Jumps, Realized Densities, and News Premia [*]

Paul Sangrey

University of Pennsylvania [†]

Job Market Paper

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, and risk premia are instantaneous covariances. 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 and associated risk premia. 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, Recursive Utility, Stochastic Volatility, Nonparametric Modeling, Semimartingales, Time Aggregation, Risk Premia, Asset Pricing

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

1. Introduction

The study of how individuals' react 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 fifteen 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 (that is, they are stochastic volatility diffusions) the diffusion volatility entirely determines the 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. Another series of classic papers shows that the instantaneous covariance between prices and investors' stochastic discount factors determine risk premia, (Merton 1973; Breeden 1979; Bollerslev, Engle, and Wooldridge 1988a).

However, hundreds of quantitatively relevant news releases strike financial markets every day and cause the prices to jump. Aït-Sahalia and Jacod (2009a, 2009b, 2012) even show that models with infinitely many jumps fit the data better than models with only finitely many jumps. Meanwhile, various papers, such as Drechsler and Yaron (2011) and Ai and Bansal (2018), show the parsimonious covariance-based characterizations of risk premia mentioned above fail when prices jump.

At present, however, no parsimonious representation with nonparametrically identified dynamics exists for jump processes. 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 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 representations for the discrete-time densities and volatilities.

To enable taking this theory to the data, I develop an estimator for the instantaneous diffusion volatility by extending Jacod et al. (2009). I identify the jump part of the dynamics, in the presence of stochastic diffusion volatility by deriving the first estimator for instantaneous jump volatility. I time-aggregate both estimators to provide estimators for the daily diffusion and jump volatilities. I then apply these estimators to high-frequency data on the S&P 500. This provides several new stylized facts. First, diffusion and jump volatility are highly positively correlated. Second, like diffusion volatility, jump volatility is highly persistent, remaining high for extended periods of time during recessions.

I then connect jump volatility to consumption-based asset pricing by nonparametrically characterizing continuous-time risk-premia in the presence of recursive utility and jumps. My characterization shows how jump and diffusion volatility jointly determine risk premia and requires both terms in general. I then take my estimators to the data and show that the diffusion volatility commands an economically and statistically significant premium, as in Brandt and Kang (2004) and Lettau and Ludvigson (2010). I further show that the jump volatility is substantially less than the diffusion volatility premium. I show that this implies that investor's preferences are not time-separable and that we need at least two factors that move at high-frequency to explain movements in risk premia.

I lay out the paper as follows. The remainder of the introduction fixes ideas and explains the close connection between discontinuous information flows and jumps in asset prices. Section 2 relates my paper to the rest of the literature. Section 3 lays out the data generating process I use, while Section 4 proves the main representation theorem. Section 5 derives the estimators, and Section 6 characterizes their finite-sample performance in simulations. Section 7 describes my dataset, and Section 8 provides a series of new stylized facts concerning the jump volatility dynamics. I derive risk premia in the presence of recursive utility and jumps in Section 9 and show that the jump volatility premium is less than the diffusion volatility premium in Section 10. Section 11 concludes. The appendices contain the proofs and robustness checks.

1.1. Stylized Features of the Data

I motivated this project by claiming that prices jump extremely often and that news frequently and dramatically affect asset prices. The literature has shown this, but it is helpful to investigate the matter ourselves to fix ideas. We need high-frequency data to identify these jumps, and so I start there. The data show jumps in price processes are ubiquitous and form a large portion of the price's variation. For example in Figure 1, I plot the daily log-return on the S&P 500 during 2012 and then zoom in on the 1-second return on April 16. The red lines are jumps in the prices identified by sampling the data once per second, and the blue lines contain the diffusion part of the process and jumps that are too small to identify easily. The behavior in this graph is entirely typical. I purposefully chose April 16, 2012, because it was a completely normal day in the markets.

As we can see in Figure 1, prices jump extremely often and drive a great deal of the variation in the price. Estimates range from as low as $\approx 7\%$ to as high as $\approx 80\%$, (Pan 2002; Huang and Tauchen 2005; Santa-Clara and Yan 2010; Ornthanalia 2014). In particular, 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. This wide divergence between various estimates likely arises from the difficulty in disentangling the infinite-activity jumps from the diffusive part. The precise percentage is not important for this paper. I estimate this proportion below, (Figure 7). Rather, the important takeaway is that jumps occur frequently enough to be important, and even 7% of the variation in the market is economically meaningful.

Almost every paper that explicitly tests for the degree of activity finds infinitely-active jumps,

Figure 1: S&P 500 Log-Return

(a) 1 Second   (b) Daily
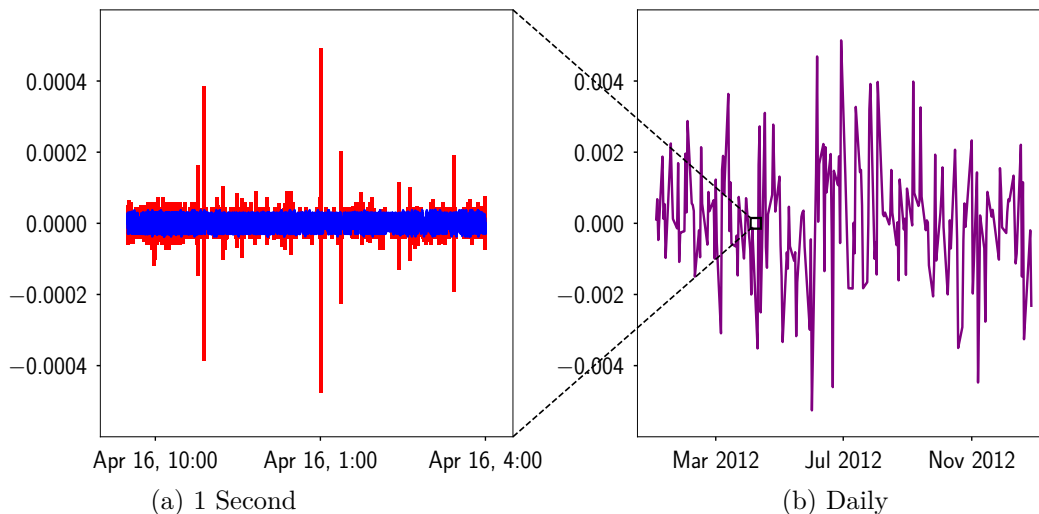
or at the very minimum a massive number, (Aït-Sahalia, Mykland, and Zhang 2005; Bakshi, Carr, and Wu 2008; Aït-Sahalia and Jacod 2009a).[1] From both a modeling and pricing perspective, a large number of jumps and infinitely many are essentially equivalent in practice, as shown in detail below. Even if the literature has not reached a consensus on the number and magnitude of the jumps, it is clear that jumps are ubiquitous and crucial to understanding price dynamics.

1.2. What Causes Jumps?

To understand Figure 1a, we need to understand what precisely a jump is. There are two equivalent characterizations. First, a jump is a discontinuity in the price process. The price changes by such a large amount over such a small period that we cannot draw a continuous line through it. However, this is a mathematical definition; we would like an economic characterization. What are jumps economically?

Various authors, such as Andersen, Bollerslev, Diebold, and Vega (2003, 2007), Beechey and Wright (2009), and Lahaye, Laurent, and Neely (2011), argue that jumps are responses of prices to news releases. Most of these papers consider the effects of macroeconomic announcements on prices. They start with a series of news items that they a priori believe to be important and show that the prices react effectively instantaneously.[2] However, in general, 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 point highlights the utter impossibility of listing all the potentially relevant events. We cannot construct investors' actual information sets. (Note, this paper uses *news* quite broadly.

---

1. The single exception is Christensen, Oomen, and Podolskij (2014), which I discuss in Section 8.3.
2. By far the most commonly studied announcements are the Federal Open Market (FOMC) announcements.

It refers to any discontinuous change in information, not just traditional news sources such as newspapers.) As these examples illustrate, news often come at unpredictable times and only a few investors may observe it, 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 the prices.

The connection between news and jumps is rather intuitive, and the empirics in the papers mentioned substantiate it. However, the connection is even more fundamental. Delbaen and Schachermayer ([1994](#)) show no-arbitrage implies prices are semimartingales.[3] In that framework, which is standard in high-frequency econometrics, jump times are times when the information contained in prices jumps. In other words, jump times are times when the representative investor's information set evolves discontinuously.

To make this claim precise, consider the following. Let $P(t)$ be a price process, and $\mathcal{F}_t^p$ be its natural filtration.[4] $\mathcal{F}_t^p$ contains the events that are known at time t to anyone observing the history of prices up to t. In other words, it is the part of the representative investor's information set relevant for pricing. Then, $P(t)$ jumps at τ if and only if $\mathcal{F}_t^p$ jumps at τ. Since standard economic intuition implies that causality runs from information to prices, $P(t)$ jumps whenever the available information evolves discontinuously, that is a news item is released. This relationship implies that we can identify news shocks by looking for jumps in the prices. Consequently, since the jump volatility is a sufficient statistic for jumps dynamics, it measures news risk.

Theorem 1 (Jump Times are News Times). *Consider a stopping time τ. Let $P(t)$ be a price process satisfying no-arbitrage. Then its natural filtration — $\mathcal{F}_t^p$ — contains all of the information in the representative investor's information set relevant for asset pricing, and $\mathcal{F}_\tau^p \neq \mathcal{F}_{\tau-}^p$ if and only if $P(t)$ jumps at τ, where $\mathcal{F}_{t-}^p$ is the associated predictable filtration.*

This result also explains why not all price changes are jumps. Prices do not always reflect new information 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 results reveal 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 buy or sell themselves. This process changes the asset's price, and it takes time.

2. Literature Review

Since questions concerning volatility, news, and risk-return trade-offs are central to finance and economics, a few different literatures study the questions considered in this paper. Consequently,

---

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

4. This paper uses functional notation to refer stochastic processes and subscript notation to refer to discrete-time objects, e.g., $P(t)$ is the price process, and P_t is the price at t. I time-index objects by the first time they enter the representative investor's information set.

I cannot hope to survey the literature adequately. I can only cover a few of the closest related papers.

2.1. Jumps in Asset Prices

The first literature that I build upon is the econometrics literature that studies jumps in asset prices. Barndorff-Nielsen and Shephard (2006) develop the bipower variation estimator to disentangle jumps and diffusive variation. Since then, several authors have shown that jumps are both frequent and economically important, including Andersen, Bollerslev, and Diebold (2007), Bollerslev, Law, and Tauchen (2008), and Aït-Sahalia and Jacod (2009b). The critical difference between my estimates of jump variation and previous bipower variation estimates is that I measure ex-ante jump variation, while previous papers measure ex-post variation. This difference is essential for two reasons. First, my density characterization relies upon an ex-ante characterization. Second, investors price ex-ante risk, and so my measure is a core object in pricing, while ex-post jump variation cannot be priced. Other authors have argued they are not just statistically significant, but economically as well. For example, we also need them to price derivatives, such as (Pan 2002; Branger, Schlag, and Schneider 2008; Todorov 2010, 2011).

In Section 1.1, I discuss the literature that measures the magnitude of jump variation and the jump intensities. I will not repeat that discussion here except to recall the twofold consensus. First, asset prices contain a vast number of jumps. Jumps are likely infinitely-active, or, at a minimum, have a very high intensity. Second, jumps constitute an economically and statistically significant portion of the price variation.

I rely on these results in three ways. First, as motivation for the project. Second, as evidence that my empirical results are reasonable. Third, and most importantly, I rely heavily on these empirical facts in that I assume that prices have infinitely-active jumps. This assumption is somewhat unusual, but not unique. For example, Gallant and Tauchen (2018) considers a similar class of processes.

Gallant and Tauchen (2018) is arguably the closest related paper in the econometrics literature. It is the only other paper that nonparametrically relates jump variation to the distribution of returns. It is a fascinating paper and provides useful estimates for the intensity of jump processes. However, their representation relies on Todorov and Tauchen (2014) and so can only handle small jumps.

2.2. Representing Price Processes

The second literature that this paper builds upon is the stochastic process representation literature. The main contribution to this literature is Theorem 4 and its corollaries. This theorem provides general conditions under which jump processes are stochastic volatility variance-gamma processes. The variance-gamma process is a Lévy process first introduced by Madan, Carr, and Chang (1998).

The first main time-change method for representing price processes is the Dambis, Dubins & Schwarz theorem, (Dambis 1965; Dubins and Schwarz 1965). Theorem 4 is the jump analog of

that theorem. Epps and Epps (1976) and various subsequent authors relate this time-change to "business-time," that is the speed at which information gets released into the market, creating the mixture-of-distributions hypothesis.

Various authors partially extend these results to the jump case. Monroe (1978) shows that any semimartingale can be embedded into Brownian motion, but did not construct this embedding explicitly. Geman, Madan, and Yor (2002) shows that this embedding is not identified. More recently, Todorov and Tauchen (2014) show how to embed the jump processes' infinitesimal jumps into an α-stable process using bipower-variation. Infinitesimal means, here, that the maximum jump size approaches zero as the increment size approaches zero. By using an ex-ante measure of jump variation, instead of an ex-post one like Todorov and Tauchen (2014) do, I can handle large jumps as well.

In Section 4.4, I time-aggregate these continuous-time representations to discrete-time under some additional assumptions. In doing this, I follow Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) who provide analogous results for diffusive processes. Barndorff-Nielsen and Shiryaev (2010) further analyze these representations, providing a useful survey of the current state of the literature.

2.3. Pricing Assets with Recursive Utility

The curvature in investors' preferences, i.e., their risk appetite, implies a negative relationship between expected returns and volatility. Consequently, many different papers estimate this relationship, and I cannot comprehensively survey this literature. Surprisingly, the empirical evidence has proven much less conclusive than the theory. Bollerslev, Engle, and Wooldridge (1988b), Harvey (1989), Ghysels, Santa-Clara, and Valkanov (2005), and Lettau and Ludvigson (2010) find a positive relationship between expected returns and volatility. Campbell (1987), Pagan and Hong (1991), Glosten, Jagannathan, and Runkle (1993), and Brandt and Kang (2004) actually find a negative relationship. Besides, many authors argue that the instantaneous correlation, which is often called a "volatility-feedback" or leverage effect is negative, both in continuous-time (Bandi and Renò 2012; Aït-Sahalia, Fan, and Li 2013) and in discrete-time (Engle and Ng 1993; Yu 2005). This negative sign is likely the main reason why estimating the risk premium has proven difficult. The researcher must disentangle two different relationships, risk-premia and volatility feedback, with opposite signs.

Investors' utility functions are not the only place their preferences can display curvature. Some examples of models with curvature in their certainty equivalence functionals (CEF) include max-min expected utility, (Gilboa and Schmeidler 1989; Epstein and Schneider 2003), models with ambiguity aversion (Hansen and Sargent 2001; Klibanoff, Marinacci, and Mukerji 2005; Ju and Miao 2012), and Epstein-Zin recursive utility (Epstein and Zin 1989; Duffie and Epstein 1992). This additional curvature leads to additional risk-return trade-offs. Ai and Bansal (2018) show premia for this curvature cause announcements to be priced differently. Hence, simple covariance-based explanations for risk-premia break down.

Arguably the closest related paper in the finance literature, Ai and Bansal (2018), is inspired by a recent surprising stylized fact presented by Lucca and Moench (2015): the majority of the equity premium occurs on the days around when the Federal Open Market Committee (FOMC) makes it announcements. This paper extends Ai and Bansal (2018) by deriving risk-premia in continuous-time for models with recursive utility and jumps. I then show that this additional term is closely related to the jump volatility. This characterization shows that, in general, the theory requires two pricing factors which move at high-frequency.

3. Data Generating Process

In this section, I describe the data generating process (DGP). Models of prices differ along two different primary 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 discrete-time representation from it. I also discuss the purely continuous DGP that my DGP nests in parallel to provide a point of comparison.

3.1. 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

$$dp(t) = \mu(t)\,dt + \sigma(t)\,dW(t), \tag{1}$$

where $W(t)$ is a Wiener process. However, as mentioned in the introduction, asset prices are not continuous processes, and so the models considered above cannot fully replicate the stylized facts in the data. For example, such a process does not have fat-tailed distributions once you condition on the volatility. This is because $W(t)$ is a Wiener process, and so conditionally on $\sigma^2(t)$ its increments are i.i.d. Gaussian random variables.

Over the last few decades, various researchers have worked very hard to add jumps to these models. The standard nonparametric way of doing 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 that it is a predictable function that multiplies the jumps. In general, the triple (δ, n, ν) is not unique, which allow us to pick a particularly useful representation later.

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

$$p(t) = p(0) + \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \int_0^t \int_X \delta(s,x)1\{\|\delta(x,s)\| \le 1\}(n-\nu)(ds,dx) \quad (2)$$
$$+ \int_0^t \int_X \delta(s,x)1\{\|\delta(x,s)\| > 1\}n(ds,dx)$$

To simplify Definition 1, I add an additional assumption:

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

Assumption 1 is relatively innocuous in practice since an easy to verify sufficient condition is that the returns themselves always have a conditional variance. Assumption 1 implies that the jump measure has a predictable compensator. 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 1 when they derive estimators. By assuming it now, I can simplify notation. I also assume without loss of generality that $p(0) = 0$, giving

$$p(t) = \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \int_0^t \int_X \delta(s,x)(n-\nu)(ds,dx). \quad (3)$$

I now assume without loss of generality that n is a standard Poisson random measure. In other words, for each open set $A \subset X$, $\int_A \delta(t,x)\,dx$ is the intensity of the Poisson process with magnitude $x \in A$. Hence, δ completely controls the process's dynamics.

This representation is quite general and can handle a great variety of different price processes. However, it is rather intractable, and not identified. For each time t, $\delta(t,\cdot)$ is a function of x. In other words, for each t, we must estimate an infinite-dimensional object using at most one realization. Also, it is not obvious how to time-aggregate this representation, i.e., parsimoniously map it to discrete-time.

3.2. 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.[5] Throughout, I use subscripts to refer to daily objects, and functional notation to refer to stochastic processes, I also adopt the convention that the time associated with a variable is when it first becomes known to the investor, i.e., measurable with respect to 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.

Definition 2. Daily Return

$$r_t := \int_{t-1}^t dp(t). \quad (4)$$

---

5. 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.

This return has a density — h — in each period given the available information at the end of the day before — $\mathcal{F}_{t-1}$.

Definition 3. Daily Density

$$r_t \mid \mathcal{F}_{t-1} \sim h\left(r_t \mid \mathcal{F}_{t-1}\right). \tag{5}$$

This predictive density fully characterizes the 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. As a consequence, this density is a primary object of interest in financial econometrics, and economists have written thousands of papers modeling it and its statistics.

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 $h\left(r_t \mid \mathcal{F}_{t-1}\right)$. Since $\mathcal{F}_{t-1}$ grows each day, $h\left(r_t \mid \mathcal{F}_{t-1}\right)$ is a function-valued time-varying parameter. Modeling such parameters is quite difficult, and so the literature has focused on representations for $h\left(r_t \mid \mathcal{F}_{t-1}\right)$ in terms of a well-behaved sufficient statistic for the dynamics, which I denote x_t, e.g., Engle (1982), Bollerslev (1986), and Nelson (1991). The most common choice for x_t is some measure of volatility.

They use x_t to separate $h\left(r_t \mid \mathcal{F}_{t-1}\right)$ 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 with some 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

$$r_t \mid \mathcal{F}_{t-1} \sim h\left(r_t \mid \mathcal{F}_{t-1}\right) = \int_{x_t} f\left(r_t \mid x_t\right) dG\left(x_t \mid \mathcal{F}_{t-1}\right), \tag{6}$$

replacing the question how should we model $h\left(r_t \mid \mathcal{F}_{t-1}\right)$ 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. As is standard, it uses volatility σ_t^2 as x_t. Here the return is a Gaussian innovation with stochastic volatility — σ_t^2, and so f is a Gaussian distribution. The σ_t^2 follows an $AR(1)$ process in logs with persistence ρ and innovation variance σ_σ^2.

Definition 4 (Simple Stochastic Volatilty Model)**.**

$$r_t \sim \sigma_t N(0,1) \tag{7}$$

$$\log\left(\sigma_t^2\right) = \rho \log\left(\sigma_{t-1}^2\right) + \sigma_\sigma N(0,1) \tag{8}$$

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

Definition 5 (Realized Density)**.**

$$RD_t := f\left(r_t \mid x\right)\Big|_{x=x_\tau} \tag{9}$$

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 Definition 4, the realized density is $f(r_t \mid x_t) = f(r_t \mid \sigma_t^2)$.

The realized density is useful because it separates the dynamic and static parts of the process. I provide conditions under which high-frequency data identifies RD_t data by providing an estimator for it in Section 5. Once we have RD_t, we only need to model G. In practice, this is much simpler than modeling $h(r_t \mid \mathcal{F}_{t-1})$ directly because x_t is usually well-behaved.

4. Modeling Jump Processes

In the previous section, I claimed that the most common choice for a sufficient statistic for the dynamics is some measure of volatility. Moving forward, I construct a new measure of volatility. This measure, unlike various realized measures in the literature, is an ex-ante measure. This distinction is fundamental to representation I construct below.

4.1. Jump Volatility

In the continuous-time data generating process of (1), I implicitly defined the instantaneous diffusion volatility $\sigma^2(t)$. It is the integrand in that representation. However, there is an equivalent representation going back as far as Merton (1973) that is more useful for our purposes. This representation gives $\sigma^2(t)$ its interpretation as an instantaneous variance; $\sigma^2(t)$ is the appropriately standardized variance of the diffusion part of the process over a shrinking interval. (I use superscript D to refer to the diffusion part of the process.)

Definition 6 (Instantaneous Diffusion Volatility)**.**

$$\sigma_t^2 := \frac{1}{\Delta} \mathbb{E}\left[\left| p^D(t+\Delta) - p^D(t) \right|^2 \,\Big|\, \mathcal{F}_{t-} \right] \tag{10}$$

One key subtlety of this definition is that we are only using the information available before time t. Variances are forward-looking operators. This subtlety is not essential in the diffusion case, and so the literature has not stressed it. In the jump case, however, it is fundamental.

The key beneficial feature of volatility is that it time-aggregates in a straightforward way. Intuitively, the daily volatility is just the integral of the high-frequency volatility.

Definition 7 (Integrated Diffusion Volatility)**.**

$$\sigma_t^2 := \int_{t-1}^{t} \sigma^2(t)\, ds. \tag{11}$$

This aggregation property is precisely what Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) use to develop the Realized Volatility estimator for σ_t^2. 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* — $\gamma^2(t)$. A volatility is a variance, and so we can construct the jump analogue to Definition 6. I substitute the diffusion part of the prices — $p^D(t)$ with the jump part — $p^J(t)$. In other words, I define the instantaneous jump volatility as the local variance of the jump part — $p^J(t)$.

Definition 8 (Instantaneous Jump Volatility)**.**

$$\gamma^2(t) := \frac{1}{\Delta}\mathbb{E}\left[\left|p^J(t+\Delta) - p^J(t)\right|^2 \,\Big|\, \mathcal{F}_{t-}\right]. \tag{12}$$

The integrated jump volatility is defined in the obvious way.

Definition 9 (Integrated Jump Volatility)**.**

$$\gamma_t^2 := \int_{t-1}^t \gamma^2(s)\, ds. \tag{13}$$

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.

Theorem 2 (Jump Volatility and the Predictable Quadratic Variation)**.** *Let $p(t)$ be an Itô semi-martingale satisfying Assumption Square-Integrable, then the following holds where $\langle p^J\rangle(t)$ is the predictable quadratic variation (angle-bracket) of $p^J(t)$:*

$$\gamma_t^2 = \int_{t-1}^t \gamma^2(s)\, ds = \int_{t-1}^t \int_{\mathcal{X}} \delta^2(s,x)\nu(dx,ds) = \langle p^J\rangle(t) - \langle p^J\rangle(t-1). \tag{14}$$

The are three main advantages of $\gamma^2(t)$ and 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 derivative of the quadratic variation like we can differentiate the predictable quadratic variation. Second, by conditioning on $\gamma^2(t)$, I construct a closed-form nonparametric continuous-time representation for $p(t)$ in Section 4.3. I do this without any truncation, Todorov and Tauchen (2014) need to truncate all of the jumps above a shrinking threshold in order to derive their results while using an ex-post measure. Third, as I show in Section 9, $\gamma^2(t)$ controls risk premia. This result is intuitive because risk-premia are ex-ante objects. As a final advantage, both $\gamma^2(t)$ and γ_t^2 are identified. I show this by constructing consistent estimators for them in Section 5.

