Internet Appendix for

"A New Model for the Joint Valuation of S&P 500 and VIX Options: Specification Analysis"

Appendix A. Affine models of VIX^2

Assume that the process of a k-dimensional state vector $\mathcal{V}_t$ satisfies the following stochastic differential equation under risk-neutral measure $\mathbb{Q}$

$$d\mathcal{V}_t \;=\; \mu(\mathcal{V}_t)dt + \Sigma(\mathcal{V}_t)dB_t + d\bar{J}_t, \tag{A.1}$$

where $\mu(\mathcal{V}_t) \in \mathbb{R}^k$ denotes the instantaneous drift function, B_t denotes a k-dimensional Brownian motion with $\Sigma(\mathcal{V}_t)\Sigma(\mathcal{V}_t)^\top \in \mathbb{R}^{k\times k}$ being the symmetric and positive definite instantaneous covariance matrix, and $d\bar{J}_t$ represents jump martingale increment. I further set $\mu(\mathcal{V}_t) = M + K\mathcal{V}_t$, $M \in \mathbb{R}^k$, $K \in \mathbb{R}^{k\times k}$. Therefore, the expected integrated state is an affine function of the state variables

$$\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T b_x^\top \mathcal{V}_s ds\right] \;=\; B_x(\tau)^\top \mathcal{V}_t + C_x(\tau), \tag{A.2}$$

where $\tau = T - t$, $b_x \in \mathbb{R}^k$, and

$$B_x(\tau) \;=\; \left(e^{K^T\tau} - I\right)\left(K^T\right)^{-1} b_x, \tag{A.3}$$

$$C_x(\tau) \;=\; \left(B(\tau)^T - b_x^T\tau\right)K^{-1}M. \tag{A.4}$$

Based on the above results, the VIX squared at time t can be written as an affine deterministic function of risk factors:

$$\begin{aligned}
\text{VIX}^2 \;&=\; \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T V_s ds + 2\int_t^T\int_{\mathbb{R}^2}\left(e^x - 1 - x\right)\mu(dx, ds)\right] \\
&=\; \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T b_v^\top \mathcal{V}_s ds + 2\int_t^T\int_{\mathbb{R}^2}\left(e^x - 1 - x\right)\mu(dx, ds)\right] \\
&=\; \frac{1}{\tau}\left(\mathcal{K}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T b_v^\top \mathcal{V}_s ds\right] + 2E^-\lambda_2^-\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T b_u^\top \mathcal{V}_s ds\right]\right) + \mathcal{U} \\
&=\; \frac{1}{\tau}\left(\mathcal{K}\left[B_v(\tau)^\top\mathcal{V}_t + C_v(\tau)\right] + 2E^-\lambda_2^-\left[B_u(\tau)^\top\mathcal{V}_t + C_u(\tau)\right]\right) + \mathcal{U} \\
&=\; A_{\text{VIX}^2} + B_{\text{VIX}^2}^\top \mathcal{V}_t,
\end{aligned} \tag{A.5}$$

with

$$A_{\text{VIX}^2} = \frac{1}{\tau}\left(\mathcal{K}C_v(\tau) + 2E^-\lambda_2^- C_u(\tau)\right) + \mathcal{U}, \tag{A.6}$$

$$B_{\text{VIX}^2} = \frac{1}{\tau}\left(\mathcal{K}B_v + 2E^-\lambda_2^- B_u\right), \tag{A.7}$$

where $\mathcal{K} = 1 + 2E^+\lambda_1^+ + 2E^-\lambda_1^-$, $\mathcal{U} = 2E^+\lambda_0^+ + 2E^-\lambda_0^-$, and $E^{+/-} = \left(\delta_1^{+/-}\right)^2 / \left(1 - \delta_1^{+/-}\right)$.

Appendix B. Characteristic functions

Assume that the characteristic functions of $\log(S_T)$ and VIX^2 take the exponential affine form

$$\begin{aligned}\phi_{s_T}(t, \mathcal{V}_t; u) &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{iu\log(S_T)}\right] \\ &= e^{-\alpha(\tau)-\beta(\tau)\log(S_t)-\gamma(\tau)V_t-\psi(\tau)m_t-\xi(\tau)Z_t-\varpi(\tau)U_t}, \qquad \text{(B.1)} \\ \phi_{\text{VIX}_T^2}(t, \mathcal{V}_t; u) &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{iuVIX_T^2}\right] \\ &= e^{-\alpha_{\text{VIX}^2}(\tau)-\gamma_{\text{VIX}^2}(\tau)V_t-\psi_{\text{VIX}^2}(\tau)m_t-\xi_{\text{VIX}^2}(\tau)Z_t-\varpi_{\text{VIX}^2}(\tau)U_t}, \qquad \text{(B.2)}\end{aligned}$$

where $u \in \mathbb{R}$ and i denotes imaginary number. According to the results from *Duffieet al.* (2000), the coefficients in the definition of $\phi_{\text{VIX}_T^2}$ satisfy the following ODEs, with $\tau = T - t$

$$\begin{aligned}\dot{\alpha}_{\text{VIX}^2}(\tau) &= \kappa_m\theta_m\psi(\tau) + \kappa_z\theta_z\xi(\tau) + \kappa_u\theta_u\varpi(\tau) \\ &\quad - \lambda_0^-\left(\frac{1}{\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right) - \xi_0\left(\frac{1}{(1+\delta_2\gamma(\tau))} - 1\right), \\ \dot{\gamma}_{\text{VIX}^2}(\tau) &= -\kappa_v\gamma_{\text{VIX}^2}(\tau) - \frac{1}{2}\sigma_v^2\gamma_{\text{VIX}^2}^2(\tau) - \lambda_1^-\left(\frac{1}{\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right) \\ &\quad - \xi_1\left(\frac{1}{(1+\delta_2\gamma(\tau))} - 1\right), \\ \dot{\psi}_{\text{VIX}^2}(\tau) &= \kappa_v\gamma_{\text{VIX}^2}(\tau) - \kappa_m\psi_{\text{VIX}^2}(\tau) - \frac{1}{2}\sigma_m^2\psi_{\text{VIX}^2}^2(\tau), \\ \dot{\xi}_{\text{VIX}^2}(\tau) &= -\kappa_z\xi_{\text{VIX}^2}(\tau) - \frac{1}{2}\varsigma^2\gamma_{\text{VIX}^2}^2(\tau) - \sigma_z\varsigma\rho_2\gamma_{\text{VIX}^2}(\tau)\xi_{\text{VIX}^2}(\tau) - \frac{1}{2}\sigma_z^2\xi_{\text{VIX}^2}^2(\tau) \\ &\quad - \xi_2\left(\frac{1}{(1+\delta_2\gamma(\tau))} - 1\right), \\ \dot{\varpi}_{\text{VIX}^2}(\tau) &= -\kappa_u\varpi_{\text{VIX}^2}(\tau) - \frac{1}{2}\sigma_u^2\varpi_{\text{VIX}^2}^2(\tau) - \lambda_2^-\left(\frac{1}{\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right),\end{aligned}$$

with boundary conditions $\alpha_{\text{VIX}^2}(0) = iuA_{\text{VIX}^2}$ and $[\gamma_{\text{VIX}^2}(0), \psi_{\text{VIX}^2}(0), \xi_{\text{VIX}^2}(0), \varpi_{\text{VIX}^2}(0)]^\top = iuB_{\text{VIX}^2}$, where the coefficients A_{VIX^2} and B_{VIX^2} are defined in Appendix A.

