Quadratic Variance Swap Models

Damir Filipović† Elise Gourier‡ Loriano Mancini§
Swiss Finance Institute ORFE Swiss Finance Institute
and EPFL Princeton University and EPFL

This version: 18 December 2014
First draft: April 2013


*For helpful comments we thank an anonymous referee, Bruno Bouchard, Peter Carr, Pierre Collin-Dufresne, Jerome Detemple, Rüdiger Fahlenbrach, Jim Gatheral, Julien Hugonnier, Jakub Jurek, Mika Kastenholz, Martin Larsson, Markus Leippold, Dilip Madan, Eckhard Platen, Leonidas Rompolis and Nizar Touzi, as well as participants at the 2013 European Finance Association meetings in Cambridge. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 307465-POLYTE, and the Swiss National Science Foundation NCCR-FinRisk.

†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.
Quadratic Variance Swap Models

Abstract

We introduce a novel class of term structure models for variance swaps. The multivariate state process is characterized by a quadratic diffusion function. The variance swap curve is quadratic in the state variable and available in closed form, greatly facilitating empirical analysis. Various goodness-of-fit tests show that quadratic models fit variance swaps on the S&P 500 remarkably well, and outperform affine models. We solve a dynamic optimal portfolio problem in variance swaps, index option, stock index and bond. An empirical analysis uncovers robust features of the optimal investment strategy.

JEL Classification: C51, G13

Keywords: stochastic volatility, variance swap, quadratic term structure, quadratic jump-diffusion, dynamic optimal portfolio
1 Introduction

A variance swap pays the difference between the realized variance of some underlying asset and the fixed variance swap rate. Variance swaps are actively traded at different maturities. This induces a term structure of variance swap rates, which reflects market expectations about future variance and provides important information for managing variance risk. Fig. 1 shows variance swap rates on the S&P 500. The term structure takes a variety of shapes and exhibits rich dynamics. During low volatility periods, such as 2005–2006, the term structure is upward sloping. During financial crises, such as fall 2008, the short-end spikes up, and the term structure becomes downward sloping. Having a model that captures such term structure movements appears to be crucial to consistently price variance swaps across different maturities or to optimally invest in such contracts. Surprisingly, the term structure of variance swap rates has received little attention in the literature.

We provide a novel class of flexible and tractable variance swap term structure models. The multivariate state variable driving the stochastic variance follows a quadratic diffusion process characterized by linear drift and quadratic diffusion functions. Variance swap rates are quadratic in the state variable. The variance swap curve is available in closed form in terms of a linear ordinary differential equation, which greatly facilitates empirical applications. Higher order polynomial specifications are possible.

We perform an exhaustive specification analysis of the univariate quadratic model and of a parsimonious bivariate extension. Model identification is provided in terms of canonical representations. We also study univariate polynomial specifications of higher order. We fit these models to the daily quadratic variation from tick-by-tick S&P 500 futures data and the term structure of variance swap rates on the S&P 500 shown in Fig. 1. Several statistical tests show that the bivariate quadratic model captures the term structure dynamics remarkably well. The quadratic state process is able to generate sudden large movements in the variance swap rates, and the quadratic variance swap model can produce a rich variety of term structure shapes, as observed empirically. Nested affine and other specifications are soundly rejected. Our quadratic model also outperforms a standard two-factor affine jump-diffusion model which is typically used in the literature. We reach this conclusion using various likelihood-based tests (e.g., Giacomini and White (2006)), information theoretic criteria (i.e., Akaike and Bayesian Information Criteria), and Diebold–Mariano tests derived from variance swap pricing errors.
We find that the bivariate quadratic model produces better forecasts of variance swap rates than the univariate quadratic and polynomial models, as well as the martingale model. The latter uses today’s variance swap rates as a prediction of future variance swap rates. Given the high persistence of variance swap rates,¹ the martingale model is a challenging benchmark. When we regress future variance swap rates on model-based predictions of variance swap rates, we find that the bivariate quadratic model has an intercept and a slope not statistically different from zero and one, respectively, and thus produces accurate forecasts. The bivariate model outperforms the martingale model, which in turn dominates the univariate quadratic and polynomial models. From an economic perspective, this suggests that the bivariate quadratic model captures well the ex-ante risk premiums embedded in variance swaps.

At least two features contribute to the popularity of variance swaps. First, hedging a variance swap is easier than hedging other volatility derivatives. In the absence of asset price jumps, the payoff of a variance swap can be replicated by dynamic trading in the underlying asset and a static position in a continuum of vanilla options with different strike prices and the same underlying and maturity date. In practice, of course, continuous trading is unfeasible and vanilla options exist only for a limited number of strike prices and could not exist at all for a given maturity date.² Second, the variance swap payoff is only sensitive to the realized variance over a desired and predetermined time horizon. Suppose an investor holds a broadly diversified portfolio and is concerned about volatility risk over the next month. Buying a variance swap on the S&P 500, with one month maturity, would provide a direct hedge against volatility risk. In contrast, taking positions on options and futures on the VIX index³ would not provide an equally direct hedge.⁴

To assess the economic relevance of variance swaps, we study a dynamic optimal portfolio problem in variance swaps, index option, stock index and risk-free bond. We use a quadratic jump-

¹First order autocorrelations are above 0.98.
²This led to a large literature analyzing and exploiting the various hedging errors when attempting to replicate a given variance swap, e.g., Neuberger (1994), Dupire (1993), Carr and Madan (1998b), Demeterfi et al. (1999), Britten-Jones and Neuberger (2000), Jiang and Tian (2005), Jiang and Oomen (2008), Carr and Wu (2009), and Carr and Lee (2010).
³Absent index jumps, the Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX) is the 30-day variance swap rate on the S&P 500 quoted in volatility units. Carr and Wu (2006) provide an excellent history of the VIX index.
⁴It is so because the VIX index is the market expectation of the S&P 500 variance over the next 30 days. Thus, as time goes by, the VIX index, and derivatives on it, are sensitive to variance expectations beyond the desired hedging horizon. In response to the need to trade volatility with more direct instruments, since December 2012 the CBOE has listed new contracts called “S&P 500 Variance Futures.” These are exchange-traded, marked-to-market variance swaps on the S&P 500 with maturities ranging up to two years. See http://www.cboe.cboe.com/Products/Spec_VA.aspx.
diffusion model for the price process of the stock index. We solve for the optimal strategy of a power utility investor who maximizes the expected utility from terminal wealth. The variance swaps are on-the-run and rolled over at pre-specified arbitrary points in time. The optimal strategy, composed of the familiar myopic and intertemporal hedging terms (Merton (1971)), is derived in quasi-closed form. A Taylor series expansion of the intertemporal hedging term involves conditional moments of the state variables, which are available in closed form. We implement the optimal portfolio using three-month and two-year variance swaps, an out-of-the-money put option, and the S&P 500. To include the put option in the investment universe, we develop a novel pricing formula for European options. The transition density of the stock price process is approximated using an Edgeworth expansion, relying on closed form expressions for joint conditional moments of the stock price and state variables.

We empirically find that the optimal portfolio weights in the variance swaps follow a short-long strategy, with a short position in the two-year variance swap (to earn the negative variance risk premium), and a long position in the three-month variance swap (to hedge volatility increases). This result is consistent with the empirical finding that long-term variance swaps carry more variance risk premium and react less to volatility increases than short-term variance swaps, e.g., Egloff et al. (2010), and Aït-Sahalia et al. (2014). We also find that optimal weights in variance swaps exhibit strong periodic patterns, which depend on the maturity and roll-over date of the contracts. Remarkably, when the stock price does not jump and the investor cannot trade index options, the optimal strategy in variance swaps remains largely the same. This suggests that the short-long strategy in variance swaps is a robust feature of the optimal investment strategy.

The optimal investment in the put option is in line with the numerical calibration in Liu and Pan (2003, Table 1). The portfolio weight is very small (less than 1% of the total wealth in most cases), increases in the jump size and/or jump intensity, and can change sign: it is positive when the jump risk premium is small (providing hedging against index jumps), while it is negative when the jump risk premium is large (to earn the jump risk premium).

We consider two relative risk aversion levels, five and one. The first is an average value in survey data. The second corresponds to logarithmic utility. Optimal portfolio weights for both

---

5Some practitioners describe such short-long trading strategies with variance swaps as “trading the spread” of variance swaps.

6Most of the survey data suggests values of the relative risk aversion between 0.23 and eight, e.g., Meyer and Meyer (2005).
levels share the patterns described above. However, the respective wealth trajectories are largely different. The more risk averse investor takes on smaller positions than the log-investor, in absolute value. This results in a smooth and steady growth of his wealth over time, which is largely unaffected by market declines. In contrast, the wealth trajectory of the log-investor is volatile and fluctuates even more than the S&P 500. This suggests that variance swaps can be used either to achieve stable wealth growth or to seek additional risk premiums, depending on the risk profile of the investor. Rebalancing the portfolio less frequently than daily, such as monthly or yearly, leads to similar results.

To summarize our findings on the optimal portfolio, the short-long strategy in variance swaps appears to be a robust feature of the optimal investment strategy. Variance swaps serve the purpose of providing exposure to volatility risk premiums and hedging volatility risk. The presence of price jumps does not significantly change the optimal investment in variance swaps, provided that the investor can optimally trade index options to hedge price jumps or earn the jump risk premium.

Our paper is related to various strands of the literature. A fast growing literature studies the variance risk premium and its impact on asset prices, e.g., Jiang and Tian (2005), Carr and Wu (2009), Bollerslev et al. (2009), Todorov (2010), Bollerslev and Todorov (2011), Drechsler and Yaron (2011), Mueller et al. (2011), Martin (2013), and Bekaert and Hoerova (2014). This line of research focuses almost exclusively on a single maturity. As mentioned above, the term structure of variance swap rates has remained unexplored until recently, e.g., Amengual (2009), Egloff et al. (2010), and Aït-Sahalia et al. (2014). Part of the reason could be that variance swap data became available only recently. We contribute to this line of research by proposing a novel quadratic term structure model, assessing its empirical performance, and studying dynamic optimal portfolios in this setting.

There is an extensive literature on term structure models for interest rates. This literature mainly focuses on affine term structure models, where the zero-coupon yield curve is affine in the state variable which follows an affine diffusion process. The loadings in turn are given in terms of

---


8 Affine diffusion processes are nested in our class of quadratic diffusion processes.
a nonlinear ordinary differential equation. Quadratic and higher order polynomial specifications of the yield curve are very limited, Filipović (2002), and Chen et al. (2004). These limitations do not exist for the variance swap curve. This allows us to define the class of generic quadratic variance swap models, where the \( Q \)-spot variance is a quadratic, or higher order polynomial, function of the state variable which follows a quadratic diffusion process. The resulting variance swap curve is quadratic, or higher order polynomial, in the state variable, and the loadings are given in terms of a linear ordinary differential equation.

Several studies investigate dynamic optimal portfolios with stochastic investment opportunity set. Most of them consider optimal investment in stock and bond only, without any derivative, often in a one-factor stochastic volatility setting, and when the price process is continuous. The focus is usually on theoretical aspects of optimal portfolios, and thus empirical analyses are not provided. In addition to stock and bond, in an affine setting, Liu and Pan (2003) extend the investment opportunity set to options, and Egloff et al. (2010) to variance swaps. We study, both from theoretical and empirical perspectives, dynamic optimal portfolios including variance swaps and index option in our quadratic setting with stock index jumps.

The structure of the paper is as follows. Section 2 presents variance swaps. Section 3 introduces quadratic variance swap models. Section 4 discusses model estimates. Section 5 studies optimal portfolios in variance swaps, index option, stock index and risk-free bond. Section 6 investigates the empirical performance of optimal portfolios. Section 7 concludes. Technical derivations and proofs are collected in an online appendix.


2 Variance swaps

Let $S_t$ denote the price process of a stock index modeled on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, Q)$, where $Q$ is a risk neutral measure. We let $S_t$ be a semimartingale of the form

$$
\frac{dS_t}{S_t} = r_t \, dt + \sigma_t \, dB_t + \int_{\mathbb{R}} \xi \left( \chi(dt, d\xi) - \nu^Q_t(d\xi)dt \right),
$$

(1)

where $r_t$ is the risk-free rate, $B_t$ is a $Q$-standard Brownian motion, and $\chi(dt, d\xi)$ denotes the random measure associated to the jumps of $S_t$. Its $Q$-compensator $\nu^Q_t(d\xi)dt$ is such that the last term in (1) is the increment of a $Q$-pure jump martingale. The diffusive component of the price volatility is $\sigma_t$.

Let $t = t_0 < t_1 < \cdots < t_n = T$ denote the trading days over a given time horizon $[t, T]$. The annualized realized variance is the annualized sum of squared log-returns over the given time horizon,

$$
RV(t, T) = \frac{252}{n} \sum_{i=1}^{n} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2.
$$

(2)

It is known that, as $\sup_{i=1,\ldots,n} (t_i - t_{i-1}) \to 0$, the realized variance converges in probability to the quadratic variation of the log-price,

$$
\sum_{i=1}^{n} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \to QV(t, T) = \int_t^T \sigma^2_s \, ds + \int_t^T \int_{\mathbb{R}} (\log(1 + \xi))^2 \chi(ds, d\xi).
$$

(3)

This approximation is commonly adopted in practice (e.g., Egloff et al. (2010)) and quite accurate at a daily sampling frequency (e.g., Broadie and Jain (2008), and Jarrow et al. (2013)), as is the case in our data set.\(^\text{11}\)

A variance swap initiated at $t$ with maturity $T$, or term $T - t$, pays the difference between the annualized realized variance $RV(t, T)$ and the variance swap rate $VS(t, T)$ fixed at $t$.\(^\text{12}\) By convention, the variance swap rate is such that the variance swap contract has zero value at inception. No arbitrage implies that

$$
VS(t, T) = \frac{1}{T-t} \mathbb{E}^Q [QV(t, T) \mid \mathcal{F}_t] = \frac{1}{T-t} \mathbb{E}^Q \left[ \int_t^T \nu^Q_s \, ds \mid \mathcal{F}_t \right],
$$

(4)

\(^\text{11}\)Market microstructure noise, while generally a concern in high frequency inference, is largely a non-issue at the level of daily returns.

\(^\text{12}\)As the difference is in variance units, the payoff is converted in dollar units via a suitable notional amount.
where $E_Q$ denotes expectation under the risk neutral measure $Q$, and

$$v_t^Q = \sigma_t^2 + \int_\mathbb{R} (\log(1 + \xi))^2 \nu_t^Q(d\xi)$$

(5)
is the $Q$-spot variance process.$^{13}$

The jump compensator $\nu_t^P(d\xi)dt$ of the index price process under the objective probability measure $P$ differs from the $Q$-jump compensator $\nu_t^Q(d\xi)dt$ in general, reflecting price jump risk premium. The $P$-spot variance $v_t^P = \sigma_t^2 + \int_\mathbb{R} (\log(1 + \xi))^2 \nu_t^P(d\xi)$ is related to the $P$-expected realized variance

$$\frac{1}{T-t} E^P [QV(t, T) \mid \mathcal{F}_t] = \frac{1}{T-t} E^P \left[ \int_t^T v_s^P ds \mid \mathcal{F}_t \right].$$

(6)

To consistently price variance swaps and capture the term structure of volatility risk, it is crucial to design models for the entire variance swap curve $T \mapsto \text{VS}(t, T)$. In view of (4), this boils down to modeling the $Q$-spot variance process $v_t^Q$. These models should be analytically tractable and yet flexible enough to reproduce the empirical features of variance swap rates. Any positive semimartingale whose $Q$-spot variance process coincides with $v_t^Q$ is then a consistent price process in the sense that $\text{VS}(t, T)$ is the corresponding variance swap rate.

It is instructive to draw an analogy between the term structure of variance swaps and interest rates. The variance swap curve reflects market expectations about future changes in $Q$-spot variance, see (4). The financial variable in interest rate models corresponding to the $Q$-spot variance $v_t^Q$ is the risk-free short rate $r_t$. Market expectations about future changes in short rates are expressed in terms of the zero-coupon yield curve

$$y(t, T) = -\frac{1}{T-t} \log E^Q \left[ e^{-\int_t^T r_s ds} \mid \mathcal{F}_t \right],$$

(7)

with short-end given by $y(t, t) = r_t$. Clearly, the yield curve is a nonlinear function of the short rate process. In contrast, the variance swap curve is a linear function of the $Q$-spot variance process. This linear relation gives greater flexibility for the specification of analytically tractable term structure models for variance swaps than for interest rates. Indeed, most common factor models for the term structure of interest rates are affine term structure models. The short rate is

$^{13}$We assume that the risk-free rate and the $Q$-spot variance are independent processes, which is certainly a tenuous assumption.
specified as an affine function of the state variable which follows an affine diffusion process. The resulting yield curve is affine in the state variable, and the loadings are given as solutions to a nonlinear ordinary differential equation, e.g., Duffie and Kan (1996), and Dai and Singleton (2000). Specifying the short rate as a quadratic function of the state variable is possible. But it generically requires that the state variable follows a Gaussian process, e.g., Ahn et al. (2002), Chen et al. (2004), and Liu (2007).\footnote{Liu (2007) considers mixtures of quadratic Gaussian and affine components in a specific setup.} Moreover, there exists no consistent polynomial specification of the yield curve beyond second order, Filipović (2002). These limitations do not exist for variance swap term structure models, and this flexibility is exploited here.

3 Quadratic variance swap models

Let \( X_t \) be a diffusion process in some state space \( \mathcal{X} \subset \mathbb{R}^m \), solving the stochastic differential equation (SDE)

\[
dX_t = \mu(X_t) \, dt + \Sigma(X_t) \, dW_t,
\]

where \( W_t \) is a standard \( d \)-dimensional Brownian motion under the risk neutral measure \( Q \), and \( \mu(x) \) and \( \Sigma(x) \) are \( \mathbb{R}^m \)- and \( \mathbb{R}^{m \times d} \)-valued functions on \( \mathcal{X} \), for some integers \( m, d \geq 1 \). The process \( X_t \) has the following quadratic structure,

**Definition 3.1.** The diffusion \( X_t \) is called quadratic if its drift and diffusion functions are linear and quadratic in the state variable,

\[
\begin{align*}
\mu(x) &= b + \beta x, \\
\Sigma(x)\Sigma(x)^\top &= a + \sum_{k=1}^m \alpha_k x_k + \sum_{k,l=1}^m A^{kl} x_k x_l,
\end{align*}
\]

for some parameters \( b \in \mathbb{R}^m, \beta \in \mathbb{R}^{m \times m}, \) and \( a, \alpha_k, A^{kl} \in \mathbb{S}^m \) with \( A^{kl} = A^{lk} \), where \( \mathbb{S}^m \) denotes the set of symmetric \( m \times m \)-matrices, and \( \top \) denotes transpose.

