Quadratic Variance Swap Models*

Damir Filipović†
Swiss Finance Institute
and EPFL

Elise Gourier‡
Swiss Finance Institute
and University of Zurich

Loriano Mancini§
Swiss Finance Institute
and EPFL

This version: 14 February 2014

*For helpful comments we thank Bruno Bouchard, Peter Carr, Pierre Collin-Dufresne, Jerome Detemple, Rüdiger Fahlenbrach, Jim Gatheral, Julien Hugonnier, Mika Kastenholz, Markus Leippold, Dilip Madan, Eckhard Platen, and Nizar Touzi, as well as participants at the 2013 European Finance Association meetings in Cambridge. Financial support from the Swiss National Science Foundation NCCR-FinRisk is gratefully acknowledged.

†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, Department of Banking and Finance, University of Zurich, Plattenstrasse 22, CH-8032 Zurich, Switzerland. E-mail: elise.gourier@bf.uzh.ch. Telephone: +41 44 634 4045.

§Loriano Mancini, Swiss Finance Institute at EPFL, Quartier UNIL-Dorigny, CH-1015 Lausanne, Switzerland. E-mail: loriano.mancini@epfl.ch. Telephone: +41 21 693 0107.
Quadratic Variance Swap Models

Abstract

We introduce a novel class of term structure models for variance swaps. The multivariate state variable follows a diffusion process characterized by a quadratic diffusion function. The variance swap curve is quadratic in the state variable, and available in closed form in terms of a linear ordinary differential equation, 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 nested specifications, including popular affine models. An empirical study of a dynamic optimal portfolio in variance swaps and the S&P 500 reveals the versatility of quadratic models, and the economic value of variance swaps.

JEL Classification: C51, G13

Keywords: variance swap, quadratic term structure model, 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. Figure 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 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 term structure of variance swap rates on the S&P 500 shown in Figure 1, with five terms ranging from 2 months to 2 years, and daily quotes spanning from January 4, 1996 to June 7, 2010. 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. 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 strong persistence of variance swap rates (first order autocorrelations are above 0.98), 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 only the bivariate quadratic model has an intercept and a slope not statistically different from zero and one, respectively, and thus produces accurate forecasts. Moreover, only 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 ex-ante risk premiums embedded in variance swaps.

Equity and variance risk premiums induced by the bivariate quadratic model are economically sizable and exhibit significant time variation, which is in line with recent studies, e.g., Bollerslev and Todorov (2011), Aït-Sahalia, Karaman, and Mancini (2012), and Martin (2013). The equity risk premium is positive and countercyclical. The variance risk premium is mostly negative and procyclical. Overall, our empirical analysis suggests that the bivariate quadratic model offers a good trade-off between tractability and fitting accuracy of the term structure dynamics.

At least two features contribute to the popularity of variance swaps. First, hedging a variance swap is relatively easier than hedging other volatility derivatives. Indeed, the payoff of a variance swap can be replicated by dynamically 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 may not exist at all for a given maturity date. Second, the variance swap payoff is only sensitive to the realized variance over a desired and predetermined time horizon. Suppose an investor, who holds a broadly diversified portfolio, 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

\footnote{This 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 (1998), Demeterfi, Derman, Kamal, and Zou (1999), Britten-Jones and Neuberger (2000), Jiang and Tian (2005), Jiang and Oomen (2008), Carr and Wu (2009), and Carr and Lee (2010).}
on the VIX index\textsuperscript{2} would not provide an equally direct hedge.\textsuperscript{3}

To assess the economic relevance of variance swaps, we study a dynamic optimal portfolio problem in variance swaps, a stock index and a risk free bond.\textsuperscript{4} 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 process, which are explicit in terms of a linear ordinary differential equation. We implement the optimal portfolio using actual 3-month and 2-year variance swap rates and S&P 500 returns. We find that the optimal portfolio weights in the variance swaps follow a short-long strategy, with a short position in the 2-year variance swap (to earn the negative variance risk premium), and a long position in the 3-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, Leippold, and Wu (2010), and Aït-Sahalia, Karaman, and Mancini (2012). We also find that optimal weights in variance swaps show strong periodic patterns, which depend on the maturity and roll-over date of the contracts, and which are mainly borne by the intertemporal hedging demand. The optimal weight in the stock index is positive (to earn the equity risk premium).

We consider two relative risk aversion levels, 5 and 1. The first is an average value in survey data.\textsuperscript{5} The second corresponds to logarithmic utility. Optimal portfolio weights for both 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 her wealth over time, which is largely unaffected by

\textsuperscript{2}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.
\textsuperscript{3}It is so because the VIX index is the market expectation of the S&P 500 variance over the next 30 days. As time goes by, the VIX index, and derivatives on it, are sensitive to the S&P 500 variance expectation beyond the desired hedging horizon. In response to the need to trade volatility with more direct instruments, as pointed out by the CBOE, 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 2 years. See http://www.cboe.com/Products/Spec_VA.aspx.
\textsuperscript{4}Egloff, Leippold, and Wu (2010) study a similar investment problem. However, there are several differences between the two studies, which are discussed in detail at the end of Section 6.1.
\textsuperscript{5}Most of the survey data suggests values of the relative risk aversion between 0.23 and 8, e.g., Meyer and Meyer (2005).
market declines. In contrast, the wealth trajectory of the log-investor exhibits large fluctuations, 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 and yearly, leads to similar results.

To further understand the performance of optimal portfolios under different economic scenarios and in terms of expected utility, we run a Monte Carlo simulation. We compare optimal portfolios in variance swaps, stock index, and bond to the stock index, and for the log-investor, to the optimal portfolio in stock index and bond. We find that the optimal portfolio including variance swaps significantly outperforms the others, both in terms of certainty equivalent and Sharpe ratio. This suggests that variance swaps have a significant economic value for risk averse investors.

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, Tauchen, and Zhou (2009), Todorov (2010), Bollerslev and Todorov (2011), Drechsler and Yaron (2011), and Mueller, Vedolin, and Yen (2011). 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, Leippold, and Wu (2010), and Aït-Sahalia, Karaman, and Mancini (2012). 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 a non-linear ordinary differential equation. Quadratic and higher order polynomial specifications


\[^7^\]Affine diffusion processes are nested in our class of quadratic diffusion processes.

of the yield curve are limited if not inexistant, Filipović (2002), and Chen, Filipović, and Poor (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 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 papers have studied dynamic optimal portfolios with stochastic investment opportunity set, e.g., Kim and Omberg (1996), Brennan and Xia (2002), Chacko and Viceira (2005), Sangvinsatios and Wachter (2005), and Liu (2007). These papers mainly focus on optimal investment in a stock and a bond. Liu and Pan (2003) extend the investment opportunity set to options, and Egloff, Leippold, and Wu (2010) to variance swaps in an affine setting. We also study optimal portfolios including variance swaps. As discussed in Section 6, our quadratic setting yields optimal strategies that are significantly different from the ones in Egloff, Leippold, and Wu (2010).

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, 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

We fix a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) where \(\mathbb{P}\) is the objective probability measure. Let \(S_t\) be a continuous semimartingale modeling the price process of a stock index with spot variance process \(v_t\). Let \(\mathbb{Q}\) be an equivalent risk neutral measure under which the risk free discounted price process follows a local martingale.

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

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

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:

\[ \sum_{i=1}^{n} \left( \log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \overset{\mathbb{P}}{\longrightarrow} \int_{t}^{T} v_s \, ds. \]

This approximation is commonly adopted in practice (e.g., Egloff, Leippold, and Wu (2010)) and quite accurate at a daily sampling frequency (e.g., Broadie and Jain (2008), and Jarrow, Kchia, Larsson, and Protter (2013)), as is the case in our dataset.\(^9\)

A variance swap initiated at \( t \) with maturity \( T \), or term \( T - t \), pays the difference between the annualized realized variance \( \text{RV}(t, T) \) and the variance swap rate \( \text{VS}(t, T) \) fixed at \( t \).\(^10\) By convention, the variance swap rate is such that the variance swap contract has zero value at inception.

No arbitrage implies that

\[ \text{VS}(t, T) = \frac{1}{T-t} \mathbb{E}_Q \left[ \int_{t}^{T} v_s \, ds \mid \mathcal{F}_t \right] \]  

where \( \mathbb{E}_Q \) denotes the expectation under \( Q \), and we assume that the risk free rate and the spot variance are independent processes under \( Q \).

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 (1), this boils down to modeling the spot variance process \( v_t = \text{VS}(t, t) \) under \( Q \). These models should be analytically tractable and yet flexible enough to reproduce the empirical features of variance swap rates. Any positive continuous semimartingale whose spot variance process coincides with \( v_t \) is then a consistent price process in the sense that \( \text{VS}(t, T) \) is the corresponding variance swap rate.

Our approach easily extends to semimartingale price processes with jumps. The spot variance is then to be set to \( v_t = \sigma_t^2 + \int_{\mathbb{R}} x^2 \, \nu_t(x) \, dx \) where \( \sigma_t^2 \) denotes the spot variance of the continuous martingale part and \( \nu_t(x) \) is the \( Q \)-compensator of the jumps of the log-price.\(^11\)

It is instructive to draw an analogy between the term structure of variance swap and interest rates. The variance swap curve reflects market expectations about future changes in spot variance,\(^9\)

\(^9\) Market microstructure noise, while generally a concern in high frequency inference, is largely a non-issue at the level of daily returns.

\(^10\) As the difference is in variance units, the payoff is converted in dollar units via a suitable notional amount.

\(^11\) Note that the spot variance \( v_t = \sigma_t^2 + \int_{\mathbb{R}} x^2 \, \nu_t(x) \, dx \) and the corresponding variance swap rates \( \text{VS}(t, T) \) can be continuous processes even if the underlying log-price exhibits jumps, i.e., \( \nu_t(x) \) is non-zero. Aït-Sahalia, Karaman, and Mancini (2012) provide empirical evidence that variance swap rates on the S&P 500 contain a non-zero \( \nu_t(x) \), and model \( v_t \) as an affine diffusion. For the empirical analysis in this paper we model variance swap rates using a general quadratic diffusion without specifying the stock index dynamics. For the optimal portfolio problem in Section 5, for tractability, we specify the stock index as a continuous process.
(1). The financial variable in interest rate models corresponding to the spot variance $v_t$ 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

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

with short-end given by $y(t, t) = r_t$. Clearly, the yield curve is a non-linear function of the short rate process. In contrast, the variance swap curve is a linear function of the spot variance process. This linear relationship gives greater flexibility for the specification of analytically tractable term structure models for variance swap 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 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 non-linear 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, Dittmar, and Gallant (2002), Chen, Filipović, and Poor (2004), and Liu (2007). 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)

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

(2)

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.

$Liu (2007)$ considers mixtures of quadratic-Gaussian and affine components in a specific setup.
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 transposition.

An \( m \)-factor quadratic variance swap model is obtained by imposing that the spot variance is a quadratic function of the state variable:

\[ v_t = g(X_t) \]

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

**Proposition 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:

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

with \( G(\tau,x) = \Phi(\tau) + \Psi(\tau)^\top x + x^\top \Pi(\tau) x, \) where the functions \( \Phi : [0, +\infty) \to \mathbb{R}, \Psi : [0, +\infty) \to \mathbb{R}^m, \) and \( \Pi : [0, +\infty) \to \mathbb{S}^m \) 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 (1) and Lemma A.2 in Appendix A with \( f(\tau,x) = \partial G(\tau,x)/\partial \tau. \) \qed
Appendix B shows that, under mild technical conditions, the converse to Proposition 3.2 also holds true: a quadratic term structure implies that the spot variance function and the state diffusion process $X_t$ be necessarily quadratic. This result implies that our quadratic model framework is exhaustive as 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 the identity $\Sigma(X_t) \Lambda_t = \Upsilon_0 + \Upsilon_1 X_t$ 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 (3)–(4) of $X_t$ and the quadratic term structure (6). 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. 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

$$dX_t = (b + \beta X_t) \, dt + \sqrt{a + \alpha X_t + AX_t^2} \, dW_t$$  \hspace{1cm} (8)  

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 (4) 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, in order that the state space 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.
on some interval \( X \) in \( \mathbb{R} \) and for some real parameters \( b, \beta, a, \alpha, \) and \( A \geq 0 \). The linear ordinary differential equations (7) 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.

