Supplemental Appendix to:
The Price of the Smile and Variance Risk Premia

PETER H. GRUBER, CLAUDIO TEBALDI, and FABIO TROJANI

This version: April 15, 2020

I. Additional Results in the Matrix AJD Model

A. Pricing Transform in the Matrix AJD Model

Under Assumption 2 and Assumption 5, the closed-form exponentially affine risk-neutral transform for $Y_T := \log(S_T)$ is given by:

$$\Psi(\tau;\gamma) := E_t\left[\exp\left(\gamma Y_T\right)\right] = \exp\Big(\gamma Y_t + tr\big[A(\tau)X_t\big] + B(\tau)\Big) \ , \tag{A-1}$$

where $\tau = T - t$, $A(\tau) = C_{22}(\tau)^{-1}C_{21}(\tau)$ and the 2×2 matrices $C_{ij}(\tau)$ are the $ij-$th blocks of the matrix exponential:

$$\begin{pmatrix} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau) \end{pmatrix} = \exp\left[\tau \begin{pmatrix} M + \gamma Q'R & -2Q'Q \\ C_0(\gamma) & -\big(M' + \gamma R'Q\big) \end{pmatrix}\right] . \tag{A-2}$$

The explicit expressions for the 2×2 matrix C_0 is:

$$C_0(\gamma) = \frac{\gamma(\gamma-1)}{2} I_2 + \Lambda\left[\Theta^Y(\gamma) - 1 - \gamma\Theta^Y(1)\right] \ , \tag{A-3}$$

and real-valued function $B(\tau)$ is given by:

$$\begin{aligned} = \ & \tau\Big\{(\gamma-1)r + \lambda_0\big[\Theta^Y(\gamma) - 1 - \gamma\Theta^Y(1)\big]\Big\} \\ & -\frac{\beta}{2} tr[\ln(C_{22}(\tau)) + \tau(M' + \gamma R'Q)] \end{aligned} \tag{A-4}$$

where $\ln(\cdot)$ is the matrix logarithm and $\Theta^Y(\gamma)$ is the univariate Laplace transform of the return jump size distribution. In the case of the double exponential distribution,

$$\Theta^Y_{DX}(\gamma) = \frac{\lambda^+\lambda^-}{\lambda^+\lambda^- + \gamma(\lambda^+ - \lambda^-) - \gamma^2} \ .$$

In the case of the lognormal distribution

$$\Theta^Y_{LN}(\gamma) = (1+\overline{k})^\gamma \exp\left(\gamma(\gamma-1)\frac{\delta^2}{2}\right) \ ,$$

see, e.g., Leippold and Trojani (2008).

B. Variance Risk Premium in the Matrix AJD Model

The affine expression for the variance risk premium in Proposition 1 is obtained by recalling the relations:

$$
\begin{aligned}
VRP_t(\tau) &= tr\left((E_t^{\mathbb{P}} - E_t^{\mathbb{Q}}) \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] \right) + (E_t^{\mathbb{P}} - E_t^{\mathbb{Q}}) \left[\frac{1}{\tau} \int_t^{t+\tau} (dS_s/S_{s-})^2 \right] \\
&= tr\left((E_t^{\mathbb{P}} - E_t^{\mathbb{Q}}) \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] \right) \\
&\quad + E^{\mathbb{Q}}[\mathcal{E}(1+k)] tr\left(\Lambda(\beta_\lambda^* E_t^{\mathbb{P}} - E_t^{\mathbb{Q}}) \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] \right) .
\end{aligned}
$$

This shows that $VRP_t(\tau)$ is the sum of two-affine functions of state X_t. To compute these functions in closed-form, we need to compute the $\mathbb{P}$ and $\mathbb{Q}$ expectation of the average integrated state X in our model. These expectations are available in closed-form:

$$
E_t^{\mathbb{Q}} \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] = X_\infty^{\mathbb{Q}} + \frac{1}{\tau} \int_0^\tau e^{Mu} (X_t - X_\infty^{\mathbb{Q}}) e^{M'u} du \ , \tag{A-5}
$$

where the long-run mean $X_\infty^{\mathbb{Q}}$ is the unique solution of the Lyapunov equation $MX_\infty^{\mathbb{Q}} + X_\infty^{\mathbb{Q}} M' = \beta Q'Q$. Similarly,

$$
E_t^{\mathbb{P}} \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] = X_\infty^{\mathbb{P}} + \frac{1}{\tau} \int_0^\tau e^{M^*u} (X_t - X_\infty^{\mathbb{P}}) e^{M^{*\prime}u} du \ , \tag{A-6}
$$

where $X_\infty^{\mathbb{P}}$ is such that $M^* X_\infty^{\mathbb{P}} + X_\infty^{\mathbb{P}} M^{*\prime} = \beta^* Q'Q$. This implies, for any 2×2 matrix D:

$$
tr\left(D E_t^{\mathbb{Q}} \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] \right) = tr\left[D \cdot \left(X_\infty^{\mathbb{Q}} + A_\tau^{\mathbb{Q}}(X_t - X_\infty^{\mathbb{Q}}) \right) \right] \ , \tag{A-7a}
$$

$$
tr\left(D E_t^{\mathbb{P}} \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right] \right) = tr\left[D \cdot \left(X_\infty^{\mathbb{P}} + A_\tau^{\mathbb{P}}(X_t - X_\infty^{\mathbb{P}}) \right) \right] \ , \tag{A-7b}
$$

where, for any 2×2 matrix H:

$$
A_\tau^{\mathbb{Q}}(H) := \frac{1}{\tau} \int_0^\tau e^{Mu} H e^{M'u} du \ ; \quad A_\tau^{\mathbb{P}}(H) := \frac{1}{\tau} \int_0^\tau e^{M^*u} H e^{M^{*\prime}u} du \ .
$$

Since these two functions are linear in H, the variance risk premium is affine in X_t. This concludes the proof. □

C. Stochastic Discount Factor in the Matrix AJD Model

Existence of a well-defined stochastic discount factor to price all shocks in our model is ensured by a proper density for an equivalent change of measure, from the physical to the risk neutral probability. To this end, we specify matrix processes $\{\Gamma_{1t}\}$, $\{\Gamma_{2t}\}$ for the market prices of Brownian shocks dW_t^*, dB_t^*, and an appropriate distribution for return jumps. Following Assumption 2, we specify a double exponential distribution for log return jumps, with parameters $\lambda^{+*}, \lambda^{-*}$ and λ^+, λ^-, respectively, under the physical and the risk neutral probabilities. We show that, under Assumption 4 in the main text, a proper density process consistent with these properties is defined for any $T \geq 0$ by:

$$
\begin{aligned}
\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_T} &= \exp\left\{ tr\left(-\int_0^T \Gamma_{1t} dW_t^* + \frac{1}{2}\int_0^T \Gamma_{1t}'\Gamma_{1t}dt - \int_0^T \Gamma_{2t} dB_t^* + \frac{1}{2}\int_0^T \Gamma_{2t}'\Gamma_{2t}dt \right)\right\} dt \\
&\quad \times \prod_{i=1}^{N_T^*} \exp\left\{ -(\lambda^- - \lambda^{*-})J_i^{*-} - (\lambda^+ - \lambda^{*+})J_i^{*+} + \ln\left(\frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+}\right)\right\} ,
\end{aligned} \tag{A-8}
$$

where

$$
\Gamma_{1t} = \sqrt{X_t}\Gamma + \frac{1}{2\sqrt{X_t}}(\beta^* - \beta)Q' , \tag{A-9}
$$

and

$$
\Gamma_{2t} = \sqrt{X_t}\Delta + \frac{\mu_0 - (r - q)}{\sqrt{X_t}} , \tag{A-10}
$$

with $\mu_0 - (r - q) \geq 0$ and Δ a 2×2 parameter matrix. The first (second) line of equality (A-8) defines a possible change of measure for diffusive (jump) shocks in our model.

Under Assumption 4, the stochastic exponential in the first line of (A-8) is a well-defined positive local martingale, and hence a supermartingale. Therefore, to show that this term is a martingale, it is enough to show that it has a constant expectation:

$$
1 = E_0^{\mathbb{P}}\left[\exp\left\{ tr\left(-\int_0^T \Gamma_{1t} dW_t^* + \frac{1}{2}\int_0^T \Gamma_{1t}'\Gamma_{1t}dt - \int_0^T \Gamma_{2t} dB_t^* + \frac{1}{2}\int_0^T \Gamma_{2t}'\Gamma_{2t}dt \right)\right\} dt \right] .
$$

In our matrix AJD setting, this property does not follow from a standard Novikov-type condition. However, it follows from a localization argument; see, e.g., Mayerhofer (2014). We now show that the second line of (A-8) also defines a martingale process. Using the independence between IID log jump sizes J^* and counting process N^* under the physical probability, it is enough to show that:

$$1 = E_0^{\mathbb{P}} \left[\exp\left\{ -(\lambda^- - \lambda^{*-}) J^{*-} - (\lambda^+ - \lambda^{*+}) J^{*+} \right\} \frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+} \right] . \tag{A-11}$$

Explicit calculations of the expectation on the right hand side yield:

$$\frac{\lambda^{*-}\lambda^{*+}}{\lambda^{*-} + \lambda^{*+}} \cdot \frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+} \int_{-\infty}^{\infty} \exp\left(-\lambda^- J^{*-} - \lambda^+ J^{*+} \right) dJ^* = 1 .$$

With respect to the risk-neutral probability $\mathbb{Q}$, log return jumps follows a double exponential distribution with parameters λ^- , λ^+. Indeed, for any $u \in \mathbb{R}$ it follows:

$$E^{\mathbb{Q}} \left[\exp(uJ) \right] = \frac{\lambda^- \lambda^+}{\lambda^- + \lambda^+} \int_{-\infty}^{\infty} e^{uJ} e^{-\lambda^- J^- - \lambda^+ J^+} dJ ,$$

which is the Laplace transform of a double exponential distribution with parameter λ^-, λ^+. This concludes the proof.