The coefficients of ϕ_{s_T} satisfy the following ODEs for $t \in (0, T]$

$$
\begin{aligned}
\dot{\alpha}(\tau) &= \left(r_t - \left(\frac{1}{1-\delta_1^+} - 1\right)\lambda_0^+ - \left(\frac{1}{1-\delta_1^-} - 1\right)\lambda_0^-\right)\beta(\tau) \\
&\quad + \kappa_m\theta_m\psi(\tau) + \kappa_z\theta_z\xi(\tau) + \kappa_u\theta_u\varpi(\tau) - \xi_0\left(\frac{1}{(1+\delta_2\gamma(\tau))} - 1\right) \\
&\quad - \lambda_0^-\left(\frac{1}{\left(1+\delta_1^-\beta(\tau)\right)\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right) - \lambda_0^+\left(\frac{1}{\left(1+\delta_1^+\beta(\tau)\right)} - 1\right) \\
\dot{\beta}(\tau) &= \left(-\frac{1}{2} - \left(\frac{1}{1-\delta_1^+} - 1\right)\lambda_1^+ - \left(\frac{1}{1-\delta_1^-} - 1\right)\lambda_1^-\right)\beta(\tau) \\
&\quad - \kappa_v\gamma(\tau) - \frac{1}{2}\beta^2(\tau) - \sigma_v\rho_1\beta(\tau)\gamma(\tau) - \frac{1}{2}\sigma_v^2\gamma^2(\tau) - \xi_1\left(\frac{1}{(1+\delta_2\gamma(\tau))} - 1\right) \\
&\quad - \lambda_1^+\left(\frac{1}{\left(1+\delta_1^+\beta(\tau)\right)} - 1\right) - \lambda_1^-\left(\frac{1}{\left(1+\delta_1^-\beta(\tau)\right)\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right) \\
\dot{\psi}(\tau) &= \kappa_v\gamma(\tau) - \kappa_m\psi(\tau) - \frac{1}{2}\sigma_m^2\psi^2(\tau) \\
\dot{\xi}(\tau) &= -\left(\frac{1}{1-\delta_1^-} - 1\right)\lambda_3^-\beta(\tau) - \kappa_z\xi(\tau) - \frac{1}{2}\varsigma^2\gamma^2(\tau) - \sigma_z\varsigma\rho_2\gamma(\tau)\xi(\tau) - \frac{1}{2}\sigma_z^2\xi^2(\tau) \\
&\quad - \xi_2\left(\frac{1}{(1+\delta_2\gamma(\tau))} - 1\right) - \lambda_3^-\left(\frac{1}{\left(1+\delta_1^-\beta(\tau)\right)\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right) \\
\dot{\varpi}(\tau) &= -\left(\frac{1}{1-\delta_1^-} - 1\right)\lambda_2^-\beta(\tau) - \kappa_u\varpi(\tau) - \frac{1}{2}\sigma_u^2\varpi^2(\tau) \\
&\quad - \lambda_2^-\left(\frac{1}{\left(1+\delta_1^-\beta(\tau)\right)\left(1-\eta_1\delta_1^-\gamma(\tau)\right)\left(1-\eta_2\delta_1^-\xi(\tau)\right)} - 1\right),
\end{aligned}
$$

with boundary conditions $\alpha(0) = 0$, $\beta(0) = -iu$, $\gamma(0) = 0$, $\psi(0) = 0$, $\xi(0) = 0$, and $\varpi(0) = 0$. The ODEs can be solved numerically using Runge-Kutta methods. To speed up the computation, I utilize the Graphics Processing Unit (GPU) with the tools provided by NVIDIA CUDA®.

Appendix C. Derivatives pricing

C.1. VIX derivatives pricing

To price futures and options on the VIX, following Branger et al. (2016), I apply the results in Lewis (2000) and Chen and Joslin (2012). Lewis (2000) offers a powerful valuation approach

that is applicable to a wide range of European options. His method requires that the fundamental transform, a generalization of the characteristic function that allows complex arguments, be available. It also requires the Fourier transform of the options payoffs. The forward and inverse Fourier transforms are

$$\begin{aligned}\hat{f}(z) &= \int_{-\infty}^{\infty} e^{izx} f(x) dx \\ f(x) &= \frac{1}{2\pi} \int_{iz_i-\infty}^{iz_i+\infty} e^{-izx} \hat{f}(z) dz,\end{aligned}$$

with transform variable $z = z_r + z_i i$, where z_r and z_i denote the real and imaginary parts of z.

The payoff of a call option on VIX_T is

$$f_1(\text{VIX}_T^2) = \left(\sqrt{\text{VIX}_T^2} - K\right)^+, \tag{C.1}$$

and its forward transform is

$$\hat{f}_1(z) = \frac{\sqrt{\pi}\left(1 - \text{erf}\left(K\sqrt{-iz}\right)\right)}{2\left(-iz\right)^{3/2}}, \tag{C.2}$$

where erf denotes the error function of a complex valued argument. A futures contract on VIX_T has the payoff function

$$f_2\left(\text{VIX}_T^2\right) = \sqrt{\text{VIX}_T^2}, \tag{C.3}$$

and its forward transform is

$$\hat{f}_2(z) = \frac{\sqrt{\pi}}{2\left(-iz\right)^{3/2}}. \tag{C.4}$$

For both payoff functions the Fourier transform is well-behaved if $z_i > 0$. Finally, I get the price of

a call option on VIX_T

$$\begin{aligned} C_{\text{VIX}}(t,T,K) &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds}\left(\sqrt{\text{VIX}_T^2}-K\right)^+\right] \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds} f_1(\text{VIX}_T^2)\right] \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r\tau}\frac{1}{2\pi}\int_{iz_i-\infty}^{iz_i+\infty} e^{-iz\text{VIX}_T^2}\hat{f}_1(z)dz\right] \\ &= \frac{e^{-r\tau}}{2\pi}\int_{iz_i-\infty}^{iz_i+\infty}\mathbb{E}_t^{\mathbb{Q}}\left[e^{-iz\text{VIX}_T^2}\right]\hat{f}_1(z)dz \\ &= \frac{e^{-r\tau}}{\sqrt{\pi}}\int_0^{\infty}\mathcal{R}\left[\psi_{\text{VIX}_T^2}\left(t,\mathcal{V}_t;-\left(z_r+iz_i\right)\right)\frac{\left(1-\text{erf}\left(K\sqrt{z_i-iz_r}\right)\right)}{2\left(z_i-iz_r\right)^{3/2}}\right]dz_r. \end{aligned} \tag{C.5}$$

Similarly, I get the price of a forward contract on VIX_T

$$\begin{aligned} F_{\text{VIX}}(t,T) &= \mathbb{E}_t^{\mathbb{Q}}\left[\sqrt{\text{VIX}_T^2}\right] \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[f_2\left(\text{VIX}_T^2\right)\right] \\ &= \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{2\pi}\int_{iz_i-\infty}^{iz_i+\infty} e^{-iz\text{VIX}_T^2}\hat{f}_2(z)dz\right] \\ &= \frac{1}{2\pi}\int_{iz_i-\infty}^{iz_i+\infty}\mathbb{E}_t^{\mathbb{Q}}\left[e^{-iz\text{VIX}_T^2}\right]\hat{f}_2(z)dz \\ &= \frac{1}{\sqrt{\pi}}\int_0^{\infty}\mathcal{R}\left[\phi_{\text{VIX}_T^2}\left(t,\mathcal{V}_t;-\left(z_r+iz_i\right)\right)\frac{1}{2\left(z_i-iz_r\right)^{3/2}}\right]dz_r, \end{aligned} \tag{C.6}$$

C.2. S&P 500 options pricing

Although fast Fourier transform (FFT) (e.g., Carr and Madan (1999)) is the most popular method in the literature, I prefer to use the direction integration (DI) approach to take advantage of GPU. Specifically, the call option can be calculated by

$$C_{\text{SPX}}(t,T,K) = \frac{\exp(-\alpha\log(K)-r(T-t))}{\pi}\int_0^{\infty}\exp(-iu\log(K))I(u,\alpha)du, \tag{C.7}$$

with an integrand function $I(u,\alpha)$

$$I(u,\alpha) = \frac{\phi_{s_T}(t,\mathcal{V}_t;u-(\alpha+1)i)}{\alpha^2+\alpha-u^2+(2\alpha+1)iu}. \tag{C.8}$$

In this paper, I set the smoothing parameter, α as 1.2, truncate the upper bound for the integration domain by 300, and use 128 integrand points, resulting in good results.

