Fast Filtering with Large Option Panels: Implications for Asset Pricing*

Arnaud Dufays, Kris Jacobs, Yuguo Liu, and Jeroen Rombouts

March 24, 2022

Abstract

The cross-section of options holds great promise for identifying return distributions and risk premia, but estimating dynamic option valuation models with latent state variables is challenging when using large option panels. We propose a particle MCMC framework with a novel filtering approach and illustrate our method by estimating workhorse index option pricing models. Estimates of the variance risk premium, variance mean reversion, and higher moments differ from the literature. We show that these differences are due to the composition of the option sample. Restrictions on the option sample's maturity dimension have the strongest impact on parameter inference and option fit in these models.

JEL Classification: G12

Keywords: Option Valuation; Particle MCMC; Posterior Density; Option Panels; Risk Premia.

*Dufays: EDHEC; Jacobs and Liu: University of Houston; Rombouts: ESSEC; Correspondence to Kris Jacobs, Bauer College of Business, University of Houston, kjacobs@bauer.uh.edu. We are very grateful to the Canadian Derivatives Institute (CDI) for financial support. We would like to thank Torben Andersen, Peter Christoffersen, Francois Cocquemas, Toby Daglish, Christian Dorion, Hitesh Doshi, Jin Duan, Bjorn Eraker, Andras Fulop, Christian Gourieroux, Rene Garcia, Mohammad Ghaderi, Bruce Grundy, Jinji Hao, Junye Li, Scott Murray, Federico Nardari, Chay Ornthanalai, Manuela Pedio, Nick Polson, Olivier Scaillet, David Schreindorfer, Gustavo Schwenkler, Sang Seo, Thijs Van Der Heijden, Viktor Todorov, Bas Werker, James Yae, Morad Zekhnini, and participants in seminars at Georgia State, George Washington, West Virginia, the Universities of Melbourne and Wellington, UNSW, the SFS Cavalcade, the ESSEC workshop on Bayesian methods in Finance, and the Vienna Workshop on the Econometrics of Option Markets for helpful discussions and comments.

1 Introduction

The cross-section of options holds great promise for identifying the return distribution and conditional risk premia, because it contains information on many different states of nature. The option pricing literature has traditionally relied on dynamic models with latent state variables to model option prices and risk premia.[1] However, there is widespread agreement in the option valuation literature that relatively complex models with stochastic volatility and multiple volatility factors as well as jumps and tail factors are needed to explain the data.[2] Moreover, long time series are needed to reliably identify the volatility dynamics (Broadie, Chernov, and Johannes, 2007) and options with different moneyness and maturity may be needed to identify various model aspects. Finally, to learn about risk premia and the structure of the stochastic discount factor, we need to estimate using both returns and options. Given these requirements and the richness of the available option data, even the estimation of relatively simple dynamic option valuation models is computationally very demanding.

To illustrate the computational challenge resulting from large option panels, consider estimation based on all available option contracts included in the OptionMetrics data used in many recent studies. Based on data from 1996 to the end of 2015, we have more than 5000 trading days for which option contracts are available. The number of available index option series per day has significantly increased over time, but on average it is approximately 200. The full sample of option contracts based on daily OptionMetrics data for 1996-2015 therefore consists of roughly one million contracts.

Even much smaller, judiciously chosen, option samples may impose significant computa-

[1] More recently, Jones (2006), Israelov and Kelly (2017), and Buchner and Kelly (2022) use cross-sectional techniques that are more akin to the factor models used to study the cross-section of stock returns.

[2] See Bates (2000) for evidence on multiple volatility factors. See Eraker (2004), Broadie, Chernov, and Johannes (2007), Pan (2002), Aït-Sahalia, Cacho-Diaz, and Laeven (2015), Fulop, Li, and Yu (2015), Andersen, Fusari, and Todorov (2017), and Bates (2012) for evidence on jumps in returns and volatility. See Andersen, Fusari, and Todorov (2015b) for evidence on a tail factor in option prices.

tional constraints. At least partly due to this computational cost, existing studies typically reduce the dimension of the data to make estimation computationally feasible.[3] In the time-series dimension, rather than using a short time series, often only one day per week is used (Bates, 2000, p.195). In the cross-sectional dimension, the sample is typically restricted to a small subset of the available option contracts. Some studies impose very stringent constraints on the cross-section and are effectively based on short-maturity at-the-money options.

The first contribution of this paper is to adapt existing estimation methods to overcome these computational constraints. We propose an orthogonal particle Markov Chain Monte Carlo (MCMC) framework to estimate dynamic option valuation models using long time series and large panels of option data. We apply an MCMC algorithm for exploring the parameter space, but we filter the latent state variables using the particle filter (Johannes, Polson, and Stroud, 2009; Johannes and Polson, 2009b). In general, particle filtering is subject to significant computational constraints when assigning weights based on the model option price, but we address this with a novel filtering approach, featuring particle weights based on the quantiles of the filtered distribution. This greatly reduces the dependence of computation time on the number of particles. This innovation enables us to estimate and test models using very large option panels. Because particle filtering using returns is computationally fast, we are also able to estimate option pricing models using returns and large option panels jointly, which facilitates inference on model parameters characterizing tail events and risk premia. We conduct a Monte Carlo experiment to demonstrate that our approach is reliable.

Our approach can in principle be applied to any option pricing model with stochastic volatility.[4] While it can therefore be used to estimate models with multiple stochastic

[3]Option samples may also be restricted for other reasons. For example, illiquid option contracts should be excluded from the sample. But regardless of these other reasons, it is well known and appreciated in the literature that computational constraints often necessitate restricting the sample size. For instance, Broadie, Chernov, and Johannes (2007, p.1461) state that "computational burdens severely constrain how much and what type of data can be used". See also Hurn, Lindsay, and McClelland (2015), Israelov and Kelly (2017), and Andersen, Fusari, and Todorov (2015a) for discussions of computational constraints and cost.

[4]With minor modifications, it can also be applied to the estimation of other dynamic models with latent

volatility factors, we limit our implementation to models with one volatility factor because the computational cost of multiple factors is higher. Rather than finding the model with the best possible fit, we want to instead emphasize the role of the time-series and cross-sectional dimension of the sample. We therefore keep those as large as possible, while using a model that captures some of the most important features of the data. Specifically, we implement our estimation method on a well-known class of models, the Heston (1993) square root and Duffie, Pan, and Singleton (2000) double-jump models. These models have some well-known shortcomings, but they provide an adequate fit and are very useful to illustrate the benefits of our approach because they have been extensively studied in the existing literature.[5] We use these existing estimates to highlight how results may differ when including larger option panels in estimation, and how restricting the sample can affect inference.[6]

In our empirical analysis, we use daily data for 1996-2015 and a large benchmark option sample that is balanced cross-sectionally and over time. The main differences and similarities between our parameter estimates and those in the available literature are as follows. We confirm the findings of Broadie, Chernov, and Johannes (2007), Eraker (2004) and Eraker, Johannes, and Polson (2003) that both jumps in variance and jumps in returns, and co-jumps in particular, matter. Estimates of the correlation between returns and variance are more negative than the existing literature, with the notable exception of Andersen et al. (2015a,b, 2020). The estimate of the variance of variance parameter based on options is smaller than the estimate based on returns, which means that the option-implied variance path is smoother than the return-implied variance path. Our estimates of the price of

state variables.

[5]Several authors have suggested model improvements such as multiple volatility components, tail factors and alternative specifications of the jump process such as time-varying jump intensities. Moreover, Bates (2019) shows that modeling of intraday returns may improve model fit. Andersen et al. (2015b) use estimates of diffusive intraday volatility obtained using high-frequency data and Bandi and Reno (2016) use high-frequency data to estimate models with jumps in returns and volatility.

[6]The problems in the literature are illustrated by the fact that even this relatively simple model is typically not estimated in its full generality using large option panels. Andersen, Fusari, and Todorov (2015a, p.1088) also note this and proceed using a different approach, which is discussed below.

diffusive variance risk are highly statistically significant and smaller than existing estimates. A related finding is that estimates of variance persistence exceed existing estimates.

The second contribution of this paper is to relate some of the differences between our estimates of these models and existing estimates to the composition of the option sample. We document the implications of restricting the option panels used in estimation, and the role of the composition of the option sample in general. To the best of our knowledge we are the first to explicitly explore this topic. The majority of existing studies impose some restrictions, but the literature does not contain evidence on how these restrictions affect inference. This complicates comparisons across studies, as well as parameter and model inference. Note that we are able to investigate this issue because our approach significantly reduces the computational burden, which allows us to obtain estimates from large option samples that serve as benchmarks and to compare them with estimates based on smaller samples.

Broadly speaking, in the face of computational constraints, the sample can be restricted in three different dimensions. First, one can restrict the length of the time series. This may be problematic because long time series may be required to reliably identify the dynamics of the state variables (Broadie, Chernov, and Johannes, 2007). The second and third dimensions are the moneyness and maturity dimensions. Many existing studies restrict both, focusing exclusively on at-the-money short-maturity options. Intuition suggests that these two types of cross-sectional restrictions will affect different parameters. Presumably different maturities are needed to identify persistence, while options with different moneyness are required to identify the parameters characterizing the higher moments. However, the existing literature does not contain any evidence on this.

We investigate the implications of data restrictions by comparing the results from the large benchmark sample with those of restricted samples.[7] Some of our findings are consistent

[7]Alternatively these issues can be studied using Monte Carlo simulation. However, even with the computational gains provided by our approach, a large-scale Monte-Carlo experiment is not feasible.

with our prior intuition, but others are sharply different. The most important finding is that maturity restrictions have the most serious implications for parameter estimates and model fit. For the models we study, maturity restrictions result in very different estimates not only for variance persistence, but also for kurtosis and risk premia. Somewhat unexpectedly, restricting the moneyness dimension does not greatly affect the estimates of parameters characterizing the higher moments, at least in our sample. Finally, and surprisingly, we find that parameters are relatively similar when estimating on different samples, even when these samples are much shorter than the benchmark 1996-2015 sample. Based on these findings, we discuss how data restrictions may have impacted parameter estimates in existing studies. We also show that the composition of the option sample is critically important for the relative weighting of returns and options in joint estimation. Some parameter estimates in the existing literature, such as the correlation between returns and variance, may largely reflect information from the underlying return series rather than from options, due to the composition of the option sample.

We conclude that when faced with computational constraints, it is critical to include options with different maturities in the sample to estimate these workhorse models. Which maturities to include in the sample, and the optimal sample composition more in general, are questions that we keep for future research. Another limitation of our empirical exercise is that these results are by design specific to the models, loss function, and sample under consideration. Specifically, for other models and/or purposes, the maturity dimension may be far less important and one can exclusively rely on short-maturity options.

Our results are relevant for an extensive literature that estimates parametric dynamic option valuation models using return and/or option data; see Singleton (2006) for a discussion of this literature. Several seminal studies combine return and option data using different econometric techniques, which are all computationally demanding. Aït-Sahalia and Kimmel (2007) estimate the model using returns and option information using approximate maximum

likelihood, but the information on options is exclusively based on the VIX. Pan (2002) uses implied-state Generalized Method of Moments (GMM) but uses one short-maturity at-the-money option per day in estimation. Chernov and Ghysels (2000) use Efficient Method of Moments (EMM) and use implied volatilities for the closest-to-maturity at-the-money calls. Eraker (2004) estimates the model in its full generality and samples the state variables using an MCMC technique. He incorporates cross-sectional information by using on average approximately three options per day. Despite the enormous advances in computing power following the publication of these studies, there is still a dearth of studies that estimate dynamic option pricing models with latent state variables using large cross sections and long time series. The computational cost is simply too high. The development of methods to overcome these constraints should therefore be a priority in option valuation and asset pricing, because such methods will expand our knowledge about option prices and returns but also more generally about risk premia in the underlying markets.

An interesting alternative approach to these computational problems is provided by Bates (2000), who manages to use big cross-sections of index options in estimation by treating spot volatilities as parameters.[8] In a series of papers, Andersen, Fusari, and Todorov (2015a,b, 2017, 2020) and Andersen, Fusari, Todorov, and Varneskov (2019, 2021) extend this approach by imposing consistency between the spot volatility implied from options and a high-frequency volatility estimate from returns.[9] The advantage of this approach compared to ours is that the extension to multifactor models is relatively more straightforward. The disadvantage is that it does not estimate the models in their full generality. In particular, the estimation approach does not explicitly take into account the filtering problem or the parameterization and the historical dynamics of the return equation. Consequently this approach

[8] See also Christoffersen et al. (2009) for an application of this approach. Bates' (2000) approach can be seen as a generalization of the approach of Bakshi, Cao, and Chen (1997), who estimate different model parameters on every day in the sample, also estimating the spot volatility as a parameter. Bates (2000) also uses a second estimation approach that imposes a dynamic constraint on the variance.

[9] These studies also develop asymptotic theory for a fixed or an increasing time dimension.

is not informative about risk premia. We therefore regard our approach as complementary, more closely related to the methods used by Eraker (2004) and Bates (2006), for instance.

Finally, several other recent papers also address the need to include information from the entire cross-section of options in estimation and propose computational improvements that make the resulting estimation problems feasible. Feunou and Okou (2018) propose the use of option-implied moments to capture the cross-sectional information. Hurn, Lindsay, and McClelland (2015) document the advantages of parallel computing when using the particle filter to estimate complex option pricing models using large cross-sections of option prices. Aït-Sahalia, Li, and Li (2020, 2021) characterize a first order approximation to the impact of multivariate stochastic volatility models with jumps on the implied volatility surface, which allows for model identification as well as estimation.

The paper proceeds as follows. Section 2 presents the option valuation model used in our empirical work and the return and option data used in estimation. Section 3 discusses the estimation of the model dynamics from returns and options. Section 4 presents the empirical results and Section 5 discusses the implications of restricting the option sample and the relative importance of returns and options for parameter estimates. Section 6 concludes.

2 Model and Data

We first discuss the return dynamic and the option valuation formula. Subsequently we discuss the return and option data.

2.1 The Model

Our approach can in principle be applied to any parametric option pricing model. To illustrate the advantages of our approach and the implications of restrictions on the option sample, we report on models that are well-known and that have been extensively studied.

This allows us to demonstrate that our methods produce reliable estimation results, while we can also relate differences in estimation results with the existing literature to the (larger) option data sets used in our analysis. We therefore base our empirical work on the SVCJ model with contemporaneous jump arrivals in return and variance (Duffie, Pan, and Singleton, 2000):

$$\frac{dS_t}{S_t} = (r_t - \delta_t + \eta_s V_t - \lambda\bar{\mu}_s)dt + \sqrt{V_t}dZ_t + (e^{J_t^s} - 1)dN_t, \tag{1}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t + J_t^v dN_t, \tag{2}$$

where S_t is the index level, r_t is the risk-free rate, δ_t is the dividend yield, κ denotes the speed of mean reversion, θ is the unconditional mean variance, and σ determines the variance of variance. dZ_t and dW_t are Brownian motions with $corr(dZ_t, dW_t) = \rho$, and $\eta_s V_t$ is the diffusive equity risk premium, which is assumed to be linear in V_t. N_t stands for a Poisson process with constant jump intensity λ, and J_t^s and J_t^v are the jump size parameters related to returns and variance, with correlation ρ_J. We assume $J_t^v \sim Exp(\mu_v)$ and $J_t^s | J_t^v \sim N(\mu_s + \rho_J J_t^v, \sigma_s^2)$. The term $\lambda\bar{\mu}_s$ is the compensation of the jump component, with $\bar{\mu}_s = \frac{e^{(\mu_s + \sigma_s^2/2)}}{1-\rho_J\mu_v} - 1$.

We assume that the risk neutral dynamic is given by:

$$\frac{dS_t}{S_t} = (r_t - \delta_t - \lambda\bar{\mu}_s^Q)dt + \sqrt{V_t}dZ_t^Q + (e^{J_t^{sQ}} - 1)dN_t^Q, \tag{3}$$

$$dV_t = \kappa^Q(\theta^Q - V_t)dt + \sigma\sqrt{V_t}dW_t^Q + J_t^{vQ}dN_t^Q, \tag{4}$$

where we assume that the diffusive variance risk premium is equal to $\eta_v V_t$, and thus $\kappa^Q = \kappa - \eta_v$ and $\theta^Q = (\kappa\theta)/\kappa^Q$. We assume that the jump risk premia are entirely attributable to the mean jump sizes of return and variance: $\eta_{J^s} = \mu_s - \mu_s^Q$ and $\eta_{J^v} = \mu_v - \mu_v^Q$.[10] The

[10] We use a simple structure of the jump risk premium because of identification concerns, following Eraker (2004) and Pan (2002), for example. Broadie, Chernov, and Johannes (2007) investigate more general entertain assumptions regarding the return jumps risk premium, but they use a very different empirical design.

jump intensity λ and the standard deviation of the return jump size σ_s do not change across measures. The total equity risk premium is therefore equal to $\eta_s V_t + \lambda(\bar{\mu}_s - \bar{\mu}_s^Q)$.