**Proposition 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 \( A = \alpha = a = 0 \) and \( a > 0 \). The canonical representation is specified by \( X = \mathbb{R}, b \geq 0, \beta \in \mathbb{R}, a = 1, \alpha = 0, A \geq 0, \) and hence
  
  \[
  dX_t = (b + \beta X_t) dt + \sqrt{1 + AX_t^2} dW_t.
  \]
  
  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
  
  \[
  dX_t = (b + \beta X_t) dt + \sqrt{AX_t^2} dW_t.
  \]
  
  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
  
  \[
  dX_t = (b + \beta X_t) dt + \sqrt{X_t + AX_t^2} dW_t.
  \]
  
  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 $\mathcal{X}$ becomes bounded. The canonical representation for this equivalence class is the Jacobi process on $\mathcal{X} = [0,1]$. 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 variance swap model. Higher dimensional extensions are conceptually straightforward, but these models would be quite 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 swap data, thus higher 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

$$
dX_{1t} = (b_1 + \beta_{11} X_{1t} + \beta_{12} X_{2t}) dt + \sqrt{a_1 + \alpha_1 X_{1t} + A_1 X_{1t}^2} dW_{1t}
$$

$$
dX_{2t} = (b_2 + \beta_{22} X_{2t}) dt + \sqrt{a_2 + \alpha_2 X_{2t} + A_2 X_{2t}^2} dW_{2t}
$$

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 spot variance function is assumed to depend on $X_{1t}$ only,

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

where $x = (x_1, x_2)$, for some real parameters $\phi$, $\psi$ and $\pi$. Hence $X_{1t}$ drives the 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 (7) simplify to (C.2) in Appendix C.

The admissible specifications for $X_{2t}$ are either Class 2 or 3 with the corresponding canonical representations given by Proposition 3.3. The diffusion function of $X_{1t}$ can be of any Class 1–3 with the corresponding canonical representations from Proposition 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,
The market price of risk process is then

$$\Lambda_t = \left( \frac{\lambda_0 + \lambda_1 X_{1t}}{\sqrt{a_1 + \alpha_1 X_{1t} + A_1 X_{1t}^2}}, 0 \right)^\top. \quad (9)$$

The parameter $\lambda_0$ may 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, Filipović, and Kimmel (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 Proposition 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 (8). The following proposition formalizes the polynomial preserving property of $X_t$. The proof is given in Appendix F.

**Proposition 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

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

where $B$ is an upper triangular $(N+1) \times (N+1)$ matrix defined in (F.2) 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 spot variance as a polynomial function of the state variable, $v_t = p_0 + p_1 X_t + \cdots + p_N X_{Nt}$, for some parameters $p_i \in \mathbb{R}, i = 0, \ldots, N$. The following corollary is an immediate consequence of Proposition 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 $X_t$:

$$VS(t, T) = \frac{1}{T-t} \left( P_0(T-t) + P_1(T-t)X_t + \cdots + P_N(T-t)X_t^N \right)$$  \hspace{1cm} (10)

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

$$\frac{dP(\tau)}{d\tau} = p + B P(\tau), \quad P(0) = 0$$  \hspace{1cm} (11)

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

It follows by inspection that the system (11) is equivalent to (7) 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 directly to variance swap rates on the S&P 500, without specifying the index dynamic. An advantage of this approach is that model estimates are not impaired by potential misspecifications of the index dynamic and allows for a thorough comparison of the variance swap models.

4.1 Dataset

Our dataset consists of daily closing over-the-counter quotes of variance swap rates on the S&P 500 index, with fixed terms at 2, 3, and 6 months, and 1 and 2 years.\textsuperscript{14} It spans 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 dataset 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.

Figure 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.

\textsuperscript{14}We thank Mika Kastenholz from Credit Suisse for providing us with the variance swap data.
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 more rarely downward sloping or $\cap$-shaped.\footnote{On some occasions, the term structure is $\sim$-shaped, but the difference between, for e.g., the 2 and 3 months variance swap rates is virtually zero and this term structure is nearly $\cup$-shaped.} The bottom and peak of the $\cup$- and $\cap$-shaped parts of the term structures, can be anywhere at the 3 or 6 months or 1 year term. The slope of the term structure, measured as the difference between the 2-year and 2-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 2 or 3 months term) rises more than the long-end, producing downward sloping term structures.

Table 1 provides summary statistics of our dataset. 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 between 0.984 and 0.995, are slightly increasing with the term and imply a mean half-life of shocks between 43 and 138 days.\footnote{The half-life $H$ is defined as the time necessary to halve a unit shock and solves $\rho^H = 0.5$, where $\rho$ is the first order autocorrelation coefficient.} 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. 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 ex-post realized variance of 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 and leptokurtic than variance swap rates. This difference highlights the profitability and riskiness of shorting variance swaps, earning 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 2-month variance swaps to 0.85 (= 2.15/2.54) for 1-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 two sections.

4.2 Model Estimates

The state process \( X_t \) driving the term structure is not observed. We use the extended Kalman filter to extract the latent state and compute the likelihood of a particular model. Duffee and Stanton (2004), among others, provide a detailed description of the method. Here we briefly discuss the implementation of the filter.

\(^{17}\)To save space, factor loadings are not reported, but are available from the authors upon request.
Let $\mathbf{VS}(t)$ denote the five-dimensional vector of variance swap rates with terms $\tau_j$ equal to 2, 3, 6 months, and 1 and 2 years observed at time $t$. Define the vector-valued function $\mathbf{H}(x)$ with $j$-th component given by $G(\tau_j, x)/\tau_j$, see (6), and denote by $D_x \mathbf{H}(x)$ its derivative. The measurement equation is then linearized as follows:

$$\mathbf{VS}(t_i) = \mathbf{H}(\hat{X}_{t_i|t_{i-1}}) + D_x \mathbf{H}(\hat{X}_{t_i|t_{i-1}}) (X_{t_i} - \hat{X}_{t_i|t_{i-1}}) + \eta_{t_i}$$

where $\hat{X}_{t_i|t_{i-1}}$ denotes the time-$t_{i-1}$ prediction of $X_{t_i}$, $\eta_{t_i}$ is a normal zero-mean error term, and $t_i - t_{i-1} = 1/252$ is one day.

The state transition equation in (2) is discretized using an Euler scheme at daily frequency and parameter estimates are obtained by maximizing the (quasi) log-likelihood function

$$\sum_{i=1}^{\mathcal{N}} -\frac{1}{2} \left[ 5 \log(2\pi) + \log |V_{t_i|t_{i-1}}| + e_{t_i}^\top V_{t_i|t_{i-1}}^{-1} e_{t_i} \right]$$

where $e_{t_i} = \mathbf{VS}(t_i) - \mathbf{H}(\hat{X}_{t_i|t_{i-1}})$ is the five-dimensional vector of time-$t_i$ variance swap rate prediction errors, which in view of (12) is distributed as $D_x \mathbf{H}(\hat{X}_{t_i|t_{i-1}}) (X_{t_i} - \hat{X}_{t_i|t_{i-1}}) + \eta_{t_i}$ with covariance matrix $V_{t_i|t_{i-1}}$, and $\mathcal{N} = 2,832$ is the sample size of daily observations.

It is known that univariate affine models cannot capture the empirical features of variance swap rates, e.g., Egloff, Leippold, and Wu (2010), and Aït-Sahalia, Karaman, and Mancini (2012). 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 these features. Intuitively, the quadratic structure of the spot variance $v_t$ relaxes the constraints imposed by an affine specification and is to some extent similar to a bivariate affine structure, when the two factors ($X_t$ and $X^2_t$) are tightly related to each other.

We begin model estimations by fitting each of the three canonical representations of the univariate quadratic model in Section 3.1 to the variance swap data. We find that the largest log-likelihood of the univariate quadratic model is achieved when the state process $X_t$ is in Class 3 (Proposition 3.3). This finding is confirmed by Akaike and Bayesian Information Criteria (AIC
and BIC). Table 2 reports the corresponding parameter estimates. The model parameters are estimated rather imprecisely, as their robust standard errors are fairly large. This may suggest that the univariate quadratic model is overparameterized, in the sense that it has too many parameters to fit available variance swap data and cannot be estimated precisely. In that case, imposing certain parameter restrictions should not deteriorate the fitting significantly. We consider four parametric restrictions that induce four alternative model specifications. Each restriction 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 (8). Specification 2 constrains the spot variance function, $v_t = \phi + \psi X_t + \pi X_t^2$, 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 $v_t$ play an important role in fitting variance swap rates.

Specification 3 restricts the functional form of the spot variance by imposing the spot variance function to have exactly one root, i.e., $\psi^2 = 4\phi\pi$. This guarantees the nonnegativity of the 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 $v_t$ 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.

We now investigate whether enriching the functional form of the 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 $N = 5$. 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. The additional parameters, $p_3$, $p_4$, $p_5$, allow for a modest increase in the log-likelihood, which is

---

18 When the state process $X_t$ is in Class 1, 2, and 3, AIC are $-97,316$, $-97,310$ and $-97,346$, and BIC are $-97,268$, $-97,262$ and $-97,298$, respectively. Both criteria achieved the minimum value when $X_t$ is in Class 3.

19 Denote $L_U$ the likelihood of the unrestricted model and $L_R$ the likelihood of the restricted model. Under the null hypothesis that the restriction holds true in the data generating process, the likelihood ratio statistic, $2 \log (L_U / L_R)$, has asymptotically a chi-square distribution with degrees of freedom equal to the number of restrictions. If the null hypothesis were marginally rejected, the outcome of the test would have to be interpreted cautiously, as (13) is the quasi log-likelihood. However, as discussed below, the four parametric restrictions are strongly rejected by LR tests.

20 Under the four null hypotheses, namely, 1) $A = 0$, 2) $\pi = 0$, 3) $\psi^2 = 4\phi\pi$ and 4) $\phi = \psi = 0$, the LR test statistics are 386, 620, 264, 286, respectively, and are all well above any conventional critical value.

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 $p_2 = \pi$. 

not statistically significant according to a LR test, and a modest reduction of the AIC and BIC. Moreover, model parameters are still estimated quite imprecisely, according to robust standard errors. Thus, the polynomial form of the spot variance helps only marginally to improve the fitting of variance swap rates.

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 values. Interestingly, nearly all parameters are estimated very precisely, as can be seen from the small robust standard errors.

The log-likelihood of the bivariate model is significantly larger than the log-likelihoods 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 27,980. The Vuong (1989) statistic of the bivariate model versus the univariate polynomial model is 46.3. These statistics are both highly significant and strongly reject the null hypothesis that the bivariate quadratic model is equivalent to any of the other two 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 10.6 and 9.8, respectively, and are both highly significant. This further supports that the bivariate quadratic model fits variance swap rates significantly better than the univariate models.

