Online Appendix
Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets

Chris Bardgett* Elise Gourier† Markus Leippold‡

March 2018

A. Joint calibration

In this section, we describe the joint calibration exercise performed using the cross-section of S&P 500 and VIX options on specific dates. This exercise gives us some guidance for model design and allows us to reduce the set of models to be estimated on a time series of options' data.

Specifically, we fix a date t and consider $\{IV^{Mkt}_{\text{SPX},i}\}_{i=1\cdots N_{\text{SPX}}}$, the set of N_{SPX} market implied volatilities of S&P 500 options for strikes $\{K_i\}$ and maturities $\{T_i\}$. We denote by $\{IV^{Mkt}_{\text{VIX},j}\}_{j=1\cdots N_{\text{VIX}}}$ the set of N_{VIX} market implied volatilities of VIX options. To estimate the parameters, we minimize the root mean squared error (RMSE) between the market and model implied volatilities:[1]

If a model is not flexible enough to jointly reproduce the implied volatility patterns of both markets on a single date, the $\mathbb{Q}$–dynamics of the model is not sufficiently rich to accurately price both the S&P 500 and VIX derivatives jointly, and we can safely discard this model from further consideration. We consider two sub-specifications of our full model (SVJ3): (i) m and u are constant (SVJ), (ii) m is stochastic but u is constant (SVJ2). In the full SVJ3 specification, we impose θ_u to 1 and $\lambda_0^{(-)} = 0$ to improve identification.

*University of Zurich, Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-4045; Email: chris.bardgett@bf.uzh.ch.

†Queen Mary University of London and Centre for Economic Policy Research (CEPR), Mile End Road, London E1 4NS, UK; Email: e.gourier@qmul.ac.uk.

‡University of Zurich, Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-5069; Email: markus.leippold@bf.uzh.ch.

[1] Alternatively, we checked that using distances taking into account the bid–ask spread of IVs as in Cont and Kokholm (2013) does not significantly change the quality of fit. Instead of the RMSE, we also looked at average relative errors (ARE). However, this does not affect our conclusions. The results using ARE are available upon request.

[**Fig.** 1 about here.]

We use two global optimizers to cope with the non-convexity of the calibration problem and the potential existence of multiple local minima, namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996), and the Differential Evolution (DE) algorithm introduced by Storn (1996).

For our calibration, we choose a date on which the markets were under stress, namely May 5, 2010 at the beginning of the European sovereign debt crisis. After cleaning our data as described previously, we have 91 VIX options at six different maturities (from 0.04 to 0.46 years) and 486 S&P 500 options at eleven different maturities (from 0.05 to 0.91 years) available. We emphasize that we perform a joint calibration. Hence, all this data is entered as input to minimize the total RMSE in Eq. (A.2) from the VIX and the S&P 500 market simultaneously across all available maturities and moneyness.

In Figure 1, we plot the market and model implied volatilities for the S&P 500 (Panels A, C, E) and the VIX (Panels B, D, F) for two maturity slices each. For the S&P 500 options, we choose the two maturities $T = 0.05$ and $T = 0.3$, and for the VIX options, $T = 0.04$ and $T = 0.36$.

$$\mathrm{RMSE}_{\mathcal{M}}(t) := \sqrt{\frac{1}{N_{\mathcal{M}}} \sum_{1 \le i \le N_{\mathcal{M}}} \left(IV^{Mkt}_{\mathcal{M},i} - IV^{Mod}_{\mathcal{M},i} \right)^2}, \quad \mathcal{M} \in \{\mathrm{SPX}, \mathrm{VIX}\}, \tag{A.1}$$

$$\mathrm{RMSE}(t) \quad := \frac{1}{2} \left(\mathrm{RMSE}_{\mathrm{SPX}}(t) + \mathrm{RMSE}_{\mathrm{VIX}}(t) \right). \tag{A.2}$$

From Panel A, Figure 1, we observe that the Heston model provides reasonable results for the S&P 500 market. However, for the VIX market (Panel B), the Heston model clearly fails to reproduce one of the stylized facts of VIX option markets, namely the positive skew of the implied volatility surface. This failure is most pronounced for the short-term options, where the Heston model generates a significant negative skew. The results for the SVJ model look much more promising. Just by adding jump components to the returns and volatility process, we can now generate the positive skew in the VIX market (Panel D), while providing an almost perfect fit for the S&P 500 options market. The SVJ model only struggles at the short end of the VIX implied volatility surface. This shortcoming

disappears when we extend the SVJ specification to the SVJ2 model by adding the factor m. Doing so gives us not only a remarkable fit for the S&P 500, but also for the VIX options market (Panel F). Looking at the RMSEs of the SVJ and SVJ2 models, we find that the SVJ provides an RMSE_{SPX} of 1.27% and an RMSE_{VIX} of 11.60%. The SVJ2 model yields 1.17% and 5.15%, respectively. Hence, while the two models are comparable in terms of their performance on the S&P 500 options market, there is an obvious difference in the VIX market on the chosen date.

In unreported results, we perform calibration exercises on other days, also including calm periods. Irrespective of the day, we observe that the SVJ and SVJ2 models perform comparably on the S&P 500 options market, both fitting the data very well. In contrast, we find that there are dates when the SVJ model struggles to fit the VIX IVs in addition to the S&P 500 IVs, whereas the SVJ2 model satisfactorily fits both.[2] Therefore, we conclude from our calibration exercise that we can discard the Heston model from further analysis and that jumps in the volatility are necessary.

Daily calibration is a multiple curve fitting exercise, which matches a model to risk-neutral distributions implied by option prices at different maturities. Some of the parameters we get from daily calibrations are unstable and vary substantially from one day to the next.[3] To achieve a more robust estimation, consistent with the whole time series of in-sample data, we choose a methodology based on particle filtering.

We find that irrespective of the day, the SVJ specification performs as well as the SVJ2 and SVJ3 specifications on the S&P 500 options market, all fitting the data very well. In contrast, we find that there are dates when the SVJ model struggles to fit the VIX IVs in addition to the S&P 500 IVs, whereas the SVJ2 and SVJ3 models satisfactorily fit both.[4] Therefore, we conclude from our calibration exercise that we can discard the Heston model from further analysis and that jumps in the volatility are necessary.

