Practical Specification of Affine Jump-Diffusion Stochastic Volatility Models

Anatoliy Belaygorod *
Washington University in Saint Louis
email: belaygorod@wustl.edu

Atilio Zardetto
Reinsurance Group of America
email: azardetto@rgare.com

September 2014 (First Draft: April 2010)

Abstract

We present econometric arguments against the popular trend to add volatility jumps to diffusion-driven models of stock price evolution on the grounds of theoretical analysis of the state-space model specification, forecasting, and the robustness of model estimation. Despite our enhanced MCMC algorithm applied in this paper to the estimation of latent volatilities, we highlight natural inference limitations stemming from general properties of stochastic volatility models. We find that, while DIC criteria indicates that volatility jumps provide small improvement of fit and likely represent a desirable feature in model specification, this improvement comes at the cost of increased model complexity leading to a decline in estimation efficiency of parameters and inferior forecasting capabilities. We provide evidence that Bates (2000) style AJD models strike the right balance between fit flexibility and estimation properties, unless the data is augmented with some observed variables carrying strong informational signal about latent volatilities.

Keywords: Affine Jump-Diffusion, Heston model, State-Space model, MCMC, Metropolis-Hastings, Bayes, randomized blocking, volatility sampler, DIC

JEL Classification Codes: G12, C11, G17

*The authors thank the participants of: SBIES conference for helpful comments and suggestions. We are grateful to Mark Jensen for helpful discussions.

I. Introduction and Literature Review

Survey papers by Bates (2000), Garcia (2010), Johannes and Polson (2009) show that estimation of continuous-time models for equity returns has become an increasingly popular area of research over the last decade. Heston-style Affine Jump-Diffusion (AJD) square-root stochastic volatility models in particular figure prominently among them for a number of reasons, including a sensible historical fit to the data as well as the existence of semi-closed form solution for option pricing (Heston (1993), Duffie et al. (2000)). In fact, many continuous time "horse-racing" papers either explicitly focus on this subclass (Eraker et al. (2003), Eraker (2004), Pan (2002), Durham (2007), Broadie et al. (2009), Forbes et al. (2007)), or at least feel compelled to include Heston-style models for comparison (Andersen et al. (2002), Chernov et al. (2003), Raggi and Bordignon (2006), Christoffersen et al. (2010), Yu et al. (2011)). AJD models are also popular amongst practitioners, engaged in volatility trading and dynamic hedging, due to their practicality and their computational speed for semi-closed form solutions of vanilla options. Given such widely generated interest, in this paper we decided to shed some more light on the issue of adding jumps to diffusion driven Heston-style models. The popular belief in the current literature is that adding jumps amounts to a significant fit improvement (Eraker et al. (2003)). This trend of adding volatility jumps has been further reinforced recently by highly volatile market movements following the 2008 sub-prime crisis, strengthening the perception that we live in discontinuous world for all state variables.

The main contribution of this paper is to provide practical and multi-directional analysis of the pros and cons of adding jumps to Heston-style specifications. While a richer model with volatility jumps will clearly be more flexible in fitting the data, we will show that the increased model complexity gives rise to econometric problems leading to the loss of estimation efficiency of model parameters and especially latent stochastic volatilities. Moreover, all jump parameters and the jump processes are very difficult to identify given how (by definition) relatively rare these events appear in the historical data. The latent nature of volatility complicates inference on volatility jumps even further. Moreover, similar to Johannes et al. (2009) we found that volatility forecasting suffers in models with volatility jumps due to the fact that the data requires a stronger mean-reversion for the periods of unusually high volatility compared to low and normal volatility environments, creating a problem for in-sample fitting of jumps in volatility. Carrying out a reliable statistical inference as reported by prior-posterior updates on all volatility process parameters and especially jump parameters becomes challenging, unless a very long time-series of data, on the scale of decades, is collected. But on such long time scales one might need to start worrying about parameter regime changes and the stability of model specification.

Moreover, adding jumps to diffusion-driven volatility process, complicates inference on jump occurrences and sizes of equity return process because, following any volatility jump, high level of volatility will likely persist, which will make it more difficult for any inference method to discern true equity jumps from diffusion noise. Consider the following theoretical argument that highlights the perils of modeling jumps in the stochastic volatility process. Econometricians are generally accustomed to relying on the following property, which is true for all non-misspecified time-homogeneous models: "any precision of inference about the unknown model parameters is achievable by increasing the size of the data set." In other words, one would expect posteriors to tighten around the true parameter values and estimation errors of time-invariant parameters to converge to zero as the data size approaches infinity. This property is often used for testing estimation algorithms for convergence as well errors in logic or programming. Unfortunately, in general, all latent state-space models by construction are time-heterogeneous and allow only limited inference about the latent variables. The thought experiment described in Appendix section B demonstrates that the volatility process of all stochastic volatility models is already so highly parameterized that no matter which estimation method we use, and no matter how perfectly specified our model is, and no matter how much data we use, we can't obtain anything better than a noisy signal on latent volatilities. Therefore, it would seem counterintuitive to continue increasing the complexity of volatility parametrization by augmenting it with a time-dependent volatility jump process.

Another way of interpreting state-space models is to view the precise estimation of the transition equation (including time-homogeneous parameters as well as the time-varying vector of volatilities) as inconsequential. Instead, what matters is the proper marginal joint density of volatilities that would enable integrating out volatilities when solving for the unconditional distribution of the observed returns. In this framework, adding jumps could be perceived as an attempt to create a right skew in the marginal joint density of volatility process. However, it is an empirical question to show that this volatility skew is a useful feature which helps to arrive at a more accurate unconditional density of the returns. Note that skewness in volatility is not required for capturing skewness in return distribution. The later density can be skewed even without introducing volatility jumps due to the presence of both: stochastic volatility and equity return jumps. In this context, our results seem to indicate that volatility jumps provide only marginal improvement in that regard.

Formal in-sample model fit comparison is conducted in this paper using Deviance Information Criteria (DIC)[1], which helps weigh the improvement in fit against the penalty for higher parametrization. Our results[2] are consistent with findings of the "horse race" literature listed above. Indeed, we find that adding jumps either only to equity return process (SVJ) or both to equity and volatility processes (SVCJ) results in a significant improvement of DIC metric over the basic SV model - fit improvement outweighs the penalty for higher parametrization. Moreover, DIC results show that SVCJ slightly outperforms SVJ model, indicating that increased model complexity is justified by the improved data fit. However, given all the other drawbacks of SVCJ model compared to SVJ described in this paper, we argue that this DIC advantage of SVCJ over SVJ must weighted against the efficiency drawbacks of estimating the model. In other words, volatility jumps are the right feature to have in affine SV models, but one needs to augment the data set with volatility-loaded observed data points in order to avoid sacrificing estimation properties.

In that venue our companion paper Belaygorod and Zardetto (2013)) argues in favor of jumps in volatility once volatility is "better observed" by adding options-derived implied volatility data in the manner similar to Jones (2003) and Duan and Yeh (2010), which is consistent with conclusions about usefulness of volatility jumps reached in Eraker (2004) and Broadie et al. (2007). Similarly, a vast literature utilizing "observed" data on realized volatility, makes a strong case for discrete jumps in volatility. (Bollerslev et al. (2009), Bollerslev et al. (2008), Jiang and Oomen (2007), Todorov (2011), Todorov et al. (2011)). Takahashi et al. (2009) treat realized volatility as an observed quantity and incorporate it as a second dimension of the measurement equation of log-stochastic volatility model, which enables them to carry out a more accurate volatility inference.

However, the information set of the models compared in this paper does not include intraday returns or observed realized/implied volatilities. Instead, the filtering problem is conditional only on daily returns of the underlying (e.g S&P500 index). Therefore, in order to explicitly distinguish the latent nature of volatility that we discuss in this paper and avoid confusing it with observed (realized or option implied quantities), we are going to refer to volatility state

[1]See Berg et al. (2004) for a detailed description and application to SV models.
[2]See Section II for detailed description of model nomenclature considered in this paper

variable in AJD models considered here as "volatility measure." The other reason we introduce this differentiation has to do with the fact that volatility measure in our models gauges the impact of the diffusion noise on the stock price, yet at each time point returns also depend on the random noise coming from the jump process. Therefore, realized volatility quantities that are typically reported (e.g. Bloomberg©) are not directly comparable with volatility measure in our model even if exactly the same data length and frequency was used to estimate them both.

An alternative econometric modeling approach aimed at capturing skew and kurtosis in the distribution of the underlying is based on a mixture of normals and doesn't require discontiniuos jump structures. Jensen and Maheu (2010) present an efficient Bayesian MCMC algorithm applied to stochastic volatility model with infinitely ordered mixture of normals driven errror innovation terms, whose component probabilities and parameters are modeled with the Dirichlet process mixture prior (DPM). This semi-parametric approach has very promising econometric properties, but unlike parametric Affine Jump Diffusion setting, it lacks semi-closed form option pricing capability, which makes it less attractive for financial applications such as dynamic hedging. Also, incorporating leverage effect in DPM setting is computationally challenging according to their footnote 3.

Another argument made in the literature in favor of SVCJ model has to do with empirically observed jump clustering. It is correctly pointed out in Eraker et al. (2003) that "clustering of jump arrivals and size reversals are extremely unlikely given the i.i.d. jump time and size specifications and the infrequent nature of jumps..." However, adding volatility jumps to AJD with jumps only in equity returns doesn't seem to reduce either jump clustering or size reversals in our results, which was one of the arguments in favor of volatility jumps in Eraker et al. (2003). While our datasets and models are not directly comparable with their paper, we speculate that arguably excessively high parametrization of models with volatility jumps gives rise to spurious results driven by small sample or model mis-specification, which could easily explain why their earlier finding are not sufficiently robust to be corroborated in our setting.

One of the main reasons affine jump diffusion models became popular was the existence of risk-neutral model counterparts, which allow for semi-closed form pricing of a wide class of exponential affine payoffs, as described in Duffie et al. (2000). Therefore, it is important to take into consideration the comparison made from Risk-Neutral modeling framework point of view. Gatheral (2006) argues that "SVJ model thus emerges as a clear winner in the comparison between Heston, SVJ and SVJJ models," citing difficulties in calibrating the higher parameterized

SVCJ (Gatheral refers to it as SVJJ) model, as well as the lack of benefit in fitting the traded option implied vol surface. In particular, he demonstrates that fitting long-term maturities and skew can be done well with any SV model, even without jumps. However, capturing short-term vol and skew requires Jumps in equity process (SVJ), while jumps in volatility process do not help to fit the short end.

Estimation methodology used in this paper is an enhanced version of Bayesian MCMC[3] algorithm of Eraker et al. (2003) with all details described in the Appendix A. We strengthen the findings in our paper by analyzing the performance of our estimation algorithm on simulated data first. Unlike Johannes et al. (2009) who implement particle filters for volatility sampling, we offer several methodological enhancements of Bayesian MCMC Metropolis-Hastings methods for sampling latent volatility, including randomized block sizes coupled with volatility transformation specifically designed to improve the efficiency of volatility sampling. The rest of this paper proceeds as follows. Section II describes in detail the three[4] key models considered in this paper: Stochastic Volatility (SV), Stochastic Volatility with Jumps in equity returns (SVJ), and Stochastic Volatility with Correlated Jumps in equity returns and volatility (SVCJ). Section III discusses estimation results and key empirical findings in details and it is followed by Conclusion. Appendices contain low-level details of multiple block Markov Chain Monte Carlo algorithm, as well as various supporting figures and tables.

II. The Model

We have followed practical considerations for constructing the econometric specification of models considered in this paper. In particular, we attempted to make sure that our model can be used to jointly forecast the key economic variables, such as market index returns and risk-free interest rates, capturing the most widely recognized stylized facts, such as flight to quality and leverage effect. Models that use constant risk-free rate assumptions are clearly misspecified, yet models that input risk free rate exogenously lack forecasting capability. Given the post-crisis interest in equity risk premium decomposition between jump premium and other components (Bollerslev and Todorov (2011), Rossi and Timmermann (2011), Yan (2011)), we believe it is important to

[3]Andersen et al. (1999) and Jacquier et al. (1994) provide comparison of EMM, GMM, MCMC and QMLE estimation methods for stochastic volatility models, which strongly favors Bayesian methodology.

[4]Stochastic Volatility with Jumps in Volatility only (SVJV) model is only briefly mentioned in this paper due to its lack of intuitive appeal and practical popularity. Therefore, in the interest of expositional brevity all details for SVJV models are omitted, but they are available from the authors upon request.

more accurately model the risk-free component of the drift. While a simple one-factor interest rate model could have been used[5], we chose the dynamic three-factor Nelson-Siegel specification for its documented empirical success (Diebold and Li (2006)) as well as its compatibility with our effort to model flight-to-quality effect as described below. Note that our flight-to-quality specification has the ability to capture large joint jump-like comovements between bond and equity markets, which is a desirable feature according to Lahaye et al. (2011). Additionally, in most Heston-style models equity returns are primarily driven by the diffusion term as opposed to the conditionally deterministic drift. We attempt to shift this balance (enable the drift to have more explanatory power) by providing a more robust model for the risk-free rate portion of the drift term[6].

