# Supplementary Material to "Jump Contagion among Stock Market Indices: Evidence from Option Markets"

H. Peter Boswijk[1,2], Roger J. A. Laeven[1,3,*], Andrei Lalu[1,2], and Evgenii Vladimirov[1,2]

[1]Amsterdam School of Economics, University of Amsterdam, The Netherlands

[2]Tinbergen Institute, The Netherlands

[3]EURANDOM and CentER, The Netherlands

January 16, 2022

**Abstract**

This supplementary material contains details concerning: the change of measure, asymptotic properties of the estimation procedure, jump-robust volatility estimation, and data selection and processing.

*Corresponding author. University of Amsterdam, Amsterdam School of Economics, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Phone: +31 6 52787501.

## Appendix A Change of Measure

This appendix provides further details on the candidate pricing kernels and the change of measure for the model specification discussed in Section 2. In particular, we show that the choice of the pricing kernel for each of the markets rules out arbitrage opportunities within each market, as well as internationally. Furthermore, we show that under the risk-neutral measures, the jump intensity dynamics are unaffected.

Similar to Pan (2002), one can show that the stochastic discount factor $M_{i,t}$ in Eqn. (5) ensures that the deflated index processes $\mathcal{S}^i_{i,t} := M_{i,t} S_{i,t} \exp(\int_0^t q_{i,s} ds)$ and the deflated money market account processes $\mathcal{B}_{i,t} := M_{i,t} \exp(\int_0^t r_{i,s} ds)$ are local martingales. In fact, applying Itô's formula, we have:

$$d\mathcal{B}_{i,t} = \mathcal{B}_{i,t} \left( -\eta_i \xi_{i,t} dW_{i,t} + \sum_{k=1}^{m} U^i_{k,t} dN_{k,t} \right),$$

with $\mathbb{E}[U^i_{k,t}] = 0$ (from the constraint $a_{i,k} + \frac{1}{2} b^2_{i,k} = 0$), and

$$d\mathcal{S}^i_{i,t} = \mathcal{S}^i_{i,t} \left[ (1-\eta_i)\xi_{i,t} dW_{i,t} - \mathbb{E}^{\mathbb{Q}_i}[J_{i,t}]\lambda_{i,t} dt + (\exp(V^i_{i,t} + Z_{i,t}) - 1) dN_{i,t} + \sum_{k \neq i} U^i_{k,t} dN_{k,t} \right],$$

where

$$\begin{aligned} \mathbb{E}[\exp(V^i_i + Z_i) - 1] &= \exp\left(a_{i,i} + \tfrac{1}{2} b^2_{i,i} + \mu_i + \rho_{i,i} b_{i,i} \sigma_i + \tfrac{1}{2}\sigma_i^2\right) - 1 \\ &= \exp\left(\mu_i^{\mathbb{Q}_i} + \tfrac{1}{2}\sigma_i^2\right) - 1 = \mathbb{E}^{\mathbb{Q}_i}[J_{i,t}], \end{aligned}$$

with $\mu_i^{\mathbb{Q}_i} = \mu_i + \rho_{i,i} b_{i,i} \sigma_i$. Therefore, the processes $\mathcal{S}^i_{i,t}$ and $\mathcal{B}_{i,t}$ are indeed local martingales.

Furthermore, in the international setting, the deflated foreign index processes and foreign money market accounts, denominated in the currency of market $i$, have to be local martingales as well. In other words, the processes $\mathcal{S}^i_{j,t} := M_{i,t} E_{ij,t} S_{j,t} \exp(\int_0^t q_{j,s} ds)$ and $\mathcal{B}^i_{j,t} := M_{i,t} E_{ij,t} \exp(\int_0^t r_{j,s} ds)$ need to be local martingales, where $E_{ij,t}$ is the exchange rate between markets $i$ and $j$, i.e., the price in currency $i$ of one unit of currency $j$. This is guaranteed, and hence arbitrage opportunities across all economies are ruled out, whenever the exchange rate dynamics $E_{ij,t}$ are such that $M_{j,t} = M_{i,t}E_{ij,t}$ (see, for example, Brandt and Santa-Clara (2002), Backus, Foresi, and Telmer (2001)).

Therefore, arbitrage-free exchange rate dynamics can be derived from the ratio of foreign to domestic pricing kernels:

$$\begin{aligned} dE_{ij,t} &= d\left(\frac{M_{j,t}}{M_{i,t}}\right) \\ &= E_{ij,t}\left[(-r_{j,t}dt - \eta_j\xi_{j,t}dW_{j,t}) - (-r_{i,t}dt - \eta_i\xi_{i,t}dW_{i,t})\right] \\ &\quad + E_{ij,t}\left[\left(\eta_i^2\xi_{i,t}^2 - \eta_i\xi_{i,t}\eta_j\xi_{j,t}\varrho_{ij,t}\right)dt + \sum_{k=1}^{m}\left(\frac{1+U_{k,t}^j}{1+U_{k,t}^i} - 1\right)dN_{k,t}\right], \end{aligned}$$

where $\varrho_{ij,t}$ is the instantaneous correlation between the Brownian motions $W_{i,t}$ and $W_{j,t}$. Using the log-normal parametrization for the relative jump sizes in the pricing kernels, that is, $U_{k,t}^i = e^{V_{k,t}^i} - 1$ with $V_k^i \sim \mathcal{N}(a_{i,k}, b_{i,k}^2)$, we have

$$\begin{aligned} \frac{dE_{ij,t}}{E_{ij,t}} &= (r_{i,t} - r_{j,t} + \eta_i^2\xi_{i,t}^2 - \eta_i\xi_{i,t}\eta_j\xi_{j,t}\varrho_{ij,t})dt + \eta_i\xi_{i,t}dW_{i,t} - \eta_j\xi_{j,t}dW_{j,t} \\ &\quad + \sum_{k=1}^{m}\left(e^{V_{k,t}^j - V_{k,t}^i} - 1\right)dN_{k,t}. \end{aligned} \tag{A.1}$$

