Empirical Asset Pricing with Nonlinear Risk Premia*

Aleksandar Mijatović† Paul Schneider‡

November 4, 2009

Abstract

In this paper we introduce a simple continuous-time asset pricing framework, based on general multi-dimensional diffusion processes, that combines semi-analytic pricing with a nonlinear specification for the market price of risk. Our framework guarantees existence of weak solutions of the nonlinear SDEs under the physical measure, thus allowing to work with nonlinear models for the real world dynamics not considered in the literature so far. It emerges that the additional flexibility in the time series modelling is econometrically relevant: a nonlinear stochastic volatility diffusion model for the joint time series of the S&P 100 and the VXO implied volatility index data shows superior forecasting power over the standard specifications for implied and realized variance forecasting.

1 Introduction

Most financial time series exhibit rapid fluctuations while being extremely persistent at the same time. Violent fluctuations are often identified as jumps caused by events such as central bank meetings or rating announcements. The economic intuition suggests that for example interest rates should be stationary. However unit-root tests often imply that interest rates are integrated and therefore exhibit extreme persistence. Ideally a model should be able to accommodate both extremes while maintaining compatibility with economic theory: random walk like behaviour in a certain region, and reversion towards a mean outside it. At first glance establishing the existence of such a model under the real world measure appears to be very difficult. A diffusion process with these characteristics would clearly need to exhibit a highly nonlinear drift under the physical measure, which implies that global Lipschitz and growth conditions,

*We are grateful to Yacine Aït-Sahalia, Valentina Corradi, and Damir Filipović for insightful comments.

†Department of Mathematics, Imperial College London, Huxley Building, 180 Queen's Gate, London SW7 2AZ, United Kingdom, a.mijatovic@imperial.ac.uk

‡Finance Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom, paul.schneider@wbs.ac.uk

typically required for the existence of a solution to a multi-dimensional SDE, are not satisfied. In a univariate diffusion setting Aït-Sahalia (1996) applies a more general method, only available in dimension one, to ascertain the existence of a model that exhibits the desired characteristics. The reason for the econometric success of his model lies in the nonlinearity of the drift. Two main obstacles to a wide applicability of such models remain. The first is the lack of closed-form, or at least semi-analytic, solutions for the prices of contingent claims within the nonlinear framework. The second is a lack of tools for proving the existence of solutions to the stochastic differential equations used when attempting to introduce nonlinearity in a multivariate setting.

Employing econometrically inconspicuous dampening functions we introduce a simple multivariate diffusion framework which exploits the existence of a solution of an SDE under a risk-neutral probability measure and guarantees the existence of a weak solution of a nonlinear SDE under the real world probability measure. From the econometric point of view our framework extends the affine approach from Cheridito et al. (2007) yielding substantially enriched dynamics. The most obvious application is a state variable formulation that entails (semi-)analytic pricing under the risk-neutral measure, which leaves flexibility for the dynamics under the physical measure similar to that of the discrete-time approach considered in Dai et al. (2006) and Bertholon et al. (2008). Recent advances in estimating the parameters of nonlinear diffusions such as the algorithms introduced in Aït-Sahalia (2001), Beskos et al. (2006) and Mijatović and Schneider (2009) ensure that reliable parameter inference can be made without explicit formulae for transition densities. An empirical application based on the joint time series of the S&P 100 and the VXO implied volatility indices reveals that our framework offers statistically significant advantages out of sample over extant model specifications in predicting implied as well as realized variance over several forecasting horizons. Furthermore we find that the size and sign of the variance risk premia implied by our model coincide with the model-independent results in Carr and Wu (2008).

The paper is organized as follows. In Section 2 we describe the main theoretical construction (see Theorem 1) for the framework we consider. Section 3 describes the nonlinear stochastic volatility model, its likelihood function and the estimation algorithm which is used to find the parameter values implied by the time series of the S&P 100 and the VXO index data. The empirical results are discussed in Subsection 3.5 and can be found in the Appendix. Section 4 concludes the paper.

2 The modelling framework

In this section we describe the theoretical basis for the modelling framework used in the present paper. As mentioned in the introduction, Theorem 1 allows us to define our model under the pricing measure $\mathbb{Q}$ and perform Girsanov's measure change to obtain any desired model under the physical measure $\mathbb{P}$ in a wide class of Itô processes where the state vector satisfies a possibly nonlinear SDE. Theorem 1 also provides a weak solution of this SDE. The central building block of our approach is provided by the following very simple observation.

Theorem 1. *Fix a time horizon $T > 0$ and suppose $X = (X_t)_{t\in[0,T]}$ is an Itô process with state space $\mathfrak{D} \subseteq \mathbb{R}^n$ that satisfies the following SDE under the pricing measure $\mathbb{Q}$*

$$dX_t = \mu^{\mathbb{Q}}(X_t)\, dt + \Sigma(X_t)\, dW_t^{\mathbb{Q}}, \qquad X_0 = x_0 \in \mathfrak{D}, \tag{1}$$

where the drift is given by the function $\mu^{\mathbb{Q}} : \mathfrak{D} \to \mathbb{R}^n$ and $W^{\mathbb{Q}} = (W_t^{\mathbb{Q}})_{t\in[0,T]}$ is a standard n-dimensional Brownian motion under $\mathbb{Q}$. We further assume that the volatility function $\Sigma : \mathfrak{D} \to \mathbb{R}^{n\times n}$ satisfies $|\det \Sigma(x)| > 0$ for all $x \in \mathfrak{D}$. Let $f : \mathfrak{D} \to \mathbb{R}^n$ be any measurable function with coordinates $f_j : \mathfrak{D} \to \mathbb{R}$, $j = 1, \ldots, n$, and define the function $D : \mathfrak{D} \to \mathbb{R}_+$ by the formula

$$D(x) := \exp\left[-\frac{c}{|\det \Sigma(x)|} - c\sum_{j=1}^{n} |f_j(x)|\right]$$

where c is some positive constant. Then the function $\Lambda : \mathfrak{D} \to \mathbb{R}_+$, defined by the formula

$$\Lambda(x) := D(x)\Sigma^{-1}(x) f(x),$$

is bounded and the process $\eta = (\eta_t)_{t\in[0,T]}$ given by

$$\eta_t = \exp\left(\int_0^t \Lambda(X_s) dW_s^{\mathbb{Q}} - \frac{1}{2}\int_0^t \Lambda(X_s)^\top \Lambda(X_s) ds\right), \quad t \le T,$$

is a $\mathbb{Q}$-martingale. Then the dynamics of $X = (X_t)_{t\in[0,T]}$ under the real world measure $\mathbb{P}$, which is defined via the Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{Q}} = \eta_T$, are given by

$$dX_t = (D(X_t) f(X_t) + \mu^{\mathbb{Q}}(X_t))\, dt + \Sigma(X_t)\, dW_t^{\mathbb{P}}, \qquad X_0 = x_0, \tag{2}$$

where $W^{\mathbb{P}} = (W_t^{\mathbb{P}})_{t\in[0,T]}$ is a standard n-dimensional Brownian motion under the measure $\mathbb{P}$, defined by $W_t^{\mathbb{P}} := W_t^{\mathbb{Q}} - \int_0^t \Lambda(X_s) ds$.

The proof of Theorem 1 follows by construction since the random variable $\Lambda(X_t)$ is bounded uniformly in $t \in [0, T]$. Therefore the Novikov criterion (see Proposition 1.15 in Chapter VIII of Revuz and Yor (1999)) applies and the density process η is a true martingale under the pricing measure $\mathbb{Q}$. The other statements in Theorem 1 follow from Girsanov's theorem (see Theorems 1.4 and 1.7 in Chapter VIII of Revuz and Yor (1999)).

The sole purpose of the dampening function D in Theorem 1 is to ensure the existence of the real world probability measure $\mathbb{P}$, which is equivalent to the pricing measure $\mathbb{Q}$. Note that the positive constant c in the function D can be made arbitrarily small. In the case the volatility function Σ and the drift function f are continuous in the state variable $x \in \mathfrak{D}$, the dampening factor D equals one in a finite precision environment (i.e. a computer) on an arbitrarily large compact subset of the domain $\mathfrak{D}$. As a consequence we have a large amount of freedom when specifying the drift function $f + \mu^{\mathbb{Q}} \approx \mu^{\mathbb{P}}$ that can achieve the desired drift behaviour of the model under the real world measure $\mathbb{P}$. The key observation here is that the constant c in the function D does not need to be estimated. It is enough to know that it exists. This by Theorem 1 implies that the solution of SDE (2) also exists and that the corresponding process behaves in the desired way under the real world measure.

