Internet Appendix for

Leverage Effect, Volatility Feedback, and Self-Exciting Market Disruptions

Peter Carr and Liuren Wu*

I. Model Summary

We fix a filtered probability space $\{\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}\}$ and assume no-arbitrage in the economy. Under certain technical conditions, there exists a risk-neutral probability measure $\mathbb{Q}$, absolutely continuous with respect to $\mathbb{P}$, such that the gains process associated with any admissible trading strategy deflated by the risk-free rate is a martingale.

A. Separating Leverage Effect from Volatility Feedback and Self-Exciting Jumps

Let F_t denote the time-t forward level of the equity index over some fixed time horizon. We separate the dynamics of the risky asset portfolio from the variation of the market's financial leverage via the following multiplicative decomposition,

$$F_t = X_t A_t, \tag{A-1}$$

where A_t denotes the time-t forward value of the risky asset and $X_t = F_t/A_t$ denotes the equity-to-risky asset ratio.

We model the equity-to-risky asset ratio X_t as a constant elasticity of variance (CEV) process

*Carr, Tandon School of Engineering, New York University; email: petercarr@nyu.edu. Wu, Baruch College, Zicklin School of Business, One Bernard Baruch Way, Box B10-225, New York, NY 10010; tel: +1-646-312-3509; fax: +1-646-312-3451; email: liuren.wu@baruch.cuny.edu.

under the risk-neutral measure $\mathbb{Q}$,

$$dX_t/X_t = \delta X^{-p} dW_t, \quad p > 0, \tag{A-2}$$

where W_t denotes a standard Brownian motion, and we model the value dynamics for the risky asset under the risk-neutral measure $\mathbb{Q}$ as,

$$dA_t/A_{t-} = \sqrt{v_t^Z} dZ_t + \int_{-\infty}^{0} (e^x - 1)\left(\mu(dx,dt) - \pi(x)dx v_t^J dt\right), \tag{A-3}$$

$$dv_t^Z = \kappa_Z\left(\theta_Z - v_t^Z\right)dt + \sigma_Z\sqrt{v_t^Z} dZ_t^v, \quad \mathbb{E}\left[dZ_t^v dZ_t\right] = \rho dt < 0, \tag{A-4}$$

$$dv_t^J = \kappa_J\left(\theta_J - v_t^J\right)dt - \sigma_J \int_{-\infty}^{0} x\left(\mu(dx,dt) - \pi(x)dx v_t^J dt\right), \tag{A-5}$$

where Z_t and Z_t^v denote two standard Brownian motions, $\mu(dx,dt)$ denotes a counting measure for jumps, $\pi(x)v_t^J$ denotes the time-t arrival rate of jumps of size x in log asset value $\ln A_t$, with

$$\pi(x) = e^{-|x|/v_J}|x|^{-1}, \tag{A-6}$$

and A_{t-} denotes the asset value at time t just prior to a jump. We assume independence between the Brownian innovation (dW_t) in financial leverage and the Brownian innovations in the asset value (dZ_t) and the asset return volatility (dZ_t^v). The innovation independence assumption allows us to model A_t and X_t as separate martingales.

B. A Reduced-Form Benchmark

In theory, a reduced-form model can accommodate any number of factors. In practice, without the discipline of an economic structure, such models often experience identification issues. As a result, most empirical studies in the literature limit the specifications to one stochastic volatility factor,

with only a few studies exploring the estimation of two volatility factors.[1] By imposing economic structures, our model not only can answer structural economic questions that cannot be addressed by a reduced-form specification, but it is also very parsimonious and highly identifiable, even with three volatility factors.

By design, the three-factor variance structure shall perform better than any one- or two-factor specifications in matching the observed option behavior. To examine the contribution of the financial leverage factor and the statistical significance of the superior performance, we create a two-factor reduced-form benchmark by setting $X_t = 1$ and hence $S_t = A_t$ to our full model. We can regard the benchmark as a restricted version of our model without the financial leverage effect. We use the term "reduced-form" to highlight the fact that the benchmark, as in most extant models in the option pricing literature, does not separately model asset value dynamics and financial leverage, and hence does not differentiate the leverage effect from the volatility feedback effect. By comparing the empirical performance of our full model with the reduced-form benchmark, we gauge the benefits of allowing distinct channels of interactions between equity returns and volatilities.

