Jumps, Realized Densities, and News Premia *

Paul Sangrey †

University of Pennsylvania

Job Market Paper

Current Version

This Version: August 21, 2018

Abstract

Announcements and other news continuously barrage financial markets causing asset prices to jump hundreds of times per day. Recursive utility implies that this jump-driven uncertainty will be priced differently than equivalent diffusive-driven uncertainty. I derive a tractable non-parametric continuous-time representation for the prices' jumps and derive the implied sufficient statistic for the jump dynamics. This statistic — *jump volatility* — is the instantaneous variance of the jump part and measures news risk. I define the realized density as the daily return density conditional on its diffusion and jump volatilities. This solves the time-aggregation problem and reduces tracking the daily return density to forecasting its volatilities. I develop estimators for the volatilities and the realized density and estimate them using high-frequency data from SPY. This nonparametrically identifies the average curvature in investor's certainty equivalence functional. I then apply these methods to high-frequency data from the S&P 500 and show that total volatility commands a positive risk premium and the proportion of volatility driven by jumps commands a negative premium. This implies that investor's certainty equivalence function is quasiconvex. I further show that volatility premia are capable of explaining the large ex-post return on FOMC announcement days.

Keywords: Asset Pricing, Jumps, News Risk, Stochastic Volatility, FOMC, High-Frequency Econometrics, Recursive Utility, Realized Volatility, Nonparametric Modeling, Semimartingales, Time Aggregation, Risk Premia

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

---

*I am indebted to my advisors, Francis X. Diebold and Frank Schorfheide as well as the other members of my committee: Amir Yaron and Xu Cheng. I have also benefited greatly from comments from Jaroslav Borovičková, Ben Connault, Frank DiTraglia, Winston Wei Dou, Jesús Fernández-Villaverde, Ron Gallant, Bryan Kelly, Asaf Manela, Yuan Liao, George Tauchen, Jessica Wachter, Dacheng Xiu, Ross Askanazi, Minsu Chang, Laura Liu and seminar participants at the University of Pennsylvania, Wharton (UPenn), the SoFiE Summer School in Chicago, the Young Economics Symposium at Yale University, and the George Washington Student Research Conference. All remaining errors are my own.

†Paul Sangrey: PhD Candidate, Department of Economics, University of Pennsylvania, The Ronald O. Perelman Center for Political Science and Economics, 133 South 36th Street, Philadelphia, PA 19104. Email: paul@sangrey.io. Web: sangrey.io

The study of individual's reactions to their time-varying risk lies at the heart of modern finance and macroeconomics. A major strand of asset pricing literature prices assets under various forms of recursive utility, (Bansal and Yaron 2004; Hansen 2014). Recently, Ai and Bansal (2018) showed announcement uncertainty is priced differently if and only if agent have recursive utility, while Tsai and Wachter (2018) showed recursive utility implies prices are not normal-times covariances when future returns jump. Meanwhile, a major strand of the financial econometrics literature measures the tails of return distributions, (Engle and Gonzalez-Rivera 1991; Patton, Ziegel, and Chen 2018). Working in a jump-free world, Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) show integrated volatility is a sufficient statistic for the return distribution's dynamics. In all of these environments, jumps in prices or information break standard covariance-based theories. This is problematic. We live in a world where prices jump all the time (Aït-Sahalia and Jacod 2012; Gallant and Tauchen 2018), and announcements frequently and dramatically affect asset prices (Lucca and Moench 2015; Law, Song, and Yaron 2018).

So the question facing us as researchers is can we generalize standard covariance-based explanations to a world with jumps? This is tricky because jumps are inherently a continuous-time phenomenon. All functions are continuous in a purely discrete world. Yet, all of the papers above study discrete-time risk, and so it is clearly possible. We must solve the time-aggregation problem and relate continuous- and discrete-time phenomena. Although many papers have solved some dimension of this problem, no one has done this in a way that nonparametrically identifies jump risk.

To put it another way, the question facing us is twofold. First, how are jumps in investors information and jumps in asset prices related? Second, is there some way of nonparametrically identifying this jump risk? Can we separate out time-variation in jump risk from other sources of risk? This paper addresses these two questions. In particular, it shows that no-arbitrage implies that asset prices jump at some time τ if and only if the representative investor's information set jumps at τ. In other words, the papers listed above all address the same issue — discontinuous information flows complicate covariance-based explanations of return dynamics. Second, it develops a new representation theory and derives a sufficient statistic for jump risk — *jump volatility* — thereby nonparametrically identifying jump risk. It further shows that this jump volatility nonparametrically identifies the curvature of investor's recursive preference aggregator. It then takes this theory to high-frequency data on the S&P 500, estimating the jump volatility, and providing a number of new stylized facts regarding jumps, jump volatility, and their relationship to expected returns.

1.1.2. Layout of the Paper

I start by showing that under some general conditions, the jump part of the prices can be represented as a integral with respect to a variance-gamma process. This implies that the instantaneous variance of the jump part of the process — the *jump volatility* — is a sufficient statistic for the jump dynamics. This is useful because it avoid the trade-off between assuming away the model-specification risk that investors face as parametric models implicitly do and tracking an infinite-dimensional unidentified object as the nonparametric literature does. In other words, I nonparametrically identify a scalar summary for the jump dynamics.

As a second contribution, I use the jump volatility and its daily analogue to solve the time-aggregation problem since volatilities aggregate over time. This enables me to convert these continuous-time measures into discrete-time ones. This is useful because on the one hand, most investors' decisions take place at some discrete-frequency and so most of the risk relevant for asset pricing occurs at that frequency, but, on the other hand, we have a great deal of data and analytical tractability in continuous-time. Being able to simultaneously handle both discrete- and continuous-time price dynamics gives us the best of both worlds, albeit at the cost of increased complexity. We can discuss the discrete-time problem facing investors while using high-frequency data to identify the processes of interest.

I define return's *realized density* — the return's daily conditional distribution given its volatilities. This reduces obtaining the return's daily densities to volatility forecasting. Since volatility dynamics are well-behaved, this works well in practice. Consequently, we can measure time-variation in tail risk in real-time because any tail-risk measure of interest, such as value-at-risk (VaR) and expected shortfall (ES), are just statistics of the forecasting densities. In the interest of space, I leave the analysis of these densities to a companion paper — Sangrey (2018).

Knowing the correct population measures is insufficient to analyze the data, and so I develop estimators for the jump and diffusion volatilities in both continuous- and discrete-time.[1] To do this I adapt a continuous-time diffusion volatility estimator from Jacod and Rosenbaum (2013).[2] Since I introduce jump volatility in this paper, the literature obviously cannot provide us with an estimator. In fact, although various authors, such as (Barndorff-Nielsen and Shephard 2004; Li 2013), have estimated daily jump variation measures, to the best of my knowledge no-one has developed a continuous-time estimator for any jump variation statistic. Furthermore, Jacod and Rosenbaum (2013) shows that we cannot use the natural estimator, the local-in-time squared variation, to estimate continuous-time jump variation. Consequently, I show how to use the absolute variation and my representation to estimate the jump volatility. I then derive estimators for the discrete (integrated) jump and diffusion volatilities by integrating the continuous-time (instantaneous) ones. As an aside, since we can consistently estimate the jump volatility at any point in time, it is nonparametrically identified.

In a influential recent paper, Lucca and Moench (2015) show that ex-post returns are high on average on the days when the Federal Open Market Committee (FOMC) makes its announcements. Ai and Bansal (2018) argue this is evidence for particular forms of recursive utility preferences. The curvature arising from the recursive time aggregation of preferences gives rise to a potential risk risk premium, similarly to how the utility function's curvature does. In an Epstein-Zin world, they

1. The diffusion volatility is volatility of the diffusion part of the process. It is often just referred to as the volatility in the literature.

2. What I mean by a continuous-time estimator is that I can estimate the value of the latent volatility process at any time τ.

show this positive premium is equivalent to a preference for early resolution of uncertainty. I show that this announcement premium is a special case of a general news premium. The key feature of announcements is that they cause investors' information set to jump.

I then use jump volatility to nonparametrically identify both the sign, as Ai and Bansal (2018) do, and the magnitude of the recursive preference aggregator's curvature. In general, risk premia has two components. The first is a covariance between marginal utility and future returns, and the second is a covariance between the derivative of the recursive aggregator and the jump part of future turns.

Now that we have estimators for the volatilities and theory telling us how they are related to risk premia, we can analyze how investors treat these risks in practice. Turning to the data, I show the *total volatility* — the sum of diffusion and jump volatilities — commands an economically and statically significant premium. This is what standard theory predicts, but the previous literature has found it hard to find empirically, (Lettau and Ludvigson 2010). I then show, using the same identification strategy, that the proportion of the volatility driven by jumps commands an economically and statistically significant negative premium. In other words, for a given level of total volatility risk, investors dislike the volatility driven by news less than they dislike other sources of volatility. This implies that the derivative of the recursive aggregator is smaller in bad times. Equivalently, the investors' certainty equivalence functional (CEF) is quasiconvex.

1.2. Discontinuous Prices and Discontinuous Information

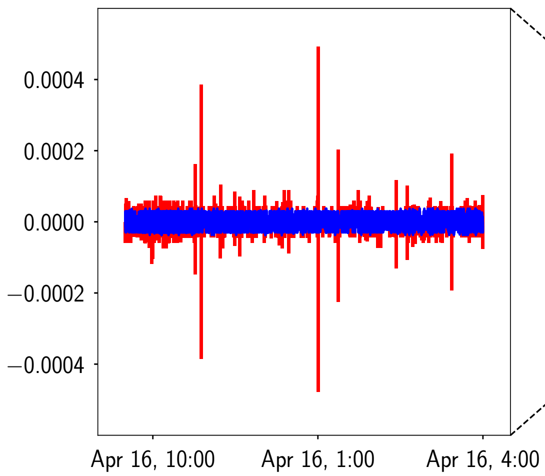
1.2.1. Stylized Features of the Data

I motivated this project by clamming that prices jump a great deal and that announcements frequently and dramatically affect asset prices. The literature has shown this, but to fix ideas it is helpful to investigate the matter ourselves, at least at a cursory level. 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 using data sampled once per second, and the blue lines contain the diffusion (continuous) part of the process and small jumps. The behavior in this graph is completely typical. I purposefully chose April 16, 2012 as a completely normal day in the markets.

As we can see in Figure 1, there clearly are many, many jumps. In addition, they 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 index using a ratio of bipower-type estimators. This wide divergence between various estimates likely arises from from the difficulty in disentangling the infinite-activity jumps from the diffusive part. The precise percentage is not important for this

Figure 1: S&P 500 Log-Return



(a) 1 Second (b) Daily

paper. In fact, I estimate this proportion below, (Figure 6). Rather, the important takeaway is that there are sufficient jumps to be economically important, even 7 % is economically meaningful.

In addition, as far as I know, every paper that explicitly tests for the degree of activity finds infinitely active jumps, or at the very minimum a very large number, (Aït-Sahalia, Mykland, and Zhang 2005; Bakshi, Carr, and Wu 2008; Aït-Sahalia and Jacod 2009a). From both a modeling and pricing perspective, a large number of jumps and infinitely many are effectively equivalent in practice, as shown in detail later. As a consequence, it is clear that jumps are ubiquitous and economically and statistically meaningful, but their the precise magnitude of the variation is less settled.

1.2.2. What Causes Jumps?

To understand Figure 1a fully, 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 argue that jumps are responses of prices to news, including Andersen, Bollerslev, Diebold, and Vega (2003, 2007), Beechey and Wright (2009), and Lahaye, Laurent, and Neely (2011). 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.

However, in general many different sources cause discontinuities in investor's information sets. Other source 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 of the potentially relevant events. We cannot construct investors' actual information sets. (Note, this paper uses *news* quite broadly. It refers to any discontinuous change in information, not just traditional news sources like newspapers.) As illustrated by the examples above, news can come at unpredictable times and be observed by only a few investors, and so picking a priori what news items are relevant will leave many relevant items out. In addition, there is no reason to assume that the resultant price change is in any way large. 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 above back it up. However, the connection is actually even more fundamental. Delbaen and Schachermayer (1994) show no-arbitrage implies prices are semimartingales.[3] In that framework, which is now standard in the literature, 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 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, i.e. a news item is released. This 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 predictable 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 also helps us understand why not all price changes are jumps. Not all new information is immediately reflected in prices in its final form. Some information takes time to process before the market participants can use it effectively. For example, after firm announces its earnings, we see many articles analyzing what this news implies about the stock in question as well as related assets. As various investors update their beliefs and buy or sell accordingly, other market participants will adjust their beliefs and also buy or sell. This drives changes in the overall price. This type of adjustment is likely to be smooth. Thinking through the evidence available takes time.

3. Throughout this paper, I use no-arbitrage to refer to no-free lunch with vanishing risk.
4. Throughout this paper, I use 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 the objects by the first time they enter the representative investor's information set.

2. Literature Review

Since questions concerning volatility, news, and risk-return trade-offs are central to finance and economics a few different literatures have studied the questions considered in this paper. Consequently, 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 (2005) develop a bipower variation estimator showing how to separate out jumps from diffusive variation. Since then, several authors have shown that 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 key difference between my estimates of jump variation and their bipower variation estimate is that I am measuring ex-ante jump variation, while they measure ex-post variation. This is important for two reasons. First, the density characterizations I provide rely upon an ex-ante characterization. Second, the investors price ex-ante risk, and so my measure is a key 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 prices derivatives, such as (Pan 2002; Branger, Schlag, and Schneider 2008; Todorov 2010, 2011).

In Section 1.2.1, 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 huge number of jumps. If jumps are not infinitely-active, they certainly have a very high intensity. Second, jumps form a 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 the 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 very interesting paper and provides useful estimates on 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 stochastic process representation literature. Arguably the most novel contribution in this paper is Theorem 3 and following corollaries. This theorem provides some 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).

This representation of asset-prices as a time-changed Lévy process is useful because it allows us to extend the results and thought patterns that have been developed for diffusion processes to jump processes. The time-change method of representing price processes has a illustrious history. The first key result is the Dambis, Dubins & Schwarz theorem, (Dambis 1965; Dubins and Schwarz 1965). Theorem 3 is the jump analogue 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 have partially extended 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) has a positive result showing how to embed the infinitesimal jumps of a jump processes into an α-stable process using bipower-variation. In their asymptotic experiment, they shrink the maximum jump size towards zero.

By using an ex-ante measure of jump variation, instead of an ex-post one like Todorov and Tauchen (2014) do, I am able to handle large jumps as well. In addition, the pricing analysis that I do requires that the volatility measure is predictable. You cannot price ex-post variation measures such as bipower-variation.

In Section 4.3, I show how to derive time-aggregate these continuous-time representations to discrete-time under some additional assumptions. In doing these, I follow Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) who provide analogous results 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' utility functions, i.e. their risk appetite, implies a negative relationship between expected returns and volatility. Consequently, many different papers estimate this relationship, and I cannot survey this literature in a comprehensive fashion. Surprisingly, the empirical evidence has proven much less conclusive than the theory. Bollerslev, Engle, and Wooldridge (1988), 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. In addition, a vast number of authors have argued that the instantaneous correlation, which is often referred to as 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 with opposite signs.

Investors' utility functions are not the only place their preferences can have curvature. A very meager selection of models with recursive preferences have 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 leads to additional risk-return trade-offs. Ai and Bansal (2018) show premia for this curvature cause announcements to be priced differently. That is 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. Since then several authors extended this result in various directions showing, for example, that the cross-sectional returns are better behaved on FOMC days, (Savor and Wilson 2014; Karnaukh 2016). Other authors have showed that it is the FOMC in particular that has this effect. Macro announcements and other central banks do not seem to to behave in the same way, (Brusa, Savor, and Wilson 2018; Law, Song, and Yaron 2018).

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 is useful because it implies that jump volatility nonparametrically identifies the CEF's curvature. It is well-known that estimating the intertemporal elasticity of substitution in a world with Epstein-Zin preferences is difficult. Obviously, estimating the curvature without parametric restrictions is even more difficult.

I explain Lucca and Moench's (2015) as compensation for uncertainty, i.e. a covariance-based explanation. This conflicts with Ai and Bansal's (2018) explanation, but is in close accord with the evidence presented Savor and Wilson (2014). They show that covariance-based explanation of the cross-sectional returns work particularly well on FOMC announcement days. It is also consistent with Mueller, Tahbaz-Salehi, and Vedolin (2017) who document a premium in foreign exchange markets for holding foreign currencies vis-á-vis the U.S. dollar on FOMC days. They explain this through increased uncertainty on those days as mediated by financial intermediaries, i.e. they use a covariance based explanation.

3. DATA GENERATING PROCESS

In the previous sections, I promised to write down a model for the continuous-time process and derive the implied discrete-time representation. I turn to this task now. To avoid overwhelming the reader with too much complexity all at once, I build up the continuous-time data generating process (DGP) step-by-step. I do this by following the same path that the literature took over the past few decades. In particular, I adopt the semimartingale framework that is standard in this literature and Delbaen and Schachermayer (1994) show no-arbitrage implies. In addition, in the following section I abstract away from any drift dynamics, bringing them back in later.

3.1. Continuous-Time DGP

The initial semimartingale models, such as Black and Scholes (1973), use diffusions. Here, the log-price $p(t)$ is portrayed in Definition 1, where $W(t)$ is a Wiener process, and $\sigma^2(t)$ is the diffusion volatility.

Definition 1. Diffusion Process

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

Almost immediately, various authors recognized that returns have time-varying volatility, and so they turned to using stochastic volatility diffusions. One of the main benefits of such an approach is that Dambis (1965) and Dubins and Schwarz (1965) proved that we can represent any continuous martingale that has a time-derivative in this fashion.

Definition 2. Stochastic Volatility Diffusion Process

$$dp(t) = \sigma(t) \, dW(t) \tag{2}$$

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. In the more recent decades, various researchers have worked very hard to add jumps to these models.

Two main methods dominate the literature. The first is a parametric approach where the author assumes a parametric form for the jump magnitudes and arrival times and magnitudes. This approach is often used the discrete-time finance literature, and so I return to it when I relate my results to theirs, e.g. (Drechsler and Yaron 2011; Tsai and Wachter 2018).

The second, a nonparametric approach, is closely related to what we are doing here. This strain of literature author models prices as Itô semimartingales. This is very general because all you are assuming is that the prices are semimartingales and each of the components of the process have time-derivatives. This implies that the jump part can be represented as an integral with respect to a Poisson random measure.

I add to this one more assumption — the price process is locally-square integrable. This implies that the jump measure has a predictable compensator, and so we can simplify our notation. Although, most high-frequency papers initially allow for jumps that are so large they do not have a compensator, they almost all restrict themselves to such processes when they derive estimators. By assuming it now, I simplify my notation.

Definition 3. Jump-Diffusion Process

$$dp(t) = \sigma(t)\, dW(t) + \int_X \delta(t,x)(\mu - \nu)(dt, dx) \tag{3}$$

Here, as above, $\sigma^2(t)$ is the diffusion volatility and $W(t)$ is a Brownian motion. I use μ to refer to a Poisson random measure with associated compensator ν. Intuitively, μ treat treats jumps with each possible magnitude as a Poisson process, and then it integrates over the various magnitudes. By subtracting off ν, we are recentering the process so that it is conditionally mean zero at every point in time. In this representation δ completely controls the dynamics of the process. For each open set $A \subset X$, $\int_A \delta(t,x)\, dx$ is the intensity of the Poisson process with magnitude $x \in A$.

This representation is very 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 want to estimate an infinite-dimensional object with at most realization. In addition, it is not obvious how to parsimoniously map this to discrete-time.

3.2. Discrete-Time DGP

To take a step back, what is a discrete-time return in this context? The return is just the change in — an increment of — the price process over some length of time, say a day.[5] Recall, throughout I use subscripts to refer to daily objects, and functional notation to refer to stochastic processes, Also, I adopt the convention that the time associated with a variable is when it first becomes known to them investor, i.e. measurable with respect to the relevant filtration. For example, r_τ is a daily return on date τ, while $p(\tau)$ is the log-price at date τ.

Definition 4. Daily Return

$$r_t \coloneqq \int_{t-1}^{t} dp(s) \tag{4}$$

This return has a density h in each period given the available information at the end of the day before $\mathcal{F}_{t-1}$. If we want to fully understand the risk market participants face, we must understand how this density evolves. In addition, understanding $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ is sufficient to understand all of the statistical risk the investor faces. In particular, any statistical measure of risk, such as Expected Shortfall or Value-at-Risk, is a statistic of this density.

Definition 5. Daily Density

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

As a consequence, this density is a key variable of interest in financial econometrics, and thousands of papers have been written modeling it and its statistics. In the next section, I consider how this is generally done, and then frame the question considered in this paper in that framework.

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

3.3. Modeling the Daily Return Density

Daily returns are not very well-behaved objects in that they are unpredictable and their distributions vary substantially over time. Furthermore, we only observe one observation for each $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$. Since $\mathcal{F}_{t-1}$ grows each day, $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ is a function-valued time-varying parameter.

Although, contemplating the dynamics of non-identified function-valued objects might provide a great deal of intellectual stimulation, it does not help much with measuring risk in practice nor with using these tools to answer other interesting questions. Consequently, the literature has looked for representations for $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ in terms of a well-behaved sufficient statistic for the dynamics, which I denote x_t. (Engle 1982; Bollerslev 1986; Nelson 1991). The most common choice for x_t is some volatility measure.

They use x_t to separate $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ 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 faced by investors but not the dynamics of the density. The third part — G — is a process governing the dynamics of x_t.

I make this precise in Equation (6). Note, both f and G are fixed across-time, and if we have chosen x_t well, G will be simple.

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

This representation replaces the question how should we model $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ 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.

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

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

3.4. Modeling the Realized Density

3.4.1. The Diffusion Case

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, before I turn to the main task at hand and derive a new sufficient statistic, we should define the measure I take from the literature — diffusion volatility.

In the continuous-time data generating process of Definition 2, 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 the diffusion part of the price.)

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

$$\sigma_t^2 \coloneqq \frac{1}{\Delta}\mathbb{E}\left[\left|p^D(t+\Delta) - p^D(t)\right|^2 \,\middle|\, \mathcal{F}_{t-}\right] \tag{9}$$

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 important in the diffusion case and so has not been highlighted by the literature. In the jump case, however, it is fundamental.

Again, to relate the continuous- and discrete-time objects, I define the integrated diffusion volatility.

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

$$\sigma_t^2 \coloneqq \int_{t-1}^{t} \sigma^2(t)\, ds \tag{10}$$

Barndorff-Nielsen and Shephard (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) exploit the aggregation property implied by the integrated diffusion volatility being the integral of the instantaneous diffusion volatility to develop the Realized Volatility — RV_t — estimator for σ_t^2. They further show that conditional on the integrated diffusion volatility, the daily density of the return is Gaussian in the pure diffusion case. This conditioning is meaningful as the integrated diffusion volatility is only known at the end of each day. In other words, $\sigma^2(t)$ varies intra-day, which is why they did not contrary-to-fact show that returns are Gaussian.