The SVCJ specification nests several models in the existing literature. See Singleton (2006) for a detailed discussion. If we set $\lambda = 0$, it reduces to the SV model of Heston (1993). If we shut down the jump in variance, it becomes the SVJR model of Bates (1996). It also nests a model with variance jumps only (SVJV) if we shut down the jumps in returns. Note that we do not estimate the more general SVSCJ model studied in Eraker, Johannes, and Polson (2003) and Pan (2002), which makes the jump intensity a function of variance. Given the computational burden of estimating the models under consideration using option data, we leave the study of this model for future work.

The model price of a European call option $C^M(V_t|\Theta)$ with maturity τ and strike price K is given by:

$$C^M(V_t|\Theta) = e^{-r_t\tau} E^Q[\max(S_{t+\tau} - K, 0)]. \tag{5}$$

In our application, we need to repeatedly calculate prices of options with different spot variances and the ability to vectorize the formula is important for computational reasons. We use an approach based on the formula of Carr and Madan (1998) to compute option prices. Because of the affine structure of the models, quasi closed-form solutions for option prices are available. See Appendix A for a detailed discussion. We denote the model price by C^M as opposed to the option's market price C. The model price is computed given the current state V_t and model parameters $\Theta(\kappa, \theta, \sigma, \rho, \eta_s, \eta_Q, \lambda, \mu_{J_s}, \sigma_{J_s}, \eta_{J_s}, \mu_v, \eta_{J_v}, \rho_J, \sigma_c)$.

2.2 Return and Option Data

We use S&P500 returns and option prices for the period January 1, 1996 to December 31, 2015, a total of 5031 trading days. The need to use a long sample to identify return

dynamics for option valuation is emphasized by Eraker, Johannes, and Polson (2003), Eraker (2004) and Broadie, Chernov, and Johannes (2007), for example. We obtain index returns and risk-free rates from CRSP, and option prices, zero coupon yields, and dividend yields from OptionMetrics. The top panel of Figure 1 plots the time series of the daily returns. The financial crisis is readily apparent, and it is characterized by large negative as well as positive returns. The bottom panel of Figure 1 plots the squared returns. This figure clearly demonstrates the challenges in modeling the twenty-year sample period. In the financial crisis the variance spikes up, but mean reverts rather quickly. The same observation applies to other periods with large variance spikes. Panel A of Table 1 provides descriptive statistics on the index returns. They exhibit negative skewness and excess kurtosis.

We use both put and call index options and impose the following standard filters on the option data:

1. Discard options with fewer than 5 days and more than 365 days to maturity.
2. Discard options with implied volatility less than 5% and greater than 150%.
3. Discard options with volume or open interest less than 5 contracts.
4. Discard options with quotes that suggest data errors. We discard options for which the best bid exceeds the best offer, options with a zero bid price, and options with negative put-call parity implied price.
5. Discard options with price less than 50 cents.

After imposing these filters, the resulting data set contains 945,110 option contracts. The sample contains more puts than calls, as expected. Moreover, the option data set obtained after imposing these filters is not balanced over time. This imbalance is substantial: we have almost eight times more options in 2015 than in 1996. In principle, this is not a problem for our estimation approach, because we can accommodate any type of balanced or unbalanced option panel. However, the results from a more balanced data set are easier to interpret.

We therefore create a more balanced panel. Put prices are converted into call prices

based on put-call parity. We use a relatively large number of moneyness-maturity bins to adequately represent the observed daily option surfaces. We use six moneyness bins and five maturity bins. Moneyness is defined as strike price divided by index price (K/S). For each moneyness-maturity bin, we include only the most liquid option, defined as the option with the highest trading volume. The data set thus has thirty options per day unless options are not available for certain bins. This procedure yields a data set with 123,399 option contracts. Panel B of Table 1 provides sample sizes for these moneyness-maturity bins. Our analysis is based on this more balanced data set.[11] Panel B of Table 1 also reports average option prices and implied volatilities for these moneyness-maturity bins.

3 Estimation

Our methodology relies on the particle MCMC framework developed in Andrieu et al. (2010). This approach reduces the dimensionality of the MCMC algorithm by using MCMC for parameter inference, while applying the particle filter to filter the latent states.

We first set up our notation and provide a generic characterization of the estimation method. Subsequently we discuss the details of our implementation of the particle filter. We briefly explain how the particle filter can be applied on return data to estimate stochastic volatility models with jumps. We then highlight the computational problems that arise when estimating these models from option data and how we address them. Next, we discuss how we combine return and option data in estimation. Finally, we use a Monte Carlo study to illustrate that our method performs well.

[11] Note that the data are still unbalanced because in the early years of the sample we do not have many observations for short-maturity out-of-the-money calls and/or in-the-money puts.

3.1 Notation

The implementation of the particle filter depends on the observables: returns, options, or both. However, conceptually the algorithm is similar. We therefore introduce notation that can be used regardless of the observables.

We first need to time-discretize the continuous-time model. Several discretization methods are available and every scheme has certain advantages and drawbacks. We use the Euler scheme, which is easy to implement and has been found to work well for this type of applications (Eraker, 2004).[12] Applying Ito's lemma and discretizing (1)-(2) gives:

$$R_{t+1} = ln(\frac{S_{t+1}}{S_t}) = r_t - \delta_t - V_t/2 + \eta_s V_t - \lambda\bar{\mu}_s + \sqrt{V_t}z_{t+1} + J^s_{t+1}B_{t+1}, \tag{6}$$

$$V_{t+1} - V_t = \kappa(\theta - V_t) + \sigma\sqrt{V_t}w_{t+1} + J^v_{t+1}B_{t+1}, \tag{7}$$

where z_{t+1} and w_{t+1} are distributed standard normal. The discrete jump frequency B_{t+1} follows the Bernoulli distribution. For each time period, there is either no jump or one jump. The corresponding discretized risk-neutral dynamics are identical but use the risk-neutral parameters. We implement the discretized model using daily returns, but we report annualized parameter estimates below.

We assume that observed option prices are equal to the model price plus error:[13]

$$C_{t,h} = C^M_{t,h}(V_t|\Theta) + \varepsilon_{t,h}, \tag{8}$$

where $h = 1, 2, .., H_t$ and H_t is the total number of options at date $t = 1, \ldots, T$. We assume

[12] For alternative simulation schemes, see for instance Broadie and Kaya (2006), Andersen (2007), and Glasserman and Kim (2011).

[13] Prices greatly vary in the cross-section of options, as can be seen from Table 1, and loss functions based on relative prices or implied volatilities may offer different insights (Hurn et al., 2015). We plan to address this issue in more detail in future work. However, we repeated our analysis using a sample of out-of-the money put options, which contains much less price variation, and the parameter estimates are relatively similar.

$\varepsilon_{t,h}$ is normally distributed $\varepsilon_{t,h} \sim N(0, \sigma_c^2)$.[14] Note that while the spot variance V_t depends on past jumps, it is sufficient to compute the model option price at time t.

It is helpful to formulate these dynamics in a state-space representation. Denote L_{t+1} as the latent states that are used to generate the observables Y_{t+1}. Based on the discretization, $L_{t+1} = (V_t, B_{t+1}, J^s_{t+1}, J^v_{t+1})$ and $Y_{t+1} = (\{C_t\}, R_{t+1})$ where $\{C_t\} = (C_{t,1}, ..., C_{t,H_t})$. Note the timing convention in the notation: Given R_t, we simulate V_t and the cross-section of option prices $\{C_t\}$; and V_t, together with B_{t+1} and J^s_{t+1}, produces the next period return R_{t+1}. Define the measurement density by $f_1(Y_{t+1}|L_{t+1})$ and the transition density by $f_2(L_{t+1}|L_t)$. The latent states evolve through the transition density function, while the observables are realizations conditional on the latent states and the measurement density. The state-space representation applies regardless of whether we observe returns, options, or both. When returns are the observables, f_1 refers to equation (6); when options are the observables, f_1 is given by equation (8).[15] For the persistent latent variance V_t, f_2 represents equation (7); for the non-persistent jump variable, f_2 is simply a random draw from the corresponding distribution.

This gives the following state-space representation:

$$\begin{array}{ccccccc} (L_{t+1}) & \xrightarrow{f_2} & (L_{t+2}) & \xrightarrow{f_2} & (L_{t+3}) & \xrightarrow{f_2} & ... \\ \downarrow f_1 & & \downarrow f_1 & & \downarrow f_1 & & \\ Y_{t+1}(\{C_t\}, R_{t+1}) & & Y_{t+2}(\{C_{t+1}\}, R_{t+2}) & & Y_{t+3}(\{C_{t+2}\}, R_{t+3}) & & ... \end{array} \tag{9}$$

Although we do not directly observe the latent states $L_{t+1}, L_{t+2}, ...$, we do observe the option prices and/or returns in each period. Equation (9) highlights that the jumps and the spot

[14] We also investigated a time-varying σ_c, but this did not lead to significant improvements in option fit.

[15] Since the option price depends only on the current spot variance and the risk-neutral expected jumps rather than the realized jumps, $f_1(\{C_t\}|L_{t+1})$ reduces to $f_1(\{C_t\}|V_t)$, which explains the timing convention in equation (8).

variances are latent variables. Similar to Eraker (2004), our estimation procedure thus accounts for both the measurement and transition densities, unlike the approach pioneered by Bates (2000), which treats the latent variables as parameters to be estimated.

In the next two sections we discuss the estimation method in general, which can apply to the case where we observe returns, options, or both. We have two sets of unknowns: 1) parameters $\Theta(\kappa, \theta, \sigma, \rho, \eta_s, \eta_Q, \lambda, \mu_{J_s}, \sigma_{J_s}, \eta_{J_s}, \mu_v, \eta_{J_v}, \rho_J, \sigma_c)$ and 2) latent states $\{L_{t+1}(V_t, B_{t+1}, J^s_{t+1}, J^v_{t+1})\}$. We use an orthogonal MCMC sampler to explore the parameter space and a particle filter to integrate the latent states.

3.2 Sampling the Parameters via Orthogonal MCMC

We use the Adaptive Metropolis-Hastings MCMC (AMH-MCMC) within the orthogonal MCMC framework to explore the parameter space. The AMH-MCMC technique is originally based on Metropolis et al. (1953) and Hastings (1970) and is widely used in the existing literature due to its flexibility, especially when dealing with high-dimensional distributions. See Johannes and Polson (2009a) for a detailed discussion. In general, at the jth iteration, AMH-MCMC randomly samples a parameter set from a proposal distribution $q_j(\Theta^p_j|\Theta_{j-1})$ and subsequently accepts a new parameter vector Θ^p_j with probability:

$$\alpha(\Theta^p_j, \Theta_{j-1}) = \min(1, \frac{f(Y_{1:T}|\Theta^p_j)p(\Theta^p_j)/q_j(\Theta^p_j|\Theta_{j-1})}{f(Y_{1:T}|\Theta_{j-1})p(\Theta_{j-1})/q_j(\Theta_{j-1}|\Theta^p_j)}), \tag{10}$$

where f denotes the total likelihood (i.e., in which the latent variables have been integrated out) and $p(\Theta)$ is the prior for Θ, which can be uninformative. We assume flat or diffuse priors for the model parameters, except for the jump frequency.[16]

[16] We found it difficult to distinguish between frequent small return jumps and more infrequent large return jumps, especially when estimating using returns only. Following Eraker, Johannes, and Polson (2003), we therefore impose an informative prior on jump frequency to favor infrequent but large return jumps when index returns are the observables, as follows: $\ln \frac{\lambda-0.5}{252-\lambda} \sim N(-6, 1)$. We use the following diffuse priors: $\ln \frac{\rho+1}{1-\rho} \sim N(-1.5, 0.5)$, and $\ln \frac{\rho_J+1}{1-\rho_J} \sim N(-0.5, 0.5)$. The other priors are flat, i.e. the density function is proportional to 1.

We set the proposal distribution $q_j(\Theta_j^p|\Theta_{j-1})$ to be multivariate normal with mean Θ_{j-1}. Tuning the parameter variances in this proposal distribution is crucial to achieve estimation efficiency. Roberts and Rosenthal (2009) and Vihola (2012) propose adaptive random walk schemes, where q_j is estimated from previous iterations $(\Theta_1, ..., \Theta_{j-1})$. To initiate the sampler, we start from some pre-specified covariance matrix for the model parameters. Then, we adjust the covariance matrix at each MCMC iteration to obtain an acceptance rate of about 30%.[17] Note that if q_j is set to be a fixed distribution q, this implementation reduces to the standard MH algorithm.

In short, the idea behind AMH-MCMC is to move to a new parameter set with probability one if it generates a higher likelihood than the previous parameter set. Note that the algorithm moves to a new parameter set with a non-zero probability even if the new likelihood is lower than the previous one.

While the AMH-MCMC sampler is in principle sufficient to estimate the model parameters, we increase computational efficiency by using the Orthogonal MCMC (O-MCMC) sampler of Martino et al. (2016). This interacting parallel MCMC scheme consists of running Z MCMCs that switch between M_V independent moves and M_H interacting steps. The independent moves, or vertical moves, use the AMH-MCMC. The M_H interacting iterations, or horizontal moves, update the current state of each MCMC using a sample Metropolis-Hastings algorithm (see Table 2 of Martino et al., 2016). The horizontal moves improve the overall mixing by exchanging information across the different MCMCs. Although the O-MCMC sampler consists of running Z MCMC samplers in parallel, it effectively implies a single ergodic MCMC algorithm in which the stationary distribution is the product of Z posterior distributions. Therefore, once the O-MCMC algorithm has converged, both the horizontal and the vertical moves can be used to obtain Z draws from the posterior

[17] This acceptance rate lies in the standard optimal rate interval given by $[23.4\%, 44\%,]$. These acceptance rates lead to an efficient balance between exploring and exploiting the posterior distribution, see e.g. Roberts and Rosenthal (2001) for more details.

distribution.

We rely on the O-MCMC scheme for two computational reasons. First, we do not know ex-ante the best way to initialize the AMH-MCMC samplers, so we can choose after some MCMC iterations which of the methods produce an acceptance rate close to our 30% target. In our implementation, half of the MCMC algorithms are initialized with the method of Roberts and Rosenthal (2009) and the other half follows Vihola (2012). Then, during the burn-in period, after each vertical move we replace the sample method and the current state of the least performing chain in terms of acceptance rate by the attributes of the chain exhibiting the closest target acceptance rate. This ensures that the two AMH-MCMC approaches are competing and that the best method is selected according to its posterior exploration performance. Second, the O-MCMC allows running Z MCMC in parallel which leads to large computational gains. In our empirical implementation, we set $Z = 10$, $M_V = 10$ and $M_H = 1$. In words, every 10th iteration, we update the 10 parallel MCMC chains using one horizontal move.

3.3 Integrating the Latent Variables Using the Particle Filter

Sampling the parameters Θ requires the total likelihood (see equation (10)). We implement a standard sampling-importance-resampling (SIR) particle filter at each time t from 1 to T using the following steps:

Step 1. Simulate the particles forward. Using the time t resampled particles L_t^i, $i = 1, \ldots, N$, where N is the total number of particles, for each particle i simulate $\tilde{L}_{t+1}^i$ from L_t^i according to $\tilde{L}_{t+1}^i = f_2(L_{t+1}|L_t^i)$.

Step 2. Compute the weight for each particle and normalize:

$$\omega_{t+1}^i = f_1(Y_{t+1}|\tilde{L}_{t+1}^i) \tag{11}$$

$$\pi_{t+1}^i = \omega_{t+1}^i / \sum_{j=1}^{N} \omega_{t+1}^j \tag{12}$$

Step 3. Resample the particles $\tilde{L}_{t+1}^i$ according to the normalized weights $\{\pi_{t+1}^i\}$, which gives L_{t+1}^i, $i = 1, \ldots, N$.

Following Malik and Pitt (2011), the total likelihood required for sampling the parameters Θ can be expressed as a function of the unnormalized weights.[18] For the SIR, this gives:

$$f_1(Y_{1:T}|\Theta) = \prod_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \omega_{t+1}^i \right\}. \tag{13}$$

In the SIR algorithm, we can think of the weights π_{t+1}^i as constituting a discrete probability distribution for L_{t+1}. After resampling, the weight for each particle changes back to $1/N$. The SIR is extremely intuitive and simple to implement. However, since new particles are simulated blindly in step 1, it may lead to the well known sample impoverishment problem (Johannes, Polson, and Stroud, 2009), especially when N is not very large. Consider a scenario where we have an extremely large negative return at $t+1$. Particles with large V_t or large negative return jump occurrence will receive large weights, while other particles will be assigned weights close to zero. As a result, the resampling might consist of repeated values of these few particles.