II. Option Valuation Via Iterated Expectations

Under our specified model dynamics, we can price European options efficiently via iterated expectations. Formally, let $c(F_t, K, T)$ denote the time-t forward value of a European call option on the equity index with strike price K and expiry date T, conditional on the time-t values of the index forward F_t and the three state variables (X_t, v_t^Z, v_t^J). Since the equity index F_t is driven by two orthogonal sources of variations A_t and X_t under the risk-neutral measure, we can perform the valuation through

[1] Empirical studies on two-factor variance structures for equity index options include, among others, Bates (2000), Huang and Wu (2004), Christoffersen, Jacobs, Ornthanalai, and Wang (2008), Christoffersen, Heston, and Jacobs (2009), Egloff, Leippold, and Wu (2010), and Santa-Clara and Yan (2010). A paper by Andersen, Fusari, and Todorov (2015) explores the identification of a three-factor variance structure from S&P 500 index options.

the law of iterated expectations,

$$
\begin{aligned}
c(F_t,K,T) &\equiv \mathbb{E}\left[(F_T-K)^+\middle|\left(F_t,X_t,v_t^Z,v_t^J\right)\right] \\
&= \mathbb{E}\left[(X_T A_T-K)^+\middle|\left(A_t,X_t,v_t^Z,v_t^J\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(\mathcal{X}A_T-K)^+\middle|(X_T=\mathcal{X},A_t,v_t^Z,v_t^J)\right]\middle|X_t\right] \\
&= \mathbb{E}\left[\mathcal{X}\cdot\mathcal{C}(A_t,K/\mathcal{X},T)|X_t\right], \qquad \text{(A-7)}
\end{aligned}
$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator under the risk-neutral measure and the function $\mathcal{C}(A_t,\mathcal{K},T)$ is defined as

$$
\mathcal{C}(A_t,\mathcal{K},T) \equiv \mathbb{E}\left[(A_T-\mathcal{K})^+\right], \qquad \text{(A-8)}
$$

which can be regarded as the forward value of a call option on the risky asset with strike $\mathcal{K}$ and expiry T. Equation (A-7) turns the calculation of the call value on the equity index into the computation of the call value on the risky asset and a numerical integration over all possible equity-to-risky asset ratios. Under the reduced-form benchmark, the call value on the risky asset will simply become the call value on the equity index, without the need for the extra layer of numerical integration.

A. Fourier Transforms and FFT Valuation of Options on Asset

To compute the forward call value on asset $\mathcal{C}(A_t,\mathcal{K},T)$, we first derive the generalized Fourier transform of the log asset return $\ln A_T/A_t$,

$$
\phi(u) \equiv \mathbb{E}_t\left[e^{iu\ln A_T/A_t}\right], \quad u\in\mathcal{D}\subseteq\mathbb{C}, \qquad \text{(A-9)}
$$

where $\mathcal{D}$ denotes a subsect of the complex plane under which the expectation in equation (A-9) is well defined. Once we obtain this transform $\phi(u)$, we can compute the option value $\mathcal{C}$ via fast Fourier transform (FFT) following the procedure proposed by Carr and Madan (1999).

Our specification for the dynamics on the log asset value can be represented as a time-changed Lévy process with affine activity rates (Carr and Wu (2004)). The generalized Fourier transform is exponential-affine in the state variables,

$$\text{(A-10)} \qquad \phi(u) = \exp\left(-a_Z(\tau) - b_Z(\tau)v_t^Z - a_J(\tau) - b_J(\tau)v_t^J\right), \quad \tau = T - t,$$

where the affine coefficients solve the following ordinary differential equations,

$$\text{(A-11)} \qquad \begin{aligned} b_Z'(\tau) &= \psi_Z(u) - \kappa_Z^{\mathbb{M}} b_Z(\tau) - \tfrac{1}{2}\sigma_Z^2 b_Z(\tau)^2, & a_Z'(\tau) &= b_Z(\tau)\kappa_Z\theta_Z, \\ b_J'(\tau) &= \psi_J(u) - (\kappa_J + \sigma_J v_J)\, b_J(\tau) + \ln\left(1 + \sigma_J b_J(\tau) v_J^{\mathbb{M}}\right), & a_J'(\tau) &= b_J(\tau)\kappa_J\theta_J, \end{aligned}$$

starting at $a_Z(0) = b_Z(0) = a_J(0) = b_J(0) = 0$, and with

$$\begin{aligned} \psi_Z(u) &= \tfrac{1}{2}(iu + u^2), \\ \psi_J(u) &= \ln(1 + iuv_J) - iu\ln(1 + v_J), \\ \kappa_Z^{\mathbb{M}} &= \kappa_Z - iu\rho\sigma_Z, \\ v_J^{\mathbb{M}} &= v_J/(1 + iuv_J). \end{aligned}$$

The ordinary differential equations governing the coefficients $(a_Z(\tau), b_Z(\tau))$ can be solved analytically,

$$\text{(A-12)} \qquad \begin{aligned} b_Z(t) &= \frac{2\psi_Z(u)\left(1 - e^{-\xi\tau}\right)}{2\xi - \left(\xi - \kappa_Z^{\mathbb{M}}\right)\left(1 - e^{-\xi\tau}\right)}, \quad \xi = \sqrt{\left(\kappa_Z^{\mathbb{M}}\right)^2 + 2\sigma_Z^2\psi_Z(u)}, \\ a_Z(t) &= \frac{\kappa_Z\theta_Z}{\sigma_Z^2}\left[2\ln\left(1 - \frac{\xi - \kappa_Z^{\mathbb{M}}}{2\xi}\left(1 - e^{-\xi\tau}\right)\right) + (\xi - \kappa_Z^{\mathbb{M}})\tau\right]. \end{aligned}$$

The ordinary differential equations governing the coefficients $(a_J(\tau), b_J(\tau))$ can be solved numerically using the standard Runge-Kutta 4th-order method.