It remains to specify a flexible model under the pricing measure $\mathbb{Q}$. We should stress here that the only assumption in Theorem 1 on the process X under the pricing measure $\mathbb{Q}$ is that it exists and satisfies SDE (1) in the theorem. Therefore the specification of the measure $\mathbb{Q}$ is in practice informed by the analytical tractability of the model in terms of the pricing of derivatives. A common choice in the multivariate diffusion setting are affine processes. The existence of this class of models is established in Duffie et al. (2003) and the algorithms for the pricing of contingent claims, which rest on the extended transform methods, are developed in Duffie et al. (2000). In the application discussed in this paper (see Section 3) we shall deviate slightly from the affine class and consider a stochastic volatility model based on a GARCH diffusion, which is in the class of polynomial models. The existence of the process is not difficult to prove and will be established in Section 3. In this model it is possible to compute analytically an approximation for the implied volatility in terms of the model parameters under $\mathbb{Q}$. This feature is crucial because the goal is to estimate the risk-neutral and the real world parameters simultaneously. We conclude this section with a simple affine example that illustrates the application of Theorem 1.

Example: Consider the following univariate short rate model

$$dr_t = (a^{\mathbb{Q}} - b^{\mathbb{Q}} r_t)\, dt + \sigma\sqrt{r_t} dW_t^{\mathbb{Q}},$$

with state space $\mathfrak{D} = (0, \infty)$ and $2a^{\mathbb{Q}} > \sigma^2$. Let

$$f(r) = a^{\mathbb{P}} - a^{\mathbb{Q}} - (b^{\mathbb{P}} - b^{\mathbb{Q}})r \quad \text{and} \quad D(r) = e^{-c\left(r + \frac{1}{\sqrt{r}}\right)}.$$

Under the real world probability measure $\mathbb{P}$ the process satisfies the SDE

$$dr_t = \left\{ a^{\mathbb{Q}} - b^{\mathbb{Q}} r_t + e^{-c\left(r_t + \frac{1}{\sqrt{r_t}}\right)} (a^{\mathbb{P}} - a^{\mathbb{Q}} - (b^{\mathbb{P}} - b^{\mathbb{Q}}) r_t) \right\} dt + \sigma \sqrt{r_t} dW_t^{\mathbb{P}}.$$

Since the domain $\mathfrak{D}$ of the process $(r_t)_{t \in [0,T]}$ is the positive real line, we can choose the constant c small enough so that for numerical purposes such as estimation we can assume that the dynamics of the process is given by

$$dr_t = \left(a^{\mathbb{P}} - b^{\mathbb{P}} r_t\right) dt + \sigma \sqrt{r_t} dW_t^{\mathbb{P}}.$$

3 Application

In this section we are going to apply Theorem 1 in order to estimate an affine (linear) and a nonlinear model on the joint time series of the S&P 100 and VXO implied volatility index. We first describe the data set and the two models that will be used for prediction and then discuss in some detail the expected maximum likelihood (EML) estimation algorithm, which is used for parameter inference of the nonlinear diffusions. Finally we perform a statistical test given in Clark and West (2007) on the estimated models with respect to their forecasting power.

3.1 Data

Models are estimated using daily S&P 100 log prices and daily VXO implied volatilities. The VXO index is defined in terms of the current value of the expected realised variance of S&P 100 over a period of one month. The CBOE computes the value of VXO using a carfully designed portfolio of exchange traded call and put options on the S&P 100 that expire in one month's time. The algorithm used by CBOE enables them to obtain a time series of model independent implied volatility. Figure 1 shows the trajectory of the VXO implied volatility index published by the CBOE and the logarithm of the S&P 100 index.

We partition our data set into two non-overlapping subsets. The *in-sample* period ranges from 2 January 1990 until 31 December 1999 and the *out-of-sample* period lasts from 3 January 2000 until 29 December 2006.

3.2 S&P 100 stochastic volatility model

We choose a stochastic GARCH diffusion variance model for the joint times series of the logarithm of the S&P 100 prices and instantaneous variance, which evolves under the pricing measure $\mathbb{Q}$ according to the

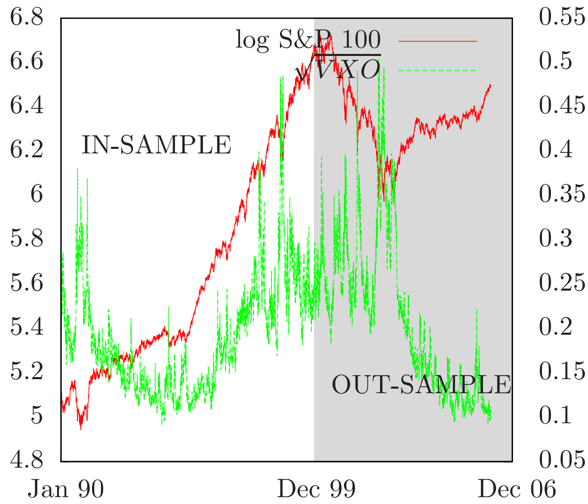


Figure 1: **Log of S&P 100 index and VXO:** The figure shows the evolution of the logarithm of the S&P 100 index (left y-axis) and of the implied volatility index VXO (right y-axis). The sample is comprised of an in-sample period (shaded in white) and an out-sample period (shaded in grey).

SDE

$$dX_t = (r - \frac{1}{2}V_t)\, dt + \rho\sqrt{V_t} dW_t^{V\mathbb{Q}} + \sqrt{1-\rho^2}\sqrt{V_t} dW_t^{X\mathbb{Q}}, \tag{3}$$

$$dV_t = (b_0^{\mathbb{Q}} + b_1^{\mathbb{Q}}\, V_t)dt + \sigma V_t\, dW_t^{V\mathbb{Q}}, \tag{4}$$

where $W^{V\mathbb{Q}} = (W_t^{V\mathbb{Q}})_{t\in[0,T]}$ and $W^{X\mathbb{Q}} = (W_t^{X\mathbb{Q}})_{t\in[0,T]}$ are two independent standard Brownian motions. Note that the risk-neutral drift $\mu^{\mathbb{Q}}$ from Theorem 1 can be expressed as

$$\mu^{\mathbb{Q}}(V_t) = \begin{pmatrix} r - \frac{1}{2}V_t \\ b_0^{\mathbb{Q}} + b_1^{\mathbb{Q}}\, V_t \end{pmatrix}.$$

It is well known that the SDE in (4) has a solution for all values of $b_0^{\mathbb{Q}}$, $b_1^{\mathbb{Q}}$ and σ. Assume that

$$b_0^{\mathbb{Q}} > 0, \tag{5}$$

and note that the comparison theorem for the solutions of SDEs (see Proposition 5.2.18 in Karatzas and Shreve (1991)), applied to $V = (V_t)_{t\in[0,T]}$ and the geometric Brownian motion that solves (4) when $b_0^{\mathbb{Q}} = 0$, implies that the process V does not leave the interval $(0, \infty)$ in finite time. It therefore follows that under condition (5) we can define the process $X = (X_t)_{t\in[0,T]}$ as a stochastic integral given by (3). This argument shows that the process $(X, V)^\top$ with state space $\mathfrak{D} = \mathbb{R} \times (0, \infty)$ exists under the pricing measure $\mathbb{Q}$ and that it follows SDE (3)–(4). It is shown in Forman and Sørensen (2008) that if in addition we have $b_1^{\mathbb{Q}} < 0$, then the variance process V is ergodic.

The power in the volatility function of SDE (4) (i.e. the CEV power) for the GARCH diffusion V is equal to one. This is a pragmatic and parsimonious educated guess between the CEV powers of 0.65 in Aït-Sahalia and Kimmel (2007) and 1.33/1.17 in Jones (2003) for similar data sets. More sophisticated volatility-of-volatility functions are also possible, e.g. $(\beta_0 + \beta_1 V_t + \beta_2 V_t^{\beta_3})^{1/2}$, but the existence of solutions of such SDEs is more difficult to establish. Here we use the simple GARCH diffusion process (see Nelson (2002) for this terminology) because our focus in this paper is on the nonlinear drift specification.