Pitt and Shephard (1999) introduce the Auxiliary Particle Filter (APF) to solve this problem by resampling before propagation.[19] In our applications, it turns out that the APF underperforms compared to the SIR due to a more volatile likelihood function. As discussed by Liu et al. (2011), the APF performs poorly when the measurement density $f_1(Y_{t+1}|L_{t+1})$ is highly peaked and the transition density $f_2(L_{t+1}|L_t)$ is diffuse. In such cases, the APF will first select repeatedly the same few particles in its resampling step, and the propagation

[18] Note that the total likelihood is a random variable as it depends on the particle weights. However, Malik and Pitt (2011) show that the SIR and the APF provide unbiased estimates of the (true) likelihood.

[19] See Carvalho, Johannes, Lopes, and Polson (2010) for an alternative approach.

step will be highly sensitive to the proposal outliers of the transition density. Due to the large option panels we use to estimate the models, the measurement density becomes highly informative and the APF becomes less reliable. We therefore use the SIR, and we implement it with a large number of particles to ensure good performance.

3.4 The Return Likelihood

As mentioned above, the algorithm can be applied to different sources of information, which corresponds to different likelihoods $f_1(Y_{t+1}|L^i_{t+1})$. First consider estimation based on returns, where we exclusively use returns as the observables. Using equations (6) and (7) this gives:

$$f_1(R_{t+1}|L^i_{t+1}) = \frac{1}{\sqrt{2\pi V^i_t}} \exp\left\{ -\frac{1}{2} \frac{\left[R_{t+1} - (r_t - \delta_t - \frac{1}{2}V^i_t + \eta_s V^i_t - \lambda\bar{\mu}_s + J^{s(i)}_{t+1} B^i_{t+1})\right]^2}{V^i_t} \right\}, \tag{14}$$

Appendix B provides additional details on particle filtering based on returns data. See also Christoffersen, Jacobs, and Mimouni (2010) for a related implementation on the SV model.

3.5 The Option Likelihood

We now consider the likelihood based on option data only, without considering the underlying returns. The existing literature that estimates option pricing models using large panels of options deals with latent states such as the spot variance broadly in two ways. The first approach is to extract the state variables from return data, either by filtering from daily returns or by calibrating from intra-day data. See for example Andersen, Fusari, and Todorov (2015a) and Christoffersen, Jacobs, and Mimouni (2010). The other approach is to treat the spot variance as a parameter to be estimated along with other parameters (Bates, 2000). Both estimation approaches are viable, but the resulting parameter estimates will

either reflect return-based information or ignore the dynamic of the spot variance. We now discuss an alternative approach that uses the particle filter to estimate model parameters exclusively based on option data. We first illustrate the computational challenges inherent in this exercise, and then we discuss how we alleviate these computational constraints.

3.5.1 Computational Constraints

The instantaneous variance follows the transition equation (7), but now the observables consist of a cross-section of H_t option prices for each day, denoted by $\{C_t\}$. The filtering problem therefore consists of evaluating the likelihood of observing the market option prices conditional on the latent states.

Conceptually this filtering procedure is as straightforward as the one using returns; however, it encounters significant computational constraints. The measurement density now corresponds to equation (8). Rather than one return for each time period t, we now have (in our sample) up to thirty option prices available at time t. The likelihood for the i^{th} particle at time t can be calculated as:

$$\begin{aligned} f_1(Y_{t+1}|L_{t+1}^i) &= \left(\prod_{h=1}^{H_t} f_1(C_{t,h}|C_{t,h}^M(V_t^i|\Theta)) \right)^{1/H_t}, \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_c} \right) \exp\left(-\frac{\sum_{h=1}^{H_t}(C_{t,h} - C_{t,h}^M(V_t^i|\Theta))^2}{2\sigma_c^2 H_t} \right). \end{aligned} \tag{15}$$

We take the square root of H_t in order to normalize the likelihood with respect to the different numbers of options on each day.[20] The total likelihood for the entire sample, summing over

[20] In joint estimation based on returns and options, the optimal relative weighting of returns and options in the likelihood is an important concern, because the abundance of option data may cause the weight of the returns in the likelihood to become negligible. See for example Bates (2003) for a discussion of this issue.

all particles, is:

$$f(Y_{1:T}|\Theta) = \prod_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\left(\frac{1}{\sqrt{2\pi}\sigma_c} \right) \exp\left(- \frac{\sum_{h=1}^{H_t} (C_{t,h} - C_{t,h}^M(V_t^i|\Theta))^2}{2\sigma_c^2 H_t} \right) \right] \right\}. \tag{16}$$

When computing this likelihood, a quasi closed-form solution for the option price is available in the affine models we consider, and each option price takes less than 0.01 seconds to evaluate. Nonetheless, this computation encounters significant computational constraints. For each function evaluation we have to evaluate option prices along three dimensions: for each option, for each particle, and for each day. Our sample period consists of 5031 trading days and up to 30 options per day. Assuming 10,000 particles (the minimum we use) then leads to approximately 1,500,000,000 computations of the option price in each function evaluation, which is computationally infeasible.

3.5.2 The State Variable Quantile Method

Given this computational burden, we propose a more efficient filtering algorithm based on the quantiles of the state variable(s), which we refer to as the state variable quantile (SVQ) method. Recall that for the models we consider, the spot variance is sufficient for option valuation, even in models with jumps. The motivating idea can loosely be thought of as reducing the three-dimensional option evaluation computation (i.e., $N \times T \times H$) into a (pseudo) two-dimensional computation (i.e., $T \times H$), where $H = \frac{1}{T} \sum_{t=1}^{T} H_t$. Specifically, the particle filter implies the computation of N option prices at each time t and for each option $h \in [1, H_t]$. The SVQ method reduces this computational burden to Q computations where Q does not depend on N. To do so, it approximates each option price $C_{t,h}^M(V_t^i|\Theta)$ as a function of the particles using a polynomial regression calibrated on Q option prices.[21] For each option price $C_{t,h}$, our approach proceeds as follows.

[21] Ferriani and Pastorello (2012) approximate model-implied option prices using flexible parametric models. Our approach is entirely different as we use the polynomial to approximate the dependence on the state variable, while relying on closed-form option pricing formulas.

1. Take Q quantiles $\{V_t^1, ..., V_t^Q\}$ evenly spaced over [0,100] from the predictive filtered spot variance distribution, i.e. $V_t|Y_{1:t-1}$.

2. Evaluate the option prices $\{C_{t,h}^M(V_t^1|\Theta), ..., C_{t,h}^M(V_t^Q|\Theta)\}$.

3. Minimize the sum of squared pricing errors using a polynomial function of order p:

$$(\hat{\beta}_1, \ldots, \hat{\beta}_{p+1}) = \underset{(\beta_1,\ldots,\beta_{p+1})}{\arg\min} \sum_{q=1}^{Q} (C_{t,h}^M(V_t^q|\Theta) - \beta_1 - \beta_2 V_t^q - \ldots - \beta_{p+1}(V_t^q)^p)^2. \quad (17)$$

4. For each particle $i \in [1, N]$, set the option price to:

$$C_{t,h}^M(V_t^i|\Theta) = \hat{\beta}_1 + \hat{\beta}_2 V_t^i + \ldots + \hat{\beta}_{p+1}(V_t^i)^p. \quad (18)$$

The implementation of the SVQ method requires a choice regarding the number Q of quantiles and the order p of the polynomial regression. We extensively experimented with these tuning parameters and we found that the approximation turns out to be extremely good for any $Q > 3p$ with $p \geq 2$. The reason is that we maximize the variance of the spot variance variable in the polynomial regression by taking the quantiles of the spot variance distribution, which ensures that we minimize the variance of the polynomial regression estimators. We set $p = 3$ and $Q = 12$ in our empirical implementation. Figure 2 shows that the root mean squared relative pricing errors for the estimates of the Heston SV model are smaller than 0.03 over the 1996-2015 sample. Estimating the Heston model on a shorter period and one option per day, we are able to compare the estimates from direct implementation of the particle filter with the ones obtained using the SVQ method. The differences are very small, the largest being 4.9% for the estimated variance risk premium parameter η_v.

To justify the use of a polynomial regression, denote the derivative of an option price with respect to the spot variance evaluated at $\bar{V}_t$ by $\frac{C_{t,h}^M(V_t|\Theta)}{dV_t}|_{V_t=\bar{V}_t} = C'(\bar{V}_t)$. Using a Taylor approximation of order 3 without loss of generality, an option price can be approximated

around a spot variance $\bar{V}_t$ as follows:

$$\begin{aligned} C^M_{t,h}(V_t|\Theta) &\approx C^M_{t,h}(\bar{V}_t|\Theta) + (V_t - \bar{V}_t)C'(\bar{V}_t) + \frac{1}{2}(V_t - \bar{V}_t)^2 C''(\bar{V}_t) + \frac{1}{6}(V_t - \bar{V}_t)^3 C'''(\bar{V}_t), \\ &= C^M_{t,h}(\bar{V}_t|\Theta) - \bar{V}_t C'(\bar{V}_t) + \frac{\bar{V}_t^2 C''(\bar{V}_t)}{2} - \frac{\bar{V}_t^3 C'''(\bar{V}_t)}{6} + \dots \\ &\quad [C'(\bar{V}_t) - \bar{V}_t C''(\bar{V}_t) + \frac{\bar{V}_t^2 C'''(\bar{V}_t)}{2}]V_t + [\frac{C''(\bar{V}_t)}{2} - \frac{\bar{V}_t C'''(\bar{V}_t)}{2}]V_t^2 + \frac{C'''(\bar{V}_t)}{6}V_t^3, \\ &= \beta_1 + \beta_2 V_t + \beta_3 V_t^2 + \beta_4 V_t^3. \end{aligned}$$

Consequently, equation (17) can be understood as a Taylor approximation around a spot variance $\bar{V}_t$ that is estimated by minimizing the error terms.

Rather than computing option prices N times, once for each particle, our approach computes the price Q times. In other words, we largely get rid of one dimension (the number of particles N) when computing option prices, thus saving more than 99% of computation time. Note also that while the option prices in equation (18) can be written in terms of the latent variable V_t, the filtered variance jumps are required in order to obtain the filtered variance path.

Thanks to this quantile spot variance method, the computational cost of our implementation of the particle filter is only mildly impacted by the number N of particles. We therefore use a large number of particles to estimate the models. In our implementation, we use $N = 10,000$ particles for the SV model and $N = 40,000$ particles for models with jumps.

Potential concerns with generalizing the quantile method to multiple factors are the curse of dimensionality and the difficulties with implementing joint quantiles. These issues can be addressed in several ways. For instance, we can rely on a Particle Gibbs approach (Andrieu et al., 2010). Specifically, we can sample each factor conditional on the others using a conditional particle filter and apply the quantile spot variance method. Sampling the factors one at a time reduces the mixing of the Markov chain, but it remains more efficient compared

to a single-move method as in Eraker (2004).[22]

3.6 The Joint Likelihood

To estimate the models using returns and options, we need to combine the two likelihoods. This seems straightforward, but it is well known that this may create a problem with the relative weights of returns and options in the likelihood. On a given day, we have a single return but a large cross-section of options. The weights affect both the likelihood of a particle in the filtering step, as well as the total likelihood of a parameter set in the AMH-MCMC. Recall though that in our option likelihood, we normalize the likelihood with respect to the number of options on each day. Therefore we effectively give equal weight to options and returns in the joint likelihood. This allows us to abstract from the size of the cross-section of options and focus on the composition of the option data. This implementation is somewhat ad-hoc, but the specification of any joint likelihood of options and returns is not entirely guided by theory and therefore to some extent ad-hoc.

Using this approach, the resulting total likelihood for each particle on date t can be expressed as:

$$f_1(Y_{t+1}|L^i_{t+1}) = f_1(R_{t+1}|L^i_{t+1})(\prod_{h=1}^{H_t} f_1(C_{t,h}|C^M_{t,h}(V^i_t|\Theta)))^{1/H_t}, \tag{19}$$

where the two components can be computed according to equations (14) and (15) respectively. The relative weights are equal in the sense that we constrain the information from returns to be equally important to the information from options, no matter how many options we have available on a given day. Given this likelihood function for each particle, the total likelihood can be calculated following the implementation discussed in Sections 3.4 and 3.5.

[22] See Kim, Shephard, and Chib (1998) for a comparison of MCMC mixing between single-moves and block sampling.

3.7 A Monte Carlo Experiment

In contrast to standard particle filter methods, the quantile spot variance method makes it possible to estimate the models we consider in this paper with both returns and large panels of option data. Because of the improvement in computational efficiency, it is also possible to study the sampling properties of the estimator in a (small-scale) Monte Carlo exercise.

Table 2 presents the results of the Monte Carlo study. We simulate fifty samples of return time series and option panels for two parameterizations of the SV model. The length of the samples is one year and the sample frequency is daily. Column 2 in Table 2 shows the two sets of parameters. They are chosen based on the empirical results below. The main difference between them is that the ρ parameters are set at -0.90 and -0.70 respectively, and we modify the σ to keep the unconditional return kurtosis constant (see Das and Sundaram (1999) for the model-implied moments). The simulated option samples are similar to the sample used in the empirical analysis and summarized in Table 1. For a one-year sample, we thus end up with 6731 options.

Table 2 shows that for both parameterizations, the medians of the parameter estimates based on the 50 replications are close to the true values, with good precision as indicated by the first and third quartiles. The estimates of the ρ and σ parameters are slightly biased downward. The price of risk parameters η_s and η_v are precisely estimated.

4 Empirical Results

Our empirical implementation focuses on joint estimation with returns and option data, following the approach in Section 3.6.[23] Our MCMC setup uses 15,000 iterations. We set the first one third of the iterations as burn-in, and report the posterior mean and standard

[23] We occasionally refer to the estimates based on return data in Table A1 in this Section. We discuss these estimates in more detail in Section 5.

deviation for each parameter from the subsequent iterations.

We discuss our results on a model-by-model basis. We start with the simplest model, the Heston (1993) stochastic volatility model (SV). We then discuss the model with return jumps (SVJR), the model with variance jumps (SVJV), and the model with correlated return and variance jumps (SVCJ). We compare our results with estimates from the existing literature. See also Singleton (2006, chapter 15) for a discussion of existing results. Subsequently we discuss and compare model fit and the model-implied higher moments.

4.1 Inference on Stochastic Volatility

The first column of Table 3 presents estimates for the Heston (1993) square root stochastic volatility (SV) model. These estimates can be compared with the existing literature in Panel C of Table 4, with the obvious caveat that those estimates are obtained using different option samples and sample periods.[24]

To motivate our empirical exercise, consider the summary of existing results in Table 4. Panel A indicates that a large number of studies report results for the Heston model under the physical measure based on index return data. However, despite the popularity of this model and its central place in the literature, relatively few papers present estimates under the risk-neutral measure based on option data, listed in Panel B, or joint estimates using both return and option data, listed in Panel C. Many of the studies in Panel C typically impose constraints due to computational complexity. For example, Aït-Sahalia and Kimmel (2007) use daily data for 1990-2004, but on every day they only use the index return and the VIX, which effectively amounts to using one short-maturity at-the-money option every day. Pan (2002) estimates the model using return and option data for 1989-1996, but the

[24]To facilitate comparisons, the estimates in Table 4 are all reported in annual units. This is similar to Pan (2002) but different from other studies such as Eraker (2004) and Broadie, Chernov, and Johannes (2007), for example. Compared to the estimates reported in Eraker (2004) and Broadie, Chernov, and Johannes (2007), the estimates of κ and σ in Table 4 are multiplied by 2.52 (multiplied by 252 and divided by 100). The estimate of θ in Table 4 is multiplied by 0.0252 (multiplied by 252 and divided by 10,000).

option cross-section is very limited.[25] The joint estimation in Eraker (2004) uses 3270 call options over 1006 trading days, averaging approximately three options per day. Chernov and Ghysels (2000) is based on the period 1985-1994, and focuses on short-maturity at-the-money calls. Bates (2000) uses large cross-sections of options but does not filter the latent state variables from options. The more recent contribution by Hurn et al. (2015) is closer to our approach and uses large option panels and filtering. In summary, many existing estimates of the Heston model reduce the cross-sectional dimension of the option data set, at least partly due to computational constraints. See Hurn et al. (2015) and Broadie et al. (2007) for related discussions on the importance of computational constraints in this literature. Some of these constraints can be overcome with additional computational resources, but in itself this is not sufficient. We propose methods that help overcome these constraints and allow us to keep the time-series and cross-sectional dimension of the option data as large as possible.

We now discuss the implications of the larger option cross-sections on model estimates. First, the posterior mean of the diffusive variance risk premium parameter η_v in Table 3 is equal to 1.1156. This estimate has the expected sign. It is similar to the values implied by nonparametric or semiparametric estimates (see for instance Bollerslev et al. (2009)), but different (smaller and statistically more significant) from most existing estimates based on parametric models estimated with options and returns jointly. The estimate in Pan (2002) is statistically significant but much larger. The estimates in Eraker (2004) and Broadie et al. (2007) are not statistically significant.[26] Our estimate of η_v is also smaller than the estimate in Bates (2000), which is obtained using options only, but additionally imposes a dynamic constraint on the spot variance. The estimate in Hurn et al. (2015) is most similar to ours. Recall that their sample also includes options with widely different moneyness and maturity.