We begin with SVCJ - the most general model considered in this paper - which relies on the continuous time specification of Pan (2002) where S_t represents the price of the underlying asset or index at time t:

$$\begin{aligned} \frac{dS_t}{S_t} &= [r_t + \eta^s V_t + \overline{\lambda}(\mu - \mu^*)]dt + \sqrt{V}_t dW_t^s + (e^{Z_t^s} - 1)dJ_t^s - \overline{\lambda}\mu dt \qquad (1) \\ dV_t &= \kappa_v(\theta - V_t)dt + \sigma_v\sqrt{V}_t dW_t^v + Z_t^v dJ_t^v \end{aligned}$$

$$< dW_t^s, dW_t^v > \quad = \quad \overline{\rho}$$

Note, that in the above expression the term r_t has been introduced for the risk free rate at time t. This specification will be further outlined later in the section. W_t^s and W_t^v are standard Brownian motion processes in $\mathbb{R}$ with correlation $\overline{\rho}$. V_t is the underlying latent volatility of the process S_t as in the traditional Bates model. Here, as in Eraker et al. (2003), the volatility process is driven by a mean reverting diffusion and a persistent jump factor. The mean reverting diffusive process is governed by mean reversion strength, κ_v, mean reversion target θ and volatility σ_v. As in Eraker et al. (2003) the moments of the distributions are affected by the jump processes that are introduced; Eraker et al. (2003) describe at length the form of the moments for the updated specification. The third term in the conditional drift of the index return process represents the premium associated with the stochastic jumps. This premium accounts for both the uncertainty

[5]Pan (2002) used one-factor CIR specification for the short rate, but concluded that it added little value to fitting a cross-section of option prices

[6]Results section shows that all modeling enhancements we have introduced to the basic Heston model turned out to be useful for improving model fit to the real data.

in jump arrivals ($\overline{\lambda}$) [7] as well as the uncertainty in jump magnitudes. The difference between μ and μ^* accounts for the different magnitudes of jumps in the real world (μ) and risk neutral (μ^*) probability spaces. Z_t^s is the magnitude of a jump in the continuous return of S_t while for the latent volatility the jump magnitude is defined by Z_t^v. The distributions of these jump magnitudes are driven by each particular specification and are defined later in the section. J_t^v and J_t^s represent Poisson processes. In this paper both Poisson processes are contemporaneous leading to a single intensity parameter $\overline{\lambda}$. The last term in the index return equation (1) is a jump compensator. The compensator term adjusts the drift of the return process by subtracting from it the long run mean of the stochastic jumps in a manner that is akin to Pan (2002). The term η^s is included in this specification as a variance risk premia even though it was found to be insignificant by Pan (2002) and Eraker et al. (2003). By application of Ito's lemma and Euler discretization we obtain the Heston-style dynamics of equity price S_t and its volatility measure V_t in discrete time:

$$
\begin{aligned}
P_{t\Delta} = ln(\frac{S_{t\Delta}}{S_{(t-1)\Delta}}) &= (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + \sqrt{V_{(t-1)\Delta}\Delta}\epsilon_{t\Delta}^s + Z_{t\Delta}^s J_{t\Delta}^s \\
V_{t\Delta} - V_{(t-1)\Delta} &= \kappa_v(\theta - V_{(t-1)\Delta})\Delta + \sqrt{V_{(t-1)\Delta}\Delta}\epsilon_{t\Delta}^v + Z_{t\Delta}^v J_{t\Delta}^v
\end{aligned}
$$

$$
(\epsilon_{t\Delta}^s, \epsilon_{t\Delta}^v)' \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \overline{\rho}\sigma_v \\ \overline{\rho}\sigma_v & \sigma_v^2 \end{pmatrix}
$$

where $\Delta = 252$ corresponds to the daily time step in our data set. Note, that the rigorous implementation of Ito's Lemma gives rise to a premium term of the form $\left(\eta^s - \frac{1}{2}\right)$. For notational ease the $\frac{1}{2}$ has been subsumed in the premium parameter. Also, the distribution of jumps in equation (1) is model specific. It is convenient to set $\bar{\lambda} = \lambda^s + \lambda^v + \lambda^c$, because it incorporates the following models:

- SV:

$$
J_{t\Delta}^s = J_{t\Delta}^v = 0
$$

[7] The specification of our jump risk premium explicitly assumes that the intensity of jump arrivals is the same for the risk neutral probability space as it is for the real world space.

- SVJ:

$$\begin{aligned} J^s_{t\Delta} &\sim Bernoulli(\lambda^s) \\ J^v_{t\Delta} &= 0 \\ Z^s_{t\Delta} &\sim \mathcal{N}(\mu_s, \sigma_s^2) \end{aligned}$$

- SVCJ:

$$\begin{aligned} J^s_{t\Delta} = J^v_{t\Delta} &\sim Bernoulli(\lambda^c) \\ Z^v_{t\Delta} &\sim \exp(\mu_v) \\ Z^s_{t\Delta}|Z^v_{t\Delta} &\sim \mathcal{N}(\mu_s + \rho_J Z^v_{t\Delta}, \sigma_s^2) \end{aligned}$$

The distribution of $Z^s_{t\Delta}$ is conditional on $Z^v_{t\Delta}$ for the SVCJ specification. This is in recognition of the fact that when the volatility jumps it is accompanied by a jump in the index return with a correlation of ρ_J. The parameters μ and μ^* are as follows:

- SV:

$$\mu = \mu^* = 0$$

- SVJ:

$$\begin{aligned} \mu &= \exp(\mu_s + \frac{\sigma_s^2}{2}) - 1 \\ \mu^* &= \exp(\mu_s^* + \frac{\sigma_s^2}{2}) - 1 \end{aligned}$$

- SVCJ:

$$\begin{aligned} \mu &= \frac{\exp(\mu_s + \frac{\sigma_s^2}{2})}{1 - \rho_J \mu_v} - 1 \\ \mu^* &= \frac{\exp(\mu_s^* + \frac{\sigma_s^2}{2})}{1 - \rho_J \mu_v} - 1 \end{aligned}$$

The lagged short risk-free interest rate $r_{(t-1)\Delta}$ can be modeled as the sum of two factors $X_{(t-1)\Delta,1} + X_{(t-1)\Delta,2}$, level and slope, as in Diebold and Li (2006) and Christensen et al. (2011), once we assume that the term-structure of interest rates follows Nelson Seigel-style dynamics.

The transition of level, slope, and curvature factors $\mathbf{X}_t$ evolves in a VAR(1) framework:

$$\mathbf{y}_{t\Delta} = \mathbf{\Lambda}\mathbf{X}_{t\Delta} + \epsilon_{t\Delta} \tag{2}$$

$$\mathbf{X}_{t\Delta} = \overline{\mu} + FtQ_{t\Delta} + \mathbf{A}(\mathbf{X}_{(t-1)\Delta} - \overline{\mu}) + \eta_t \tag{3}$$

Here $\mathbf{y}_{t\Delta}$ is the vector of zero coupon yields at time t. $\overline{\mu}$ is a vector of the long run target of the latent term structure factors, level, slope and curvature. The mean reversion matrix $\mathbf{A}$ is restricted as per the model estimation section. Flight-to-Quality, the $FtQ_{t\Delta}$ in the above expression, is modeled by reflecting the stylized fact that long and medium maturity yields decrease due to the increased demand for treasuries on the days when equity markets drop:

$$FtQ_{t\Delta} = \begin{pmatrix} c_1 \\ -c_1 \\ c_2 \end{pmatrix} ln(\frac{S_{t\Delta}}{S_{(t-1)\Delta}})$$

c_1 and c_2 are Flight-to-Quality parameters. The measurement equation of the term structure of interest rates follows Nelson-Seigel Dynamics of Diebold and Li (2006) with λ exponential decay rate for each maturity τ_p.

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_p}}{\lambda\tau_p} & \frac{1-e^{-\lambda\tau_p}}{\lambda\tau_p} - e^{-\lambda\tau_p} \end{pmatrix},$$

In the above measurement equation $\lambda > 0$. The measurement error is $\epsilon_t|\mathbf{\Sigma_y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma_y})$ where $\mathbf{\Sigma_y}$ is a diagonal matrix; the transition error is $\eta_t|\mathbf{\Omega} \sim \mathcal{N}_3(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega}$ is a full covariance matrix.

III. Estimation Results

Our estimation is based on daily data from Bloomberg on S&P500 and swap rates (with maturities 1, 3, 6, 12mo; 2, 3, 5, 7, 10, 20, 30yr) between Jan'07 and April'12. We intentionally included a sizable sample from pre-2008 crisis period, as well as the crisis itself in order to enable the parameter learning from several types of recent environments. While adding additional historical data is possible, we believe the size of our data set strikes the right balance between several considerations. On the one hand, the data set must be sufficiently large to minimize small sample estimation issues. On the other hand, it is typically important to emphasize the most recent period in order to alleviate concerns about parameter stability over excessively long time horizons, especially given typical short-term forecasting applications of this model (e.g. for dynamic hedging quarterly strategy analysis). Finally, there is a practical/operational concern of increased estimation run-time for long data sets.

Parameter estimates are presented in Table 1 for SV, SVJ, and SVCJ models. The table provides supporting evidence for more efficient estimation of parameters in the SVJ model over the SVCJ[8]. In particular, note the large estimation efficiency improvements for the volatility parameters κ and θ. The additional structure imposed by volatility jumps makes inference on these particular parameters far more difficult in the SVCJ model than in the SVJ. Moreover, the jump parameters themselves are estimated with lower efficiency in the SVCJ. This feature is also driven by the presence of additional structure in the latent process. Table 1 and Figure 3 also show that in the absence of jumps, the SV model is forced to fit large market moves using a generally higher latent volatility process. The standard deviation around the mean estimate of the variance parameter σ_v was the highest in SV model, which is also indicative of inferior fit of SV model.

The intensity of jump parameters is comparable between the SVJ and SVCJ model. Moreover, the magnitude of mean index jump priors is similar between both specifications[9]. However, Figure 7 indicates that SVCJ equity jumps drawn from SVCJ posterior are larger and can be as high as -8%. This occurs because the posterior of equity jump size is sampled conditionally on the actual draw of volatility jump size magnitude, which is drawn from a truncated normal posterior (see Appendix A), introducing a right skew in the marginal posterior distribution of equity jump sizes. Moreover, volatility jumps that occur in the SVCJ model (and persist at roughly the same mean reversion strength as in the SVJ model), will lead to far more realized volatility in the time periods index return process immediately following a volatility jump. This feature can be observed in Figure 1. Note how the observed index return process in the SVCJ

[8] One might argue that a larger parameter inefficiency is not necessarily a sign of an inferior model and could be a reflection that a more complicated model requires a "better" sampler since its posteriors are more complex. While a "better" sampler does exist, to the best of our knowledge, this paper has already utilized state-of-the-art methods to improve on sampling efficiency for all models, which allows for a fair assessment of the estimation complexity. Other things held constant, SVCJ is harder to estimate, which raises even higher the bar for proving its added value.

[9] $\mu_s = -0.03$ in SVJ vs. $\mu_s + \mu_v * \rho_J = -0.01 - 0.53 * 0.04 = -0.031$ in SVCJ model

specification is practically wholly contained within the lower 5% of the in sample estimate of the distribution. Compare this unrealistic result with the SVJ results also in Figure 1. The presence of excessive realized volatility is an undesirable forecasting feature for applications such as dynamic hedging because it can lead to overstated gains/losses through Gamma in a Delta hedged portfolio. Comparison of mean estimated volatility processes in Figure 3 shows that, while generally similar, SVCJ model results in higher volatility measure spikes and consequently further contributes to excess realized volatility. In fact, SVCJ model generates volatility measure process as high as SV model at the time of crisis, but for a different reason. While SV model needs higher volatility measure to fit equity jumps, SVCJ creates higher volatility measure because it allows volatility to jump, which subsequently persists. SVJ model produces notably lower volatility measure around Oct'08 - Apr'09, which corresponds to more realistic in-sample fit during this period on Figure 1.

Out-of-sample 13-month long forecast of daily returns under all three models in Figure 2 supports the conclusion that SVCJ model tends to generate excessive amount of realized volatility, because over 270 daily returns are all contained deep inside the $5^{th} - 95^{th}$ percentile range, while SV and SVJ models produce a more reasonable out-of-sample fit. Note that separation between SV and SVJ models in out-of-sample graphs will become more visible only after plotting far left tail percentiles, capturing equity return jumps. On the contrary, volatility jumps manage to propagate through higher percentiles due to persistence of volatility by raising overall volatility level in the time points following a volatility jump. In analyzing out-of-sample performance, we emphasize the ability of each model to reproduce a realistic realized volatility coverage, because S&P500 returns are mostly driven by the diffusion (and jump, if any) at each time point due to the absence of any known (lagged) predictive covariates that would help identify a strong drift component for S&P500 return process to forecast it *out-of-sample*.

However, one can extract information about the risk-premium from *in-sample* estimation of the drift component, which justifies our efforts in modeling the risk-free term-structure feeding into the drift. While the fit of risk free rate term-structure is not the primary focus of this paper, it is worth noting that parameter estimates reported in the bottom half of Table 1 are quite stable across all three models, because most of the inference about term-structure parameters comes from the Nelson-Siegel modeling component and interest rate data, which are shared by all three models. Estimates of $\overline{\mu}$ imply an intuitive upward sloping yield curve with a long-term target of 5% for long rate. Flight to Quality parameter estimates imply approximately 11bps positive shock to 30yr yield and 13bps to 10yr yield for every 1% change in $S\&P$ index price. Interestingly, Figure 4 shows even larger impact on mid-maturity yields - upto 23bps for 1yr and 19bps for 3yr points. Negligible impact of FtQ on short 1mo rate is consistent with our preliminary analysis of the joint daily co-movement of the data and the imposed parametric structure. The large spike in FtQ impact on mid-maturity yields compared to long maturity yields implies that market participants favor mid-maturity bonds for "parking" their money on temporary basis in current economic conditions, because short-term bond yields are too low, while long-term bonds present too much duration risk, which can generate significant capital losses given the historical upward mean-reverting pressure on yields in our current nearly all time low interest rate environment. These results based on the refined interest rate structure enable us to model the drift more accurately than in most of the literature that typically utilizes simpler drift modeling approaches. As an important consequence, our inference about equity risk premium, which is buried inside the drift, should be more accurate.

