

ONLINE APPENDIX TO

Quadratic Variance Swap Models

Damir Filipović*
Swiss Finance Institute
and EPFL

Elise Gourier†
ORFE
Princeton University

Loriano Mancini‡
Swiss Finance Institute
and EPFL

This version: 19 March 2015

First draft: April 2013

Journal of Financial Economics, forthcoming

This appendix provides technical derivation and proofs.

***Corresponding author:** Damir Filipović, Swiss Finance Institute at EPFL, Quartier UNIL-Dorigny, CH-1015 Lausanne, Switzerland. E-mail: damir.filipovic@epfl.ch. Telephone: +41 21 693 0108. Fax: +41 21 693 0110.

†Elise Gourier, Operations Research and Financial Engineering (ORFE), Princeton University, 98 Charlton St, Princeton, NJ 08540, U.S. E-mail: egourier@princeton.edu.

‡Loriano Mancini, Swiss Finance Institute at EPFL, Quartier UNIL-Dorigny, CH-1015 Lausanne, Switzerland. E-mail: loriano.mancini@epfl.ch.

Appendix A. Proof of Theorem 3.2

This section provides the proof of Theorem 3.2 along with some technical results on diffusion processes which form the background of some other proofs in the paper. As at the beginning of Section 3, let X_t be a diffusion process taking values in some state space $\mathcal{X} \subset \mathbb{R}^m$ and satisfying the SDE (8) where W_t is a standard d -dimensional Brownian motion under the risk neutral measure \mathbb{Q} . The following assumption is obviously met by all quadratic processes in this paper.

Assumption A.1. *The drift and dispersion functions $\mu(x)$ and $\Sigma(x)$ are assumed to be continuous maps from \mathcal{X} to \mathbb{R}^m and $\mathbb{R}^{m \times m}$ satisfying the linear growth condition*

$$\|\mu(x)\|^2 + \|\Sigma(x)\|^2 \leq K(1 + \|x\|^2), \quad x \in \mathcal{X}, \quad (\text{A.1})$$

for some finite constant K .

Lemma A.2. *Let $g(x)$ be some C^2 -function on \mathcal{X} , let $k(x)$ be some continuous function bounded from below on \mathcal{X} , and suppose $f(\tau, x)$ is a $C^{1,2}$ -function on $[0, +\infty) \times \mathcal{X}$ whose x -gradient satisfies a polynomial growth condition*

$$\|\nabla_x f(\tau, x)\| \leq K(1 + \|x\|^p), \quad \tau \leq T, \quad x \in \mathcal{X}, \quad (\text{A.2})$$

for some finite constant $K = K(T)$ and some $p \geq 1$, for all finite T .

If $f(\tau, x)$ satisfies the partial differential equation

$$\begin{aligned} \frac{\partial f(\tau, x)}{\partial \tau} &= \sum_{i=1}^m \mu_i(x) \frac{\partial f(\tau, x)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^m \left(\Sigma(x) \Sigma(x)^\top \right)_{ij} \frac{\partial^2 f(\tau, x)}{\partial x_i \partial x_j} - k(x) f(\tau, x), \\ f(0, x) &= g(x), \end{aligned} \quad (\text{A.3})$$

for all $\tau \geq 0$ and $x \in \mathcal{X}$, then

$$f(T - t, X_t) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^T k(X_s) ds} g(X_T) \mid \mathcal{F}_t \right] \quad \text{for all } t \leq T < \infty. \quad (\text{A.4})$$

Proof. Fix some finite T . Itô's formula applied to $M_t = e^{-\int_0^t k(X_s)ds} f(T-t, X_t)$ gives

$$dM_t = e^{-\int_0^t k(X_s)ds} D_t dt + e^{-\int_0^t k(X_s)ds} \nabla_x f(T-t, X_t) \Sigma(X_t) dW_t, \quad (\text{A.5})$$

with drift term given by

$$D_t = -\frac{\partial f}{\partial \tau} + \sum_{i=1}^m \mu_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^m \left(\Sigma \Sigma^\top \right)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} - kf, \quad (\text{A.6})$$

where we omitted the arguments $T-t$ and X_t for simplicity, which vanishes by assumption. Hence M_t is a \mathbb{Q} -local martingale with $M_T = g(X_T)$. It remains to be shown that M_t is a true \mathbb{Q} -martingale. Assumption (A.2) implies

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T e^{-2\int_0^t k(X_s)ds} \|\nabla_x f(T-s, X_s) \Sigma(X_s)\|^2 ds \right] &\leq K_1 \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \|\nabla_x f(T-s, X_s)\|^2 \|\Sigma(X_s)\|^2 ds \right] \\ &\leq K_2 \left(1 + \mathbb{E}^{\mathbb{Q}} \left[\sup_{s \leq T} \|X_s\|^{2p} \right] \right), \end{aligned} \quad (\text{A.7})$$

for some finite constants K_1, K_2 . Lemma A.3 below now yields the assertion. \square

The following useful lemma follows from Karatzas and Shreve (1991, Problem V.3.15). For the convenience of the reader we provide a self-contained short proof.

Lemma A.3. *The above diffusion process X_t with $X_0 = x \in \mathcal{X}$ satisfies $\mathbb{E}^{\mathbb{Q}} [\sup_{s \leq T} \|X_s\|^{2p}] < \infty$, for all $p \geq 1$ and finite T .*

Proof. Let $n \geq 1$ and define the finite stopping time $T_n = \inf\{t \mid \|X_t\| \geq n\}$. The stopped process $X_t^{T_n} = X_{t \wedge T_n}$ satisfies

$$X_t^{T_n} = x + \int_0^t \mu(X_s^{T_n}) 1_{\{s \leq T_n\}} ds + \int_0^t \Sigma(X_s^{T_n}) 1_{\{s \leq T_n\}} dW_s =: x + D_t + M_t. \quad (\text{A.8})$$

We fix a finite T . In what follows, K_1, K_2, \dots denote some universal finite constants, which only depend on T . First, observe that the linear growth condition (A.1) implies the pathwise inequality

$$\sup_{s \leq t} \|D_s\|^{2p} \leq K_1 \int_0^t \|\mu(X_u^{T_n})\|^{2p} du \leq K_2 \int_0^t \left(1 + \sup_{s \leq u} \|X_s^{T_n}\|^{2p} \right) du. \quad (\text{A.9})$$

Next, the Burkholder–Davis–Gundy inequality, Karatzas and Shreve (1991, Theorem III.3.28), applied to the continuous local martingale M_t , combined with (A.1), yields

$$\mathbb{E}^{\mathbb{Q}} \left[\sup_{s \leq t} \|M_s\|^{2p} \right] \leq K_3 \int_0^t \mathbb{E}^{\mathbb{Q}} \left[\|\Sigma(X_u^{T_n})\|^{2p} \right] du \leq K_4 \int_0^t \left(1 + \mathbb{E}^{\mathbb{Q}} \left[\sup_{s \leq u} \|X_s^{T_n}\|^{2p} \right] \right) du. \quad (\text{A.10})$$

Combining these inequalities, we obtain

$$\mathbb{E}^{\mathbb{Q}} \left[\sup_{s \leq t} \|X_s^{T_n}\|^{2p} \right] \leq K_5 \left(x^{2p} + t + \int_0^t \mathbb{E}^{\mathbb{Q}} \left[\sup_{s \leq u} \|X_s^{T_n}\|^{2p} \right] du \right). \quad (\text{A.11})$$

By dominated convergence, the nonnegative function $[0, T] \ni t \mapsto \mathbb{E}^{\mathbb{Q}} [\sup_{s \leq t} \|X_s^{T_n}\|^{2p}]$ is continuous. Applying Gronwall’s inequality, Karatzas and Shreve (1991, Problem V.2.7), to it yields

$$\mathbb{E}^{\mathbb{Q}} \left[\sup_{s \leq T} \|X_s^{T_n}\|^{2p} \right] \leq K_5 \left(x^{2p} + T + \int_0^T (x^{2p} + u) K_5 e^{K_5(T-u)} du \right). \quad (\text{A.12})$$

The right hand side does not depend on n . Letting $n \rightarrow \infty$, monotone convergence thus proves the claim. \square

Theorem 3.2 now follows from (4) and Lemma A.2 with $f(\tau, x) = \partial G(\tau, x)/\partial \tau$.