□

D. Mapping of our notation to the Bates (2000) notation

Several well-studied affine option pricing models with independent factors are nested in our framework, if we allow β to be a diagonal matrix instead of a scalar. In this case, the independent volatility factors can be written as diagonal elements of X_t. Below, we show the equivalence of the processes and how the parameters can be converted from the notation in the original papers into our notation. For the sake of legibility, we suppress the time index on all components of state variables and Brownian motions.

The return dynamics of the $SV_{2,0}$ two-factor model of Christoffersen, Heston and Jacobs (2009) is

$$\frac{dS}{S} = (r - q)dt + \sqrt{V_1} dz_1 + \sqrt{V_2} dz_2 \tag{A-12}$$

where r is the risk-free rate, q the dividend yield and V_i are two independent stochastic volatility factors with the following dynamics:

$$dV_i = (a_i - b_i V_i)dt + \sigma_i \sqrt{V_i} dw_i \qquad i = 1, 2 \tag{A-13}$$

where the correlation between dz_i and dw_j is $\delta_{ij}\rho_i$.

If we write $X = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$ and $dZ = \begin{pmatrix} dz_1 & dZ_{12} \\ dZ_{21} & dz_2 \end{pmatrix}$, then (A-12) can be written as

$$\frac{dS}{S} = (r-q)dt + tr[\sqrt{X_t}dZ],$$

which is exactly the diffusive part of our return equation.

In order to show the equality of the volatility factors, we first need to establish that the diagonal elements of X_t in (8) are independent CIR processes if the parameter matrices M, R, Q are diagonal. We start by explicitly writing the diagonal elements of X_t in this case:

$$dX_{ii} = \left(\beta Q_{ii}^2 + 2M_{ii}X_{ii}\right)dt + \sum_k \sqrt{X_{ki}}\, dB_{ki} \qquad \text{(A-14)}$$

To eliminate the seeming interdependence of the diagonal elements, we introduce n new independent Brownian motions dW_i:

$$dW_i = \frac{1}{\sqrt{X_{ii}}} \sum_k \sqrt{X_{ki}}\, dB_{ki}$$

This allows us to express (A-14) as n independent CIR processes:

$$dX_{ii} = \left(\beta Q_{ii}^2 + 2M_{ii}X_{ii}\right)dt + 2Q_{ii}\sqrt{X_{ii}}dW_i \qquad \text{(A-15)}$$

To convert our notation into the notation of (A-13), simply set

$$a_i = \beta_{ii}Q_{ii}^2, \quad b_i = -2M_{ii}, \quad \sigma_i = 2Q_{ii}, \text{ and } \rho_i = R_{ii}.$$

Remark 1 *Our state matrix X_t will generally not remain diagonal, even if all parameter matrices and the initial state X_0 are diagonal. This does not void the nesting argument, because $X_{12,t}$ does not enter the pricing equation. There is no economic interpretation for the process $X_{12,t}$, it is a mere artifact of writing a two-dimensional CIR process in matrix form.*

The jump intensity in Bates (2000) is given as $\lambda_t = \lambda_0 + \lambda_1 V_{1t} + \lambda_2 V_{2t}$, which is already identical to our jump intensity $\lambda_t = \lambda_0 + tr(\Lambda X_t)$, if we write $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Our definitions of the jump size distribution is the same as in Bates (2000).

II. Estimation

A. Extended Kalman Filter

The first step in out estimation approach is based on an extended Kalman filter for our matrix AJD model, using the full panel of option-implied volatilities as observables in the observation equation of the filter. Hence, we estimate all risk-neutral parameters in the model from the observation equation and the physical parameters in the matrix state dynamics from the transition equation. Let us denote the set of all parameters estimated in this first step by $\theta := (M, Q, R, \beta, \lambda^+, \lambda^-, \Lambda; M^*, \beta^*)$. The physical dynamics of our state variable is given in (17):

$$dX_t = \left[\beta^* Q'Q + M^* X_t' + X_t M^{*\prime}\right] dt + \sqrt{X_t} dB_t^* Q + Q' dB_t^{*\prime} \sqrt{X_t} \ .$$

We discretize this process on a weekly grid with $\Delta_k = 7$ calendar days. When there is no data for a given Wednesday, we skip the respective week and set $\Delta_k = 14$. We initialize the filter at the steady state $X_0 = X_\infty^{\mathbb{P}}$, which is computed by solving the Lyapunov equation $M^* X_\infty^{\mathbb{P}} + X_\infty^{\mathbb{P}} M^{*\prime} = \beta^* Q'Q$. We work with half vectorized states $\widehat{X}_t$ and initialize the variance of state $\widehat{X}_0$ as $\widehat{\Sigma}_0 = 0$. At each step, we compute from the Laplace transform (A-1) the exact expectation ($\overline{X}_{t+\Delta}$) and covariance matrix ($\overline{V}_{t+\Delta}$) of hidden state $\widetilde{X}_{t+\Delta}$ conditional on filtered state $\widehat{X}_t$:

$$\begin{aligned} \overline{X}_{t+\Delta} &= \beta\,\overline{\mu} + \Phi \widehat{X}_t \Phi' &\text{(A-16)} \\ \overline{V}_{t+\Delta} &= (I_4 + K_4)\left(\Phi \widehat{X}_t \Phi' \otimes \overline{\mu} + \beta\ \overline{\mu} \otimes \overline{\mu} + \overline{\mu} \otimes \Phi \widehat{X}_t \Phi'\right) &\text{(A-17)} \end{aligned}$$

where

$$\begin{aligned} \overline{\mu} &= -\frac{1}{2} C_{12} C_{11}' \\ \Phi &= e^{\Delta M^*} \\ C &= \exp\left[\Delta \begin{pmatrix} M^* & -2Q'Q \\ 0 & -M^{*\prime} \end{pmatrix}\right] = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \end{aligned}$$

with square 2×2 matrices $C_{11}, C_{12}, C_{21}, C_{22}$ and 4×4 commutation matrix K_4. In this way, we obtain the the set of observation equations

$$\widehat{O}_{t+\Delta,i} = O_{t+\Delta,i}(\widetilde{X}_{t+\Delta}; \theta) + \varepsilon_{t+\Delta,i}, \ ; \ i = 1, \dots, N_{t+\Delta} \ . \qquad \text{(A-18)}$$

Here, $\widehat{O}_{t+\Delta,i}$ is the Black-Scholes implied volatility of the $i-$th option on day $t + \Delta$, $N_{t+\Delta}$ the total number of options observed on that day, $O_{t+\Delta,i}(\widetilde{X}_{t+\Delta}; \theta)$ the model-induced Black-

Scholes implied volatility and $\varepsilon_{t+\Delta,i}$ a cross-sectionally iid noise with zero mean and variance σ_r^2. We allow observation errors to be correlated in the time series, by allowing for a degree of autocorrelation in average pricing errors over time:

$$corr(\overline{\varepsilon}_{t+\Delta}, \overline{\varepsilon}_t) = \rho_r \ ,$$

where $\overline{\varepsilon}_t$ is the average pricing error over all options on day t. Given the linearity of the transition equation, we need to linearize only the observation equation, by computing numerically the Jacobian matrix:

$$G_t = \frac{\partial O_{t+\Delta}}{\partial \widetilde{X}_t} \ . \tag{A-19}$$

Using matrix

$$F = \Phi \otimes \Phi \ , \tag{A-20}$$

we update the state variance covariance matrix as

$$\widetilde{\Sigma}_{t+\Delta} = F\widehat{\Sigma}_t F' + \overline{V}_{t+\Delta} \ . \tag{A-21}$$

Finally, we update the filtered state $\widehat{X}_{t+\Delta}$ and the variance covariance matrix for the next iteration as follows:

$$\begin{aligned} S_t &= G_t\widetilde{\Sigma}_{t+\Delta}G_t' + \sigma_r^2 I_2 \\ H_t &= \widetilde{\Sigma}_{t+\Delta}G_t' S_t^{-1} \\ \widehat{X}_{t+\Delta} &= \widetilde{X}_{t+\Delta} + H_t\left(\widehat{O}_{t+\Delta,i} - O_{t+\Delta,i}(\widetilde{X}_{t+\Delta}, \theta)\right) \\ \widehat{\Sigma}_{t+\Delta} &= (I_2 - H_t G_t)\widetilde{\Sigma}_{t+\Delta} \end{aligned}$$

Given parameter vector θ, we compute the time-series of predicted states $\{\widetilde{X}_t\}$ and the corresponding log-likelihood function

$$\mathcal{L}(\theta) = \sum_{i=1}^{N}\left[\log\det(S) + \left(\widehat{O}_{t+\Delta,i} - O_{t+\Delta,i}(\widetilde{X}_{t+\Delta},\theta)\right)' S_t^{-1}\left(\widehat{O}_{t+\Delta,i} - O_{t+\Delta,i}(\widetilde{X}_{t+\Delta},\theta)\right)\right] \ . \tag{A-22}$$

The estimated parameter $\widehat{\theta}$ is the maximizer of $\mathcal{L}(\theta)$, where the maximization is performed using differential evolution of Storn and Price (1997).

B. Model Identification

B.1. Admissibility conditions

The set of admissibility conditions are necessary for the infinitesimal generator of the Wishart process to define a regular Markov process $X_t^{x_0}$ with evolution in the set of positive semidefinite symmetric square 2×2 matrices, where x_0 represents the initial condition. The condition required to apply Theorem 2.4 of Cuchiero, Filipović, Mayerhofer, Teichmann (2011) in our framework is:

$$\Omega\Omega' \succeq QQ'$$

($A \succeq B$ means that the difference $A - B$ is positive semidefinite).[1] This condition grants the existence of a weak solution if x_0 is a symmetric, positive-definite 2×2 matrix and a strong solution if x_0 symmetric positive definite and nonsingular.