[25] Pan (2002) also evaluates model fit using a wider cross-section based on the estimates from the smaller sample.

[26] Broadie et al. (2007) argue that insignificant estimates in the existing literature may be due to a flat volatility term structure of volatility and/or the absence of options with longer maturities in the data. Panel B of Table 1 indicates that the term structure of implied volatility in our sample varies by moneyness.

The estimate of η_v has important implications because it defines the relation between the physical and risk neutral mean reversion and long-run variance. The estimate of the long-run physical variance of returns θ in Table 3 is 0.0334. This gives a risk-neutral long-run variance of $\theta^Q = \kappa\theta/\kappa^Q = \kappa\theta/(\kappa - \eta_v) = 0.09717$. Consistent with the literature, the risk-neutral variance considerably exceeds the physical variance due to a positive η_v estimate. As mentioned above, this finding is consistent with nonparametric evidence, see for example Bollerslev, Tauchen, and Zhou (2009). The estimate of the mean reversion κ in Table 3 is 1.6999. The risk-neutral mean reversion parameter $\kappa^Q = \kappa - \eta_v$ from Table 3 is thus equal to 1.6999-1.1156=0.5843. Our finding that the mean reversion estimated from options is smaller than the physical mean reversion from returns (in Table A1) confirms existing findings. However, our estimate of mean reversion is much smaller than existing estimates in Panel C of Table 4, again with the exception of Hurn et al. (2015).[27] Note that the return-based estimate of κ in the first column of Table A1 is consistent with existing estimates.

Our estimates of the parameters characterizing the tails of the distribution also differ from existing results. In Table 3, the estimate of ρ, the correlation between the return and variance innovations, is -0.9085, much more negative than estimates in existing studies in Panels B and C of Table 4. Our estimate of ρ based on returns in Table A1 is -0.7886, suggesting that our findings are driven by (the composition of) the option sample. The estimate in Hurn et al. (2015) is -0.741, but Andersen, Fusari, and Todorov (2015a) report an estimate of ρ of -0.934 when estimating the SVCJ model. Our findings are thus consistent with estimates from a study that relies on large cross-sections of option contracts.

The estimate of σ in column 1 of Table 3 is 0.3715. Note from Table A1 that the estimate from returns is larger (0.5121). Figure 5 plots the filtered variance paths based on returns and joint estimation and clearly shows that the filtered variance path from joint estimation

[27] Note that κ^Q is negative in Hurn et al. (2015) as well as other studies, which implies a negative long-run variance θ^Q.

is smoother than that from returns. In contrast, the option-based and joint estimates of σ in Panels B and C of Table 4 are generally larger than the return-based estimates in Panel A , and studies that allow for a direct comparison, such as Eraker (2004) and Christoffersen, Jacobs, and Mimouni (2010), confirm this finding. However, our finding is consistent with the results in Andersen, Fusari, and Todorov (2015a), who find that the option-implied variance path is less variable than the physical variance path estimated nonparametrically using high-frequency returns data.[28] The correlation between the two variance paths in Figure 5 is 0.896, which is of course high but generally lower than in existing studies.

We conclude that our estimates of the parameters characterizing the tails of the distribution in the Heston model, σ and ρ, as well as the mean reversion parameter κ, significantly differ from much of the existing literature. This may be due to the sample period. Table 5 therefore compares the results for the 1996-2015 sample (column 1) with those for 1996-2000 (column 2), 1996-2006 (column 3) and 2011-2015 (column 4). These results strongly suggest that the sample period does not explain why our results differ from the existing literature. While there are some differences between the posterior means and variances across the sample periods, the estimates (other than the long-run variance) are similar across sample periods, even though the two alternative samples do not contain the financial crisis and the sample in column 2 is relatively short (1996-2000). One notable difference is the ρ parameter, which is more negative for the most recent sample (2011-2015).

Finally, note from Table 3 that the posterior standard deviations are rather small compared to the posterior means. Table A1 indicates that the posterior standard deviations are higher when using returns only in estimation. Figures 3 and 4 plot the trace for the return-based and joint estimation. The figures clearly indicate convergence of the MCMC for both estimation exercises and for all parameters.

[28] Note that our findings on σ may be related to our findings on ρ. Skewness is determined by ρ but kurtosis is increasing in $|\rho|$ and σ (Das and Sundaram, 1999).

4.2 Inference on Jumps in Returns

We now discuss our estimates of the stochastic volatility model with return jumps (SVJR) model in column 2 of Table 3. Table A2 in the Appendix summarizes parameter estimates for this model from the existing literature. The parameters in both tables are annualized.[29]

First consider the SV parameters κ, θ, ρ, and σ. Note that jumps in returns capture higher moments in the return distribution, and it would therefore not be surprising if the tail parameters ρ and σ were different from the SV model in column 1. However, the estimates of these parameters are similar to those for the SV model. Our estimates indicate the presence of relatively large, but infrequent jumps, which are negative on average, with risk-neutral average jump sizes that are more negative than physical jump sizes. The physical jump size μ_s is equal to -0.0064, the jump risk premium η_{J^s} is equal to 0.0300, and the risk-neutral average jump size $\mu_s^Q = \mu_s - \eta_{J^s}$ is therefore equal to -0.0364. These jumps occur on average twice every three years (λ=0.638). The standard deviation of the jumps σ_s is 9.37%.

Table A2 summarizes estimates of the SVJR model from the existing literature. The literature contains several estimation results for this model, but in most cases they use different information. First, as with existing estimates of the SV model, many of the existing estimates are obtained from returns. Panel A of Table A2 indicates that when estimating based on returns, some studies find jumps that occur more frequently, while others document larger but more infrequent jumps. Our results are closer to the latter studies.

Relatively few studies offer evidence based on options or options and returns jointly. Pan (2002) finds evidence for relatively frequent jumps with a large risk-neutral mean and large standard deviation. Eraker (2004) finds a risk-neutral average jump size of -5% with a

[29] As in the SV model, parameters are often expressed differently in the papers referenced in these tables. For the SV parameters, see the discussion in the previous subsection. For the jump parameters, many studies express the mean and standard deviation of the jump in percentages, which means they are equal to the parameters in Table A2 multiplied by 100. In the case of the jump intensity, some papers, such as Pan (2002) express it as in Table A2; others, such as Eraker, Johannes, and Polson (2003) express the intensity in daily units and need to be multiplied by 252 to obtain the estimates in Table A2.

standard deviation of 16.7%, and these jumps occur on average once every two years. Our jump results are close to those of Eraker (2004), but the estimated standard deviation of the jump size is smaller. Broadie, Chernov, and Johannes (2007) proceed differently: They estimate the jump parameters from (futures) options but keep the SV parameter constrained by theory at their values estimated from returns. Despite these differences in implementation, our findings on return jumps are rather similar to those of Broadie, Chernov, and Johannes (2007).

Table 3 also presents the average equity risk premium for the different models. Recall that the total equity risk premium is given by $\eta_s V_t + \lambda(\bar{\mu}_s - \bar{\mu}_s^Q)$, where $\eta_s V_t$ is due to diffusive risk and $\lambda(\bar{\mu}_s - \bar{\mu}_s^Q)$ is due to jump risk. The average risk premium in the SVJR model is larger than in the SV model. In the SVJR model, approximately 4.5% of the equity risk premium is due to jump risk, with the remaining 6% due to diffusive risk. Broadie, Chernov, and Johannes (2007) report that in their sample, price jump risk premia contribute about 3% per year to an overall equity premium of 8%. These results are obtained using futures data, a different sample and an entirely different approach. Our overall risk premium is a bit higher but the relative contributions of diffusive and jump risk are similar. Pan (2002) also reports comparable estimates, with price jump risk premia that contribute about 3.5% per year to an overall equity premium of 9%.

A variance decomposition shows that in the SVJR model, approximately 13.60% of the variation in returns is due to jumps. Consistent with the existing literature, our results therefore indicate that return jumps are relatively more important for risk premia than for explaining overall return variation.

4.3 Inference on Jumps in Variance

The third and fourth columns of Table 3 present estimates of the SVJV and SVCJ models, respectively. The SVCJ model in column 4 contains jumps in returns and variance that are

correlated. Table A3 summarizes the results of several existing studies that report on this model.[30] Column 3 reports on the SVJV model, which is nested by the SVCJ model: it contains jumps in variance but not in returns, and as a result it has three fewer parameters.

For the SVJV model, the estimate of the risk neutral average jump size $\mu_v^Q = \mu_v - \eta_{J^v}$ is equal to 3.71%. The estimate of the risk premium η_{J^v} is small and negative, and the estimate of the frequency of the jumps λ is equal to 0.6716, implying these jumps occur approximately twice every three years. When adding jumps in returns in the SVCJ model in column 4, the frequency of the jumps slightly increases. The physical jump sizes differ from the magnitudes in the SVJV and SVJR models, but the risk-neutral jump sizes are similar. The estimate of the correlation between the return and variance jump is negative, as expected, at -0.7382. Not surprisingly the posterior mean of η_v is smaller when adding variance jumps. Table A3 indicates that our estimates of μ_v^Q and the jump correlation are within the range of existing studies.

We conclude that our results are consistent with the existing literature (Eraker, Johannes, and Polson, 2003; Eraker, 2004), which finds evidence for jumps in volatility as well as jumps in returns and co-jumps in returns and volatility. Our estimates of variance jump parameters are also for the most part consistent with existing estimates.

4.4 Model Fit

The log likelihoods in Table 3 can be used to assess the importance of return and variance jumps for modeling returns and options, and indicate that the more complex models are supported by the data. The results from returns in Table A1 also confirm the importance of accounting for jumps in returns as well as jumps in variance. These findings are consistent with existing findings on the importance of jumps in returns (Pan, 2002; Bates, 2000) and jumps in variance (Eraker et al., 2003).

[30] See, among others, Eraker, Johannes, and Polson (2003), Eraker (2004), Broadie, Chernov, and Johannes (2007), and Andersen, Fusari, and Todorov (2015a).

Table 3 indicates that jumps in returns and co-jumps in returns and variance result in improvements in the log likelihood. A comparison with Table A1 indicates that jumps are much more important when jointly modeling returns and options. This suggests that these jumps are especially useful to model risk premia, consistent with the findings in Pan (2002).

Finally, it is worth noting that despite the improvements in log likelihood, Table 3 indicates that the richer models with jumps do not substantially outperform the simple SV model in terms of RMSE. The RMSE for the SV model is \$3.134, and the jump models only improve the fit by 4-10 cents, dependent on the model. This confirms the results of Eraker (2004). Singleton (2006) notes that these findings may be due to the fact that we optimize the likelihood rather than minimizing root mean squared error.

Figure 6 provides additional insight into these different measures of model performance by providing scatterplots of the realized pricing kernels. We obtain a time series of the realized pricing kernel for each model by inserting the realized returns and posterior means of the state variables on each day into the pricing kernels, defined as the ratio of the risk-neutral to the physical density implied by the respective model discounted by the risk-free rate. The resulting paths look similar for the four models we study for most of the sample, but there are significant outliers in crisis periods. Figure 6 shows how these outliers vary across models. Panel (a) scatter plots the realized kernel for the SVJR model against the realized kernel for the SV model, and Panel (b) plots the SVCJ kernel against the SV model. It is clear that the paths of the realized kernels for the SVJR and especially the SVCJ models contain many fewer outliers than the path of the SV model. This illustrates the value of the jump models, but because the realized kernels are similar across models on most days, average fit is very similar across models. We leave a more detailed investigation of this issue for future research.

4.5 Conditional Moments

A useful way to highlight the differences between the estimated models is to study the conditional moments. We present results for the conditional variance of returns, the conditional covariance between returns and variance, and the conditional variance of variance. Panel A of Table 6 provides the analytical expressions for these moments.

Panel B of Table 6 presents the averages over our sample for the daily moments for the SV, SVJR, SVJV and SVCJ models. The differences between the models are small for the conditional variance and conditional covariance between returns and variance. However, the models with jumps in variance, especially the SVCJ model, are characterized by a much higher variance of variance. Despite these unconditional differences in the variance of variance across models, the paths of the variance of variance in these models are by construction highly correlated.

5 The Information in Returns and the Cross-Section of Options

We have used the particle MCMC framework and the quantile spot variance method to estimate the option valuation models using the underlying returns and large panels of options. We now report on estimation exercises for different samples. We first compare the posterior density for the dataset that contains options and returns with the posterior density from returns only. Then we investigate how restricting the option sample in the moneyness and/or maturity dimension affects the posterior density and option fit. For brevity we limit our discussion to the SV model.

5.1 Posterior Densities from Returns and Options

Figure 7 compares the posterior density for the parameters of the stochastic volatility estimated on returns only and returns and options jointly. We discuss the posterior densities of the five parameters in the SV model that can be identified from returns as well as options: κ, θ, σ, ρ, and the price of diffusive equity risk η_s. The posterior means and standard deviations are listed in Tables 3 and A1 respectively. The blue solid line labeled "P" in Figure 7 shows the posterior density estimated from returns and the black dashed line labeled "P&Q - all" shows the posterior density for the joint estimation.

Figure 7 and Tables 3 and A1 illustrate that while the posterior means for the two samples are similar for the θ and η_s parameters, they are significantly different for κ, σ, and ρ. Including options in estimation yields a more persistent variance, a more negative skewness parameter ρ, and a lower kurtosis parameter σ. The posterior densities in Figure 7 also suggest that the option data are more informative about κ, σ, and ρ than the return data. This is confirmed by the posterior standard deviations in Tables 3 and A1. For η_s, the posterior standard deviations are very similar. This finding is not surprising as the option data do not contain additional information on this parameter. For θ, the return-based posterior standard deviation is actually a bit lower.

5.2 The Composition of the Option Sample and the Posterior Density

The motivation for our estimation method is that while cross-sectional information is helpful for identification of certain model parameters, conventional implementations of the particle filter, as well as other existing estimation methods, are computationally very expensive when using long time series and large option panels. As previously discussed, this is one of the reasons why many existing studies have restricted the time-series and cross-

sectional dimension of the data. A natural question is therefore if these data restrictions affect inference.

We focus on the five parameters in the SV model discussed above (κ, θ, σ, ρ, η_s) and the price of diffusive variance risk η_v. Figure 7 presents the posterior densities for three option samples. The black dashed line labeled "P&Q - all" represents the posterior density for the joint return and option sample discussed in the previous section. We also plot the posterior densities for two restricted option samples. Following the existing literature, we restrict either the moneyness dimension or the maturity dimension. The green solid line labeled "ATM" plots the posterior density for the sample with ATM options but all maturities, and the red dotted line labeled "SM" is based on the sample with short maturity options but all moneyness.[31] Panel C of Table 1 presents the sample sizes as well as descriptive statistics for these restricted samples. Columns 5 and 6 in Table 5 present the posterior means and standard deviations. The ATM and SM datasets are smaller than the one used in our benchmark analysis ("ALL"), but because the likelihood is scaled back by the number of options in the dataset, this does not affect the relative weight of returns and options in the posterior density. The critical difference between the datasets that determines the relative weight of returns and options is the cross-sectional composition of the option sample.

The top two panels of Figure 7 show that the posterior densities of the κ and θ parameters are very similar for the ATM and ALL samples. However, the posterior density for κ based on the SM sample is very different and more similar to the posterior density based on returns. The two panels in the middle row of Figure 7 show the posterior density for the σ and ρ parameters. The posterior mean of ρ is more negative for all three option samples compared to the return sample, and the posterior standard deviation is smaller in the option samples. For the σ parameter on the other hand, the SM sample leads to a larger posterior mean

[31] For the ATM sample, we exclude options with moneyness outside the 0.98-1.02 range. We select up to six options for each of the five maturity bins in Table 1. For the SM sample, we rely on a sample with options with maturity less than 30 days in Panel B of Table 1. If options with maturity less than 30 days are not available we include the options for the shortest maturity longer than 30 days.

than the return sample, while for the two other samples the posterior mean of σ is smaller than the one from returns. These findings regarding σ are not entirely surprising, because we know that the implied volatility surface changes a lot more over time at short maturities.

Finally, the bottom panel presents the results for the prices of risk η_s and η_v. For η_s, the SM sample yields very different posterior means compared to the other two option samples (ALL and ATM). Surprisingly, the posterior density for the SM sample also differs from the one obtained from returns. For η_v, the posterior density for the SM sample is once again the outlier. Note that we do not have a posterior density for η_v when estimating on returns only.

5.3 Discussion

As mentioned above, many existing studies use samples that consist exclusively of short-maturity and/or at-the money options because of computational constraints. Our findings show that those constraints affect parameter inference. This may explain some of the differences between model parameter estimates in the existing literature (See Tables 4, A2, and A3).

Our most important finding is that option samples that exclusively contain short-maturity options result in very different inference. Recall from our discussion of Table 5 that parameter estimates were fairly similar for different sample periods, even when the sample period is short. Taken together, these findings strongly suggest that when restricting the sample to address computational constraints, priority should be given to including a large cross-section of different option maturities.