With the generalized Fourier transform $\phi(u)$ on the risky asset return, we first re-scale the forward call value on the risky asset $c(k) = \mathcal{C}(A_t, \mathcal{K}, T)/A_t$ to represent the forward call value in percentages of the forward asset value as a function of moneyness defined as the log strike over forward $k \equiv \ln \mathcal{K}/A_t$. Then, we derive the Fourier transform on the re-scaled forward call $c(k)$ in terms of the Fourier transform on the risky asset return,

$$\chi(u) \equiv \int_{-\infty}^{\infty} e^{iuk} c(k) dk = \frac{\phi(u-i)}{(iu)(iu+1)}, \tag{A-13}$$

which is well-defined when u contains an imaginary component $u = u_r - i\alpha$, with u_r being real and α being a real positive number. With the transform in (A-13), the call value can be computed via the following inversion,

$$c(k) = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} e^{-iu_r k} \chi(u_r - i\alpha) du_r. \tag{A-14}$$

We perform the inversion numerically by discretizing the integral using the trapezoid rule:

$$c(k) \approx \frac{e^{-\alpha k}}{\pi} \sum_{m=0}^{N} \delta_m e^{-iu_m k} \chi(u_m - i\alpha) \Delta u, \tag{A-15}$$

where $\delta_m = \frac{1}{2}$ when $m = 0$ and 1 otherwise. We cast the operation in (A-15) in the form of discrete fast Fourier transform (FFT), which is an efficient algorithm for computing discrete Fourier coefficients. The discrete Fourier transform is a mapping of $\mathbf{f} = (f_0, \cdots, f_{N-1})^\top$ on the vector of Fourier

coefficients $\mathbf{d} = (d_0, \cdots, d_{N-1})^\top$, such that

$$d_j = \frac{1}{N} \sum_{m=0}^{N-1} f_m e^{-jm\frac{2\pi}{N}i}, \quad j = 0, 1, \cdots, N-1. \tag{A-16}$$

FFT allows the efficient calculation of $\mathbf{d}$ if N is an even number, say $N = 2^n, n \in \mathbb{N}$. The algorithm reduces the number of multiplications in the required N summations from an order of 2^{2n} to that of $n2^{n-1}$, a very considerable reduction.

To map the operation in equation (A-15) to the FFT form in (A-16), we set the summation grid by $\eta = \Delta u$ and $u_m = \eta m$, and we set the relative strike grid by $k_j = -b + \lambda j$ with $\lambda = 2\pi/(\eta N)$ and $b = \lambda N/2$. Then, the call value becomes

$$c(k_j) \approx \frac{1}{N} \sum_{m=0}^{N-1} f_m e^{jm\frac{2\pi}{N}i}, \quad f_m = \delta_m \frac{N}{\pi} e^{-\alpha k_j + i u_m b} \chi(u_m - i\alpha)\eta, \tag{A-17}$$

with $j = 0, 1, \cdots, N-1$. The inversion has the FFT form and can hence be computed efficiently across the whole spectrum of strikes k_j.

B. Numerical Integration with Gauss–Hermite Quadrature

Once we have computed the forward call value on asset across the whole spectrum of strikes using the FFT method, we approximate the integration in equation (A-7) by a weighted sum of a finite number (M) of forward asset call values at different equity-to-asset ratio values,

$$c(F_t, K, T) = \int_0^\infty f(\mathcal{X}|X_t)\mathcal{X}C(A_t, K/\mathcal{X}, T)d\mathcal{X} \approx \sum_{j=1}^{M} \mathcal{W}_j \mathcal{X}_j C(A_t, K/\mathcal{X}_j, T), \tag{A-18}$$

where $f(\mathcal{X}|X_t)$ denotes the transition density of X from X_t at time t to $\mathcal{X}$ at time T. The points $\mathcal{X}_j$ and their corresponding weights in the approximation are chosen according to the Gauss–Hermite

quadrature rule.

The constant elasticity of variance process in equation (A-2) is related to a standard Bessel process of order $\nu = 1/(2p)$ through the change of variable, $z_t = X_t^p/(\delta p)$. From the well-known expression for the transition density of the Bessel process (see Borodin and Salminen (1996) and Revuz and Yor (1999) for details on Bessel processes), we can derive the probability transition density as,

$$f(\mathcal{X}|X_t) = \frac{\mathcal{X}^{2p-\frac{3}{2}} X_t^{\frac{1}{2}}}{\delta^2 p(T-t)} \exp\left(-\frac{X_t^{2p}+\mathcal{X}^{2p}}{2\delta^2 p^2(T-t)}\right) I_v\left(\frac{X_t^p \mathcal{X}^p}{\delta^2 p^2(T-t)}\right), \tag{A-19}$$

where $I_\nu(x)$ is the modified Bessel function of the first kind of order ν. Since the value of the modified Bessel function increases quickly once its argument x becomes large, a modified version of the function $J_\nu(x) = I_\nu(x)e^{-x}$ can be calculated with more numerical stability, especially when x is large. Accordingly, the density function can be rewritten as,

$$f(X_T|X_t) = \frac{\mathcal{X}^{2p-\frac{3}{2}} X_t^{\frac{1}{2}}}{\delta^2 p(T-t)} \exp\left(-\frac{(X_t^p-\mathcal{X}^p)^2}{2\delta^2 p^2(T-t)}\right) J_v\left(\frac{X_t^p \mathcal{X}^p}{\delta^2 p^2(T-t)}\right). \tag{A-20}$$