From Table 2 we see that the *jump risk premium* is higher in SVCJ due to larger jump intensity. In addition, higher unconditional expected volatility caused by volatility jumps also results in higher *volatility risk premium* in SVCJ compared to SVJ model. Therefore overall risk premium is higher in the SVCJ model. However the premium increase is being driven by parameters which were estimated with less efficiency than in the SVJ model. While stable and reliable risk premium estimates in practice are very hard to obtain, we can see that SV model results in arguably low total risk premium of 3%, while SVCJ and SVJ produce 6.2% and 4% respectively, which make up a more realistic range with the caveat for SVCJ mentioned previously. SVJ risk premium estimate of 4% is primarily driven by the price investors charge for the potential losses due to the expected 2 jumps per year. It is not surprising that the same data could give rise to somewhat varying inference about the drift of the underlying under these models. Although we have parameterized the drift to be a function of variance, it is harder[10] to learn about the former compared to the later. This fact highlights the significance of correctly specifying and estimating the volatility process dynamics, because it will drive our inference about the equity premium. Prior-posterior plots in Figure 6 indicate approximately the same amount of learning from the data about jump parameters in both models, supporting the argument brought up in Section I, stating that identifying jump parameters in general is typically quite challenging. Essentially, these two models provide us with alternative ways of

[10]See Andersen and Benzoni (2008)

fitting the same return data almost equally well, but with somewhat different implication for risk premium. Given our concerns about SVCJ jump/volatility estimation presented above, we would argue that SVJ model paints a more accurate picture of equity risk premium than SVCJ.

In order to address the jump clustering based argument in favor of SVCJ models in EJP, our results in Figure 7 indicate that SVJ model infers the time series of jumps that doesn't appear to exhibit as much clustering, while SVCJ model arguably has more visible jump clustering characteristics. This fact cannot be attributed to a large difference in jump frequencies between the two models in our estimation. Therefore we must conclude, that in our results, clustering appears more prevalent in the SVCJ configuration. However, some of our conclusions could be driven by specific limitations of distributional assumptions of the joint jump process. In fact, Johannes et al. (2009) use sequential likelihood ratios to argue that "...model differentiation occurs primarily during market stress periods, showing the importance of accurate jump modeling for overall model specification." In particular, allowing for negative volatility jumps, non-concurrent jumps, or more flexible (bi-modal) specification for equity jump process could result in different conclusions. In fact, Belaygorod and Zardetto (2013) paper argues in favor of regime-changing SVCJ model estimated using the same data augmented with VIX historical time-series.

Formal in-sample fit comparison is conducted using DIC results presented in Table 2. The results clearly indicate inferiority of SV model and a slight advantage of SVCJ over SVJ in terms of data fit. In order to ensure that our DIC-based conclusions are not driven by estimation noise or the choice of priors, we ran another estimation of SVCJ model where priors were re-centered to match SVJ priors[11]. We found that DIC result is unchanged and while there is some further degradation in parameter efficiency, overall parameter estimates are similar and don't change the conclusion of the paper. Detailed estimation report for the prior robustness exercise run is available from the authors upon request.

Estimation efficiency comparison can be further analyzed by looking at inefficiency factors of sampled volatility measure. Table 3 summarizes volatility inefficiencies at each point in time. These results strongly support the premise that jumps make volatility inference more

[11]In our experience, practical estimation of these models is best accomplished using "Empirical Bayes" approach (for formal treatment see Chapter 3 of Carlin (2000)), which relies on recentering all priors by sequentially re-estimating the model aiming for consistency between the data and hyperparameters controlling the priors. While "purists" might critisize this approach for forming the priors in part by observing the data, from practical perspective of calibrating such highly parameterized SV models, Empirical Bayes is the best alternative in the opinion of these authors and sometimes might be the only choice for selecting prior hyperparameters that results in efficient MCMC algorithms.

difficult, because SVJ model has on average higher and more volatile right skewed distribution of inefficiency factors compared to SV. Similarly, volatility inference in SVCJ model is on average less efficient and more volatile than either SV or SVJ.

Prior-posterior updates of volatility diffusion process parameters in Figure 5 indicate similar amount of learning about the drift and diffusion portion of volatility measure equation in all models, which means that our MCMC algorithm can successfully recover these parameters in the context of all three models. This is not surprising given how similar the inferred posterior volatility paths appear to look. Given all these similarities coupled with the numerous concerns about various features of SVCJ model raised above, we argue that the slight advantage on DIC scale is not sufficient to justify upgrading SVJ model to SVCJ.

IV. Conclusion

Adding jump processes to virtually any diffusion-driven model specification has become almost universally accepted desirable feature. In fact, at this point the burden of proof rests with a researcher who chooses to omit jumps from his model. The usual criticism of diffusion-only driven models consist of fat tail arguments, observed historical events labeled as 'shocks', discreteness of observations and other technical discontinuities, richer, more general, and flexible modeling framework, etc. Nowadays, the typical defense of diffusion-only models is that they simply present the first modeling attempt and jumps could (and should!) be added as an extension and future work. However, in this paper we set out to demonstrate that, contrary to the widely established opinion in the recent literature, depending on the specific modeling application, and at least in the case of widely used AJD setting, adding jumps to volatility process comes with a slew of issues that make us question the trade-off. In particular, at least for applications where volatility process remains latent and cannot be precisely pinned down at each time (volatilities could arguably be isolated by including intraday data, VIX, or some other option data in the estimation set), adding jumps to the diffusion-driven latent volatility process, while slightly improving the fit based on DIC, will result in reduced efficiency of parameter estimation. Estimation results are generally similar, yet show two important differences, the first of which is the estimated long-term equity risk-premium, which is 2% higher with jumps in volatility. The second, the apparent increase in realized volatility in the SVCJ model accompanied by higher inefficiencies and unrealistic in sample distribution of returns.

We conclude by cautioning researchers not to take the benefits of adding jumps for granted and weigh the additional modeling flexibility against various practical considerations. If the presence of volatility jumps is deemed to be an indispensable feature for a given modeling effort or application, we recommend augmenting the information set with any observable data that could shed light on latent volatility process, as in Belaygorod and Zardetto (2013).

APPENDIX

A. MCMC Algorithm

The model is estimated using a Markov Chain Monte Carlo (MCMC) methodology. The target density is sampled by means of a block sampling algorithm whereby the conditional marginal distributions for a set of parameters are sampled by conditioning on the output from other blocks. Parameters are divided into two groups Λ^J and Θ. Λ^J comprises all of the parameters associated with jump arrivals and magnitudes, namely $\Lambda^J = (\lambda^s, \lambda^v, \lambda^c, \mu_s^*, \mu_s, \mu_v, \rho_J, \sigma_s)'$. Θ accounts for all other parameters, namely $\Theta = (\eta^s, \kappa, \theta, \sigma_v, \overline{\rho}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Omega}, \mathbf{\Sigma_y})$.

The Model section also describes the distribution of the jump arrivals and their magnitudes. After the initialization of all parameters, including volatilities, these are the first quantities to be sampled.

1. Jump Occurrence Sampler $\{J_{t\Delta}\}_{t=1}^T$. The distribution of jumps in equation (1) is model specific. It is convenient to set $\bar{\lambda} = \lambda^s + \lambda^v + \lambda^c$ to account for the three possible configurations of the model SVCJ, SV and SVJ.

   - SVCJ: In the case of concurrent jumps we have $J_{t\Delta} = J_{t\Delta}^s = J_{t\Delta}^v$. $J_{t\Delta}$ can be sampled directly from a two point discrete distribution with the following probabilities:

$$\Pr(J_{t\Delta} = 1 | V_{t\Delta}, P_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, Z_{t\Delta}^s, Z_{t\Delta}^v, \Theta, \Lambda^J) \propto$$
$$\overline{\lambda}\Delta f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z_{t\Delta}^s \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta + Z_{t\Delta}^v \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right)$$

$$\Pr(J_{t\Delta} = 0 | V_{t\Delta}, P_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, Z_{t\Delta}^s, Z_{t\Delta}^v, \Theta, \Lambda^J) \propto$$
$$(1 - \overline{\lambda}\Delta) f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right)$$

$$\Sigma = \left(\begin{array}{cc} 1 & \overline{\rho}\sigma_v \\ \overline{\rho}\sigma_v & \sigma_v^2 \end{array}\right)$$

   - SV: In this configuration there are no jumps, therefore $J_{t\Delta} = J_{t\Delta}^s = J_{t\Delta}^v = 0$.

- SVJ: In this configuration jumps occur only on the index return, therefore $J^v_{t\Delta} = 0$. $J^s_{t\Delta}$ is obtained from the following discrete probabilities.

$$\Pr(J^s_{t\Delta} = 1 | V_{t\Delta}, P_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, Z^s_{t\Delta}, \Theta, \Lambda^J) \propto$$
$$\overline{\lambda}\Delta f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right)$$

$$\Pr(J^s_{t\Delta} = 0 | V_{t\Delta}, P_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, Z^s_{t\Delta}, \Theta, \Lambda^J) \propto$$
$$(1 - \overline{\lambda}\Delta) f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right)$$

$$\Sigma = \left(\begin{array}{cc} 1 & \overline{\rho}\sigma_v \\ \overline{\rho}\sigma_v & \sigma_v^2 \end{array}\right)$$

2. Volatility Jump size $\{Z^v_{t\Delta}\}_{t=1}^T$ sampler.

   - SVCJ model:

     For the cases where $J_{t\Delta} = 0$ sampling is performed directly from the prior.

$$Z^v_{t\Delta} \quad \sim \quad \exp(\mu_v)$$

     Otherwise, (i.e. if $J_{t\Delta} = 1$) by Bayes theorem,

     the conditional posterior $Z^v_{t\Delta} | J_{t\Delta} = 1, \Theta, \Lambda^J, V_{t\Delta}, V_{(t-1)\Delta}, P_{t\Delta}, \mathbf{X_{(t-1)\Delta}}, Z^s_{t\Delta}$ is proportional to:

$$\exp\left\{-\frac{Z^v_{t\Delta}}{\mu_v}\right\} f_{\mathcal{N}}(Z^s_{t\Delta} | \mu_s + \rho_J Z^v_{t\Delta}, \sigma_s^2)$$
$$\times f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta + Z^v_{t\Delta} \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right).$$

     Draws from $Z^v_{t\Delta} | J_{t\Delta} = 1, \Theta, \Lambda^J, V_{(t-1)\Delta}, Z^s_{t\Delta}, \mathbf{X_{(t-1)\Delta}}, P_{t\Delta}, V_{t\Delta}$ are thus made by sampling from the truncated normal, $\mathcal{N}(\alpha^v_{t\Delta}, \omega^v_{t\Delta})\mathcal{I}_{[0,\infty)}$, where $\mathcal{I}_{[0,\infty)}$ is an indicator function over the set $[0,\infty)$:

$$\alpha^v_{t\Delta} = \frac{a_{t\Delta}\sigma_s^2\mu_v - \left(\rho_J\mu_v(\mu_s - Z^s_t) + \sigma_s^2\right)V_{(t-1)\Delta}\Delta\sigma_v^2(1-\overline{\rho}^2) - \overline{\rho}\sigma_v b_{t\Delta}\sigma_s^2\mu_v}{\mu_v(\rho_J^2 V_{(t-1)\Delta}\Delta\sigma_v^2(1-\overline{\rho}^2)+\sigma_s^2)}$$

$$\omega^v_{t\Delta} = \frac{\sigma_s^2 V_{(t-1)\Delta}\Delta\sigma_v^2(1-\overline{\rho}^2)}{\rho_J^2 V_{(t-1)\Delta}\Delta\sigma_v^2(1-\overline{\rho}^2)+\sigma_s^2}$$

with $a_{t\Delta} = V_{t\Delta} - V_{(t-1)\Delta} - \kappa(\theta - V_{(t-1)\Delta})\Delta$ and $b_{t\Delta} = P_{t\Delta} - (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta - Z^s_{t\Delta}$.

- SV model:

  No jumps. Nothing to sample. $Z^v_{t\Delta} = 0$

- SVJ model:

  No jumps in Volatility. Nothing to sample. $Z^v_{t\Delta} = 0$

3. Index Jump Size $\{Z^s_{t\Delta}\}_{t=1}^T$ sampler.

- SVCJ model:

  If $J_{t\Delta} = 0$ then sample from the prior: $Z^s_{t\Delta}|Z^v_{t\Delta} \sim \mathcal{N}(\mu_s + \rho_J Z^v_{t\Delta}, \sigma_s^2)$.

  Otherwise, (i.e. if $J_{t\Delta} = 1$) by Bayes theorem,

  the conditional posterior $Z^s_{t\Delta}|Z^v_{t\Delta}, J_{t\Delta} = 1, \Theta, \Lambda^J, V_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, P_{t\Delta}$ is proportional to:

$$f_{\mathcal{N}}(Z^s_{t\Delta}|\mu_s + \rho_J Z^v_{t\Delta}, \sigma_s^2) \times f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta}\end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta + Z^v_{t\Delta}\end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right).$$

  Therefore, draws from $Z^s_{t\Delta}|Z^v_{t\Delta}, J_{t\Delta} = 1, \Theta, \Lambda^J, V_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, P_{t\Delta}$ are made from $\mathcal{N}(\alpha^s_{t\Delta}, \omega^s_{t\Delta})$ where:

$$\alpha^s_{t\Delta} = \frac{\sigma_s^2(\sigma_v a_{t\Delta} - \overline{\rho} b_{t\Delta}) + \Delta\sigma_v V_{(t-1)\Delta}(1-\overline{\rho}^2)(\mu_s + \rho_J Z^v_{t\Delta})}{\sigma_v(\Delta V_{(t-1)\Delta}(1-\overline{\rho}^2) + \sigma_s^2)}$$

$$\omega^s_{t\Delta} = \frac{\sigma_s^2\Delta V_{(t-1)\Delta}(1-\overline{\rho}^2)}{\Delta V_{(t-1)\Delta}(1-\overline{\rho}^2)+\sigma_s^2}$$

with $a_{t\Delta} = V_{t\Delta} - V_{(t-1)\Delta} - \kappa(\theta - V_{(t-1)\Delta})\Delta - Z^v_{t\Delta}$ and $b_{t\Delta} = P_{t\Delta} - (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta$.