Within our estimation sample this more general parametrization did not produce any advantage with respect to the more parsimonious $\Omega\Omega' = \beta QQ'$ with $\beta \geq 1$ that is henceforth adopted in the text. The stricter condition $\beta \geq 3$ is required for the state to remain positive with probability 1. The estimated parameters $1 < \beta, \beta^* < 3$ imply that the time series of initial conditions includes periods of time where the state variable is singular (non-positive definite) and in particular over periods of time with low volatility the rank of the state variable is statistically not different from 1. The dynamics of the process is well defined also when the state matrix has reduced rank one, in fact the process is locally equivalent to a one-dimensional square root process and the admissibility condition grants that the process is constrained within the the cone of positive semi-definite symmetric matrices.

Two additional conditions are required to have a well defined joint return-variance process: these are i) In order to have a well defined correlated brownian motion matrices it is necessary that $RR' \preceq \mathbb{I}_2$. ii) In order to have a positive jump intensity the conditions $\lambda_0 \geq 0$ and $Tr\left[\Lambda X\right] \geq 0$. Assuming a lower triangular expression $\Lambda = [\Lambda_{11}\ 0;\ \Lambda_{12}\ \Lambda_{22}]$, it is sufficient to require that the symmetric matrix $L = [\Lambda_{11}\ \Lambda_{12}/2;\ \Lambda_{12}/2\ \Lambda_{22}]$ is semi-positive definite, then: $Tr\left[\Lambda X\right] = Tr\left[LX\right] \geq 0$.

B.2. Identification conditions

During an estimation procedure in the presence of unobserved factors, it is necessary to determine jointly parameters and state variables.

[1]The condition 2.11 pg. 404 in Cuchiero et al. (2011) does not imply any restriction on M in the absence of the jump component.

In the following we illustrate the basic restrictions that have to be imposed on the parameters to guarantee that the mapping between observed time series and the level of the latent state variables is uniquely defined.

As previously stated the universe of tradable securities that form the market span includes the full panel of options in addition to a fund tracking the underlying index level and a the risk-free security available in any quantity in the market. Recalling that option prices are determined by the transform method approach as an integral convolution of the characteristic function, we assume that the econometrician observes a large panel of traded option prices (and model free replications of variance swap prices) so that the characteristic function of the underlying data generating process can be uniquely reconstructed.

Characteristic function of the process is uniquely determined by the solution of a set of Riccati ODE whose flow, in turn, is uniquely determined by the parameters that define the infinitesimal generator.

As a consequence the mapping between observed prices and state variables is non-unique if the model admits invariant transformations, i.e. joint transformations of the state variables and of the set of parameters that leave unaffected the final expression of the infinitesimal generator.

In the following we derive the identification conditions that remove these ambiguities. Recall that the infinitesimal generator for the risk neutral process has the following expression:

$$\begin{aligned}\mathcal{A}^{\mathbb{Q}}_{Y,X} f(Y,X) &= \left(r - q - \frac{Tr[X]}{2}\right)\frac{\partial f}{\partial Y} + \frac{Tr[X]}{2}\frac{\partial^2 f}{\partial Y^2} + Tr[XR'QD]\frac{\partial f}{\partial Y} \\ &+ Tr[\Omega'\Omega + MX + XM' + XDQ'QD]f \\ &+ (\lambda_0 + Tr[\Lambda X])\int d\nu(z)\,[f(Y+z,X) - f(Y,X)] \qquad \text{(A-23)}\end{aligned}$$

where $D = \left(\frac{\partial}{\partial X_{ij}}\right)_{1\leq i,j\leq 2}$, $Y = \log(S)$ and $\nu(z)$ is the double-exponential density. The generator of the process under the $\mathbb{P}$-measure has the same functional expression with a different selection of the parameter M that is shifted by the change of measure as described in (18) in the main paper.

To remove the ambiguity in the change of measure induced by the fact both factors of the product $\lambda\overline{k}$ may be affected by the change of measure, we use the Assumption 5 that forces the jump intensity to be unaffected by the change of measure.[2] As an implication, the vector including all the parameters that are to be uniquely determined in terms of observed data is given by:

$$\left(\underbrace{M, M^*, Q, R, \Lambda, \Delta}_{2\times 2 \text{ Matrices}}, \underbrace{\lambda_0, \lambda_+, \lambda_-, \mu, r, q, \beta_\lambda, \beta, \beta^*}_{\text{Scalars}}\right)$$

while the state variable X is a symmetric, positive definite 2×2 matrix.

In order to remove all the ambiguities on the matrix diffusion process we proceed as follows: first, we classify all relevant invariant transformations. These are listed below:

- The permutation matrix that exchanges row and column 1 with row and column 2. We remove the permutation matrix degeneracy imposing that the eigenvalues of the matrix X are ordered in increasing order.

- The transposition of line and rows, in fact X is symmetric. Transposition invariance would simply exchange upper with lower triangular matrices, so the above definitions requiring the matrices to be lower triangular remove also the transposition invariance.

- Rotation matrices. This can be proved as follows: the trace is a spectral invariant, this implies that the set of invariant transformation is determined by a joint transformation of the state variable X and of the matrix parameter P:

$$\begin{aligned} X &\rightarrow \mathcal{D} X \mathcal{D}^{-1} \\ P &\rightarrow \mathcal{D}^{-1} P \mathcal{D} \end{aligned}$$

  where P is a generic matrix parameter while $\mathcal{D}$ may be an element of the group $GL(2)$ of general linear transformations. Note however that the state variable has to be a symmetric matrix $X \in Sym_+(2)$, hence invariant transformations are only those orthogonal transformations $O(2) \subset GL(2)$, i.e. $\mathcal{D}^{-1} = \mathcal{D}'$, that preserve the symmetry of the state variable matrix. This is a direct implication of the spectral theorem: two symmetric matrices sharing the same set of eigenvalues may differ at

[2] Note that our estimation procedure does not produce a unique identification of the conditional return distribution under the historical measure. In fact a unique parameter β_λ is determined from eq.(30) by considering model free payoffs of the variance swaps while the historical jump size distribution is determined by two parameters λ_+^*, λ_-^*. Hence complete identification would be achieved forcing an additional parametric restriction to make the mapping $(\beta_\lambda, \lambda_+, \lambda_-) \longleftrightarrow (\lambda_+^*, \lambda_-^*)$ one-to-one. The explicit determination of this mapping is inconsequential for our results.

most by an orthogonal transformation:

$$X \in Sym_+(2) \Rightarrow \mathcal{D}X\mathcal{D}^{-1} \in Sym_+(2) \text{ iff } \mathcal{D} \in O(2)$$

In the following we list two Corollaries that will be used to select the parametric restrictions:

Corollary 2 *If X is any 2×2 symmetric matrix then the value of $Tr[AX]$ depends only on the sum of the out of diagonal elements $(A_{12} + A_{21})$ of A.*

Proof.

$$Tr[AX] = A_{11}X_{11} + (A_{12} + A_{21})X_{12} + A_{22}X_{22}$$

■

Corollary 3 *Under an orthogonal transformation $\mathcal{D}$ the matrix differential operator D transforms as: $\mathcal{D}^{-1}D\mathcal{D}$.*

Proof. The result follows from the application of the total differential rule applied to the change of variables:

$$X' = \mathcal{D}^{-1}X\mathcal{D}$$

■

We are now ready to state:

Proposition 1 *Consider the infinitesimal generator (A-23) of an admissible process (X_t, Y_t) where X_t is a symmetric, positive definite 2×2 matrix and $Y_t = \log(S_t)$. There are no transformations that leave it invariant if the following parametric conditions apply:*

- *M, $M^* = M + \Gamma Q$ are non singular lower triangular matrices with positive elements along the diagonal.*
- *Q and Ω (in the general specification) are non singular upper triangular matrices with positive elements along the diagonal.*
- *R, Λ and Δ are upper triangular matrices.*

Proof.

The Trace terms appearing in the infinitesimal generator of the risk neutral process are:

(1) $\mathbb{Q}$ measure: $Tr[\Omega'\Omega + MX + XM' + XDQ'QD]$ ($\mathbb{P}$ measure: $Tr[\Omega'\Omega + M^*X + X(M^*)' + XDQ'QD]$),

(2) $Tr[XR'QD]$ where $D = \left(\frac{\partial}{\partial x_{ij}}(\cdot)\right)_{i,j=1,2}$,

(3) $Tr[X\Lambda]$.

Consider first term (1): the Choleski decomposition theorem states that one can derive the existence of two unique upper triangular non singular matrices Ω, Q. Prescription on Ω and Q to be non singular upper triangular matrices with negative elements along the diagonal remove the invariance $\Omega'\Omega = (\mathcal{D}\Omega)'\mathcal{D}\Omega$ and $Q'Q = (\mathcal{D}Q)'(\mathcal{D}Q)$.

To remove the invariance determined by the invariant transformation $\mathcal{D}^{-1}X\mathcal{D}$, we exploit the unique 'QR' decomposition of a non singular M as a product of an orthogonal matrix and an lower triangular matrix with negative elements along the diagonal $M = \mathcal{D}^M M^U$. Selecting M lower triangular and $\mathcal{D}^M = \mathbb{I}$ any rotational invariance is removed. The same identification condition works also for M^*.

