# Empirical Asset Pricing with Nonlinear Risk Premia *

Aleksandar Mijatović† Paul Schneider‡

April 4, 2013

**Abstract**

We introduce a new model for the joint dynamics of the S&P 100 index and the VXO implied volatility index. The nonlinear specification of the variance process is designed to simultaneously accommodate extreme persistence and strong mean reversion. This grants superior forecasting power over the standard (linear) specifications for implied variance forecasting. We obtain statistically significant predictions in an out-of-sample exercise spanning several market crashes starting 1986 and including the recent subprime crisis. The model specification is possible through a simple continuous-time no-arbitrage asset pricing framework that combines semi-analytic pricing with a nonlinear specification for the market price of risk.

## 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 in continuous time 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, 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 this model lies in the nonlinearity of the drift and the diffusion

*We are grateful to Yacine Aït-Sahalia, Valentina Corradi, Damir Filipović, Eric Renault and two anonymous referees for insightful comments.

†Department of Mathematics, Imperial College, London SW7 2AZ, United Kingdom, a.mijatovic@imperial.ac.uk

‡Institute of Finance, University of Lugano, Via Buffi 13, CH-6900 Lugano, paul.schneider@usi.ch

function. 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.

In this paper we develop a multivariate, nonlinear model for the joint time series of the S&P 100 and the VXO implied volatility index. The model is designed to feature the aforementioned traits and turns out to be empirically successful. The specification offers statistically significant advantages in out-of-sample forecasting over extant models in predicting both implied and realized variance over several horizons. Results are particularly pronounced for long-term predictions. 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 (2009). The study adds to the empirics of Bakshi et al. (2006) in that the VXO is modelled in terms of instantaneous variance (and not directly) so that information from the index price dynamics can also be taken into account. Our data ranges from 1986 until 2012 and spans multiple market crashes. In order to demonstrate the robustness to the presence (and absence) of extreme market conditions, we perform two out-of-sample studies using this data set: one including both the crash of 1987 and the extreme market conditions in the years 2008–2010 and the other excluding the two time intervals. Since the prediction results are qualitatively very similar in both cases, this suggests that our specification is robust even to extreme market conditions and that it performs equally well in their absence.

The model is justified theoretically by an econometrically inconspicuous dampening functions which we introduce into the Radon-Nikodym density process. This technique is developed generically for a 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 (2008), Beskos et al. (2006) and Mijatović and Schneider (2010) ensure that reliable parameter inference can be made without explicit formulae for transition densities.

The paper is organized as follows. Section 2 introduces the linear (LN) and nonlinear (NL) GARCH stochastic volatility models as joint models for the S&P 100 and the VXO implied volatility index, which are used for estimation and prediction. Section 3 describes the likelihood function which is used to find the parameter values of the GARCH stochastic volatility models from Section 2 implied by the time series of the S&P 100 and the VXO index data. The empirical results are discussed in Section 4. Appendix A gives a simple theoretical construction (Theorem 1) for the framework we consider. Appendix B describes in detail the steps to evaluate the likelihood function. Section 5 concludes the paper.

## 2 S&P 100 stochastic volatility model

The task in this section is to specify a flexible model for the S&P 100 index. We start by defining the model under the pricing measure $\mathbb{Q}$. The dynamics under the real world measure $\mathbb{P}$ are then obtained by Girsanov's change of measure technique.

The specification of the dynamics of the process under 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 from the affine class and consider a stochastic volatility model based on a GARCH diffusion, which is in the class of polynomial models in the sense of Cuchiero et al. (2010). The existence of the process is not difficult to establish as can be seen below.[1] Furthermore, the approximation to the model-free implied volatility (MFIV), defined in Neuberger (1994), can in our model be analytically computed, using a result in Meddahi and Renault (2004), in terms of the model parameters under $\mathbb{Q}$. This feature of the model is crucial because the goal is to estimate the risk-neutral and the real world parameters simultaneously.

More precisely, 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 SDE

$$
\begin{aligned}
dX_t &= (r - \frac{1}{2}V_t)\,dt + \rho\sqrt{V_t}dW_t^{V\mathbb{Q}} + \sqrt{1-\rho^2}\sqrt{V_t}dW_t^{X\mathbb{Q}}, && (1)\\
dV_t &= (b_0^{\mathbb{Q}} + b_1^{\mathbb{Q}}\,V_t)dt + \sigma V_t\,dW_t^{V\mathbb{Q}}, && (2)
\end{aligned}
$$

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 and the constant $\rho$ lies in the open interval $(-1,1)$. This specification for the instantaneous variance process was introduced in Meddahi and Renault (2004) in a time series context.

Note that the risk-neutral drift $\mu^{\mathbb{Q}}$ (cf. notation in Theorem 1 of Appendix A) of the process that satisfies (1)–(2) is given by

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

It is well known that the SDE in (2) has a solution for all values of $b_0^{\mathbb{Q}}$, $b_1^{\mathbb{Q}}$ and $\sigma$. Assume that

$$
b_0^{\mathbb{Q}} > 0, \tag{3}
$$

[1] It is even possible to obtain a series representations of the transition density, see Wong (1964).

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 (2) 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 (3) we can define the process $X = (X_t)_{t\in[0,T]}$ as a stochastic integral given by (1). 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 (1)–(2). 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 (2) (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 data sets similar to the one considered in this paper (see also the discussion in Heston et al. (2007)). 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 the focus in this paper is the nonlinear drift specification.

Under the physical measure $\mathbb{P}$ we shall consider two kinds of dynamics: one with a linear drift function $\mu_{\mathrm{LN}}^{\mathbb{P}}$ and another with a nonlinear drift function $\mu_{\mathrm{NL}}^{\mathbb{P}}$. The first is given by the formula

$$\mu_{\mathrm{LN}}^{\mathbb{P}}(V_t) := \begin{pmatrix} a_0 + a_1\, V_t \\ b_0^{\mathbb{Q}} + b_1\, V_t \end{pmatrix}. \tag{4}$$

There is no canonical choice for the market price of risk for a GARCH diffusion stochastic volatility model. In Aït-Sahalia and Kimmel (2007), the market price of variance risk is set to zero. Following Jones (2003), we define a linear market price of risk, such that the drift $\mu_{\mathrm{LN}}^{\mathbb{P}}$ can be expressed as $\mu_{\mathrm{LN}}^{\mathbb{P}}(V_t) = f_{\mathrm{LN}}(V_t) + \mu_{\mathrm{LN}}^{\mathbb{Q}}(V_t)$, where the function $f_{\mathrm{LN}}$ is given by

$$f_{\mathrm{LN}}(V_t) = \begin{pmatrix} a_0 - r + (a_1 + \frac{1}{2})V_t \\ (b_1 - b_1^{\mathbb{Q}})\, V_t \end{pmatrix}. \tag{5}$$

Note that the drifts under $\mathbb{Q}$ and $\mathbb{P}$ in the model LN take the same functional form (i.e. they are both linear in the state variable). An easy application of Girsanov's theorem (with the Novikov condition) implies the existence of the real world probability measure $\mathbb{P}$, equivalent to $\mathbb{Q}$, and under which the process $(X,V)$ has the dynamics given by the drift function in (4).

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

$$\mu_{\mathrm{NL}}^{\mathbb{P}}(V_t) := \begin{pmatrix} a_0 + a_1\, V_t \\ b_0 + b_1\, V_t + b_2\, V_t^2 + b_3/V_t \end{pmatrix}. \tag{6}$$

The corresponding function $f_{\mathrm{NL}}$, which captures the difference between the drifts under $\mathbb{P}$ and $\mathbb{Q}$, is given

as

$$f_{\mathrm{NL}}(V_t) = \begin{pmatrix} a_0 - r + (a_1 + \frac{1}{2})V_t \\ b_0 - b_0^{\mathbb{Q}} + (b_1 - b_1^{\mathbb{Q}})\, V_t + b_2\, V_t^2 + b_3\, /\, V_t \end{pmatrix}. \tag{7}$$

In this case, however, a dampening function $D_{\mathrm{NL}}$, defined in (8), is required to guarantee the existence of the process under the real world measure $\mathbb{P}$. The mathematical reason for the inclusion of the dampening factor lies in the fact that without it, it is not possible in general to ensure that the martingale property of the corresponding stochastic exponential holds in our non-linear model. In Theorem 1 of Appendix A we provide a theoretical basis for this purpose in the setting of general non-linear multi-dimensional SDEs. The dampening factor, which depends on the drift and the diffusion functions of the SDE, is constructed to prevent explosions of the density process in (20) under the measrue $\mathbb{Q}$, thereby guaranteeing Novikov's condition. Its key ingredient is a strictly positive tuning constant $c$ (cf. formula (8)), which can be chosen to be arbitrarily small. If the volatility function $\Sigma$ and the drift function $f$ in Theorem 1 are continuous in the state variable $x \in \mathfrak{D}$, as is the case in the proposed nonlinear model (see (1)–(2) and (7)), the constant $c$ can be chosen in such a way that the dampening factor $D$ in Theorem 1 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 this gives the modeler a large amount of freedom when specifying the real world drift function $\mu^{\mathbb{P}}$, since the approximate equality $f + \mu^{\mathbb{Q}} \approx \mu^{\mathbb{P}}$ on large compact subsets of the state space 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 the SDEs under both $\mathbb{P}$ and $\mathbb{Q}$ exists and that the density process (20) gives rise to a valid pricing kernel. In order to empirically justify the equality $f + \mu^{\mathbb{Q}} \approx \mu^{\mathbb{P}}$ used in the estimation, we perform a Monte Carlo study in Section 4.4 below, to validate that the process stays within the compact subset where the equality $f + \mu^{\mathbb{Q}} \approx \mu^{\mathbb{P}}$ holds true with high probability. We would like to stress here that, in any application of the method described here, it is the modeler's task to perform an empirical test, analogous to the one in Section 4.4, which demonstrates the robustness of the method with respect to the choice of the constant $c$ and justifies the claim that $c$ does not need to be estimated from the data.