[2]Our findings are consistent with Gatheral (2008), who shows that the Heston model is incapable of reproducing the positive skew in VIX IVs, and with Sepp (2008a,b), who finds that incorporating positive jumps in the volatility dynamics into the Heston model removes this shortcoming.

[3]Parameters obtained when calibrating to daily options prices are not stable over time, as explained in Broadie et al. (2007) and Lindström et al. (2008).

[4]Our findings are consistent with Gatheral (2008), who shows that the Heston model is incapable of reproducing the positive skew in VIX IVs, and with Sepp (2008a,b), who finds that incorporating positive jumps in the volatility dynamics into the Heston model removes this shortcoming.

B. Particle filter

B.1. Specific data treatment for the particle filter

As the datasets comprise a large number of options (up to 600 a day), it is unfeasible to calculate the option prices every day for every particle. As a consequence, we follow Pan (2002) and Johannes et al. (2009), among others, and use weekly (Wednesday) options data. Furthermore, this eliminates beginning-of-week and end-of-week effects. Our particle filter uses daily time steps and incorporates information on the underlying indexes on a daily basis (i.e., only options are considered weekly).

Moreover, the S&P 500 options dataset contains a large amount of ATM options compared to OTM and deep OTM options. If we use the filter (within the maximum likelihood procedure) on this entire dataset, the fitting of ATM options will be its priority rather than (deep) OTM options. This results in fitting the body of the S&P 500 returns distribution rather than the tails, which is not what we want. We need information about the extreme events contained in the data to be incorporated into the models. For this reason, we interpolate the S&P 500 IV slices and re-sample the option prices from the resulting parametric fit uniformly with respect to moneyness.[5] Other advantages of our use of interpolation is that the resulting data is arbitrage free and we have fewer points for each slice (but still accurately representing the information of each slice), thus reducing the computational complexity.[6]

For the interpolation, we use the efficient mixture of log-normals approach of Rebonato and Cardoso (2004) to have a parametric fit for each S&P 500 implied volatility slice. The RMSE of the S&P 500 implied volatilities parametric fits are on average around 0.25% and we therefore do not lose information, especially given the market bid–ask spread. Finally, using the parametric fit, we can sample a fixed number (we have chosen 15) of "market option prices" for the desired strikes. We have chosen to resample the option prices from each parametric slice uniformly in the strike (or, equivalently, the moneyness). We however do not resample the options for which the strike is smaller than 40% or larger than 140% of the current futures price. The reason is that there are usually only one or two options outside this interval of moneyness and we do not wish to re-sample options where

[5] It is common to interpolate data, see, e.g., Broadie et al. (2007). This eliminates arbitrage opportunities in the data and removes the accumulation of options around the ATM region.

[6] Since we have considered mid-prices and because of synchronization issues between the underlying and the options, implied volatility slices are not guaranteed to be arbitrage free.

the interpolation results could be driven by an outlier.

We do not perform any interpolation for the VIX options dataset, as most VIX options are OTM and therefore contain information about the tails of the VIX distribution (i.e., variance and central tendency processes). Therefore all available VIX option prices are used.

Finally, we decompose the time series of observations into two periods. The first period is from March 1st, 2006 to Feb 28, 2009 (shortly after the VIX index increased to its highest point). This was a rather calm period that we will use as the in-sample estimation period.[7] Our out-of-sample period starts on March 1st, 2009 and ends on April 30, 2016. This period includes very high levels of volatility (for implied volatilities from the S&P 500 and VIX options as well as the VIX index values). The last column of Table 5 of the paper presents the number of options within each moneyness and maturity range in both periods. In particular, in the in-sample period our data contains 4'997 close-to-maturity OTM options on the S&P 500 and 2'283 OTM call options on the VIX. These options have maturities shorter than two months. Analogously, in the out-of-sample period, the dataset contains 27'615 close-to-maturity options on the S&P 500 and 6'994 on the VIX. As highlighted in Bollerslev and Todorov (2011), these options provide valuable information on jumps as they have little value unless a large movement in the S&P 500 is possible.

B.2. Measurement equations

We discretize the continuous-time model on a uniform time grid composed of $M+1$ points $t \in \{t_0 = 0, t_1 = \Delta t, ..., t_k = k\Delta t, ..., t_M = M\Delta t\}$, for some $M \in \mathbb{N}^*$. Since we use daily data, Δt corresponds to one day. In discrete time, the model evolves under $\mathbb{P}$ as follows:

$$\begin{aligned}
\Delta Y_t &= [-\lambda^{(-)}(v_t, m_t, u_t)(\theta_Z^{\mathbb{P}}(0,1,0,0) - 1) - \lambda^{(+)}(v_t, m_t)(\theta_Z^{\mathbb{P}}(1,0,0,0) - 1) - \frac{1}{2}v_t + \gamma_t]\Delta t \\
&\quad + \sqrt{v_t}\Delta W_t^{Y,\mathbb{P}} + Z_t^{Y(+),\mathbb{P}}\Delta N_t^{(+)} + Z_t^{Y(-),\mathbb{P}}\Delta N_t^{(-)}, \qquad \text{(B.1)} \\
\Delta v_t &= \kappa_v^{\mathbb{P}}\left(\frac{\kappa_v}{\kappa_v^{\mathbb{P}}}m_t - v_t\right)\Delta t + \sigma_v\sqrt{v_t}\Delta W_t^{v,\mathbb{P}} + Z_t^{v(-),\mathbb{P}}\Delta N_t^{(+)} + Z_t^{v(+),\mathbb{P}}\Delta N_t^{(-)}, \qquad \text{(B.2)} \\
\Delta m_t &= \kappa_m^{\mathbb{P}}(\theta_m^{\mathbb{P}} - m_t)\Delta t + \sigma_m\sqrt{m_t}\Delta W_t^{m,\mathbb{P}}, \qquad \text{(B.3)} \\
\Delta u_t &= \kappa_u(\theta_u - u_t)\Delta t + \sigma_u\sqrt{m_t}\Delta W_t^{u,\mathbb{P}}, \qquad \text{(B.4)}
\end{aligned}$$

[7] We have decided to include the beginning of the financial crisis so that the in-sample period actually includes several dates with extreme events.

where the notation ΔX_t for some process X represents the increment $X_{t_{k+1}} - X_{t_k}$ with $t = t_k \in \{t_0, ..., t_{M-1}\}$. In what follows, we will assume that the long-term mean of the process u_t is normalized to one, i.e., $\theta_u = 1$.