An \( m \)-factor quadratic variance swap model is obtained by specifying the \( Q \)-spot variance as a quadratic function of the state variable,

\[
v_t^Q = g(X_t),
\]
with \( g(x) = \phi + \psi^\top x + x^\top \pi x \), for some parameters \( \phi \in \mathbb{R} \), \( \psi \in \mathbb{R}^m \), and \( \pi \in S^m \). The following theorem justifies the terminology of quadratic variance swap model.

**Theorem 3.2.** Under the above assumptions, the quadratic variance swap model admits a quadratic term structure. That is, the variance swap rates are quadratic in the state variable,

\[
VS(t, T) = \frac{1}{T-t} G(T - t, X_t),
\]

with \( G(\tau, x) = \Phi(\tau) + \Psi(\tau)^\top x + \pi(\tau) x \), where the functions \( \Phi : [0, +\infty) \to \mathbb{R} \), \( \Psi : [0, +\infty) \to \mathbb{R}^m \), and \( \Pi : [0, +\infty) \to S^m \) satisfy the linear ordinary differential equations

\[
\begin{align*}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b^\top \Psi(\tau) + \text{tr}(a \Pi(\tau)), \quad \Phi(0) = 0, \\
\frac{d\Psi(\tau)}{d\tau} &= \psi + \beta^\top \Psi(\tau) + 2\Pi(\tau)b + \alpha \cdot \Pi(\tau), \quad \Psi(0) = 0, \\
\frac{d\Pi(\tau)}{d\tau} &= \pi + \beta^\top \Pi(\tau) + \Pi(\tau)\beta + A \cdot \Pi(\tau), \quad \Pi(0) = 0,
\end{align*}
\]

where we define the tensor operations \( (\alpha \cdot \Pi)_k = \text{tr}(\alpha^k \Pi) \) and \( (A \cdot \Pi)_{kl} = \text{tr}(A_{kl} \Pi) \).

**Proof.** The assertion follows from (4) and Lemma A.2 in Appendix A with \( f(\tau, x) = \partial G(\tau, x) / \partial \tau \).

Appendix B shows that, under mild technical conditions, the converse to Theorem 3.2 also holds true: a quadratic term structure implies that the \( \mathbb{Q} \)-spot variance function and the state diffusion process \( X_t \) be necessarily quadratic. This result implies that our quadratic model framework is exhaustive in the sense that we do not miss any other diffusion specification which is consistent with a quadratic term structure.

We also specify an \( \mathbb{R}^d \)-valued process for the market price of risk, \( \Lambda \), such that \( dW_t^P = dW_t - \Lambda_t dt \) is a \( \mathbb{P} \)-Brownian motion, and \( \Sigma(X_t) \Lambda_t = \Upsilon_0 + \Upsilon_1 X_t \) holds for some parameters \( \Upsilon_0 \in \mathbb{R}^m \) and \( \Upsilon_1 \in \mathbb{R}^{m \times m} \). This implies that the \( \mathbb{P} \)-dynamics of \( X_t \) are of the form

\[
dX_t = (b + \Upsilon_0 + (\beta + \Upsilon_1) X_t) dt + \Sigma(X_t) dW_t^P.
\]

Thus, the process \( X_t \) follows a quadratic diffusion under \( \mathbb{P} \) as well. The properties of \( X_t \) derived from the quadratic structure hold under \( \mathbb{Q} \) as well as under \( \mathbb{P} \).
It follows by inspection that an affine transformation of the state, \( X_t \mapsto c + \gamma X_t \), preserves the quadratic property (9)–(10) of \( X_t \) and the quadratic term structure (12). From an econometric viewpoint, this implies that the above general model is not identifiable. This calls for a canonical representation. A full specification analysis of general multi-factor quadratic models is beyond the scope of this paper.\(^{15}\) In the following sections, we first provide an exhaustive specification analysis for the univariate quadratic model. We then study a bivariate extension and univariate polynomial specifications of higher order. Model identification is asserted in terms of canonical representations.

### 3.1 Univariate quadratic model

In this section, let \( m = d = 1 \) and consider a univariate quadratic diffusion

\[
dX_t = (b + \gamma X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t
\]  

(15)

on some interval \( \mathcal{X} \) in \( \mathbb{R} \) and for some real parameters \( b, \beta, a, \alpha, \) and \( A \geq 0 \). The linear ordinary differential equations (13) simplify to (C.1) in Appendix C.

The invariance of quadratic processes with respect to affine transformations allows us to distinguish exactly three equivalence classes of quadratic processes on unbounded intervals with a canonical representation each. In other words, any univariate quadratic process (on unbounded intervals and possibly after an affine transformation) necessarily falls in one of the three equivalence classes. The three canonical representations are identifiable, and thus can be estimated using variance swap data. The proof is given in Appendix D.

**Theorem 3.3.** Denote the discriminant of the diffusion function of \( X_t \) by \( D = \alpha^2 - 4Aa \). The quadratic process \( X_t \) falls in one of the following three equivalence classes,

- **Class 1:** either \( A > 0 \) and \( D < 0 \), or \( A = \alpha = 0 \) and \( a > 0 \). The canonical representation is

\(^{15}\)This would require to find necessary and sufficient conditions on the model parameters and the state space \( \mathcal{X} \) such that the multivariate quadratic diffusion \( X_t \) be well-defined in \( \mathcal{X} \). The matrix-valued quadratic form on the right hand side of (10) needs to be positive semi-definite for all \( x \in \mathcal{X} \). Moreover, it has to vanish in the direction orthogonal to the boundary at all boundary points, for the state space to be invariant under the dynamics of \( X_t \). Hence the state space \( \mathcal{X} \) is specified by the zeros of quadratic forms on \( \mathbb{R}^m \). The zero level sets of quadratic forms on \( \mathbb{R}^m \) are complex geometric objects, and the canonical classification of quadratic diffusions would at least require an exhaustive classification of such zero level sets. Filipović and Larsson (2014) provide a related study.
specified by $X = \mathbb{R}$, $b \geq 0$, $\beta \in \mathbb{R}$, $a = 1$, $\alpha = 0$, $A \geq 0$, and hence
\[ dX_t = (b + \beta X_t) \, dt + \sqrt{1 + AX_t^2} \, dW_t. \] (16)

Note that for $A = 0$ we obtain a Gaussian process.

- **Class 2:** either $A > 0$ and $D = 0$, or $A = \alpha = a = 0$. The canonical representation is specified by $X = (0, +\infty)$, $b = 1$ or $0$, $\beta \in \mathbb{R}$, $a = 0$, $\alpha = 0$, $A \geq 0$, and hence
\[ dX_t = (b + \beta X_t) \, dt + \sqrt{AX_t^2} \, dW_t. \] (17)

Note that for $A = 0$ we obtain a deterministic process.

- **Class 3:** either $A > 0$ and $D > 0$, or $A = 0$ and $\alpha \neq 0$. The canonical representation is specified by $X = [0, +\infty)$, $b \geq 0$, $\beta \in \mathbb{R}$, $a = 0$, $\alpha = 1$, $A \geq 0$, and hence
\[ dX_t = (b + \beta X_t) \, dt + \sqrt{X_t + AX_t^2} \, dW_t. \] (18)

The boundary point 0 is not attainable if and only if $b \geq 1/2$, in which case we can choose $X = (0, +\infty)$. Note that for $A = 0$ we obtain an affine process.

**Remark 3.4.** For $A < 0$ and $D > 0$, the state space $X$ becomes bounded. The canonical representation for this equivalence class is the Jacobi process on $X = [0, 1]$. We do not consider this case, as here we focus on state processes on unbounded state spaces.

### 3.2 Bivariate quadratic model

In this section, we consider a bivariate extension of the above univariate quadratic model. Higher dimensional extensions are conceptually straightforward, but these models would be difficult to estimate because of the large number of parameters. Our empirical analysis below shows that a bivariate model provides a good fit to variance swaps and quadratic variation, thus higher order dimensional extensions do not appear to be practically relevant.
Let $m = 2$ and consider a bivariate quadratic diffusion $X_t = (X_{1t}, X_{2t})^\top$ of the form

\[
\begin{align*}
    dX_{1t} &= (b_1 + \beta_{11} X_{1t} + \beta_{12} X_{2t}) \, dt + \sqrt{a_1 + \alpha_1 X_{1t}^2 + A_1 X_{2t}^2} \, dW_{1t}, \\
    dX_{2t} &= (b_2 + \beta_{22} X_{2t}) \, dt + \sqrt{a_2 + \alpha_2 X_{2t}^2 + A_2 X_{2t}^2} \, dW_{2t},
\end{align*}
\]

(19)

with $\beta_{12} \geq 0$ and $X_{2t} \geq 0$. The components $X_{1t}$ and $X_{2t}$ are instantaneously uncorrelated and only interact via the drift term. The $\mathbb{Q}$-spot variance function is assumed to depend on $X_{1t}$ only,

\[g(x) = \phi + \psi x_1 + \pi x_1^2,\]

(20)

where $x = (x_1, x_2)$, for some real parameters $\phi, \psi$ and $\pi$. Hence $X_{1t}$ drives the $\mathbb{Q}$-spot variance, while $X_{2t}$ determines the stochastic mean reversion level, $-(b_1 + \beta_{12} X_{2t})/\beta_{11}$, of $X_{1t}$. The linear ordinary differential equations (13) simplify to (C.3) in Appendix C.

The admissible specifications for $X_{2t}$ are either Class 2 or 3 with the corresponding canonical representations given by Theorem 3.3. The diffusion function of $X_{1t}$ can be of any Class 1–3 with the corresponding canonical representations from Theorem 3.3. Imposing $b_1 = 0$ when the diffusion function of $X_{1t}$ is in Class 1 or 2, and $b_1 = 0$ or $1/2$ when it is in Class 3, ensures that the bivariate quadratic model is identified. This is proved in Appendix E. The univariate quadratic model is nested in the bivariate model, setting $X_{2t}$ to a positive constant value.

To keep the model parsimonious, a risk premium is attached only to the first Brownian motion, $W_{1t}$. The market price of risk process is then

\[
\Lambda_t = \left( \frac{\lambda_0 + \lambda_1 X_{1t}}{\sqrt{a_1 + \alpha_1 X_{1t}^2 + A_1 X_{2t}^2}}, 0 \right)^\top.
\]

(21)

The parameter $\lambda_0$ can take any real value if the diffusion function of $X_{1t}$ is in Class 1, $\lambda_0 \geq 0$ if the diffusion function of $X_{1t}$ is in Class 2 or in Class 3 along with $b_1 = 1/2$, and $\lambda_0 = 0$ otherwise. It follows from Cheridito et al. (2007) that the change of measure $\mathbb{P} \sim \mathbb{Q}$ is well defined under these conditions.
3.3 Univariate polynomial model

An important property of quadratic diffusion processes is that their conditional $n$th moments are available in closed form as polynomials of degree $n$ in the state variables. This is in fact the reason why in Theorem 3.2 we obtain the closed form quadratic expression for $G(T - t, X_t)$. Indeed, $\partial G(T - t, X)/\partial T$ is simply the $\mathcal{F}_t$-conditional moment of the quadratic polynomial $g(X_T)$ in $X_T$. This polynomial preserving property of $X_t$ suggests a natural extension of the quadratic variance swap models, namely higher order polynomial variance swap models. Here we discuss the univariate case. The multivariate case is a straightforward but notationally cumbersome extension.

As in Section 3.1, we consider the univariate quadratic diffusion process (15). The following theorem formalizes the polynomial preserving property of $X_t$. The proof is given in Appendix F.

**Theorem 3.5.** The $(N + 1)$ row vector of the first $N$ $\mathcal{F}_t$-conditional moments of $X_{t+\tau}$ with $\tau \geq 0$ is given by

$$
\left( 1, \mathbb{E}^Q[X_{t+\tau} | \mathcal{F}_t], \ldots, \mathbb{E}^Q[X_{t+\tau}^N | \mathcal{F}_t] \right) = (1, X_t, \ldots, X_t^N) e^{B \tau},
$$

(22)

where $B$ is an upper triangular $(N + 1) \times (N + 1)$ matrix defined in (F.3) in Appendix F, and $e^{B \tau}$ denotes the matrix exponential of $B \tau$.

A polynomial variance swap model is then obtained by specifying the $\mathbb{Q}$-spot variance as a polynomial function of the state variable, $v^\mathbb{Q}_t = p_0 + p_1 X_t + \cdots + p_N X_t^N$, for some parameters $p_i \in \mathbb{R}, i = 0, \ldots, N$. The following corollary is an immediate consequence of Theorem 3.5.

**Corollary 3.6.** Under the above assumptions, the polynomial variance swap model admits a polynomial term structure. That is, the variance swap rates are polynomial of degree $N$ in $X_t$:

$$
VS(t, T) = \frac{1}{T - t} \left( P_0(T - t) + P_1(T - t)X_t + \cdots + P_N(T - t)X_t^N \right),
$$

(23)

where the functions $P_i : [0, +\infty) \to \mathbb{R}$ satisfy the linear ordinary differential equations

$$
\frac{dP(\tau)}{d\tau} = p + B P(\tau), \quad P(0) = 0,
$$

(24)

where $P(\tau) = (P_0(\tau), P_1(\tau), \ldots, P_N(\tau))^\top$ and $p = (p_0, p_1, \ldots, p_N)^\top$.

System (24) is equivalent to (13) for $N = 2$, with loadings $\Phi(\tau) = P_0(\tau)$, $\Psi(\tau) = P_1(\tau)$, and
\( \Pi(\tau) = P_2(\tau). \)

4 Model estimation

In this section, we fit the variance swap models in Sections 3.1–3.3 to variance swap rates on the S&P 500 and its quadratic variation computed from tick-by-tick S&P 500 futures prices. An advantage of this estimation approach is that model estimates are not impaired by potential mis-specifications of the index dynamics and allows for a thorough comparison of the variance swap models.

4.1 Data set

Our data set includes daily over-the-counter quotes of variance swap rates on the S&P 500, with fixed terms at two, three, and six months, and one and two years.\(^{16}\) It ranges from January 4, 1996 to June 7, 2010, and includes 3,626 observations for each term. Standard statistical tests do not detect any day-of-the-week-effect, so we use all available daily data. An interesting feature of this data set is that terms, rather than maturities, are fixed. This facilitates the comparison of the term structure over time, without using any interpolation method to recover variance swap rates for a specific term. Our data set also includes daily quadratic variation computed from tick-by-tick S&P 500 futures prices using the two-scale estimator of Zhang et al. (2005).

Fig. 1 shows the term structure of variance swap rates over time and suggests that variance swap rates are mean-reverting, volatile, with spikes and clustering during the major financial crises over the last 15 years, and historically high values during the financial crisis in fall 2008. While most term structures are upward sloping (48\% of our sample), they can also be \( \cup \)-shaped (23\% of our sample) and rarely downward sloping or \( \cap \)-shaped.\(^{17}\) The bottom and peak of the \( \cup \)- and \( \cap \)-shaped parts of the term structures, can be anywhere at the three or six months or one year term. The slope of the term structure, measured as the difference between the two-year and two-month variance swap rates, shows a strong negative relation to the contemporaneous level of volatility. Thus, in high volatility periods, the short-end of the term structure (variance swap rates with two

\(^{16}\)We thank Mika Kastenholz from Credit Suisse for providing us with the variance swap data.

\(^{17}\)On some occasions, the term structure is \( \sim \)-shaped, but the difference between, e.g., the two and three months variance swap rates is virtually zero and this term structure is nearly \( \cup \)-shaped.
or three months terms) rises more than the long-end, producing downward sloping term structures.

Table 1 provides summary statistics of our data set. We split the sample in two parts. The first part ranges from January 4, 1996 to April 2, 2007, includes 2,832 daily observations (about 3/4 of the whole sample), and will be used for in-sample analysis and model estimation. The second part ranges from April 3, 2007 to June 7, 2010, includes 794 daily observations, and will be used for out-of-sample analysis, including model validation. The out-of-sample analysis appears to be particularly interesting as the sample period covers the recent financial crisis, a period of unprecedented market turmoil, which was not experienced in the in-sample period.

For the sake of interpretability, we follow market practice and report variance swap rates in volatility percentage units, i.e., $\sqrt{\text{VS}(t,T)} \times 100$. Various empirical regularities emerge from Table 1. The mean level of variance swap rates is slightly but strictly increasing with term. The standard deviation, skewness and kurtosis of variance swap rates are decreasing with term. Unreported first order autocorrelations of variance swap rates range from 0.984 to 0.995, are slightly increasing with the term, and imply a mean half-life of shocks between 43 and 138 days.\(^{18}\) This confirms that mean reversion is present in the time series and suggests that long-term variance swap rates are more persistent than short-term rates. Comparing in- and out-of-sample statistics reveals a significant increase in level and volatility of variance swap rates, mainly due to the market turmoil in fall 2008.

A Principal Component Analysis (PCA) shows that the first principal component explains about 95.3% of the total variance of variance swap rates and can be interpreted as a level factor, while the second principal component explains an additional 3.8% and can be interpreted as a slope factor.\(^{19}\) This finding is somehow expected because PCA of several other term structures, such as bond yields, produces qualitatively similar results. Less expected is that two factors explain nearly all the variance of variance swap rates, i.e., 99.1%. Repeating the PCA for various subsamples produces little variation in the first two factors and explained total variance.

Table 1 also shows summary statistics of variance swap floating legs, i.e., the realized variance of daily S&P 500 returns for various terms. All statistics of realized variances share qualitatively the same features as those of the variance swap rates. The main difference is that, especially during the in-sample period, realized variances tend to be lower and more volatile, positively skewed

\(^{18}\) The half-life $H$ is defined as the time necessary to halve a unit shock and solves $\varrho^H = 0.5$, where $\varrho$ is the first order autocorrelation coefficient.