A second important conclusion is that the composition of the option sample seems to affect inference on some parameters more than others. Specifically, the posterior mean for κ and σ is more sensitive to the option sample than the posterior mean for θ and ρ. Also, while the posterior mean for θ based on joint estimation is similar to the one based on returns,

this is not the case for κ, σ, and ρ.

While some of these findings are perhaps not surprising, others are very different from what we expected. Most notably, we expected data restrictions in the moneyness dimension to have more serious consequences for inference on σ and ρ, and data restrictions in the maturity dimension to affect inference on κ and θ. Instead, inference on θ and ρ is similar across the three options samples, and the implications of restricting the sample in the maturity dimension have implications for many model parameters.

Table 7 provides additional evidence on the implications of these different sample restrictions. It documents the implications of the sample composition for option fit by maturity and moneyness. The RMSEs are computed based on the model parameters and 5000 simulations of the posterior spot variance. The first column documents (in-sample) fit, using estimates based on the 1996-2015 sample period with all options.[32] The other columns report RMSEs for the same option sample, but they use the parameters from the subsamples in Table 5. Table 7 confirms that the strongest implications result from restricting the maturity dimension of the option sample used in estimation (the column labeled "SM"). This results in a much higher average pricing error ($5.091), which is largely due to the very large pricing errors for long-maturity options. Restricting the sample to short-maturity options also leads to higher average pricing errors, but the deterioration in fit is modest. Table 7 also reports on the option fit for the 1996-2015 sample period when using the parameter estimates from the shorter samples in Table 5. The deterioration in fit when estimating on shorter time periods is surprisingly small, confirming that the model parameters can be estimated rather precisely using shorter sample periods.

An important caveat to our findings on the relative importance of the moneyness and maturity dimension for inference and option fit is that these results may depend on the choice

[32]Note that the overall average RMSE σ_c in the first column of Table 7 is $2.997, whereas it is $3.134 in Table 3. Table 3 reports the posterior mean, whereas the fit in Table 7 is based on simulations of the spot variance, similar to the method used for the other columns.

of loss function. Another caveat is that our restricted samples still include a large number of options and several options on a given day. Our conclusion regarding the implications of data restrictions used in the existing literature may therefore be conservative, because existing studies sometimes use smaller samples. This has important implications. For example, Panel (d) of Figure 7 indicates that the posterior mean for ρ is very similar for the three option samples, close to -0.9. This contrasts with the estimates from the existing literature in Table 4, which indicates that many option-based estimates of ρ are close to the posterior mean from returns data. We verified that the results in Panel (d) are contingent on sufficiently large option samples. For very small option samples, we obtain posterior means of ρ that are similar to the return-based estimates.

6 Conclusion

The use of option panels holds great promise for identifying the return distribution and conditional risk premia, because the cross-section of options contains information on different states of nature. However, the estimation of dynamic option valuation models with latent state variables is challenging due to the complexity of the models and the richness of the available option data. To overcome these computational challenges, we propose an orthogonal particle MCMC framework with a novel filtering approach, featuring particle weights based on the quantiles of the filtered distribution. We use a Monte Carlo study to demonstrate that our approach is reliable.

We illustrate our approach by estimating a class of models with jumps in returns and variance (Duffie, Pan, and Singleton, 2000) using twenty years of daily data, and almost thirty option contracts per day with a wide cross-section of maturities and moneyness. It was previously impossible to estimate these models using large option panels in full generality, while solving the filtering problem. We confirm that both return and variance jumps, and co-jumps in particular, are important. We obtain more precise estimates of the parameter

that determines the diffusive price of variance risk. These estimates are smaller compared to existing parametric studies, but more consistent with existing nonparametric evidence. Our estimates of the diffusive parameters characterizing skewness, kurtosis, and the persistence of the variance process also differ from most existing studies.

We next use our approach to show that the composition of the option sample affects estimation results, and that this explains the differences between our estimates and the existing literature. In the face of computational constraints, the sample can be restricted in several ways. One can restrict the length of the time series, or one can restrict the moneyness and/or maturity dimensions of the option samples. Little is known about the implications of these data restrictions for parameter inference and option fit. We find that to identify the models we consider in this paper, it is most critical to include options with different maturities in the sample. Somewhat surprisingly, restricting the moneyness dimension of the option sample affects inference much less. Parameter estimates are similar when shorter samples are used in estimation, suggesting that the length of the sample period may also be less critical for inference than commonly thought. Restricting the maturities of the options used in estimation also has the largest impact on model fit, at least for the loss function we use in estimation. We leave a more detailed study of the optimal composition of the option sample and the trade-off between the length of the sample period and the size of the option cross-sections for future research.

In future work we intend to use the computational advantage of our approach, combined with a Particle Gibbs approach as discussed in Section 3.5.2, to study richer models with multiple volatility factors (Bates, 2000), time varying jump intensities (Pan, 2002; Bates, 2006; Christoffersen et al., 2021), tail factors (Andersen, Fusari, and Todorov, 2015a), and different parametric specifications of the jump processes (Bates, 2012; Andersen, Fusari, and Todorov, 2017). Our approach can also be combined with the use of high-frequency returns (Bates, 2019) or volatility estimates based on high-frequency data (Andersen, Fusari,

and Todorov, 2015b). We also plan to investigate in more detail the implications of the choice of error specification (Andersen, Fusari, Todorov, and Varneskov, 2021) and loss function (Hurn, Lindsay, and McClelland, 2015). Finally, a detailed comparison between the computational efficiency and properties of our approach and that of existing methods in Eraker (2004), Bates (2000), Andersen, Fusari, and Todorov (2015a,b, 2017, 2020) and Andersen, Fusari, Todorov, and Varneskov (2019) is needed.

References

Aït-Sahalia, Yacine, Julio Cacho-Diaz, and Roger Laeven, 2015, Modeling Financial Contagion Using Mutually Exciting Jump Processes, *Journal of Financial Economics* 117, 585–606.

Aït-Sahalia, Yacine, and Robert Kimmel, 2007, Maximum Likelihood Estimation of Stochastic Volatility Models, *Journal of Financial Economics* 83, 413–452.

Aït-Sahalia, Yacine, Chenxu Li, and Chen Xu Li, 2020, Implied Stochastic Volatility Models, *The Review of Financial Studies* 34, 394–450.

Aït-Sahalia, Yacine, Chenxu Li, and Chen Xu Li, 2021, Closed-form Implied Volatility Surfaces for Stochastic Volatility Models with Jumps, *Journal of Econometrics* 222, 364–392.

Andersen, Leif, 2007, Efficient Simulation of the Heston Stochastic Volatility Model, Working Paper.

Andersen, Torben G., Luca Benzoni, and Jesper Lund, 2002, An Empirical Investigation of Continuous-Time Models for Equity Returns, *Journal of Finance* 57, 1239–1284.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015a, Parametric Inference and Dynamic State Recovery From Option Panels, *Econometrica* 83, 1081–1145.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2015b, The Risk Premia Embedded in Index Options, *Journal of Financial Economics* 117, 558–584.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2017, Short-Term Market Risks Implied by Weekly Options, *Journal of Finance* 72, 1335–1386.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2020, The Pricing of Tail Risk and the Equity Premium: Evidence From International Option Markets, *Journal of Business & Economic Statistics* 38, 662–678.

Andersen, Torben G., Nicola Fusari, Viktor Todorov, and Rasmus T. Varneskov, 2019, Unified Inference for Nonlinear Factor Models from Panels with Fixed and Large Time Span, *Journal of Econometrics* 212, 4–25.

Andersen, Torben G., Nicola Fusari, Viktor Todorov, and Rasmus T. Varneskov, 2021, Spatial Dependence in Option Observation Errors, *Econometric Theory* 37, 205247.

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein, 2010, Particle Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72, 269–342.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical Performance of Alternative Option Pricing Models, *Journal of Finance* 52, 2003–2049.

Bandi, F. M., and R. Reno, 2016, Price and Volatility Co-Jumps, *Journal of Financial Economics* 119, 107–146.

Bates, David S., 1996, Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutschemark Options, *Review of Financial Studies* 9, 69–107.

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

Bates, David S., 2003, Empirical Option Pricing: A Retrospection, *Journal of Econometrics* 116, 387–404.

Bates, David S., 2006, Maximum Likelihood Estimation of Latent Affine Processes, *Review of Financial Studies* 19, 909–965.

Bates, David S., 2012, U.S. Stock Market Crash Risk, 1926-2010, *Journal of Financial Economics* 105, 229–259.

Bates, David S., 2019, How Crashes Develop: Intradaily Volatility and Crash Evolution, *The Journal of Finance* 74, 193–238.

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

Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2007, Model Specification and Risk Premia: Evidence From Futures Options, *Journal of Finance* 62, 1453–1490.

Broadie, Mark, and Ozgur Kaya, 2006, Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes, *Operations Research* 54, 217–231.

Buchner, Matthias, and Bryan Kelly, 2022, A Factor Model for Option Returns, *Forthcoming, Journal of Financial Economics* .

Carr, Peter, and Dilip B. Madan, 1998, Option Valuation Using the Fast Fourier Transform, *Journal of Computational Finance* 2, 61–73.

Carvalho, Carlos, Michael Johannes, Hedibert Lopes, and Nicholas Polson, 2010, Particle Learning and Smoothing, *Statistical Science* 25, 88–106.

Chacko, George, and Luis M. Viceira, 2003, Spectral GMM Estimation of Continuous-Time Processes, *Journal of Econometrics* 116, 259–292.

Chernov, Mikhail, A. Ronald Gallant, Eric Ghysels, and George Tauchen, 2003, Alternative Models for Stock Price Dynamics, *Journal of Econometrics* 116, 225–257.

Chernov, Mikhail, and Eric Ghysels, 2000, A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation, *Journal of Financial Economics* 56, 407–458.

Christoffersen, Peter, Bruno Feunou, Yoontae Jeon, and Chayawat Ornthanalai, 2021, Time-Varying Crash Risk Embedded in Index Options: The Role of Stock Market Liquidity, *Review of Finance* 25, 1261–1298.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well, *Management Science* 55, 1914–1932.

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

Crisóstomo, Ricardo, 2018, Speed and biases of fourier-based pricing choices: a numerical analysis, *International Journal of Computer Mathematics* 95, 1565–1582.

Das, Sanjiv Ranjan, and Rangarajan K. Sundaram, 1999, Of Smiles and Smirks: A Term Structure Perspective, *Journal of Financial and Quantitative Analysis* 34, 211–239.

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

Eraker, Bjørn, 2004, Do Stock Prices and Volatility Jump? Evidence From Spot and Option Prices, *Journal of Finance* 59, 1367–1403.

Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The Impact of Jumps in Equity Index Volatility and Returns, *The Journal of Finance* 58, 1269–1300.

Ferriani, Fabrizio, and Sergio Pastorello, 2012, Estimating and Testing Nonaffine Option Pricing Models with a Large Unbalanced Panel of Options, *The Econometrics Journal* 15, 171–203.

Feunou, Bruno, and Cedric Okou, 2018, Risk-neutral moment-based estimation of affine option pricing models, *Journal of Applied Econometrics* 33, 1007–1025.

Fulop, Andras, Junye Li, and Jun Yu, 2015, Self-Exciting Jumps, Learning, and Asset Pricing Implications, *Review of Financial Studies* 28, 876–912.

Glasserman, Paul, and Kyoung-Kuk Kim, 2011, Gamma Expansion of the Heston Stochastic Volatility Model, *Finance and Stochastics* 15, 267–296.

Hastings, W K, 1970, Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika* 57, 97–109.

Heston, Steven L., 1993, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 327–343.

Hurn, A. S., K. A. Lindsay, and A. J. McClelland, 2015, Estimating the Parameters of Stochastic Volatility Models Using Option Price Data, *Journal of Business and Economic Statistics* 33, 579–594.

Israelov, Roni, and Bryan Kelly, 2017, Forecasting the Distribution of Option Returns, *Working Paper, University of Chicago* .

Johannes, Michael, and Nicholas Polson, 2009a, MCMC Methods for Continuous-Time Financial Econometrics, in *Handbook of Financial Econometrics, Vol 2*, 1–72.

Johannes, Michael, and Nicholas Polson, 2009b, Particle Filtering, in *Handbook of Financial Time Series*, 1015– 1031.

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

Jones, Christopher, 2006, A Nonlinear Factor Analysis of S&P 500 Index Option Returns, *Journal of Finance* 61, 2325–2363.

Jones, Christopher S., 2003, The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets, *Journal of Econometrics* 116, 181–224.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, 1998, Stochastic volatility: likelihood inference and comparison with arch models, *The review of economic studies* 65, 361–393.

Liu, Jie, Wilson Wang, and Fai Ma, 2011, A Regularized Auxiliary Particle Filtering Approach for System State Estimation and Battery Life Prediction, *Smart Materials and Structures* 20, 075021.

Malik, Sheheryar, and Michael K. Pitt, 2011, Particle Filters for Continuous Likelihood Evaluation and Maximisation, *Journal of Econometrics* 165, 190–209.

Martino, Luca, Víctor Elvira, David Luengo, Jukka Corander, and Francisco Louzada, 2016, Orthogonal Parallel MCMC Methods for Sampling and Optimization, *Digital Signal Processing* 58, 64–84.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, 1953, Equation of State Calculations by Fast Computing Machines, *Journal of Chemical Physics* 21, 1087–1092.

Pan, Jun, 2002, The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-series Study, *Journal of Financial Economics* 63, 3–50.

Pitt, Michael K., and Neil Shephard, 1999, Filtering via Simulation: Auxialiary Particle Filters, *Journal of the American Statistical Association* 94, 590–599.

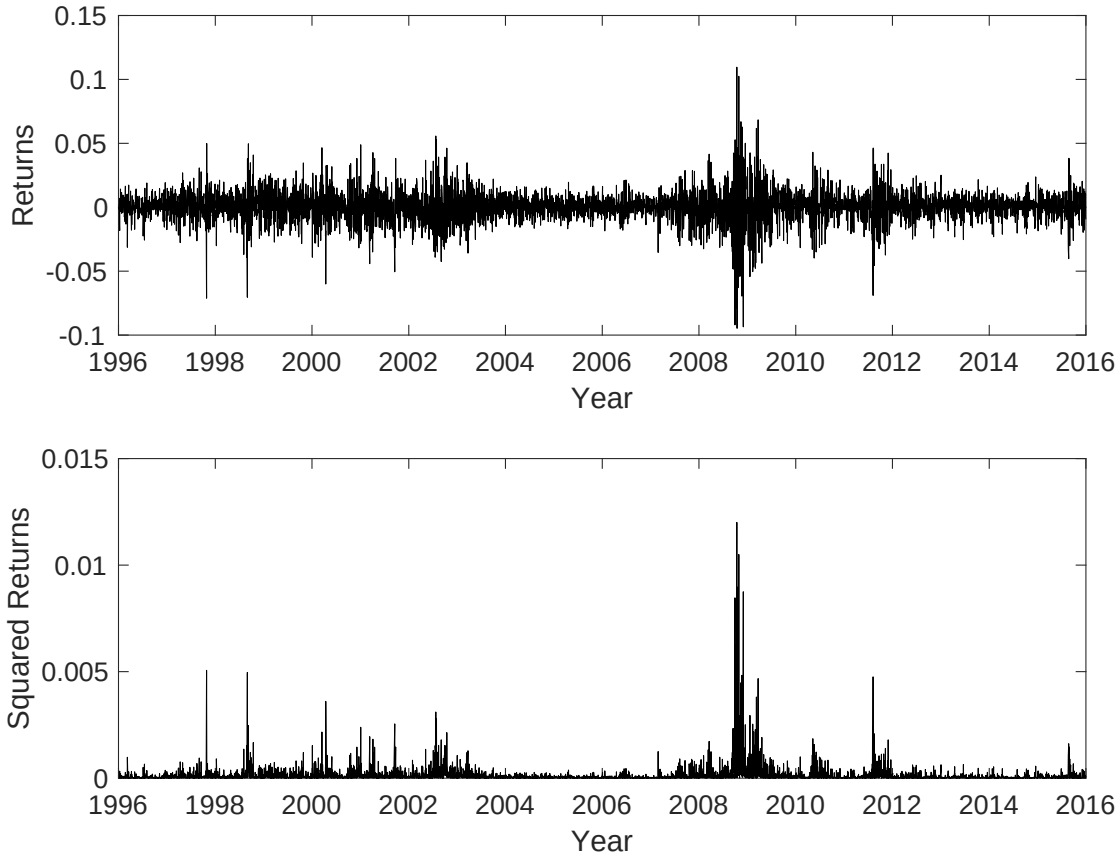
Roberts, Gareth O, and Jeffrey S Rosenthal, 2001, Optimal Scaling for Various Metropolis-Hastings Algorithms, *Statistical science* 16, 351–367.