As the log-returns are observable, Eq. (B.1) is the first measurement equation. The second measurement equation comes from the observation of the VIX index level with error:

$$\mathrm{VIX}_t^2 - (\alpha_{\mathrm{VIX}^2} v_t + \beta_{\mathrm{VIX}^2} m_t + \gamma_{\mathrm{VIX}^2} u_t + \delta_{\mathrm{VIX}^2}) = \epsilon_t^{\mathrm{VIX}}. \tag{B.5}$$

Jiang and Tian (2007) point to systematic biases in the calculation of the VIX index, such as model misspecification or data limitations. For example, in practice, the index is calculated using a finite number of options thereby inducing an error in the computation of the integral defining VIX^2. These biases are captured by the error term $\epsilon_t^{\mathrm{VIX}}$, which is assumed to follow a normal distribution with mean zero and variance $s > 0$, $\epsilon_t^{\mathrm{VIX}} \sim N(0, s)$.

To better identify the total variance of S&P 500 returns under the $\mathbb{P}$ measure, we add a measurement equation, which links the logarithm of the daily Realized Variance (RV_t) of S&P 500 returns[8] to the logarithm of the total spot variance under $\mathbb{P}$, as in Filipović et al. (2016). The associated measurement error ϵ_t is conditionally normally distributed with mean $\rho_\epsilon \epsilon_{t-1}$ and variance $c_0 + c_1 \mathrm{RV}_{t-1}$. The rationale behind this component of the measurement equation is the following. Andersen et al. (2001), among others, provide empirical evidence that $\log(\mathrm{RV}_t)$ is approximately normally distributed. The conditional mean specification of ϵ_t allows for autocorrelation in the measurement error, which can be induced by clustering of price jumps caused by persistence of the price jump intensity and/or microstructure noise in the estimates of daily realized variance. Autocorrelation in the measurement error is also reported in Wu (2011). The conditional variance specification of ϵ_t captures in a parsimonious way the heteroscedasticity of the measurement error due to the volatility of realized variance.

The last measurements are the prices of S&P 500 and VIX options. We assume that the option prices are observed with an error. This error represents several sources of noise, such as bid–ask spreads, timing, and processing errors. We define these errors as the relative differences between

[8]The Realized Variance (RV) of the S&P 500 index is obtained from the website of the Oxford-Man Institute Realized Library.

market $O_t^{\mathcal{M},Mkt}$ and model prices $O_t^{\mathcal{M},Mod}$, $\mathcal{M} \in \{\text{SPX}, \text{VIX}\}$:

$$\frac{O_{t,i}^{\text{SPX},Mod}(Y_t, v_t, m_t, \Theta, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,i}^{\text{SPX},Mkt}}{O_{t,i}^{\text{SPX},Mkt}} = \epsilon_{t,i}^{\text{SPX},options}, \quad i = 1, \ldots, N_{\text{SPX},t}, \tag{B.6}$$

$$\frac{O_{t,j}^{\text{VIX},Mod}(v_t, m_t, \Theta, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,j}^{\text{VIX},Mkt}}{O_{t,j}^{\text{VIX},Mkt}} = \epsilon_{t,j}^{\text{VIX},options}, \quad j = 1, \ldots, N_{\text{VIX},t}, \tag{B.7}$$

where $N_{\mathcal{M},t}$ is the number of contracts available in the corresponding market and the Θ's are the sets of parameters to estimate:

$$\begin{aligned}
\Theta^{\mathbb{P}} &= \{\kappa_v^{\mathbb{P}}, \kappa_m^{\mathbb{P}}, \theta_m^{\mathbb{P}}, \nu_m^{\mathbb{P}}, \nu_v^{(+)\mathbb{P}}, \nu_v^{(-)\mathbb{P}}, \mu_Y^{(+)\mathbb{P}}, \mu_Y^{(-)\mathbb{P}}, \eta_Y\}, \\
\Theta &= \{\kappa_v, \kappa_m, \theta_m, \nu_v^{(+)}, \nu_v^{(-)}, \mu_Y^{(+)}, \mu_Y^{(-)}\}, \\
\Theta^{\mathbb{P},\mathbb{Q}} &= \{\lambda_0^{(-)}, \lambda_1^{(-)}, \lambda_2^{(-)}, \lambda_3^{(-)}, \lambda_0^{(+)}, \lambda_1^{(+)}, \lambda_2^{(+)}, \sigma_m, \sigma_v, \sigma_u, \kappa_u, \rho_{Yv}\}.
\end{aligned}$$

We assume the error terms to be normally distributed and heteroscedastic:

$$\epsilon_{t,i}^{\text{SPX},options} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{t,i}^{\text{SPX}}}), \quad \epsilon_{t,j}^{\text{VIX},options} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{t,j}^{\text{VIX}}}), \tag{B.8}$$

The variance of the errors is

$$\sigma^2_{\epsilon_{t,i}^{\text{SPX}}} = \exp\left(\phi_0 \cdot \text{bid-ask spread}_i + \phi_1 \left|\log\left(\frac{K_{\text{VIX},i}}{F_t^{\text{SPX}}(T_i)}\right)\right| + \phi_2(T_i - t) + \phi_3\right), \tag{B.9}$$

$$\sigma^2_{\epsilon_{t,j}^{\text{VIX}}} = \exp\left(\psi_0 \cdot \text{bid-ask spread}_j + \psi_1 \left|\log\left(\frac{K_{\text{SPX},j}}{F_t^{\text{VIX}}(T_j)}\right)\right| + \psi_2(T_j - t) + \psi_3\right), \tag{B.10}$$

with ϕ_i and ψ_i in $\mathbb{R}$, $i \in \{0, 1, 2, 3\}$.[9]

[9]The fact that the option pricing errors are normally distributed does not constitute a restriction. The reason is that the errors are heteroscedastic and the coefficients generating the heteroscedasticity are driven by the data, i.e., we optimize over the parameters $\{\phi_i, \psi_i\}_{0 \le i \le 3}$.