\(^{19}\) To save space, factor loadings are not reported, but are available from the authors upon request.
and leptokurtic than variance swap rates. This difference highlights the profitability and riskiness of shorting variance swaps, earning the large negative variance risk premiums embedded in such contracts. The ex-post variance risk premium is defined as the average realized variance minus the variance swap rate, which is simply the average payoff of a long position in the respective variance swap. The corresponding summary statistics are reported in the last panel of Table 1. In the in-sample period, ex-post variance risk premiums are negative and, except for the longest maturity, increasing in absolute value with the term. Notably, ex-post Sharpe ratios from shorting variance swaps also increase with their term, ranging from 0.60 (\(= 1.67/2.80\)) for two-month variance swaps to 0.85 (\(= 2.15/2.54\)) for one-year variance swaps. This suggests that it is more profitable on average to sell long-term than short-term variance swaps. In the out-of-sample period, the opposite holds as short-term variance swap rates increase proportionally more than long-term variance swap rates, making it more profitable, ex-post, to buy long-term variance swaps.

To summarize, the term structure of variance swap rates exhibits rich dynamics, challenging any term structure model. Whether our quadratic models are flexible enough to fit variance swap rates is an empirical question that we address in the following sections.

4.2 Model estimates

The state process \(X_t\) driving the term structure is not observed and variance swap rates and quadratic variation are nonlinear functions of \(X_t\). To extract the latent state we use the nonlinear unscented Kalman filter, which is found by Christoffersen et al. (2012) to have good finite sample properties in the context of estimating dynamic term structure models. We then estimate the model parameters using maximum likelihood.

The measurement equation entails a six-dimensional observation vector. The first five components are the variance swap rates with terms equal to two, three, and six months, and one and two years. The last component is the logarithm of the daily quadratic variation computed from tick-by-tick S&P 500 futures prices, entering the measurement equation as an affine function of \(\log(v_t^Q)\). We therefore use information both from the variance swap and S&P 500 futures markets to estimate the models. Appendix G provides details on the estimation method.

It is known that univariate affine models cannot capture the empirical features of variance swap rates, e.g., Egloff et al. (2010), and Aït-Sahalia et al. (2014). These models, for example, can
only produce upward or downward sloping term structures, and variance swap rates have all the same persistence. Such model-based features of variance swap rates are in sharp contrast with the empirical features summarized in Table 1. In principle our univariate quadratic model in Section 3.1 could capture the empirical features of variance swap rates. Intuitively, the quadratic features of the Q-spot variance \( v_t^Q \) and of the state process diffusion function relax the constraints imposed by an affine specification.

We begin model estimations by fitting each of the three canonical representations of the univariate quadratic model in Section 3.1 to variance swap rates and quadratic variation. We find that the largest log-likelihood of the univariate quadratic model is achieved when the state process \( X_t \) is in Class 3 (Theorem 3.3). This finding is confirmed by Akaike and Bayesian Information Criteria (AIC and BIC).\(^{20}\) Table 2 reports the corresponding parameters, which are estimated rather precisely.

We further investigate this model by considering four parametric restrictions that induce four alternative model specifications. Each specification is tested via a likelihood ratio (LR) test. Specification 1 imposes that \( X_t \) has an affine dynamic by setting the quadratic coefficient \( A = 0 \) in (15). Specification 2 constrains the Q-spot variance function, \( v_t^Q = \phi + \psi X_t + \pi X_t^2 \), to be linear in \( X_t \) by setting \( \pi = 0 \). The corresponding LR tests strongly reject both restrictions, suggesting that the quadratic features of \( X_t \) and \( v_t^Q \) play an important role in fitting variance swap rates and quadratic variation. Specification 3 restricts the functional form of the Q-spot variance by imposing the Q-spot variance function to have exactly one root, i.e., \( \psi^2 = 4 \phi \pi \). This guarantees the nonnegativity of the Q-spot variance for any realization of \( X_t \). Specification 4 further restricts Specification 3 by testing whether the root is at \( X_t = 0 \), i.e., \( \phi = \psi = 0 \). The corresponding LR tests strongly reject both restrictions, confirming that a flexible quadratic link between \( v_t^Q \) and \( X_t \) is statistically important to fit variance swap rates. To summarize, these statistical tests suggest that the full flexibility of the univariate quadratic model is necessary to fit variance swap rates and quadratic variation.

We now investigate whether enriching the functional form of the Q-spot variance can improve the fitting of the data. We estimate the univariate polynomial variance swap model in Section 3.3 when the state process \( X_t \) follows a quadratic diffusion and the degree of the polynomial is \( N = 5 \).

\(^{20}\)When the state process \( X_t \) is in Class 1, 2, and 3, AIC are \(-93,127\), \(-94,163\) and \(-94,264\), and BIC are \(-93,066\), \(-94,102\) and \(-94,180\), respectively. Both criteria achieved the minimum value when \( X_t \) is in Class 3.
The choice \( N = 5 \) asserts that the univariate polynomial model has the same number of parameters as the bivariate quadratic model, estimated next. Table 2 reports the parameter estimates.\(^{21}\) The additional parameters, \( p_3, p_4, p_5 \), allow only for a modest increase in the log-likelihood and a modest decrease of the AIC and BIC. Moreover, the economic magnitude of such parameters appears to be rather small. Thus, the polynomial form of the \( Q \)-spot variance helps only marginally to improve the fitting of variance swap rates and quadratic variation.

We now turn to the bivariate extension of the quadratic model in Section 3.2. We estimate all the identifiable equivalence class combinations of \( X_{1t} \) and \( X_{2t} \), and find that the best fit, in terms of likelihood, AIC and BIC, is obtained when \( X_{1t} \) is in Class 1 and \( X_{2t} \) is in Class 3. Table 2 reports the parameter estimates, as well as AIC and BIC. All the parameters are estimated precisely, as can be seen from the small standard errors.

The log-likelihood of the bivariate model is significantly larger than the log-likelihood of univariate models and the values of the BIC and AIC are significantly lower. The LR statistic of the bivariate model versus the univariate quadratic model is 28,757. The Vuong (1989) statistic of the bivariate model versus the univariate polynomial model is 48.5. Both statistics are highly significant and strongly reject the null hypothesis that the bivariate quadratic model is equivalent to any of the other two univariate models.\(^{22}\) Following Giacomini and White (2006), we also compare the bivariate model and the univariate models using scoring-type rules. The test statistic is the log-likelihood under the bivariate model minus the log-likelihood under the univariate quadratic or polynomial model. If the two models are equivalent, the test statistic has zero mean, which can be tested using a simple t-test.\(^{23}\) The t-statistics are 10.1 and 9.8, respectively, and are both highly significant. These tests further support that the bivariate quadratic model fits variance swap rates and quadratic variation significantly better than the univariate models.

Fig. 2 shows the filtered trajectories of the state process \( X_t \) in the bivariate model and suggests a

---

\(^{21}\)The relation between model parameters in Section 3.3 and those in Table 2 is straightforward, namely \( p_0 = \phi \), \( p_1 = \psi \) and \( p_2 = \pi \).

\(^{22}\)The asymptotic distribution of the test statistics under the null hypotheses is the chi-square with five degrees of freedom and standard normal, respectively. Recall that the bivariate quadratic model nests the univariate quadratic model. Setting \( b_2 = \beta_2 = a_2 = \alpha_2 = A_2 = 0 \) in the bivariate model, i.e., imposing five parameter restrictions, implies that \( X_{2t} \) is constant and can be normalized to one for identification purposes. Thus, \( \beta_1 \) in the bivariate model corresponds to \( b_1 \) in the univariate model.

\(^{23}\)We view this test as a robustness check of the previous LR and Vuong’s tests. Given the autocorrelation and heteroskedasticity in the log-likelihood differences, robust standard errors are computed using the Newey and West (1987) variance estimator with the number of lags optimally chosen according to Andrews (1991).
natural interpretation of its components. \( X_{1t} \) is more volatile and mimics the time series trajectories of short-term variance swap rates, mainly capturing sudden movements in those rates. \( X_{2t} \) is more persistent and mainly captures long-term movements in variance swap rates.

### 4.3 Two-factor affine jump-diffusion model

We now introduce a two-factor affine jump-diffusion (AJD) model that provides a challenging benchmark to assess the accuracy of our quadratic models in subsequent goodness-of-fit tests. The \( \mathbb{Q} \)-dynamics (1) of the index are specified as

\[
\begin{align*}
\frac{dS_t}{S_{t-}} &= r_t \, dt + \sqrt{X_{1t}} \, dB_t + \xi_t \, dN_t - \mathbb{E}^\mathbb{Q}[\xi_t] \nu_t \, dt, \\
\frac{dX_{1t}}{X_{1t}} &= \beta_1 (X_{1t} - X_{2t}) \, dt + \sqrt{\alpha_1 X_{1t}} \, dW_{1t} + Z_{1t} \, dN_t, \\
\frac{dX_{2t}}{X_{2t}} &= (b_2 + \beta_2 X_{2t}) \, dt + \sqrt{\alpha_2 X_{2t}} \, dW_{2t},
\end{align*}
\]

for some standard Brownian motions \( B_t \) and \( (W_{1t}, W_{2t}) \). The first factor \( X_{1t} \) is the diffusive component of the \( \mathbb{Q} \)-spot variance and follows a two-factor mean reverting process in which the second factor \( X_{2t} \) controls its stochastic long run mean and \( \beta_1 < 0 \) the speed of mean reversion. The second factor \( X_{2t} \) follows its own stochastic mean reverting process and mean reverts to \(-b_2/\beta_2 > 0\), with speed of mean reversion \( \beta_2 < 0 \). The relative index jump size \( \xi_t > -1 \) can be any integrable random variable. Only the second moment of the log-jump size, \( \mathbb{E}^\mathbb{Q}[(\log(1 + \xi_t))^2] = \mu_{2S} \), enters the \( \mathbb{Q} \)-spot variance. The variance jump size \( Z_{1t} \) is exponentially distributed with parameter \( \mu_{1Q} \), ensuring that variance stays positive. Jump sizes \( \xi_t \) and \( Z_{1t} \) are independent from Brownian motions and jump times.

Jumps in returns and variance occur contemporaneously, triggered by \( dN_t \), as in the double-jump model introduced by Duffie et al. (2000).\(^{24}\) The intensity of the counting jump process \( N_t \) is stochastic and given by \( \nu_t = \nu_0 + \nu_1 X_{1t} \), where \( \nu_0 \) and \( \nu_1 \) are nonnegative constants.\(^{25}\)

As in our bivariate quadratic model (Section 3.2), we use the market price of risk specification in (21) and attach the risk premium \( \lambda_1 \sqrt{X_{1t}/\alpha_1} \) to the \( \mathbb{Q} \)-Brownian motion \( W_{1t} \). Thus, the

\(^{24}\)Eraker et al. (2003) fit models with contemporaneous and independent jumps in returns and variance to S&P 500 data. They find that the two models perform similarly, but the model with contemporaneous jumps is estimated more precisely. Eraker (2004), Broadie et al. (2007), Chernov et al. (2003), and Todorov (2010) provide further evidence for contemporaneous jumps in returns and variance.

\(^{25}\)Rewriting the dynamics of \( X_{1t} \) in terms of the compensated jump component, i.e., \( Z_{1t} \, dB_t - \mu_{1Q}(\nu_0 + \nu_1 X_{1t}) \, dt \), shows that the speed of mean reversion is \((-\beta_1 - \mu_{1Q}^2 \nu_1)\), and the stochastic long run mean is \((-\beta_1 X_{2t} + \mu_{1Q}^2 \nu_0)/(-\beta_1 - \mu_{1Q}^2 \nu_1)\).
Girsanov-transformed $\mathbb{P}$-Brownian motion is $dW_t^\mathbb{P} = dW_t^1 - \lambda_1 \sqrt{X_{1t}}/\sqrt{\alpha_1} \, dt$. As variance jumps are the main feature of AJD models to generate volatility of volatility, we also allow for a variance jump risk premium. Specifically, under $\mathbb{P}$ the variance jump size $Z_{1t}$ is exponentially distributed with parameter $\mu_1^P$.

The logarithm of the $\mathbb{P}$-spot variance is an affine function of $\log(v_t^Q)$, where the $Q$-spot variance $v_t^Q = X_{1t} + \mu_2 S(\nu_0 + \nu_1 X_{1t})$, which is an affine function of $X_{1t}$. A number of empirical studies have shown that diffusive affine models do not generate enough volatility of volatility. The jump component $Z_{1t} \, dN_t$ in $X_{1t}$ can produce quick upward movements of the spot variance, increasing volatility of volatility relative to a diffusive affine specification. In our bivariate quadratic model, the factors have a quadratic, rather than affine, diffusion, and the $Q$-spot variance $v_t^Q$ is a quadratic function of $X_{1t}$, and does not exhibit jumps; see (20). The quadratic features of the factors and $Q$-spot variance generate more volatility of volatility relative to diffusive affine specifications. Which modeling approach is more suitable for fitting variance swap rates and quadratic variation is an empirical question that we address below.

Model (25) is a challenging benchmark and subsumes many existing stochastic volatility models along most dimensions. Nearly no study allows at the same time for stochastic long run mean, stochastic jump intensity and jumps in returns and variance. Bakshi et al. (1997), Bates (2000, 2006), Pan (2002), Eraker et al. (2003), Eraker (2004), Broadie et al. (2007, 2009) set $X_{2t}$ to a constant positive value. Almost all studies assume either constant jump intensities (e.g., Eraker et al. (2003) and Broadie et al. (2007)) or jumps in returns but not in variance (e.g., Pan (2002) and Broadie et al. (2009)). Recently, Andersen et al. (2014) introduce a three-factor stochastic volatility model with a flexible price jump specification, which is shown to outperform the price jump specification in (25).

Model (25) is fitted to variance swap rates and quadratic variation, as all other models in Section 4.2. Given the presence of jumps in the spot variance, the model is estimated using the particle filter method in Bardgett et al. (2013). Table 3 reports the parameter estimates. The diffusive variance $X_{1t}$ is more volatile and fast mean reverting than the second factor $X_{2t}$ that controls its long run mean. Jumps are estimated to be rare events, as one jump occurs on average once every 3.8 years ( = $1/(\nu_0 + \nu_1 \mathbb{E}^\mathbb{P}[X_{1t}]$)). These findings are broadly consistent with estimates

\[ \mathbb{E}^\mathbb{P}[X_{1t}] = (\beta_2 b_2/\beta_2 + \mu_1^P \nu_0)/(-\beta_1^P), \]

where $\beta_1^P = \beta_1 + \mu_1^P \nu_1 + \lambda_1$.  

\[ \text{26} \]
of similar models reported in the literature. Importantly, in terms of likelihood, AIC and BIC, our bivariate quadratic model outperforms the two-factor affine jump-diffusion model, which in turn outperforms the univariate models.

4.4 Goodness-of-fit tests

To corroborate the above likelihood-based analysis, we now analyze the variance swap pricing errors for the various models and run various goodness-of-fit tests.

Table 4 summarizes the pricing errors, which are defined as model-based minus actual variance swap rates, both in volatility units. Consistently with the likelihood-based analysis, the bivariate quadratic model nearly always, significantly outperforms the other models in terms of bias and root mean square error (RMSE). For example, in the out-of-sample period, the RMSE of the bivariate quadratic model for the two-month variance swap rates is 60% lower than the RMSE of the univariate model. The comparison among the bivariate quadratic, univariate polynomial and two-factor AJD models is particularly interesting, as the three models have the same number of parameters. In most cases, the RMSE of the bivariate model is less than half the RMSE of the polynomial model. In the in-sample period, the bivariate quadratic model largely outperforms the two-factor AJD model. In the out-of-sample period, the bivariate quadratic model tends to outperform the two-factor AJD model that proves to be a challenging benchmark and dominates univariate models for most variance swap terms.

Fig. 3 shows actual and model-based trajectories under the bivariate quadratic model of the two-month and two-year variance swap rates, which are respectively the most and least volatile rates. The good performance of the model is evident throughout the in-sample and out-of-sample periods. A small lack of fit of the highest values of the two-year variance swap rates in the out-of-sample period is noticeable and occurs during the market tumults of fall 2008.

To assess the statistical differences of the model pricing errors, we run various Diebold–Mariano (DM) tests.\textsuperscript{27} For each model and each term, the time-\(t\) loss function is given by the absolute pricing error, \(L(e_t) = |e_t|\), where \(e_t = \sqrt{G(\tau, X_t)/\tau} - \sqrt{VS(t, t+\tau)}\).\textsuperscript{28} Denote the time-\(t\) loss differential

\textsuperscript{27}We follow standard practice and use Diebold–Mariano tests to draw conclusions about models, rather than about model forecasts; see Diebold (2012) for a discussion of this point.
\textsuperscript{28}The time-\(t\) pricing error considered here uses the time-\(t\) filtered value of \(X_t\), not its prediction as in the log-likelihood (G.11), which makes the DM tests complementary to the likelihood-based analysis in the previous section.
between the univariate and bivariate quadratic models by $d_t^{(u,b)} = L(e_t^{(u)}) - L(e_t^{(b)})$. The loss differential between the polynomial and bivariate models, $d_t^{(p,b)}$, and between the two-factor AJD and bivariate models, $d_t^{(a,b)}$, are similarly defined. Under the null hypothesis that the two models have pricing errors of equal magnitude, $E^P[d_t^{(u,b)}] = 0$. If the bivariate model outperforms the univariate model, then $E^P[d_t^{(u,b)}] > 0$. The DM statistic is the t-statistic for this test. Table 4 reports the results. DM tests strongly confirm that the bivariate model largely outperforms the univariate quadratic and polynomial models. In the in-sample period the bivariate quadratic model consistently outperforms the two-factor AJD model. In the out-of-sample period the bivariate model outperforms the two-factor AJD model in three out of five terms, and has statistically the same performance in the other two terms. As a robustness check, we also run DM tests using pricing errors in variance units, rather than volatility units, i.e., $e_t = G(\tau, X_t)/\tau - VS(t, t + \tau)$, and using quadratic loss functions, rather than absolute loss functions. These additional DM tests largely confirm the results in Table 4.