Consider now term (2): rewriting it as $Tr[R'QDX_t]$, the degeneracy is completely removed by the condition that the matrix R' is lower triangular. This can be proved as follows: first we observe that the term DX_t is symmetric. In fact, by the symmetry of the matrix X_t follows that:

$$DX_t = \sum_{1\leq j\leq 2} \frac{\partial}{\partial x_{ij}}(\cdot)\, x_{jk} = \sum_{1\leq j\leq 2} x_{kj}\frac{\partial}{\partial x_{ji}}(\cdot) = (DX_t)^T$$

Then, by the above Corollary, one can immediately conclude that the sum depends only on the sum $(R'Q)_{12} + (R'Q)_{21}$. In light of the above considerations, Q is uniquely specified as upper triangular. This implies that, to have $R'Q$ uniquely determined and upper triangular with $(R'Q)_{21} = 0$, also $R' = (R'Q)\,Q^{-1}$ must be lower triangular. Finally, consider term (3), uniqueness of the specification of Λ is achieved by imposing that it is lower triangular. ■

C. Definition of Level $\mathcal{L}_t$, Skew $\mathcal{S}_t$ and Term Structure $\mathcal{M}_t$ Proxies

To analyze our results in terms of observable properties of the implied volatility surface, such as in Figure 3, we define the following proxies[3]

[3] We have evaluated the regression $IV(\tau, K)_t = \mathcal{L}_t + \mathcal{S}_t \cdot K + \mathcal{M}_t \cdot \tau$ as an alternative specification. We have found similar, but more noisy results. We have also performed robustness checks with respect to our definition. The alternative term structure measure $\mathcal{M}_t^6 := \frac{1}{\frac{6}{12}-\frac{1}{12}}\left[IV(\tau = \frac{6}{12}, \Delta = 0.5) - IV(\tau = \frac{1}{12}, \Delta = 0.5)\right]$ is, for example, 92% correlated with our term structure measure.

$$
\begin{array}{lll}
\text{level} & \mathcal{L}_t & := IV(\tau = \tfrac{1}{12}, \Delta = 0.5) \\
\text{short term skew} & \mathcal{S}_t & := \tfrac{1}{0.6-0.4}\left[IV(\tau = \tfrac{1}{12}, \Delta = 0.6) - IV(\tau = \tfrac{1}{12}, \Delta = 0.4)\right] \\
\text{long term skew} & \mathcal{S}_t^{long} & := \tfrac{1}{0.6-0.4}\left[IV(\tau = \tfrac{3}{12}, \Delta = 0.6) - IV(\tau = \tfrac{3}{12}, \Delta = 0.4)\right] \\
\text{term structure} & \mathcal{M}_t & := \tfrac{1}{\frac{3}{12}-\frac{1}{12}}\left[IV(\tau = \tfrac{3}{12}, \Delta = 0.5) - IV(\tau = \tfrac{1}{12}, \Delta = 0.5)\right] \\
\text{skew term structure} & \mathcal{M}_t^{skew} & := \tfrac{1}{\frac{3}{12}-\frac{1}{12}}\left[\mathcal{S}_t^{long} - \mathcal{S}_t\right]
\end{array}
$$

where IV and Δ stand for the Black-Scholes implied volatility and delta. The time to maturity τ is measured in years. In the data, we obtain the required implied volatilities through two-dimensional interpolation of the volatility surface. In the model, we calculate these quantities exactly.

III. Review of the Literature

Our work borrows from an enormous literature that has studied the economic sources of volatility variations, the dynamics of the option-implied volatility smile and the origins of a negative variance premium. We contribute to this literature along several dimensions.

First, we use a novel specification of stochastic volatility, which parsimoniously jointly identifies three multi-frequency volatility risk factors, the price of the smile and the term structure of the variance risk premium. Following Heston's (1993) seminal model, Bates (2000) was the first to recognize that volatility is a multi-frequency object dependent on factors with distinct persistence and variability properties. More recent three-factor specifications such as Carr and Wu (2017), Gruber, Tebaldi and Trojani (2010) and Andersen, Fusari and Todorov (2015a), among others, improve on the fit of the volatility smile provided by benchmark two-factor models. Andersen, Fusari and Todorov (2015b) study the predictive power of option risk factors for future index and index volatility returns, but remain agnostic about the price of the smile. Bardgett, Gourier and Leippold (2019) estimate a three factor model with jumps in volatility in a standard affine setting using the information in both SPX and VIX options and a particle filter.

Our model is complementary to these approaches by adding several new ingredients. We relax the assumption of factor independence with a new specification of interdependent volatility factors that follow a matrix AJD.[4] This state space yields three economically

[4]See, among others, Gourieroux (2006), da Fonseca, Grasselli and Tebaldi (2008) and Buraschi, Porchia and Trojani (2010) for examples and applications of affine matrix-valued diffusions, as well as Leippold and Trojani (2008) for a broad class of affine matrix jump diffusion processes.

interpretable volatility risk factors, which feature distinct persistence features and are priced very differently. It also implies dynamic factor correlations, which parsimoniously embed a dynamic skewness component disconnected from the spot volatility.[5] In contrast to Andersen et al. (2015b) or Bardgett et al. (2019), we specify and estimate our model under the physical and the pricing measures, thus providing a coherent no-arbitrage framework for variance risk factors, their prices (i.e. the price of the smile) as well as the variance risk premium and the equity premium. Relaxing the factor independence assumption furthermore allows for a richer structure of the price of the smile, because the price of each risk factor is not exclusively spanned by the factor itself. We adopt a matrix jump diffusion process with jumps in index returns allowing for a parsimonious specification of the state dynamics. We document that our model does not require jumps in volatility to provide a sufficiently realistic description of the SPX volatility dynamics. While it would be in principle possible to add jumps to our volatility process, a parsimonious specification would require non-innocuous additional identification assumptions.[6]

Second, our paper borrows from a large literature that has studied the trading of variance risk factors, the market price of volatility and the term structure of variance risk premia. In a first strand of this literature, Dupire (1993) and Neuberger (1994) were among the first to propose option portfolio strategies for trading proxies of realized variance, followed by Carr and Madan (1998), Demeterfi, Derman, Kamal and Zou (1999) and Britten-Jones and Neuberger (2000), among others. From the price of such portfolios, the price of variance can be measured in a model-free way, giving rise to a variety of synthetic variance swap contracts. Recent papers have focused on the properties of variance swaps in presence of jumps and on swap contracts for trading higher-order risks, such as, e.g., skewness and kurtosis.[7] A key insight of this literature, which motivates part of our work, is the tradeability of variance, skewness and higher-order unspanned risks by means of appropriate option portfolios. Given the no-arbitrage constraints prevailing in liquid option markets, it is natural to expect that the prices of these risks are interconnected and difficult to study in isolation. Therefore,

[5] This helps to avoid the puzzling skew sensitivities of benchmark arbitrage-free models noted in Constantinides and Lian (2015).

[6] For instance, Carr and Wu (2017) assume that the probability of a co-jump in returns and volatility follows a pure-jump single-factor dynamics. Such an assumption restricts the jump variance risk premia to be perfectly correlated across horizons, which we feel excessively constrains the term structure of variance risk premia for our analysis.

[7] Martin (2012), Neuberger (2012) and Bondarenko (2014) introduce definitions of variance swap payoffs robust to jumps. Kozhan, Neuberger and Schneider (2013) propose a synthetic skew swap to study skewness vs. variance risk premia, while Schneider and Trojani (2014) trade and price fear using skew swaps. More broadly, Schneider and Trojani (2019) introduce divergence swaps and characterize in a model-free way the premia for trading general nonlinear risks of different orders.

we endow our model with a joint arbitrage-free specification of the price of variance risk together with the risk factors traded in option markets and the price of the smile.

Another strand of this literature has established the existence of a negative risk premium for market volatility. Buraschi and Jackwerth (2001) test the spanning properties of option markets and conclude in favour of models with priced unspanned risks, such as stochastic volatility or return jumps. Bakshi and Kapadia (2003) provide first direct evidence on a negative variance risk premium using delta-hedge call option positions. Similar evidence is obtained by Carr and Wu (2009), using synthetic variance swaps. Todorov (2010) and Bollerslev and Todorov (2011) conclude that variance risk premia are dominated by a premium for jump variance risk.

A third strand of this literature studies the term structure of variance risk premia. Ait-Sahalia, Karaman and Mancini (2012) and Filipovic, Gourier and Mancini (2016) estimate an affine and a quadratic two-factor volatility model, based on OTC variance swaps. Their focus is on the term structure of equity vs. variance risk premia and on the optimal portfolio choice with variance swaps, respectively. The first paper estimates a downward sloping term structure of variance risk premia. The second paper documents that the optimal portfolio contains an important position in variance swap calendar spreads, which earns a premium in the decreasing term structure of variance risk premia and simultaneously limits portfolio losses when volatility rises. Li and Zinna (2016) estimate a three factor jump diffusion using returns and variance swaps data. In their model, the term structure of variance risk premia is inverted for short periods of time, when volatility is sufficiently large, while it is steeply downward sloping in periods of low volatility. Dew-Becker, Giglio, Le and Rodriguez (2016) estimate a discrete-time version of Ait-Sahalia et al.' s 2012 model, using different sets of variance swaps with maturities from 1 month to 14 years. They document that a steep unconditional term structure of variance risk premia at the short end poses a strong puzzle for recent parametrizations of structural long-run risk models, such as Drechsler and Yaron (2011) and Wachter (2013), but less so for models with a time-varying exposure to rare disasters, such as Gabaix (2012). Feunou, Jahan-Parvar and Tedongap (2013) propose a new methodology for modeling and estimating time-varying downside risk and upside uncertainty, motivated by a structural model with disappointment aversion, and focus on the assessment of risk-return tradeoffs in financial markets. Empirically, they find that relative downside risk is compensated through a higher conditional mode of returns and that conditional skewness is a priced risk factor. Feunou, Fontaine, Taamouti and Tedongap (2014) use reduced rank regressions to extract hidden volatility states from a panel of model-free implied variance measures of different maturities, under the assumption of an affine hidden state space dynamics. With this approach, they identify two common volatility factors as linear combinations of the original model-free implied variance measures and they document the resulting joint predictive power for equity, bond and variance excess returns. Our approach is different, as it is based on a new tractable three-factor arbitrage-free model for the whole implied volatility surface, which allows us to jointly estimate the structure of the hidden volatility states, the pricing parameters and variance risk premia in a coherent framework. In this way, the information in the whole implied volatility surface is used to identify three hidden state variables with different persistence features, which span implied volatilities and variance risk premia consistently with the restrictions implied by no arbitrage.