This conditional Gaussiianity separates the daily return distribution into a well-behaved component in terms of the volatility and a Gaussian noise component. To relate it to the previous discussion, we have the following decomposition for $h\left(r_t \,\middle|\, \mathcal{F}_{t-1}\right)$.

$$f\left(r_t \,\middle|\, \sigma_t^2\right) = f\,\Bigg|_{x_t = \left[\int_{t-1}^{t} \sigma^2(s)\, ds\right]} = N\left(0, \int_{t-1}^{t} \sigma^2(s)\, ds\right) \tag{11}$$

Clearly, if we track $f\left(r_t \,\middle|\, \sigma_t^2\right)$, we know everything about the dynamics and can compute any risk measure of interest. Therefore, I define the realized density as f evaluated at the realized value of x_t.

Definition 8 (Realized Density)**.**

$$RD_t \coloneqq f\left(r_t \,\middle|\, x_t\right) = f\left(r_t \,\middle|\, \sigma_t^2\right) \tag{12}$$

The realized density is the density generalization of realized volatility. It is useful because it allows us to convert the complicated variation in $h\left(r_t \,|\, \mathcal{F}_{t-1}\right)$ into much simpler variation in x_t. In addition, we can derive estimators for it using infill asymptotic arguments.

3.4.2. The Jump-Diffusion Case

If we could derive a realized density in the jump-diffusion case as we did in the diffusion case we would be able to transplant many of the useful results from the diffusion case to the jump case. To do this, we need to determine what to use for x_t. To do that, I define the *jump volatility*—$\gamma^2(t)$. A volatility is a variance, and so recall the definition of the instantaneous diffusion volatility.

$$\sigma^2(t) \coloneqq \frac{1}{\Delta}\mathbb{E}\left[\left|p^D(t+\Delta) - p^D(t)\right|^2 \,\middle|\, \mathcal{F}_t\right] \tag{13}$$

The most obvious definition of the jump volatility is the analogous expression for the jumps, except we substitute the diffusion part of the prices — $p^D(t)$ with the jump part—$p^J(t)$. I define the instantaneous jump volatility as the local variance of the jump part $p^J(t)$.

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

$$\gamma^2(t) \coloneqq \frac{1}{\Delta}\mathbb{E}\left[\left|p^J(t+\Delta) - p^J(t)\right|^2 \,\middle|\, \mathcal{F}_{t-}\right] \tag{14}$$

The integrated jump volatility is defined in the obvious way.

Definition 10 (Integrated Jump Volatility)**.**

$$\gamma_t^2 \coloneqq \int_{t-1}^{t} \gamma^2(s)\, ds \tag{15}$$

We can also define the jump volatility in terms of Equation (6). 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)**.**

$$\gamma_t^2 = \int_{t-1}^{t} \gamma^2(s)\, ds = \int_{t-1}^{t}\int_{\mathcal{X}} \delta^2(s,x)\nu(dx,ds) = \langle p^J\rangle(t) - \langle p^J\rangle(t-1) \tag{16}$$

The instantaneous jump volatility is its time derivative. The diffusion volatility has the same relationship to the predictable quadratic variation of the diffusion part. This relationship is usually discussed in terms of the quadratic variation, not the predictable quadratic variation, but in the diffusion case these two measures coincide. In the jump case, however, they do not. If we were try to remove the expectation from the definition of the jump volatility, the resulting object is not well-defined because the quadratic variation of the jump part is not absolutely continuous.

I now discus the realized density in the jump case. In this case, the return has two parts: $dp(t) = \sigma(t)\, d(t) + \int_X \delta(t,x)(\mu - \nu)(dx, dt)$ Conditional on the values of $\sigma^2(t)$ and $\delta(t,\cdot)$, the jumps and diffusion parts are independent. Consequently, we can think of returns as 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 we can view the diffusion part as a Gaussian density with variance equal to the integrated diffusion volatility. If we can develop a parametric expression for the jump part,we get a representation for the price effectively for free.

In the following sections, I develop the representation theory necessary to do this. Surprisingly, we can derive a parametric family for the jump realized density under some economically innocuous assumptions. Here, it is Laplace instead of Gaussian. Before I explain where this comes from, we have the following expression for the realized density of a jump-diffusion. (It may be useful to recall that $*$ is the convolution symbol.)

$$RD_t \coloneqq f\left(r_t \,\middle|\, \sigma_t^2, \gamma_t^2\right) = f\Bigg|_{x_t=\left[\int_{t-1}^t \sigma^2(s)\,ds,\ \int_{t-1}^t \gamma^2(s)\,ds\right]} = N\left(0, \int_{t-1}^t \sigma^2(s)\,ds\right) * \mathcal{L}\left(0, \int_{t-1}^t \gamma^2(s)\,ds\right) \tag{17}$$

4. Modeling Jump Processes

In the previous section, I stated that the realized density in the jump-diffusion case was a convolution of a Gaussian and Laplace densities. In this section, I explain the novel contribution part of that result. To be precise, I explain why the jump part is conditionally Laplace.

4.1. Static Jump Processes

To build up the intuition, I show what happens when we have a simple jump process where the locations of the jumps are Poisson distributed and the magnitudes are i.i.d. Gaussian variables. I will then show how we can reduce the general case to this one.

For the ease of exposition define $N(t)$ as the process that determining when $p^J(t)$ jumps. Clearly $N(t) - N(t-)$ equals one if and only if we have a jump at time t.

Definition 11. Location Process

$$N(t) \coloneqq \sum_{s \leq t} 1\left\{\left|p^J(t) - p^J(t-)\right| > 0\right\} \tag{18}$$

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

$$p^D(t) = \sum_{s \leq t} \kappa(s)|N(s) - N(s-)| \tag{19}$$

The representation in Equation (19) has variability arising 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.

$$\gamma_t^2 = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}\left[\left| p_{t+\Delta} - p_t \right|^2 \middle| \mathcal{F}_t \right] \tag{20}$$

Applying the law of iterated expectations.

$$= \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}_{N(t+\Delta)-N(t)} \left[\sum_{i=1}^{N(t+\Delta)-N(t)} \mathbb{V}\mathrm{ar}(\kappa(t)) \middle| \mathcal{F}_t, N(t+\Delta) - N(t) \right] \tag{21}$$

$$= \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}[N(t+\Delta) - N(t)] \mathbb{E}[\kappa(t)^2] \tag{22}$$

$$= \frac{1}{\Delta} \Delta = 1 \tag{23}$$

To put Equation (23) 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 is true in general if the intensities and magnitudes are independent. We are combining the variation from the jump locations and the jump magnitude into one 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 exactly the same way. For example, if we double the variance of the magnitudes or the intensity, the volatility doubles. This is the random analogue to the statement that variances of sums of i.i.d random variables are sums of variances.

4.1.1. Variance-Gamma Process

In the previous section, I encouraged you to consider a process with Poisson distributed locations and Gaussian distributed jump magnitudes. Moving forward, we want to simultaneously generalize this process to allow for dynamics and find processes where the jump volatility is a sufficient statistic for the dynamics. We do not want to use the intensity and magnitude functions because they are not identified. The jump volatility, on the other hand, can be estimated from high-frequency data.

One obvious generalization of previous example is the compound Poisson case. A compound Poisson process is composed as the sum of finitely-many independent Poisson processes with different intensities/magnitudes. This representation will not quite work for our purposes. The variation I use to identify the jump volatility comes from the data's infinitely-active jumps. However, a compound Poisson process can have at most finitely many jumps in any finite interval. This comes from the requirement that jump processes $N(t)$ must converge.

However, if we consider the limiting case when $N(t) \to \infty$, we can work around this problem. This brings its own problem. We cannot have infinitely jumps of magnitudes greater than any fixed $\epsilon > 0$ in a finite interval otherwise the jump will diverge. Unlike in the finite $N(t)$ case, where there the structure between the jumps and magnitudes is completely unrelated, as $N(t) \to \infty$ to maintain local square-integrability we need to place some constraints on how the magnitudes act.

In particular, we can shrink the size of the increments towards zero. Intuitively, we can have infinite sums of arbitrarily small summands can converge to a finite value. We can allow for infinite intensities. In fact, one common pure-jump process — the variance-gamma process — is an infinite-activity "Poisson process" with arbitrarily small Gaussian-distributed summands.

This process is often used in modeling options. 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.

Definition 12 (Variance-Gamma P:ocess)**.**

$$\text{Variance-Gamma}(t) \coloneqq W(\Gamma(t)) \tag{24}$$

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.[6] The exponential distribution is a special case of the Gamma distribution. (Just 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.[7]

To get some intuition on where this Laplace distribution is coming from, it is helpful to recall the following characterization of a Laplace 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.

$$z \sim \mathcal{L}(\text{mean} = 0, \text{variance} = 1) \iff z \sim \frac{\sigma}{\sqrt{2}} N(0, 1), \ \sigma^2 \sim \exp(1) \tag{25}$$

Equation (25) is the discrete analogue of Definition 12. 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 exponentially distributed.

However, one way of characterizing a Poisson process is as a process where the time between jumps is exponentially distributed. This is well-defined because the length of the intervals take

6. I introduce the notion of a standard variance-gamma process here to facilitate exposition because it has some nice aggregation properties that the general case does not.

7. This property does not hold for general variance-gamma processes.

positive real values. This brings us back to the initial discussion of a variance-gamma process as a "compound Poisson" process with an infinite intensity.

The previous discussion focused exclusively on a process without any dynamics, but one might hope that by allowing the volatility to vary over time, we can approximate a wide variety of processes. This turns out to be the case as I show later.

$$dp(t) = \int_{\mathcal{X}} \delta(t,x)(\mu - \nu)(dt, dx) \text{ becomes } dp(t) = \frac{\gamma(t)}{\sqrt{2}}\, d\mathcal{L}(t) \tag{26}$$

At each time, I replace a function $\delta(t,\cdot)$ with a single scalar $\gamma^2(t)$. In addition the integrator is switched from a Poisson random measure to a standard variance-gamma process. This simplifies analysis generally, as $\delta(t,\cdot)$ is not identified, while γ_t^2 is.

4.2. Dynamic Price Processes

In the previous section, I introduced a model for the jump part of the process. In this section, I characterize the class of price processes that have price processes that can be simplified in the way Equation (26) is. To do this, I introduce the class of processes I consider and derive the representation from there.

As is standard in the high-frequency literature, I assume that prices are absolutely continuous semimartingales. This allows me to write out the following representation for the price process. This Grigelionis form of an Itô semimartingale is the standard nonparametric model in the high-frequency literature. I start by simplifying this process. Here, $p(0)$ is the price at time 0, and $\mu(t)$ is a finite-variation predictable process (a drift term).

Definition 13. Grigelionis Form

$$\begin{aligned} p(t) = p(0) + \int_0^t \mu(s)\, ds + \int_0^t \sigma(s)\, dW(s) + \int_0^t \int_X \delta(s,x) 1\{\|\delta(x,s)\| \leq 1\}(\mu - \nu)(ds, dx) \\ + \int_0^t \int_X \delta(s,x) 1\{\|\delta(x,s)\| > 1\} \mu(ds, dx) \end{aligned} \tag{27}$$

By assuming that the returns have variances, or equivalently that $p(t)$ is locally-square integrable, we can combine the last two terms.[8]

$$p(t) = p(0) + \int_0^t \mu(s)\, ds + \int_0^t \sigma(s)\, dW(s) + \int_0^t \int_X \delta(s,x)(\mu - \nu)(ds, dx) \tag{28}$$

8. To see why this is the case, note the following. We need a predictable compensator of the large jumps, i.e. the set where $1\{\|\delta(s,x)\| > 1\}$. Since squaring values greater than 1 makes them larger, and the compensator of the squared process is well-defined, the compensator of the non-squared process must be as well.

Then we can switch the variance-gamma integral representation for the jump part under some conditions and assume without loss of generality that $p(0) = 0$.

$$= \int_0^t \mu(s)\, ds + \int_0^t \sigma(s)\, dW(s) + \frac{1}{\sqrt{2}} \int_0^t \gamma(s)\, d\mathcal{L}(s) \tag{29}$$

The previous paragraphs make explicit the return's data generating process and how it relates to $\gamma^2(t)$. The validity of the representation in Equation (29) and the structure of its proof are closely analogous to the diffusion case, and so I will start by reviewing that case. The predicable quadratic variation, or angle-bracket, is the integral of the squared process projected onto the information set available up to that time, i.e. $\mathcal{F}_{t-}$, i.e the predictable compensator of the quadratic variation. Intuitively, it is the time-varying conditional mean of the quadratic variation. In the diffusion case, it is the integrated diffusion volatility.

$$\left\langle p^D \right\rangle(t) = \int_0^t \sigma^2(s)\, ds \tag{30}$$

4.2.1. Time-Changes Theorems Versus Central Limit Theorems

The validity of diffusion representation for general continuous processes can be traced back the Dambis-Dubins-Schwarz theorem, which shows that any continuous martingale time-changed by its predicable quadratic variation is a Brownian motion (Dambis 1965; Dubins and Schwarz 1965). (I use an equal sign with an $\mathcal{L}$ above it to refer to equality in law.) Here a time-change means to evaluate at a random clock. The expression on the right-hand of the equals sign in Equation (31) is to be understood as a functional composition of the time arguments.

$$p^D(t) \overset{\mathcal{L}}{=} W\left(\left\langle p^D \right\rangle(t)\right) \tag{31}$$

The Dambis-Dubins-Schwarz theorem is a very nice result and underlies the representation of continuous martingales as stochastic-volatility diffusions. However, it only applies to the diffusion part of the process. Monroe (1978) shows you can can represent arbitrary semimartingales as time-changed Wiener processes. Conceptually, this is a very nice result, but since the representation is not explicit and the time-change is not identified, this result has not proved particularly useful in practice.

The key difference between the jump part and the continuous part of a semimartingale is that all of the ex-ante variation in the continuous part comes from variation in magnitudes, while in the jump part we have two sources of ex-ante variation — the magnitudes and the locations. Intuitively, the Dambis, Dubins, and Schwarz theorem shows how to separate the variation in any continuous martingale into a predictable part (the volatility) and an i.i.d. remainder. By doing this the martingale becomes a sum of appropriately scaled independent random variables. In other words, it is a "central limit theorem".[9] In fact one method of characterizing standard central limit

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

theorems is as special cases of this result.

In the jump case though, things are slightly more complicated. Not only do we have variation in magnitudes, we also have variation in 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 this number of summands is geometrically-distributed, various geometric-stable central limit theorems exist that tell us how the sum behaves as the expected number of summands approaches infinity, (Mittnik and Svetlozar 1993; Kozubowski and Svetlozar 1994).

So the problem at hand is is there an analogue to the Dambis,Dubins and Schwarz theorem that has the same relationship to the geometric stable central limit theorems that the Dambis, Dubins and Schwarz theorem does to the standard central limit theorem.

4.2.2. Jump Process Representation Theorem

Before we do this, we need to make two economically innocuous assumptions that are not completely standard in the literature. It turns out that the predicable quadratic variation $\langle p \rangle(t)$ is not necessarily well-defined. We need to assume that the process is locally-square integrable. In other words, the process has finite variance over finite intervals. This is a constraint on the very large jumps because the only way it can fail is if there is a large probability of many very large jumps. Although is not a standard modeling assumption, it is almost always assumed by the theorems proving laws of large numbers or central limit theorems.

Assumption SQ. The purely-discontinuous part of $p(t)$ is locally-square integrable.

The other restriction is on how frequently jumps occur—we need the jump part to have infinite activity. In other words, we need there to be at least one jump in every finite interval. This does two things, 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.

The assumption above sounds very restrictive 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. Although, the precise percentage of variation in the returns is more contentious. In addition, standard variance-gamma processes are limits of compound Poisson process. Hence, if we had only finitely-many jumps the predictive density would be a mixture of a Laplace and a point mass of zero. We would be approximating it with standard variance-gamma process with the same variance as the mixture above As long as the probability of a jump is reasonably high, this will work well in practice.

Assumption I. The process $p(t)$ has infinite-activity jumps.

The last assumption is that it does not have any predictable jumps. What this means is that there does not exist any times τ such that event $p(t)$ jumping at τ is contained in an information set $\mathcal{F}_{\tau-}$.

Having made those assumption, we can now state the following theorem. Interestingly, we do not need to assume that that $p(t)$ has a time derivative.

Theorem 3 (Time-Changing Jump Martingales)**.** *Let $p^J(t)$ be a purely discontinuous, infinite-activity, locally-square integrable martingale that can be represented as $H * (\mu - \nu)$ where $H(t)$ is a predictable process, μ a Poisson random measure, and ν its predictable compensator with Lebesgue base Levy measure. Further assume that it has no predictable jumps.*

Then $p^J(t)$ time-changed by its predictable quadratic variation is a standard variance-gamma process. In other words, $p^J(t) \stackrel{\mathcal{L}}{=} \mathcal{L}\left(\left\langle p^J \right\rangle(t)\right)$.[10]

4.2.3. Processes with Finite-Activity Jumps

Arguably the most controversial assumption that I made was Assumption I. 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, we get a point mass at zero if it does not jump, and if it does we are back in the previous case. 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.

To to do this the first thing 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 conditionally independent.

I condition on the number of jumps and show that the magnitudes are a continuous process in that space. Thus we can apply the Dambis, Dubins & Schwarz theorem there, getting 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 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.

Corollary 3.1 (Time-Changing Finite-Activity Jump Martingales)**.** *Let $Y(t)$ be a purely discontinuous, locally-square integrable martingale that can be represented as $H * (\mu - \nu)$ where $H(t)$ is a predictable process, μ a Poisson random measure, and ν its predictable compensator with Lebesgue base Levy measure λ. Further assume that it has no predictable jumps. Then $Y(t)$ time-changed by its predictable quadratic variation is a mixture of the 0 process — δ_0 – and the standard standard variance-gamma process where the mixing weights are the intensity of the jump process.*

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

Time-changed results are not particularly intuitive, and so we would like an integral representation as well. Completely analogously to how the time-changed Brownian motion results imply that you can represent continuous martingales as integrals with respect to Brownian motion as long as the relevant characteristics are absolutely continuous, you can represent jump martingales as integrals with respect a standard variance-gamma process under the same conditions.

Corollary 3.2 (Jumps Processes as Integrals)**.** *Let $p(t)t$ be a purely-discontinuous Itô semimartingale that is locally-square integrable and has infinite-activity jumps. Then $p(t) = \frac{1}{\sqrt{2}} \int_0^t \gamma(s) \, d\mathcal{L}(s)$, where $\mathcal{L}$ is a standard standard variance-gamma process, and γ is a predictable function.*

4.2.4. Representing the Entire Process

All of the above results were only concerned the jump martingale part of the process. To take it do the data, we need a representation for the entire process. To get that we just combine it with previous representations in the literature.

Theorem 4 (Locally Square-Integrable Itô Semimartingales as Integrals)**.** *Let $p(t)$ be an Itô semimartingale that is locally-square integrable and has infinity-activity jumps. Then we can represent $p(t)$ as follows.*

$$p(t) = \int_0^t \mu(s) \, ds + \int_0^t \sigma(s) \, dW(s) + \frac{1}{\sqrt{2}} \int_0^t \gamma(s) \, d\mathcal{L}(s) \tag{32}$$

Proof. This is a straightforward implication of the standard representation of an Itô semimartingale and Corollary 3.2 □

4.3. Discrete-time Price Processes

I now derive a discrete-time representation from the time-change representation in Theorem 3. If the volatilities were constant, we could rewrite the distributions of increments as the scale times the density, e.g. $N(0, \sigma^2) = \sigma N(0, 1)$.[11] The question becomes when can we do this if the volatilities are not constant? We need some form of independence between the volatilities and the innovations. In other words, we a restriction on the leverage. Since the jump part is purely discontinuous, it is orthogonal to the continuous 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 motion and Laplace motion the martingale property is sufficient to imply that the integrators are independent of everything else except possibly their own volatility. To aggregate we need to separate the volatility from the innovation part. Consequently, we must assume the innovations to the volatilities are independent from the innovations to their respective integrators.

I am not assuming away all dependence. I place no restrictions on the predictable relationship between the mean and volatilities. As long as it a positive amount of time for feedback from the volatilities to affect the level of the prices or vice-versa, this assumption is satisfied.

11. The Laplace density also can be rewritten in scale form, except we have a $\sqrt{2}$ factor as well. $\mathcal{L}(0, \gamma^2) = \frac{\gamma}{\sqrt{2}} \mathcal{L}(0, 1)$. This is where the $\sqrt{2}$ was coming from in the integral representations.

In addition, 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 will leave to future work the correct way of bringing it into our framework. This could be done 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.

As long as that assumption holds, the correct discrete-time representation of the return is as follows, where I divide the innovation distribution of the log-return into two parts. The first part arises from the continuous part of the price process, and so it is a scale-mixture of Gaussian random-variables. The second part arises from the jump part, and so it is a scale-mixture of Laplace random variables.

$$RD_t = N\left(\int_{t-1}^{t} \mu(s)\, ds, \int_{t-1}^{t} \sigma^2(s)\, ds\right) * \mathcal{L}\left(0, \int_{t-1}^{t} \gamma^2(s)\, ds\right) \tag{33}$$

Then to recover the predictive density, we need to integrate out the drift μ_t and volatilities: σ_t^2 and γ_t^2.

$$h\left(r_t \,|\, \mathcal{F}_{t-1}\right) = \int_{\mu_t, \sigma_t^2, \gamma_t^2} \left[N(\mu_t, \sigma_t^2) * \mathcal{L}(0, \gamma_t^2)\right] dG\left(\mu_t, \sigma_t^2, \gamma_t^2 \middle| \mathcal{F}_{t-1}\right) \tag{34}$$

5. Estimation

In the previous section, I derived a representation of the processes in terms of $\sigma^2(t), \gamma^2(t)$, and $\mu(t)$ and their discrete-time analogues. If we want to use this framework in practice, we must estimate these parameters. As is well known, μ_t is not identified. On the other hand, we know that the diffusion volatility is identified, and I show that the jump volatility is identified by deriving an estimator for it now using infill asymptotics. In particular, I show that we can consistently estimate $\gamma^2(\tau)$ for any fixed τ.[12]

5.1. Assumptions

I start by fixing some notation and defining some assumptions. The way that the instantaneous volatility estimators work is that you consider an appropriately defined average over an increasing number of increasing number of increments over a shrinking interval. In other words, for a given index — n, we have a triangular array of increments. In addition, we have both a true D.G.P. with time-varying volatility and an approximate D.G.P. where I assume that the volatility is locally constant.

This makes the notation rather complicated because I have to keep track of both triangular arrays as I take limits. I adopt the notation used in Jacod and Protter (2012) for the most part.

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

Specifically, I use $\Delta_i^n X$ to refer to a increment i in process $X(t)$ of length Δ^n, and I take limits with respect to n, that is $\{\Delta_i^n X\}$ is a triangular array of increments of $X(t)$.

Having done that I can lay out a series of assumptions that we will need to prove the consistency results that follow. They are very similar to the standard assumptions used in the literature. When relevant 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).

This is basically the same as assumption H in the literature. I replace the assumption on the jumps with a slightly more general and much more straightforward one. I also slightly modify the literature's assumption SH. (Here ω is used to index the underlying probability space Ω.)

Assumption SHL. We have HL and there is a constant A such that the following hold.

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

These two assumptions have a close relationship, Assumption HL is the local version of Assumption SHL. In other words 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 estimator we can under the other as well. This means that in many of 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. This is the type of convergence that you need to derive valid confidence intervals in this literature.