Finally, Figure 2 shows the filtered trajectories of the state process \( X_t \) in the bivariate model. It suggests a natural interpretation of its components. \( X_{1t} \) is more volatile and mimics the time series trajectories of the short term variance swap rates, mainly capturing sudden movements in

\(^{22}\)The asymptotic distribution of the test statistics under the null hypotheses are the chi-square with 5 degrees of freedom and standard normal, respectively. Recall that the bivariate quadratic model nests the univariate quadratic model. Setting \( b_2 = \beta_{22} = a_2 = \alpha_2 = A_2 = 0 \) in the bivariate model, i.e., imposing 5 parameter restrictions, implies that \( X_{2t} \) is constant and can be normalized to 1 for identification purposes. Thus, \( \beta_{12} \) in the bivariate model parametrization corresponds to \( b_1 \) in the univariate model parametrization.

\(^{23}\)We view this test as a main robustness check of the previous LR and Vuong’s tests. Given the autocorrelation and heteroscedasticity 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).
those rates. $X_{2t}$ is more persistent and mainly captures long term movements in variance swap rates.

### 4.3 Goodness-of-fit Tests

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

Table 3 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 outperforms the other models in terms of bias and root mean square error (RMSE), and often to a large extent. For example, in the out-of-sample period, the RMSE of the bivariate quadratic model for the 2-month variance swap rates is 65% lower than the RMSE of the other models. The comparison between the bivariate quadratic model and the univariate polynomial model is particularly interesting, as the two 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, both in-sample and out-of-sample.

Figure 3 shows actual and model-based trajectories under the bivariate quadratic model of the 2-month and 2-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 2-year variance swap rates is noticeable in the out-of-sample period, which includes the market tumults of Fall 2008.

To assess the statistical differences of pricing errors of the different models, we run various Diebold–Mariano (DM) tests.\footnote{We follow the standard practice in the literature of using Diebold–Mariano tests to draw conclusions about models, rather than about model forecasts; see Diebold (2012) for a discussion of this point.} For each model and each term, the time-$t$ loss function is given by the absolute pricing error, $L(e_t) = |e_t|$, where $e_t = \sqrt{G(\tau, X_t)/\tau} - \sqrt{\text{VS}(t, t + \tau)}$.\footnote{Note that the time-$t$ pricing error considered here uses the time-$t$ filtered value of $X_t$, not its prediction as in (13), which makes the DM tests complementary to the likelihood-based analysis in the previous section.} Denote the time-$t$ loss differential 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, $d_t^{(p,b)}$, is similarly defined. Under the null hypothesis that the two models have pricing errors of equal magnitude, $\mathbb{E}_P[d_t^{(u,b)}] = 0$. If the bivariate model outperforms the univariate model, then $\mathbb{E}_P[d_t^{(u,b)}] > 0$. The DM statistic is
the t-statistic for this test.\textsuperscript{26} Table 3 reports the results. DM tests strongly confirm that the bivariate model significantly outperforms the univariate quadratic and polynomial models.\textsuperscript{27} As a robustness check, we also run DM tests using pricing errors in variance units, rather than volatility units, i.e., \( e_t = G(\tau, X_t)/\tau - \text{VS}(t, t + \tau) \), and using quadratic loss functions, rather than absolute loss functions. These additional DM tests strongly confirm the results in Table 3.

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

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

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 a benchmark, 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. This 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 4 reports the regression results.\textsuperscript{28} Interestingly, 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 \( H_0 : \gamma_1 = 1 \). Only the bivariate quadratic model passes all these tests and outperforms the martingale model. The univariate quadratic and polynomial models provide biased and inefficient forecasts in most cases. The martingale model provides relatively accurate forecasts for the 1-day horizon, but its forecasting accuracy deteriorates when moving to the 10-day horizon. The univariate quadratic model provides the least accurate forecasts. To summarize, also predictive regressions strongly confirm that the bivariate quadratic model captures well the variance swap term structure dynamics.

\textsuperscript{26} The standard error is computed using the Newey and West (1987) autocorrelation and heteroscedasticity consistent variance estimator with the number of lags optimally chosen according to Andrews (1991).

\textsuperscript{27} The DM test statistics are positive but not significant only for the 6 month and 1 year variance swaps in the out-of-sample period, which may be due to the limited sample size, i.e., 794 daily observations.

\textsuperscript{28} 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.
even outperforming the martingale model.

5 Optimal Portfolios: Theoretical Setup

In this section, we formalize and solve an optimal portfolio problem for variance swaps, stock index, and risk free bond. 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 (2) where $W_t$ is a standard $d$-dimensional Brownian motion under the risk neutral measure $Q$. The spot variance, $v_t$, and variance swap rates, $\text{VS}(t, T)$, are given as functions of the state variable, $X_t$, by (5) and (6), respectively.

5.1 Investing in Variance Swaps

We 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 notational long position in this variance swap is given by

$$\Gamma_t = \mathbb{E}_Q \left[ e^{-r(T^*-t)} \frac{1}{\tau} \left( \int_{t^*}^{T^*} v_s \, ds - \tau \text{VS}(t^*, T^*) \right) | \mathcal{F}_t \right]$$

$$= \frac{e^{-r(T^*-t)}}{\tau} \left( \int_{t^*}^{t} v_s \, ds + (T^*-t)\text{VS}(t, T^*) - \tau \text{VS}(t^*, T^*) \right)$$

where $r$ is the constant risk free rate. In stochastic differential form, we obtain $d\Gamma_t = \Gamma_t \cdot \tau dt + dM_t$ with the $Q$-martingale increment excess return

$$dM_t = \frac{e^{-r(T^*-t)}}{\tau} (v_t dt + d((T^*-t)\text{VS}(t, T^*))) = \frac{e^{-r(T^*-t)}}{\tau} \nabla_x G(T^* - t, X_t)^\top \Sigma(X_t) dW_t$$

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. This makes 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

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

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 3-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

\[
dV_t = r dt + n_t \frac{e^{-r(T^*(t) - t)}}{\tau} \nabla_x G(T^*(t) - t, X_t)^\top \Sigma(X_t) dW_t \quad (14)
\]

where the excess return on the right hand side is a \( Q \)-martingale increment.

### 5.2 Optimal Portfolio Problem

We consider an investment universe consisting of the risk free bond, stock index \( S \), 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 stock index price process has \( Q \)-dynamics

\[
dS_t \over S_t = r dt + \sqrt{g(X_t)} R(X_t)^\top dW_t \quad (15)
\]

where \( R = (R_1, \ldots, R_d)^\top : \mathcal{X} \to \mathbb{R}^d \) is some function with constant norm \( \|R\| \equiv 1 \), modeling the correlation between stock returns and spot variance changes.\(^{29}\)

Let \( w_t \) denote the fraction of wealth invested in the stock index 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

\[
1 - w_t - n_t^\top \Gamma_t,
\]

where \( \Gamma_t \) is the vector of the variance swap spot values. Combining (14) and (15),

\(^{29}\)The price process dynamic in (15) is tantamount to \( dS_t / S_t = r dt + \sqrt{v_t} dB_t \) for the scalar \( Q \)-Brownian motion \( B_t \) defined as \( dB_t = R(X_t)^\top dW_t \). That is, \( B_t \) and \( W_t \) have correlation \( d(B_t, W_t) / dt = R_b(X_t) \).
the resulting wealth process $V_t$ has $\mathbb{Q}$-dynamics

$$\frac{dV_t}{V_t} = r \, dt + \left( n_t^\top, w_t \right) \mathcal{G}(t, X_t) \, dW_t$$

with the $(n + 1) \times d$-volatility matrix $\mathcal{G}(t, X_t)$ defined by

$$\mathcal{G}(t, x) = \begin{pmatrix} 
D(t, x) & 0_{n \times 1} \\
0_{1 \times m} & \sqrt{g(x)} \\
\end{pmatrix} \begin{pmatrix} 
\Sigma(x) \\
R(x)^\top \\
\end{pmatrix}$$

where $D(t, x)$ is the $n \times m$-matrix whose $i$th row is given by $(e^{-r(T_i^*(t) - t)/\tau_i}) \nabla_x G(T_i^*(t) - t, x)^\top$.

We now formulate the optimal portfolio problem. We fix a finite time horizon $T$, and maximize expected utility from terminal wealth of an investor with power utility function $u(V) = V^{1-\eta}/(1-\eta)$ and constant relative risk aversion $\eta > 0$. That is, we solve the optimization problem

$$\max_{n, w} \mathbb{E}_P [u(V_T)]$$

for some given initial wealth $V_0$. The investor takes the market price of risk as given, which we specify as follows.\textsuperscript{30} The objective probability measure $\mathbb{P}$ is related to the risk neutral measure $\mathbb{Q}$ on $\mathcal{F}_T$ via the Radon–Nikodym density

$$\frac{dQ}{dP}\big|_{\mathcal{F}_T} = \exp \left( -\int_0^T \Lambda(X_t)^\top dW_t + \frac{1}{2} \int_0^T \|\Lambda(X_t)\|^2 dt \right)$$

for some sufficiently regular market price of risk function $\Lambda : \mathcal{X} \rightarrow \mathbb{R}^d$.\textsuperscript{31}

Since the number $n$ of on-the-run variance swaps available in the market can be chosen arbitrarily large, it is no essential loss in generality to assume market completeness.

**Assumption 5.1.** The market is complete with respect to the stock index and the $n$ on-the-run $\tau_i$-variance swaps. Specifically, we assume that the filtration $\mathcal{F}_t = \mathcal{F}_t^W$ is generated by $W_t$, and that

\textsuperscript{30}By exogenously specifying the market price of risk we take a partial equilibrium view. This approach is standard in the optimal allocation literature, e.g., Liu and Pan (2003), Chacko and Viceira (2005), and Liu (2007).

\textsuperscript{31}If we denote the corresponding Girsanov transformed $\mathbb{P}$-Brownian motion by $dW_t^\mathbb{P} = dW_t - \Lambda(X_t) dt$, we obtain the familiar stochastic exponential representation

$$\frac{dQ}{dP}\big|_{\mathcal{F}_T} = \exp \left( -\int_0^T \Lambda(X_t)^\top dW_t^\mathbb{P} - \frac{1}{2} \int_0^T \|\Lambda(X_t)\|^2 dt \right).$$
the \((n + 1) \times d\)-volatility matrix \(G(t, X_t)\) is injective \(dt \otimes dQ\)-a.s.

Appendix H shows that, as a consequence of Assumption 5.1, the dimension \(d\) of the Brownian motion cannot exceed the number \(m\) of factors and the number \(n\) of variance swaps by more than one, i.e., \(d \leq m + 1\) and \(d \leq n + 1\). Moreover, the maturity date functions \(T_i^*(t)\) have to be mutually different for all \(t\).

We now state the existence and characterization result for the optimal strategy with standard technical assumptions and proof given in Appendix G.

**Theorem 5.2.** Under Assumptions 5.1 and G.1–G.3 in Appendix G there exists an optimal strategy \(n_t^*, w_t^*\) given as solution of the linear equation

\[
G(t, X_t) ^\top \begin{pmatrix} n_t \\ w_t \end{pmatrix} = \frac{1}{\eta} \Lambda(X_t) + \Sigma(X_t) ^\top \nabla x h(T - t, X_t) \tag{20}
\]

where the function \(h(\tau, x)\) is defined in (G.3) in Appendix G.

The optimal strategy is thus 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 volatility of the stock index. We discuss the computation of \(\nabla x h(T - t, X_t)\) in Appendix I.

The following corollary shows that variance swaps can be used to span volatility risk. The optimal investment in the stock index is thus only seeking its risk premium. In their affine setting, Egloff, Leippold, and Wu (2010) reach the same conclusion. Corollary 5.3 extends this result to a general multivariate diffusion setting. The proof is given in Appendix H.