Finally, we run predictive regressions for each model and each term. We regress the actual future variance swap rate $VS(t, t + \tau)$ on a constant and the $d$-day ahead, model-based prediction, $E^P[G(\tau, X_t)/\tau | \mathcal{F}_{t-d}]$, obtained at time $t - d$, i.e.,

$$VS(t, t + \tau) = \gamma_0 + \gamma_1 E^P[G(\tau, X_t)/\tau | \mathcal{F}_{t-d}] + \text{error}_t.$$  \hfill (26)

If the model captures well the variance swap term structure dynamics, then it should provide unbiased, $\gamma_0 = 0$, and efficient, $\gamma_1 = 1$, forecasts of future variance swap rates. As an additional benchmark in the context of predictive regressions, we consider the martingale model that uses the actual variance swap rate at time $t - d$ as a predictor of the future variance swap rate. The martingale model is a challenging benchmark because of the strong persistence of variance swap rates; first order autocorrelations of variance swap rates range from 0.984 to 0.995, Section 4.1.

We consider two forecasting horizons, $d = 1$ day and $d = 10$ days. Table 5 reports the regression

---

29 The standard error is computed using the Newey and West (1987) autocorrelation and heteroskedasticity consistent variance estimator with the number of lags optimally chosen according to Andrews (1991).

30 The DM test statistics are positive but not significant only for the six month and one year variance swaps in the out-of-sample period, which can be due to the limited sample size, i.e., 794 daily observations.
Notably, for both forecasting horizons and nearly all terms, the bivariate quadratic model provides unbiased and efficient variance swap rate forecasts, as can be seen from the high p-values of the null hypotheses $H_0 : \gamma_0 = 0$, and $H_0 : \gamma_1 = 1$. The univariate quadratic and polynomial models provide biased and inefficient forecasts in most cases, as can be seen from the low p-values. The martingale model provides relatively accurate forecasts for the one-day horizon, most persistent, long-term variance swap rates, but its forecasting accuracy largely deteriorates when moving to the ten-day horizon. The two-factor AJD proves again to be a challenging model, passing most tests. However, only the bivariate quadratic model passes all tests at 10% confidence level. To summarize, also predictive regressions strongly confirm that the bivariate quadratic model captures well the variance swap term structure dynamics.

5 Optimal portfolios: Theoretical setup

In this section, we study dynamic optimal investment in variance swaps, index option, stock index and risk-free bond. Because the stock index can jump and variance swaps are only sensitive to the quadratic variation, variance swaps cannot be used to hedge index jumps. Options on the stock index, such as out-of-the-money put options, are typically used to hedge this jump risk. We therefore allow the investor to dynamically trade variance swaps, index option, stock index and risk-free bond. As becomes clear later, all these securities are necessary to span the risk in the economy and achieve market completeness. While our primary interest is on variance swap investment, studying optimal portfolios in all these securities allows us to have a comprehensive view on the optimal portfolios in Egloff et al. (2010) and Liu and Pan (2003).32

We now formalize and solve the dynamic optimal portfolio problem. As at the beginning of Section 3, we consider a diffusion process $X_t$ in some state space $\mathcal{X} \subset \mathbb{R}^m$, solving the SDE (8), where $W_t$ is a standard $d$-dimensional $\mathbb{Q}$-Brownian motion. The $\mathbb{Q}$-spot variance, $v_t^\mathbb{Q}$, and variance swap rates, VS($t, T$), are given as functions of the state variable, $X_t$, by (11) and (12), respectively.

31 Also in these regressions, robust standard errors are computed using the Newey and West (1987) covariance matrix estimator with the number of lags optimally chosen according to Andrews (1991). Given the strong persistence of variance swap rates, all $R^2$ of predictive regressions are high, between 70% and 99%, and not reported.

5.1 Investing in variance swaps

We first compute the return of an investment in variance swaps. Fix a term \( \tau > 0 \), and consider a \( \tau \)-variance swap issued at some inception date \( t^* \). Denote its maturity \( T^* = t^* + \tau \). The nominal spot value \( \Gamma_t \) at date \( t \in [t^*, T^*) \) of a one dollar notional long position in this variance swap is given by

\[
\Gamma_t = \mathbb{E}^Q \left[ e^{-r(T^*-t)/\tau} \left( \int_{t^*}^{T^*} v_s^Q \, ds - \tau VS(t^*, T^*) \right) \bigg| \mathcal{F}_t \right] = e^{-r(T^*-t)/\tau} \left( \int_{t^*}^{t} v_s^Q \, ds + (T^*-t)VS(t, T^*) - \tau VS(t^*, T^*) \right),
\]

(27)

where the risk-free rate \( r \) is constant for simplicity. In stochastic differential form, we obtain

\[
d\Gamma_t = \Gamma_t r \, dt + dM_t \quad \text{with the } \mathbb{Q}\text{-martingale increment excess return}
\]

\[
dM_t = e^{-r(T^*-t)/\tau} \left( v_t^Q \, dt + \frac{e^{-r(T^*-t)/\tau}}{\tau} \nabla_x G(T^* - t, X_t) \Sigma(X_t) \, dW_t, \right.
\]

(28)

where \( \nabla_x \) denotes the gradient. Now fix a date \( t \in [t^*, T^*) \), and consider an investor with positive wealth \( V_t \) who takes a position in this variance swap with relative notional exposure of \( n_t \). The cost of entering such a position is \( n_t V_t \Gamma_t \). The remainder of the wealth, \( V_t - n_t V_t \Gamma_t \), is invested in the risk-free bond, making the investment self-financing. At a later instant \( t + dt \), the wealth has grown to \( V_{t+dt} = (V_t - n_t V_t \Gamma_t) (1 + r \, dt) + n_t V_t \Gamma_{t+dt} \). The resulting rate of return is

\[
\frac{dV_t}{V_t} = \frac{V_{t+dt} - V_t}{V_t} = (1 - n_t \Gamma_t) r \, dt + n_t \, d\Gamma_t = r \, dt + n_t \, dM_t.
\]

(29)

Consider now \( \tau \)-variance swaps that are issued at a sequence of inception dates \( 0 = t^*_0 < t^*_1 < \cdots \), with \( t^*_{k+1} - t^*_k \leq \tau \), for example three-month variance swaps issued every month. At any date \( t \in [t^*_k, t^*_{k+1}) \) the investor takes a position in the respective on-the-run \( \tau \)-variance swap with maturity \( T^*(t) = t^*_k + \tau \). In the limit case where a new \( \tau \)-variance swap is issued at any date \( t \), we obtain a “sliding” variance swap investment, and we set \( T^*(t) = t + \tau \). Iterating the above reasoning shows that the resulting wealth process \( V_t \) evolves according to

\[
\frac{dV_t}{V_t} = r \, dt + n_t \frac{e^{-r(T^*(t)-t)/\tau}}{\tau} \nabla_x G(T^*(t) - t, X_t) \Sigma(X_t) \, dW_t,
\]

(30)

where the excess return on the right hand side is a \( \mathbb{Q}\)-martingale increment.
5.2 Optimal portfolio problem

We now consider an investment universe consisting of stock index $S$, risk-free bond, index option $O$, and $n$ on-the-run variance swaps with different terms $\tau_1 < \cdots < \tau_n$ and respective issuance dates encoded by $n$ maturity functions $T^*_1(t), \ldots, T^*_n(t)$, as defined above.

The $\mathbb{Q}$-dynamics (1) of the index are specified as

$$\frac{dS_t}{S_{t-}} = r dt + \sigma(X_t) R(X_t)^\top dW_t + \xi (dN_t - \nu^Q(X_t) dt), \tag{31}$$

where $\sigma(X_t)^2$ is the diffusive component of the $\mathbb{Q}$-spot variance $\nu^Q_t = \sigma(X_t)^2 + (\log(1 + \xi))^2 \nu^Q(X_t)$, $R = (R_1, \ldots, R_d)^\top : \mathcal{X} \to \mathbb{R}^d$ is some function with constant norm $\|R\| \equiv 1$, modeling the correlation between stock returns and diffusive variance changes. The last term in (31) is a $\mathbb{Q}$-compensated jump component. The random arrival of jump events is induced by the counting process $N_t$, which has a stochastic intensity $\nu^Q(X_t)$. Following Liu and Pan (2003), we adopt a deterministic relative index jump size $\xi > -1$. That is, conditional upon a jump arrival, the stock index jumps from $S_{t-}$ to $S_{t-}(1 + \xi)$. This specification of a deterministic jump size simplifies our analysis in the sense that only one index option is needed to complete the market with respect to the jump component. This formulation, though simple, is capable of capturing the sudden and high-impact nature of index jumps that cannot be produced by diffusions. More generally, one could introduce random jump size with multiple values and use multiple options to complete the market.

The index option is left unspecified at this stage, but to fix ideas one can think of it as an out-of-the-money put option on the stock index, as will be the case in our empirical analysis of optimal portfolios. Let $O_t = O(S_t, X_t)$ be the time-$t$ value of the option. The $\mathbb{Q}$-dynamics of $O_t$ are

$$dO_t = r O_t dt + \left( \partial_s O_t S_t \sigma(X_t) R(X_t)^\top + \nabla_x O_t^\top \Sigma(X_t) \right) dW_t + \Delta O_t (dN_t - \nu^Q(X_t) dt), \tag{32}$$

where $\partial_s O_t$ and $\nabla_x O_t$ measure the sensitivity of the option price to infinitesimal changes in the stock index and state variables, respectively, and $\Delta O_t$ measures the change in the option price.

---

33 The index price dynamics in (31) are equivalent to $dS_t/S_{t-} = r dt + \sigma(X_t) dB_t + \xi (dN_t - \nu^Q(X_t) dt)$ for the scalar $\mathbb{Q}$-Brownian motion $B_t$ defined as $dB_t = R(X_t)^\top dW_t$. That is, $B_t$ and $W_t$ have correlation $d\langle B, W_t \rangle_t/dt = R_k(X_t)$. 26
when the underlying stock index jumps,

\[
\partial_s O_t = \frac{\partial}{\partial s} O(S_t, X_t), \quad \nabla_x O_t = \nabla_x O(S_t, X_t), \quad \Delta O_t = O(S_t - (1 + \xi), X_t) - O(S_{t-}, X_t). \tag{33}
\]

When the option has nonzero sensitivities \(\partial_s O_t, \nabla_x O_t\) and \(\Delta O_t\), it provides exposure to the fundamental sources of risk, \(W_t\) and \(N_t\), and access to their risk premiums.

Let \(w_t\) denote the fraction of wealth invested in the stock index, \(\phi_t\) the fraction of wealth invested in the option, and \(n_t = (n_{1t}, \ldots, n_{nt})^\top\) the vector of relative notional exposures to each on-the-run \(\tau_i\)-variance swap, \(i = 1, \ldots, n\). To make the investment self-financing, the fraction of wealth invested in the risk-free bond is given by \(1 - n_t^\top \Gamma_t - w_t - \phi_t\), where \(\Gamma_t\) is the vector of the variance swap spot values. Combining (30), (31) and (32), the resulting wealth process \(V_t\) has \(Q\)-dynamics

\[
\frac{dV_t}{V_{t-}} = n_t^\top d\Gamma_t + w_t \frac{dS_t}{S_t} + \phi_t \frac{dO_t}{O_{t-}} + (1 - n_t^\top \Gamma_t - w_t - \phi_t) r \, dt
\]

\[
= r \, dt + \theta_t^W dW_t + \theta_t^N \xi (dN_t - \nu Q(X_t) \, dt),
\]

where \(\theta_t^W\) and \(\theta_t^N\) are defined, for given portfolio weights \(n_t, w_t,\) and \(\phi_t\), by

\[
\begin{pmatrix}
\theta_t^W \\
\theta_t^N
\end{pmatrix} = G_t \begin{pmatrix} n_t \\
w_t \\
\phi_t
\end{pmatrix},
\tag{35}
\]

the \((d+1) \times (n+2)\) matrix \(G_t\) is defined by

\[
G_t = \begin{pmatrix}
\Sigma(X_t)^\top & \sigma(X_t) R(X_t) & 0_{d \times 1} \\
0_{1 \times m} & 0 & 1
\end{pmatrix}
\begin{pmatrix}
D_t & 0_{m \times 1} & 0_{d \times 1} \nabla_x O_t \\
0_{1 \times n} & 1 & \frac{\partial O_t}{\partial t} \\
0_{1 \times n} & 1 & \frac{\Delta O_t}{\xi}
\end{pmatrix},
\tag{36}
\]

and \(D_t\) is the \(m \times n\) matrix whose \(i\)th column is given by \(\left(e^{-r(T_i^*(t) - t)/\tau_i}\right) \nabla_x G(T_i^*(t) - t, X_t)\). Effectively, by taking positions \(n_t, w_t\) and \(\phi_t\) on the risky assets, the investor invests \(\theta_t^W\) in the diffusive shocks \(W_t\), and \(\theta_t^N\) in the jump risk \(N_t\), controlling the portfolio exposure to the fundamental risks.

We now formulate the optimal portfolio problem. We consider an investor who has a fixed finite time horizon \(T\), maximizes his expected utility from terminal wealth, and has a power utility
function with constant relative risk aversion $\eta$. That is, the investment objective is

$$\max_{\{m, w_t, \phi_t, 0 \leq t \leq T\}} \mathbb{E}_P \left[ \frac{V_T^{1-\eta}}{1-\eta} \right],$$

(37)

for some given initial wealth $V_0$. The objective probability measure $P$ is related to the risk neutral measure $Q$ via the pricing kernel

$$\frac{d\pi_t}{\pi_t} = -r dt - \Lambda(X_t)^T dW_t^P + \left( \frac{\nu^Q(X_t)}{\nu^P(X_t)} - 1 \right) \left( dN_t - \nu^P(X_t) dt \right),$$

(38)

where $\nu^P(X_t)$ is the jump intensity of $N_t$ under $P$, and $dW_t^P = dW_t - \Lambda(X_t) dt$ is a $P$-Brownian motion. The pricing kernel $\pi_t$ sets the risk premiums in the economy, with $\Lambda(X_t)$ and $\nu^Q(X_t)/\nu^P(X_t)$ controlling the premium for diffusive and jump risks, respectively. As usual in the optimal portfolio literature, we exogenously specify the risk premiums in (38), and our analysis of optimal portfolios is of a partial equilibrium nature. That is, the investor solving (37) takes the risk premiums as given. As pointed out by Liu and Pan (2003, Page 403), “this is very much the spirit of the asset allocation problem: a small investor takes prices (both risks and returns) as given and finds for himself the optimal trading strategy.” Chacko and Viceira (2005), Liu (2007), Aït-Sahalia et al. (2009), Detemple and Rindisbacher (2010), among others, provide studies of optimal portfolios in partial equilibrium settings.

By choosing the number $n$ of on-the-run variance swaps available in the market according to the number $d$ of driving Brownian motions, it allows to achieve market completeness. Market completeness in turn allows us to solve the optimal portfolio problem analytically.

**Assumption 5.1.** The market is complete with respect to the stock index, the index option, and the $n$ on-the-run $\tau_i$-variance swaps. Specifically, we assume that $n = m = d - 1$, and that the $(d + 1) \times (d + 1)$ matrix $G_t$ is invertible $dt \otimes dQ$-a.s.

From (36) we see that $G_t$ is invertible $dt \otimes dQ$-a.s. if and only if the $d \times d$ matrix $\left( \Sigma(X_t)^T, \sigma(X_t)R(X_t) \right)$ and the $(d - 1) \times (d - 1)$ matrix $D_t$ are invertible $dt \otimes dQ$-a.s. and

$$\frac{\partial_s O_t}{\xi S_t} \neq \frac{\Delta O_t}{\xi S_t} \ dt \otimes dQ\text{-a.s.}$$

(39)

\footnote{Here we use the fact that $\frac{\Delta O_t}{\xi S_t} = \frac{\Delta O_t}{\xi S_{t-}}$ and $\partial_s O_t = \partial_s O_{t-} \ dt \otimes dQ\text{-a.s.}$}
The matrix $\mathcal{D}_t$ is invertible $dt \otimes d\mathbb{Q}$-a.s. only if the maturity date functions $T^*_i(t)$ are mutually different for all $t$. This means that none of the $n = d - 1$ on-the-run $\tau_i$-variance swaps is redundant. Condition (39) states that the option price has to exhibit different sensitivities with respect to large and small index price changes. For convex option prices $\frac{\Delta O_t}{\xi_{S_t}} > \partial_s O_t$ and (39) holds.

Because of market completeness, the control variables $n_t$, $w_t$, $\phi_t$ in the optimal portfolio problem (37) can be replaced by $\theta_t^W$, $\theta_t^N$. The solution of (37) then consists of two logical steps. First, find the optimal exposures $\theta_t^{W*}$ and $\theta_t^{N*}$ to the fundamental risk factors $W_t$ and $N_t$ to support the optimal wealth dynamics. Second, invert (35) to obtain the optimal positions $n_t^*$, $w_t^*$, and $\phi_t^*$ in variance swaps, stock index, and option. The indirect utility function for (37) is

$$J(t,v,x) = \max_{\{\theta_s^W, \theta_s^N, t \leq s \leq T\}} \mathbb{E}^\mathbb{P}\left[ \frac{V^1 - \eta}{1 - \eta} \bigg| V_t = v, X_t = x \right],$$

which satisfies the Hamilton–Jacobi–Bellman (HJB) equation

$$0 = \max_{\theta^W, \theta^N} \left\{ \frac{\partial J}{\partial t} + \frac{\partial J}{\partial v} \left( r + \theta^{W^\top} \Lambda(x) - \theta^N \xi \nu \mathbb{Q}(x) \right) + \frac{1}{2} \frac{\partial^2 J}{\partial v^2} v^2 \theta^{W^\top} \theta^W ight.$$ 

$$+ \nabla_x J^\top (\mu(x) + \Sigma(x) \Lambda(x)) + \frac{1}{2} \sum_{i,j=1}^m \frac{\partial^2 J}{\partial x_i \partial x_j} \left( \Sigma(x) \Sigma(x)^\top \right)_{ij}$$

$$+ \theta^{W^\top} \nu \Sigma(x)^\top \nabla_x \left( \frac{\partial J}{\partial v} \right) + \left( J(t,v(1+\theta^N \xi),x) - J(t,v,x) \right) \nu \mathbb{P}(x) \right\},$$

where arguments of $J$ are omitted for simplicity. We now state the existence and characterization result for the optimal strategy with standard technical assumptions and proof given in Appendix H.