Lemma 5 (HL implies SHL)**.** *If $X^n(t) \xrightarrow{\mathcal{L}\text{-s}} X(t)$ under Assumption SHL, then $X^n(t) \xrightarrow{\mathcal{L}\text{-s}} X(t)$ under Assumption HL, and the equivalent statement holds for convergence in probability.*

To reduce the amount of notation, I adopt the following notational convention so that our processes are well-defined over the entire line, not just the place we are estimating them. Essentially, I set the processes equal to zero outside of the relevant window.

$$i \in \mathbb{Z}, i \leq 0 \implies \Delta_i^n X = 0 \tag{35}$$

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

$$I(i,n) \coloneqq [(i - k_n - 1)\Delta^n, (i-1)\Delta^n] \tag{36}$$

If we choose a sequence $i \to \tau$, then the interval clearly approaches τ from the left. Also, since I assume $X(t)$ is one-dimensional, I can assume without loss of generality that the driving Wiener and variance-gamma processes are one-dimensional.

5.2. Estimating the Instantaneous Volatilities

Now, that I have finished specifying the framework, I state the estimators themselves. The intuition behind the convergence of the following estimator is that we are averaging the volatilities over shrinking intervals that are approaching τ from the left. As long as the infill asymptotics imply the number of increments are averaging over is increasing faster than the length of the interval is shrinking we can precisely estimate the volatility. Since we estimating the process from the left, we are approximating the value before τ, i.e we are estimating $\sigma^2(\tau)$ or $\gamma^2(\tau-)$.

I first drive 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 — $\left\langle p^D \right\rangle(t)$ — is straightforward. We can use the integrated volatilities sample analogue. In particular, we can use $\widehat{\langle p^D \rangle(t)}$ to estimate σ^2_τ by taking the time derivative.

The main difficulty here is separating the jump and diffusion variation. We do this by treating the large increments as jumps, where large is defined in terms of an asymptotic rate. Asymptotically, this eliminates large jumps, and the small jumps do not affect the asymptotic distribution.

I prove this theorem under the assumptions in Section 5.1 to simplify the exposition of the paper. The conclusion is the same as in Jacod and Rosenbaum (2013) I focus on the univariate case.

Theorem 6 (Estimating the Instantaneous Diffusion Volatility)**.** *Let $X(t)$ satisfy Assumptions HL and SQ. Let k_n, Δ^n be sequences such that $k_n \to \infty$, $\Delta^n \to 0$, and $k_n \Delta^n \to 0$. Let τ be a deterministic time with $0 < \tau < \infty$. Let $v_1^n = a(\Delta^n)^z$, where $z < \frac{1}{2}$, and $v_2^n \to 1$. Then we have the following convergence in probability.*

$$\widehat{\sigma^2(i^2)}^n(k_n, \tau, X) \coloneqq \frac{1}{k_n \Delta_n} \sum_{m=0}^{k_n - 1} v_2^n \left|\Delta_{i_n}^n X\right|^2 1\{\left|\Delta_{i_n}^n x\right| \leq v_1^n\} \xrightarrow{\mathbb{P}} \sigma^2(\tau-) \tag{37}$$

One might think that we can create a similar estimator for $\gamma^2(t)$ — form an estimator of $\left\langle p^J \right\rangle(t)$ and take its time derivative. 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, the reason for this is that by considering a specific time τ we are implicitly conditioning on it. Doing this removes some of the variation due to the locations, and over a small enough window the variation from the large jumps disappears. If we also truncate away variation arising from the small jumps, we have no variation left that we can use to identify the jump volatility.

It is well known that over a fixed interval, the quadratic variation of jump processes and diffusive processes behaves similarly when we use infill asymptotics, (Jacod, Podolskij, Vetter, et al. 2010).

This is not true if we consider other exponents. For example, if we consider powers greater than 2, only the jump part contributes in the limit. This helps explain why the literature has focused so intently on the quadratic variation.

If we consider shrinking intervals, this is no longer the case. Instead, it is the absolute value, not the square, where variance-gamma and diffusive processes have similar asymptotic properties.[13]

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 — $\sqrt{\Delta}$.[14] In addition, I show the limit contains both $\gamma^2(\tau)$ and $\sigma^2(\tau)$. On the negative side, it does not separate the two parts, even asymptotically.

On the positive side, 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 $\gamma^2(\tau)$. It is the mean of a convolution of $|\mathcal{L}|(t)$ and $|W|(t)$. Then we can plug $\sqrt{\widehat{\sigma^2(\tau)}}$ into this function, and numerically solve for $\widehat{\gamma^2(\tau)}$. When we do this, the sample argmax converges to the population argmax, which is the true value as the moment condition is suitably convex. In other words, I construct an extremum estimator for $\gamma^2_(\tau)$ in terms of the local means of $|\mathcal{L}|(t)$ and $|W|(t)$.

Theorem 7 (Estimating the Instantaneous Absolute Volatility)**.** *Let $p(t)$ be Itô semimartingale.*

Let $p(t)$ satisfy Assumptions HL, I, and SQ, and 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 $\operatorname{erfcx} := \frac{2\exp(x^2)}{\sqrt{\pi}} \int_x^\infty \exp(-s^2)\, ds$.[15]

$$\frac{1}{k_n\sqrt{\Delta_n}} \sum_{m=0}^{k_n-1} |\Delta^n_{i_n+m} p| \xrightarrow{\mathbb{P}} \mathbb{E}|N(0,1)|\sigma(\tau-) + \frac{\gamma(\tau-)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right) \tag{38}$$

The expression on the right is the mean of the convolution of $\left|N\left(0, \sigma^2_{\tau-}\right)\right|$ and $\left|\mathcal{L}\left(0, \gamma^2_{\tau-}\right)\right|$. The result shows is that the appropriately standardized local sample mean converges to this population mean.

Since we can estimate the absolute volatility and the diffusion volatility, we can combine them to get an estimator for $\gamma^2(\tau)$. To do this, we need to weight the difference between the approximate moment as a function of γ and the sample moment. In general, any convex weighting function of the differences will work. I use the absolute value of the difference between two values that are converging to the same value when $\widehat{\gamma\tau}$ equals the true value because it works well in practice. We could also use various other convex weighting functions for the difference.

Theorem 8 (Estimating the Instantaneous Jump Volatility)**.** *Let $X(t)$ satisfy Assumptions HL, I, and SQ, and 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 $\widehat{\sigma(\tau)_n}$ converge in probability to $\sigma(\tau-)$. Let $\gamma(\tau) > 0$ and g be strictly-increasing, convex, and continuous then we have the following.*

13. It is an interesting open question whether this equivalence in rates extends to other jump processes.
14. As an aside, neither of the processes are martingales. They are semimartingales.
15. erfcx is the scaled complementary error function. It is a reparameterization of Mill's ratio. Efficient, numerically stable implementations are provided by most scientific programming suites.

$$\widehat{\gamma(k_n, \tau, X)} := \operatorname*{argmin}_{\gamma} g\left(\left|\frac{1}{k_n\sqrt{\Delta}}\sum_{m=0}^{k_n-1}|\Delta^n_{i_n+m}X| - \mathbb{E}|N(0,1)|\widehat{\sigma(\tau)_n} - \frac{\gamma}{\sqrt{2}}\operatorname{erfcx}\left(\frac{\widehat{\sigma(\tau)_n}}{\gamma}\right)\right|\right) \tag{39}$$
$$\xrightarrow{\mathbb{P}} \gamma(\tau-)$$

5.3. Estimating the Integrated Volatilities

So, we have an estimator for the instantaneous jump and integrated volatilities. We want to estimate discrete increments of the volatilities. To do this, we use the obvious procedure and simply average the instantaneous estimators over the course of the day. The diffusion estimator defined this way will coincide with standard diffusion estimators in the literature up to edge effects.

Definition 14. Integrated Diffusion Volatility Estimator

$$\widehat{\sigma_t^2} := \frac{1}{\# i_n \in [t-1, t]} \sum_{t-1 < t_n \leq t} \widehat{\sigma_n^2}(k_n, t_n, X) \tag{40}$$

Definition 15. Integrated Jump Volatility Estimator

$$\widehat{\gamma_t^2} := \frac{1}{\# i_n \in [t-1, t]} \sum_{t-1 < t_n \leq t} \gamma_n^2\widehat{(k_n, t_n}, X) \tag{41}$$

Since we are averaging over estimates of the local volatility, if the instantaneous estimates are consistent then then in the integrated ones will be as well. Squaring various estimators does not affect consistency by the continuous mapping theorem.

6. SIMULATIONS

One key advantages of my representation is that we can simulate it easily if we can simulate the volatilities in continuous-time. 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 know as a Heston model.) 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. Here, θ is asymptotic mean, κ is the mean-reversion rate, and ω is a scale parameter.

$$dx(t) = \kappa(\theta - (x(t)) + \omega\sqrt{x(t)}\, dW(t) \tag{42}$$

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.

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}

Once, I have $\sigma^2(t)$ and $\gamma^2(t)$ I can plug them into the following continuous-time D.G.P.

$$dp(t) = \sqrt{\sigma^2(t)}\, dW(t) + \frac{\sqrt{\gamma^2(t)}}{\sqrt{2}}\, d\mathcal{L}(t) \tag{43}$$

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

6.2. Simulation Results

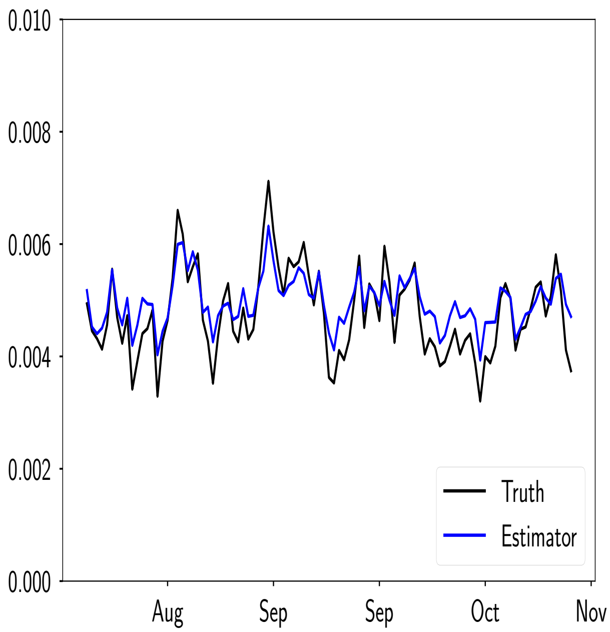
I focus on the volatility results below as they are sufficient statistics for all of the unknown objects. As long as they are bounded away from zero, which they are, the density estimators will perform well if they do. I chose the parameters for the Heston model to reflect the discrete-time dynamics observed in SPY. I report estimates at the 1 min and 300 ms frequencies to reflect the frequencies used in the data section and because they are far enough apart to observe how the estimators converge.

The attentive reader might notice that the two estimators deviations from the truth in Table 2 are almost perfectly negative correlated. This is because estimating $\sigma_t^2 + \gamma_t^2$ is significantly easier than solving the convolution problem and separating the two of them. It is frustrating because it makes interpreting the empirical results more difficult, but is just the nature of the problem. Also, in interpreting this graph, it is worth noting that the estimates are completely independent from day to day, and you never see the truth. I set the scale on the y-axis to cover reasonable values for the root volatilities.

In Table 2, I report the mean-square error of the previous estimates average of a year's worth of simulations. As can bee seen, as we increase the sampling frequency, the errors shrink quite significantly. This is precisely what the theory tells us. We are estimating the values consistently, and so the errors must shrink.

Figure 2: Simulation Results

(a) σ_t



(b) γ_t

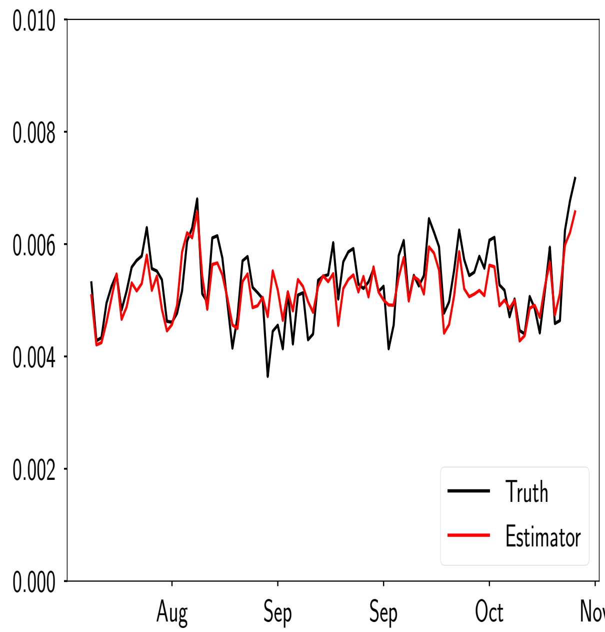


Table 2: $\frac{\text{RMSE}}{\mathbb{E}[\text{Root Volatility}]}$

Average over 250 days.	$\frac{\mathbb{E}[(\widehat{\sigma_t}-\sigma_t)^2]}{\mathbb{E}[\sigma_t]}$	$\frac{\mathbb{E}[(\widehat{\gamma_t}-\gamma_t)^2]}{\mathbb{E}[\gamma_t]}$
$\approx$ 180 obs. per minute	0.09	0.08
$\approx$ 1 obs. per minute	0.86	0.82

7. Data

I use three main sources of data in my analysis. First, we need an asset with that is traded at high-frequency to apply the method to. For the analysis to be interesting, we need a dataset that faces a dense stream of relevant news. So I chose SPY, (SPDR S&P 500 ETF Trust), which is an exchange traded fund that mimics the S&P 500. The S&P 500 is arguably the most important index of financial activity in the world. It is likely the most closely watched equity index and several extremely popular index funds track it directly. Consequently, the economics and finance literature has studied it extensively, often using it as a proxy for the market.

7.1. High-Frequency Data from TAQ

Since I am only using one asset, and it is among the most liquid assets available, 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

300-millisecond frequency, i.e. about 3 times per second. I use data starting in 2003 and going through September 2017. Since the asset is only traded during business hours, this leads to 3713 days of data with an average of $\approx 78\,000$ observations per day. The dataset takes up about 4.4 GiB of memory. It is also worth noting that SPY is by far the most liquid ETF, especially in recent years, reducing the effect of market microstructure: bid-ask spreads, bounces, rounding error, and so on.

This market microstructure makes the asset fail to be a semimartingale in practice. Thankfully, a substantial literature has developed to deal with precisely this issue. The two leading methods for dealing with it are sampling rather sparsely, say at a 5-minute frequency as argued for in Liu, Patton, and Sheppard (2015) and pre-averaging, where one takes appropriately weighted averages of the price over small (shrinking) intervals. Since, we need to separate the jump volatility from the diffusion volatility, we need to sample much more finely than once very 5 minutes. This is because effectively the only information available to separate the jumps and diffusive component nets come 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 is known to not affect consistency of the resultant estimators. The basic idea is rather simple, we average the price over a small interval to remove the noise. If we are careful at the rate at which we shrink the interval in relation to the rate at which we shrink the observation frequency, we can remove the effect of the noise from our estimates.

7.2. FOMC Announcement Dates and News Reports

The second source of data is a list of FOMC announcement dates. I obtain the precise times they make their announcements from Bloomberg. In my analysis I only use the information on the minute when the announcement was made. I am not trying to identify jumps that occur arbitrarily close to the announcement and so this should not affect my analysis. There are a total of 163 dates between May 1997 and September 2017.

For the third dataset, I perform a case study analyzing how $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ moved during the Fall of 2008 and how it relates to the news released then in Section 8.3.1. To do this, I need to a list of interesting news events during this period. I use the daily market reports put out by CNN Money's New York Office. These reports contain an expert-curated list of the interesting economic news released that day. This list often includes various market movements for both the market as a whole and individual important assets, earnings announcements, and various statements by highly-watched figures such as the FOMC Chairman.

8. VOLATILITY: EMPIRICS

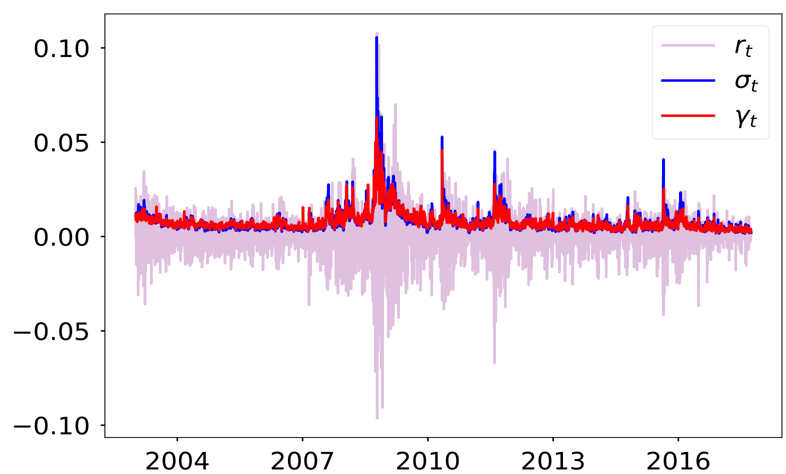
I separate this empirical part into four subsections. In the first section, I characterize the static properties of the volatilities. In the second, I characterize their dynamic properties. I show that both volatilities are highly persistent, displaying long-memory. To isolate the effect of the jump volatility and remove the effect of trends, I consider the jump proportion — $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$. This ratio is a measure for the proportion of the investors' new information driven by news. To get a better understanding of this jump proportion, I undertake a case study on what $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$'s dynamics in the most interesting subsample contained in my data — Fall 2008.

8.1. Statics

My results concerning the diffusion volatility are consistent, for the most part, with previous work on the topic. Since I am the first person to estimate the jump volatility, I cannot directly compare these results to previous estimates. Thankfully, I am able to show that the jump volatility is quite similar to the diffusion volatility, and so the intuition developed there can be carried over to the jump case.

I start by graphing the results in Figure 3 to give a high-level understanding of the volatilities' dynamics. As can be seen there, the volatilities are very closely related, their correlation coefficient equals 0.93. In addition, as one would expect from volatility measures, they both greatly increase during crises/recessions. The diffusion volatility does spike more than the overall volatility.

Figure 3: Root Volatilities



As mentioned in the introduction, various authors have argued that realized volatilities are approximately log-Gaussian, (Andersen, Bollerslev, Diebold, and Labys 2001; Andersen, Bollerslev, Diebold, and Ebens 2001). One might expect this to continue to be the case here. At a qualitative level, as can been in Figure 4, this is true. The log volatilities are more bell-shaped than the root-volatilities, but they are not log-Gaussian.[16]

16. The black lines on the graphs of the marginals are Gaussian densities fit to the data. They are there for comparison purposes only. They clearly do not fit well.

Figure 4: Log-Volatility Densities

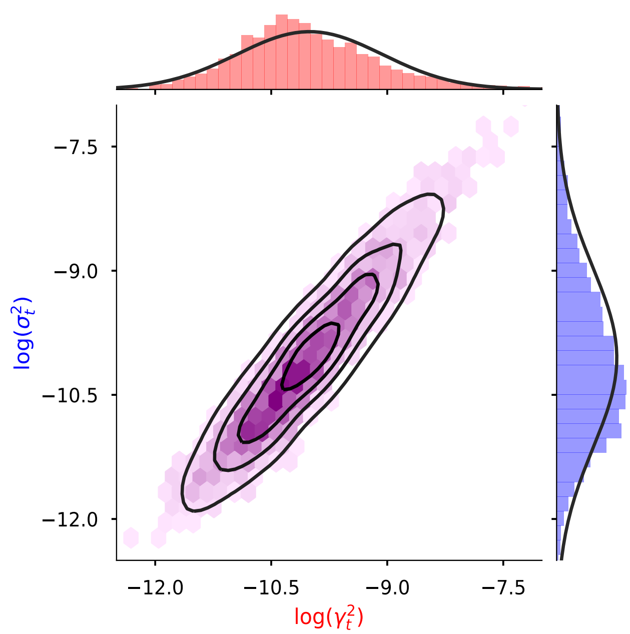


Figure 4 plots the two log-volatility distributions along with their joint distribution. As can be seen from the graph both marginal distributions are skewed right, and the joint distribution is just as skewed as the marginal densities. [17] It is worth noting that being skewed right means that the volatilities are more likely to take on abnormally large values than take on abnormally small ones. Abnormally high volatilities are associated with crises, and so the distributions are skewed in a direction that increases the investors' risk relative to an unskewed distribution. This is particularly noteworthy as these are distributions of log-volatilities, and taking the logarithm already removes a large amount of skewness.

We can further see from Figure 4 that the two volatilizes are very closely related. The joint distribution clusters very heavily around the 45-degree line. The correlation, for example, between the log-diffusion volatility and the log-jump volatility is 0.95.[18] To further understand the log volatility distributions, I report the first few standardized moments in the Table 3. As we can see, the two distributions are quite similar.

8.2. Dynamics

Having considered the data's static properties, I now consider the dynamic properties. I start with the univariate case. Throughout, I will focus on the log-volatilities because they behave in a more clean fashion as shown in Section 8.1, and so the true conditional expectations are likely closer to the approximations that I will do below. Here, I replicated standard stylized features for the diffusion volatility, and show that the jump volatility behaves similarly. I then turn to the

17. The only reason that the diffusion density might appear to be skewed left is because it is plotted sideways.
18. This correlation coefficient is not the same as the for the volatilities themselves because Pearson's correlation coefficient only measures the linear relationship.

Table 3: Volatility Summary Statistics

	σ_t^2	γ_t^2	$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$	$\log(\sigma_t^2)$	$\log(\gamma_t^2)$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$
Mean	9.98×10^{-5}	8.13×10^{-5}	0.51	−10.05	−10.00	−9.32	−0.68
Std. Dev.	3.21×10^{-4}	1.69×10^{-4}	0.08	1.09	0.94	1.00	0.17
Skewness	16.50	9.43	0.19	0.82	0.76	0.85	−0.41
Kurtosis	439.70	131.08	3.56	4.24	4.17	4.33	3.66
Min	2.34×10^{-6}	3.14×10^{-6}	0.21	−12.96	−12.67	−12.11	−0.41
Max	1.11×10^{-2}	3.99×10^{-3}	0.92	−4.50	−5.52	−4.19	−0.08

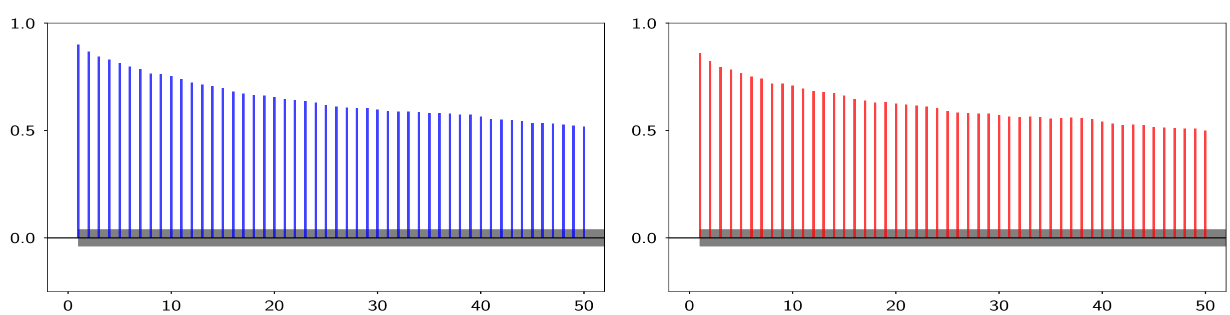
joint analysis, where the two volatilities no longer perform similarly. Only in the joint analysis by examining the variation driven by each of the volatilities independent of this component do we see do obvious differences.