Under the physical measure $\mathbb{P}$ we shall consider a linear and a nonlinear drift specifications. The former $\mathbb{P}$-drift is linear in the state variables (LN model) and is given by the formula

$$\mu_{\mathrm{LN}}^{\mathbb{P}}(V_t) := \begin{pmatrix} a_0 + a_1 V_t \\ b_0^{\mathbb{Q}} + b_1 V_t \end{pmatrix}. \tag{6}$$

This drift corresponds to a usual market price of risk assumption as stated in Jones (2003) and in Aït-Sahalia and Kimmel (2007). In the language of Section 2 the drift $\mu^{\mathbb{P}}$ can be expressed using the function f given by

$$f_{\mathrm{LN}}(V_t) = \begin{pmatrix} a_0 - r + (a_1 + \frac{1}{2})V_t \\ (b_1 - b_1^{\mathbb{Q}})\, V_t \end{pmatrix}. \tag{7}$$

The linear model LN will serve as a benchmark for econometric relevance of the nonlinear model (NL model) whose drift under the physical measure $\mathbb{P}$ is given by

$$\mu_{\mathrm{NL}}^{\mathbb{P}}(V_t) := \begin{pmatrix} a_0 + a_1 V_t \\ b_0 + b_1 V_t + b_2 V_t^2 + b_3/V_t \end{pmatrix}. \tag{8}$$

The corresponding function f from Section 2 is given as

$$f_{\mathrm{NL}}(V_t) = \begin{pmatrix} a_0 - r + (a_1 + \frac{1}{2})V_t \\ b_0 - b_0^{\mathbb{Q}} + (b_1 - b_1^{\mathbb{Q}})\, V_t + b_2 V_t^2 + b_3 \,/\, V_t \end{pmatrix}. \tag{9}$$

As described in Section 2 the dampening functions $D_{\mathrm{NL}}, D_{\mathrm{LN}}$ can be made arbitrarily close to one through the choice of the constant c. Hence for numerical purposes and econometric implementation it suffices to work directly with the drifts $\mu_{\mathrm{NL}}^{\mathbb{P}}$ and $\mu_{\mathrm{LN}}^{\mathbb{P}}$ given in (6) and (8) respectively.

For implementation it is convenient to consider the process $Y = (Y_t)_{t \in [0,T]}$, given by $Y_t := \gamma(V_t)$, where

	θ^{σ}	$\theta^{\mathbb{Q}}$	$\theta^{X\mathbb{P}}$	$\theta^{V\mathbb{P}}$
Linear Spec. (6)	$\sigma, \rho, b_0^{\mathbb{Q}}$	$b_1^{\mathbb{Q}}$	a_0, a_1	b_1
Nonlinear Spec. (8)	σ, ρ	$b_0^{\mathbb{Q}}, b_1^{\mathbb{Q}}$	a_0, a_1	b_0, b_1, b_2, b_3

Table 1: **Parameter sets for the linear and the nonlinear model:** the table displays the partition of the parameter vector θ into the following groups: θ^{σ} (the parameters that influence the dynamics under both the risk-neutral measure $\mathbb{Q}$ as well as the physical measure $\mathbb{P}$); $\theta^{\mathbb{Q}}$ (parameters that influences the only the risk-neutral dynamics); $\theta^{X\mathbb{P}} \cup \theta^{V\mathbb{P}}$ (parameters that appear only under the physical measure $\mathbb{P}$).

the transformation $\gamma : (0, \infty) \to \mathbb{R}$ of the variance process is defined by the formula

$$\gamma(v) = \frac{\log v}{\sigma}. \tag{10}$$

The evolution of the process Y under the physical measure $\mathbb{P}$ is given by

$$dY_t = \left\{ \left(b_0^{\mathbb{Q}} + b_1 V_t \right) \frac{1}{\sigma V_t} - \frac{\sigma}{2} \right\} dt + dW_V^{\mathbb{P}}(t), \tag{11}$$

in linear model (6) and by

$$dY_t = \left\{ \left(b_0 + b_1 V_k + b_2 V_t^2 + \frac{b_3}{V_k} \right) \frac{1}{\sigma V_k} - \frac{\sigma}{2} \right\} dt + dW_V^{\mathbb{P}}(t), \tag{12}$$

in nonlinear model (8).

For estimation purposes we partition the parameter vector θ into four classes. The first class θ^{σ} contains the parameters that influence the dynamics under both the physical measure $\mathbb{P}$ and the pricing measure $\mathbb{Q}$. The second class $\theta^{\mathbb{Q}}$ contains the parameters that arise only under the pricing measure $\mathbb{Q}$. The third set $\theta^{X\mathbb{P}}$ contains the parameters that influence the dynamics of the process X only under the physical measure $\mathbb{P}$ and the fourth class $\theta^{V\mathbb{P}}$ contains the parameters that arise only under the measure $\mathbb{P}$ in the SDE for the variance process V. It is clear that we can express $\theta = \theta^{\sigma} \cup \theta^{\mathbb{Q}} \cup \theta^{X\mathbb{P}} \cup \theta^{V\mathbb{P}}$ and that these four classes are pairwise disjoint.

3.3 Likelihood function

The *instantaneous* stochastic variance is a latent variable even though a time series of the *implied* variance is available through the VXO index. Note that the drift of the variance V in our model, given in SDE (4) under the pricing measure $\mathbb{Q}$, is affine. Therefore the current price of the variance swap is linear in the current value of the variance V_t in our model as the following simple calculation, based on Fubini's theorem, demonstrates

$$\frac{1}{\Delta}\,\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\Delta} V_s\,ds\right] = A(\theta^{\mathbb{Q}},\Delta) + B(\theta^{\mathbb{Q}},\Delta)\,V_t, \qquad \Delta > 0, \tag{13}$$

where the coefficients $A(\theta^{\mathbb{Q}},\Delta)$ and $B(\theta^{\mathbb{Q}},\Delta)$ are given by

$$B(\theta^{\mathbb{Q}},\Delta) = \frac{1}{b_1^{\mathbb{Q}}\Delta}\left(\exp(b_1^{\mathbb{Q}}\Delta) - 1\right), \quad A(\theta^{\mathbb{Q}},\Delta) = -\frac{b_0^{\mathbb{Q}}}{b_1^{\mathbb{Q}}}(1 - B(\theta^{\mathbb{Q}},\Delta)). \tag{14}$$

We define IV_t as the squared VXO index (described in Section 3.1) observed at time t. It is directly related to the expected variance over the period of 22 days (i.e. $\Delta = 22/262$) by the formula

$$IV_t \approx \frac{1}{\Delta}\,\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\Delta} V_s\,ds\right]. \tag{15}$$

This approximation is very good and the error stems solely from the fact that the algorithm that computes the value of VXO uses finitely many options.

We now exploit the relationship in (15) to express the log-likelihood function for both the linear and the nonlinear model described by the real world drifts given in (6) and (8) respectively and by SDE (3)–(4) under the pricing measure $\mathbb{Q}$. By the Markov property property we can in both models decompose the log-likelihood into a sum of log-transition densities (for ease of notation we henceforth denote IV_{t_i} by IV_i and X_{t_i} by X_i) as follows

$$\ell(X_1, IV_1, \ldots, X_N, IV_N \mid X_0, IV_0, \theta) = \sum_{i=1}^{N} \log p^{IV}(X_i, IV_i \mid X_{i-1}, IV_{i-1}, \theta), \tag{16}$$

where $p^{IV}(X_i, IV_i \mid X_{i-1}, IV_{i-1}, \theta)$ denotes the conditional transition density of the random vector $(X_{t_i}, IV_{t_i})^\top$. The linear transformation

$$V_t = \frac{IV_t - A(\theta^{\mathbb{Q}},\tau)}{B(\theta^{\mathbb{Q}},\tau)}, \tag{17}$$

which follows from (15), implies that we can express the log-likelihood as

$$\sum_{i=1}^{N} \log p^{V}(X_i, V_i \mid X_{i-1}, V_{i-1}, \theta) - N\,\log B(\theta^{\mathbb{Q}},\tau), \tag{18}$$

where $p^{V}(X_i, V_i \mid X_{i-1}, V_{i-1}, \theta)$ denotes the conditional transition density of the random vector $(X_{t_i}, V_{t_i})^\top$

given the values of X_{i-1} and V_{i-1}. The final change of variable $Y_t = \gamma(V_t)$ given in (10) yields the log-likelihood which takes the form

$$\ell(\theta) = \sum_{i=1}^{N} \{\log p(X_i, Y_i \mid X_{i-1}, Y_{i-1}, \theta) - \sigma Y_i\} - N\,(\log B(\theta^{\mathbb{Q}}, \tau) - \sigma), \tag{19}$$

where $p(X_i, Y_i \mid X_{i-1}, Y_{i-1}, \theta)$ denotes the conditional transition density of the random vector $(X_{t_i}, Y_{t_i})^\top$.

