Institut canadien des dérivés
Canadian Derivatives Institute

L'Institut bénéficie du soutien financier de l'Autorité des marchés financiers ainsi que du ministère des Finances du Québec

Document de recherche

DR 19-12

Estimation and Filtering with Big Option Data : Implications for Asset Pricing

Publié Août 2019

Ce document de recherche a été rédigée par :

Kris Jacobs, University of Houston
Yuguo Liu, University of Houston

L'Institut canadien des dérivés n'assume aucune responsabilité liée aux propos tenus et aux opinions exprimées dans ses publications, qui n'engagent que leurs auteurs. De plus, l'Institut ne peut, en aucun cas être tenu responsable des conséquences dommageables ou financières de toute exploitation de l'information diffusée dans ses publications.

Estimation and Filtering with Big Option Data: Implications for Asset Pricing*

Kris Jacobs Yuguo Liu
University of Houston

August 27, 2019

Abstract

The computational cost of estimating option valuation models is very high. We propose to address these constraints by filtering the state variables using particle weights based on model-implied spot volatilities rather than model prices. We illustrate our method by estimating stochastic-volatility and double-jump models. Using long time series and large cross-sections has important implications for option pricing and asset pricing more generally. The variance risk premium parameter is somewhat smaller and much more precisely estimated. Inference on parameters characterizing skewness, kurtosis, and risk aversion changes. Moneyness and especially maturity restrictions may result in identification problems for the models we study.

JEL Classification: G12

Keywords: Option Valuation; Big Data; Particle Filter; Risk Premia; Identification.

*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, Jinji Hao, Rene Garcia, Mohamad Ghaderi, Bruce Grundy, Junye Li, Scott Murray, Federico Nardari, Manuela Pedio, Nick Polson, David Schreindorfer, Gustavo Schwenkler, Sang Seo, Thijs Van Der Heijden, Viktor Todorov, James Yae, Morad Zekhnini, and participants in seminars at Georgia State, West Virginia, the Universities of Melbourne and Wellington, UNSW, the SFS Cavalcade, the ESSEC workshop on Bayesian methods in Finance, and the Midwest Finance Meetings for helpful discussions and comments.

1 Introduction

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.[1] It is also well-known that 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. Moreover, to learn about risk premia and the structure of the stochastic discount factor, we need to estimate using both returns and options. Given the richness of the available option data, the resulting estimation of state-of-the-art option valuation models can be computationally very demanding.

This paper studies estimation and filtering dynamic option valuation models using big option data. It also documents the implications of restricting the size of the option panels. To illustrate the computational challenge resulting from big option panels, consider estimation using the OptionMetrics data used in many recent studies, and for simplicity consider using all available option data. 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 computational constraints. At least partly due to this computational cost, existing studies typically reduce the dimension of the data to make estimation computationally feasible.[2] 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, typically a small subset of the

[1] 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.

[2] 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.

available option contracts is used. Some studies impose very stringent constraints on the cross-section of options and are effectively based on short-maturity at-the-money options.

The first objective of this paper is to adapt existing estimation methods to overcome these computational constraints. We apply Markov Chain Monte Carlo (MCMC) for the parameter search, 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 by assigning weights based on model-implied spot volatilities instead. This enables us to estimate and test complex models using very large option panels. Because particle filtering using returns is relatively straightforward, we are also able to perform joint estimation of state-of-the-art option pricing models based on returns and large option panels. This facilitates inference on model parameters characterizing tail events and risk premia. We show that the likelihood for our implementation of the particle filter is very similar to the likelihood associated with the traditional implementation of the particle filter, which is computationally much more intensive.

Our approach can in principle be applied to any parametric option pricing model. To illustrate its advantages, we implement our estimation method using daily data on a well-known class of models, the workhorse Heston (1993) square root and Duffie, Pan, and Singleton (2000) double-jump models. The limitations of these models are by now well understood, and several authors have suggested model improvements such as multiple volatility components, tail factors and alternative specifications of the jump process. Moreover, Bates (2018) shows that modeling of intraday returns may improve model fit. Related, 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. We use daily data and the double-jump model because for this model implementation, the literature on dynamic option valuation models offers a wealth of evidence. By comparing our estimation results to existing studies and interpreting the differences, we are able to illustrate the advantages of our approach and the implications of restricting the sample. Moreover, the parameter estimates we obtain are useful in their own right.[3]

[3]For example, Andersen, Fusari, and Todorov (2015a, p.1088) note that, prior to their study, the double-jump model had not been estimated in its full generality because of the computational constraints discussed above.

We use daily data for 1996-2015 in our empirical analysis. We do not explicitly address the optimal structure of the option sample. We start with the simple intuition that large cross-sections may be helpful to identify model parameters, especially those governing the tails of the distribution. However, it may not always be optimal or preferable to use the entire cross-section of options, for instance because of liquidity concerns. Moreover, one may prefer cross-sectional data that are balanced across maturity and moneyness, or balanced over time. We therefore do not use all available option contracts but instead we use a large benchmark option sample that is balanced cross-sectionally and over time.

The second contribution of this paper is to document the implications of the composition of the option sample. Because the reduction in the computational burden of our approach allows us to estimate the models using large option samples, we are able to document the consequences of restricting the option sample. We focus on two sets of implications. First, the composition of the option sample may affect inference, in the sense that the likelihoods for different option samples have different optima. A second implication is that the likelihood may be very flat for certain samples, which may lead to identification problems.

Some of our findings simply confirm existing intuition. For example, we find that in the Heston model, it is difficult to reliably estimate the parameters that capture skewness and kurtosis when exclusively using at-the-money-options. We also confirm that it is critical to include options with different maturities to estimate the persistence of the variance. A more surprising finding is that exclusively relying on options with similar maturities also makes it more difficult to estimate the parameters characterizing the tails. We generally find that it is critical to include options with different maturities in the sample to reliably estimate the spot variance and variance mean reversion, which in turn impacts estimation of all other parameters, especially risk premia.

Regarding the moneyness dimension of the data and the tail parameters, we find that including in- and out-of-the-money options is even more critical to reliably estimate the skewness parameter than the kurtosis parameter. Including different maturities also facilitates inference on the skewness parameter and the parameters characterizing the jump processes. However, we find that inference on the price of return jump risk is challenging even when the sample includes large cross-sections of options.

When estimating the Heston (1993) and Duffie, Pan, and Singleton (2000) models using daily data for 1996-2015 and big cross-sections of options, the main differences and similari-

ties 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 matter, but that variance jumps are relatively more important. Estimates of the risk-neutral correlation between returns and variance are more negative than the existing literature, with the notable exception of Andersen, Fusari, and Todorov (2015a). Related to this, our 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 findings suggest 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.[4]

Because we estimate the models using a joint likelihood based on options and returns, we can also compare our estimates of risk premia with the existing literature. Our estimates of the price of diffusive variance risk are smaller than existing ones in the literature, but the most important difference is that they are highly statistically significant. This finding is mainly due to the inclusion of options with different maturities in our sample. Related to the smaller estimates of the price of variance risk, estimates of variance persistence exceed existing estimates. Once again this finding is due to the composition of the option sample. Finally, our estimation based on returns and big panels of option data also allows us to reliably estimate jump risk parameters. In models with return jumps, approximately 2% of the equity risk premium is due to jump risk, with the remaining 6% due to diffusive risk. This finding is remarkably consistent with several existing studies.

To the best of our knowledge we are the first to explicitly address the implications of restricting the option sample. This is important because the majority of existing studies impose some restrictions, and it is not clear from the existing literature how this affects empirical results. This complicates comparisons across studies and conclusions about model

[4]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. We eliminate the impact of increasing numbers of option contracts over time by scaling the option likelihood by the inverse of the number of option contracts in the sample. See for example Bates (2003) for a discussion of this issue.

performance. Moreover, our results also provide valuable insights on how to restrict the sample, if and when such restrictions are required for computational reasons. For instance, we find that for the models studied in this paper, certain restrictions in the moneyness dimension do not have very adverse consequences, while limiting the sample to a narrow range of maturities is likely to cause identification problems. This insight is not obvious a priori.

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) uses Markov Chain Monte Carlo (MCMC) technique and incorporates cross-sectional information by using on average approximately three options per day.

In the existing literature, Bates (2000) and Andersen, Fusari, and Todorov (2015a) are related to our approach because they use big cross-sections of index options in estimation. Bates (2000) uses a large dynamic panel for 1988-1993, but treats spot volatilities as parameters.[5] This 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. Andersen, Fusari, and Todorov (2015a) proceed differently and impose consistency between the spot volatility implied from options and a high-frequency volatility estimate from returns. Our approach differs because of our treatment of the filtering problem and the information used to filter the latent states.

Finally, several 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 (2017) 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

[5] Bates (2000) also uses a second estimation approach that imposes the dynamic constraint on the variance.

filter to estimate complex option pricing models using large cross-sections of option prices.

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 our estimates for parameter inference and identification, as well as 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 explain differences in estimation results by referring 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):

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

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

where S_t is the index level, r_t is the risk-free rate, δ_t is the dividend yield, γ_t is the total risk premium, κ denotes the speed of mean reversion, θ the unconditional mean variance, and σ determines the variance of variance. dZ_t and dW_t are Brownian motions with $corr(dZ_t, dW_t) = \rho$. N_t is 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 premiums are entirely attributable to the mean jump sizes of return and variance: $\eta_{J^s} = \mu_s - \mu_s^Q$ and $\eta_{J^v} = \mu_v - \mu_v^Q$.[6] The jump intensity λ and the standard deviation of the return jump size σ_s do not change across measures. Therefore, the total equity risk premium can be written as $\gamma_t = \eta_s V_t + \lambda(\bar{\mu}_s - \bar{\mu}_s^Q)$, where $\eta_s V_t$ is the diffusive equity risk premium, which is assumed to be linear in V_t.

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 volatility. 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:

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

We use the Fast Fourier Transform (FFT) in 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 more 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

[6] We use a simple structure of the jump risk premium because of concerns regarding identification, 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.

the current state V_t and model parameters $\Theta(\kappa, \theta, \sigma, \rho, \lambda, ...)$. 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.

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. Panel A 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. Panel B 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, but 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 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 129,182 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.[7] Panel B of Table 1 also reports average option prices and implied volatilities for these moneyness-maturity bins.

3 Estimation

We first discuss how the particle filter can be used to estimate models with latent states, and stochastic volatility models with jumps in particular. We then briefly discuss how this framework can be applied to estimate these models using return data. We discuss the computational problems that arise when applying these models to option data and how we address these problems. Finally we outline how we combine return and option data in estimation.

3.1 Estimation Framework and Notation

We first discuss the estimation method in general, because the algorithm is conceptually similar when different data sources are used. The implementation of the algorithm however differs dependent on the observables: returns, options, or both. 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). Applying Ito's lemma and discretizing (1)-(2) gives

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

[7]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.

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

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:

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

where $h = 1, 2, .., H_t$ and H_t is the total number of options at date t. We assume $\epsilon_{t,h}$ is normally distributed and $\epsilon_{t,h} \sim N(0, \sigma_c^2)$.[8]

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 O_{t+1}. Based on the discretization, $L_{t+1} = (V_t, B_{t+1}, J^s_{t+1}, J^v_{t+1})$ and $O_{t+1} = (\{C_t\}, R_{t+1})$ where $\{C_t\} = (C_{t,1}, ..., C_{t,H_t})$.[9] Define the measurement density by $f_1(O_{t+1}|L_{t+1})$ 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).[10] For the persistent latent variance, V_t, f_2 represents equation (7) and for the non-persistent jump variable, f_2 is simply a random draw from the corresponding distribution.

The discretized dynamics can therefore be described as follows:

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

[9] Note the timing convention in the notation: V_t generates 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}.

[10] 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).

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

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.

We now 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 $\Theta(\kappa, \theta, \sigma, \rho, \lambda, ...)$ and 2) latent states $\{L_{t+1}(V_t, B_{t+1}, J^s_{t+1}, J^v_{t+1})\}$. We use particle filtering to filter the latent states and adaptive Metropolis-Hastings sampling to perform the parameter search.

3.2 Particle Filtering

For now, think of the parameters Θ as given. A standard sampling-importance resampling (SIR) particle filter can be implemented at each time t using the following steps:

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

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

$$\omega^i_{t+1} = f_1(O_{t+1}|\tilde{L}^i_{t+1}) \tag{10}$$

$$\pi^i_{t+1} = \omega^i_{t+1} / \sum_{j=1}^{N} \omega^j_{t+1} \tag{11}$$

Step 3. Resample the particles $\tilde{L}^i_{t+1}$ according to the normalized weights $\{\pi^i_{t+1}\}$, which gives L^i_{t+1}, $i = 1 : N$. Now go back to Step 1.