8.2.1. Measuring the Persistence

I start by considering the autocorrelation of the two series. In Figure 5, we can see that the two series are both extremely persistent, hence the possibility of trends remarked on in the previous section.[19] Clearly, the dynamics are quite similar.

Figure 5: Autocorrelation Functions

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



Obviously, both series are extremely persistent. In fact, one might presume that the processes have a unit root. This is not the case. I test this results reporting the results in Table 4. The standard Augmented Dickey-Fuller test rejects at any reasonable level of significance (Dickey and Fuller 1981). Since it does not have unit root, one might think that it is short-memory process, i.e. have a geometrically decaying autocorrelation function. Perhaps less surprisingly, the KwiatkowskiPhillipsSchmidtShin (KPSS) test also rejects this hypothesis, (Kwiatkowski et al. 1992).

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

To those readers familiar with the empirical volatility literature, this result should not be surprising. The diffusive 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. In Table 4 I report estimates for the long-memory coefficient using the Geweke Porter-Hudak (GPH) estimator (Geweke and Porter-Hudak 1983). The smoothed periodogram estimator developed by Reisen (1994) gives essentially identical results.

Table 4: Persistence Statistics

	$\log \sigma_t^2$	$\log \gamma_t^2$	$\log \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$
		p-value	
ADF Test (Unit-Root Null)	0.02 %	0.01 %	0.00 %
KPSS Test (Short-Memory Null)	$\ll 1$ %	$\ll 1$ %	9.3 %
		statistic	
1st-Order Autocorrelation	0.91	0.88	0.24
Fractional Integration Coefficient (d)	0.61 (0.44, 0.77)	0.63 (0.45, 0.81)	0.42 (0.19, 0.65)

8.2.2. Univariate Dynamics

Turning now to the dynamics of the data, I run independent $AR(1)$ regressions on each of the two volatilities to gain some high-level understanding of the dynamics we get the following. Both series are quite persistent and predictable. However, we still have economically large changes from one day to another. In other words, the autocorrelation and innovation variance are both high.[20]

I now turn to considering univariate AR models for both series. I use Schwarz Information Criterion (SIC) to select the lag order.[21] This is not the ideal thing to do as it assumes away the long-memory I just demonstrated. However, it still can be 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 5 lags for $\log \sigma_t^2$ and 10 lags for $\log \gamma_t^2$. The two series are both quite predictable, with $\mathbb{R}^2$'s of approximately 80 %. These numbers are are likely higher than that found in the literature, which are often in the neighborhood of 40 % to 50 %, because I am able to do a good job at separating out the diffusion and jump volatilities, (Bollerslev,

20. 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 this should not affect my conclusion, but in some of the more border cases, it might be an issue.

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

Patton, and Quaedvlieg 2016, 8). Effectively, my regressand and regressors have less measurement error than is commonly used in the literature. In addition, volatility became more predictable during the Great Recession, which is a large portion of my sample.

The coefficient on the first lag in an AR(1) regression equals 0.87 for $\log \sigma_t^2$ and equals 0.86 for $\log \gamma_t^2$. In the SIC model, the first coefficient is smaller, 0.62 and 0.48, respectively, but the sum of all of the coefficients is similar. I report all of the coefficients and associated confidence intervals in Table 11, which is in the appendix. The innovation standard deviations are 0.49 for $\log \sigma_t^2$ and 0.42 for $\log \gamma_t^2$. To understand the previous numbers recall that these are log-volatilities, and so the innovation standard deviations are in log-deviations.

8.2.3. Joint Dynamics

Having considered the univariate dynamics, I now turn to the joint dynamics. The goal here is to understand how the two series relate in driving their joint dynamics. I start by considering whether the two volatility series Granger cause each other. I conclusively reject the null of no causality in either direction. The sum-of-squared residuals (SSR) test for $\log \sigma_t^2$ causing $\log \gamma_t^2$ with one lag returns a $\chi^2(df = 1)$ value of 148.77.[22] Conversely, the SSR test for $\log \gamma_t^2$ causing $\log \sigma_t^2$ with one lag returns a $\chi^2(df = 1)$ value of 111.34. These results are robust to the number of lags chosen or the specific version of the test I use, overwhelmingly rejecting 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, I consider a vector autoregression (VAR) specification. Here, the Schwarz Information Criterion (SIC) chooses 6 lags. I report these results in Table 12 which is in Section D. I report the results for a VAR(1) specification in Table 5 to aid in understanding these dynamics. The results for the more general specification are consistent with these results. The results are as one would expect from the Granger-causality results above. Both volatilities depend on the lags of both volatilities. The magnitudes of the coefficients for diffusion volatility are larger though.

Table 5: VAR(1) Results

	constant	$\log \sigma_{t-1}^2$	$\log \gamma_{t-1}^2$
$\log \sigma_t^2$	-0.66	0.64	0.30
	$(-0.84, -0.48)$	(0.59, 0.69)	(0.24, 0.31)
$\log \gamma_t^2$	-1.51	0.27	0.58
	$(-1.67, -1.35)$	(0.23, 0.35)	(0.53, 0.63)

If we look at the $\mathbb{R}^2$ — 79 % — or the innovation covariance matrix — $\begin{pmatrix} 0.25 & 0.18 \\ 0.18 & 0.20 \end{pmatrix}$ — we see

22. Since a $\chi^2(df = 1)$ is the distribution of $|N(0, 1)|^2$, this is equivalent to a t-static of 12.20.

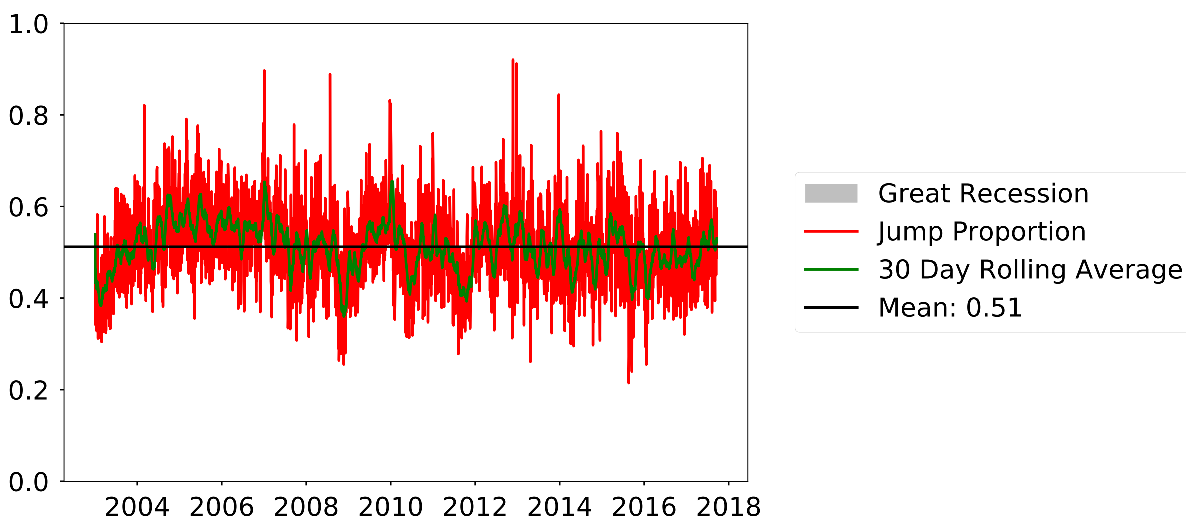
that they are similar in magnitude to the univariate case. The correlation between the innovations — 0.80 — is still quite high, but lower than the unconditional correlation. In other words, the common component drives a large of the amount of the variation in the innovations.

8.3. Jump Proportion

In Section 8.2.3 I showed that $\log(\gamma_t^2)$ and $\log(\sigma_t^2)$ share a persistent component that drives a large amount of the variation variation both over longer and shorter horizons. We would like to isolate the effect of the jumps and examine their dynamics directly. (This will be quite important when we consider the pricing implications.) To do this I define the jump proportion — $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$. I briefly alluded to this previously but did not investigate it in any detail.

To understand what $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ is, I start by plotting its variation over time in Figure 6. The mean of $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ and the Great Recession are plotted for reference. I plotted the rolling average to better visualize the series' low-frequency variation. Clearly, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ has substantial low- and high-frequency variation.

Figure 6: Time-Varying Jump Proportion



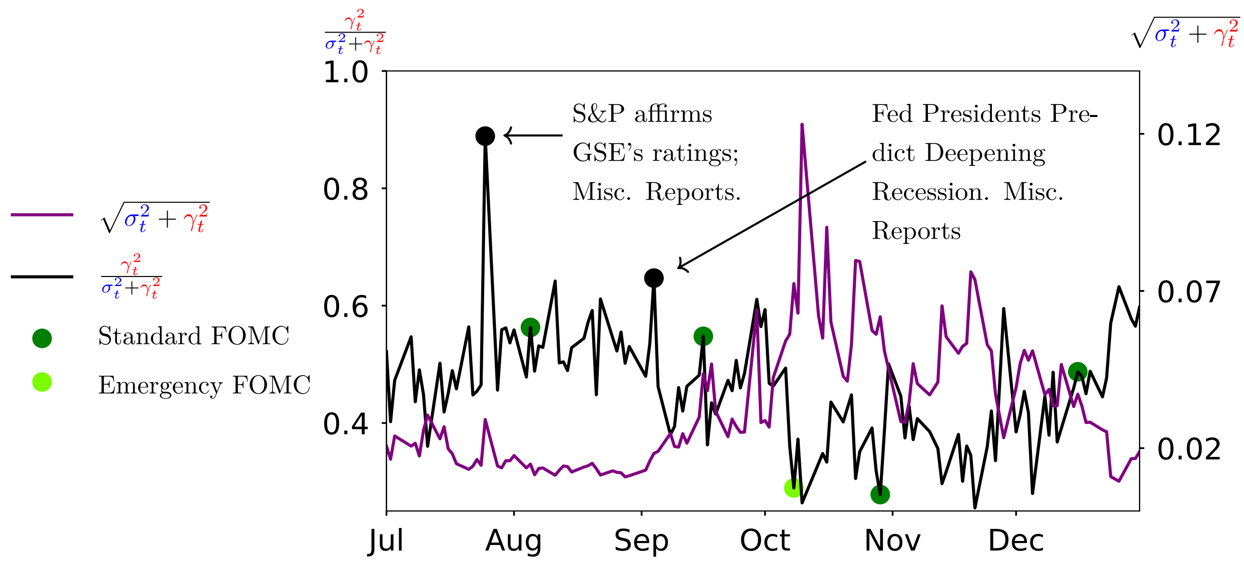
Since, I plot daily data Figure 6, and the dataset spans several years, the graph is at too low a resolution to understand exactly what is going on. Consequently, I consider the behavior of $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ during the most interesting subsample in my data — Fall 2008.

8.3.1. Jump Proportion: Fall 2008

They key to understanding the results in this section, is to remember that jumps are responses of price to news and the volatilities are instantaneous ex-ante measures of variability. Consequently, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ is a measure of the proportion price variance driven by the news. (It is a true proportion and lies in $[0, 1]$). Its denominator — $\sigma_t^2 + \gamma_t^2$ is the total instantaneous variability faced by the marginal investor. It is the natural volatility measure used in much of the literature. Its numerator isolates the proportion of that driven by news.

Without further ado, I consider the behavior of $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ in Fall 2008. In Figure 7, I plot $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ in black on the left axis and $\sigma_t^2+\gamma_t^2$ in purple on the right axis to provide a point of reference. I also label a few important dates. In particular, I label the days when $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ reaches its highest values and days on which the Federal Open Market Committee (FOMC) issued announcements.

Figure 7: $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ in Fall 2008



I now consider the behavior of $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ on days when it was extremely high. I did this by examining the news released on the days where $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ spikes. I discuss the sources I used in the data section. Importantly, these days were found by looking at when $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ spikes, not by looking at the news reports and backwards engineering the estimates. Hence, they provide a level of external verification that $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ genuinely measures news risk as the theory predicts.

$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ reached its highest point in the Fall of 2008 on July 25. In fact, this is the highest value $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ reaches in the entire sample except for a couple of days right around the end of the year, where it usually spikes. Several different reports were released on this day. On the financial side, S&P reaffirmed the Fannie Mae and Freddie Mac's AAA credit ratings but put them on a negative watch list; Wachovia slipped 7.60 % after a downgrade by a Morgan Keegan analyst, and Washington Mutual dove 13 % and then recovered most of its losses after confirming that it had boosted its liquidity to $50 billion. On the macroeconomic side, durable manufactured orders, housing sales, and consumer confidence all came in higher than expected. All in all, a large amount of news was released, particularly in sectors that would prove quite important in the ensuing financial crisis. Interestingly, during this time people were were worried a potential meltdown, but it had not happened yet. Lehman Brothers filed for bankruptcy about 6 weeks later on September 15.

Ex-post, the news on this day was moderately positive. The S&P 500 ended the day up 0.40 %. However, as discussed in the previous paragraph there were a large number of important announcements made on that day, and it was far from ex-ante obvious that they would be positive. This is precisely what we would expect a day with a very high proportion of news risk to look like.

The second highest day was September 4, where again the market was clearly worried about a potential meltdown. Various news reports released on this day include falling oil prices, mixed retail growth, and dour labor market readings. In addition, Dallas Fed president Richard Fisher and San Francisco Fed President Janet Yellen both discussed the current bad economic conditions and their expectation that conditions would get worse before they get better. Again, we have a day with a large amount of news risk.

As descended further into the Financial Crisis, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ actually declines. One might think this is surprising at first; most volatility measures such as realized volatility and the VIX increased as the crisis progress. At further inspection this makes sense. $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ fell because σ_t^2 increased faster than γ_t^2 did, not because γ_t^2 fell. In other words, non-news risk increased faster than news risk did. In the aftermath of the Lehman Brothers collapse, investors were likely more worried about what the previous negative news shocks meant and how they would play out, than they were about a new news shock. In other words, in the summer before we entered the financial death spiral, investors faced a huge amount of news driven risk, once we descended into it, the outworking of those shocks played a dominant role in their risk.

A large portion of the literature's discussion of news risk focuses on FOMC announcements, (Andersen, Bollerslev, Diebold, and Vega 2003, 2007; Ai and Bansal 2018). Consequently, I now consider how $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ behaves on the days when FOMC makes its announcements. The green dots on Figure 7 are the meetings of the FOMC. On three of them $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ reaches a local maximum.

On the other hand, one might be surprised that on two of the days, including the Emergency FOMC meeting, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ is actually quite low. However, it is well-known that the predictability of FOMC announcements varies over time. In fact, current market reports at the time point out that the Fed's actions were expected in both of those instances. In addition, $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ measures the proportion of risk driven by news and so will also be low if non-news risk is particularly high. During the two FOMC days when $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ was low, total volatility was quite high, as can be seen when looking at $\sigma_t^2 + \gamma_t^2$.

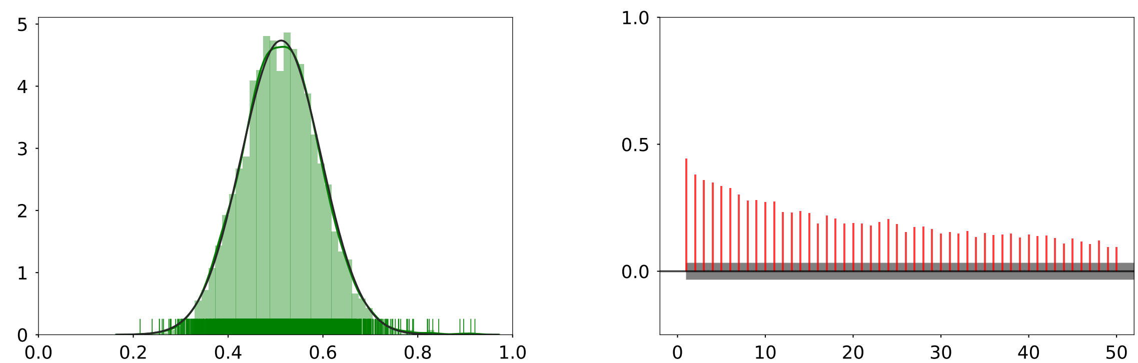
In addition, the VIX was quite high during this period, setting several records, which have yet to be overturned. One day it did so was October 10, 2008 when it hit 77 at one point. This is the day identified with the purple dot in Figure 7, and is the day during my sample when $\sigma_t^2 + \gamma_t^2$ peaked. This day, a Friday, rounded out the worst week on the S&P 500 since the Wall Street Crash of 1929.

8.3.2. Jump Proportion: Stylized Features

Now that I have shown that $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ is economically interesting and measures the proportion of risk driven by news, we can turn to understanding its stylized features. I start by plotting its histogram in Figure 8a. The green line is a kernel density estimate, and the back line is a Gaussian distribution fit to the data. As we can see the density is roughly Gaussian.[23] $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$'s mean is 63 %.

In other words, over one-half of the volatility is driven by jumps. This may sound high to those not familiar with the literature, but is consistent with the literature, as discussed in Section 2. It also varies quite a bit. Its standard deviation is 10 %, and it goes as low as 30 % and as high as 91 %.

Figure 8: $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$



(a) Density

(b) Autocorrelation Function

It is also very persistent as can seen in Figure 8b. I also showed this in Table 4, where I showed that the log $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ had long-memory with a fractional integration coefficient $d = 0.42$.

$\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ tends to be high on days when there a significant amount of new information in the market as discussed in Section 8.3.1, as the theory predicts. Perhaps more surprisingly, it also is high on the last few trading days of the year. Whether this is real, in that the reshuffling of portfolios as investors who trade rarely causes discontinuous changes in the representative investor's information set or spurious because of different market microstructure effects on these days is an interesting question but outside the scope of this paper.

9. News Premia: Theory

9.1. Characterizing Risk with Recursive Utility and Jumps

Recalling the discussion in the introduction, discontinuous prices or information flows break the derivation of CAPM-style results where risk-premia are covariances with marginal utility. In particular, risk premia are no longer proportional to the integrated diffusive co-volatility between prices and stochastic discount factors. In Ai and Bansal's (2018) world, we have an announcement SDF (A-SDF) whose covariation is also priced, while in Tsai and Wachter's (2018) world, it is the covariance during extreme events that matters. Equivalently, the investors demans compensation for

23. The histogram is scaled so the densities integrate to 1.

Poisson risk, (Tsai and Wachter 2018, 25). The level of compensation they require is not solely determined by the curvature of their utility functions, i.e. their risk aversion, but also by the curvature of their recursive preference aggregator.

We further know from Theorem 1 that these two models have the same structure. If investors have recursive utilities and are subject to discontinuous information flows, they price jumps variation dfferently than diffusive variation. To be as explicit as possible, Tsai and Wachter (2018) think about jumps risks as jumps in consumption and view them as extreme events, while I follow Ai and Bansal (2018) and assume that consumption is continuous. Consequently, jumps in investor's value functions are solely driven by discontinuous information flows. In that sense, the results are different, but their relationship to discontinuities in the value functions is the same.

Given that we live in a world where prices jump all the time, and preferences are likely not time-separable, nonparametrically identifying how this curvature interacts with price dynamics would provide a benchmark to discipline our models. However, this is known to be quite difficult. Even separating the coefficient of relative risk aversion and the intertemporal elasticity of substitution is quite hard, and this is just one special case of identifying the CEF's curvature. As another example, only one paper claims to measure ambiguity attitudes using revealed preference arguments on real data (Baillon et al. 2018). Again, this is only one special case of the problem at hand, as ambiguity aversion is just one way the CEF exhibits curvature.

In the remainder of this section, I generalize Ai and Bansal's (2018) results and derive the risk premia as a function of the jump and diffusion volatilities and the curvature of the agent's value function and CEF. Having done this, we can easily see that $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$'s risk-premia in the data identifies this curvature.

To do this, I need to introduce some notation. I adopt the setup that Ai and Bansal (2018) use. In particular, I follow them and consider intertemporal preferences that can be represented as in Strzalecki (2013). Let $V(t)$ be the representative investor's value function at time t, and $u(\cdot)$ be the associated utility function for current period consumption $c(t)$. I denote the CEF by $\mathcal{I}\left[\cdot \,\middle|\, \mathcal{F}_{t-}\right]$ and let κ denote the rate of time-preference. I assume that κ is constant for notation convenience, but this can be easily generalized.

Definition 16 (Certainty Equivalence Functional)**.**

$$dV(t) = u\left(c\left(t\right)\right)dt + d\mathcal{I}\left[\int_{s\geq t} \exp\left(-\kappa\left(s-t\right)\right)u(c(s))\,ds \,\middle|\, \mathcal{F}_{t-}\right] \tag{44}$$

I make the filtration $\mathcal{F}_{t-}$ explicit to emphasize the fact that certainty equivalence functionals maps information sets to utility. For example, if preferences are time-separable $\mathcal{I}$ is just the expectations operator $\mathbb{E}\left[\cdot \,\middle|\, \mathcal{F}_{t-}\right]$.

I focus on CEF's with the following form for some strictly increasing function ϕ. These include, for example, the second-order expected utility of Ergin and Gul (2009) and the recursive preferences of Kreps and Porteus (1978) and Epstein and Zin (1989).

$$\mathcal{I}\left[V(t) \,|\, \mathcal{F}_t\right] = \phi\left(\mathbb{E}\left[\phi^{-1}\left(V(t)\right) \,\middle|\, \mathcal{F}_{t-}\right]\right) \tag{45}$$

I focus on this class of preferences because it covers many of the leading classes of preferences and provides a clean characterization of they way that risk-aversion and the CEF's curvature interact in driving risk premia. The curvature of ϕ controls the level of news premia. If we have time-separable preferences, than ϕ is the identity function and has no curvature. Conversely, if we are in a purely diffusive world, $V(t)$ is a predictable function, and so we can pull ϕ through the expectation and ϕ cancels with its inverse. Again, we have no additional risk premia. However, if V jumps, it is not predictable, and we cannot do this.

9.2. Deriving the Asset Pricing Equation

9.2.1. The Investor's Portfolio Optimization Problem

In this section, I derive a formula for risk-premia when investors have recursive utility and information jumps. Prices are still semimartingales, and so prices are still expectations with respect to some risk-neutral measure, (Delbaen and Schachermayer 1994). In other words, we have some pricing kernel $\mathcal{M}$ that is as a change of measure between the physical and risk-neutral distributions for the discounted prices $\widetilde{P}(t)$. In other words, Equation (46) holds for all stopping times $\tau > t$ and prices $\widetilde{P}(t)$.[24]

$$\widetilde{P}(t) = \mathbb{E}\left[\mathcal{M}(\tau)\widetilde{P}(\tau) \,\middle|\, \mathcal{F}_t\right] \tag{46}$$

However, $\mathcal{M}(\tau)$ is no longer proportional to $DV(\tau)$. Consequently, we must figure out what it actually is. To do this, we need to place some more structure on the environment.

Assumption. Assumptions on the Market Environment

1. Both u and $\mathcal{I}$ are Lipschitz continuous, Fréchet differentiable with Lipschitz derivatives.
2. $u : \mathbb{R} \to \mathbb{R}$ has strictly positive first-order derivatives and $\mathcal{I}$ is increasing with respect to first-order stochastic dominance.
3. The consumption process $c(t)$ is continuous.
4. There is a representative investor who prices all assets.