B.3. Filtering methodology

We follow the notation used in Section 4 of the paper. The log-likelihood of a time series of $n+1$ observations with joint density p, conditional on a set of parameters Θ, is equal to

$$\log p(y^{t_n}|\Theta) = \log p(y_{t_0}, ..., y_{t_n}|\Theta) = \sum_{k=1}^{n} \log p(y_{t_k}|y^{t_{k-1}}, \Theta) + \log p(y_{t_0}|\Theta), \tag{B.11}$$

where, by the Law of Total Probability,

$$p(y_{t_k}|y^{t_{k-1}}, \Theta) = \int p(y_{t_k}|L_{t_k}, \Theta) p(L_{t_k}|y^{t_{k-1}}, \Theta) dL_{t_k}. \tag{B.12}$$

Given an initial density $p(L_{t_0}|\Theta)$, the transition density of the state variables $p(L_{t_k}|L_{t_{k-1}}, \Theta)$ and the likelihood function $p(y_{t_k}|L_{t_k}, \Theta)$, filtering methods allow us to estimate the distribution $p(L_{t_k}|y^{t_k}, \Theta)$ of the current state at time $t_k = k\Delta t$, given all observations up to that time. In the following, we simplify notation and drop the subscript for the conditioning on the parameters Θ. The filtering density is given by Bayes's formula,

$$p(L_{t_k}|y^{t_k}) \propto p(y_{t_k}|L_{t_k}) p(L_{t_k}|y^{t_{k-1}}), \tag{B.13}$$

where $\propto$ means proportional to.

The likelihood function is known, but the predictive distribution of the state is not. It is given by the following integral, which involves the previous filtering density:

$$p(L_{t_k}|y^{t_{k-1}}) = \int p(L_{t_k}|L_{t_{k-1}}) p(L_{t_{k-1}}|y^{t_{k-1}}) dL_{t_{k-1}}. \tag{B.14}$$

The key idea is to approximate the posterior density function of the latent variables $p(L_{t_k}|y^{t_k})$ by a sum of point masses positioned at strategic points, called particles, $\{L_{t_k}^{(i)}\}_{1 \leq i \leq n_p}$:

$$\hat{p}(L_{t_k}|y^{t_k}) = \sum_{i=1}^{n_p} \pi_{t_k}^{(i)} \delta(L_{t_k} - L_{t_k}^{(i)}), \tag{B.15}$$

where $\pi_{t_k}^{(i)}$ denotes the normalized importance weight for particle i, $\delta(.)$ is the Dirac delta function,

and n_p is the number of support points (particles) for $\hat{p}(L_{t_k}|y^{t_k})$. Then, we can recursively calculate the filtering density by

$$\begin{aligned}\hat{p}(L_{t_k}|y^{t_k}) &\propto \int p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}})\hat{p}(L_{t_{k-1}}|y^{t_{k-1}})dL_{t_{k-1}} \\ &= \sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}}^{(i)})\pi_{t_{k-1}}^{(i)}. \end{aligned} \tag{B.16}$$

To implement the particle filter, we need to simulate at every time t_k a number n_p of particles $L_{t_k}^{(i)}, i = 1, ..., n_p$ from $p(L_{t_k}|y^{t_{k-1}})$ and to evaluate $p(y_{t_k}|L_{t_k}^{(i)})$. Based on these simulated particles, we can approximate $p(y_{t_k}|y^{t_{k-1}})$ by

$$p(y_{t_k}|y^{t_{k-1}}) \approx \frac{1}{n_p}\sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k}^{(i)}). \tag{B.17}$$

We used $n_p = 30,000$ particles on days when the observations contain option prices and $n_p = 10,000$ when the observations are only composed of the S&P500 returns and VIX index levels. Larger numbers of particles did not change our estimates, but increased the computational burden.

The filtering algorithm can be decomposed into the following steps.

Step 1: Initialization. We simulate n_p initial particles for the latent variables $\{v_{t_0}^{(i)}, m_{t_0}^{(i)}, u_{t_0}^{(i)}\}_{i=1,...,n_p}$, which are compatible with the initial value of the VIX squared, i.e., given the specification in Eq. (B.5). The following steps are repeated for each time step t_k in the grid from $k = 0$ to $k = M - 1$.

Step 2: First-stage resampling. At this point, we assume that we have n_p particles (i.e., possible values of m_t, v_t, and u_t) at time t_k given all observations y^{t_k} up to t_k. At time t_{k+1}, there are new observations $y_{t_{k+1}}$. The goal of this step is to retain, from the previous sample of particles $\{v_{t_k}^{(i)}, m_{t_k}^{(i)}, u_{t_k}^{(i)}\}_{1\leq i\leq n_p}$, only those which are likely to generate the new observations $y_{t_{k+1}}$. For this purpose, we assign a weight (the so called "first-stage weights") to each particle, which is proportional to the likelihood of new market observations $y_{t_{k+1}}$ given the value of the particle L_{t_k} at time t_k. Intuitively, particles that are compatible with the new observations will be assigned larger weights than other particles. To increase the speed of the first-stage resampling, we do not consider options as part of the observations $y_{t_{k+1}}$ (only in this step) and limit $y_{t_{k+1}}$ to the values of the indexes.