4.2. Static Jump Processes (Variance-Gamma Process)

In the next section, I construct the static model that my model reduces to when there are no dynamics. It will also be the integrator in the general case. I start with a simple jump process where the locations of the jumps are Poisson distributed, and the magnitudes are i.i.d. Gaussian variables and then take limits to construct the general case.

Define $N(t)$ as the process that determines when $p^J(t)$ jumps, i.e., $N(t) - N(t-) = 1$ if and only if $p(t)$ jumps at time t.

Definition 10. Location Process

$$N(t) := \sum_{s \le t} \mathbf{1} \left\{ \left| p^J(t) - p^J(t-) \right| > 0 \right\} \tag{15}$$

Let $\kappa(t) := \left\{ p^D(t) \,\middle|\, N(t) \neq N(t-) \right\}$ be a process that controls the jump magnitudes. Note, $\kappa(t)$ is not a Wiener process, as its variance does not depend on the length of the interval. It is just an ordered collection of $N(0,1)$ random variables, one for each t. In this case, the jump part of the price process has the following relatively simple form:

$$p^J(t) = \sum_{s \le t} \kappa(s)|N(s) - N(s-)|. \tag{16}$$

The variability in (16) arises from two places: the number of jumps and their magnitudes. Since we are in a time series context, the number of jumps before some time t and their locations contain the same information. In this case, we can rewrite the jump volatility as follows using the law of iterated expectations:

$$\gamma_t^2 = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}\left[|p_{t+\Delta} - p_t|^2 \middle| \mathcal{F}_t \right] = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}_{N(t+\Delta)-N(t)} \left[\sum_{i=1}^{N(t+\Delta)-N(t)} \mathbb{V}\text{ar}(\kappa(t)) \middle| \mathcal{F}_t, N(t+\Delta) - N(t) \right]. \tag{17}$$

We can then use (16) to simplify this expression.

$$\gamma_t^2 = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}[N(t+\Delta) - N(t)] \mathbb{E}[\kappa(t)^2] = \frac{1}{\Delta} \Delta = 1. \tag{18}$$

To put (18) into words, the variance of the jump process is the mean of the intensity multiplied by the variance of the magnitude and rescaled appropriately. This characterization holds in general if the intensities and magnitudes are independent. I am combining the variation from the jump locations and the jump magnitude into a single parameter. If we change either the intensity of the jumps or the variance of the jump magnitudes, the variance of $p^J(t)$ changes in precisely the same way. This irrelevance is useful because the intensity and magnitude functions are not identified. The jump volatility, on the other hand, is.

One obvious generalization of (16) is the compound Poisson process. These processes are comprised of the sum of finitely-many independent Poisson processes with different intensities/magnitudes. This representation does not work for our purposes. The variation I use to identify the jump volatility comes from the data's infinitely-active jumps. A compound Poisson process can have at most finitely many jumps in any finite interval because $N(t)$ must converge.

However, this problem no longer affects our results if we consider the limiting case when $N(t) \to$

∞, We must model $\kappa(t)$ properly for the $p(t)$ to remain square-integrable.[6] In particular, we cannot have infinitely jumps of magnitudes greater than any fixed $\epsilon > 0$ in a finite interval otherwise the jump will diverge. Consequently, we must shrink the size of the increments towards zero as we let $N(t) \to \infty$.

One common pure-jump process — the variance-gamma process — is an infinite-activity "compound Poisson process" with arbitrarily small Gaussian-distributed summands. My model for the jumps reduces to this if it does not have any dynamics. The option pricing literature often uses the variance-gamma process in its models. For example, European option prices are available in closed form, (Madan, Carr, and Chang 1998).

A gamma process — $\Gamma(t)$ — is a process with gamma distributed increments, and a variance-gamma process is a Wiener process time-changed (subordinated) by a gamma process. Equivalently, a gamma process is a pure-jump Lévy process where the jumps that lie in an interval $[x, x + \Delta x)$ are Poisson distributed with intensity $x^{-1} \exp(-x)\Delta x$ for any x and Δ small.

Definition 11 (Variance-Gamma Process)**.**

$$\text{Variance-Gamma}(t) \coloneqq W(\Gamma(t)) \tag{19}$$

Throughout this paper, I exclusively use the standard variance-gamma process, which is the variance-gamma process whose increments are mean zero with all the scale parameters equal to one.[7] The exponential distribution is a special case of the Gamma distribution. (Take a Gamma random variable and set all of its scale parameters equal to one.) If we consider the special case of a standard Wiener process time-changed by a gamma process with rate=1 exponentially-distributed increments, we get a standard variance-gamma process. I use the symbol $\mathcal{L}(t)$ to refer to the standard variance-gamma process because the increments of this process are Laplace random variables.

In order to understand why the increments are Laplace-distributed, consider the following characterization of a standard variance-gamma process. A Laplace distribution is as a Gaussian distribution with random variance, where the random variable is exponentially distributed. The $\sqrt{2}$ in the expression is an adjustment to convert the standard deviation into a scale parameter. In other words, we have the following characterization of the Laplace density:

$$z \sim \mathcal{L}(\text{mean} = 0, \text{variance} = 1) \iff z \sim \frac{\sigma}{\sqrt{2}} N(0, 1), \; \sigma^2 \sim \exp(1). \tag{20}$$

This is the discrete analogue of Definition 11. Each increment of a variance-gamma process has two sources of variation: the number of jumps, which "is" exponentially distributed, and the magnitudes, which are Gaussian distributed. This characterization is not quite accurate because exponential random variables are real-valued, not integer-valued. The number of jumps cannot be

---

6. Just letting $p(t)$ be an ordered collection of $N(0, 1)$ variables does not work.

7. I introduce the notion of a standard variance-gamma process here to facilitate exposition because it aggregates in ways that the general case does not.

exponentially distributed.

However, one way to characterize a Poisson process is as a process where the waiting time between jumps is exponentially distributed. This characterization is well-defined because the intervals' lengths take positive real values, and it returns us to the initial discussion of a variance-gamma process as a "compound Poisson" process with an infinite intensity.

4.3. Jump Process Representation Theorem

4.3.1. Itô Semimartingales

Having defined the DGP, I state the first primary result. Recall the simplified Grigelionis form of the semimartingale, (3):

$$p(t) = \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \int_0^t \int_X \delta(s,x)(n-\nu)(ds,dx). \tag{21}$$

We can 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.

I need the jump part to have infinite activity. In other words, we need at least one jump in every finite interval. This assumption implies two results. First, it implies that we do not need to keep track of the probability that there are no jumps in a specific interval, and second it identifies $\gamma^2(t)$. If we consider an interval without jumps, we obviously cannot estimate $\gamma^2(t)$ because we have no variation to identify it with.

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

Assumption Infinite-Activity Jumps might sound sounds very restrictive at first and contradicts the compound Poisson assumption often used in the literature. However, in practice, it is rather 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. Besides, standard variance-gamma processes are limits of compound Poisson process. Consequently, as long as we have a sufficient number of jumps, the representation will work well in practice. I discuss this further in Section 4.3.3.

The last assumption requires jump times to be unpredictable.

Assumption No Unpredictable Jumps. There does not exist any stopping times τ such that the event $p(\tau) \neq p(\tau-)$ is contained in the information set $\mathcal{F}_{\tau-}$.

Having set out the assumptions, I state the main theorem. I will later prove a more general proposition, Theorem 4. However, I have already described the environment sufficiently to make the result understandable. By stating the result that will likely be used in practice, it is easier to see where I am headed.

Theorem 3 (Locally Square-Integrable Itô Semimartingales as Integrals)**.** *Let $p(t)$ be an Itô semimartingale satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Unpredictable*

Jumps. Then we can represent $p(t)$ as

$$p(t) = \int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dW(s) + \frac{1}{\sqrt{2}} \int_0^t \gamma(s)\,d\mathcal{L}(s). \tag{22}$$

Proof. We can replace the jump part of (3) with an integral with respect to the standard-variance gamma process where the root jump volatility is the integrator using Corollary 4.1. □

At each time, I am replacing the function, $\delta(t, \cdot)$, with a single scalar $\gamma^2(t)$. In addition the integrator is switched from a compensated Poisson random measure, $(n-\nu)$, to a standard variance-gamma process, $\mathcal{L}(t)$. This greatly simplifies analysis, as $\delta(t, \cdot)$ is not identified, while $\gamma^2(t)$ is.

4.3.2. Time-Change Representation

The proof of Theorem 3 relies on Corollary 4.1, which I have not yet proven. In practice, Theorem 4 is the main result. The other results, such as Corollary 4.1, are straightforward implications of it. I prove this theorem now. Theorem 4 is a time-change representation for jump processes and hence is closely related to the time-change representations in the diffusion case. Consequently, I start by recalling those results.

The validity of diffusion representation for general continuous processes is ultimately implied by the Dambis-Dubins-Schwarz theorem, which shows that any continuous martingale time-changed by its predictable quadratic variation is a Wiener process, (Dambis 1965; Dubins and Schwarz 1965):

$$p^D(t) \stackrel{\mathcal{L}}{=} W\left(\langle p^D \rangle(t)\right). \tag{23}$$

(I use an equals sign with an $\mathcal{L}$ above it to refer to equality in law. The right-hand side of (23) evaluates the Wiener process at the random-clock determined by $\langle p^D(t) \rangle$.)

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, and 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] One method of proving standard central limit theorems is deriving them from this result.

In the jump case, though, the dynamics are more complicated. Not only do we have variation in magnitudes, but also we have variation in the locations, or, equivalently, in the number of the jumps. When we take the infill asymptotics, both of these sources of variation are still present. 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

---

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

tell us how the sum behaves as the expected number of summands approaches infinity, (Mittnik and Svetlozar 1993; Kozubowski and Svetlozar 1994).

I can generalize the Itô semimartingale assumption in Theorem 3. I only need to be able to represent the prices as integral with respect to a Poisson random measure. In particular, the $p(t)$'s characteristics do not need time-derivatives.

Theorem 4 (Time-Changing Jump Martingales)**.** *Let $p^J(t)$ be a purely discontinuous, martingale satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Unpredictable 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 Levy 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]

The proof of this theorem is in Section 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 $\mathcal{F}_{t-}$.

I condition on the number of jumps and show that the magnitudes are a continuous process in that space. Thus, I can apply the Dambis, Dubins & Schwarz theorem there, which results in a time-changed Wiener process. The standard representations further imply that each hitting times for each open set of magnitudes is a compound Poisson process. We can time-change these locations by their predictable quadratic variation, getting a standard Poisson process. However, since 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 predictable quadratic variation of $p(t)$ is the composition of quadratic variation arising from each of two time-changes. Therefore, the original process is a time-changed standard variance-gamma process.

Time-changed results are not particularly intuitive, and so we would like an integral representation as well. So we can assume that $p(t)$'s characteristics are absolutely continuous.

Corollary 4.1 (Jumps Processes as Integrals)**.** *Let $p^J(t)$ be a Itô semimartingale satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Unpredictable 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.*

Corollary 4.1 is completely analogous to how we can represent continuous martingales as stochastic volatility diffusion as shown in Dambis, Dubins, and Schwarz theorem by assuming the relevant characteristics are absolutely continuous.

---

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

4.3.3.   Processes with Finite-Activity Jumps

Arguably the most controversial assumption I make is Assumption Infinite-Activity Jumps. Various authors have argued that we have a large, but finite, number of jumps in each period. The natural question is what happens to the distributional result in this case? In any given interval, the price process is a point mass at zero if it does not jump. If the price does jump, we can represent it as done above. In other words, the ex-ante distribution over each interval is a mixture of a point mass at zero and a Laplace distribution where the mixing weights are the probability of the jump in that interval.

Corollary 4.2 (Time-Changing Finite-Activity Jump Martingales)**.** *Let $p^J(t)$ be a purely discontinuous martingale satisfying Assumptions Square-Integrable and No Unpredictable 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 Levy measure.*

Then $p^J(t)t$ time-changed by its predictable quadratic variation is a mixture of the 0 process — δ_0 — and the standard variance-gamma process where the mixing weights are the intensity of the jump process.

Corollary 4.2 implies that Theorem 4 is the limiting case of a finite-activity process as the intensity approaches infinity. Consequently, Theorem 3 approximates the true DGP well if the intensity is relatively large.

4.4.   Deriving the Realized Density

Having derived the continuous-time representation in Theorem 3 and proved it the previous section, I now solve the time-aggregation problem and derive the realized density. Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) simultaneously derived the realized density when prices have continuous paths, although, they did not call it thus. In particular, they show that if volatility and prices are correlated, σ_t^2 is a sufficient statistic for the dynamics under some technical conditions. They further show that conditional on the integrated diffusion volatility, the daily density of the return is Gaussian in the pure diffusion case.

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

$$f\left(r_t \,\big|\, \sigma_t^2\right) = f\left.\right|_{x_t = \left[\int_{t-1}^t \sigma^2(s)\,ds\right]} = N\left(0, \int_{t-1}^t \sigma^2(s)\,ds\right). \tag{24}$$

I now discuss the realized density in the jump case. In this case, the return has two parts: $dp(t) = \sigma(t)\,d(t) + \int_X \delta(t,x)(n-\nu)(dx, dt)$. Conditional on the values of $\sigma^2(t)$ 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 diffusion volatility. Hence, we only need to develop a parametric expression for 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 the following discrete-time representation holds.

Theorem 5 (Realized Density Representation)**.** *Let $p(t)$ be an Itô semimartingale satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Unpredictable Jumps. Let $\sigma^2(t)$ and $\gamma^2(t)$ be semimartingales whose martingale components are independent of the martingale components of $p(t)$. Then*

$$RD_t = N \left(\int_{t-1}^{t} \mu(s)\, ds, \int_{t-1}^{t} \sigma^2(s)\, ds \right) * \mathcal{L} \left(0, \int_{t-1}^{t} \gamma^2(s)\, ds \right), \tag{25}$$

and the predictive density is

$$h\left(r_t \,|\, \mathcal{F}_{t-1} \right) = \int_{\mu_t, \sigma_t^2, \gamma_t^2} RD_t(\mu_t, \sigma_t^2, \gamma_t^2)\, dG\left(\mu_t, \sigma_t^2, \gamma_t^2 | \mathcal{F}_{t-1} \right). \tag{26}$$

The intuition behind Theorem 5 is as follows. If $\gamma^2(t)$ was constant, we could pull it out of the integral without affecting the distribution: $\int_{t-1}^{T} \gamma_{t-1}^2\, d\mathcal{L}(s) \overset{\mathcal{L}}{=} \frac{\gamma_{t-1}^2}{\sqrt{2}} \int_{t-1}^{t} d\mathcal{L}(s)$. Since increments of the standard variance-gamma process are Laplace distributed, the second component is just $\mathcal{L}(0, 1)$. 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 the independence conditions between the martingale components to recover the general case.

To recover (26) from (25), I integrate the realized density out using the G that controls their dynamics. 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. Many of those models can be extended straightforwardly to the jump-diffusion case because the stylized features of the γ_t^2 and σ_t^2 are quite similar, as I show in Section 8.

The primary assumption that I added in Theorem 5 was the independence between the martingale components of the various terms. We need this assumption to justify the time-aggregation because we need the marginal and conditional distributions given the volatilities of $p(t)$ to coincide. In other words, I restrict the leverage effect but do not assume away all dependence. The volatilities and drift can be arbitrarily related.

To go into a more detail, since the jump part is purely discontinuous, it is orthogonal to the diffusion part. In other words, if we condition on the one process, the other process is still a martingale. Since we are integrating with respect to Brownian and Laplace motions the martingale property is sufficient to imply that the integrators of all predictable components of the representation are independent. To aggregate we need to separate the volatilities from the martingale components. Consequently, we must assume martingale components of the volatilities are independent of the

martingale components of $p(t)$.

The predictable relationship between the drift and volatilities is entirely unrestricted as is the relationship between the volatilities themselves. As long as it takes a positive amount of time for feedback from 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." There is some evidence that this is an artifact of the estimation procedure, and so I leave to future work the optimal way of bringing it into my framework. One way to do this is by keeping track of this correlation and using tools similar to those developed by the above paper and by Neuberger (2012) and Kalnina and Xiu (2017) and making Gaussian and Laplacian conditioning arguments.

5.   Estimation

I now construct estimators for $\sigma^2(t)$ and $\gamma^2(t)$ and their daily analogs. As is standard, the data do not identify $\mu(t)$, and so we cannot estimate it. The estimator I propose 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. It was not even a priori obvious that $\gamma^2(t)$ is nonparametrically identified. In particular, I develop a consistent estimator for $\gamma^2(\tau)$ for any fixed τ.[10] Also, this is the first consistent estimator for any instantaneous measure of jump dynamics.

5.1.   Assumptions

To start, I fix some notation and state some assumptions. The way that the instantaneous volatility estimators work is by taking 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 D.G.P. with time-varying volatility and an approximate D.G.P., 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 that I use are very similar to the standard ones used in the literature. When possible, I simplify them using the representation theory I have developed thus far.

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

   2. $\sigma(t)$ is càdlàg (or càglàd).

   3. $\gamma(t)$ is càdlàg (or càglàd).

---

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

Assumption 4 is essentially Jacod and Protter's (2012) assumption H. The assumption on the jumps is slightly more general and more straightforward. I also slightly modify the literature's assumption SH. (Here ω indexes the underlying probability space Ω.)

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

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

These two assumptions have a close relationship, 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, if we can prove consistency under the one assumption, the estimator automatically converges under the other assumption as well. This result implies that in the proofs below we can assume SHL without loss of generality. To make this statement explicit, we have the following lemma whose proof is in the appendix. The arrow with $\mathcal{L}$-s above it refers to stable convergence in law, which is the type of convergence necessary for confidence intervals to be valid in this setting.

Lemma 6 (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.*

To reduce the amount of notation, I adopt the following notational convention from the literature so that the processes are well-defined over the entire line, not just the place we are estimating them:

$$i \in \mathbb{Z}, i \le 0 \implies \Delta_i^n p = 0. \tag{27}$$

I am setting the processes equal to zero outside of the relevant window.

To estimate the instantaneous volatility, I must approximate $\sigma^2(\tau-)$ and $\gamma^2(\tau-)$. Thus, we need to choose a sequence of i_n, $k_n \Delta_{i_n}$, so that we are averaging the variation, either squared or absolute, over smaller and smaller intervals to the left of τ. Consider the following interval:

$$I(i,n) := [(i - k_n - 1)\Delta^n, (i-1)\Delta^n]. \tag{28}$$

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 are one-dimensional without loss of generality.

5.2. Instantaneous Volatility Estimators

Having specified the framework, I state the estimators themselves. The intuition behind their convergence is that we are averaging the volatilities over shrinking intervals that approach τ from

the left. As long as the infill asymptotics imply the number of increments being averaging over is increasing faster than the length of the interval is shrinking we precisely estimate the volatility. Since we are estimating the process from the left, we are approximating the value before τ, i.e., we are estimating $\gamma^2(\tau-)$.

I first derive an estimator for $\sigma^2(\tau-)$. There are few such estimators in the literature that do this, including Mancini (2001) and Jacod and Rosenbaum (2013). They do this by that noting that estimating the integrated diffusion volatility — $\langle p^D \rangle(t)$ — is straightforward. We can use the integrated volatilities sample analog. In particular, we can use time-derivative of $\widehat{\langle p^D \rangle}(t)$ to estimate σ_τ^2.

The main difficulty in practice is separating the jump and diffusion variation. I do this by truncating away the large increments, where large is in terms of an asymptotic rate. Asymptotically, this eliminates large jumps, and the small jumps do not affect the asymptotic value.

Theorem 7 (Estimating the Instantaneous Diffusion Volatility). *Let $p(t)$ be an Itô semimartingale 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 $c_1(\Delta^n)^{1/4} \leq v_1^n \leq= c_2\sqrt{\Delta^n}$ for some constants c_1, c_2 and $v_2^n \to 1$. Then*

$$\widehat{\sigma}_{i_n}^2(k_n, \tau-, p) := \frac{1}{k_n \Delta^n} \sum_{m=0}^{k_n-1} v_2^n \big|\Delta_{i_n}^n p\big|^2 1\{\big|\Delta_{i_n}^n p\big| \leq v_1^n\} \xrightarrow{\mathbb{P}} \sigma^2(\tau-). \tag{29}$$

One might think a similar estimation strategy would work to estimate $\gamma^2(t)$, i.e., form an estimator of $\langle p^J \rangle(t)$ 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 would converge to zero in their proof of the validity of their estimator for $\sigma^2(t)$. Intuitively, 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.

Over a fixed interval the quadratic variation of jump processes and diffusive processes are of the same asymptotic order as we shrink Δ, (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 Laplace and diffusive processes that have similar asymptotic properties. [11] 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: $\sqrt{\Delta^n}$.[12] In addition, the $\lim_{\Delta^n \to 0}\big|\Delta_{i_n}^n p(t)\big|$ contains both $\gamma^2(\tau)$ and $\sigma^2(\tau)$.

Theorem 8 (Estimating the Instantaneous Absolute Volatility). *Let $p(t)$ be an Itô semimartingale 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 $i_n = i - k_n - 1$.*

---

11. It is an interesting open question to what other jump processes this result extends.

12. As an aside, neither of the processes are martingales. They are semimartingales.

Then the following holds, where $\text{erfcx} := \frac{2\exp(x^2)}{\sqrt{\pi}} \int_x^\infty \exp(-s^2)\,ds$:[13]

$$\frac{1}{k_n\sqrt{\Delta^n}} \sum_{m=0}^{k_n-1} |\Delta_{i_n+m}^n p| \xrightarrow{\mathbb{P}} \mathbb{E}|N(0,1)|\sigma(\tau-) + \frac{\gamma(\tau-)}{\sqrt{2}} \text{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right). \tag{30}$$

As long as $\sigma^2(t)$ and $\gamma^2(t)$ are locally constant around τ, we can use the implied parametric form to compute the limiting value as a function of $\sigma^2(\tau)$ and $\gamma^2(\tau)$. The expression on the right of (30) is the mean of the convolution of $|N\left(0,\sigma_{\tau-}^2\right)|$ and $|\mathcal{L}\left(0,\gamma_{\tau-}^2\right)|$.

We can 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. In general, any convex weighting function of the differences will work. I use the absolute value of the difference between the two values because it works well in simulation.

Theorem 9 (Estimating the Instantaneous Jump Volatility)**.** *Let* $p(t)$ *be an Itô semimartingale 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:*

$$\widehat{\gamma}(k_n, \tau-, p) := \underset{\gamma}{\arg\min}\, g\left(\left|\frac{1}{k_n\sqrt{\Delta^n}}\sum_{m=0}^{k_n-1}|\Delta_{i_n+m}^n p| - \mathbb{E}|N(0,1)|\widehat{\sigma}_n(\tau-) - \frac{\gamma\,\text{erfcx}\left(\frac{\widehat{\sigma}_n(\tau-)}{\gamma}\right)}{2}\right|\right) \xrightarrow{\mathbb{P}} \gamma(\tau-). \tag{31}$$

5.3. Implementation

We now have an estimator for the instantaneous jump 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)). To do this, I define the function:

$$g(x) := \left(1 - (2x-1)^2\right)(x >= 0)(x <= 1). \tag{32}$$

The pre-averaged data is the rolling average of the true data:

$$\bar{p}_{i_n} := \frac{1}{\kappa_n\sqrt{\int_0^1 g^2(s)\,ds}} \sum_{m=1}^{\kappa_n-1} g\left(\frac{m-1}{\kappa_n}\right)\Delta_{i_n+m}^n p. \tag{33}$$

The g function is there to correct for the error introduced by the pre-averaging.

---

13. 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.

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 $\kappa_n = \lfloor \frac{\theta}{(\Delta^n)^{.55}} \rfloor$. 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.[14] I set $\theta = 0.5$, which is a values recommend by Hautsch and Podolskij (2013), and works well in my simulations.

I apply Theorem 7 to estimate $\sigma^2(\tau-)$. To do this, we need to choose v_2^n to be converge to 1, I let $v_2^n = 1$. More importantly, I need to choose the truncation threshold v_1^n. We need v_1^n to asymptotically upper bound the absolute diffusion part. In the literature, it is usually chosen $v_1^n = c\widetilde{\sigma}(\tau-)\Delta_n^{0.49}$ where $\widetilde{\sigma}(\tau-)$ is a preliminary estimator for σ and c is a number of standard deviations chosen by the econometrician.