**Corollary 5.3.** If \(d > m\) then the optimal investment in the stock index, \(w_t^*\), is fully determined by the myopic term and does not depend on the choice of the variance swaps.
5.3 Bivariate Quadratic Model Specification

We now resume the bivariate quadratic variance 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

$$
\Sigma(x) = \begin{pmatrix}
\sqrt{1 + A_1 x_1^2} & 0 & 0 \\
0 & \sqrt{x_2 + A_2 x_2^2} & 0
\end{pmatrix},
$$

To account for the widely documented correlation between index returns and spot variance changes, e.g., Broadie, Chernov, and Johannes (2007), and Aït-Sahalia and Kimmel (2007), the correlation vector function is chosen to be of the form $R(x) = \left( R_1(x), 0, \sqrt{1 - R_1(x)^2} \right) ^\top$. The correlation between index returns and variance changes is then given by

$$
\text{Corr}\left( \frac{dS_t}{S_t}, dv_t \right) = \frac{\nabla_x g(X_t)^\top \Sigma(X_t)}{\| \nabla_x g(X_t)^\top \Sigma(X_t) \|} R(X_t) = \text{sign} (\psi + 2\pi X_{1t}) \ R_1(X_t).
$$

We set $R_1(x) = -\text{sign} (\psi + 2\pi x_1) \times 0.7$ to achieve a constant correlation of $-0.7$, in line with the literature. As a consequence, we obtain $R_3(x) = \sqrt{1 - 0.7^2} = 0.714$.

Consistently with (9), we specify the market price of risk function as

$$
\Lambda(x) = \begin{pmatrix}
\lambda_0 + \lambda_1 x_1 \\
\frac{1}{\sqrt{1 + A_1 x_1^2}} \\
0
\end{pmatrix} \Lambda_3(x) ^\top
$$

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 $R_3(X_t)\Lambda_3(X_t)$ has a direct impact on the equity risk premium, which is given by

$$
\frac{\mathbb{E}_P [dS_t / S_t | \mathcal{F}_t] - \mathbb{E}_Q [dS_t / S_t | \mathcal{F}_t]}{dt} = \sqrt{g(X_t)} R(X_t)^\top \Lambda(X_t).
$$

Based on our estimations, $R_3(X_t)\Lambda_3(X_t)$ is much 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 (22). Clearly, $\|\Lambda(x)\|^2$ needs to be specified so that the argument in the square root in (22) is nonnegative for all $x \in X$. We specify it as proportional to spot variance

$$\|\Lambda(x)\|^2 = \kappa g(x)$$  \hspace{1cm} (24)

with $\kappa \geq \kappa^* = \max_{x \in X} \Lambda_1(x)^2 / g(x)$.\footnote{Alternatively, we could specify $\|\Lambda(x)\|^2 = c$, for some constant $c \geq \max_{x \in X} \Lambda_1(x)^2$. This specification implies that $\nabla_x h(\tau, x) = 0$, because the function $h$ defined in (G.3) in Appendix G no longer depends on $x$. Hence, in this case the optimal investment in variance swaps and stock index in (20) consists of myopic demand alone and is available in closed form.} Since $\Lambda_1(x)$ is uniformly bounded in $x$, it follows that the spot variance $g(x)$ and the equity risk premium (23) 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 (24) of $\|\Lambda(x)\|^2$.

We set $\kappa = 1.58$ in (24), which corresponds to a sample average equity risk premium of 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, Engle, and Mancini (2008).}. Figure 4 shows the induced model-based time series of equity risk premium (23), which exhibits significant time variation and countercyclical behavior. Figure 4 also shows the induced time series of variance risk premium, which is given by

$$\frac{\mathbb{E}_P[dv_t | \mathcal{F}_t] - \mathbb{E}_Q[dv_t | \mathcal{F}_t]}{dt} = (\psi + 2\pi X_{1t})(\lambda_0 + \lambda_1 X_{1t}).$$  \hspace{1cm} (25)

The variance risk premium is procyclical, takes both positive and negative values, and is negative most of the time. Both model-based equity and variance risk premiums are economically sizable and follow plausible dynamics which are in line with recent studies, e.g., Bollerslev and Todorov (2011), Aït-Sahalia, Karaman, and Mancini (2012), and Martin (2013). These features lend further empirical support to our quadratic variance swap model.

### 5.4 Optimal Portfolios in the Bivariate Quadratic Model

We assume that $n = 2$ variance swaps are available for investment, 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 and Corollary 5.3 are satisfied. Appendix J sketches the arguments.

The optimal fraction of wealth invested in the stock index is given by

\[
w^*_t = \frac{\Lambda_3(X_t)}{\eta \sqrt{g(X_t) R_3(X_t)}}
\]

which is recovered by setting \( \mathbf{v} = (0, 0, 1)^\top \) in (H.2) in Appendix H. As stated in Corollary 5.3, it is fully determined by the myopic term and does not depend on the choice of the variance swaps. It follows that the optimal weight \( w^*_t \), while being state-dependent, is uniformly bounded from below and above with sharp bounds given by

\[
\frac{\sqrt{\kappa - \kappa^*}}{\eta \sqrt{1 - 0.7^2}} \leq w^*_t \leq \frac{\sqrt{\kappa}}{\eta \sqrt{1 - 0.7^2}}.
\]

The intertemporal hedging demand is fully borne by the optimal investment in the variance swaps. Plugging (26) in (20) 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

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

We provide a closed form approximation of \( \nabla_x h(T - t, X_t) \) in (J.1) in Appendix J.

## 6 Optimal Portfolios: Empirical Findings

We perform an empirical analysis of optimal portfolios in the above bivariate quadratic model. The investment universe consists of the risk free bond, the stock index, and 3-month and 2-year variance swaps, rolled over monthly and yearly, respectively. The initial wealth is normalized to 100. 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.\(^{34}\) For a comparison we also consider \( \eta = 1 \), which corresponds to logarithmic utility. When the log-investor has only access to the stock

\(^{34}\)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.
index and the risk free bond, it is well known that the optimal weight in the stock index is given by the ratio of equity risk premium and spot variance, \( R(X_t)^\top \Lambda(X_t)/\sqrt{g(X_t)} \), e.g., Filipović and Platen (2009). Optimal portfolios are rebalanced daily. That is, each day optimal portfolio weights are adjusted according to (26) and (28). We also consider proxy portfolios with lower rebalancing frequencies. We first study the optimal portfolios using historical data. We then perform a Monte Carlo analysis of optimal portfolios. Section 6.3 discusses several robustness checks that largely confirm our results.

### 6.1 Optimal and Proxy Portfolios

Figures 5 and 6 display the optimal portfolio weights in the stock index, and in on-the-run 3-month and 2-year variance swaps, for \( \eta = 5 \) and \( \eta = 1 \), respectively. The optimal weights in variance swaps follow a short-long strategy, with a short position in the 2-year variance swap and a long position in the 3-month variance swap. As the negative variance risk premium in 2-year variance swaps is larger in absolute value than the risk premium in 3-month variance swaps (Section 4.1), going short in 2-year variance swaps allows to reap the large risk premium. Short positions in 2-year variance swaps are partially hedged via long positions in 3-month variance swaps, limiting portfolio losses when volatility increases. The 3-month variance swap is more sensitive to volatility increases than the 2-year variance swap, and is thus an effective hedging instrument.

The optimal weights in variance swaps exhibit significant periodic patterns, with increasing weights in absolute value when their maturities are approaching. Intuitively, close to maturity, most realized variance has 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 (26) 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 exhibits no periodic pattern, which is in line with Corollary 5.3. The bounds (27) have a stabilizing effect on the optimal fraction of wealth invested in the stock index. The optimal weights in the stock index and the 3-month variance swap are significantly larger for \( \eta = 1 \) than for \( \eta = 5 \). The log-investor seeks significantly more exposure to the stock index and to a lesser extent to the 2-year variance swap, and needs a larger position in the 3-month variance swap to hedge this
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 $D(t, X_t)$ in equation (28) for $n_t$ close to singular. However, low volatility also implies small actual returns in the stock index and variance swaps. This in turn annihilates the impact of oscillating portfolio weights on the wealth process, resulting in non-oscillating wealth trajectories, as shown below in Figures 8 and 9.

In view of Corollary 5.3, the optimal weight in the stock index includes myopic demand only. Figure 7 decomposes the optimal weights in variance swaps into myopic and intertemporal hedging demands according to (20), for $\eta = 5$. Myopic and intertemporal hedging demands are positive for the 3-month, and negative for the 2-year variance swaps, reflecting the risk premium versus hedging trade-off in these contracts. The periodic patterns in the optimal weights in variance swaps are mainly borne by the intertemporal hedging demand, while the myopic demand only exhibits little periodicity. The hedging demand gets closer to zero as the terminal investment date approaches. As the investment horizon shrinks, the need to hedge against future fluctuations in the investment opportunity set becomes less important and the investor behaves more myopically. For the log-optimal portfolio, $\eta = 1$, it is well known that the intertemporal hedging demand is zero.\footnote{It follows from (G.3) in Appendix G that the function $h(\tau, x)$ in equation (20) is zero for $\eta = 1$.}

Figure 8 shows the wealth trajectory of the optimal portfolio for $\eta = 5$. The wealth trajectory exhibits low volatility and steady growth. This results in a Sharpe ratio of 1.46%, which is larger than the Sharpe ratio of 1.20% of the S&P 500. Thus, optimally investing in variance swaps 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 S&P 500 yields a higher terminal wealth than the optimal portfolio. Indeed, the optimal portfolio is not designed to maximize terminal wealth. Compared to the stock index, the optimal portfolio may exhibit lower returns on some occasions but it has always a lower volatility. This implies that including variance swaps in the portfolio of a risk averse investor brings more utility than investing in the stock index only.

Figure 9 shows the wealth trajectories of the optimal portfolios with and without variance swaps for $\eta = 1$. The log-optimal wealth process including variance swaps has a Sharpe ratio of 1.54%, and exhibits significantly larger fluctuations than the S&P 500, which contrasts with the optimal

exposure.
wealth trajectory of the more risk averse investor with \( \eta = 5 \). This suggests 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. The trajectory of the log-optimal portfolio in stock and bond is very similar to the S&amp;P 500, as the optimal weight in the stock index turns out to be close to one.

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 and 3-month variance swap positions are rebalanced monthly, and the 2-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 3-month and 2-year variance swaps, respectively, according to the portfolio weights \( \pi_{it}^* \) given as exponentially weighted average of past optimal portfolio weights,

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

where \( \omega_{it} = e^{-(t_{ik}^* - t)} \).\(^{36}\) These portfolio weights attempt to capture the periodic pattern of the optimal weights over the lifetime of the variance swaps. The rationale for assessing the performance of this proxy portfolio is twofold. First, low rebalancing frequencies obviously reduce transaction costs when implementing the portfolio strategy in practice. Second, the portfolio gains can be evaluated using market data, without resolving to model-based variance swap rates.\(^{37}\) Figures 8 and 9 show that the wealth trajectories of these proxy portfolios are similar to the ones of the optimal portfolios, for \( \eta = 5 \) and \( \eta = 1 \). This is rather remarkable. Although this is mainly an in-sample result based on one historical realization, it suggests that our optimal portfolio strategies have potential to be implemented in practice.

The results documented above differ from those in Egloff, Leippold, and Wu (2010) in a number of ways. In their 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 one can invest in newly issued variance swaps at zero spot value (“sliding” variance swap

\(^{36}\)We set \( \pi_{i0} = n_{i0}^* \) for the initial holding period.