One advantage of DI over FFT is that I do not need to interpolate option prices over the strikes, leading to a more accurate result with dramatically reduced time. The most important is that with FFT, I have to employ a large number of strikes (e.g., $2^{12} = 4096$), which leads to a broad strike band. However, in most cases, I only need 100 strikes to do the calibration, which means that only 100 of 4096 calculated option prices are really used. Therefore, with DI, I can use the strike vector method to accelerate computation as I only need to calculate characteristic function once for multiple option prices with different strikes.

To further accelerate the computation, I utilize parallel programming on GPU. The main effort is partitioning the computational requirement into thousands of small computations that can be executed simultaneously. These computations are assigned to thousands of threads of the GPU which are executed concurrently on different cores. The GPU hardware consists of a number of streaming multiprocessors which in turn consist of multiple cores. Threads are organized in blocks, where one or more block runs on a streaming multiprocessor (see, e.g., Cook (2012)).

Appendix D. Estimation Procedure

This section details the estimation procedure. The estimation procedure applied in this paper follows Huang and Wu (2004), Christoffersen et al. (2009), Andersen et al. (2015), and others. One primary reason for utilizing this approach is that the estimation procedure can be largely paralleled, significantly reducing the computing time. The procedure mainly involves two parts: 1) global optimization over structural parameters and 2) local optimization over state variables.

I denote the structural parameters of the 4F-ICJ model by Θ and state vector at time t by $\mathcal{V}_t = (V_t, m_t, Z_t, U_t)$. Further, the model-implied Black-Scholes IV is given by $IV_X(t, T, K, \mathcal{V}_t, \Theta)$, where X represents the underlying asset and K is the strike price. For simplicity, thereafter, I use IV_X to represent $IV_X(t, T, K, \mathcal{V}_t, \Theta)$. The estimation proceeds by jointly optimized over parameters and state vector realizations. Note that the parameters are kept constant and only the state variables are solved for period-by-period. Figure D.1 outlines the overall estimation procedure.

Structural parameters optimization. To estimate the structural parameters, I construct an

estimator which takes form

$$\hat{\Theta} \;=\; \arg\min \frac{1}{2}\left(\text{error}_{\text{SPX}} + \text{error}_{\text{VIX}}\right),$$

where error_X is defined as

$$\text{error}_X \;=\; \frac{1}{N_X}\sum_{n,t}^{N_X}\left(\frac{IV_X(t,n) - IV_X^M(t,n)}{IV_X^M(t,n)}\right)^2, \qquad X \in \{\text{SPX}, \text{VIX}\},$$

where $N_X = \sum_{t=1}^{T} N_X(t)$. $N_X(t)$ denotes the total number of option contracts at time t. IV_X^M represents the market-implied volatilities.

Sates optimization. For a given set of structural parameters, Θ, I trivially obtain the corresponding state vector $\mathcal{V}_t$ by solving T equal-weighted pricing error optimization problems of the form

$$\mathcal{V}_t \;=\; \arg\min \frac{1}{2}\left(\text{error}_{\text{SPX}}(t) + \text{error}_{\text{VIX}}(t)\right), \qquad t \;=\; 1, 2, \ldots, T,$$

where $\text{error}_X(t)$ is defined as

$$\text{error}_X(t) \;=\; \frac{1}{N_X(t)}\sum_{n=1}^{N_X(t)}\left(\frac{IV_X(t,n) - IV_X^M(t,n)}{IV_X^M(t,n)}\right)^2, \qquad X \in \{\text{SPX}, \text{VIX}\}.$$

Notably, the local optimization problems are independent, and thus they can be solved in parallel. Furthermore, the local optimization problem can be solved quickly as it only involves four state variables in each period. In addition, the highly numerical-demanding part of options pricing is offloaded to the Graphics Processing Unit (GPU), supported by the powerful parallel computing tool – CUDA, as discussed in Appendix C. Therefore, the model can be estimated very fast, with an overall computing time around 20 hours on a high-performance computing cluster consisting of 6 nodes with 28 CPU cores and 4 GPUs each.

Following Bardgett et al. (2019), this paper adopts a widely used global optimizer, namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996). The CMA-ES is an evolutionary (stochastic or randomized) algorithm for difficult non-linear non-convex black-box optimization problems in the continuous domain. The algorithm

has been adopted as one of the standard tools for continuous optimization research labs and industrial environments worldwide. It is typically applied to unconstrained or bonded constraint optimization problems and search space dimensions between three and a hundred. Furthermore, it has been shown to be a particularly reliable and highly competitive evolutionary algorithm for global optimization (e.g., Hansen and Kern (2004)). Therefore, the CMA-ES is particularly suitable for the option pricing model estimation and can be extremely useful in coping with the potential existence of multiple local minima.

In the following, I briefly introduce the algorithm of CMA-ES. Please refer to Hansen and Ostermeier (1996) and Hansen and Kern (2004) for a thorough discussion. CMA-ES samples solutions from a multivariate gaussian distribution. After evaluating all solutions, the solutions are sorted by evaluation values. Then the algorithm updates the distribution parameters (e.g., the mean vector and the covariance matrix) based on the ranking of evaluation values. The procedure is outlined in algorithm 1. The critical feature of CMA-ES is that it solves randomized black-box optimization as it does not use approximate gradients, which contrasts quasi-Newton methods. Suppose we consider the black-box search scenario, where we want to minimize an objective function:

$$\begin{aligned} f \quad &: \quad \mathbb{R}^n \to \mathbb{R} \\ &\qquad \mathbf{x} \mapsto f(\mathbf{x}). \end{aligned}$$

The objective is to find search points, $\mathbf{x} \in \mathbb{R}^n$, with a function value, $f(\mathbf{x})$, as small as possible. Black-box optimization refers to the situation, where function values of evaluated search points are the only accessible information on f. The search points to be evaluated can be freely chosen. A randomized black-box search algorithm mainly consists of three parts: 1) sampling of new solutions, 2) re-ordering of the sampled solutions based on their fitness, and 3) update of the internal state variables based on the re-ordered samples.

For the local optimization, I follow Andersen et al. (2015) and use the NLopt optimizer to solve for state variables each period. That is, given a set of structural parameters generated from a global optimizer, I obtain the corresponding state vector by solving T local optimization problems. In particular, I utilize the Nelder-Mead simplex algorithm as described in Nelder and Mead (1965) and Lagarias, Reeds, Wright, and Wright (1998). The algorithm is simple and has

demonstrated enduring popularity. Furthermore, it is a direct search method often applied to non-linear optimization problems.[1]

Finally, after solving for state vectors, the estimation method goes back to global optimization and updates a new set of structural parameters based on which the local optimizer generates a series of state variables. The technique progresses until no further significant decreases in the overall objective are obtained.

