An Empirical Comparison of Transformed Diffusion Models for VIX and VIX Futures

Ruijun Bu*
University of Liverpool, UK

Fredj Jawadi
University of Evry, France

Yuyi Li
University of Liverpool, UK

*Corresponding author: Management School, Chatham Street, Liverpool, L69 7ZH, UK, Tel: +44-151-795-3122, Fax: +44-151-795-3004, Email: RuijunBu@liv.ac.uk (Ruijun Bu).

An Empirical Comparison of Transformed Diffusion Models for VIX and VIX Futures

Abstract

Transformed diffusions (TDs) are nonlinear functions of continuous-time affine diffusion processes. Since they are flexible models with tractable analytic properties, financial modelling with TDs has become increasing popular in recent years. We first provide a formal classification of TD models into drift-driven, diffusion-driven, and distribution-driven according to their empirical emphases and specification strategies. Motivated by the stylized distributional features of VIX such as skewness and excess kurtosis, we then propose a pair of new distribution-driven TDs for modelling VIX dynamics and pricing VIX futures by directly incorporating such information into the specification of the transformation. We conduct a comprehensive empirical investigation into the relative performance of the three classes of models against several empirically relevant criteria. Our focus is on the in-sample goodness-of-fit measure and the out-of-sample forecast accuracy for modelling VIX and pricing VIX futures, as well as the stock return predictability of the implied Variance Risk Premium. Our findings demonstrate that the newly proposed distribution-driven models have clear advantages over well-established alternatives in most of our exercises.

JEL Classification: C13, C32, G13, G15

Keywords: Transformation Model; Nonlinear Diffusion; Skewed Student-t Distribution; Volatility Index; VIX Futures.

1. Introduction

As a measure of market volatility implied by traded S&P 500 index option prices, the Volatility Index (VIX), also known as the "investor fear gauge", has attracted enormous attention in recent years. For financial market participants, it is of the utmost importance to understand the dynamics of market volatility which is a crucial determinant of investment decisions. One of the most important strands of the literature focuses on the data generating process of VIX, since a realistic model for the VIX dynamics is vital for correct inference and accurate derivative pricing.

Continuous-time diffusion models are particularly useful for modelling financial variables, not only because they are flexible yet parsimonious analytical tools, but also because derivative pricing crucially relies on a "continuous-time" no-arbitrage argument (i.e. change of measure). For this reason, a growing number of studies have emerged on modelling the dynamics of VIX and price VIX futures and options using continuous-time models. Whaley (1993) used a Geometric Brownian Motion model which does not have the mean-reverting feature for pricing volatility futures contracts. Grunbichler and Longstaff (1996) considered pricing VIX futures and options assuming that VIX follows the Cox et al. (1985) square-root process (CIR). Detemple and Osakwe (2000) considered the log-normal Ornstein-Uhlenbeck (OU) model. Goard and Mazur (2013) advocated the so-called 3/2 model, which is the same as the Ahn and Gao (1999) model, for VIX and VIX options. More recently, Eraker and Wang (2015) proposed a new nonlinear diffusion model with a cubic drift term to study the Variance Risk Premium (VRP) implied by VIX futures prices.

One of the greatest challenges in diffusion modelling is to construct models that are sufficiently flexible to describe complex nonlinear dynamics in reality and sufficiently tractable to allow efficient inference such as the Maximum Likelihood (ML) and convenient and accurate derivative pricing. For this purpose, one useful approach is to consider transformation models in continuous time. Bu et al. (2011) promoted the idea of modelling nonlinear diffusion, say Y, as a transformation of a more tractable basic affine process, say X, for which the closed-form transition density is available. Prominent examples of tractable underlying diffusions include, but not restricted to, the OU and the CIR processes. For a given transformation, i.e. $Y = V(X)$, the transition density of the transformed diffusion (TD) Y is simply the distribution transformation of the transition density of X under mild conditions. Most importantly, if V is nonlinear and flexible, then Y would also be a nonlinear and conceivably more flexible diffusion process.

Financial modelling with TD models has been increasingly popular in recent years[1]. Bu et al. (2011) considered transformed OU and CIR processes with Constant Variance Elasticity (CEV) diffusion terms (denoted as OUCEV and CIRCEV) for modelling short-term

[1] In a non-financial context, Forman and Sørensen (2014) considered a transformed OU process (henceforth denoted as OUFS) with a bimodal marginal distribution for modelling molecular dynamics.

interest rates[2]. They showed that their models can generate similar nonlinearities in both the drift and diffusion terms to those estimated nonparametrically (e.g. Aït-Sahalia 1996a). More importantly, Bu et al. (2016b) showed that the CIRCEV model can provide much better fit to the VIX data than the Nonlinear Drift CEV (NLDCEV) model of Aït-Sahalia (1996b) and Conley et al. (1997). Detemple and Osakwe (2000) used the exponential transformation of the OU process (henceforth denoted as OUDO) for pricing volatility options. Goard and Mazur (2013) considered the reciprocal transformation of the CIR process for modelling VIX and pricing VIX options. More recently, Eraker and Wang (2015) considered a transformation of the CIR process which leads to a cubic drift function (henceforth denoted as CIREW). Using the Fourier transformation method, they derived a pricing formula for VIX futures and studied the VRP implied by the VIX futures prices.

The first contribution of this paper is to provide a formal classification of TDs in the literature, according to their empirical emphases and specification strategies. The first class are considered as "drift-driven" (e.g. CIREW and OUDO) where for an underlying process X, the transformation is derived so that the resulting TD Y has a desired drift function. The second class are "diffusion-driven" (e.g. OUCEV and CIRCEV) where the users specify a desired diffusion term of Y and the drift term is determined simultaneously. The third class are "distribution-driven" (e.g. OUFS) where the users specify the marginal distribution of Y to be a member a class of parametric distributions, based on which the transformation function is then derived. It can be shown that theoretically all three approaches are interchangeable since there is a intrinsic relationship between the drift, the diffusion and the marginal distribution of any stationary diffusion process. In practice, however, the empirical relevance and also the analytic tractability of the three strategies can be fairly different depending on the users' preferences.

Although TD models are increasingly popular in financial modelling, their full potential in modelling financial derivatives have not been fully explored. The second contribution of this paper is to propose a pair of new distribution-driven nonlinear TD processes purposefully designed to fully incorporate stylized features such as the skewness and excess kurtosis in the distribution of the VIX data while at the same time deliver a closed-form transition density for efficient likelihood inference and closed-form VIX futures pricing. Our new TDs (named as CIRSKST and OUSKST) are constructed as the transformed CIR and OU process, respectively, where we propose to use the Skewed Student-t distribution (SKST) of Hansen (1994) to directly explore the information in the marginal distribution of VIX. Following a simple argument of change of measure in continuous time, we derive a closed-form formula for the prices of the VIX futures contract.