In contrast to this literature, we fully exploit the information in the implied volatility surface to identify three volatility risk factors with distinct persistence properties. We find that the interaction of these risk factors is key to understand the dynamics of the term structure of variance risk premia and is hardly identifiable from variance swap data alone. Moreover, thanks to our specification of interacting factor risk premia, we obtain a more flexible term structure of variance risk premia, both in states of high and low volatilities. Our joint estimation of volatility factors and variance risk premia highlights the essential role of option-implied skewness, (i) as risk premium factor for variance risk premia and the price of the smile and (ii) as indespensable state variable for an accurate representation of the volatility surface.

Our analysis can contribute to understand different structural mechanisms for variance risk premia. While long-run risk models might help to explain the persistent dynamics of the long end of the term structure of variance risk premia, we find that the dynamics in periods of distress may be better explained by high-frequency volatility shocks that do not affect the price of the smile. Such a mechanism may be rationalized by rare disaster models with a time-varying market exposure only weakly correlated with aggregate consumption shocks, as in Gabaix (2012). From a different angle, the multi-frequency variance risk premia in our model are consistent with a price of volatility risk that depends on high frequency shocks in situations of financial distress. Adrian and Rosenberg (2008) decompose market volatility into two weakly persistent components, which are priced in the cross-section of stock returns. They interpret the highest frequency volatility component as a proxy of skewness risk reflecting the tightness of financial constraints. Muir (2013) emphasizes the high-frequency character of financial crises and explains in a theoretical model with financial intermediation why the term structure of variance risk premia can be inverted in phases of financial turmoil. The dynamics of variance risk premia estimated by our model, in particular the high-frequency inverted term structure of variance risk premia in periods of distress, is compatible with the economic intuition in this literature.

IV. Additional Figures

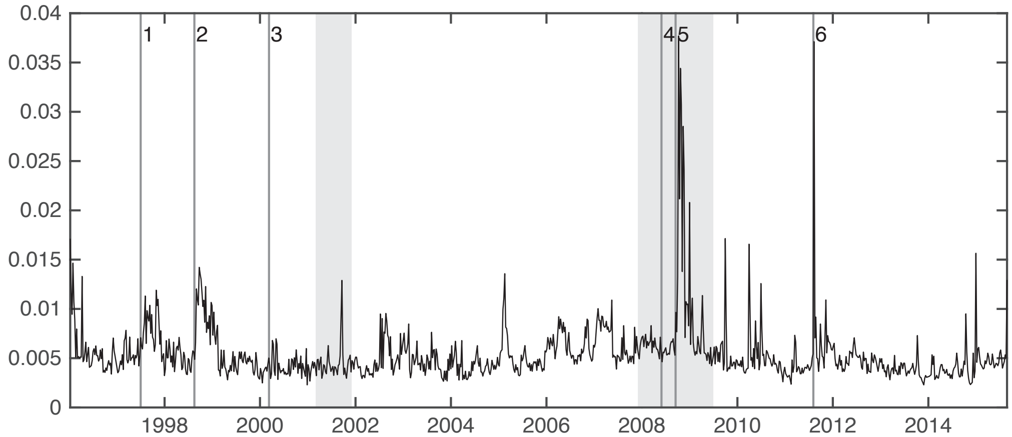
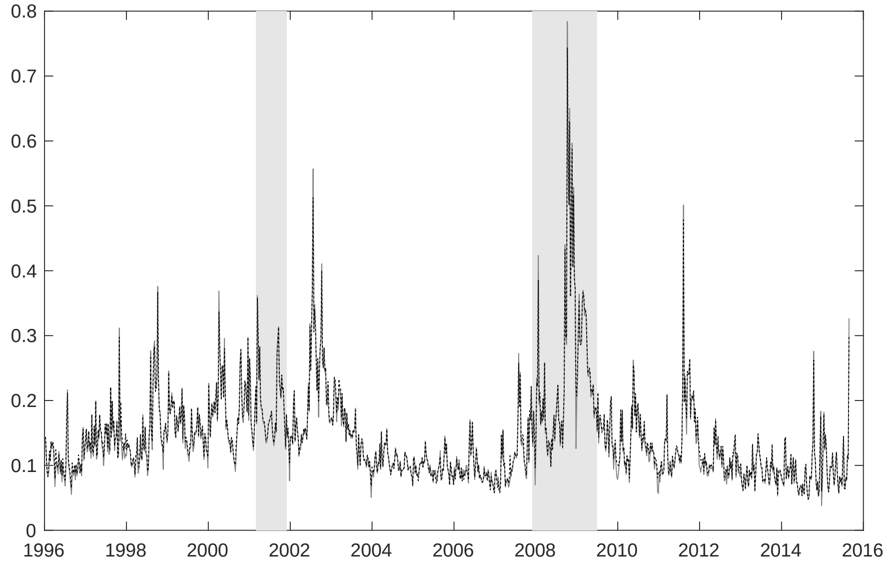


Figure 1: Time series of mean absolute implied volatility errors ($MAIVE$) for our model (GT^2). For every day t in our sample, we plot the $MAIVE$ on that day, defined by $MAIVE_t := \frac{1}{N_t}\sum_{i=1}^{N_t} |IV_i - \widehat{IV}_i|$, where N_t is the number of available options on that day. Grey areas highlight NBER recessions; vertical lines indicate important crisis events as listed in the caption of Figure 2 of the main paper.

Panel A: Model GT^2



Panel B: Model AFT

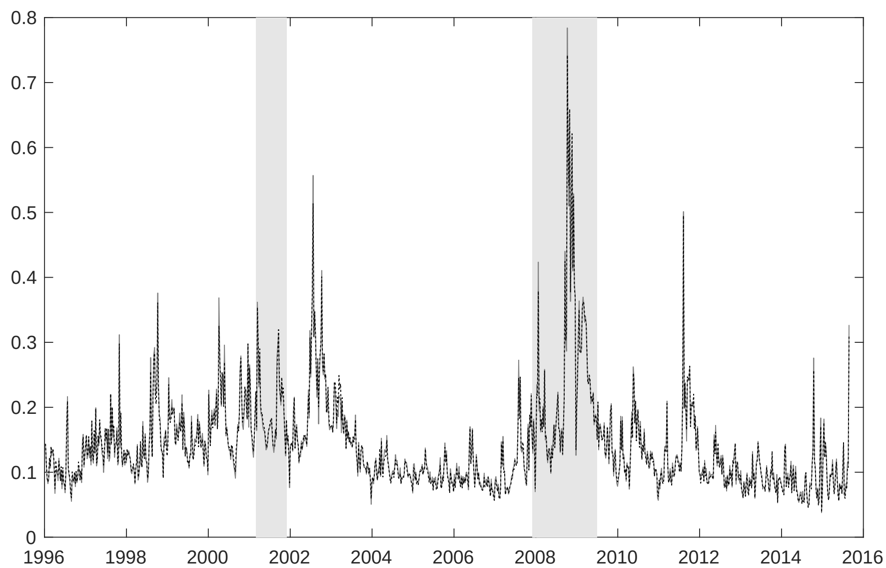
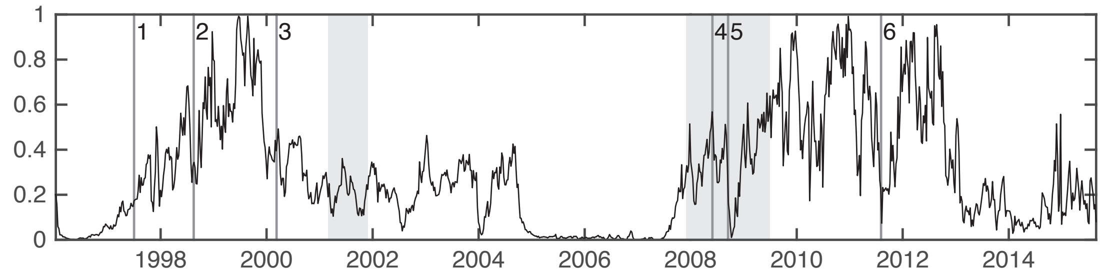
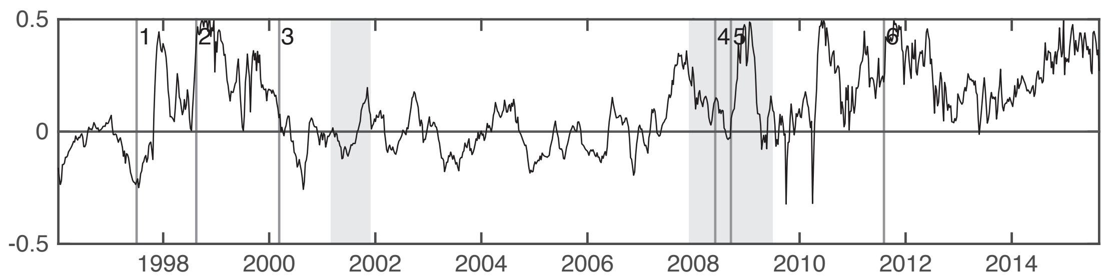


Figure 2: Time series of realized and model-implied spot volatilities. We plot the time series of realized volatilities in the data, computed according to Appendix B of Bollerslev and Todorov (2011) (grey lines), together with the time series of estimated model-implied spot volatilities in models AFT and model GT^2 (black dashed lines). Model-implied spot volatilities are obtained using hidden states and model parameters estimated with the penalized nonlinear least squares approach of Section II.*F.* in the main paper.