3.4 Limited information expected maximum likelihood estimation

The transition densities for the transformed variance processes (11) and (12) that arise in the log-likelihood (19) are not available in closed form. To overcome this issue and that of the supposedly flat likelihood function we apply expected maximum likelihood (EML) estimation algorithm from Mijatović and Schneider (2009). This technique makes use of the closeness of the law of the Brownian bridge to the true law of the diffusion bridge, and of the Euler scheme approximation for the transition density when the time interval between observations is small. However, the EML algorithm cannot be directly applied to the present econometric problem, as it only works for the estimation of one-dimensional diffusions. We therefore propose an efficient three-step limited-information maximum likelihood procedure, described below. The efficiency of the algorithm arises from the fact that EML can be used to express the globally optimal drift parameters $\theta^{V\mathbb{P}\star}$ and $\theta^{X\mathbb{P}\star}$ (optimal parameters are denoted with a superscript *) as complicated, yet closed-form functions of the parameters $\theta^\sigma \cup \theta^{\mathbb{Q}}$ and the data. In other words, for fixed values $\theta^\sigma \cup \theta^{\mathbb{Q}}$ the data implies optimal parameter values $\theta^{V\mathbb{P}\star}$ and $\theta^{X\mathbb{P}\star}$. The EML algorithm therefore effectively reduces the parameter space from $\theta^{X\mathbb{P}} \cup \theta^{V\mathbb{P}} \cup \theta^\sigma \cup \theta^{\mathbb{Q}}$ to $\theta^\sigma \cup \theta^{\mathbb{Q}}$. As a result a conventional likelihood search using standard optimization techniques over equation (19) is necessary only for $\theta^\sigma \cup \theta^{\mathbb{Q}}$.

To make our processes suitable for EML estimation we first introduce $M-1$ auxiliary data $U_{i,1}, \ldots, U_{i,M-1}$ between each observed data pair $(X_i, Y_i)^\top, (X_{i+1}, Y_{i+1})^\top$ with the convention that $U_{i,0} := U_i := (X_i, Y_i)^\top$, and $U_{i,M} := U_{i+1} := (X_{i+1}, Y_{i+1})^\top$. This augmentation leads to a total of $MN+1$ data pairs. To lighten notation we switch for the below equations to a single-index notation U_k, $k = 0, \ldots, MN$. We set $\delta := \frac{\Delta}{M}$ and write down the discretized version of the continuous-time SDE eliminating heteroskedasticity in the innovations for the linear variance model (LN)

$$\begin{gathered}\frac{X_{k+1} - X_k - \rho\sqrt{e^{\sigma Y_k}}\varepsilon^V_{k+1}}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}} = \left\{ \left(a_0 + a_1\, e^{\sigma Y_k}\right) \frac{1}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}} \right\} \delta + \varepsilon^X_{k+1} \\ Y_{k+1} - Y_k + \frac{\sigma}{2}\delta = \left(b_0^{\mathbb{Q}} + b_1\, e^{\sigma Y_k}\right) \frac{1}{\sigma\, e^{\sigma Y_k}}\delta + \varepsilon^V_{k+1},\end{gathered} \tag{20}$$

and the nonlinear variance model (NL)

$$\begin{aligned}\frac{X_{k+1}-X_k-\rho\sqrt{e^{\sigma Y_k}}\varepsilon^V_{k+1}}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}} &= \left\{\left(a_0+a_1\,e^{\sigma Y_k}\right)\frac{1}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}}\right\}\delta+\varepsilon^X_{k+1}\\ Y_{k+1}-Y_k+\frac{\sigma}{2}\delta &= \left(b_0+b_1\,e^{\sigma Y_k}+b_2\,e^{2\sigma Y_k}+\frac{b_3}{e^{\sigma Y_k}}\right)\frac{1}{\sigma\,e^{\sigma Y_k}}\delta+\varepsilon^V_{k+1}.\end{aligned} \tag{21}$$

It can be seen that the difference equations (20) and (21) above for both, log stock prices, as well as stochastic variance can be written in the form

$$\begin{aligned}g_X(U_{k+1},U_k) &= \left(f^X_0(U_k)+f^X_1(U_k)\right)\delta+\varepsilon^X_{k+1}\\ g_{\mathcal{M}}(U_{k+1},U_k) &= \sum_{l=0}^{L_{\mathcal{M}}} f^{\mathcal{M}}_l(U_k)\,\delta+\varepsilon^V_{k+1},\ \mathcal{M}\in\{\text{LN, NL}\}\,,\end{aligned}$$

where the functions g and f are displayed in tables 2b and 2a. For the linear variance model we have $L_{\text{LN}}=1$, and for the nonlinear model we have $L_{\text{NL}}=3$. Innovations ε^V_k and ε^X_k are both identically independently $N(0,\delta)$-distributed random variables for $k=1,\dots,MN$. Note the appearance of ε^V_{k+1} in equ. (20) and (21), respectively. This is the reason why we need to estimate $\theta^{V\mathbb{P}}$ first (the variance dynamics do not depend on the log stock price) and $\theta^{X\mathbb{P}}$ subsequently, conditional on $\theta^{V\mathbb{P}\star}$. Hence the terminology 'limited information' EML estimation. For the below algorithm denote with $\theta^{V^{\mathcal{M}}\mathbb{P}},\mathcal{M}\in\{\text{LN, NL}\}$ the parameters of the linear, respectively nonlinear variance process. If there is no ambiguity we will just write $\theta^{V\mathbb{P}}$.

Optimizing $\theta^{V\mathbb{P}}\mid\theta^{\sigma},\theta^{\mathbb{Q}}$: Plugging the current values of θ^{σ} and $\theta^{\mathbb{Q}}$ into eq. (15) the observed data implies a time series of Y. An estimate of the parameters of the transformed variance process Y can now be obtained by means of EML (Mijatović and Schneider, 2009). For this purpose we introduce functions f and g from difference equations (20) and (21) which are displayed in Table 2. For a given variance model $\mathcal{M}$ we put the variance drift parameters in a vector $x^{\mathcal{M}}:=\left(b^{\mathcal{M}}_0,\dots,b^{\mathcal{M}}_{L_{\mathcal{M}}}\right)^{\top}$. EML then yields the

	$\mathcal{M} = \text{LN}$	$\mathcal{M} = \text{NL}$
$f_0^{\mathcal{M}}(u_k)$	$1/(\sigma e^{\sigma y_k})$	$1/(\sigma e^{\sigma y_k})$
$f_1^{\mathcal{M}}(u_k)$	$1/\sigma$	$1/\sigma$
$f_2^{\mathcal{M}}(u_k)$		$e^{\sigma y_k}/\sigma$
$f_3^{\mathcal{M}}(u_k)$		$1/(\sigma\, e^{2\sigma y_k})$
$g_{\mathcal{M}}(u_k, u_{k-1})$	$y_k - y_{k-1} + (\frac{1}{2}\sigma - b_0^{\mathbb{Q}})\delta$	$y_k - y_{k-1} + \frac{1}{2}\sigma\delta$

(a) Variance drift functions

$f_0^X(u_k)$	$1/(\sqrt{1-\rho^2}\sqrt{e^{\sigma y_k}})$
$f_1^X(u_k)$	$\sqrt{e^{\sigma y_k}}/\sqrt{1-\rho^2}$
$g_X(u_k, u_{k-1})$	$\left(x_k - x_{k-1} - \rho\sqrt{e^{\sigma y_{k-1}}}\varepsilon_k^V\right) / \left(\sqrt{1-\rho^2}\sqrt{e^{\sigma y_{k-1}}}\right)$

(b) Stock drift functions for model $\mathcal{M} \in \{\text{NL, LN}\}$

Table 2: **Function specification for EML estimation:** The tables contain the functions that appear as summands in the respective drifts of the LN and NL models, which need to be evaluated in the conditional expectations in (22) and (23), expressed in terms of the variable $u_k := (x_k, y_k)$.

optimal drift coefficients $\theta^{V^{\mathcal{M}}\mathbb{P}_\star}$ as the unique solution of the linear system $x^{\mathcal{M}} = (\varXi^{\mathcal{M}})^{-1}\,\varpi^{\mathcal{M}}$ with

$$\varXi^{\mathcal{M}} = \delta \sum_{n=1}^{N-1}\sum_{m=0}^{M-1} \begin{pmatrix} \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n,U_{n+1}}}\left[f_0^{\mathcal{M}}(U_{n,m}) f_0^{\mathcal{M}}(U_{n,m})\right] & \cdots & \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n,U_{n+1}}}\left[f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m}) f_0^{\mathcal{M}}(U_{n,m})\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n,U_{n+1}}}\left[f_0^{\mathcal{M}}(U_{n,m}) f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m})\right] & \cdots & \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n,U_{n+1}}}\left[f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m}) f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m})\right] \end{pmatrix}, \tag{22}$$

$$\varpi^{\mathcal{M}} = \sum_{n=1}^{N-1}\sum_{m=0}^{M-1} \begin{pmatrix} \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n,U_{n+1}}}\left[g(U_{n,m+1}, U_{n,m})\, f_0^{\mathcal{M}}(U_{n,m})\right] \\ \vdots \\ \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n,U_{n+1}}}\left[g(U_{n,m+1}, U_{n,m})\, f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m})\right] \end{pmatrix}. \tag{23}$$