Appendix B. X_t is necessarily quadratic

The aim of this section is to show that, under some mild technical conditions, a quadratic term structure of variance swap rates implies that the state process X_t be quadratic. In addition to Assumption A.1 in Appendix A, we assume the following:

Assumption B.1. *The SDE (8) is well posed in \mathcal{X} . That is, for any $x \in \mathcal{X}$ there exists a \mathcal{X} -valued weak solution $X = X^x$ of (8) with $X_0 = x$ which is unique in law. We let X_t be realized on the canonical space of continuous paths $\omega : [0, \infty) \rightarrow \mathcal{X}$. It is well known that in this case X_t has the strong Markov property, e.g., Karatzas and Shreve (1991, Chapter V).*

Assumption B.2. *The \mathbb{Q} -spot variance is given by $v_t^{\mathbb{Q}} = g(X_t)$ for some C^2 -function $g(x)$ on \mathcal{X} .*

Assumption B.3. *The law $\mathbb{Q} = \mathbb{Q}_x$ of the state process $X = X^x$ is risk neutral for any initial state $X_0 = x \in \mathcal{X}$, and the variance swap curve is given by $\text{VS}(t, T) = \frac{1}{T-t} \int_t^T \mathbb{E}^{\mathbb{Q}} [v_s^{\mathbb{Q}} | X_t] ds$.*

Hence $VS(t, T)$ is a function of the prevailing state X_t and term $T - t$. It is well known that, under suitable regularity conditions, this function can be characterized by a partial differential equation. The following lemma makes this explicit.

Lemma B.4. *Suppose $f(\tau, x)$ is a $C^{1,2}$ -function on $[0, +\infty) \times \mathcal{X}$, and let $k(x)$ be some continuous function on \mathcal{X} . Then, under the above assumptions, the converse of Lemma A.2 holds true: validity of (A.4) for all initial states $X_0 = x \in \mathcal{X}$ implies that $f(\tau, x)$ satisfies the partial differential equation (A.3).*

Proof. By assumption, $M_t = e^{-\int_0^t k(X_s) ds} f(T - t, X_t)$ is a \mathbb{Q} -martingale. Hence its drift, given by (A.6), has to vanish a.s. for all $t \leq T < \infty$ and for all initial states $x \in \mathcal{X}$. This is equivalent to (A.3). \square

We are ready to state and prove the converse of Theorem 3.2.

Theorem B.5. *Assume that the variance swap model admits a quadratic term structure. That is, $G(\tau, x)$ in (12) is a quadratic function in x , $G(\tau, x) = \Phi(\tau) + \Psi(\tau)^\top x + x^\top \Pi(\tau) x$, for some C^2 -functions $\Phi : [0, +\infty) \rightarrow \mathbb{R}$, $\Psi : [0, +\infty) \rightarrow \mathbb{R}^m$, and $\Pi : [0, +\infty) \rightarrow \mathbb{S}^m$. Then the \mathbb{Q} -spot variance function is quadratic, $g(x) = \phi + \psi^\top x + x^\top \pi x$, with parameters given by $\phi = d\Phi(0)/d\tau$, $\psi = d\Psi(0)/d\tau$, and $\pi = d\Pi(0)/d\tau$. Moreover, the following holds:*

- (i) *Suppose $\Psi_i(\tau)$ and $\Pi_{ij}(\tau)$, $1 \leq i \leq j \leq m$, are linearly independent functions. Assume the state space \mathcal{X} contains $\{\lambda x \mid x \in O, \lambda \geq 1\}$ for some open set O in \mathbb{R}^m . Then the process X_t is quadratic with drift and diffusion functions of the form (9)–(10). The functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ satisfy the linear ordinary differential equations (13).*
- (ii) *If $\Pi(\tau) \equiv 0$, and if $\Psi_i(\tau)$, $1 \leq i \leq m$, are linearly independent functions, then the drift function of the state process X_t is affine of the form (9). The functions $\Phi(\tau)$ and $\Psi(\tau)$ satisfy the linear ordinary differential equations*

$$\begin{aligned} \frac{d\Phi(\tau)}{d\tau} &= \phi + b^\top \Psi(\tau), & \Phi(0) &= 0, \\ \frac{d\Psi(\tau)}{d\tau} &= \psi + \beta^\top \Psi(\tau), & \Psi(0) &= 0. \end{aligned} \tag{B.1}$$

Proof. Notice that the assumptions of Lemma B.4 are satisfied by the function $f(\tau, x) = \partial G(\tau, x)/\partial \tau$. Moreover, note that by assumption, $g(x) = f(0, x) = \phi + \psi^\top x + x^\top \pi x$ for $\phi = d\Phi(0)/d\tau$,

$\psi = d\Psi(0)/d\tau$, and $\pi = d\Pi(0)/d\tau$. We denote by $c(x) = \Sigma(x)\Sigma(x)^\top$ the diffusion function of X_t . Integrating the partial differential equation (A.3) for $f(\tau, x)$ in τ leads to

$$\begin{aligned}
& \frac{d\Phi(\tau)}{d\tau} - \phi + \left(\frac{d\Psi(\tau)}{d\tau} - \psi \right)^\top x + x^\top \left(\frac{d\Pi(\tau)}{d\tau} - \pi \right) x \\
&= \sum_{i=1}^m \Psi_i(\tau) \mu_i(x) + \sum_{i,j=1}^m \Pi_{ij}(\tau) (\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)) \\
&= \sum_{i=1}^m \Psi_i(\tau) \mu_i(x) + \sum_{i=1}^m \Pi_{ii}(\tau) (2\mu_i(x)x_i + c_{ii}(x)) + 2 \sum_{i<j} \Pi_{ij}(\tau) (\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)),
\end{aligned} \tag{B.2}$$

for all τ and $x \in \mathcal{X}$. On the left hand side of this equation there is quadratic polynomial.

If $\Psi_i(\tau)$ and $\Pi_{ij}(\tau)$, $i \leq j$, are linearly independent, we obtain that $\mu_i(x)$ and $\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)$ are polynomials in x of degree less than or equal two. If, moreover, \mathcal{X} contains $\{\lambda x \mid x \in O, \lambda \geq 1\}$ for some open set O in \mathbb{R}^m then the linear growth condition (A.1) implies that $\mu_i(x)$ is in fact affine in x , that is of the form (9). Plugging this in $\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)$ yields (10). Plugging these expressions back in (B.2), and separating the powers of x , we arrive at the linear ordinary differential equations (13). This proves part (i). Part (ii) follows using a similar argument. \square

Appendix C. Univariate and bivariate quadratic term structures

The functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ for the univariate quadratic model in Section 3.1 satisfy the linear ordinary differential equations

$$\begin{aligned}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b\Psi(\tau) + a\Pi(\tau), & \Phi(0) &= 0, \\
\frac{d\Psi(\tau)}{d\tau} &= \psi + \beta\Psi(\tau) + (2b + \alpha)\Pi(\tau), & \Psi(0) &= 0, \\
\frac{d\Pi(\tau)}{d\tau} &= \pi + (2\beta + A)\Pi(\tau), & \Pi(0) &= 0,
\end{aligned} \tag{C.1}$$

for real parameters ϕ, ψ, π .