The first assumption is a regularity condition justifying solving the investor's maximization problem using derivatives. The second is a restriction that implies the preferences are reasonable. The third implies that all of the discontinuities in the environment are related to information flows or changes in wealth. It is relatively innocuous as I do not assume the underlying technology or production processes are continuous, and we can mimic any discrete-time process exactly. It would be implied by a friction introducing a lag between the investor's decision to consume and the actual

24. Recall, that I use $P(t)$ to refer to the price and $p(t)$ to refer to the log-price.

consumption, because this will make consumption predictable, and equivalently continuous. The fourth 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 it to simplify the exposition.

Consider a representative investor who is facing a portfolio optimization problem. Her preferences are given by Definition 16. She has access to a vector of assets $\Xi(t) := \xi_1(t), \ldots$. I assume for simplicity that she has no other sources of income. Over some small length Δ, we can represent the investor's problem as follows. She enters into the period with asset allocation $\Xi(t-\Delta)$, and prices are $P(t)$.[25] She need to solve for consumption $c(t)$ and asset allocation $\Xi(t)$. I do everything cum-dividend to avoid introducing even more notation, the extension to the ex-dividend case is straightforward.[26]

Problem 1. Consumer's Portfolio Allocation

$$V\left(\Xi(t-\Delta), P(t)\right) = \max_{c(t),\, \Xi(t)} \int_t^{t+\Delta} u(c(s))\, ds + \mathcal{I}\left[\exp(-\kappa\Delta) V\left(\Xi(t), P(t+\Delta)\right) \,|\, \mathcal{F}_t\right] \tag{47}$$

$$c(t) + \sum_i P_i(t)\xi_i(t-\Delta) = \sum_i P_i(t)\xi_i(t) \tag{48}$$

The continuous-time problem is the limit of Problem 1 as $\Delta \to 0$. 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.[27]

Theorem 9 (Asset-Pricing Equation)**.** *Let the assumptions in Assumption 5 hold, prices 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 prices satisfy the following asset pricing equation for all stopping times $\tau > t$, where $M(t) := \frac{DV(t)}{\mathbb{E}[DV(t) \,|\, \mathcal{F}_{t-}]}$ and $M_*(t) = \frac{D\mathcal{I}(V(t))}{\mathbb{E}[D\mathcal{I}(V(t)) \,|\, \mathcal{F}_{t-}]}$.*

$$\widetilde{P}(t) = \mathbb{E}\left[M(\tau)M_*(\tau)\widetilde{P}(\tau) \,\middle|\, \mathcal{F}_{t-}\right] \tag{49}$$

Conceptually, Theorem 9 is straightforward. The pricing kernel $\mathcal{M}$ has two parts: $M(t)$, which is a normalized derivative of the value function, and $M_*(t)$ which is a normalized derivative of the certainty equivalence functional. If the CEF is the expectations operator, then $M_*(t)$ is identically one. Consequently, jump variation is priced identically to diffusive variation. Even though Theorem 9 is straightforward it is hard to understand as it is very abstract. Consequently, we make a few assumptions that let us derive more concrete expressions. It will also give a clean characterization of when news command positive (or negative) risk premia.

25. The timing notation may seem somewhat strange here because I am maintaining the convention where time arguments denote when the objects first enter the representative investor's information set.

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

27. This is essentially the headline result in Ai and Bansal (2018, Theorem 2).

9.2.2. Example: Epstein-Zin Preferences

I start by considering what $M_*(t)$ is in the most commonly used form of recursive preferences — Epstein-Zin. These preferences have been quite popular ever since Bansal and Yaron (2004) showed you can use them to help resolve the equity premium puzzle. I adopt the notation Bansal and Yaron (2004) use and let ρ refer risk aversion and ψ refer to the intertemporal elasticity of substitution (IES). You can represent Epstein-Zin preferences over an length of time Δ as follows for a value function U.

Definition 17 (Epstein-Zin Utility)**.**

$$U_t = \left[c_t^{1-1/\psi} + \exp(-\kappa\Delta)\mathbb{E}\left[U_{t+\Delta}^{1-\rho} \,\middle|\, \mathcal{F}_t \right]^{\frac{1-1/\psi}{1-\rho}} \right]^{\frac{1}{1-1/\psi}} \tag{50}$$

This formulation of Epstein-Zin utility is not in the form given in Definition 16, and so is not particularly useful for our purposes. Define $V_t \coloneqq U_t^{1-1/\psi}$. Then we can reparameterize Equation (50) as follows.

$$V_t = \left[c_t^{1-1/\psi} + \exp(-\kappa\Delta)\mathbb{E}\left[V_{t+\Delta}^{\frac{1-\rho}{1-1/\psi}} \,\middle|\, \mathcal{F}_t \right]^{\frac{1-1/\psi}{1-\rho}} \right] \tag{51}$$

In fact, not only is it in the form of Definition 16, but also it is in the form of Equation (45) if we define $\phi(V) \coloneqq \frac{1-\rho}{1-1/\psi} V^{\frac{1-\rho}{1-1/\psi}}$,[28]

$$V_t = \left[c_t^{1-1/\psi} + \exp(-\kappa\Delta)\phi^{-1}\left(\mathbb{E}\left[\phi\left(V_{t+\Delta}\right) \mid \mathcal{F}_t \right] \right) \right] \tag{52}$$

From Theorem 9, we know that the price today will be high if $M_*(\tau)$ and $P(\tau)$ co-move, i.e. are conditionally positively correlated. If asset i has positive price, $P_i(t)$ will co-move positively with $V(t)$. Consequently, jump variation will command a higher (lower) premium if and only if $M_*(t)$ positively (negatively) co-moves with future utility V_τ. We do not need to worry about co-movement between $M(t)$ and $M_*(t)$ because $M(t)$ is continuous and $M_*(t)$ is purely discontinuous, and hence they are orthogonal.

$M_*(t)$ is a re-normalized derivative of $\mathcal{I}$.[29] That is it equals the following where D is the Fréchet derivative operator.

$$M_*(t) = \frac{\mathcal{I}(DV(t))}{\mathbb{E}\left[D\mathcal{I}(V(t)) \mid \mathcal{F}_{t-} \right]} \tag{53}$$

If the CEF has the form given in Equation (45), this derivative is proportional to $\phi'(V(\tau))$. To see the intuition behind this consider the following. We can expand the derivative of $\mathcal{I}$ as a function of $V(t)$ using the chain rule.

28. The constant in front cancels between ϕ and ϕ^{-1}, and so does not affect the level of utility. It it is there to ensures that ϕ is always an increasing function.

29. This is how I define $M_*(t)$ in the proof of Theorem 9.

$$\mathcal{D}\phi^{-1}\left(\mathbb{E}\left[\phi(V(\tau))\,|\,\mathcal{F}_t\right]\right) = \frac{1}{\phi'\left(\mathbb{E}\left[\phi(V(\tau))\,|\,\mathcal{F}_{t-}\right]\right)}\mathbb{E}\left[\phi'(V(\tau))\,|\,\mathcal{F}_{t-}\right] \propto \mathbb{E}\left[\phi'(V(\tau))\,|\,\mathcal{F}_{t-}\right] \tag{54}$$

Hence, we have a positive risk premia if $\phi'(V)$ is an increasing function, and a negative risk premia if it is a decreasing function. Equivalently, we have a positive (negative) premia if ϕ is concave (convex).

In the Epstein-Zin world, we know ϕ. Consequently, we can compute the sign of the premium as a function of the parameters. In particular, we have a positive premium if and only if $\rho > \frac{1}{\psi}$ — the risk-aversion is greater than the inverse of the intertemporal elasticity of substitution. That is investors prefer early resolution of uncertainty. Conversely, we have a negative premium if and only if investors prefer late resolution of uncertainty.

9.3. Deriving Risk Premia

The end of the previous section is essentially where Ai and Bansal (2018) stop. I, however, want to nonparametrically identify news risk. To do this, I characterize risk premia in the environment. This is useful because we can, at least in principle, estimate risk premia, and hence we can use them to identify the curvature of the relevant functions. The question facing us is how do we derive risk premia from Theorem 9.

9.3.1. An Itô's Lemma for the Expectation of a Square-Integrable Semimartingale

If prices were continuous, we could use Itô's lemma to solve for the expected log-return in terms of the covariance between the $M(t)$ and $P(t)$. However, the standard formulation of Itô's lemma where the convexity correction is proportional to the quadratic variation does not hold in this environment. There are generalized Itô's lemmas that work for more general semimartingales, but they no longer take the form of a second-derivative multiplied by a quadratic variation.

It turns out if we use the predictable quadratic variation instead of the quadratic variation, similar to how I constructed the representation for the prices themselves in Section 4, we can derive a version of Itô's lemma that has the standard form but applies to jump processes.

Lemma 10 (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 the following.*

$$d\mathbb{E}\left[f(\widetilde{Z})\,\middle|\,\mathcal{F}_{t-}\right] = \mathbb{E}\left[f'(\tilde{Z}(t-))\,d\tilde{Z}(t)\,\middle|\,\mathcal{F}_{t-}\right] + \frac{1}{2}f''(\tilde{Z}(t-))\,d\langle\tilde{Z}\rangle(t) \tag{55}$$

The above equality's assumptions and conclusion are both weaker than Itô's lemma. We do not need continuous processes, but then then the equality only holds in expectation. However, this is sufficient for our purposes as risk premia are expectations.

9.3.2. Risk Premia are Proportional to Predictable Quadratic Variations

Having derived asset pricing equation, I now compute risk premia using the Itô's lemma in Lemma 10.

Theorem 11 (Asset-Pricing Equation)**.** *Let the assumptions in Assumption 5 hold, prices 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 *are as follows.*

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_{rf}(t)}{P_{rf}(t-)}\,\middle|\,\mathcal{F}_{t-}\right] = -\,d\langle m, p\rangle(t) - d\langle m_*, p^J\rangle(t) \tag{56}$$

The first term in Equation (56) should be completely familiar to my readers. The discounted risk-premia is a negative covariance between the log-SDF and the price. For example, consider a stochastic volatility diffusion. Now the second term is identically zero, and the first term can be factored as $-\sigma_m(t)\sigma_{p_i}$. Here we have a risk price: $-\sigma_m(t)$ and a risk-quantity σ_{p_i}. This is Merton's (1973) result that justifies factor-based asset pricing models for the cross-section. In my more general environments, this term includes both covariation between the diffusive and jump components. We can split it up as $-\,d\langle m, p\rangle(t) = -\,d\langle m^D, p^D\rangle(t) - d\langle m^J, p^J\rangle(t)$. The risk-premia is proportional to the total covariation. This is a generalization of Tsai and Wachter (2018, Corollary 6), which shows that if preferences are time-separable, then what they call a rare-event consumption CAPM holds. Risk-premia are determined by covariance with consumption and covariance with marginal utilities when prices jump, (10).

The second term is more unusual. It reflects compensation for bearing news risk, which is why it only affects the jump part of the prices. Some basic intuition is as follows. We would expect the risk-premia to be a covariation between the normalized derivative of the CEF — $M_*(t)$ — and the price from Theorem 9. Since $M_*(t)$ only arises from non-predictable (jump) changes in $V(t)$, investors only care about covariation between it and p_i^J. By using Lemma 10 I show that this additional term is a local covariance between m_* and p_i^J, just like the standard risk-aversion based risk premium is. Consequently, we can still talk about risk-prices and quantities of risk.

To sum up, the jump part requires compensation for two sources of risk. It requires compensation for the curvature of the value function, $M(t)$, which shows up in the first term. It also requires compensation for curvature of the CEF. This shows up in the second term, $M_*(t)$.

Equation (56) expands upon the literature in two main ways. First, it shows that the risk-premia for jumps as in Tsai and Wachter (2018) and Ai and Bansal (2018) arise from the same fundamental property of asset pricing. News and jump risk are the same thing in this environment, and so they show up in the same place in investor's risk premia. Second, it provides a fully nonparametric relationship between investors' preferences and news premia. The characterization in Tsai and Wachter (2018) only holds under some parametric assumptions. The central result of Ai and Bansal (2018) — Theorem 2 — is fully nonparametric. However, it does not quantify the precise relationship between risk premia and preferences. It only determines the sign of the relationship, not the level.

10. NEWS PREMIA: EMPIRICS

In this section, I disentangle the news risk from standard risk premia. To avoid adding complexity to this admittedly already complex paper, I do not set down a fully specified model for the various stochastic discount factors. Instead, I make a CAPM-style approximation where I assume that all of the terms are proportional to market wealth.

Essentially, I am assuming that the only ICAPM factor is wealth and the log-derivative of the CEF is proportional to the log-value function. This is admittedly a severe simplification. To gain a full understanding of the relationship between news risk and investors' preferences, you would need to specify a full general equilibrium model. However, this special case is the logical place to start and has a vast history in economics and finance, (Brandt and Kang 2004; Bollerslev, Tauchen, and Zhou 2009; Lettau and Ludvigson 2010; Drechsler and Yaron 2011). Trying to understand the relationship between volatility and risk premia is an important place to start, and so that is what I do.

10.1. Excess Return and Volatility: Contemporaneous Relationship

I start by considering the contemporaneous relationship between the volatility and the return. In this section, I replicate the standard results that volatility and returns are contemporaneously negatively correlated. They key difference between these results and those in the literature, is that I split this relationship up into a total volatility — $\sigma_t^2 + \gamma_t^2$ — and a jump proportion — $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$. The reason I do this is because, as discussed in Section 9.3.2, standard intuition in terms of the relationship between returns and volatility is about the relationship between the total volatility and the returns, i.e. the first part of the decomposition in Theorem 11.

My analysis below uses the daily excess return — rx_t, again to make the results more easily comparable with those in the literature. I construct rx_t by taking r_t and subtracting the log yield on the 10 year treasury bill, which I obtain from FRED. I annualize rx_t (multiplied it by 252) to make the results more interpretable. I also use logarithmic transformation of the volatilities because they are closer to Gaussian as discussed in Section 8.1.

In particular, I run a series of regressions of the following form, using OLS. 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).

$$rx_t = \beta_0 + \beta_1 \mathbf{1}\{\text{FOMC}\}_t + \beta_2 \log\left(\sigma_t^2 + \gamma_t^2\right) + \beta_3 \log\left(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}\right) + \epsilon_t \tag{57}$$

As can clearly be seen in Table 6, $\log \sigma_t^2 + \gamma_t^2$ and rx_t are strongly negatively correlated. This is what we expected given the discussion in the literature. I have a new result where $\log \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ and rx_t are strongly positively contemporaneously correlated. These results hold at most reasonable significant levels and are unaffected by the inclusion of FOMC days and interaction terms. I also report a weighted least squares regression where I correct for heteroskedasticity by weighting the

Table 6: $\mathbb{E}\left[rx_t \,\middle|\, \sigma_t^2+\gamma_t^2, \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}, \mathbf{1}\{\mathrm{FOMC}\}_t\right]$ (OLS)

Constant	$\mathbf{1}\{\mathrm{FOMC}\}_t$	$\log\left(\sigma_t^2+\gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\log\left(\sigma_t^2+\gamma_t^2\right)\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\bar{\mathbb{R}}^2$
0.01	0.98				0.6 %
[0.15]	[3.64]				
−3.76		−0.41			2.4 %
[−5.40]		[−5.70]			
1.20			1.67		1.2 %
[6.05]			[5.51]		
−2.72		−0.35	0.74		2.6 %
[−3.11]		[−4.36]	[2.29]		
−1.38		−0.21	2.48	0.19	2.6 %
[−0.48]		[−0.70]	[0.67]	[0.50]	
−2.86	1.18	−0.37	0.80		3.3 %
[−3.30]	[4.36]	[−4.46]	[2.45]		
−1.18	1.17	−0.25	2.22	0.16	3.3 %
[−0.55]	[4.27]	[−0.75]	[0.51]	[0.34]	

datapoints by the inverse of the total volatility on each day as a heteroskedasticity correction in Section E. We know that weighted least square regressions are more efficient. $\sigma_t^2+\gamma_t^2$ is not quite the innovation variance because this is a contemporaneous regression but it should be substantially correlated with it, and so I continue to use robust standard errors. In this case, the results do not noticeably differ. The heteroskedasticity adjustment will be important in the measurement of risk premia.

It is also worth noting that I am running a contemporaneous regression, and so the $\bar{\mathbb{R}}^2$'s that I report in Table 6 are reasonable. We can explain a measurable amount of the variation in the data, but even using contemporaneous data we can only explain a small part of the variation in the excess market return.

10.2. Excess Return and Volatility: Risk Premia

10.2.1. Separating the Predictable and Contemporaneous Variation

In the discussion above, both volatilities and returns occur contemporaneously. An investor cannot trade on contemporaneous information. In principle, since the jump proportion is a known function of predictable variables $\sigma^2(t)$ and $\gamma^2(t)$ it is tradable in continuous-time. However, because of estimation error in the estimates above and estimation error in implementing this strategy in practice, this is likely not feasible in practice.

Consequently, we want to isolate the predictable variation in the volatilities and estimate our risk premia as functions of this variation. To make the problem more feasible and easily comparable with the previous literature, I do the implied discrete-time regressions instead of the continuous time ones. Intuitively, we want to regress returns on expected volatilities.[30]

How can we do this? The literature commonly proxies for these expectations by using the lagged regressors, i.e. replace $\sigma_t^2 + \gamma_t^2$ and $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ with $\sigma_{t-1}^2 + \gamma_{t-1}^2$ and $\frac{\gamma_{t-1}^2}{\sigma_{t-1}^2+\gamma_{t-1}^2}$. However, since we know that these volatilities are not AR(1) processes, we introduce a great deal of unnecessary noise when we do this. In addition, the coefficient estimates calculated from using the lagged variables themselves as regressors will combine the true relationship between expected returns and volatilities and the autoregressive parameters for the volatility processes.

What we want to do is approximately span the part of $\mathcal{F}_{t-1}$ relevant for predicting $\sigma_t^2 + \gamma_t^2$ and $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$. In other words, we want to use instruments that form a good predictive system for the volatilities. I constructed a reasonable predictive system in a companion paper, (Sangrey 2018). To put it in the standard instrumental variable language a good instrumental variable is independent of the ϵ_t in Equation (57) but relevant for predicting the regressors. Those two conditions are precisely the conditions that define a good forecasting variable.

10.2.2. News Premia Estimates

In what follows I use an approximate heterogeneous autoregressive (HAR) specification for the 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. To make my results easily comparable with the previous literature I also use some pre-scheduled announcements as instruments. However, the additional $\bar{\mathbb{R}}^2$ provided by the announcements is very small and so including the additional instruments does not significantly affect the estimates. (The instruments are not weak. Their $\bar{\mathbb{R}}^2$'s are 20 % to 30 % for $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ and 75 % to 80 % for $\sigma_t^2 + \gamma_t^2$. See Table 14)

Since the signal-to-noise ratio is rather small here, to get good estimates of the basic underlying parameters we must be as efficient as possible. In highly heteroskedastic environment such as this one, adjusting for heteroskedasticity appropriately significantly improves estimator's efficiency. Another key benefit of using $\sigma_t^2 + \gamma_t^2$ and $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ as regressors directly instead of proxying for them is that the correct heteroskedasticity adjustment is $\mathbb{V}\mathrm{ar}\left(rx_t \,|\, \mathcal{F}_t\right)^{-1}$. This is essentially $\sigma_t^2 + \gamma_t^2$. If μ_t has time-varying volatility this is not quite correct. However, as the proportion of the rx_t's time-varying volatility driven by μ_t's time-varying volatility is almost certainly quite small, this will likely not affect the resultant estimates significantly. To account for this potential error, I continue to use heteroskedasticity and autocorrelation robust standard errors (Newey and West 1994).

Again I run a regression in the form in Equation (58).

$$rx_t = \beta_0 + \beta_1 \mathbf{1}\{\text{FOMC}\}_t + \beta_2 \log\left(\sigma_t^2 + \gamma_t^2\right) + \beta_3 \log\left(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}\right) + \epsilon_t \tag{58}$$

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

I present the basic estimates conceding the volatility and news premia in Section 10.2.2. I consider several other specification in Section E. Again, the volatility coefficients are robust with respect to heteroskedasticity correction and to the particular instruments chosen. If you run the regression over a subsample, the results either agree or are not statistically significant.

Table 7: News Premia Estimates

Regressors			
Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log \sigma_t^2 + \gamma_t^2$	$\log \frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2}$
0.29	0.31		
[10.14]	[1.92]		
3.22		0.29	
[6.40]		[5.86]	
-1.61			-2.96
[-5.31]			[-6.34]
0.81		0.19	-2.16
[1.10]		[3.37]	[-4.48]
0.83	0.09	0.19	-2.15
[1.14]	[0.51]	[3.43]	[-4.46]

Since I am regressing 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 $\log \sigma_t^2 + \gamma_t^2$ is 0.31. 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 0.31 %. [31] For comparison, the average year-to-year difference in average $\sigma_t^2 + \gamma_t^2$ in my sample is $\approx 80\,\%$. It increased by $\approx 160\,\%$ between 2007 and 2008.

The average annual absolute difference in $\sigma_t^2 + \gamma_t^2$ is only 10.30 %, 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 $-2.15\,\%$. Again, when you combine these two numbers, you get very large movements in risk premia.

However, there is something quite surprising about the sign of $\log(\frac{\gamma_t^2}{\sigma_t^2 + \gamma_t^2})$. It is negative. As shown in Section 9, this implies that the CEF is quasiconvex. As discussed in Ai and Bansal (2018) this is contrary to many economic models such as long-run risks model and models with time-separable utility imply that this sign should be either positive or zero, respectively. One notable exception that does imply a negative additional news premium are internal habits models, (Constantinides 1990; Boldrin, Christiano, and Fisher 2001).

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

10.2.3. Robustness Checks

Estimating risk premia is difficult because the signal-to-noise ratio is quite small. The literature has pointed out a number of issues that can bias the empirical estimates. Perhaps the most important is the Stambaugh bias, (Stambaugh 1999). He shows in cases where we have stochastic regressors, a finite-sample bias can inflate coefficient estimates even if we have a few hundred datapoints. However, since I am running daily regressions and not monthly, as is commonly done, and hence have approximately 3700 datapoints and this bias decreases at a $\frac{1}{\text{\# datapoints}}$ rate, this source of bias should not noticeable affect my estimates.

The other important sources of bias noted in the literature are also not near as much of an issue in my case because I am using daily data. For example, if you regress long-horizon returns on persistent regressors, the $\mathbb{R}^2$ 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 leading to various issues as well. Again I am not using overlapping returns, and so this also does not apply.

There is one main source of error that is worth pointing out. The regressors that I am using are estimated from high-frequency data. Consequently, we may have a error-in-regressors problem. This should not be a significant issue for three reasons. First, since I have a great deal of intraday data, they should be estimated rather 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 is does entirely on the magnitude of the high-frequency returns, not their sign. In addition, it does not even apply to estimating $\sigma_t^2 + \gamma_t^2$. Consequently, the bias would shrink the coefficient estimates towards zero. Third, and most importantly, since I am instrumenting for the returns by their lags and the estimation error is likely independent across time, both the coefficient estimates and their standard errors should be asymptotically valid.