Panel A: Volatility factor $X_{11t}/tr(X_t)$



Panel B: Volatility factor $X_{12t}/tr(X_t)$



Panel C: Diffusive variance $tr(X_t) := X_{11t} + X_{12t}$

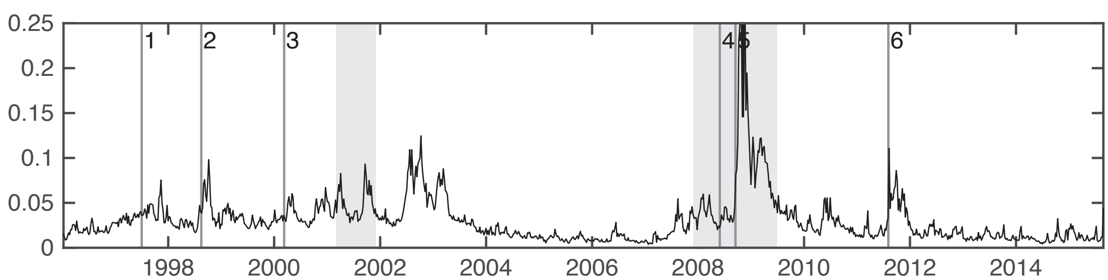


Figure 3: Scaled volatility factors $X_{11t}/tr(X_t), X_{12t}/tr(X_t)$ and diffusive variance $tr(X_t) := X_{11t} + X_{12t}$. Grey areas highlight NBER recessions; vertical lines indicate the crisis events as indicated in the caption to Figure 2.

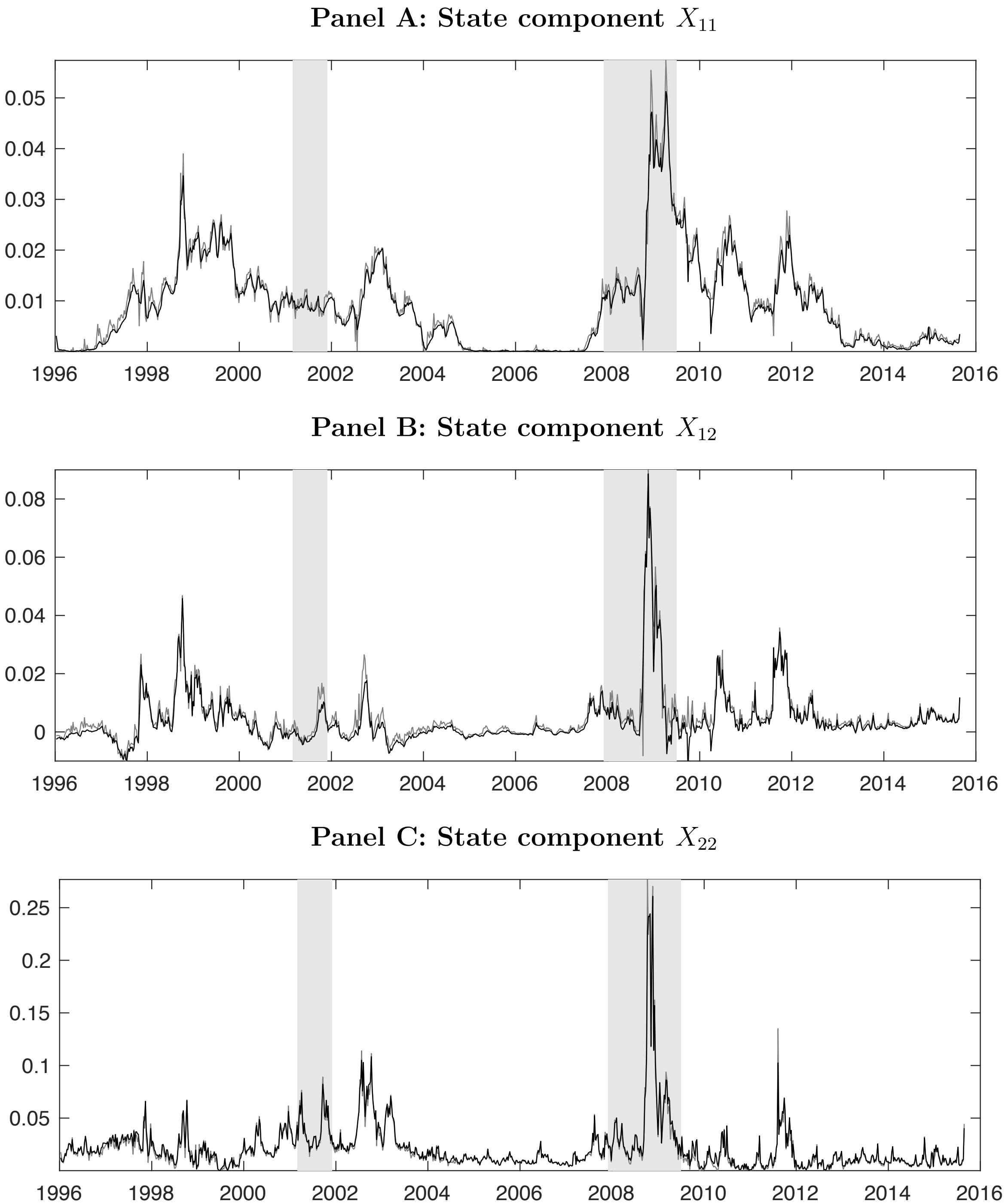


Figure 4: Comparison of state recovery methods. Black line: state obtained from the extended Kalman Filter. Grey line: state recovered using nonlinear least squares (NLLS). For both recovery methods, we have used the same parameters, as specified in Tab.1. For the NLLS estimation, we have used criterion (35) with $\lambda = 0$.

V. Additional Tables

r	q	**Pure diffusion models**	**Jump-diffusion models**
1	0	SV_{10} ($N = 6$) Heston (1993)	SVJ_{10} ($N = 8$) Bates (1996)
2	0	SV_{20} ($N = 12$) Christoffersen et al. (2009)	SVJ_{20} ($N = 18$) Bates (2000)
3	0	SV_{30} ($N = 18$) (this paper)	SVJ_{30} ($N = 25$) (this paper)
3	1	SV_{31} ($N = 14$) da Fonseca et al. (2008)	GT^2 ($N = 21$) Leippold and Trojani (2008)

Table 1: Models related to Assumption 2. r is the number of model state variables and q the number of skewness components disconnected from volatility. N is the number of model parameters. Our model GT^2 corresponds to the jump diffusion model GT^2 in the table.

Panel A: Summary statistics of the data

	In-sample	Out-of sample	Total
Time frame	1996-2002	2003-2015/08	1996-2015/08
Sampling frequency		weekly	
Trading days T	359	656	1015
Number of observations	37'281	139'982	177'263
Average time to maturity (days)	141.5	115.2	120.7
Average moneyness (S/K)	1.00	0.98	0.98

Panel B: Number of observations by duration and delta

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\Delta < 0.2$	2'523	7'229	5'222	4'880	19'854
$0.2 < \Delta < 0.4$	3'645	11'354	8'890	8'537	32'426
$0.4 < \Delta < 0.6$	3'707	12'575	10'107	9'392	35'781
$0.6 < \Delta < 0.8$	5'220	17'531	13'507	13'029	49'287
$0.8 < \Delta$	4'714	15'271	10'542	9'388	39'915
all	19'809	63'960	48'268	45'226	177'263

Table 2: Main characteristics of our S&P500 option panel. We use out-of the money calls and puts.

Panel A: Diffusion parameters

	SV_{20}	SV_{30}	SV_{31}	SVJ_{20}	SVJ_{30}	GT^2
M_{11}	-0.3121	-0.0844	-1.0716	-0.3242	-0.1231	-0.0079
M_{22}	-5.0719	-5.4283	-4.9213	-4.4564	-4.2041	-2.6808
M_{33}		-1.4410			-0.5517	
M_{21}			-14.3050			1.0265
Q_{11}	0.2370	0.1957	0.0556	0.0903	0.0742	0.0698
Q_{22}	0.4209	0.4498	0.5256	0.4204	0.2853	0.2924
Q_{33}		0.0718			0.0738	
Q_{12}			0.1440			-0.0770
R_{11}	-1.0000	-1.0000	-0.0431	-1.0000	-0.9997	-0.2970
R_{22}	-0.5348	-1.0000	-0.6405	-0.3823	-0.7111	-0.4057
R_{33}		0.9633			-0.1178	
R_{12}			0.7672			-0.8708
β_{11}	1.0000	1.0031	1.0000	1.0006	1.0064	1.0012
β_{22}	1.0000	1.0007		1.0000	1.0042	
β_{33}		1.0162			1.0146	
M^*_{11}	-1.4051	-1.2204	-0.6378	-0.7395	-0.8289	-0.5467
M^*_{22}	-1.8593	-2.2558	-2.7528	-1.9462	-1.2661	-2.6808
M^*_{33}		-0.4869			-0.5539	
M^*_{21}			-1.9200			0.3982
β^*_{11}	1.0000	1.0017	1.0000	1.0006	1.0064	1.0012
β^*_{22}	1.0000	1.0046		1.0000	1.0042	
β^*_{33}		1.0693			1.0146	

Panel B: Jump parameters

	SVJ_{20}	SVJ_{30}	GT^2
λ_0	0.0000	0.0003	0.0000
Λ_{11}	43.8971	57.3248	25.6671
Λ_{22}	1.0566	11.9429	15.9795
Λ_{33}		0.0454	
Λ_{12}			40.4278
$\bar{k}$	-0.1500	-0.1500	
δ	0.1500	0.1500	
λ^-			7.1518
λ^+			58.3547

Panel C: Physical jump parameter

		GT^2
β_λ	0.053	0.324

Table 3: In-sample (1996/01-2002/12) parameter estimates. Panel A: diffusion parameters. Panel B: jump parameters. For consistency and for brevity, all parameter values are reported using a notation based on matrix AJD, i.e., by considering Bates- and Heston-type models as nested diagonal matrix AJD models.