The Gauss-Hermite quadrature rule is designed to approximate the integral $\int_{-\infty}^{\infty} h(x)\, e^{-x^2} dx$, where $h(x)$ is an arbitrary smooth function. After some re-scaling, the integral can be regarded as an expectation of $h(x)$ where x is a normally distributed random variable with zero mean and variance of one half. See Davis and Rabinowitz (1984) for details.

To apply the quadrature rule, we need to map the quadrature nodes and weights $\{x_i, w_j\}_{j=1}^M$ to our choice of $\mathcal{X}_j$ and the weights $\mathcal{W}_j$. Given the constant elasticity of variance dynamics, one reasonable choice is,

$$\mathcal{X}(x) = X_t e^{\sqrt{2V_X}x - \frac{1}{2}V_X}, \quad V_X = X_t^{-2p}(T-t). \tag{A-21}$$

The choice is motivated by a log-normal approximation of the density of $\mathcal{X}$ by assuming that the instantaneous return variance $\delta^2 X_t^{-2p}$ is fixed. Then, given the Gauss-Hermite quadrature $\{w_j, x_j\}_{j=1}^M$, we choose the $\mathcal{X}_j$ points as

$$\mathcal{X}_j = X_t e^{\sqrt{2V_X}x_j - \frac{1}{2}V_X}, \tag{A-22}$$

and the summation weights as

$$\mathcal{W}_j = \frac{f(\mathcal{X}_j|X_t)\mathcal{X}'(x_j)}{e^{-x_j^2}} w_j = \frac{f(\mathcal{X}_j|X_t)\mathcal{X}_j\sqrt{2V_X}}{e^{-x_j^2}} w_j. \tag{A-23}$$

III. Model Estimation and State Identification Methodology

The model uses three state variables $(\widetilde{X}_t, v_t^Z, v_t^J)$ to capture the variation of the implied volatility surface over time. To identify the values of the structural parameters that govern the financial leverage and risky asset dynamics, and to extract the levels of the three state variables at different time periods, we cast the model into a state-space form by treating the three state variables as hidden states, and the option observations as measurements with errors. We employ a nonlinear filtering technique to extract the levels of the states at each date from the implied volatility observations. The model parameters are estimated by maximizing the likelihood defined on the model forecasting errors on the options. The online appendix provides the technical details on the estimation procedure.

Let $V_t \equiv \left[\widetilde{X}_t, v_t^Z, v_t^J\right]^\top$ denote the state vector at time t. We specify the state propagation equation based on an Euler approximation of their statistical dynamics,

$$V_t = f(V_{t-1};\Theta) + \sqrt{Q_{t-1}}\,\varepsilon_t, \tag{A-24}$$

where ε_t denotes the standardized forecasting error vector, $f(V_{t-1};\Theta)$ denotes the conditional forecasts as a function of state vector V_{t-1} and the parameter set Θ, given by,

$$f(V_{t-1};\Theta) = \begin{bmatrix} \widetilde{X}_t + \widetilde{X}_t^{1-p}(\widetilde{a}_X - \kappa_L^\top V_{t-1})\Delta t \\ \kappa_Z\theta_Z\Delta t + (1-\kappa_Z^{\mathbb{P}}\Delta t)v_{t-1}^Z \\ \kappa_J\theta_J\Delta t + (1-\kappa_J^{\mathbb{P}}\Delta t)v_{t-1}^J \end{bmatrix}, \tag{A-25}$$

with $\Delta t = 7/365$ denoting the weekly frequency of the data, $\kappa_L = [\widetilde{\kappa}_{XX}, \widetilde{\kappa}_{XZ}, \widetilde{\kappa}_{XJ}]^\top$, and Q_{t-1} denotes the forecasting error covariance matrix, which is a diagonal matrix with the three diagonal elements given by

$$Q_{t-1} = \begin{bmatrix} \widetilde{X}_{t-1}^{2-2p}\Delta t \\ \sigma_Z^2 v_{t-1}^Z \Delta t \\ \sigma_J^2 (v_{J^-}^{\mathbb{P}})^2 v_{t-1}^J \Delta t \end{bmatrix}. \tag{A-26}$$

The measurement equations are specified on the option observations, with additive, normally-distributed measurement errors:

$$y_t = h(V_t;\Theta) + \sqrt{R}e_t, \tag{A-27}$$

where y_t denotes the time-t forward value of the out-of-the-money options computed from the implied volatility, scaled by the Black-Scholes vega of the option,[2] $h(V_t;\Theta)$ denotes the corresponding model value as a function of the state vector V_t and the parameter set Θ. We assume that the pricing errors on the scaled option prices are i.i.d. normal with zero mean and constant variance.