10.3. FOMC (Announcement) Premia

10.3.1. GMM versus OLS Decomposition

The third interesting finding concerns returns on FOMC dates. In the last row of Section 10.2.2, I report an estimate for the excess returns on FOMC days. Unlike previous results, this coefficient is not statistically significant at any reasonable significance level. In addition, Ai and Bansal (2018) uses this high average return to identify the sign of the curvature of the CEF. Essentially, they use $\mathbf{1}\{\text{FOMC}\}_t$ as a proxy for $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$. The negative coefficients that I find for $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ leads to different conclusion regarding the CEF's curvature. In the previous section, I let these results and the tension the pose with the literature pass by unremarked, I turn to analyzing this now.

Similar to Lucca and Moench (2015) I find that days on which the FOMC makes an announcement have surprisingly high returns, as we can see in Table 6. [32] The question at hand is why do the coefficients in Section 10.2.2 and Table 6 differ?

32. Regressing on a constant is numerically equivalent to taking a sample mean.

I focus on the difference between the last rows of these two tables.

Table 8: $\mathbb{E}\left[rx_t \,\middle|\, \sigma_t^2 + \gamma_t^2, \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}, \mathbf{1}\{\text{FOMC}\}_t\right]$ (WLS)

	Constant	$\mathbf{1}\{\text{FOMC}\}_t$	$\log\left(\sigma_t^2 + \gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$
OLS	−2.86	1.18	−0.37	0.80
	[−3.30]	[4.36]	[−4.46]	[2.45]
	−3.97	1.15	−0.43	
	[−5.20]	[4.27]	[−5.40]	
WLS	−1.76	0.44	−0.24	0.53
	[−3.22]	[2.62]	[−4.86]	[2.47]
	−2.38	0.42	−0.27	
	−5.12	2.49	−5.68	
Unweighted-IV	−1.59	0.88	−0.06	−1.59
	[−1.34]	[3.12]	[−0.62]	[−2.53]
	0.52	0.92	0.05	
	0.64	3.35	0.64	
GMM	0.83	0.09	0.19	−2.15
	[1.14]	[0.51]	[3.43]	[−4.46]
	3.23	0.18	0.30	
	[6.43]	[1.10]	[5.92]	

As can be seen in Table 8, both performing the appropriate reweighing, and conditioning on the volatility, reduces the effect of the FOMC days. We need both of them to fully absorb the effect. You do not, interestingly, need to condition on $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$, $\sigma_t^2 + \gamma_t^2$ is sufficient. As can be seen in Section D, the particular way which we do this is not particularly relevant.

10.3.2. Issues with FOMC Day Identification

Readers familiar with Lucca and Moench (2015) might find the results in Table 8 surprising. Lucca and Moench (2015) argue that realized volatility was lower on FOMC days in their sample. It turns out in my extended sample, this is no longer true. However, this is not the fundamental reason our results differ.

To see this in detail, I reproduce the stylized fact that Lucca and Moench (2015) demonstrated as reported by Ai and Bansal (2018). Ai and Bansal's (2018) sample ended in December 2013. I then expanded the sample to the end of my period. I downloaded the precise time of the announcement from Bloomberg and the returns are were computed as the end of the hour returns from TAQ.

To create Table 9, I combined the average the return over the various hours, shifting the time so that the FOMC makes announcement at the end of hour zero in each period. The first (−5) hour is somewhat special because it only shows up in days when the FOMC releases its announcement at non-standard times. There are not 5 hours between 9:30 a.m. and 2:00 p.m. This is why the average return during this period does not change in the larger sample.

Table 9: FOMC: Hourly Return

	-5	-4	-3	-3	-1	0	1	2
May 1997 – December 2013								
FOMC	10.80	5.53	0.12	3.16	10.40	-5.17	0.20	2.81
136 obs.	[3.32]	[2.28]	[0.06]	[1.64]	[1.45]	[-0.71]	[0.09]	[0.51]
non-FOMC	-1.63	-1.65	-1.09	-1.47	-1.14	-1.11	-1.09	1.85
3862 obs.	[-2.08]	[-2.81]	[-2.26]	[-2.86]	[-1.91]	[-1.41]	[-5.13]	[-1.58]
May 1997 – September 2017								
FOMC	10.80	2.33	1.33	1.90	8.71	-2.49	0.16	1.07
156 obs.	[3.32]	[2.06]	[0.71]	[1.14]	[1.47]	[-0.40]	[0.07]	[0.48]
non-FOMC	-1.64	-1.35	-0.93	-1.23	-1.11	-0.89	-0.87	-1.15
4772 obs.	[-2.08]	[-2.70]	[-2.27]	[-2.87]	[-2.24]	[-1.36]	[-3.77]	[-2.48]

As we can see in Table 9 in the hours leading up to an FOMC announcement, the average returns are positive and statistically significant at normal levels. I also report the average return on non-FOMC days for compassion. In the expanded sample, the average returns are smaller for the most part, but they do not go away entirely.

To use the data in Table 9 to identify the sign of additional news premia, we need to remove the standard SDF term — $d\langle m, p\rangle$. There are two ways we can do this. We can either consider the risk-premium over an arbitrarily small interval. As the length of an interval goes to zero, $SDF \to 1$. Hence, any risk premium arising from risk aversion converges to zero. This is the primary argument that Ai and Bansal (2018) make (20). They assume that a FOMC premium occurs over a small enough interval that the SDF effectively equals 1.

There is one big issue with this argument. As can be seen in Table 9, the excess returns do not occur in the minutes around the announcement. Rather they occur a few hours before, not over some arbitrarily small window. This additional premium could potentially still be a news premia. Maybe the news is leaked ahead of the announcement. This is the argument that Ai and Bansal (2018) make. It might be valid; we cannot rule out this channel with the information presented thus far.

Actually, to be fair, we do not need volatility to equal zero for their identification strategy to be valid, we just need it be no higher than average. This is testable. Lucca and Moench (2015, 358) show the intra-day realized volatility was lower on FOMC days in their sample.

This does not validate Ai and Bansal's (2018) identification strategy for two reasons. First, as discussed in Section 10.1 risk premia are not functions of contemporaneous variables. We would expect FOMC days to have higher premia if they conditionally predict future volatility. This is true even if they actually have lower volatility ex-post over the sample in question.

In Table 10, I regress $\log \sigma_t^2 + \gamma_t^2$ onto its lag, lagged jump proportion, and an FOMC indicator. As we can see, $\mathbf{1}\{\text{FOMC}\}_t$ positively predicts volatility at any reasonable significance level.

Table 10: FOMC Conditionally Predicts Volatility

Regressand	Regressors			
$\log \sigma_t^2 + \gamma_t^2$	Constant	$\mathbf{1}\{\text{FOMC}\}$	$\log \sigma_{t-1}^2 + \gamma_{t-1}^2$	$\log \frac{\gamma_{t-1}^2}{\sigma_{t-1}^2+\gamma_{t-1}^2}$
2003–2018/9	-1.11	0.44	0.88	-0.02
	[-6.31]	[8.76]	[54.35]	[-0.33]
2013–2013	-1.25	0.36	0.88	-0.16
	[-6.25]	[5.92]	[47.66]	[-2.33]

The analysis in Table 10 is rather cursory. Maybe if we included additional regressors, the $\mathbf{1}\{\text{FOMC}\}_t$'s coefficient's sign would flip again. However, the important thing is that we cannot reject the hypothesis that $\mathbf{1}\{\text{FOMC}\}_t$ conditionally positively predicts volatility, even over the shorter sample.

This makes the using $\mathbf{1}\{\text{FOMC}\}_t$ dates as a proxy for $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ quite suspect. Lucca and Moench's (2015) headline result here is a t-static, and so its asymptotic distribution is of first-order importance. The literature has assumed the t-statistics are asymptotically Gaussian.

We have a relatively small sample. The FOMC only makes announcements 8 times per year, and so there are only 136 FOMC announcement even in the expanded sample. We also have a small signal-to-noise ratio. Consequently, the risk premia are weakly identified. This type of scenario is precisely when Gaussian distributional approximations perform poorly.

In particular, it is well known, (Andrews and Guggenberger 2009; Andrews and Cheng 2012), that pre-testing can substantially alter the post-test statistics' distributions. In this scenario, pre-testing for volatility, and then estimating risk-premia will likely lead to non-Gaussian distributed statistics in the second stage. In other words, the pre-test Lucca and Moench (2015) perform does not validate Ai and Bansal's (2018) identification strategy.

In addition, once you condition on total volatility and the jump premia and appropriately adjust for heteroskedasticity, the FOMC premia is indistinguishable from zero. It has a t-static of 0.53, Section 10.2.2. To be as forthright was possible, this does not mean the there is nothing special about FOMC days. First, we can explain their observed high ex-post returns by appropriately conditioning on volatility. Explaining a stylized fact is not explaining it away. Second, my sample starts in 2003 because I need a lot of data to separate the jump and diffusive volatilities. This does mean that I cannot say anything about the behavior of the returns before that period.[33] Both Lucca and Moench (2015) and Ai and Bansal (2018) consider substantially longer samples. FOMC

33. The data in Table 9 goes back to 1997, but that is it.

days might have commanded large premia before this period that cannot be explained in terms of volatility terms.

11. Concluding Remarks

In this paper, I investigate how jumps affect investors' risk. I start by showing that standard no-arbitrage based pricing theory implies that jumps are asset price responses to news. 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, as 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 the diffusion volatility 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 this to the S&P 500 using high-frequency data from SPY. I find approximately one-half of the squared variation is jump variation and that this proportion varies significantly over time. I also evaluate 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 taking logs. In addition, γ_t^2 is very persistent and has long-memory. It is also very heavily correlated with the diffusion volatility.

I then analyze jumps' economic risk. I do this by showing that risk premia have the following form $d\langle m, p\rangle(t) + d\left\langle m_*, p^J\right\rangle$, where $m(t)$ is the log SDF and measures curvature in investors' value functions, and $m_*(t)$ is the log Announcement SDF and measures curvature in the investor's recursive preference aggregator.

I then consider the empirical relationship between $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$, which is a nonparametric measure of news risk, and the excess return rx_t. I show that $\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$ commands an economically and statistically significant negative premium. This implies that $m_*(t)$ is high in good times, and hence the certainty equivalence functional (CEF) is quasiconvex.

This negative premium is somewhat surprising as the literature has heretofore argued that uncertainty driven by news commands a larger premium than other uncertainty. The primary evidence for this is the large average ex-post return on days the FOMC makes announcements. I show the volatility on these days and the heteroskedasticity in the noise terms is capable of explaining this large return without any additional news premia.

As this work is the first to analyze the jump volatility, a great deal of work still needs to be done. The obvious first issue 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 well will require figuring out what the "correct" multivariate Laplace distribution is. There are several in the literature.

Second, previous authors have shown that the stylized features of diffusion volatility are relatively stable across different assets, is this also true for the jump volatility? For example, people have argued that news risk is fundamental in understanding foreign exchange markets. How does γ_t^2 act in those environments.

On the financial side, a great deal more empirical and theoretical work is needed to fully understand the relationship between the curvature of the CEF and news premia. A fully specified general equilibrium model would be useful to replace the CAPM-style market volatility terms in Section 10 with covariances with the stochastic discount factors. In addition, further empirical work precisely measuring this curvature and relating the CEF's local and global curvature would be quite useful.

To wrap up, this paper provides a sufficient statistic for jump dynamics — γ_t^2 — and shows that the jump proportion nonparametrically identifies the curvature of the CEF. It also provides a number of new stylized facts and new estimators to enable taking these models to the data. However, like most papers that introduce measures, it leaves a great number of interesting questions to future work.

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.

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.

Andrews, Donald W.K., and Xu Cheng. 2012. "Estimation and Inference With Weak, Semi-Strong, and Strong Identification." *Econometrica* 80 (5): 2153–2211.

Andrews, Donald W.K., and Patrik Guggenberger. 2009. "Incorrect Asymptotic Size of Subsampling Procedures Based on Post-Consistent Model Selection Estimators." *Journal of Econometrics* 152 (1): 19–27.

Baillon, Aurélien, Zhenxing Huang, Asli Selim, and Peter P. Wakker. 2018. "Measuring Ambiguity Attitudes for all (Natural) Events." *Econometrica* (Forthcoming).

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.

———. 2005. "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation." *Journal of Financial Econometrics* 4 (1): 1.

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.

Black, Fischer, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (3): 637–654.

Boldrin, Michele, Larence J. Christiano, and Jonas D. M. Fisher. 2001. "Habit Persistence, Asset Returns, and the Business Cycle." *The American Economic Review* 91 (1): 149–166.

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

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. "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.

Bollerslev, Tim, George Tauchen, and Hao Zhou. 2009. "Expected Stock Returns and Variance Risk Premia." *The Review of Financial Studies* 22 (11): 4463–4492.

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.

Brusa, Francesca, Pavel G. Savor, and Mungo I. Wilson. 2018. *One Central Bank to Rule Them All.* Working Paper. Temple Univeristy, March.

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

Constantinides, George M. 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy* 98 (3): 519–543.

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 Gloria Gonzalez-Rivera. 1991. "Semiparametric ARCH Models." *Journal of Business & Economic Statistics* 9 (4): 345–359.

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.

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. 2014. "Nobel Lecture: Uncertainty Outside and Inside Economic Models." *Journal of Political Economy* 122 (5): 945–987.

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.

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

Jacod, Jean, Mark Podolskij, Mathias Vetter, et al. 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 Rozovksi 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.* Forthcoming.

Karnaukh, Nina. 2016. *The Dollar Ahead of FOMC Target Rate Changes.* Working Paper. Columbus, OH: Ohio State University.

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.

Law, Tzuo Hann, Dongho Song, and Amir Yaron. 2018. *Fearing the Fed: How Wall Street Reads Main Street.* Working Paper. Social Science Research Network, May.

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. 2013. "Robust Estimation and Inference for Jumps in Noisy High Frequency Data: A Local-to-Continuity Theory for the Pre-Averaging Method." *Econometrica* 81 (4): 1673–1693.

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 dellIstituto Italiano degli Attuari* 64:19–47.

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.

Mueller, Philippe, Alireza Tahbaz-Salehi, and Andrea Vedolin. 2017. "Exchange Rates and Monetary Policy Uncertainty." *The Journal of Finance* 72 (3): 1213–1252.

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.

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 James Barnett William .A. 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.

Patton, Andrew J., Johanna F. Ziegel, and Rui Chen. 2018. *Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk).* Working Paper. Duke University, April.

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.

Sangrey, Paul. 2018. *Jumps, Tail Risk, and the Distribution of Stock Returns.* Working Paper. University of Pennsylvania.

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.

Savor, Pavel, and Mungo Wilson. 2014. "Asset Pricing: A Tale of Two Days." *Journal of Financial Economics* 113 (2): 171–201.

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 24641. May.

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

APPENDIX A REPRESENTATION THEOREMS

Theorem 3 (Time-Changing Jump Martingales)**.** *Let $p^J(t)$ be a purely discontinuous, infinite-activity, locally-square integrable martingale that can be represented as $H * (\mu - \nu)$ where $H(t)$ is a predictable process, μ a Poisson random measure, and ν its predictable compensator with Lebesgue base Levy measure. Further assume that it has no predictable jumps.*

Then $p^J(t)$ time-changed by its predictable quadratic variation is a standard variance-gamma process. In other words, $p^J(t) \overset{\mathcal{L}}{=} \mathcal{L}\left(\left\langle p^J \right\rangle(t)\right)$.[34]

Proof. To prove the result, we proceed in a number of steps. Since we are staring 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. Therefore, there are two key parts to the result above. First, we need to handle the dynamics contained in H, and second we need to reduce the two-dimensional representation to a one-dimensional one.

First, 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, we switch the base Levy measure to one that has this property. Second, we make a time-change argument in each strip to deal with its dynamics. Third, we then switch from an integral with respect to a Poisson process to one with respect to a Poisson random measure by taking the appropriate sum of these processes.

We also use capital letters to refer to processes as is standard in the literature. Since we are not doing any discretization, there should be no confusion here. Define $1^z = 1\{x \in [z, z + dz]\}$ where $z \in \mathbb{R}$, and $dz \in \mathbb{R}_+$, where we 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 $[z, z = dz]$.

We now turn to switching 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, $(Y^z)^\tau = \sum_{s \leq t, x \in [z, z+dz]} \delta(x, s)$, where δ is a predictable Dirac delta. Also, define $\langle W \rangle^{-1}(t) = \inf\{\tau : \langle W \rangle = t\}$ for any process W. This is the standard inverse definition when the process may be zero, and is innocuous here because if $\langle W \rangle \overset{a.s.}{=} 0 \implies W \overset{a.s.}{=} 0$.

Recall that we assume 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, we will use tilde's to refer to measures associated with this random measure. For example, $\tilde{\nu}$ is its associated compensator. Note, since we are using compensated random measures, each strip $[z, z + dz]$ is a martingale.

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

The benefit of using this representation is that it implies the associated predictable integrator $\tilde{H}$ 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, we are 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 we will use in the sequel.

We 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.

$$\hat{\mathcal{F}}_t^{X^n} \coloneqq \cap_{\epsilon>0}\mathcal{F}_{t-\epsilon}^{\tilde{X}^z} \cup \sigma(\{\rho < t\}) \tag{59}$$

It is worth noting that when you progressively enlarge a filtration with an honest time, semimartingales with respect to the original filtration are still semimartingales with respect to to the new filtration (Barlow 1978). However, this enlargement does not necessarily preserve the martingale structure. Since we are 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 since 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 $Z_t^\rho \coloneqq \Pr[\rho > t \,|\, \mathcal{F}_{t-}]$, chosen by to be cádlág. The $\mathcal{F}_t$ dual optional projection of the process $1\{\rho \le t\}$ is denoted by $A^\rho(t)$. Importantly $\hat{X}$ is a martingale with respect to $\widehat{\mathcal{F}}_t$. Also, define $\mu^\rho(t) = \mathbb{E}[A^\rho(\infty) \,|\, \mathcal{F}_{t-}] = A^\rho(t) + Z^\rho(t)$.

$$X(t) = \widehat{X}(t) + \int_0^{t\wedge\rho} \frac{d\langle X, \mu^\rho(t)\rangle(s)}{Z^\rho(s-)} - \int_\rho^t \frac{d\langle X, A^\rho(t) + Z^\rho(t)\rangle(s)}{1 - Z^\rho(s-)} \tag{60}$$

Since $\mu^\rho(t)$ is $\mathcal{F}_{t-}$ measurable, and the jumps are distributed according to a Poisson process, and hence $\mu^\rho(t)$ is a constant. Consequently, the predictable quadratic variation terms in Equation (60) 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 we will do.

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 the relationship 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 $[H^z * X^z] = \sum_{s \leq t} H^2(\hat{\tau})$, 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, and so the relevant filtrations are independent here. Consequently, we have the 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.

$$\lim_{I\to\infty}\sum_{i\leq I} Y^z\left(\langle Y^z\rangle^{-1}\right) \overset{\mathcal{L}}{=} \lim_{I\to\infty} W(\exp^{\sum \nu^i}) = W(\Gamma(t)) = \mathcal{L} \tag{61}$$

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.

□

Corollary 3.1 (Time-Changing Finite-Activity Jump Martingales). *Let $Y(t)$ be a purely discontinuous, locally-square integrable martingale that can be represented as $H*(\mu-\nu)$ where $H(t)$ is a predictable process, μ a Poisson random measure, and ν its predictable compensator with Lebesgue base Levy measure λ. Further assume that it has no predictable jumps. Then $Y(t)$ time-changed by its predictable quadratic variation is a mixture of the* 0 *process — δ_0 – and the standard 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 wen can make the argument we 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.

$$Y(t) = \begin{cases} \mathcal{L}\left(\langle Y\rangle(t)\right) & \text{with intensity } \nu \\ \delta_0(t) & \text{with intensity } 1-\nu \end{cases} \tag{62}$$

□

Theorem 2 (Jump Volatility and the Predictable Quadratic Variation).

$$\gamma_t^2 = \int_{t-1}^{t} \gamma^2(s)\,ds = \int_{t-1}^{t}\int_{\mathcal{X}} \delta^2(s,x)\nu(dx,ds) = \langle p^J\rangle(t) - \langle p^J\rangle(t-1) \tag{16}$$

Proof.

$$[p]^J(t) = \sum_{s\leq t}\Delta p(s)^2 \tag{63}$$

$$= \int_0^t\int_{\mathcal{X}} \delta^2(s,x)\mu(ds,dx) \tag{64}$$

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.

$$\implies \langle p \rangle^J(t) = \mathbb{E}\left[[p]^j(t) \,\middle|\, \mathcal{F}_{t-} \right] \tag{65}$$

$$= \mathbb{E}\left[\int_0^t \int_X \delta^2(s,x)\mu(ds,dx) \,\middle|\, \mathcal{F}_{t-} \right] \tag{66}$$

$$= \int_0^t \int_X \delta^2(s,x)\nu(ds,dx) \tag{67}$$

We also need to show that the limit in the expectation form approaches $\gamma^2(t)$.
Define $\gamma^2(t) \coloneqq \int_X \delta^2(t,x)\nu(dx,dt)$.

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[\left| p^J(t+\Delta) - p^J(t) \right|^2 \,\middle|\, \mathcal{F}_{t-} \right] = \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[\left| \int_t^{t+\Delta} \delta(s,x)(\mu-\nu)(ds\,dx) \right|^2 \,\middle|\, \mathcal{F}_{t-} \right] \tag{68}$$

By the Itô Isometry, we can rewrite Equation (68) as follows.

$$\text{Equation (68)} = \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[\int_t^{t+\Delta} \delta^2(s,x)\mu(ds\,dx) \,\middle|\, \mathcal{F}_{t-} \right] \tag{69}$$

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

$$= \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[\int_t^{t+\Delta} \int_{\mathcal{X}} \delta^2(s,x)(dx\,dx) \,\middle|\, \mathcal{F}_{t-} \right] \tag{70}$$

$$= \lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}\left[\Delta\gamma^2(t) + \int_t^{t+\Delta} \left(\gamma^2(t) - \gamma^2(s)\right) ds \,\middle|\, \mathcal{F}_{t-} \right] \tag{71}$$

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

$$= \lim_{\Delta\to 0}\gamma^2(t) + \frac{1}{\Delta}\Delta O\left(\mathbb{E}\left[\left| \sup_{t\le s\le t+\Delta} \gamma^2(t) - \gamma^2(s) \right| \,\middle|\, \mathcal{F}_{t-} \right] \right) \tag{72}$$

$$= \gamma^2(t) \tag{73}$$

□

Corollary 3.2 (Jumps Processes as Integrals)**.** *Let $p(t)t$ be a purely-discontinuous Itô semimartingale that is locally-square integrable and has infinite-activity jumps. Then $p(t) = \frac{1}{\sqrt{2}}\int_0^t \gamma(s)\, d\mathcal{L}(s)$, where $\mathcal{L}$ is a standard standard variance-gamma process, and γ is a predictable function.*

Proof. Since $Y(t)$ is an Itô semimartingale, $Y(t) = \int_0^t \int_{\mathbb{R}} \delta(s,x)\, ds$, where we use standard notation.

This implies that its predicable quadratic variation, $K(t) \coloneqq \int_0^t \int_{\mathbb{R}} \delta(s,x)^2 \, dx \, ds$, with time-derivative $k(t) \coloneqq \int_{\mathbb{R}} \delta^2(s,x) \, dx$.