We can think of the weights π^i_{t+1} 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.[11] The steps of the APF can be described as follows:

Step 1. Compute the first stage weights based on the predictive likelihood for each particle L_t^i. We need to first calculate the time t expected values for each particle.

$$\hat{L}_{t+1}^i = E(f_2(L_{t+1}|L_t^i)) \tag{12}$$

Then we compute the weights evaluated at this conditional expectation

$$_1\omega_{t+1}^i = f_1(O_{t+1}|\hat{L}_{t+1}^i) \tag{13}$$

Finally we normalize the weights, similar to what we do in the SIR.

$$_1\pi_{t+1}^i = {}_1\omega_{t+1}^i / \sum_{j=1}^{N} {}_1\omega_{t+1}^j \tag{14}$$

Step 2. Resample L_t^i according to the weights $\{_1\pi_{t+1}^i\}$, which gives $\breve{L}_t^i$.

Step 3. Simulate particles forward by taking into account new observations, $\tilde{L}_{t+1}^i = f_2(L_{t+1}|\breve{L}_t^i, O_{t+1})$.

To simulate the latent states in $\tilde{L}_{t+1}^i = (\tilde{V}_t^i, \tilde{B}_{t+1}^i, \tilde{J}_{t+1}^{s,i}, \tilde{J}_{t+1}^{v,i})$, we simulate the jump arrivals $\tilde{B}_{t+1}^i$, the return jump size $\tilde{J}_{t+1}^{s,i}$, the spot variance $\tilde{V}_t^i$ and the variance jump size $\tilde{J}_{t+1}^{v,i}$ sequentially, while marginalizing out everything else. Note that $\tilde{J}_{t+1}^{v,i}$ does not affect the observable O_{t+1}, but only affects V_{t+1}^i in the next period. We therefore omit $\tilde{J}_{t+1}^{v,i}$ in the

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

notation of the conditioning information set. We proceed as follows. First, $\tilde{B}_{t+1}$ takes the value of 1 with probability:

$$p(B^i_{t+1} = 1|\hat{J}^{s,i}_{t+1}, \hat{V}^i_t, O_{t+1}) \propto f_1(O_{t+1}|\hat{J}^{s,i}_{t+1}, \hat{V}^i_t, B^i_{t+1} = 1) \tag{15}$$

where $\hat{J}^{s,i}_{t+1} = \mu_s$ is the expected return jump size and $\hat{V}^i_t = E(f_2(V_t|\breve{V}^i_{t-1}, \breve{B}^i_t \breve{J}^{v,i}_t))$ is the expected spot variance at time t. Next, we simulate $\tilde{J}^{s,i}_{t+1}$ conditional on $\tilde{B}_{t+1}$, $\hat{V}^i_t$ and O_{t+1}, we simulate $\tilde{V}^i_t$ conditional on $\tilde{B}_{t+1}$, $\breve{B}^i_t \breve{J}^{v,i}_t$ and O_{t+1}, and we simulate $\tilde{J}^{v,i}_{t+1}$ conditional on $\tilde{J}^{s,i}_{t+1}$.

Step 4. Compute the second stage weight for each particle $\tilde{L}^i_{t+1}$:

$${}_2\omega^i_{t+1} = f_1(O_{t+1}|\tilde{L}^i_{t+1})/f_1(O_{t+1}|E(f_2(L_{t+1}|\breve{L}^i_t))) \tag{16}$$

$${}_2\pi^i_{t+1} = {}_2\omega^i_{t+1}/\sum_{j=1}^{N} {}_2\omega^j_{t+1} \tag{17}$$

Step 5. Another resampling on $\tilde{L}^i_{t+1}$ according to the weights $\{{}_2\pi^i_{t+1}\}$, which gives L^i_{t+1}, $i = 1 : N$.

The APF differs from the standard SIR in two ways: 1) the APF resamples before simulating new latent states according to the predictive likelihood and thus only particles with high likelihood to have generated O_{t+1} move forward; 2) the APF takes new observations into account to propagate new latent states. These two key characteristics of the APF are crucial for capturing rare events and help correct potential sample impoverishment problems associated with the standard SIR particle filter. See Pitt and Shephard (1999) and Johannes, Polson, and Stroud (2009) for more detailed discussions.

3.3 Adaptive Metropolis-Hastings MCMC

We use the Adaptive Metropolis-Hastings MCMC (AMH-MCMC) for the parameter search. 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.[12] In general, AMH-MCMC randomly samples a parameter set from a proposal distribution $q_j(\Theta_j^p|\Theta_{j-1})$ and subsequently accepts a new parameter vector Θ_j^p with probability:

$$\alpha(\Theta_j^p, \Theta_{j-1}) = min(1, \frac{f_1(O_{1:T}|\Theta_j^p)p(\Theta_j^p)/q_j(\Theta_j^p|\Theta_{j-1})}{f_1(O_{1:T}|\Theta_{j-1})p(\Theta_{j-1})/q_j(\Theta_{j-1}|\Theta_j^p)}) \tag{18}$$

where f_1 denotes the (simulated) likelihood and $p(\Theta)$ is the prior for Θ, which can be uninformative. We assume a flat prior and thus $p(\Theta_j^p) = p(\Theta_{j-1})$.[13]

We assume the proposal distribution $q_j(\Theta_j^p|\Theta_{j-1})$ to be multivariate normal with mean zero. Tuning the parameter variances in this proposal distribution is crucial to achieve estimation efficiency. Roberts and Rosenthal (2009) propose the adaptive random walk scheme, where q_j is estimated from previous iterations $(\Theta_1, ..., \Theta_{j-1})$ after a given amount of iterations. For example, we start from some pre-specified variance for each parameter and run 1000 iterations. Then, we adjust the new variance for each parameter of $q_j, j = 1001 : 2000$ to be the variance calculated from the first 1000 iterations, and so on. 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 search for optimal parameters by moving to a new parameter set with probability one if it generates a higher likelihood than the previous parameter set. To avoid local optima, the algorithm moves to new parameter set with a non-zero probability even if the new likelihood is lower than the previous one.

The total likelihood in equation (18) is simply a by-product of particle filtering. According to Malik and Pitt (2011), the total likelihood associated with the particle filter conditional on a set of parameters can be expressed as a function of the unnormalized weights:

$$f_1(O_{1:T}|\Theta) = \prod_{t=1}^{T} \left\{ \left(1/N \sum_{j=1}^{N} {}_1\omega_{t+1}^j \right) \left(1/N \sum_{j=1}^{N} {}_2\omega_{t+1}^j \right) \right\} \tag{19}$$

As a special case, for the SIR we only have first stage weights. The total likelihood therefore reduces to:

[12] See Johannes and Polson (2009a) for an in-depth discussion.

[13] 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 only observables.

$$f_1(O_{1:T}|\Theta) = \prod_{t=1}^{T} \left\{ 1/N \sum_{j=1}^{N} \omega_{t+1}^{j} \right\} \tag{20}$$

Generally, in particle filtering we need to keep all the historical information (the ancestors) of every particular particle at time t. This becomes increasingly time consuming ($O(T^2)$) as T increases, especially when we have a large N. However, since the likelihood function (20) does not require us to record the latent variable path, we proceed by simply conducting the parameter search first without recording the latent processes, and then back out latent states with the result of the parameter search. In our application, we noticed that although the APF performed better for filtering latent states given a set of parameters, it performed worse than the SIR in parameter search due to the more volatile likelihood. Therefore, we apply SIR for the parameter search and then filter the latent states using APF. In our implementation, we use 10,000 particles.

3.4 Estimation Using Returns Data

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

$$\begin{aligned} f_1(O_{t+1}|L_{t+1}^i) &= f_1(R_{t+1}|L_{t+1}^i) \\ &= \frac{1}{\sqrt{2\pi V_t^i}} \exp\left\{ -\frac{1}{2} \frac{\left[R_{t+1} - (r_t - \delta_t - \frac{1}{2}V_t^i + \eta_s V_t^i - \lambda\bar{\mu}_s + J_{t+1}^s B_{t+1})\right]^2}{V_t^i} \right\} \end{aligned} \tag{21}$$

The Appendix 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 Estimation Using Option Data

We now consider model estimation using option data only, that is, estimation that does not consider 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 these 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.

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:

$$\begin{aligned} f_1(O_{t+1}|L^i_{t+1}) &= \left(\prod_{h=1}^{H_t} f_1(C_{t,h}|C^M_{t,h}(V^i_t|\Theta))\right)^{1/H_t} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_c}\right) exp\left(-\frac{\sum_{h=1}^{H_t}(C_{t,h}-C^M_{t,h}(V^i_t|\Theta))^2}{2\sigma_c^2 H_t}\right) \end{aligned} \tag{22}$$

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. The total likelihood for the entire sample summing over all particles is:

$$f_1(O_{1:T}|\Theta) = \prod_{t=1}^{T} \left\{ 1/N \sum_{i=1}^{N} \left[\left(\frac{1}{\sqrt{2\pi}\sigma_c} \right) exp \left(- \frac{\sum_{h=1}^{H_t} (C_{t,h} - C_{t,h}^M(V_t^i|\Theta))^2}{2\sigma_c^2 H_t} \right) \right] \right\} \tag{23}$$

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, and 10,000 particles leads to approximately 1,500,000,000 computations of the option price in each function evaluation, which is computationally infeasible.

3.5.2 The Implied Spot Variance Method

Given this computational burden, we propose a more efficient filtering algorithm based on the implied spot variance. The motivating idea can loosely be thought of as reducing the three-dimensional option evaluation computation into a (pseudo) two-dimensional computation.

According to equation (22), a particle's likelihood is inversely proportional to the sum of squared pricing errors (SSE) for that particle. Define the implied spot variance (ISV) on day t as the spot variance that results in the smallest SSE:

$$ISV_t = \arg\min_{ISV_t} \sum_{h=1}^{H_t} (C_{t,h} - C_{t,h}^M(ISV_t|\Theta))^2 \tag{24}$$

To illustrate how this sum of squared errors changes with particle values, we select a typical day in our sample, December 1, 2015. Figure 2 scatter plots the SSE against the particle values using the following parameter values for the Heston model: $\kappa = 3$, $\theta = 0.25$, $\rho = -0.7$, and $\sigma = 0.4$. We have repeated this analysis for other days, models and parameter values and the conclusions are very similar.

By definition, the ISV is the lowest point on the parabola. The further a particle is away from ISV, the larger the SSE. Thus, we can fit this parabola and compute the SSE for any particle according to its distance to ISV. This simplifies an $O(T * N * H)$ problem to roughly an $O(T*H)$ problem, where $H = \frac{1}{T}\sum_{t=1}^{T} H_t$. In reality the reduction in computational complexity is less spectacular because we need an extra ISV search step. Fortunately, this step is extremely fast. For given parameters, the ISV search step is performed by a two stage grid search: in the first stage, we divide the domain for the variance into 20 grids, and evaluate the SSE for each grid point; in the second stage, we divide the space on both sides of the lowest SSE grid point into another 20 grids, and we evaluate the SSE once more. The ISV is given by the grid point with the lowest SSE. The accuracy of the ISV search may of course be affected by the discreteness of the grid, but we can increase the number of grid points to improve accuracy. We have extensively experimented with this. We have also compared our approach to the gradient search method and found that 20 grids provide a satisfactory approximation.

After we obtain ISV_t, we fit a parabola using the grid points, and then we compute the SSE for each particle: $SSE_{i,t} = a_{1,t} + a_{2,t} * (V_t^i - ISV_t)^2$. The likelihood for particle i can now be calculated as follows:

$$f_1(O_{t+1}|L_{t+1}^i) = \left(\frac{1}{\sqrt{2\pi}\sigma_c}\right) exp(-\frac{SSE_{ISV,t}(\hat{b}_t * (V_t^i - ISV_t)^2 + 1)}{2\sigma_c^2 H_t}) \tag{25}$$

which gives

$$f_1(O_{1:T}|\Theta) = \prod_{t=1}^{T}\left\{1/N\sum_{i=1}^{N}\left[\left(\frac{1}{\sqrt{2\pi}\sigma_c}\right) exp(-\frac{SSE_{ISV}(\hat{b}_t * (V_t^i - ISV_t)^2 + 1)}{2\sigma_c^2 H_t})\right]\right\} \tag{26}$$

where we use $SSE_{ISV,t} = a_{1,t}$ and $\frac{SSE_{i,t}}{SSE_{ISV,t}} = 1 + b_t(V_t^i - ISV_t)^2$ with $b_t = \frac{a_{2,t}}{a_{1,t}}$. Rather than computing option prices for $N = 10,000$ particles, we now only need to compute 40 grids in the ISV search step and then for each particle we calculate its likelihood based on its distance to ISV. In other words, we get rid of one dimension (the number of particles N) when computing option prices, thus saving more than 99% of computation time. Moreover, the ISV search step is sequentially independent, and therefore we can parallelize the computation

on $t = 1 : T$. Note also that while the likelihood in equations (25) and (26) 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.