- SV model

  No jumps. Nothing to sample. $Z^s_{t\Delta} = 0$

- SVJ model

  If $J_{t\Delta} = 0$ then sample from the prior: $Z^s_{t\Delta} \sim \mathcal{N}(\mu_s, \sigma_s^2)$.

  Otherwise, (i.e. if $J_{t\Delta} = 1$) by Bayes theorem,

  the conditional posterior $Z^s_{t\Delta}|J_{t\Delta} = 1, \Theta, \Lambda^J, V_{(t)\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, P_{t\Delta}$ is proportional to:

$$
\begin{aligned}
&f_{\mathcal{N}}(Z^s_{t\Delta}|\mu_s, \sigma_s^2) \\
&\quad \times f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right) \middle| \left(\begin{array}{c} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right).
\end{aligned}
$$

  Therefore, draws from $Z^s_{t\Delta}|J_{t\Delta} = 1, \Theta, \Lambda^J, V_{t\Delta}, V_{(t-1)\Delta}, \mathbf{X_{(t-1)\Delta}}, P_{t\Delta}$ are made from $\mathcal{N}(\alpha^s_{t\Delta}, \omega^s_{t\Delta})$ where:

$$
\begin{aligned}
\alpha^s_{t\Delta} &= \frac{\sigma_s^2(\sigma_v a_{t\Delta} - \overline{\rho} b_{t\Delta}) + \Delta\sigma_v V_{(t-1)\Delta}(1-\overline{\rho}^2)\mu_s}{\sigma_v(\Delta V_{(t-1)\Delta}(1-\overline{\rho}^2) + \sigma_s^2)} \\
\omega^s_{t\Delta} &= \frac{\sigma_s^2 \Delta V_{(t-1)\Delta}(1-\overline{\rho}^2)}{\Delta V_{(t-1)\Delta}(1-\overline{\rho}^2) + \sigma_s^2}
\end{aligned}
$$

  with $a_{t\Delta} = V_{t\Delta} - V_{(t-1)\Delta} - \kappa(\theta - V_{(t-1)\Delta})\Delta$ and $b_{t\Delta} = P_{t\Delta} - (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta$.

4. Jump parameters (Λ^J) are sampled with the tailored Metropolis-Hastings sampler of Chib and Greenberg (1994), Chib and Greenberg (1995) where the target density $\pi(\Lambda^J|\Theta, P, V, y, J, Z, X)$ is sampled by Bayesian methods based on the following proportionality:

$$
\begin{aligned}
\pi(\Lambda^J|\Theta, y, P, V, J, Z, \mathbf{X}) &\propto \pi(\Theta, \Lambda^J) f(P, V, y|\mathbf{X}, J, Z, \Theta) \\
&\quad \times f(J|\Lambda^J) f(Z|\Lambda^J)
\end{aligned}
$$

   Construction of the proposal density is done using standard MLE approaches. An interior point optimization method is used for estimation of the proposal mean. The negative inverse Hessian obtained from this optimization is then used to construct the variance of the normal proposal. The objective function for optimization is the posterior derived from Bayes rule as shown above. Once a draw is made from the proposal density it is subjected to an accept reject algorithm (Chib (2001)). All parameters are initialized at prior means with a suitable burn in period implemented before draws are considered acceptable. Target densities were constructed as follows.

- SV

  No jumps. Nothing to sample. $Z^s_{t\Delta} = 0$

- SVJ

$$
\begin{aligned}
&\pi(\Lambda^J|\Theta, y, P, V, J, Z, \mathbf{X}) \propto \pi(\overline{\lambda})\pi(\mu_s)\pi(\mu_s^*)\pi(\sigma_s) \\
&\quad \prod_{t=1}^{T} f_{\mathcal{N}}\left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{pmatrix}, V_{(t-1)\Delta}\Delta\Sigma\right) \times \\
&\quad \prod_{t=1}^{T} \left\{ \begin{array}{c} \Delta\overline{\lambda} \\ 1 - \Delta\overline{\lambda} \end{array} \right\} \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(\frac{(Z^s_{t\Delta} - \mu_s)^2}{2\sigma_s^2}\right)
\end{aligned}
$$

  The priors for λ and σ_s were selected as Gamma Distributions. The priors for μ_s and μ_s^* were selected as Normal Distributions. For all prior parameters refer to table at the end of the section.

- SVCJ

$$
\begin{aligned}
&\pi(\Lambda^J|\Theta, y, P, V, J, Z, \mathbf{X}) \propto \pi(\lambda)\pi(\mu_s)\pi(\mu_s^*)\pi(\sigma_s)\pi(\rho_J)\pi(\mu_v) \\
&\quad \prod_{t=1}^{T} f_{\mathcal{N}}\left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta + Z^v_{t\Delta} \end{pmatrix}, V_{(t-1)\Delta}\Delta\Sigma\right) \times \\
&\quad \prod_{t=1}^{T} \left\{ \begin{array}{c} \Delta\overline{\lambda} \\ 1 - \Delta\overline{\lambda} \end{array} \right\} \frac{1}{\mu_v} \exp\left(-\frac{Z^v_{t\Delta}}{\mu_v}\right) \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(\frac{(Z^s_{t\Delta} - \mu_s - \rho_J Z^v_{t\Delta})^2}{2\sigma_s^2}\right)
\end{aligned}
$$

  The priors for μ_v, λ and σ_s were selected as Gamma Distributions with the appropriate parameters. The priors for ρ_J, μ_s and μ_s^* were selected as Normal Distributions with the appropriate parameters.

5. Parameter Θ. The parameters of Θ sampled in the first sub-block are $\{\eta^s, \kappa, \theta\}$. These parameters are sampled from the following target density.

$$\pi(\eta^s, \kappa, \theta | \Lambda^J, \sigma_v, \overline{\rho}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Omega}, \mathbf{\Sigma_y}, y, P, V, J, Z, \mathbf{X}) \propto \pi(\eta^s)\pi(\kappa)\pi(\theta) f(P, V | J, Z, \mathbf{X}, \Theta, \Lambda^J)$$

All priors were chosen as Normal distributions with appropriate parameters. The likelihood function $f(P, V|J, Z, \Theta)$ is constructed as follows.

- SV

$$f(P, V|\mathbf{X}, J, Z, \Theta, \Lambda^J) = \prod_{t=1}^{T} f_{\mathcal{N}} \left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{pmatrix}, V_{(t-1)\Delta}\Delta\Sigma \right)$$

- SVJ

$$f(P, V|\mathbf{X}, J, Z, \Theta, \Lambda^J) = \prod_{t=1}^{T} f_{\mathcal{N}} \left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \end{pmatrix}, V_{(t-1)\Delta}\Delta\Sigma \right)$$

- SVCJ

$$f(P, V|\mathbf{X}, J, Z, \Theta, \Lambda^J) = \prod_{t=1}^{T} f_{\mathcal{N}} \left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta + Z^v_{t\Delta} \end{pmatrix}, V_{(t-1)\Delta}\Delta\Sigma \right)$$

The Parameters of Θ sampled in the second sub-block are $\{\lambda, \overline{\mu}, \mathbf{A}, c_1, c_2\}$ are sampled next. These parameters are sampled using Kalman Filtering to filter out the factors $\mathbf{X}_{t\Delta,i}$

using standard Predictive Error Decomposition (PED) as described in Kim and Nelson (1999) and Hamilton (1994). The target density for these parameters is as follows.

$$\pi(\lambda, \overline{\mu}, \mathbf{A}, c_1, c_2 | \Lambda^J, \eta^s, \kappa, \theta, \sigma_v, \overline{\rho}, \mathbf{\Omega}, \mathbf{\Sigma_y}, y, P, V, J, Z) \propto \pi(\lambda)\pi(\overline{\mu})\pi(\mathbf{A})\pi(c_1)\pi(c_2) f(P, V, \mathbf{y} | J, Z, \Theta)$$

The prior of λ was chosen to be log-normal. All other priors in this block were chosen normal. Using the notation of the Kalman filter, the first distribution is the measurement of the data and the second is the transition of the factors. The transition equation in the case of all three models is as follows for a single time point in discrete form.

$$\begin{bmatrix} X_{t\Delta,1} \\ X_{t\Delta,2} \\ X_{t\Delta,3} \\ X_{t-1\Delta,1} \\ X_{t-1\Delta,2} \end{bmatrix} = \begin{bmatrix} \overline{\mu}_1 \\ \overline{\mu}_2 \\ \overline{\mu}_3 \\ \overline{\mu}_1 \\ \overline{\mu}_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ -c_1 \\ c_2 \\ 0 \\ 0 \end{bmatrix} ln(\frac{S_{t\Delta}}{S_{(t-1)\Delta}}) +$$

$$\begin{bmatrix} a_{1,1} & a_{1,2} & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & 0 \\ 0 & 0 & a_{3,3} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \left(\begin{bmatrix} X_{(t-1)\Delta,1} \\ X_{(t-1)\Delta,2} \\ X_{(t-1)\Delta,3} \\ X_{(t-1)\Delta,1} \\ X_{(t-1)\Delta,2} \end{bmatrix} - \begin{bmatrix} \overline{\mu}_1 \\ \overline{\mu}_2 \\ \overline{\mu}_3 \\ \overline{\mu}_1 \\ \overline{\mu}_2 \end{bmatrix} \right) + \eta_t$$

The elements of the matrix **A** were restricted as per Diebold et al. (2005), Diebold and Li (2006). Also, $\eta_t | \mathbf{\Omega} \sim [\mathcal{N}_3(\mathbf{0}, \mathbf{\Omega}), 0, 0]^T$. The last two terms of the transition are deterministic (known) at time t. The measurement equation for the three models are as follows.

- SV

$$f(P, V, y, \mathbf{X} | J, Z, \Theta, \Lambda^J) =$$

$$\prod_{t=1}^{T} f_{\mathcal{N}}\left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \\ \mathbf{y}_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (X_{t-1\Delta,1} + X_{t-1\Delta,2} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \\ \mathbf{\Lambda X}_{t\Delta} \end{pmatrix}, \overline{\mathbf{\Sigma}}\right)$$

- SVJ

$$\begin{aligned} &f(P, V, y, \mathbf{X}|J, Z, \Theta, \Lambda^J) = \\ &\prod_{t=1}^{T} f_{\mathcal{N}}\left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \\ \mathbf{y}_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (X_{t-1\Delta,1} + X_{t-1\Delta,2} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta \\ \mathbf{\Lambda X}_{t\Delta} \end{pmatrix}, \overline{\mathbf{\Sigma}}\right) \end{aligned}$$

- SVCJ

$$\begin{aligned} &f(P, V, y, \mathbf{X}|J, Z, \Theta, \Lambda^J) = \\ &\prod_{t=1}^{T} f_{\mathcal{N}}\left(\begin{pmatrix} P_{t\Delta} \\ V_{t\Delta} \\ \mathbf{y}_{t\Delta} \end{pmatrix} \middle| \begin{pmatrix} (X_{t-1\Delta,1} + X_{t-1\Delta,2} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta + Z^s_{t\Delta} \\ V_{(t-1)\Delta} + \kappa(\theta - V_{(t-1)\Delta})\Delta + Z^v_{t\Delta} \\ \mathbf{\Lambda X}_{t\Delta} \end{pmatrix}, \overline{\mathbf{\Sigma}}\right) \end{aligned}$$

Note that for all three cases $\overline{\Sigma} = f(\Delta V_{(t-1)\Delta}, \mathbf{\Sigma}, \mathbf{\Sigma_y})$. The factors $\mathbf{X}_{t\Delta}$ are subsequently sampled for all time points using standard Kalman Smoothing conditional on the parameters Θ. The smoothed factors are then used in the conditioning sets for all other blocks. The next sub-block samples $\mathbf{\Sigma_y}$ from the following density:

$$\pi(\mathbf{\Sigma_y}|\Lambda^J, \eta^s, \kappa, \theta, \sigma_v, \overline{\rho}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Omega}, y, P, V, J, Z, \mathbf{X}) \propto \pi(\mathbf{\Sigma_y}) f(P, V, \mathbf{y}|\mathbf{X}, J, Z, \Theta, \Lambda^J, \mathbf{\Sigma_y})$$

This distribution is proportional to:

$$\propto \pi(\mathbf{\Sigma_y}) \prod_{t=1}^{T} f_{\mathcal{N}}\Big(\mathbf{y}_{t\Delta} \Big| \Big(\mathbf{\Lambda X}_{t\Delta}, \mathbf{\Sigma_y} \Big)\Big)$$

Since $\mathbf{\Sigma_y}$ is an uncorrelated diagonal matrix, we can sample its elements one at a time from the following target density.

$$\propto \pi(\sigma_{y,k}) \prod_{t=1}^{T} f_{\mathcal{N}}\left(\; y_{t\Delta,k} \; \Big| \Big(\; \lambda_{\mathbf{k}} \mathbf{X}_{t\Delta} \; , \sigma_{y,k} \Big) \right)$$

The counter k is for the different maturities of y. Therefore λ_k represents the k^{th} row of $\mathbf{\Lambda}$. By selecting the prior of $\sigma_{y,k}$ to be Inverse Gamma distributed as $\sim \mathcal{IG}(\frac{\alpha_{y,k}}{2}, \frac{\beta_{y,k}}{2})$ we can derive the following closed form conditional target density using conjugate priors.