As our third contribution of this paper, we provide a comprehensive empirical investigation into the comparative performance of the three classes of models for modelling time

[2] The CIRCEV model nests the Ahn and Gao (1999) and Goard and Mazur (2013) models as special cases.

series of VIX data (i.e. under the physical P-measure) and for pricing VIX futures (i.e. under the risk-neutral Q-measure). Our data consist of 6352 daily observations of VIX from January 2, 1990 to March 20, 2015 and 19215 observations of VIX futures closing prices from March 26, 2004 to February 17, 2015. Our comparison is based on a set of empirically important criteria. Firstly, we compare the three classes of models in terms of their in-sample goodness-of-fit and out-of-sample forecasting accuracy for modelling the VIX dynamics under the physical measure. We then examine our competing models in terms of their in-sample and out-of-sample performance for pricing VIX futures, which is carried out jointly under both the physical and risk-neutral measures. Finally, following the arguments of Bollerslev et al. (2009) and Eraker and Wang (2015), we extract the time-varying VRP jointly inferred from the VIX and VIX futures data for each of our competing models and examine their predictability for S&P 500 index returns.

Our empirical analysis provides a number of important results. For modelling the VIX dynamics under the physical measure, the newly proposed distribution-driven models dominated well-established models in the literature in terms of both in-sample goodness-of-fit measures and out-of-sample forecast accuracy. While both the CIRSKST and the OUSKST models fitted the VIX data equally well, the latter produced the smallest average forecast errors. For modelling VIX and pricing VIX futures under the joint measures, very similar outcomes emerged. While the CIRSKST model dominated all other models in terms of in-sample performances, the distribution-driven class dominated the other two classes in terms of superior out-of-sample performances. In particular, the OUSKST model produced significantly smaller average forecast errors than any other competing model. All three classes of models performed rather similarly in terms of stock return predictabilities of the extracted VRP. Nevertheless, with non-substantial margins, the CIRSKST model produced the highest predictability for the shortest forecasting horizon and the OUSKST model produced the highest predictability for the majority of the longer horizons. This outcome is broadly consistent with earlier results about their respective superior in-sample and out-of-sample abilities. It is worth mentioning that no models from the other classes ever dominated the rest of the field against any of our empirically important criteria. These findings demonstrate quite robustly that the new distribution-driven models have clear advantages over well-established alternatives. In the meantime, we found that the CIREW, the CIRCEV and the OUCEV models performed rather similarly in all categories, and the OUDO model performed better out-of-sample than in-sample.

The rest of the paper is organized as follows: In Section 2, we briefly review the transformation approach for modelling continuous-time diffusions and propose a formal classification. In Section 3, we propose our new distribution-driven specifications purposefully designed for financial time series. In Section 4, we discuss the pricing of VIX futures and our estimation strategy. In Section 5, we present a thorough empirical comparison of competing TDs for modelling VIX dynamics and pricing VIX futures against several empirically vital criteria. Concluding remarks are included in Section 6.

2. Nonlinear Transformed Diffusion Models

2.1. General Framework

The dynamics of a continuous-time diffusion Y is typically written as

$$dY_t = \mu_Y(Y_t;\phi)\,dt + \sigma_Y(Y_t;\phi)\,dW_t$$

where $\mu_Y(y;\phi)$ and $\sigma_Y^2(y;\phi)$ are the instantaneous drift and diffusion functions respectively, and W is a standard Brownian motion. The focus of financial modelling is on the specification of $\mu_Y(y;\phi)$ and $\sigma_Y^2(y;\phi)$. Well known examples include Merton (1973), Black and Scholes (1973), Vasicek (1977), Cox et al. (1985), Duffie and Kan (1996), Aït-Sahalia (1996b), Conley et al. (1997), Ahn and Gao (1999), Bu et al. (2011).

Bu et al. (2011) proposed to model stochastic financial variables by TDs where Y is assumed to be a strictly monotone transformation of some basic affine diffusion X, i.e.

$$Y_t = V(X_t;\vartheta)$$

where

$$dX_t = \mu_X(X_t;\omega)\,dt + \sigma_X(X_t;\omega)\,dW_t, \tag{1}$$

which depends on parameter ω. $V(x;\vartheta)$ is the transformation function with parameter ϑ satisfying $\partial V(x;\vartheta)/\partial x \neq 0$ for all $x \in D_X$. Ito's Lemma determines that

$$\mu_Y(y;\phi) = \frac{\mu_X(U(y;\vartheta);\omega)}{U'(y;\vartheta)} - \frac{\sigma_X^2(U(y;\vartheta);\omega)\,U''(y;\vartheta)}{2U'(y;\vartheta)^3} \tag{2}$$

$$\sigma_Y^2(y;\phi) = \frac{\sigma_X^2(U(y;\vartheta);\omega)}{U'(y;\vartheta)^2} \tag{3}$$

where $U(y;\vartheta) = V^{-1}(y;\vartheta)$ is the unique inverse of $V(x;\vartheta)$, and $U'(y;\vartheta)$ and $U''(y;\vartheta)$ are its first and second derivatives. Let $p_X(x|x_0;\omega)$ and $p_Y(y|y_0;\phi)$ be the transition probability density function (PDF) of X and Y, respectively. It follows immediately that

$$p_Y(y|y_0;\phi) = |U'(y;\vartheta)|\,p_X(U(y;\vartheta)\,|U(y_0;\vartheta);\omega)$$

The choice of (1) is typically restricted to the affine class so that $p_X(x|x_0;\omega)$ and hence $p_Y(y|y_0;\phi)$ are in closed form. In the literature, the CIR and the OU processes are the most common choices. The dynamics of the CIR process is written as

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t$$

where $\kappa > 0$, $\theta > 0$ and $2\kappa\theta \geq \sigma^2 > 0$ and that of the OU process is given by

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t$$

where $\kappa > 0$ and $\sigma^2 > 0$. The resulting closed-form transition PDF $p_Y(y|y_0;\phi)$ forms a very useful basis for ML inference and pricing VIX futures.

2.2. Classification of Transformed Diffusion Models

The most important task in modelling TDs is the specification of $V(x;\vartheta)$. Bu et al. (2011) noted that for any given specification of X, the knowledge of the functional form of either $\mu_Y(y;\phi)$ or $\sigma_Y^2(y;\phi)$ leads to the unique solution of $V(x;\vartheta)$. In the meantime, under the stationarity assumption of X, the marginal Cumulative Distribution Function (CDF) of X and Y both exist, from which $V(x;\vartheta)$ can also be inferred uniquely. In this section, we provide a formal classification of TD models depending on their empirical emphases and hence their corresponding specification strategies.

2.2.1. Drift-Driven Class

We define the class of TD models which focus on a desired drift function $\mu_Y(y;\phi)$ as drift-driven. Note that for a given $\mu_Y(y;\phi)$, the transformation function $U(y;\vartheta)$ is the solution to the system of Ordinary Differential Equations (ODE) of (2) and (3), which depends on both $\mu_X(x;\omega)$ and $\sigma_X^2(x;\omega)$. For an arbitrary $\mu_Y(y;\phi)$, the closed-form expression for $U(y;\vartheta)$ is usually unavailable.