3.5.3 How Reliable is the ISV Particle Filter?

While our implementation of the particle filter is computationally much less expensive, the question arises how accurately it approximates the log likelihood for a conventional implementation of the particle filter. The problem is of course that to investigate the quality of this approximation, the computational costs of the conventional particle filter again become an obstacle.

We therefore investigate the quality of the approximation using a subsample. We limit ourselves to the option sample for the year 2015. Figure 3 compares the log likelihood for our implementation using ISV and the traditional implementation of the particle filter. We plot the likelihood as a function of each of the five parameters in the SV model. We use 1,000 particles in this exercise, which makes the likelihood less smooth but makes the exercise computationally feasible. The results suggest that our approach is very reliable and provides a good approximation to a traditional, much more computationally expensive, implementation of the particle filter. We have also conducted smaller-scale comparisons using different sample periods and the results are very similar.

3.6 Joint Estimation Using Return and Option Data

In order to estimate the models using both sources of information, we combine the likelihood from returns and options. The relative weights of returns and options in the likelihood are critically important, especially because on every 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 parameter search. Recall 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:

$$f_1(O_{t+1}|L^i_{t+1}) = f_1(R_{t+1}|L^i_{t+1}) * (\prod_{h=1}^{H_t} f_1(C_{t,h}|C^M_{t,h}(V^i_t|\Theta)))^{1/H_t} \tag{27}$$

where the two components can be computed according to equations (21) and (22) 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 similar to the approach in Sections 3.4 and 3.5.

4 Empirical Results

We present and discuss estimates from returns and option data jointly, using the approach in Section 3.6. We also obtained estimates exclusively based on return data, as well as estimates exclusively based on option data, using the approaches in Sections 3.4 and 3.5 respectively. To save space we do not discuss these results in detail. Because we occasionally refer to them, we include them in the Appendix.

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. We present results for the moments implied by the estimated option pricing models, and we use these moments to illustrate important differences between the models. Finally, we discuss preference parameters implied by our estimates.

4.1 The SV Model

We start by discussing our estimates for the Heston (1993) square root stochastic volatility (SV) model, which is an important building block in the option valuation literature. The first column of Table 2 presents estimates of the SV model. These estimates can be compared with estimates from the existing literature in Panel C of Table 3, with the caveat that the estimates in Table 3 are obtained using different sample periods.[14] Note that to facilitate comparisons, the estimates in Table 3 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.[15]

To further motivate our empirical results, consider these existing estimates of the Heston model in Table 3. Panel A of Table 3 indicates that a large number of studies report results for the Heston model under the physical measure based on index return data. In contrast, 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. The studies in Panel C typically face important trade-offs due to computational complexity. For example, Aït-Sahalia and Kimmel (2007) estimate the model using return and option data using approximate maximum likelihood. They 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) uses implied-state Generalized Method of Moments (GMM) to estimate the model using return and option data for 1989-1996. However, note that while the parameter estimates in Pan (2002) are obtained using a very limited cross-section of options, model fit is also evaluated using a wider cross-section. Eraker (2004) uses Markov Chain Monte Carlo (MCMC) based on return and option data. The sample is limited to 1987-1990 and contains 3270 call options over 1006 trading days, averaging just over three options per day. Chernov and Ghysels (2000) use Efficient Method of Moments (EMM) for the period 1985-1994, and focus on short-maturity at-the-money calls.

[14] In Table 3, Panel A contains existing estimates for the Heston model based on returns and Panel B contains estimates based on options. The results in Panel A and B can be compared with our results in the first column of Tables A1 and A2 respectively in the Internet Appendix.

[15] Compared to the estimates reported in Eraker (2004) and Broadie, Chernov, and Johannes (2007), the estimates of κ and σ in Table 3 are multiplied by 2.52 (multiplied by 252 and divided by 100). The estimate of θ in Table 3 is multiplied by 0.0252 (multiplied by 252 and divided by 10,000).

We conclude that existing estimates of the Heston model that use option data face computational constraints, and that they reduce the cross-sectional dimension of the option data set, presumably partly due to these constraints. See Hurn et al. (2015) and Broadie et al. (2007) for related discussions. Clearly some of these constraints can be addressed 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 show that these methods are helpful to estimate risk premia as well as other model parameters. Consider our estimate of the variance risk premium. The posterior mean of η_v, the parameter that characterizes the diffusive variance risk premium, is equal to 1.0836 in Table 2. This estimate has the expected sign and it is highly statistically significant. The existing literature does not contain many estimates of η_v that are based on options and returns jointly, but nevertheless our estimate differs from the existing literature in important ways. It is smaller and statistically significant. Several existing studies report estimates of η_v that are not statistically significant, see for instance Eraker (2004) and Broadie et al. (2007). Pan (2002) and Hurn et al. (2015) report statistically significant estimates. The sample used in Hurn et al. (2015), who also obtain statistically significant estimates of η_v, is similar to ours in the sense that it includes options with widely different moneyness and maturity. Broadie et al. (2007) note 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.

Our estimate of η_v is also smaller than existing estimates based on options and returns.[16] The small magnitude of the estimate of η_v has important implications because the variance risk premium η_v defines the relation between physical and risk neutral mean reversion and long-run variance. The estimate of the long-run physical variance of returns θ in Table 2 is 0.0351. This gives a risk-neutral long-run variance of $\theta^Q = \kappa\theta/\kappa^Q = = \kappa\theta/(\kappa - \eta_v) = 0.0706$. Consistent with the literature, the risk-neutral variance considerably exceeds the physical variance due to a positive estimate of η_v. This finding is also consistent with nonparametric evidence, see for example Bollerslev, Tauchen, and Zhou (2009).

[16] Our estimate 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.

Now consider mean reversion in the variance. The estimate of κ in Table 2 is 2.156. The risk-neutral mean reversion parameter $\kappa^Q = \kappa - \eta_v$ from Table 2 is thus equal to 2.156-1.084=1.072. The finding that the physical and risk-neutral mean reversion estimated from options are smaller than the physical mean reversion from returns confirms existing findings. However, our estimate of physical mean reversion based on joint estimation is much smaller than the estimates in Panel C of Table 2.[17] Note that this difference is unlikely to be due to the sample period: the return-based estimate of κ in the first column of Table A1 in the Internet Appendix is consistent with existing estimates in the literature. The option-based estimate in column 1 of Table A2 suggests that our estimate of κ in Table 2 is due to our option sample. One possible explanation of our findings is that in a joint estimation exercise, the variance process is estimated to be much more persistent under the physical measure when option samples include options with longer maturities.

We conclude that our estimate of the price of diffusive variance risk η_v differs from the existing literature both with respect to its magnitude and statistical significance, and this has important implications for the estimates of the mean reversion in variance. Moreover, our estimates of κ^Q and θ^Q are more plausible than in some existing studies.

Our estimates also differ from existing results with respect to the parameters characterizing the tails of the distribution. In Table 2, the estimate of ρ, the correlation between the return and variance innovations, is -0.9161, much more negative than estimates in existing studies in Panels B and C of Table 3. Once again the estimates based on returns and options separately in Tables A1 and A2 in the Internet Appendix show that this finding is driven by the option sample. Andersen, Fusari, and Todorov (2015a) report an estimate of ρ of -0.934 when estimating the SVCJ model. Our findings are thus more consistent with estimates from a study that mainly relies on option-based information.

The estimate of σ in column 1 of Table 2 is equal to 0.4262. For this parameter, the difference with the existing literature is somewhat more subtle. Note from Tables A1 and A2 in the Internet Appendix that the estimate of σ based on options is 0.4091 and the estimate from returns is 0.5430. The estimate of σ based on options and returns jointly in Table 2 is therefore closer to the option-based estimate, which is smaller than the return-based estimate. Now consider the estimates of σ in the existing literature in Table 3. The option-based and joint estimates in Panels B and C are generally larger than the return-

[17] Note that κ^Q is negative in some existing papers, which implies a negative long-run variance θ^Q.

based estimates in Panel A. In studies that allow for a direct comparison, such as Eraker (2004), Bakshi, Cao, and Chen (1997) and Christoffersen, Jacobs, and Mimouni (2010), the option-implied variance process is more variable than the return-implied variance process. Our findings suggest that if the variance dynamics are imposed and large option panels are used in estimation, this finding does not necessarily obtain. This 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. Note that these findings on σ are related to our findings on ρ. Skewness is determined by ρ but kurtosis is increasing in $|\rho|$ and σ (Das and Sundaram, 1999). Hence, our more negative estimates of ρ also make the smaller estimates of σ possible. This again confirms the findings in Andersen, Fusari, and Todorov (2015a).

Note from Table 2 that most parameters are rather precisely estimated, not only the parameter η_v which characterizes the variance risk premium. Tables A1 and A2 in the Internet Appendix indicate that we also get relatively small posterior standard deviations when estimating the model using options only, but that the confidence intervals are wider when using returns only in estimation. Figures 4 and 5 provide more insight into this finding. These figures plot the trace for the return-based and option-based estimation respectively of the SV model. Our implementation of the particle filter uses 10,000 particles. For the return-based estimation, we use 50,000 iterations. For the option-based estimation, we use 2,500 iterations. We set the first one fourth of the iterations as burn-in, and report the posterior mean and standard deviation for each parameter from the subsequent iterations. The figure clearly indicates rather rapid convergence for both estimation exercises and for all parameters. The figures also clearly indicate that the estimation problem is much more stable for the option-based estimation, which explains the smaller posterior standard deviations.

Figure 6 provides additional evidence on the SV model. It compares the filtered variance paths based on returns, options and joint estimation. The filtered variance path from options is smoother than that from returns, which is consistent with the lower values of option-based κ and σ. The filtered variance path based on returns and options jointly in the third panel is more similar to the option-implied path in the second panel.

We conclude that our estimates of the parameters characterizing the tails of the distribution in the Heston model, σ and ρ, also differ in important ways from much of the existing literature. Table A3 in the Internet Appendix shows that this finding is not due to the

sample period.

4.2 The SVJR Model

We now discuss our estimates of the stochastic volatility model with return jumps (SVJR) model in column 2 of Table 2. Table A4 in the Internet Appendix summarizes parameter estimates for this model from the existing literature. The parameters in both tables are annualized.[18]

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. However, the estimates of these parameters are similar to those for the SV model in column 1.[19]

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 -0.0134, the jump risk premium η_{J^s} is equal to 0.0236, and the risk-neutral average jump size $\mu_s^Q = \mu_s - \eta_{J^s}$ is therefore equal to -0.0370. These jumps occur on average less than once per year (λ=0.895). The standard deviation of the jumps σ_s is 4.91%.

Table A4 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 compared to our approach. First, as with existing estimates of the SV model, many of the existing estimates are obtained from returns. Panel A of Table A4 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 group of studies.

Relatively few studies offer evidence based on options or options and returns. Pan (2002),

[18] 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 A4 multiplied by 100. In the case of the jump intensity, some papers, such as Pan (2002) express it as in Table A4; 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 A4.

[19] Table A1 in the Internet Appendix shows that our return-based estimates of σ and κ in the SVJR model are smaller than in the SV model. This finding is consistent with the results in Eraker, Johannes, and Polson (2003). The return jumps help capture the tails of the return distribution.

using a method of moments technique based on options and returns data, 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 standard deviation of 16.7%. These jumps occur on average once every two years. Our results are close to those of Eraker (2004), but the estimated standard deviation of the jump size is smaller. Finally, Broadie, Chernov, and Johannes (2007) proceed somewhat 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 2 also presents the average equity risk premium for the different models. Recall that the total risk premium is given by $\gamma_t = \eta_s V_t + \lambda(\bar{\mu}_s - \bar{\mu}_s^Q)$, 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 2% of the equity risk premium is due to jump risk, with the remainder 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, but they are remarkably similar. Pan also reports roughly similar 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 6% 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 premiums than for explaining overall return variation.