$$\pi(\sigma_{y,k}|\Lambda^J, \eta^s, \kappa, \theta, \sigma_v, \overline{\rho}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Omega}, y, P, V, J, Z, \mathbf{X}) \sim \mathcal{IG}\left(\frac{\alpha_{y,k} + T}{2}, \frac{\beta_{y,k} + \sum_{t=1}^{T} (y_{t\Delta,k} - \lambda_{\mathbf{k}} \mathbf{X}_{t\Delta})^2}{2} \right)$$

The next sub-block is for $\mathbf{\Omega}$ which is sampled from the following target density.

$$\pi(\mathbf{\Omega}|\Lambda^J, \eta^s, \kappa, \theta, \sigma_v, \overline{\rho}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Sigma_y}, y, P, V, J, Z, \mathbf{X}) \propto \pi(\mathbf{\Omega}) f(P, V, \mathbf{y}|\mathbf{X}, J, Z, \Theta, \Lambda^J, \mathbf{\Omega})$$

This distribution is proportional to:

$$\propto \pi(\mathbf{\Omega}) \prod_{t=1}^{T} f_{\mathcal{N}}\left(\; \overline{\mathbf{X}}_{t\Delta} \; \Big| \Big(\; FtQ_{t\Delta} + \mathbf{A}\overline{\mathbf{X}}_{(t-1)\Delta} \; , \mathbf{\Omega} \Big) \right)$$

Note that in the above expression the factors have been demeaned such that $\overline{\mathbf{X}}_{t\Delta} = \mathbf{X}_{t\Delta} - \overline{\mu}$. To facilitate sampling of $\mathbf{\Omega}$ a Wishart prior is selected for Ω^{-1} such that $\mathbf{\Omega^{-1}} \sim \mathcal{W}_3(\alpha_\Omega, \mathbf{R_\Omega})$. Given this choice, the following result is obtained by conjugate priors.

$$\begin{aligned} \pi(\mathbf{\Omega^{-1}} \quad | \quad & \Lambda^J, \eta^s, \kappa, \theta, \sigma_v, \overline{\rho}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Sigma_y}, y, P, V, J, Z, \mathbf{X}) \sim \\ & \mathcal{W}_3\left(\alpha_\Omega + T, \left\{ \mathbf{R_\Omega}^{-1} + \sum_{t=1}^{T} (\overline{\mathbf{X}}_{t\Delta} - FtQ_{t\Delta} - \mathbf{A}\overline{\mathbf{X}}_{(t-1)\Delta})(\overline{\mathbf{X}}_{t\Delta} - FtQ_{t\Delta} - \mathbf{A}\overline{\mathbf{X}}_{(t-1)\Delta})' \right\}^{-1} \right) \end{aligned}$$

Finally, the last sub-block of Θ to be sampled is for $\overline{\rho}$ and σ_v. To sample these parameters it is easier to sample the full covariance matrix Σ. The full covariance matrix can be sampled from the following target density.

$$\pi(\boldsymbol{\Sigma}|\Lambda^J,\eta^s,\kappa,\theta,\boldsymbol{\Omega},\lambda,\overline{\mu},\mathbf{A},c_1,c_2,\boldsymbol{\Sigma}_{\mathbf{y}},y,P,V,J,Z,\mathbf{X}) \propto \pi(\Sigma) f(P,V,\mathbf{y}|\mathbf{X},J,Z,\Theta,\Lambda^J,\boldsymbol{\Sigma})$$

This distribution is proportional to the prior of Σ times the likelihood function of the data. Note that in this case, the likelihood of the data is dependant on the model under consideration due to the jump process. Therefore the sampler for Σ will vary accordingly.

- SV

  The target density is proportional to:

$$\propto \pi(\Sigma)\prod_{t=1}^{T} f_{\mathcal{N}}\left(\left(\begin{array}{c} P_{t\Delta} \\ V_{t\Delta} \end{array}\right)\middle|\left(\begin{array}{c} (r_{(t-1)\Delta}+\eta^s V_{(t-1)\Delta}-\overline{\lambda}\mu^*)\Delta \\ V_{(t-1)\Delta}+\kappa(\theta-V_{(t-1)\Delta})\Delta \end{array}\right), V_{(t-1)\Delta}\Delta\Sigma\right)$$

  To facilitate sampling of Σ a Wishart prior is selected for Σ^{-1} such that $\Sigma^{-1} \sim \mathcal{W}_2(\alpha_\Sigma,\mathbf{R}_{\boldsymbol{\Sigma}})$. Given this choice, the following result is obtained by conjugate priors.

$$\begin{aligned}
\pi(\boldsymbol{\Sigma}^{-1} \quad | \quad & \Lambda^J,\eta^s,\kappa,\theta,\boldsymbol{\Omega},\lambda,\overline{\mu},\mathbf{A},c_1,c_2,\boldsymbol{\Sigma}_{\mathbf{y}},y,P,V,J,Z,\mathbf{X}) \sim \\
& \mathcal{W}_2\left(\alpha_\Sigma+T,\left\{\mathbf{R}_{\boldsymbol{\Sigma}}{}^{-1}+\sum_{t=1}^{T}\frac{1}{\Delta V_{(t-1)\Delta}}\left(\begin{array}{c} P_{t\Delta}-(r_{(t-1)\Delta}+\eta^s V_{(t-1)\Delta}-\overline{\lambda}\mu^*)\Delta \\ V_{t\Delta}-V_{(t-1)\Delta}-\kappa(\theta-V_{(t-1)\Delta})\Delta \end{array}\right)\right.\right. \\
& \left.\left. \times \left(\begin{array}{c} P_{t\Delta}-(r_{(t-1)\Delta}+\eta^s V_{(t-1)\Delta}-\overline{\lambda}\mu^*)\Delta \\ V_{t\Delta}-V_{(t-1)\Delta}-\kappa(\theta-V_{(t-1)\Delta})\Delta \end{array}\right)'\right\}^{-1}\right)
\end{aligned}$$

- SVJ

$$\begin{aligned}
\pi(\boldsymbol{\Sigma}^{-1} \quad | \quad & \Lambda^J,\eta^s,\kappa,\theta,\boldsymbol{\Omega},\lambda,\overline{\mu},\mathbf{A},c_1,c_2,\boldsymbol{\Sigma}_{\mathbf{y}},y,P,V,J,Z,\mathbf{X}) \sim \\
& \mathcal{W}_2\left(\alpha_\Sigma+T,\left\{\mathbf{R}_{\boldsymbol{\Sigma}}{}^{-1}+\sum_{t=1}^{T}\frac{1}{\Delta V_{(t-1)\Delta}}\left(\begin{array}{c} P_{t\Delta}-(r_{(t-1)\Delta}+\eta^s V_{(t-1)\Delta}-\overline{\lambda}\mu^*)\Delta-Z^s_{t\Delta} \\ V_{t\Delta}-V_{(t-1)\Delta}-\kappa(\theta-V_{(t-1)\Delta})\Delta \end{array}\right)\right.\right. \\
& \left.\left. \times \left(\begin{array}{c} P_{t\Delta}-(r_{(t-1)\Delta}+\eta^s V_{(t-1)\Delta}-\overline{\lambda}\mu^*)\Delta-Z^s_{t\Delta} \\ V_{t\Delta}-V_{(t-1)\Delta}-\kappa(\theta-V_{(t-1)\Delta})\Delta \end{array}\right)'\right\}^{-1}\right)
\end{aligned}$$

- SVCJ

$$
\begin{aligned}
\pi(\mathbf{\Sigma^{-1}} \quad | \quad & \Lambda^J, \eta^s, \kappa, \theta, \mathbf{\Omega}, \lambda, \overline{\mu}, \mathbf{A}, c_1, c_2, \mathbf{\Sigma_y}, y, P, V, J, Z, \mathbf{X}) \sim \\
& \mathcal{W}_2 \left(\alpha_\Sigma + T, \left\{ \mathbf{R_\Sigma}^{-1} + \sum_{t=1}^{T} \frac{1}{\Delta V_{(t-1)\Delta}} \left(\begin{array}{c} P_{t\Delta} - (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta - Z^s_{t\Delta} \\ V_{t\Delta} - V_{(t-1)\Delta} - \kappa(\theta - V_{(t-1)\Delta})\Delta - Z^v_{t\Delta} \end{array} \right) \right. \right. \\
& \left. \left. \times \left(\begin{array}{c} P_{t\Delta} - (r_{(t-1)\Delta} + \eta^s V_{(t-1)\Delta} - \overline{\lambda}\mu^*)\Delta - Z^s_{t\Delta} \\ V_{t\Delta} - V_{(t-1)\Delta} - \kappa(\theta - V_{(t-1)\Delta})\Delta - Z^v_{t\Delta} \end{array} \right)' \right\}^{-1} \right)
\end{aligned}
$$

Note, that once the covariance matrix has been sampled it must be normalized so that $\sigma_{1,1} = 1$. The Σ matrix is not jointly identified with $V_{t\Delta}$ unless the $\sigma_{1,1}$ is fixed. In the Wishart sampler identification is weakly imposed through the selection of an appropriately restricted prior ($\sigma_{1,1}$ centered around 1). However, due to Gibbs sampling of the Wishart distribution the $\sigma_{1,1}$ element will sample to the prior mean value with some variance around the estimates. Normalization is performed as a way of enforcing the identification, $\sigma_{1,1} = 1$. Nonetheless this methodology has a scaling effect on the posterior distribution of Σ. The normalization approach used is such that, on average, the scaling factor of the latent volatilities is equal to one and does not hinder Markov Chain convergence. An alternative to this methodology is to sample the parameters $\overline{\rho}$ and σ_v directly using a Metropolis Hastings (MH) sub-block. This approach was tested for this model achieving comparable results at the cost of sampling efficiency. Given the low variance around 1 for $\sigma_{1,1}$ draws, the Gibbs-Wishart updating is favored over the MH approach from an operational standpoint. Finally, an alternative method was suggested by McCulloch et al. (2000) in which the algorithm is evaluated without scaling of the restricted parameter. The distributions of parameters and factors are then scaled ex-post. This methodology would be applicable to the SV and SVJ models. However, implementation is not feasible for specifications where the volatility diffusion process is augmented by a jump process such as in the SVCJ model.

6. Last to be sampled is the latent volatilities $V_{t\Delta}$ for all time periods $t = 1$ to T. We begin by defining the following vectors, $\Psi_{t\Delta} = \{\mathbf{y_{t\Delta}}', P_{t\Delta},\}'$ and $\Xi_{t\Delta} = \left\{J_{t\Delta}, Z^s_{t\Delta}, Z^v_{t\Delta}, \mathbf{X_{t\Delta}}, \mathbf{X_{(t-1)\Delta}}\right\}$. Taking advantage of this notation we can directly construct the following joint distribution

for volatility at times 1 and 0 (i.e. V_1, V_0).

$$\pi(V_0, V_1, \Psi_1|\Lambda^J, \Theta, \Xi_1) = \pi(V_0|\Lambda^J, \Theta)\pi(V_1, \Psi_1|V_0, \Lambda^J, \Theta, \Xi_1)$$

This logic can be extended to obtain the joint of V_2, V_1 and V_0.

$$\pi(V_0, V_1, V_2, \Psi_1, \Psi_2|\Lambda^J, \Theta, \Xi_1, \Xi_2) =$$
$$\pi(V_0|\Lambda^J, \Theta)\pi(V_1, \Psi_1|V_0, \Lambda^J, \Theta, \Xi_1)\pi(V_2, \Psi_2|V_1, \Psi_1, V_0, \Lambda^J, \Theta, \Xi_1, \Xi_2)$$

In general:

$$\pi(V_0, V_1, \ldots, V_T, \Psi_1, \ldots, \Psi_T|\Lambda^J, \Theta, \Xi_1, \ldots, \Xi_T) =$$
$$\pi(V_0|\Lambda^J, \Theta)\prod_{t=1}^{T}\pi(V_{t\Delta}, \Psi_{t\Delta}|V_{(t-1)\Delta}, \ldots, V_1, \Psi_{(t-1)\Delta}, \ldots, \Psi_1, V_0, \Lambda^J, \Theta, \Xi_t, \ldots, \Xi_1)$$

To improve sampling efficiency in volatility, it was determined that a change of variable to the square root of $V_{t\Delta}$ was helpful. Efficiency is improved, somewhat unorthodoxly, by matching the target density to the proposal density. Analysis of a low-dimensional conditional target density of volatility sub-blocks reveals that the this density function exhibits left side skewness which cannot be matched by symmetric proposals like normal or student-t densities. In the effort to obtain a better match of target to proposal a variable change is required which will impart more symmetry unto the volatility likelihood. Three variable transformations were tested: $\sqrt{V_{t\Delta}}, \ln V_{t\Delta}$ and $\frac{1}{V_{t\Delta}}$. The $\frac{1}{V_{t\Delta}}$ was found to increase the left side skew while the $\ln V_{t\Delta}$ was found to introduce right skewness. The $\sqrt{V_{t\Delta}}$ was found to impart the most symmetry to the target density. Once the transformation was performed, a student-t proposal proved very effective as the proposal density for latent volatility sampling. The following target density, reflects the variable change.

$$\pi(V_0, \sqrt{V_1}, \ldots, \sqrt{V_T}, \Psi_1, \ldots, \Psi_T|\Lambda^J, \Theta, \Xi_1, \ldots, \Xi_T) \propto$$
$$\prod_{t=1}^{T} 2\sqrt{V_{t\Delta}} \times \pi(V_{t\Delta}, \Psi_{t\Delta}|V_{(t-1)\Delta}, \ldots, V_1, \Psi_{(t-1)\Delta}, \ldots, \Psi_1, V_0, \Lambda^J, \Theta, \Xi_t, \ldots, \Xi_1)$$

This last joint distribution can not be sampled directly due to computational limitations in the number of observations, T. However, it is possible to break up the vector and sample it piecewise using Randomized MCMC Blocking as described by Pitt and Shephard (1997). First, subsets are generated of random sizes to improve MCMC mixing. Second, the joint likelihood in the subset becomes the target density, to obtain a proposal from which to draw; a line search method is used to find the peak of the function. This peak and the negative inverse Hessian at that point are utilized to construct a normal joint proposal from which to make draws. A standard accept reject algorithm is used in the end to determine if the draws are acceptable.