The vector- and matrix-valued functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ for the bivariate quadratic model

in Section 3.2 satisfy the linear ordinary differential equations

$$\begin{aligned}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b^\top \Psi(\tau) + a_1 \Pi_{11}(\tau) + a_2 \Pi_{22}(\tau), & \Phi(0) &= 0, \\
\frac{d\Psi(\tau)}{d\tau} &= \begin{pmatrix} \psi \\ 0 \end{pmatrix} + \beta^\top \Psi(\tau) + 2\Pi(\tau) b + \begin{pmatrix} \alpha_1 \Pi_{11}(\tau) \\ \alpha_2 \Pi_{22}(\tau) \end{pmatrix}, & \Psi(0) &= 0, \\
\frac{d\Pi(\tau)}{d\tau} &= \begin{pmatrix} \pi & 0 \\ 0 & 0 \end{pmatrix} + \beta^\top \Pi(\tau) + \Pi(\tau) \beta + \begin{pmatrix} A_1 \Pi_{11}(\tau) & 0 \\ 0 & A_2 \Pi_{22}(\tau) \end{pmatrix}, & \Pi(0) &= 0,
\end{aligned} \tag{C.2}$$

for real parameters ϕ, ψ, π . For the purpose of solving these ordinary differential equations, it is useful to vectorize them by setting $Q(\tau) = (\Phi(\tau), \Psi_1(\tau), \Psi_2(\tau), \Pi_{11}(\tau), \Pi_{12}(\tau), \Pi_{22}(\tau))^\top$. The above system then reads (for $\beta_{21} = 0$):

$$\frac{dQ(\tau)}{d\tau} = \begin{pmatrix} \phi \\ \psi \\ 0 \\ \pi \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & b_1 & b_2 & a_1 & 0 & a_2 \\ 0 & \beta_{11} & \beta_{21} & 2b_1 + \alpha_1 & 2b_2 & 0 \\ 0 & \beta_{12} & \beta_{22} & 0 & 2b_1 & 2b_2 + \alpha_2 \\ 0 & 0 & 0 & 2\beta_{11} + A_1 & 2\beta_{21} & 0 \\ 0 & 0 & 0 & \beta_{12} & \beta_{11} + \beta_{22} & \beta_{21} \\ 0 & 0 & 0 & 0 & 2\beta_{12} & 2\beta_{22} + A_2 \end{pmatrix} Q(\tau), \quad Q(0) = 0. \tag{C.3}$$

Appendix D. Proof of Theorem 3.3

It follows by inspection that the quadratic property is invariant with respect to affine transformations $\mathcal{X} \rightarrow c + \gamma \mathcal{X}$, $x \mapsto c + \gamma x$ of the state variable, for any real parameters c and $\gamma \neq 0$. Indeed, the transformed process $\hat{X}_t = c + \gamma X_t$ is quadratic with drift and diffusion functions

$$\begin{aligned}
\hat{b}(\hat{x}) &= b\gamma - \beta c + \beta \hat{x} \equiv \hat{b} + \hat{\beta} \hat{x}, \\
\hat{a}(\hat{x}) &= a\gamma^2 - \alpha\gamma c + A c^2 + (\alpha\gamma - 2Ac) \hat{x} + A \hat{x}^2 \equiv \hat{a} + \hat{\alpha} \hat{x} + \hat{A} \hat{x}^2.
\end{aligned} \tag{D.1}$$

The discriminant of $\hat{a}(\hat{x})$ satisfies $\hat{D} = \gamma^2 D$. This proves that Classes 1–3 in Theorem 3.3 form equivalence classes with respect to affine transformations of X_t . It remains to be shown that for any class there exists an affine transformation such that the drift and diffusion functions are of the desired form.

Class 1: Assume first that $A > 0$ and $D < 0$. Any affine transformation with $c = \frac{\alpha}{2A}\gamma$ and $\gamma = \pm\sqrt{\frac{4A}{-D}}$ yields $\hat{a}(\hat{x}) = 1 + A\hat{x}^2$. The sign of γ can be chosen such that $\hat{b} = (b - \beta\frac{\alpha}{2A})\gamma$ be nonnegative as desired. Since the diffusion function has no real zeros, the canonical state space is $\hat{\mathcal{X}} = \mathbb{R}$, e.g., Filipović (2009, Lemma 10.11). If $A = \alpha = 0$ and $a > 0$, we set $\gamma = 1/\sqrt{a}$, and note that c can be chosen such that \hat{b} becomes zero.

Class 2: Assume first that $A > 0$ and $D = 0$. Any affine transformation with $c = \frac{\alpha}{2A}\gamma$ yields $\hat{a}(\hat{x}) = A\hat{x}^2$. The factor γ can be chosen such that $\hat{b} = (b - \beta\frac{\alpha}{2A})\gamma$ is either 1 or 0. A standard comparison result for diffusion processes, Karatzas and Shreve (1991, Proposition V.2.18), shows that \hat{X}_t is bounded from below by the positive geometric Brownian motion $dZ_t = \beta Z_t + \sqrt{A}Z_t dW_t$. Hence the canonical state space is $\hat{\mathcal{X}} = (0, +\infty)$. If $A = \alpha = a = 0$, we can chose γ and c so that \hat{b} becomes zero.

Class 3: Assume first that $A > 0$ and $D > 0$. Any affine transformation with $c = \frac{\alpha \pm \sqrt{D}}{2A}\gamma$ and $\gamma = \frac{\pm 1}{\sqrt{D}}$ yields $\hat{a}(\hat{x}) = \hat{x} + A\hat{x}^2$. The sign of γ can be chosen such that $\hat{b} = (b - \beta\frac{\alpha \pm \sqrt{D}}{2A})\gamma$ be nonnegative. Standard stochastic invariance results for diffusion processes, e.g., Filipović (2009, Lemma 10.11), then show that $\hat{X}_t \geq 0$ for all t whenever $\hat{X}_0 \geq 0$. We now claim that $\hat{b} \geq \frac{1}{2}$ is necessary and sufficient for the canonical state space $\hat{\mathcal{X}}$ not to contain 0. Indeed, elementary calculations show that the scale function of \hat{X}_t is

$$p(\hat{x}) = \int_1^{\hat{x}} \left(\frac{(1+A)y}{1+Ay} \right)^{-2\hat{b}} \left(\frac{1+Ay}{1+A} \right)^{-\frac{2\hat{b}}{A}} dy. \quad (\text{D.2})$$

It satisfies $p(\hat{x}) = p(r)\mathbb{P}[\tau_r < \tau_R] + p(R)\mathbb{P}[\tau_r > \tau_R]$ for any $0 \leq r < \hat{X}_0 = \hat{x} < R$, and hitting times defined by $\tau_c = \inf\{t \geq 0 \mid \hat{X}_t = c\}$, see Karatzas and Shreve (1991, Section V.5.C). Since $\tau_R \uparrow \infty$ for $R \uparrow \infty$, it follows that $\mathbb{P}[\tau_0 = \infty] = 1$ if and only if $p(0) = -\infty$, e.g., Filipović (2009, Exercise 10.12). The latter is equivalent to $2\hat{b} \geq 1$, which proves the claim. If $A = 0$ and $\alpha \neq 0$, we set $\gamma = 1/\alpha$, and chose c such that \hat{a} becomes zero. Note that the conditions on \hat{b} hold necessarily if X_t be well defined, e.g., Filipović (2009, Lemma 10.11) and the arguments above. This completes the proof of Theorem 3.3.