Nevertheless, special cases where the drift-driven strategy does lead to elegant closed-form $U(y;\vartheta)$ have been proposed in the literature. For example, when X is the CIR and the desired drift function is a cubic polynomial[3], Eraker and Wang (2015) showed that the required $U(y;\vartheta)$ is given by

$$U(y;\vartheta) = 1/(y-\varphi) - \alpha$$

and the resulting drift function is given by

$$\mu_Y(y;\phi) = \kappa(y-\varphi) + \left[\sigma^2 - \kappa(\theta+\alpha)\right](y-\varphi)^2 - \alpha(y-\varphi)^3$$

Similarly, Detemple and Osakwe (2000) showed that when X is the OU, the logarithmic transformation $U(y;\vartheta) = \ln y$ implies the following linear drift for $\ln y$

$$\mu_{\ln Y}(\ln y;\phi) = \kappa(\theta - \ln y)$$

or equivalently

$$\mu_Y(y;\phi) = \left(\kappa\theta + \sigma^2/2\right)y - \kappa y \ln y$$

which gives rise to the so-called Mean-Reverting Log process (i.e. OUDO). As leading examples of TDs for VIX in the literature, both the OUDO model and the CIREW model are considered in our empirical comparison.

[3] Ahn and Gao (1999) model for short-term interest rates is a special case where the underlying basic affine process is the CIR and the drift function has a quadratic form.

2.2.2. Diffusion-Driven Class

The diffusion-driven class of TD models begin with a desired diffusion function $\sigma_Y^2(y;\phi)$. It follows that $U(y;\vartheta)$ is the solution to the 1st-order ODE in (3). Solving (3) is comparatively simple and in some cases analytical solutions exist. A prominent example frequently promoted in financial modelling is the following CEV specification

$$\sigma_Y^2(y;\phi) = \sigma_0^2 y^{2\gamma} \text{ for } \gamma \in [0,+\infty)$$

It was introduced by Chan et al. (1992) who considered a linear drift and subsequently studied by Aït-Sahalia (1996b) who promoted a nonlinear drift to improve the mean reversion effect. Bu et al. (2011) proposed two TD models with CEV diffusion terms, the CIRCEV and the OUCEV, where X is either the CIR or the OU, respectively. They showed that for the CIRCEV model the transformation function is

$$U(y;\vartheta) = \begin{cases} \left(y^{1-\gamma}/(1-\gamma)\right)^2/4 & \text{for} \quad \gamma \in [0,1) \cup (1,+\infty) \\ (\log y)^2/4 & \text{for} \quad \gamma = 1 \end{cases}$$

and for the OUCEV model it is

$$U(y;\vartheta) = \begin{cases} y^{1-\gamma}/(1-\gamma) & \text{for} \quad \gamma \in [0,1) \cup (1,+\infty) \\ \log y & \text{for} \quad \gamma = 1 \end{cases}$$

It is worth stressing that the drift and diffusion terms of both models are nonlinear. In addition to the CEV diffusion term, the drift also exhibits a much stronger pull at high and low levels of the state variable than the linear drift. Both properties are consistent with empirical findings about the two functions reported in Aït-Sahalia (1996a,b), Conley et al. (1997), Stanton (1997) and others. The CIRCEV model encompasses the CIR model with $\gamma = 0.5$ and the Ahn and Gao (1999) model with $\gamma = 1.5$, and the OUCEV model nests the OU model with $\gamma = 0$. Clearly, both models are more general, which not only provide the nonlinearity in both terms but also allow extra degrees of freedom in the data-driven choice of γ. Bu et al. (2016a,b) considered some extensions of the CIRCEV model for empirical applications.

In practice, it is always useful to test the linear or affine restrictions when a nonlinear model is deemed more suitable. However, although the CIREW model is a transformed CIR, it does not nest CIR model itself. Consequently, Eraker and Wang (2015) had to rely on a parametric bootstrap method to test the affine restriction. In contrast, the CIRCEV model strictly nests the CIR model and also the Ahn and Gao (1999) model. Therefore, testing existing specifications nested in the model only requires the very simple standard Likelihood Ratio testing procedure. Finally, despite its desirable flexibility and tractability, to our best knowledge, the CIRCEV and the OUCEV have yet to be exploited for pricing financial assets such as the VIX derivatives.

2.2.3. Distribution-Driven Class

The third class of TD models are designed to have a desired marginal distribution. Under the stationarity assumption, the marginal PDFs and CDFs exist. Let $F_Y(y;\vartheta)$ and $F_X(x;\omega)$ be the marginal CDF of Y and X, respectively. Under the increasing monotonicity assumption, we have

$$F_Y(y;\vartheta) = F_X[U(y;\vartheta);\omega]$$

which implies the following transformation function

$$U(y;\vartheta) = F_X^{-1}[F_Y(y;\vartheta);\omega] \tag{4}$$

and the Jacobian

$$\frac{\partial U(y;\vartheta)}{\partial y} = \frac{f_Y(y;\vartheta)}{f_X\left\{F_X^{-1}[F_Y(y;\vartheta);\omega];\omega\right\}}$$

where $f_Y(y;\vartheta)$ and $f_X\{x;\omega\}$ are the marginal PDF of Y and X. The transition PDF of Y is then

$$p_Y(y|y_0;\omega,\vartheta) = \frac{f_Y(y;\vartheta)}{f_X\{U(y;\vartheta);\omega\}} p_X(U(y;\vartheta)|U(y_0;\vartheta);\omega) \tag{5}$$

This specification strategy was considered by Forman and Sørensen (2014) for modelling molecular dynamics, where they assumed that the marginal distribution of Y is a mixture of two normal distributions. The motivation of their specification is that it is a stylized feature that the marginal distribution of their protein folding data exhibits bimodality and existing models failed to model this marginal feature and time series dynamics adequately. They assumed that the underlying process is the OU process.

3. A New Transformed Diffusion for VIX

Bu et al. (2011) noted that there exists little direct guidance on how to specify $U(y;\vartheta)$ such that the resulting TDs are not only tractable but also realistically flexible to capture important distributional features of the data such as skewness and fat tails. For this reason, the distribution-driven approach has the clear advantage over the other two classes in that researchers can directly specify a flexible functional form for the marginal density $f_Y(y;\vartheta)$ to directly incorporate distributional information in the data. Moreover, choices of desired drift function $\mu_Y(y;\phi)$ or desired diffusion function $\sigma_Y(y;\phi)$ that will lead to closed-form solutions of $U(y;\vartheta)$ are rather limited. In contrast, there is an enormous literature on density specification and estimation in statistics, and (4) and (5) show that any desired specification can lead to closed-form transformation function $U(y;\vartheta)$ and hence transition PDF $p_Y(y|y_0;\omega,\vartheta)$.