The resulting exchange rate processes feature both diffusive components with stochastic volatility and compound jump process components. In our set-up, we allow the exchange rate processes to jump simultaneously with jumps in any of the markets, and the jump sizes depend on how these jumps are perceived in the markets $i$ and $j$. More specifically, due to the parametrization assumption, the exchange rate $E_{ij,t}$ jumps simultaneously with a jump in a market $k$ with log-jump size $V_k^j - V_k^i \sim \mathcal{N}(a_{j,k} - a_{i,k}, b_{j,k}^2 - b_{i,k}^2)$.

Define the equivalent martingale measure $\mathbb{Q}_i$ in market $i$ from the Radon-Nikodym density process $\psi_{i,t}$, satisfying

$$\frac{d\psi_{i,t}}{\psi_{i,t}} = -\eta_i\xi_{i,t}dW_{i,t} + \sum_{k=1}^{m} U_{k,t}^i dN_{k,t}. \tag{A.2}$$

Under $\mathbb{Q}_i$, the processes

$$W_{j,t}^{\mathbb{Q}_i} = W_{j,t} + \int_0^t \eta_i \xi_{i,s} \varrho_{ij,s} ds, \qquad j = 1, \dots, m,$$

are standard Brownian motions with the same instantaneous correlations as the original Brownian motions under $\mathbb{P}$. Note that $\varrho_{ii,t} = 1$, so that $W_{i,t}^{\mathbb{Q}_i} = W_{i,t} + \int_0^t \eta_i \xi_{i,s} ds$.

Under the defined equivalent measure $\mathbb{Q}_i$, the discounted foreign asset prices denominated in currency $i$ are $\mathbb{Q}_i$-martingales. To see this, define $\tilde{B}_{j,t}^i := \exp(-\int_0^t r_{i,s} ds) E_{ij,t} \exp(\int_0^t r_{j,s} ds)$ and $\tilde{S}_{j,t}^i := \exp(-\int_0^t r_{i,s} ds) E_{ij,t} S_{j,t} \exp(\int_0^t q_{j,s} ds)$. By applying Itô's formula, the dynamics of these processes under $\mathbb{Q}_i$ can be characterized as follows:

$$\begin{aligned}
\frac{d\tilde{B}_{j,t}^{\mathbb{Q}_i}}{\tilde{B}_{j,t}^{\mathbb{Q}_i}} &= \eta_i \xi_{i,t} dW_{i,t}^{\mathbb{Q}_i} - \eta_j \xi_{j,t} dW_{j,t}^{\mathbb{Q}_i} + \sum_{k=1}^m \left( e^{V_{k,t}^j - V_{k,t}^i} - 1 \right) dN_{k,t}, \\
\frac{d\tilde{S}_{j,t}^{\mathbb{Q}_i}}{\tilde{S}_{j,t}^{\mathbb{Q}_i}} &= (1 - \eta_j) \xi_{j,t} dW_{j,t}^{\mathbb{Q}_i} + \eta_i \xi_{i,t} dW_{i,t}^{\mathbb{Q}_i} + \left( e^{Z_{j,t} + V_{j,t}^j - V_{j,t}^i} - 1 \right) dN_{j,t} \\
&\quad - \mathbb{E}^{\mathbb{Q}_j}[J_{j,t}] \lambda_{j,t} dt + \sum_{k \neq j}^m \left( e^{V_{k,t}^j - V_{k,t}^i} - 1 \right) dN_{k,t}.
\end{aligned}$$

Define $G_t^k := \int_0^t \left( e^{V_{k,s}^j - V_{k,s}^i} - 1 \right) dN_{k,s}$ and $H_t^j := \int_0^t \left( e^{Z_{j,s} + V_{j,s}^j - V_{j,s}^i} - 1 \right) dN_{j,s}$. Then, given the assumptions on the zero mean relative jump sizes in the pricing kernels, i.e., $a_{i,k} + \frac{1}{2} b_{i,k}^2 = 0$ for $k, i = 1, \dots, m$, it follows that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}_i}[G_s^k] &= \mathbb{E}\left[ \psi_{i,t} G_t^k \right] \\
&= \mathbb{E}\left[ - \int_0^t \eta_i \xi_{i,s} G_s^k \psi_{i,s} dW_{i,s} + \int_0^t \psi_{i,s} \left( e^{V_{k,s}^j} - e^{V_{k,s}^i} + G_s^k \left( e^{V_{k,s}^i} - 1 \right) \right) dN_{k,s} \right] \\
&= 0,
\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{E}^{\mathbb{Q}_i}[H_t^j] &= \mathbb{E}\left[\psi_{i,t} H_t^j\right] \\
&= \mathbb{E}\left[-\int_0^t \eta_i \xi_{i,s} H_s^j \psi_{i,s} dW_{i,s} + \int_0^t \psi_{i,s}\left(e^{Z_{j,s}+V_{j,s}^j} - e^{V_{j,s}^i} + H_s^j\left(e^{V_{j,s}^i} - 1\right)\right) dN_{j,s}\right] \\
&= \mathbb{E}\left[\int_0^t \mathbb{E}^{\mathbb{Q}_j}[J_{j,t}]\psi_{i,s}\lambda_{j,s} ds\right].
\end{aligned}
$$

Given that

$$
\mathbb{E}^{\mathbb{Q}_i}\left[\int_0^t \mathbb{E}^{\mathbb{Q}_j}[J_{j,s}]\lambda_{j,s} ds\right] = \mathbb{E}\left[\psi_{i,t}\int_0^t \mathbb{E}^{\mathbb{Q}_j}[J_{j,s}]\lambda_{j,s} ds\right] = \mathbb{E}\left[\int_0^t \mathbb{E}^{\mathbb{Q}_j}[J_{j,s}]\psi_{i,s}\lambda_{j,s} ds\right],
$$

it follows that the discounted processes $\tilde{B}_{j,t}^i$ and $\tilde{S}_{j,t}^i$ are indeed local martingales under $\mathbb{Q}_i$. Therefore, the pricing kernels rule out international arbitrage opportunities.