Appendix E. Identification of the bivariate quadratic model

The identification of the bivariate quadratic model in Section 3.2 follows from the proof of Theorem 3.3 in Section D. When X_{1t} is of Class 3, the boundary point 0 is not attainable if and only if $b_1 \geq 1/2$. To prove the necessity of this statement assume that $b_1 < 1/2$. Conditioning on $\beta_{12}X_{2t} < 1/2 - b_1$ for all $t \leq 1$, and using a comparison argument for diffusion processes, see Karatzas and Shreve (1991, Section V.2.C), one can show similarly as in the proof of Theorem 3.3, Class 3, that $X_{1t} = 0$ for some $t \leq 1$ with nonzero probability. To prove the sufficiency assume that $b_1 \geq 1/2$. The comparison argument for diffusion processes, along with the arguments for Class 3 in the proof of Theorem 3.3, implies that $X_{1t} > 0$ for all t whenever $X_{10} > 0$.

Appendix F. Proof of Theorem 3.5

Let $0 \leq n \leq N$. In view of Lemma B.4, the n th \mathcal{F}_t -conditional moment function $f_n(\tau, X_t) = \mathbb{E}^{\mathbb{Q}} [X_{t+\tau}^n | \mathcal{F}_t]$ formally solves the partial differential equation

$$\begin{aligned} \frac{\partial}{\partial \tau} f_n(\tau, x) &= \mathcal{A}f_n(\tau, x), \\ f_n(0, x) &= x^n, \end{aligned} \tag{F.1}$$

where $\mathcal{A} = (b + \beta x) \frac{\partial}{\partial x} + \frac{1}{2}(a + \alpha x + Ax^2) \frac{\partial^2}{\partial x^2}$ denotes the infinitesimal generator of the quadratic diffusion X_t . We solve (F.1) by the guess $f_n(\tau, x) = \sum_{k=0}^N M_{kn}(\tau)x^k$, for some $(N+1) \times (N+1)$ matrix valued function $M(\tau) = (M_{kn}(\tau))$. Plugging this guess in (F.1), noting that

$$\mathcal{A}x^k = k(k-1)\frac{a}{2}x^{k-2} + k\left(b + (k-1)\frac{\alpha}{2}\right)x^{k-1} + k\left(\beta + (k-1)\frac{A}{2}\right)x^k, \tag{F.2}$$

and matching coefficients in x , we obtain the $N + 1$ linear systems of $N + 1$ ordinary differential equations

$$\frac{d}{d\tau} \begin{pmatrix} M_{0n}(\tau) \\ M_{1n}(\tau) \\ M_{2n}(\tau) \\ M_{3n}(\tau) \\ \vdots \\ M_{Nn}(\tau) \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & b & 2\frac{\alpha}{2} & 0 & \cdots & 0 \\ 0 & \beta & 2(b + \frac{\alpha}{2}) & 3 \cdot 2\frac{\alpha}{2} & 0 & \vdots \\ 0 & 0 & 2(\beta + \frac{A}{2}) & 3(b + 2\frac{\alpha}{2}) & \ddots & 0 \\ 0 & 0 & 0 & 3(\beta + 2\frac{A}{2}) & \ddots & N(N-1)\frac{\alpha}{2} \\ \vdots & & & 0 & \ddots & N(b + (N-1)\frac{\alpha}{2}) \\ 0 & \dots & & & 0 & N(\beta + (N-1)\frac{A}{2}) \end{pmatrix}}_{=B} \begin{pmatrix} M_{0n}(\tau) \\ M_{1n}(\tau) \\ M_{2n}(\tau) \\ M_{3n}(\tau) \\ \vdots \\ M_{Nn}(\tau) \end{pmatrix}, \quad (\text{F.3})$$

along with the initial condition

$$M_{kn}(0) = \begin{cases} 1, & \text{if } k = n \\ 0, & \text{otherwise.} \end{cases} \quad (\text{F.4})$$

In matrix notation, denote by B the $(N + 1) \times (N + 1)$ matrix in (F.3), the system (F.3)–(F.4) reads

$$\frac{d}{d\tau} M(\tau) = BM(\tau), \quad M(0) = Id, \quad (\text{F.5})$$

where Id is the identity matrix. Its solution is given by the matrix exponential $M(\tau) = e^{B\tau}$. It remains to be verified that this provides indeed the n th \mathcal{F}_t -conditional moments of $X_{t+\tau}$. Clearly, $f_n(\tau, x) = \sum_{k=0}^n (e^{B\tau})_{kn} x^k$ is a $C^{1,2}$ -function whose x -gradient satisfies the polynomial growth condition (A.2). Hence, Theorem 3.5 follows from the above arguments and Lemma A.2, noting that $\text{VS}(t, T) = \frac{1}{T-t} \int_0^{T-t} f(\tau, X_t) d\tau$.

Appendix G. Maximum likelihood estimation and unscented Kalman filter

We provide more details about the model estimation approach used in Section 4.2. We first cast the model in state space form, and then describe the nonlinear unscented Kalman filter.

G.1. The state space form

We cast the model in state space form, which consists of a transition equation and a measurement equation. The transition equation describes the discrete-time dynamics of the state process. The measurement equation describes the relation between the state variable X_t and variance swap rates and quadratic variation at time t . To simplify notation, in this section, the time step $t - 1$ to t is one day. The transition equation is obtained from discretizing the SDE (8) using an Euler scheme at daily frequency, which leads to

$$X_t = \Phi_0 + \Phi_X X_{t-1} + \omega_t, \quad \omega_t \sim N(0, Q_t), \quad (\text{G.1})$$

with Φ_0 , Φ_X , and Q_t given in closed form. The measurement equation is given by

$$Z_t = H(X_t) + u_t, \quad u_t \sim N(0, \Omega_t), \quad (\text{G.2})$$

where Z_t is the six-dimensional observation vector. The first five components are given by the variance swap rates with terms two, three, six months, and 1 and two years. The corresponding components in $H(X_t)$ are the model-based variance swap rates in (12). The associated measurement errors are normally distributed, cross-sectionally uncorrelated, and with constant variance σ_{VS}^2 . The sixth component of the observation vector is the logarithm of the daily quadratic variation, $\log(\text{QV}_t)$. The corresponding component in $H(X_t)$ is $\log(v_t^{\mathbb{P}})$, where we specify the \mathbb{P} -spot variance by $\log(v_t^{\mathbb{P}}) = c_0 + c_1 \log(v_t^{\mathbb{Q}})$, and $v_t^{\mathbb{Q}} = g(X_t)$ is the \mathbb{Q} -spot variance in (11). This specification does not follow the typical modeling of the price jump risk premium in affine models, where $v_t^{\mathbb{P}}$ and $v_t^{\mathbb{Q}}$ would both be affine functions of X_t . The associated measurement error ϵ_t is conditionally normally distributed with mean $\rho_\epsilon \epsilon_{t-1}$ and variance $c_2 + c_3 \text{QV}_{t-1}$. The rationale behind the sixth component of the measurement equation is the following. Andersen, Bollerslev, Diebold, and Ebens (2001), among others, provide empirical evidence that $\log(\text{QV}_t)$ is approximately normally distributed. The conditional mean specification of ϵ_t allows for autocorrelation in the measurement error, which can be induced by clustering of price jumps caused by persistence of the price jump intensity and/or microstructure noise in the estimates of daily quadratic variation. Autocorrelation in the measurement error of similar magnitude is also reported in Wu (2011). The conditional variance

specification of ϵ_t captures in a parsimonious way the heteroskedasticity of the measurement error due to the volatility of quadratic variation.

The daily quadratic variation is computed using tick-by-tick data from the S&P 500 futures and applying the two-scale estimator of Zhang, Mykland, and Ait-Sahalia (2005) with fast and slow time scales given by two-tick and 20-tick, respectively. We experimented also other consistent measures of quadratic variation, such as the two-scale estimator based on 1-tick and 10-tick time scales and the multi-scale estimator of Zhang (2006), and results were virtually unchanged. Model estimates are presented in Table 2.