The symbol $\mathbb{Q}_{\mathcal{M}}^{x,y}$ denotes the (unknown) law of the diffusion bridge pertaining to model $\mathcal{M}$ conditioned on the endpoints x and y, respectively. We approximate the law of the true diffusion bridge $\mathbb{Q}_{\mathcal{M}}^{x,y}$ with the law of a Brownian bridge $\mathbb{W}_{\mathcal{M}}^{x,y}$. It is shown in Mijatović and Schneider (2009) that $\mathbb{Q}_{\mathcal{M}}^{x,y}$ is absolutely continuous with respect to $\mathbb{W}_{\mathcal{M}}^{x,y}$, and that there is in fact very little deviation between the two even for long time intervals. Exact draws from the Brownian bridge are obtained from the stochastic difference equation (Stramer and Yan, 2007)

$$U_{i-1,m+1} = U_{i-1,m} + \frac{U_{i-1,M} - U_m}{M-m} + \sqrt{\frac{M-m-1}{M-m}}\,\varepsilon_{i-1,m+1}, \tag{24}$$

with $\varepsilon_{i,m} \sim N(0,\delta)$, $i = 1,\dots,N-1, m = 1,\dots,M-1$.

Optimizing $\theta^{X\mathbb{P}} \mid \theta^{\sigma},\theta^{\mathbb{Q}},\theta^{V\mathbb{P}\star}$: Conditional on the optimal variance drift parameters $\theta^{V\mathbb{P}\star}$ the f and g functions (the g function depends on the drift parameters of the variance through ε^V) from Table 2a can now be swapped with the functions from Table 2b to estimate optimal stock drift parameters $\theta^{X\mathbb{P}\star}$ through the solution of the linear system (22) – (23).

There is no direct EML estimator for $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$. To find $\theta^{\sigma\star}$ and $\theta^{\mathbb{Q}\star}$ we therefore need to perform a conventional likelihood search using likelihood (19) as the objective function. Since for any value of θ^{σ} and $\theta^{\mathbb{Q}}$ EML yields optimal $\theta^{X\mathbb{P}\star}$ and $\theta^{V\mathbb{P}\star}$, we see likelihood function (19) only as a function of $\theta^{\sigma},\theta^{\mathbb{Q}}$, and the data. To approximate the unknown transition densities which appear in (19) we use the simulation-based estimator from Pedersen (1995) in connection with the Brownian bridge importance sample from Durham and Gallant (2002)

$$\sum_{i=1}^{N}\log p^{\mathcal{M}}(X_i,Y_i \mid X_{i-1},Y_{i-1},\theta) \approx \sum_{i=1}^{N}\log\left\{\frac{1}{S}\sum_{s=1}^{S}\frac{\prod_{m=1}^{M} p^{E\mathcal{M}}(U_{i-1,m}\mid U_{i-1,m-1},\theta)}{\prod_{m=1}^{M-1} q(U_{i-1,m}\mid U_{i-1,m-1},U_{i-1,M})}\right\}. \tag{25}$$

Here, $p^{\mathcal{M}}$ refers to the true transition density arising from Heston dynamics with variance specification (11) (LN) and (12) (NL), respectively. The density $p^{E\mathcal{M}}$ denotes a normal distribution arising from the Euler discretization of the corresponding SDE. Auxiliary state variables $U_{i-1,m},\dots,U_{i-1,m}$, $i = 1,\dots,N, m = 1,\dots,M-1$ are simulated according to the stochastic difference equation

$$U_{i-1,m+1} = U_{i-1,m} + \frac{U_{i-1,M}-U_m}{M-m} + \sqrt{\frac{M-m-1}{M-m}}\,\Sigma(U_{i-1,m})\,\varepsilon_{i-1,m+1}, \tag{26}$$

where

$$\Sigma(U_{i-1,m}) = \begin{pmatrix}\sqrt{1-\rho^2}e^{\sigma Y_{i-1,m}} & \rho\sigma e^{\sigma Y_{i-1,m}}\\ 0 & 1\end{pmatrix},\ \varepsilon_{i-1,m+1} = \begin{pmatrix}\varepsilon^X_{i-1,m+1}\\ \varepsilon^V_{i-1,m+1}\end{pmatrix}. \tag{27}$$

Both $p^{E\mathcal{M}}$ and q are multivariate normal densities:

$$q(U_{i-1,m+1}\mid U_{i-1,m},U_{i-1,M}) = \phi\left(U_{i-1,m+1};U_{i-1,m}+\frac{U_{i-1,M}-U_m}{M-m},\frac{M-m-1}{M-m}\,\Sigma(U_{i-1,m})\Sigma(U_{i-1,m})^{\top}\delta\right)$$

$$p^{E\mathcal{M}}(U_{i-1,m+1}\mid U_{i-1,m},\theta) = \phi\left(\begin{pmatrix}X_{i-1,m+1}-X_{i-1,m}\\ g_{\mathcal{M}}(U_{i-1,m+1},U_{i-1,m})\end{pmatrix};\begin{pmatrix}a_0+a_1e^{\sigma Y_{i-1,m}}\\ \sum_{l=0}^{L_{\mathcal{M}}} f_l^{\mathcal{M}}(U_{i-1,m})\end{pmatrix}\delta,\Sigma(U_{i-1,m})\Sigma(U_{i-1,m})^{\top}\delta\right).$$

Following Stramer and Yan (2007) we set $S = M^2 = 576$. Note that the ε variates appearing in (24) from EML estimation may be reused in this step.

3.5 Empirical results

We assess the quality of the linear and nonlinear models introduced in Section 3.2 by investigating forecasts of realized variance, implied variance and stock returns for various maturities. The forecasting exercise is performed both in and out of sample. For the out-of-sample period the model is re-estimated each time a new datapoint is added. Figure 1 gives a visual impression of the in-sample as well as the out-of-sample period. The corresponding estimation paths for a selection of parameters can be seen in Figures 3a to 3d.

Point estimates and standard errors for the parameters of the nonlinear model NL (cf. (12)) and the linear model LN (cf. (11)) obtained from the limited information EML algorithm described in Section 3.4 can be found in Table 3. These parameter estimates are based on the entire sample. The $\mathbb{Q}$ mean-reversion parameter $b_1^{\mathbb{Q}}$ is large and positive for the linear and the nonlinear model. This is consistent with the explosive coefficients estimated in Jones (2003) and Pan (2002) and with the negative variance risk premia observed in Carr and Wu (2008). The positive estimates result in a time series for instantaneous variance that is located consistently below the time series of implied variance through relation (15) (see Figure 2a). The correlation and diffusion parameters ρ and σ as well as $b_0^{\mathbb{Q}}$ are also comparable in scale for both specifications. However under the physical measure $\mathbb{P}$ the linear and the nonlinear specifications predict different behaviour. Figure 2d shows that during calm times (in the region between 0.01 and 0.04) the nonlinear specification predicts that the instantaneous variance behaves as a random walk or a process that diverts at an even faster rate. Figure 2b suggests that there is a strong pull away from the zero boundary and from very high values in the case of the nonlinear drift. Such behaviour cannot be reproduced with a linear drift specification (cf. also the drift function estimated from the time series of the VIX index in Bandi and Renó (2009), which is of a shape similar to that of the drift function in Figure 2b).

Recall that the *risk premium* (i.e.the *market price of risk*) at time $t \in [0, T]$ in the model $\mathcal{M} \in \{\text{LN, NL}\}$ is given by

$$\Lambda_{\mathcal{M}}(V_t) = \Sigma(V_t)^{-1} f_{\mathcal{M}}(V_t), \qquad \text{where} \quad \Sigma(V_t) = \sqrt{V_t} \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & \sigma\sqrt{V_t} \end{pmatrix} \tag{28}$$

and $f_{\mathcal{M}}$ is defined in (7) and (9) for $\mathcal{M} = \text{LN}$ and $\mathcal{M} = \text{NL}$ respectively. The risk premium for the stochasticity of the variance is given by the second component $\Lambda^V_{\mathcal{M}}(V_t)$ of the market price of risk vector $\Lambda_{\mathcal{M}}(V_t)$. Note that in the case $\mathcal{M} = \text{LN}$, the variance risk premium $\Lambda^V_{\text{LN}}(V_t)$ is a non-zero constant given by $(b_1 - b_1^{\mathbb{Q}})/\sigma$. The resulting time series of the risk premia reflect the difference in the estimated real world drifts described in the previous paragraph: while the unconditional mean of the risk premium on the $W^{V\mathbb{Q}}$ Brownian motion (see SDE (4)) is similar for both specifications, the nonlinear model exhibits time-variability in the market prices of risk (see Figure 2c), in contrast to the constant risk premium, given by $(b_1 - b_1^{\mathbb{Q}})/\sigma$, in the linear model.