It is important to note that the jump intensity processes have the same dynamics under the defined equivalent measure $\mathbb{Q}_i$ as under the physical probability measure. To see this, denote the compensated compound Hawkes processes by

$$
\chi_{k,t} = \int_0^t J_{k,t} dN_{k,t} - \int_0^t \mathbb{E}[J_{k,s}]\lambda_{k,s} ds, \qquad k = 1, \ldots, m. \tag{A.3}
$$

The processes $\chi_{k,t}$ are local martingales under $\mathbb{P}$ by definition. Therefore, by the predictable version of the Girsanov-Meyer theorem (see Theorem 41 in Protter (2005)),

$$
\begin{aligned}
\chi_{k,t} - \int_0^t \frac{1}{\psi_{i,s}} d\langle \chi_k, \psi_i \rangle_s &= \chi_{k,t} - \int_0^t \mathbb{E}[J_{k,s} U_{k,s}^i]\lambda_{k,s} ds \\
&= \int_0^t J_{k,t} dN_{k,t} - \int_0^t \left(\mathbb{E}[J_{k,s}] + \mathbb{E}[J_{k,s} U_{k,s}^i]\right)\lambda_{k,s} ds
\end{aligned}
$$

is a local martingale under $\mathbb{Q}_i$. Using again $J_k = e^{Z_k} - 1$ with $Z_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ and $U_k^i = e^{V_k^i} - 1$

with $V_k^i \sim \mathcal{N}(a_{i,k}, b_{i,k}^2)$, we have

$$\begin{aligned}\mathbb{E}[J_{k,s}] + \mathbb{E}[J_{k,s}U_{k,s}^i] &= \mathbb{E}\left[e^{Z_{k,s}+V_{k,s}^i} - e^{V_{k,s}^i}\right] \\ &= \exp\left(a_{i,k} + \tfrac{1}{2}b_{i,k}^2 + \mu_i + \rho_{i,k}b_{i,k}\sigma_k + \tfrac{1}{2}\sigma_k^2\right) - \exp\left(a_{i,k} + \tfrac{1}{2}b_{i,k}^2\right) \\ &= \exp\left(\mu_k^{\mathbb{Q}_i} + \tfrac{1}{2}\sigma_k^2\right) - 1 = \mathbb{E}^{\mathbb{Q}_i}[J_{k,s}],\end{aligned}$$

with $\mu_k^{\mathbb{Q}_i} = \mu_k + \rho_{i,k}b_{i,k}\sigma_k$. Therefore,

$$\int_0^t J_{k,t}dN_{k,t} - \int_0^t \mathbb{E}^{\mathbb{Q}_i}[J_{k,s}]\lambda_{k,s}ds, \qquad k = 1, \dots, m,$$

are $\mathbb{Q}_i$-local martingales, which implies, by the martingale characterization of jump intensities, that $\lambda_{k,t}$ are intensity processes for the corresponding Hawkes processes $N_{k,t}$ under the risk-neutral probability measure as well. In other words, the measure change in economy $i$ does not affect the dynamics of the jump intensities $\lambda_{k,t}$ for $k = 1, \dots, m$, and thus does not change jump times.

# Appendix B Asymptotic Properties of the Estimation Procedure

In this appendix, we derive in detail the asymptotic properties of our estimators. This ultimately leads to expressions for asymptotic standard errors of the parameter estimates in our partial-information implied-state C-GMM procedure.

We start by introducing the required Hilbert space. Let $\pi$ be a probability density function on $\mathbb{R}^d$. We denote by $L^2(\pi)$ the Hilbert space of complex-valued functions such that

$$L^2(\pi) := \left\{ f : \mathbb{R}^d \to \mathbb{C} : \int |f(\tau)|^2 \pi(\tau) d\tau < \infty \right\}.$$

The inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$ on $L^2(\pi)$ are defined as

$$\langle f, g \rangle := \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau, \quad \text{and} \quad \|f\| := \langle f, f \rangle^{\frac{1}{2}},$$

where $\overline{g(\tau)}$ denotes the complex conjugate of $g(\tau)$.

Let us further extend the notion of inner product for vectors of functions in $L^2(\pi)$. For this purpose, we first define the $L^2(\pi)^k$ space of vector functions as

$$L^2(\pi)^k := \left\{\mathbf{f} = (f_1, \ldots, f_k)' : f_i \in L^2(\pi)\right\}.$$

Then the inner product of two (column) vector functions $\mathbf{f} = (f_1, \ldots, f_k)'$ and $\mathbf{g} = (g_1, \ldots, g_k)'$ is defined as

$$\langle \mathbf{f}, \mathbf{g} \rangle := \int \mathbf{f}(\tau)' \overline{\mathbf{g}(\tau)} \pi(\tau) d\tau = \sum_{i=1}^{k} \int f_i(\tau) \overline{g_i(\tau)} \pi(\tau) d\tau.$$

Similarly, for matrices $\mathbf{F}$ and $\mathbf{G}$ of $L^2(\pi)$ functions, with dimensions $k \times p$ and $k \times d$, respectively, $\langle \mathbf{F}, \mathbf{G} \rangle := \int \mathbf{F}(\tau)' \overline{\mathbf{G}(\tau)} \pi(\tau) d\tau$, a $p \times d$ matrix.