The first stage weight $\omega_{t_{k+1}}^{(i)}$ assigned to the i^{th} particle $L_{t_k}^{(i)}$ at time t_{k+1} is given by

$$\omega_{t_{k+1}}^{(i)} = p(L_{t_k}^{(i)}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|L_{t_k}^{(i)}),$$

where $p(y_{t_{k+1}}|L_{t_k}^{(i)})$ is the density of the observation vector $y_{t_{k+1}}$ given the values of the particle vector $L_{t_k}^{(i)}$. The importance weights $\{\omega_{t_{k+1}}^{(i)}\}_{1\leq i\leq n_p}$ add up to 1, so that they define a proper probability mass function. Conditioning on the number of jumps in ΔY_t (or equivalently in Δv_t) and Δm_t gives

$$\omega_{t_{k+1}}^{(i)} \propto \sum_{j,l\in\mathbb{N}} p(y_{t_{k+1}}|L_{t_k}^{(i)}, \Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)})\mathbb{P}(\Delta N_{t_k}^{(+)} = j, \Delta N_{t_k}^{(-)} = l).$$

Given that we use daily observations, we limit the possible number of jumps of the Poisson random variables $\Delta N_{t_k}^{(+)}$ and $\Delta N_{t_k}^{(-)}$ to zero or one.[10] We recall that the new observation is composed of the S&P 500 returns and the VIX level $y_{t_{k+1}} = (\Delta Y_{t_k} = Y_{t_{k+1}} - Y_{t_k}, \text{VIX}_{t_{k+1}}^2)$. Since the $\text{VIX}_{t_{k+1}}^2$ is a sum of normal distributions and no more than two exponential distributions, there is no closed form for this bivariate density in the general case. To preserve tractability, we approximate the exponentially distributed jump sizes by a categorical distribution (a generalization of a Bernoulli distribution), which is supported in a certain number of (the corresponding exponential distribution's) quantiles.[11] As a consequence, the weight $\omega_{t_{k+1}}^{(i)}$ is a sum of weighted bivariate normal densities.

To eliminate the particles $\{L_{t_k}^{(i)}\}_{1\leq i\leq n_p}$ that are not likely to generate the new observations $y_{t_{k+1}}$, we resample (with replacement) the particles according to a stratified resampling scheme:[12]

$$z(i) \sim StratRes(n_p, \omega_{t_{k+1}}^{(1)}, ..., \omega_{t_{k+1}}^{(n_p)}).$$

We have now new sample of n_p latent factors $\{L_{t_k}^{(j)}\}_{1\leq j\leq n_p}$, which are now equally likely. Indeed, particles representing m_{t_k}, v_{t_k}, and u_{t_k} are shuffled into a new set of particles: $\{m_{t_k}^{(j)}, v_{t_k}^{(j)}, u_{t_k}^{(j)}\}_{j=1..n_p} = \{m_{t_k}^{z(i)}, v_{t_k}^{z(i)}, u_{t_k}^{z(i)}\}_{i=1..n_p}$. We resample the same number of particles, although this is in principle not necessary.

The next step of the particle filter consists in propagating the latent factors according to their

[10] This Bernoulli approximation is found to be very accurate in Johannes et al. (2009).
[11] Robustness tests were performed on simulated data to check that the choice of quantiles was appropriate.
[12] We checked that using a multinomial or stratified resampling scheme gives similar results.

conditional density given the previous values $L_{t_k}^{(i)}$ and the new observations $y_{t_{k+1}}$:

$$L_{t_{k+1}}^{(i)} \sim p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}}).$$

Because the distribution $p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}})$ is not known in closed form, we use a proposal density $q(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}})$. Propagating v_t, m_t, and u_t requires preliminary knowledge on the jump components, so we first focus on the jumps.

Step 3: Generating the jumps. We calculate the joint probability of jumps in ΔY_t (or equivalently in Δv_t) between t_k and t_{k+1} using

$$\mathbb{P}(\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)})\mathbb{P}(\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)}). \tag{B.18}$$

Conditionally on the jump sizes, the first part of the right-hand side has already been calculated in the first-stage weights. Using Bayes's rule, we get an approximation for $\mathbb{P}(\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)}|y_{t_{k+1}})$.

We simulate the jump sizes for Δv_t and Δm_t according to their exponential law.

Step 4: Propagating the latent factors. The latent factors v, m, and u are propagated following a Milstein discretization scheme of the SDE. See Kloeden and Platen (1992) for details. We use the full truncation method to prevent them from taking negative values.

Step 5: Computing the filtering density. At this point, the newly generated particles $\{L_{t_{k+1}}^{(i)}\}_{1\leq i\leq n_p}$ are a sample of $p(L_{t_{k+1}}|y^{t_{k+1}})$. We now calculate the *second-stage weights* $\{\pi_{t_{k+1}}^{(i)}\}_{1\leq i\leq n_p}$ which approximate the probabilities $p(L_{t_{k+1}}^{(i)}|y^{t_{k+1}})$, and give an approximation for the filtering density at time t_{k+1}. These weights are proportional to the likelihood of observations at time t_{k+1} given the propagated particles $L_{t_{k+1}}^{(i)}$, with a correction related to the proposal density

$$\pi_{t_{k+1}}^{(i)} \propto \frac{p(L_{t_{k+1}}^{(i)}|L_{t_k}^{(i)})p(y_{t_{k+1}}|L_{t_{k+1}}^{(i)})}{\omega_{t_{k+1}}^{z(i)} q(L_{t_{k+1}}^{(i)}|L_{t_k}^{(i)}, y_{t_{k+1}})}.$$

The posterior distribution of the state variables is approximated by

$$\hat{p}(L_{t_{k+1}}|y^{t_{k+1}}) = \sum_{i=1}^{n_p} \pi_{t_{k+1}}^{(i)} \delta(L_{t_{k+1}} - L_{t_{k+1}}^{(i)}).$$

We choose the most likely value of a given factor by taking the expectation of the estimated filtering density, e.g., $\hat{v}_{t_{k+1}} = \mathbb{E}_{\hat{p}}[v_{t_{k+1}}^{(i)}]$.