In the present case the dampening factor takes the following form

$$D_{\mathrm{NL}}(V_t) = \exp\left(-c\left(V_t^2 + 1/(V_t^{3/2}\sigma\sqrt{1-\rho^2})\right)\right), \tag{8}$$

since Theorem 1 of Appendix A can be applied in this instance with the function $g(x,v) = v^2$ (note that $\det\Sigma(x,v) = v^{3/2}\sigma\sqrt{1-\rho^2}$ and hence both coordinates of the function $f_{\mathrm{NL}}(v)$ are bounded above by $A(g(x,v) + 1/|\det\Sigma(x,v)|)$ for all $(x,v)$ and a large constant $A > 0$). As argued in Section 4.4, in the setting of our GARCH stochastic volatility model with a nonlinear real world drift, for numerical purposes and econometric implementation it suffices to work directly with the drifts $\mu^{\mathbb{P}}_{\mathrm{NL}}$ and $\mu^{\mathbb{P}}_{\mathrm{LN}}$ given by the formulae in (4) and (6) respectively. In other words in the setting of our model the dampening factor

$D_{\mathrm{NL}}(V_t)$ can be made arbitrarily close to one on a set of probability $1-\varepsilon$, for any fixed but arbitrarily small $\varepsilon > 0$, through the choice of the positive constant $c$. This intuitive argument is corroborated by the numerical tests carried out in Section 4.4, where the impact of the dampening factor $D_{\mathrm{NL}}(V_t)$ on the likelihood ratio $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is shown to be insignificant for $c \leq 10^{-6}$. The resulting risk premia are non-explosive and satisfy the conditions in Heston et al. (2007).

# 3 Estimation

The goal of this section is to describe the estimation algorithm for the linear and the nonlinear models introduced in Section 2. The data set, the joint time series of the S&P 100 and VXO implied volatility index, used in the estimation procedure will be described in the following section. In this section we define the log-likelihood function (Section 3.1) and give a summary of the estimation procedure used to maximise it (Section 3.2). A key step in this algorithm is the EML procedure form Mijatović and Schneider (2010) (see Appendix B for details), which crucially allows us to reduce the number of parameters that need to be estimated.

## 3.1 Likelihood function

The *instantaneous* stochastic variance is a latent variable even though a time series of *implied* variance is available through the VXO index. Note that the drift of the variance $V$ in our model, given by SDE (2) under the pricing measure $\mathbb{Q}$, is affine. The fact that in this case the price of the variance swap at a current time $t$ is linear in the current value of the instantaneous variance (of the asset return) $V_t$ was first established in Meddahi and Renault (1996) in a time series context under the $\mathbb{P}$ measure. The authors also derived the following formula

$$\frac{1}{\Delta}\,\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\Delta} V_s\,ds\right] = A(\theta^{\mathbb{Q}},\Delta) + B(\theta^{\mathbb{Q}},\Delta)\,V_t, \qquad \Delta > 0, \tag{9}$$

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

$$B(\theta^{\mathbb{Q}},\Delta) = \frac{1}{b_1^{\mathbb{Q}}\Delta}\left(\exp(b_1^{\mathbb{Q}}\Delta) - 1\right), \quad A(\theta^{\mathbb{Q}},\Delta) = -\frac{b_0^{\mathbb{Q}}}{b_1^{\mathbb{Q}}}(1 - B(\theta^{\mathbb{Q}},\Delta)). \tag{10}$$

We define $IV_t$ as the squared VXO index (described in Section 4.1) 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[2]

$$IV_t \approx \frac{1}{\Delta}\,\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\Delta} V_s\,ds\right]. \tag{11}$$

We now exploit the relationship in (11) to express the log-likelihood function for both the linear and the nonlinear model described by the real world drifts given in (4) and (6) respectively and by SDE (1)–(2) under the pricing measure $\mathbb{Q}$. By the Markov 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

$$\ell(X_1, IV_1, \ldots, X_N, IV_N \mid X_0, IV_0, \theta) = \sum_{i=1}^{N} \log p^{IV}(X_i, IV_i \mid X_{i-1}, IV_{i-1}, \theta), \tag{12}$$

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

$$V_t = \frac{IV_t - A(\theta^{\mathbb{Q}}, \tau)}{B(\theta^{\mathbb{Q}}, \tau)}, \tag{13}$$

which follows from (11), implies that we can express the log-likelihood as

$$\sum_{i=1}^{N} \log p^{V}(X_i, V_i \mid X_{i-1}, V_{i-1}, \theta) - N\,\log B(\theta^{\mathbb{Q}}, \tau), \tag{14}$$

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 $Y_t = \log(V_t)/\sigma$ (see also (22)) yields the log-likelihood which takes the form

$$\ell(\theta) = \sum_{i=1}^{N} \{\log p(X_i, Y_i \mid X_{i-1}, Y_{i-1}, \theta) - \sigma Y_i\} - N\,(\log B(\theta^{\mathbb{Q}}, \tau) - \sigma), \tag{15}$$

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

[2]With a continuum of option prices traded, the formula in (11) would be exact for diffusion processes through the Carr and Madan (2001) formula applied to the log-payoff, since, under those assumptions, the log-contract is the forward price of the expected quadratic variation (cf. for example Neuberger (2012, Proposition 5)),

$$-2\mathbb{E}_t^{\mathbb{Q}}\left[\log X_T - \log X_t\right] = \mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T V_s\,ds\right] = 2\int_0^{F_{t,T}} \frac{P_{t,T}(K)}{K^2}dK + 2\int_{F_{t,T}}^{\infty} \frac{C_{t,T}(K)}{K^2}dK,$$

where $F_{t,T}$ is the forward price of $X$ and $P_{t,T}(K)$ (resp. $C_{t,T}(K)$) is the put (resp. call) option struck at $K$ and expiring at $T$. With only finitely many options traded at the exchange, the VXO is only an approximation of the model-free implied variance, which explains the use of $\approx$ sign in equation (11).

| | $\theta^{\sigma}$ | $\theta^{\mathbb{Q}}$ | $\theta^{X\mathbb{P}}$ | $\theta^{V\mathbb{P}}$ |
|---|---|---|---|---|
| Linear Spec. (4) | $\sigma, \rho, b_0^{\mathbb{Q}}$ | $b_1^{\mathbb{Q}}$ | $a_0, a_1$ | $b_1$ |
| Nonlinear Spec. (6) | $\sigma, \rho$ | $b_0^{\mathbb{Q}}, b_1^{\mathbb{Q}}$ | $a_0, a_1$ | $b_0, b_1, b_2, b_3$ |

Table 1: **Parameter sets for the linear and the nonlinear model:** the table displays the partition of the parameter vector $\theta$ into the following groups: $\theta^{\sigma}$ (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 $\mathbb{P}$).

### 3.2 Summary of the estimation algorithm

For estimation purposes we partition the parameter vector $\theta$ (of the models given by SDE (2) under $\mathbb{Q}$ and with either linear (4) or nonlinear (6) drifts under $\mathbb{P}$) into four classes. The first class $\theta^{\sigma}$ 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.