Let $J(t)$ be the purely-discontinuous martingale part of $Y(t)$, then Theorem 3 implies that $J(K^{-1})(t) \overset{\mathcal{L}}{=} \mathcal{L}(t)$, or equivalently, $J(t) \overset{\mathcal{L}}{=} \int_0^{K^{-1}(t)} d\mathcal{L}(s)$. Then since $k\mathcal{L}(1) = \mathcal{L}(k^2)$, where $\mathcal{L}(1)$ is a standard Laplace random variable, and $k(t)$ is a predicable process, this implies that $J(t) = \int_0^t k(s) \, d\mathcal{L}(s)$. This is completely analogous to how the time-changed theorem for continuous processes and absolute continuity imply the integral representation of continuous martingales.

□

APPENDIX B VOLATILITY ESTIMATION

Lemma 5 (HL implies SHL)**.** *If $X^n(t) \xrightarrow{\mathcal{L}\text{-s}} X(t)$ under Assumption SHL, then $X^n(t) \xrightarrow{\mathcal{L}\text{-s}} X(t)$ under Assumption HL, and the equivalent statement holds for convergence in probability.*

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

Step 1

We can assume without loss of generality that $b(0) = 0$, and so there is a localizing sequence T_p such that $\|\mu(t)\| \leq p$ if $0 \leq t \leq T_p$. Define the stopping times $R_p = \inf\left(t : \|X(t)\| + \|\sigma(t)\| \geq p\right)$ and the stopping times $Q_p = \inf\left(t : \|X(t)\| + \|\gamma(t)\| \geq p\right)$. These increase to $+\infty$ as well. Therefore, we can set $S_p = T_p \wedge R_p \wedge Q_p$.

Then we can define the following processes.

$$b^{(p)}(t) = b(t \wedge S_p), \;\; \sigma^{(p)}(t) = \sigma(t \wedge S_p), \;\; \gamma^{(p)}(t) = \gamma(t \wedge S_p) \tag{74}$$

$$X^{(p)}(t) = \begin{cases} 0 & \text{if } S_p = 0 \\ X(0) + \int_0^t b^{(p)}(s) \, ds + \int_0^t \sigma^{(p)}(s) \, dW(s) + \int_0^t \gamma^{(p)} \, d\mathcal{L}(s) & \text{if } S_p > 0 \end{cases} \tag{75}$$

Now, local characteristics of $X^{(p)}$ agree when $t < S_p$ as they are defined to be the same. If $S_p = 0$, then $\|X(t)\| = 0$, 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, $X^{(p)}(t)$ satisfies Assumption SHL, since $\|X^{(p)}(t)\| \leq 3p$.

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 $X(t)$ does not satisfy Assumption I, 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 $X^{(p)}(t)$ with the 0 process when necessary, and so the result even holds if Assumption I does not hold.

□

Theorem 7 (Estimating the Instantaneous Absolute Volatility)**.** *Let $p(t)$ be Itô semimartingale.*

Let $p(t)$ satisfy Assumptions HL, I, and SQ, and 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 $\operatorname{erfcx} := \frac{2\exp(x^2)}{\sqrt{\pi}} \int_x^\infty \exp(-s^2)\, ds$.[35]

$$\frac{1}{k_n\sqrt{\Delta_n}} \sum_{m=0}^{k_n-1} |\Delta^n_{i_n+m} p| \xrightarrow{\mathbb{P}} \mathbb{E}|N(0,1)|\sigma(\tau-) + \frac{\gamma(\tau-)}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma(\tau-)}{\gamma(\tau-)}\right) \tag{38}$$

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.

Step 1

In this section, I start by applying Itô's Formula for convex functions to $|X|(t)$ to separate its variation into its jump and continuous components. Recall the left-derivative of the absolute value function.

$$f'_- = \operatorname{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x \leq 0 \end{cases} \tag{76}$$

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

35. erfcx is the scaled complementary error function. It is a reparameterization of Mill's ratio. Efficient, numerically stable implementations are provided by most scientific programming suites.

$$\begin{aligned} |X(t)|) &= \int_0^t \operatorname{sign}(X(s-))\,dX(s) + A(t) && (77)\\ &= \int_0^t \operatorname{sign}(X(s-))\,dW(s) + \int_0^t \operatorname{sign}(X(s-))\,d\mathcal{L}(s) + A(t) && (78) \end{aligned}$$

$A(t)$ is a finite-valued increasing process and so it can be absorbed into the drift term of $X(t)$ and vanishes as $\Delta \to 0$. If the Laplace part and the Diffusion parts have the same sign, $|X|(t) - A(t)$ 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, $\operatorname{sign}(X(s-))$ 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 $|X(t)| - A(t)$ 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.

$$\begin{aligned} |X(t)| - A(t) = &\int_0^t \operatorname{sign}(W(s-))1_{\Omega^W}(s-)\sigma(s)\,dW(s) - \int_0^t \operatorname{sign}(\mathcal{L}(s-))1_{\Omega^W}(s-)\gamma(s)\,d\mathcal{L}(s) \\ &+ \int_0^t \operatorname{sign}(\mathcal{L}(s-))1_{\Omega^{\mathcal{L}}}(s-)\gamma(s)\,d\mathcal{L}(s) - \int_0^t \operatorname{sign}(W(s-))1_{\Omega^{\mathcal{L}}}(s-)\sigma(s)\,dW(s) \end{aligned} \tag{79}$$

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 Equation (79) as follows condition on the signs differing as follows..[36]

$$\begin{aligned} &\int_0^\infty \int_x^\infty (y-x)\psi_{\gamma,\Delta}(x)|\phi|_{\sigma,\Delta}(y)\,dx\,dy + \int_0^\infty \int_y^\infty (x-y)\psi_{\gamma,\Delta}(x)|\phi|_{\sigma,\Delta}\,dy\,dx && (80)\\ =&\frac{\sqrt{\Delta}}{\sqrt{2}}\left(-\gamma + \frac{2}{\sqrt{\pi}}\sigma + \gamma \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right)\right) + \frac{\gamma\sqrt{\Delta}}{\sqrt{2}}\operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right) && (81)\\ =&\sqrt{\Delta}\left(m_1\sigma + \frac{\gamma}{\sqrt{2}}\left(2\operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right) - 1\right)\right) && (82) \end{aligned}$$

In the part where they both have the same sign, we can the absolute value is just the sum of the absolute values and so we can rewrite Equation (79) given that the signs are the same as follows.

$$m_1\sigma\sqrt{\Delta} + \frac{\gamma}{\sqrt{2}}\sqrt{\Delta} \tag{83}$$

Then by taking the average of the Equation (82) and Equation (83), we can solve for Equation (79).

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

$$\mathbb{E}|X(t)| - A(t) = m_1 \sigma \sqrt{\Delta} + \frac{\gamma\sqrt{\Delta}}{\sqrt{2}} \operatorname{erfcx}\left(\frac{\sigma}{\gamma}\right) \tag{84}$$

The first part of this equation is the expectation of the absolute value of the diffusion part. If erfcx($\frac{\sigma}{\gamma}$ were replaced with 1, the second part would be the absolute value of the jump part. Consequently,erfcx $\left(\frac{\sigma}{\gamma}\right)$ reweighs 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 exactly, what we would want.

Step 3

In this section, we consider the asymptotic behavior of the estimator.

We 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 $\{T\}$ such that $\sigma(t)$ and $\gamma(t)$ are constant over the intervals between the variance T. follows.

$$X(t) = \sum_j \int_{T_j}^{T_{j+1}} \sigma(T_j)\, dW(s) + \int_{T_j}^{T_{j+1}} \gamma(T_j)\, d\mathcal{L}(s) \tag{85}$$

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 T goes to zero, (which it may not at all). Consequently, we let T depend upon n in our notation.

$$\mathbb{E}\left[\left\|\frac{1}{k_n\sqrt{\Delta_n}} \sum_{m=0}^{n} |\Delta_{i_n+m}^n X| - \left|(m_1\sigma(T_n) + \frac{1}{2}\gamma(T_n)\right|\right\|^2\right] \tag{86}$$

$$= \frac{1}{k_n^2\Delta_n}\mathbb{E}\left[\left\| \sum_{m=0}^{k_n} \left| \int_{T(n,m+1)}^{T(n,m)} \sigma(T_n) dW(s) + \int_{T(n,m+1)}^{T(n.m)} \gamma(T_n) d\mathcal{L}(s)\right| \right.\right. \tag{87}$$

$$\left.\left. - k_n\sqrt{\Delta_n}\left|m_1\sigma(T_n) + \frac{1}{2}\gamma(T_n)\right|\right\|^2\right]$$

We now split the integral into two parts. One over Ω_n and one over Ω_n^c.

$$\leq \frac{1}{k_n^2 \Delta^n} \mathbb{E}\Bigg[\Bigg\| \sum_{\Omega_n} \left| \int \sigma(T_n)\, dW(s) \right| + \left| \int \gamma(T_n)\, d\mathcal{L}(s) \right| \tag{88}$$
$$+ \sum_{\Omega_n^c} \left| \left| \int \sigma(T_n)\, dW(s) \right| - \left| \int \gamma(T_n)\, d\mathcal{L}(s) \right| \right|$$
$$- \frac{\lambda_n \sqrt{\Delta^n}}{2} |m_1 \sigma(T_n) + \gamma(T_n)| - \frac{k_n \lambda_n^c \sqrt{\Delta^n}}{2} |m_1 \sigma(T_n)| + O_p(\Delta^n) \Bigg\|^2 \Bigg]$$

By the last two terms are the expectations of the first terms up to finite-variation terms, and so we have the difference of two martingales. So by the Burkholder-Davis-Gundy inequality and the triangular inequality, we can simplify the above expression as follows.

$$= \qquad \frac{1}{k_n^2 \Delta^n} \left(O(k_n \Delta_n) + O(k_n \Delta^n) + o(k_n \Delta^n) \right) \to 0 \tag{89}$$

Step 5

To finish deriving the theorem, we show that approximating the volatility functions by step functions is innocuous. Consider a sequence $T_n \to \tau$, and define $\tilde{\sigma}(t) = \sigma(\max T_n : T_n \leq t)$, 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 Equation (85). Let $t_1, t_2 < \tau$.

$$\mathbb{E}\Bigg[\Bigg\| \int_{t_1}^{t_2} \mu(s)\, ds + \int_{t_1}^{t_2} \sigma(s)\, dW(s) + \frac{1}{2} \int_{t_1}^{t_2} \gamma(s) d\mathcal{L}(s) \tag{90}$$
$$- \left(\int_{t_1}^{t_2} \tilde{\sigma}(s)\, dW(s) + \frac{1}{2} \int_{t_1}^{t_2} \tilde{\gamma}(s)\, d\mathcal{L}(s) \right) \Bigg\|^2 \Bigg]$$

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

$$\leq \mathbb{E}\left[\int_{t_1}^{\tau} \mu(s)^2(s) + \int_{t_1}^{\tau} |\tilde{\sigma}(s) - \sigma(s)|^2\, ds + \frac{1}{2} \int_{t_1}^{\tau} |\tilde{\gamma}(s) - \gamma(s)|^2\, ds \right] \tag{91}$$

Then we can bound each of the terms.

$$= (O(1)\gamma_b^2(\tau) + O(1)\gamma_\sigma^2(\tau) + O(1)\gamma_\gamma^2(\tau))(\tau - t_2) \tag{92}$$
$$= O(1)(\tau - t_2) \tag{93}$$

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 Equation (85). So consider the mean-square deviation of the estimator from its limiting value.

$$\frac{1}{k_n^2 \Delta_n} \mathbb{E}\left[\left\| \sum_{m=0}^{k_n - 1} |\Delta_{i_n + m}^n X| - Y'(t) + Y'(t) - \left(m_1 \sigma(T-) k_n \Delta_n + \frac{1}{2} \gamma(T-) k_n \Delta_n \right) \right\|^2 \right] \tag{94}$$

By splitting the term into two parts and using the bounds from Equation (89) and Equation (93).

$$= \frac{1}{k_n^2 \Delta_n} \left(O(\Delta k_n) + O(\Delta k_n) \right) \to 0 \tag{95}$$

$\square$

Theorem 6 (Estimating the Instantaneous Diffusion Volatility)**.** *Let $X(t)$ satisfy Assumptions HL and SQ. Let k_n, Δ^n be sequences such that $k_n \to \infty$, $\Delta^n \to 0$, and $k_n \Delta^n \to 0$. Let τ be a deterministic time with $0 < \tau < \infty$. Let $v_1^n = a(\Delta^n)^z$, where $z < \frac{1}{2}$, and $v_2^n \to 1$. Then we have the following convergence in probability.*

$$\widehat{\sigma^2(i^2)}^n(k_n, \tau, X) \coloneqq \frac{1}{k_n \Delta_n} \sum_{m=0}^{k_n - 1} v_2^n \left|\Delta_{i_n}^n X\right|^2 1\{|\Delta_{i_n}^n x| \leq v_1^n\} \xrightarrow{\mathbb{P}} \sigma^2(\tau-) \tag{37}$$

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 3.1 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, $|\Delta^n_{i_n+m}X| = O_p(\Delta^n)$ for all i, since $x(t)$ is an integral with bounded integrands and integrators with quadratic variation proportional to Δ^n. We start by considering the jump part of the variation. To prove consistency of the original process, we need to show that the jump part converges to zero. Let $Z(t)$ be the jump part of the $X(t)$, and $B(t)$ be the drift part.

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

$$\Big|(x+y+z+w)1\{|x+y+z+w| < v\} - x^2\Big| \leq K\frac{|x|^4}{v^2} + \epsilon x^2 + \frac{K}{\epsilon}((v^2 \wedge y^2) + z^2 + w^2) \tag{96}$$

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. The drift and jump parts just remain as they are.

$$Y^n(t) = \sigma(T_n)(W_t - W_{T_n}1\{T_n \leq t\} \tag{97}$$

$$Y'^n(t) = \int_{T_n\wedge t}^t (\sigma(s) - \sigma(T_n)\, dW(s) \tag{98}$$

$$Z^n(t) = \int_{T_n\wedge t}^t \gamma(s)\, d\mathcal{L}(s) \tag{99}$$

$$B^n(t) = \int_{T_n\wedge t}^t \mu(s) ds \tag{100}$$

Note, $X(T_n \wedge t) = Y^n(t) + Y'^n(t) + Z^n(t) + B^n t$. Now, we can use Equation (96), with $x = \frac{\Delta^n_{i_n+m}Y^n}{\sqrt{\Delta^n}}$, $y = \frac{\Delta^n_{i_n+m}Z^n_{i_n+m}}{\sqrt{\Delta^n}}$, and $w = \frac{\Delta^n_{i_n+m}B^n_{i_n+m}}{\sqrt{\Delta^n}}$. The main issue here is showing that all of the parts except for $Y^n(t)$ converge to zero because then we are essentially just taking the variance of that part. Take $v = \frac{v_n}{\sqrt{\Delta^n}} = a\Delta_n^{-\omega}$, where $\omega > 0$.

Then we have the following inequality.

$$\begin{aligned}\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}\Big|(Y^n_t)^2 - (X^n_t)^2\Big| \leq \frac{1}{k_n}\sum_{m=0}^{k_n-1}\Bigg(K(\Delta^n)^{2\omega}\left|\frac{\Delta^n_{i_n+m}Y^n}{\sqrt{\Delta^n}}\right|^4 + \epsilon\left|\frac{\Delta^n_{i_n+m}Y^n}{\sqrt{\Delta^n}}\right|^2 \\ + \frac{K}{\epsilon}\Delta^\omega\left|\frac{\Delta^n_{i_n+m}Z}{(\Delta^n)^{1/2-\omega}}\right|^2 + \frac{K}{\epsilon}\left|\frac{\Delta^n_{i_n+m}Y'^n}{\sqrt{\Delta^n}}\right|^n + \frac{K}{\epsilon}\left|\frac{\Delta^n_{i_n+m}B}{\sqrt{\Delta^n}}\right|^2\Bigg)\end{aligned} \tag{101}$$

Set $\gamma_n = \sum_{s\in|T_n, T_n+(k_n+2)\Delta^n|}|\sigma(s) - \sigma(T_n)|^2$, which is bounded and converges to zero, and $\phi_n = \sum_{s\in|T_n, T_n+(k_n+2)\Delta^n|}|\gamma(s)|^2$ The key hard part is bounding $\Delta^n_{i_n+m}Z$. Clearly, $E|\Delta^n_{i_n+m}| \leq \phi_n\sqrt{\Delta^n}$. 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.

$$\mathbb{E}|\mathcal{L}(0,\phi_n)\wedge 1| = \phi_n\sqrt{\Delta^n} - \exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right)(\phi_n\sqrt{\Delta^n}+1) \leq O\left(\frac{1}{\sqrt{\Delta^n}}\right)\exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right) \quad (102)$$

In addition, since T_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, $\frac{\Delta^n_{i_n+m}}{(\Delta^n)^{1/2-\omega}} \overset{a.s.}{\in} O_p\left(\frac{1}{(\Delta^n)^{1+\omega}}\right)\exp\left(-\frac{1}{\phi_n\sqrt{\Delta^n}}\right) = o_p\left(\Delta^{\omega}\right)$.

Note, I am using K to refer to an arbitrary constant here, which may change. $\Delta^n_{i_n+m}B$ is the drift term, and so $|\Delta^n_{i_n+m}B| \leq K\Delta^n$. $\mathbb{E}[|\Delta^n_{i_n+m}Y^n|^4|\mathcal{F}_{(i_n+m-1)\Delta^n}] \leq K(\Delta^n)^2$. $\mathbb{E}[\left|\Delta^n_{i_n+m}Y'^n\right|^n|\mathcal{F}_{(i_n+m-1)\Delta^n}] \leq K\Delta^2\mathbb{E}[\gamma_n|\mathcal{F}_{(i_n+m-1)\Delta^n} \leq K\Delta^N$.

As a consequence, we have the following where ξ_n is some sequence converging to zero.

$$\mathbb{E}\left[\left|(Y^n_t)^2 - (X^n_t)^2\right|\right] \leq K\epsilon + \frac{K}{\epsilon}\left((\Delta^n)^{2\omega} + \xi_n + \mathbb{E}[\gamma_n]\right) \quad (103)$$

So, 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 $\lim_{n\to\infty}\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}|Y^n_t|^2$ is. If we recall its definition, we note that converges to the variance of the increment.

$$\frac{1}{k_n\Delta^n}\sum_{m=0}^{k_n-1}|\sigma_{T_n}(W_t - W_{T_n})1\{T_n\leq t\}|^2 \quad (104)$$

$$=\sigma(T_n)^2\frac{1}{k_n}\sum_{m=0}^{k_n-1}\left|\frac{\Delta^n_{i_n+m}W}{\sqrt{\Delta^n}}\right|^2 \to \sigma(T_n)^2 \quad (105)$$

Since the square is a convex function, we can combine these two previous limits, and we get that the original expression converges to $\sigma(T_n)^2$. However, this is the local integrated volatility evaluated at T_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.

□

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).

$$\operatorname{erfcx}(x) \coloneqq \mathbb{E}|N(0,1)|r\left(x\sqrt{2}\right) \quad (106)$$

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

It also implies the following bounds, (Theorem 2.3).

$$\frac{2}{\sqrt{\pi}(\sqrt{x^2+2}+x)} < \operatorname{erfcx}(x) < \frac{4}{\sqrt{\pi}(\sqrt{x^2+4}+3x)} \tag{107}$$

Theorem 8 (Estimating the Instantaneous Jump Volatility)**.** *Let $X(t)$ satisfy Assumptions HL, I, and SQ, and 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 $\widehat{\sigma(\tau)_n}$ converge in probability to $\sigma(\tau-)$. Let $\gamma(\tau) > 0$ and g be strictly-increasing, convex, and continuous then we have the following.*

$$\begin{aligned}\gamma(\widehat{k_n, \tau, X}) &\coloneqq \operatorname*{argmin}_{\gamma} g\left(\left|\frac{1}{k_n\sqrt{\Delta}}\sum_{m=0}^{k_n-1}|\Delta^n_{i_n+m}X| - \mathbb{E}|N(0,1)|\widehat{\sigma(\tau)_n} - \frac{\gamma}{\sqrt{2}}\operatorname{erfcx}\left(\frac{\widehat{\sigma(\tau)_n}}{\gamma}\right)\right|\right) \\ &\xrightarrow{\mathbb{P}} \gamma(\tau-)\end{aligned} \tag{39}$$

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

$$\widehat{Q}_n(\gamma) \coloneqq g\left(\left|\frac{1}{k_n\sqrt{\Delta}}\sum_{m=0}^{k_n-1}|\Delta^n_{i_n+m}X| - m_1\hat{\sigma}(T-) - \gamma\operatorname{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right)\right|\right) \tag{108}$$

We can start by noting that $\widehat{Q}_n(\gamma)$ is implicitly a continuous function of $\hat{\sigma}_n(T-)$. However, since, by assumption, $\hat{\sigma}_n(T-) \xrightarrow{\mathbb{P}} \sigma_0$, 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 $\widehat{Q}_n(\gamma)$.

$$Q_0(\gamma) \coloneqq g\left(\left|\gamma_0\operatorname{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right) - \gamma\operatorname{erfcx}\left(\frac{\sigma_0}{\gamma\sqrt{2}}\right)\right|\right) \tag{109}$$

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 $A(\sigma,\gamma) \coloneqq \gamma\operatorname{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right)$. Showing $A(\sigma,\gamma)$ is strictly increasing for all σ is sufficient to show this convexity because of properties on g and the absolute-value function. This statement is likely to hold as erfcx is almost constant as a function of γ, and so we have to make rigorous what is meant by almost.

$$\frac{\partial}{\partial\gamma}\gamma\operatorname{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right) = \operatorname{erfcx}\left(\frac{\sigma}{\gamma\sqrt{2}}\right) - \frac{\sigma}{\gamma^2\sqrt{2}}\frac{\partial}{\partial x}\operatorname{erfcx}(x)\bigg|_{x=\frac{\sigma}{\gamma\sqrt{2}}} \tag{110}$$

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 $\hat{\gamma}_n \xrightarrow{\mathbb{P}} \gamma_0$.

$\square$

APPENDIX C NEWS PREMIA THEOREMS

Theorem 1 (Jump Times are News Times)**.** *Consider a predictable 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 ${}^p_p(t)$ is pre-visible, ${}^p_p(\tau)$ 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 $\rceil$ 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 $\rceil$. Let $P(t)$ be the price, and $M(t)$ be the representative investor's pricing kernel.