The algorithm described above extracts latent factors, if one assumes that the model parameters are known. Pitt (2002) builds on Gordon et al. (1993) to show that the parameters can be estimated using the Maximum Likelihood Importance Sampling Criterion, defined as the product over time of the averages of the second-stage weights. The likelihood of observations given the values of the particles is then estimated by the average of the second-stage weights over the particles

$$\hat{p}(y^{t_n}|\Theta, \mathcal{M}) = \prod_{k=1}^{M} \hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) \hat{p}(y_{t_0}|\Theta, \mathcal{M}),$$

where $\hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) = \frac{1}{n_p} \sum_{i=1}^{n_p} \pi_{t_k}^{(i)}$.

Finally, we use the Weighted Likelihood function analyzed in Hu and Zidek (2002) to assign comparable weights to the different datasets and to ensure that the estimation is not dominated by S&P 500 options. Such procedure has been used by Ornthanalai (2014) to estimate a model with Lévy jumps to S&P 500 options and returns.

C. Additional model diagnostics

This appendix provides additional diagnostics of model specification for the SVJ3 model, including VIX options in the estimation dataset. In Figure 2, Panels A and B display the mean of the estimated posterior distribution of the Brownian motions driving the S&P 500 returns and the short-term variance factor v. The filtered Brownian motion driving the returns exhibits excess negative skewness, which is probably due to the fact that the jumps captured by the model are large jumps (around -10%). Due to the financial crisis being in our in-sample period, the model puts emphasis on these catastrophic jumps, which causes it to neglect small negative jumps. These jumps are captured by the Brownian motion instead.

[**Fig.** 2 about here.]

We find a correlation of -0.75 between the two filtered Brownian motions, consistently with the estimated value of the leverage coefficient (-0.74). The correlation between the filtered Brownian motions driving v and m is -0.09. Panel C of Figure 2 compares the model-implied VIX values to the data, and shows that the model provides an excellent fit to the VIX data. This is confirmed by small RMSEs. The in-sample RMSE is 0.026 (0.016 if excluding the period starting in September 2008). The out-of-sample RMSE is 0.020. Panel D represents the error between the true VIX squared and the model-implied value across time. The assumption of independent and identically distributed errors is clearly violated. In particular, errors are much larger, consistently with intuition, during high volatility periods. However, even during these periods, the errors remain very small. Panel E of Figure 2 compares the trajectory of the realized variance of S&P 500 returns to the estimated trajectory of the model-implied quadratic variation, i.e., total spot variance of returns. There is a satisfactory overlap of the two curves.

Figure 3 represents, for every month, two measures of realized variation (annualized) computed from high-frequency data, versus the model $\mathbb{P}$-expectation of their limiting quantity at the beginning of the month. Panel A plots the ex-post realized variance and compares it to the expected total variance. Due to averaging over a month, the ex-post variance is smoother than its expectation, which adjusts every day to new information. The two curves are reasonably close. Similarly, Panel B displays the tri-power quarticity versus the expected integrated quarticity.[13] The tri-power quarticity is an estimate of the integrated quarticity which is robust to jumps, see Barndorff-Nielsen and Shephard (2004). The ex-post realized quarticity tends to be higher than its first conditional moment during the crisis, but smaller otherwise. Overall, these plots do not underline obvious shortcomings of the model, considering the measurement errors made when calculating measures of variation from high-frequency returns.

[**Fig.** 3 about here.]

[13] Our dataset for the tri-power quarticity ends in 2010. Therefore, we only display the comparison between the data- and model implied quarticity until this date.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, The distribution of realized stock return volatility, Journal of Financial Economics 61, 43–76.

Barndorff-Nielsen, O. E., and N. Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, Journal of Financial Econometrics 2, 1–48.

Bollerslev, T., and V. Todorov, 2011, Tails, fears, and risk premia, Journal of Finance 66, 2165–2211.

Broadie, M., M. Chernov, and M. Johannes, 2007, Model specification and risk premia: evidence from futures options, Journal of Finance 62, 1453–1490.

Cont, R., and T. Kokholm, 2013, A consistent pricing model for index options and volatility derivatives, Mathematical Finance 23, 248–274.

Filipović, D., E. Gourier, and L. Mancini, 2016, Quadratic variance swap models, Journal of Financial Economics 119, 44–68.

Gatheral, J., 2008, Consistent modeling of SPX and VIX options, Bachelier Congress 37.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proceedings F, Radar and Signal Processing 140, 107–113.

Hansen, N., and A. Ostermeier, 1996, Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation, Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC 96) 312–317.

Hu, F., and J. V. Zidek, 2002, The weighted likelihood, Canadian Journal of Statistics 30, 347–371.

Jiang, G. J., and Y. S. Tian, 2007, Extracting model-free volatility from option prices: an examination of the VIX index, Journal of Derivatives 14, 35–60.

Johannes, M. S., N. G. Polson, and J. R. Stroud, 2009, Optimal filtering of jump diffusions: extracting latent states from asset prices, Review of Financial Studies 22, 2759–2799.

Kloeden, P. E., and E. Platen, 1992, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, Germany).

Lindström, E., J. Ströjby, M. Brodén, M. Wiktorsson, and J. Holst, 2008, Sequential calibration of options, Computational Statistics and Data Analysis 52, 2877–2891.

Ornthanalai, C., 2014, Lévy jump risk: evidence from options and returns, Journal of Financial Economics 112, 69–90.

Pan, J., 2002, The jump-risk premia implicit in options: evidence from an integrated time-series study, Journal of Financial Economics 63, 3–50.

Pitt, M. K., 2002, Smooth particle filters for likelihood evaluation and maximisation, Unpublished working paper 651, University of Warwick.

Rebonato, R., and T. Cardoso, 2004, Unconstrained fitting of implied volatility surfaces using a mixture of normals, Journal of Risk 7, 55–74.

Sepp, A., 2008a, Pricing options on realized variance in the Heston model with jumps in returns and volatility, Journal of Computational Finance 11, 33–70.

Sepp, A., 2008b, VIX option pricing in a jump-diffusion model, Risk Magazine 84–89.