Recall that, in the full-information setting, we consider the moment function based on the CCF of the state vector $Y_t$ and its empirical counterpart:

$$h_t(\tau; \hat{v}_t, \theta) := h(\tau, Y_t^{\theta}, Y_{t+1}^{\theta}; \hat{v}_t, \theta) = m(r, Y_t) \left(e^{\mathrm{i}s \cdot g Y_{t+1}} - \phi(s, Y_t, \Delta_t; \hat{v}_t, \theta)\right),$$

where $\tau = (r, s)'$ with $r, s \in \mathbb{R}^{2m}$, and $m(r, Y_t) = e^{\mathrm{i}r \cdot Y_t}$ is an "instrument" function. However, in the partial-information setting, we have $k$ sets of "marginal" moment conditions stacked in the vector

$$\mathbf{h}_t(\tau; \hat{v}_t, \theta) = \begin{pmatrix} h_t^{(1)}(\tau; \hat{v}_t, \theta) \\ \vdots \\ h_t^{(k)}(\tau; \hat{v}_t, \theta) \end{pmatrix},$$

with

$$h^{(i)}(\tau; \hat{v}_t, \theta) = m(r, Y_t^{(i)}) \left(e^{\mathrm{i}s \cdot Y_{t+1}^{(i)}} - \phi^{(i)}(s, Y_t, \Delta_t; \hat{v}_t, \theta)\right), \qquad \text{for } i = 1, \ldots, k,$$

where $r, s \in \mathbb{R}^2$, and where $Y_t^{(i)}$ and $\phi^{(i)}(\cdot)$ are the marginal states and marginal CCFs, respectively.

Before we state our formal convergence result, we first introduce some assumptions. We start by imposing the following assumptions on our stochastic process and moment functions:

Assumption **B.1** *The stochastic process $Y_t$ is a stationary Markov process.*

Assumption **B.2** *The moment functions $\mathbf{h}_t(\tau; \hat{v}_t, \theta)$ satisfy the following conditions:*

*(i)* $\mathbf{h}_t(\tau; v, \theta)$ *is continuously differentiable w.r.t.* $\theta$ *and* $v$*;*

*(ii)* $\mathbf{h}_t(\tau; v, \theta) \in L^2(\pi)^k, \forall \theta \in \Theta$ *and* $\forall v \in \mathbb{R}^m_+$*;*

*(iii) The equation* $\mathbb{E}^{\theta_0}[\mathbf{h}_t(\tau; v_t, \theta_0)] = \mathbf{0}$, $\forall \tau \in \mathbb{R}^{2\times 2m}$ *$\pi$-almost everywhere, has a unique solution* $\theta_0$ *in the interior of* $\Theta$*.*

For the next assumption, recall that the sample analogue of the moment conditions, given $T+1$ observations, is given by

$$\mathbf{h}_T(\tau; \hat{v}, \theta) := \frac{1}{T}\sum_{t=1}^{T} \mathbf{h}(\tau, Y_t^\theta, Y_{t+1}^\theta; \hat{v}_t, \theta).$$

Assumption **B.3** *The sample moment conditions satisfy, as* $T \to \infty$*:*

*(i)* $\sup_{\theta\in\Theta} \|\mathbf{h}_T(\cdot, v, \theta) - \mathbb{E}^{\theta_0}[\mathbf{h}_t(\cdot, v_t, \theta)]\| \xrightarrow{P} \mathbf{0}$*;*

*(ii)* $\sqrt{T}\mathbf{h}_T(\tau; v, \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{K})$ *on* $L^2(\pi)^k$*, where* $\mathcal{N}(0, \mathbf{K})$ *is the distribution of an $n$-dimensional Gaussian random element of* $L^2(\pi)^k$ *with mean zero and covariance operator* $\mathbf{K}$*, the Hilbert-Schmidt operator, defined by*

$$\mathbf{K} : L^2(\pi)^k \to L^2(\pi)^k, \qquad \mathbf{K}\mathbf{f}(\tau_1) := \int \mathbf{k}(\tau_1, \tau_2)\mathbf{f}(\tau_2)\pi(\tau_2)d\tau_2, \tag{B.1}$$

*with kernel* $\mathbf{k}(\tau_1, \tau_2) := \mathbb{E}^{\theta_0}\left[\mathbf{h}_t(\tau_1; v_t, \theta_0)\overline{\mathbf{h}_t(\tau_2; v_t, \theta_0)}\right]$.

Note that in the partial-information setting, the kernel $\mathbf{k}(\tau_1, \tau_2)$ is an $k \times k$ matrix function with $(i, j)^{\text{th}}$ element $\mathbb{E}^{\theta_0}\left[h_t^{(i)}(\tau_1; v_t, \theta_0)\overline{h_t^{(j)}(\tau_2; v_t, \theta_0)}\right]$.

Finally, we impose the following condition on the non-parametric spot volatility estimator:

Assumption **B.4** *Let the non-parametric volatility estimator* $\hat{v}_t$ *be defined from* $n$ *high-frequency returns prior to time* $t$*, and*

*(i)* $\hat{v}_t \xrightarrow{P} v_t$*, as* $n \to \infty$*;*

*(ii)* $n \to \infty$ *as* $T \to \infty$*, such that* $T/n \to 0$.