The tails of the Laplace and Gaussian random variables are very similar. The Gaussian density is proportional to $\exp(-x^2/2)$ in the tails, while the Laplace density is 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 6. On the other hand, the law of the iterated logarithm tightly bounds the deviations of a Gaussian variable, and so I use $v_1^n = \sqrt{2}\widetilde{\sigma}(\tau-)\sqrt{\Delta}\sqrt{\log(\log(1/\Delta_n))}$.

To form a preliminary estimator, I start with the 1.25 times bipower variation and then iterate until convergence. We need to start by overestimating the volatility to avoid incorrectly setting $\widehat{\sigma}(\tau-) = 0$ since that would truncate away all the increments. It is worth noting that this volatility estimator relies on neither $\gamma^2(\tau)$ nor the qualitative properties of the Laplace representation.

In addition, we must choose k_n, where $1/k_n$ controls the length of the interval over which the volatilities are treated as approximately constant. Theory tells us that $k_n \to \infty$ and $k_n\sqrt{\Delta^n} \to 0$, I choose $k_n = 1000 + (\Delta^n)^{1/4}$ because that seems to work well in the simulations with market microstructure noise.

Now that I have an estimator for $\sigma^2(\tau-)$, I need an estimator for the local absolute value. I plug the pre-averaged data into (30). 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 that has not been proven. An interesting extension for future work would be to extend these results to cover the noise case as well and to figure out the various biases arising there.

5.4. Integrated Volatilities

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

Theorem 10 (Consistentcy of the Integrated Estimators)**.** *Let $p(t)$ be Itô semimartingale satisfying Assumptions HL, Infinite-Activity Jumps, and Square-Integrable. Let k_n, Δ^n satisfy $k_n \to \infty$ and*

---

14. The transformation creating $\bar{p}_{in}$ does not affect the volatilities but does affect the mean.

$k_n\sqrt{\Delta_n} \to 0$. *Define* $i_n = i - k_n - 1$. *Then*

$$\widehat{\sigma}_t^2 := \frac{1}{\#i_n \in [t-1,t]} \sum_{t-1 < t_n \leq t} \widehat{\sigma}^2(k_n, t_n, p) \xrightarrow{\mathbb{P}} \int_{t-1}^t \sigma^2(s)\,ds, \tag{34}$$

and

$$\widehat{\gamma}_t^2 := \frac{1}{\#i_n \in [t-1,t]} \sum_{t-1 < t_n \leq t} \widehat{\gamma}^2(k_n, t_n, p) \xrightarrow{\mathbb{P}} \int_{t-1}^t \gamma^2(s)\,ds. \tag{35}$$

Proof. I am averaging estimates 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, I take the daily averages of the instantaneous volatilities. To estimate the realized density, I plug the daily estimates into (25). Since this function is uniformly continuous, as long the volatilities are bounded away from zero, the resulting estimators should work well.

6. Simulations

One key advantage of my representation is that it can be simulated from easily as long as we can simulate the instantaneous volatilities. Perhaps the most commonly used such model for the diffusion volatility is the Cox-Ingersoll-Ross (CIR) process. (A diffusion model whose volatility follows a CIR process is known as a Heston model.) 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, and so I adopt this model for the jump volatility as well. Once we have the volatilities, we can simulate the price as the sum of the diffusion and jump parts directly.

6.1. Simulation Data Generating Process

The Cox-Ingersoll-Ross (CIR) process, also known as the square-root process, has the following form:

$$dx(t) = \kappa(\theta - (x(t)) + \omega\sqrt{x(t)}\,dW(t), \tag{36}$$

where θ is asymptotic mean, κ is the mean-reversion rate, and ω is a scale parameter.

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. Note, the asymptotic standard deviation for a CIR process equals $\frac{2\theta\omega^2}{\kappa}$. I chose the specific parameter values displayed below to match the discrete-time dynamics of the price processes.

Once I obtain $\sigma^2(t)$ and $\gamma^2(t)$, I plug them into the following continuous-time DGP:

$$dp(t) = \sigma(t)\,dW(t) + \frac{\gamma(t)}{\sqrt{2}}\,d\mathcal{L}(t). \tag{37}$$

Table 1: Volatility Parameters

| Parameter	θ	κ	ω	$\frac{2\theta\omega^2}{\kappa}$
$\sigma^2(t)$	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}

This gives me a sequence of prices, which I use to estimate the volatilities.

6.2. Simulation Results

I focus on the daily volatility results below as they are sufficient statistics for all of the daily objects, which is what I use in the applications.[15] To provide a comparison, I also report 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). In the jump case, the estimators above do not converge to γ_t^2 but rather to the jump part of the quadratic variation. However, since γ_t^2 is the 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 5 without the microstructure correction described there. Figure 2 reports the results when I sample at the one-second frequency. Some of the jump variation estimators are not easy to see on the plot because I truncated them to zero. In the tables, I report averages over 250 days. It is worth noting that each of the daily estimates are independent, i.e., I do not smooth across days.

Figure 2: Simulation Results without Microstructure Noise

(a) σ_t (b) γ_t
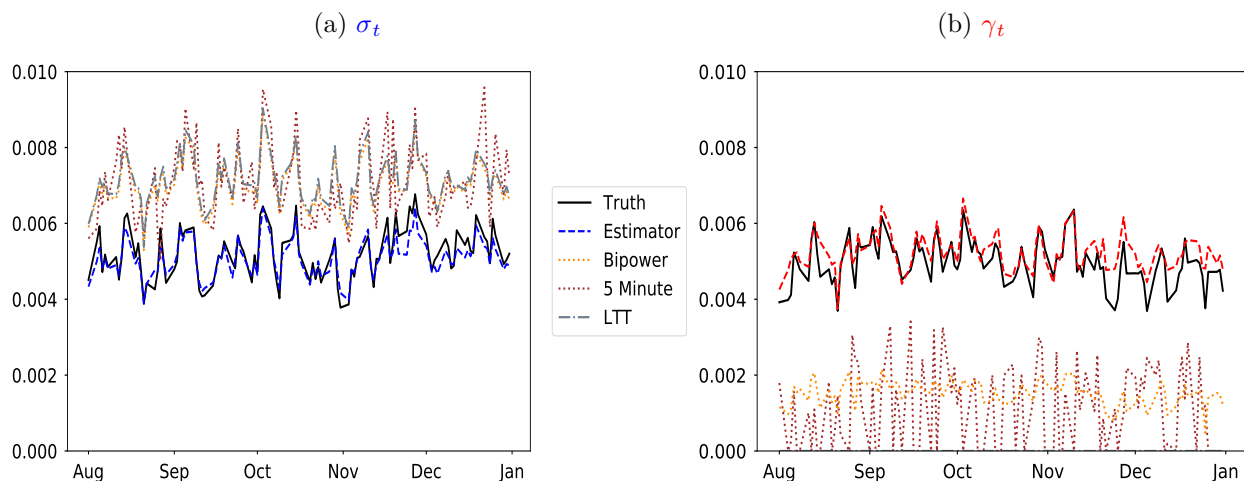
As can be seen in Figure 2, the estimators in the literature for σ_t^2 are badly biased upwards in finite-samples when the jump activity is high. This bias even holds in simulations without market microstructure noise at the one-second frequency, which gives approximately $24\,000$ observations

---

15. I report continuous-time results in Section E.

per day. This bias for σ_t^2 causes the literature's estimators for γ_t^2 to be severely biased as well, contrary to the theory.[16]

Table 2: Relative Simulation Error without Microstructure

Observations per Minute	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$				$\frac{\mathbb{E}[(\widehat{\gamma_t}-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$			
	Bipower	LTT	5 Minute	Proposed	Bipower	LTT	5 Minute	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

The proposed estimators, however, perform quite well at this frequency. Table 2 reports the average root mean square errors of a year's worth of various estimators. As can be seen from this table, the proposed estimators outperform the other estimators in the literature by approximately an order of magnitude in this simulation.

Figure 3: Simulation Results with Microstructure Noise

(a) σ_t

(b) γ_t
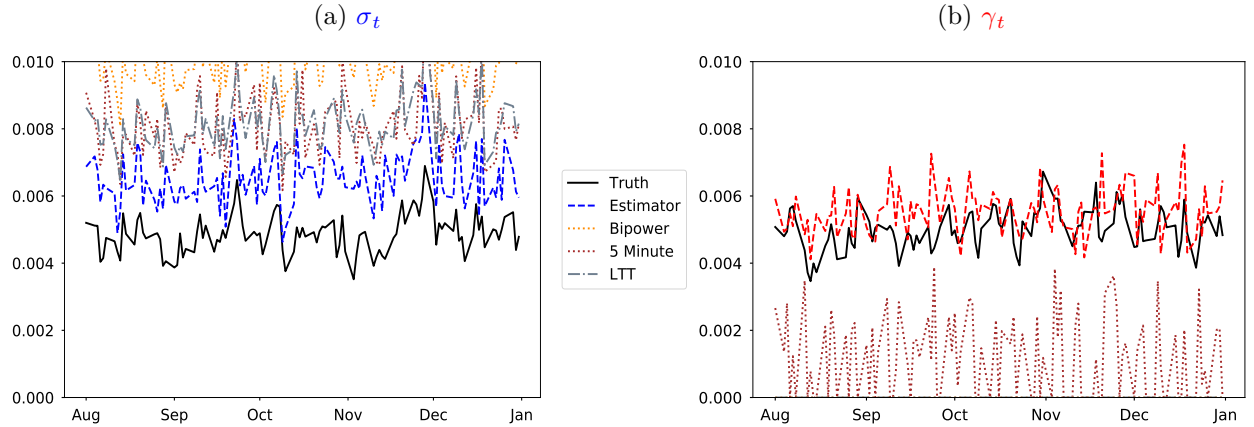
The data has substantial market microstructure noise. To mimic its effect, I follow Christensen, Oomen, and Podolskij (2014) and consider a scenario where we observe $r_{i_n} + u_{i_n}$ and u_{i_n} follows

$$u_{i_n} = \beta u_{i_n-1} + \epsilon_{i_n}, \quad \epsilon_{i_n} \overset{i.i.d.}{\sim} N(0, \omega^2(1-\beta^2)). \tag{38}$$

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): $\frac{1}{T}\sum_{t=1}^{T-1} \frac{\Delta^n}{2} \sum_{t-1<i_n,i_n-1<t} |\Delta_{i_n}^n p| |\Delta_{i_n-1}^n p|$. I set $\beta = 0.77$, which is the value used in Christensen, Oomen, and Podolskij (2014). They set it to match the trade sign of the S&P 500 futures contract on the day of the 2010 Flash Crash.

---

16. This bias likely explains why I find significantly higher jump variation than Christensen, Oomen, and Podolskij (2014) do, which is the other paper to use pre-averaging and ultra high-frequency returns to measure jump variation. Because they use bipower variation to measure the jump proportion, their estimators for the jump proportion are likely highly-biased downwards.

I now add the market-microstructure correction described in Section 5. 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.

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 2, I report the mean-square error of the previous estimates average of 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. Again, 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 are both biased and inconsistent.

Table 3: Relative Simulation Error with Microstructure

Observations per Minute	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$				$\frac{\mathbb{E}[(\widehat{\gamma_t}-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$			
	Bipower	LTT	5 Minute	Proposed	Bipower	LTT	5 Minute	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

The last simulation result I report is a simulation with only a few jumps. To be precise, I simulate the volatilities using the DGP described in Table 1. Then instead of simulating the prices using (37), 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. I also add the microstructure noise as described in (38).

Table 4: Relative Simulation Error with Microstructure and Poisson Jumps

Observations per Minute	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$				$\frac{\mathbb{E}[(\widehat{\gamma_t}-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$			
	BNS	LTT	5 Minute	Proposed	BNS	LTT	5 Minute	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

As we can see in Table 4, the proposed method works well even in cases with only a few jumps. It uniformly dominates the other methods by a significant amount when estimating γ_t^2. The results

are closer if we look at estimating σ_t^2. In that case, the estimation error is only smaller when we sample quite finely. However, for liquid stocks, this is the empirically-relevant case. Also, the proposed method outperforms by more when estimating the γ_t^2 then it underperforms when estimating σ_t^2. To see, this note that the average RMSE over the two targets is lowest with the proposed method at all frequencies. Consequently, I would suggest using the proposed method in all cases except when we are sure we only have a few jumps and we cannot sample very finely.[17]

7. 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, which I obtain from the Trade and Quotes (TAQ) database at Wharton Research Data Services (WRDS). The S&P 500 is arguably the most important index of financial activity. It is likely the most closely watched equity index and several heavily subscribed index funds (including SPY) track it directly. Consequently, the economics and finance literature has studied it extensively, often using it as a proxy for the market.

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 resultant 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 $4.4\,\mathrm{GiB}$ 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.

This market microstructure causes the asset to fail to be a semimartingale in practice. Thankfully, a substantial literature has developed to deal with precisely this issue. The two leading methods are sampling rather sparsely, for example at a 5 minute frequency as Liu, Patton, and Sheppard (2015) argue for, and pre-averaging, where one takes appropriately weighted averages of the price over small (shrinking) intervals. We need to separate the jump volatility from the diffusion volatility, and so we must sample much more finely than once very 5 minutes. This requirement arises because the only information the estimators use to separate the jumps and diffusive component come from the tails, and tails by definition are times without much data. Consequently, any deconvolution procedure we use here is inherently low-powered.

Consequently, I preprocess the data using the pre-averaging approach as in Podolskij and Vetter (2009) and Aït-Sahalia, Jacod, and Li (2012). This procedure is known not to affect the consistency of the estimation procedure. The basic idea is rather simple. We average the price over a small interval to remove the noise. If we pick the rates at which we shrink the interval to appropriately balance averaging away the noise and estimating the instantaneous variation, the estimators will

---

17. If we consider the satiation without microstructure noise, the results are similar. The previous methods do perform better when estimating σ_t^2, but the proposed method still substantially outperforms in estimating γ_t^2.

be consistent even in the presence of noise.

In the analysis for news premia, I predict log-returns on SPY, which I compute using the high-frequency data. I include the overnight returns, and so the return on day τ starts at the close of business on day $\tau - 1$ and ends at the close of business on day τ.

8. VOLATILITY: EMPIRICS

I separate this empirical part into three subsections. The first section characterizes the static properties of the volatilities. The second characterizes their dynamic properties. In particular, it shows that both volatilities are highly persistent, displaying long-memory. The third section introduces a new measure of jump variability — $\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$ — in order to isolate the effect of γ_t^2 in the presence of σ_t^2. This ratio is a measure for the proportion of the investors' new information driven by news.

8.1. 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, σ_t^2 and γ_t^2 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 4, the volatilities are very closely related; their correlation coefficient equals 0.93. As one would expect from previous volatility measures, they both significantly increase during crises/recessions. Interestingly, σ_t^2 spiked more than γ_t^2 during the Financial Crisis and seems to spike more during recessions.

Figure 4: Root Volatilities
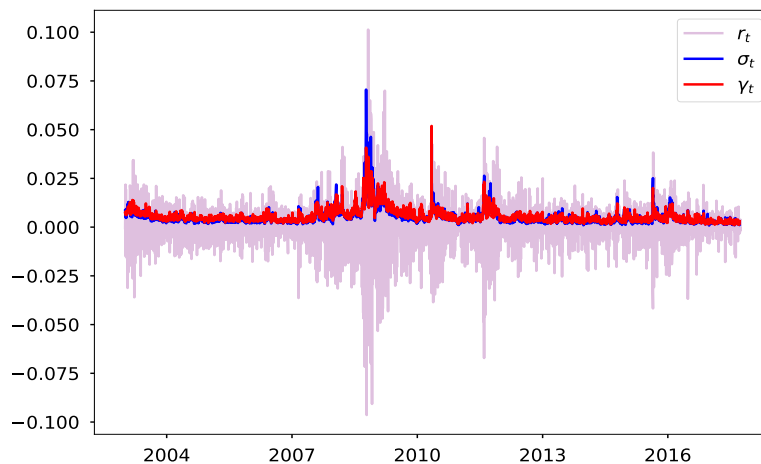
Figure 5 plots the two log-volatility distributions along with their joint distribution. As can

be seen from the graph both marginal distributions are skewed right, and the joint distribution is just as skewed as the marginal densities. It is worth noting that being skewed right means that the volatilities are more likely to take on abnormally large values than take on abnormally small ones. Volatilities usually spike during crises, and so the distributions are skewed in a direction that increases the investors' risk relative to an unskewed distribution. This fact is particularly noteworthy as these are distributions of log-volatilities, and taking the logarithm already removes a large amount of skewness.

Figure 5: Log-Volatility Densities
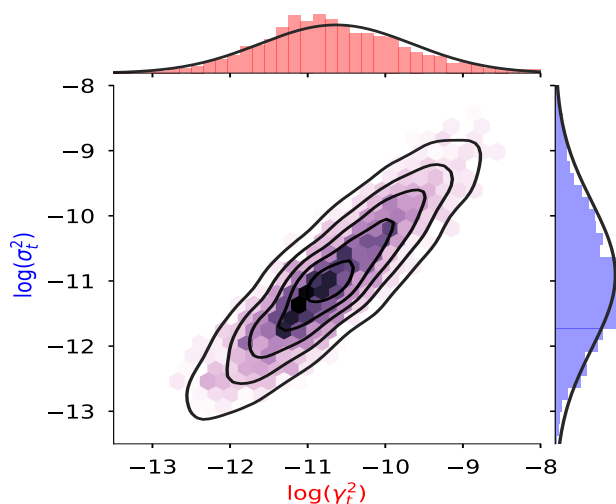
A few of the original realized volatility papers, (Andersen, Bollerslev, Diebold, and Labys 2001; Andersen, Bollerslev, Diebold, and Ebens 2001), argue that realized volatilities are approximately log-Gaussian. One might expect this to continue to hold in this case. The black lines in Figure 5 are Gaussian 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 5.[18]

Table 5: 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

We are interested not just in the univariate dynamics, but also their relationships. We know from Figure 5 that the two volatilities move together. To investigate this further, Table 6 reports

---

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

the correlations between the various volatility measures and daily excess returns.

Table 6 also includes an indicator — $\mathbf{1}\{\text{FOMC}\}_t$ — for days when the Federal Open Market Committee (FOMC) releases its announcements. As discussed in the literature review, much of the previous literature on the effect of news on asset prices has focused on the effect of FOMC announcements.

Table 6: 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	$\mathbf{1}\{\text{FOMC}\}_t$
σ_t^2	1.00	0.74	0.96	-0.29	-0.11	0.01
γ_t^2	0.74	1.00	0.89	-0.10	-0.13	0.06
$\sigma_t^2 + \gamma_t^2$	0.96	0.89	1.00	-0.23	-0.13	0.05
$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$	-0.11	-0.13	-0.13	1.00	0.12	0.05

Clearly, σ_t^2 and γ_t^2 are highly positively correlated. Table 6 also reports the correlations between the logarithms of the parameters above because Pearson's correlation coefficients only measure linear relationships. On the other hand, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ is weakly negatively correlated with the other volatility measures. Note, this is possible because it is a non-linear transformation of γ_t^2 and $\sigma_t^2+\gamma_t^2$. Interestingly, $\mathbf{1}\{\text{FOMC}\}_t$ is positively correlated with all of the volatility measures even though they are not all positively correlated with each other. The standard negative contemporaneous relationship between volatility and returns also holds.

Since the volatilities are closer to log-Gaussian than they are to Gaussian, Table 7 reports the correlations reported in Table 6 with the volatilities measured in terms of their logarithms.

Table 7: 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

The signs of the relationships are the same in both tables, but the magnitudes are larger in Table 7. This result should not be too surprising given the evidence above. The volatilities' distributions are closer to log-Gaussian than they are to Gaussian.

8.2. Dynamics

Having considered the data's static properties, I now consider the dynamic properties starting with the univariate case. Throughout, I focus on the log-volatilities because they behave are closer to Gaussian as shown in Section 8.1, and so the true conditional expectations are likely closer to

approximately linear. I first replicate the standard stylized features for the diffusion volatility and show that the jump volatility behaves similarly. I then perform a joint analysis.

8.2.1.   Measuring the Persistence

Figure 6 plots the volatilities' autocorrelation functions. Both series are extremely persistent.[19] We can also see that both series have a similar univariate autocorrelation structure.

Figure 6: Autocorrelation Functions

(a) $\log(\sigma_t^2)$

(b) $\log(\gamma_t^2)$
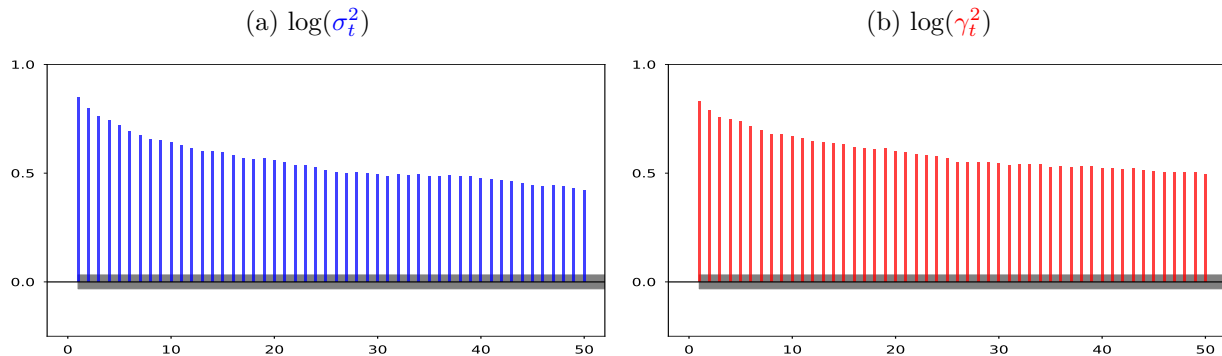
Since the series are so persistent, one might wonder if they have a unit root. Table 8 rejects this hypothesis. In particular, the standard Augmented Dickey-Fuller test rejects at any reasonable level of significance, (Dickey and Fuller 1981). Since the volatilities do not have a unit root, one might think that they are short memory processes, that is, their autocorrelation functions decay geometrically. Perhaps less surprisingly given Figure 6, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test also rejects this hypothesis, (Kwiatkowski et al. 1992).

Those readers familiar with the empirical volatility literature should not find this result too surprising. The diffusion volatility's long memory is a key stylized fact in the literature (Andersen, Bollerslev, Diebold, and Labys 2003). Perhaps more surprisingly, the jump volatility also has long memory. Table 8 reports estimates for the long-memory coefficient, (d), using the Geweke Porter-Hudak (GPH) estimator (Geweke and Porter-Hudak 1983). The smoothed periodogram estimator developed by Reisen (1994) gives almost identical results.

Notably, the point estimates for d are in the infinite-variance region ($d > 1/2$). These estimates imply the volatility itself has an infinite unconditional variance. [20] However, we cannot reject the hypothesis that the $d < 1/2$ in any of the cases.

8.2.2.   Univariate Dynamics

Table 9 reports independent $AR(1)$ regressions on each volatility to gain some high-level understanding of the dynamics. Both series are quite persistent and predictable. However, we still have

---

19. The gray bars are the standard Bartlett bands, i.e., confidence sets for the null of independent and identically distributed data.

20. Having an infinite unconditional variance does not imply that the volatilities have an infinite conditional variance. A process can be locally square-integrable even if has infinite variance. Prices exhibit this behavior.

Table 8: Persistence Statistics

	$\log(\sigma_t^2)$	$\log(\gamma_t^2)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$
	p-value		
ADF Test (Unit-Root Null)	1.90×10^{-5}	3.61×10^{-4}	8.01×10^{-15}
KPSS Test (Short-Memory Null)	$\ll 1\%$	$\ll 1\%$	$\ll 1\%$
	statistic		
1st-Order Autocorrelation	0.85	0.83	0.26
Fractional Integration Coefficient (d)	0.57 (0.45, 0.79)	0.66 (0.50, 0.82)	0.47 (0.31, 0.64)

economically significant innovations. In other words, the autocorrelation and innovation variance are both high.[21]

I now consider univariate autoregressive models for both series. I use Schwarz Information Criterion (SIC) to select the lag order.[22] This is not the ideal thing to do as it assumes away the long-memory that I just demonstrated. However, it still is useful to understand the short-memory dynamics of the two series. The two series both exhibit substantial autocorrelation as shown above, with the AR coefficients declining slowly. SIC chooses 9 lags for both series. The two series are both quite predictable in terms of $\mathbb{R}^2$ as well. The regressions chosen by SIC give an $\mathbb{R}^2$ of 76% for $\log(\sigma_t^2)$ and 69% for $\log(\gamma_t^2)$. These numbers are likely higher than that found in the literature, which are often in the neighborhood of 40% to 50%, because I am doing a better job at separating out the diffusion and jump volatilities, (Bollerslev, Patton, and Quaedvlieg 2016, 8). Effectively, my variables have less measurement error than is commonly used in the literature. In addition, volatility appears to be more predictable during the Great Recession, which is a large portion of my sample.