\(^{37}\)The actual 2-month and 1-year variance swap rates needed to evaluate the gains in the 3-month and 2-year variance swaps at the rebalancing dates, respectively, are available in our dataset. The realized variance is given by the sum of squared daily log-returns of the S&amp;P 500.
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 $\eta = 1$. 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.\(^{38}\)

### 6.2 Monte Carlo Analysis

The optimal and proxy portfolios above are based on the historical realization of variance swap rates and the S&P 500. In this section we perform a Monte Carlo analysis of optimal and proxy portfolios. The goal is to evaluate their performance under different economic scenarios, and to assess the utility effect of a combined investment in variance swaps, stock index, and bond versus investing in stock index and bond, or holding the stock index only.

We simulate 10,000 trajectories of the bivariate quadratic model in Section 5.3, and measure the performance of the optimal and proxy portfolios. The investment horizon is $T = 2$ years. Table 5 reports certainty equivalent, implied rate, average Sharpe ratio, and average terminal wealth. The certainty equivalent is defined as the initial amount of money to be invested in the risk free bond that would yield the same terminal utility as the respective portfolio. The implied rate is defined as the constant annual rate of return on the initial wealth of 100 that would be needed to achieve the same terminal utility. Sharpe ratios are computed using daily changes of the portfolio value.

Optimal and proxy portfolios including variance swaps systematically outperform the stock index, as well as the log-optimal stock-bond portfolio. This holds true in terms of certainty equivalent and Sharpe ratio, irrespective of rebalancing frequency and risk aversion, highlighting the added economic value of variance swaps. The differences are economically important. For $\eta = 5$, the optimal portfolio has an implied rate of 2.53%. This is much higher than the respective implied rate of $-2.73\%$ of the stock index. For $\eta = 1$, the implied rate of the optimal portfolio including variance swaps is 4.34%, which is 61 and 64 basis points higher than the implied rate of the log-optimal

\(^{38}\)As mentioned above, our optimal strategies take short positions in the long term variance swap (to earn the variance risk premium), long positions in the short term variance swap (to hedge volatility increases) and long positions in the stock index (to earn the equity risk premium). Egloff, Leippold, and Wu (2010) find opposite trading directions in their optimal strategy.
stock-bond portfolio and of the stock index, respectively. Proxy portfolios have smaller certainty equivalents than the respective optimal portfolios, but still outperform the stock index, as well as the optimal stock-bond portfolio for $\eta = 1$.

Sharpe ratios of optimal and proxy portfolios including variance swaps are higher than Sharpe ratios of the stock index. This is interesting because the optimal portfolio is not designed to maximize the Sharpe ratio. It confirms our previous empirical findings, which were based on a single historical trajectory of variance swap rates and the stock index.

The expected terminal wealth corresponds to the utility of a risk neutral investor. Such an investor prefers the log-optimal portfolio including variance swaps, that yields an average terminal wealth of 114.06, over the stock index. In contrast, a risk averse investor with $\eta = 5$ prefers to include variance swaps in her portfolio at the cost of a lower expected terminal wealth of 106.08.

To summarize, variance swaps have a significant economic value for risk averse investors. Adding variance swaps to a portfolio improves its performance, also when rebalanced infrequently.

6.3 Robustness Checks

We performed several robustness checks that largely confirm our optimal portfolio results.

Optimal portfolios above are based on 3-month and 2-year variance swaps. Optimal portfolios based on variance swaps with other term combinations (such as 3-month and 1-year, 6-month and 1-year, 6-month and 2-year) have similar performance. The same holds when using different roll-over periods (such as daily, half term, or term of the variance swaps). 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 and roll-over periods has no impact on the optimal wealth trajectory. In particular, the optimal portfolio weight in the stock index neither depends on variance swap terms nor on roll-over periods, Corollary 5.3.

Besides the risk aversion levels of $\eta = 5$ and 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.

We also considered other investment horizons, such as 5 and 10 years. The pattern of optimal portfolio weights is only marginally affected by the choice of the investment horizon.
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 (24) accordingly. This leads to smaller portfolio weights in the stock index, as theory predicts, and the pattern of the optimal weights in the variance swaps are essentially unaffected.

Risk aversion has a nonlinear impact on myopic and intertemporal hedging demands.\textsuperscript{39} For the logarithmic utility case, $\eta = 1$, the intertemporal hedging demand is zero. When $\eta$ increases, the intertemporal hedging demand first increases in absolute value, peaks when $\eta$ is between 2 and 3, and then decreases and approaches zero when $\eta$ increases further. The investor becomes more and more risk averse and eventually holds only the risk free bond. Even for high values of risk aversion, such as $\eta = 30$, intertemporal hedging demands still exhibit period patterns similar to Figure 7.

7 Conclusion

We introduce a novel class of quadratic term structure models for variance swaps, which are among the most important volatility derivatives. The multivariate state variable follows a quadratic diffusion process. The variance swap curve is quadratic in the state variable, and available in closed form in terms of a linear ordinary differential equation, greatly facilitating empirical applications. Various goodness-of-fit tests show that quadratic models fit variance swap rates remarkably well and largely outperform nested specifications, including popular affine models. We also study dynamic optimal portfolios in variance swaps, stock index and risk free bond. Optimal portfolio weights are available in quasi-closed form in terms of a Taylor series expansion involving conditional moments of the state process, which are available in closed form. The empirical analysis of optimal portfolios shows that optimal portfolio weights in variance swaps follow 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 hedge volatility increases). Such portfolio weights exhibit strong periodic patterns, which depend on the roll-over period and maturity of the variance swaps, and which are mainly borne by the intertemporal hedging demand. 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. A Monte Carlo study shows that in both cases the added economic value of variance swaps, in terms of expected utility, is substantial.

\textsuperscript{39}Liu (2007) discusses this point for optimal portfolios in stock and bond.
<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>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 A: Variance swap rates

<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>4.54</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 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>−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>

Panel C: Realized variance swap payoffs

Table 1: Dataset 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 index 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, and includes 2,832 daily observations. Out-of-sample period is from April 3, 2007 to June 7, 2010, and includes 794 daily observations.
<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>2.005</td>
<td>32.240</td>
<td>0.577</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>-0.742</td>
<td>25.440</td>
<td>-0.450</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>4.232</td>
<td>0.216</td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_1$</td>
<td>0.402</td>
<td>0.897</td>
<td>1.0610^{-4}</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.182</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-0.248</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_2$</td>
<td>0.010</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>0.023</td>
<td>1.641</td>
<td>0.007</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.243</td>
<td>25.070</td>
<td>-0.002</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.016</td>
<td>0.017</td>
<td>0.009</td>
</tr>
<tr>
<td>$\psi$</td>
<td>-0.002</td>
<td>0.011</td>
<td>0.003</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.002</td>
<td>0.044</td>
<td>0.017</td>
</tr>
<tr>
<td>$p_3$</td>
<td></td>
<td></td>
<td>-0.001</td>
</tr>
<tr>
<td>$p_4$</td>
<td></td>
<td></td>
<td>-0.006</td>
</tr>
<tr>
<td>$p_5$</td>
<td>0.002</td>
<td>0.349</td>
<td></td>
</tr>
</tbody>
</table>

Log-likelihood 48,681 48,933 62,671
AIC -97,346 -97,844 -125,310
BIC -97,298 -97,778 -125,245

Table 2: Model estimates. Entries are parameter estimates (Est.) for the univariate quadratic, univariate polynomial and bivariate quadratic models and corresponding robust standard errors (S.E.). Identifiable, thus restricted, versions of the following model are estimated: dynamics $dX_{1t} = (b_1 + \beta_{11}X_{1t} + \beta_{12}X_{2t}) \, dt + \sqrt{a_1 + \alpha_1X_{1t} + A_1X_{1t}^2} \, dW_{1t}$, $dX_{2t} = (b_2 + \beta_{22}X_{2t}) \, dt + \sqrt{a_2 + \alpha_2X_{2t} + A_2X_{2t}^2} \, dW_{2t}$; spot variance $v_t = \phi + \pi X_{1t} + \psi X_{1t}^2 + p_3 X_{1t}^3 + p_4 X_{1t}^4 + p_5 X_{1t}^5$; market price of risk $(\lambda_0 + \lambda_1 X_{1t})/\sqrt{a_1 + \alpha_1X_{1t} + A_1X_{1t}^2}$ for the Brownian motion $W_{1t}$. An empty entry means that the parameter is set to zero because of model identification. AIC and BIC are Akaike and Bayesian Information Criteria, respectively. Sample data are variance swap rates on the S&P 500, with terms of 2, 3, 6, 12, 24 months, from January 4, 1996 to April 2, 2007, and include 2,832 daily observations.
<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>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.10</td>
<td>1.69</td>
<td>9.34</td>
<td>0.13</td>
<td>1.67</td>
<td>9.03</td>
<td>-0.01</td>
<td>0.49</td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>1.14</td>
<td>9.36</td>
<td>0.16</td>
<td>1.09</td>
<td>9.22</td>
<td>0.12</td>
<td>0.40</td>
</tr>
<tr>
<td>6</td>
<td>-0.08</td>
<td>0.57</td>
<td>5.88</td>
<td>0.01</td>
<td>0.56</td>
<td>7.03</td>
<td>0.08</td>
<td>0.44</td>
</tr>
<tr>
<td>12</td>
<td>-0.17</td>
<td>1.13</td>
<td>6.18</td>
<td>-0.10</td>
<td>1.14</td>
<td>5.86</td>
<td>-0.08</td>
<td>0.29</td>
</tr>
<tr>
<td>24</td>
<td>0.21</td>
<td>1.55</td>
<td>6.22</td>
<td>0.03</td>
<td>1.46</td>
<td>5.48</td>
<td>0.07</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Panel B: Out-of-sample

<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>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.12</td>
<td>2.42</td>
<td>6.30</td>
<td>-0.08</td>
<td>2.14</td>
<td>5.98</td>
<td>0.13</td>
<td>0.80</td>
</tr>
<tr>
<td>3</td>
<td>-0.26</td>
<td>1.48</td>
<td>6.06</td>
<td>-0.16</td>
<td>1.33</td>
<td>5.20</td>
<td>-0.03</td>
<td>0.61</td>
</tr>
<tr>
<td>6</td>
<td>-0.19</td>
<td>1.35</td>
<td>0.39</td>
<td>0.02</td>
<td>1.41</td>
<td>0.29</td>
<td>-0.13</td>
<td>1.36</td>
</tr>
<tr>
<td>12</td>
<td>0.42</td>
<td>1.96</td>
<td>0.13</td>
<td>0.60</td>
<td>1.91</td>
<td>0.06</td>
<td>0.12</td>
<td>1.96</td>
</tr>
<tr>
<td>24</td>
<td>0.68</td>
<td>4.08</td>
<td>4.15</td>
<td>0.44</td>
<td>3.67</td>
<td>4.31</td>
<td>0.20</td>
<td>2.26</td>
</tr>
</tbody>
</table>