**Theorem 5.2.** Under Assumptions 5.1, and H.1 and H.2 in Appendix H, there exists an optimal strategy $n_t^*$, $w_t^*$, $\phi_t^*$ given by inverting (35) where

$$\theta_t^{W*} = \frac{1}{\eta} \Lambda(X_t) + \Sigma(X_t)^\top \nabla_x h(T-t,X_t),$$

$$\theta_t^{N*} = \frac{1}{\xi} \left( \frac{\nu \mathbb{P}(X_t)}{\nu \mathbb{Q}(X_t)} \right)^{1/\eta} - 1,$$

and the function $h(\tau,x)$ is defined in (H.3) in Appendix H.

The optimal exposure $\theta_t^{W*}$ to the diffusive risk is composed of the familiar myopic and intertemporal hedging terms, as discussed in Merton (1971). The myopic demand, coming from $\Lambda(X_t)/\eta$,
would be the mean-variance optimal investment over the next instant not accounting for future investments, or assuming a constant investment opportunity set. The intertemporal hedging demand, coming from \( \Sigma(X_t)\top \nabla_x h(T-t, X_t) \), arises due to the need to hedge against fluctuations in the investment opportunities. These fluctuations are induced, inter alia, by the stochastic diffusive component of the volatility of the stock index. We discuss the computation of \( \nabla_x h(T-t, X_t) \) in Appendix I. The optimal exposure \( \theta^*_t \) to the jumps only consists of a myopic term.

The following corollary shows that variance swaps can be used to span diffusive volatility risk. The optimal investments in the stock index and index option are thus only seeking the diffusive and jump risk premiums. The proof is given in Appendix J.

**Corollary 5.3.** The optimal investment in the stock index and index option, \( w^*_t \) and \( \phi^*_t \), is fully determined by the myopic terms and does not depend on the choice of the variance swaps.

The optimal exposure to jump risk, \( \theta^*_t \), is the same as the optimal exposure derived by Liu and Pan (2003), with the only difference being that in our setting the jump intensity is stochastic. If the jump risk is not priced, i.e., \( \nu^Q(X_t)/\nu^P(X_t) = 1 \), then \( \theta^*_t = 0 \) and the optimal wealth process does not jump. The investor optimally decides not to invest in the jump risk, because it does not carry any risk premium. In this case, (35) implies that

\[
\phi^*_t = -\frac{\xi}{\Delta O_t/O_t^-} w^*_t.
\] (43)

The fraction on the right hand side is the ratio of the relative jump sizes, \( \xi \) and \( \Delta O_t/O_t^- \), of the index and option prices. Suppose that \( \xi < 0 \), and the option is a put, \( \Delta O_t > 0 \). Then \( \phi^*_t \) has the same sign as \( w^*_t \). If \( w^*_t > 0 \), the investor takes a long position in the stock index, earning the equity risk premium, and a long position in the put option, hedging the jump risk. The optimal long position in the put option \( \phi^*_t \) is increasing in the absolute relative index jump size, \( -\xi \), and is decreasing in the relative option price change upon a jump, \( \Delta O_t/O_t^- \). If the latter is small, a large fraction of wealth needs to be allocated to the put option to hedge the jump risk.

If the jump risk is priced, \( \nu^Q(X_t)/\nu^P(X_t) > 1 \), then \( \theta^*_t \) is negative, and the optimal wealth process exhibits negative jumps. The investor optimally takes on jump risk to earn its risk premium. Because \( \theta^*_t \xi > -1 \), the optimal wealth level is always one jump away from being negative. Suppose
again that $\xi < 0$, and the option is a put, $\Delta O_t > 0$. The optimal investment in the put option is

$$
\phi_t^* = -\frac{\xi}{\Delta O_t/O_t} (w_t^* - \theta_t^{N*}).
$$

As can be seen in (42), $\theta_t^{N*}$ is increasing in the jump risk premium $\nu^Q(X_t)/\nu^P(X_t)$ and in the risk tolerance, $1/\eta$. If $w_t^* > 0$ and the jump risk premium and/or risk tolerance are low, then $w_t^* - \theta_t^{N*} > 0$, and the investor still takes a long position in the put option to hedge index jump risk. If instead the jump risk premium and/or risk tolerance are high, then $w_t^* - \theta_t^{N*} < 0$, and the investor optimally takes a short position in the put option to earn the jump risk premium.

### 5.3 Bivariate quadratic model specification

We now resume the bivariate quadratic variance swap model in Section 3.2. Our empirical analysis in Section 4 shows that the best fit is attained when $X_{1t}$ is in Class 1 and $X_{2t}$ is in Class 3. We focus on this specification in the following. The dimension of the Brownian motion $W_t$ is $d = 3$, and the $2 \times 3$ dispersion matrix $\Sigma(x)$ takes the form

$$
\Sigma(x) = \begin{pmatrix}
\sqrt{1 + A_1x_1^2} & 0 & 0 \\
0 & \sqrt{x_2 + A_2x_2^2} & 0
\end{pmatrix}.
$$

(45)

We specify the $P$-jump intensity as $\nu^P(x) = \nu^P \sigma(x)^2$ where $\nu^P$ is a positive constant. This specification allows for more jumps to occur during more volatile periods, as shown in the empirical literature. Similarly, the $Q$-jump intensity is set equal to $\nu^Q(x) = \nu^Q \sigma(x)^2$. The diffusive spot variance is thus proportional to the $Q$-spot variance function,

$$
\sigma(x)^2 = \frac{g(x)}{1 + (\log(1 + \xi))^2 \nu^Q}.
$$

(46)

In our empirical analysis of optimal strategies, we set the jump intensities $\nu^P = 0.5$, $\nu^Q = 0.7$, and the index jump size $\xi = -25\%$, similarly to Liu and Pan (2003). These parameters imply that one large index jump occurs on average once every 50 years. In Section 6.2 we experiment with other jump configurations, namely smaller and more frequent index jumps. Our conclusions on the empirical features of optimal strategies remain largely unchanged.
To account for the widely documented correlation between index returns and diffusive variance changes, e.g., Broadie et al. (2007), and Ait-Sahalia and Kimmel (2007), the correlation vector function is chosen to be of the form \( \mathbf{R}(x) = \left( R_1(x), 0, \sqrt{1 - R_1(x)^2} \right)^\top \). The correlation between index returns and diffusive variance changes is then given by

\[
\text{Corr} \left( \frac{dS_t}{S_t}, d\sigma(X_t)^2 \right) = \frac{\nabla_x g(X_t)^\top \Sigma(X_t)}{\|\nabla_x g(X_t)^\top \Sigma(X_t)\|} \mathbf{R}(X_t) = \text{sign} (\psi + 2\pi X_{1t}) \, R_1(X_t). \tag{47}
\]

Consistently with (21), we specify the market price of diffusive risk function as

\[
\Lambda(x) = \left( \frac{\lambda_0 + \lambda_1 x_1}{\sqrt{1 + A_1 x_1}}, 0, \Lambda_3(x) \right)^\top,
\]

where \( \Lambda_3(x) \) is implicitly defined, up to its sign, by

\[
\Lambda_3(x) = \pm \sqrt{\|\Lambda(x)\|^2 - \Lambda_1(x)^2}.
\]

The sign of \( R_3(x) \Lambda_3(x) \) has a direct impact on the equity risk premium, which is given by

\[
\frac{\mathbb{E}^P[dS_t/S_t | \mathcal{F}_t] - \mathbb{E}^Q[dS_t/S_t | \mathcal{F}_t]}{dt} = \sigma(X_t) \mathbf{R}(X_t)^\top \Lambda(X_t) + \xi (\nu^P - \nu^Q) \sigma(X_t)^2.
\]

Based on our estimates, \( R_3(X_t) \Lambda_3(X_t) \) is larger in absolute value than \( R_1(X_t) \Lambda_1(X_t) \). Since \( R_3(x) \) is positive, a negative \( \Lambda_3(x) \) would lead to a negative equity risk premium, which would be economically odd, so we take the positive square root in (49). Clearly, \( \|\Lambda(x)\|^2 \) needs to be specified so that the argument in the square root in (49) is nonnegative for all \( x \in \mathcal{X} \). We specify it as proportional to the \( \mathcal{Q} \)-spot variance

\[
\|\Lambda(x)\|^2 = \kappa g(x), \tag{51}
\]

with \( \kappa \geq \kappa^* = \max_{x \in \mathcal{X}} \Lambda_1(x)^2/g(x). \) Since \( \Lambda_1(x) \) is uniformly bounded in \( x \), it follows that the \( \mathcal{Q} \)-spot variance \( g(x) \) and the equity risk premium (50) are increasing functions in \( x_1 \), for \( x_1 \) large enough. This means that the equity risk premium increases in bad times, i.e., when variance increases and stock index falls due to the leverage effect. Such a countercyclical equity risk premium is certainly a desirable feature of our model and motivates the chosen specification (51) of \( \|\Lambda(x)\|^2 \).
We set $\kappa$ in (51) to achieve a sample average of the equity risk premium (50) equal to 6%.\footnote{The equity risk premium is notoriously difficult to estimate. Merton (1980) even argues that a positive risk premium should be explicitly modeled, and various studies have followed this approach, e.g., Jackwerth (2000), and Barone-Adesi et al. (2008).}

The stock index (31) exhibits quadratic stochastic variance and jumps, which is outside the standard affine setting. To study the empirical features of the optimal trading strategy we develop a novel pricing formula for European options. The transition density of the index $S_t$ is approximated using an Edgeworth expansion, relying on closed form expressions for joint conditional moments of $S_t$ and state variables $X_t$. Appendix K provides the derivation of the option pricing formula and discusses the computation of the option price sensitivities $\partial_s O_t$, $\nabla_x O_t$ and $\Delta O_t$ in (33).

### 5.4 Optimal portfolios in the bivariate quadratic model

We assume that an index put option and $n = 2$ variance swaps are available for investment. The latter are specified by their maturity date functions $T_1^*(t)$ and $T_2^*(t)$. We allow for various roll-over strategies. In all cases the maturity date functions differ, $T_1^*(t) \neq T_2^*(t)$, for all $t$, which is important in view of Assumption 5.1. It is a tedious but routine exercise to check that all assumptions underpinning Theorem 5.2 are satisfied. Appendix L sketches the arguments.

The optimal fractions of wealth invested in the stock index and index option are given by

$$
\begin{align*}
    w_t^* &= \frac{1}{\eta} A_3(X_t) - \theta_t^{N*} \sigma(X_t) R_3(X_t) \partial_s O_t \xi S_t / \Delta O_t \\
    \phi_t^* &= \frac{-\xi O_t^-}{\Delta O_t} \times \left[ \frac{1}{\eta} A_3(X_t) - \theta_t^{N*} \sigma(X_t) R_3(X_t) \right] / \sigma(X_t) R_3(X_t) (1 - \partial_s O_t \xi S_t / \Delta O_t),
\end{align*}
$$

(52)

which is recovered by setting $v = (0, 0, 1)^T$ in (J.3) in Appendix J.

For a put option, the ratio $0 \leq \partial_s O_t \xi S_t^- / \Delta O_t < 1$, due to monotonicity and convexity. Thus the denominators are positive. If the jump risk premium is small then $\theta_t^{N*}$ is small, and the numerators are also positive. In that case, the investor optimally takes long positions in the stock index and put option, earning the equity risk premium and hedging the jump risk, respectively. If the jump risk premium is large then $\theta_t^{N*}$ is large, and the investor can optimally short the put option, earning the jump risk premium, and hedging the short put with a short position in the stock index.

The intertemporal hedging demand is fully borne by the optimal investment in variance swaps.
Plugging (52) in (35) shows that the optimal vector of relative notional exposures to the respective on-the-run variance swaps is given as solution \( \mathbf{n}_t^* = \mathbf{n}_t \) of the linear equation

\[
\Sigma(X_t)^\top D_t \mathbf{n}_t = \frac{1}{\eta} \left( \Lambda(X_t) - \frac{\Lambda_3(X_t)}{R_3(X_t)} \mathbf{R}(X_t) \right) - \frac{\phi_t^*}{O_t} \Sigma(X_t)^\top \nabla_x \mathbf{O}_t + \Sigma(X_t)^\top \nabla_x h(T - t, X_t). \tag{53}
\]

We provide a closed form approximation of \( \nabla_x h(T - t, X_t) \) in (L.3) in Appendix L.

### 6 Optimal portfolios: Empirical findings

We now perform an empirical analysis of optimal portfolios using the above bivariate quadratic model. The investment universe consists of the stock index, risk-free bond, out-of-the-money put option with strike price 0.95\( S_t \), and three-month and two-year variance swaps, rolled over monthly and yearly, respectively. The initial wealth is normalized to 100. The risk-free rate is set to 2%. The investment horizon is \( T = 14.4 \) years, which is the time span of our sample. The risk aversion is set to \( \eta = 5 \), which is an average value in survey data. We also consider the risk aversion \( \eta = 1 \), which corresponds to logarithmic utility. An investor with \( \eta = 1 \) is significantly less risk averse than an investor with \( \eta = 5 \). Optimal portfolios are rebalanced daily. That is, each day optimal portfolio weights are adjusted according to (52) and (53). We also consider proxy portfolios with lower rebalancing frequencies. Section 6.2 discusses several robustness checks that largely confirm our results.

#### 6.1 Optimal and proxy portfolios

Fig. 4 and Fig. 5 display the optimal portfolio weights in on-the-run three-month and two-year variance swaps, stock index, and put option, for \( \eta = 5 \) and \( \eta = 1 \), respectively. The optimal weights in variance swaps induce a short-long strategy, with a short position in the (long-term) two-year variance swap, and a long position in the (short-term) three-month variance swap. As the negative variance risk premium for two-year variance swaps is larger in absolute value than the risk premium for three-month variance swaps (Section 4.1), going short in two-year variance swaps allows to reap a larger risk premium. These short positions are partially hedged via long positions.

\[36\]Meyer and Meyer (2005) survey some of the key studies by economists of how the coefficient of relative risk aversion varies across the population. Most of the survey data suggests values between 0.23 and 8.
in three-month variance swaps, limiting portfolio losses when volatility increases. The three-month variance swap is more sensitive to volatility increases than the two-year variance swap, and it is thus a more effective hedging instrument.

The optimal weights in variance swaps exhibit significant periodic patterns, with increasing portfolio weights in absolute value when their maturities are approaching. Intuitively, close to maturity, most realized variance has been accumulated, inducing little volatility in spot value and thus reducing the risk premium carried by the variance swap. To keep an optimal level of portfolio risk exposure and earn risk premiums, the optimal weights in variance swaps need to increase in absolute value.

The optimal weight in the stock index (52) is positive, which is consistent with the positive equity risk premium to be earned. In contrast to the weights in variance swaps, the stock index weight does not exhibit any periodic pattern. The optimal weights in the stock index and the three-month variance swap are significantly larger for $\eta = 1$ than for $\eta = 5$. The log-investor substantially increases the wealth allocation to the stock index. When the stock index falls and volatility increases, the large positions in three-month variance swaps effectively prevent large drops of the portfolio value.

The optimal weight in the out-of-the-money put option is positive, very small, and around 0.2%, for the investor with $\eta = 5$. In their calibration exercise, Liu and Pan (2003) report similar portfolio weights for out-of-the-money put options. As the jump risk premium is small ($\nu^Q / \nu^P = 1.4$) and the index jump size is large ($\xi = -25\%$), the investor optimally uses the put option to hedge index jumps, rather than to earn the jump risk premium. During low volatility periods, index jumps are less likely to occur, and the investor optimally reduces the put option weight essentially to zero. As noted above, the log-investor takes larger positions in the stock index than the more risk averse investor with $\eta = 5$. The log-investor’s portfolio is therefore significantly exposed to index jumps, and the optimal portfolio weight in the put option increases to around 2%, during relatively volatile periods.

Some oscillations in portfolio weights are observed during the low volatility period 2005–2006. Because volatility reaches historically low values, variance swap rates are also low. This renders the matrix $\mathcal{D}_t$ in (36) close to singular. However, low volatility also implies small changes in variance swap values. This in turn annihilates the impact of oscillating portfolio weights on the wealth
process, resulting in non-oscillating wealth trajectories, as shown below in Fig. 6 and Fig. 7.

Fig. 6 shows the wealth trajectory of the optimal portfolio for an investor with risk aversion \( \eta = 5 \). The wealth trajectory exhibits low volatility and steady growth. Thus, optimally investing in variance swaps, put option and stock index allows for a smooth wealth growth, which is far less sensitive to market falls than investing in the stock index only. The corresponding Sharpe ratio is 1.45%, which is larger than the Sharpe ratio of 1.20% of the S&P 500. The S&P 500 yields a higher terminal wealth than the optimal portfolio. This can occur because the optimal portfolio is not designed to maximize terminal wealth. Compared to the stock index, the optimal portfolio can exhibit lower returns on some occasions but it has always lower volatility. Optimally including variance swaps and put options in the portfolio of a risk averse investor brings more utility than investing in the stock index only because the risk averse investor dislikes large wealth fluctuations.

Fig. 7 shows the wealth trajectory of the optimal portfolios for an investor with risk aversion \( \eta = 1 \). The log-optimal wealth process has a Sharpe ratio of 1.56%, and exhibits significantly larger fluctuations than the S&P 500. This is in sharp contrast with the optimal wealth trajectory of the more risk averse investor with \( \eta = 5 \). It appears that variance swaps can be used either to seek additional risk premiums or achieve stable wealth growth, depending on the risk profile of the investor.

In separate work, we consider a special case of the current optimal portfolio problem. We study optimal investment in variance swaps, stock index and bond, when the price process of the stock index is continuous and the investor has no access to index options, in the bivariate quadratic setting (Section 5.3). Remarkably, power utility investors follow very similar optimal trading strategies, taking short-long positions in variance swaps, and long positions in the stock index. Even though the settings are different in terms of index dynamics and investment universe, there is a striking similarity of the optimal weights in variance swaps. This suggests that short-long positions in variance swaps are a robust feature of the optimal trading strategy. Furthermore, optimal wealth trajectories for \( \eta = 5 \) and \( \eta = 1 \) share very similar patterns as wealth trajectories in Fig. 6 and Fig. 7. This lends further empirical support to our finding that variance swaps can be used either to seek additional risk premiums or achieve stable wealth growth.