Our estimation algorithm uses likelihood-based inference. The separation of the parameter vector $\theta$ described above, will enable us to reduce the estimation problem for $\theta$ to estimating only a subset of parameters $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$. The key idea is to treat the optimal parameter values for $\theta^{X\mathbb{P}}$ and $\theta^{V\mathbb{P}}$ as a function of (i.e. conditional on) $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$, via a non-analytic and yet computationally efficient expected maximum likelihood (EML) algorithm from Mijatović and Schneider (2010). In particular, this will enable us to perform a standard likelihood search over the reduced parameter space $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$, as for each proposed parameter vector in $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$ we have, through the EML algorithm, the optimal $\theta^{X\mathbb{P}\star} \cup \theta^{V\mathbb{P}\star} \mid \theta^{\sigma} \cup \theta^{\mathbb{Q}}$ (optimal parameters are denoted with a superscript $^{\star}$). It is important to note that our estimation algorithm is *not* an iterative procedure which alternates between estimating in turn $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$ and $\theta^{X\mathbb{P}} \cup \theta^{V\mathbb{P}}$, as was for example proposed in Pastorello et al. (2003) or Song et al. (2005). On the contrary, our approach is a standard likelihood search using the *concentrated likelihood*.[3]

The EML estimation step makes use of the fact that the law of the Brownian bridge and the law of a

[3]Expressed in the notation of Fan et al. (2006, Introduction), we are interested in estimating a parameter $\xi$, which can be naturally partitioned into two sub-vectors $\theta$ and $\nu$ as $\xi = (\theta, \nu)$. The relation to the parameter names in the present paper is as follows: their $\theta$ is our $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$ and their $\nu$ is our $\theta^{X\mathbb{P}} \cup \theta^{V\mathbb{P}}$. If $Q(\theta, \nu)$ is the sample objective function, then, given any $\theta$, one can find the optimal value of $\nu$ as a function of $\theta$ by solving

$$\nu(\theta) = \arg\max_{\nu} Q(\theta, \nu).$$

In our case this step is achieved by the EML algorithm. The *concentrated objective function* (or profile objective function) is then defined as a function of $\theta$ alone, $Q(\theta, \nu(\theta))$, and its maximisation with respect to $\theta$ gives the estimator of interest:

$$\theta^{\star} = \arg\max_{\theta} Q(\theta, \nu(\theta)) \Rightarrow [\theta^{\star}, \nu(\theta^{\star})] = \arg\max_{\theta, \nu} Q(\theta, \nu)$$

diffusion bridge are, in a certain precise sense, close to each other, making the Euler scheme approximation for the transition densities also close to each other, when the time interval between observations is small. The EML algorithm itself cannot be directly applied to the present econometric problem, as this approximation in law works only for one-dimensional diffusions. In the present paper we show how to generalise the EML algorithm to a multivariate context in the special case where the variance process is Markovian in its own right (i.e. its future law does not depend on the current level of the asset return process).

The concentrated likelihood function is evaluated in three steps. In the *first* step, we convert (for a given $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$), using formula (13), the VXO time series to the time series for the latent instantaneous variance process $V$. Note that this step is the same for the linear as well as the nonlinear model, since under $\mathbb{Q}$ the variance process has an affine drift in both cases. In the *second* step, based on this time series, we perform the EML algorithm, yielding the optimal parameters $\theta^{V\mathbb{P}\star}$ as a function of $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$. This step involves simulating from a Brownian bridge and the solution of a least squares problem for a quadratic form. The precise description of this step is given in Appendix B.2. In the same step, conditioning on the same values $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$ and the optimal $\theta^{V\mathbb{P}\star}$ obtained in the second step, we perform another EML algorithm to obtain $\theta^{X\mathbb{P}\star}$ as a function $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$ and $\theta^{V\mathbb{P}\star}$. In its implementation, this step is completely analogous to step two and the details are described in Appendix B.3. The computational efficiency of the algorithm is based on the fact that EML can be used to express the globally optimal drift parameters $\theta^{V\mathbb{P}\star}$ and $\theta^{X\mathbb{P}\star}$ as complicated, yet easily computable functions of the parameters $\theta^{\sigma} \cup \theta^{\mathbb{Q}}$ and the data. The concentrated likelihood function is then given directly in the *third* step via the simulated likelihood approximation from Pedersen (1995).

We conclude this section by remarking that we compared the estimation based on the likelihood function described above with the likelihood expansion of Aït-Sahalia (2008). We found that the latter is numerically much harder to handle but qualitatively they both deliver similar results.

# 4 Empirical results

The aim of this section is to assess the performance of the linear and nonlinear models introduced in Section 2 by investigating forecasts of realized variance, implied variance and stock returns for various maturities, based on the joint time series of the S&P 100 and VXO implied volatility index. We start by describing the data set (Section 4.1). The forecasting exercise is then performed out of sample for the entire data set, and also for a subsample excluding the 1987 crash and the credit crisis starting from 2007. For the out-of-sample period the model is re-estimated each time a new data point is added. Figure 1 gives a visual impression of the data sample. We then perform a statistical test given in Clark and West (2007) on the estimated models with respect to their forecasting abilities. This test corrects for the upward bias in mean squared errors (MSE) which occurs with more flexibly parameterized models. The test is designed for nested models and allows direct comparison.

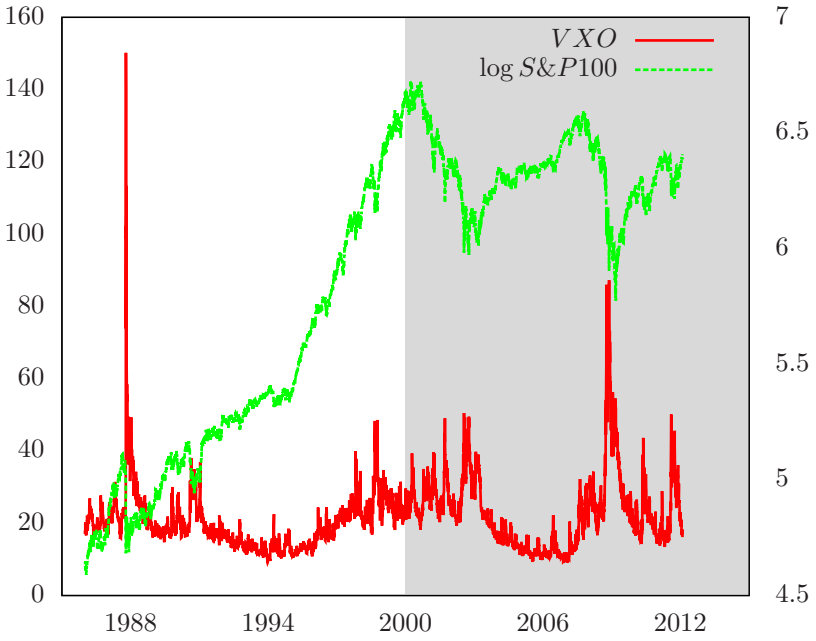



Figure 1: **Log of S&P 100 index and VXO:** The figure shows the evolution of the logarithm of the S&P 100 index (right y-axis) and of the implied volatility index VXO (left y-axis in %). For our out-of-sample analysis we use a burn-in period (shaded in white), and the forecasting is then done within the sample shaded in grey.

### 4.1 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 carefully 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[4] enables them to obtain a time series of (averaged over strikes) implied volatility, highly correlated with model-free implied volatility (VIX). We use the VXO instead of the VIX index to include the crash in 1987. The data set ranges from 2 January 1986 until 3 February 2012. The data is obtained from Bloomberg. The market crashes in 1987 and the subprime crisis pose extreme scenarios, challenging the specifications tested in this paper. 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 subsets. For our *out-of-sample* analysis we use a burn-in period from 2 January 1986 until 11 November 1999 and then the remaining sample for forecasting. For robustness we also use a second sub sample ranging from January 1988 until January 2007, excluding the big crashes in 1987 and the credit crisis starting from 2008.

[4]Originally designed by Whaley (1993).

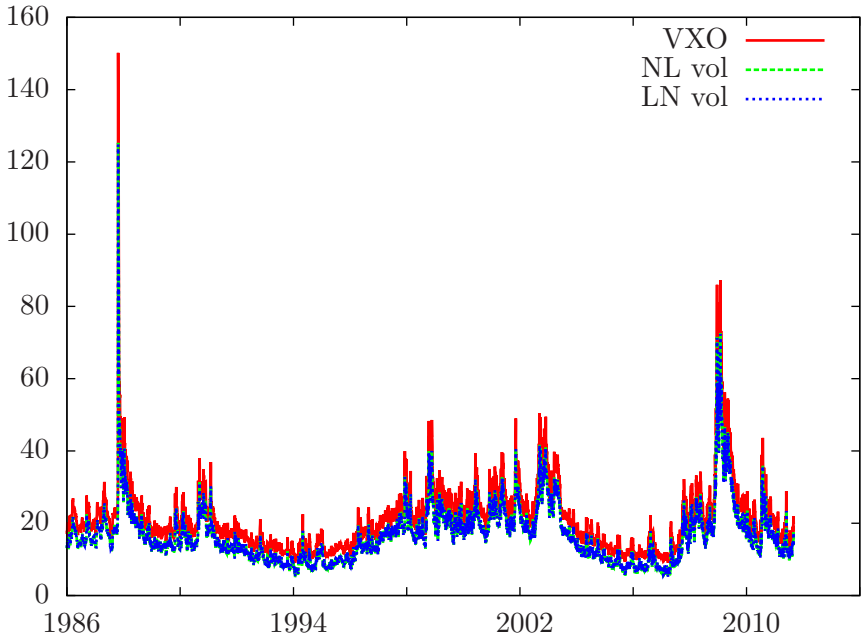


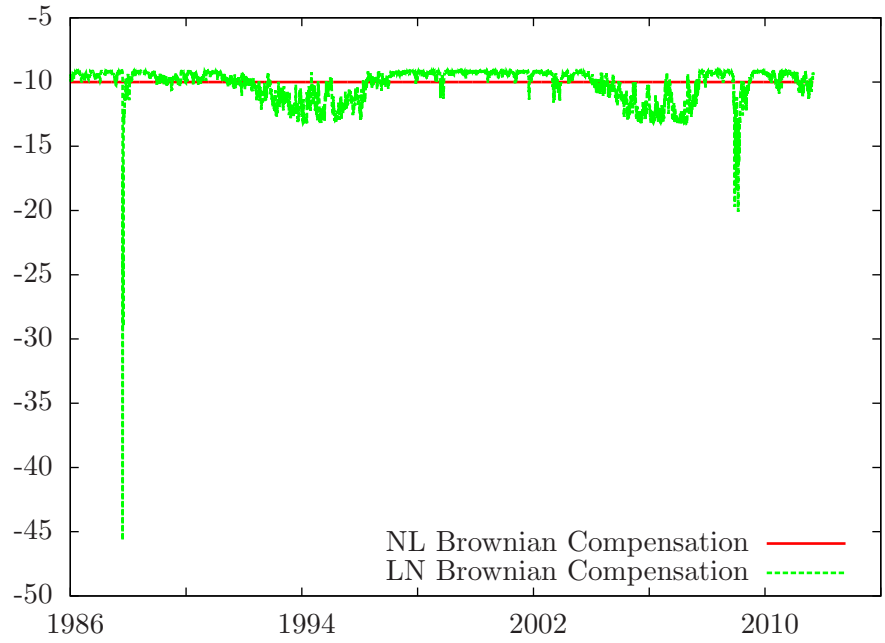

Figure 2: **VXO index and Implied Instantaneous S&P 100 Volatility:** this figure displays the VXO implied volatility index along with the instantaneous variance implied by the $\mathbb{Q}$ parameters from Table 2. The data are reported in percent.