3.5.1 Forecasts

In this subsection we consider, in addition to the nonlinear model (NL) and the linear model (LN), a random walk martingale model (RW), where the prediction for any future value is taken to be the current value. The forecasting power of the models is tested with 2395 in-sample, and 1630 out-of-sample observations of implied variance forecasts. Each observation of forecast errors is comprised of a cross section of residuals pertaining to 1 day, 1 week, 4 weeks, 12 weeks (quarter trading year) and 26 weeks (half trading year) forecasting errors for stock returns and implied variance. Realized variance forecast errors are computed for horizons of 1 week, 4 weeks, 12 weeks and 26 weeks, where realized variance at time t_i computed over N days is defined as

$$RV_i(N) := \frac{262}{N} \sum_{j=i-N}^{i} (X_j - X_{j-1})^2. \tag{29}$$

For the model $\mathcal{M} \in \{\text{RW, LN, NL}\}$ we compute the realized variance using the model-implied instantaneous variance, which is annualized by construction

$$RV_i^{\mathcal{M}}(N) := \frac{1}{N} \sum_{j=i-N}^{i} V_j. \tag{30}$$

Conditional expectations for the LN and NL models are computed by Monte Carlo integration using $2 \cdot 10^4$ paths with hourly discretization of the SDE.[1] Tables 4, 5 and 6 report the mean absolute error (MAE) and the root mean squared error (RMSE) of the sampling distribution of forecasting residuals for the realized variance, the implied variance and the stock returns, respectively. In addition directional forecasts as well as p-values of the Clark and West (2007) test statistics for nested models are reported.

Figure 2a indicates structural breaks in the implied variance time series. The in-sample period spans these regimes, and the out-of-sample period contains both very rough and very calm periods. Both the linear and the nonlinear specification fail to capture the time-dependent long-run mean of the implied

[1] With approximation (15) forecasts for the implied variance can be computed as linear functions of conditional instantaneous variance expectations. Expectations for both the linear and nonlinear model are evaluated by Monte Carlo integration. This is despite the availability of an analytic expression for the conditional expectation in the linear model so that both specifications are subject to the same simulation error (the same set of random numbers is used for the integration).

variance. The Clark and West (2007) statistics indicate that both the linear, and the nonlinear model have significant advantages over the random walk. These results are in line with the risk premia approach in Chernov (2007). The statistics furthermore indicate that the nonlinear model has a statistically (highly) significant advantage in forecasting over the linear model. This observation holds for the realized and implied variance for all forecasting horizons, both in sample and out of sample. A tremendous improvement for the realized variance over the RW specification can be attributed to mean reversion, which the NL and the LN model accommodate. The forecast residuals for all three competing models are heavily negatively skewed, however, and the distribution is fat-tailed as a consequence of a few heavy outliers. Table 4 additionally reports normalized MSE (NMSE) for comparison with the results in Sizova (2008). For stock returns excellent in-sample results, most likely obtained through explicit modeling of the leverage effect, cannot be reproduced out of sample. The likely reason is a change in the drift regime (cf. Figure 1), which is not accounted for by the variance modeling.

The results for the directional forecasts of the realized and implied variance and of stock returns are mixed. The NL and LN models show very good results in sample and out of sample for the realized variance. For the implied variance the directional forecasts are slightly worse than the ones in Ahoniemi (2006), where forecasts are given for one maturity only. Directional out-of-sample forecasts for stock returns also suffer from the change in the drift during the out-of-sample period mentioned above (cf. Figure 1).

4 Conclusion

We introduce a simple continuous-time diffusion framework that combines semi-analytic pricing formulae with flexible nonlinear time series modeling. Using an econometrically inconspicuous dampening function we ensure that a solution to the nonlinear stochastic differential equation under the physical measure exists. We estimate a nonlinear stochastic volatility model on the joint time series of the S&P 100 and the VXO implied volatility index. Forecast tests show that the nonlinear model has superior forecasting power over the random walk and the linear model for short prediction horizons; the results are statistically significantly better both in sample and out of sample. This suggests that a nonlinear specification of the drift under the physical measure could potentially be very useful in trading and risk management.

References

Ahoniemi, K. (2006). Modeling and forecasting implied volatility – an econometric analysis of the vix index. Working paper, Helsinki Center of Economic Research.

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9(2):385–426.

Aït-Sahalia, Y. (2001). Closed-form likelihood expansions for multivariate diffusions. Working paper, Princeton University and NBER.

Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83:413–452.

Bandi, F. and Renó, R. (2009). Nonparametric stochastic volatility. Working paper, University of Chicago and Università de Siena.

Bertholon, H., Montfort, A., and Pegoraro, F. (2008). Econometric asset pricing modelling. *Journal of Financial Econometrics*, 6(4):407–458.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(333–382).

Carr, P. and Wu, L. (2008). Variance risk premiums. *Review of Financial Studies*. forthcoming.

Cheridito, P., Filipović, D., and Kimmel, R. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83(1):123–170.

Chernov, M. (2007). On the role of risk premia in volatility forecasting. *Journal of Business & Economic Statistics*, 25(4).

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138:291–311.

Dai, Q., Le, A., and Singleton, K. (2006). Discrete-time dynamic term structure models with generalized market prices of risk. Working paper, Stanford University.

Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability*, 13:984–1053.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Durham, G. B. and Gallant, R. A. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–316.

Forman, J. L. and Sørensen, M. (2008). The pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35:438–465.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New York.

Jones, C. S. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics*, 116:181–224.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 2 edition.

Mijatović, A. and Schneider, P. (2009). Globally optimal parameters for non-linear diffusions. *Annals of Statistics*. forthcoming.

Nelson, D. B. (2002). Arch models as diffusion approximations. *Journal of Econometrics*, 45:7–38.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63:3–50.

Pedersen, A. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22:55–71.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaft*. Springer, Berlin.

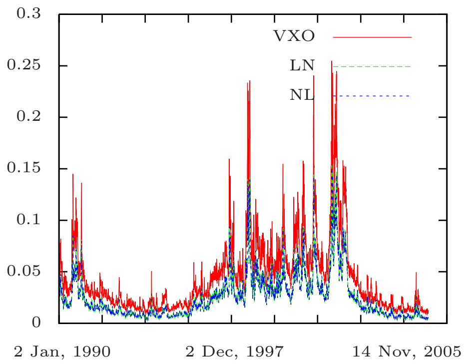
Sizova, N. (2008). Variance forecast performance measures: An economic approach. Working paper, Duke University.

Stramer, O. and Yan, J. (2007). On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. *Journal of Computational & Graphical Statistics*, 16(3):672–691.

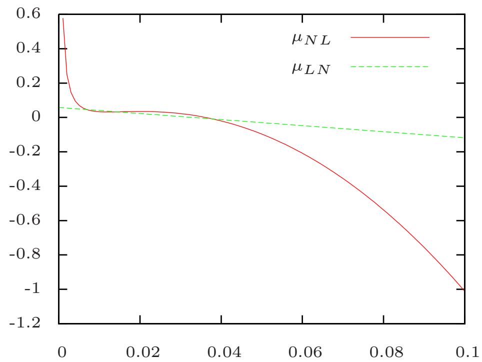
A Figures and Tables

	Linear	Nonlinear
σ	2.2047	2.1734
	(0.0374)	(0.0554)
ρ	-0.6768	-0.6803
	(0.01577)	(0.0165)
$b_0^{\mathbb{Q}}$	0.05817	0.0500
	(0.0046)	(0.0098)
$b_1^{\mathbb{Q}}$	10.9858	11.3260
	(0.00648)	(0.08008)
a_0	0.0748	0.0284
	(0.0308)	(0.0359)
a_1	3.3370	6.0870
	(0.0119)	(0.0622)
b_0		-0.1064
		(0.0221)
b_1	-1.7645	8.9591
	(0.0355)	(0.0689)
b_2		-180.7473
		(0.22490)
b_3		0.00068
		(0.00016)

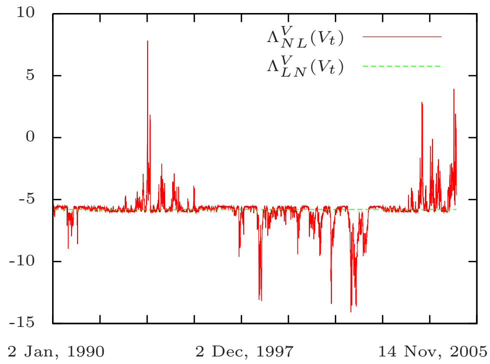
Table 3: **Parameter Estimates:** The table displays parameter estimates for the linear model (11) and the nonlinear model (12). Huber (sandwich) standard errors are computed from the asymptotic covariance matrix pertaining to the likelihood in (19) with the transition density approximation in in (25). The asymptotic covariance matrix of the estimated parameter vector $\hat{\theta}$ is computed according to the formula in Hamilton (1994, page 145, formula 5.8.7).