G.2. The unscented Kalman filter

If the function $H(X_t)$ were linear, i.e., $H(X_t) = H_0 + H_X X_t$, the Kalman filter would provide efficient estimates of the conditional mean and variance of the state vector. Let $\hat{X}_{t|t-1} = E_{t-1}[X_t]$ and $\hat{Z}_{t|t-1} = E_{t-1}[Z_t]$ denote the expectation of X_t and Z_t , respectively, using information up to and including time $t - 1$, and let $P_{t|t-1}$ and $F_{t|t-1}$ denote the corresponding error covariance matrices. Furthermore, let $\hat{X}_t = E_t[X_t]$ denote the expectation of X_t including information at time t , and let P_t denote the corresponding error covariance matrix. The Kalman filter consists of two steps: prediction and update. In the prediction step, $\hat{X}_{t|t-1}$ and $P_{t|t-1}$ are given by

$$\begin{aligned}\hat{X}_{t|t-1} &= \Phi_0 + \Phi_X \hat{X}_{t-1}, \\ P_{t|t-1} &= \Phi_X P_{t-1} \Phi_X^\top + Q_t,\end{aligned}\tag{G.3}$$

and $\hat{Z}_{t|t-1}$ and $F_{t|t-1}$ are in turn given by

$$\hat{Z}_{t|t-1} = H_0 + H_X \hat{X}_{t|t-1},\tag{G.4}$$

$$F_{t|t-1} = H_X P_{t|t-1} H_X^\top + \Omega_t.\tag{G.5}$$

In the update step, the estimate of the state vector is refined based on the difference between observed and predicted quantities, with $\hat{X}_t = E_t[X_t]$ and P_t given by

$$\hat{X}_t = \hat{X}_{t|t-1} + W_t(Z_t - \hat{Z}_{t|t-1}),\tag{G.6}$$

$$P_t = P_{t|t-1} - W_t F_{t|t-1} W_t^\top,\tag{G.7}$$

where $W_t = P_{t|t-1} H_X^\top F_{t|t-1}^{-1}$.

In our setting, the function $H(X_t)$ is nonlinear, and the Kalman filter has to be modified. Non-linear state space systems have traditionally been handled with the extended Kalman filter, which effectively linearizes the measure equation around the predicted state. However, in recent years the unscented Kalman filter has emerged as an attractive alternative. Rather than approximating the measurement equation, it uses the true nonlinear measurement equation and instead approximates the distribution of the state vector with a deterministically chosen set of sample points, called “sigma points” that capture the true mean and covariance of the state vector. When propagated through the nonlinear pricing function, the sigma points capture the mean and covariance of the data accurately to the 2nd order (3rd order for true Gaussian states) for any nonlinearity.

More specifically, a set of $2L + 1$ sigma points and associated weights are selected according to the following scheme

$$\begin{aligned} \hat{\chi}_{t|t-1}^0 &= \hat{X}_{t|t-1}, & w^0 &= \frac{\kappa}{L+\kappa}, \\ \hat{\chi}_{t|t-1}^i &= \hat{X}_{t|t-1} + \left(\sqrt{(L+\kappa)P_{t|t-1}} \right)_i, & w^i &= \frac{1}{2(L+\kappa)}, \quad i = 1, \dots, L, \\ \hat{\chi}_{t|t-1}^i &= \hat{X}_{t|t-1} - \left(\sqrt{(L+\kappa)P_{t|t-1}} \right)_i, & w^i &= \frac{1}{2(L+\kappa)}, \quad i = L+1, \dots, 2L, \end{aligned} \quad (\text{G.8})$$

where L is the dimension of $\hat{X}_{t|t-1}$, κ is a scaling parameter, w^i is the weight associated with the i -th sigma point, and $\left(\sqrt{(L+\kappa)P_{t|t-1}} \right)_i$ is the i -th column of the matrix square root. Then, in the prediction step, (G.4) and (G.5) are replaced by

$$\begin{aligned} \hat{Z}_{t|t-1} &= \sum_{i=0}^{2L} w^i H(\hat{\chi}_{t|t-1}^i), \\ F_{t|t-1} &= \sum_{i=0}^{2L} w^i (H(\hat{\chi}_{t|t-1}^i) - \hat{Z}_{t|t-1})(H(\hat{\chi}_{t|t-1}^i) - \hat{Z}_{t|t-1})^\top + \Omega_t. \end{aligned} \quad (\text{G.9})$$

The update step is still given by (G.6) and (G.7), but with W_t computed as

$$W_t = \sum_{i=0}^{2L} w^i (\hat{\chi}_{t|t-1}^i - \hat{X}_{t|t-1})(H(\hat{\chi}_{t|t-1}^i) - \hat{Z}_{t|t-1})^\top F_{t|t-1}^{-1}. \quad (\text{G.10})$$

Finally, the log-likelihood function is given by

$$\sum_{t=1}^{\mathcal{N}} -\frac{1}{2} \left[6 \log(2\pi) + \log |F_{t|t-1}| + (Z_t - \hat{Z}_{t|t-1})^\top F_{t|t-1}^{-1} (Z_t - \hat{Z}_{t|t-1}) \right], \quad (\text{G.11})$$

where $\mathcal{N} = 2,832$ is the sample size of daily observations.

Appendix H. Proof of Theorem 5.2

We first list the technical assumptions that will enable us to prove Theorem 5.2.

Assumption H.1. *The functions $\mu(x)$, $\Sigma(x)$, $\Lambda(x)$, $\nu^{\mathbb{P}}(x)$, $\nu^{\mathbb{Q}}(x)$, and*

$$k(x) = -\frac{1}{2\eta} \left(\frac{1}{\eta} - 1 \right) \|\Lambda(x)\|^2 - \nu^{\mathbb{Q}}(x) \left(\left(\frac{\nu^{\mathbb{P}}(x)}{\nu^{\mathbb{Q}}(x)} \right)^{\frac{1}{\eta}} - 1 \right) - \frac{1}{\eta} \left(\nu^{\mathbb{Q}}(x) - \nu^{\mathbb{P}}(x) \right) \quad (\text{H.1})$$

are continuous on \mathcal{X} .

We note that $k(x) \equiv 0$ if $\eta = 1$, and $k(x) \geq 0$ for all $x \in \mathcal{X}$ if $\eta \geq 1$.

Assumption H.2. *The SDE*

$$d\hat{X}_t = \left(\mu(\hat{X}_t) + \frac{1}{\eta} \Sigma(\hat{X}_t) \Lambda(\hat{X}_t) \right) dt + \Sigma(\hat{X}_t) dW_t \quad (\text{H.2})$$

is well posed in \mathcal{X} (see Assumption B.1), and the function

$$h(\tau, x) = \log \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_0^\tau k(\hat{X}_s) ds} \mid \hat{X}_0 = x \right] \quad (\text{H.3})$$

is of class $C^{1,2}$ on $[0, T] \times \mathcal{X}$, where $k(x)$ is given in (H.1).

Lemma B.4 implies that $H(\tau, x) = e^{h(\tau, x)}$ satisfies the linear PDE

$$\frac{\partial H(\tau, x)}{\partial \tau} = \frac{1}{2} \sum_{i,j=1}^m \left(\Sigma(x) \Sigma(x)^\top \right)_{ij} \frac{\partial^2 H(\tau, x)}{\partial x_i \partial x_j} + \left(\mu(x) + \frac{1}{\eta} \Sigma(x) \Lambda(x) \right)^\top \nabla_x H(\tau, x) - k(x) H(\tau, x), \quad (\text{H.4})$$

along with the initial condition $H(0, x) = 1$. Hence $h(\tau, x)$ satisfies the nonlinear PDE

$$\begin{aligned} \frac{\partial h(\tau, x)}{\partial \tau} = & \frac{1}{2} \sum_{i,j=1}^m \left(\Sigma(x) \Sigma(x)^\top \right)_{ij} \frac{\partial^2 h(\tau, x)}{\partial x_i \partial x_j} + \left(\mu(x) + \frac{1}{\eta} \Sigma(x) \Lambda(x) \right)^\top \nabla_x h(\tau, x) \\ & + \frac{1}{2} \nabla_x h(\tau, x)^\top \Sigma(x) \Sigma(x)^\top \nabla_x h(\tau, x) - k(x), \end{aligned} \quad (\text{H.5})$$

along with the initial condition $h(0, x) = 0$.