Algorithm 1 CMA-ES: Covariance Matrix Adaptation Evolution Strategy

Require: α_μ, α_σ, α_{cp}, α_{c1}, $\alpha_{c\lambda}$ ▷ Learning rates

Require: d_σ ▷ Damping factor

Require: $t = 0$ ▷ Generation counter

Require: $\mu^{(0)} \in \mathbb{R}^n$, $\sigma \in \mathbb{E}_+$ ▷ Inputs: initial mean vectors and step size

Require: $C^{(0)} = I$, $p_\sigma^{(0)} = 0$, $p_c^{(0)} = 0$ ▷ Initialize covariance matrix and evolution paths.

repeat

Sample $x_i^{(t+1)} = \mu^{(t)} + \sigma^{(t)} y_i^{(t+1)}$ where $y_i^{(t)} \sim \mathcal{N}\left(0, C^{(t)}\right)$, $i = 1, \ldots, \Lambda$

Select top λ samples with the best performance $x_i^{(t+1)}$, $i = 1, \ldots, \lambda$

$\mu^{(t+1)} \leftarrow \mu^{(t)} + \alpha_\mu \frac{1}{\lambda} \sum_{i=1}^{\lambda} \left(x_i^{(t+1)} - \mu^{(t)}\right)$

$p_\sigma^{(t+1} \leftarrow (1-\alpha_\sigma) p_\sigma^{(t)} + \sqrt{\alpha_\sigma(2-\alpha_\sigma)\lambda} C^{(t)^{-\frac{1}{2}}} \frac{\mu^{(t+1)} - \mu^{(t)}}{\sigma^{(t)}}$

$\sigma^{(t+1)} \leftarrow \sigma^{(t)} \exp\left(\frac{\alpha_\sigma}{d_\sigma}\left(\frac{||p_\sigma^{(t+1}||}{\mathbb{E}||\mathcal{N}(0,I)||} - 1\right)\right)$

$p_c^{(t+1)} \leftarrow (1-\alpha_{cp}) p_c^{(t)} + \sqrt{\alpha_{cp}(2-\alpha_{cp})\lambda} \frac{\mu^{(t+1)} - \mu^{(t)}}{\sigma^{(t)}}$

$C^{(t+1)} \leftarrow (1-\alpha_{c1}-\alpha_{c\lambda}) C^{(t)} + \alpha_{c1} p_c^{(t+1)} p_c^{(t+1)^\top} + \alpha_{c\lambda} \frac{1}{\lambda} \sum_{i=1}^{\lambda} y_i^{(t+1)} y_i^{(t+1)^\top}$

$t \leftarrow t + 1$

until hit stopping criteria

return $\mu^{(t)}$, $\sigma^{(t)}$, $C^{(t)}$

Symbol	Meaning	Symbol	Meaning
$x_i^{(t)} \in \mathbb{R}^n$	the i-th samples at the generation (t)	$\mu^{(t)}$	mean of the generation (t)
$y_i^{(t)} \in \mathbb{R}^n$	$x_i^{(t)} = \mu^{(t-1)} + \sigma^{(t-1)} y_i^{(t)}$	$\sigma^{(t)}$	step size
$C^{(t)}$	Covariance matrix	d_α	damping factor for $\sigma' s$ update
$p_\sigma^{(t)}$	evaluation path for σ at the generation (t)	α_μ	learning rate for $\mu' s$ update
$p_c^{(t)}$	evaluation path for C at the generation (t)	α_σ	learning rate for p_σ
α_{c1}	learning rate for $C' s$ rank-1 update	α_{cp}	learning rate for p_c
$\alpha_{c\lambda}$	learning rate for $C' s$ rank-min(λ, n) update		

[1]Specifically, the algorithm uses a simplex of n+1 points for n-dimensional vectors $\mathbf{x}$. The algorithm first makes a simplex around the initial guess $\mathbf{x}_0$ by adding 5% of each component $\mathbf{x}_0(i)$ to $\mathbf{x}_0$. The algorithm uses these n vectors as elements of the simplex in addition to $\mathbf{x}_0$. Then the algorithm modifies the simplex repeatedly according to the specific procedure discussed in Lagarias et al. (1998).

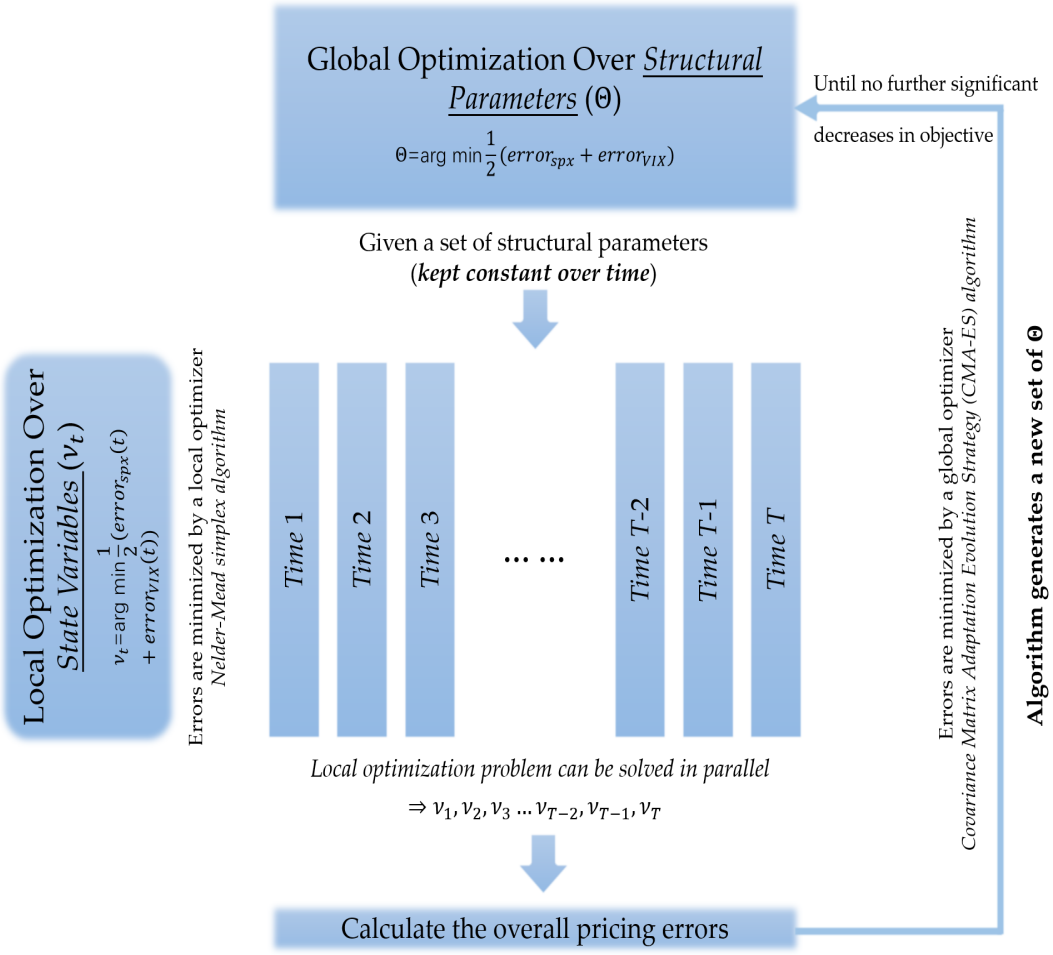

Fig. D.1. Estimation procedure.

Appendix E. Structures in pricing errors

Another way to investigate the performance of different model specifications is to check for remaining structures in the pricing errors of these models. I define pricing errors as the difference between the model-implied IV and the market-observed IV. If a model is specified reasonably well, we should find minimal structures in the pricing errors on the S&P 500 and VIX options. I check for remaining structures in the mean pricing error at each moneyness and maturity. The mean pricing error of a good model should be close to zero and show no apparent structures along both the moneyness and the maturity dimensions.

Since an option's days to maturity and moneyness change every day, I estimate the pricing

error at fixed moneyness and maturity by using nonparametric smoothing. A positive pricing error indicates model overpricing compared to market data, and a negative pricing error indicates model underpricing. The following studies are based on the full sample.

E.1. S&P 500 index options analysis

Figure E.1 plots the smoothed pricing errors of S&P 500 options at different moneyness and maturities under each of the six model specifications. Within each panel, the five lines represent pricing errors for five maturities: 30 (solid), 60 (dashed), 90 (dotted), 150 (dot-dashed), and 300 days (dash-dash-dotted).