Roberts, Gareth O., and Jeffrey S. Rosenthal, 2009, Examples of Adaptive MCMC, *Journal of Computational and Graphical Statistics* 18, 349–367.

Singleton, Kenneth J., 2006, *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment, Princeton University Press*.

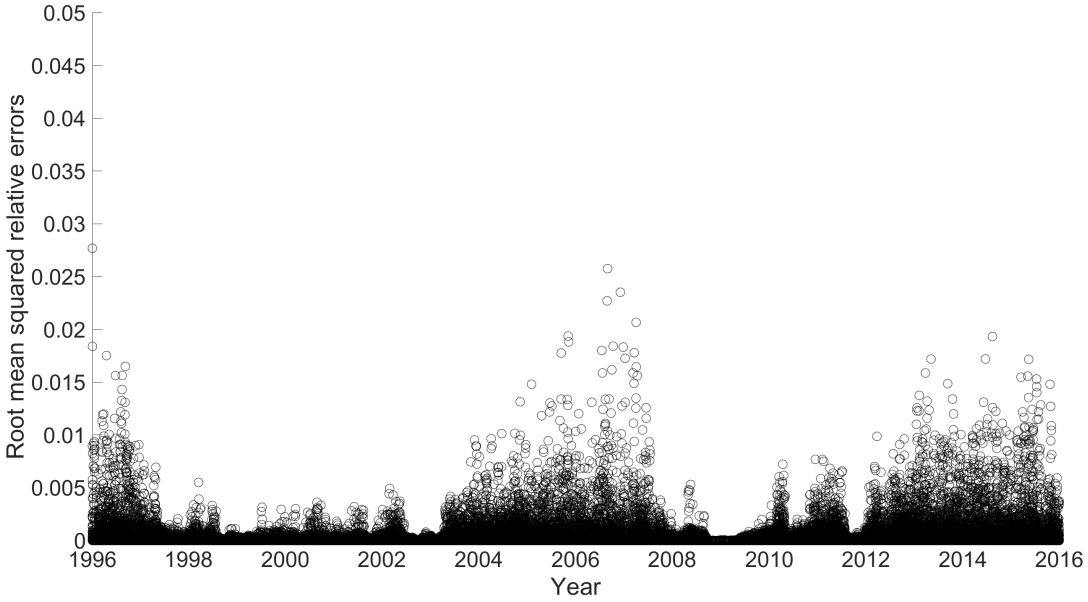
Vihola, Matti, 2012, Robust Adaptive Metropolis Algorithm with Coerced Acceptance Rate, *Statistics and Computing* 22, 997–1008.

Figure 1: Daily S&P500 Returns and Squared Returns 1996-2015



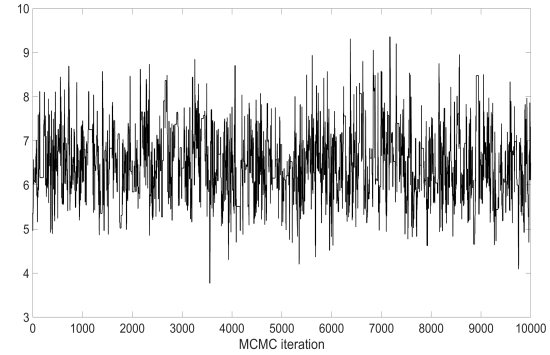
Notes: We plot daily log returns in the top panel. In the bottom panel, we plot squared daily log returns. The sample period is from January 1, 1996 until December 31, 2015.

Figure 2: QSV Pricing Errors for the SV Model.

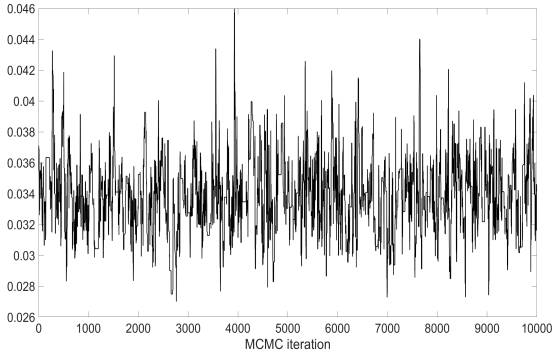


Notes: We compute Root Mean Squared relative option pricing errors for the QSV method computed over the twelve quantiles, based on the implementation of the SV model. The model parameters are fixed at the posterior parameter means of the SV model in Table 3.

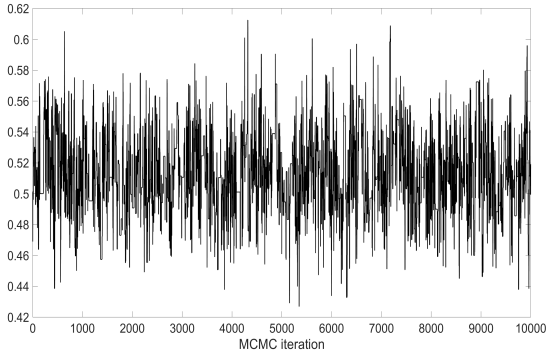
Figure 3: Parameter Trace for SV Model Parameters. Return-Based Estimation



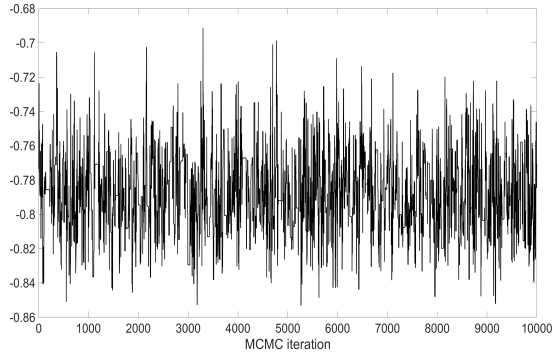
(a) κ



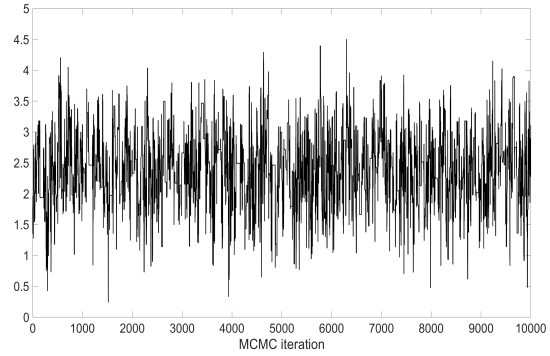
(b) θ



(c) σ



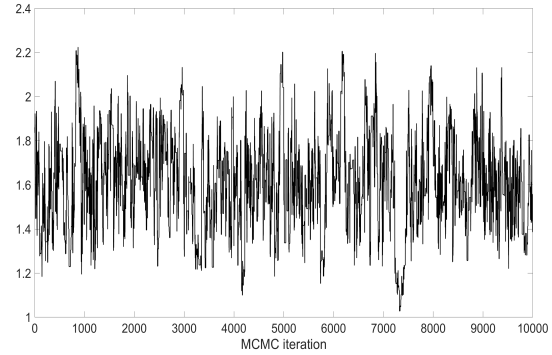
(d) ρ



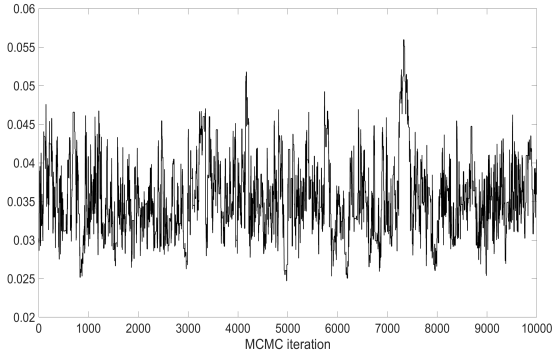
(e) η_s

Notes: We plot the traces for each parameter in the SV model. We use 15,000 iterations. The first 5,000 of the iterations are treated as burn-in.

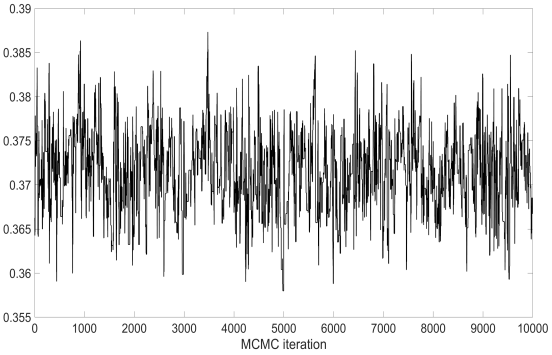
Figure 4: Parameter Trace for SV Model Parameters. Joint Estimation



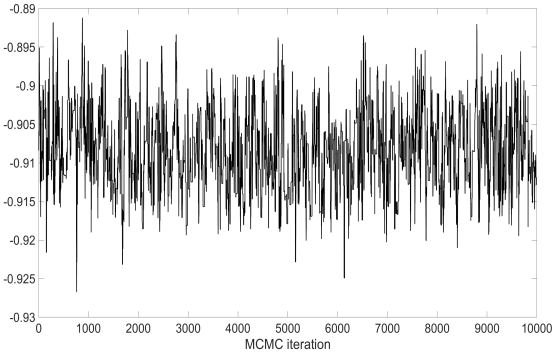
(a) κ



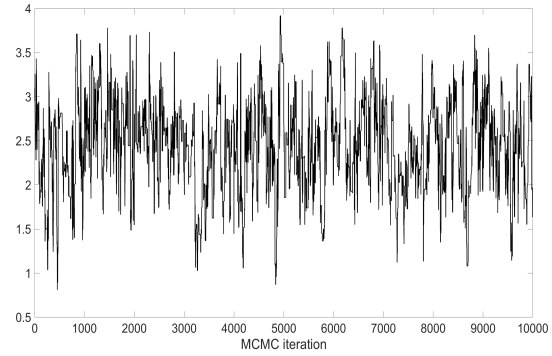
(b) θ



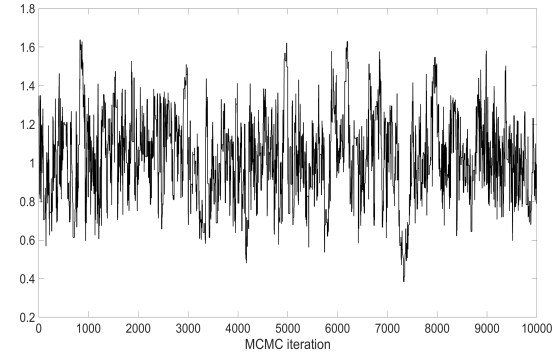
(c) σ



(d) ρ



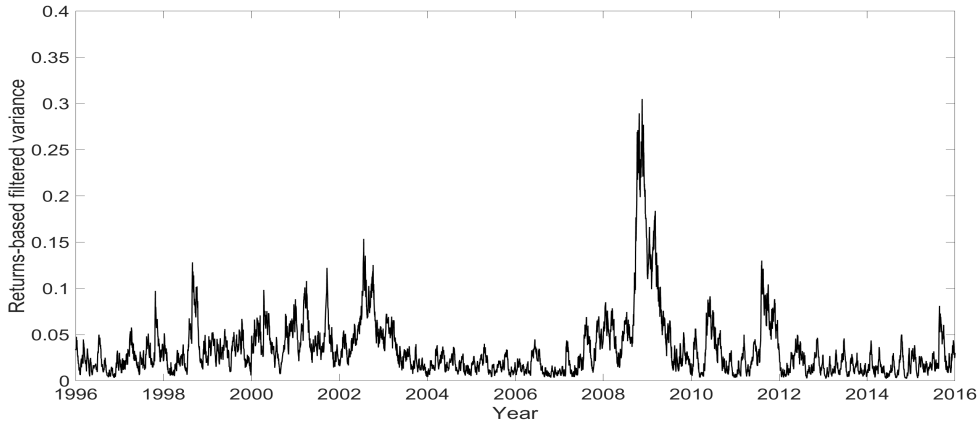
(e) η_s



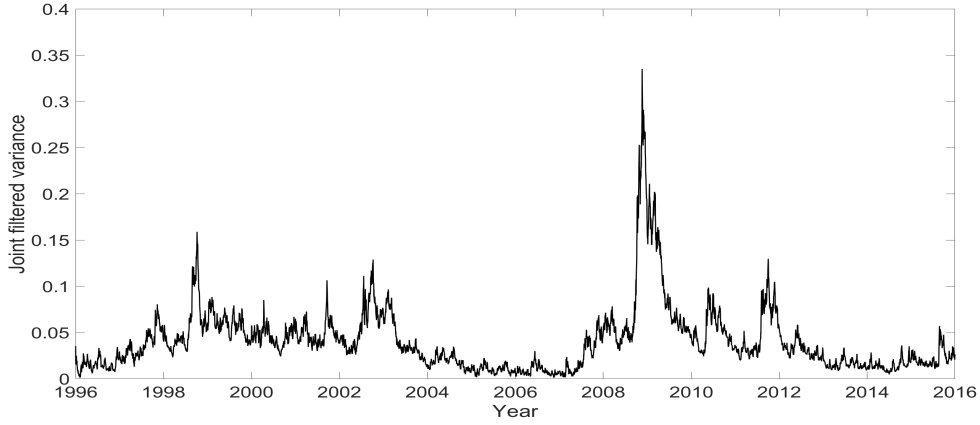
(f) η_v

Notes: We plot the traces for each parameter in the SV model. We use 15,000 iterations. The first 5,000 of the iterations are treated as burn-in.

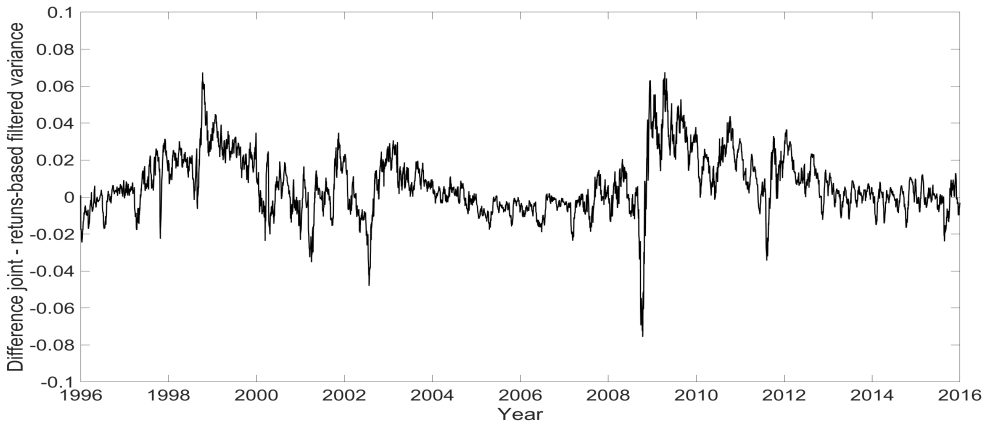
Figure 5: Filtered Variance Paths. SV Model



(a) Returns-based



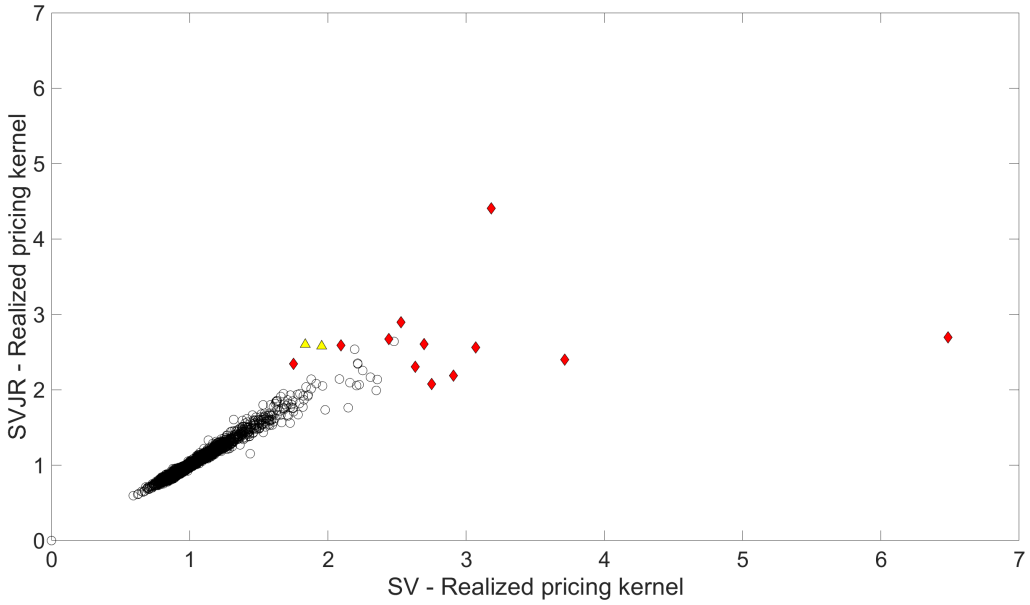
(b) Joint



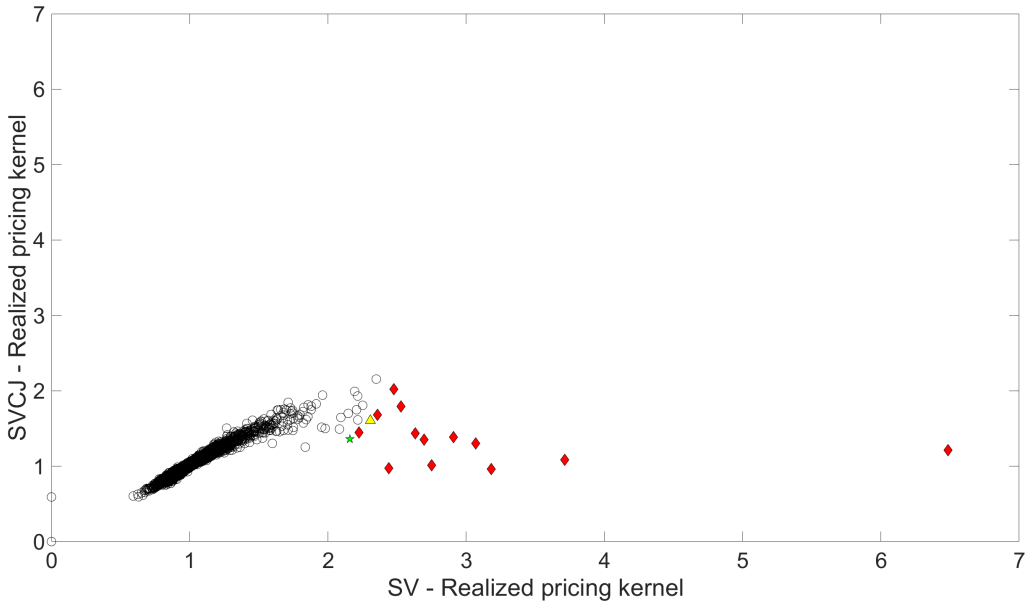
(c) Difference Joint - Returns-based

Notes: We plot the filtered variance path estimated from returns in the top panel, and the variance path from joint estimation in the second panel. The bottom panel plots the difference between the variance estimated from returns and the variance from joint estimation.