### 4.2 Risk premia

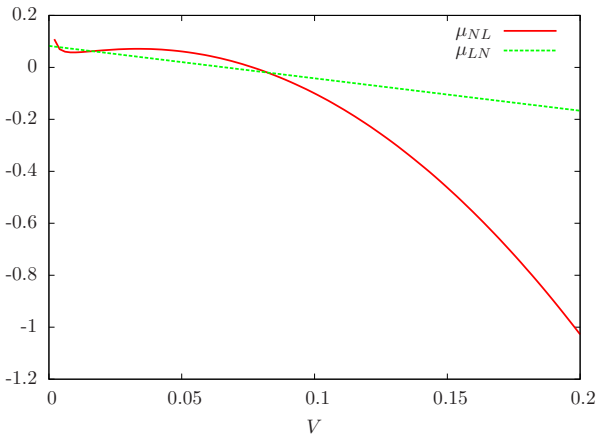
Conventional stationarity tests are based on linear model assumptions and financial time series regularly fail these tests contrary economic intuition. The VXO time series is no exception: Depending on the sub sample used augmented Dickey-Fuller tests sometimes report a unit root, and sometimes stationarity. The nonlinear model introduced in this paper is potentially capable of reflecting mean-diverting behaviour in certain regions while displaying strong pull-back to the unconditional mean for very high or very low variance regimes. In this section we investigate this question and also consider implications for risk premia induced through nonlinearities. We start by estimating the parameters, the procedure of which is described in detail in Appendix B. Point estimates and standard errors for the parameters of the nonlinear model NL (cf. (24)) and the linear model LN (cf. (23)) can be found in Table 2.

The $\mathbb{Q}$ mean-reversion parameter $b_1^{\mathbb{Q}}$ is large and positive for both 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 (2009), for the VIX index.[5] The positive estimates result in a time series for instantaneous variance that is located consistently below the time series of the VXO through relation (11). This can be seen in Figure 2 to be the case for both the linear and the nonlinear models under the real world measure $\mathbb{P}$. Since both models are linear under the risk-neutral measure $\mathbb{Q}$ the implied variance series are virtually indistinguishable. The correlation and diffusion parameters $\rho$ and

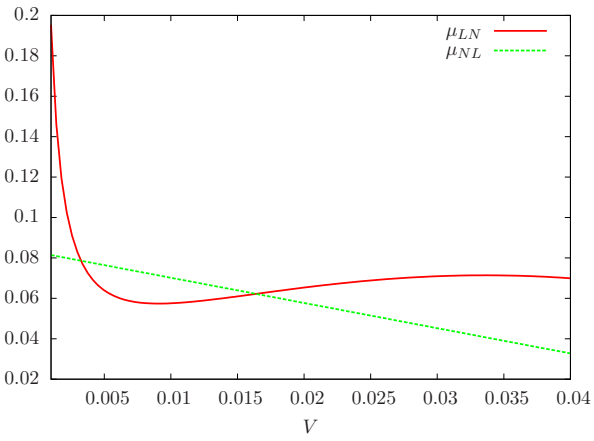
[5]The VIX and the VXO indices are very highly correlated and very similar in level.




(a) Variance premia




(b) Estimated drift




(c) Estimated drift (Zoom)

Figure 3: **Variance premia and the nonlinear Drift:** Risk premia for the variance (see Section 4.2 for the definition) are displayed in Figure 3a. Figure 3b and 3c display the nonlinear drift in variance model (6), and the linear drift in model (4) (both at the estimated parameter values).

| | $\sigma$ | $\rho$ | $b_0^{\mathbb{Q}}$ | $b_1^{\mathbb{Q}}$ | $a_0$ | $a_1$ | $b_0$ | $b_1$ | $b_2$ | $b_3$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Linear | 2.4196 | -0.7146 | 0.0778 | 8.3141 | 0.0296 | 0.9384 | | -1.6904 | | |
| | (0.1066) | (0.0108) | (0.0081) | (0.0521) | (0.0248) | (0.2543) | | (0.0105) | | |
| Nonlinear | 2.4407 | -0.7137 | 0.0837 | 8.1319 | 0.0421 | 0.6438 | 0.0351 | 1.1075 | -24.5954 | 9.704e-05 |
| | (0.0908) | (0.0069) | (0.0082) | (0.0097) | (0.0220) | (0.0069) | (0.0142) | (0.0183) | (0.0040) | (8.769e-05) |

Table 2: **Parameter Estimates:** The table displays parameter estimates for the linear model (23) and the nonlinear model (24). Huber (sandwich) standard errors are computed from the asymptotic covariance matrix pertaining to the likelihood in (15) with the transition density approximation in in (30). 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).

$\sigma$ as well as $b_0^{\mathbb{Q}}$ are also comparable in scale for both specifications.

Under the physical measure $\mathbb{P}$, however, the linear and the nonlinear specifications predict very different behaviour. Figure 3c shows that during calm times (in the region between 0.01 and 0.04) the nonlinear specification predicts that the instantaneous variance behaves like a random walk or a process that mean-diverts at an even faster rate. Figure 3b 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 generated with a linear drift specification. The drift function estimated from the time series of the VIX index non-parametrically in Bandi and Renó (2009) and Bakshi et al. (2006) is of a shape similar to that of the drift function in Figure 3b. This suggests that the nonlinear drift specification is flexible enough to accommodate the necessary functional form.

Next we investigate the nonlinear specification through the lense of risk premia. Specifically we look at the compensation for Brownian motion risk driving the variance state variable. Recall that the *risk premium* (i.e. the *market price of risk*) at time $t \in [0, T]$ in the model $\mathcal{M} \in \{\text{LN}, \text{NL}\}$ is given by

$$\Lambda_{\mathcal{M}}(V_t) \quad = \quad \Sigma(V_t)^{-1} f_{\mathcal{M}}(V_t), \qquad \text{where} \quad \Sigma(V_t) = \sqrt{V_t} \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & \sigma\sqrt{V_t} \end{pmatrix}, \tag{16}$$

and $f_{\mathcal{M}}$ is defined in (5) and (7) 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 $\Lambda_{\mathcal{M}}(V_t)$. 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 (2)) is similar for both specifications, the nonlinear model exhibits time-variability in the market prices of risk (see Figure 3a), in contrast to the constant risk premium, given by $(b_1 - b_1^{\mathbb{Q}})/\sigma$, in the linear model. Consistent with the literature on variance risk premia measured through variance swaps, the instantaneous compensation from Brownian motion risk is, on average, large and negative.

### 4.3 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 out-of-sample with predictions for implied variance, realized variance and stock returns. 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

$$RV_i(N) := \frac{262}{N} \sum_{j=i-N}^{i} \left( \frac{X_j - X_{j-1}}{X_{j-1}} \right)^2. \tag{17}$$

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

$$RV_i^{\mathcal{M}}(N) := \frac{1}{N} \sum_{j=i-N}^{i} V_j. \tag{18}$$

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.[6] Tables 3, 4 and 5 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) (CW) test statistics for nested models are reported. Table 3 reports normalized MSE (NMSE) for comparison with the results in Sizova (2008).

Figure 2 indicates structural breaks in the implied variance time series. The sample period spans these regimes, and the out-of-sample period contains both very rough and very calm periods. The visual impression suggests exactly what the nonlinear model is designed to accomplish: Very high volatility states persist only very briefly, the negative pull in the drift growing quadratically and directing the process down to smaller levels. An analogous phenomenon occurs for very low volatility states. During normal market periods the process itself becomes very persistent. We therefore expect the nonlinear model to do at least as good as the RW model. The effect of mean reversion, which the LN model also accommodates is likely to show advantages.