Table 3: 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., $\left(\frac{\sqrt{G(\tau, X_t)} - \sqrt{VS(t, t+\tau)}}{\tau}\right) \times 100$. Entries are mean (Bias) and root mean square error (RMSE) of pricing errors for variance swap rates under the univariate quadratic, univariate polynomial and bivariate quadratic models. DM$_u$ (respectively, DM$_p$) is the Diebold–Mariano test statistic of the univariate quadratic (respectively, polynomial) model versus the bivariate quadratic model, Section 4.3. Under the null hypothesis that the univariate quadratic (respectively, polynomial) 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 competing 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, and includes 2,832 daily observations. Panel B shows pricing error statistics for the out-of-sample period, which is from April 3, 2007 to June 7, 2010, and includes 794 daily observations.
<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Martingale $\gamma_0$ $\gamma_1$</th>
<th>Univ. quad. $\gamma_0$ $\gamma_1$</th>
<th>Univ. poly. $\gamma_0$ $\gamma_1$</th>
<th>Biv. quad. $\gamma_0$ $\gamma_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel A: 1-day ahead prediction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.14 0.98 2.11 -1.07 -0.36 1.07</td>
<td>0.00 0.00 0.00 0.00 0.01 0.04</td>
<td>0.09 0.98 0.14 0.22</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.09 0.98 2.83 -1.33 -0.06 1.00</td>
<td>0.00 0.01 0.00 0.00 0.29 0.75</td>
<td>0.12 0.97 0.02 0.01</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.06 0.99 15.55 -9.24 0.26 0.95</td>
<td>0.01 0.04 0.00 0.00 0.00 0.00</td>
<td>-0.08 1.01 0.26 0.37</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.04 0.99 0.56 0.90 0.54 0.89</td>
<td>0.05 0.13 0.00 0.01 0.00 0.00</td>
<td>-0.03 1.01 0.78 0.70</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.03 0.10 2.21 0.33 0.51 0.91</td>
<td>0.05 0.14 0.00 0.00 0.16 0.20</td>
<td>0.10 0.98 0.59 0.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Panel B: 10-day ahead prediction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.68 0.88 2.70 -1.16 0.35 0.89</td>
<td>0.00 0.00 0.00 0.00 0.17 0.06</td>
<td>0.38 0.93 0.06 0.18</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.53 0.91 3.49 -1.55 0.49 0.86</td>
<td>0.00 0.01 0.00 0.00 0.00 0.00</td>
<td>0.30 0.94 0.03 0.07</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.42 0.93 -51.00 34.08 0.66 0.84</td>
<td>0.00 0.02 0.00 0.00 0.00 0.00</td>
<td>0.07 0.99 0.65 0.78</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.33 0.95 0.89 0.79 0.83 0.83</td>
<td>0.00 0.03 0.00 0.00 0.00 0.00</td>
<td>0.10 0.99 0.54 0.74</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.30 0.96 2.47 0.30 0.88 0.84</td>
<td>0.00 0.02 0.00 0.00 0.03 0.03</td>
<td>0.34 0.94 0.11 0.22</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: 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., $\text{VS}(t, t+\tau) = \gamma_0 + \gamma_1 \mathbb{E}_t [G(\tau, X_t)/\tau | F_{t-d}] + \text{error}_t$, where $d$ is either 1-day (Panel A) or 10-day (Panel B), and $\mathbb{E}_t [G(\tau, X_t)/\tau | F_{t-d}]$ 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, and includes 3,626 daily observations.
<table>
<thead>
<tr>
<th>Panel A: Risk aversion $\eta = 5$</th>
<th>C.E.</th>
<th>Rate</th>
<th>S.R.</th>
<th>$\mathbb{E}_p[V_T]$</th>
<th>Rebalance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal portfolio</td>
<td>101.06</td>
<td>2.53</td>
<td>1.59</td>
<td>106.08</td>
<td>daily</td>
</tr>
<tr>
<td>Proxy portfolio</td>
<td>100.95</td>
<td>2.47</td>
<td>1.47</td>
<td>106.08</td>
<td>monthly, yearly</td>
</tr>
<tr>
<td>Stock index</td>
<td>90.98</td>
<td>−2.73</td>
<td>1.18</td>
<td>110.42</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Risk aversion $\eta = 1$</th>
<th>C.E.</th>
<th>Rate</th>
<th>S.R.</th>
<th>$\mathbb{E}_p[V_T]$</th>
<th>Rebalance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal portfolio</td>
<td>104.79</td>
<td>4.34</td>
<td>1.44</td>
<td>114.06</td>
<td>daily</td>
</tr>
<tr>
<td>Proxy portfolio</td>
<td>104.42</td>
<td>4.16</td>
<td>1.51</td>
<td>114.42</td>
<td>monthly, yearly</td>
</tr>
<tr>
<td>Stock-bond portfolio</td>
<td>103.52</td>
<td>3.73</td>
<td>1.39</td>
<td>110.99</td>
<td>daily</td>
</tr>
<tr>
<td>Stock index</td>
<td>103.47</td>
<td>3.70</td>
<td>1.18</td>
<td>110.42</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 5: Monte Carlo study of optimal portfolios. Results are based on 10,000 simulated trajectories of bivariate state process $X_t$ (Section 5.3), variance swap rates, and stock index. Optimal portfolio: wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio: wealth is invested as in the optimal portfolio, but positions are rebalanced less frequently, namely stock index and 3-month variance swap positions are rebalanced monthly, 2-year variance swap position is rebalanced yearly. Stock-bond portfolio: for the risk aversion $\eta = 1$, wealth is optimally invested in the risk free bond and stock index. Stock index: initial wealth is invested in the stock index and positions are not rebalanced. Initial wealth is 100. C.E. is the certainty equivalent defined as $e^{-rT}\mathcal{C}$, and $\mathcal{C}$ is such that $u(\mathcal{C}) = \mathbb{E}_p[u(V_T)]$, where $u(V) = V^{1-\eta}/(1-\eta)$ for $\eta \neq 1$ and $u(V) = \log(V)$ for $\eta = 1$, $T$ is the investment horizon of 2 years, $r$ is the risk free rate set to 2%, and $V_T$ is the terminal wealth. Rate is the percentage implied annual rate of return, i.e., $\log(\mathcal{C}/100)/T \times 100$. S.R. is the Sharpe ratio of the corresponding portfolio. $\mathbb{E}_p[V_T]$ is the average terminal wealth, obtained by averaging terminal wealths across sample paths. Panel A is for the risk aversion $\eta = 5$. Panel B is for the risk aversion $\eta = 1$, i.e., the logarithmic utility case.
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 2, 3, 6, 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; Proposition 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 2, 3, 6, 12, 24 months. 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 $X_{1t}$ in Class 1 and $X_{2t}$ in Class 3; Proposition 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 2, 3, 6, 12, 24 months. Variance swap rates are in volatility percentage units, i.e., $\sqrt{\text{VS}(t,T)} \times 100$. Upper graph: variance swap rates with 2-month term (shortest term in our sample). Lower graph: variance swap rates with 2-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: Equity and variance risk premium. Upper graph: equity risk premium defined as \( \frac{(\mathbb{E}[dS_t/S_t|\mathcal{F}_t] - \mathbb{E}_Q[dS_t/S_t|\mathcal{F}_t])/dt}{S_t} \), where \( S_t \) is the stock index. Lower graph: variance risk premium, in units of volatility, defined as \( \frac{(\mathbb{E}[dv_t|\mathcal{F}_t] - \mathbb{E}_Q[dv_t|\mathcal{F}_t])/dt}{\sqrt{v_t}} \), where \( v_t \) is the spot variance. Both risk premiums are derived from the bivariate quadratic model in Section 5.3. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 5: Optimal portfolio. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. The risk aversion is $\eta = 5$. $n_{1t}$ is the optimal fraction of wealth invested in the 3-month variance swap. $n_{2t}$ is the optimal fraction of wealth invested in the 2-year variance swap. Upper graph: optimal portfolio weight in the stock index, $w_t$. Lower graph: optimal portfolio weights in variance swaps, $n_{1t}$ and $n_{2t}$. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 6: Optimal portfolio for log-investor. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. The risk aversion is \( \eta = 1 \). \( n_{1t} \) is the optimal fraction of wealth invested in the 3-month variance swap. \( n_{2t} \) is the optimal fraction of wealth invested in the 2-year variance swap. Upper graph: optimal portfolio weight in the stock index, \( w_t \). Lower graph: optimal portfolio weights in variance swaps, \( n_{1t} \) and \( n_{2t} \). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 7: Myopic and intertemporal hedging demands. Wealth is optimally invested in the risk-free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. The risk aversion is $\eta = 5$. $n_{1t}$ is the optimal fraction of wealth invested in the 3-month variance swap. $n_{2t}$ is the optimal fraction of wealth invested in the 2-year variance swap. Upper graph: myopic component. Lower graph: intertemporal hedging demand. The myopic component and intertemporal hedging demand in the risky assets are in (20). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.
Figure 8: Wealth process. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio is rebalanced less frequently: stock index and 3-month variance swap positions are rebalanced monthly, 2-year variance swap position is rebalanced yearly. The 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 9: Wealth process for log-investor. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio is rebalanced less frequently: stock index and 3-month variance swap positions are rebalanced monthly, 2-year variance swap position is rebalanced yearly. “S&P500-bond portfolio” optimally invests in risk free bond and stock index, and is rebalanced daily. The 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.
References


Revuz, D., and M. Yor, 1994, *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin, Germany, second edn.


ONLINE APPENDIX TO
Quadratic Variance Swap Models

This appendix provides technical derivation and proofs.
A Kolmogorov Backward Equation

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 (2) where $W_t$ is a standard $d$-dimensional Brownian motion under the risk neutral measure $\mathbb{Q}$. The following assumption is obviously met by all quadratic processes in this paper.

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

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

for some finite constant $K$.

**Lemma A.2.** Let $g(x)$ be some $C^2$-function 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

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

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

If $f(\tau, x)$ satisfies the Kolmogorov backward equation

$$
\frac{\partial f(\tau, x)}{\partial \tau} = \sum_{i=1}^m \mu_i(x) \frac{\partial f(\tau, x)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^m \left( \Sigma(x) \Sigma(x)^\top \right)_{ij} \frac{\partial^2 f(\tau, x)}{\partial x_i \partial x_j} \tag{A.3}
$$

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

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

$$
f(T - t, X_t) = \mathbb{E}_Q[g(X_T) \mid \mathcal{F}_t] \quad \text{for all } t \leq T < \infty. \tag{A.4}
$$

**Proof.** Fix some finite $T$. Itô’s formula applied to $M_t = f(T - t, X_t)$ gives

$$
dM_t = D_t dt + \nabla_x f(T - t, X_t) \Sigma(X_t) dW_t
$$
with drift term

\[
D_t = -\frac{\partial f(T-t,X_t)}{\partial T} + \sum_{i=1}^{m} \mu_i(X_t) \frac{\partial f(T-t,X_t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{m} \left( \Sigma(X_t) \Sigma(X_t)^\top \right)_{ij} \frac{\partial^2 f(T-t,X_t)}{\partial x_i \partial x_j}
\]  

(A.5)

which vanishes by assumption. Hence \(M_t\) is a \(\mathbb{Q}\)-local martingale with \(M_T = g(X_T)\). It remains to be shown that \(M_t\) is a true \(\mathbb{Q}\)-martingale. Assumption (A.2) implies

\[
\mathbb{E} \left[ \int_0^T \|\nabla_x f(T-s,X_s) \Sigma(X_s)\|^2 \, ds \right] \leq \mathbb{E} \left[ \int_0^T \|\nabla_x f(T-s,X_s)\|^2 \|\Sigma(X_s)\|^2 \, ds \right] \leq K \left( 1 + \mathbb{E} \left[ \sup_{s \leq T} \|X_s\|^{2p} \right] \right)
\]

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

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

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

\[
\sup_{s \leq t} \|D_s\|^{2p} \leq K_1 \int_0^t \|\mu(X_{u T_n}^T)\|^{2p} \, du \leq K_2 \int_0^t \left( 1 + \sup_{s \leq u} \|X_{s T_n}^T\|^{2p} \right) \, du.
\]

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

\[
\mathbb{E} \left[ \sup_{s \leq t} \|M_s\|^{2p} \right] \leq K_3 \int_0^t \mathbb{E} \left[ \|\Sigma(X_{u T_n}^T)\|^{2p} \right] \, du \leq K_4 \int_0^t \left( 1 + \mathbb{E} \left[ \sup_{s \leq u} \|X_{s T_n}^T\|^{2p} \right] \right) \, du.
\]
Combining these inequalities, we obtain

\[ E \left[ \sup_{s \leq t} \| X_{s \wedge T_n} \|^2 \right] \leq K_5 \left( x^{2p} + t + \int_0^t E \left[ \sup_{s \leq u} \| X_{s \wedge T_n} \|^2 \right] du \right). \]

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

\[ E \left[ \sup_{s \leq T} \| X_{s \wedge T_n} \|^2 \right] \leq K_5 \left( x^{2p} + T + \int_0^T (x^{2p} + u) K_5 e^{K_5(T-u)} du \right). \]

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

B \(X_t\) is Necessarily Quadratic

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

Assumption B.1. The SDE (2) 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 (2) 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 spot variance is given by \(v_t = g(X_t)\) for some \(C^2\)-function \(g(x)\) on \(\mathcal{X}\).