3.1. Skewed Student-t Marginal Distribution

Most financial data exhibit skewness and fat tails that affine diffusions are unable to model sufficiently (c.f. Aït-Sahalia 1996b). The distribution-driven specification strategy allows for more flexibility in the marginal distributions without sacrificing the dynamic structure. To directly incorporate distributional information of the data, we propose two new distribution-driven TD models where the marginal distribution of Y is assumed to follow the Skewed Student-t Distribution (SKST) of Hansen (1994). The location-scale version of the SKST distribution has the following density function:

$$f_Y(y;\vartheta) = \begin{cases} \dfrac{bq}{\varsigma}\left(1+\dfrac{1}{v-2}\left(\dfrac{\frac{b}{\varsigma}(y-m)+a}{1-\lambda}\right)^2\right)^{-(v+1)/2} & \text{if} \quad y < m - a\varsigma/b, \\ \dfrac{bq}{\varsigma}\left(1+\dfrac{1}{v-2}\left(\dfrac{\frac{b}{\varsigma}(y-m)+a}{1+\lambda}\right)^2\right)^{-(v+1)/2} & \text{if} \quad y \geq m - a\varsigma/b, \end{cases}$$

where $\varsigma > 0$, $2 < v < \infty$, $-1 < \lambda < 1$, and

$$a = 4\lambda q\left(\frac{v-2}{v-1}\right), b^2 = 1 + 3\lambda^2 - a^2, q = \frac{\Gamma\left((v+1)/2\right)}{\sqrt{\pi(v-2)}\Gamma(v/2)}.$$

Note that the location-scale SKST has four parameters, i.e., $\vartheta = (m, \varsigma, \lambda, v)'$, While m and ς are the mean and standard deviation of the distribution, λ and v control the skewness and the degrees of freedom (i.e. fat-tailedness) of the distribution. The SKST distribution reduces to the usual student-t distribution when $\lambda = 0$. Due to its flexibility in modelling skewness and kurtosis, the SKST distribution is often used in financial modelling (c.f. Patton 2004, Jondeau and Rockinger 2006).

3.2. CIRSKST and OUSKST Models

The distribution-driven specification strategy requires a suitable normalization of the underlying diffusion X to exclude unidentified parameters. For the CIR process, it can be easily shown that we can only identify θ and σ^2 up to the ratio θ/σ^2. For ease of model estimation and pricing VIX futures subsequently, we choose to set $\sigma^2 = 1$ and consequently our normalized CIR process is given by

$$dX_t = \kappa(\theta - X_t)dt + \sqrt{X_t}dW_t$$

The resulting TD model is referred to as the CIRSKST model with $\phi = (m, \varsigma, \lambda, v, \kappa, \theta)'$. Similarly, for the OU process, it can be shown that θ and σ^2 are unidentified and can be set to any admissible constant. For ease of estimation and pricing, we choose to set $\theta = 0$ and $\sigma^2 = 1$ and thus the normalized OU process is given by

$$dX_t = -\kappa X_t dt + dW_t$$

The resulting TD model is referred to as the OUSKST model with $\phi = (m, \varsigma, \lambda, \upsilon, \kappa)'$.

4. Valuation of VIX Futures

4.1. VIX Futures Price

In the absence of arbitrage opportunities in a complete market, the VIX futures price is the conditional mean of the unique risk-neutral Martingale measure. Since our underlying diffusion X has closed-form transition density, the VIX futures price can be easily obtained in explicit form. The unique risk-neutral Q-measure that is equivalent to the observed physical P-measure can be established by applying Girsanov's theorem. Let $\Lambda(Y_t)$ be the Market Price of Risk (MPR) with respect to the Brownian motion. Under the Q-measure, the risk-neutral process of Y can be expressed as

$$dY_t = [\mu_Y(Y_t; \phi) - \Lambda(Y_t)\sigma_Y(Y_t; \phi)]\,dt + \sigma_Y(Y_t; \phi)\,dW_t^Q \tag{6}$$

Following the convention, the parametric specification of Y is assumed to be the same under both measures. It is important to note that for TDs, the specification in (6) is determined jointly by the specification of X and $U(y; \vartheta)$. Thus, in order for $\sigma_Y(Y_t; \phi)$ to be identical under both measures, the parameters in the diffusion function of X and $U(y; \vartheta)$ must also be identical under both measures. Consequently, the diffusion function $\sigma_Y(Y_t; \phi)$ remains the same under both measures, but the drift parameters will differ under the two measures.

Define ω^Q as the parameter of X under the Q-measure. Then, the price of VIX futures at time t is simply the time-t conditional expectation of the value of VIX at a future maturity date T under the Q-measure, i.e.

$$F\left(t, T, y_t, \omega^Q, \vartheta\right) = \mathbb{E}_t^Q\left[y_T | y_t, \omega^Q, \vartheta\right] = \int_0^\infty y_T p_Y\left(y_T | y_t; \omega^Q, \vartheta\right) dy_T$$

By construction of our TDs, the risk-neutral transition density $p_Y\left(y|y_0; \omega^Q, \vartheta\right)$ is in closed-form and thus VIX futures pricing involves only a 1-dimensional numerical integration.

4.2. Joint Measure Estimation

Let $\{Y_{i\Delta}, i = 0, ..., n_{VIX}\}$ be a sample of VIX data, where Δ is the sampling interval. Define ω as the parameter of X under the P-measure. Then, the log-likelihood (LL) function under the physical P-measure is given by

$$LL_{VIX}(\omega,\vartheta)=\sum_{i=1}^{n_{VIX}}\ln p_Y\left(Y_{i\Delta}|Y_{(i-1)\Delta};\omega,\vartheta\right)$$

Meanwhile, let $\{F_j(t,T,Y_t), j=1,...,n_F\}$ be a sample of VIX futures prices and assume that the VIX futures pricing error has the following distribution

$$e_j(\omega^Q,\vartheta)=F_j(t,T,Y_t)-F_j\left(t,T,Y_t,\omega^Q,\vartheta\right)\sim N\left(0,\sigma_F^2\right)$$

We can profile out σ_F which can be estimated by

$$\hat{\sigma}_F=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left[F_j(t,T,Y_t)-F_j\left(t,T,Y_t,\omega^Q,\vartheta\right)\right]^2} \tag{7}$$

which is actually the Root Mean Squared Error (RMSE) of our pricing model. Consequently, our profile LL for VIX futures can be written as

$$LL_F\left(\omega^Q,\vartheta\right)=\sum_{j=1}^{n_F}\ln\varphi(e_j)$$

where $\varphi(\cdot)$ is normal density with mean zero and standard deviation $\hat{\sigma}_F$ given in (7). Finally, the joint LL based on both VIX and VIX futures data is given by the following sum

$$LL_{Total}\left(\omega,\omega^Q,\vartheta\right)=LL_{VIX}(\omega,\vartheta)+LL_F\left(\omega^Q,\vartheta\right)$$

5. Empirical Analysis

5.1. Data and Models

We compare the empirical performance of the three classes of TD models discussed above for modelling the VIX dynamics and pricing VIX futures. Our data obtained from the CBOE website include the daily VIX closing index from January 2, 1990 to March 20, 2015 (6352 observations) and VIX future closing prices from March 26, 2004 to February 17, 2015 (19215 observations). Following Eraker and Wang (2015), we constructed 7 series of daily hypothetical constant maturity ($1,2,3,...,7$ month) future prices by linear interpolation. Each series contains 2742 observations. We use data up to December 31, 2014 for in-sample analysis and the remaining data for out-of-sample comparison. This gives us 6298 observations of VIX index, 2711 observations of VIX futures for each constant maturity, and a total of 19141 observations of natural maturity VIX futures for the estimation of our models.