We now study the performance of proxy portfolios when the number of contracts in the portfolio is rebalanced at lower frequencies than daily. Specifically, the stock index, put option and three-
month variance swap positions are rebalanced monthly, and the two-year variance swap position is
rebalanced yearly. Between rebalancing dates, positions are kept constant. At rebalancing dates
\( t_{ik}^* \), \( i = 1, 2 \), variance swap investments are rolled over to newly issued three-month and two-year
variance swaps, respectively, according to the portfolio weights \( \pi_{it}^* \) given as exponentially weighted
average of past optimal portfolio weights,

\[
\pi_{it}^* = \frac{\sum_{t_{ik}^*-1 < t \leq t_{ik}^*} \pi_{it}^* \omega_{it}}{\sum_{t_{ik}^*-1 < t \leq t_{ik}^*} \omega_{it}},
\]

where \( \omega_{it} = e^{-(t_{ik}^*-t)} \).37 These portfolio weights attempt to capture the periodic pattern of the
optimal weights over the lifetime of the variance swaps. The reason for assessing the performance
of proxy portfolios is that low rebalancing frequencies reduce transaction costs when implementing
the portfolio strategy in practice. Interestingly, Fig. 6 and Fig. 7 show that the wealth trajectories
of the proxy portfolios are similar to the ones of the optimal portfolios. Although this is only an
in-sample result, it suggests that our optimal portfolio strategies have potential to be implemented
in practice.

The results above differ from those in Egloff et al. (2010) in a number of ways. In their diffusive
affine setting, the optimal weight in the stock index is constant over time and the optimal weights in
variance swaps are state-independent. In our quadratic setting, optimal portfolio weights depend on
state variables and exhibit the rich dynamics discussed above. Thus, the two optimal strategies are
fundamentally different. Furthermore, they assume that at any time the investor can trade newly
issued variance swaps at zero spot value ("sliding" variance swap investment). This is a special
case of our framework in which we take into account investments in on-the-run variance swaps.
This allows us to uncover periodic patterns in the optimal variance swap weights. Moreover, their
empirical implementation of optimal portfolios is static, while we implement dynamic strategies.
They use a risk aversion of \( \eta = 200 \) while we use \( \eta = 5 \) and \( \eta = 1 \). In our setting, the stock index
can jump and the investor can trade put options to hedge jump risk. This is not the case in their
setting. Finally, market price of risk specifications are different in the two studies. This implies
that optimal portfolio weights are significantly different and actually mirror each other.38

37 We set \( \pi_{i0} = n_{i0}^* \) for the initial holding period.
38 As mentioned above, our optimal trading strategy is to go short in long-term variance swaps (to earn the variance
risk premium), long in short-term variance swaps (to hedge volatility increases), and long in the stock index (to earn
the equity risk premium). Egloff et al. (2010) find opposite trading directions in their optimal trading strategy.
6.2 Robustness checks

We performed several robustness checks that largely confirm our empirical analysis of optimal portfolios.

Optimal portfolios above are based on three-month and two-year variance swaps. Optimal portfolios based on variance swaps with other term combinations (such as three-month and one-year; six-month and one-year; six-month and two-year) have similar performance. The same holds true when using different roll-over periods (such as daily, half term, or term of the variance swaps). In the analysis above we use 95% out-of-the-money put options. We also used at-the-money put options and the resulting wealth processes were very similar. For example, when the risk aversion is \( \eta = 5 \), the optimal wealth process always grows steadily over time and is significantly smoother than the trajectory of the stock index. Indeed, since we are in a complete market setup, in theory, the choice of variance swap terms, roll-over periods and index derivatives has no impact on the optimal wealth trajectory.

We experimented with other values of index jump size and intensity, and jump risk premium. When index jumps are smaller or carry more risk premium, the optimal investment in the put option switches from long to short, as theory predicts. For example, when the index jump size is \( \xi = -10\% \), one index jump occurs on average once every ten years, the jump risk premium is \( \nu_Q/\nu_P = 1.2 \), and the risk aversion is \( \eta = 5 \), the optimal investment \( \phi_t \) in the out-of-the-money put option is negative, around \(-1\%\), and somewhat mirrors the \( \phi_t \) in Fig. 4. Optimal investments in variance swaps and stock index largely share the same patterns as in Fig. 4. The optimal wealth trajectory is smooth and very similar to the wealth trajectory in Fig. 6.

We also considered other investment horizons, such as five and ten years. The pattern of optimal portfolio weights is only marginally affected by the choice of the investment horizon.

Besides the risk aversion levels of \( \eta = 5 \) and \( \eta = 1 \), we also experimented with higher values, such as \( \eta = 30 \). The optimal portfolio weights in the risky assets follow the same pattern. The weights are smaller in absolute value, which is consistent with the investor being more risk averse.

The above empirical analysis is based on a sample average equity risk premium of 6%. We redid the analysis for a sample average equity risk premium set to 4% by changing the parameter \( \kappa \) in (51) accordingly. This simply leads to smaller portfolio weights in the stock index, as theory predicts.
Finally, we discuss the impact of transaction costs on wealth trajectories. We analyzed actual bid-ask spreads of variance swap rates from a large broker-dealer. Bid-ask spreads relative to variance swap rates tend to be smaller in pre-crisis than crisis periods, and to decrease with term. The average relative bid-ask spread for two-month and one-year variance swap are 2.3% and 1.2%, respectively. We used these average bid-ask spreads to assess the impact of transaction costs on the proxy portfolios in Section 6.1, which are rebalanced at lower frequency than daily. As newly issued variance swaps have zero value at inception, bid-ask spreads are paid when liquidating existing variance swap positions. We find that such bid-ask spreads have only a minor impact on wealth trajectories. These results are not reported but are available from the authors upon request. Optimal portfolios also include put options, stock index, and risk-free bond. Optimal portfolio weights in put options are very tiny. Bid-ask spreads for liquidly traded stock index and risk-free bonds are very small. Thus, these transaction costs have practically no impact on wealth trajectories.

7 Conclusion

We introduce a novel class of quadratic term structure models for variance swaps. The multivariate state variable driving the stochastic variance follows a quadratic diffusion process. The variance swap curve is quadratic in the state variable and available in closed form, greatly facilitating empirical applications. Various goodness-of-fit tests show that quadratic models fit variance swap rates remarkably well and significantly outperform affine specifications. The quadratic features of the stochastic variance and of the state process diffusion function appear to generate enough volatility of volatility to fit the empirical dynamics of variance swap rates and quadratic variation.

We study dynamic optimal portfolios in variance swaps, put option, stock index, and risk-free bond, when the stock index can jump. Optimal portfolio weights are available in terms of a Taylor series expansion involving conditional moments of the state variables, which in turn are available in closed form. The empirical analysis of optimal portfolios reveals that optimal portfolio weights in variance swaps induce a short-long strategy, with a short position in long-term variance swaps (to earn the negative variance risk premium), and a long position in short-term variance swaps (to

39 The S&P 500 index can be traded via exchange-traded funds (ETF’s) at very low bid-ask spreads. As an example on October 11, 2013 the relative bid-ask spread of SPDR S&P 500 ETF (SPY) was 0.01%, according to Fidelity.com.
hedge volatility increases). This short-long strategy in variance swaps is a robust feature of the optimal trading strategy. Portfolio weights exhibit strong periodic patterns, which depend on the roll-over period and maturity of the variance swaps. The optimal investment in variance swaps can be used either to achieve stable wealth growth or to seek additional risk premium, depending on the risk profile of the investor. Depending on the index jump risk premium, the optimal investment in put options is used either to hedge index jumps or to earn the jump risk premium. Optimal portfolio weights in put options appear to be very small relative to portfolio weights in variance swaps.

Future research can take various directions. Variance swaps on different underlying assets, such as commodities, exchange and interest rates, are actively traded over-the-counter. Quadratic models can easily be applied to these contracts. The recently listed S&P 500 Variance Futures at the CBOE (see Footnote 4) will provide additional data for further studies on variance swaps. Studying derivatives on variance swaps in our quadratic setup can also be an interesting direction for future research.
References


<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-sample</td>
<td></td>
<td></td>
<td></td>
<td>Out-of-sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20.76</td>
<td>6.80</td>
<td>0.87</td>
<td>4.09</td>
<td>27.55</td>
<td>11.05</td>
<td>1.39</td>
<td>4.49</td>
</tr>
<tr>
<td>3</td>
<td>20.90</td>
<td>6.54</td>
<td>0.78</td>
<td>3.87</td>
<td>27.78</td>
<td>10.21</td>
<td>1.21</td>
<td>3.93</td>
</tr>
<tr>
<td>6</td>
<td>21.48</td>
<td>6.32</td>
<td>0.78</td>
<td>3.93</td>
<td>27.94</td>
<td>9.14</td>
<td>1.00</td>
<td>3.51</td>
</tr>
<tr>
<td>12</td>
<td>22.25</td>
<td>6.06</td>
<td>0.62</td>
<td>3.19</td>
<td>27.66</td>
<td>8.12</td>
<td>0.72</td>
<td>3.08</td>
</tr>
<tr>
<td>24</td>
<td>22.86</td>
<td>5.90</td>
<td>0.55</td>
<td>2.75</td>
<td>27.71</td>
<td>7.03</td>
<td>0.27</td>
<td>2.71</td>
</tr>
</tbody>
</table>

Panel B: Realized variances

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>16.42</td>
<td>6.38</td>
<td>0.86</td>
<td>3.20</td>
<td>26.42</td>
<td>14.70</td>
<td>1.80</td>
<td>5.45</td>
</tr>
<tr>
<td>3</td>
<td>16.53</td>
<td>6.08</td>
<td>0.69</td>
<td>2.82</td>
<td>26.92</td>
<td>13.89</td>
<td>1.53</td>
<td>4.08</td>
</tr>
<tr>
<td>6</td>
<td>16.67</td>
<td>5.79</td>
<td>0.51</td>
<td>2.42</td>
<td>27.26</td>
<td>13.14</td>
<td>1.30</td>
<td>3.22</td>
</tr>
<tr>
<td>12</td>
<td>17.02</td>
<td>5.20</td>
<td>0.07</td>
<td>1.77</td>
<td>27.87</td>
<td>11.12</td>
<td>0.60</td>
<td>1.60</td>
</tr>
<tr>
<td>24</td>
<td>18.04</td>
<td>5.28</td>
<td>0.22</td>
<td>2.92</td>
<td>28.21</td>
<td>7.52</td>
<td>-0.35</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Panel C: Realized variance swap payoffs

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-1.67</td>
<td>2.80</td>
<td>-1.49</td>
<td>12.27</td>
<td>0.34</td>
<td>11.38</td>
<td>2.71</td>
<td>10.99</td>
</tr>
<tr>
<td>3</td>
<td>-1.69</td>
<td>2.79</td>
<td>-1.26</td>
<td>10.34</td>
<td>0.42</td>
<td>10.77</td>
<td>1.96</td>
<td>6.59</td>
</tr>
<tr>
<td>6</td>
<td>-1.90</td>
<td>2.77</td>
<td>-1.53</td>
<td>11.55</td>
<td>0.51</td>
<td>10.48</td>
<td>1.40</td>
<td>4.12</td>
</tr>
<tr>
<td>12</td>
<td>-2.15</td>
<td>2.54</td>
<td>-1.73</td>
<td>9.95</td>
<td>0.69</td>
<td>9.11</td>
<td>0.32</td>
<td>1.85</td>
</tr>
<tr>
<td>24</td>
<td>-2.04</td>
<td>3.10</td>
<td>0.36</td>
<td>6.75</td>
<td>0.71</td>
<td>6.87</td>
<td>-0.11</td>
<td>1.42</td>
</tr>
</tbody>
</table>

Table 1: Summary statistics. Mean, standard deviation (Std), skewness (Skew) and kurtosis (Kurt) of variance swap rates in Panel A, realized variances in Panel B, and realized variance swap payoffs on the S&P 500 in Panel C. Variance swap rates and realized variances are in volatility percentage units, i.e., \( \sqrt{\text{VS}(t,T)} \times 100 \) and \( \sqrt{\text{RV}(t,T)} \times 100 \), respectively. Variance swap payoffs are \( (\text{RV}(t,T) - \text{VS}(t,T)) \times 100 \). Term \( \tau \) is in months. In-sample period is from January 4, 1996 to April 2, 2007, a total of 2,832 observations for each series. Out-of-sample period is from April 3, 2007 to June 7, 2010, a total of 794 observations for each series.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Univ. quad.</th>
<th>Univ. poly.</th>
<th>Biv. quad.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>S.E.</td>
<td>Est.</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.305</td>
<td>0.035</td>
<td>0.712</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.370</td>
<td>0.018</td>
<td>-0.309</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>1.672</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.037</td>
<td>0.008</td>
<td>2.720</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_1$</td>
<td>0.312</td>
<td>0.038</td>
<td>0.626</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-0.205</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>-0.010</td>
<td>0.002</td>
<td>0.017</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.023</td>
<td>0.005</td>
<td>0.144</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.015</td>
<td>0.001</td>
<td>0.013</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.027</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>$p_3$</td>
<td>4.10^{-7}</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$p_4$</td>
<td>-3.10^{-6}</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$p_5$</td>
<td>3.10^{-8}</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{VS}^2$</td>
<td>5.14</td>
<td>0.000</td>
<td>0.507</td>
</tr>
<tr>
<td>$c_0$</td>
<td>-0.801</td>
<td>0.090</td>
<td>-0.698</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.998</td>
<td>0.029</td>
<td>1.038</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.174</td>
<td>0.010</td>
<td>0.179</td>
</tr>
<tr>
<td>$c_3$</td>
<td>0.290</td>
<td>0.066</td>
<td>0.302</td>
</tr>
<tr>
<td>$p_\epsilon$</td>
<td>0.614</td>
<td>0.039</td>
<td>0.617</td>
</tr>
</tbody>
</table>

Table 2: Model estimates. Entries are parameter estimates (Est.) for the univariate quadratic, univariate polynomial, and bivariate quadratic models, and corresponding standard errors (S.E.). Identifiable, thus restricted, versions of the following model are estimated: Q-dynamics of the state process $dX_t = (b_1 + \beta_{11}X_{1t} + \beta_{12}X_{2t}) dt + \sqrt{a_1 + \alpha_1X_{1t} + A_1X_{1t}^2} dW_{1t}$, $dX_{2t} = (b_2 + \beta_{22}X_{2t}) dt + \sqrt{a_2 + \alpha_2X_{2t} + A_2X_{2t}^2} dW_{2t}$; Q-spot variance $\nu_t^Q = \phi + \psi X_{1t} + \pi X_{2t}^2 + p_3X_{1t}^3 + p_4X_{1t}^4 + p_5X_{1t}^5$; market price of risk $(\lambda_0 + \lambda_1X_{1t})/\sqrt{a_1 + \alpha_1X_{1t} + A_1X_{1t}^2}$ for the Q-Brownian motion $W_{1t}$. An empty entry means that the parameter is set to zero to achieve model identification. Models are estimated using maximum likelihood with unscented Kalman filter. The measurement equation is six-dimensional. The first five components are variance swap rates with variance of measurement error $\sigma_{VS}^2$. The sixth component is the logarithm of the daily quadratic variation, $\log(QV_t)$, with expectation $c_0 + c_1\log(\nu_t^Q)$, and conditionally normal measurement error $\epsilon_t$ with mean $\rho_\epsilon \epsilon_{t-1}$ and variance $c_2 + c_3QV_{t-1}$. Appendix G provides details on the estimation approach. AIC and BIC are Akaike and Bayesian Information Criteria, respectively. Sample data are daily variance swap rates on the S&P 500, with terms of two, three, six, 12, 24 months, and daily quadratic variation computed from tick-by-tick S&P 500 futures prices, from January 4, 1996 to April 2, 2007, a total of 2,832 observations for each series.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Est.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-5.486</td>
<td>0.560</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.200</td>
<td>0.076</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.172</td>
<td>0.051</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.011</td>
<td>0.003</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.018</td>
<td>0.022</td>
</tr>
<tr>
<td>$\nu_0$</td>
<td>0.177</td>
<td>0.039</td>
</tr>
<tr>
<td>$\nu_1$</td>
<td>1.364</td>
<td>0.250</td>
</tr>
<tr>
<td>$\mu_{2S}$</td>
<td>0.072</td>
<td>0.032</td>
</tr>
<tr>
<td>$\mu^\mathbb{P}_1$</td>
<td>0.075</td>
<td>0.014</td>
</tr>
<tr>
<td>$\mu^\mathbb{Q}_1$</td>
<td>0.147</td>
<td>0.034</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.335</td>
<td>0.054</td>
</tr>
<tr>
<td>$\sigma^2_{VS}$ $10^4$</td>
<td>0.078</td>
<td>0.028</td>
</tr>
<tr>
<td>$c_0$</td>
<td>-0.704</td>
<td>0.160</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.986</td>
<td>0.088</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.161</td>
<td>0.066</td>
</tr>
<tr>
<td>$c_3$</td>
<td>0.453</td>
<td>0.211</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.501</td>
<td>0.198</td>
</tr>
</tbody>
</table>

Log-likelihood | 58,510 |
AIC | -116,986 |
BIC | -116,885 |

Table 3: Two-factor affine jump-diffusion model. Entries are parameter estimates (Est.) for the two-factor affine jump-diffusion model (Section 4.3), and corresponding standard errors (S.E.). 