First order conditions in the HJB equation (41) yield the optimal control law

$$\begin{aligned} \theta^{W^*}(t, v, x) &= \frac{-\frac{\partial J}{\partial v} \Lambda(x) - \Sigma(x)^\top \nabla_x \left(\frac{\partial J}{\partial v} \right)}{\frac{\partial^2 J}{\partial v^2} v}, \\ \theta^{N^*}(t, v, x) &= \frac{1}{\xi} \left[\left(\frac{\nu^{\mathbb{P}}(x) J(1 - \eta)}{\nu^{\mathbb{Q}}(x) \frac{\partial J}{\partial v} v} \right)^{\frac{1}{\eta}} - 1 \right]. \end{aligned} \quad (\text{H.6})$$

Guessing the functional form $J(t, v, x) = e^{\eta h(T-t, x)} \frac{(e^{-r(T-t)} v)^{1-\eta}}{1-\eta}$ allows recovering (42). Straight-forward verification of the HJB equation (41) then completes the proof of Theorem 5.2.

Appendix I. Computation of the intertemporal hedging demand

We now discuss the computation of the intertemporal hedging term $\nabla_x h(\tau, x)$ in (42). In view of (H.1) and (H.3), the intertemporal hedging demand is zero, $\nabla_x h(\tau, x) = 0$, if $k(x) \equiv c$ is constant. This occurs in the myopic logarithmic utility case, $\eta = 1$, or in the absence of jump risk premium, $\nu^{\mathbb{Q}}(x) = \nu^{\mathbb{P}}(x)$, and constant norm $\|\Lambda(x)\|$ of the diffusive market price of risk. In general $\nabla_x h(\tau, x)$ needs to be computed numerically. We assume that

$$k(x) = c + \epsilon P(x), \quad (\text{I.1})$$

for some constant c , some continuous function $P(x)$, and some ϵ with small absolute value. The first order expansion of $\nabla_x h(\tau, x) = \nabla_x h(\tau, x, \epsilon)$ around $\epsilon = 0$ is

$$\begin{aligned} \nabla_x h(\tau, x, \epsilon) &= \nabla_x h(\tau, x, 0) + \nabla_x \partial_\epsilon h(\tau, x, 0) \epsilon + o(\epsilon) \\ &= -\nabla_x \int_0^\tau \mathbb{E}^{\mathbb{Q}} \left[\epsilon P(\widehat{X}_s) \mid \widehat{X}_0 = x \right] ds + o(\epsilon). \end{aligned} \quad (\text{I.2})$$

If the diffusion \widehat{X}_t is quadratic and $P(x)$ is a polynomial then the conditional moments on the right hand side of (I.2) are available in closed form. Therefore, closed form expressions for $\nabla_x h(\tau, x, \epsilon)$ are available for Taylor expansions of arbitrary order in ϵ .

Appendix J. Proof of Corollary 5.3

Taking account of the factorization (36), we can rewrite (35) as

$$\theta_t^W = \Sigma(X_t)^\top \left(\mathcal{D}_t \mathbf{n}_t + \frac{\nabla_x O_t}{O_{t-}} \phi_t \right) + \sigma(X_t) \mathbf{R}(X_t) \left(w_t + \frac{\partial_s O_t S_t}{O_{t-}} \phi_t \right), \quad (\text{J.1})$$

$$\theta_t^N = w_t + \frac{\Delta O_t}{\xi O_{t-}} \phi_t. \quad (\text{J.2})$$

As stated below Assumption 5.1, the $d \times d$ matrix $(\Sigma(X_t)^\top, \sigma(X_t) \mathbf{R}(X_t))$ is invertible $dt \otimes d\mathbb{Q}$ -a.s. Hence there exists a random vector $\mathbf{v} = \mathbf{v}(X_t)$ with $\mathbf{v}^\top \Sigma(X_t)^\top = 0$ and $\mathbf{v}^\top \mathbf{R}(X_t) \sigma(X_t) \neq 0$ $dt \otimes d\mathbb{Q}$ -a.s. Projecting both sides of (J.1) onto \mathbf{v} , and using the optimal $\theta_t^W = \theta_t^{W*}$ and $\theta_t^N = \theta_t^{N*}$ in (42), we obtain from (J.1)–(J.2) the linear equations

$$\begin{aligned} \frac{1}{\eta} \mathbf{v}^\top \Lambda(X_t) &= \mathbf{v}^\top \mathbf{R}(X_t) \sigma(X_t) \left(w_t + \frac{\partial_s O_t S_t}{O_{t-}} \phi_t \right), \\ \frac{1}{\xi} \left(\left(\frac{\nu^{\mathbb{P}}(X_t)}{\nu^{\mathbb{Q}}(X_t)} \right)^{1/\eta} - 1 \right) &= w_t + \frac{\Delta O_t}{\xi O_{t-}} \phi_t, \end{aligned} \quad (\text{J.3})$$

for w_t and ϕ_t hold $dt \otimes d\mathbb{Q}$ -a.s. Hence the solution $w_t^* = w_t$ and $\phi_t^* = \phi_t$ of (35) for the optimal θ_t^{W*} and θ_t^{N*} in (42) is fully determined by the myopic term and does not depend on the choice of the variance swaps. This proves Corollary 5.3.

Appendix K. Computation of option price and sensitivities

For a fixed time $t_0 \geq 0$, from (31) the log price process $L_t = \log S_t$, for $t \geq t_0$, can be decomposed into $L_t = Y_t + X_{3t}$, where

$$Y_t = \int_{t_0}^t \left((\log(1 + \xi) - \xi) \nu^{\mathbb{Q}}(X_s) - \frac{1}{2} \sigma(X_s)^2 \right) ds = K_Y \int_{t_0}^t g(X_s) ds, \quad (\text{K.1})$$

with $K_Y = -\frac{(\xi - \log(1+\xi))\nu^Q + 1/2}{1 + (\log(1+\xi))^2\nu^Q} < 0$, as in (46), and X_{3t} has the \mathbb{Q} -dynamics, for $t \geq t_0$,

$$dX_{3t} = r dt + \sigma(X_t)\mathbf{R}(X_t)^\top dW_t + \log(1 + \xi)(dN_t - \nu^{\mathbb{Q}}(X_t)dt), \quad X_{3t_0} = \log S_{t_0}. \quad (\text{K.2})$$

With a slight abuse of notation we henceforth denote the three-dimensional jump-diffusion process $X_t = (X_{1t}, X_{2t}, X_{3t})^\top$. Its diffusion matrix $A(x)$ is

$$A(x) = \begin{pmatrix} 1 + A_1x_1^2 & 0 & R_1(x)\sigma(x)\sqrt{1 + A_1x_1^2} \\ 0 & x_2 + A_2x_2^2 & 0 \\ R_1(x)\sigma(x)\sqrt{1 + A_1x_1^2} & 0 & \sigma(x)^2 \end{pmatrix}. \quad (\text{K.3})$$