[Insert Figure E.1 near here]

Panels A and B show that the two-factor models (2F and 2F-CJ) exhibit large mean pricing errors along with both the maturity and the moneyness dimensions. At short maturities, two-factor models overprice OTM puts relative to OTM calls. At long maturities, the pattern is reversed. OTM puts are underpriced relative to OTM calls. Compared with the 2F model, the 2F-CJ model gradually reduces pricing errors to zero as we move from deep OTM put to ATM options. The 2F model, however, consistently overprices options with K/F smaller than 1.

Consistent with Andersen et al. (2015), I find that including U can substantially improve the model fit on OTM puts, as shown in Panel C. The 3FU-CJ model shows much better performance, except at very long maturities (300 days), where it still significantly underprices OTM puts. The IV skew is a direct result of conditional non-normality in asset returns. The downward slope of the skew reflects asymmetry (negative skewness) in the risk-neutral distribution. The positive curvature of the skew reflects the fat-tails (leptokurtosis) of this distribution. The central limit theorem, moreover, implies that under very general conditions, the conditional return distribution should converge to normality as the maturity increases. As a result, the IV skew should flatten out when the maturity increases. The market data, however, shows that the IV skew steepens slightly as maturity increases. The biases shown in Panels A to C therefore imply that the IV skew of the corresponding model flattens out faster than observed in the data.

Panel D shows that including Z can increase the curvature of IV smile or kurtosis of risk-neutral distribution. Consistent with Table 6, the 3FZ-ICJ model performs well for ATM options.

Furthermore, as shown in Table 4, Z is a persistent factor, which can help to slow down the convergence to normality. Panel E confirms the previous finding that co-jumps is a critical feature of pricing models. The 4F-IJ model consistently underprices OTM puts for medium- to long-term options.

Importantly, Panel F shows that the 4F-ICJ model looks promising in generating a persistent IV skew across the maturity horizon. The mean pricing errors of the 4F-ICJ model are close to zero and show no apparent structures along both the moneyness and the maturity dimensions.

E.2. VIX options analysis

Figure E.2 plots the smoothed pricing errors of VIX options. Within each panel, the four lines represent pricing errors for five maturities: 30 (solid), 60 (dashed), 90 (dotted), and 120 days (dot-dashed).

[Insert Figure E.2 near here]

For comparison, I use the same scale for all panels except for Panel A, where I use a larger scale to accommodate the larger pricing errors from the 2F model. Panel A shows that the 2F model has difficulty capturing the positive IV skew of VIX options across all maturities. The model consistently overprices (underprices) short-term (long-term) options across the moneyness dimension. Volatility jumps are thus a necessary feature for pricing VIX options. The 2F-CJ model, as shown in Panel B, exhibits much better performance. I nonetheless still observe prominent error structures in Panel B. The 2F-CJ model consistently underprices OTM VIX calls across all maturities. The 2F-CJ model is thus still struggling to capture positive IV skew patterns of VIX options.

Panel C illustrates that the positive IV skew can be effectively captured by incorporating U. As a result, the 3FU-CJ model substantially improves the model fit of VIX options. The 3FU-CJ model, however, generates steeper IV skew than observed in the data. The 3FU-CJ model overprices OTM call options relative to OTM put options. Panel D shows that the 3FZ-ICJ model can largely improve the performance of the 2F-CJ model by including factor Z. Unlike the 3FU-CJ model, the 3FZ-ICJ model generates a relatively flatter IV skew. Furthermore, the 3FZ-ICJ model shows a much better fit with long-term options.

Panel E shows the results of the 4F-IJ model, which incorporates both Z and U but excludes co-jumps in V. As a result, the jump intensity of volatility jumps is driven by V and Z. Factor U, however, has a major impact on return jumps, which can substantially increase the flexibility of the model for pricing S&P 500 options in the joint valuation framework. The limited model's flexibility for pricing S&P 500 options can also deteriorate the performance on VIX options. For example, the model structures or factor levels are forced to capture the features of S&P 500 options at the expense of the fitness of VIX options, and the reverse is also true. Like the 3FZ-ICJ model, the 4F-IJ model also relatively underprices OTM calls. U is therefore crucial for capturing a high skew pattern.

Panel F shows the results of the 4F-ICJ model. As expected, the 4F-ICJ model produces the smallest pricing errors along both the moneyness dimension and the maturity dimension. Compared with the 4F-IJ model, the success of the 4F-ICJ model stems from incorporating co-jumps in V. Consequently, U can directly affect the volatility jumps and thus the risk-neutral distribution of variance. Not surprisingly, including co-jumps can largely improve the model fit on OTM calls across all maturities. As shown in Panel F, the right ends of all of the curves are very close to zero. Combining all the evidence, I conclude that the 4F-ICJ model significantly outperforms all other nested models in jointly pricing S&P 500 and VIX options.