Figure 6: Scatter Plots of Realized Pricing Kernels



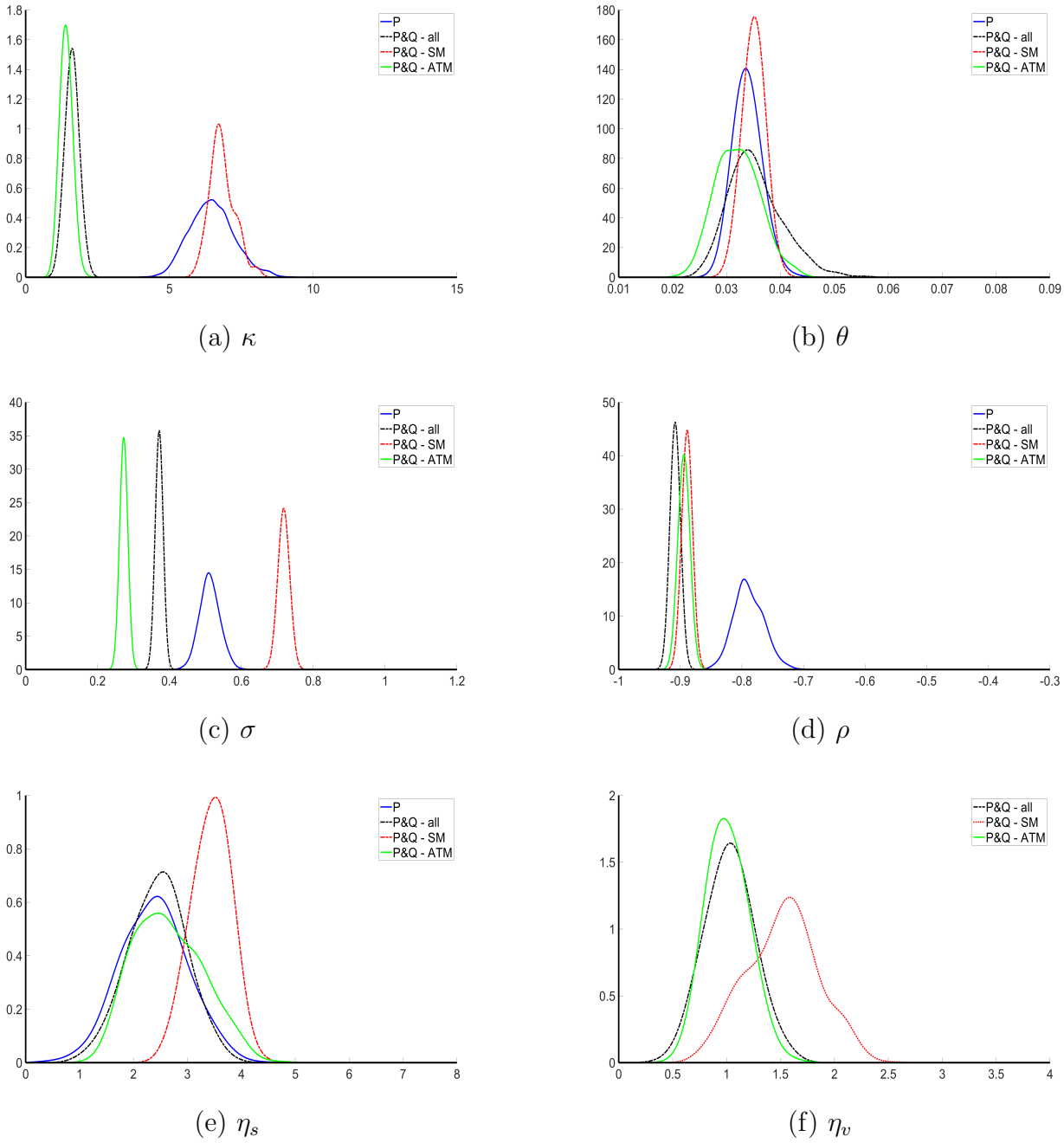
(a) SV versus SVJR



(b) SV versus SVCJ

Notes: We scatter plot the realized pricing kernel for the SV model versus the SVJR (top) and SVCJ (bottom) models respectively. The sample period is from January 1, 1996 until December 31, 2015. We highlight outliers according to different crisis periods: yellow triangles for the Dotcom crisis (2000 - 2002), red diamonds for the Global Financial Crisis (2007-2008), and a green star for July 20, 1998, which is in the period following the Russian crisis and during the LTCM meltdown. The other sample days are indicated using black circles.

Figure 7: Posterior Densities from Returns and Various Option Samples



(a) κ (b) θ

(c) σ (d) ρ

(e) η_s (f) η_v

Notes: Posterior density functions for the SV model parameters based on joint estimation with returns and either the full or a restricted option dataset. The posterior density functions from return-based estimation are given in plain blue. The joint estimation is represented in black dashed. The densities with ATM options but all maturities and the sample with short maturity options but all moneyness are displayed in green solid and red dotted, respectively.

Table 1: Return and Option Data

Panel A: Return Data

	Mean	StdDev	Skewness	Kurtosis	Max	Min
Index Returns	0.0797	0.1958	-0.0535	10.7980	0.1158	-0.0904

Panel B: Option Data

	Maturity (days)					
Moneyness (K/S)	5-30	30-60	60-90	90-180	180-365	All
Number of Option Contracts						
0.85-0.90	2941	4837	4271	4490	4177	20716
0.90-0.95	3828	4942	4615	4753	4542	22680
0.95-1.00	3986	4949	4717	4837	4831	23320
1.00-1.05	3984	4971	4687	4839	4780	23261
1.05-1.10	2450	4758	4441	4575	4416	20640
1.10-1.15	694	2098	2479	3646	3865	12782
All	17883	26555	25210	27140	26611	123399
Average Call Prices						
0.85-0.90	157.35	160.47	168.64	175.77	194.47	171.34
0.90-0.95	93.06	101.65	112.06	123.90	146.15	115.36
0.95-1.00	37.51	48.84	59.14	75.61	100.70	64.36
1.00-1.05	9.87	20.12	30.37	44.57	71.46	35.28
1.05-1.10	2.17	4.91	9.73	19.64	40.54	15.40
1.10-1.15	1.81	2.75	4.91	9.46	24.00	8.59
All	50.30	56.46	64.14	74.83	96.22	68.39
Average Implied Volatilities						
0.85-0.90	0.34	0.28	0.26	0.25	0.24	0.28
0.90-0.95	0.27	0.24	0.23	0.23	0.22	0.24
0.95-1.00	0.21	0.20	0.20	0.21	0.21	0.21
1.00-1.05	0.18	0.18	0.18	0.19	0.19	0.18
1.05-1.10	0.20	0.16	0.16	0.17	0.18	0.17
1.10-1.15	0.32	0.21	0.18	0.17	0.17	0.21
All	0.25	0.21	0.20	0.20	0.20	0.22

Panel C: Restricted Samples. Descriptive Statistics

	Total No. of Options	Avg.Maturity($days$)	Avg.Moneyness(K/S)
ALL	123399	111.58	0.9899
ATM	25670	104.9550	1.0002
SM	19099	19.1151	0.9779

Notes: Panel A reports descriptive statistics for the sample of index returns. The mean and standard deviation are annualized. Panel B reports the number of contracts, average call price, and average implied volatility in the option data set where we choose the most liquid (highest trading volume) option within each moneyness-maturity range. Moneyness is defined as K/S. Due to the fact that OTM options are more heavily traded, this data set mainly consists of OTM call and OTM put options. Panel C presents the sample size for the restricted samples used in Figures 7 and Table 5. ALL represents the sample with all options. ATM represents the sample with ATM options but all maturities and SM represents the sample with short maturity options but all moneyness.

Table 2: Monte Carlo Results

Parameter	True value	Median	Q25	Q75
κ	2.0000	2.0144	1.9089	2.1613
θ	0.0350	0.0361	0.0333	0.0378
σ	0.3800	0.3669	0.3647	0.3693
ρ	-0.9000	-0.9277	-0.9324	-0.9226
η_s	2.5000	2.5302	2.3948	2.6702
η_v	1.0000	0.9956	0.8855	1.1384
Parameter	True value	Median	Q25	Q75
κ	2.0000	2.0537	1.9047	2.1757
θ	0.0350	0.0358	0.0341	0.0385
σ	0.4900	0.4642	0.4551	0.4682
ρ	-0.7000	-0.7288	-0.7397	-0.7233
η_s	2.5000	2.5074	2.3946	2.5709
η_v	1.0000	0.9911	0.8603	1.0759

Notes: We report Monte Carlo results based on two configurations of the SV model. For each of the two parameter settings, we simulate 50 samples of returns and option panels samples for a one-year period. The maturity and moneyness structure of the simulated option panels are based on the 2015 option sample, implying 6731 options used for estimation of each SV model.

Table 3: Parameter Estimates Based on Joint Estimation Using Returns and Options

	SV	SVJR	SVJV	SVCJ
κ	1.6999	1.5382	1.1777	0.9613
	(0.2010)	(0.1096)	(0.1421)	(0.0596)
θ	0.0334	0.0236	0.0310	0.0282
	(0.0040)	(0.0015)	(0.0041)	(0.0032)
σ	0.3715	0.3682	0.3651	0.3586
	(0.0047)	(0.0046)	(0.0038)	(0.0075)
ρ	-0.9085	-0.9187	-0.9242	-0.9191
	(0.0045)	(0.0039)	(0.0047)	(0.0026)
η_s	2.6623	2.5562	2.3032	2.7095
	(0.5168)	(0.4504)	(0.3659)	(0.4556)
η_v	1.1156	1.2466	0.5569	0.6239
	(0.2017)	(0.1123)	(0.1425)	(0.0895)
λ		0.6380	0.6716	0.7249
		(0.0347)	(0.0941)	(0.0316)
μ_s		-0.0064		0.0320
		(0.0106)		(0.0094)
σ_s		0.0937		0.0872
		(0.0053)		(0.0057)
η_{J^s}		0.0300		0.0562
		(0.0097)		(0.0092)
μ_v			0.0314	0.0727
			(0.0092)	(0.0026)
η_{J^v}			-0.0057	0.0562
			(0.0095)	(0.0051)
ρ_J				-0.7382
				(0.1785)
σ_c	3.1345	3.0308	3.0963	3.0978
	(0.0345)	(0.0221)	(0.0286)	(0.0266)
Diffusive ERP	0.0889	0.0603	0.0714	0.0764
Jump ERP		0.0450		0.0101
Loglikelihood	2723.58	2866.25	2764.37	2878.71

Notes: We report parameter estimates based on the joint likelihood from returns and options for the SV, SVJR, SVJV and SVCJ models. Parameters are annualized and under the physical measure. The posterior standard deviations for each parameter are reported in parentheses. σ_c is the option RMSE in dollars. Diffusive ERP and Jump ERP represent the average equity risk premia due to the diffusive and jump components.

Table 4: Parameter Estimates in Existing Studies: The Heston SV Model.

Panel A: Based on Returns

Author	Period	κ	θ	σ	ρ
ABL	1953-1996	4.032	0.017	0.202	-0.380
CV	1980-2000	14.282	0.033	5.193	-0.629
CGGT	1953-1999	3.276	0.015	0.151	-0.279
EJP	1980-2000	5.821	0.023	0.361	-0.397
Jones	1986-2000	3.704	0.026	0.524	-0.603
Eraker	1987-1990	4.284	0.022	0.277	-0.373
Bates_1	1953-1996	5.940	0.016	0.315	-0.579
CJM	1996-2004	6.520	0.035	0.460	-0.771

Panel B: Based on Options

Author	Period	κ	θ	σ	ρ	η_v
BCC	1988-1991	1.150	0.040	0.390	-0.640	
Bates_2	1988-1993	1.490	0.067	0.742	-0.571	
	1988-1993	1.260	0.071	0.694	-0.587	2.28
BCJ	1987-2003	7.056	0.019	0.361	-0.397	
CJM	1996-2004	2.879	0.063	0.537	-0.704	

Panel C: Based on Returns and Options

Author	Period	κ	θ	σ	ρ	η_v
Pan	1989-1996	7.100	0.014	0.320	-0.530	7.600
		-0.500*	-0.195*			
Eraker	1987-1990	4.788	0.049	0.554	-0.569	2.520
		2.268*	0.103*			
ASK	1990-2004	5.070	0.046	0.480	-0.767	
HLM	1990-2007	1.879	0.037	0.386	-0.741	1.9894
		-0.111*	-0.630*			

Notes: We report parameters for the SV model in existing studies. Estimates in Panel A are physical values. Estimates in Panel B are risk-neutral values. In Panel C, estimates with a star (*) indicate risk-neutral values; the others are physical values. All parameters are annualized. BCC: Bakshi, Cao, and Chen (1997), based on the S&P 500; ABL: Andersen, Benzoni, and Lund (2002), based on the S&P 500; CV: Chacko and Viceira (2003), based on the S&P 500; CGGT: Chernov, Gallant, Ghysels, and Tauchen (2003), based on the DJIA; EJP: Eraker, Johannes, and Polson (2003), based on the S&P 500; Jones: Jones (2003), based on the S&P 100; Eraker: Eraker (2004), based on the S&P 500; Bates_1: Bates (2006), based on the S&P 500; Bates_2: Bates (2000), based on the S&P 500. The second row of Bates_2 presents the estimates with dynamic constraint on the spot variance; CJM: Christoffersen, Jacobs, and Mimouni (2010), based on the S&P 500; BCJ: Broadie, Chernov, and Johannes (2007), based on the S&P 500; Pan: Pan (2002), based on the S&P 500; ASK: Aït-Sahalia and Kimmel (2007), based on the S&P 500; HLM: Hurn, Lindsay, and McClelland (2015), based on the S&P 500.

Table 5: Parameter Estimates for the SV model. Alternative Sample Periods and Option Subsamples.

	1996-2015	1996-2000	1996-2006	2011-2015	SM	ATM
κ	1.6999	1.3061	1.4799	1.9240	6.8354	1.4078
	(0.2010)	(0.2913)	(0.2119)	(0.2771)	(0.4304)	(0.1839)
θ	0.0334	0.0401	0.0296	0.0399	0.0350	0.0321
	(0.0040)	(0.0086)	(0.0041)	(0.0056)	(0.0017)	(0.0040)
σ	0.3715	0.3700	0.3194	0.4231	0.7181	0.2728
	(0.0047)	(0.0104)	(0.0061)	(0.0100)	(0.0131)	(0.0053)
ρ	-0.9085	-0.8867	-0.8968	-0.9303	-0.8891	-0.8941
	(0.0045)	(0.0110)	(0.0077)	(0.0087)	(0.0058)	(0.0070)
η_s	2.6623	3.0017	2.7025	2.5583	3.4414	2.6509
	(0.5168)	(0.6172)	(0.5761)	(0.6378)	(0.3259)	(0.6334)
η_v	1.1156	1.0161	1.0008	0.9836	1.5182	1.0106
	(0.2017)	(0.2812)	(0.2095)	(0.2654)	(0.3271)	(0.1822)
σ_c	3.1345	3.1403	2.7663	2.9803	1.4415	2.9978
	(0.0345)	(0.0651)	(0.0401)	(0.0664)	(0.0187)	(0.0386)
Index Mean	0.0797	0.0926	0.1716	0.1093	0.0797	0.0797
Index Variance	0.0384	0.0315	0.0341	0.0239	0.0384	0.0384

Notes: We report parameters estimated for subsample periods and restricted option panels for the SV model based on joint estimation. Parameters are annualized and reported under the physical measure. In parentheses, we report the posterior standard deviation for each parameter. σ_c is the option RMSE in dollars. Table 1 provides descriptive statistics for the restricted option panels. We also report the average index returns and variances for each of the samples used in estimation.