Storn, R., 1996, On the usage of differential evolution for function optimization, in M. H. Smith, M. A. Lee, J. Keller, and J. Yen, eds., Proceedings of the 1996 Biennial Conference of the North American Fuzzy Information Processing Society, 519–523 (IEEE Press, New York).

Wu, L., 2011, Variance dynamics: joint evidence from options and high-frequency returns, Journal of Econometrics 160, 280–287.

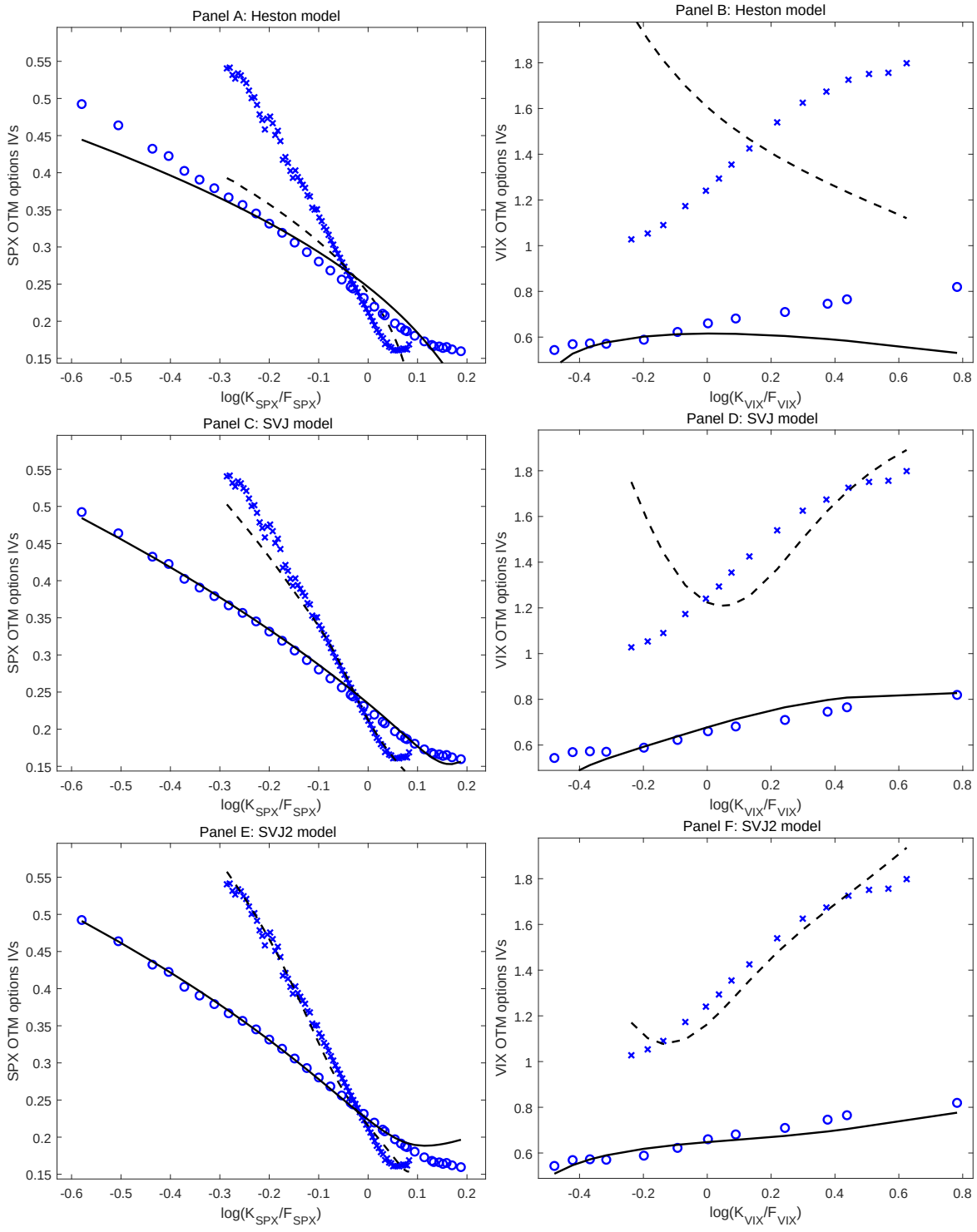

Fig. 1. This figure represents fitted IVs on May 5, 2010, obtained by a joint calibration on the S&P 500 and VIX options. Circles represent the market IV for $T = 0.05$ (S&P 500) and $T = 0.04$ (VIX). Crosses represent the market IV for $T = 0.3$ (S&P 500) and $T = 0.36$ (VIX). The dashed line corresponds to the model fit for $T = 0.05$ (S&P 500) and $T = 0.04$ (VIX). The solid line corresponds to the model fit for $T = 0.3$ (S&P 500) and $T = 0.36$ (VIX). Panels A (S&P 500) and B (VIX) plot the model IVs for the Heston model, Panels C and D for the SVJ model, and Panels E and F for the SVJ2 model.