4.3 Jumps in Variance

The third and fourth columns of Table 2 present estimates of the SVJV and SVCJ models, respectively. The SVCJ model in column 4 contains jumps in returns and variance that are correlated. Several existing studies report estimates for this model (see, among others, Eraker, Johannes, and Polson, 2003; Eraker, 2004; Broadie, Chernov, and Johannes, 2007; Andersen, Fusari, and Todorov, 2015a). We also report on the SVJV model in column

3 because we encountered some identification problems when implementing the SVCJ model. The SVJV model is nested by the SVCJ model: it contains jumps in variance but not in returns, and as a result it has three fewer parameters. This model also allows us to further comment on the relative importance of jumps in returns and jumps in variance for modeling returns and options.

For the SVJV model, the estimate of the risk neutral average jump size $\mu_v^Q = \mu_v - \eta_{J^v}$ is equal to 6.46%. 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 0.919, implying these jumps occur just less than once per year. When adding jumps in returns in the SVCJ model in column 4, the jump size is very similar but the frequency of the jumps decreases. The estimate of the correlation between the return and variance jump is negative, as expected, at -0.5030.

Table A5 presents parameter estimates for the SVCJ model from existing studies. Our estimate of μ_v^Q is very close to that of Andersen, Fusari, and Todorov (2015a) and well within the range of existing studies. The same remark applies to our correlation estimate. Furthermore, existing estimates of variance jumps seem to be much more consistent in the literature, even though these studies use very different sample periods, compared to the existing estimates of jumps in returns. A plausible reason for the robustness of these estimates is that variance jumps are better identified in the data.

Our estimates also provide additional insight into the diffusive variance risk premium η_v. Recall that in the SV model, we find that while our positive estimate is consistent with the literature, it is also estimated precisely, which contradicts many existing findings. When adding return jumps, our estimate of η_v is still positive and statistically significant, but smaller. However, when adding variance jumps, the sign of η_v may change and the statistical significance decreases. This suggests that identification of the diffusive variance risk premium is not a problem in the absence of variance jumps. However, it is difficult to separately identify diffusive and jump variance risk premiums. This conclusion may be due to our use of daily data. Bandi and Reno (2016) show that identification in models with jumps in returns and volatility is facilitated when using high-frequency intraday data.

We conclude that our estimates of variance jump parameters are overall consistent with existing estimates. The decomposition of the average equity risk premium in diffusive and jump components in the SVCJ model is similar to the one in the SVJR model. Approximately 2% of the equity risk premium is due to jump risk, and 7% is due to diffusive risk.

4.4 Model Fit

The log likelihoods in Table 2 are very useful as indicators of the importance of return and variance jumps for modeling returns and options. Using a likelihood ratio test, the more complex models are always statistically supported by the data. The results in Tables A1 and A2 in the Internet Appendix indicate that we also need to account for jumps in returns as well as jumps in variance when returns and options are modeled separately. These findings confirm the importance of jumps in returns (Pan, 2002; Bates, 2000) and jumps in variance (Broadie et al., 2007).

Table 2 indicates that while the jumps in returns in the SVJR model lead to statistically significant improvements in fit compared with the SV model, it is the jumps in variance in the SVJV and SVCJ models that result in very large improvements in the log likelihood. Moreover, a comparison of Table 2 and Table A2 in the Appendix indicates that jumps are especially 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 large improvements in log likelihood, the richer models with jumps do not substantially outperform the simple SV model in terms of RMSE. The RMSE for the SV model is approximately \$3, and the jump models only improve the fit by 3 cents. 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. We leave a more detailed investigation for future research.

4.5 Option-Implied 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. These moments are provided in Panel A of Table 4.

Panel B of Table 4 presents the averages over our sample for the daily moments for the SV, SVJR, SVJV and SVCJ models. Clearly 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 are characterized by a much higher variance of variance.

It is also useful to consider how these moments change over time. Figure 7 presents the

time series of the conditional variance of variance, which peaks in the financial crisis for all four models.[20] Figure 7 indicates that while there are large differences across models in the average of the conditional variance of variance, as evident from Panel B of Table 4, the paths of the conditional variance of variance are highly correlated across models.

4.6 Implied Preference Parameters

We now provide a more structural interpretation of the estimated equity and variance risk premiums. This can be done for both the diffusive and jump risk; here we limit our analysis to the diffusive risk. Recall that the diffusive equity premium in the models we estimate is given by $\gamma_t = \eta_s V_t$. This equity risk premium and the corresponding risk neutralization can be motivated by a pricing kernel derived from the preferences of a representative agent.

In the case of a representative agent with power utility, the diffusive parameter η_s that characterizes the equity risk premium is equal to the rate of relative risk aversion. Our estimate of η_s for the SV model in Table 2 is equal to 2.502. This estimate is similar in magnitude to estimates obtained in other studies that use both options and returns, such as Hurn et al. (2015) and Eraker (2004), for example. However, it is also well-known (see for example Bates, 2000) that in the power utility case, the parameter η_v that characterizes the reduced form diffusive risk premium $\eta_v V_t$ is given by $\eta_v = -\phi\rho\sigma$, where ϕ is the rate of relative risk aversion. It is therefore possible to use the model estimates to calibrate another, variance-based, estimate of the risk aversion parameter, which should in principle be equal to that implied by the equity premium. Using our estimates of η_v, ρ, and σ for the SV model in Table 2, we obtain an estimate of risk aversion equal to 2.775, very close to the 2.502 estimate implied by the equity premium. We repeated this exercise using the estimates in other studies, and in all cases we found that the estimates of risk aversion implied by the diffusive variance risk premium are a multiple of the estimate implied by the diffusive equity risk premium. This finding suggests that our estimates may benefit from improved identification due to the larger cross-sections of options used in estimation.

Christoffersen et al. (2013) consider a more general utility specification in which the

[20] We do not report results for the time series of the conditional variance and the conditional covariance between returns and variance to save space. We simply note that these paths look very similar across models. The path of the conditional variance can be gleaned from Figure 6. For the conditional covariance, note from the equations in Table 4 that the time variation in the path is determined by the conditional variance. The conditional covariance therefore reaches a minimum during the financial crisis for all models.

utility of the representative agent is directly affected by the variance. In this setup the reduced form η_s and η_v parameters are related to the rate of relative risk aversion ϕ and the positive parameter that captures the variance aversion ξ as follows:

$$\eta_s = \phi - \xi\sigma\rho \tag{28}$$

$$\eta_v = -\eta_s\sigma\rho + (1-\rho^2)\sigma^2\xi \tag{29}$$

Solving for ϕ and ξ using the parameter estimates for the SV model in Table 2, we find that the implied ϕ and ξ are equal to 1.073 and 3.660 respectively, which means that both parameters have the theoretically expected sign. The values are also plausible. We repeated this exercise using parameter estimates from existing papers. Most existing estimates imply that one of the parameters has a sign that is inconsistent with economic theory. This finding again suggests that our estimates may benefit from improved identification due to the presence of options with different moneyness and maturity in the sample.

5 The Information in Options and Returns

Using the ISV particle filter, we were able to obtain model estimates using big panels of options. We now discuss how restricting the option sample in the moneyness and/or maturity dimension affects parameter inference and identification, and how it impacts the relative weighting of options and returns.

5.1 Likelihoods for Different Option Samples

The motivation for our newly proposed implementation of the particle filter 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 large option panels. As previously discussed, this is one of the reasons why existing studies have restricted the time-series and cross-sectional dimension of the data. A natural question is if restricting the data set in this way affects inference, model properties, and identification.

These implications can in principle be studied in various ways. One approach is to use Monte Carlo techniques: simulate repeated option data samples for a given model and then investigate if we can retrieve the known parameters. This approach is very instructive but also has some clear disadvantages. We therefore use a more direct approach. We simply inspect the log likelihood as a function of the parameters of interest. Because of the high cost of computing the log likelihood, this approach is not feasible using many existing estimation techniques, such as the conventional implementation of the particle filter. However, with our implementation of the particle filter, it is relatively straightforward to evaluate the likelihood.

Figure 8 illustrates these implications using the parameters of the SV model. We present the log likelihood as a function of the parameters in the SV model: κ, θ, σ, ρ, and the price of diffusive variance risk η_v. In each of the panels in Figure 8, the solid line represents the log likelihood for the option sample we use in our empirical exercise in Section 4, which we refer to as ALL. This dataset is described in Panel B of Table 1. The two other lines present the log likelihoods as a function of the parameters for more restricted option samples. Following the existing literature, we restrict either the moneyness dimension or the maturity dimension. ATM represents the sample with ATM options but all maturities and SM represents the sample with short maturity options but all moneyness. 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.[21] Note that the ATM and SM datasets are somewhat smaller than the one used in our benchmark analysis, but because the likelihood is scaled back by the number of options in the dataset, this does not affect the comparison in Figure 8. The critical difference between the datasets is the characteristics of the options in the sample. Panel C of Table 1 presents the sample sizes as well as descriptive statistics.

The top two panels of Figure 8 show that the log likelihood for the ATM sample is very similar to the log likelihood for the ALL sample as a function of the κ and θ parameters. However, when only using short maturities (the SM sample), the log likelihood is very flat, which may complicate inference.[22] The two panels in the middle row of Figure 8 show the log

[21] Because the sample does not contain options with maturity less than 5 days, if options with maturity less than 30 days are not available we include the options for the shortest maturity longer than 30 days.

[22] This finding is not necessarily due to the use of short maturities, but rather to the absence of multiple maturities in the sample. We verified that similar problems obtain when only long-maturity options are included in the sample. When using a single maturity, it is difficult to estimate mean reversion, which also

likelihood as a function of the σ and ρ parameters. Overall, the results confirm our intuition. The likelihood is flatter for the ATM sample than for the ALL sample. However, Figure 8 clearly indicates that this problem is much more serious for the skewness parameter ρ than the kurtosis parameter σ. Moreover, including different maturities also clearly facilitates inference on the skewness parameter.

Finally, the bottom panel presents the results for the price of diffusive variance risk η_v. The results are consistent with those for the other parameters.

Figure 9 reports on a similar exercise for the jump parameters. The two top panels report on jumps in returns in the SVJR model. The two bottom parameters report on jumps in variance in the SVJV model. In each case, the left panel reports on jump frequency and the right panel reports on jump size. The most important observation is that Figure 9 confirms the conclusions from Figure 8. Limiting the option sample makes it more difficult to conduct inference on the jump parameters. This is especially the case when using options with similar (short) maturities. Figure 9 also suggests that the parameters associated with variance jumps are easier to identify than the parameters associated with return jumps. The risk premium associated with the return jump is especially hard to identify.

These findings have important implications for the existing literature. As mentioned above, many existing studies use samples that consist exclusively of short-maturity and/or at-the money options. One of the reasons for these restrictions is that estimation is prohibitively costly with large option panels. Our findings show that sample composition impacts inference, and using larger samples requires new estimation approaches such as the one we propose. Moreover, our findings suggest that with very restricted samples, identification of some of the parameters may be a problem. This may explain the substantial differences between parameter estimates in the existing literature (See Tables 3, A4, and A5), although differences in sample period as well as other variations in the empirical setup may also play a role.

5.2 Returns and the Cross-Section of Options

It is well-known that when using both return and option data in a joint estimation exercise, in some cases the data heavily weigh the information in the options. In the existing

affects the estimation of other parameters. We present results for samples with short maturities only because of the existing literature.

literature, this is mainly due to the fact that we have only one return per day but a large number of options. However, recall that our approach effectively gives equal weights to returns and options to avoid this imbalance. In our implementation, it is therefore the cross-sectional composition of the option sample that determines the relative weight of returns and options. Figure 10 illustrates the relative information in the option and return data in a joint estimation exercise, and the impact that the choice of option sample can have on this trade-off. We present the log likelihood as a function of four parameters in the SV model: κ, θ, σ, and ρ. In each of the panels, the broken line represents the log likelihood for the option sample we use in our empirical exercise in Section 4. The solid line represents the likelihood for returns.

Figure 10 indicates that the option data are much more informative about the model parameters than the return data. The return likelihood is flatter than the option likelihood for all parameters, with the exception of ρ. The return likelihood clearly indicates that values of ρ close to -1 are not supported by the return data, whereas the option data have difficulty differentiating between values of ρ between -0.9 and -1.

Now consider the likelihoods in Figure 8, based on samples that are restricted in the moneyness or maturity dimension. These samples result in much flatter option likelihoods as a function of some of the parameters, and this will affect the relative importance of the return and option data in joint estimation. This may explain why existing studies based on returns and options often report estimates of ρ closer to -0.6, whereas our estimate in Table 2 is approximately -0.9.