For X_t to be a quadratic jump-diffusion process, the cross-term needs to be a quadratic function of x . We thus aim to find suitable coefficients q_0 , q_1 and q_2 such that

$$R_1(x)\sigma(x)\sqrt{1 + A_1x_1^2} = R_1(x_1)\sqrt{K_{\sigma^2}g(x_1)}\sqrt{1 + A_1x_1^2} = q_0 + q_1x_1 + q_2x_1^2, \quad (\text{K.4})$$

where we define $K_{\sigma^2} = (1 + (\log(1 + \xi))^2\nu^Q)^{-1} > 0$ as in (46). To capture the leverage effect in (47), $R_1(X_t)$ should be approximately equal to $-\text{sign}(\psi + 2\pi X_{1t}) \times 0.7$ in the empirical range of X_{1t} , which is shown in Fig. 2. Hence we choose q_0 , q_1 and q_2 such that

$$(q_0 + q_1x_1 + q_2x_1^2)^2 \approx 0.7^2 K_{\sigma^2} g(x_1)(1 + A_1x_1^2), \quad (\text{K.5})$$

so that the three highest order terms, x_1^2 , x_1^3 and x_1^4 , match. This is equivalent to setting

$$q_2^2 = K_{\sigma^2} 0.7^2 \pi A_1, \quad 2q_1q_2 = K_{\sigma^2} 0.7^2 \psi A_1, \quad 2q_0q_2 + q_1^2 = K_{\sigma^2} 0.7^2 (\pi + \phi A_1). \quad (\text{K.6})$$

The coefficient q_2 is chosen to be the negative root as we want $R_1(X_t) \leq 0$ in the empirical range of X_{1t} . Hence

$$q_2 = -0.7\sqrt{K_{\sigma^2}}\sqrt{\pi A_1}, \quad q_1 = -0.7\sqrt{K_{\sigma^2}}\frac{\psi\sqrt{A_1}}{2\sqrt{\pi}}, \quad q_0 = -0.7\sqrt{K_{\sigma^2}}\frac{\pi + \phi A_1 - \psi^2 A_1/(4\pi)}{2\sqrt{\pi A_1}}. \quad (\text{K.7})$$

The function $R_1(x_1)$ is then specified by (K.4) and (K.7). Fig. 1 shows that, for most values x_1 in the empirical range of X_{1t} , $R_1(x_1)$ is between -0.8 and -0.7 . Inspection shows that $R_1(x_1)^2 < 1$

for all $x_1 \in \mathbb{R}$.

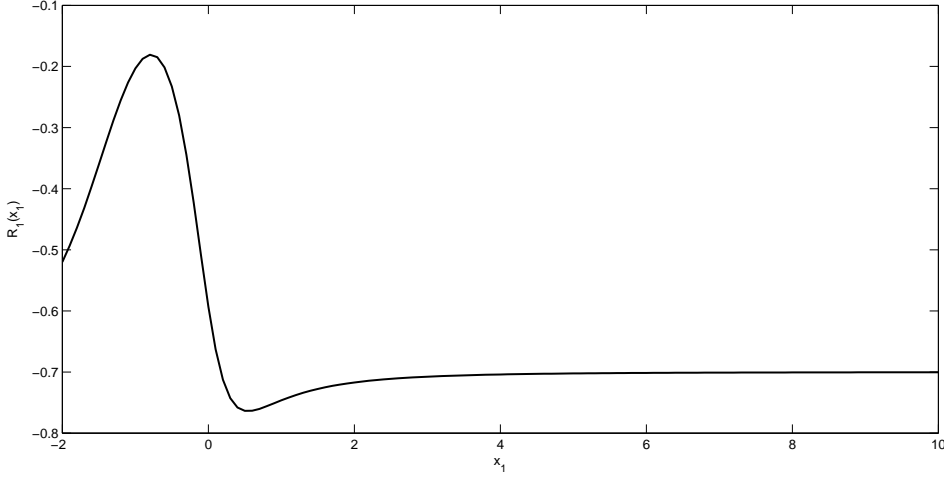


Figure 1: First element $R_1(x_1)$ of the correlation vector between index returns and diffusive variance changes specified by (K.4) and (K.7).

With this specification, X_t is a quadratic jump-diffusion process and hence is polynomial preserving. Its conditional moments are polynomials of its current value with coefficients given in closed form as shown in Theorem 3.5 for the univariate diffusion case.

Theorem K.1. Let $D = \frac{(3+N)(2+N)(1+N)}{6}$ denote the dimension of the space of polynomials in X_T of degree N or less. The D -row vector of the mixed \mathcal{F}_{t_0} -conditional moments of X_T of order N or less with $T \geq t_0$ is given by

$$\begin{aligned} & \left(1, \mathbb{E}^{\mathbb{Q}} [X_{1T} | \mathcal{F}_{t_0}], \dots, \mathbb{E}^{\mathbb{Q}} [X_{2T} X_{3T}^{N-1} | \mathcal{F}_{t_0}], \mathbb{E}^{\mathbb{Q}} [X_{3T}^N | \mathcal{F}_{t_0}] \right) \\ & = \left(1, X_{1t_0}, \dots, X_{2t_0} X_{3t_0}^{N-1}, X_{3t_0}^N \right) e^{\tilde{B}(T-t_0)}, \quad (\text{K.8}) \end{aligned}$$

where \tilde{B} is an upper block triangular $D \times D$ matrix derived similarly to the matrix B in Appendix F and $e^{\tilde{B}(T-t_0)}$ denotes the matrix exponential of $\tilde{B}(T-t_0)$.

Proof. We proceed as in the proof of Theorem 3.5, but for the generator of the univariate quadratic

diffusion replaced by the generator of the three-dimensional quadratic jump-diffusion X_t ,

$$\begin{aligned} \mathcal{A}f(x) = & \begin{pmatrix} \beta_{11}x_1 + \beta_{12}x_2 \\ b_2 + \beta_{22}x_2 \\ r \end{pmatrix}^\top \nabla_x f(x) + \frac{1}{2} \sum_{i,j=1}^3 A_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \\ & + \left(f(x + \log(1 + \xi)e_3) - f(x) - \nabla_x f(x)^\top e_3 \log(1 + \xi) \right) \nu^\mathbb{Q}(x), \end{aligned} \quad (\text{K.9})$$

where $e_3 = (0, 0, 1)^\top$. It is easily verified that \mathcal{A} is polynomial preserving: for $f(x)$ being a polynomial in x of degree $0 \leq n \leq N$, $\mathcal{A}f(x)$ is a polynomial in x of degree n or less. This property allows us to build the matrix \tilde{B} similarly as B in Appendix F by applying the generator \mathcal{A} to the mixed powers $1, x_1, \dots, x_2 x_3^{N-1}, x_3^N$ and collecting terms. The rest of the proof follows literally as in Appendix F. \square

We use an Edgeworth expansion of the characteristic function of $L_T | \mathcal{F}_{t_0}$, which allows us then to apply standard Fourier inversion to infer the option price. Let $z \in \mathbb{C}$, the characteristic function of $L_T | \mathcal{F}_{t_0}$ can be expanded as follows, where C_n refers to the n^{th} cumulant of $L_T | \mathcal{F}_{t_0}$,

$$\mathbb{E}^\mathbb{Q} [e^{zL_T} | \mathcal{F}_{t_0}] = \exp \left(\sum_{n=1}^{\infty} C_n \frac{z^n}{n!} \right) = \exp \left(C_1 z + C_2 \frac{z^2}{2} \right) \left(1 + C_3 \frac{z^3}{3!} + O(z^4) \right). \quad (\text{K.10})$$

The cumulants of $L_T | \mathcal{F}_{t_0}$ are inferred from the moments. The n^{th} power of L_T can be decomposed as follows

$$\begin{aligned} L_T^n &= (Y_T + X_{3T})^n = \sum_{k=0}^n \binom{n}{k} Y_T^k X_{3T}^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} K_Y^k \int_{t_0}^T \int_{t_0}^T \dots \int_{t_0}^T g(X_{t_1}) \dots g(X_{t_k}) dt_k \dots dt_1 X_{3T}^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} K_Y^k (k!) \int_{t_0}^T \int_{t_1}^T \dots \int_{t_{k-1}}^T g(X_{t_1}) \dots g(X_{t_k}) dt_k \dots dt_1 X_{3T}^{n-k}. \end{aligned} \quad (\text{K.11})$$