Assumption B.4.(ii) is required for the estimation error in $\hat{v}_t$ to be negligible in the large-$T$ asymptotic properties of the estimator.

Recall that the criterion function for the C-GMM estimator $\hat{\theta}$ is given by

$$Q_T(\hat{v}, \theta) = \|\mathbf{h}_T(\cdot, \hat{v}, \theta)\|^2 = \int \mathbf{h}_T(\tau, \hat{v}, \theta)' \overline{\mathbf{h}_T(\tau, \hat{v}, \theta)} \pi(\tau) d\tau.$$

We are now equipped to state the following proposition:

**Proposition 1** *Under Assumptions B.1–B.4, as* $T \to \infty$,

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}),$$

*where*

$$\mathbf{A} := \left\langle \mathbb{E}^{\theta_0}[\nabla_\theta \mathbf{h}_t(\cdot, v, \theta_0)], \mathbb{E}^{\theta_0}[\nabla_\theta \mathbf{h}_t(\cdot, v, \theta_0)] \right\rangle,$$
$$\mathbf{B} := \left\langle \mathbb{E}^{\theta_0}[\nabla_\theta \mathbf{h}_t(\cdot, v, \theta_0)], \mathbf{K}\mathbb{E}^{\theta_0}[\nabla_\theta \mathbf{h}_t(\cdot, v, \theta_0)] \right\rangle,$$

*with* $\mathbf{K}$ *as defined in* (B.1).

*Proof*: The consistency of the C-GMM procedure follows from Carrasco and Florens (2000) and Boswijk, Laeven, and Lalu (2015). Based on this, a mean value expansion of $\mathbf{h}_T(\tau, \hat{v}, \hat{\theta})$ yields

$$\mathbf{h}_T(\tau, \hat{v}, \hat{\theta}) = \mathbf{h}_T(\tau, v, \theta_0) + \nabla_\theta \mathbf{h}_T(\tau, \bar{v}, \bar{\theta})(\hat{\theta} - \theta_0) + \nabla_v \mathbf{h}_T(\tau, \bar{v}, \bar{\theta})(\hat{v} - v),$$

where $\bar{\theta}$ and $\bar{v}$ are mean values. Note that in our implied-state GMM setting we have to take into account both "direct" and "indirect" effects in the moment functions, i.e.,

$$\begin{aligned}
\nabla_\theta \mathbf{h}_T(\tau, v, \theta) &= \frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \mathbf{h}(\tau, Y_t^\theta, Y_{t+1}^\theta, v_t, \theta) \\
&= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \mathbf{h}(\tau, Y_t, Y_{t+1}, v_t, \theta)}{\partial \theta'} + \frac{\partial \mathbf{h}(\tau, Y_t, Y_{t+1}, v_t, \theta)}{\partial Y_t}\frac{\partial Y_t(\theta)}{\partial \theta'} \\
&\quad + \frac{\partial \mathbf{h}(\tau, Y_t, Y_{t+1}, \xi, \theta)}{\partial Y_{t+1}}\frac{\partial Y_{t+1}(\theta)}{\partial \theta'}, \\
\nabla_v \mathbf{h}_T(\tau, v, \theta) &= \frac{1}{T}\sum_{t=1}^{T} \nabla_v \mathbf{h}(\tau, Y_t^\theta, Y_{t+1}^\theta, v_t, \theta) \\
&= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \mathbf{h}(\tau, Y_t, Y_{t+1}, v_t, \theta)}{\partial v'} + \frac{\partial \mathbf{h}(\tau, Y_t, Y_{t+1}, v_t, \theta)}{\partial Y_t}\frac{\partial Y_t(v_t)}{\partial v'} \\
&\quad + \frac{\partial \mathbf{h}(\tau, Y_t, Y_{t+1}, v_t, \theta)}{\partial Y_{t+1}}\frac{\partial Y_{t+1}(v_t)}{\partial v'},
\end{aligned}$$

where the first elements on the right-hand sides of both equations capture only the direct dependence of the moment function on $\theta$ and $v$, while the remaining terms are due to the implied-state procedure.

Employing the mean value expansion in the first-order condition for optimality, we obtain

$$\begin{aligned}
0 &= \left\langle \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}), \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}) \right\rangle \\
&= \left\langle \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}), \mathbf{h}_T(\tau, v, \theta_0) + \nabla_\theta \mathbf{h}_T(\tau, \bar{v}, \bar{\theta})(\hat{\theta} - \theta_0) + \nabla_v \mathbf{h}_T(\tau, \bar{v}, \bar{\theta})(\hat{v} - v) \right\rangle,
\end{aligned}$$

so that

$$\begin{aligned}
\sqrt{T}(\hat{\theta} - \theta_0) &= -\left\langle \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}), \nabla_\theta \mathbf{h}_T(\tau, \bar{v}, \bar{\theta}) \right\rangle^{-1} \left\langle \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}), \sqrt{T}\mathbf{h}_T(\tau, v, \theta_0) \right\rangle \\
&\quad - \left\langle \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}), \nabla_\theta \mathbf{h}_T(\tau, \bar{v}, \bar{\theta}) \right\rangle^{-1} \left\langle \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta}), \nabla_v \mathbf{h}_T(\tau, \bar{v}, \bar{\theta}) \right\rangle \sqrt{T}(\hat{v} - v).
\end{aligned}$$

The second term on the right-hand side of the expression above vanishes asymptotically by Assumption B.4. For the first term, Assumption B.3 implies that

$$\left\langle \mathbb{E}^{\theta_0}[\nabla_\theta \mathbf{h}_t(\cdot, v_t, \theta_0)], \sqrt{T}\mathbf{h}_T(\tau, v, \theta_0) \right\rangle \xrightarrow{d} \mathcal{N}(0, \mathbf{B}).$$