8.2.3. Joint Dynamics

The joint analysis starts by considering whether the two volatility series Granger cause each other. Standard tests conclusively reject the null of no-causality in either direction. The sum-of-squared residuals (SSR) test for $\log(\gamma_t^2)$ Granger-causing $\log(\sigma_t^2)$ with one lag returns a $\chi^2(df = 1)$ value of 298[23] Conversely, the SSR test for $\log(\sigma_t^2)$ Granger-causing $\log(\gamma_t^2)$ with one lag returns a $\chi^2(df = 1)$ value of 398. These results are robust to the number of lags chosen and to the specific version of the

---

21. 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.

22. Other selection criteria such as Akaike information criteria (AIC) choose similar models. As one would expect, AIC chooses a few more lags.

23. Since a $\chi^2(df = 1)$ is the distribution of $|N(0,1)|^2$, this is equivalent to a t-static of 17.27.

Table 9: Univariate Autoregressive Models

	$\log \sigma_t^2$		$\log \gamma_t^2$	
	AR(1)			
Intercept	-1.63	$(-1.82,\ -1.45)$	-1.78	$(-1.97,\ -1.59\ )$
	0.85	$(0.83,\ 0.87)$	0.83	$(0.82,\ 0.85\ )$
$\mathbb{R}^2$	$72\,\%$		$69\,\%$	
	AR(BIC)			
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	

test. The tests overwhelmingly reject no-causality in every case. In other words, adding information about the jumps helps us to predict the diffusive variation, and vice-versa.

To make this operational, consider a vector autoregression (VAR). Here, the Schwarz Information Criterion (SIC) chooses 6 lags. Table 14, which is in Section F, reports the results. Table 10 reports the results for a VAR(1). The results for the more general specification are consistent with these results. The results are consistent with the Granger-causality results above. Both volatilities depend on the lags of both volatilities. The coefficients for diffusion volatility are larger in magnitude, however.

Table 10: VAR(1) Results

	Intercept	$\log\left(\sigma_{t-1}^2\right)$	$\log\left(\gamma_{t-1}^2\right)$	Innovation Variance	$\mathbb{R}^2$
$\log \sigma_t^2$	-0.84	0.56	0.38	0.33	$74\,\%$
$\log \gamma_t^2$	-1.80	0.34	0.48	0.27	$72\,\%$

The correlation between the innovations is 0.63. Since both the unconditional correlation and the innovation correlation between the two series are high, there appears to be a shared component that drives a large amount of the variation in both series.

8.3.   Jump Proportion

The previous sections showed that the volatilities share a component that drives a large portion of each of their variations. We would like to isolate the effect of the jumps and examine its dynamics directly. (This will be quite important when we consider the pricing implications.) To do this, define the jump proportion — $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ — which the previous sections briefly alluded to but did not investigate in detail.

Figure 7: Time-Varying Jump Proportion
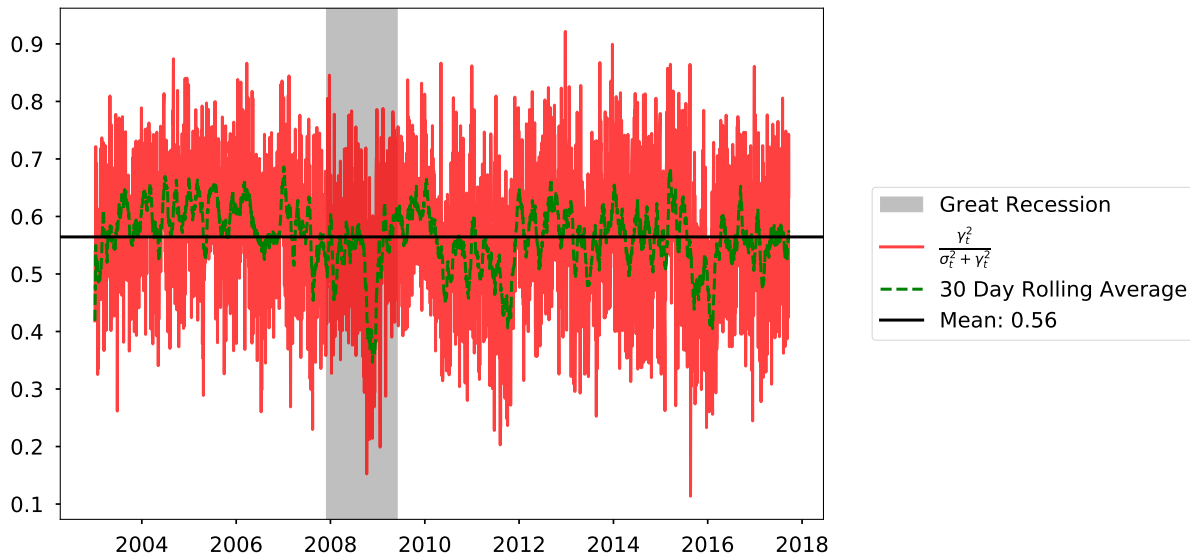
To understand $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$, Figure 7 plots its variation over time. Its mean and the Great Recession are plotted for reference. Figure 7 also displays the rolling average to visualize the series' low-frequency variation better. Clearly, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ has substantial low- and high-frequency variation.

Figure 8: $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$
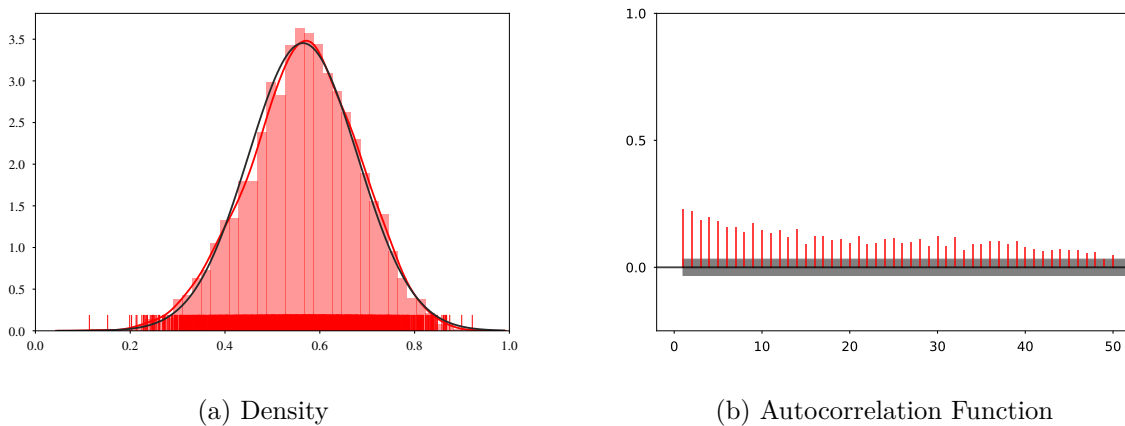
(a) Density

(b) Autocorrelation Function

Figure 8a plots $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$'s histogram. The red line is a kernel density estimate, and the back line is a Gaussian distribution fit to the data. As we can see, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$'s density is roughly log-Gaussian.

Since Figure 7 plots daily data, and the dataset spans several years, the graph is at too low a resolution to be easily comprehensible. Hence, Figure 9 plots $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ and $\mathbf{1}\{\text{FOMC}\}_t$ on the same graph in the most interesting sub-period in the data: 2008–2009. As we might expect given previous work, such as Andersen, Bollerslev, Diebold, and Vega (2003), Faust et al. (2007), and Beechey and Wright (2009), $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ often spikes when the FOMC makes its announcements. However, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ varies significantly more than $\mathbf{1}\{\text{FOMC}\}_t$ does. If you regress $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ on $\mathbf{1}\{\text{FOMC}\}_t$, the resulting coefficient is 0.50 with associated t-statistic equal to 7.90. Even though this relationship is highly statistically significant, The $\bar{\mathbb{R}}^2$ from this regression is only 0.78 %.

Figure 9: $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ versus $\mathbf{1}\{\text{FOMC}\}_t$
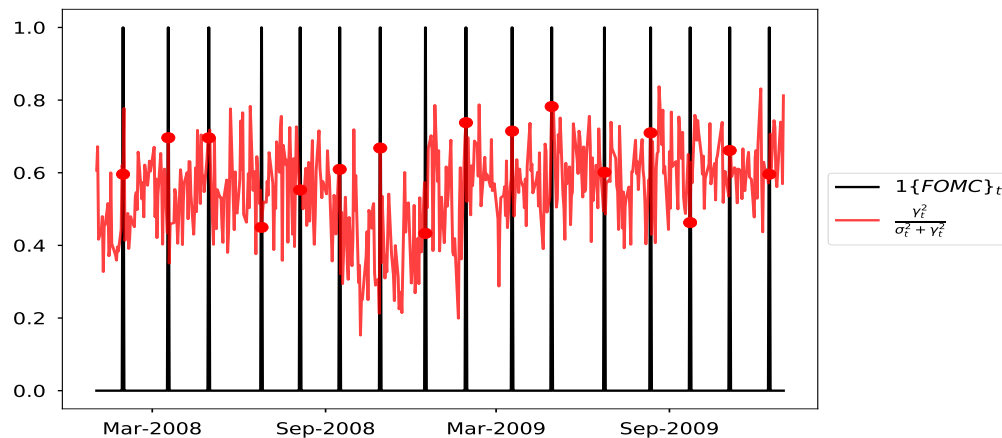
8.4. Realized Density Evaluation

In Section 6, I showed you 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 5 is a valid conditional density, and we can consistently estimate the conditioning variables. Consequently, techniques developed to analyze conditional densities work well here.

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 its inverse-CDF through simulation. I then apply this inverse-CDF to the demeaned daily return. This procedure jointly evaluates the density representation, the time-aggregation procedure, and the estimation of σ_t^2 and γ_t^2.

As can be seen in Figure 10, 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. 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. I am estimating 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

Figure 10: Realized Density Evaluation
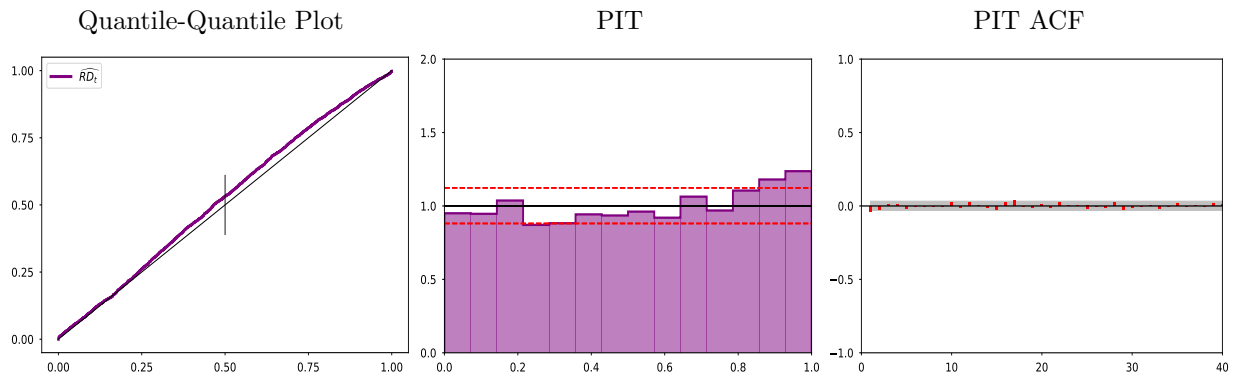
deviations here do not invalidate that representation at all. If we look at the Figure 10, we can see that the far right tail is also measured relatively well. It is only the 80th to 95th percentiles that I am missing. Furthermore, Figure 10 the deviations are most perfectly uncorrelated across time. This lack of correlation implies the densities dynamics are estimated quite well.

9.   News Premia: Theory

Discontinuous prices or information flows that this paper considers break the derivation of CAPM-style results where risk-premia are instantaneous covariances with marginal utility. In particular, risk premia are no longer proportional to the integrated diffusive covariation between prices and stochastic discount factors. For example, in Ai and Bansal's (2018) world, we have an announcement SDF (A-SDF) whose covariation with returns is also priced, while in Tsai and Wachter's (2018) world, it is the covariation between the SDF and returns during extreme events that matters. This section unifies these two theories by decomposing the covariation between the prices and the investor's pricing kernels into their predictable and unpredictable components. In particular, it shows that risk premia have two components in the general case. Contrary to the discussion in Tsai and Wachter (2018, 31), both of these terms are, appropriately-defined, covariances, and so we would expect returns to exhibit a factor structure.

9.1.   Preferences

To avoid introducing more notation than necessary, I characterize the investor's decision problem over a short period Δ and take $\Delta \to 0$. Following Ai and Bansal (2018), I adopt the intertemporal preferences represented as in Strzalecki (2013). Let $V(t)$ be the representative investor's value function at time t, and $u(\cdot)$ be the associated flow utility over current period consumption — $C(t)$. Let κ denote the rate of time-preference. I assume that κ is constant for notation convenience, but this can be easily generalized.

Definition 12 (Certainty Equivalence Functional)**.**

$$V(t) = \int_t^{t+\Delta} u\left(c\left(t\right)\right) dt + \mathcal{I}\left[\int_{s \geq t+\Delta} \exp\left(-\kappa\left(s-t\right)\right) u(C(s)) \, ds \,\middle|\, \mathcal{F}_{t-}\right].$$ (39)

I immediately specialize to the form given in Definition 12, which is (Ai and Bansal 2018, observation ii, p. 1401). Most of the results still go through in the general case, albeit with a loss of interpretability. In this case, investors preferences are represented as

$$\mathcal{I}\left[V(t) \,\middle|\, \mathcal{F}_{t-}\right] = \phi^{-1}\left(\mathbb{E}\left[\phi(V(t)) \,\middle|\, \mathcal{F}_{t-}\right]\right),$$ (40)

for some strictly increasing function ϕ. Preferences having this form include the recursive utility of Kreps and Porteus (1978) and Epstein and Zin (1989) and the second-order expected utility of Ergin and Gul (2009). If ϕ is the identity function, preferences are time-separable. They also cleanly characterize the problem at hand. Ai and Bansal (2018) show that these form of preferences lead to an announcement premium if and only if ϕ is concave.

I make filtration, $\mathcal{F}_t$, explicit in (40) to emphasize that certainty equivalence functionals map information sets to utility. In particular, $\mathcal{I}$ is just the expectations operator, $\mathbb{E}\left[\cdot \,\middle|\, \mathcal{F}_{t-}\right]$, if preferences are time-separable.

Note, if $V(t)$ is continuous, it is predictable, i.e., $V(t) \in \mathcal{F}_{t-}$, then

$$\mathcal{I}\left[V(t) \,\middle|\, \mathcal{F}_{t-}\right] = \phi^{-1}\left(\mathbb{E}[\phi(V(t)) \,\middle|\, \mathcal{F}_{t-}\right) = \phi^{-1}\left(\phi(V(t))\mathbb{E}[1 \,\middle|\, \mathcal{F}_{t-}\right) = \phi^{-1}(\phi(V(t))) = V(t).$$ (41)

In other words, the recursive utilities can be appropriately reparameterized in terms of a time-separable preferences. This validity of this reparameterization is why Tsai and Wachter (2018, Theorem 5) find a single risk price is sufficient even in the presence of recursive utility as long as the underlying shocks are continuous.

To make this more concrete, consider the example of Epstein-Zin preferences. I adopt the notation used in Bansal and Yaron (2004). Let ρ denote risk aversion and ψ denote to the intertemporal elasticity of substitution (IES). Then Epstein-Zin Utility can be represented as:

$$U_t = \left[C_t^{1-1/\psi} + \exp(-\kappa\Delta)\mathbb{E}\left[U_{t+\Delta}^{1-\rho} \,\middle|\, \mathcal{F}_t\right]^{\frac{1-1/\psi}{1-\rho}}\right]^{\frac{1}{1-1/\psi}}$$ (42)

The formulation in (42) is not in the form given in (40), and so is not particularly useful for our purposes. Define $V_t := U_t^{1-1/\psi}$ and reparameterize (42) as

$$V_t = \left[C_t^{1-1/\psi} + \exp(-\kappa\Delta)\mathbb{E}\left[V_{t+\Delta}^{\frac{1-\rho}{1-1/\psi}} \,\middle|\, \mathcal{F}_t\right]^{\frac{1-1/\psi}{1-\rho}}\right].$$ (43)

Let $\phi(V) := \frac{1-\rho}{1-1/\psi} V^{\frac{1-\rho}{1-1/\psi}}$ and $U(C(t)) := C_t^{1-1/\psi}$, then we have: [24]

$$V_t = \left[u(C_t) + \exp(-\kappa\Delta)\phi^{-1}\left(\mathbb{E}\left[\phi\left(V_{t+\Delta}\right)\mid\mathcal{F}_t\right]\right)\right]. \tag{44}$$

9.2. The Investor's Portfolio Optimization Problem

To fix intuition, consider a one-period version of the model. The investor can continually trade between time 0 and time 1 and consumes her wealth at time 1 as displayed in Figure 11. At some time $\tau \in (0,1)$, a news item is released, on which the investor can trade. The investor's preferences satisfy the following utility recursion for any time $\tau > t$.

$$V_t(W_t) = u(C_t) + \phi^{-1}\left(\mathbb{E}\left[\phi\left(V_\tau(W_\tau)\right)\mid\mathcal{F}_{t-}\right]\right). \tag{45}$$

Assume that the investor has access to three assets. 1) A risk-less asset, $\chi_{f,t}$, that pays off 1 unit in every period, and whose price equals 1 because investors do not discount the future. Essentially, it is a costless storage technology. 2) An asset, ζ_t, whose payout, R_ζ, is announced by the news release. 3) An asset ξ_t whose payout R_ξ realizes as a Brownian motion, i.e., its variance and mean are both proportional to the length of the remaining interval. Figure 11 displays the timing. I maintain the convention where the time subscript refers to when the variable first enters the investor's information set.

Figure 11: Timing
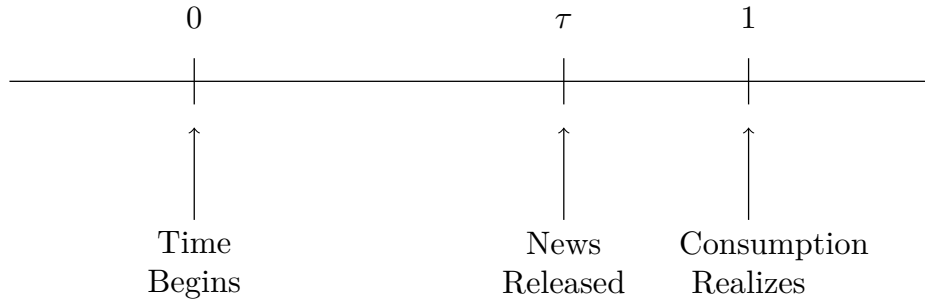
Since this is a finite-horizon problem, we can solve it by working backward. At time 1, all uncertainty has been resolved, and the representative agent eats all of her wealth:

$$V_1(W_1) = u(W_1). \tag{46}$$

Let $\tau < t < 1$, then the investor can trade ξ_t and the risk-less asset χ_t. However, since the news was already released, we know the payout of ζ_t, and so it is a risk-free asset. Consequently, the

---

24. The constant in front cancels between ϕ and ϕ^{-1}, and so does not affect the level of utility. It is there to ensure that ϕ is an increasing function regardless of the values of the parameters.

value function equals

$$V_t(W_t) = \max_{\xi} \phi^{-1}(\mathbb{E}\left[\phi\left(V_1\left(W_1\right)\right)|\mathcal{F}_t\right]), \text{ where } W_1 = W_t + (1-t)(R_\xi - 1)\xi_t, \qquad (47)$$

because she gets return 1 from the risk-less asset and R_ξ from the risky asset over the course of the entire interval. By substituting the constraint into the problem, and noting that $V_1(W_1) = u(W_1)$, this simplifies to

$$V_t(W_t) = \max_{\xi} \phi^{-1}\left(\mathbb{E}\left[\phi\left(u\left(W_t + (1-t)\xi_t(R_\xi - 1)\right)\right)|\mathcal{F}_t\right]\right). \qquad (48)$$

The first-order condition is

$$0 = \mathbb{E}\left[\phi'(u(W_1))u'(W_1)(1-t)(R_{\xi_t} - 1)\,\big|\,\mathcal{F}_t\right], \qquad (49)$$

since the term arising from the derivative of ϕ^{-1} is always positive.

The investor's problem for some time t in $(0, \tau)$ has similar structure except now she trades both assets.

$$V_t(W_t) = \max_{\zeta_t, \xi_t} \phi^{-1}(\mathbb{E}\left[\phi(V_\tau(W_\tau))|\mathcal{F}_t\right]) \qquad (50)$$

$$W_\tau = W_t + (R_\zeta - 1)\zeta_t + (\tau - t)(R_\xi - 1)\xi_t \qquad (51)$$

Substituting the constraints into the problem gives

$$V_t(W_t) = \max_{\zeta_t, \xi_t} \phi^{-1}\left(\mathbb{E}\left[\phi\left(V_\tau\left(W_t + (R_\zeta - 1)\zeta_t + (\tau - t)(R_\xi - 1)\xi_t\right)\right)|\mathcal{F}_t\right]\right). \qquad (52)$$

Taking first-order conditions with respect to the x for $x \in \{\zeta, \xi\}$ and simplifying gives

$$0 = \mathbb{E}\left[\phi'\left(V_\tau\left(W_\tau\right)\right)V_\tau'(W_\tau)\left(R_x - 1\right)\big|\mathcal{F}_t\right]. \qquad (53)$$

Consider some time immediately before τ, $\tau-$, and some time right after τ, $\tau+$. Then, substitute (49) into (52) and consider the derivative with respect to ξ_t:[25]

$$0 = \mathbb{E}\left[\mathbb{E}\left[\phi'\left(V_\tau\left(W_\tau\right)\right)V_\tau'(W_\tau)\big|\mathcal{F}_{t+}\right](d\xi)(R_\xi - 1)\big|\mathcal{F}_{\tau-}\right]. \qquad (54)$$

Since ξ is continuous, it is orthogonal to all discontinuous process, and so we can replace the expectation with respect to τ^+ with an expectation with respect to τ^-:

$$0 = \mathbb{E}\left[\mathbb{E}\left[\phi'\left(V_\tau\left(W_\tau\right)\right)V_\tau'(W_\tau)\big|\mathcal{F}_{t-}\right](R_\xi - 1)\big|\mathcal{F}_{\tau-}\right] \qquad (55)$$

---

25. I am taking limits here with respect to time loosely here to provide intuition. I make the statements rigorous in the theorems below.

Consequently, the investor only cares about the predictable part of the co-variation. In order for the returns to have finite variances, this must be proportional to the length of the interval. Their variance is proportional to $(\tau^+) - (\tau^-) \approx 0$. Hence, the Brownian asset ξ is risk-less over short enough intervals.

I substitute (49) into (52), and consider the derivative with respect to ζ_t. The jump asset, ξ_t, is not risk-less over short enough intervals:

$$0 = \mathbb{E}\left[\mathbb{E}\left[\phi'\left(V_\tau\left(W_\tau\right)\right)V_\tau'(W_\tau)(R_\zeta - 1)(d\zeta)\,\middle|\,\mathcal{F}_{\tau-}\right]\,\middle|\,\mathcal{F}_{\tau-}\right]. \tag{56}$$

We cannot pull R_ζ outside of the inner expectation because it is not predictable, i.e., it is not contained in the $\mathcal{F}_{\tau-}$ information set. Hence, there is no reason to expect the ex-post variation to be proportional to the length of the interval.