The plots of the VIX and the constant maturity VIX future prices are given in Figures 1 and 2, respectively. Some summary statistics are reported in Table 1. The mean of VIX is 20.61 with standard deviation 10.19. The large skewness 2.21 and kurtosis 9.25 suggest strong deviation from normality. The evolution of VIX indicates that the mean reversion is weak when VIX is low but much stronger when it is high. This suggests that the most suitable diffusion model for VIX may have a drift function that is close to zero when VIX is low and strongly negative when VIX is high. Meanwhile, the local volatility of VIX is also low when VIX is low and substantially higher when VIX is high. This suggests that the most suitable diffusion model may also need a diffusion term that increases rapidly as VIX increases.

[Figure 1 and 2]
[Table 1]

The evolutionary paths of constant maturity VIX future prices with different maturities are highly correlated. The largest eigenvalue of the correlation matrix is almost 50 times larger than the second largest, suggesting that the first factor explains around 98% of the variation in the 7 series. To a large extent, this justifies the use of single-factor models for our data. Meanwhile, the spread between short and long maturities are fairly narrow. This is indicative of a rather flat term structure of VIX futures. Augmented Dickey-Fuller tests on these time series all rejected the unit root hypothesis with 4 lags at 5% significance level, which justifies the use of stationary models.

A total of six models are considered in our empirical analysis. For each class, two models are considered, one being a transformed CIR model and the other being a transformed OU model. The two models from the drift-driven class are the CIREW and the OUDO models. The two models from the diffusion-driven class are the CIRCEV and the OUCEV models. Finally and the most importantly, the distribution-driven class is represented by the newly proposed CIRSKST and OUSKST models.

5.2. Analysis of Time Series of VIX

We first examine the performance these models for modelling VIX dynamics. This amounts to comparing models under the physical P-measure. We investigate both in-sample goodness-of-fit measures and out-of-sample forecasting accuracy. As discussed earlier, one of the biggest advantages of TDs is the availability of closed-form transition PDFs. For this reason, ML is obviously the preferred choice estimation method. Table 2 reports our estimation results. Firstly, in terms of goodness-of-fit measured by the LL, the newly proposed distribution-driven models outperformed the other two classes very clearly. This may not be too surprising, because the CIRSKST and the OUSKST models directly capture the information in the stylized skewness and kurtosis of the marginal distribution of the data and therefore use the information more effectively than the other two classes. Within this class, the CIRSKST fitted the data better than the OUSKST, which is also expected since the former has one more parameter describing the underlying dynamics than the latter. However, in terms of AIC, they are effectively the same, whereas the BIC even favors the OUSKST due to its parsimony. Based on this evidence, we may conclude that the newly proposed distribution-driven models outperformed the other two classes in terms of the in-sample performance for modelling VIX under the P-measure. Comparing between the other two classes, the drift-driven CIREW and OUDO models are inferior to the two diffusion-driven CEV models quite significantly. This reflects existing empirical evidence that the diffusion specification is more important than the drift specification for most financial series (c.f. Aït-Sahalia 1996a, Mencia and Santana 2013).

[Table 2]

We then compare the models in terms of out-of-sample forecast accuracy. For each model, we produce a series of rolling sample one-period-ahead forecast for VIX and compute the Root Mean Squared Forecast Error (RMSFE) also reported in Table 2. Again, the two distribution-driven models outperformed the other classes rather clearly. This is quite revealing results because normally there is a trade of between in-sample goodness-of-fit and out-of-sample forecasting accuracy, i.e. models fit the in-sample data well tend to perform weakly in out-of-sample forecasting. However, this is not the case here, suggesting no evidence of over-fitting. This is a quite strong indication of the effectiveness and superiority of the newly proposed models for modelling the VIX dynamics. Based on this evidence, we can further conclude that the newly proposed distribution-driven models also outperformed the other two classes in terms of the out-of-sample performance under the P-measure. It is worth mentioning that the OUSKST model performed better than the CIRSKST model. This together with the smaller BIC value leads us to conclude that in terms of overall performance the distribution-driven OUSKST is probably the best model for the VIX dynamics under the P-measure.

5.3. Analysis of VIX Futures Pricing

We now examine the performance of these models for pricing VIX futures. We estimate model parameters under both measures jointly by ML using a combination of the VIX data and one of the two sets of VIX future data. The first set includes 19141 observations of observed natural maturity VIX futures data in our sample period. Since this combination only involves directly observed data, the parameter estimates are most realistic. In addition, as a robustness check, we also used a set of hypothetical constant maturity futures. Following Eraker and Wang (2015), we include $1, 3, 5$ and 7-month maturity futures ($2711 \times 4 = 10844$

observations) in the second set. Tables 3 and 4 present the estimation results for these two cases. It turns out that the two sets of data produced rather similar results. As we argued above, the strong correlation between hypothetical constant maturity VIX futures prices is a reasonable justification for single-factor models. Meanwhile, the fact that the estimation results from natural maturity futures are so close to those from hypothetical constant maturity futures is a further confirmation of the general suitability of single-factor models for our data.

We again compare the in-sample and out-of-sample performance of the three classes for pricing VIX futures. Note that the LL, AIC and BIC are joint measures based on VIX and VIX futures, and the RMSE measures the in-sample average futures pricing errors. From the results based on natural maturity futures in Table 3, we can see that the CIRSKST model produced the largest LL and the smallest AIC and BIC values, as well as the smallest RMSE across all six models. This suggests that the distribution-driven CIRSKST is the all-round best model in terms of in-sample performance, leading the second besting performing model, which is the drift-driven CIREW model, by a significant margin. Meanwhile, the OUSKST model produced the best results within the transformed OU class of models. In contrast, the drift-driven OUDO model has the worst results, suggesting that the exponential transform, which is parameter free, may be too simple to produce necessary flexibility for VIX data. The results based on constant maturity futures in Table 4 delivered a rather similar picture, i.e. the CIRSKST produced the best in-sample performance, while the OUSKST outperformed the other models in the transformed OU class.

[Table 3 and 4]

To examine out-of-sample performance for pricing futures, we calculated the RMSFE for pricing natural maturity futures in our forecasting sample. From Table 3, the distribution-driven class dominated the other two classes overwhelmingly, with the OUSKST and the CIRSKST being the best two performing models. In particular, within the transformed CIR class, the RMSFEs from the CIREW and CIRCEV models are 44% and 40% higher than that of the CIRSKST model, respectively. In the transformed OU class, the RMSFEs from the OUDO and OUCEV models are 24% and 46% higher than that of the OUSKST model, respectively. More importantly, as in the case of forecasting VIX under the P-measure, the OUSKST model is also by far the best forecasting model for VIX future prices under the Q-measure. In the case where constant maturity futures are used, we reported in Table 4 the RMSFEs based on all futures in the forecasting sample (RMSFE_F) and those based on futures with different maturities (RMSFE_1, RMSFE_3, RMSFE_5, RMSFE_7). Firstly, in terms of every single measure, the distribution-driven class dominated the other two classes, where the OUSKST is always better than the CIRSKST. In particular, the relative performance of the OUSKST model is even stronger. For example, RMSFE_F of the second best CIRSKST model is now 102% higher than that of the OUSKST and the RMSFE_F of the worst performing CIREW model is even 150% higher. Secondly, in terms of RMSFE for individual maturities, the difference across all six models is the smallest for 1-month maturity, where the OUSKST outperforms the second best CIRSKST by 27% and the CIREW model by 58%. The difference is the largest for 3-month maturity with the OUSKST outperforming the CIRSKST by 187% and the CIREW model by a staggering 317%.