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

$$P(t) = \mathbb{E}\left[M(\tau)P(\tau) \,|\, \mathcal{F}_t^r\right] \forall \tau \geq t \tag{111}$$

In addition, e being relevant for asset pricing implies that there exists a stopping time τ such that the following inequality holds.

$$\mathbb{E}\left[M(\tau)P(\tau) \,|\, \mathcal{F}_t^r\right] \neq \mathbb{E}\left[M(\tau)P(\tau) \,|\, \mathcal{F}_t^p\right] \tag{112}$$

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 10 (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 the following.*

$$d\mathbb{E}\left[f(\widetilde{Z}) \,\middle|\, \mathcal{F}_{t-}\right] = \mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) \,\middle|\, \mathcal{F}_{t-}\right] + \frac{1}{2} f''(\tilde{Z}(t-))\, d\langle \tilde{Z} \rangle(t) \tag{55}$$

Proof. The argument below is a standard application of Itô's formula for non-continuous processes applied to processes of founded 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).

$$d\mathbb{E}\left[f(\tilde{Z}(t)) \,\middle|\, \mathcal{F}_{t-}\right] = d\mathbb{E}\left[\sum_{i=1}^{d} \frac{\partial f}{\partial z_i} \tilde{Z}(t-)\, d\tilde{Z}_i(t) + \frac{1}{2}\sum_{i,j} \frac{\partial f}{\partial z_i \partial z_j} f(\tilde{z}(t-)) \left\langle \tilde{z}_i{}^D, \tilde{z}_j{}^D \right\rangle(t) \right. \tag{113}$$
$$\left. + \left(\Delta f(\tilde{Z}(t)) - \sum_{i=1}^{d} \frac{\partial f}{\partial z_i} f(\tilde{Z}(t-))\Delta \tilde{Z}_i(t)\right) \,\middle|\, \mathcal{F}_{t-}\right]$$

Rearranging and combining terms, we have.

$$= d\mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) + \left(\Delta f(\tilde{Z}(t)) + f'(\tilde{Z}(t-))\Delta \tilde{Z}(s)\right) \right. \tag{114}$$
$$\left. - \frac{1}{2} f''(\tilde{Z}(t-)) \left\langle \tilde{Z}^D \right\rangle(t) \,\middle|\, \mathcal{F}_{t-}\right]$$

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.

$$= d\mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) + \frac{1}{2} f''(\tilde{Z}(t-))\, d\left\langle \tilde{Z}^D \right\rangle(t) \,\middle|\, \right] \tag{115}$$
$$+ \frac{1}{2}\, d\mathbb{E}\left[f''(\tilde{Z}(t-))\Delta \tilde{Z}(t)^2 + O((\Delta \tilde{Z}(t)^3) \,\middle|\, \mathcal{F}_{t-}\right] \tag{116}$$

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

$$= d\mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) \,\middle|\, \mathcal{F}_{t-1}\right] + \frac{1}{2} f''(\tilde{Z}(t-))\, d[\tilde{Z}](t) + \mathbb{E}\left[O((\Delta \tilde{Z}(t)^3)) \,\middle|\, \mathcal{F}_{t-}\right] \tag{117}$$

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

$$= d\mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) \,\middle|\, \mathcal{F}_{t-1}\right] + \frac{1}{2}\mathbb{E}\left[f''(\tilde{Z}(t-))\, d[\tilde{Z}](t) \,\middle|\, \mathcal{F}_{t-}\right] \tag{118}$$
$$+ \mathbb{E}\left[c_1 O([\tilde{Z}]^{3/2}) \,\middle|\, \mathcal{F}_{t-}\right]$$

Since we are considering local changes in time.

$$= d\mathbb{E}\left[f'(\tilde{Z}(t-))\, d\tilde{Z}(t) \,\middle|\, \mathcal{F}_{t-}\right] + \frac{1}{2} f''(\tilde{Z}(t-))\, d\left\langle \tilde{Z} \right\rangle(t) \tag{119}$$

□

Theorem 9 (Asset-Pricing Equation)**.** *Let the assumptions in Assumption 5 hold, prices 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 prices satisfy the following asset pricing equation for all stopping times $\tau > t$, where $M(t) := \frac{DV(t)}{\mathbb{E}[DV(t) \,|\, \mathcal{F}_{t-}]}$ and $M_*(t) = \frac{D\mathcal{I}(V(t))}{\mathbb{E}[D\mathcal{I}(V(t)) \,|\, \mathcal{F}_{t-}]}$.*

$$\widetilde{P}(t) = \mathbb{E}\left[M(\tau)M_*(\tau)\widetilde{P}(\tau) \,\middle|\, \mathcal{F}_{t-}\right] \tag{49}$$

Proof. First define the discounted price: $\widetilde{P}(t) := \exp(-\kappa t)P(t)$. I start by substituting out the consumption in terms of the constraint. By doing that we can rewrite, Problem 1 as follows. This is an optimization problem and so we can characterize the optimum using first-order conditions.

$$V(\Xi(t), P(t)) = \max_{\Xi(t), c(t)} \int_t^{t+\Delta} u(c(s))\, ds + \exp(-\kappa\Delta)\mathcal{I}\left[V(\Xi(t+\Delta), P(t+\Delta)) \,|\, \mathcal{F}_t\right] \tag{120}$$

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.

$$V(\Xi(t-\Delta), \widetilde{P}(t)) = \max_{\Xi(t), c(t)} u(c(t)) + \exp(-\kappa\Delta)\mathcal{I}\left[V(\Xi(t), \widetilde{P}(t+\Delta)) \,\middle|\, \mathcal{F}_t\right] \tag{121}$$

$$c(t) + \sum_i P_i(t)\xi_i(t) = \sum_i P_i(t)\xi_i(t+\Delta) \tag{122}$$

We can substitute in the constraints for $c(t)$. Then the first-order conditions for each $\xi_i(t)$ have the following form, at the optimal level of consumption $c^*(t)$, where I substitute prices for discounted ones.

$$u'(c^*(t))P_i(t) = \mathbb{E}\left[\mathcal{DI}\left[V(\Xi(t), \widetilde{P}(t+\Delta)) \,\middle|\, \mathcal{F}_t\right] \mathcal{D}V^*(\Xi(t), \widetilde{P}(t+\Delta))\widetilde{P}_i(t+\Delta) \,\middle|\, \mathcal{F}_t\right] \tag{123}$$

We can rearrange this equation as follows. A discounted price at time t is just itself. We are discounting over a length-zero interval.

$$\widetilde{P}_i(t) = \mathbb{E}\left[\mathcal{DI}\left[V(\Xi^*(t+\Delta), \widetilde{P}(t+\Delta)) \,\middle|\, \mathcal{F}_t\right] \frac{\mathcal{D}V(\Xi^*(t+\Delta), \widetilde{P}(t+\Delta))}{u'(c(t))}\widetilde{P}_i(t+\Delta) \,\middle|\, \mathcal{F}_t\right] \tag{124}$$

Define the two SD F's are as follows: the regular SDF — $M(\tau) := \frac{\mathcal{D}V(\Xi^*(\tau), \widetilde{P}(\tau))}{u'(c(t))}$ — and A-SDF — $M_*(\tau) := \mathcal{DI}\left[V(\Xi^*(\tau), \widetilde{P}(\tau)) \,\middle|\, \mathcal{F}_t\right]$. Then since Δ is arbitrary, Equation (124) is equivalent to the following.

$$\widetilde{P}_i(t) = \mathbb{E}\left[M_*(\tau)M(\tau)\widetilde{P}_i(\tau)\,\middle|\,\mathcal{F}_t\right] \tag{125}$$

□

Theorem 11 (Asset-Pricing Equation). *Let the assumptions in Assumption 5 hold, prices 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 *are as follows.*

$$\mathbb{E}\left[\frac{dP_i(t)}{P_i(t-)} - \frac{dP_{rf}(t)}{P_{rf}(t-)}\,\middle|\,\mathcal{F}_{t-}\right] = -\,d\langle m, p\rangle(t) - d\langle m_*, p^J\rangle(t) \tag{56}$$

Proof. The goal here is to replace the asset pricing equation in Theorem 9 with a stochastic logarithm of $P_i(t)$ I abuse notation here and let $M(t)$ be the non-discounted stochastic discount factor. In Theorem 9 it was mean 1, here it will have the same mean as the price of a risk-free asset.

$$\widetilde{P}_i(t) = \mathbb{E}\left[M(\tau)M_*(\tau)\widetilde{P}(\tau)\,\middle|\,\mathcal{F}_t\right] \tag{126}$$

Since $M(t), M_*(t)$ given $\mathcal{F}_t$ equal 1, we can pre-multiply by them.

$$M(t)M_*(t)P(t) = \mathbb{E}\left[M(\tau)M_*(\tau)P(\tau)\,|\,\mathcal{F}_t\right] \tag{127}$$

In other words, $M(t)M_*P(t)$ 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.) This comes from the Jensen inequality term when you expand the logarithm, which needs to be appropriately dealt with.

Before, I derive the risk premia below, I consider a few properties of the stochastic logarithm. 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.

$$\begin{aligned}\mathcal{L}og(X \cdot Y \cdot Z) = {} & \mathcal{L}og(X) + \mathcal{L}og(Y) + \mathcal{L}og(Z) + [\mathcal{L}og(X), \mathcal{L}og(Z)] + [\mathcal{L}og(X), \mathcal{L}og(Z)] \\ & + [\mathcal{L}og(Y), \mathcal{L}og(Z)]\end{aligned} \tag{128}$$

As noted above, since the original expression is a logarithm, the stochastic logarithm is as well.

$$0 = \mathbb{E}\left[\int_t^{\tau} d\mathcal{L}og\left(MM_*P\right)(s) \,\middle|\, \mathcal{F}_t\right] \tag{129}$$

We can expand this using Equation (128). We can also replace the integrals with differentials without loss of generality because τ is arbitrary.

$$\begin{aligned} \implies 0 = \mathbb{E}\,[&d\mathcal{L}og(M)(t) + d\mathcal{L}og(M_*)(t) + d\mathcal{L}og(P)(t) \\ &+ d[\mathcal{L}og(M), \mathcal{L}og(P)](t) + d[\mathcal{L}og(M_*), \mathcal{L}og(P)](t) + d[\mathcal{L}og(M), \mathcal{L}og(M_*)](t) \,|\, \mathcal{F}_{t-}] \end{aligned} \tag{130}$$

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

$$\begin{aligned} = \mathbb{E}\,[&d\mathcal{L}og(M)(t) + d\mathcal{L}og(M_*)(t) + d\mathcal{L}og(P)(t) \\ &+ d[\log(M), \log(P)](t) + d[\log(M_*), \log(P)](t) + d[\mathcal{L}og(M), \mathcal{L}og(M_*)](t) \,|\, \mathcal{F}_{t-}] \end{aligned} \tag{131}$$

We can combine M and M_* together.

$$= \mathbb{E}\left[d\mathcal{L}og\left(M \cdot M_*\right)(t) + d\mathcal{L}og(P)(t) + d[\log(M), \log(P)](t) + d[\log(M_*), \log(P)](t) \,|\, \mathcal{F}_{t-}\right] \tag{132}$$

The stochastic logarithm satisfies the following stochastic differential equation.

$$\mathcal{L}og(X)(t) = \int_0^t \frac{1}{X(s-)}\, dX(s) \tag{133}$$

Consequently, we can rewrite Equation (132) as follows. I also replace the quadratic variation terms with predictable quadratic variation terms.

$$0 = \mathbb{E}\left[\frac{d\left(M \cdot M_*\right)(t)}{M(t-)M_*(t-)} + \frac{dP(t)}{P(t-)} \,\middle|\, \mathcal{F}_{t-}\right] + d\langle \log(M), \log(P)\rangle(t) + d\langle \log(M_*), \log(P)\rangle(t) \tag{134}$$

If $M_*(t)$ is identically 1, the all of the terms containing it disappear, and we have the following. Also, we

$$\mathbb{E}\left[\frac{dP(t)}{P(t-)} + \frac{dM(t)}{M(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = -\, d\langle m, p\rangle(t) \tag{135}$$

Equation (135) is the standard asset pricing equation.

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

$$\mathbb{E}\left[\frac{dP(t)}{P(t-)} + \frac{d\left(M \cdot M_*\right)(t)}{M(t-)M_*(t-)} \,\middle|\, \mathcal{F}_{t-}\right] = -\, d\langle m, p\rangle(t) - d\langle m_*, p\rangle(t) \tag{136}$$

Since the expression above must price all assets, if we consider a risk-neutral asset, we have all of the of the quadratic variation terms being equal to zero.

$$\frac{dP_{rf}(t)}{P_{rf}(t-)} = -\mathbb{E}\left[\frac{d\left(M \cdot M_*\right)(t)}{M(t-)M_*(t-)}\,\middle|\,\mathcal{F}_{t-}\right] \tag{137}$$

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

$$\frac{dP_i(t)}{P_i(t-)} - \frac{dP_{rf}(t)}{P_{rf}(t-)} = -\,d\langle m, p\rangle(t) - d\langle m_*, p\rangle(t) \tag{138}$$

$\square$

Table 11: Univariate Autoregressive Models

	const	lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7	lag 8	lag 8	lag 10	$\mathbb{R}^2$
$\log \sigma_t^2$	−1.18	0.88										78 %
	(−1.33, −1.02)	(0.87, 0.90)										
	−0.63	0.62	0.13	0.04	0.04	0.08						80 %
	(−0.78, −0.47)	(0.58, 0.65)	(0.09, 0.16)	(0.01, 0.08)	(0.04, 0.11)	(0.05, 0.12)						
$\log \gamma_t^2$	−1.27	0.87										76 %
	(−1.43, −1.11)	(0.86, 0.89)										
	−0.45	0.51	0.16	0.06	0.08	0.05	0.00	0.01	−0.04	0.06	0.05	80 %
	(−0.61, −0.29)	(0.48, 0.55)	(0.13, 0.20)	(0.03, 0.10)	(0.04, 0.12)	(0.01, 0.09)	(−0.03, 0.04)	(−0.08, −0.01)	(−0.08, −0.01)	(0.03, 0.10)	(0.02, 0.09)	

Table 12: Vector Autoregression Models

	$\log \sigma_t^2$		$\log \gamma_t^2$	
VAR(1)				
const	−0.66	(−0.84, −0.49)	−1.51	(−1.67, −1.35)
$\log \sigma_{t-1}^2$	0.64	(0.59, 0.69)	0.27	(0.22, 0.35)
$\log \gamma_{t-1}^2$	0.30	(0.24, 0.31)	0.58	(0.53, 0.63)
$\mathbb{R}^2$	78.53 %		77.14 %	
Innovation Covariance	$\begin{pmatrix} 0.25 & 0.18 \\ 0.18 & 0.20 \end{pmatrix}$			
VAR(6) — Chosen by SIC				
const	−0.23	(−0.42, −0.03)	−0.58	(−0.75, −0.41)
$\log \sigma_{t-1}^2$	0.48	(0.43, 0.53)	0.23	(0.18, 0.28)
$\log \gamma_{t-1}^2$	0.18	(0.11, 0.24)	0.31	(0.25, 0.36)
$\log \sigma_{t-2}^2$	0.07	(−0.01, 0.12)	−0.03	(−0.08, 0.02)
$\log \gamma_{t-2}^2$	0.07	(0.01, 0.14)	0.16	(0.11, 0.22)
$\log \sigma_{t-3}^2$	0.02	(−0.03, 0.08)	−0.07	(−0.11, −0.02)
$\log \gamma_{t-3}^2$	0.02	(−0.04, 0.08)	0.13	(0.07, 0.18)
$\log \sigma_{t-4}^2$	0.03	(−0.02, 0.09)	−0.07	(−0.12, −0.02)
$\log \gamma_{t-4}^2$	0.05	(−0.01, 0.11)	0.16	(0.10, 0.21)
$\log \sigma_{t-5}^2$	0.08	(0.26, 0.14)	0.01	(−0.04, 0.06)
$\log \gamma_{t-5}^2$	−0.03	(−0.10, 0.03)	0.07	(0.02, 0.13)
$\log \sigma_{t-6}^2$	0.04	(−0.01, 0.10)	−0.04	(−0.09, 0.00)
$\log \gamma_{t-6}^2$	−0.03	(−0.10, 0.03)	0.08	(0.03, 0.14)
$\mathbb{R}^2$	80.30 %		80.14 %	
Innovation Covariance	$\begin{pmatrix} 0.23 & 0.16 \\ 0.16 & 0.18 \end{pmatrix}$			

Note, this table is the transpose of most of the tables in this document. We do this in order to fit the document in a reasonable fashion.

APPENDIX E NEWS PREMIA: EMPIRICAL RESULTS

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

Constant	$\mathbf{1}\{\mathrm{FOMC}\}_t$	$\log\left(\sigma_t^2 + \gamma_t^2\right)$	$\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\log\left(\sigma_t^2 + \gamma_t^2\right)\log\left(\frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}\right)$	$\bar{\mathbb{R}}^2$
−2.28		−0.26			1.3 %
[−5.01]		[−5.58]			
−2.28		−0.26			1.4 %
[−5.01]		[−5.58]			
0.81	0.80				0.6 %
[6.15]		[3.94]			
−1.68		−0.23	0.51		1.5 %
[−3.11]		[−4.80]	[2.36]		
−1.76	0.44	−0.24	0.53		1.6 %
[−3.22]	[2.62]	[−4.86]	[2.47]		
3.06	0.45	0.24	7.56	0.71	1.9 %
[1.96]	[2.72]	[1.58]	[3.51]	[3.29]	
3.08		0.25	7.45	0.70	1.7 %
[1.98]		[1.62]	[3.47]	[3.26]	

Table 14: Instrument Variables: First Stage Regression

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

Regressand	Constant	$\mathbf{1}\{\text{FOMC}\}$	ϕ_{t-1}	ϕ_{t-2}	ϕ_{t-5}	ϕ_{t-25}	ψ_{t-1}	ψ_{t-2}	ψ_{t-5}	ψ_{t-25}	$\psi_{t-1}\phi_{t-1}$	$\bar{\mathbb{R}}^2$
	−0.38		0.44									0.20
	[−17.97]		[13.48]									
$\log \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$	−0.22		0.30	0.18	0.15	0.05						0.25
	[−10.63]		[13.85]	[9.18]	[8.56]	[3.28]						
	−0.51		0.22	0.13	0.14	0.08	−0.05	−0.01	0.01	0.02		0.29
	[−0.88]		[11.07]	[6.11]	[7.18]	[4.37]	[−8.05]	[−0.95]	[2.21]	[5.19]		
	−0.36	−0.07	0.41	0.13	0.14	0.08	−0.03	−0.01	0.01	0.02	0.02	0.29
	[−3.81]	[−3.96]	[3.91]	[6.11]	[7.27]	[4.34]	[−3.08]	[−0.87]	[2.24]	[5.33]	1.82	
	−1.08						0.88					0.78
	[−5.91]						[46.95]					
$\log \sigma_t^2 + \gamma_t^2$	−0.47						0.62	0.18	0.10	0.05		0.80
	[−4.97]						[27.55]	[8.32]	[6.11]	[5.01]		
	−0.33		−0.12	0.12	0.09	0.01	0.61	0.19	0.11	0.05		0.80
	[−2.27]		[−1.92]	[2.03]	[1.61]	[0.18]	[27.38]	[8.60]	[6.37]	[4.24]		
	−1.39	0.43	−1.49	0.13	0.09	0.03	0.50	0.19	0.11	0.05	−0.15	0.81
	[−4.55]	[9.39]	[−4.10]	[2.09]	[1.66]	[0.48]	[13.52]	[8.56]	[6.29]	[4.24]		

Table 15: News Premia Estimates: $\psi_t \coloneqq \log \sigma_t^2 + \gamma_t^2$, $\phi_t \coloneqq \gamma_t^2 / \sigma_t^2 + \gamma_t^2$

	Regressors				Instruments			
	Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log \psi_t$	$\log \phi_t$	$\mathbf{1}\{\text{FOMC}\}_t$	$\log \psi_{t-l} \ldots$	$\log \phi_{t-l} \ldots$	$\log \psi_{t-l} \log \phi_{t-l}$
	3.26		0.29			✓		
	[6.32]		[5.78]					
	3.22		0.29		✓	✓	✓	
	[6.40]		[5.86]					
	-1.23			-2.38			✓	
	[-3.82]			[-4.74]				
$l \in \{1, 2, 5, 25\}$	-1.61			-2.96	✓	✓	✓	
	[-5.31]			[-6.34]				
	0.64		0.18	-2.29	✓	✓	✓	
	[0.85]		[3.17]	[-4.66]				
	0.81		0.19	-2.16	✓	✓	✓	✓
	[1.10]		[3.37]	[-4.48]				
	0.83	0.09	0.19	-2.15	✓	✓	✓	✓
	[1.14]	[0.51]	[3.43]	[-4.46]				
	3.34		0.30		✓			
	[6.21]		[5.69]					
$l = 1$	-1.08			-2.14			✓	
	[-2.79]			[-3.55]				
	1.39		0.21	-1.61	✓	✓		
	[1.45]		[3.29]	[-2.41]				
	1.53	0.12	0.22	-1.50	✓	✓	✓	✓
	[1.68]	[0.72]	[3.51]	[-2.38]				

Table 16: News Premia Estimates: Levels
(Volatility is measured in yearly terms. (252*daily)).
$l \in \{1, 2, 5, 25\}$

Regressors					Instruments			
Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\sigma_t^2 + \gamma_t^2$	γ_t^2	$(\sigma_t^2 + \gamma_t^2)(\gamma_t^2)$	$\mathbf{1}\{\text{FOMC}\}_t$	$\sigma_{t-l}^2 + \gamma_{t-l}^2 \dots$	$\gamma_{t-l}^2 \dots$	$(\sigma_{t-l}^2 + \gamma_{t-l}^2)(\gamma_{t-l}^2)$
0.19		6.86			✓			
[4.62]		[3.09]						
0.17			15.18			✓		
[4.08]			[3.55]					
0.23		16.83	−24.59		✓	✓		
[4.46]		[1.11]	[−0.74]					
0.22	0.24	16.94	−24.87		✓	✓	✓	
[4.29]	[1.56]	[1.12]	[−0.75]					
0.19		69.87	−119.63	−88.95	✓	✓	✓	✓
[3.59]		[2.58]	[−2.23]	[−2.23]				
0.19	0.20	63.13	−107.76	−78.21	✓	✓	✓	✓
[3.63]	[1.20]	[2.34]	[−2.02]	[−2.52]				

Table 17: News Premia Estimates: Robustness $\psi_t \coloneqq \log \sigma_t^2 + \gamma_t^2,\ \phi_t \coloneqq \frac{\gamma_t^2}{\sigma_t^2+\gamma_t^2}$, $l \in \{1, 2, 5, 25\}$

	Regressors				Instruments			
	Intercept	$\mathbf{1}\{\text{FOMC}\}_t$	$\log \psi_t$	$\log \phi_t$	$\mathbf{1}\{\text{FOMC}\}_t$	$\log \psi_{t-l} \ldots$	$\log \phi_{t-l} \ldots$	$\log \psi_{t-l} \log \phi_{t-l}$
Sub-period Analysis								
2003–2007	3.42	0.39	0.31	0.37	✓	✓	✓	✓
	[1.15]	[1.33]	[1.43]	[0.37]				
2008–2012	1.86	0.72	0.18	−0.36	✓	✓	✓	✓
	[0.96]	[1.99]	[1.29]	[−0.33]				
2013–2007/9	1.74	−0.23	0.34	−3.05	✓	✓	✓	✓
	[1.00]	[−0.89]	[2.64]	[−3.08]				
Unweighted Analysis								
	1.14		0.11		✓	✓	✓	✓
	[1.17]		[1.66]					
	−1.00			−1.59	✓	✓	✓	✓
	[−3.19]			[−3.33]				
	−1.67		−0.05	−1.82	✓	✓	✓	✓
	[−1.38]		[−2.91]	[−0.57]				
	−1.59	0.88	−0.06	−1.59	✓	✓	✓	✓
	[−1.34]	[3.12]	[−0.62]	[−2.53]				