To facilitate comparing the two equations, isolate the unpredictable variation in the SDF by multiplying and dividing through by its left-limit, which we can pull through the inner expectation:

$$0 = \mathbb{E}\left[\phi'\left(V_{\tau-}\left(W_{\tau-}\right)\right)V_{\tau-}'(W_{\tau-})\mathbb{E}\left[\frac{\phi'\left(V_\tau\left(W_\tau\right)\right)V_\tau'(W_\tau)}{\phi'\left(V_{\tau-}\left(W_{\tau-}\right)\right)V_\tau'(W_\tau)}(R_\zeta - 1)(d\zeta)\,\middle|\,\mathcal{F}_{t-}\right]\,\middle|\,\mathcal{F}_{\tau-}\right] \tag{57}$$

Note, $\phi'(V_{\tau-}(W_{\tau-}))V_{\tau-}'(W_{\tau-})$ is predictable, while $\frac{\phi'(V_\tau(W_\tau))V_\tau'(W_\tau)}{\phi'(V_{\tau-}(W_{\tau-}))V_\tau'(W_\tau)}$ is purely unpredictable.

9.3. Deriving the Asset-Pricing Equation

The market environment is mostly standard. We need a series of technical conditions that ensure that preferences are reasonable and first-order conditions uniquely characterize the optimum.

Assumption. Market Environment

1. Both u and ϕ are Lipschitz continuous with Lipschitz derivatives.

2. $u : \mathbb{R} \to \mathbb{R}$ has strictly positive first-order derivatives, and ϕ is increasing with first- and second-order derivatives that are bounded away from zero and infinity.

3. A representative investor prices all assets.

4. Consumption — $C(t)$ — is an Itô semimartingale.

5. All of the stochastic processes do not contain predictable jumps.

The fourth assumption is the principal distinction from the setup in Ai and Bansal (2018). Assumption 6 generalize their assumptions by allowing consumption to jump. I will discuss later how my results slightly simplify if we require consumption to be continuous. The third assumption is likely unnecessarily restrictive, most of the results in this section would go through in terms of a marginal investor's preferences. I make this assumption to simplify the exposition.

Consider an representative investor with preferences given by (40). She has access to a (potentially infinite) vector of assets $\Xi(t) \coloneqq \xi_1(t), \dots$. Assume for simplicity that she has no other sources

of income. Over some small length of time Δ, the investor's problem is as follows. She enters into the period with asset allocation $\Xi(t - \Delta)$, and prices are $p(t)$.[26] She need to solve for consumption $C(t)$ and an asset allocation $\Xi(t)$. The results are reported cum-dividend to avoid introducing even more notation. The extension to the ex-dividend case is straightforward.[27]

Problem 1. Consumer's Portfolio Allocation

$$V\left(\Xi(t-\Delta), p(t)\right) = \max_{C(t),\Xi(t)} \int_t^{t+\Delta} u(C(s))\,ds + \phi^{-1}\left(\mathbb{E}\left[\exp(-\kappa\Delta)\phi\left(V\left(\Xi(t), P(t+\Delta)\right)\right)\mid\mathcal{F}_t\right]\right)$$

$$(58)$$

$$C(t) + \sum_i P_i(t)\xi_i(t-\Delta) = \sum_i P_i(t)\xi_i(t) \tag{59}$$

The continuous-time problem is the limit of Problem 1 as Δ approaches 0. The trade-offs are slightly easier to see in the discrete-time problem. The investor must purchase consumption, $C(t)$, and assets, $\Xi(t)$, at prices, $P_i(t)$, using wealth, $\sum_i P_i(t)\Xi(t-\Delta)$. Let $\widetilde{P}(t) = \exp\left(-\kappa(t)\right)p(t)$ be the appropriately discounted price. We are interested in excess returns, not returns themselves. Then we can derive the following result, where $p(t)$ refers to the price.[28]

Theorem 11 (Asset-Pricing Equation). *Let Assumption 6 hold, prices be Itô semimartingales, and the representative consumer face Problem 1 as $\Delta \to 0$. Assume preferences are such that optimal consumption is strictly positive. Define*

$$M^{UP}(t) := \frac{\phi'(V\left(W(t)\right)}{\phi'(V\left(W(t-\right))} \text{ and } M(t) := \frac{\phi'(V\left(W(t-\right))}{\phi'\left(\phi^{-1}\left(\mathbb{E}\left[\phi(V(W(t)))\mid\mathcal{F}_{t-}\right]\right)\right)} \frac{V'\left(W(t)\right)}{u'(c(t-))}. \tag{60}$$

Then $M^{UP}(t)$ is a purely discontinuous martingale, and for all stopping times $\tau > t$,[29]

$$\widetilde{P}(t) = \mathbb{E}\left[M(\tau)M^{UP}(\tau)\widetilde{P}(\tau)\,\Big|\,\mathcal{F}_t\right] \tag{61}$$

Conceptually, Theorem 11 is straightforward. Prices are semimartingales, and so we have a pricing kernel — $\mathcal{M}(t)$ — that prices all assets:

$$\widetilde{P}(t) = \mathbb{E}\left[\mathcal{M}(\tau)\widetilde{P}(\tau)\,\Big|\,\mathcal{F}_t\right]. \tag{62}$$

However, $\mathcal{M}(t) \not\propto V'(W(t))$. Instead, it has two parts: $M(t)$, which reflects compensation for consumption risk and $M^{UP}(t)$ which reflects compensation for discontinuities in the investor's information set.

---

26. The timing notation may seem somewhat strange here because it maintains the convention used elsewhere in the paper where time arguments denote when the objects first enter the representative investor's information set.

27. Cum-dividend means before dividend. Assets here behave like Bitcoin or gold and never pay out dividends.

28. This is a generalization of Ai and Bansal (2018, Theorem 2) to allow for jumps in consumption.

29. I use the UP superscript because M^{UP} is an unpredictable process.

9.4. Deriving Risk Premia

The end of the previous section is essentially where Ai and Bansal (2018) stop. Section 10 estimates
risk premia, and so I must derive risk premia from Theorem 11. If prices were continuous, Itô's
formula lets us solve for the expected log-return in terms of the covariance between the $M(t)$ and $p(t)$.
However, the generalized Itô's formula in the literature that applies to general semimartingale does
not have a simple form in terms of covariances. To resolve this impasse, I derive a generalized Itô's
formula in terms of predictable quadratic covariation, (integrated diffusive and jump volatilities)
that has the standard form but applies to jump processes.

Lemma 12 (An Itô's Formula for the Expectation of a Square Integrable Semimartingale)**.** *Let
f be a twice-differentiable function and $\widetilde{Z}$ be a vector-valued semimartingale with locally bounded
predictable $\langle Z \rangle(t)$. Then the differential of f satisfies*

$$d\mathbb{E}\left[f(\widetilde{Z}) \,\Big|\, \mathcal{F}_{t-} \right] = \mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) \,\Big|\, \mathcal{F}_{t-} \right] + \frac{1}{2} f''(\tilde{Z}(t-))\, d\langle \tilde{Z} \rangle(t). \tag{63}$$

The assumptions and conclusion in Lemma 12 are both weaker than Itô's formula for continuous
processes. We do not need continuous processes, but the equality only holds in expectation. How-
ever, this is sufficient for our purposes as risk premia are expectations. Importantly, the convexity
correction has the same form as it does in the standard Itô's formula.

I now compute risk premia by applying Lemma 12 to the logarithm. Let $m(t) \coloneqq \log(M(t))$ and
$m^{UP}(t) \coloneqq \log(M^{UP}(t))$. Recall that throughout, $p(t)$ refers the log-price. Let P_f denote the price
of the risk-free asset.

Theorem 13 (Asset-Pricing Equation)**.** *Let the assumptions in Assumption 6 hold, $P_i(t)$ be an
Itô semimartingales, and the representative consumer face Problem 1 as $\Delta \to 0$. Assume that
preferences are such that optimal consumption is strictly positive. Then risk-premia for some asset
i is*

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)} \,\bigg|\, \mathcal{F}_{t-} \right] = -\, d\langle m, p^D + p^J \rangle(t) - d\langle m^{UP}, p^J \rangle(t). \tag{64}$$

The cost of the assumptions' generality is that Theorem 13 is rather abstract. To make the
representation more concrete, consider a few specializations. First, assume that preferences are time-
separable and consumption is continuous. Then $M^J(t)$ is identically one, and $M(t) = \frac{u'(C(t))}{u'(C(t-))}$ by
the envelop theorem. Consequently, $\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)} \,\big|\, \mathcal{F}_{t-} \right] = -\sigma'_m \sigma_{p_i}$. This is the consumption-
CAPM model of Breeden (1979).

In addition, since $m(t)$ is continuous, which is implied by $u'(\cdot)$ being a smooth function and $C(t)$
being continuous, jumps do not command a premium. This is because time-separability implies
$m^{UP}(t)$ is identically zero and so $d\langle m, p^D + p^J \rangle(t) + d\langle m^{UP}, p^J \rangle(t) = d\langle m, p^D \rangle$.

If we allow for jumps and recursive utility, (64) generalizes Tsai and Wachter (2018, Theorem 5).
The key difference is that it is apparent that the second term is a covariance. Even in the presence
of jumps, risk premia can be split into a risk price and risk quantity. It is not immediately apparent

in Tsai and Wachter's (2018) environment that their formula is a covariance, but as long as returns have finite variance, we can rewrite their expression as a predictable quadratic variation.[30]

If we are in a world like Ai and Bansal (2018), where consumption is continuous and the envelop theorem holds (which implies $V'(W(t))$ is a continuous process), (64) simplifies to

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)}\,\middle|\,\mathcal{F}_{t-}\right] = -d\langle m, p^D\rangle(t) - d\langle m^{UP}, p^J\rangle(t). \tag{65}$$

If we further assume that high-frequency consumption movements can be ignored, i.e., $u'(c(t-)) = u'(C(t))$, we can combine the two terms using the law of iterated expectations.

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)}\,\middle|\,\mathcal{F}_{t-}\right] = -d\left\langle \frac{\phi'(V(Wt))}{\phi'(V(W(t-)))}, p(t)\right\rangle. \tag{66}$$

Again, we have a single risk-price and risk-premia equal $-\sigma'_m\sqrt{(\sigma_t^2 + \gamma_t^2)_i}$. Doing this is equivalent to assuming that the market wealth portfolio is the only risk factor and ignoring movements in the wealth-consumption ratio. However, as will be shown below, the data require a two-factor model.

We could instead assume that consumption is continuous and the envelop theorem holds, but high-frequency movements in consumption cannot be ignored. In that case, the risk premia equation is

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)}\,\middle|\,\mathcal{F}_{t-}\right] = -d\langle m, p^D\rangle(t) - d\langle m^{UP}, p^J\rangle(t). \tag{67}$$

In this case, we could isolate the effects of each of the two factors by estimating the risk-premia as a bivariate function of σ_t^2 and γ_t^2. In that case, we could view the second term as a measure of announcement premia, similar to how Ai and Bansal (2018) use the excess returns on FOMC days. However, then the regressions done below imply that ϕ is convex because γ_t^2 predicts lower risk premia once we condition in σ_t^2. Regardless, the data demand we have two factors that move at high-frequency. They also require the news risk premium to be less than the diffusion volatility premium.

10. News Premia: Empirics

Recall the formula for risk-premia in the presence of recursive utility and jumps derived in Theorem 13:

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)}\,\middle|\,\mathcal{F}_{t-}\right] = -d\langle m, p^D\rangle(t) - d\langle m^{UP}, p^J\rangle(t). \tag{68}$$

In general, we must to specify a full model for both $m(t)$ and $m^{UP}(t)$ in order to take (68) to the data. There are two leading cases. One, make a CAPM-style approximation that assumes market

---

30. This is implied by Lemma 12.

wealth is the only factor, i.e., $V_t = V(W_t)$.[31] In this case, we have

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_{rf}(t)}{P_{rf}(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = \beta_1(\sigma_t^2 + \gamma_t^2) + \beta_2\gamma_t^2 \qquad (69)$$

because $V(t)$ is perfectly correlated with $p(t)$. Two, assume that the news structure and underlying productivity shocks are continuous as in Ai and Bansal (2018). Then $V(t)$ is a continuous process and $\langle m, p^J \rangle = 0$. This gives

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_{rf}(t)}{P_{rf}(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = \beta_1\sigma_t^2 + \beta_2\gamma_t^2. \qquad (70)$$

10.1. Excess Return and Volatility: Contemporaneous Relationship

The question now facing us is how should we estimate (69) and (69). In practice, $\sigma_t^2 + \gamma_t^2$ and γ_t^2 are very heavily correlated, (89 %), 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 use $\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$ instead of γ_t^2. To make the results more Gaussian and avoid the need for interaction terms, I report elasticities, i.e., I apply a log transformation. Hence, the preferred specification is

$$rx_t = \beta_0 + \beta_1 \log\left(\sigma_t^2 + \gamma_t^2\right) + \beta_2 \log\left(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}\right) + \epsilon_t. \qquad (71)$$

I report robustness results in Section D. The results in the other specifications either agree with the main specification or are insignificant.

Consider the contemporaneous relationship between the volatility and the return. This section starts by replicating the standard result that volatility and returns are contemporaneously negatively correlated, (Lettau and Ludvigson 2010). The crucial difference between the results reported here and those in the literature is that Table 11 splits contemporaneous relationship up into relationships with $\sigma_t^2 + \gamma_t^2$ and with $\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$.

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 is obtained 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 the square brackets. I use Bartlett's kernel with the optimal bandwidth, per Newey and West (1994).

As can clearly be seen in Table 11, $\log(\sigma_t^2 + \gamma_t^2)$ and rx_t are strongly negatively correlated. This is what the literature has found Brandt and Kang (2004) and Lettau and Ludvigson (2010). The unconditional positive relationship between $\log(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2})$ and rx_t is new, however.

Weighted least squares is more efficient than ordinary least squares if we choose the weights appropriately. Section D reports weighted regressions that weight each datapoint by the inverse of that day's total volatility. This weighting is optimal up to unpredictable terms because $\sigma_t^2 + \gamma_t^2$

---

31. Under the assumptions in Assumption 6, this implies V_t jumps since W_t does.

Table 11: $\mathbb{E}\left[rx_t \mid \sigma_t^2 + \gamma_t^2, \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right]$ (OLS)

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

equals variance of the martingale part of $p(t)$ in expectation. This martingale part is the innovation in (71). In order to deal with any residual heteroskedasticity, Section D continues to use robust standard errors.

It is also worth noting that since the regressions are contemporaneous, the $\bar{\mathbb{R}}^2$'s that Table 11 reports are reasonable. The volatility explains a notable, but small, part of the variation in the excess return.

10.2. News Premia

The regressions in Table 11 are contemporaneous, and so they conflate risk premia and volatility feedback effects. If we try to interpret the coefficients as measures of risk premia, we have the classic endogenous regressors problem because the regressors and error terms are correlated.

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.[32] The most common way of handling endogenous regressors is using instrumental variables, and that is what I do.

In particular, I use the lagged regressors as instruments. This procedure gives better estimates than regressing on the lagged volatilities directly for three reasons. First, $\sigma_t^2 + \gamma_t^2$ and $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ are not AR(1) processes. Hence, regressing directly on $\sigma_{t-1}^2 + \gamma_{t-1}^2$ and $\frac{\gamma_{t-1}^2}{\sigma_{t-1}^2+\gamma_{t-1}^2}$ unnecessarily throws away useful information. Second, the coefficients from these this regression conflate predictability of the volatilities and risk premia. Consequently, they only identify the sign, not the magnitude of the risk premia. Third, as discussed in Section 10.1, returns are highly heteroskedastic. Since I consistently estimate $\sigma_t^2 + \gamma_t^2$, I can adjust for heteroskedasticity in the instrumented contemporaneous relationship. It is not obvious how to do this appropriately if you regress rx_t on $\sigma_{t-1}^2 + \gamma_{t-1}^2$ and $\frac{\gamma_{t-1}^2}{\sigma_{t-1}^2+\gamma_{t-1}^2}$.

The lagged volatilities are valid instruments. First, they explain a large amount of the variation

---

32. This is equivalent to regressing expected returns on volatilities, but we do not observe expected returns.

in the regressors. I adopt an approximate heterogeneous autoregressive (HAR) specification to choose lags used as instruments, (Corsi 2009). To be precise, I use $\sigma_{t-l}^2 + \gamma_{t-l}^2$, $\frac{\gamma_{t-l}^2}{\sigma_{t-l}^2 + \gamma_{t-l}^2}$ for $l \in \{1, 2, 5, 25\}$ as instruments. I report the results from the first-stage regressions in Table 17. The $\bar{\mathbb{R}}^2$ for the $\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$ regression equals 14.63 % with an associated F-statistic of 248.07. The $\bar{\mathbb{R}}^2$ for the $\sigma_t^2 + \gamma_t^2$ regression equals 79.43 % with an associated F-statistic of 20 140. Both of these are comfortably within the strong instruments region. Second, they are predetermined. Consequently, they are, by definition, independent of the date t innovation. Innovations cannot be predicted.

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.

Table 12: News Premia Estimates

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

The question facing us is how do we interpret the coefficients reported in Table 12. If $\log(V_t)$ is proportional to wealth and both consumption and wealth move at high-frequency, then the coefficient on $\sigma_t^2 + \gamma_t^2$ measures risk aversion. The coefficient on $\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$ measures the curvature of the CEF as parameterize by ϕ. Per the discussion in Section 9.4, this implies the CEF is convex, and if investors have Epstein-Zin preferences they prefer late resolution of uncertainty.

Conversely, if consumption is the only factor and is continuous, the coefficient on $\log(\sigma_t^2)$ measures risk aversion. The coefficient on $\log(\gamma_t^2)$ measures the curvature of the CEF. Here, though, the sign on that term changes if we include $\log(\sigma_t^2)$.

The obvious question is why is this? This is likely because the univariate regression on γ_t^2 suffers from the classic endogenous regressors problem. I showed in Table 7 that $\log(\sigma_t^2)$ and $\log(\gamma_t^2)$ are highly positively correlated. Since risk aversion implies that σ_t^2 commands a premia and the two volatilities are highly correlated, the univariate regression misattributes risk premia driven by risk aversion to news premia.[33]

This implies that the correctly specified regressions are the bivariate ones. In both cases we have that risk aversion results in diffusive risk commanding a large, positive premium. The news

---

33. The regression in terms of $\log(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2})$ does not suffer from this exogeneity problem to as near a large extent because it is not nearly as heavily correlated.

risk premium is substantially smaller in both cases.

We want to interpret the magnitude of the coefficients, not just the sign. Since I regress annualized excess log-return on the log total volatility and log jump proportion, the estimates are elasticities. These elasticities are highly statistically and economically significant. For example, consider the first row. The elasticity of rx_t with respect to $\sigma_t^2 + \gamma_t^2$ is 0.24. In other words, a $1\,\%$ increase in $\sigma_t^2 + \gamma_t^2$ for the course of an entire year increases the expected yearly return by $0.24\,\%$.[34] 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 only $6.13\,\%$, but the regression coefficient is significantly larger. A $1\,\%$ change in $\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$ over the course of changes expected yearly returns by $-5.01\,\%$. In both cases, the implied movements in risk premia from year to year are very large. I am not the first researcher to find movements in risk premia that are this large; Martin (2017) reports similarly-sized changes.

I consider several other specification in Section D. The volatility coefficients are robust to the heteroskedasticity correction and the particular instruments chosen, (Table 18). Results from running the regression over a subsample either agree with the main results or are not statistically significant, (Table 20).

10.3.   Robustness Checks

Estimating risk premia is difficult because the signal-to-noise ratio is quite low. The literature has pointed out some issues that can bias the empirical estimates. Perhaps the most important is the Stambaugh bias, (Stambaugh 1999). He shows that a finite-sample bias can inflate coefficient estimates if the regressors are stochastic. However, the regressions are run at the daily frequency, not the monthly frequency as is commonly done. Hence, I have approximately 3700 datapoints. Since this bias decreases at a $\frac{1}{\#\text{ datapoints}}$ rate, this bias should not noticeably affect my estimates.

The other significant sources of bias noted in the literature are also not nearly as significant here because I use daily data. For example, regressing long-horizon returns on persistent regressors causes the $\mathbb{R}^2$ to spuriously increase with the horizon under certain conditions. However, I am not using long-horizon returns, and so this does not apply. Various authors also have used overlapping returns to increase their effective sample size, which can invalidate the inference. I do not use overlapping returns, and so this also does not apply.

There is one primary source of error that is worth pointing out. The regressors that I use are estimated from high-frequency data. Consequently, we may have an error-in-regressors problem. This problem should not be a significant issue for three reasons. First, since I have a great deal of intraday data, the regressors should be estimated precisely. Second, the main empirical source of estimation error is separating the diffusion and jump components, and this should be independent of the expected returns because it only depends on the magnitude of the high-frequency returns, not

---

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

their sign. Besides, it does not even affect estimating $\sigma_t^2 + \gamma_t^2$. Third, and most importantly, as I am instrumenting for the volatilities by their lags and the estimation error is likely independent across time, both the coefficient estimates and their standard errors should be asymptotically valid.

11.   Conclusion

This paper investigates how jumps affect investors' risk. I first show that standard no-arbitrage based pricing theory implies that jumps are price responses to news shocks. When a news shock hits causing the representative investor's information set to jump, she responds by pricing assets differently. Having done that, I introduce jump volatility — γ_t^2 — which is a sufficient statistic for the jump part of price dynamics. I then introduce the realized density, RD_t, to reduce tracking the returns' predictive density — $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ – to forecasting γ_t^2 and σ_t^2. I do this by providing a new representation for infinite-activity jump processes as integrals with respect to a variance-gamma process. I then develop nonparametric estimators for the instantaneous and integrated jump and diffusion volatilities and for the realized density to enable taking these representations to the data.

I apply these estimators to the S&P 500 using high-frequency data from SPY. I find that jumps drive approximately one-half of the ex-post squared variation and that this proportion varies substantially over time. I also evaluate the performance of the estimators in simulations and find that my estimators perform well in estimating the volatilities. I then consider the behavior of these estimators in the data providing several new stylized facts. I show that the jump volatility is relatively well-behaved and has a bell-shaped distribution after applying a logarithmic transformation. In other words, the volatilities are roughly log-Gaussian. Finally, I show that γ_t^2 is both very persistent, having long-memory, and highly correlated with σ_t^2.

I next analyze how jumps affect expected returns. In particular, I show that risk premia have the following form — $-d\langle m, p\rangle(t) - d\langle m^{UP}, p^J\rangle$ — where $m(t)$ is the predictable part of the log-SDF and $m^{UP}(t)$ is the unpredictable part. I further relate $m(t)$ and $m^{UP}(t)$ to the curvature of the utility function and the certainty equivalence functional. The theory requires two factors that move at high-frequency in general. I show that the premium associated with γ_t^2 is statically and economically significantly less than the one associated with σ_t^2. This divergence implies that investors preferences are not time-separable and that the data require two factors that move at high-frequency as well.

As this work introduces the jump volatility, a great deal of work still needs to be done. One prominent question is how to generalize the theory and empirics to higher dimensions. Can we derive a similar multivariate representation and estimators for the jump processes? Doing this will require figuring out what the appropriate multivariate Laplace distribution is. Presently, several multivariate Laplace distributions exist, but it is not apparent any of them have the proper relationship to Poisson and Gaussian processes. Moreover, this paper shows the proposed estimators are consistent in the noise-free case. Deriving the relevant inference theory in the presence of market microstructure noise would be quite useful.

Second, previous authors have shown that the stylized features of σ_t^2 are relatively stable across

different assets. Is this also true for γ_t^2? For example, people have argued that news risk is fundamental in understanding foreign exchange markets. How does γ_t^2 act in those environments? Third, on the financial side, a great deal more empirical and theoretical work is needed to fully understand the relationship between the premia associated with σ_t^2 and γ_t^2. A fully specified general equilibrium model that determines the correct underlying risk factors would be useful to rationalize the new empirical evidence.

Fourth, since this paper reduces forecasting returns' distributions to forecasting the volatilities, it greatly simplifies tracking time-varying tail risk. Consequently, building a joint dynamic model for both volatilities and the drift and analyzing the resulting models' performance in tracking tail risk would be extremely useful.[35]

References

Ai, Hengjie, and Ravi Bansal. 2018. "Risk Preferences and the Macro Announcement Premium." *Econometrica* 86 (4): 1383–1430.

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.

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

Aït-Sahalia, Yacine, Jean Jacod, and Jia Li. 2012. "Testing for Jumps in Noisy High Frequency Data." *Journal of Econometrics* 168:207–222.

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.

---

35. This is the project I tackle in "Jumps, Tail Risk, and the Distribution of Stock Returns."

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2001. "The Distribution of Realized Exchange Rate Volatility." *Journal of the American Statistical Association* 96 (453): 42–55.