Estimating the model on the OTC index options data involves 40 measurement equations

[2]See, for example, Bakshi, Carr, and Wu (2008) for a detailed discussion on the rationale for the option pricing transformation and scaling for model estimation.

built on the 40 implied volatility series across five relative strikes at each maturity and eight time to maturities. When we estimate the model on listed options for the five selected companies, the dimension of the measurement equation varies over time as the number of option observations, as well as their relative strikes and time to maturities, varies over time.

When the state propagation and the measurement equation are Gaussian linear, the Kalman (1960) filter provides efficient forecasts and updates on the mean and covariance of the state vector and observations. Our state-propagation equations and measurement equations do not satisfy the Gaussian and linear conditions. We use an extended version of the Kalman filter, the unscented Kalman filter, to handle the deviations.

Let $\overline{V}_t, \overline{y}_t, \overline{\Sigma}_{xy,t}$ denote the time-$(t-1)$ ex ante forecasts of time-t values of the state vector, the measurement series, and the covariance between series x and y; let $\widehat{V}_t, \widehat{y}_t, \widehat{\Sigma}_{xy,t}$ denote the corresponding ex post update on the state vector, the measurement, and the covariances. The unscented Kalman filter uses a set of deterministically chosen (sigma) points to approximate the state distribution. At each time t, if we use k to denote the number of states (three in our model) and use $\eta > 0$ to denote a control parameter, we first generate a set of $2k+1$ sigma vectors χ_{t-1} from the time $(t-1)$ updated mean $\widehat{V}_{t-1}$ and covariance $\widehat{\Sigma}_{VV,t-1}$ of the state vector according to the following equations,

$$\begin{aligned} \chi_{t-1,0} &= \widehat{V}_{t-1}, \\ \chi_{t-1,i} &= \widehat{V}_{t-1} \pm \sqrt{(k+\eta)(\widehat{\Sigma}_{VV,t-1})_j}, \quad j = 1,\ldots,k; \quad i = 1,\ldots,2k, \end{aligned} \tag{A-28}$$

with the corresponding weights w_i given by,

$$w_0 = \eta/(k+\eta), \quad w_i = 1/[2(k+\eta)], \quad i = 1,\ldots,2k. \tag{A-29}$$

These sigma vectors form a discrete distribution with w_i being the corresponding probabilities.

We propagate these sigma points through the propagation equation (A-24) to compute the forecasted mean and covariance of the state vector at time t,

$$\begin{aligned} \overline{\chi}_{t,i} &= f(\chi_{t-1,i};\Theta), \quad \overline{V}_t = \textstyle\sum_{i=0}^{2k} w_i \overline{\chi}_{t,i}, \\ \overline{\Sigma}_{VV,t} &= \textstyle\sum_{i=0}^{2k} w_i (\overline{\chi}_{t,i} - \overline{V}_t)(\overline{\chi}_{t,i} - \overline{V}_t)^\top + Q_{t-1}. \end{aligned} \tag{A-30}$$

We then re-generate the sigma points $\widetilde{\chi}_t$ based on the forecasted mean $\overline{V}_t$ and covariance $\overline{\Sigma}_{VV,t}$, and compute the forecasted mean and covariances of the measurements,

$$\begin{aligned} \overline{\xi}_{t,i} &= h(\widetilde{\chi}_{t,i};\Theta), \quad \overline{y}_t = \textstyle\sum_{i=0}^{2k} w_i \overline{\xi}_{t,i}, \\ \overline{\Sigma}_{yy,t} &= \textstyle\sum_{i=0}^{2k} w_i \left(\overline{\xi}_{t,i} - \overline{y}_t\right)\left(\overline{\xi}_{t,i} - \overline{y}_t\right)^\top + R, \\ \overline{\Sigma}_{Vy,t} &= \textstyle\sum_{i=0}^{2k} w_i \left(\widetilde{\chi}_{t,i} - \overline{V}_t\right)\left(\overline{\xi}_{t,i} - \overline{y}_t\right)^\top. \end{aligned} \tag{A-31}$$

With these moment conditions, we perform the filtering step the same as in the the Kalman filter,

$$\widehat{V}_t = \overline{V}_t + \mathcal{K}_t (y_t - \overline{y}_t), \quad \widehat{\Sigma}_{VV,t} = \overline{\Sigma}_{VV,t} - \mathcal{K}_t \overline{\Sigma}_{yy,t} \mathcal{K}_t^\top, \tag{A-32}$$

where the Kalman gain is

$$\mathcal{K}_t = \overline{\Sigma}_{Vy,t} \left(\overline{\Sigma}_{yy,t}\right)^{-1}. \tag{A-33}$$

We refer the reader to Wan and van der Merwe (2001) for general treatments of the unscented Kalman filter.