B. Thought Experiment on Volatility Inference

Consider the following simulated data-based test of the estimation algorithm of a (Markov) stochastic volatility model[12]. Let's assume that we simulated 10,000 data points, and asked the following question: given a perfect knowledge of time-homogeneous parameters, the initial vol at time zero, volatility at time $t = 11$: v_{11}, and index returns between times $t = (1, ..., 11)$, what is the posterior distribution of the vector of first ten latent volatilities $V_{1,10}$? Next, ask the same question, but allow conditioning on all 10,000 return observations. The conditional joint posterior distribution of $V_{1,10}$ in both cases will be identical. The fact that the latent volatility factor drives the diffusion component of the process exacerbates the lack of inference in estimation. Therefore, we conclude that *regardless of estimation method used* the precision of our inference about latent volatilities is bounded and cannot be improved without limit by simply increasing the number of time-series observations (like we typically can for other time-invariant model unknowns). This argument still works by induction even if, as is usually the case, we can't condition on volatility v_{11} during estimation. Indeed, more time-series return data would improve our inference about v_{11}, but the best case scenario would be a perfect knowledge of v_{11}, which, as we already know, still results in limited amount of inference about the preceding vector of latent volatilities $V_{1,10}$. The limit on our inference ability appears to be driven by the magnitude of volatility in the volatility process, which effectively functions as a prior distribution for individual volatilities: the less volatile the unobserved process happens to be, the more precisely we can estimate its individual realizations. Lower *volatility level*, leading to lower noise in observed return data, will also result in more accurate inference on individual latent volatilities. The arguments presented above would imply that the volatility process of all stochastic volatility models is already so highly parameterized that no matter which estimation method we use, and no matter how perfectly specified our model is, and no matter how much data we use, we can't obtain anything better than a noisy signal on latent volatilities. Therefore, it would seem counterintuitive to continue increasing the complexity of volatility parametrization by augmenting it with a time-dependent volatility jump process.

[12] Think of a Heston model for simplicity, or a version of it we call SV below

C. Tables and Figures

MCMC Posterior Means, Standard Deviations, and Inefficiency Factors for all Model Parameters

Model Parameter	SV Mean	SV StDev	SV Ineff	SVJ Mean	SVJ StDev	SVJ Ineff	SVCJ Mean	SVCJ StDev	SVCJ Ineff
$\overline{\rho}$	-0.95	0.007	11.67	-0.95	0.003	4.63	0.95	0.005	5.34
θ	0.05	0.003	4.46	0.05	0.003	4.10	0.04	0.002	9.57
κ_v	2.63	0.18	2.06	2.32	0.14	1.96	2.37	0.15	10.70
σ_v	0.44	0.008	21.61	0.38	0.004	6.12	0.34	0.004	7.71
$\bar{\lambda}$	N/A	N/A	N/A	1.86	0.29	12.03	2.25	0.32	15.68
σ_s	N/A	N/A	N/A	0.04	0.001	6.43	0.04	0.001	11.90
μ_s	N/A	N/A	N/A	-0.03	0.002	4.58	-0.011	0.002	10.63
μ_s^*	N/A	N/A	N/A	-0.04	0.001	3.33	-0.02	0.001	4.51
μ_v	N/A	N/A	N/A	N/A	N/A	N/A	0.04	0.002	7.09
ρ_J	N/A	N/A	N/A	N/A	N/A	N/A	-0.53	0.04	6.16
η_s	0.57	0.02	1.00	0.57	0.02	1.00	0.57	0.02	1.00
c_1	0.001	0.0002	1.00	0.001	0.0002	1.00	0.001	0.0002	1.00
c_2	0.006	0.001	1.00	0.006	0.001	1.04	0.006	0.001	1.00
λ	0.64	0.003	1.37	0.64	0.003	1.26	0.64	0.003	1.40
$\overline{\mu_1}$	0.05	0.003	1.27	0.05	0.003	1.27	0.05	0.003	1.22
$\overline{\mu_2}$	-0.04	0.004	1.28	-0.04	0.004	1.30	-0.04	0.004	1.20
$\overline{\mu_3}$	-0.08	0.01	1.99	-0.08	0.01	1.84	-0.08	0.01	1.98
$a_{1,1}$	0.99	0.002	1.06	0.99	0.002	1.05	0.99	0.002	1.15
$a_{1,2}$	0.01	0.002	1.02	0.01	0.002	1.05	0.01	0.002	1.08
$a_{2,1}$	0.001	0.0002	1.00	0.001	0.0002	1.00	0.001	0.0002	1.03
$a_{2,2}$	0.997	0.0005	1.12	0.997	0.0005	1.15	0.997	0.0005	1.18
$a_{3,3}$	0.997	0.001	1.19	0.997	0.001	1.94	0.997	0.001	1.96

Table 1: This table summarizes daily data based parameter estimates using 20,000 MH multi-core iterations with 50% burn-in. The top half of the table lists parameters driving the equity process, while the bottom half is reserved for the parameters of the risk free rate term-structure. Posterior summaries for the elements of $\mathbf{\Sigma_y}$ (ranging between 1 and 30 bps) and $\mathbf{\Omega}$ matrices are omitted, but they are available from the authors upon request, similar to prior hyperparameters used in estimation and any other tuning parameters or standard output diagnostics. All inefficiency factors are calculated based on the recepie in Carlin (2000) using the first 5% autocorrelation cut-off.

DIC and Equity Risk Premiums

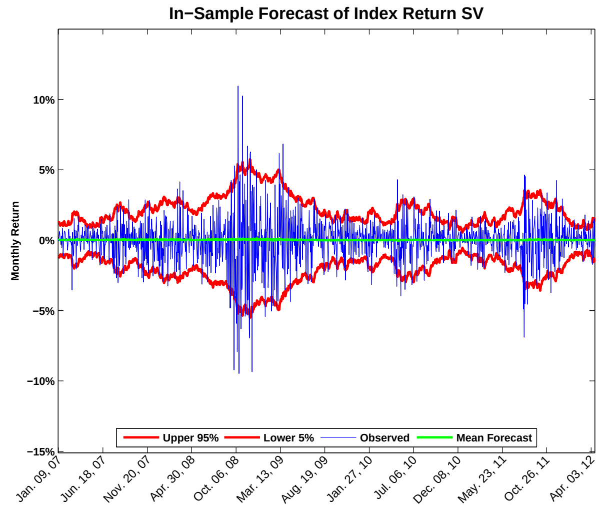
Model	DIC	Diffusion Premium	Jump Premium	Equity Premium Combined
SV	-164,424	3.0%	0	3.0%
SVCJ	-164,479	4.2%	2.0%	6.2%
SVJ	-164,469	2.6%	1.4%	4.0%

Table 2: This table reports the Deviance Information Criteria (DIC) used to compare in-sample fit of all three models. In addition the last 3 columns break down the total equity risk premium (last column) into its constituents: diffusion and jump components. Note that diffusion premium was calculated using steady-state value of the volatility process, using the formula provided in Eraker (2004).

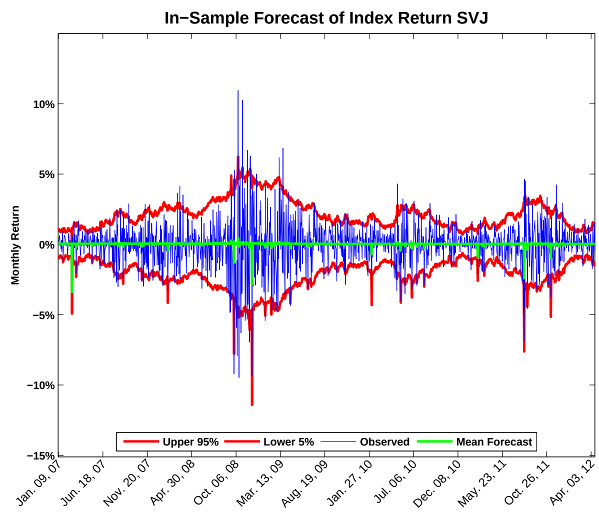
Summary Of Volatility Inefficiency Factors

Model	Mean	St Dev	Max	Skew
SV	119.2	46.0	232.8	0.2
SVJ	156.2	58.7	326.4	0.5
SVCJ	177.1	69.6	346.2	0.7

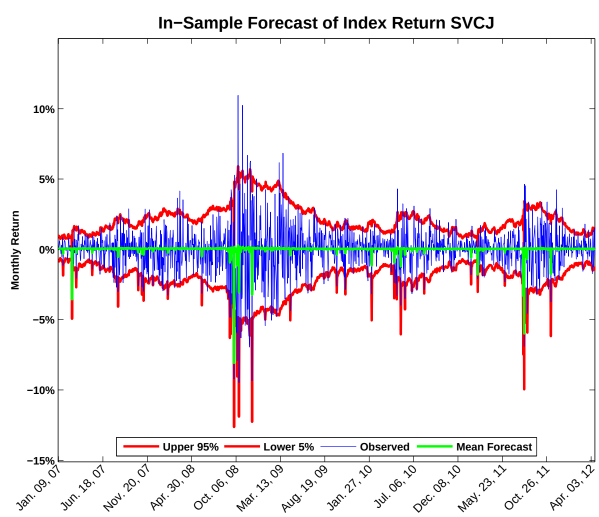
Table 3: We compute inefficiency factors for volatility measure V_t at each point in time. This table provides distribution summary of these inefficiency factors. SV has clearly the lowest average inefficiency, standard deviation of inefficiency factors over time, and the max. SVJ model is second best.



(a) SV

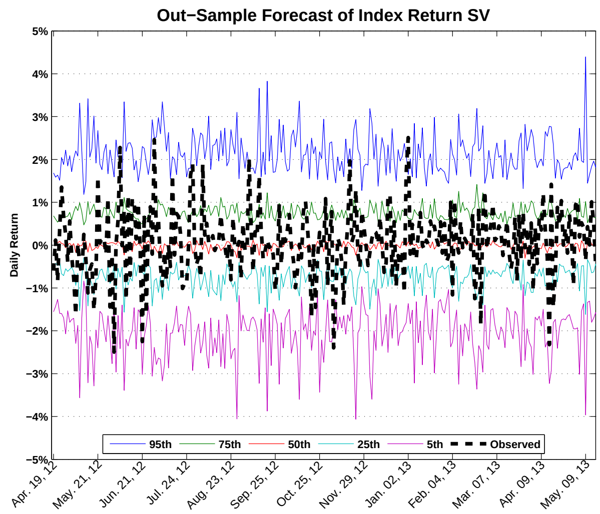


(b) SVJ

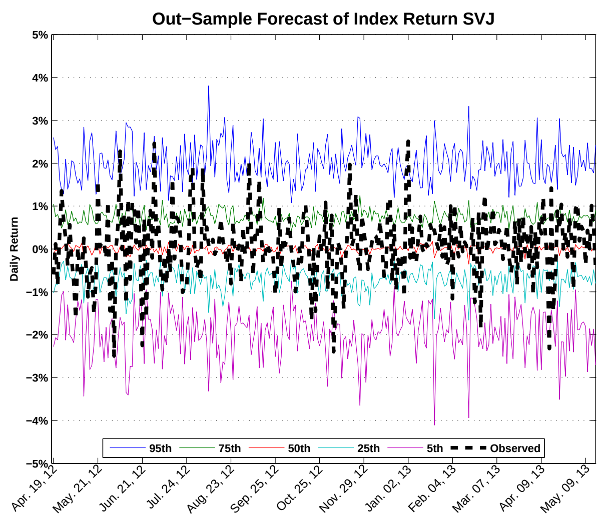


(c) SVCJ

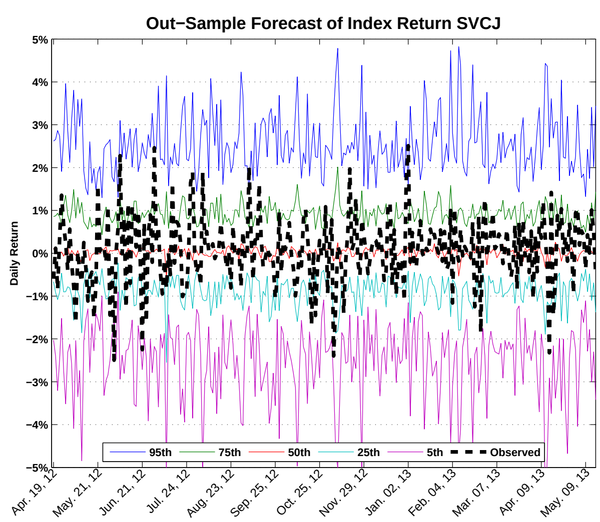
Figure 1. In-sample fit of index returns. Blue line is actual, green is estimated mean, and red line marks 5th and 95th percentile range.



(a) SV



(b) SVJ



(c) SVCJ

Figure 2. Out-of-sample 13-month forecast of daily index returns. Black dotted line is actual return realization, solid lines denote various percentiles, as marked.