With realized variance being path-dependent we would also expect the effect of mean reversion to amplify the advantage of the LN and the NL model over the RW model. Table 3, which contains the

[6] With approximation (11) 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).

| OUT-SAMPLE (short) | | 1w | 4w | 12w | 26w |
|---|---|---|---|---|---|
| RMSE | RW | 0.0622 | 0.0681 | 0.0711 | 0.0729 |
| | LN | 0.0326 | 0.0242 | 0.0226 | 0.0231 |
| | NL | 0.0334 | 0.0258 | 0.0239 | 0.0240 |
| NMSE | RW | 186% | 351% | 543% | 718% |
| | LN | 51% | 44% | 55% | 72% |
| | NL | 53% | 50% | 61% | 78% |
| MAE | RW | 0.0304 | 0.0329 | 0.0348 | 0.0359 |
| | LN | 0.0163 | 0.0135 | 0.0144 | 0.0162 |
| | NL | 0.0165 | 0.0139 | 0.0143 | 0.0165 |
| DIR | RW | 49% | 47% | 49% | 48% |
| | LN | 70% | 70% | 65% | 60% |
| | NL | 70% | 70% | 64% | 59% |
| CW | LN vs. RW | $0.0000^{(***)}$ | $0.0000^{(***)}$ | $0.0001^{(***)}$ | $0.0012^{(***)}$ |
| | NL vs. RW | $0.0000^{(***)}$ | $0.0000^{(***)}$ | $0.0001^{(***)}$ | $0.0014^{(***)}$ |
| | NL vs. LN | 0.9926 | 0.8100 | 0.2386 | 0.1118 |
| OUT-SAMPLE | | | | | |
| RMSE | RW | 0.1157 | 0.1265 | 0.1373 | 0.1426 |
| | LN | 0.0603 | 0.0574 | 0.0593 | 0.0547 |
| | NL | 0.0611 | 0.0591 | 0.0600 | 0.0547 |
| NMSE | RW | 160% | 242% | 379% | 582% |
| | LN | 43% | 49% | 70% | 85% |
| | NL | 44% | 52% | 72% | 85% |
| MAE | RW | 0.0444 | 0.0490 | 0.0534 | 0.0564 |
| | LN | 0.0243 | 0.0224 | 0.0250 | 0.0275 |
| | NL | 0.0245 | 0.0229 | 0.0250 | 0.0267 |
| DIR | RW | 49% | 48% | 47% | 47% |
| | LN | 70% | 68% | 61% | 60% |
| | NL | 70% | 68% | 62% | 60% |
| CW | LN vs. RW | $0.0003^{(***)}$ | $0.0082^{(***)}$ | $0.0386^{(**)}$ | $0.0578^{(*)}$ |
| | NL vs. RW | $0.0003^{(***)}$ | $0.0081^{(***)}$ | $0.0396^{(**)}$ | $0.0594^{(*)}$ |
| | NL vs. LN | 0.9830 | 0.8761 | 0.7033 | $0.0910^{(*)}$ |

Table 3: **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 (17) and the random variable $RV_i^{\mathcal{M}}(\tau)$ is given in (18) 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 (23) (resp. in (24)). 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. The top panel is based on the joint S&P 100 and VXO sample excluding the '87 crash and the credit crisis '08-. The bottom panel is based on the entire sample ranging from '86 until '12.

| OUT-SAMPLE (short) | | 1d | 1w | 4w | 12w | 26w |
|---|---|---|---|---|---|---|
| MSE | RW | 0.0086 | 0.0169 | 0.0266 | 0.0367 | 0.0421 |
| | LN | 0.0086 | 0.0168 | 0.0261 | 0.0356 | 0.0426 |
| | NL | 0.0085 | 0.0162 | 0.0256 | 0.0348 | 0.0406 |
| MAE | RW | 0.0048 | 0.0095 | 0.0157 | 0.0229 | 0.0264 |
| | LN | 0.0048 | 0.0095 | 0.0163 | 0.0257 | 0.0337 |
| | NL | 0.0048 | 0.0092 | 0.0153 | 0.0234 | 0.0300 |
| DIR | RW | 48% | 48% | 46% | 45% | 45% |
| | LN | 49% | 51% | 53% | 55% | 52% |
| | NL | 50% | 53% | 56% | 52% | 52% |
| CW | LN vs. RW | $0.0027^{(***)}$ | $0.0001^{(***)}$ | $0.0001^{(***)}$ | $0.0022^{(***)}$ | $0.0407^{(**)}$ |
| | NL vs. RW | $0.0196^{(**)}$ | $0.0082^{(***)}$ | $0.0055^{(***)}$ | $0.0108^{(**)}$ | $0.0218^{(**)}$ |
| | NL vs. LN | $0.0266^{(**)}$ | $0.0125^{(**)}$ | $0.0116^{(**)}$ | $0.0285^{(**)}$ | $0.0500^{(**)}$ |
| OUT-SAMPLE | | | | | | |
| MSE | RW | 0.0184 | 0.0311 | 0.0495 | 0.0747 | 0.0902 |
| | LN | 0.0184 | 0.0307 | 0.0471 | 0.0656 | 0.0722 |
| | NL | 0.0182 | 0.0298 | 0.0460 | 0.0629 | 0.0700 |
| MAE | RW | 0.0073 | 0.0135 | 0.0226 | 0.0358 | 0.0458 |
| | LN | 0.0073 | 0.0134 | 0.0223 | 0.0339 | 0.0419 |
| | NL | 0.0073 | 0.0132 | 0.0218 | 0.0319 | 0.0386 |
| DIR | RW | 0.4754 | 0.4706 | 0.4442 | 0.4241 | 0.4327 |
| | LN | 0.5188 | 0.5306 | 0.5547 | 0.5665 | 0.6045 |
| | NL | 0.4994 | 0.5148 | 0.5425 | 0.5719 | 0.6031 |
| CW | LN vs. RW | 0.1134 | $0.0504^{(*)}$ | $0.0371^{(**)}$ | $0.0369^{(**)}$ | $0.0642^{(*)}$ |
| | NL vs. RW | 0.1210 | $0.0495^{(**)}$ | $0.0506^{(*)}$ | $0.0477^{(**)}$ | $0.0717^{(*)}$ |
| | NL vs. LN | 0.1449 | $0.0754^{(*)}$ | $0.0986^{(*)}$ | $0.0709^{(*)}$ | $0.0903^{(*)}$ |

Table 4: **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 (13), 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 (23) (resp. in (24)). 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. The top panel is based on the joint S&P 100 and VXO sample excluding the '87 crash and the credit crisis '08-. The bottom panel is based on the entire sample ranging from '86 until '12.

| OUT-SAMPLE (short) | | 1d | 1w | 4w | 12w | 26w |
|---|---|---|---|---|---|---|
| MSE | RW | 0.0115 | 0.0240 | 0.0457 | 0.0719 | 0.1067 |
| | LN | 0.0115 | 0.0243 | 0.0475 | 0.0804 | 0.1315 |
| | NL | 0.0115 | 0.0242 | 0.0476 | 0.0823 | 0.1366 |
| MAE | RW | 0.0082 | 0.01744 | 0.0336 | 0.0553 | 0.0835 |
| | LN | 0.0082 | 0.0175 | 0.0343 | 0.0604 | 0.0992 |
| | NL | 0.0082 | 0.0175 | 0.0342 | 0.0609 | 0.1004 |
| DIR | RW | 51% | 51% | 54% | 51% | 52% |
| | LN | 50% | 51% | 51% | 47% | 46% |
| | NL | 51% | 51% | 54% | 51% | 52% |
| CW | LN vs. RW | 0.9691 | 0.9825 | 0.9826 | 0.9699 | 0.9617 |
| CW | NL vs. RW | 0.6856 | 0.7710 | 0.8225 | 0.9167 | 0.9214 |
| CW | NL vs. LN | 0.2449 | 0.2659 | 0.3010 | 0.5275 | 0.5805 |
| OUT-SAMPLE | | | | | | |
| MSE | RW | 0.0137 | 0.0274 | 0.0521 | 0.0868 | 0.1395 |
| | LN | 0.0137 | 0.0276 | 0.0536 | 0.0951 | 0.1586 |
| | NL | 0.0137 | 0.0276 | 0.0534 | 0.0930 | 0.1560 |
| MAE | RW | 0.0093 | 0.0194 | 0.0381 | 0.0647 | 0.1024 |
| | LN | 0.0093 | 0.0194 | 0.0385 | 0.0688 | 0.1145 |
| | NL | 0.0093 | 0.0194 | 0.0384 | 0.0678 | 0.1120 |
| DIR | RW | 52% | 54% | 58% | 59% | 60% |
| | LN | 49% | 50% | 52% | 52% | 59% |
| | NL | 50% | 52% | 52% | 56% | 60% |
| CW | LN vs. RW | 0.4136 | 0.7086 | 0.8686 | 0.9166 | 0.8658 |
| CW | NL vs. RW | 0.8145 | 0.9277 | 0.9318 | 0.9264 | 0.8880 |
| CW | NL vs. LN | 0.7641 | 0.4299 | 0.1299 | 0.1190 | $0.0909^{(*)}$ |

Table 5: **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 $\mathcal{M} \in \{\text{NL}, \text{LN}\}$ 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 (23) (resp. in (24)). 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. The top panel is based on the joint S&P 100 and VXO sample excluding the '87 crash and the credit crisis '08-. The bottom panel is based on the entire sample ranging from '86 until '12.

results for realized variance forecasting, indeed demonstrates that the RW model is dominated both by the LN and the NL models. The directional forecasts ranging from 60%-70% are encouraging for both the LN and NL models, recommending their use in connection with a variance trading strategy.[7] The NMSE statistic given in Table 3 is comparable to the one reported in Sizova (2008) even though the forecasts in Sizova (2008) are based on a single time horizon. The numbers suggest that both the LN and the NL fare well with her SV-CJ jump diffusion model for the VIX index for short forecast maturities. The top panel in Table 3 indicates that the extreme returns in market crashes have no immediate effect on the predictability of realized variance through the LN or the NL model.