The dynamics of the stock index $dS_t/S_t = r_t dt + \sqrt{X^1_t} dB_t + \xi_t dN_t - \mathbb{E}^\mathbb{Q}[\xi_t] \nu_t dt$, where $r_t$ is the risk-free rate, the diffusive spot variance $X^1_t$ evolves as $dX^1_t = \beta_1 (X^1_t - X^2_t) dt + \sqrt{\alpha_1} X^1_t dW^1_t + Z_1^t dN_t$, its stochastic long run mean is controlled by $X^2_t$, which evolves as $dX^2_t = (b_2 + \beta_2 X^2_t) dt + \sqrt{\alpha_2} X^2_t dW^2_t$, for some standard Brownian motions $B_t$ and $(W^1_t, W^2_t)$. The second moment of the log-price jump size is $\mu^S_2$. The variance jump size $Z^t$ is exponentially distributed with parameter $\mu^\mathbb{Q}_1$. Jump sizes $\xi_t$ and $Z^t$ are independent from Brownian motions and jump times. Jumps in returns and variance occur contemporaneously and are triggered by $dN_t$. The intensity of $N_t$ is $\nu_t = \nu_0 + \nu_1 X^1_t$, where $\nu_0$ and $\nu_1$ are constants. The $\mathbb{Q}$-spot variance is $v^\mathbb{Q}_t = X^1_t + \mu^\mathbb{Q}_2 (\nu_0 + \nu_1 X^1_t)$, which is an affine function of $X^1_t$. The risk premium $\lambda_1$ is attached to the $\mathbb{Q}$-Brownian motion $W^\mathbb{Q}_t$, and gives the $\mathbb{P}$-Brownian motion $dW^\mathbb{P}_t = dW^\mathbb{Q}_t - \lambda_1 \sqrt{X^1_t} / \sqrt{\alpha_1} dt$. Under the objective measure $\mathbb{P}$, the variance jump size $Z^t$ is exponentially distributed with parameter $\mu^\mathbb{P}_1$. The model is estimated using particle filter. The measurement equation is six-dimensional. The first five components are variance swap rates with variance of measurement error $\sigma^2_{VS}$. The sixth component is the logarithm of the daily quadratic variation, $\log(QV_t)$, with expectation $c_0 + c_1 \log(v^\mathbb{Q}_t)$, and conditionally normal measurement error $\epsilon_t$ with mean $\rho \epsilon_{t-1}$ and variance $c_2 + c_3 QV_{t-1}$. AIC and BIC are Akaike and Bayesian Information Criteria, respectively. Sample data are the same as in Table 2.
Table 4: Variance swap pricing errors. The pricing error is defined as the model-based minus observed variance swap rate, both in volatility percentage units, i.e., \( (\sqrt{G(\tau, X_t)} / \tau - \sqrt{\text{VS}(t, t + \tau)}) \times 100 \). Entries are mean (Bias) and root mean square error (RMSE) of pricing errors for variance swap rates under the univariate quadratic, univariate polynomial, two-factor affine jump-diffusion (AJD) and bivariate quadratic models. DM_u (respectively, DM_p and DM_a) is the Diebold–Mariano test statistic of the univariate quadratic (respectively, polynomial and AJD) model versus the bivariate quadratic model, Section 4.4. Under the null hypothesis that the univariate quadratic (respectively, polynomial and AJD) model and the bivariate quadratic model have pricing errors of equal magnitude, the DM test statistic is a standard normal. A positive value means that the bivariate quadratic model outperforms the univariate model. Term \( \tau \) is in months. Panel A shows pricing error statistics for the in-sample period, used to estimate the models, which is from January 4, 1996 to April 2, 2007, a total of 2,832 observations for each series. Panel B shows pricing error statistics for the out-of-sample period, which is from April 3, 2007 to June 7, 2010, a total of 794 observations for each series.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Bias</th>
<th>RMSE</th>
<th>DM_u</th>
<th>Bias</th>
<th>RMSE</th>
<th>DM_p</th>
<th>Bias</th>
<th>RMSE</th>
<th>DM_a</th>
<th>Bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univ. quad.</td>
<td></td>
<td>Univ. poly.</td>
<td></td>
<td>AJD</td>
<td></td>
<td>Biv. quad.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>1.63</td>
<td>9.85</td>
<td>0.39</td>
<td>1.86</td>
<td>12.62</td>
<td>-0.02</td>
<td>0.60</td>
<td>11.73</td>
<td>-0.06</td>
<td>0.43</td>
</tr>
<tr>
<td>3</td>
<td>0.09</td>
<td>1.10</td>
<td>9.60</td>
<td>0.34</td>
<td>1.35</td>
<td>13.07</td>
<td>0.13</td>
<td>0.53</td>
<td>12.01</td>
<td>0.08</td>
<td>0.37</td>
</tr>
<tr>
<td>6</td>
<td>-0.06</td>
<td>0.58</td>
<td>7.81</td>
<td>0.01</td>
<td>0.85</td>
<td>8.02</td>
<td>0.11</td>
<td>0.55</td>
<td>9.88</td>
<td>0.04</td>
<td>0.43</td>
</tr>
<tr>
<td>12</td>
<td>-0.15</td>
<td>1.16</td>
<td>5.85</td>
<td>-0.29</td>
<td>1.30</td>
<td>8.99</td>
<td>-0.04</td>
<td>0.37</td>
<td>7.88</td>
<td>-0.11</td>
<td>0.28</td>
</tr>
<tr>
<td>24</td>
<td>0.15</td>
<td>1.57</td>
<td>5.33</td>
<td>-0.22</td>
<td>1.60</td>
<td>7.25</td>
<td>0.12</td>
<td>0.59</td>
<td>8.02</td>
<td>0.06</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Panel A: In-sample

|          | Univ. quad. |       | Univ. poly. |       | AJD |       | Biv. quad. |       |
| 2        | -0.14 | 2.35 | 6.89  | 0.58 | 4.01 | 4.59  | 0.26  | 0.95 | 0.07  | 0.06  | 0.94 |
| 3        | -0.27 | 1.43 | 7.17  | 0.17 | 2.93 | 4.57  | 0.04  | 0.79 | 6.69  | -0.09 | 0.60 |
| 6        | -0.17 | 1.39 | 1.95  | -0.59| 2.17 | 1.11  | -0.17 | 1.53 | 2.89  | -0.20 | 1.32 |
| 12       | 0.44  | 1.96 | 0.29  | -1.67| 2.65 | 1.77  | 0.08  | 1.86 | -0.31 | 0.10  | 1.89 |
| 24       | 0.62  | 4.01 | 3.89  | -1.08| 4.24 | 3.37  | 0.16  | 2.67 | 4.06  | 0.21  | 2.18 |

Panel B: Out-of-sample
Table 5: Variance swap predictive regressions. For each model and term, entries report time series regressions of future actual variance swap rates on a constant and a \(d\)-day ahead model-based prediction, i.e., \(\text{VS}(t, t + \tau) = \gamma_0 + \gamma_1 \mathbb{E}^P[G(\tau, X_t) / \tau | \mathcal{F}_{t-d}] + \text{error}_t\), where \(d\) is either one-day (Panel A) or ten-day (Panel B), and \(\mathbb{E}^P[G(\tau, X_t) / \tau | \mathcal{F}_{t-d}]\) is the time \(t-d\) model-based, conditional prediction of the \(\tau\)-variance swap rate observed at time \(t\). Variance swap rates are in variance percentage units, i.e., \(\text{VS}(t, t + \tau) \times 100\). For each term \(\tau\), the first row reports estimates of \(\gamma_0\) and \(\gamma_1\), the second row reports the p-value of the null hypotheses \(H_0 : \gamma_0 = 0\), and \(H_0 : \gamma_1 = 1\), respectively. If model-based variance swap rate predictions are unbiased, then \(\gamma_0 = 0\). If model-based variance swap rate predictions are efficient, then \(\gamma_1 = 1\). Robust standard errors are computed using Newey and West (1987) covariance matrix estimator with the number of lags optimally chosen according to Andrews (1991). The martingale model is a benchmark model in which the future actual \(\text{VS}(t, t + \tau)\) is predicted using the past actual \(\text{VS}(t-d, t-d+\tau)\). Term \(\tau\) is in months. The sample period is from January 4, 1996 to June 7, 2010, a total of 3,626 observations for each series.
Figure 1: Term structure of variance swaps rates. Variance swap rates on the S&P 500 in volatility percentage units, $\sqrt{\text{VS}(t,T)} \times 100$. Terms are two, three, six, 12, 24 months. Sample period is from January 4, 1996 to June 7, 2010.
Figure 2: Time series evolution of state process. In the bivariate quadratic model in Section 3.2, $X_{1t}$ is in Class 1 and $X_{2t}$ is in Class 3; Theorem 3.3. The model is fitted to daily variance swap rates on the S&P 500, from January 4, 1996 to April 2, 2007, and terms of two, three, six, 12, 24 months, and daily quadratic variation computed from tick-by-tick S&P 500 futures data. Table 2 reports model estimates. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 3: Actual and model-based variance swap rates. Model-based variance swap rates are from the bivariate quadratic model in Section 3.2, with $X_1$ in Class 1 and $X_2$ in Class 3; Theorem 3.3. The model is fitted to daily variance swap rates on the S&P 500 with terms of two, three, six, 12, 24 months, and daily quadratic variation from tick-by-tick S&P 500 futures prices, from January 4, 1996 to April 2, 2007. Table 2 reports model estimates. Variance swap rates are in volatility percentage units, i.e., $\sqrt{VS(t, T)} \times 100$. Upper graph: variance swap rates with two-month term (shortest term in our sample). Lower graph: variance swap rates with two-year term (longest term in our sample). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 4: Optimal portfolio. Wealth is optimally invested in three-month and two-year variance swaps, index put option, stock index, and risk-free bond. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Risk aversion is $\eta = 5$. $n_{1t}$ is the optimal fraction of wealth invested in the three-month variance swap, and $n_{2t}$ in the two-year variance swap (upper graph); $w_t$ in the stock index (middle graph); $\phi_t$ in the index put option (lower graph). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 5: Optimal portfolio for log-investor. Wealth is optimally invested in three-month and two-year variance swaps, index put option, stock index, and risk-free bond. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Risk aversion is \( \eta = 1 \). \( n_{1t} \) is the optimal fraction of wealth invested in the three-month variance swap, and \( n_{2t} \) in the two-year variance swap (upper graph); \( w_t \) in the stock index (middle graph); \( \phi_t \) in the index put option (lower graph). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 6: Wealth process. Wealth is optimally invested in three-month and two-year variance swaps, index put option, stock index, and risk-free bond. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio is rebalanced less frequently: three-month variance swap, index put option, and stock index positions are rebalanced monthly, two-year variance swap position is rebalanced yearly. Risk aversion is $\eta = 5$. S&P 500 is normalized to 100. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

Figure 7: Wealth process for log-investor. Wealth is optimally invested in three-month and two-year variance swaps, index put option, stock index, and risk-free bond. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio is rebalanced less frequently: three-month variance swap, index put option, and stock index positions are rebalanced monthly, two-year variance swap position is rebalanced yearly. Risk aversion is $\eta = 1$. S&P 500 is normalized to 100. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
ONLINE APPENDIX TO
Quadratic Variance Swap Models

This appendix provides technical derivation and proofs.

A Diffusions and partial differential equations

This section provides some technical results on diffusion processes which form the background of several proofs in this 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 $Q$. 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

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

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

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

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

$$
\frac{\partial f(\tau, x)}{\partial \tau} = \sum_{i=1}^m \mu_i(x) \frac{\partial f(\tau, x)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^m \left( \Sigma(x) \Sigma(x)^T \right)_{ij} \frac{\partial^2 f(\tau, x)}{\partial x_i \partial x_j} - k(x)f(\tau, x),
$$

$$
f(0, x) = g(x),
$$

(A.3)
for all \( \tau \geq 0 \) and \( x \in \mathcal{X} \), then

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

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

with drift term given by

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

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

\[
\mathbb{E}^Q \left[ \int_0^T e^{-\int_0^s k(X_u) du} \| \nabla_x f(T - s, X_s) \Sigma(X_s) \|^2 ds \right] \leq K_1 \mathbb{E}^Q \left[ \int_0^T \| \nabla_x f(T - s, X_s) \Sigma(X_s) \|^2 ds \right]
\]

\[
\leq K_2 \left( 1 + \mathbb{E}^Q \left[ \sup_{s \leq T} \| X_s \|^{2p} \right] \right),
\] (A.7)

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

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 \( \mathbb{E}^Q \left[ \sup_{s \leq T} \| X_s \|^{2p} \right] < \infty \), 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 \( X_{t \wedge T_n} = X_{t \wedge T_n} \) satisfies

\[
X_{t \wedge T_n} = x + \int_0^t \mu(X_{s \wedge T_n}) 1_{\{s \leq T_n\}} ds + \int_0^t \Sigma(X_{s \wedge T_n}) 1_{\{s \leq T_n\}} dW_s =: x + D_t + M_t.
\] (A.8)

We fix a finite \( T \). In what follows, \( K_1, K_2, \ldots \) denote some universal finite constants, which only
depend on $T$. First, observe that the linear growth condition (A.1) implies the pathwise inequality
\[
\sup_{s \leq t} \| D_s \|^{2p} \leq K_1 \int_0^t \| \mu(X_u^{T_n}) \|^{2p} du \leq K_2 \int_0^t \left( 1 + \sup_{s \leq u} \| X_s^{T_n} \|^{2p} \right) du.
\] (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
\[
\mathbb{E}^Q \left[ \sup_{s \leq t} \| M_s \|^{2p} \right] \leq K_3 \int_0^t \mathbb{E}^Q \left[ \| \Sigma(X_u^{T_n}) \|^{2p} \right] du \leq K_4 \int_0^t \left( 1 + \mathbb{E}^Q \left[ \sup_{s \leq u} \| X_s^{T_n} \|^{2p} \right] \right) du.
\] (A.10)

Combining these inequalities, we obtain
\[
\mathbb{E}^Q \left[ \sup_{s \leq t} \| X_s^{T_n} \|^{2p} \right] \leq K_5 \left( x^{2p} + t + \int_0^t \mathbb{E}^Q \left[ \sup_{s \leq u} \| X_s^{T_n} \|^{2p} \right] du \right).
\] (A.11)

By dominated convergence, the nonnegative function $[0, T] \ni t \mapsto \mathbb{E}^Q \left[ \sup_{s \leq t} \| X_s^{T_n} \|^{2p} \right]$ is continuous. Applying Gronwall’s inequality, Karatzas and Shreve (1991, Problem V.2.7), to it yields
\[
\mathbb{E}^Q \left[ \sup_{s \leq T} \| X_s^{T_n} \|^{2p} \right] \leq K_5 \left( x^{2p} + T + \int_0^T (x^{2p} + u) K_5 e^{K_5(T-u)} du \right).
\] (A.12)

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

\[\Box\]

**B \hspace{1em} 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 $\omega : [0, \infty) \to \mathcal{X}$. 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 $v_t^Q = g(X_t)$ 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 $\text{VS}(t, T) = \frac{1}{T-t} \int_t^T \mathbb{E}^\mathbb{Q}[V^\mathbb{Q}_s | X_t] \, ds$.

Hence $\text{VS}(t, T)$ 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).

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 $\Phi : [0, +\infty) \to \mathbb{R}$, $\Psi : [0, +\infty) \to \mathbb{R}^m$, and $\Pi : [0, +\infty) \to \mathbb{S}^m$. 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 $\Pi_{ij}(\tau)$, $1 \leq i, j \leq m$, 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

\[
\frac{d\Phi(\tau)}{d\tau} = \phi + b^\top \Psi(\tau), \quad \Phi(0) = 0,
\]

\[
\frac{d\Psi(\tau)}{d\tau} = \psi + \beta^\top \Psi(\tau), \quad \Psi(0) = 0.
\]
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 \( \tau \) leads to

\[
\frac{d\Phi(\tau)}{d\tau} - \phi + \left( \frac{d\Psi(\tau)}{d\tau} - \psi \right)^\top x + x^\top \left( \frac{d\Pi(\tau)}{d\tau} - \pi \right) x \\
= \sum_{i=1}^{m} \Psi_i(\tau)\mu_i(x) + \sum_{i,j=1}^{m} \Pi_{ij}(\tau) (\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)) \\
= \sum_{i=1}^{m} \Psi_i(\tau)\mu_i(x) + \sum_{i=1}^{m} \Pi_{ii}(\tau) (2\mu_i(x)x_i + c_{ii}(x)) + 2 \sum_{i<j}^{m} \Pi_{ij}(\tau) (\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)) \\
\]  

for all \( \tau \) 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.

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

\[
\frac{d\Phi(\tau)}{d\tau} = \phi + b\Psi(\tau) + a\Pi(\tau), \quad \Phi(0) = 0, \\
\frac{d\Psi(\tau)}{d\tau} = \psi + \beta\Psi(\tau) + (2b + \alpha)\Pi(\tau), \quad \Psi(0) = 0, \\
\frac{d\Pi(\tau)}{d\tau} = \pi + (2\beta + A)\Pi(\tau), \quad \Pi(0) = 0, \\
\]  

for real parameters \( \phi, \psi, \pi \).

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

\[
\frac{d\Phi(\tau)}{d\tau} = \phi + \beta^\top \Psi(\tau) + a_1 \Pi_{11}(\tau) + a_2 \Pi_{22}(\tau), \quad \Phi(0) = 0,
\]

\[
\frac{d\Psi(\tau)}{d\tau} = \begin{pmatrix} \psi \\ 0 \end{pmatrix} + \beta^\top \Psi(\tau) + 2\Pi(\tau) b + \begin{pmatrix} \alpha_1 \Pi_{11}(\tau) \\ \alpha_2 \Pi_{22}(\tau) \end{pmatrix}, \quad \Psi(0) = 0,
\]

\[
\frac{d\Pi(\tau)}{d\tau} = \begin{pmatrix} \pi \\ 0 \end{pmatrix} + \beta^\top \Pi(\tau) + \Pi(\tau) \beta + \begin{pmatrix} A_1 \Pi_{11}(\tau) & 0 \\ 0 & A_2 \Pi_{22}(\tau) \end{pmatrix}, \quad \Pi(0) = 0,
\]

for real parameters \( \phi, \psi, \pi \). 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 \( \beta_{21} = 0 \)):

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

\[\text{(C.3)}\]

### D Proof of Theorem 3.3

It follows by inspection that the quadratic property is invariant with respect to affine transformations \( X \rightarrow c + \gamma X, \ x \mapsto c + \gamma x \) 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

\[
\begin{align*}
\hat{b}(\hat{x}) &= b\gamma - \beta c + \beta \hat{x} = \hat{b} + \hat{\beta} \hat{x}, \\
\hat{a}(\hat{x}) &= a\gamma^2 - \alpha \gamma c + A c^2 + (\alpha \gamma - 2 A c) \hat{x} + A \hat{x}^2 = \hat{a} + \hat{\alpha} \hat{x} + \hat{A} \hat{x}^2.
\end{align*}
\]

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 $\gamma$ 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 $\hat{X} = \mathbb{R}$, 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 $\gamma$ 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 $\hat{X} = (0, +\infty)$. If $A = \alpha = a = 0$, we can chose $\gamma$ 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 $\gamma = \pm \frac{1}{\sqrt{D}}$ yields $\hat{a}(\hat{x}) = \hat{x} + A\hat{x}^2$. The sign of $\gamma$ 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 $\hat{X}$ not to contain 0. Indeed, elementary calculations show that the scale function of $\hat{X}_t$ is

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

It satisfies $p(\hat{x}) = p(r)P[\tau_r < \tau_R] + p(R)P[\tau_r > \tau_R]$ for any $0 \leq r < \hat{X}_0 = \hat{x} < R$, and hitting times defined by $\tau_c = \inf\{t \geq 0 | \hat{X}_t = c\}$, see Karatzas and Shreve (1991, Section V.5.C). Since $\tau_R \uparrow \infty$ for $R \uparrow \infty$, it follows that $P[\tau_0 = \infty] = 1$ 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.
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 $X_{10} > 0$.