6 Conclusion

Estimating state-of-the art option valuation models is challenging due to the complexity of the models and the richness of the available option data. We implement an estimation approach that partly overcomes these constraints, by filtering the state variables using weights based on model-implied spot volatilities rather than model prices.

We first demonstrate that this approach is reliable. Then 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 different maturities and widely different moneyness. We confirm that both return and variance

jumps are important, but variance jumps more so. We obtain more precise and somewhat smaller estimates of the parameter that determines the diffusive price of variance risk, and the diffusive parameters characterizing skewness, kurtosis, risk aversion, and the persistence of the variance process also differ from most existing studies.

We then use our approach to show that sample composition affects estimation results. Restrictive option samples may give rise to flat likelihoods, which complicates inference. When both returns and options are used in estimation, the composition of the option sample may affect the relative importance of returns and options for parameter estimates. Restricting the option sample may affect not only comparisons between existing studies, but also our overall assessment of these models and the option valuation literature.

Moreover, our results provide guidance on the optimal way to restrict the sample, if and when restrictions are required. For instance, we find that for the models studied in this paper, certain restrictions in the moneyness dimension do not have serious consequences, while limiting the sample to a narrow range of maturities is likely to cause identification problems. We leave a more detailed study of the optimal composition of the option sample for future research.

In future work we intend to use the computational advantage of our approach to study richer models with multiple volatility factors (Bates, 2000), time varying jump intensities (Pan, 2002), 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, 2018) or volatility estimates based on high-frequency data (Andersen, Fusari, and Todorov, 2015b). A more detailed comparison of the computational efficiency and properties of our approach and that of existing methods is also needed. Finally, we plan to investigate the implications of the choice of error specification and loss function (Hurn et al., 2015) for parameter estimates and model performance.

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.

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.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical Performance of Alternative Option Pricing Model, *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 Options, *Journal of Econometrica* 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., 2018, How Crashes Develop: Intradaily Volatility and Crash Evolution, Forthcoming, Journal of Finance .

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.

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, Steven Heston, and Kris Jacobs, 2013, Capturing Option Anomalies With a Variance-Dependent Pricing Kernel, *Review of Financial Studies* 26, 1962–2006.

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.

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.

Feunou, Bruno, and Cédric Okou, 2017, Risk-Neutral Moment-Based Estimation of Affine Option Pricing Models, *Working Paper, Bank of Canada* .

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

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

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

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.

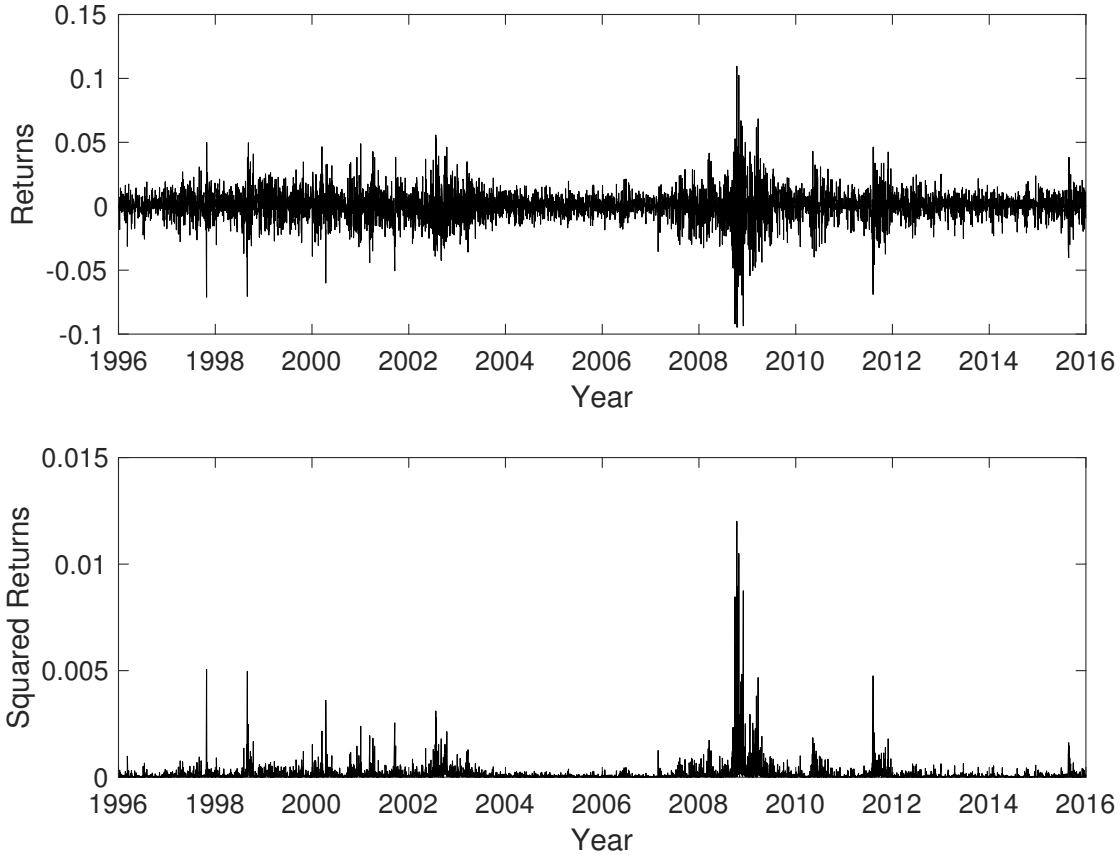
Pan, Jun, 2002, The Jump-risk Premia Implicit in Options: Eevidence from an iItegrated Time-series Study, *Journal of Financial Economics* 63, 3–50.

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, 2009, Examples of Adaptive MCMC, *Journal of Computational and Graphical Statistics* 18, 349–367.

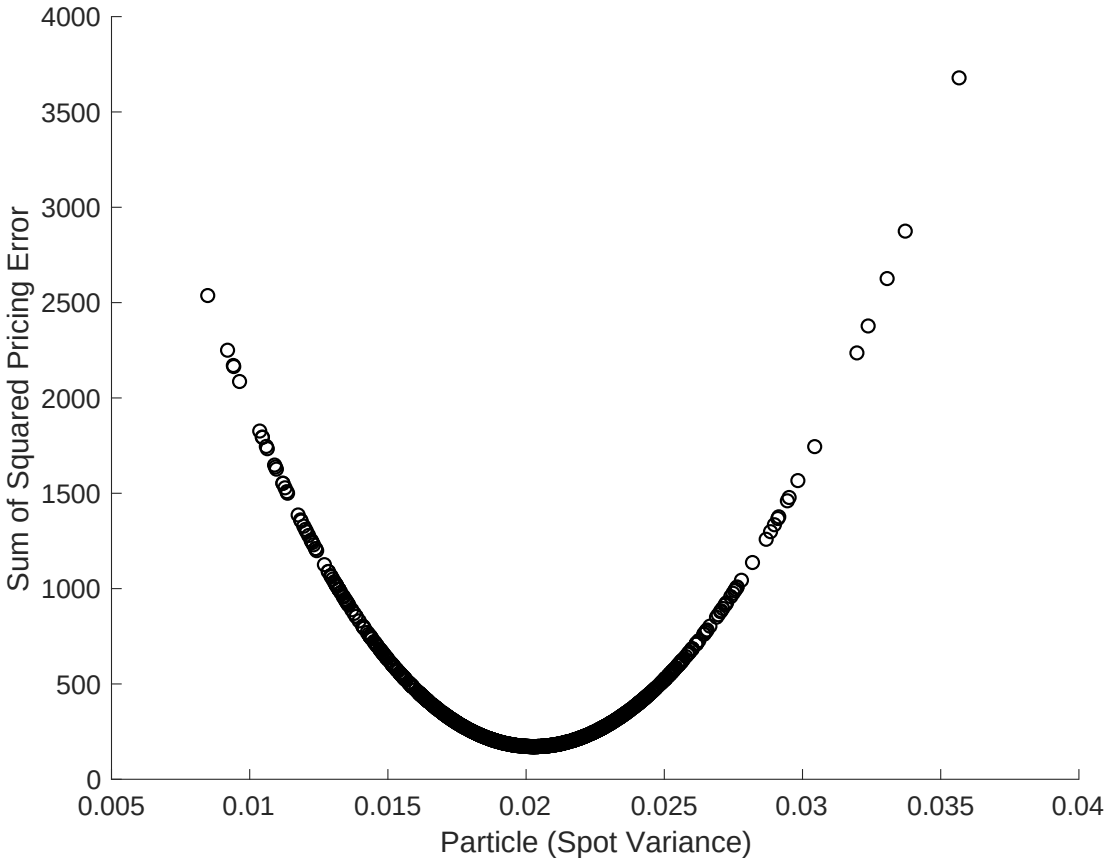
Singleton, Kenneth J., 2006, *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment, Princeton University Press*.

Figure 1: Daily Returns and Squared Returns 1996-2015



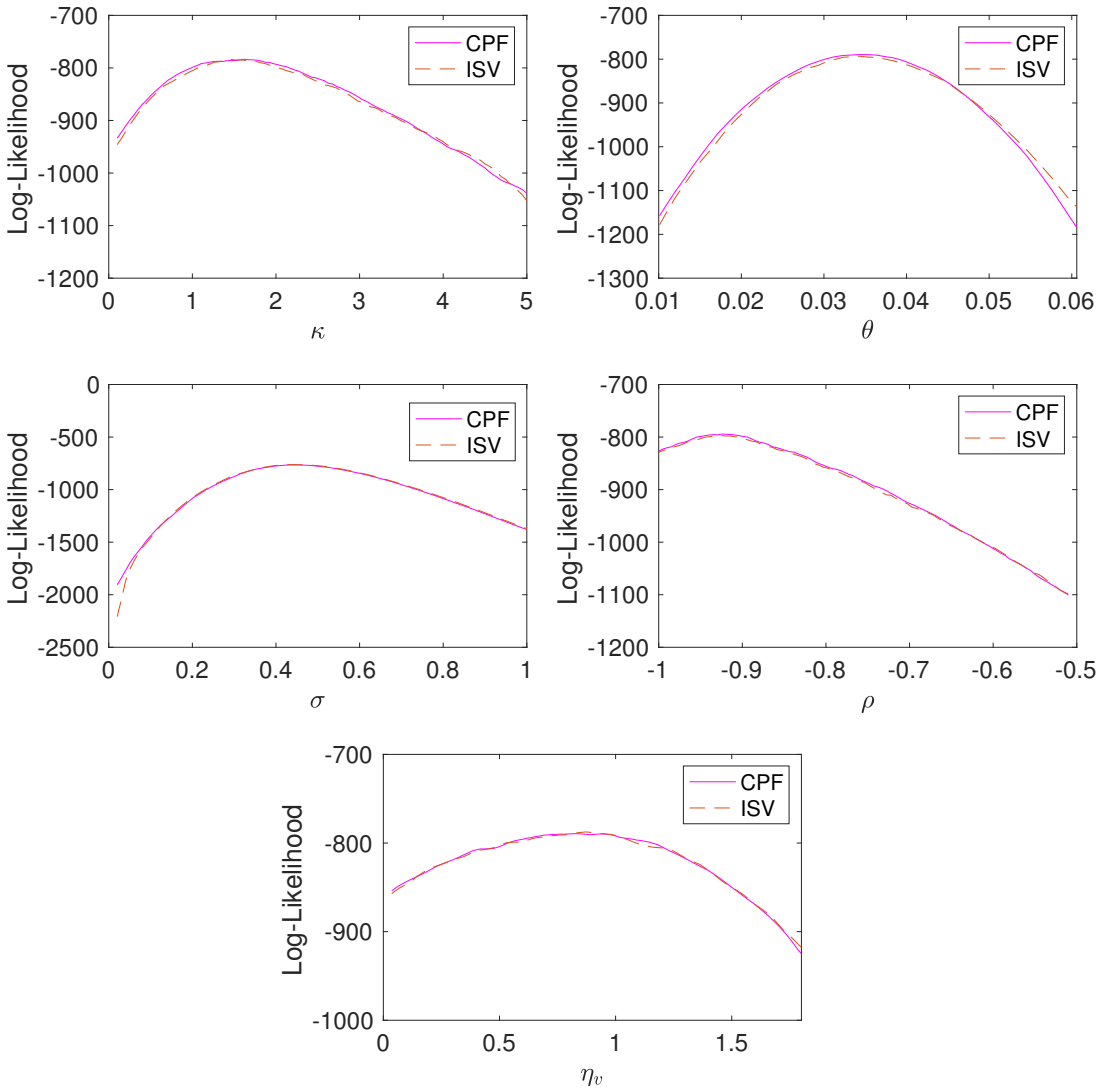
Notes: We plot daily log returns in the top panel. In the bottom panel, we plot squared daily log returns. The sample period is from January 1, 1996 until December 31, 2015.