Overall, these evidence suggest that the two newly proposed distribution-driven models performed very well in our joint measure analysis, where the CIRSKST performed well in both in-sample fit and out-of-sample forecast and the OUSKST model is particularly suitable for out-of-sample exercises.

5.4. Variance Risk Premium and Return Predictability

Recent studies including Bollerslev et al. (2009) and Eraker and Wang (2015) documented that the VRP predicts stock returns. Bollerslev et al. (2009) using a model free framework showed that the predictive R^2 peaked at 6.8% at the 3-month horizon. Eraker and Wang (2015), using their drift-driven model and a flexible regression model, documented the highest adjusted-R^2 of 8.41% at the 9-month horizon. Since the three classes of models performed quite differently in capturing the VIX dynamics under both the physical and risk-neutral measures, it will be useful to also examine the stock return predictability of the VRP extracted from competing models.

Eraker and Wang (2015) showed in the context of general TDs that the market price of VIX-squared risk is the same as the market price of risk for the spot variance, and hence the VRP defined as the difference between the drift of the VIX-squared under the Q-measure and that under the P-measure is the same as in the extant literature (e.g. Bollerslev et al. 2009)[4]. Therefore, for each estimated model, we can extract a time series VRP by first applying the Ito's Lemma to the estimated dynamics for VIX to get the corresponding dynamics for VIX-squared and then taking the difference between the drift under the two measures. We use parameters calibrated from the VIX and natural maturity VIX futures.

Following the literature, we consider VIX and S&P 500 Index return data at the monthly frequency and a multi-period predictive regression of the following form

$$\frac{1}{h}\sum_{j=1}^{h} r_{t+j} = \beta_0(h) + \boldsymbol{x}_t'\boldsymbol{\beta}(h) + \varepsilon_t$$

where the dependent variable is the scaled h-month return, $\boldsymbol{x}_t$ is a predictor vector, and ε_t is a white noise process. As in Eraker and Wang (2015), the current framework does not imply any specific functional form between VRP and stock returns. For Eraker and Wang (2015), the best predictability was achieved by including VIX, VRP and VRP^2. Following this

[4]See Eraker and Wang (2015) for a detailed discussion.

suggestion, we include VIX, VIX^2, VRP and VRP^2 as our candidate predictors and then apply the stepwise regression technique to choose the best predictive models. Our sample period is from January 1990 to December 2014.

[Table 5]

Results in terms of the adjusted R^2 from all six models for forecasting horizon ranging from 1 month to 2 years are reported in Table 5. The magnitudes of the adjusted R^2's at all forecasting horizons are fairly consistent with those in Eraker and Wang (2015). Also consistent with Eraker and Wang (2015) is that the adjusted R^2's peaked at the 9-month forecasting horizon uniformly across all models. It is important to mention that our results are slightly higher than theirs, 8.71% for the CIREW model compared to theirs 8.41%. However, we conjecture that this is most likely due to the difference in the data we used for calibrating the pricing models as we covered different sample period. The highest value achieved in this study is the 8.75% from the OUSKST model.

Generally speaking, all three classes of models performed similarly, with no single model dominating the others. However, at the 1-month horizon, the CIRSKST model produced the highest predictability but its performance is much weaker than other models for longer horizons. In contrast, the OUSKST model produced the highest predictability for the majority of the horizons from 3-month to 18-month. Nonetheless, the advantages are marginal. Thus, we can only cautiously conclude that the CIRSKST model may have some advantage in the very short forecasting horizon possibly due to its consistently superior in-sample fit to both VIX and VIX future data. We may also cautiously conclude that the OUSKST model has some marginal advantages over existing models for longer horizons, which is also consistent with its convincingly superior out-of-sample performance for VIX and VIX futures.

6. Conclusion

This paper made three contributions to the literature of modelling VIX and pricing VIX futures using continuous-time diffusions. We systematically studied existing TD models and more importantly we suitably classified all TD models into three clearly defined classes in terms of how the transformation functions are specified. Our classification may help researchers to understand the relative advantages and disadvantages of different TD models and hence provides a useful guide for creating new TD models for practice in the future. Taking the advantage of the distribution-driven approach, we proposed two new models, namely the CIRSKST and the OUSKST models, purposefully designed to directly exploit stylized features in the marginal distribution of VIX. We also derived a closed-form formula for the prices of VIX future contracts, which allowed us to study the VIX and VIX futures markets at the minimum computational expenses. The main contribution of this paper is our fairly systematic and comprehensive empirical comparison of the three classes of TDs for modelling VIX and VIX futures. Our comparison focused on in-sample goodness-of-fit measures and out-of-sample forecasting accuracy for modelling VIX dynamics under both the physical and risk-neutral measures, as well as the stock return predictability of the VRP extracted from competing models, which are all empirical vital criteria for modelling market volatility. As we expected, the two newly proposed distribution-driven models demonstrated their advantages in most of the comparisons we carried out. While both models can fit the data well with the CIRSKST model always producing the best in-sample results, the OUSKST model was particularly powerful in its out-of-sample ability. Our framework allows the pricing of VIX options in very similar fashion. Thus, future work may consider the pricing of VIX futures and options jointly, since the option markets are expected to contain more abundant ex ante information about the market volatility and the financial risk.

Acknowledgment

The authors wish to thank all session participants at the Fourth International Symposium in Computational Economics and Finance (Paris, April 2016) for their constructive comments. The corresponding author gratefully acknowledges the financial support for this project from the ESRC through grant ES/J00622X/1 and the British Academy through grant SG131649.

References

Ahn, D. H., and Gao, B. (1999). A Parametric Non-linear Model of Term Structure Dynamics. Review of Financial Studies, 12, 721-762.

Aït-Sahalia, Y. (1996a). Nonparametric Pricing of Interest Rate Derivatives. Econometrica, 64, 527-560.

Aït-Sahalia, Y. (1996b). Testing Continuous-Time Models of the Spot Interest Rate. Review of Financial Studies, 9, 385-426.

Black, F., and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities, Journal of Political Economy, 81, 637-654.

Bollerslev, T., Tauchen, G., and Zhou H. (2009). Expected Stock Returns and Variance Risk Premia. The Review of Financial Studies, 22, 4463-4492.