Assumption B.3. The law \(Q = Q_x\) 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 \(VS(t, T) = \frac{1}{T-t} \int_t^T E_Q[v_s | X_t] ds\).

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

Lemma B.4. Suppose \(f(\tau, x)\) is a \(C^{1,2}\)-function on \([0, +\infty) \times \mathcal{X}\) whose \(x\)-gradient satisfies a polynomial growth condition (A.2) for some finite constant \(K = K(T)\) and some \(p \geq 1\), for all
finite $T$. 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 Kolmogorov backward equation (A.3).

Proof. By assumption, $M_t = f(T - t, X_t)$ is a $\mathcal{Q}$-martingale. Hence its drift, given in (A.5), 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 Proposition 3.2.

Proposition B.5. Assume that the variance swap model admits a quadratic term structure. That is, $G(\tau, x)$ in (6) 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 spot variance function is quadratic, $g(x) = \phi + \psi^\top x + x^\top \pi x$, with parameters given by $\phi = d\Phi(0)/d\tau$, $\psi = d\Psi(0)/d\tau$, and $\pi = d\Pi(0)/d\tau$. Moreover, the following holds:

(i) Suppose $\Psi_i(\tau)$ and $\Pi_{ij}(\tau)$, $1 \leq i \leq j \leq m$, are linearly independent functions. Assume the state space $\mathcal{X}$ contains $\{\lambda x \mid x \in O, \lambda \geq 1\}$ for some open set $O$ in $\mathbb{R}^m$. Then the process $X_t$ is quadratic with drift and diffusion functions of the form (3)–(4). The functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ satisfy the linear ordinary differential equations (7).

(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 (3). 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
Integrating the Kolmogorov backward equation (A.3) for $f(\tau, x)$ in $\tau$ leads to

$$\frac{d\Phi(\tau)}{d\tau} - \phi + \left( \frac{d\Psi(\tau)}{d\tau} - \psi \right)^T x + x^T \left( \frac{d\Pi(\tau)}{d\tau} - \pi \right) x$$

$$= \sum_{i=1}^{m} \Psi_i(\tau) \mu_i(x) + \sum_{i,j=1}^{m} \Pi_{ij}(\tau) (\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x))$$

$$= \sum_{i=1}^{m} \Psi_i(\tau) \mu_i(x) + \sum_{i=1}^{m} \Pi_{ii}(\tau) (2\mu_i(x)x_i + c_{ii}(x)) + 2 \sum_{i<j} \Pi_{ij}(\tau) (\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x))$$

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 2. 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 (3). Plugging this in $\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)$ yields (4). Plugging these expressions back in (B.1), and separating the powers of $x$, we arrive at the linear ordinary differential equations (7). This proves part (i). Part (ii) follows using a similar argument.

\[\square\]

## 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

\[
\begin{align*}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b^T \Psi(\tau) + a_1 \Pi_{11}(\tau) + a_2 \Pi_{22}(\tau), & \Phi(0) = 0 \\
\frac{d\Psi(\tau)}{d\tau} &= \begin{pmatrix} \psi \\ 0 \end{pmatrix} + \beta^T \Psi(\tau) + 2\Pi(\tau) b + \begin{pmatrix} \alpha_1 \Pi_{11}(\tau) \\ \alpha_2 \Pi_{22}(\tau) \end{pmatrix}, & \Psi(0) = 0 \\
\frac{d\Pi(\tau)}{d\tau} &= \begin{pmatrix} \pi \\ 0 \\ 0 \end{pmatrix} + \beta^T \Pi(\tau) + \Pi(\tau) \beta + \begin{pmatrix} A_1 \Pi_{11}(\tau) & 0 \\ 0 & A_2 \Pi_{22}(\tau) \end{pmatrix}, & \Pi(0) = 0
\end{align*}
\]

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

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

D Proof of Proposition 3.3

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

\[
\begin{align*}
\hat{b}(\hat{x}) &= b_0 - \beta c + \beta \hat{x} \equiv \hat{b} + \hat{\beta}\hat{x} \\
\hat{a}(\hat{x}) &= a_0 - \alpha c + 2Ac + (\alpha - 2Ac) \hat{x} + A\hat{x}^2 \equiv \hat{a} + \hat{\alpha}\hat{x} + \hat{A}\hat{x}^2.
\end{align*}
\]

The discriminant of \(\hat{a}(\hat{x})\) satisfies \(\hat{D} = \gamma^2 D\). This proves that Classes 1–3 in Proposition 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 $c = \frac{\alpha \gamma}{2A}$ 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 $\hat{b} = (b - \beta \frac{\alpha \gamma}{2A})\gamma$ 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 $c = \frac{\alpha \gamma}{2A}$ yields $\hat{a}(\hat{x}) = A\hat{x}^2$. The factor $\gamma$ can be chosen such that $\hat{b} = (b - \beta \frac{\alpha \gamma}{2A})\gamma$ 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 $d\hat{Z}_t = \beta \hat{Z}_t + \sqrt{A}\hat{Z}_t dW_t$. 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 $c = \frac{\alpha \gamma}{2A}$ and $\gamma = \pm 1/\sqrt{D}$ yields $\hat{a}(\hat{x}) = \hat{x} + A\hat{x}^2$. The sign of $\gamma$ can be chosen such that $\hat{b} = (b - \beta \frac{\alpha \gamma}{2A})\gamma$ 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

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

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 $\tau_c = \inf\{t \geq 0 \mid \hat{X}_t = c\}$, see Karatzas and Shreve (1991, Section V.5.C). Since $\tau_R \uparrow \infty$ for $R \uparrow \infty$, it follows that $\mathbb{P}[\tau_0 = \infty] = 1$ if and only if $p(0) = -\infty$, e.g., Filipović (2009, Exercise 10.12). The latter is equivalent to $2\hat{b} \geq 1$, which proves the claim. If $A = 0$ and $\alpha \neq 0$, we set $\gamma = 1/\alpha$, and chose $c$ such that $\hat{a}$ becomes zero. Note that the conditions on $\hat{b}$ hold necessarily if $X_t$ be well defined, e.g., Filipović (2009, Lemma 10.11) and the arguments above. This completes the proof of Proposition 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 Proposition 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 Proposition 3.3, Class 3, that $X_{1t} = 0$ for some $t \leq 1$ with non-zero 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 Proposition 3.3, implies that $X_{1t} > 0$ for all $t$ whenever $X_{10} > 0$.

F Proof of Proposition 3.5

Let $0 \leq n \leq N$. The $n$th $\mathcal{F}_t$-conditional moment function $f_n(\tau, X_t) = \mathbb{E}_{\mathbb{Q}}[X_{t+\tau}^n | \mathcal{F}_t]$ formally solves the Kolmogorov backward equation

$$
\frac{\partial}{\partial \tau} f_n(\tau, x) = Af_n(\tau, x)
$$

$$
f_n(0, x) = x^n
$$

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

$$\mathcal{A} x^k = k(k-1) \frac{a}{2} x^{k-2} + k \left( b + (k-1) \frac{\alpha}{2} \right) x^{k-1} + k \left( \beta + (k-1) \frac{A}{2} \right) x^k$$
and matching coefficients in $x$, we obtain the $N + 1$ linear systems of $N + 1$ ordinary differential equations

$$
\frac{d}{d\tau} \begin{pmatrix}
M_{0n}(\tau) \\
M_{1n}(\tau) \\
M_{2n}(\tau) \\
\vdots \\
M_{Nn}(\tau)
\end{pmatrix} = \begin{pmatrix}
0 & b & 2\frac{3}{2} & 0 & \cdots & 0 \\
0 & \beta & 2(b + \frac{3}{2}) & 3\cdot 2\frac{3}{2} & 0 & \vdots \\
0 & 0 & 2(\beta + \frac{3}{2}) & 3(b + 2\frac{3}{2}) & \ddots & 0 \\
0 & 0 & 0 & 3(\beta + 2\frac{3}{2}) & \ddots & N(N-1)\frac{3}{2} \\
0 & \cdots & 0 & \ddots & \ddots & N(b + (N-1)\frac{3}{2}) \\
0 & \cdots & 0 & \ddots & \ddots & N(\beta + (N-1)\frac{3}{2})
\end{pmatrix} \begin{pmatrix}
M_{0n}(\tau) \\
M_{1n}(\tau) \\
M_{2n}(\tau) \\
\vdots \\
M_{Nn}(\tau)
\end{pmatrix}
$$

along with the initial condition

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

(F.3)

In matrix notation, denote by $B$ the $(N + 1) \times (N + 1)$ matrix in (F.2), the system (F.2)–(F.3) reads

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

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

### G Proof of Theorem 5.2

We first list the technical assumptions that will enable us to prove Theorem 5.2.
Assumption G.1. The Radon–Nikodym density (19) is integrable in the following sense

\[ \mathbb{E}_Q \left[ \left( \frac{dQ}{dP} |_{\mathcal{F}_T} \right)^{-\frac{1}{\eta}} \right] < \infty. \]

Assumption G.2. The exponential \( Q \)-local martingale

\[ D_t = \exp \left( \frac{1}{\eta} \int_0^t \Lambda(X_s) \top dW_s - \frac{1}{2\eta^2} \int_0^t \|\Lambda(X_s)\|^2 ds \right), \quad t \in [0,T] \quad \text{(G.1)} \]

is a true martingale. That is, \( \mathbb{E}[D_T] = 1 \), and we can define the auxiliary equivalent probability measure \( \hat{Q} \sim Q \) on \( \mathcal{F}_T \) by

\[ \frac{d\hat{Q}}{dQ} |_{\mathcal{F}_T} = D_T. \quad \text{(G.2)} \]

Assumption G.3. The state process \( X_t \) is a well-defined diffusion under \( \hat{Q} \), and the function

\[ h(\tau, x) = \log \mathbb{E}_{\hat{Q}} \left[ \exp \left( \frac{1}{2\eta} \left( \frac{1}{\eta} - 1 \right) \int_0^\tau \|\Lambda(X_s)\|^2 ds \right) \mid X_0 = x \right] \quad \text{(G.3)} \]

is of class \( C^{1,2} \) on \([0,T] \times \mathcal{X}\). A partial differential equation (PDE) for \( H(\tau, x) = e^{h(\tau, x)} \) is provided in (I.3).