Figure 2: Sum of Option Pricing Errors vs Particles



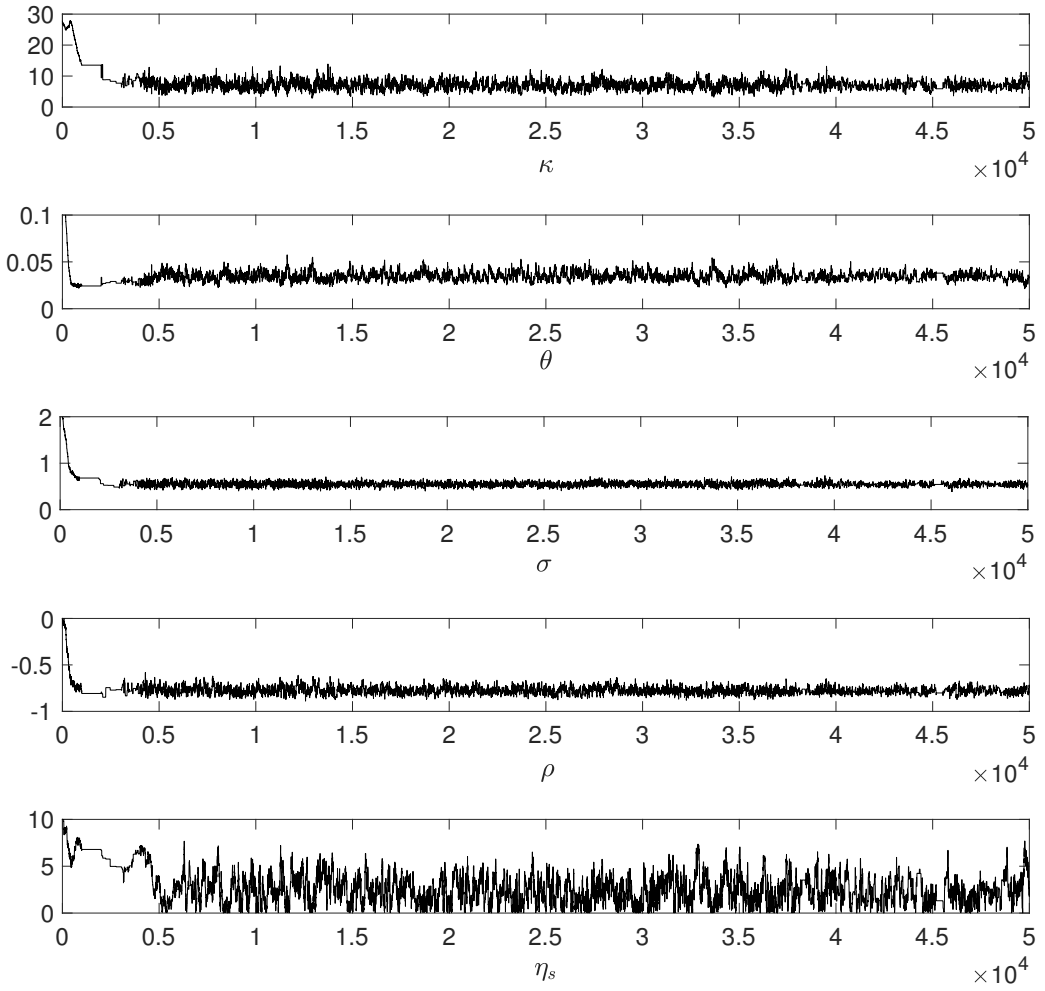
Notes: We show how the sum of squared pricing errors changes with the particle values in the case of the SV model. For each of the particle values (spot variances), we calculate the sum of squared pricing errors. We use option data for December 1, 2015 and the following parameter values: $\kappa = 3$, $\theta = 0.25$, $\sigma = 0.4$, and $\rho = -0.7$.

Figure 3: Log-Likelihoods Based on the ISV Particle Filtering Method



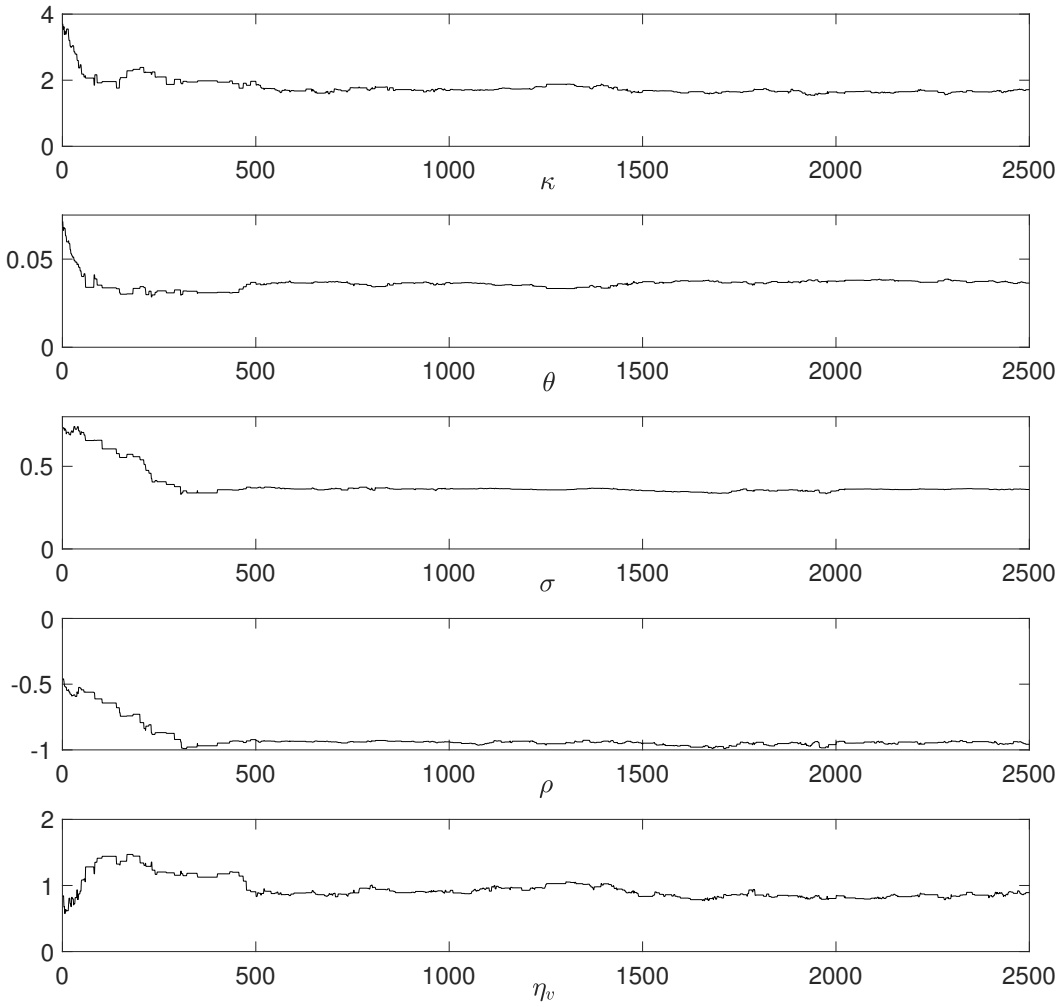
Notes: Using all option data in the sample for the year 2015, we plot the log likelihood based on particle filtering with the ISV method (ISV), and we compare it to the log likelihood using the conventional implementation of the particle filter (CPF).

Figure 4: Parameter Trace for SV Model Parameters. Return-Based Estimation



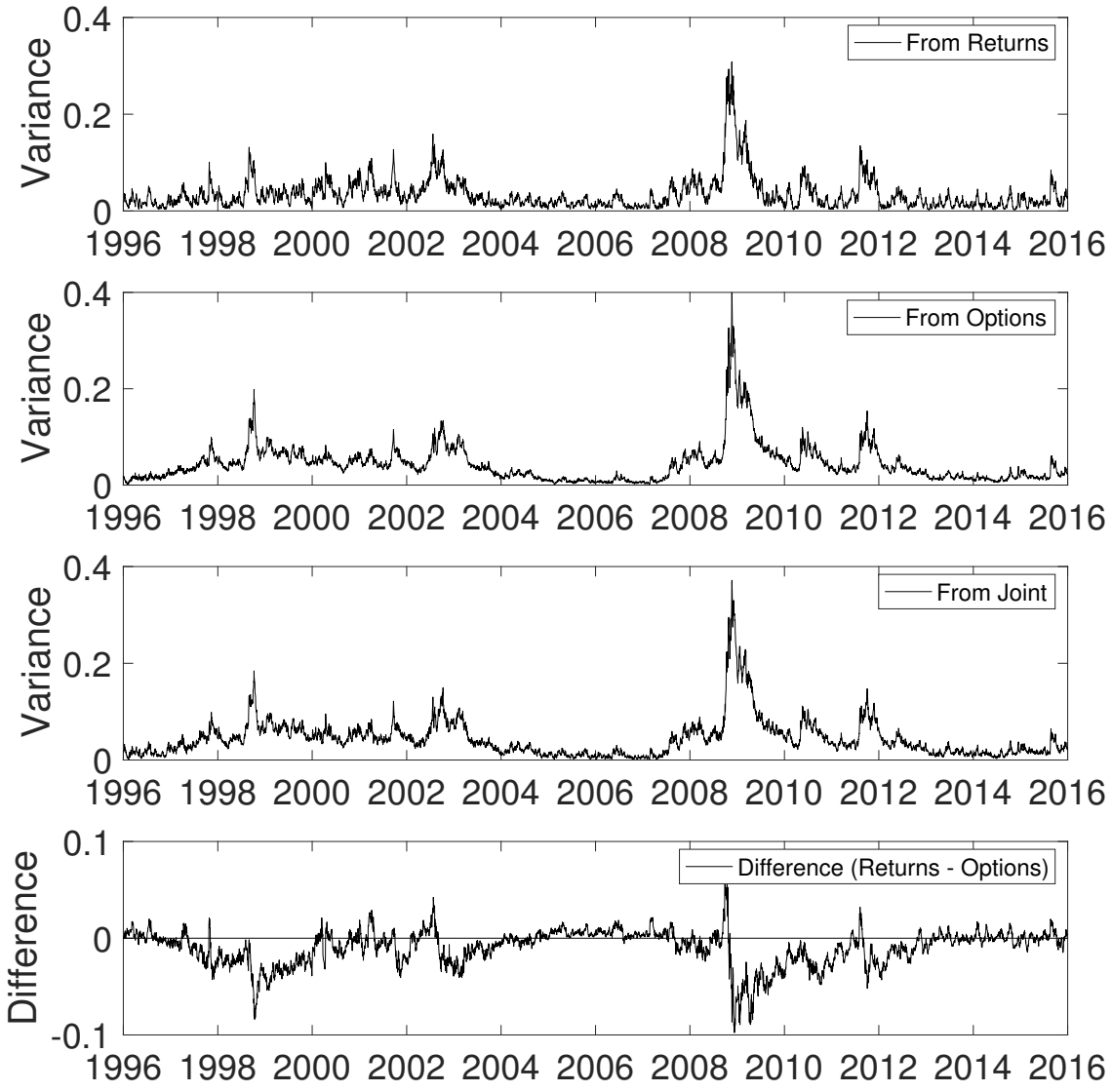
Note: We plot the full traces for each parameter in the SV model. We use 50,000 iterations. The first 1/4 of the iterations are treated as burn-in.