Taking the conditional expectation gives the n^{th} moment of $L_T | \mathcal{F}_{t_0}$

$$\mathbb{E}^\mathbb{Q} [L_T^n | \mathcal{F}_{t_0}] = \sum_{k=0}^n \binom{n}{k} K_Y^k (k!) \int_{t_0}^T \int_{t_1}^T \dots \int_{t_{k-1}}^T \mathbb{E}^\mathbb{Q} [g(X_{t_1}) \dots g(X_{t_k}) X_{3T}^{n-k} | \mathcal{F}_{t_0}] dt_k \dots dt_1. \quad (\text{K.12})$$

We compute the conditional expectation in the integral using nested conditional expectations

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}} \left[g(X_{t_1}) \cdots g(X_{t_k}) X_{3T}^{n-k} \mid \mathcal{F}_{t_0} \right] &= \mathbb{E}^{\mathbb{Q}} \left[g(X_{t_1}) \cdots g(X_{t_k}) \underbrace{\mathbb{E}^{\mathbb{Q}} \left[X_{3T}^{n-k} \mid \mathcal{F}_{t_k} \right]}_{P_0(t_k, X_{t_k})} \mid \mathcal{F}_{t_0} \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[g(X_{t_1}) \cdots g(X_{t_{k-1}}) \underbrace{\mathbb{E}^{\mathbb{Q}} \left[g(X_{t_k}) P_0(t_k, X_{t_k}) \mid \mathcal{F}_{t_{k-1}} \right]}_{P_1(t_k, t_{k-1}, X_{t_{k-1}})} \mid \mathcal{F}_{t_0} \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[g(X_{t_1}) P_{k-1}(t_k, t_{k-1}, \dots, t_1, X_{t_1}) \mid \mathcal{F}_{t_0} \right] \\
&= P_k(t_k, t_{k-1}, \dots, t_0, X_{t_0}).
\end{aligned} \tag{K.13}$$

The notation $P_j(t_k, \dots, t_{k-j}, X_{t_{k-j}})$, for $0 \leq j \leq k$, refers to a polynomial in $X_{t_{k-j}}$ of order $n - k + 2j$ or less, obtained recursively in closed form from Theorem K.1.¹

Option prices are computed using standard Fourier inversion, following Carr and Madan (1998) for at-the-money options and Fang and Oosterlee (2008) for out-of-the-money options. Infinitesimal option sensitivities, $\partial_s O_t$ and $\nabla_x O_t$, are calculated numerically by bumping the differentiation variable by $\delta\%$ times its current value. In practice we choose $\delta = 0.2$.²

Appendix L. Arguments for the bivariate quadratic model

Following up on Section 5.4, we provide a sketch of the arguments that all assumptions underpinning Theorem 5.2 are satisfied for the bivariate quadratic model in Section 5.3.

The function $g(X_t)$ as well as the determinant of the 2×2 matrix \mathcal{D}_t are nonzero polynomials in X_t , for all $t \in [0, T]$, and with smooth t -dependent coefficients. On the other hand, for any $C^{1,2}$ -function $\ell(t, x)$ it follows from the occupation times formula that

$$1_{\{\ell(t, X_t) = 0\}} \nabla_x \ell(t, X_t)^\top \Sigma(X_t) \Sigma(X_t)^\top \nabla_x \ell(t, X_t) = 0 \quad dt \otimes d\mathbb{Q}\text{-a.s.} \tag{L.1}$$

see Revuz and Yor (1994, Corollary (1.6), Chap. VI) and Filipović (2001, Lemma 3.3.1). Since $\Sigma(X_t) \Sigma(X_t)^\top$ is positive definite $dt \otimes d\mathbb{Q}$ -a.s., we infer that $\ell(t, X_t) \neq 0$ if $\nabla_x \ell(t, X_t) \neq 0$, $dt \otimes d\mathbb{Q}$ -a.s. Applying this to $g(X_t)$ and the determinant of \mathcal{D}_t , we find that $g(X_t) \neq 0$ and \mathcal{D}_t is invertible,

¹The integrals were computed using numerical integration. Resulting moments were benchmarked with those obtained with Monte Carlo simulations.

²We checked the robustness of the option prices and sensitivities with respect to the numerical value of δ .

$dt \otimes d\mathbb{Q}$ -a.s. In Appendix K we saw that $R_1(x_1)^2 < 1$ for all $x_1 \in \mathbb{R}$. Hence the 3×3 matrix $(\Sigma(X_t)^\top, \sigma(X_t)\mathbf{R}(X_t))$ is invertible $dt \otimes d\mathbb{Q}$ -a.s. For an index put option $\frac{\Delta O_t}{\xi S_{t-}} > \partial_s O_{t-}$ and (39) holds. Validity of Assumption 5.1 follows. Assumption H.1 holds by inspection. Assumption H.2 follows from general theory about diffusion processes, see Filipović and Larsson (2014) for the existence and uniqueness of the quadratic diffusion X_t , and Feehan and Pop (2014) for the $C^{1,2}$ regularity of $h(\tau, x)$.³

Following up on Appendix I, we provide an approximation of $\nabla_x h(\tau, x)$ in (53). From (H.1) and (51) we obtain $k(x) = K_k g(x)$ where

$$K_k = -\frac{1}{2\eta} \left(\frac{1}{\eta} - 1 \right) \kappa - \nu^Q K_{\sigma^2} \left(\left(\frac{\nu^P}{\nu^Q} \right)^{\frac{1}{\eta}} - 1 \right) - \frac{1}{\eta} (\nu^Q - \nu^P) K_{\sigma^2}, \quad (\text{L.2})$$

and we define $K_{\sigma^2} = (1 + (\log(1 + \xi))^2 \nu^Q)^{-1} > 0$ as in (46). This is of the form (I.1) with $c = K_k \phi$ and $\epsilon P(x) = K_k (\psi x_1 + \pi x_1^2)$. The first order Taylor expansion (I.2) then reads

$$\nabla_x h(\tau, x) \approx -K_k \nabla_x \int_0^\tau \mathbb{E}_{\mathbb{Q}} \left[\psi \widehat{X}_{1s} + \pi \widehat{X}_{1s}^2 \mid \widehat{X}_0 = x \right] ds, \quad (\text{L.3})$$

which is available in closed form.

³Feehan and Pop (2014) require that $k(x)$ is bounded from below, which is satisfied if $\eta \geq 1$ as noted below Assumption H.1.

References

- Andersen, T., Bollerslev, T., Diebold, F., Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Carr, P., Madan, D., 1998. Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2, 61–73.
- Fang, F., Oosterlee, C. W., 2008. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing* 31, 826–848.
- Feehan, P. M. N., Pop, C., 2014. On the martingale problem for degenerate-parabolic partial differential operators with unbounded coefficients and a mimicking theorem for Ito processes. *Transactions of the AMS* Forthcoming.
- Filipović, D., 2001. Consistency Problems for Heath-Jarrow-Morton Interest Rate Models, vol. 1760 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.
- Filipović, D., 2009. *Term-Structure Models*. Springer-Verlag, Berlin.
- Filipović, D., Larsson, M., 2014. Polynomial preserving diffusions and applications in finance, unpublished working paper, Swiss Finance Institute.
- Karatzas, I., Shreve, S. E., 1991. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.
- Revuz, D., Yor, M., 1994. *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin, Germany, second ed.
- Wu, L., 2011. Variance dynamics: Joint evidence from options and high-frequency returns. *Journal of Econometrics* 160, 280–287.
- Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019–1043.
- Zhang, L., Mykland, P. A., Ait-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.