Given the forecasted option prices $\overline{y}_t$ and their conditional covariance matrix $\overline{\Sigma}_{yy,t}$ obtained from the unscented Kalman filtering, we compute the quasi-log likelihood value for each week's

observation on the option prices assuming normally distributed forecasting errors,

$$l_t(\Theta) = -\frac{1}{2}\log\left|\overline{\Sigma}_{yy,t}\right| - \frac{1}{2}\left(\left(y_t - \overline{y}_t\right)^\top \left(\overline{\Sigma}_{yy,t}\right)^{-1}\left(y_t - \overline{y}_t\right)\right). \tag{A-34}$$

We estimate the model parameters by numerically maximizing the sum of the conditional log likelihood value on each date,

$$\Theta \equiv \arg\max_{\Theta} \mathcal{L}(\Theta, \{y_t\}_{t=1}^N), \quad \text{with} \quad \mathcal{L}(\Theta, \{y_t\}_{t=1}^N) = \sum_{t=1}^{N} l_t(\Theta), \tag{A-35}$$

where N denotes the number of weeks in the sample.

The model has nine parameters $(p, \kappa_Z, \theta_Z, \sigma_Z, \rho, \kappa_J, \theta_J, \sigma_J, v_J)$ and three state variables (X_t, v_t^Z, v_t^J) to price the equity and equity index options. The model parameters are estimated to match the average shape of the option implied volatility surfaces via the measurement equation (A-27), with the three states capturing the time variation of the volatility surface. In addition, the model has six parameters $(\widetilde{a}, \widetilde{\kappa}_{XX}, \widetilde{\kappa}_{XZ}, \widetilde{\kappa}_{XJ}, \kappa_v^{\mathbb{P}}, \kappa_J^{\mathbb{P}})$ to control the statistical dynamics, which dictate the state propagation equation in (A-24) and are hence identified by the time-series behavior of the option implied volatility series. The differences between $(\kappa_v^{\mathbb{P}}, \kappa_J^{\mathbb{P}})$ and (κ_v, κ_J) determine the market prices of the diffusion and jump variance risk (γ^v, γ^J), respectively.

IV. Option Pricing Performance

Table A1 reports the summary statistics of the pricing errors from the two estimated models, the full model and the reduced-form two-factor benchmark without the financial leverage effect. Panel A reports the sample averages of the pricing errors, defined as the difference between the implied volatility quotes and the corresponding model values. The mean pricing errors from our model are

mostly small except at the one-month maturity and do not show any obvious patterns. The mean pricing errors from the reduced-form benchmark are larger overall and show some remaining pattern along the strike dimension. In particular, without the financial leverage factor, the benchmark model has difficulties fitting the long-term negative implied volatility skew. To compensate, model estimation increases the contribution from the negative jump component, and ends up generating too much negative skew at short maturities. By allowing a distinct channel for the financial leverage variation, the full model can readily generate strong negative implied volatility skews at long maturities and therefore mitigates the tension for matching short-term and long-term implied volatility skews.

[Table A1 about here.]

Panel B reports the mean absolute pricing error in implied volatility points. The estimates from the full model are mostly smaller than those from the reduced-form benchmark. The average mean absolute pricing error from the 40 implied volatility series is 0.69 for our model and 0.95 for the reduced-form benchmark. For both models, the mispricing is the most severe for the one-month 120%-strike series. If we divide the log relative strike by $IV\sqrt{\tau}$, we can see that at one month maturity, 120% strike is over four standard deviations aways from the spot level. The implied volatility quotes thus contain large measurement errors. Furthermore, short-term high-strike implied volatility series tend to show the least co-movements with other implied volatility series. Out-of-money put options and long-dated contracts capture more of institutional needs for hedging against market crashes, whereas far out-of-money call options at short maturities attract more retail activities, act more like lottery tickets, and show more idiosyncratic movements. The identified states capture more of the systematic variations than such idiosyncratic movements.

Panel C reports the weekly autocorrelation of each pricing error series. The pricing errors from both models are quite persistent, with the weekly autocorrelation estimates averaging at 0.91 for the full model and even larger at 0.93 for the two-factor benchmark. Persistence in pricing errors

makes economic sense. If the pricing errors are caused by temporary supply-demand shocks, their dissipation takes time. If we assume a first-order autoregressive structure, the average autocorrelation of 0.91 for the full model implies a half life (i.e., the time it takes for the autocorrelation to be reduced by half) of about eight weeks. The larger average correlation for the benchmark model at 0.93 implies a longer half life of 10.5 weeks. In general, systematic market movements tend to be more persistent than supply-demand shocks. Thus, lower persistence for the pricing errors is an indication of better performance for the model in separating systematic market movements from idiosyncratic supply-demand shocks.[3]

The last row of Table A1 reports the maximized log likelihood values from the two models. The reduced-form benchmark can be regarded as a constrained version of our model, with five fewer parameters and one fewer state variable. One can construct a likelihood ratio static between the two models as twice the difference between the log likelihood values, which has a Chi-squared distribution with 935 degrees of freedom. The p-value from this likelihood ratio test is virtually zero. The reduced-form benchmark is strongly rejected.