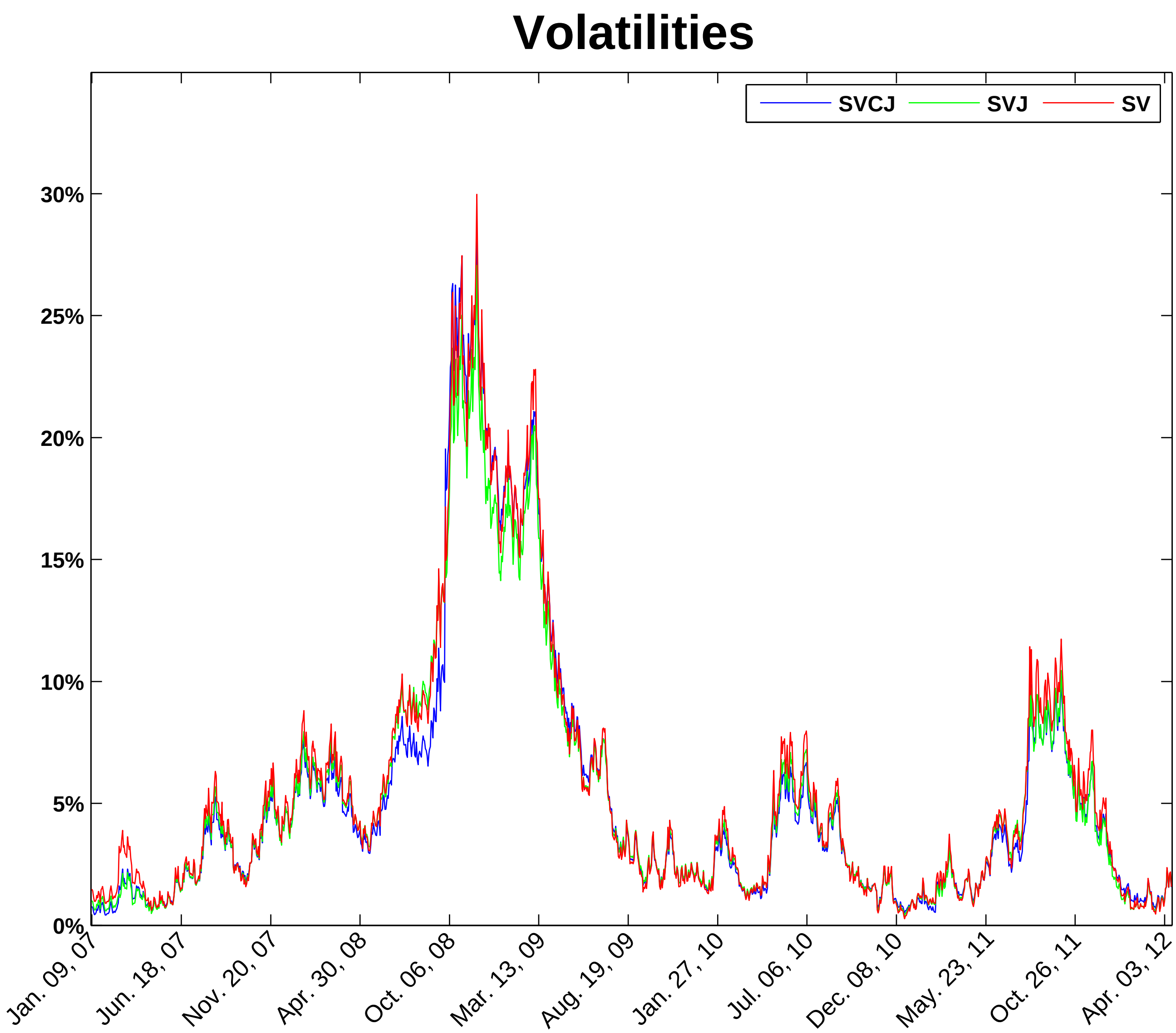


Figure 3. Filtered mean volatilities for all models.

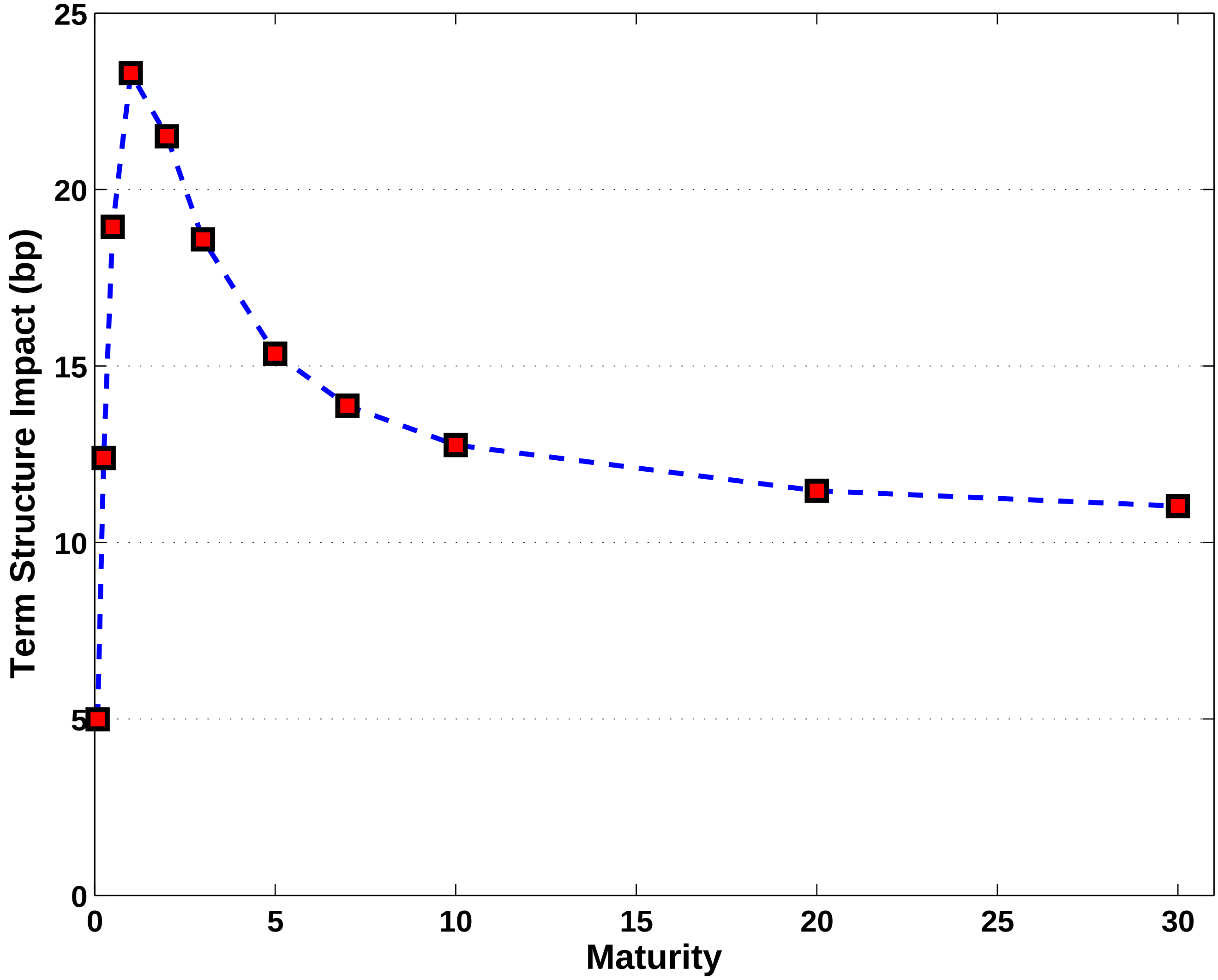
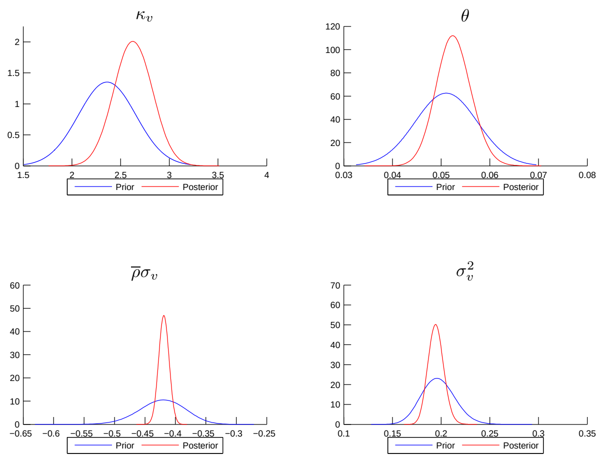
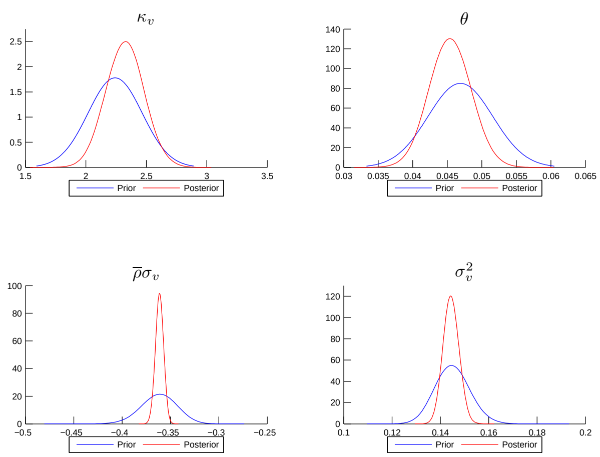


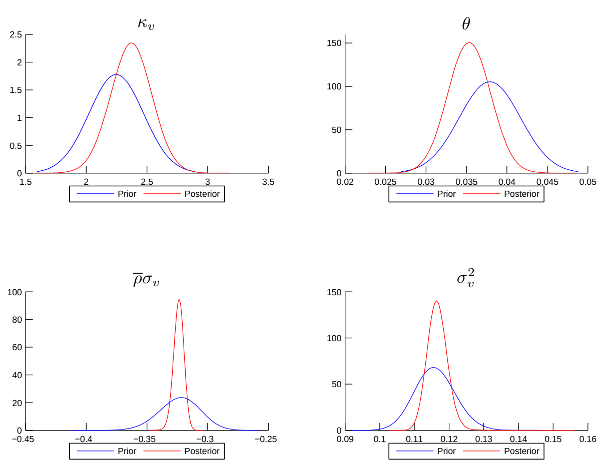
Figure 4. This figure displays the average impact Flight-to-Quality structure has on the risk free yield curve following 1% increase in equity. It shows that market participants move money in (out) of equity from (to) treasuries each day. Mid-maturity yields are affected the most.



(a) SV

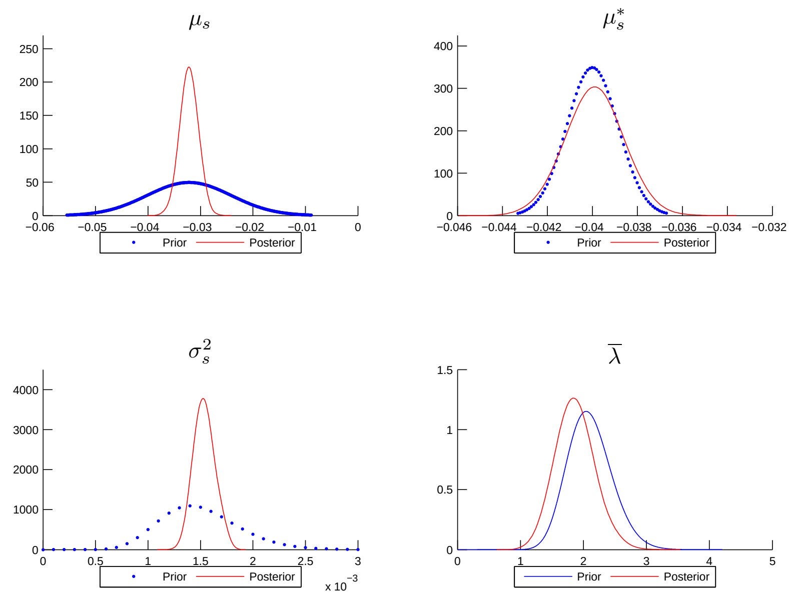


(b) SVJ

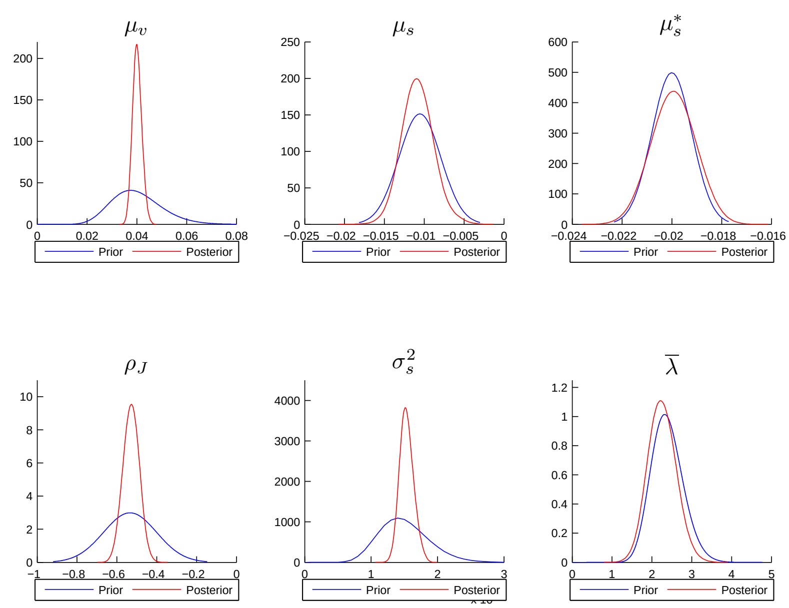


(c) SVCJ

Figure 5. Prior-Posterior update of stochastiv volatility evolution equation parameters.