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 $f_n(\tau, X_t) = \mathbb{E}^Q[X^n_{t+\tau} \mid \mathcal{F}_t]$ formally solves the partial differential equation

$$
\begin{align*}
\frac{\partial}{\partial \tau} f_n(\tau, x) &= \mathcal{A}f_n(\tau, x), \\
 f_n(0, x) &= x^n,
\end{align*}
$$

(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

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

(F.2)
and matching coefficients in $x$, we obtain the $N + 1$ linear systems of $N + 1$ ordinary differential equations

\[
\frac{d}{d\tau} \begin{pmatrix}
M_{0n}(\tau) \\
M_{1n}(\tau) \\
M_{2n}(\tau) \\
M_{3n}(\tau) \\
\vdots \\
M_{Nn}(\tau)
\end{pmatrix} = \begin{pmatrix}
0 & b & 2\frac{a}{2} & 0 & \cdots & 0 \\
0 & \beta & 2\left(b + \frac{a}{2}\right) & 3\cdot 2\frac{a}{2} & 0 & \vdots \\
0 & 0 & 2\left(\beta + \frac{a}{2}\right) & 3\left(b + 2\frac{a}{2}\right) & \ddots & 0 \\
0 & 0 & 0 & 3\left(\beta + 2\frac{a}{2}\right) & \ddots & N(N - 1)\frac{a}{2} \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & \ldots & 0 & N\left(\beta + (N - 1)\frac{a}{2}\right)
\end{pmatrix} \begin{pmatrix}
M_{0n}(\tau) \\
M_{1n}(\tau) \\
M_{2n}(\tau) \\
M_{3n}(\tau) \\
\vdots \\
M_{Nn}(\tau)
\end{pmatrix},
\]  
(F.3)

along with the initial condition

\[
M_{kn}(0) = \begin{cases}
1, & \text{if } k = n \\
0, & \text{otherwise}.
\end{cases}
\]  
(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

\[
\frac{d}{d\tau} M(\tau) = BM(\tau), \quad M(0) = Id,
\]  
(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 $F_t$-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 $VS(t, T) = \frac{1}{T-t} \int_{t}^{T-t} f(\tau, X_t) \, d\tau$.

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 $\Phi_0$, $\Phi_X$, 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 $\sigma_{VS}^2$. The sixth component of the observation vector is the logarithm of the daily quadratic variation, $\log(QV_t)$. The corresponding component in $H(X_t)$ is $\log(v_P^t)$, where we specify the $P$-spot variance by $\log(v_P^t) = c_0 + c_1 \log(v_Q^t)$, and $v_Q^t = g(X_t)$ is the $Q$-spot variance in (11). This specification does not follow the typical modeling of the price jump risk premium in affine models, where $v_P^t$ and $v_Q^t$ would both be affine functions of $X_t$. The associated measurement error $\epsilon_t$ is conditionally normally distributed with mean $\rho_t \epsilon_{t-1}$ and variance $c_2 + c_3 QV_{t-1}$. The rationale behind the sixth component of the measurement equation is the following. Andersen et al. (2001), among others, provide empirical evidence that $\log(QV_t)$ is approximately normally distributed. The conditional mean specification of $\epsilon_t$ 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 $\epsilon_t$
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 et al. (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

$$\hat{X}_{t|t-1} = \Phi_0 + \Phi_X \hat{X}_{t-1},$$
$$P_{t|t-1} = \Phi_X P_{t-1} \Phi_X^T + Q_t,$$  \hspace{1cm} (G.3)

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

$$\hat{Z}_{t|t-1} = H_0 + H_X \hat{X}_{t|t-1},$$
$$F_{t|t-1} = H_X P_{t|t-1} H_X^T + \Omega_t.$$  \hspace{1cm} (G.4)

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

$$\hat{X}_t = \hat{X}_{t|t-1} + W_t (Z_t - \hat{Z}_{t|t-1}),$$
$$P_t = P_{t|t-1} - W_t F_{t|t-1} W_t^T.$$  \hspace{1cm} (G.6)

\hspace{1cm} (G.7)
where \( W_t = P_{t|t-1} H^T X F_{t|t-1}^{-1} \).

In our setting, the function \( H(X_t) \) is nonlinear, and the Kalman filter has to be modified. Nonlinear 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

\[
\begin{align*}
\hat{X}^0_{t|t-1} &= \hat{X}_{t|t-1}, \\
\hat{X}^i_{t|t-1} &= \hat{X}_{t|t-1} + (\sqrt{(L + \kappa) P_{t|t-1}})_i, & w^0 &= \frac{\kappa}{L + \kappa}, \\
\hat{X}^i_{t|t-1} &= \hat{X}_{t|t-1} - (\sqrt{(L + \kappa) P_{t|t-1}})_i, & w^i &= \frac{1}{2(L + \kappa)}, \quad i = 1, \ldots, L, \\
\hat{X}^{L+1}_{t|t-1} &= \hat{X}_{t|t-1} + (\sqrt{(L + \kappa) P_{t|t-1}})_i, & w^i &= \frac{1}{2(L + \kappa)}, \quad i = L + 1, \ldots, 2L,
\end{align*}
\]

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

\[
\begin{align*}
\hat{Z}_{t|t-1} &= \sum_{i=0}^{2L} w^i H(\hat{X}^i_{t|t-1}), \\
F_{t|t-1} &= \sum_{i=0}^{2L} w^i (H(\hat{X}^i_{t|t-1}) - \hat{Z}_{t|t-1})(H(\hat{X}^i_{t|t-1}) - \hat{Z}_{t|t-1})^T + \Omega_t.
\end{align*}
\]

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

\[
W_t = \sum_{i=0}^{2L} w^i (\hat{X}^i_{t|t-1} - \hat{X}_{t|t-1})(H(\hat{X}^i_{t|t-1}) - \hat{Z}_{t|t-1})^T F_{t|t-1}^{-1}.
\]
Finally, the log-likelihood function is given by
\[
\sum_{t=1}^{N} -\frac{1}{2} \left[ 6 \log(2\pi) + \log |F_{t|t-1}| + (Z_t - \hat{Z}_{t|t-1})^\top F_{t|t-1}^{-1} (Z_t - \hat{Z}_{t|t-1}) \right],
\] (G.11)
where \( N = 2,832 \) is the sample size of daily observations.

\[H\] 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 \( \mu(x), \Sigma(x), \Lambda(x), \nu^p(x), \nu^q(x), \) and
\[
k(x) = -\frac{1}{2\eta} \left( \frac{1}{\eta} - 1 \right) ||\Lambda(x)||^2 - \nu^q(x) \left( \left( \frac{\nu^p(x)}{\nu^q(x)} \right)^\frac{1}{\eta} - 1 \right) - \frac{1}{\eta} (\nu^q(x) - \nu^p(x))
\] (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
\[
d\hat{X}_t = \left( \mu(\hat{X}_t) + \frac{1}{\eta} \Sigma(\hat{X}_t) \Lambda(\hat{X}_t) \right) dt + \Sigma(\hat{X}_t) dW_t
\] (H.2)
is well posed in \( \mathcal{X} \) (see Assumption B.1), and the function
\[
h(\tau, x) = \log \mathbb{E}^q \left[ e^{-\int_0^\tau k(\hat{X}_s) ds} \mid \hat{X}_0 = x \right]
\] (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
\[
\frac{\partial H(\tau, x)}{\partial \tau} = \frac{1}{2} \sum_{i,j=1}^{m} \left( \Sigma(x) \Sigma(x)^\top \right)_{ij} \frac{\partial^2 H(\tau, x)}{\partial x_i \partial x_j} + \left( \mu(x) + \frac{1}{\eta} \Sigma(x) \Lambda(x) \right)^\top \nabla_x H(\tau, x) - k(x) H(\tau, x),
\] (H.4)
along with the initial condition \( H(0, x) = 1 \). Hence \( h(\tau, x) \) satisfies the nonlinear PDE

\[
\frac{\partial h(\tau, x)}{\partial \tau} \= \frac{1}{2} \sum_{i,j=1}^{m} \left( \Sigma(x) \Sigma(x)^\top \right)_{ij} \frac{\partial^2 h(\tau, x)}{\partial x_i \partial x_j} + \left( \mu(x) + \frac{1}{\eta} \Sigma(x) \Lambda(x) \right)^\top \nabla_x h(\tau, x)
\]

\[
+ \frac{1}{2} \nabla_x h(\tau, x)^\top \Sigma(x) \Sigma(x)^\top \nabla_x h(\tau, x) - k(x),
\]

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

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

\[
\theta^{W^*}(t, v, x) = -\frac{\partial J}{\partial v} \Lambda(x) - \Sigma(x)^\top \nabla_x \left( \frac{\partial J}{\partial v} \right),
\]

\[
\theta^{N^*}(t, v, x) = \frac{1}{\xi} \left[ \left( \frac{\nu^p(x) J(1-\eta)}{\nu^q(x) \frac{\partial J}{\partial v}} \right)^{\frac{1}{\eta}} - 1 \right].
\]

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

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, \( \nu^q(x) = \nu^p(x) \), 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

\[
k(x) = c + \epsilon P(x),
\]

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

\[
\nabla_x h(\tau, x, \epsilon) = \nabla_x h(\tau, x, 0) + \nabla_x \partial_\epsilon h(\tau, x, 0) \epsilon + o(\epsilon)
\]

\[
\= -\nabla_x \int_0^\tau \mathbb{E}^Q \left[ \epsilon P(\hat{X}_s) \mid \hat{X}_0 = x \right] ds + o(\epsilon).
\]
If the diffusion $\tilde{X}_t$ 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 $\epsilon$.

J Proof of Corollary 5.3

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

$$\theta^W_t = \Sigma(X_t)^\top \left( D_t n_t + \frac{\nabla_x O_t}{O_{t^-}} \phi_t \right) + \sigma(X_t) R(X_t) \left( w_t + \frac{\partial_x O_t S_t}{O_{t^-}} \phi_t \right),$$

$$\theta^N_t = w_t + \frac{\Delta O_t}{\xi O_{t^-}} \phi_t. \tag{J.1}$$

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

$$\frac{1}{\eta} v^\top \Lambda(X_t) = v^\top R(X_t) \sigma(X_t) \left( w_t + \frac{\partial_x O_t S_t}{O_{t^-}} \phi_t \right),$$

$$\frac{1}{\xi} \left( \left( \frac{\nu^\Sigma(X_t)}{\nu^2(X_t)} \right)^{1/\eta} - 1 \right) = w_t + \frac{\Delta O_t}{\xi O_{t^-}} \phi_t, \tag{J.3}$$

for $w_t$ and $\phi_t$ hold $dt \otimes dQ$-a.s. Hence the solution $w^*_t = w_t$ and $\phi^*_t = \phi_t$ of (35) for the optimal $\theta^W*$ and $\theta^N*$ 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.

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

$$Y_t = \int_{t_0}^t \left( \log(1 + \xi) - \xi \nu^Q(X_s) - \frac{1}{2} \sigma(X_s)^2 \right) ds = K_Y \int_{t_0}^t g(X_s) ds, \tag{K.1}$$
with \( K_Y = -\frac{(\xi - \log(1 + \xi))\nu Q + 1/2}{1 + (\log(1 + \xi))\nu Q^2} < 0 \), as in (46), and \( X_{3t} \) has the \( \mathbb{Q} \)-dynamics, for \( t \geq t_0 \),

\[
dX_{3t} = r \, dt + \sigma(X_t)R(X_t)\, dW_t + \log(1 + \xi)(dN_t - \nu Q(X_t)\, dt), \quad X_{3t_0} = \log S_{t_0}. \tag{K.2}
\]

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

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

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

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

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

\[
(q_0 + q_1x_1 + q_2x_1^2)^2 \approx 0.7^2K_{\sigma^2}g(x_1)(1 + A_1x_1^2), \tag{K.5}
\]

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

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

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

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

The function \( R_1(x_1) \) is then specified by (K.4) and (K.7). Fig. 8 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 8: 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

$$
\begin{pmatrix}
1, \mathbb{E}[X_{1T} | \mathcal{F}_{t_0}], \ldots, \mathbb{E}[X_{2T}X_{3T}^{N-1} | \mathcal{F}_{t_0}], \mathbb{E}[X_{3T}^N | \mathcal{F}_{t_0}] \n\end{pmatrix} = \begin{pmatrix} 1, X_{1t_0}, \ldots, X_{2t_0}, X_{3t_0}^{N-1}, X_{3t_0}^N \end{pmatrix} e^{\tilde{B}(T-t_0)}, \quad (K.8)
$$

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 $X_t$,

$$A f(x) = \left( \begin{array}{c} \beta_{11} x_1 + \beta_{12} x_2 \\ b_2 + \beta_{22} x_2 \\ b_3 \end{array} \right) \top \nabla_x f(x) + \frac{1}{2} \sum_{i,j=1}^3 A_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

(K.9)

$$+ \left( f(x + \log(1 + \xi)e_3) - f(x) - \nabla_x f(x) \top e_3 \log(1 + \xi) \right) \nu^Q(x),$$

where $e_3 = (0, 0, 1)^\top$. It is easily verified that $A$ is polynomial preserving: for $f(x)$ being a polynomial in $x$ of degree $0 \leq n \leq N$, $A f(x)$ 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 $A$ to the mixed powers $1, x_1, \ldots, x_2 x_3^{N-1}, x_3^N$ and collecting terms. The rest of the proof follows literally as in Appendix F.

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

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

(K.10)

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

$$L_T^n = (Y_T + X_{3T})^n = \sum_{k=0}^n \binom{n}{k} Y_T^k X_{3T}^{n-k}$$

(K.11)

$$= \sum_{k=0}^n \binom{n}{k} K_k^T \int_{t_0}^T \int_{t_0}^T \cdots \int_{t_0}^T g(X_{t_1}) \cdots g(X_{t_k}) \, dt_k \cdots dt_1 \, X_{3T}^{n-k}$$

$$= \sum_{k=0}^n \binom{n}{k} K_k^T (k!) \int_{t_0}^T \int_{t_1}^T \cdots \int_{t_{k-1}}^T g(X_{t_1}) \cdots g(X_{t_k}) \, dt_k \cdots dt_1 \, X_{3T}^{n-k}.$$
We compute the conditional expectation in the integral using nested conditional expectations
\[
\mathbb{E}_Q \left[ g(X_{t_1}) \cdots g(X_{t_k}) X_{n-k} | \mathcal{F}_{t_0} \right] = \mathbb{E}_Q \left[ g(X_{t_1}) \cdots g(X_{t_k}) \mathbb{E}_Q \left[ X_{n-k}^0 | \mathcal{F}_{t_k} \right] | \mathcal{F}_{t_0} \right] = \mathbb{E}_Q \left[ g(X_{t_1}) \cdots g(X_{t_{k-1}}) \mathbb{E}_Q \left[ g(X_{t_k}) P_0(t_k, X_{t_k}) | \mathcal{F}_{t_{k-1}} \right] | \mathcal{F}_{t_0} \right] = \mathbb{E}_Q \left[ g(X_{t_1}) P_{k-1}(t_k, t_{k-1}, \ldots, t_1, X_{t_1}) | \mathcal{F}_{t_0} \right] \tag{K.13}
\]

The notation \( P_j(t_k, \ldots, t_{k-j}, X_{t_{k-j}}) \), 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.\(^{40}\)

Option prices are computed using standard Fourier inversion, following Carr and Madan (1998a) 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 \( \delta = 0.2 \).\(^{41}\)

**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 \( 2 \times 2 \) 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
\[
1_{\{\ell(t, X_t) = 0\}} \nabla_x \ell(t, X_t) \Sigma(X_t) \Sigma(X_t)^\top \nabla_x \ell(t, X_t) = 0 \quad dt \otimes dQ\text{-a.s.} \tag{L.1}
\]

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 \( dt \otimes dQ\text{-a.s.} \), we infer that \( \ell(t, X_t) \neq 0 \) if \( \nabla_x \ell(t, X_t) \neq 0 \), \( dt \otimes dQ\text{-a.s.} \). 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,

\(^{40}\)The integrals were computed using numerical integration. Resulting moments were benchmarked with those obtained with Monte Carlo simulations.

\(^{41}\)We checked the robustness of the option prices and sensitivities with respect to the numerical value of \( \delta \).
$dt \otimes dQ$-a.s. In Appendix K we saw that $R_1(x_1)^2 < 1$ for all $x_1 \in \mathbb{R}$. Hence the $3 \times 3$ matrix $(\Sigma(X_t)\sigma(X_t)\mathbf{R}(X_t))$ is invertible $dt \otimes dQ$-a.s. For an index put option $\frac{AO_t}{\xi_{\infty}} > \partial_t O_t$ 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 $h(\tau, x)$.

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

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

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

$$\nabla_x h(\tau, x) \approx -K_k \nabla_x \int_0^\tau \mathbb{E}_Q \left[ \psi \hat{X}_{1s} + \pi \hat{X}_{1s}^2 \big| \hat{X}_0 = x \right] ds,$$

which is available in closed form.

--

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