Because supply-demand shocks in option contracts mainly dissipate via hedging with nearby contracts (Wu and Zhu (2016)), we also expect the pricing errors of nearby contracts to show positive correlation, but expect such correlation to decline as the contract terms grow further apart. To verify this hypothesis, we compute the cross-correlation of the pricing error series and measure the distance of the different series in terms of their distance in relative strikes and time to maturities. Figure A1 plots the correlation estimates of the pricing errors from the full model against the two distance measures,[4] with Graph A plotting the correlation estimates of same-maturity pairs against the relative strike distance, and Graphs B plotting the correlation estimates of same-strike pairs against the maturity distance. Within the same maturity, the correlation estimates decline clearly

[3] See Bali, Heidari, and Wu (2009) for a detailed illustration of this point in the context of term structure models.

[4] The patterns on the pricing errors from the benchmark model are similar.

with the strike distance. For paris with adjacent strikes, i.e., with relative strike distance of 10%, the correlation estimates are all positive. By contrast, for pairs that are the farthest apart in strike with a distance of 40%, the correlation estimates are all negative. The patterns along the maturity dimension in Graph B are similar but noisier, mainly because hedging with contracts at the same expiry are much more commonly used than across different expiries.

[FIGURE 1 about here.]

Despite the observations on the pricing error persistence and cross correlation, equation (A-27) assumes an iid measurement error structure for model estimation. Bakshi and Wu (2010) and Bates (2000), among others, propose to use more general measurement error structures to accommodate these serial and contemporaneous correlations. Our experience suggests that imposing a diagonal measurement error variance structure for model estimation often brings more numerical stability to the estimation procedure and the extracted states. The intuition is similar in spirit to the idea of ridge regression. The state updating weights involve the inversion of the covariance matrix of the observation as shown in equation (A-33). When the observations are highly correlated with each other, the covariance matrix can become multi-collinear, and the inversion can generate numerically unstable results just as what happens to regressions on highly correlated variables. The ridge regression modifies the covariance matrix of the regressors by adding a small diagonal component to improve the stability of the numerical inversion and to shrinkage the regression coefficients toward zero. By imposing a diagonal structure on the measurement error covariance matrix, the covariance matrix as estimated from equation (A-31) are less likely to become multi-collinear, thus leading to more stable inversions and hence more stable Kalman weighting. Furthermore, the iid assumption amounts to impose an equal importance prior weighting to the 40 implied volatility series.

References

Andersen, T. G.; N. Fusari; and V. Todorov. "The Risk Premia Embedded in Index Options." *Journal of Financial Economics*, 117 (2015), 558–584.

Bakshi, G.; P. Carr; and L. Wu. "Stochastic Risk Premiums, Stochastic Skewness in Currency Options, and Stochastic Discount Factors in International Economies." *Journal of Financial Economics*, 87 (2008), 132–156.

Bakshi, G., and L. Wu. "The Behavior of Risk and Market Prices of Risk over the Nasdaq Bubble Period." *Management Science*, 56 (2010), 2251–2264.

Bali, T.; M. Heidari; and L. Wu. "Predictability of Interest Rates and Interest-Rate Portfolios." *Journal of Business and Economic Statistics*, 27 (2009), 517–527.

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

Borodin, A. N., and P. Salminen. *Handbook of Brownian Motion*. Birkhauser, Boston, MA (1996).

Carr, P., and D. B. Madan. "Option Valuation Using the Fast Fourier Transform." *Journal of Computational Finance*, 2 (1999), 61–73.

Carr, P., and L. Wu. "Time-Changed Lévy Processes and Option Pricing." *Journal of Financial Economics*, 71 (2004), 113–141.

Christoffersen, P.; K. Jacobs; C. Ornthanalai; and Y. Wang. "Option Valuation with Long-Run and Short-Run Volatility Components." *Journal of Financial Economics*, 90 (2008), 272–297.

Christoffersen, P. F.; S. L. Heston; and K. Jacobs. "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well." *Management Science*, 55 (2009), 1914–1932.

Davis, P. J., and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, New York (1984).

Egloff, D.; M. Leippold; and L. Wu. "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments." *Journal of Financial and Quantitative Analysis*, 45 (2010), 1279–1310.

Huang, J.-Z., and L. Wu. "Specification Analysis of Option Pricing Models Based on Time-Changed Lévy Processes." *Journal of Finance*, 59 (2004), 1405–1440.

Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." *Transactions of the ASME–Journal of Basic Engineering*, 82 (1960), 35–45.

Revuz, D., and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, Berlin, Germany, 3rd edition (1999).

Santa-Clara, P., and S. Yan. "Crashes, Volatility, and the Equity Premium: Lessons from S&P 500 Options." *Review of Economics and Statistics*, 92 (2010), 435–451.

Wan, E. A., and R. van der Merwe. "The Unscented Kalman Filter." In S. Haykin, editor, "Kalman Filtering and Neural Networks," chapter 7, 221–280. Wiley & Sons Publishing, New York (2001).

Wu, L., and J. Zhu. "Simple Robust Hedging with Nearby Contracts." *Journal of Financial Econometrics*, 15 (2016), 1–35.