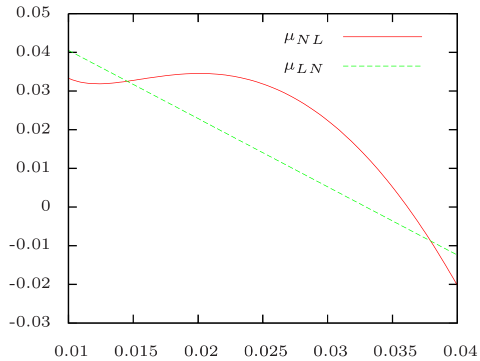
(a) VXO



(b) Estimated Drift

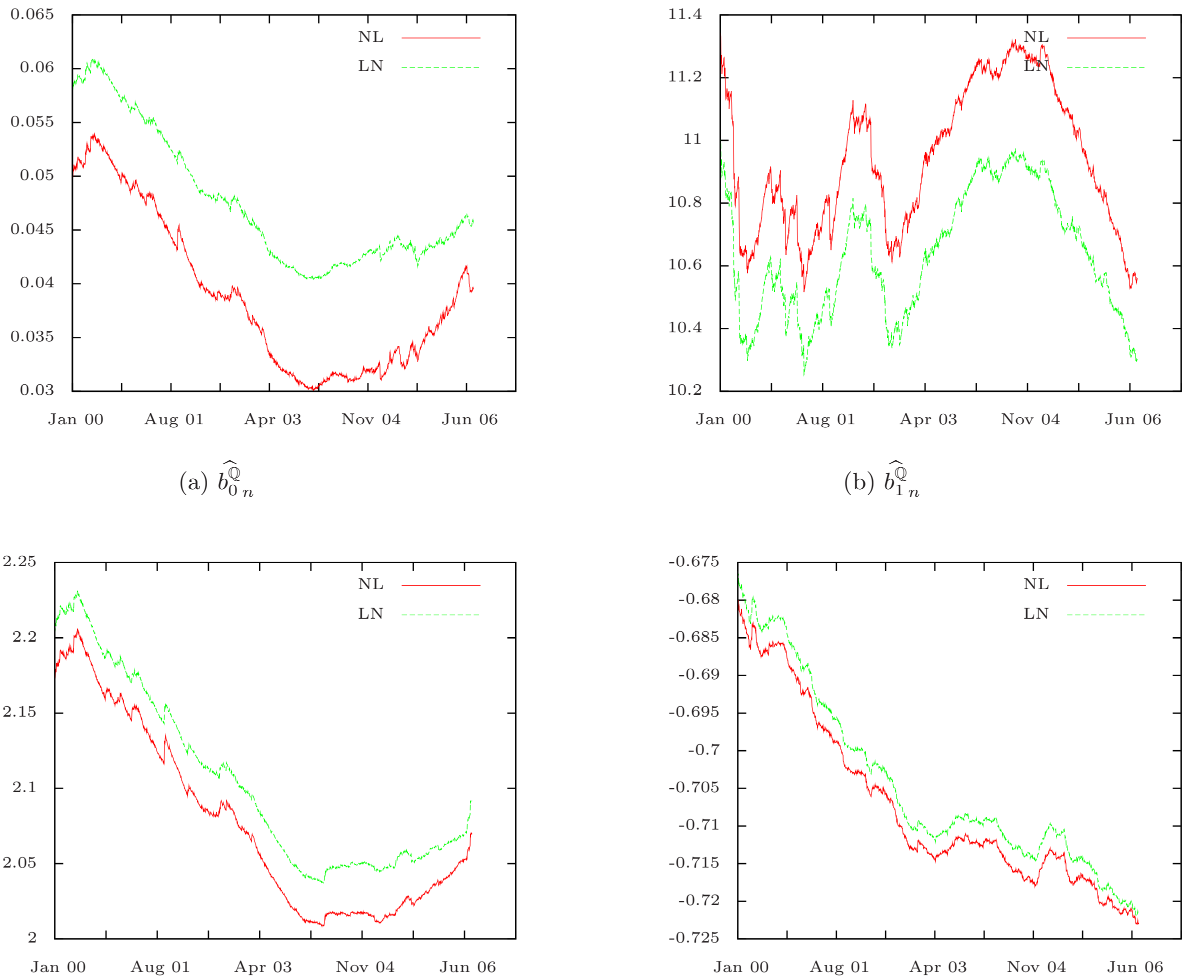


(c) Variance Premia



(d) Estimated Drift (Zoom)

Figure 2: **VXO and instantaneous variance:** Figure 2a displays the VXO along with the instantaneous variance implied by the $\mathbb{Q}$ parameters from Table 3. Figure 2b displays the nonlinear drift at the parameter estimates for variance model (8), and the linear drift for model (6). The implied time series for the risk premia of the stochastic variance, given by the second coordinate of the market price of risk vector (28), in the linear and the nonlinear model is displayed in Figure 2c.



(a) $\widehat{b_0^{\mathbb{Q}}}_n$

(b) $\widehat{b_1^{\mathbb{Q}}}_n$

(c) $\widehat{\sigma}_n$

(d) $\widehat{\rho}_n$

Figure 3: **Parameter paths:** Figure 3 displays estimates for the parameters $b_0^{\mathbb{Q}}, b_1^{\mathbb{Q}}, \sigma$, and ρ, respectively, as the sample window is updated on a daily basis. For each date on the x-axis model (11) and (12) are re-estimated using the EML methodology from Section 3.4.

IN-SAMPLE		1w	4w	12w	26w
RMSE	RW	0.05201	0.05731	0.05908	0.05986
	LN	0.02758	0.01825	0.01494	0.01348
	NL	0.02776	0.01814	0.01433	0.01364
NMSE	RW	238%	567%	943%	1182%
	LN	66%	57%	60%	59%
	NL	67%	56%	55%	61%
MAE	RW	0.02191	0.02425	0.02480	0.02537
	LN	0.01174	0.00914	0.00862	0.00991
	NL	0.01172	0.00897	0.00807	0.00968
DIR	RW	0.49227	0.51609	0.50230	0.49979
	LN	0.70497	0.74133	0.68784	0.62599
	NL	0.70121	0.74384	0.69118	0.61429
CW	LN vs. RW	$0.00281^{(***)}$	$0.00612^{(***)}$	$0.00989^{(***)}$	$0.01087^{(**)}$
	NL vs. RW	$0.00289^{(***)}$	$0.00638^{(***)}$	$0.01053^{(**)}$	$0.01166^{(**)}$
	NL vs. LN	1	0.26299	$0.09399^{(*)}$	$0.09795^{(*)}$
OUT-SAMPLE					
RMSE	RW	0.06789	0.07478	0.07767	0.07958
	LN	0.03596	0.02714	0.02600	0.02585
	NL	0.03690	0.02899	0.02723	0.02597
NMSE	RW	192%	361%	546%	751%
	LN	53%	47%	61%	79%
	NL	56%	54%	67%	79%
MAE	RW	0.03389	0.03718	0.03892	0.03946
	LN	0.01857	0.01561	0.01705	0.01876
	NL	0.01875	0.01606	0.01678	0.01825
DIR	RW	0.49049	0.48557	0.49847	0.41989
	LN	0.70534	0.69061	0.62247	0.57274
	NL	0.70166	0.68447	0.61387	0.56967
CW	LN vs. RW	$0^{(***)}$	$0^{(***)}$	$0.00014^{(***)}$	$0.00107^{(***)}$
	NL vs. RW	$0^{(***)}$	$0^{(***)}$	$0.00017^{(***)}$	$0.00126^{(***)}$
	NL vs. LN	1	1	0.22124	$0.09065^{(*)}$