Figure 5: Parameter Trace for SV Model Parameters. Option-Based Estimation



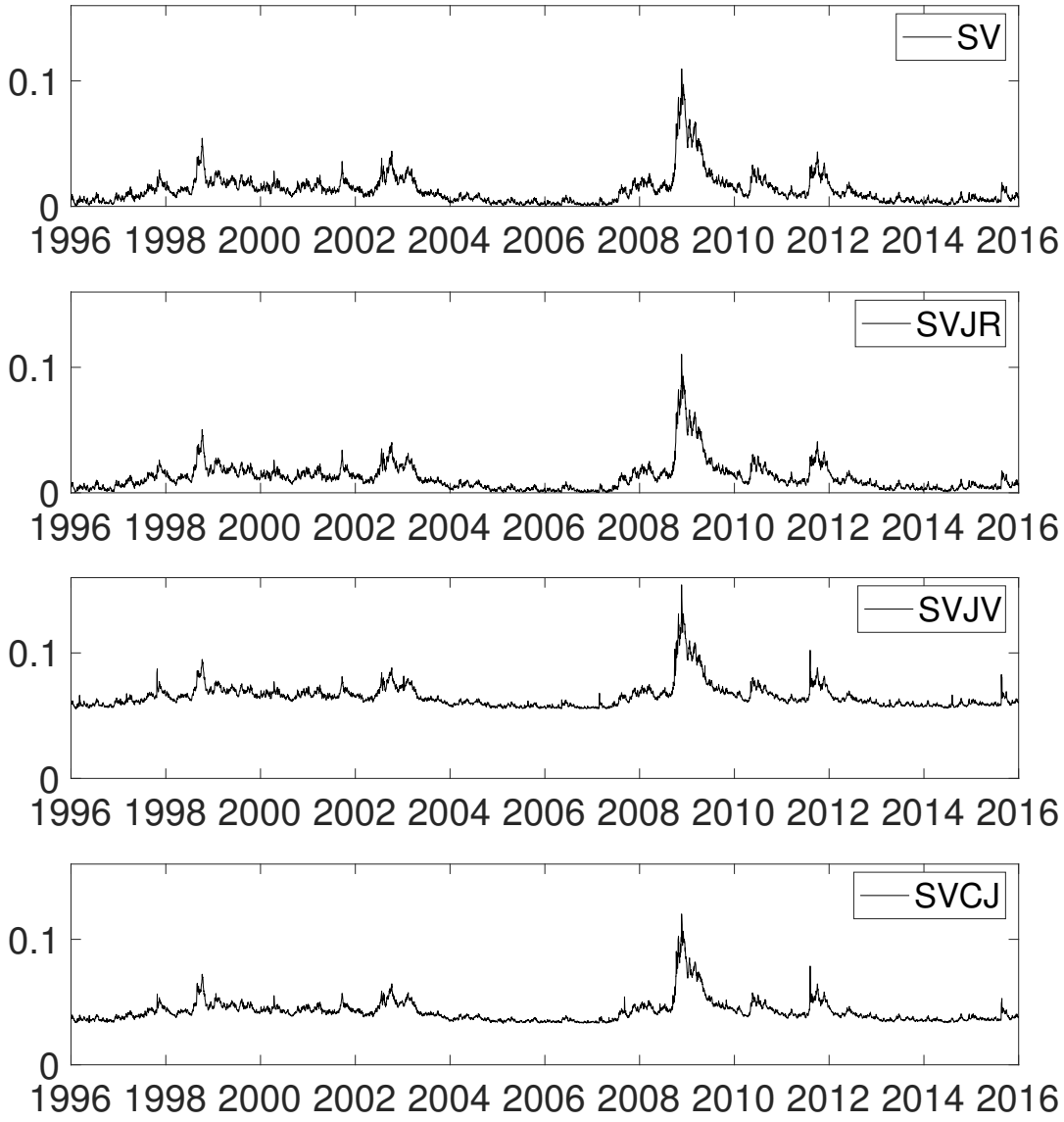
Note: We plot the full traces for each parameter in the SV model. We use 2500 iterations. The first 1/4 of the iterations are treated as burn-in.

Figure 6: Filtered Variance Paths. SV Model



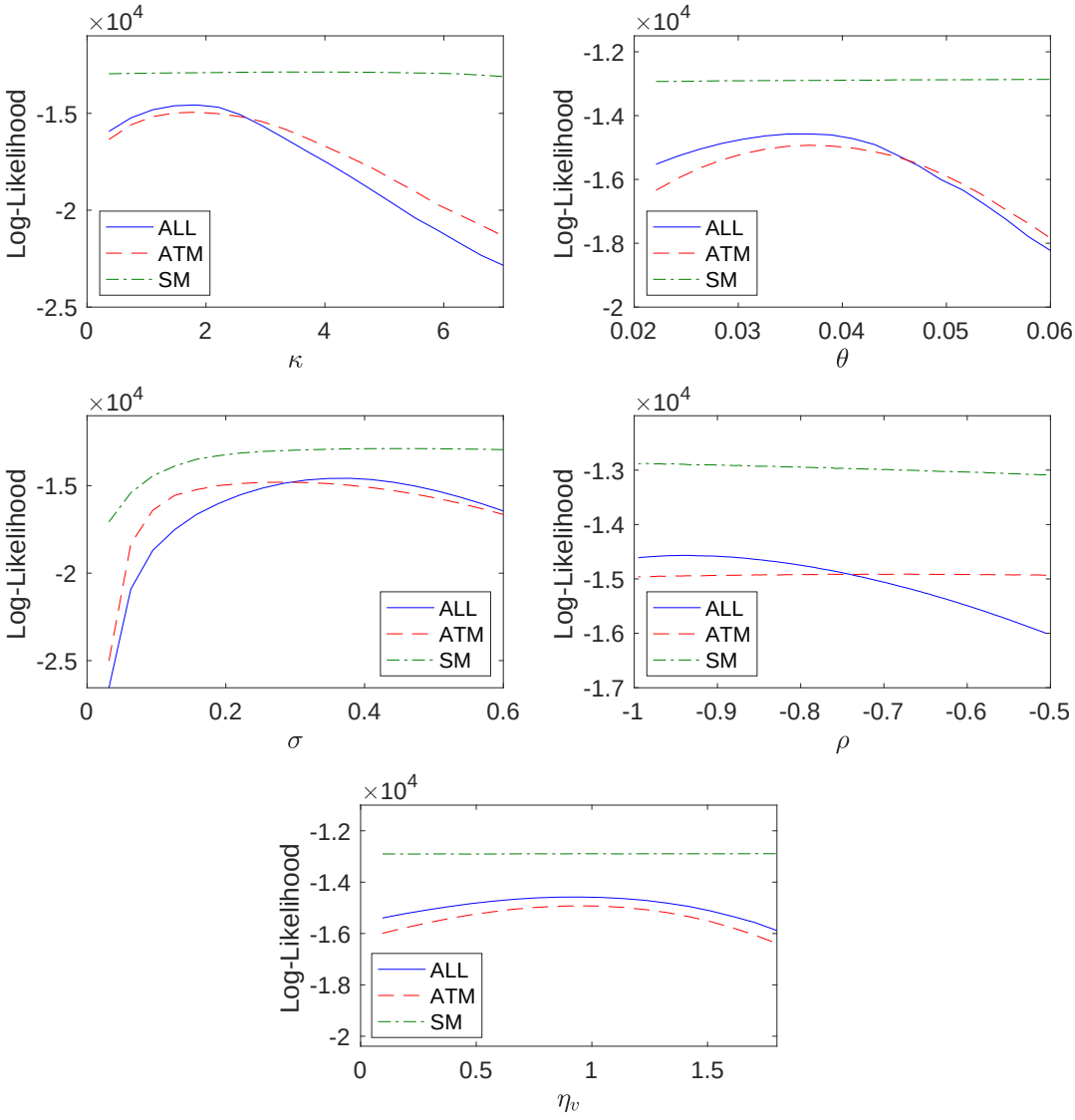
Notes: We plot the filtered variance path estimated from returns in the top panel, the variance estimated from options in the second panel, and the variance from joint estimation in the third panel. The bottom panel plots the difference between the variance estimated from returns and the variance estimated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

Figure 7: Conditional Variance of Variance. Various Models



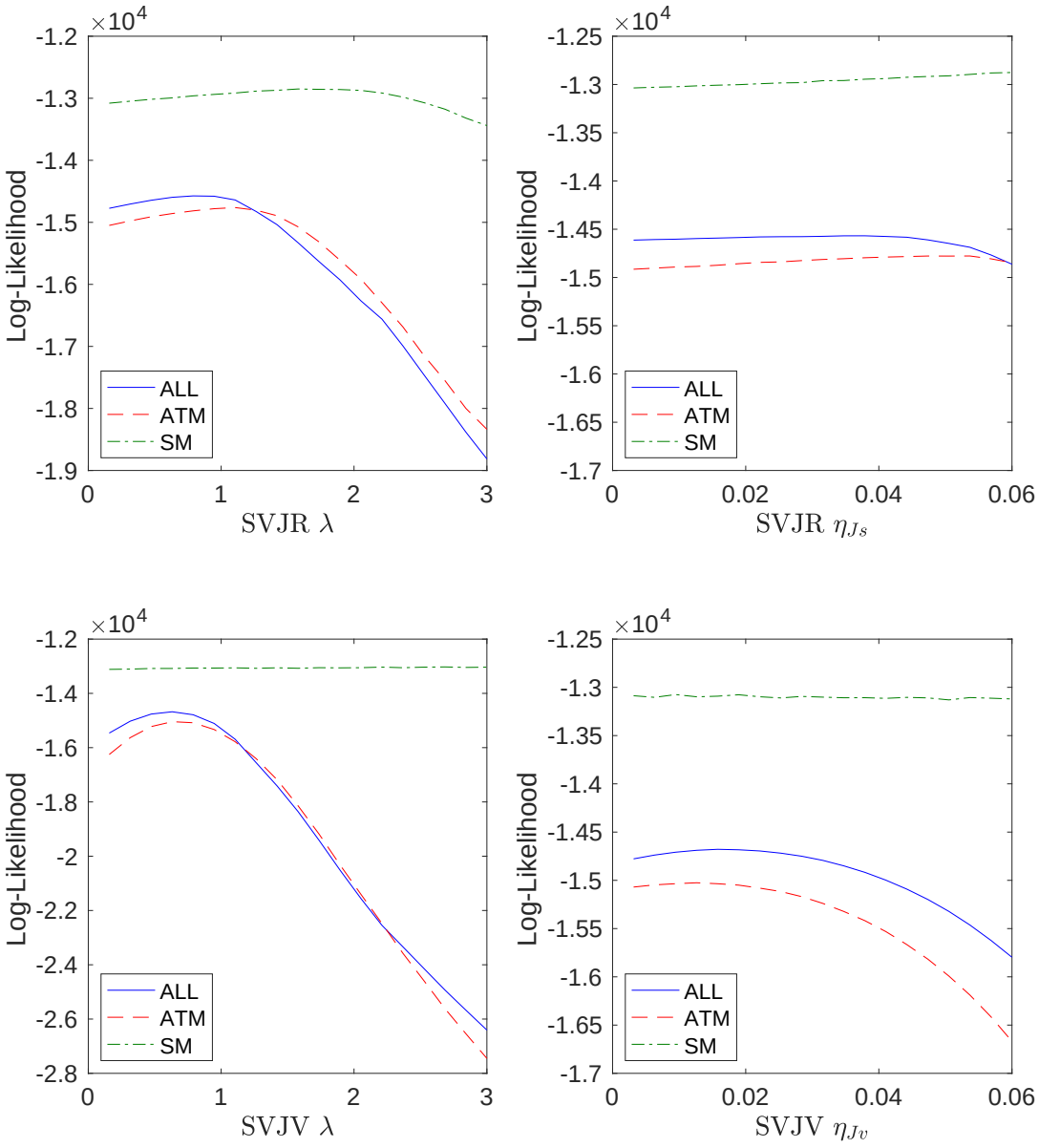
Notes: We plot the conditional variance of variance for four option pricing models.