Panel A: Model GT^2

M_{11}	-0.036	M_{21}	0.793	M_{22}	-5.654
R_{11}	-0.945	R_{21}	-0.250	R_{22}	-0.241
Q_{11}	0.087	Q_{21}	-0.277	Q_{22}	0.205
β	1.083	λ^{+}	474.621	λ_{-}	7.611
Λ_{11}	68.755	Λ_{21}	75.757	Λ_{22}	20.895

Panel B: Model AFT

ρ_1	-0.896	ρ_2	-0.996	ρ_u	0.025
k_1	19.827	k_2	0.009	k_u	7.964
v_1	0.005	v_2	0.204	μ_u	8.195
σ_1	0.471	σ_2	0.119	c_2^{+}	0.290
μ_{v1}	45.154	c_1^{+}	503.217	c_2^{-}	35.108
c_0^{+}	29.073	c_1^{-}	71.823	c_3^{-}	827.816
λ^{-}	45.411	$\lambda+$	229.933		

Table 4: Parameter estimates obtained from the NLLS estimation for the in-sample period (1996-2002) as described in Section I. *F.*

	X_{11}	X_{12}	X_{22}
Min	0.0000	-0.0135	0.0001
Max	0.0513	0.0886	0.2609
Mean	0.0090	0.0040	0.0217
Median	0.0080	0.0016	0.0154
Positive	1.0000	0.6847	1.0000
Stdv	0.0089	0.0096	0.0245
Skewness	1.4686	3.8837	4.8909
Kurtosis	5.9205	24.4185	39.0020
AR(1)	0.9891	0.9480	0.8843
Half life	1.2118	0.2496	0.1084

Table 5: Summary statistics of weekly filtered volatility factors X_{11}, X_{12} and X_{22} for the entire sample. "Positive" denotes the fraction of positive realizations. Half lives are given in years.

Panel A1: $MAIVE$ **for** SVJ_{20} **model, in-sample**

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\lvert\Delta\rvert < 0.2$	0.934	0.601	0.655	0.595	0.663
$0.2 < \lvert\Delta\rvert < 0.4$	0.781	0.476	0.637	0.631	0.607
$0.4 < \lvert\Delta\rvert < 0.6$	0.693	0.509	0.536	0.515	0.537
$0.6 < \lvert\Delta\rvert < 0.8$	0.810	0.654	0.481	0.547	0.587
$0.8 < \lvert\Delta\rvert$	1.284	1.046	0.806	1.026	1.002
all	0.889	0.660	0.605	0.651	0.669

Panel A2: $MAIVE$ **for** GT^2 **model, in-sample**

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\lvert\Delta\rvert < 0.2$	0.899	0.544	0.488	0.501	0.571
$0.2 < \lvert\Delta\rvert < 0.4$	0.842	0.416	0.479	0.466	0.501
$0.4 < \lvert\Delta\rvert < 0.6$	0.715	0.464	0.553	0.385	0.485
$0.6 < \lvert\Delta\rvert < 0.8$	0.730	0.499	0.486	0.378	0.476
$0.8 < \lvert\Delta\rvert$	0.890	0.565	0.626	0.876	0.722
all	0.804	0.493	0.529	0.502	0.541

Panel B1: $MAIVE$ **for** SVJ_{20} **model, out of sample**

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\lvert\Delta\rvert < 0.2$	0.851	1.201	1.376	1.656	1.309
$0.2 < \lvert\Delta\rvert < 0.4$	0.626	0.631	0.870	1.222	0.839
$0.4 < \lvert\Delta\rvert < 0.6$	0.862	0.439	0.448	0.793	0.569
$0.6 < \lvert\Delta\rvert < 0.8$	1.254	0.753	0.444	0.713	0.711
$0.8 < \lvert\Delta\rvert$	1.360	0.950	0.597	0.878	0.891
all	1.046	0.772	0.661	0.967	0.817

Panel B2: $MAIVE$ **for** GT^2 **model, out of sample**

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\lvert\Delta\rvert < 0.2$	0.690	0.665	0.745	1.022	0.771
$0.2 < \lvert\Delta\rvert < 0.4$	0.640	0.385	0.474	0.592	0.488
$0.4 < \lvert\Delta\rvert < 0.6$	0.764	0.418	0.485	0.396	0.468
$0.6 < \lvert\Delta\rvert < 0.8$	0.863	0.445	0.461	0.538	0.516
$0.8 < \lvert\Delta\rvert$	0.815	0.534	0.468	0.780	0.602
all	0.771	0.477	0.502	0.627	0.551

Table 6: $MAIVE$ stratified by maturity in calendar days and moneyness in absolute Black-Scholes deltas. We report the mean absolute implied volatility error across maturity and moneyness bins for our model (GT^2) and for the benchmark Bates (2000) model (SVJ_{20}), for the in-sample period (1996/01-2002/12) and the out of sample period (2003/01-2015/08).

Panel A1: Fraction of prices within bid/ask spread SVJ_{20} model, in-sample

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\Delta < 0.2$	0.368	0.398	0.320	0.367	0.365
$0.2 < \Delta < 0.4$	0.436	0.534	0.346	0.323	0.401
$0.4 < \Delta < 0.6$	0.610	0.575	0.414	0.324	0.449
$0.6 < \Delta < 0.8$	0.721	0.612	0.552	0.363	0.518
$0.8 < \Delta$	0.706	0.559	0.489	0.224	0.444
all	0.589	0.557	0.448	0.320	0.450

Panel A2: Fraction of prices within bid/ask spread GT^2 model, in-sample

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\Delta < 0.2$	0.371	0.449	0.453	0.422	0.429
$0.2 < \Delta < 0.4$	0.372	0.617	0.446	0.408	0.472
$0.4 < \Delta < 0.6$	0.605	0.598	0.316	0.395	0.454
$0.6 < \Delta < 0.8$	0.771	0.717	0.492	0.488	0.582
$0.8 < \Delta$	0.847	0.848	0.601	0.275	0.589
all	0.616	0.669	0.462	0.405	0.518

Panel B1: Fraction of prices within bid/ask spread SVJ_{20} model, out of sample

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\Delta < 0.2$	0.317	0.207	0.176	0.127	0.194
$0.2 < \Delta < 0.4$	0.445	0.400	0.264	0.211	0.321
$0.4 < \Delta < 0.6$	0.469	0.611	0.527	0.305	0.500
$0.6 < \Delta < 0.8$	0.509	0.504	0.659	0.432	0.531
$0.8 < \Delta$	0.716	0.690	0.782	0.579	0.694
all	0.518	0.518	0.532	0.362	0.486

Panel B2: Fraction of prices within bid/ask spread GT^2 model, out of sample

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\Delta < 0.2$	0.381	0.365	0.349	0.199	0.325
$0.2 < \Delta < 0.4$	0.446	0.627	0.508	0.396	0.518
$0.4 < \Delta < 0.6$	0.521	0.577	0.458	0.514	0.522
$0.6 < \Delta < 0.8$	0.708	0.737	0.599	0.457	0.629
$0.8 < \Delta$	0.902	0.912	0.872	0.594	0.833
all	0.635	0.688	0.586	0.457	0.601

Table 7: Fraction of model-implied option prices within bid-ask spread for the benchmark SVJ_{20} model and our model (GT^2), across maturity and moneyness bins for the in-sample period (1996/01-2002/12) and the out of sample period (2003/01-2015/08).

Panel A: Adjusted R^2

	1mo	2mo	3mo	4mo	5mo	6mo	7mo	8mo	9mo
GT^2									
all	0.04	1.19	2.09	4.94	6.25	5.32	5.12	4.80	4.50
X_{11}, X_{12}	0.04	0.67	0.96	2.22	3.40	3.88	4.14	3.95	3.83
AFTs									
all	0.28	1.53	3.67	6.03	7.62	6.89	6.26	6.38	6.82
U_t	0.07	0.72	1.31	2.07	3.26	4.21	4.37	4.32	4.42