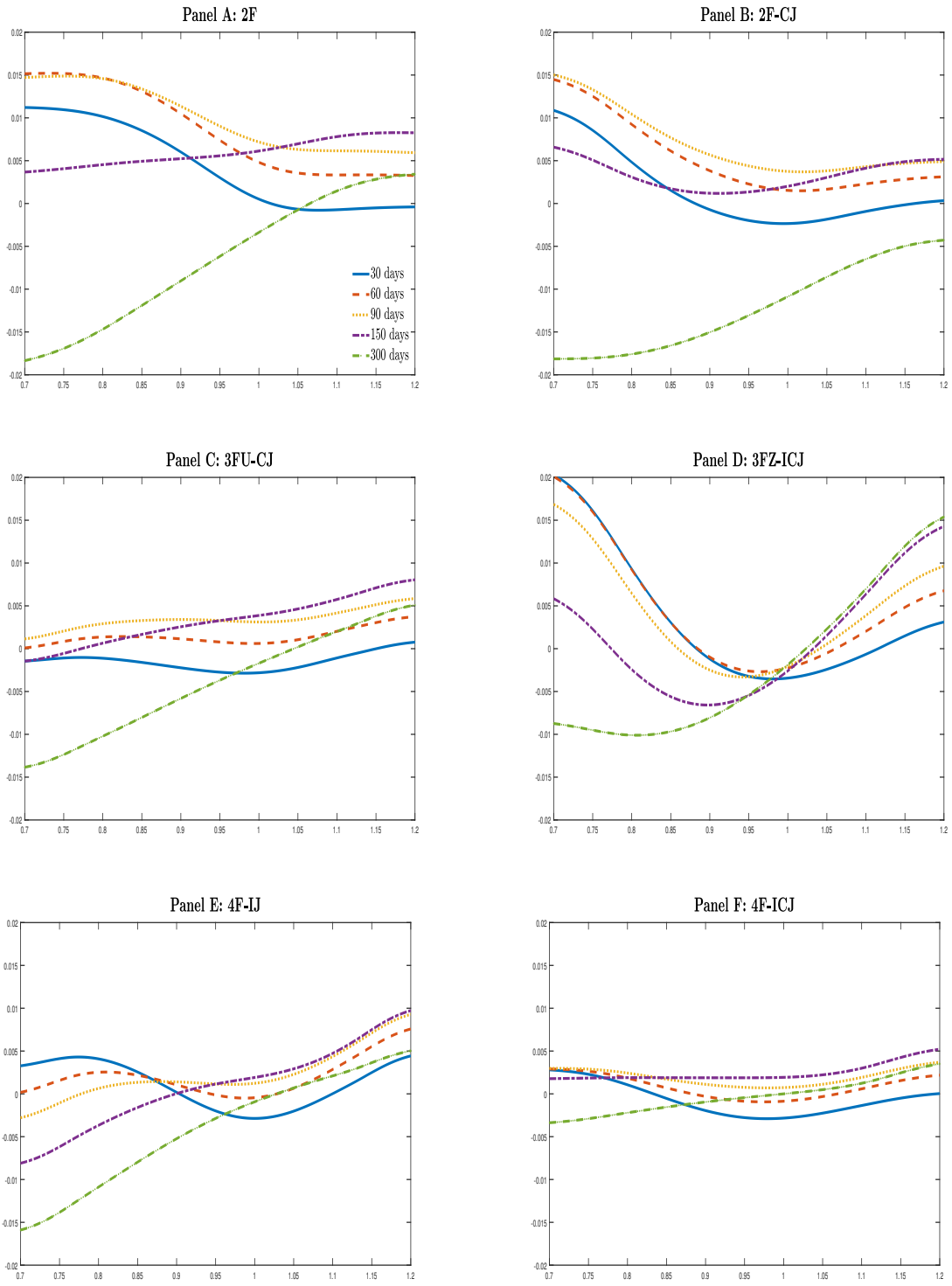

Fig. E.1. This figure shows mean pricing errors of S&P 500 options. The pricing errors are defined as the difference between the model-implied implied volatility (IV) and the market-observed IV. Using independent Gaussian kernels, I estimate the mean pricing errors at fixed moneyness and maturities. Each panel denotes one model. The different lines in each panel denote different maturities, as shown in Panel A.

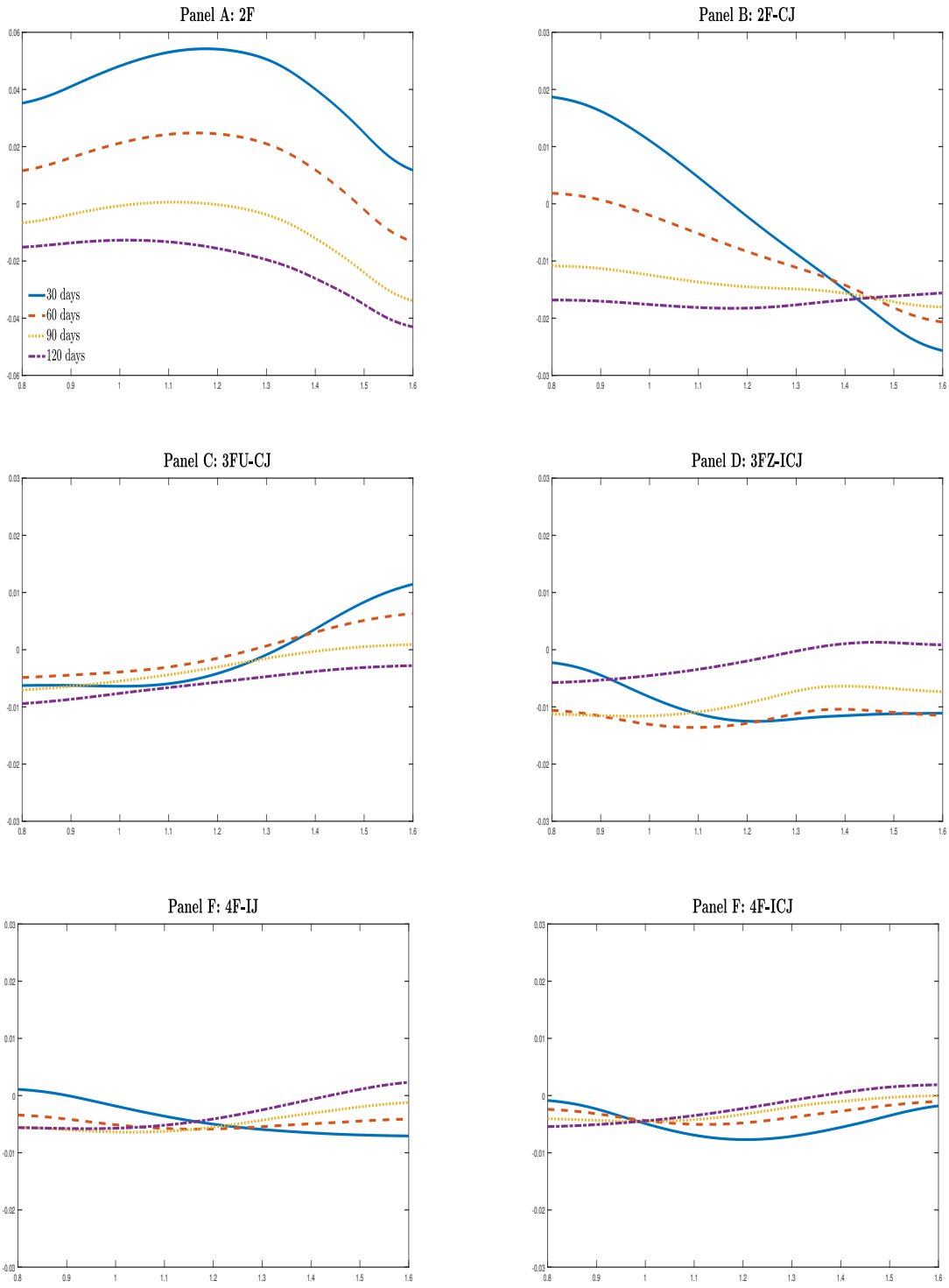

Fig. E.2. This figure shows smoothed mean pricing errors of VIX options. The pricing errors are defined as the difference between the model-implied implied volatility (IV) and the market-observed IV. Using independent Gaussian kernels, I estimate the mean pricing errors at fixed moneyness and maturities. Each panel denotes one model. The different lines in each panel denote different maturities, as shown in Panel A.

Appendix F. Alternative Train-Test Sample Splits

In this section, I estimate the models using alternative train-test sample splits as a robustness check. Specifically, I switch the training and testing sets used in the main text. I estimate the models with the sample from April 2, 2015, to December 27, 2017, and then test the model performance using the sample from April 1, 2007, to April 1, 2015. Notably, the major crises occurred during the testing sample, highly challenging the model performance.

The parameters reported in Table F.1 are very similar to those in Table 4. Therefore, the model structures are pretty stable over the entire sample period. Table F.2 reports the pricing errors for the models considered in this paper. Panel A shows the in-sample results. Importantly, consistent with findings in Table 5, the 4F-ICJ model can largely reduce the pricing errors by 25.95% relative to the benchmark 2F-CJ. Furthermore, 4F-ICJ shows superior performance over all other alternative models, highlighting the importance of the model features of 4F-ICJ in the joint pricing framework. In addition, the three-factor model with either U or Z can also reduce errors by about 17.6%. Lastly, co-jumps are crucial in the four-factor models as 4F-IJ model, which excludes co-jumps, performs much worse than 4F-ICJ. The findings remain similar for out-of-sample results in Panel B. Specifically, 4F-ICJ shows dominant pricing power and reduces errors by 27.10% relative to the benchmark. Surprisingly, the overall pricing error of 4F-ICJ slightly increases to 0.0608 out-of-sample from 0.0582 in-sample. Given the major crises in the out-of-sample period, the evidence strongly supports the superior performance of 4F-ICJ.

Table F.3 compares models by using AIC/BIC criteria. Consistent with the results in Table F.2, 4F-ICJ continues to be the most successful model for jointly pricing S&P 500 and VIX options after accounting for degrees of freedom. Overall, the results from alternative sample split confirm the findings in the main analysis.

Table F.1: In-sample model parameter estimates. This table reports the model parameter estimates and their standard errors (in parentheses). Models are estimated using S&P 500 and VIX options data sampled every Wednesday over the period April 4, 2007, to April 1, 2015.