TABLE A1
Model Pricing Performance Comparison

Entries report the summary statistics of the pricing errors from our model (left side) and the reduced-form benchmark (right side). The pricing errors are defined as the difference between market observations and model values in implied volatility points. The last row of the table reports the maximized log likelihood values for the two models.

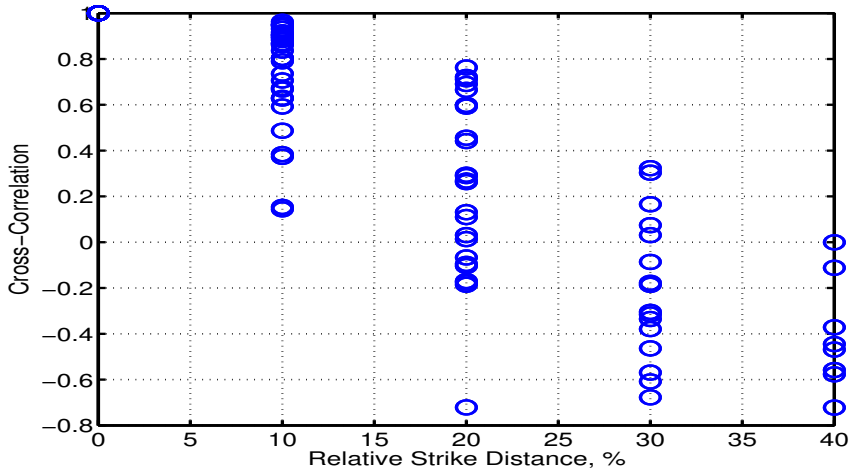
	Full model								Reduced-form benchmark							
$\frac{K}{S}$/m	1	3	6	12	24	36	48	60	1	3	6	12	24	36	48	60
Panel A. Mean Pricing Errors																
80	-0.67	0.69	0.48	0.22	-0.16	-0.34	-0.31	-0.13	0.34	0.59	0.35	0.26	0.23	0.25	0.35	0.50
90	0.44	0.33	0.30	0.22	-0.03	-0.13	-0.06	0.14	-0.74	-0.14	0.08	0.14	0.08	0.10	0.21	0.37
100	-0.33	-0.04	0.13	0.12	-0.05	-0.06	0.06	0.28	-0.09	0.04	0.13	0.05	-0.10	-0.09	0.03	0.21
110	-0.39	-0.08	0.09	0.05	-0.10	-0.06	0.10	0.35	1.94	0.48	0.27	0.01	-0.29	-0.28	-0.16	0.03
120	-0.71	0.66	0.59	0.24	-0.08	-0.07	0.10	0.36	0.26	0.09	-0.07	-0.16	-0.43	-0.47	-0.36	-0.16
Average					**0.05**								**0.10**			
Panel B. Mean Absolute Errors																
80	1.91	0.96	1.01	0.77	0.57	0.59	0.67	0.80	2.46	1.45	1.30	1.03	0.76	0.74	0.89	1.13
90	0.88	0.62	0.60	0.53	0.32	0.26	0.35	0.58	1.71	0.77	0.60	0.53	0.41	0.45	0.66	0.94
100	0.82	0.30	0.43	0.46	0.33	0.22	0.26	0.56	0.95	0.66	0.69	0.56	0.33	0.28	0.48	0.77
110	1.01	0.70	0.68	0.63	0.52	0.40	0.37	0.59	2.65	1.15	1.17	1.00	0.59	0.42	0.42	0.65
120	2.50	1.19	1.08	0.88	0.67	0.55	0.49	0.65	2.74	1.69	1.23	1.16	0.86	0.64	0.49	0.60
Average					**0.69**								**0.95**			
Panel C. Weekly Autocorrelation of Pricing Errors																
80	0.78	0.92	0.97	0.96	0.94	0.95	0.96	0.97	0.84	0.95	0.97	0.97	0.96	0.95	0.96	0.97
90	0.74	0.93	0.94	0.94	0.89	0.86	0.92	0.96	0.83	0.92	0.92	0.92	0.91	0.91	0.95	0.96
100	0.82	0.78	0.89	0.93	0.91	0.86	0.89	0.95	0.82	0.91	0.94	0.94	0.87	0.86	0.93	0.95
110	0.76	0.92	0.95	0.95	0.95	0.94	0.93	0.94	0.89	0.92	0.96	0.98	0.95	0.90	0.92	0.95
120	0.75	0.86	0.94	0.96	0.96	0.96	0.95	0.95	0.78	0.93	0.94	0.98	0.98	0.95	0.93	0.94
Average					**0.91**								**0.93**			
Likelihood					70,264								58,582			

FIGURE A1

Dependence of Pricing Error Cross-Correlations on Strike and Maturity Distance.

Circles in the graphs represent pair-wise cross-correlation estimates of different pricing error series obtained from the full model, plotted against the relative strike distance for same-maturity pairs in Graph A and against maturity distance for same-strike pairs in Graph B.

Graph A. Full Model Same Maturity Pairs



Graph B. Full Model Same Strike Pairs