———. 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.

———. 2007. "Real-time Price Discovery in Global Stock, Bond and Foreign Exchange Markets." *Journal of International Economics* 73 (2): 251–277.

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.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *The Journal of Finance* 59 (4): 1481–1509.

Baricz, Árpád. 2008. "Mills' Ratio: Monotonicity Patterns and Functional Inequalities." *Journal of Mathematical Analysis and Applications* 340 (2): 1362–1370.

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.

———. 2006. "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation." *Journal of Financial Econometrics* 4 (1): 1–30.

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.

Beechey, Meredith J., and Jonathan H. Wright. 2009. "The High-Frequency Impact of News on Long-Term Yields and Forward Rates: Is it Real?" *Journal of Monetary Economics* 56 (4): 535–544.

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31 (3): 307–327.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988a. *Journal of Political Economy* 96 (1): 116–131.

———. 1988b. "A Capital Asset Pricing Model with Time-Varying Covariances." *Journal of Political Economy* 96 (1): 116–131.

Bollerslev, Tim, Tzuo Hann Law, and George Tauchen. 2008. "Risk, Jumps and Diversification." *Journal of Econometrics* 144:234–256.

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg. 2016. "Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting." *Journal of Econometrics* 192 (1): 1–18.

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.

Breeden, Douglas T. 1979. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." *Journal of Financial Economics* 7 (3): 265–296.

Campbell, John Y. 1987. "Stock Returns and the Term Structure." *Journal of Financial Economics* 18 (2): 373–399.

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.

Dickey, David A., and Wayne A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica:* 1057–1072.

Drechsler, Itamar, and Amir Yaron. 2011. "What's Vol Got to Do with It." *The Review of Financial Studies* 24 (1): 1–45.

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.

Duffie, Darrell, and Larry G. Epstein. 1992. "Stochastic Differential Utility." *Econometrica* 60 (2): 353–394.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica:* 987–1007.

Engle, Robert F., and Victor K. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *The Journal of Finance* 48 (5): 1749–1778.

Epps, Thomas W., and Mary Lee Epps. 1976. "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis." *Econometrica* 44 (2): 305–321.

Epstein, Larry G., and Martin Schneider. 2003. "Recursive Multiple-Priors." *Journal of Economic Theory* 113 (1): 1–31.

Epstein, Larry G., and Stanley E. Zin. 1989. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* 57 (4): 937–969.

Ergin, Haluk, and Faruk Gul. 2009. "A Theory of Subjective Compound Lotteries." *Journal of Economic Theory* 144 (3): 899–929.

Faust, Jon, John H. Rogers, Shing-Yi B. Wang, and Jonathan H. Wright. 2007. "The High-Frequency Response of Exchange Rates and Interest Rates to Macroeconomic Announcements." *Journal of Monetary Economics* 54 (4): 1051–1068.

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. 2002. "Stochastic Volatility, Jumps and Hidden Time Changes." *Finance and Stochastics* 6 (1): 63–90.

Geweke, John, and Susan Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4 (4): 221–238.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2005. "There is a Risk-Return Trade-off after All." *Journal of Financial Economics* 76 (3): 509–548.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with Non-Unique Prior." *Journal of Mathematical Economics* 18 (2): 141–153.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *The Journal of Finance* 48 (5): 1779–1801.

Hansen, Lars Peter, and Thomas J. Sargent. 2001. "Robust Control and Model Uncertainty." *The American Economic Review* 91 (2): 60–66.

Harvey, Campbell R. 1989. "Time-Varying Conditional Covariances in Tests of Asset Pricing Models." *Journal of Financial Economics* 24 (2): 289–317.

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

Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. 2009. "Microstructure Noise in the Continuous Case: The Pre-averaging Approach." *Stochastic Processes and their Applications* 119 (7): 2249–2276.

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.

Ju, Nengjiu, and Jianjun Miao. 2012. "Ambiguity, Learning, and Asset Returns." *Econometrica* 80 (2): 559–591.

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.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica* 73 (6): 1849–1892.

Kozubowski, Tomasz J., and Rachev T. Svetlozar. 1994. "The Theory of Geometric Stable Distributions and Its Use in Modeling Financial Data." *European Journal of Operational Research* 74 (2): 310–324.

Kreps, David M., and Evan L. Porteus. 1978. "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica* 46 (1): 185–200.

Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54 (1–3): 159–178.

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.

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard. 2015. "Does Anything Beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes." *Journal of Econometrics* 187 (1): 293–311.

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.

Lucca, David O., and Emanuel Moench. 2015. "The Pre-FOMC Announcement Drift." *The Journal of Finance* 70 (1): 329–371.

Madan, Dilip B., Peter P. Carr, and Eric C. Chang. 1998. "The Variance Gamma Process and Option Pricing." *Review of Finance* 2 (1): 79.

Mancini, Cecilia. 2001. "Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion." *Giornale dell'Istituto Italiano degli Attuari* 64:19–47.

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.

Merton, Robert C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41 (5): 867–887.

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

Monroe, Itrel. 1978. "Processes that can be Embedded in Brownian Motion." *The Annals of Probability* 6 (1): 42–56.

Nelson, Daniel B. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica:* 347–370.

Neuberger, Anthony. 2012. "Realized Skewness." *The Review of Financial Studies* 25 (11): 3423.

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.

Nikeghbali, Ashkan. 2007. "Non-stopping Times and Stopping Theorems." *Stochastic Processes and their Applications* 117 (4): 457–475.

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

Ornthanalia, Chayawat. 2014. "Lévy Jump Risk: Evidence from Options and Returns." *Journal of Financial Economics* 112 (1): 69–90.

Pagan, Adrian R., and Y.S. Hong. 1991. "Nonparametric Estimation and the Risk Premium." Chap. 2 in *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics,* edited by William A. Barnett, James Powell, and George Tauchen, 51–75. Cambridge University Press.

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.

Reisen, Valderio A. 1994. "Estimation of the Fractional Difference Parameter in the ARIMA(p, d, q) Model Using the Smoothed Periodogram." *Journal of Time Series Analysis* 15 (3): 335–350.

Santa-Clara, Pedro, and Shu Yan. 2010. "Crashes, Volatility, and the Equity Preimum: Lessons from S&P 500 Options." *The Review of Economics and Statistics* 92 (2): 435–451.

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

Strzalecki, Tomasz. 2013. "Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion." *Econometrica* 81 (3): 1039–1074.

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

———. 2011. "Econometric Analysis of Jump-Driven Stochastic Volatility Models." *Journal of Econometrics* 160 (1): 12–21.

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.

Tsai, Jerry, and Jessica A. Wachter. 2018. *Pricing Long-Lived Securities in Dynamic Endowment Economies.* Working Paper, Working Paper Series 24641. National Bureau of Economic Research, May.

Yu, Jun. 2005. "On Leverage in a Stochastic Volatility Model." *Journal of Econometrics* 127 (2): 165–178.

Appendix A Representation Theorems

Theorem 2 (Jump Volatility and the Predictable Quadratic Variation)**.** *Let $p(t)$ be an Itô semimartingale satisfying Assumption Square-Integrable, then the following holds where $\langle p^J \rangle(t)$ is the predictable quadratic variation (angle-bracket) of $p^J(t)$:*

$$\gamma_t^2 = \int_{t-1}^t \gamma^2(s)\,ds = \int_{t-1}^t \int_{\mathcal{X}} \delta^2(s,x)\nu(dx,ds) = \langle p^J \rangle(t) - \langle p^J \rangle(t-1). \tag{14}$$

Proof.

$$[p]^J(t) = \sum_{s \leq t} \Delta p(s)^2 = \int_0^t \int_{\mathcal{X}} \delta^2(s,x)\mu(ds,dx) \tag{72}$$

This comes from the view of the jumps as integrals with respect to Poisson random measures and there being no predictable jumps. Intuitively, the compensator ν does not jump and realizations of μ are equal 1 which does not change when squared.

$$\implies \langle p \rangle^J(t) = \mathbb{E}\left[[p]^j(t)\,\big|\,\mathcal{F}_{t-}\right] = \mathbb{E}\left[\int_0^t \int_X \delta^2(s,x)\mu(ds,dx)\,\bigg|\,\mathcal{F}_{t-}\right] = \int_0^t \int_X \delta^2(s,x)\nu(ds,dx) \tag{73}$$

We also need to show that the limit in the expectation form approaches $\gamma^2(t)$.

Define $\gamma^2(t) := \int_X \delta^2(t,x)\nu(dx,dt)$, then

$$\lim_{\Delta \to 0} \frac{1}{\Delta}\mathbb{E}\left[\left|p^J(t+\Delta) - p^J(t)\right|^2\,\Big|\,\mathcal{F}_{t-}\right] = \lim_{\Delta \to 0} \frac{1}{\Delta}\mathbb{E}\left[\left|\int_t^{t+\Delta} \delta(s,x)(\mu - \nu)(ds\,dx)\right|^2\,\bigg|\,\mathcal{F}_{t-}\right]. \tag{74}$$

By the Itô Isometry, we can rewrite (74) as

$$= \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E} \left[\int_t^{t+\Delta} \delta^2(s, x) \mu(ds\,dx) \,\Big|\, \mathcal{F}_{t-} \right] \tag{75}$$

Then by choosing δ so that dx, ds are independent, and the projection of ν onto the Lebesgue measure is constant.

$$= \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E} \left[\int_t^{t+\Delta} \int_{\mathcal{X}} \delta^2(s, x)(dx\,dx) \,\Big|\, \mathcal{F}_{t-} \right] \tag{76}$$

$$= \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E} \left[\Delta \gamma^2(t) + \int_t^{t+\Delta} \left(\gamma^2(t) - \gamma^2(s) \right) ds \,\Big|\, \mathcal{F}_{t-} \right] \tag{77}$$

We can split this into the value of the jump volatility at t and deviations from it.

$$= \lim_{\Delta \to 0} \gamma^2(t) + \frac{1}{\Delta} \Delta O \left(\mathbb{E} \left[\left| \sup_{t \le s \le t+\Delta} \gamma^2(t) - \gamma^2(s) \right| \,\Big|\, \mathcal{F}_{t-} \right] \right) = \gamma^2(t). \tag{78}$$

$\square$

Theorem 4 (Time-Changing Jump Martingales). *Let $p^J(t)$ be a purely discontinuous, martingale satisfying Assumptions* Square-Integrable, Infinite-Activity Jumps, *and* No Unpredictable 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 Levy measure.*

Then $p^J(t)$ time-changed by its predictable quadratic variation is a standard variance-gamma process. In other words, $p^J(t) \stackrel{\mathcal{L}}{=} \mathcal{L} \left(\langle p^J \rangle(t) \right)$.[36]

Proof. To prove the result, I start with a representation of a purely-discontinuous martingale as an integral with respect to a Poisson random measure. This is a two-dimensional representation of the jump process with all of the dynamics contained in the predictable process H. There are two key parts to the theorem above. First, we must handle the dynamics contained in H, and second we need to reduce the two-dimensional representation to a one-dimensional one.

We know that there are only finitely-many jumps in any strip that is bounded away from 0, but infinitely-many in any interval containing 0. To maintain this intuition, I switch the base Levy measure to one that has this property. Second, I make a time-change argument in each strip to deal with its dynamics. Third, I switch from an integral with respect to infinitely-many Poisson processes to one with respect to a Poisson random measure by taking the appropriate sum of these processes. I use capital letters to refer to processes as is standard in the literature. Since I do not discretize, there should be no confusion. Define $1^z = 1\{x \in [z, z+dz]\}$ where $z \in \mathbb{R}$, and $dz \in \mathbb{R}_+$, where I suppress dz in the notation. Similarly, for a process X define $X^z = X * 1^z$. In words, X^z is the process X restricted to the strip $[z, z = dz]$.

---

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

Denote $p(t)$ by $Y(t)$. I now switch the representation of Y as an integral with respect to a Poisson random measure with more intuitive properties. Y is locally-square integrable, and hence $\langle Y \rangle$ is well-defined, that is for any stopping-time τ, the stopped-process $\langle Y \rangle^\tau$ is almost surely finite. Since Y is a purely-discontinuous process, Y^z is a two-dimensional sum. To put in mathematical notation, $(Y^z)^\tau = \sum_{s \le t, x \in [z, z+dz]} \delta(x, s)$, where δ is a predictable Dirac delta. Also, define $\langle X \rangle^{-1}(t) = \inf\{\tau : \langle X \rangle = t\}$ for any process X. This is the standard inverse definition when the process may be zero, and is innocuous here because if $\langle X \rangle \overset{a.s.}{=} 0 \implies X \overset{a.s.}{=} 0$.

Recall that I assumed that the base measure of μ was the Lebesgue measure λ. $\left(\frac{1}{z}\right)$ is an infinite-measure and is absolutely continuous with respect to Lebesgue measure in any interval not containing zero. Let $\tilde{\mu}$ be a Poisson random measure with associated Levy measure $\left(\frac{1}{z}\right)$. Throughout the rest of proof, I use tilde's to refer to measures associated with this random measure. For example, $\tilde{\nu}$ is its associated compensator. Note, since I am using compensated random measures, each strip $[z, z+dz]$ is a martingale.

The benefit of using this representation is that it implies the associated predictable integrator, $\widetilde{H}$, is $O_p(1)$. In the original case, the local square-integrability of Y implies that $H(x, t)$ as a function of x is $O_p\left(\frac{1}{x}\right)$. Effectively, I am moving the necessary reduction in the intensity of the process as the jump size increases into the Poisson random measure instead of the integrator.

It is worth noting that in general we cannot choose $\tilde{H}$ to be proportional to a constant; it might be zero. However, since we have an infinite-activity process, we can without loss of generality. In addition, the Poisson processes formed by restricting the Poisson random measure to a strip in $\mathbb{R}$, $\tilde{X}^z$, have intensity measures, $\nu(x) = x^{-1} \exp(-x)\, dx$, which I use in the sequel.

I now turn to using a time-change argument to handle the dependence of Y^z, or equivalently, H^z. Since $\tilde{X}^z$ is a finite-activity Poisson process, its intrinsic filtration is the filtration generated by the jump locations. Let $t^{\tilde{X}^z}$ be a jump time for the process $\tilde{X}^z$, and consider the set $\{t < t^{\tilde{X}^z}\}$. This set is optional, but not predictable, and its ending time $\hat{t}$, is not a stopping time with respect to the predictable filtration. (It is what is known in the literature as an honest time.) This allows us to define the minimal enlargement of the filtration of Y^z, $\mathcal{F}_t^{Y^z}$ so that the $\hat{t}$ are stopping times.

$$\hat{\mathcal{F}}_t^{X^n} := \cap_{\epsilon > 0} \mathcal{F}_{t-\epsilon}^{\tilde{X}^z} \cup \sigma(\{\rho < t\}) \tag{79}$$

It is worth noting that when you progressively enlarge a filtration with an honest time, semi-martingales with respect to the original filtration are still semimartingales with respect to the new filtration (Barlow 1978). However, this enlargement does not necessarily preserve the martingale structure. Since I am doing this almost surely only finitely many times and jumps of the original process are almost surely unique, it is without loss of generality to consider the case with only one jump.

Consider X stopped at some time ρ that is a stopping time with respect to the expanded filtration, not to the original one. It is worth noting that we are expanding the predictable filtration $\mathcal{F}_{t-}$, not the original filtration. So using the Nikeghbali (2007, eqn 2.3), we can define the martingale on the new space. Then $X(t)$ has the following form, where $Z_t^\rho := \Pr[\rho > t \mid \mathcal{F}_{t-}]$, is chosen by to

be cádlág. The $\mathcal{F}_t$ dual optional projection of the process $1\{\rho \leq t\}$ is denoted by $A^\rho(t)$. Let $\hat{X}$ be a martingale with respect to $\widehat{\mathcal{F}}_t$. Also, define $\mu^\rho(t) = \mathbb{E}\left[A^\rho(\infty)\,|\,\mathcal{F}_{t-}\right] = A^\rho(t) + Z^\rho(t)$. Then

$$X(t) = \widehat{X}(t) + \int_0^{t \wedge \rho} \frac{d\langle X, \mu^\rho(t)\rangle(s)}{Z^\rho(s-)} - \int_\rho^t \frac{d\langle X, A^\rho(t) + Z^\rho(t)\rangle(s)}{1 - Z^\rho(s-)}. \tag{80}$$

Since $\mu^\rho(t)$ is $\mathcal{F}_{t-}$ measurable, and the jumps are distributed according to a Poisson process, $\mu^\rho(t)$ is a constant. Consequently, the predictable quadratic variation terms in (80) terms are almost surely zero.

Consider the process $\hat{X}^z$, where $\tilde{X}^z$ and $\hat{X}^z$ are equal pathwise, but we change the filtration from $\mathcal{F}_t$ to $\hat{\mathcal{F}}_t$. Since the stopping times of $\widetilde{X}^z$ are sufficient to generate its filtration, and $\widetilde{H}$ is predictable, we can choose $\widehat{\mathcal{F}}_t^{X^z}$ to be generated by the predictable σ-algebra. Equivalently, it is generated by the continuous processes. As a result, for any process adapted to this filtration there exists a continuous process that is equal to it in probability. Since equality in distribution is weaker than equality in probability, it is without loss of generality to assume that the process is continuous, and so I do so.

By the Dambis, Dubins & Schwarz theorem, we know that a continuous process is a Wiener process when time-changed by its quadratic variation. Therefore, $\hat{Y}^z([Y^z]) \overset{\mathcal{L}}{=} W$, where W is the standard Wiener process. Intuitively, we can view the jump magnitudes as appropriately rescaled Gaussian random variables.

However, this is not the filtration generated by the data, and so we need to consider what this representation implies about the original filtration. We start by considering the precise relationship between the predictable and quadratic variations both within and between each of the filtrations.

$\langle \hat{Y}^z \rangle \overset{a.s.}{=} [\hat{Y}^z]$ because all of the adapted processes in $\hat{\mathcal{F}}_t$ are predictable. In addition, changing the filtration does not change the quadratic variation because the process is optional and adapted, and all the change of filtration is doing is turning optional processes into predicable ones.

Therefore, the key question is what is the relationship between the $\langle Y^z \rangle$ and $[Y^z]$ in the original filtration. The quadratic variation of an integral with respect to a finite-activity Poisson process is $[H^z * X^z] = \sum_{s \leq t}(H^Z)^2(\hat{\tau})$, where the $\hat{\tau}$ are the jump locations.

Since $\tilde{X}^z$ is a Poisson process, the amount of time between jumps, that is the length of the intervals define above, is an exponential random variable with intensity $\hat{\nu}^z$. Since $\hat{\nu}^z$ is a deterministic function, $\hat{F}^n$ is an exponential-time change of $\tilde{\mathcal{F}}$. Therefore, $Y^z = H^z * X^z$ is Wiener process after both an exponential time-change and then a continuous-time change in the transformed space.

There are two main limitations of this result. First, the exponential time-change is not identified, and so we cannot use it for inference. Second, we want an expression for Y not just for each of the Y^z.

The first problem can be resolved by recalling that if the expectations of a sufficiently general class of functions are the same between two processes, then the processes equal in distribution. A sequence of nested expectations does not change if we reorder the nesting as long as the σ-algebras we are conditioning on are independent. However, because the exponential-time change was with

respect to a Poisson process with a deterministic compensator and the other time-change was with respect to a predictable process, the relevant filtrations are independent here. Consequently, we have that if we time-change Y^z by $\langle Y^z \rangle$, then we have a Wiener process with an exponential subordinator.

To resolve the second problem, that is aggregate over the strips correctly, note what happens if we aggregate all of the μ^z together. $\tilde{\mu}^z$ is a Poisson random variable with intensity measure $\tilde{\nu}(z) = z^{-1} \exp(-z) dz$. However, the definition of the Gamma process is that its intensity measure over strips is precisely the expression above.

For a countable partition of $\mathbb{R}$, $z_1, z_2 \ldots$, $Y = \sum_{z_i} Y^{z_i}$, and $\tilde{H} = \sum_{z_i} \tilde{H}^{z_i}$, and $\tilde{\mu} = \sum_{z_i} \tilde{\mu}^{z_i}$. Furthermore, Wiener processes are stable under countable sums as long as the variance remains finite, which it will in this context because the initial process is locally-square integrable. Consequently, we can do the following.

$$\lim_{I \to \infty} \sum_{i \leq I} Y^z \left(\langle Y^z \rangle^{-1} \right) \stackrel{\mathcal{L}}{=} \lim_{I \to \infty} W(\exp^{\sum \nu^i}) = W(\Gamma(t)) = \mathcal{L} \tag{81}$$

To wrap it up, if we time-change a purely-discontinuous, jump process with infinite-variation by its predictable quadratic variation, we get the variance-gamma process, also known as a standard variance-gamma process.

$\square$

Corollary 4.2 (Time-Changing Finite-Activity Jump Martingales)**.** *Let $p^J(t)$ be a purely discontinuous martingale satisfying Assumptions Square-Integrable and No Unpredictable 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 Levy measure.*

Then $p^J(t)t$ time-changed by its predictable quadratic variation is a mixture of the 0 process — δ_0 — and the standard variance-gamma process where the mixing weights are the intensity of the jump process.

Proof. Since Y is a finite-activity jump process, $(\mu - \nu) * 1^z$ is almost-surely zero as a function of z for all but a finite-subset of $\mathbb{R}$. For a segment of time when there are no jumps, the process is identically 0. You cannot time-change a process by the 0 process. Therefore, the proof of the main theorem where we take the limit of the number of strips to infinity is no longer valid.

However, if we split event-space Ω into spaces where $[Y] > 0$ and $[Y] = 0$, then in the first subset we can make the argument I made above, while in the second subset we have the 0 process. δ_0 is not affected by time changes, and so if we time-change both subsets by $\langle Y \rangle$, we do not affect the distribution. As a result, the time-changed distribution is a mixture of δ_0 and $\mathcal{L}$ where the mixture weight depends upon the intensity of the process ν. That is we have the following.

$$Y(t) = \begin{cases} \mathcal{L}\left(\langle Y \rangle(t)\right) & \text{with intensity } \nu \\ \delta_0(t) & \text{with intensity } 1 - \nu \end{cases} \tag{82}$$

$\square$

Corollary 4.1 (Jumps Processes as Integrals)**.** *Let $p^J(t)$ be a Itô semimartingale satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Unpredictable 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.*

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

Let $J(t)$ be the purely-discontinuous martingale part of $Y(t)$, then Theorem 4 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 $\mathcal{L}$), $J(t) = \int_0^t k(s) \, d\mathcal{L}(s)$. This is completely analogous to how the time-changed theorem for continuous processes and absolute continuity imply the integral representation of continuous martingales.

$\square$

Theorem 5 (Realized Density Representation)**.** *Let $p(t)$ be an Itô semimartingale satisfying Assumptions Square-Integrable, Infinite-Activity Jumps, and No Unpredictable Jumps. Let $\sigma^2(t)$ and $\gamma^2(t)$ be semimartingales whose martingale components are independent of the martingale components of $p(t)$. Then*

$$RD_t = N\left(\int_{t-1}^t \mu(s) \, ds, \int_{t-1}^t \sigma^2(s) \, ds\right) * \mathcal{L}\left(0, \int_{t-1}^t \gamma^2(s) \, ds\right), \tag{25}$$

and the predictive density is

$$h\left(r_t \mid \mathcal{F}_{t-1}\right) = \int_{\mu_t, \sigma_t^2, \gamma_t^2} RD_t(\mu_t, \sigma_t^2, \gamma_t^2) \, dG\left(\mu_t, \sigma_t^2, \gamma_t^2 \mid \mathcal{F}_{t-1}\right). \tag{26}$$

Proof. Consider the diffusion part of the process.

$$h\left(p^D(t) - p^D(t-1) \mid \mathcal{F}_{t-1}\right) = h\left(\sum_{n \in \frac{1}{\Delta}, \dots, 0} \int_{t-n\Delta}^{t-(n+1)\Delta} \sigma(s) \, dW(s) \,\middle|\, \mathcal{F}_t\right) \tag{83}$$

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.

$$= h\left(\sum_{n \in 1, \frac{1}{\Delta}} \sigma(t - n\Delta) \int_{t-n\Delta}^{t-(n+1)\Delta} dW(s) \,\middle|\, \mathcal{F}_{t-1}\right) \tag{84}$$

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.

$$\stackrel{\mathcal{L}}{=} h\left(\sqrt{\int_{t-1}^{t}\sigma^2(s)\,ds}\int_{t-1}^{t}dW(s)\,\bigg|\,\mathcal{F}_{t-1}\right) \tag{85}$$

$$\stackrel{\mathcal{L}}{=} \int_{\sigma_t^2}N\left(0,\int_{t}^{t+\Delta}\sigma^2(s)\,ds\right)dG\left(\sigma_t^2\,\big|\,\mathcal{F}_{t-1}\right) \tag{86}$$

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 is the square-root of one-half the variance. Consequently, when you pull the variance outside of the integral, you get an additional $\sqrt{2}$ in the denominator.