The results for implied variance forecasting in Table 4 are more clear-cut. The RW forecast cannot compete with either the LN, or the NL model. In this case also the pattern emerges that NL gives significant advantages over LN. Except for the 1 day forecast the NL outperforms the LN model uniformly with respect to MAE, MSE, and in particular the CW criterion across all maturities. In the shorter sample excluding the big crises, the effects are even more pronounced. There is evidence for predictability on the one-day horizon, where the NL model outperforms the LN model, showing that a nonlinear drift can have impact over short time horizons. Directional forecasts also exhibit an interesting pattern. The quality of the LN and NL forecasts improve with the forecasting horizon, while RW forecasts deteriorate. This is a strong indication of reversion to an unconditional mean and thus hints at the weaknesses in unit root tests, which the VXO and VIX indices regularly fail depending on the sub sample under consideration.

The results in Bollerslev et al. (2009) suggest that variance risk premia, measured by the difference between S&P 500 realized variance and the VIX index, have predictive power for S&P 500 returns. One source for this predictability may well be the pronounced negative correlation between S&P 500 returns and their variance, i.e. the leverage effect. With explicit variance models, we can re-assess this question through the lens of the VXO. Inspecting Table 5, we find no evidence for predictability through any of the models. In particular, neither the NL or the LN model have anything to add over the RW model.

Consolidating the results from realized variance, implied variance, and index return forecasting, the nonlinear model suggests itself as an alternative to the prevailing purely affine models used in the literature. Its specification is suited for the peculiar behaviour of index variance, which exhibits strong mean reversion and extreme persistence at the same time, and its suitability carries over to predicting instantaneous variance-related quantities such as realized variance and implied variance. Overall the prediction power of the nonlinear model is not worse than that of the LN and RW models in the case of the realized variance and index returns. However, the prediction of the VXO implied variance in the nonlinear model outperforms significantly the LN and RW models out-of-sample.

[7] With VIX futures traded very liquidly, the forecasts could be used as buy or sell signals.

### 4.4 How significant is the dampening factor $D_{\mathbf{NL}}(V_t)$?

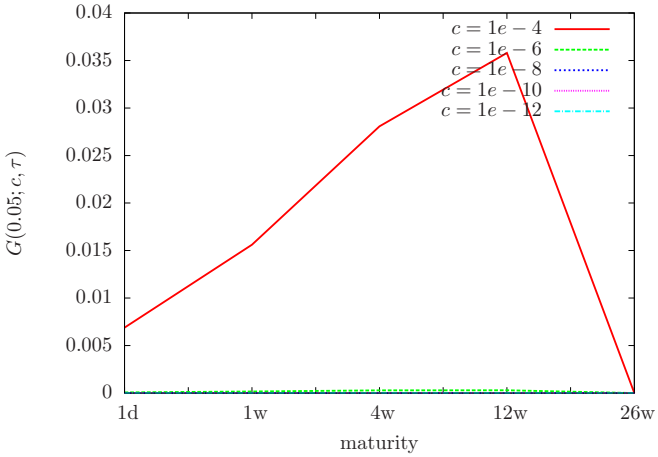
In the present section we investigate how much the dampening factor $D_{\mathrm{NL}}(V_t)$ affects the likelihood ratio $\eta$ from eq. (20). In the estimations the dampening factor was treated with a value identically equal to one. While this is of no concern for predicting out of sample, a justification for this choice is nevertheless important. Using the full-sample parameter estimates reported in Table 2, we evaluate the likelihood ratio for different time horizons and for different choices of $c$ and compare quantiles of the distributions.

Specifically we simulate the system (1)-(2) under the $\mathbb{Q}$ ($\mathbb{P}$) measure, using high-frequency data augmentation to reduce the discretization bias, to evaluate the stochastic integrals in (20) as accurately as possible. We generate 5,000,000 sample paths and estimate the probability distribution function $\mathbb{Q}(\eta_\tau(c) \leq x)$ ($\mathbb{P}(\eta_\tau(c) \leq x)$), where we choose $x \in \{0.05, 0.5, 0.95\}$, $c \in \left\{10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}, 10^{-12}\right\}$, and compare the quantiles of the parameterized distributions.

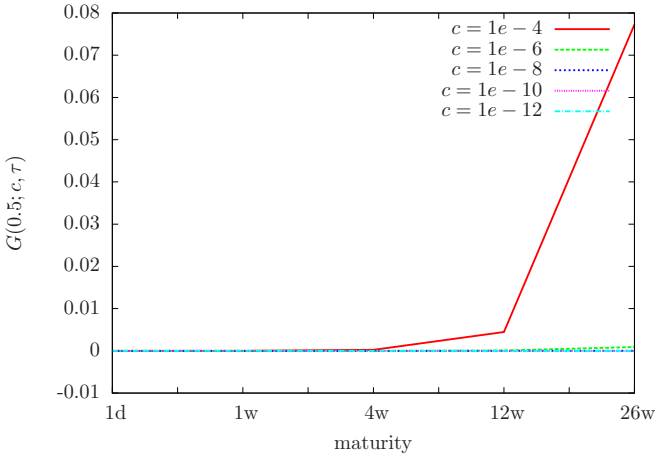
The results of this numerical exercise are consolidated into figures 4a, 4b, and 4c for the $\mathbb{Q}$ simulation and figures 5a, 5b, and 5c for the $\mathbb{P}$ simulation. Under both measures, the distribution of the likelihood ratio shows very little sensitivity to the dampening parameter for $c$ small enough. While the percentage difference between $\mathbb{Q}(\eta_\tau(10^{-4}) \leq x)$ and $\mathbb{Q}(\eta_\tau(0) \leq x)$ ($\mathbb{P}(\eta_\tau(10^{-4}) \leq x)$ and $\mathbb{P}(\eta_\tau(0) \leq x)$) grows with the time horizon, for $c \leq 10^{-6}$, there is virtually no difference between the quantiles. This numerical exercise demonstrates that for the nonlinear model used in this paper the procedure is robust to the choice of $c \leq 10^{-6}$. Put differently, the observed path of the process lies well within the compact sets that correspond to $c \leq 10^{-6}$ and the model specification given by (1)–(2) and (7).

## 5 Conclusion

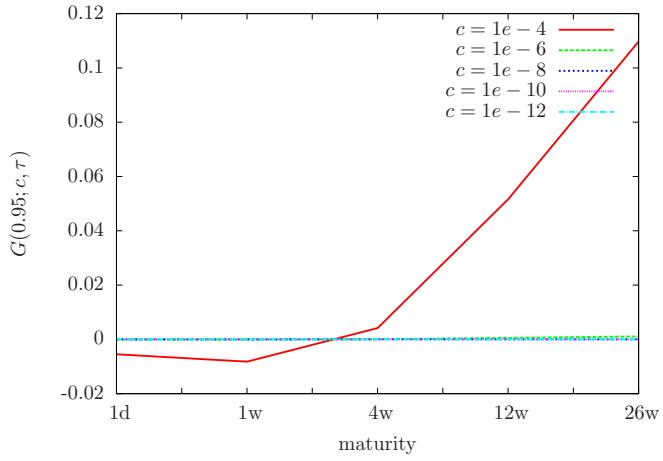
We introduce a simple continuous-time diffusion framework that combines semi-analytic pricing formulae with flexible nonlinear time series modeling. Using an econometrically inconspicuous dampening function we ensure that a solution to the nonlinear stochastic differential equation under the physical 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. Out-of-sample forecast tests show that the nonlinear model has superior forecasting power over the random walk and the linear model in particular for long prediction horizons in predicting the VXO. This suggests that a nonlinear specification of the drift under the physical measure could potentially be very useful in trading and risk management.