Under Assumptions 5.1 and G.1 it is well known (e.g., Karatzas and Shreve (1998, Theorem 3.7.6)) that the optimal terminal wealth is given by

\[ V^*_T = (u')^{-1} \left( \lambda e^{-rT} \frac{dQ}{dP} |_{\mathcal{F}_T} \right) \quad \text{(G.4)} \]

for some Lagrangian \( \lambda = \lambda(V_0) \) such that \( \mathbb{E}_Q [e^{-rT} V^*_T] = V_0 \). Notice that \( (u')^{-1}(z) = z^{-\frac{1}{\eta}} \).

Straightforward manipulations together with (19) give

\[ e^{-rT} V^*_T = \lambda^{-\frac{1}{\eta}} e^{\left( \frac{1}{\eta} - 1 \right) rT} \exp \left( \frac{1}{\eta} \int_0^T \Lambda(X_s) \top dW_s - \frac{1}{2\eta} \int_0^T \|\Lambda(X_s)\|^2 ds \right) \]

\[ = \lambda^{-\frac{1}{\eta}} e^{\left( \frac{1}{\eta} - 1 \right) rT} D_T \exp \left( \frac{1}{2\eta} \left( \frac{1}{\eta} - 1 \right) \int_0^T \|\Lambda(X_s)\|^2 ds \right) \quad \text{(G.5)} \]

for the exponential \( Q \)-martingale \( D_t \) defined in (G.1). Since the discounted optimal wealth process

\[ 40 \text{In view of (G.5), this is automatically satisfied for relative risk aversion } \eta \geq 1. \]
Since \( X_t \) is a time-homogeneous diffusion under \( \hat{Q} \), we have

\[
\mathbb{E}_{\hat{Q}}\left[ e^{-rt} V_t^* \right] = \mathbb{E}_{\hat{Q}}\left[ e^{-rT} V_T^* \right] = \lambda - \frac{1}{\eta} e^{\left( \frac{1}{\eta} - 1 \right) rt} \exp \left( \frac{1}{2} \eta \left( \frac{1}{\eta} - 1 \right) \int_0^T \| \Lambda(X_s) \|^2 ds \right).
\]

From the factorization (1), it follows that \( \mathbb{E}_{\hat{Q}}\left[ \exp \left( \frac{1}{2} \eta \left( \frac{1}{\eta} - 1 \right) \int_0^T \| \Lambda(X_s) \|^2 ds \right) \right] = e^{h(T-t,X_t)} \).

\[\text{Proof of Theorem 2.2} \]

Hence \( G(t,x) \) is injective if and only if ker \( (\partial_t + \partial_x) \Sigma(X_t) \cap \ker \sqrt{ \mathcal{G}(X_t) R(X_t)^T } = \{0 \} \). In view of Assumption 5.1, on one hand this implies that

\[ \ker \mathbf{Q}_1 \cap \ker \mathbf{Q}_2 = \{0 \} \]

on \( \mathcal{F}_t \), which yields the linear system (14). By Assumption 2.1, \( \mathcal{L}_1 \mathcal{L}_2 \) yields the linear system (14). Theorem 2.2 is proved.

\[\text{Proof of Corollary 2.3} \]

From the factorization (17), it follows that \( G(t,x) \) is injective if and only if ker \( \partial_t + \partial_x \Sigma(X_t) \cap \ker \sqrt{ \mathcal{G}(X_t) R(X_t)^T } = \{0 \} \). In view of Assumption 5.1, on one hand this implies that

\[ \ker \mathbf{Q}_1 \cap \ker \mathbf{Q}_2 = \{0 \} \]

on \( \mathcal{F}_t \), which is justified by Assumption 2.1. Hence theorem 2.2 is proved. Hence there exists a (possibly non-unique) solution \( \mu \) to the factorization (17). Theorem 2.2 yields the linear system (14). By Assumption 2.1, \( \mathcal{L}_1 \mathcal{L}_2 \) yields the linear system (14). Theorem 2.2 is proved.

\[\text{Proof of Theorem 2.2} \]

Proposed

Therefore we obtain via the formula (G.3) the following equality:

\[
\mathbb{E}_Q \left[ e^{\left( \frac{1}{2} \eta \left( \frac{1}{\eta} - 1 \right) \int_0^T \| \Lambda(X_s) \|^2 ds \right)} \right] = e^{h(T-t,X_t)}.
\]

\[\text{Proof of Theorem 5.2} \]

Proposed

Therefore we obtain via the formula (G.3) the following equality:

\[
\mathbb{E}_Q \left[ e^{\left( \frac{1}{2} \eta \left( \frac{1}{\eta} - 1 \right) \int_0^T \| \Lambda(X_s) \|^2 ds \right)} \right] = e^{h(T-t,X_t)}.
\]
Whence \( \dim \ker \Sigma(X_t) \leq 1 \ dt \otimes dQ \)-a.s. in particular, and we conclude that \( d \leq n+1 \) and \( d \leq m+1 \). On the other hand it also implies that the maturity date functions \( T^*_t(t) \) have to be mutually different for all \( t \).

Now recall the elementary fact from linear algebra that the kernel of \( \Sigma(X_t) \) is the orthogonal complement of the image of \( \Sigma(X_t)^\top \) in \( \mathbb{R}^d \). That is, \( \mathbb{R}^d = \ker \Sigma(X_t) \oplus \text{im} \Sigma(X_t)^\top \). Taking account of the factorization (17), we can rewrite the left hand side of (20) as \( \Sigma(X_t)^\top D(t, X_t)^\top n_t + w_t \sqrt{g(X_t)} R(X_t) \). Since \( d > m \), the \( m \times d \)-matrix \( \Sigma(X_t) \) cannot be injective, and there exists a non-zero vector \( v = v(X_t) \) in \( \ker \Sigma(X_t) \). In view of (H.1) it satisfies \( \sqrt{g(X_t)} R(X_t)^\top v \neq 0 \ dt \otimes dQ \)-a.s.

Projecting both sides of (20) onto \( v \), we then obtain

\[
 w_t \sqrt{g(X_t)} R(X_t)^\top v = \frac{1}{\eta} \Lambda(X_t)^\top v, \quad dt \otimes dQ \text{-a.s.} \tag{H.2}
\]

Hence the solution \( w^*_t = w_t \) of (20) is fully determined by the myopic term and does not depend on the choice of the variance swaps. This proves Corollary 5.3.

I Computation of the Intertemporal Hedging Demand

We now discuss the computation of \( \nabla_x h(\tau, x) \) in (20). In view of (G.3), if the market price of risk has a constant norm, \( \|\Lambda(x)\|^2 \equiv c \), then \( h(\tau, x) \) does not depend on \( x \), which implies \( \nabla_x h(\tau, x) = 0 \), and thus there is no intertemporal hedging demand at all. The same would obviously hold true for the myopic logarithmic utility case \( \eta = 1 \). In general, \( \nabla_x h(\tau, x) \) needs to be computed numerically, e.g., via Taylor expansion. Suppose, for example, that

\[
 \|\Lambda(x)\|^2 = c + \epsilon^\top P(x) \tag{I.1}
\]
for some constant $c$, some $\mathbb{R}^k$-valued function $P(x)$, and some parameter $\epsilon \in \mathbb{R}^k$ with small norm.

The first order expansion of $\nabla_x h(\tau, x) = \nabla_x h(\tau, x, \epsilon)$ around $\epsilon = 0$ is\footnote{We omit the effect of $\epsilon$ on $\hat{Q}$ given by (G.2). This is in particular justified in the bivariate quadratic model, where, in view of (21) and (22), the $\hat{Q}$-law of $X_t$ is invariant with respect to constant shifts of $\|\Lambda(x)\|^2$.}

$$
\nabla_x h(\tau, x, \epsilon) = \nabla_x h(\tau, x, 0) + \nabla_x \left( \nabla_x h(\tau, x, \epsilon) |_{\epsilon = 0} \right) + o(\|\epsilon\|)
$$

(1.2)

If the diffusion $X_t$ is quadratic under $\hat{Q}$ and $P(x)$ are polynomials in $x$, then the conditional moments in the right hand side are available in closed form. Indeed, closed form expressions for $\nabla_x h(\tau, x, \epsilon)$ are available for Taylor expansions of arbitrary order in $\epsilon$.

We next provide a PDE for the function $H(\tau, x) = e^{h(\tau, x)}$, which could serve as an alternative procedure to compute $\nabla_x h(\tau, x)$. Consider the auxiliary measure $\hat{Q}$ defined in (G.2), and denote the respective Girsanov transformed $\hat{Q}$-Brownian motion by $d\hat{W}_t = dW_t - \frac{1}{\eta} \Lambda(X_t) dt$. Then the $\hat{Q}$-dynamics of $X_t$ reads $dX_t = \left( \mu(X_t) + \frac{1}{\eta} \Sigma(X_t) \Lambda(X_t) \right) dt + \Sigma(X_t) d\hat{W}_t$. The Feynman–Kac theorem thus yields the following linear PDE for $H(\tau, x)$:

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

$$
+ \frac{1}{2\eta} \left( \frac{1}{\eta} - 1 \right) \|\Lambda(x)\|^2 H(\tau, x)
$$

(1.3)

$$
H(0, x) = 1.
$$

### J 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. It follows from Cheridito, Filipović, and Kimmel (2007) that the processes on the right hand side of (19) and (G.1) are true martingales for $t \in [0, T]$, and hence define equivalent probability measures $P \sim Q$ and $\hat{Q} \sim Q$ on $\mathcal{F}_T$. This asserts validity of Assumption G.2. Next, we note that the function $g(x)$ as well as the determinant of the $2 \times 2$ matrix $D(t, x)$ are non-zero polynomials in $x$, 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 \( 1_{\{\ell(t,X_t)=0\}} \nabla_x \ell(t,X_t) \Sigma(X_t) \Sigma(X_t)^\top \nabla_x \ell(t,X_t) = 0, dt \otimes d\mathbb{Q}-\text{a.s.}; \)

see Revuz and Yor (1994, Corollary (1.6), Chap. VI) and Filipović (2001, Lemma 3.3.1). Since \( \Sigma(X_t) \Sigma(X_t)^\top \) is positive definite \( dt \otimes d\mathbb{Q}-\text{a.s.}, \) we infer that \( \ell(t,X_t) \neq 0 \) if \( \nabla_x \ell(t,X_t) \neq 0, dt \otimes d\mathbb{Q}-\text{a.s.} \)

Applying this to \( g(x) \) and the determinant of \( \mathcal{D}(t,x) \), we find that \( g(X_t) \neq 0 \) and \( \mathcal{D}(t,X_t) \) is injective, \( dt \otimes d\mathbb{Q}-\text{a.s.} \)

Whence validity of Assumption 5.1 follows. Assumption G.1 holds since we shall consider relative risk aversion \( \eta \geq 1 \) only. Finally, the \( C^{1,2} \)-regularity of \( h(\tau,x) \) and validity of the PDE (I.3) follows, e.g., from Heath and Schweizer (2000). This asserts Assumption G.3.

We finally provide an approximation of \( \nabla_x h(\tau,x) \) in (28). Following up on Appendix I, we note that the specification (24) is of the form (I.1) with \( c = \kappa \varphi_0, \mathbf{P}(x) = (\kappa x_1, \kappa x_2)^\top, \) and \( \epsilon = (\psi_0, \pi_0)^\top. \)

The first order Taylor expansion (I.2) then reads

\[
\nabla_x h(\tau,x) \approx \frac{1}{2\eta} \left( \frac{1}{\eta} - 1 \right) \nabla_x \int_0^\tau \mathbb{E}_{\hat{\mathbb{Q}}} \left[ \kappa \psi_0 X_{1s} + \kappa \pi_0 X_{2s} \mid X_0 = x \right] ds, \tag{J.1}
\]

which is available in closed form.