You can just carry the mean through the analysis, and then add it back in when you are done. To combine the jump and diffusion realized densities, note that density of a independent variables are convolutions of the densities. The integrators are pure-jump and diffusive martingales, and so they are automatically 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 because it is simply the function inside the integral.

$\square$

Appendix B   Volatility Estimation

Lemma 6 (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. I use $U^n(p)(t)$, and $U(p)(t)$ to refer to two processes defined as functions 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 $p(\tau)$. 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

I let $\omega \in \Omega$ index the event space. 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 $R_j = \inf\left(t : \|p(t)\| + \|\sigma(t)\| \geq p\right)$ and the stopping times $Q_j = \inf\left(t : \|p(t)\| + \|\gamma(t)\| \geq p\right)$. These increase to $+\infty$ as well. Therefore, we can set $S_j = \tau_j \wedge R_j \wedge Q_j$.

Then we can define the following processes:

$$\mu^{(j)}(t) = \mu(t \wedge S_j), \ \sigma^{(j)}(t) = \sigma(t \wedge S_j), \ \gamma^{(j)}(t) = \gamma(t \wedge S_j) \tag{87}$$

and

$$p^{(j)}(t) = \begin{cases} 0 & \text{if } S_j = 0 \\ p(0) + \int_0^t \mu^{(j)}(s) \, ds + \int_0^t \sigma^{(j)}(s) \, dW(s) + \int_0^t \gamma^{(j)} \, d\mathcal{L}(s) & \text{if } S_j > 0. \end{cases} \tag{88}$$

Now, local characteristics of $p^{(j)}$ agree when $t < S_j$ as 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 $\|p^{(j)}(t)\| \leq 3p$.

Step 2

By the proof of Jacod and Protter 2012, Lemma 4.4.9, 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 clearly holds for the stopped versions above. In addition, 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 original result was for convergence in probability, the new one is 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. 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.

$\square$

Theorem 8 (Estimating the Instantaneous Absolute Volatility). *Let $p(t)$ be an Itô semimartingale 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 $i_n = i - k_n - 1$.*

Then the following holds, where $\mathrm{erfcx} := \frac{2 \exp(x^2)}{\sqrt{\pi}} \int_x^\infty \exp(-s^2) \, ds$:[37]

$$\frac{1}{k_n\sqrt{\Delta^n}} \sum_{m=0}^{k_n-1} |\Delta_{i_n+m}^n p| \xrightarrow{\mathbb{P}} \mathbb{E}|N(0,1)|\sigma(\tau-) + \frac{\gamma(\tau-)}{\sqrt{2}} \mathrm{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right). \tag{30}$$

Proof. This proof is divided into a number of steps. I start 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.

---

37. 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.

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.

$$f'_- = \text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x \leq 0 \end{cases} \tag{89}$$

Then using Medvegyev (2007, Theorem 6.65), where $A(t)$ is a finite-valued increasing process, we can rewrite $||p(t)$ as

$$|p(t)| = \int_0^t \text{sign}(p(s-))\, dp(s) + A(t) = \int_0^t \text{sign}(p(s-))\, dW(s) + \int_0^t \text{sign}(p(s-))\, d\mathcal{L}(s) + A(t). \tag{90}$$

$A(t)$ is a finite-valued increasing process and so it 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 difficult. In that case, $\text{sign}(p(s-))$ is the same as the sign of the larger, in magnitude, of the two processes. Since the two processes have different signs, the smaller process has the opposite sign. Consequently, the part of $|p(t)| - A(t)$ where the two process has different signs can be rewritten as follows. Let $\Omega^{\mathcal{L}}$ be the set where the Laplace part in magnitude is larger and Ω^W the part where the diffusion part is.

$$|p(t)| - A(t) = \int_0^t \text{sign}(W(s-))1_{\Omega^W}(s-)\sigma(s)\, dW(s) - \int_0^t \text{sign}(\mathcal{L}(s-))1_{\Omega^W}(s-)\gamma(s)\, d\mathcal{L}(s)$$
$$+ \int_0^t \text{sign}(\mathcal{L}(s-))1_{\Omega^{\mathcal{L}}}(s-)\gamma(s)\, d\mathcal{L}(s) - \int_0^t \text{sign}(W(s-))1_{\Omega^{\mathcal{L}}}(s-)\sigma(s)\, dW(s) \tag{91}$$

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 (91) as follows condition on the signs differing as follows.[38]

$$\int_0^\infty \int_x^\infty (y-x)\psi_{\gamma,\Delta}(x)|\phi|_{\sigma,\Delta}(y)\, dx\, dy + \int_0^\infty \int_y^\infty (x-y)\psi_{\gamma,\Delta}(x)|\phi|_{\sigma,\Delta}\, dy\, dx \tag{92}$$

$$= \frac{\sqrt{\Delta}}{\sqrt{2}}\left(-\gamma + \frac{2}{\sqrt{\pi}}\sigma + \gamma\, \text{erfcx}\left(\frac{\sigma}{\gamma}\right)\right) + \frac{\gamma\sqrt{\Delta}}{\sqrt{2}}\, \text{erfcx}\left(\frac{\sigma}{\gamma}\right) \tag{93}$$

$$= \sqrt{\Delta}\left(m_1\sigma + \frac{\gamma}{\sqrt{2}}\left(2\, \text{erfcx}\left(\frac{\sigma}{\gamma}\right) - 1\right)\right) \tag{94}$$

---

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

In the part where they both have the same sign, the absolute value is just the sum of the absolute values and so we can rewrite (91) given that the signs are the same as follows.

$$m_1 \sigma \sqrt{\Delta} + \frac{\gamma}{\sqrt{2}} \sqrt{\Delta} \tag{95}$$

Then by taking the average of the (94) and (95), we can solve for (91).

$$\mathbb{E}|p(t)| - A(t) = m_1 \sigma \sqrt{\Delta} + \frac{\gamma \sqrt{\Delta}}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right) \tag{96}$$

The first part of this equation is the expectation of the absolute value of the diffusion part. If $\operatorname{erfcx}(\sigma/\gamma)$ were replaced with 1, the second part would be the absolute value of the jump part. Consequently, $\operatorname{erfcx}(\sigma/\gamma)$ reweights the jumps appropriately. It is also worth noting that $\lim_{x \to 0} \operatorname{erfcx}(x) = 1$, and $\lim_{x \to \infty} \operatorname{erfcx}(x) = 0$. Consequently, as σ vanishes we recover the mean of absolute value of the jumps, while as γ vanishes we recover the mean of the absolute value of the diffusion part. This is precisely what we need.

Step 3

In this section, I consider the asymptotic behavior of the estimator. I prove 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.

Since $\sigma(t)$ and $\gamma(t)$ are step functions, there exists a sequence $\{\tau_j\}$ such that $\sigma(t)$ and $\gamma(t)$ are constant over the intervals between the various τ_j.

$$p(t) = \sum_j \int_{\tau_j}^{\tau_{j+1}} \sigma(\tau_j) \, dW(s) + \int_{\tau_j}^{\tau_{j+1}} \gamma(\tau_j) \, d\mathcal{L}(s) \tag{97}$$

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 reference 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 τ 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

$$\mathbb{E}\left[\left\| \frac{1}{k_n \Delta^n} \sum_{m=0}^{n} |\Delta_{i_n+m}^n p| - \left|m_1 \sigma(\tau_n) + \frac{\gamma(\tau_n)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau_n)}{\gamma(\tau_n)}\right)\right| \right\|^2 \right] \tag{98}$$

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

$$\left. \left. - k_n \sqrt{\Delta^n} \left|m_1 \sigma(\tau_n) + \frac{\gamma(\tau_n)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau_n)}{\gamma(\tau_n)}\right)\right| \right\|^2 \right]. \tag{99}$$

Then by applying (96), we have

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

$$\left. - k_n\sqrt{\Delta^n}\left|m_1\sigma(\tau_n) + \frac{\gamma(\tau_n)}{\sqrt{2}}\operatorname{erfcx}\left(\frac{\sigma(\tau_n)}{\gamma(\tau_n)}\right)\right|\right\|^2\right]. \tag{100}$$

Simplifying implies this is

$$= O_p\left(\frac{k_n^2(\Delta^n)^2}{k_n^2\Delta_n}\right), \tag{101}$$

since $A(t)$ is a finite-variation term.

Step 5

To finish deriving the theorem, we show that approximating the volatility functions by step functions 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 $\gamma_x^2(t) = \sup_{s_1,s_2<t\wedge\tau}|x(s_1) - \tilde{x}(s_2)|^2$ for x equal to σ and γ, while let $\gamma_b^2(t) = \sum_{s_1,s_2<t\wedge\tau}|b(s_1) - b(s_2)|$. 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 (97). Let $t_1, t_2 < \tau$.

$$\mathbb{E}\left[\left\|\int_{t_1}^{t_2}\mu(s)\,ds + \int_{t_1}^{t_2}\sigma(s)\,dW(s) + \frac{1}{2}\int_{t_1}^{t_2}\gamma(s)d\mathcal{L}(s)\right.\right. \tag{102}$$

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

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

$$\leq \mathbb{E}\left[\int_{t_1}^{\tau}\mu(s)^2(s) + \int_{t_1}^{\tau}|\tilde{\sigma}(s) - \sigma(s)|^2\,ds + \frac{1}{2}\int_{t_1}^{\tau}|\tilde{\gamma}(s) - \gamma(s)|^2\,ds\right] \tag{103}$$

Then we can bound each of the terms.

$$= (O(1)\gamma_b^2(\tau) + O(1)\gamma_\sigma^2(\tau) + O(1)\gamma_\gamma^2(\tau))(\tau - t_2) \tag{104}$$

$$= O(1)(\tau - t_2) \tag{105}$$

In other words, if we choose a sequence of meshes so that the supremum of their magnitudes $\Delta^n \to 0$ and the minimal value $\tau - k_n\Delta^n \to 0$, 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 preceding parts 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 (97). Consider the mean-square deviation of the estimator from its limiting value.

$$\frac{1}{k_n^2 \Delta_n} \mathbb{E}\left[\left\|\left\|\sum_{m=0}^{k_n-1}|\Delta_{i_n+m}^n p| - Y'(t) + Y'(t) - \left(m_1\sigma(\tau-)k_n\sqrt{\Delta^n} + \frac{\gamma(\tau-)}{\sqrt{2}}\operatorname{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right)k_n\sqrt{\Delta^n}\right)\right\|\right\|^2\right]$$

(106)

By splitting the term into two parts and using the bounds from (101) and (105).

$$= \frac{1}{k_n^2 \Delta_n}\left(O(\Delta k_n^2) + O(\Delta k_n)\right) \to 0$$

(107)

$\square$

Theorem 7 (Estimating the Instantaneous Diffusion Volatility). *Let $p(t)$ be an Itô semimartingale 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 $c_1(\Delta^n)^{1/4} \leq v_1^n \leq = c_2\sqrt{\Delta^n}$ for some constants c_1, c_2 and $v_2^n \to 1$. Then*

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

(29)

Proof. The intuition behind the proof is straightforward. We separate the large jumps from the continuous part by truncating, and then note that the small jumps do not matter asymptotically because by squaring the remainder they get pushed even closer to zero. Consequently, we only pick up the middle range of the distribution, which is dominated by the continuous variation. Effectively, we are considering $\lim_{s\to 0} I\hat{V}(\tau) - I\hat{V}(\tau - s)$, and since we are estimating the left-limit of its time-derivative, $\sigma^2(\tau-)$, this works.

By localization we can strengthen some assumptions. Specifically, we can replace Assumption HL with Assumption SHL. In addition, the jump martingale part of the process is a sum of an integral with respect to Laplace motion $\mathcal{L}(t)$ and the zero process $\delta_0(t)$ where the weights depend upon the intensity of the jumps by Corollary 4.2 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 that the jump part is an integral with respect to $\mathcal{L}(t)$. The part of the proof regarding the continuous part of the process will not change in either part.

Step 1

I proceed by showing convergence in mean square, which implies convergence in probability. Note, $\left|\Delta_{i_n+m}^n p\right|^2 = O_p(\Delta^n)$ 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 consistency of the original process, I must show that the jump part converges to zero.

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

$$\left|(x+y+z+w)1\{|x+y+z+w| < v\} - x^2\right| \leq K\frac{|x|^4}{v^2} + \epsilon x^2 + \frac{K}{\epsilon}((v^2 \wedge y^2) + z^2 + w^2). \quad (108)$$

Define the following four processes, where I split the process up. The continuous variation is split into two parts, one with locally constant volatility and the other being the additional deviation coming from the change in the volatility.

$$Y^n(t) = \sigma(\tau_n)(W_t - W_{\tau_n})1\{\tau_n \leq t\} \quad (109)$$

$$Y'^n(t) = \int_{\tau_n \wedge t}^t (\sigma(s) - \sigma(\tau_n))\, dW(s) \quad (110)$$

$$Z^n(t) = \int_{\tau_n \wedge t}^t \gamma(s)\, d\mathcal{L}(s) \quad (111)$$

$$B^n(t) = \int_{\tau_n \wedge t}^t \mu(s)ds \quad (112)$$

Note, $p(\tau_n \wedge t) = Y^n(t) + Y'^n(t) + Z^n(t) + B^n t$. Now, we can use (108), with $x = \frac{\Delta_{i_n+m}^n Y^n}{\sqrt{\Delta^n}}$, $y = \frac{\Delta_{i_n+m}^n Z_{i_n+m}^n}{\sqrt{\Delta^n}}$, and $w = \frac{\Delta_{i_n+m}^n B_{i_n+m}^n}{\sqrt{\Delta^n}}$. The main issue here is showing that all of the parts except for $Y^n(t)$ converge to zero because then we are essentially just taking the variance of that part. 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 we have the following inequality:

$$\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}\left|(Y_t^n)^2 - (p_t^n)^2\right| \leq \frac{1}{k_n}\sum_{m=0}^{k_n-1}\left(\frac{K}{\omega_n^2}\left|\frac{\Delta_{i_n+m}^n Y^n}{\sqrt{\Delta^n}}\right|^4 + \epsilon\left|\frac{\Delta_{i_n+m}^n Y^n}{\sqrt{\Delta^n}}\right|^2 \right. \quad (113)$$

$$\left. + \frac{K}{\omega_n^2\epsilon}\left|\frac{\Delta_{i_n+m}^n Z}{\sqrt{\Delta^n}\omega_n}\right|^2 + \frac{K}{\epsilon}\left|\frac{\Delta_{i_n+m}^n Y'^n}{\sqrt{\Delta^n}}\right|^2 + \frac{K}{\epsilon}\left|\frac{\Delta_{i_n+m}^n B}{\sqrt{\Delta^n}}\right|^2\right).$$

Set $\gamma_n = \sum_{s\in|\tau_n,\tau_n+(k_n+2)\Delta^n|}|\sigma(s) - \sigma(\tau_n)|^2$, 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 $\Delta_{i_n+m}^n Z$. Clearly, $E\left|\Delta_{i_n+m}^n\right| \leq \phi_n\sqrt{\Delta^n}$. 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.

$$\mathbb{E}|\mathcal{L}(0, \phi_n) \wedge 1| = \phi_n\sqrt{\Delta^n} - \exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right)(\phi_n\sqrt{\Delta^n} + 1) \leq O\left(\frac{1}{\sqrt{\Delta^n}}\right)\exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right). \quad (114)$$

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, $\frac{1}{\omega_n^2}\frac{\Delta_{i_n+m}^n Z}{\sqrt{\Delta^n}\omega_n} \overset{a.s.}{\in} O_p\left(\frac{1}{\Delta^n\omega_n^3}\right)\exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right) = o_p(1)$ as exponential functions decay faster than polynomials increase.

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

$$\mathbb{E}\left[\left|(Y_t^n)^2 - (p_t^n)^2\right|\right] \le K\epsilon + \frac{K}{\epsilon}\left(o_p(1) + o_p(1) + \mathbb{E}[\gamma_n]\right). \tag{115}$$

If we take $n \to \infty$, and then $\epsilon \to 0$, the left hand side of the above equation converges to zero.

Step 2

To complete the proof, we have to consider what $\lim_{n\to\infty}\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}|Y_t^n|^2$ is. If we recall its definition, we note that converges to the variance of the increment:

$$\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}|\sigma_{\tau_n}(W_t - W_{\tau_n})1\{\tau_n \le t\}|^2 = \sigma(\tau_n)^2\frac{1}{k_n}\sum_{m=0}^{k_n-1}\left|\frac{\Delta_{i_n+m}^n W}{\sqrt{\Delta^n}}\right|^2 \to \sigma(\tau_n)^2 \tag{116}$$

Since the square is a convex function, we can combine these two previous limits, and we get 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. Clearly, if we multiply the expression by a value that is almost surely converging to 1, none of the results change, and we are done.

$\square$

Properties of the Scaled Complementary Error Function

We first note that the scaled complementary error function is a reparameterization of the Mills' ratio $r(x)$, (Baricz 2008):

$$\text{erfcx}(x) := \mathbb{E}|N(0,1)|r\left(x\sqrt{2}\right). \tag{117}$$

As a result, we can easily adopt the known features of that function bounds for it. In particular, this implies that $\text{erfcx}(x)$ is a convex, strictly-decreasing function. It also implies the following bounds, (Theorem 2.3):

$$\frac{2}{\sqrt{\pi}(\sqrt{x^2+2}+x)} < \text{erfcx}(x) < \frac{4}{\sqrt{\pi}(\sqrt{x^2+4}+3x)}. \tag{118}$$

Theorem 9 (Estimating the Instantaneous Jump Volatility)**.** *Let $p(t)$ be an Itô semimartingale 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:*

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

$$(31)$$

Proof. In the following proof, I use 0 subscripts to denote population objects.

$$\widehat{Q}_n(\gamma) := g\left(\left|\frac{1}{k_n\sqrt{\Delta}}\sum_{m=0}^{k_n-1}\left|\Delta_{i_n+m}^n p\right| - m_1\hat{\sigma}(\tau-) - \gamma\operatorname{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right)\right|\right) \qquad (119)$$

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

$$Q_0(\gamma) := g\left(\left|\gamma_0\operatorname{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right) - \gamma\operatorname{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right)\right|\right) \qquad (120)$$

Clearly, this equals zero when $\gamma = \gamma_0$. Moving forward, we will show that both $\widehat{Q}_n(\gamma)$ and $Q_0(\gamma)$ are both strictly convex, which will imply the minimum is unique. Define $A(\sigma, \gamma) := \gamma\operatorname{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right)$. 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. This statement is likely to hold as erfcx is almost constant as a function of γ, but I must make rigorous what is meant by almost.

$$\frac{\partial}{\partial\gamma}\gamma\operatorname{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right) = \operatorname{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right) - \frac{\sigma}{\gamma^2\sqrt{2}}\frac{\partial}{\partial x}\operatorname{erfcx}(x)\bigg|_{x=\frac{\sigma}{\gamma\sqrt{2}}} \qquad (121)$$

Since erfcx is a decreasing function, the last term is negative, and so the entire equation is strictly positive. This 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 we assumed that $\gamma_0 > 0$, γ_0 is in the interior of a convex set. Consequently, by Newey and McFadden (1994, Theorem 2.7), $\hat{\gamma}_n$ is well-defined in the sense of being a unique minimizer, and $\hat{\gamma}_n \xrightarrow{\mathbb{P}} \gamma_0$.

$\square$

:w

Appendix C   News Premia Theorems

Theorem 1 (Jump Times are News Times)**.** *Consider a stopping time* τ. *Let* $P(t)$ *be a price process satisfying no-arbitrage. Then its natural filtration —* $\mathcal{F}_t^p$ *— contains all of the information*

in the representative investor's information set relevant for asset pricing, and $\mathcal{F}_\tau^p \neq \mathcal{F}_{\tau-}^p$ if and only if $P(t)$ jumps at τ, where $\mathcal{F}_{t-}^p$ is the associated predictable filtration.

Proof. Since, $P(t)$ satisfies no-arbitrage in the sense of no-free lunch with vanishing risk, by Delbaen and Schachermayer (1994), it is a semimartingale. First we prove that if $P(t)$ jumps at τ, then the two filtrations are not equal. Note, $\mathcal{F}_{t-}^p = \cup_{s<t}\mathcal{F}_s^p$. Clearly, $p(\tau) \notin \mathcal{F}_s^p$ for all $s < t$, and so it is not contained in their union, and so $\mathcal{F}_{\tau-} \neq \mathcal{F}_\tau$.

To prove the reverse direction, let $^p_P(t)$ by the predictable projection of $P(t)$, but then since $^pP(t)$ is pre-visible, $^pP(\tau)$ is measurable with respect to $\mathcal{F}_{\tau-}^p$, but $p(\tau)$ is not by assumption. Hence, it cannot equal its predictable projection with probability 1. However, this implies that τ is a jump time of $P(t)$.

The only other thing that we need to prove is that $\mathcal{F}_t^p$ contains all of the information that the representative investor knows that is relevant for asset pricing. Assume not. Then there exists an event $\mathcal{E}$ contained in the representative investor's information set $\mathcal{F}_t^r$ that is relevant for asset pricing, but is not measurable with respect to $\mathcal{E}$. Let $P(t)$ be the price, and $\mathcal{M}(t)$ be the representative investor's pricing kernel.

Then we know the following, where $P(t)$ is the cum-dividend price.

$$P(t) = \mathbb{E}\left[\mathcal{M}(\tau)P(\tau) \,|\, \mathcal{F}_t^r\right] \forall \tau \geq t \tag{122}$$

In addition, $\mathcal{E}$ being relevant for asset pricing implies that there exists a stopping time τ such that the following inequality holds.

$$\mathbb{E}\left[\mathcal{M}(\tau)P(\tau) \,|\, \mathcal{F}_t^r\right] \neq \mathbb{E}\left[\mathcal{M}(\tau)P(\tau) \,|\, \mathcal{F}_t^p\right] \tag{123}$$

However, $P(t)$ is measurable with respect to $\mathcal{F}_t^p$ by definition, and it equals the value on the left. Hence, we have a contradiction.