(a) Low Quantile

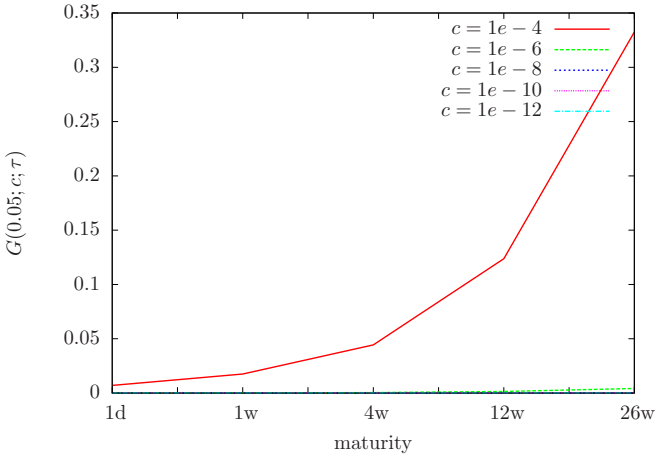



(b) Mean

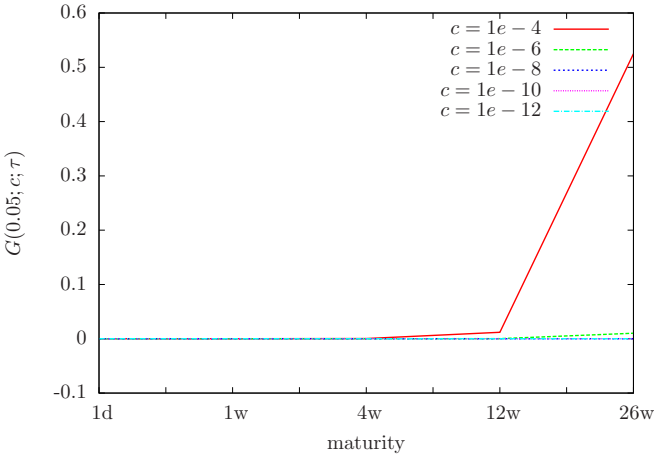



(c) High Quantile

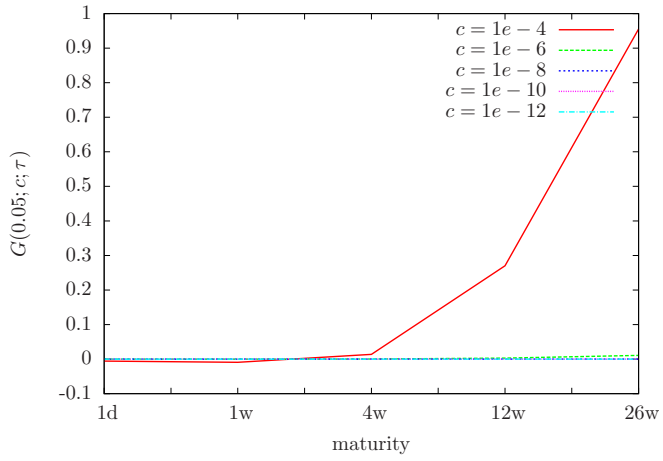
Figure 4: **Sensitivity of Likelihood Ratio $\eta$ to the Dampening Factor:** Denoting by $F(x; c, \tau) := \mathbb{Q}(\eta_\tau(c) \leq x)$, where the likelihood ratio $\eta_\tau(c) = \frac{d\mathbb{P}}{d\mathbb{Q}}(c)$ is defined in eq. (20), the panels display $G(p; c, \tau) := \log\left(F^{-1}(p; c, \tau)/F^{-1}(p; 0, \tau)\right)$. The quantiles are computed from Monte Carlo integrating the likelihood ratio using high-frequency data augmentation with 5,000,000 sample paths. The time horizon is chosen to match exactly the times used in the forecasting exercise.




(a) Low Quantile




(b) Mean




(c) High Quantile

Figure 5: **Sensitivity of Likelihood Ratio $\eta$ to the Dampening Factor:** Denoting by $F(x; c, \tau) := \mathbb{P}\left(\eta_\tau(c) \leq x\right)$, where the likelihood ratio $\eta_\tau(c) = \frac{d\mathbb{P}}{d\mathbb{Q}}(c)$ is defined in eq. (20), the panels display $G(p; c, \tau) := \log\left(F^{-1}(p; c, \tau)/F^{-1}(p; 0, \tau)\right)$. The quantiles are computed from Monte Carlo integrating the likelihood ratio using high-frequency data augmentation with 5,000,000 sample paths. The time horizon is chosen to match exactly the times used in the forecasting exercise.

## References


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

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36(4):906–937.

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

Bakshi, G., Ju, N., and Ou-Yang, H. (2006). Estimation of continuous-time models with an application to equity volatility dynamics. *Journal of Financial Economics*, 82:227 – 249.

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)*, 68:725–382.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22:4463–4492.

Carr, P. and Madan, D. (2001). Towards a theory of volatility trading. In Jouini, E., Cvitanic, J., and Musiela, M., editors, *Option Pricing, Interest Rates and Risk Management*, Handbooks in Mathematical Finance, pages 458–476. Cambridge University Press.

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22:1311–1341.

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.

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

Cuchiero, C., Teichmann, J., and Keller-Ressel, M. (2010). Polynomial processes and their application to mathematical finance. *Finance & Stochastics*. forthcoming.

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.

Fan, Y., Pastorello, S., and Renault, E. (2006). Maximization by parts in extremum estimation. Working paper, Vanderbilt University, Universitá di Bologna, and UNC.

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.

Heston, S., Lowenstein, M., and Willard, G. (2007). Options and bubbles. *Review of Financial Studies*, 20(2):359–390.

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.

Meddahi, N. and Renault, E. (1996). Aggregations and marginalization of garch and stochastic volatility models. Working paper, University of Montreal and CREST-INSEE.

Meddahi, N. and Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics*, 119:355–379.

Mijatović, A. and Schneider, P. (2010). Globally optimal parameters for non-linear diffusions. *Annals of Statistics*, 38(1):215–245.

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

Neuberger, A. (1994). The log contract. *Journal of Portfolio Management*, 20(2):74–80.

Neuberger, A. (2012). Realized skewness. *Review of Financial Studies*.

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

Pastorello, S., Patilea, V., and Renault, E. (2003). Iterative and recursive estimation in structural non-adaptive models. *Journal of Business & Economic Statistics*, 21:449–509.

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.

Song, P. X.-K., Fan, Y., and Kalbfleisch, J. D. (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, 100:1145–1167.

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.

Whaley, R. E. (1993). Derivatives on market volatility: Hedging tools long overdue. *Journal of Derivatives*, 1:71–84.

Wong, E. (1964). The construction of a class of stationary markoff processes. *Stochastic Processes in Mathematical Physics and Engineering*, pages 264–276.

# A Equivalence of the risk-neutral and real world measures in the models of Section 2

In this section we describe the theoretical basis for the modelling framework used in this paper. Theorem 1 below allows us to define a 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.

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

$$dX_t = \mu^{\mathbb{Q}}(X_t)\,dt + \Sigma(X_t)\,dW_t^{\mathbb{Q}}, \qquad X_0 = x_0 \in \mathfrak{D}, \tag{19}$$

*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*

$$D(x) := \exp\left[-\frac{c}{|\det \Sigma(x)|} - cg(x)\right],$$

*where $g : \mathfrak{D} \to \mathbb{R}$ is a function that satisfies $\sum_{j=1}^n |f_j(x)| \le A \cdot (g(x) + 1/|\det \Sigma(x)|)$ for all $x \in \mathfrak{D}$ and a constant $A > 0$, and $c$ 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*

$$\eta_t = \exp\left(\int_0^t \Lambda(X_s)dW_s^{\mathbb{Q}} - \frac{1}{2}\int_0^t \Lambda(X_s)^\top \Lambda(X_s)ds\right), \quad t \le T, \tag{20}$$

*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*

$$dX_t = (D(X_t)f(X_t) + \mu^{\mathbb{Q}}(X_t))\,dt + \Sigma(X_t)\,dW_t^{\mathbb{P}}, \qquad X_0 = x_0, \tag{21}$$

*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 $W_t^{\mathbb{P}} := W_t^{\mathbb{Q}} - \int_0^t \Lambda(X_s)ds$.*

*Proof.* 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)) can be applied to the local martingale

$$\int_0^t \Lambda(X_s) dW_s^{\mathbb{Q}}$$

and hence the density process $\eta$ 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)). □

# B Concentrated likelihood estimation

A summary of the estimation algorithm is given in Section 3.2. This section describes in detail the steps to evaluate the concentrated likelihood function for the models introduced in Section 2. One evaluation of the likelihood function (as a function of $\theta^\sigma \cup \theta^{\mathbb{Q}}$) is based on 5 steps

1. Invert the latent variance state variable $V$ from the VXO through equation (13).
2. Transform the variance state variable to a unit-diffusion process, through equation (22).
3. Compute $\theta^{V\mathbb{P}\star} \mid \theta^\sigma, \theta^{\mathbb{Q}}$ through EML, described in section B.2.
4. Compute $\theta^{X\mathbb{P}\star} \mid \theta^\sigma, \theta^{\mathbb{Q}}$ through EML, described in section B.3.
5. Evaluate transition density approximation, described in section B.4.

## B.1 Preparations for EML estimation

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

$$\gamma(v) = \frac{\log v}{\sigma}. \tag{22}$$

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

$$dY_t = \left\{ \left( b_0^{\mathbb{Q}} + b_1 V_t \right) \frac{1}{\sigma V_t} - \frac{\sigma}{2} \right\} dt + dW_V^{\mathbb{P}}(t), \tag{23}$$

in linear model (4) and by

$$dY_t = \left\{ \left( b_0 + b_1 V_k + b_2 V_t^2 + \frac{b_3}{V_k} \right) \frac{1}{\sigma V_k} - \frac{\sigma}{2} \right\} dt + dW_V^{\mathbb{P}}(t), \tag{24}$$

in nonlinear model (6).

