	[1.68]	[−3.56]				

Table 15: News Premia Estimates: Robustness

	Regressors				Instruments			
	Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log(\sigma_t^2+\gamma_t^2)$	$\log(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2})$	$\mathbf{1}\{\text{FOMC}\}_t$	$\log(\frac{\gamma_{t-l}^2}{\sigma_{t-l}^2+\gamma_{t-l}^2})\dots$	$\log(\sigma_{t-l}^2+\gamma_{t-l}^2)\dots$	$\log(\frac{\gamma_{t-l}^2}{\sigma_{t-l}^2+\gamma_{t-l}^2})\cdot \log(\sigma_{t-l}^2+\gamma_{t-l}^2)$
Sub-period Analysis								
2003–2007	−2.59	0.69	0.08	−6.93	✓	✓	✓	✓
	[−0.67]	[1.58]	[0.36]	[−2.05]				
2008–2012	0.17	0.79	0.06	−1.56	✓	✓	✓	✓
	[0.10]	[1.82]	[0.55]	[−1.33]				
2013–2007/9	3.71	−0.26	0.40	−1.94	✓	✓	✓	✓
	[2.50]	[−1.05]	[4.40]	[−1.66]				
Unweighted Analysis								
	0.63		0.06		✓	✓	✓	✓
	[0.82]		[0.77]					
	0.03			−0.04	✓	✓	✓	✓
	[0.06]			[−0.04]				
	−0.77		−0.02	−1.14	✓	✓	✓	✓
	[−0.62]		[−0.17]	[−1.50]				
	−1.42	0.93	−0.09	−0.81	✓	✓	✓	✓
	[−1.13]	[3.02]	[−0.99]	[−0.89]				

Table 16: Instrument Variables: First Stage Regression

$$\psi_t \coloneqq \log\left(\sigma_t^2 + \gamma_t^2\right),\ \phi_t \coloneqq \log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$$

Regressand	Intercept	ϕ_{t-1}	ϕ_{t-2}	ϕ_{t-5}	ϕ_{t-25}	ψ_{t-1}	ψ_{t-2}	ψ_{t-5}	ψ_{t-25}	$\psi_{t-1}\phi_{t-1}$	$\bar{\mathbb{R}}^2$	$\widehat{F}$
	−0.44	0.26									6.58 %	62.2
	[−25.84]	[7.88]										
$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	−0.28	0.18	0.16	0.12	0.06						11.53 %	110.0
	[−9.48]	[8.95]	[7.14]	[6.37]	[3.39]							
	−0.61	0.14	0.13	0.10	0.07	−0.06	−0.00	0.01	0.02		14.63 %	248.1
	[−8.39]	[8.38]	[6.86]	[5.51]	[4.07]	[−6.89]	[−0.25]	[1.91]	[3.14]			
	−0.23	0.73	0.11	0.10	0.07	−0.02	−0.00	0.01	0.02	0.06	15.49 %	525.4
	[−2.35]	[7.05]	[6.51]	[5.43]	[3.90]	[−1.51]	[−0.07]	[1.87]	[3.34]	[5.73]		
	−2.10					0.19					66.28 %	1986.4
	[−10.85]					[44.57]						
$\log\left(\sigma_t^2 + \gamma_t^2\right)$	−0.57					0.61	0.17	0.13	0.04		79.19 %	7712.2
	[−4.94]					[26.42]	[8.96]	[8.55]	[4.07]			
	−0.59	−0.15	0.00	0.07	0.06	0.60	0.16	0.13	0.05		79.27 %	9517.4
	[−4.99]	[−3.47]	[0.11]	[1.79]	[1.65]	[27.95]	[8.80]	[8.92]	[4.40]			
	−1.31	−1.23	0.02	0.07	0.07	0.53	0.16	0.13	0.05	−0.11	79.43 %	20 140
	[−5.84]	[−5.05]	[0.63]	[1.91]	[1.92]	[18.91]	[8.70]	[8.92]	[4.70]	[−4.43]		