Bu, R., Cheng, J., and Hadri, K. (2016a). Reducible Diffusions with Time-Varying Transformations with Application to Short-Term Interest Rates. Economic Modelling, 52, 266-277.

Bu, R., Cheng, J., and Hadri, K. (2016b). Specification Analysis in Regime-Switching Continuous-Time Diffusion Models for Market Volatility. Studies in Nonlinear Dynamics and Economerics, forthcoming.

Bu, R., Giet, L., Hadri, K., and Lubrano, M. (2011). Modelling Multivariate Interest Rates Using Time-Varying Copulas and Reducible Non-linear Stochastic Differential Equations. Journal of Financial Econometrics, 9(1), 198-236.

Chan, K. C., Karolyi, A., Longstaff, F., and Sanders, A. (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. Journal of Finance, 47, 1209-1227.

Choi, S. (2009). Regime-Switching Univariate Diffusion Models of the Short-Term Interest Rate. Studies in Nonlinear Dynamics and Econometrics, 13(1), Article 4.

Conley, T. L., Hansen, L., Luttmer, E., and Scheinkman, J. (1997). Short Term Interest Rates as Subordinated Diffusion. Review of Financial Studies, 10, 525-578.

Cox, J., Ingersoll, J., and Ross, S. (1980). An Analysis of Variable Rate Loan Contracts. Journal of Finance 35, 389-403.

Cox, J., Ingersoll, J., and Ross, S. (1985). In Intertemporal General Equilibrium Model of Asset Prices. Econometrica, 53, 363-384.

Detemple, J., and Osakwe, C. (2000). The Valuation of Volatility Option, European Finance Review, 4, 21-50.

Duffie, D., and Kan, R. (1996). A Yield Factor Model of Interest Rates. Mathematical Finance, 6, 379-406.

Durham, G., and Gallant, R. (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. Journal of Business and Economic Statistics. 20(3) 297-338.

Eraker, B., and Wang, J. (2015). A Non-linear Dynamic Model of the Variance Risk Premium. Journal of Econometrics, 187, 547-556.

Forman, J. L., and Sørensen, M., (2014). A Transformation Approach to Modelling Multi-Modal Diffusions. Journal of Statistical Planning and Inference, 146, 56-69.

Goard, J., and Mazur, M. (2013). Stochastic Volatility Models and the Pricing of VIX Options. Mathematical Finance, 23, 439-458.

Grunbichler, A., and Longstaff, F. (1996). Valuing Futures and Options on Volatility, Journal of Banking Finance, 20, 985-1001.

Hansen, B. (1994). Autoregressive Conditional Density Estimation. International Economic Review, 35, 705-730.

Jondeau, E., and Rockinger, M. (2006). The Copula-GARCH Model of Conditional Dependencies - An International Stock Application. Journal of International Money and Finance, 25, 827-853.

Mencia, J., and Sentana, E. (2013). Valuation of VIX Derivatives. Journal of Financial Economics, 108, 367-391.

Merton, R. C. (1973). Theory of Rational Option Pricing. Bell Journal of Economics and Management Science, 4, 141-183.

Patton, A. (2004). On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation. Journal of Financial Econometrics, 2, 130-168.

Stanton, R. (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. Journal of Finance, 52, 1973-2002.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure, Journal of Financial Economics, 5, 177-188.

Whaley, R. E. (1993). Derivatives on Market Volatility: Hedging Tools Long Overdue, Journal of Derivatives, 1, 71-84.

Figure 1: Time Series of VIX

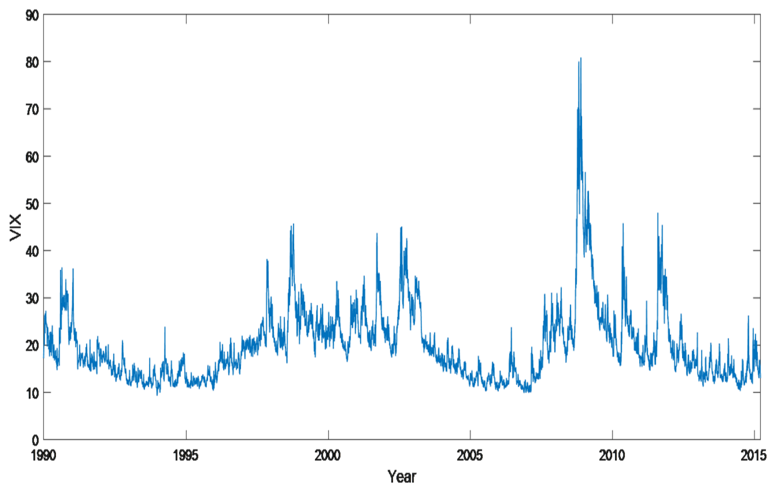

Figure 2: Time Series of Constant Maturity VIX Future Prices

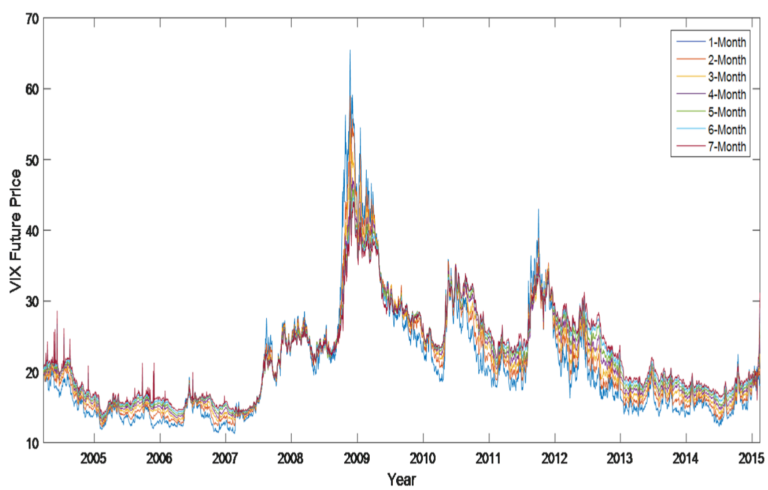

Table 1: Summary of VIX and Constant Maturity VIX Future Prices

	VIX Futures							VIX
	1M	2M	3M	4M	5M	6M	7M	
Correlation	1.00	0.99	0.97	0.96	0.94	0.92	0.91	
	0.99	1.00	1.00	0.99	0.97	0.96	0.95	
	0.97	1.00	1.00	1.00	0.99	0.98	0.97	
	0.96	0.99	1.00	1.00	1.00	0.99	0.99	
	0.94	0.97	0.99	1.00	1.00	1.00	0.99	
	0.92	0.96	0.98	0.99	1.00	1.00	1.00	
	0.91	0.95	0.97	0.99	0.99	1.00	1.00	
Eigenvalue	6.85	0.14	0.01	0.00	0.00	0.00	0.00	
Maximum	65.46	59.00	53.78	49.73	46.71	44.64	44.08	80.86
Minimum	11.29	12.23	12.54	12.87	13.20	13.53	13.69	9.89
Mean	21.41	22.02	22.37	22.64	22.88	23.09	23.25	20.61
Median	19.41	20.81	21.68	22.16	22.70	23.16	23.41	17.63
Std. Dev.	8.77	8.02	7.46	7.10	6.85	6.67	6.51	10.19
Skewness	1.62	1.31	1.05	0.84	0.71	0.60	0.54	2.21
Kurtosis	6.14	5.09	4.13	3.44	3.07	2.84	2.72	9.25