Together with consistency and Slutsky's lemma, this yields the desired result. $\square$

We finally discuss the estimation of the standard errors. First, given the consistent estimators $\hat{\theta}$ and $\hat{v}$, we obtain a consistent estimator of the matrix $\mathbf{A}$:

$$\begin{aligned}\hat{\mathbf{A}}_T &= \left\langle \nabla_\theta \mathbf{h}_T(\cdot, \hat{v}, \hat{\theta}), \nabla_\theta \mathbf{h}_T(\cdot, \hat{v}, \hat{\theta}) \right\rangle \\ &= \int \nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta})' \overline{\nabla_\theta \mathbf{h}_T(\tau, \hat{v}, \hat{\theta})} \pi(\tau) d\tau \\ &= \sum_{i=1}^{k} \int \nabla_\theta h_T^{(i)}(\tau, \hat{v}, \hat{\theta}) \overline{\nabla_\theta h_T^{(i)}(\tau, \hat{v}, \hat{\theta})} \pi(\tau) d\tau.\end{aligned}$$

Next, let us denote the estimator of the covariance operator by

$$\mathbf{K}_T \mathbf{f}(\tau_1) = \int \mathbf{k}_T(\tau_1, \tau_2) \mathbf{f}(\tau_2) \pi(\tau_2) d\tau_2, \tag{B.2}$$

with kernel

$$\mathbf{k}_T(\tau_1, \tau_2) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{h}_t(\tau_1; \hat{v}_t, \hat{\theta}) \overline{\mathbf{h}_t(\tau_2; \hat{v}_t, \hat{\theta})}.$$

Then, asymptotic standard errors of our parameter estimates are obtained as the square root of the diagonal elements of

$$T^{-1} \hat{\mathbf{A}}_T^{-1} \hat{\mathbf{B}}_T \hat{\mathbf{A}}_T^{-1},$$

where

$$\begin{aligned}\hat{\mathbf{B}}_T &= \left\langle \nabla_\theta \mathbf{h}_T(\cdot, \hat{v}, \hat{\theta}), \mathbf{K}_T \nabla_\theta \mathbf{h}_T(\cdot, \hat{v}, \hat{\theta}) \right\rangle \\ &= \int \nabla_\theta \mathbf{h}_T(\tau_1, \hat{v}, \hat{\theta})' \overline{\mathbf{K}_T \nabla_\theta \mathbf{h}_T(\tau_1, \hat{v}, \hat{\theta})} \pi(\tau_1) d\tau_1 \\ &= \int \nabla_\theta \mathbf{h}_T(\tau_1, \hat{v}, \hat{\theta})' \overline{\int \mathbf{k}_T(\tau_1, \tau_2) \nabla_\theta \mathbf{h}_T(\tau_2, \hat{v}, \hat{\theta}) \pi(\tau_2) d\tau_2} \pi(\tau_1) d\tau_1 \\ &= \sum_{i=1}^{k} \sum_{j=1}^{k} \int \nabla_\theta h_T^{(i)}(\tau_1, \hat{v}, \hat{\theta}) \overline{\int k_T^{(ij)}(\tau_1, \tau_2) \nabla_\theta h_T^{(j)}(\tau_2, \hat{v}, \hat{\theta}) \pi(\tau_2) d\tau_2} \pi(\tau_1) d\tau_1.\end{aligned}$$

## Appendix C Jump-Robust Volatility Estimation

This appendix provides the details of the jump-robust volatility estimation procedure. We assume that for each day $t = 1, \ldots, T$, we observe $n+1$ intra-day equity prices at equidistant time points: $S_{t-1+j/n}$, $j = 0, \ldots, n$ (implying that the opening price of day $t$ equals the closing price of day $t-1$). Omitting the market-specific subscripts for notational convenience, we denote the intra-day log-returns by

$$\Delta_j^{t,n} S = \log(S_{t-1+j/n}) - \log(S_{t-1+(j-1)/n}).$$

We use the so-called threshold estimator for realized variance, originally proposed by Mancini (2001):

$$\hat{v}_t^2 := \sum_{j=1}^{n} \left(\Delta_j^{t,n} S\right)^2 \mathbb{1}\{|\Delta_j^{t,n} S| \leq r_n\}, \tag{C.1}$$

where $r_n$ is some deterministic sequence, converging to 0 as $n \to \infty$, used as a threshold to disentangle continuous variation from the jump contribution.

This threshold estimator has been shown to be consistent for the piece-wise constant variance $v_t^2$; its efficiency depends on the choice of the threshold $r_n$. Following Bollerslev and Todorov (2011), we consider an adaptive thresholding with $r_n = \alpha n^{-\bar{\omega}}$ and set $\bar{\omega} = 0.49$ and $\alpha = 3\sqrt{\frac{1}{5}\sum_{i=1}^{5} RV_{t-i}}$, where $RV_t$ is the realized variance estimator imposing no threshold. We base the parameter $\alpha$ on the average of the previous five days' estimates for better option pricing performance.[1]

The non-parametric jump-robust volatility estimator (C.1) allows us to forego a parametric representation of the volatility processes, and focus on the estimation of the jump parameters in our multivariate option pricing model. Hence, in the estimation procedure, described in Section 3, we consider a semi-nonparametrically approximated representation of the model with "frozen" spot volatilities. In our empirical analysis, we obtain the spot volatility estimates based on high-frequency data of the equity indices just prior to the observation time of the

[1]For the first day in the sample, we use $\alpha = 3\sqrt{\min(BV_t, RV_t)}$, where $BV_t$ is the bipower variation estimator proposed by Barndorff-Nielsen and Shephard (2004).

option panel.