Panel B: Factor loadings

	1mo	2mo	3mo	4mo	5mo	6mo	7mo	8mo	9mo
GT^2 (all)									
const	0.014	0.021	0.026	0.024	0.020	0.016	0.014	0.016	0.017
	(0.78)	(1.64)	(2.63)	(2.83)	(2.46)	(2.09)	(2.07)	(2.50)	(2.73)
X_{11}	2.42	2.10	2.01	3.00	2.93	2.05	1.81	1.50	1.17
	(1.30)	(1.60)	(2.03)	(3.44)	(3.63)	(2.61)	(2.54)	(2.23)	(1.87)
X_{12}	0.49	2.78	2.93	3.27	3.71	3.89	3.59	3.40	3.32
	(0.20)	(1.79)	(1.97)	(2.95)	(4.12)	(4.69)	(4.70)	(4.99)	(5.27)
X_{22}	-0.63	-1.10	-1.27	-1.68	-1.57	-1.07	-0.84	-0.74	-0.64
	(-0.68)	(-2.02)	(-3.31)	(-4.77)	(-4.73)	(-4.38)	(-4.06)	(-3.78)	(-3.35)
GT^2 (only X_{11}, X_{12})									
const	0.004	0.003	0.006	-0.002	-0.005	-0.001	0.001	0.004	0.007
	(0.27)	(0.30)	(0.70)	(-0.27)	(-0.73)	(-0.22)	(0.14)	(0.78)	(1.25)
X_{11}	2.42	2.10	2.01	3.00	2.93	2.05	1.81	1.49	1.17
	(1.30)	(1.63)	(2.08)	(3.53)	(3.78)	(2.73)	(2.66)	(2.34)	(1.96)
X_{12}	-0.45	1.13	1.04	0.76	1.36	2.28	2.34	2.30	2.36
	(-0.22)	(0.92)	(0.77)	(0.66)	(1.66)	(3.31)	(3.76)	(4.22)	(5.10)
AFT (all)									
const	0.006	0.015	0.025	0.020	0.009	0.012	0.013	0.018	0.021
	(0.25)	(0.98)	(2.10)	(1.83)	(0.95)	(1.26)	(1.52)	(2.18)	(2.57)
$V_{1,t}$	-0.81	-0.81	-0.99	-1.17	-1.16	-0.77	-0.55	-0.48	-0.46
	(-1.37)	(-2.65)	(-3.41)	(-3.74)	(-6.28)	(-4.50)	(-3.45)	(-3.38)	(-3.74)
$V_{2,t}$	-0.34	-2.03	-3.24	-3.06	-2.48	-2.81	-2.58	-2.76	-2.95
	(-0.16)	(-1.36)	(-2.76)	(-2.88)	(-2.58)	(-2.98)	(-3.03)	(-3.42)	(-3.75)
U_t	2.54	3.17	3.58	4.06	4.35	3.87	3.37	3.09	2.99
	(1.55)	(2.67)	(3.86)	(5.30)	(8.12)	(7.82)	(7.82)	(8.23)	(8.75)
AFT (only U_t)									
const	0.008	0.001	0.001	-0.002	-0.007	-0.010	-0.008	-0.005	-0.005
	(0.43)	(0.10)	(0.08)	(-0.17)	(-0.89)	(-1.42)	(-1.25)	(-0.93)	(-0.87)
U_t	1.19	1.84	1.98	2.15	2.44	2.63	2.50	2.34	2.27
	(0.87)	(1.92)	(2.37)	(3.25)	(5.10)	(6.72)	(7.32)	(7.99)	(8.87)

Table 8: Predictive regressions. We regress index returns on the states of models GT^2 and AFT for horizons from 1 to 9 months. The regressions are performed separately for each horizon and over the full sample (1996-2015/08). The label "all" corresponds to regression with all estimated states as predictors. The labels X_{11}, X_{12} (U_t) for GT^2 (AFT) correspond to the constrained affine specification without the high-frequency volatility factor. Panel A: Adjusted R^2s. Panel B: Factor loadings and t-statistics in brackets.

References

Adrian, Tobias and Joshua Rosenberg, "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *The Journal of Finance*, 2008, *63* (6), 2997–3030.

Ait-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, "The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis," *SSRN Working Paper*, 2012.

Andersen, Torben, Nicola Fusari, and Viktor Todorov, "Parametric Inference and Dynamic State Recovery from Option Panels," *Econometrica*, 2015, *83* (3), 1081–1145.

___ , ___ , and ___ , "The Risk Premia Embedded in Index Options," *Journal of Financial Economics*, 2015, *117* (3), 558–584.

Bakshi, Gurdip and Nikunj Kapadia, "Delta-Hedged Gains and the Negative Market Volatility Risk Premium," *Review of Financial Studies*, 2003, *16(2)*, 527–566.

Bardgett, C., E. Gourier, and M. Leippold, "Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets.," *Journal of Financial Economics*, 2019, *131* (3), 593–618.

Bates, David, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 2000, *94* (1), 181–238.

Bollerslev, Tim and Viktor Todorov, "Tails, Fears, and Risk Premia," *The Journal of Finance*, 2011, *66* (6), 2165–2211.

Bondarenko, Oleg, "Variance Trading and Market Price of Variance Risk," *Journal of Econometrics*, 2014, *180* (1), 81–97.

Britten-Jones, Mark and Anthony Neuberger, "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 2000, *55*, 839–866.

Buraschi, Andrea and Jens C. Jackwerth, "The Price of a Smile: Hedging and Spanning in Option Markets," *Review of Financial Studies*, 2001, *14*, 495–527.

___ , Paolo Porchia, and Fabio Trojani, "Correlation Risk and Optimal Portfolio Choice," *Journal of Finance*, 2010, *65* (1), 393–420.

Carr, Peter and Dilip B. Madan, "Towards a Theory of Volatility Trading," *in R. Jarrow (ed) Risk Book on Volatility. New York: Risk, pp 417-27*, 1998.

—— **and Liuren Wu**, "Variance Risk Premiums," *Review of Financial Studies*, 2009, *22* (3), 1311–1341.

—— **and** —— , "Leverage Effect, Volatility Feedback, and Self-Exciting Market Disruptions: Disentangling the Multi-dimensional Variations in S&P500 Index Options," *Journal of Financial and Quantitative Analysis*, 2017, *52* (5), 2119–2156.

Christoffersen, Peter, Steve Heston, and Kris Jacobs, "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well," *Management Science*, 2009, *55*, 1914–1932.

Constantinides, George M. and Lei Lian, "The Supply and Demand of S&P 500 Put Options," *NBER Working Paper No. 21161*, 2015.

Cont, Rama and Sana Ben Hamida, "Recovering volatility from option prices by evolutionary optimization," *Journal of Computational Finance, Vol. 8, No. 4, Summer 2005*, 2005, *8 (4).*

Cuchiero, C., D. Filipovic, E. Mayerhofer, and J. Teichmann, "Affine processes on positive semidefinite matrices," *The Annals of Applied Probability*, 2011, *21(2)*, 397–463.

—— **, M. Keller-Ressel, E. Mayerhofer, and J. Teichmann**, "Affine processes on symmetric cones," *Journal of Theoretical Probability*, 2016, *29* (2), 359–422.

da Fonseca, Jose, Martino Grasselli, and Claudio Tebaldi, "A Multifactor Volatility Heston Model," *Quantitative Finance*, 2008, *8 (6)*, 591–604.

Demeterfi, Kresimir, Emanuel Derman, Michael Kamal, and Joseph Zou, "A Guide to Volatility and Variance Swaps," *The Journal of Derivatives*, 1999, *6(4)*, 9–32.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, "The Price of Variance Risk," *Journal of Financial Economics, f*, 2016.

Drechsler, Itamar and Amir Yaron, "What's Vol Got to Do With It," *Review of Financial Studies*, 2011, *24* (1), 1–45.

Dupire, Bruno, "Model Art," *Risk*, 1993, *September*, 118–120.

Feunou, Bruno, Jean-Sebastien Fontaine, Abderrahim Taamouti, and Romeo Tedongap, "Risk Premium, Variance Premium and the Maturity Structure of Uncertainty," *Review of Finance*, 2014, *18*, 219–269.

——, **Mohammad R. Jahan-Parvar, and Romeo Tedongap**, "Modeling Market Downside Volatility," *Review of Finance*, 2013, *17* (1), 443–481.

Filipovic, Damir, Elise Gourier, and Loriano Mancini, "Quadratic Variance Swap Models," *Journal of Financial Economics*, 2016, *119* (1), 44–68.

Gabaix, Xavier, "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *The Quarterly Journal of Economics*, 2012, *127* (2), pp. 645–700.

Gourieroux, Christian, "Continuous Time Wishart Process for Stochastic Risk," *Econometric Reviews*, 2006, *25*, 177–217.

Gruber, Peter H., Claudio Tebaldi, and Fabio Trojani, "Three Make a Smile – Dynamic Volatility, Skewness and Term Structure Components in Option Valuation," *Working Paper. University of Lugano and Bocconi University*, 2010.

Heston, Steven L., "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 1993, *6*, 327–343.

Kozhan, Roman, Anthony Neuberger, and Paul G. Schneider, "The Skew Risk Premium in the Equity Index Market," *Review of Financial Studies*, 2013, *26*, 2174–2203.

Leippold, Markus and Fabio Trojani, "Asset Pricing with Matrix Jump Diffusions," *SSRN eLibrary paper no. 1274482*, 2008.

Li, Junye and Gabriele Zinna, "The Variance Risk Premium: Components, Term Structures, and Stock Return Predictability," *SSRN working paper 2445828*, 2016.

Martin, Ian, "Simple Variance Swaps," *NBER Working Paper Series 16884*, March 2012.

Mayerhofer, Eberhard, "Wishart processes and Wishart Distributios: An affine Process Point of View," *Lecture Notes*, 2014.

Muir, Tyler, "Financial Crises, Risk Premia, and the Term Structure of Risky Assets," *Job Market Paper*, 2013.

Neuberger, Anthony, "The Log Contract," *Journal of Portfolio Management*, 1994, *20* (2), 74–80.

——, "Realized Skewness," *Review of Financial Studies*, 2012, *25* (11), 3423–3455.

Schneider, Paul and Fabio Trojani, "Fear Trading," *SSRN Working Paper 1994454*, 2014.

___ **and** ___ , "Divergence And The Price Of Uncertainty," *Journal of Financial Econometrics*, 2019, *17* (3), 341396.

Storn, Rainer and Kenneth Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous," *Journal of Global Optimization*, 1997, *11*, 341–359.

Todorov, Viktor, "Variance Risk-Premium Dynamics: The Role of Jumps," *Review of Financial Studies*, 2010, *23* (1), 345–383.

Wachter, Jessica A., "Can Time-Varying Risk of Rare Disasters explain Aggregate Stock Market Volatility?," *Journal of Finance*, 2013, *68*, 987–1035.