$\square$

Lemma 12 (An Itô's Formula for the Expectation of a Square Integrable Semimartingale)**.** *Let f be a twice-differentiable function and $\widetilde{Z}$ be a vector-valued semimartingale with locally bounded predictable $\langle Z \rangle(t)$. Then the differential of f satisfies*

$$d\mathbb{E}\left[f(\widetilde{Z}) \,\Big|\, \mathcal{F}_{t-}\right] = \mathbb{E}\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t) \,\Big|\, \mathcal{F}_{t-}\right] + \frac{1}{2}f''(\tilde{Z}(t-))\,d\langle\tilde{Z}\rangle(t). \tag{63}$$

Proof. The argument below is a standard application of Itô's formula for non-continuous processes applied to processes of bounded variation. In addition, the notation below should be interpreted in vector form. For example, $d\tilde{Z}(t)$ is the vector of dZ_i for all i, and $\langle \tilde{Z}^D \rangle$ is a matrix. We start by writing expanding the differential inside the expectation using Itô's formula for non-continuous

semimartingales, (Medvegyev [2007], Theorem 6.46):

$$dE\left[f(\tilde{Z}(t))\,\middle|\,\mathcal{F}_{t-}\right] = dE\left[\sum_{i=1}^{d}\frac{\partial f}{\partial z_i}\tilde{Z}(t-))\,d\tilde{Z}_i(t) + \frac{1}{2}\sum_{i,j}\frac{\partial f}{\partial z_i\partial z_j}f(\tilde{z}(t-))\langle\tilde{z}_i^D,\tilde{z}_j^D\rangle(t)\right. \tag{124}$$
$$\left. + \left(\Delta f(\tilde{Z}(t)) - \sum_{i=1}^{d}\frac{\partial f}{\partial z_i}f(\tilde{Z}(t-))\Delta\tilde{Z}_i(t)\right)\,\middle|\,\mathcal{F}_{t-}\right].$$

Rearranging and combining terms, we have.

$$= dE\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t) + \left(\Delta f(\tilde{Z}(t)) + f'(\tilde{Z}(t-))\Delta\tilde{Z}(s)\right)\right. \tag{125}$$
$$\left. - \frac{1}{2}f''(\tilde{Z}(t-))\langle\tilde{Z}^D\rangle(t)\,\middle|\,\mathcal{F}_{t-}\right].$$

Then by Taylor's theorem, canceling terms and noting that continuity implies bounded for the derivatives of f as long as $\tilde{Z}$ is bounded:

$$= dE\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t) + \frac{1}{2}f''(\tilde{Z}(t-))\,d\langle\tilde{Z}^D\rangle(t)\,\middle|\,\right] \tag{126}$$
$$+ \frac{1}{2}\,dE\left[f''(\tilde{Z}(t-))\Delta\tilde{Z}(t)^2 + O((\Delta\tilde{Z}(t)^3)\,\middle|\,\mathcal{F}_{t-}\right]. \tag{127}$$

Since the quadratic variation and the predictable quadratic variation coincide for continuous processes.

$$= dE\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t)\,\middle|\,\mathcal{F}_{t-1}\right] + \frac{1}{2}f''(\tilde{Z}(t-))\,d[\tilde{Z}](t) + E\left[O((\Delta\tilde{Z}(t)^3))\,\middle|\,\mathcal{F}_{t-}\right]. \tag{128}$$

By the Davis-Burkholder-Gundy inequality, for some constant c:

$$= dE\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t)\,\middle|\,\mathcal{F}_{t-1}\right] + \frac{1}{2}E\left[f''(\tilde{Z}(t-))\,d[\tilde{Z}](t)\,\middle|\,\mathcal{F}_{t-}\right] \tag{129}$$
$$+ E\left[c_1 O([\tilde{Z}]^{3/2})\,\middle|\,\mathcal{F}_{t-}\right].$$

Since we are considering local changes in time,

$$= dE\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t)\,\middle|\,\mathcal{F}_{t-}\right] + \frac{1}{2}f''(\tilde{Z}(t-))\,d\langle\tilde{Z}\rangle(t). \tag{130}$$

$\square$

Theorem 11 (Asset-Pricing Equation). *Let Assumption 6 hold, prices be Itô semimartingales, and the representative consumer face Problem 1 as $\Delta \to 0$. Assume preferences are such that optimal*

consumption is strictly positive. Define

$$M^{UP}(t) := \frac{\phi'(V(W(t))}{\phi'(V(W(t-))} \ and \ M(t) := \frac{\phi'(V(W(t-))}{\phi'(\phi^{-1}(\mathbb{E}[\phi(V(W(t)))|\mathcal{F}_{t-}]))} \frac{V'(W(t))}{u'(c(t-))}. \tag{60}$$

Then $M^{UP}(t)$ is a purely discontinuous martingale, and for all stopping times $\tau > t$,[39]

$$\widetilde{P}(t) = \mathbb{E}\left[M(\tau)M^{UP}(\tau)\widetilde{P}(\tau)\,\middle|\,\mathcal{F}_t\right] \tag{61}$$

Proof. Define the discounted price: $\widetilde{P}(t) := \exp(-\kappa t)P(t)$. This is a concave maximization problem and so first-order conditions characterize the optimum. Assume, for now, that the investor can only adjust his portfolio at a discrete grid of points $t, t + \Delta, t + 2\Delta, \ldots$ Then consumption and prices are effectively constant within each period, and the investor is faced with the following problem:

$$V(W(t)) = \max_{\Xi(t),C(t)} u(C(t)) + \exp(-\kappa\Delta)\phi^{-1}([\phi(V(W(t+\Delta)))|\mathcal{F}_t]) \tag{131}$$

$$C(t) + \sum_i P_i(t)\xi_i(t) = W(t) \tag{132}$$

$$W(t+\Delta) = \sum_i P_i(t+\Delta)\xi_i(t) \tag{133}$$

Submitting in the constraints gives

$$V(W(t)) = \max_{\Xi(t)} u(W(t) - \sum_i P_i(t+\Delta) + \exp(-\kappa\Delta)\phi^{-1}\left(\left[\phi\left(V(\sum_i P_i(t+\Delta)\xi_i(t))\right)\,\middle|\,\mathcal{F}_t\right]\right). \tag{134}$$

The discounted and original prices coincide at t, and we can equate $\exp(-\kappa\Delta)P_i(t+\Delta)$ with $\widetilde{P}_i(t+\Delta 0$. Hence, by using chain rule, and the formula for the derivative of an inverse, the first-order condition for (134) is

$$u'(c(t))\widetilde{P}_i(t) = \mathbb{E}\left[\frac{\phi'(V(W(t+\Delta))}{\phi'(\phi^{-1}(\mathbb{E}[\phi(V(W(t+\Delta)))|\mathcal{F}_t]))}V'(W(t+\Delta))\widetilde{P}_i(t+\Delta)\,\middle|\,\mathcal{F}_t\right], \tag{135}$$

at the optimal level of consumption and optimal asset shares. We can rearrange (135) as:

$$\widetilde{P}_i(t) = \mathbb{E}\left[\frac{\phi'(V(W(t+\Delta))}{\phi'(\phi^{-1}(\mathbb{E}[\phi(V(W(t+\Delta)))|\mathcal{F}_t]))}\frac{V'(W(t+\Delta))}{u'(c(t))}\widetilde{P}_i(t+\Delta)\,\middle|\,\mathcal{F}_t\right]. \tag{136}$$

If we plug in the risk-free rate $\widetilde{P}_f(t)$ in to (136), the prices on each side of the equal side are the same, and we can divide through by them. This gives

$$1 = \mathbb{E}\left[\frac{\phi'(V(W(t+\Delta))}{\phi'(\phi^{-1}(\mathbb{E}[\phi(V(W(t+\Delta)))|\mathcal{F}_t]))}\frac{V'(W(t+\Delta))}{u'(c(t))}\,\middle|\,\mathcal{F}_t\right]. \tag{137}$$

---

39. I use the UP superscript because M^{UP} is an unpredictable process.

In other words, the two terms in the inside the expectation are a martingale. Consequently, prices are a martingale with respect to the change of measure they induce. We can take limits as $\Delta \to 0$ in (136), which gives

$$\widetilde{P}_i(t) = \mathbb{E}\left[\frac{\phi'(V(W(t))}{\phi'\left(\phi^{-1}\left(\mathbb{E}\left[\phi(V(W(t)))\,|\,\mathcal{F}_{t-}\right]\right)\right)}\frac{V'(W(t))}{u'(c(t-))}\widetilde{P}_i(t)\,\bigg|\,\mathcal{F}_{t-}\right] \tag{138}$$

Now, we want to separate these two terms into a pure jump component and the remainder.

To do this, multiply and divide the first expression by $\phi'(V(W(t-))$:

$$\widetilde{P}_i(t) = \mathbb{E}\left[\frac{\phi'(V(W(t))}{\phi'(V(W(t-))}\frac{\phi'(V(W(t-))}{\phi'\left(\phi^{-1}\left(\mathbb{E}\left[\phi(V(W(t)))\,|\,\mathcal{F}_{t-}\right]\right)\right)}\frac{V'(W(t))}{u'(c(t-))}\widetilde{P}_i(t)\,\bigg|\,\mathcal{F}_{t-}\right] \tag{139}$$

Note, the first term here is simply $M^{UP}(t)$. Claim: $M^{UP}(t)$ is purely discontinuous. By Ai and Bansal (2018, Theorem 1), we know that the value function is a differentiable, and hence continuous, function of wealth. In addition, I am taking limits locally in time, and ϕ' is strictly positive. Consider $\lim_{\Delta \to 0} M^{UP}(t-\Delta)$.

$$\lim_{\Delta \to 0}\frac{\phi'(V(W(t-\Delta)))}{\phi'(V(W((t-\Delta)-)))} = \frac{\lim_{\Delta \to 0}\phi'(V(W(t-\Delta)))}{\lim_{\Delta \to 0}\phi'(V(W((t-\Delta)-)))} = \frac{\phi'(V(\lim_{\Delta \to 0}W(t-\Delta)))}{\phi'(\lim_{\Delta \to 0}V(W((t-\Delta)-)))}$$

$$\tag{140}$$

$$= \frac{\phi'(V(W(t-))}{\phi'(W(t-))} = 1$$

This implies that $M^{UP}(t)$ is a pure-jump process because its continuous part is identically one. In addition, I assumed there were no-predictable jumps, hence any drift (finite-variation, predictable) terms in the environment must be continuous. Consequently, $M^{UP}(t)$ is a pure-jump martingale.

$\square$

Theorem 13 (Asset-Pricing Equation)**.** *Let the assumptions in* Assumption 6 *hold,* $P_i(t)$ *be an Itô semimartingales, and the representative consumer face* Problem 1 *as* $\Delta \to 0$*. Assume that preferences are such that optimal consumption is strictly positive. Then risk-premia for some asset* i *is*

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)}\,\bigg|\,\mathcal{F}_{t-}\right] = -d\langle m, p^D + p^J\rangle(t) - d\langle m^{UP}, p^J\rangle(t). \tag{64}$$

Proof. The goal here is to replace the asset pricing equation in Theorem 11 with a stochastic logarithm of $P_i(t)$. Let $\widetilde{M}(\tau) = \exp(-\kappa(\tau))M(\tau)$ be the discounted stochastic discount factor. In this derivation, it is more useful to place the deterministic discounting into the discount factor than into the prices.

Then the asset-pricing equation is:

$$\widetilde{P}_i(t) = \mathbb{E}\left[\widetilde{M}(\tau)M^{UP}(\tau)\widetilde{P}(\tau)\,\Big|\,\mathcal{F}_t\right]. \tag{141}$$

Since $M(t)M^{UP}(t)$ given $\mathcal{F}_t$ equal 1, we can pre-multiply by it,

$$\widetilde{M}(t)M^{UP}(t)\widetilde{P}(t) = \mathbb{E}\left[\widetilde{M}(\tau)M^{UP}(\tau)\widetilde{P}(\tau)\,\Big|\,\mathcal{F}_t\right]. \tag{142}$$

In other words, $\widetilde{M}(t)M^{UP}P(t)$ is a martingale. This is the standard SDF type result. Discounted prices are martingales. I now take the stochastic logarithm of both sides. Taking the stochastic logarithm (as opposed to the regular logarithm) is useful because it preserves the martingale property. (The stochastic logarithm — $\mathcal{L}og(X)$ — is the inverse of the Doléans-Dade exponential.)

Before, I do this, it is useful to consider a few of the stochastic logarithms' properties. First, the following holds: $\mathcal{L}og(X \cdot Y) = \mathcal{L}og(X) + \mathcal{L}og(Y) + [\mathcal{L}og(X), \mathcal{L}og(Y)]$. We can also handle triple-products. You just need to apply the expression twice, and note that finite-variation terms do not affect the quadratic variation.

$$\mathcal{L}og(X \cdot Y \cdot Z) = \mathcal{L}og(X) + \mathcal{L}og(Y) + \mathcal{L}og(Z) + [\mathcal{L}og(X), \mathcal{L}og(Z)] + [\mathcal{L}og(X), \mathcal{L}og(Z)] \tag{143}$$
$$+ [\mathcal{L}og(Y), \mathcal{L}og(Z)]$$

As noted above, since (142) is a martingale its stochastic logarithm is as well.

$$0 = \mathbb{E}\left[\int_t^\tau d\mathcal{L}og\left(MM^{UP}P\right)(s)\,\Big|\,\mathcal{F}_t\right]. \tag{144}$$

We can expand this equation using (143). We can also replace the integrals with differentials without loss of generality because τ is arbitrary:

$$\implies 0 = \mathbb{E}\Big[d\mathcal{L}og(\widetilde{M})(t) + d\mathcal{L}og(M^{UP})(t) + d\mathcal{L}og(P)(t) \tag{145}$$
$$+ d[\mathcal{L}og(\widetilde{M}), \mathcal{L}og(P)](t) + d[\mathcal{L}og(M^{UP}), \mathcal{L}og(P)](t) + d[\mathcal{L}og(\widetilde{M}), \mathcal{L}og(M^{UP})](t)\,\Big|\,\mathcal{F}_{t-}\Big].$$

The stochastic logarithm equals the regular logarithm up to finite-variation terms:

$$= \mathbb{E}\Big[d\mathcal{L}og(\widetilde{M})(t) + d\mathcal{L}og(M^{UP})(t) + d\mathcal{L}og(P)(t) \tag{146}$$
$$+ d[\log(\widetilde{M}), \log(P)](t) + d[\log(M^{UP}), \log(P)](t) + d[\mathcal{L}og(\widetilde{M}), \mathcal{L}og(M^{UP})](t)\,\Big|\,\mathcal{F}_{t-}\Big].$$

We can combine M and M^{UP} together:

$$= \mathbb{E}\Big[d\mathcal{L}og\left(M \cdot M^{UP}\right)(t) + d\mathcal{L}og(P)(t) + d[\log(\widetilde{M}), \log(P)](t) + d[\log(M^{UP}), \log(P)](t)\,\Big|\,\mathcal{F}_{t-}\Big]. \tag{147}$$

The stochastic logarithm satisfies the following stochastic differential equation:

$$\mathcal{L}og(X)(t) = \int_0^t \frac{1}{X(s-)} \, dX(s). \tag{148}$$

Consequently, we can rewrite (147) as follows, where I replace the quadratic variation terms with predictable quadratic variation terms,

$$0 = \mathbb{E}\left[\frac{d\left(M \cdot M^{UP}\right)(t)}{\widetilde{M}(t-)M^{UP}(t-)} + \frac{dP(t)}{P(t-)} \,\middle|\, \mathcal{F}_{t-}\right] + d\langle \log(\widetilde{M}), \log(P)\rangle(t) + d\langle \log(M^{UP}), \log(P)\rangle(t). \tag{149}$$

If $M^{UP}(t)$ is identically 1, the all of the terms containing it disappear, which gives the standard asset pricing equation:

$$\mathbb{E}\left[\frac{dP(t)}{P(t-)} + \frac{d\widetilde{M}(t)}{\widetilde{M}(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = -d\langle m, p\rangle(t), \tag{150}$$

where $m = \log(M)$. We can ignore the discounting because it only cases a mean shift, and so will not affect quadratic covariation terms.

In the recursive case with jumps through, it is more complicated. An announcement SDF term is a pure-jump process so it only have non-zero covariation with the jump part of the prices:

$$\mathbb{E}\left[\frac{dP(t)}{P(t-)} + \frac{d\left(\widetilde{M} \cdot M^{UP}\right)(t)}{\widetilde{M}(t-)M^{UP}(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = -d\langle m, p\rangle(t) - d\langle m^{UP}, p\rangle(t), \tag{151}$$

where $m^{UP}(t) = \log(M^{UP}(t))$. Since (151) prices all assets, if we consider a risk-neutral asset, we have all of the of the quadratic variation terms equaling zero:

$$\frac{dP_f(t)}{P_f(t-)} = -\mathbb{E}\left[\frac{d\left(\widetilde{M} \cdot M^{UP}\right)(t)}{\widetilde{M}(t-)M^{UP}(t-)} \,\middle|\, \mathcal{F}_{t-}\right]. \tag{152}$$

Consequently, the risk premium on a asset i with discounted price P_i is

$$\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)} = -d\langle m, p\rangle(t) - d\langle m^{UP}, p\rangle(t) \tag{153}$$

Since $M^{UP}(t)$ and hence $m^{UP}(t)$ are purely discontinuous processes, the second quadratic variation does not depend upon $p^D(t)$. That is

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_f(t)}{P_f(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = -d\langle m, p^D + p^J\rangle(t) - d\langle m^{UP}, p^J\rangle(t). \tag{154}$$

$\square$

Appendix D News Premia: Empirical Results

Table 13: $\mathbb{E}\left[rx_t \,\middle|\, \sigma_t^2+\gamma_t^2, \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}, \mathbf{1}\{\text{FOMC}\}_t\right]$ (OLS)

Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\log\left(\sigma_t^2+\gamma_t^2\right)\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\bar{\mathbb{R}}^2$
0.01	0.88				0.26 %
[0.23]	[2.94]				
−4.55		−0.46			2.67 %
[−5.45]		[−5.58]			
1.02			1.65		1.61 %
[5.81]			[6.48]		
−3.17		−0.39	1.13		3.35 %
[−3.94]		[−4.12]	[4.06]		
−1.27		−0.19	3.91	0.29	3.42 %
[−0.54]		[−0.85]	[1.07]	[0.80]	
−4.73	1.09	−0.47			3.09 %
[−5.88]	[3.55]	[−6.11]			
−3.40	1.00	−0.40	1.07		3.70 %
[−3.96]	[3.38]	[−5.08]	[3.94]		
	0.98	−0.23	3.55	0.26	3.74 %
[−0.68]	[3.30]	[−0.95]	[0.88]	[0.65]	

Appendix E Simulation Results

Figure 12: Continuous-Time Simulation Results without Microstructure

(Average every 5 minutes)

(a) $\mathbb{E}[|p(t)|]$ (b) $\sigma(t)$ (c) $\gamma(t)$
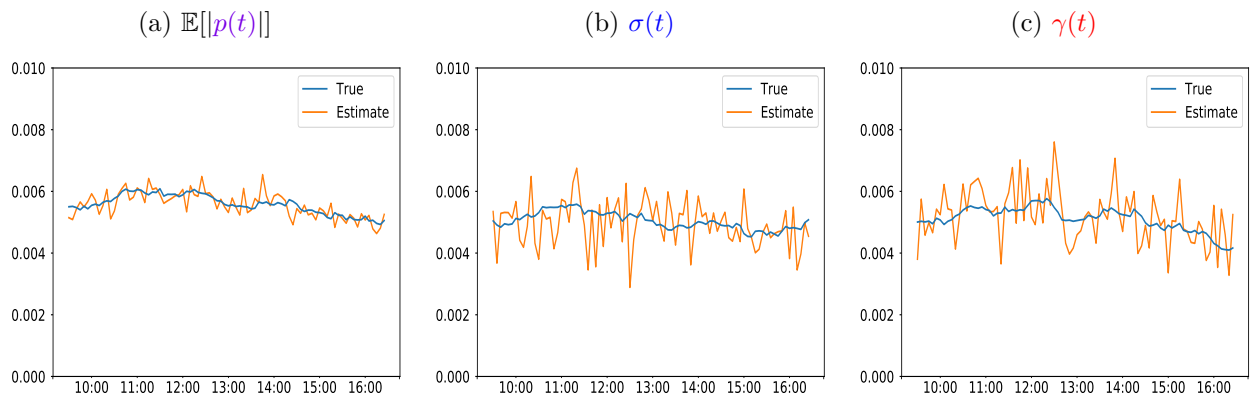
Figure 13: Continuous-Time Simulation Results with Microstructure

(Average every 5 minutes)

(a) $\mathbb{E}[|p(t)|]$ (b) $\sigma(t)$ (c) $\gamma(t)$
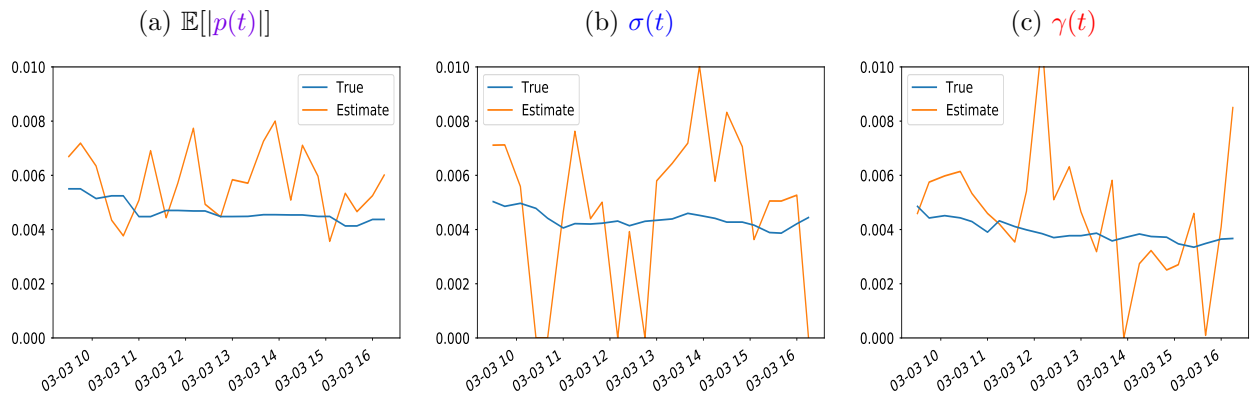
APPENDIX F   VOLATILITY: EMPIRICAL RESULTS

Table 14: Vector Autoregression Models

	$\log(\sigma_t^2)$		$\log(\gamma_t^2)$	
	VAR(1)			
Intercept	-0.84	$(-1.04, -0.64)$	-1.80	$(-1.98, -1.62)$
$\log(\sigma_{t-1}^2)$	0.56	$(0.52,\ 0.59)$	0.34	$(0.31,\ 0.38)$
$\log(\gamma_{t-1}^2)$	0.38	$(0.33,\ 0.42)$	0.48	$(0.44,\ 0.52)$
$\mathbb{R}^2$	24%		72%	
Innovation Covariance		$\begin{pmatrix} 0.33 & 0.19 \\ 0.19 & 0.27 \end{pmatrix}$		
	VAR(6) — Chosen by SIC			
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 15: $\mathbb{E}\left[rx_t \,\middle|\, \mathbf{1}\{\text{FOMC}\}_t, \sigma_t^2+\gamma_t^2, \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right]$ (WLS)

Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\log\left(\sigma_t^2+\gamma_t^2\right)\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\bar{\mathbb{R}}^2$
0.34 [14.43]	0.26 [1.44]				0.05 %
−2.68 [−6.93]		−0.28 [−7.77]			1.83 %
0.68 [8.49]			0.59 [4.14]		0.53 %
−2.29 [−5.46]		−0.27 [−7.19]	0.51 [4.59]		2.20 %
2.62 [2.48]		0.18 [1.89]	8.62 [4.92]	0.75 [4.63]	2.85 %
−2.76 [−6.76]	0.41 [2.13]	−6.76 [−7.47]			1.93 %
−2.37 [−5.51]	0.39 [2.06]	−0.28 [−7.17]	0.51 [3.60]		2.28 %
2.38 [1.66]	0.36 [2.55]	0.17 [1.21]	8.50 [3.65]	0.74 [3.39]	2.91 %

Table 16: News Premia Estimates

Regressors			
Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$
0.34 [14.43]	0.26 [1.44]		
2.95 [6.61]		0.24 [5.88]	
−2.45 [−5.12]			−5.01 [−5.86]
−5.04 [−0.58]		0.14 [2.68]	−4.15 [−4.93]
2.95 [6.62]	0.15 [0.85]	0.24 [5.90]	
−5.10 [−5.10]	0.33 [1.54]		−5.02 [−5.81]
0.14 [0.20]	0.25 [1.25]	0.16 [3.66]	−3.52 [−5.35]

Table 17: Instrument Variables: First Stage Regression

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

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		$79.27\,\%$	
	$[-4.99]$	$[-3.47]$	$[0.11]$	$[1.79]$	$[1.65]$	$[27.95]$	$[8.80]$	$[8.92]$	$[4.40]$			
	-1.31	-1.23	0.02	0.07	0.07	0.53	0.16	0.13	0.05	-0.11	$79.43\,\%$	$20\,140$
	$[-5.84]$	$[-5.05]$	$[0.63]$	$[1.91]$	$[1.92]$	$[18.91]$	$[8.70]$	$[8.92]$	$[4.70]$	$[-4.43]$		

Table 18: News Premia Estimates: Other Instruments

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

Table 19: News Premia Estimates in Levels

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

Regressors						Instruments			
Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	σ^2_t	γ^2_t	$(\sigma^2_t)(\gamma^2_t)$	$\mathbf{1}\{\text{FOMC}\}_t$	$\sigma^2_{t-l}\cdots$		$\gamma^2_{t-l}\cdots$	$(\sigma^2_{t-l})(\gamma^2_{t-l})$
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]					
0.26	0.59	−0.34	−0.01		✓	✓		✓	✓
[5.69]	[2.78]	[−2.00]	[−3.18]						
0.27	0.34	0.71	−0.45	−0.01	✓	✓		✓	✓
[5.97]	[1.72]	[3.07]	[−2.41]	[−3.38]					

Table 20: News Premia Estimates: Robustness

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