# Appendix D Data Selection and Processing

This appendix provides details of the various data selection criteria and transformations applied to spot, futures and options data. First, we describe the full set of filters used to decide which option data observations were included in each reference interval. Next, we give additional details about the approach used to back out forward prices using the put-call parity. Finally, we discuss the interpolation of the Black-Scholes implied volatility surfaces.

## D.1 Option Data Selection

To select the set of options in a reference interval, we apply the following filter rule sequence:

(i) retain recordings with message type “Trade” or “Quote”;

(ii) retain recordings with a positive Transaction price or recordings with positive Bid and Ask prices;

(iii) for each distinct Reuters Instrument Code (RIC) symbol retain the last Bid, Ask and Transaction price in the reference interval;

(iv) select the Transaction price if available, otherwise calculate the mid Bid-Ask price.

The first two rules trivially filter out incomplete or erroneous recordings. The last two rules are similar to “last close” price series published by stock exchanges, which also typically prioritize trade data over submitted quotes.

To further reduce the presence of noise in the selected data (which can come from wide bid-ask spreads, or synchronicity mismatches between bid and ask quote timings), we consider a few additional filters. Complementing the aforementioned rule (iii), we have also determined for each distinct RIC the median Bid and median Ask recorded during the reference interval in order to calculate a “median spread” equal to the difference between median Ask and median Bid. We then employ the following additional filters:

(i) drop RIC symbols only if all of the following four conditions are met (concurrently):

(a) the number of either Bid or Ask quotes recorded in the interval is less than or equal to 2;

(b) there are no trade observations available in the interval;

(c) the elapsed time between the last Bid and Ask is larger than 10 seconds;

(d) the spread between last Bid and Ask is larger than 95% × median spread.

(ii) for each RIC symbol replace last Bid/Ask with the corresponding median Bid/Ask if all of the following three conditions are satisfied (concurrently):

(a) spread between last Bid and Ask is three times larger than the median spread;

(b) spread between last Bid and Ask is larger than 8 currency units;

(c) time difference between last Bid and Ask is larger than 5 seconds.

The first filter removes infrequently traded instruments which we deem likely to have illiquid quotes. The second filter aims to strike a balance between data synchronization and quote reliability.

## D.2 Implying Forward Prices from Put-Call Parity Pairings

To circumvent potential issues which would arise if we were to make explicit modeling choices for future dividend yields, we follow the route described in Aït-Sahalia and Lo (1998) and back out forward prices using the put-call parity relationship and estimate our model based on log-forward returns instead of log-index returns.

More specifically, to imply forward prices, we collect for each day all the put-call pairs with the same strike price and maturity, subject to an additional constraint that there are at least two Bid and two Ask quotes for each option during the reference interval. The additional constraint on the number of quotes filters out illiquid options and ensures we obtain reliable forwards. After implying forward prices from all the available put-call pairs, we take the average of the forward prices implied from pairs with the same option maturity and use the resulting term structure of forward prices to calculate Black-Scholes implied volatilities. For this last step, we require risk-free interest rates for each market. In principle, these could also be backed out from box spreads built from the option sets available in each interval, but this would have required an overly complicated option pairing algorithm. We therefore opted to use publicly available datasets with daily LIBOR-US, LIBOR-GBP and EURIBOR interest rate fixings. We have used linear interpolation for these fixings where needed to match the considered option's maturity.

We also need to interpolate the forward prices implied from put-call parity pairs of observed options for each maturity. We do that by exploiting a raw interpolation of discount factors, i.e., a linear interpolation between the log of discount factors yields that $\log D_\tau = \alpha \log D_{\tau_1} + (1-\alpha)\log D_{\tau_2}$, where $D_\tau = e^{(r-q)\tau}$ and $\alpha = \frac{\tau_2-\tau}{\tau_2-\tau_1}$. Therefore, an interpolated forward price for maturity $\tau = 40$ can be obtained as

$$F_t(\tau) = D_\tau S_t = (D_{\tau_1} S_t)^\alpha \left(D_{\tau_2} S_t\right)^{1-\alpha} = F_t(\tau_1)^\alpha F_t(\tau_2)^{1-\alpha}.$$

Given that E-Mini S&P 500 future options are American style options, we extract forward prices for these by matching put and call volatilities calculated using a binomial tree pricer which, up to a modest degree of residual pricing noise, can account for early exercise pricing premiums. We note that although our estimation procedure uses option pricing methods designed for European options, the inputs are Black-Scholes implied volatilities. Therefore, having implied volatilities from a binomial tree for American style E-Mini options, the estimation can make use of these volatilities.

## D.3 Volatility Surface Interpolation

This sub-section provides details of the standard interpolation technique we use to construct the implied volatility data panel, used as input in the estimation procedure. We first provide details of the filters employed to select the option price quotes from which implied volatilities are calculated. Next, we provide more information about the interpolation procedure and summary statistics for the resulting implied volatility surfaces.

Defining the moneyness level, $k$, as the strike-to-forward ratio, i.e., $k = K/F$, we designate an option as an out-of-the-money (OTM) option if it has moneyness level $k > 1.02$ for call options and $k < 0.98$ for put options. We consider options to be close to at-the-money (ATM) if $0.98 \leq k \leq 1.02$. We designate an option as in-the-money (ITM) if it is not OTM or close to ATM. We use call options to imply volatilities when $k > 1$, unless a particular call option has a spread which is more than twice as large as its put counterpart, or the put counterpart was quoted closer to the temporal reference point. A mirrored condition is applied for $k \leq 1$. These conditions trade off the liquidity of relevant options against the synchronicity of the data points used as inputs for building volatility smiles. When building implied volatility smiles, we make sure that for each volatility smile the call (put) prices (calculated for all options using put-call parity) are monotonically decreasing (increasing) functions of $k$.