Table 2: ML Estimation Results for VIX

	EW	CIRCEV	CIRSKST	OUDO	OUCEV	OUSKST
κ	3.7553	3.8678	2.9601	4.0219	4.1021	3.4136
	(0.5562)	(0.5605)	(0.4376)	(0.5718)	(0.5778)	(0.4730)
θ	0.0508	0.1627	0.9340	2.9295	-0.8009	0
	(0.0031)	(0.0462)	(0.4615)	(0.0489)	(0.1140)	
σ	0.2175	0.3027	1	0.9838	0.3030	1
	(0.0190)	(0.0234)		(0.0088)	(0.0233)	
γ		1.3958			1.3955	
		(0.0262)			(0.0261)	
φ	0.7315					
	(0.7003)					
α	0.0033					
	(0.0031)					
m			18.0955			17.8926
			(0.6148)			(0.6629)
ς			7.3174			7.3328
			(0.6803)			(0.9703)
ν			6.8309			4.8395
			(1.4024)			(0.6726)
λ			0.6242			0.6371
			(0.0472)			(0.0328)
LL ($\times 10^3$)	-9.6909	-9.6896	-9.6662	-9.8128	-9.6886	-9.6673
AIC ($\times 10^4$)	1.9392	1.9387	1.9344	1.9632	1.9385	1.9345
BIC ($\times 10^4$)	1.9426	1.9414	1.9385	1.9652	1.9412	1.9378
$RMSFE$	1.2784	1.2778	1.2744	1.2753	1.2764	1.2722

Table 3: Joint ML Estimation Results (Natural Maturity VIX Futures)

	EW	CIRCEV	CIRSKST	OUDO	OUCEV	OUSKST
κ	3.6111	4.1315	2.5786	4.7369	4.3672	2.5341
	(0.5608)	(0.5776)	(0.1283)	(0.6251)	(0.5951)	(0.0473)
θ	0.0558	0.0277	0.3597	2.9293	-0.2798	0
	(0.0041)	(0.0022)	(0.0175)	(0.0451)	(0.0133)	
σ	0.3069	0.1838	1	1.0682	0.1656	1
	(0.0027)	(0.0042)		(0.0123)	(0.0052)	
γ		1.5766			1.6105	
		(0.0067)			(0.0087)	
φ	-2.0272					
	(0.0003)					
α	0.0100					
	(0.0000)					
m			27.0558			23.5662
			(0.5075)			(0.4662)
ς			11.9904			30.7818
			(0.2824)			(0.3663)
ν			28.3530			2.1919
			(0.1885)			(0.0259)
λ			0.6918			0.7490
			(0.0591)			(0.0117)
κ^Q	0.7281	0.9486	0.6346	1.3353	1.0709	1.1892
	(0.0198)	(0.0166)	(0.0267)	(0.0107)	(0.0167)	(0.0155)
θ^Q	0.0855	0.0367	0.7882	2.9874	-0.2942	0.1222
	(0.0024)	(0.0021)	(0.0204)	(0.0080)	(0.0120)	(0.0173)
LL ($\times 10^4$)	-5.6257	-5.6537	-5.6042	-5.8057	-5.6702	-5.6765
AIC ($\times 10^5$)	1.1253	1.1309	1.1210	1.1612	1.1342	1.1354
BIC ($\times 10^5$)	1.1258	1.1314	1.1217	1.1616	1.1347	1.1360
$RMSE$	2.7529	2.7931	2.7173	3.0025	2.8162	2.8226
$RMSFE\_F$	1.5933	1.5426	1.1038	1.2798	1.5098	1.0312

Table 4: Joint ML Estimation Results (Constant Maturity VIX Futures)

	EW	CIRCEV	CIRSKST	OUDO	OUCEV	OUSKST
κ	3.9252	3.7733	2.0326	4.1419	4.0618	3.1945
	(0.5181)	(0.5500)	(0.0433)	(0.5810)	(0.5732)	(0.0663)
θ	0.0476	0.0381	0.4465	2.9294	-0.3017	0
	(0.0028)	(0.0038)	(0.0054)	(0.0482)	(0.0166)	
σ	0.2737	0.1942	1	0.9984	0.1675	1
	(0.0026)	(0.0056)		(0.0095)	(0.0060)	
γ		1.5415			1.5939	
		(0.0090)			(0.0105)	
φ	-1.1504					
	(0.0022)					
α	0.0100					
	(0.0001)					
m			26.2446			19.4947
			(0.3617)			(0.4528)
ς			11.8907			10.1070
			(0.3826)			(0.2587)
ν			27.1275			3.1463
			(0.3867)			(0.0551)
λ			0.6558			0.6078
			(0.0093)			(0.0098)
κ^Q	0.7831	1.0227	0.7108	1.2880	1.1403	1.2931
	(0.0211)	(0.0185)	(0.0153)	(0.0125)	(0.0194)	(0.0173)
θ^Q	0.0702	0.0459	0.7632	3.0127	-0.3116	-0.0304
	(0.0018)	(0.0036)	(0.0110)	(0.0068)	(0.0151)	(0.0334)
LL $(\times 10^4)$	-3.5515	-3.5665	-3.5430	-3.6495	-3.5754	-3.5783
AIC $(\times 10^4)$	7.1044	7.1342	7.0877	7.3001	7.1520	7.1579
BIC $(\times 10^4)$	7.1098	7.1388	7.0939	7.3039	7.1566	7.1634
$RMSE$	2.6161	2.6507	2.5853	2.8335	2.6698	2.6807
$RMSFE\_F$	11.8824	11.7330	9.6216	9.7756	11.5153	4.7618
$RMSFE\_1M$	1.4497	1.3996	1.1600	1.2370	1.3658	0.9156
$RMSFE\_3M$	2.8066	2.7686	1.9350	2.0893	2.7500	0.6738
$RMSFE\_5M$	3.6145	3.5867	2.8626	2.8811	3.5190	1.1220
$RMSFE\_7M$	4.0117	3.9781	3.6640	3.5682	3.8805	2.0504

Table 5: Market Return Regression Adjusted R^2

	1M	3M	6M	9M	12M	15M	18M	24M
EW	0.52	5.66	7.08	8.71	8.37	6.89	5.42	5.26
CIRCEV	0.52	5.65	7.08	8.70	8.35	6.87	5.41	5.24
CIRSKST	0.80	5.36	6.78	8.42	8.11	6.63	5.14	5.14
OUDO	0.22	6.05	7.37	8.51	8.00	6.56	5.30	4.69
OUCEV	0.39	5.88	7.27	8.70	8.28	6.81	5.43	5.05
OUSKST	0.48	6.21	7.48	8.75	8.33	6.95	5.50	5.16