Table 6: Instantaneous Conditional Moments

Panel A: Conditional Moments

	Conditional Variance(R)	Conditional Variance(V)	Conditional Covariance(R, V)
SV	V_t	$\sigma^2 V_t$	$\rho\sigma V_t$
SVJR	$V_t + \lambda(\mu_s^2 + \sigma_s^2)$	$\sigma^2 V_t$	$\rho\sigma V_t$
SVJV	V_t	$\sigma^2 V_t + \lambda\mu_v$	$\rho\sigma V_t$
SVCJ	$V_t + \lambda E(\xi^2)$	$\sigma^2 V_t + \lambda\mu_v$	$\rho\sigma V_t + \lambda\rho_J\mu_v^2$

Panel B: Conditional Moments. Sample Averages

	Conditional Variance(R)	Conditional Variance(V)	Conditional Covariance(R, V)
SV	0.0406	0.0056	-0.0137
SVJR	0.0417	0.0049	-0.0122
SVJV	0.0407	0.0258	-0.0137
SVCJ	0.0418	0.0582	-0.0144

Notes: Panel A presents closed-form expressions, following Eraker, Johannes, and Polson (2003), for three instantaneous conditional moments for each of the four models we study. Panel B presents the sample averages for these moments. Note that $E(\xi^2) = \mu_s^2 + 2\mu_s\mu_v\rho_J + \rho_J^2\mu_v^2 + \sigma_s^2$.

Table 7: RMSEs based on Parameter Estimates from Subsamples.

(K/S, Mat)	1996-2015	1996-2000	1996-2006	2011-2015	SM	ATM
(<0.90,<30)	1.888	1.966	1.948	1.781	1.262	2.047
(0.90-0.95,<30)	2.090	2.227	2.164	1.959	1.899	2.369
(0.95-1.00,<30)	2.685	2.707	2.807	2.651	3.865	2.812
(1.00-1.05,<30)	2.939	2.767	3.200	2.897	4.502	2.958
(1.05-1.10,<30)	2.074	1.941	2.209	2.068	3.378	2.142
(>1.10,<30)	1.949	1.901	1.986	1.976	3.301	1.971
(<0.90,30-60)	2.822	3.020	3.013	2.534	1.532	3.362
(0.90-0.95,30-60)	2.803	3.083	2.977	2.539	2.415	3.495
(0.95-1.00,30-60)	2.915	3.008	3.086	2.870	4.039	3.180
(1.00-1.05,30-60)	3.272	3.108	3.598	3.278	4.708	3.370
(1.05-1.10,30-60)	2.460	2.244	2.740	2.439	3.723	2.668
(>1.10,30-60)	2.209	2.047	2.412	2.215	3.621	2.392
(<0.90,60-90)	3.202	3.483	3.562	2.774	1.642	4.239
(0.90-0.95,60-90)	2.743	3.107	3.016	2.426	2.164	3.815
(0.95-1.00,60-90)	2.514	2.704	2.685	2.507	3.243	2.891
(1.00-1.05,60-90)	2.945	2.906	3.216	3.005	3.631	3.082
(1.05-1.10,60-90)	2.731	2.522	3.045	2.768	3.289	3.025
(>1.10,60-90)	2.372	2.128	2.693	2.369	3.201	2.745
(<0.90,90-180)	3.235	3.451	3.781	2.805	3.058	4.723
(0.90-0.95,90-180)	2.585	2.881	2.963	2.392	3.514	3.847
(0.95-1.00,90-180)	2.065	2.356	2.205	2.187	4.100	2.536
(1.00-1.05,90-180)	2.460	2.604	2.633	2.568	4.169	2.640
(1.05-1.10,90-180)	2.983	2.981	3.095	3.076	3.188	3.233
(>1.10,90-180)	2.651	2.564	2.819	2.798	2.415	3.030
(<0.90,>180)	4.039	4.010	4.419	4.445	8.267	4.941
(0.90-0.95,>180)	3.930	4.104	3.901	4.652	8.907	4.128
(0.95-1.00,>180)	3.875	4.144	3.508	4.824	9.796	3.606
(1.00-1.05,>180)	3.910	4.288	3.600	4.722	9.899	3.937
(1.05-1.10,>180)	3.717	4.035	3.614	4.200	9.020	4.155
(>1.10,>180)	3.836	3.953	3.831	4.103	6.952	4.244
RMSE	2.997	3.110	3.149	3.128	5.091	3.426

Notes: We report root mean squared errors (in dollars) based on the 1996-2015 sample period and all options, using parameter estimates based on subsample periods and restricted option panels. σ_c is the overall average option RMSE for the entire sample. All results are based on the SV model and joint estimation, as in Table 5.

Appendix A Option Valuation

We use a strike-optimized method based on the formula of Carr and Madan (1998). The time-t price of a call option with strike K and maturity τ can be written as:

$$C_t = \int_k^\infty e^{-r\tau}(e^{s_{t+\tau}} - e^k) f_s(s_{t+\tau}) ds_{t+\tau}, \tag{A.1}$$

where $s_t = log(S_t)$ and $k = log(K)$. In the above formula, as k approaches $-\infty$, the call option price converges to S_t rather than zero and thus is not square-integrable. Carr and Madan introduce a dampening factor α to solve the problem and let:

$$c = e^{\alpha k} C. \tag{A.2}$$

Now the Fourier transform can be applied to c:

$$\begin{aligned}
\psi(u) &= \int_{-\infty}^{\infty} e^{iuk} c dk, \\
&= \int_{-\infty}^{\infty} e^{iuk} e^{\alpha k} \int_k^\infty e^{-r\tau}(e^{s_{t+\tau}} - e^k) f_s(s_{t+\tau}) ds_{t+\tau} dk, \\
&= \int_{-\infty}^{\infty} e^{-r\tau} f_s(s_{t+\tau}) \int_{-\infty}^{s_{t+\tau}} [e^{s_{t+\tau}+(\alpha+iu)k} - e^{(\alpha+1+iu)k}] dk ds_{t+\tau}, \\
&= \int_{-\infty}^{\infty} e^{-r\tau} f_s(s_{t+\tau}) [\frac{e^{s_{t+\tau}(\alpha+1+iu)}}{(\alpha+iu)(\alpha+1+iu)}] ds_{t+\tau}, \\
&= \frac{e^{-r\tau}}{(\alpha+iu)(\alpha+1+iu)} \int_{-\infty}^{\infty} e^{(\alpha+1+iu)s_{t+\tau}} f_s(s_{t+\tau}) ds_{t+\tau}, \\
&= \frac{e^{-r\tau}\phi_{s_{t+\tau}}(\alpha+1+iu)}{(\alpha+iu)(\alpha+1+iu)},
\end{aligned} \tag{A.3}$$

where ϕ_{s_t} denotes the risk neutral characteristic function of the log-price. Dropping the notation for dependence on t, the call option value is given by:

$$C = e^{\alpha k} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iuk} \psi(u) du. \tag{A.4}$$

Since the imaginary part of $\psi(u)$ is odd and the real part is even, Equation (A.4) can be further simplified as

$$C = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} Re[e^{-iuk}\psi(u)]du. \tag{A.5}$$

Carr and Madan (1998) evaluate Equation (A.5) with the fast Fourier transform algorithm. However, with todays' computer power we solve directly Equation (A.5) with a numerical integration method (see, Crisóstomo, 2018, for a comparison of the two methods in terms of speed and accuracy). In practice, we vectorize Equation (A.5) with respect to the strikes and we apply the Simpson's rule for computing the Call prices.

Appendix B Particle Filtering Using Returns

The particle filtering algorithm relies on the approximation of the true density of the state L_{t+1} by a set of N discrete points or particles that are updated iteratively through equation (7). Here we outline how SIR particle filtering is implemented using the return data.

Step 1: Simulating the State Forward
For $i = 1, \ldots, N$, we first simulate all shocks from their corresponding distribution:

$$(w_{t+1}, B_{t+1}, J^s_{t+1}, J^v_{t+1})^i \tag{B.1}$$

where the correlation between the state variables needs to be taken into account.[33] Then, new particles are simulated according to equation (7):

$$V_t = V_{t-1} + \kappa(\theta - V_{t-1}) + \sigma\sqrt{V_{t-1}}w_t + J^v_t B_t \tag{B.2}$$

Note that $t+1$ shocks affect R_{t+1} and V_{t+1}, and thus to simulate V_t, we in fact need w_t, J^v_t and B_t from the previous period. We then record w_{t+1}, J^v_{t+1} and B_{t+1} for the next period for each particle.

Step 2: Computing and Normalizing the Weights
Now, we compute the weights according to the likelihood for each particle $i = 1, \ldots, N$:

[33] The Brownian shocks z_{t+1} and w_{t+1} are correlated with coefficient ρ, and J^s_{t+1} and J^v_{t+1} are correlated with coefficient ρ_J.

$$\begin{aligned}\omega_{t+1}^i &= f_1(Y_{t+1}|L_{t+1}^i) \\ &= \frac{1}{\sqrt{2\pi V_t^i}} \exp\left\{ -\frac{1}{2} \frac{\left[R_{t+1} - (r_t - \delta_t - \frac{1}{2}V_t^i + \eta_s V_t^i - \lambda\bar{\mu}_s + J_{t+1}^{s(i)} B_{t+1}^i)\right]^2}{V_t^i} \right\}\end{aligned} \tag{B.3}$$

The normalized weights π_{t+1}^i are calculated as:

$$\pi_{t+1}^i = \omega_{t+1}^i / \sum_{j=1}^{N} \omega_{t+1}^j \tag{B.4}$$

Step 3: Resampling

The set $\left\{\pi_{t+1}^i\right\}_{i=1}^N$ can be viewed as a discrete probability distribution of $L_{t+1} = (V_t, J_{t+1}^v, B_{t+1})$ from which we can resample. The resampled $\left\{L_{t+1}^i\right\}_{i=1}^N$ as well as its ancestors are stored for the next period.

The filtering for period $t+1$ is now done. The filtering for period $t+2$ starts over from step 1 by simulating based on resampled particles and shocks for period $t+1$. By repeating these steps for all $t = 1, \ldots, T$, particles that are more likely to generate the observed return series tend to survive till the end, yielding a discrete distribution of filtered spot variances for each day.

Table A1: Parameter Estimates: Return-Based Estimation

	SV	SVJR	SVJV	SVCJ
κ	6.4802	5.8990	8.0885	7.4387
	(0.7732)	(0.7451)	(0.9158)	(0.8418)
θ	0.0339	0.0336	0.0263	0.0244
	(0.0025)	(0.0024)	(0.0020)	(0.0015)
σ	0.5121	0.4933	0.4616	0.4387
	(0.0270)	(0.0267)	(0.0281)	(0.0268)
ρ	-0.7886	-0.8024	-0.8180	-0.8232
	(0.0241)	(0.0234)	(0.0251)	(0.0245)
η_s	2.3818	2.2695	2.4483	3.2508
	(0.6123)	(0.4518)	(0.5748)	(0.2510)
λ		1.2089	0.8060	0.8128
		(0.4845)	(0.2183)	(0.0907)
μ_s		-0.0167		-0.0261
		(0.0063)		(0.0108)
σ_s		0.0173		0.0221
		(0.0066)		(0.0050)
μ_v			0.1252	0.0822
			(0.0611)	(0.0297)
ρ_J				-0.0960
				(0.1124)
Loglikelihood	16122	16131	16137	16149

Notes: We report parameters estimated using returns only for the SV, SVJR, SVJV and SVCJ models. Parameters are annualized. In parentheses, we report the posterior standard deviation for each parameter.

Table A2: Parameter Estimates in Existing Studies: The SVJR Model

Panel A: Based on Returns

Author	Period	κ	θ	σ	ρ	λ	μ_s	σ_s
ABL	1953-1996	3.704	0.013	0.184	-0.620	5.040	0.000	0.012
CV	1980-2000	12.187	0.019	4.274	-0.271	0.372	0.051	$exp(+)$
						5.379	0.026	$exp(-)$
CGGT	1953-1999	3.276	0.015	0.151	-0.279			
EJP	1980-2000	3.226	0.021	0.240	-0.467	1.512	-0.025	0.041
Eraker	1987-1990	3.024	0.021	0.202	-0.468	0.756	-0.037	0.066
Bates_1	1953-1996	4.380	0.014	0.244	-0.612	0.744	-0.010	0.052
CJM	1996-2004	6.589	0.032	0.450	-0.777	2.790	-0.013	0.013

Panel B: Based on Options

Author	Period	κ	θ	σ	ρ	λ	μ_s	σ_s
BCC	1988-1991	2.030	0.040	0.380	-0.570	0.590	-0.050	0.070
BCJ	1987-2003	5.796	0.012	0.240	-0.467	1.512	-0.100	0.041
CJM	1996-2004	2.638	0.063	0.448	-0.782	2.832	-0.015	0.006

Panel C: Based on Returns and Options

Author	Period	κ	θ	σ	ρ	λ	μ_s	σ_s
Pan	1989-1996	6.400	0.015	0.300	-0.530	12.300	-0.008	0.039
		3.300*	0.030*				-0.192*	
Eraker	1987-1990	4.788	0.042	0.512	-0.586	0.504	-0.010	0.167
		2.772*	0.072*				-0.050*	
HLM	1990-2007	1.711	0.049	0.653	-0.740	2.332	-0.021	0.019
		0.415*	0.203*			2.831*		

Notes: We report parameter estimates for the SVJR model from existing studies. Estimates in Panel A are physical values. Estimates in Panel B are risk-neutral values. In Panel C, estimates with a star (*) indicate risk-neutral values and the rest are physical values. All parameters are annualized. BCC: Bakshi, Cao, and Chen (1997), based on the S&P 500; ABL: Andersen, Benzoni, and Lund (2002), based on the S&P 500; CV: Chacko and Viceira (2003), based on the S&P 500; CGGT: Chernov, Gallant, Ghysels, and Tauchen (2003), based on the DJIA; EJP: Eraker, Johannes, and Polson (2003), based on S&P 500; Eraker: Eraker (2004), based on S&P 500; Bates_1: Bates (2006), based on S&P 500; CJM: Christoffersen, Jacobs, and Mimouni (2010), based on the S&P 500; BCJ: Broadie, Chernov, and Johannes (2007), based on the S&P 500; Pan: Pan (2002), based on the S&P 500; HLM: Hurn, Lindsay, and McClelland (2015), based on the S&P 500.

Table A3: Parameter Estimates in Existing Studies: The SVCJ Model

Panel A: Based on Returns

Author	Period	κ	θ	σ	ρ	λ	μ_s	σ_s	μ_v	ρ_J
EJP	1980-2000	6.552	0.014	0.199	-0.484	1.663	-0.018	0.029	0.037	-0.601
Eraker	1987-1990	4.032	0.014	0.146	-0.461	1.008	-0.032	0.049	0.032	0.312

Panel B: Based on Options

Author	Period	κ	θ	σ	ρ	λ	μ_s	σ_s	μ_v	ρ_J
BCJ	1987-2003	14.112	0.006	0.199	-0.484	1.663	-0.066	0.029	0.108	-0.601
AFT	1996-2010	2.049	0.033	0.354	-0.934	4.435	0.005	0.004	0.052	-0.502

Panel C: Based on Returns and Options

Author	Period	κ	θ	σ	ρ	λ	μ_s	σ_s	μ_v	ρ_J
Eraker	1987-1990	5.796	0.034	0.411	-0.582	0.504	-0.061	0.036	0.041	-0.693
		2.772*	0.071*				-0.075*			

Notes: We report parameter estimates for the SVCJ model from existing studies. Estimates in Panel A are physical values. Estimates in Panel B are risk-neutral values. In Panel C, estimates with a star (*) indicate risk-neutral values and the rest are physical values. All parameters are annualized. EJP: Eraker, Johannes, and Polson (2003), based on the S&P 500; Eraker: Eraker (2004), based on the S&P 500; BCJ: Broadie, Chernov, and Johannes (2007), based on the S&P 500; AFT: Andersen et al. (2015a), based on the S&P 500.