Table 4: **Realized variance forecasting:** This table displays mean absolute error MAE, given by $\frac{1}{N-\tau}\sum_{i=\tau}^{N}|\epsilon_i(\tau)|$, root mean squared forecast error RMSE, given by $\sqrt{\frac{1}{N-\tau}\sum_{i=\tau}^{N}\epsilon_i(\tau)^2}$, and normalized MSE (NMSE), defined as $\sum_{i=\tau}^{N}\epsilon_i(\tau)^2/\left(\sum_{i=\tau}^{N}(RV_i(\tau)-\overline{RV}_i(\tau))^2\right)$, where $\epsilon_i(\tau) := RV_i(\tau) - \mathbb{E}^{\mathbb{P}}_{t_{i-\tau}}\left[RV_i^{\mathcal{M}}(\tau)\right]$ and $\tau \in \{5, 22, 66, 131\}$. The realized varieance $RV_i(\tau)$ is defined in (29) and the random variable $RV_i^{\mathcal{M}}(\tau)$ is given in (30) for any $\mathcal{M} \in \{\text{RW, LN, NL}\}$, where RW denotes the random walk model and LN (resp. NL) stands for the linear (resp. nonlinear) model given in (11) (resp. in (12)). DIR shows the percentage of correct directional forecasts. CW denotes p-values for the Clark and West (2007) test for nested models. Asterisks $^{(***),(**),(*)}$ denote significance at the 1%, 5% and 10% confidence level respectively.

IN-SAMPLE		1d	1w	4w	12w	26w
MSE	RW	0.00762	0.01380	0.02029	0.02765	0.02967
	LN	0.00760	0.01362	0.01945	0.02518	0.02745
	NL	0.00748	0.01308	0.01833	0.02353	0.02688
MAE	RW	0.00388	0.00748	0.01110	0.01456	0.01691
	LN	0.00388	0.00751	0.01149	0.01656	0.02140
	NL	0.00385	0.00732	0.01085	0.01517	0.01934
DIR	RW	0.49269	0.49687	0.49310	0.47305	0.49812
	LN	0.52069	0.53364	0.55328	0.58379	0.61220
	NL	0.51692	0.53197	0.55161	0.55871	0.54952
CW	LN vs. RW	$0.00413^{(***)}$	$0.00954^{(***)}$	$0.02119^{(**)}$	$0.01957^{(**)}$	$0.01366^{(**)}$
	NL vs. RW	$0.00809^{(***)}$	$0.02778^{(**)}$	$0.06111^{(*)}$	$0.03683^{(**)}$	$0.02815^{(**)}$
	NL vs. LN	$0.01307^{(**)}$	$0.04338^{(**)}$	0.11345	$0.05275^{(*)}$	$0.01554^{(**)}$
OUT-SAMPLE						
MSE	RW	0.00903	0.01789	0.02810	0.03919	0.04401
	LN	0.00902	0.01781	0.02786	0.03926	0.04831
	NL	0.00891	0.01699	0.02679	0.03684	0.04260
MAE	RW	0.00513	0.01014	0.01676	0.02470	0.02678
	LN	0.00514	0.01019	0.01749	0.02847	0.03826
	NL	0.00507	0.00970	0.01605	0.02507	0.03148
DIR	RW	0.48128	0.47759	0.45181	0.41866	0.36710
	LN	0.51013	0.51320	0.51688	0.50460	0.43892
	NL	0.50890	0.52977	0.53898	0.48803	0.46593
CW	LN vs. RW	$0.02588^{(**)}$	$0.00106^{(***)}$	$0.01306^{(**)}$	0.10687	0.41651
	NL vs. RW	$0.01775^{(**)}$	$0.06154^{(*)}$	$0.01735^{(**)}$	$0.01821^{(**)}$	$0.0328^{(**)}$
	NL vs. LN	$0.02178^{(**)}$	$0.07032^{(*)}$	$0.02653^{(**)}$	$0.03312^{(**)}$	$0.04385^{(**)}$

Table 5: **Implied variance forecasting:** This table displays mean absolute error MAE, given by $\frac{1}{N-\tau}\sum_{i=1}^{N-\tau}|\epsilon_i(\tau)|$, and root mean squared forecast error RMSE, defined by $\sqrt{\frac{1}{N-\tau}\sum_{i=1}^{N-\tau}\epsilon_i(\tau)^2}$, where $\epsilon_i(\tau) := IV_{t_i+\tau} - \mathbb{E}^{\mathbb{P}}_{t_i}\left[IV^{\mathcal{M}}_{t_i+\tau}\right]$ and $\tau \in \{1, 5, 22, 66, 131\}$. The random variable $IV^{\mathcal{M}}_t$ is defined as a linear transformation, given in (17), of the instantaneous variance in the model $\mathcal{M} \in \{\text{NL, LN}\}$ and IV_t denotes the square of the VXO index at time t. As in the previous table RW denotes the random walk model and LN (resp. NL) stands for the linear (resp. nonlinear) model given in (11) (resp. in (12)). DIR shows the percentage of correct directional forecasts. CW denotes p-values for the Clark and West (2007) test for nested models. Asterisks $^{(***)},^{(**)},^{(*)}$ denote significance at the 1%, 5% and 10% confidence level respectively.

IN-SAMPLE		1d	1w	4w	12w	26w
MSE	RW	0.00936	0.02041	0.03966	0.07192	0.11658
	LN	0.00932	0.02006	0.03702	0.05697	0.06950
	NL	0.00932	0.02002	0.03693	0.05690	0.07029
MAE	RW	0.00667	0.01540	0.03058	0.05492	0.09428
	LN	0.00664	0.01501	0.02812	0.04245	0.05708
	NL	0.00664	0.01495	0.02798	0.04225	0.05717
DIR	RW	0.53698	0.58629	0.65650	0.79440	0.88048
	LN	0.53698	0.58629	0.65650	0.79440	0.88048
	NL	0.53698	0.58629	0.65650	0.79440	0.88048
CW LN vs. RW		$0.00011^{(***)}$	$0.00594^{(***)}$	$0.00265^{(***)}$	$0.00012^{(***)}$	$0^{(***)}$
CW NL vs. RW		$0.00022^{(***)}$	$0.00722^{(***)}$	$0.00417^{(***)}$	$0.00032^{(***)}$	$0^{(***)}$
CW NL vs. LN		$0.06384^{(*)}$	0.17865	$0.03174^{(**)}$	0.33946	1
OUT-SAMPLE						
MSE	RW	0.01218	0.02555	0.04865	0.07302	0.10893
	LN	0.01221	0.02580	0.05075	0.08293	0.13686
	NL	0.01221	0.02576	0.05062	0.08386	0.13894
MAE	RW	0.00881	0.01852	0.03566	0.05529	0.08311
	LN	0.00882	0.01866	0.03702	0.06202	0.10240
	NL	0.00882	0.01865	0.03673	0.06161	0.10046
DIR	RW	0.50460	0.48987	0.50583	0.49233	0.52732
	LN	0.50153	0.49662	0.48312	0.46163	0.45734
	NL	0.50460	0.48987	0.50583	0.49233	0.52732
CW	LN vs. RW	1	1	1	1	1
	NL vs. RW	1	1	1	1	1
	NL vs. LN	0.22935	0.3444	0.33081	0.38529	0.37186

Table 6: **Log stock forecasting:** This table displays mean absolute error MAE, given by $\frac{1}{N-\tau}\sum_{i=1}^{N-\tau}|\epsilon_i(\tau)|$, and root mean squared forecast error RMSE, defined by $\sqrt{\frac{1}{N-\tau}\sum_{i=1}^{N-\tau}\epsilon_i(\tau)^2}$, where $\epsilon_i(\tau) : X_{t_i+\tau} - \mathbb{E}^{\mathbb{P}}_{t_i}\left[X^{\mathcal{M}}_{t_i+\tau}\right]$ and $\tau \in \{1, 5, 22, 66, 131\}$. The random variable $X^{\mathcal{M}}_t$ represents the log stock in the model $\mathcal{M} \in \{\text{NL, LN}\}$ and X_t denotes the recorded value of the logarithm of the S&P 100 at time t. As in the previous table RW denotes the random walk model and LN (resp. NL) stands for the linear (resp. nonlinear) model given in (11) (resp. in (12)). DIR shows the percentage of correct directional forecasts. CW denotes p-values for the Clark and West (2007) test for nested models. Asterisks $^{(***)},^{(**)},^{(*)}$ denote significance at the 1%, 5% and 10% confidence level respectively.