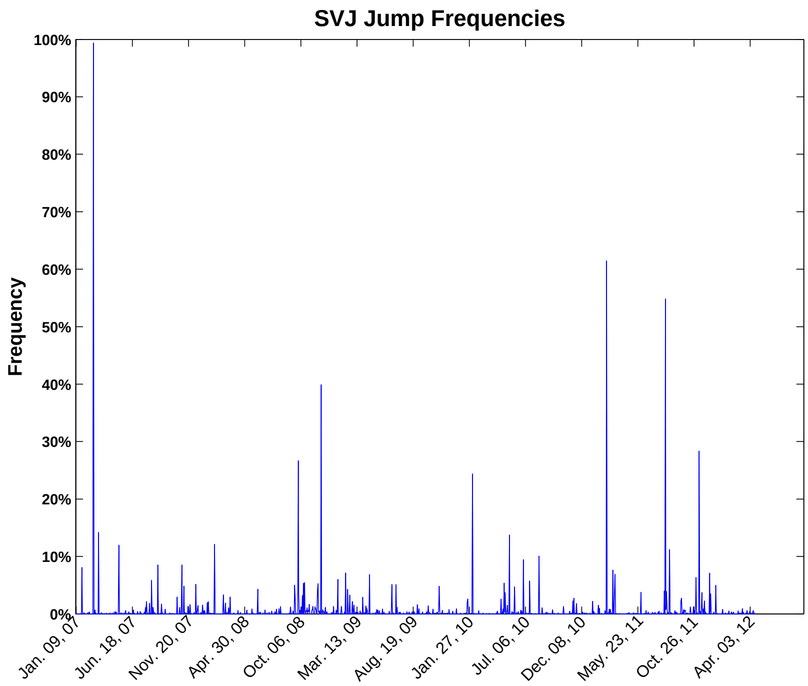


(a) SVJ

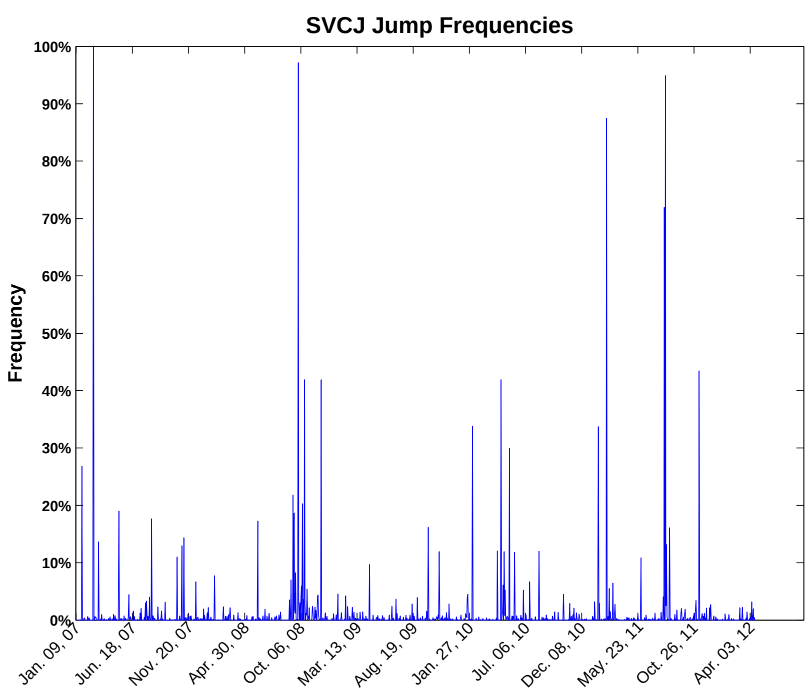


(b) SVCJ

Figure 6. Prior-Posterior update of jump parameters.

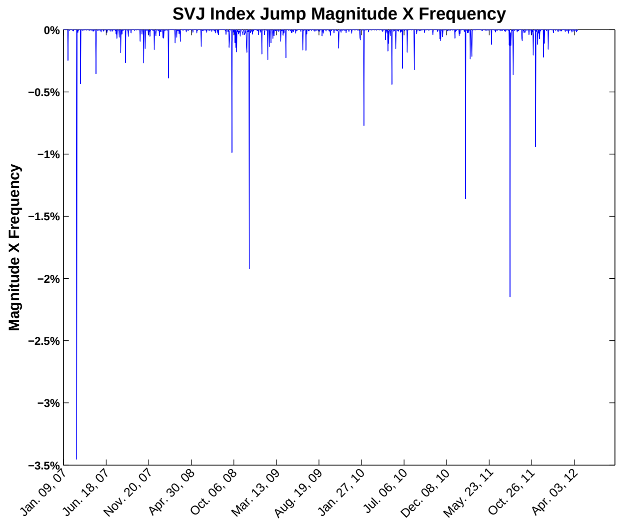


(a) SVJ

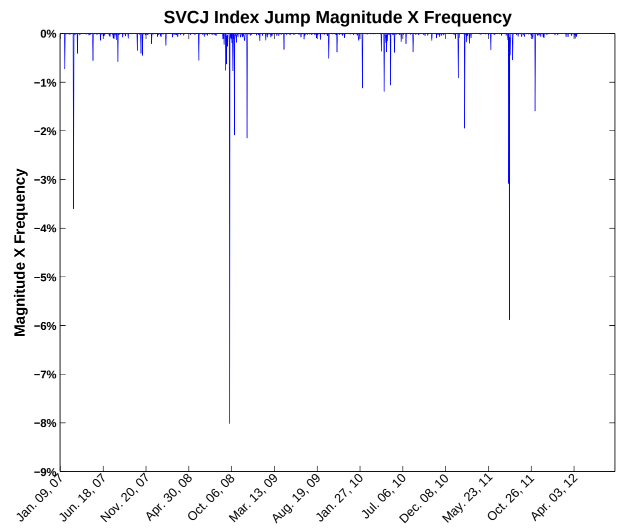


(b) SVCJ

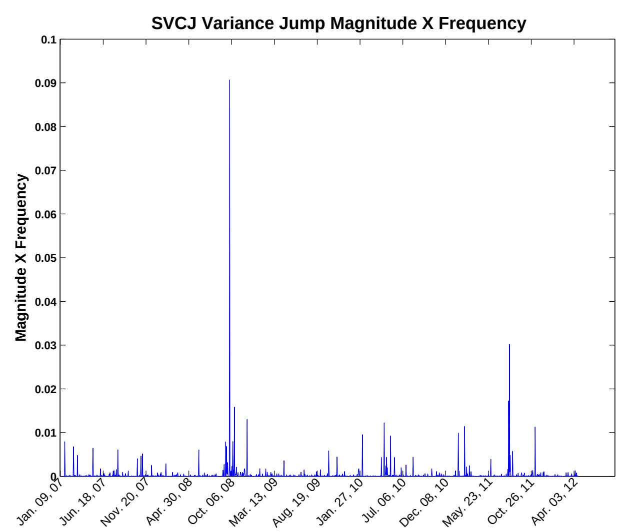
Figure 7. Posterior jump probabilities.



(a) SVJ



(b) SVCJ Equity



(c) SVCJ Volatility Measure

Figure 8. Posterior mean jump sizes times jump occurence.

References

Andersen, Torben and Luca Benzoni, "Realized Volatility," 2008. Working Paper, Federal Reserve Bank of Chicago.

_ **, Hyung-Jin Chung, and Bent Sorensen**, "Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study," *Journal of Econometrics*, 1999, *1*, 6187.

_ **, Luca Benzoni, and Jesper Lund**, "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 2002, *3*, 1239–1284.

Bates, David, "Post-87 Crash Fears in S&P 500 Futures Options," *Journal of Econometrics*, 2000, *94*, 181–238.

Belaygorod, Anatoliy and Atilio Zardetto, "Joint Regime Changing Jump-Diffusion Model of S&P and VIX," 2013. Working Paper.

Berg, Andreas, Renate Meyer, and Jun Yu, "Deviance Information Criterion for comparing stochastic volatility models," *Journal of Business & Economic Statistics*, 2004, *22*, 107–120.

Bollerslev, Tim and Victor Todorov, "Tails, Fears, and Risk Premia," *Journal of Finance*, 2011, *66*, 2165–2211.

_ **, Tzuo Law, and George Tauchen**, "Risk, jumps, and diversification," *Journal of Econometrics*, 2008, *144*, 234–256.

_ **, Uta Kretschmer, Christian Pigorsch, and George Tauchen**, "A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects," *Journal of Econometrics*, 2009, *150*, 151–166.

Broadie, Mark, Mikhail Chernov, and Michael Johannes, "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance*, 2007, *3*, 1453–1490.

_ **, _ , and _** , "Understanding Index Option Returns," *Review of Financial Studies*, 2009, *22*, 4493–4529.

Carlin, Bradley, *Bayes and Empirical Bayes Methods for Data Analysis*, 2 ed., Chapman and Hall, 2000.

Chernov, Mikhail, Ronald Gallant, Eric Ghyselsb, and George Tauchen, "Alternative models for stock price dynamics," *Journal of Econometrics*, 2003, *116*, 225–257.

Chib, Siddhartha, "Markov Chain Monte Carlo Methods: Computation and Inference," *Handbook of Econometrics,*, 2001, *5*, 3569–3648.

_ **and Edward Greenberg**, "Bayes Inference in Regression Models With ARMA (p,q) Errors," *Journal of Econometrics*, 1994, *64*, 183–206.

_ **and** _ , "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 1995, *49* (4), 327–335.

Christensen, Jens, Frank Diebold, and Glenn Rudebusch, "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," *Journal of Econometrics*, 2011, *164*, 4–20.

Christoffersen, Peter, Kris Jacobs, and Karim Mimouni, "Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns, and Option Prices," *Review of Financial Studies*, 2010, *23*, 3141–3189.

Diebold, Francis X. and Canlin Li, "Forecasting the term structure of government bonds," *Journal of Econometrics*, 2006, *130*, 337–364.

_ **, Glenn D. Rudebusch, and S. Boragan Aruoba**, "The macroeconomy and the yield curve: a dynamic latent factor approach," *Review of Financial Studies*, 2005, *forthcoming.*

Duan, Jin-Chuan and Chung-Ying Yeh, "Jump and volatility risk premiums implied by VIX," *Journal of Economic Dynamics and Control*, 2010, *34*, 2232–2244.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, "Transform Analysis and Asset Pricing for Affine Jump-diffusions," *Econometrica*, 2000, *68*, 1343–1376.

Durham, Garland, "SV mixture models with application to S&P 500 index returns," *Journal of Financial Economics*, 2007, *3*, 822–856.

Eraker, Bjorn, "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 2004, *3*, 1367–1403.

_ **, Michael Johannes, and N. Polson**, "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 2003, *3*, 1269–1298.

Forbes, Catherine, Gael Martin, and Jill Wright, "Inference for a class of stochastic volatility models using option and spot prices: application of a bivariate kalman filter," *Econometric Reviews*, 2007, *26*, 387–418.

Garcia, Rene, "Econometrics of Option Pricing," *Encyclopedia of Quantitative Finance*, 2010.

Gatheral, Jim, *The Volatility Surface: A Practitioner's Guide*, 1 ed., John Wiley and Sons, Inc, 2006.

Heston, Steven, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, 1993, *6*, 327–343.

Jacquier, Eric, Nicholas Polson, and Peter Rossi, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 1994, *12*, 371–389.

Jensen, Mark and John Maheu, "Bayesian semiparametric stochastic volatility modeling," *Journal of Econometrics*, 2010, *157*, 306–316.

Jiang, George and Roel Oomen, "Estimating Latent Variables and Jump Diffusion Models Using High Frequency Dat," *Journal of Financial Econometrics*, 2007, *5*, 130.

Johannes, Michael and Nicholas Polson, "MCMC Methods for Continuous-Time Financial Econometrics," *Handbook of Financial Econometrics*, 2009, *2*.

_ , _ , and Jonathan Stroud, "Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices," *Review of Financial Studies*, 2009, *22*, 2759–2799.

Jones, Christopher, "The dynamics of stochastic volatility: evidence from underlying and options markets," *Journal of Econometrics*, 2003, *116*, 181–224.

Lahaye, Jerome, Sebastien Laurent, and Christopher Neelye, "Jumps, cojumps and macro announcements," *Journal of Applied Econometrics*, 2011, *26*, 893–921.

McCulloch, Robert, Peter Rossi, and Peter Rossi, "A Bayesian analysis of the multinomial probit model with fully identi"ed parameters," *Journal of Econometrics*, 2000, *99*, 173–193.

Pan, Jun, "The jump-risk premia implicit in option prices: evidence from an integrated time-series study," *Journal of Financial Economics*, 2002, *63*, 3–50.

Pitt, Michael and Neil Shephard, "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, 1997, *84*, 653–667.

Raggi, David and Silvano Bordignon, "Comparing stochastic volatility models through Monte Carlo simulations," *Computational Statistics & Data Analysis*, 2006, *50*, 1678–1699.

Rossi, Alberto and Allan Timmermann, "What is the Shape of the Risk-Return Relation?," 2011. Working Paper, UCSD.

Takahashi, Makoto, Yasuhiro Omori, and Toshiaki Watanabe, "Estimating stochastic volatility models using daily returns and realized volatility simultaneously," *Computational Statistics and Data Analysis*, 2009, *53*, 2404–2426.

Todorov, Viktor, "Econometric analysis of jump-driven stochastic volatility models," *Journal of Econometrics*, 2011, *160*, 181–224.

_ , George Tauchen, and Iaryna Grynkiv, "Realized Laplace transforms for estimation of jump diffusive volatility models," *Journal of Econometrics*, 2011, *164*, 367–381.

Yan, Shu, "Jumprisk, stockreturns, and slope of implied volatility smile," *Journal of Financial Economics*, 2011, *99*, 216–233.

Yu, Cindy, Haitao Li, and Martin Wells, "MCMC estimation of lvy jump models using stock and option prices," *Mathematical Finance*, 2011, *21*, 383–422.