	2F		2F-CJ		3FU-CJ		3FZ-ICJ		4F-IJ		4F-ICJ	
κ_v	12.4769	(1.7234)	15.8875	(3.7816)	9.6135	(0.7433)	19.5251	(0.9386)	5.5074	(0.6343)	6.6748	(0.4673)
σ_v	1.6595	(0.5325)	1.1955	(0.2318)	0.6692	(0.0934)	1.0695	(0.0305)	0.5979	(0.0480)	0.4710	(0.0321)
ρ_1	-0.9385	(0.0835)	-0.9217	(0.0523)	-0.9774	(0.0883)	-0.9699	(0.0323)	-0.9228	(0.0865)	-0.9636	(0.0613)
η_1			4.1720	(0.7359)	4.0274	(0.4321)	4.6722	(0.4285)			2.0558	(0.2104)
ς							0.0223	(0.0055)	0.7111	(0.0912)	0.3345	(0.0213)
κ_m	1.3903	(0.0339)	1.2439	(0.1231)	1.0598	(0.0231)	0.9906	(0.0474)	2.4606	(0.0432)	1.5686	(0.0622)
θ_m	0.0254	(0.0022)	0.0143	(0.0031)	0.0079	(0.0008)	0.0111	(0.0023)	0.0213	(0.0021)	0.0230	(0.0041)
σ_m	0.0938	(0.0112)	0.1154	(0.0214)	0.1056	(0.0123)	0.1192	(0.0232)	0.1224	(0.0214)	0.1128	(0.0177)
κ_z							1.5913	(0.0362)	0.7226	(0.0343)	0.0253	(0.0057)
σ_z							4.0784	(0.0121)	6.0319	(0.0133)	2.9813	(0.2313)
ρ_2							0.9773	(0.0621)	0.9272	(0.0850)	0.8860	(0.1322)
η_2							3.5470	(0.0723)			0.3929	(0.0219)
κ_u					0.3120	(0.0432)			0.7394	(0.0520)	0.5677	(0.0978)
σ_u					0.3781	(0.0212)			0.2175	(0.0432)	0.4102	(0.0768)
δ_1^-	-0.1059	(0.0113)	-0.1625	(0.0166)	-0.0842	(0.0075)	-0.1639	(0.0424)	-0.1162	(0.0332)	-0.0860	(0.0056)
δ_1^+	0.0128	(0.0013)	0.0117	(0.0021)	0.0138	(0.0022)	0.0145	(0.0035)	0.0168	(0.0023)	0.0139	(0.0052)
δ_2							0.1274	(0.0357)	0.3251	(0.0533)	0.1151	(0.0312)
λ_0^+	0.1023	(0.3212)			0.1002	(0.0458)	0.1123	(0.0482)				
λ_1^+	1.7488	(0.0129)	1.2489	(0.2565)	0.8450	(0.1233)	1.6883	(0.0412)	0.1402	(0.0264)	0.1834	(0.0413)
λ_2^+	0.3322	(0.1033)	0.4144	(0.1721)								
λ_0^-							0.0422	(0.0173)				
λ_1^-	4.8548	(0.5432)	4.9690	(0.7321)	3.1860	(0.4121)	4.9910	(0.3561)	1.6072	(0.3787)	2.9209	(0.3447)
λ_2^-	1.0612	(0.4214)	1.2331	(0.4671)								
λ_3^-					0.2103	(0.0722)			1.3211	(0.0372)	1.2339	(0.0451)
ξ_0									0.1332	(0.0412)		
ξ_1									0.3612	(0.1322)	0.2612	(0.0422)
ξ_2							3.0412	(0.4128)	0.1020	(0.0182)	0.0910	(0.0213)

Table F.2: The pricing performance. This table reports the pricing errors, defined in Equations (18) and (19), for in-sample and out-of-sample. I use the 2F-CJ model as the benchmark. The error reductions relative to the benchmark (2F-CJ) are in italics below each model. The estimation period is from April 2, 2015 to December 27, 2017, and the out-of-sample period is from April 4, 2007 to April 1, 2015.

	2F	2F-CJ	3FU-CJ	3FZ-ICJ	4F-IJ	4F-ICJ
Panel A: In-sample RMSREs						
Overall	0.1002	0.0786	0.0641	0.0648	0.0647	0.0582
	27.48%		*-18.45%*	*-17.56%*	*-17.68%*	*-25.95%*
SPX	0.0864	0.0731	0.0639	0.0634	0.0649	0.0625
	18.19%		*-12.59%*	*-13.32%*	*-11.22%*	*-14.50%*
VIX	0.1140	0.0841	0.0642	0.0663	0.0645	0.0538
	35.55%		*-23.66%*	*-21.13%*	*-23.31%*	*-36.03%*
Panel B: Out-of-sample RMSREs						
Overall	0.1102	0.0834	0.0687	0.0721	0.0719	0.0608
	32.13%		*-17.63%*	*-13.55%*	*-13.79%*	*-27.10%*
SPX	0.1039	0.0822	0.0677	0.0727	0.0700	0.0668
	26.40%		*-17.64%*	*-11.56%*	*-14.84%*	*-18.73%*
VIX	0.1164	0.0846	0.0698	0.0715	0.0739	0.0547
	37.59%		*-17.49%*	*-15.48%*	*-12.65%*	*-35.34%*

Table F.3: Comparison of information criterion values. This table compares Akaike or Bayesian information criterion for each candidate model. The estimation period is from April 2, 2015 to December 27, 2017, and the out-of-sample period is from April 4, 2007 to April 1, 2015. The criteria values are divided by the number of periods in the corresponding sample.

	2F	2F-CJ	3FU-CJ	3FZ-ICJ	4F-IJ	4F-ICJ
Panel A: In-sample						
AIC	-1202.38	-1357.32	-1483.83	-1475.96	-1475.54	-1540.03
BIC	-1182.56	-1337.48	-1454.18	-1446.21	-1436.01	-1500.50
Panel B: Out-of-sample						
AIC	-3641.28	-4190.13	-4567.14	-4473.88	-4475.30	-4798.37
BIC	-3620.67	-4169.45	-4536.47	-4442.94	-4434.43	-4757.50

Appendix G. Risk Premium

The traditional approach for studying market prices of risk (MRP) is by estimating the model under both physical ($\mathbb{P}$) and risk-neutral ($\mathbb{Q}$) measures. Since this paper only examines the model under $\mathbb{Q}$ measure, some critical parameters are hard to be (accurately) identified (e.g., jump sizes under $\mathbb{P}$ measure). Therefore, estimating risk premium or option returns would be more accurate by calibrating the model under both physical and risk-neutral measures. I leave this analysis in future work. If we assume the latent states are well identified, MRP can be partially inferred.[2]

I calculate the mean-reversion and the mean of the process to infer the MRP (in particular, the MRP of diffusion parts). Based on the assumption that the market prices of diffusive risks are proportional to the corresponding risk level, the coefficient of proportionality, Λ, can be estimated by comparing the mean-reversion under different measures. For example, the coefficient of proportionality for risk m is $\Lambda_m = \kappa_m - \kappa_m^{\mathbb{P}}$. And the same applies to other risk factors. I first get the autocorrelation of each risk by regressing the states at time t on their values at time t-1. The resulting coefficients for states V, m, Z, U are 0.8742, 0.9623, 0.9334, and 0.9882, respectively.