To make our processes suitable for EML estimation we first introduce $M-1$ auxiliary data $U_{i,1},\dots,U_{i,M-1}$ 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 $U_k$, $k=0,\dots,MN$. 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)

$$\begin{aligned}\frac{X_{k+1}-X_k-\rho\sqrt{e^{\sigma Y_k}}\varepsilon^V_{k+1}}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}} &= \left\{\left(a_0+a_1\,e^{\sigma Y_k}\right)\frac{1}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}}\right\}\delta+\varepsilon^X_{k+1}\\ Y_{k+1}-Y_k+\frac{\sigma}{2}\delta &= \left(b^{\mathbb{Q}}_0+b_1\,e^{\sigma Y_k}\right)\frac{1}{\sigma\,e^{\sigma Y_k}}\delta+\varepsilon^V_{k+1},\end{aligned} \tag{25}$$

and the nonlinear variance model (NL)

$$\begin{aligned}\frac{X_{k+1}-X_k-\rho\sqrt{e^{\sigma Y_k}}\varepsilon^V_{k+1}}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}} &= \left\{\left(a_0+a_1\,e^{\sigma Y_k}\right)\frac{1}{\sqrt{e^{\sigma Y_k}}\sqrt{1-\rho^2}}\right\}\delta+\varepsilon^X_{k+1}\\ Y_{k+1}-Y_k+\frac{\sigma}{2}\delta &= \left(b_0+b_1\,e^{\sigma Y_k}+b_2\,e^{2\sigma Y_k}+\frac{b_3}{e^{\sigma Y_k}}\right)\frac{1}{\sigma\,e^{\sigma Y_k}}\delta+\varepsilon^V_{k+1}.\end{aligned} \tag{26}$$

It can be seen that the difference equations (25) and (26) 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 6b and 6a. For the linear variance model we have $L_{\text{LN}}=1$, and for the nonlinear model we have $L_{\text{NL}}=3$. Innovations $\varepsilon^V_k$ and $\varepsilon^X_k$ are both identically independently $N(0,\delta)$-distributed random variables for $k=1,\dots,MN$. Note the appearance of $\varepsilon^V_{k+1}$ in equ. (25) and (26), 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 $\theta^{V\mathbb{P}\star}$. 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 $\theta^{V\mathbb{P}}$.

### B.2 Optimal $\theta^{V\mathbb{P}}\mid\theta^\sigma,\theta^{\mathbb{Q}}$:

Plugging the current values of $\theta^\sigma$ and $\theta^{\mathbb{Q}}$ into eq. (11) 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, 2010). For this purpose we introduce functions $f$ and $g$ from difference equations (25) and (26) which are displayed in Table 6. For a given variance model $\mathcal{M}$ we put the variance

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

(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 $\mathcal{M} \in \{\text{NL}, \text{LN}\}$

Table 6: **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 (27) and (28), expressed in terms of the variable $u_k := (x_k, y_k)$.

drift parameters in a vector $x^{\mathcal{M}} := \left(b_0^{\mathcal{M}}, \ldots, b_{L_{\mathcal{M}}}^{\mathcal{M}}\right)^\top$. EML then yields the optimal drift coefficients $\theta^{V^{\mathcal{M}}\mathbb{P}_\star}$ as the unique solution of the linear system $x^{\mathcal{M}} = (\Xi^{\mathcal{M}})^{-1}\, \varpi^{\mathcal{M}}$ with

$$\Xi^{\mathcal{M}} = \delta \sum_{n=1}^{N-1} \sum_{m=0}^{M-1} \begin{pmatrix} \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n, U_{n+1}}}\left[f_0^{\mathcal{M}}(U_{n,m}) f_0^{\mathcal{M}}(U_{n,m})\right] & \cdots & \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n, U_{n+1}}}\left[f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m}) f_0^{\mathcal{M}}(U_{n,m})\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n, U_{n+1}}}\left[f_0^{\mathcal{M}}(U_{n,m}) f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m})\right] & \cdots & \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n, U_{n+1}}}\left[f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m}) f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m})\right] \end{pmatrix}, \tag{27}$$

$$\varpi^{\mathcal{M}} = \sum_{n=1}^{N-1} \sum_{m=0}^{M-1} \begin{pmatrix} \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n, U_{n+1}}}\left[g(U_{n,m+1}, U_{n,m})\, f_0^{\mathcal{M}}(U_{n,m})\right] \\ \vdots \\ \mathbb{E}_{\mathbb{Q}_{\mathcal{M}}^{U_n, U_{n+1}}}\left[g(U_{n,m+1}, U_{n,m})\, f_{L_{\mathcal{M}}}^{\mathcal{M}}(U_{n,m})\right] \end{pmatrix}. \tag{28}$$

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 (2010) 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)

$$U_{i-1,m+1} = U_{i-1,m} + \frac{U_{i-1,M} - U_m}{M-m} + \sqrt{\frac{M-m-1}{M-m}}\, \varepsilon_{i-1,m+1}, \tag{29}$$

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

### B.3 Optimal $\theta^{X\mathbb{P}} \mid \theta^{\sigma}, \theta^{\mathbb{Q}}, \theta^{V\mathbb{P}\star}$:

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 $\varepsilon^V$) from Table 6a can now be swapped with the functions from Table 6b to estimate optimal stock drift parameters $\theta^{X\mathbb{P}\star}$ through the solution of the linear system (27) – (28).

### B.4 Transition Density Evaluation

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 (15) as the objective function. Since for any value of $\theta^{\sigma}$ and $\theta^{\mathbb{Q}}$ EML yields optimal $\theta^{X\mathbb{P}\star}$ and $\theta^{V\mathbb{P}\star}$, we see likelihood function (15) only as a function of $\theta^{\sigma}, \theta^{\mathbb{Q}}$, and the data. To approximate the unknown transition densities which appear in (15) we use the simulation-based estimator from Pedersen (1995) in connection with the Brownian bridge importance sample from Durham and Gallant (2002)

$$\begin{aligned}
&\sum_{i=1}^{N} \log p^{\mathcal{M}}(X_i, Y_i \mid X_{i-1}, Y_{i-1}, \theta^{\sigma}, \theta^{\mathbb{Q}}, \theta^{X\mathbb{P}\star}(\theta^{\sigma}, \theta^{\mathbb{Q}}), \theta^{V\mathbb{P}\star}(\theta^{\sigma}, \theta^{\mathbb{Q}})) \\
&\approx \sum_{i=1}^{N} \log \left\{ \frac{1}{S} \sum_{s=1}^{S} \frac{\prod_{m=1}^{M} p^{E\mathcal{M}}(U_{i-1,m} \mid U_{i-1,m-1}, \theta^{\sigma}, \theta^{\mathbb{Q}}, \theta^{X\mathbb{P}\star}(\theta^{\sigma}, \theta^{\mathbb{Q}}), \theta^{V\mathbb{P}\star}(\theta^{\sigma}, \theta^{\mathbb{Q}}))}{\prod_{m=1}^{M-1} q(U_{i-1,m} \mid U_{i-1,m-1}, U_{i-1,M})} \right\}.
\end{aligned} \tag{30}$$

Here, $p^{\mathcal{M}}$ refers to the true transition density arising from Heston dynamics with variance specification (23) (LN) and (24) (NL), respectively. The density $p^{E\mathcal{M}}$ denotes a normal distribution arising from the Euler discretization of the corresponding SDE. Auxiliary state variables $U_{i-1,m}, \ldots, U_{i-1,m}$, $i = 1, \ldots, N, m = 1, \ldots, M-1$ are simulated according to the stochastic difference equation

$$U_{i-1,m+1} = U_{i-1,m} + \frac{U_{i-1,M} - U_m}{M-m} + \sqrt{\frac{M-m-1}{M-m}}\, \Sigma(U_{i-1,m})\, \varepsilon_{i-1,m+1}, \tag{31}$$

where

$$\Sigma(U_{i-1,m}) = \begin{pmatrix} \sqrt{1-\rho^2} e^{\sigma Y_{i-1,m}} & \rho\sigma e^{\sigma Y_{i-1,m}} \\ 0 & 1 \end{pmatrix}, \; \varepsilon_{i-1,m+1} = \begin{pmatrix} \varepsilon^X_{i-1,m+1} \\ \varepsilon^V_{i-1,m+1} \end{pmatrix}. \tag{32}$$

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

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

$$p^{E\mathcal{M}}(U_{i-1,m+1} \mid U_{i-1,m}, \theta) = \phi\left(\begin{pmatrix} X_{i-1,m+1} - X_{i-1,m} \\ g_{\mathcal{M}}(U_{i-1,m+1}, U_{i-1,m}) \end{pmatrix}; \begin{pmatrix} a_0 + a_1 e^{\sigma Y_{i-1,m}} \\ \sum_{l=0}^{L_{\mathcal{M}}} f_l^{\mathcal{M}}(U_{i-1,m}) \end{pmatrix} \delta, \Sigma(U_{i-1,m})\Sigma(U_{i-1,m})^{\top}\delta\right).$$

Following Stramer and Yan (2007) we set $S = M^2 = 576$. Note that the $\varepsilon$ variates appearing in (29) from EML estimation may be reused in this step.