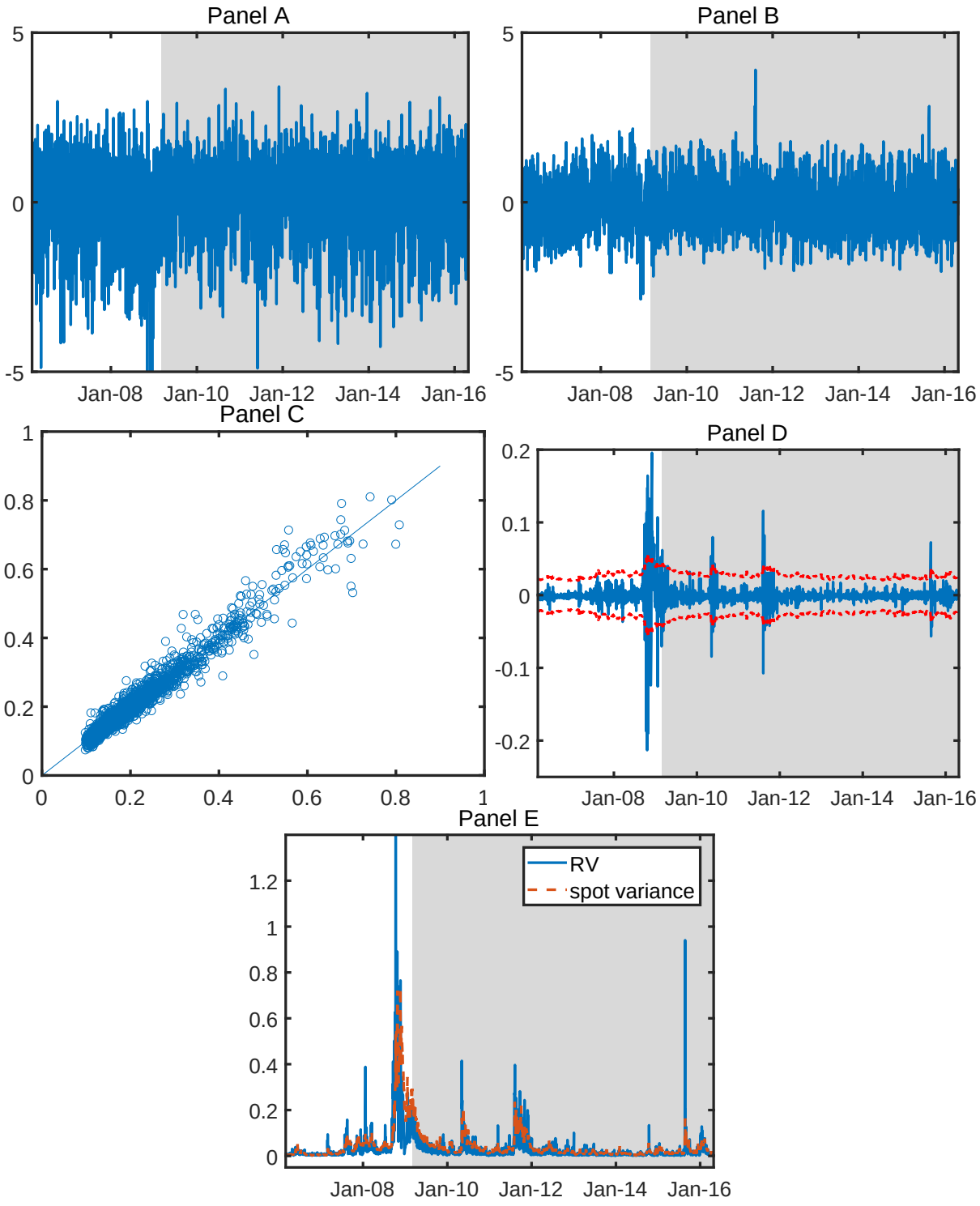

Fig. 2. This figure provides model diagnostics to evaluate the SVJ3 specification, when the estimation dataset includes VIX options. Panels A and B represent the filtered Brownian motions driving the dynamics of the S&P 500 returns and their variance factor v. Panel C is a scatterplot comparing model-implied VIX values to the data across the (in- and out-of-sample) time series. Panel D plots the error between the true VIX squared and the filtered value throughout time. Panel E represents the realized variance computed from high-frequency data and the estimated trajectory of the model-implied quadratic variation under $\mathbb{P}$.

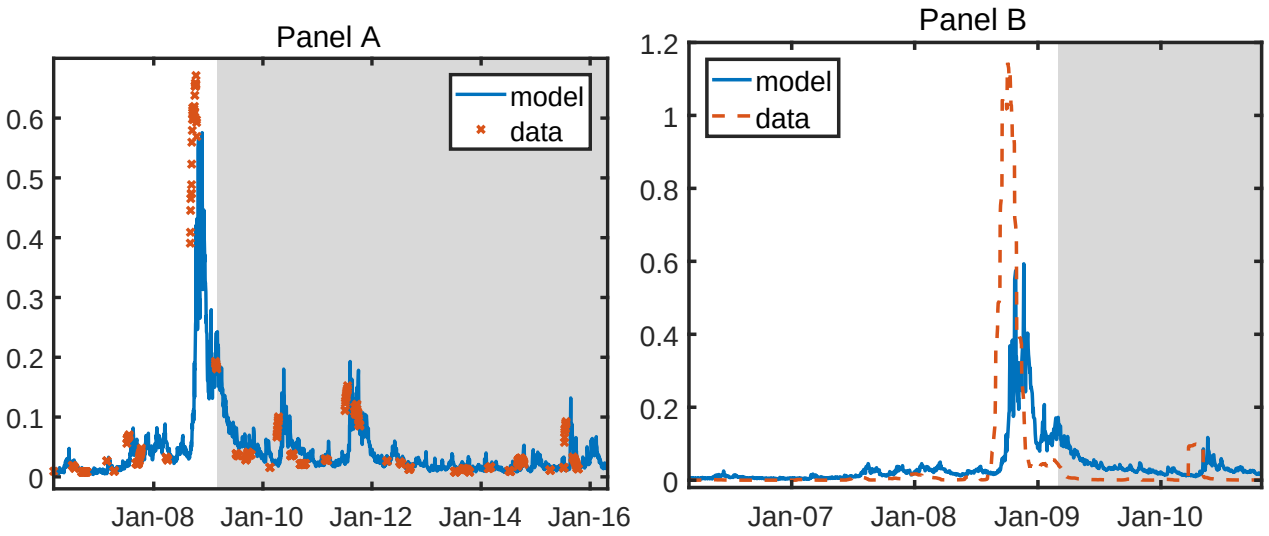

Fig. 3. This figure compares, for every month, future measures of variation calculated from high frequency returns, to the model-implied expectation of their counterpart. Panel A represents the total variance realized during each month, versus the model $\mathbb{P}$-expectation of the total variance at the beginning of the month. Panel B represents the ex-post tri-power quarticity versus the expectation of the integrated quarticity (based on data available up to end of 2010).