Furthermore, if we assume the jumps sizes are the same under both measures, then we get the estimated mean-reversion of risk factors under P are: $\kappa_v^{\mathbb{P}} = 6.9033$, $\kappa_m^{\mathbb{P}} = 2.4041$, $\kappa_z^{\mathbb{P}} = 3.3566$, and $\kappa_u^{\mathbb{P}} = 0.5947$. As such, the inferred coefficient of proportionality of risk factors are $\Lambda_v = -1.0257$, $\Lambda_m = -0.2162$, $\Lambda_z = -3.3441$, and $\Lambda_u = -0.2770$. In addition, the average equity risk premium is estimated at 0.0583. Therefore, in line with the literature, the diffusive part of risk factors is found to be negative, which supports the existence of a negative variance and variance-of-variance risk premium. To put the values in context, with different model specifications, Eraker (2004) estimates Λ_v as a range from -0.009 to -0.017; Li and Zinna (2018) estimate Λ_v and Λ_m at -1.7486 and -0.3238, respectively; Bardgett et al. (2019) estimate Λ_v and Λ_m at -1.80 and -0.11, respectively; Huang, Schlag, Shaliastovich, and Thimme (2019) estimate Λ_v and Λ_z at -4.8 and -5, respectively; and Branger et al. (2016) estimate Λ_z with a range from -1.1420 to -6.1919. Furthermore, Bardgett et al. (2019) set Λ_u at 0 as the trajectory of the factor U is difficult to estimate. Therefore, the estimates in this paper are close to the corresponding values reported in the literature.

To further investigate the model-implied average rates of return for SPX and VIX options, I

[2] I thank an anonymous referee for suggesting this study.

conduct Monte Carlo simulations based on the estimated parameters in Table 4 and the above estimated diffusive risk premium. Provided the model specifications under both $\mathbb{P}$ and $\mathbb{Q}$ measures, the expected returns for holding the options to expiration are given by

$$\begin{aligned} \mathbb{E}_t\left[R_{X,put}\right] &= \frac{\mathbb{E}_t^{\mathbb{P}}\left[\max\left(K_X - X_T, 0\right)\right]}{e^{-\int_t^T r_t dt}\mathbb{E}_t^{\mathbb{Q}}\left[\max\left(K_X - X_T, 0\right)\right]} \\ \mathbb{E}_t\left[R_{X,call}\right] &= \frac{\mathbb{E}_t^{\mathbb{P}}\left[\max\left(X_T - K_X, 0\right)\right]}{e^{-\int_t^T r_t dt}\mathbb{E}_t^{\mathbb{Q}}\left[\max\left(X_T - K_X, 0\right)\right]} \end{aligned} \tag{G.1}$$

where T denotes the expiration date and K_X is the strike price for asset X (SPX or VIX futures). Therefore, the expected option return is determined by the gap between the $\mathbb{P}$ and $\mathbb{Q}$ probability measures as a result of market prices of different sources of risk. Following Eraker and Yang (2021), I simulate monthly (buy-and-hold) returns on SPX puts and VIX calls with different maturities and strikes. SPX puts and VIX calls are essential tools for hedging downside risk in the market. Consistent with Eraker and Yang (2021), I find that both SPX puts and VIX calls yield negative returns.

Figure G.1 shows average returns on holding SPX put and VIX call options to maturities. Notably, the simulated return patterns are very similar to those observed from market. First, the model can generate a negative premium for all SPX puts and VIX calls. For example, the model implies negative returns around -10%, -20%, and -45% for ITM ($k = 1.05$), ATM ($k = 1$), and OTM ($k = 0.8$) 1-month SPX puts. In addition, the model yields negative returns around -12%, -23%, and -40% for ITM ($k = 0.8$), ATM ($k = 1$), and OTM ($k = 1.4$) 1-month VIX calls. Second, short-term SPX puts and VIX calls yield a much more negative premium. Therefore, short-term SPX puts and VIX calls are much more expensive than long-term ones. Third, SPX puts (VIX calls) are increasing (decreasing) with moneyness. As such, out-of-the-money SPX puts and VIX calls ask for a much higher price for holding them. Finally, the simulated returns are relatively higher than market-implied returns. One possible reason is that the simulation under-estimates the magnitude of jump premium as I assume the jump sizes are the same under both measures. The model could yield a more negative return for SPX puts and VIX calls if the jump risk premium is also negative (e.g., jumps sizes are smaller under the physical measure).

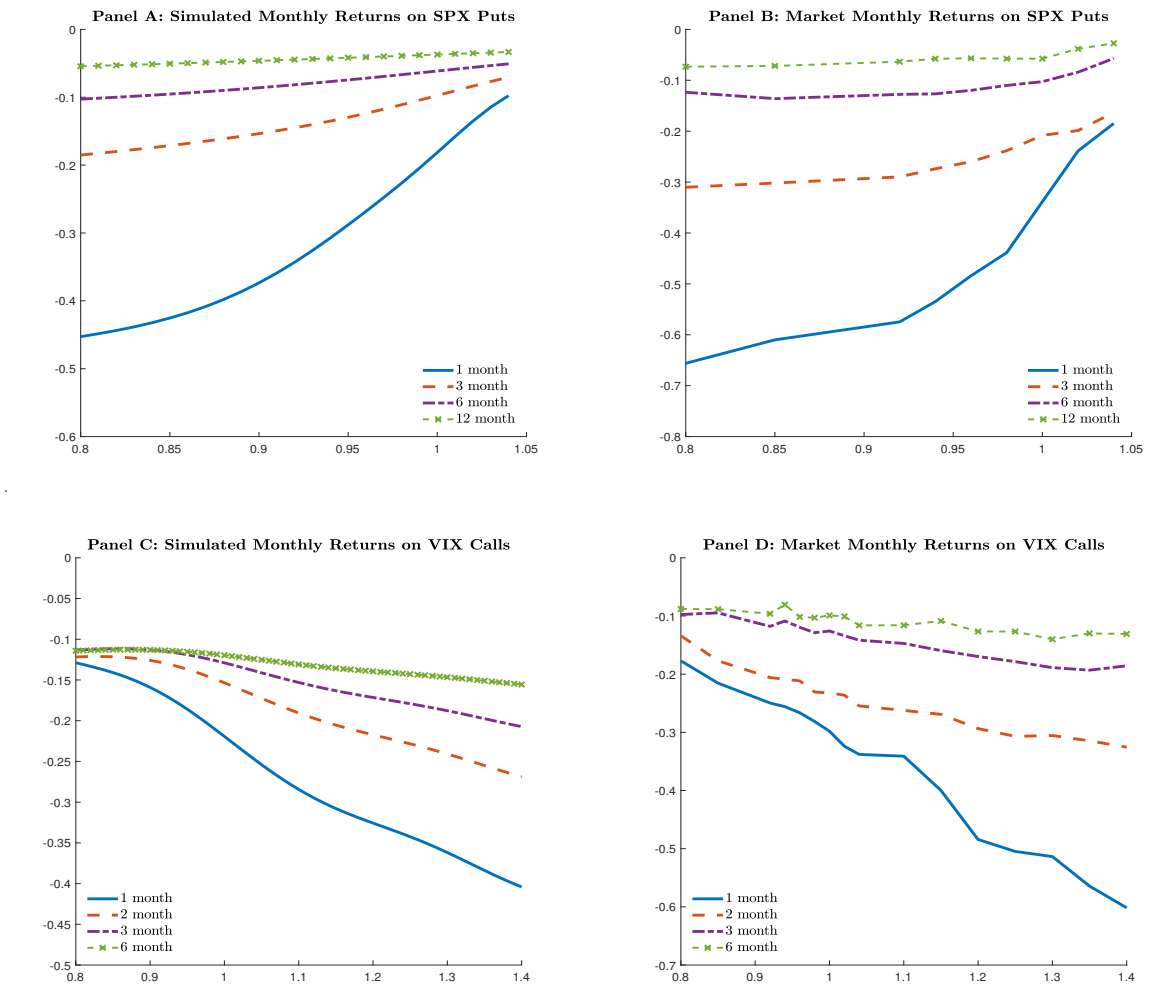

Fig. G.1. This figure compares market (right) and model-implied (left) monthly returns on SPX puts and VIX calls options. For SPX options, we consider options with four maturities of 1, 3, 6, and 12 month. For VIX options, we consider options with four maturities of 1, 2, 3, and 6 month. The x-axis denotes the relative strike price, k, defined as the strike price divided by underlying asset price.