The standard SVI parametrization of implied total variance, $w(x, \tau)$, with time-to-expiry $\tau$ is given as a function of log-moneyness $x = \log(k) = \log(K/F)$ and a parameter set $\chi = \{a, b, \rho, m, \sigma\}$:

$$w(x, \tau) = \sigma_{BS}^2(x)\tau = a + b\left(\rho(x - m) + \sqrt{(x - m)^2 + \sigma^2}\right), \tag{D.1}$$

where $a \in \mathbb{R}, b \geq 0, |\rho| < 1, m \in \mathbb{R}, \sigma > 0$ and $a + b\sigma\sqrt{1 - \rho^2} \geq 0$. Our application regards the stochastic volatility inspired (SVI) model as an interpolation method akin to a polynomial fit. In fact, when testing different approaches, we also considered a quadratic function to fit volatility smiles. However, the SVI parametrization most of the times displayed a better fit compared to the quadratic function. We do not treat SVI as an option pricing model per se in the sense that we do not calibrate it to all option data using a single set of parameter values. Instead we fit the functional form (D.1) independently for every reference interval and for every option maturity. This allows us to compromise between interpolating with fully flexible non-parametric approaches such as kernel smoothing and calibrating a parametric option pricing model.

To build the input for our estimation procedure, we calibrate the SVI model at every time point for two volatility slices using a quasi-explicit calibration approach as per De Marco and Martini (2009). For each day we choose two volatility slices such that times-to-maturity for the first slice $\tau_1 \leq \tau$ and for the second $\tau_2 > \tau$, and $\tau_1, \tau_2$ are the closest available maturities

Table 1: SVI interpolation RMSEs

| | FTSE 100 | | DAX 30 | | S&P 500 | |
|---|---|---|---|---|---|---|
| | $5 < \tau \leq 40$ | $40 < \tau \leq 75$ | $5 < \tau \leq 40$ | $40 < \tau \leq 75$ | $5 < \tau \leq 40$ | $40 < \tau \leq 75$ |
| $0.75 < k \leq 0.85$ | 0.68 | 0.37 | 0.81 | 0.35 | 0.56 | 0.30 |
| $0.85 < k \leq 0.92$ | 0.17 | 0.09 | 0.20 | 0.10 | 0.41 | 0.14 |
| $0.92 < k \leq 0.98$ | 0.13 | 0.07 | 0.20 | 0.09 | 0.29 | 0.17 |
| $0.98 < k \leq 1.03$ | 0.15 | 0.07 | 0.24 | 0.10 | 0.33 | 0.11 |
| $1.03 < k \leq 1.10$ | 0.22 | 0.11 | 0.34 | 0.14 | 0.44 | 0.16 |
| $1.10 < k \leq 1.20$ | 0.29 | 0.15 | 0.44 | 0.23 | 0.41 | 0.22 |
| Total | 0.19 | 0.12 | 0.37 | 0.18 | 0.40 | 0.19 |

This table reports the SVI interpolation RMSEs, reported as a percentage, for the filtered samples of options written on the FTSE 100, DAX 30 and S&P 500 indices. The sample consists of the daily options data covering the period 1 January 2006 to 13 August 2015. The data are interpolated for each market, each day, and each maturity slice separately.

to $\tau$. After having calibrated an SVI fit for these two volatility smiles, we interpolate between these slices linearly in total variance to $\tau$, which we set equal to 40 days.

Table 1 reports the RMSEs for implied volatility data based on SVI interpolations for each of the markets we consider and for different data buckets. The results show that the SVI interpolation generally has very small approximation errors, with RMSEs less than 0.5% for options with moneyness levels between 0.85 and 1.1.

The moneyness range we use for our standardized option panel at each time point is determined by the following interval rule:

$$\max\{\min\{k_1, k_2\} - 0.05, 0.85\} \leq k \leq \min\{\max\{k_1, k_2\} + 0.01, 1.1\}.$$

Although it would be better to have a fully homogeneous option panel with fixed moneyness range at every time point, there are days when the observed range is considerably narrower than it is on other days. Extrapolating these narrow ranges to obtain a wider fixed moneyness range would generate unreliable information. Therefore, we limit extrapolations to a maximum of up to 5% on the left wing (relative to the ATM point) and only 1% on the right wing of each implied volatility smile. For the estimation procedure we sample from the resulting interpolated volatility fit up to 13 option implied volatilities evenly spaced between 0.85 and 1.09 moneyness levels.

## References


Aït-Sahalia, Y., & Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, *53*(2), 499–547.

Backus, D. K., Foresi, S., & Telmer, C. I. (2001). Affine term structure models and the forward premium anomaly. *The Journal of Finance*, *56*(1), 279–304.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, *2*(1), 1–37.

Bollerslev, T., & Todorov, V. (2011). Estimation of jump tails. *Econometrica*, *79*(6), 1727–1783.

Boswijk, H. P., Laeven, R. J., & Lalu, A. (2015). Asset returns with self-exciting jumps: Option pricing and estimation with a continuum of moments. Working paper, University of Amsterdam and Tinbergen Institute.

Brandt, M. W., & Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, *63*(2), 161–210.

Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, *16*(6), 797–834.

De Marco, S., & Martini, C. (2009). Quasi-explicit calibration of Gatheral's SVI model. *Zeliade White Paper*, 1–15.

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, *64*, 19–47.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, *63*(1), 3–50.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations* (2nd ed.). Springer.