Figure 8: Log-Likelihoods for Restricted Option Samples: The SV Model



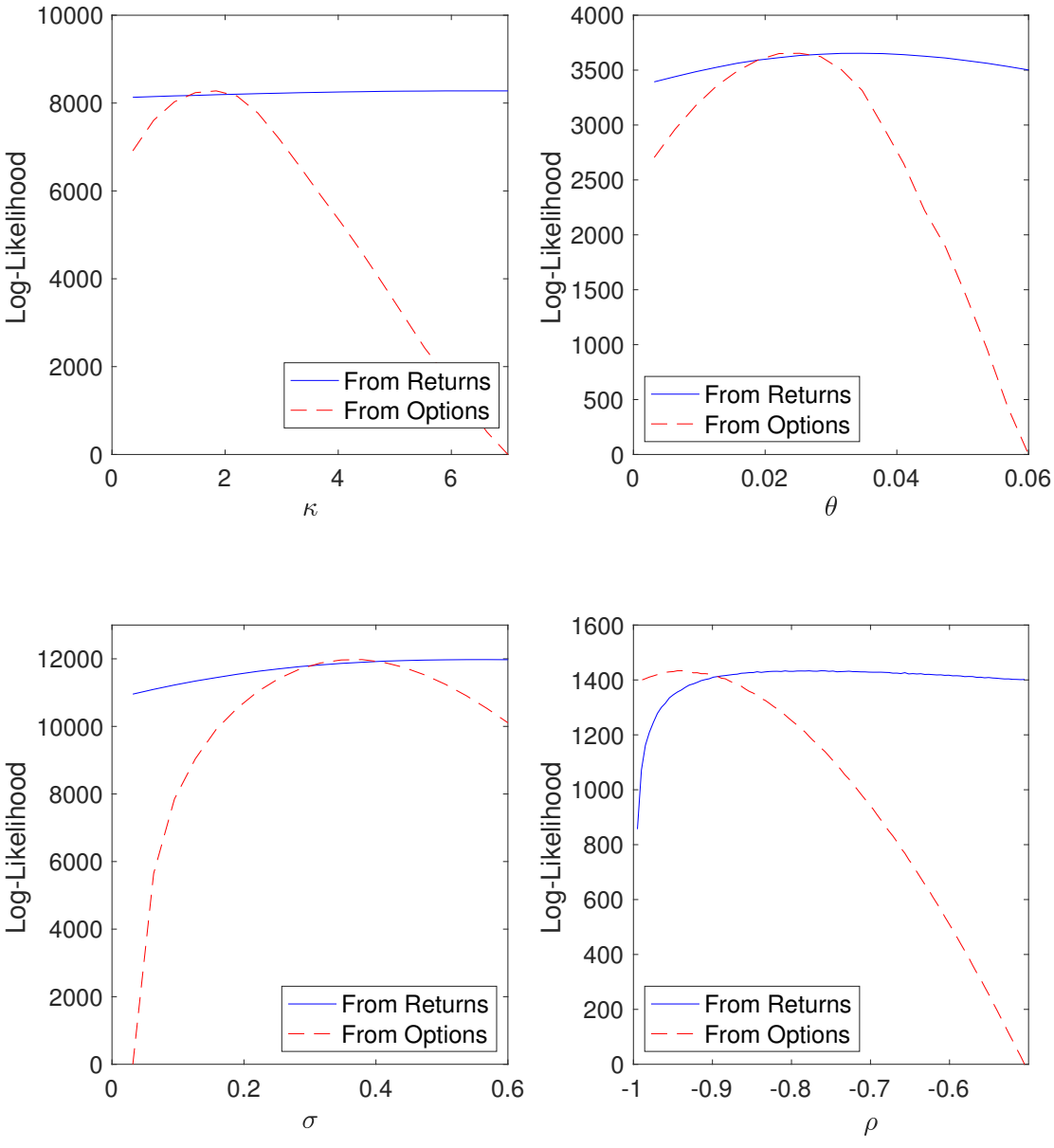
Notes: We plot the log-likelihood for the SV model as a function of the model parameters. We either use the option sample used in our estimation exercise or samples that are restricted based on maturity and/or moneyness. ALL represents the sample with all options. ATM represents the sample with ATM options but all maturities and SM represents the sample with short maturity options but all moneyness.

Figure 9: Log-Likelihoods for Restricted Option Samples: Jump Parameters



Notes: We plot the log-likelihood for the jump parameters in the SVJR and SVJV models. We either use the option sample used in our estimation exercise or samples that are restricted based on maturity and/or moneyness. 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.

Figure 10: Log-Likelihood for Index Returns and Options



Notes: We plot the log-likelihood for index returns vs index options as a function of model parameters. For ease of presentation, we shift the level of the likelihood to have the same value at the maximum. Note that we do not scale the gradient of the likelihood.

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	4312	4750	4249	4621	4186	22118
0.90-0.95	4377	4824	4532	4839	4540	23112
0.95-1.00	4383	4838	4626	4922	4823	23592
1.00-1.05	4383	4842	4633	4949	4823	23630
1.05-1.10	4046	4741	4380	4774	4596	22537
1.10-1.15	1432	2639	2443	3711	3968	14193
All	22933	26634	24863	27816	26936	129182
Average Call Prices						
0.85-0.90	149.35	150.81	158.71	165.44	184.66	161.79
0.90-0.95	100.03	105.41	113.95	125.17	146.97	118.31
0.95-1.00	40.72	50.48	60.28	76.45	101.39	65.87
1.00-1.05	11.16	20.82	30.91	45.14	71.75	35.95
1.05-1.10	2.62	4.92	9.42	19.37	40.57	15.38
1.10-1.15	2.47	2.92	5.09	9.29	23.44	8.64
All	51.06	55.89	63.06	73.48	94.80	67.66
Average Implied Volatilities						
0.85-0.90	0.34	0.28	0.26	0.25	0.24	0.27
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	0.31	0.20	0.18	0.17	0.17	0.18
All	0.24	0.21	0.21	0.20	0.20	0.21

Panel C: Restricted Samples. Descriptive Statistics

	Total No. of Options	Avg.Maturity($days$)	Avg.Moneyness(K/S)
ALL	129182	112.02	0.9886
ATM	109987	94.61	1.0001
SM	28831	22.7978	0.9786

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 generally 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 8 and 9. 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: Parameter Estimates Based on Joint Estimation Using Returns and Options

	SV	SVJR	SVJV	SVCJ
κ	2.1564	1.5531	1.0158	1.1248
	(0.0567)	(0.0404)	(0.0522)	(0.0489)
θ	0.0351	0.0359	0.0282	0.0241
	(0.0008)	(0.0008)	(0.0009)	(0.0008)
σ	0.4262	0.4152	0.3892	0.3450
	(0.0087)	(0.0085)	(0.0092)	(0.0074)
ρ	-0.9161	-0.9378	-0.9417	-0.9237
	(0.0129)	(0.0134)	(0.0118)	(0.0107)
η_s	2.5016	2.3513	2.7833	3.0401
	(0.2805)	(0.3081)	(0.3077)	(0.3486)
η_v	1.0836	0.5753	-0.1376	0.0498
	(0.0472)	(0.0454)	(0.0421)	(0.0355)
λ		0.8949	0.9188	0.6005
		(0.0239)	(0.0363)	(0.0359)
μ_s		-0.0134		-0.0104
		(0.0008)		(0.0005)
σ_s		0.0491		0.0426
		(0.0005)		(0.0006)
η_{J^s}		0.0236		0.0361
		(0.0007)		(0.0003)
μ_v			0.0594	0.0608
			(0.0011)	(0.0013)
η_{J^v}			-0.0052	0.0018
			(0.0012)	(0.0015)
ρ_J				-0.5030
				(0.0076)
Diffusive ERP	0.0878	0.0844	0.0785	0.0733
Jump ERP		0.0206		0.0210
Loglikelihood	1265	1281	1352	1397

Notes: We report parameter estimates jointly estimated using both returns and options for the SV, SVJR, SVJV and SVCJ models. Parameters are annualized and under the physical measure. In parentheses, we report the posterior standard deviation for each parameter. Diffusive ERP and Jump ERP represent the equity risk premium due to the diffusive and jump components.

Table 3: 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 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 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 4: 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	0.0438	0.0080	-0.0171
SVJR	0.0455	0.0075	-0.0169
SVJV	0.0450	0.0610	-0.0165
SVCJ	0.0435	0.0416	-0.0143

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

Appendix A Option Valuation

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

$$C = \int_k^\infty e^{-r\tau}(e^{s_T} - e^k) f_s(s_T) ds_T \tag{A.1}$$

where $s_T = log(S_T)$ and $k = log(K)$. In the above formula, as k approaches $-\infty$, C_k 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:

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

here ϕ_{s_T} denotes the risk neutral characteristic function of log-price. The call option value is given by:

$$\begin{aligned} C &= e^{\alpha k}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iuk}\psi(u)du \\ &= \frac{e^{-\alpha k}}{\pi}\int_{0}^{\infty} Re[e^{-iuk}\psi(u)]du \end{aligned} \tag{A.4}$$

The second equality holds because the imaginary part of $\psi(u)$ is odd while the real part is even.

$$C^M = \frac{e^{-\alpha k}}{\pi}\int_{0}^{\infty} Re[e^{-iuk}\psi(u)]du \tag{A.5}$$

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 $i = 1 : N$, we first simulate all shocks from their corresponding distribution:

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

where the correlation between the state variables needs to be taken into account.[23] 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.

[23] 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.

Step 2: Computing and Normalizing the Weights

Now, we compute the weights according to the likelihood for each particle $i = 1 : N$:

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

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 $t = 1 : T$, 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.

The implementation for the APF is quite similar, except some extra simulation and weighting steps are needed, as discussed in Section 3.

Internet Appendix for Jacobs and Liu "Estimation and Filtering with Big Option Data: Implications for Asset Pricing"

Table A1: Parameter Estimates: Return-Based Estimation

	SV	SVJR	SVJV	SVCJ
κ	6.9691	6.3308	8.0033	7.5660
	(0.9034)	(1.0469)	(1.1433)	(1.1116)
θ	0.0359	0.0350	0.0273	0.0271
	(0.0034)	(0.0042)	(0.0031)	(0.0029)
σ	0.5430	0.5290	0.4754	0.4680
	(0.0286)	(0.0332)	(0.0344)	(0.0337)
ρ	-0.7906	-0.7967	-0.8217	-0.8214
	(0.0240)	(0.0296)	(0.0279)	(0.0294)
η_s	2.5374	2.3737	2.5311	3.1167
	(1.1318)	(1.2476)	(1.2247)	(1.2206)
λ		0.9713	0.9845	0.9953
		(0.1257)	(0.1103)	(0.1189)
μ_s		-0.0132		-0.0101
		(0.0074)		(0.0076)
σ_s		0.0203		0.0212
		(0.0087)		(0.0102)
μ_v			0.0662	0.0547
			(0.0110)	(0.0103)
ρ_J				-0.4159
				(0.1067)
Loglikelihood	16123	16130	16135	16141

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: Option-Based Estimation

	SV	SVJR	SVJV	SVCJ
κ	1.9153	1.3812	1.2853	0.6183
	(0.0431)	(0.0758)	(0.0781)	(0.3120)
θ	0.0356	0.0354	0.0274	0.0264
	(0.0009)	(0.0007)	(0.0012)	(0.0005)
σ	0.4091	0.3922	0.3741	0.2940
	(0.0120)	(0.0221)	(0.0170)	(0.0121)
ρ	-0.9330	-0.9562	-0.9515	-0.9518
	(0.0253)	(0.0236)	(0.0777)	(0.0441)
η_v	0.8010	0.4961	0.0468	-0.3640
	(0.0448)	(0.0684)	(0.0375)	(0.0515)
λ		0.8554	0.8284	0.6719
		(0.0414)	(0.0543)	(0.0585)
$\mu_s - \eta_{J^s}$		-0.0376		-0.0474
		(0.0011)		(0.0014)
σ_s		0.0492		0.0396
		(0.0012)		(0.0015)
μ_v			0.0644	0.0623
			(0.0043)	(0.0015)
η_{J^v}			-0.0013	-0.0025
			(0.0023)	(0.0014)
ρ_J				-0.5740
				(0.0051)
Loglikelihood	-15459	-15398	-15308	-15239

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

Table A3: Parameter Estimates for the SV model. Alternative Sample Periods.

	1996-2015	1996-2006	1996-2000
κ	2.1564	1.6470	2.0934
	(0.0567)	(0.0486)	(0.0498)
θ	0.0351	0.0283	0.0294
	(0.0008)	(0.0010)	(0.0013)
σ	0.4262	0.3201	0.3744
	(0.0087)	(0.0058)	(0.0105)
ρ	-0.9161	-0.9219	-0.9030
	(0.0129)	(0.0139)	(0.0090)
η_s	2.5016	3.8294	6.3855
	(0.2805)	(0.5374)	(1.0639)
η_v	1.0836	0.9232	1.5809
	(0.0472)	(0.0453)	(0.0565)
Index Mean Return	0.0797	0.0926	0.1716
Index Variance	0.0384	0.0315	0.0341

Notes: We report parameters estimated for subsample periods for the SV model based on joint estimation. Parameters are annualized and under the physical measure. In parentheses, we report the posterior standard deviation for each parameter. The subsample periods are 1996-2006 and 1996-2000. We also report the index mean returns and variances for each of the subsamples.

Table A4: 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	2.790	0.016	0.101	0.430	1.764	-0.030	0.018
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 A5: 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.