Diffusion Copulas: Identification and Estimation

Ruijun Bu[*] Kaddour Hadri[†] Dennis Kristensen[‡]

April 2020

Abstract

We propose a new semiparametric approach for modelling nonlinear univariate diffusions, where the observed process is a nonparametric transformation of an underlying parametric diffusion (UPD). This modelling strategy yields a general class of semiparametric Markov diffusion models with parametric dynamic copulas and nonparametric marginal distributions. We provide primitive conditions for the identification of the UPD parameters together with the unknown transformations from discrete samples. Likelihood-based estimators of both parametric and nonparametric components are developed and we analyze the asymptotic properties of these. Kernel-based drift and diffusion estimators are also proposed and shown to be normally distributed in large samples. A simulation study investigates the finite sample performance of our estimators in the context of modelling US short-term interest rates. We also present a simple application of the proposed method for modelling the CBOE volatility index data.

JEL Classification: C14, C22, C32, C58, G12
Keywords: Continuous-time model; diffusion process; copula; transformation model; identification; nonparametric; semiparametric; maximum likelihood; sieve; kernel smoothing.

[*]Department of Economics, Management School, University of Liverpool, Liverpool, UK. Email: ruijunbu@liv.ac.uk.

[†]Queen's University Management School, Queen's University Belfast, Belfast, Northern Ireland, UK. Email: k.hadri@qub.ac.uk.

[‡]Department of Economics, University College London, London, UK. Email: d.kristensen@ucl.ac.uk.

1 Introduction

Most financial time series have fat tails that standard parametric models are not able to generate. One forceful argument for this in the context of diffusion models was provided by Aït-Sahalia (1996b) who tested a range of parametric models against a nonparametric alternative and found that most standard models were inconsistent with observed features in data.

One popular semiparametric approach that allows for more flexibility in terms of marginal distributions, and so allowing for fat tails, is to use the so-called copula models, where the copula is parametric and the marginal distribution is left unspecified (nonparametric). Joe (1997) showed how bivariate parametric copulas could be used to model discrete-time stationary Markov chains with flexible, nonparametric marginal distributions. The resulting class of semiparametric models are relatively easy to estimate; see, e.g. Chen and Fan (2006). However, most parametric copulas known in the literature have been derived in a cross-sectional setting where they have been used to describe the joint dependence between two random variables with known joint distribution, e.g. a bivariate t-distribution. As such, existing parametric copulas may be difficult to interpret in terms of the dynamics they imply when used to model Markov processes. This in turn means that applied researchers may find it difficult to choose an appropriate copula for a given time series.

One could have hoped that copulas with a clearer dynamic interpretation could be developed by starting with an underlying parametric Markov model and then deriving its implied copula. This approach is unfortunately hindered by the fact that the stationary distributions of general Markov chains are not available on closed-form and so their implied dynamic copulas are not available on closed form either. This complicates both the theoretical analysis (such as establishing identification) and the practical implementation of such models.

An alternative approach to modelling fat tails using Markov diffusions is to specify flexible forms for the so-called drift and diffusion term. Such non-linear features tend to generate fat tails in the marginal distribution of the process. This approach has been widely used to, for example, model short-term interest rates; see, e.g., Aït-Sahalia (1996a,b), Conley et al. (1997), Stanton (1997), Ahn and Gao (1999) and Bandi (2002). These models tend to either be heavily parameterized or involve nonparametric estimators that suffer from low precision in small and moderate samples.

We here propose a novel class of dynamic copulas that resolves the above-mentioned issues: We show how copulas can easily be generated from parametric diffusion processes. The copulas have a clear interpretation in terms of dynamics since they are constructed from an underlying dynamic continuous-time process. At the same time, a given copula-based diffusion can exhibit strong non-linearities in its drift and diffusion term even if the underlying copula is derived from, for example, a linear model. Furthermore, primitive conditions for identification of the parameters are derived; and this despite the fact that the copulas are implicit. Finally, the models can easily be implemented in practice using existing numerical methods for parametric diffusion processes. This in turn implies that estimators are easy to compute and do not involve any smoothing parameters; this is in contrast to existing semi- and nonparametric estimators of diffusion models.

The starting point of our analysis is to show that there is a one-to-one correspondence between any given semiparametric Markov copula model and a model where we observe a nonparametric transformation of an underlying parametric Markov process. We then restrict attention to parametric Markov diffusion processes which we refer to as underlying parametric diffusions (UPD's). Copulas generated from a given UPD has a clear interpretation in terms of dynamic properties. In particular, standard results from the literature on diffusion models can be employed to establish mixing properties and existence of moments for a given model; see, e.g. Chen et al. (2010). Moreover, we are able to derive primitive conditions for the parameters of the copula to be identified together with the unknown transformation.

Once identification has been established, estimation of our copula diffusion models based on a discretely sampled process proceeds as in the discrete-time case. One can either estimate the model using a one-step or two-step procedure: In the one-step procedure, the marginal distribution and the parameters of the UPD are estimated jointly by sieve-maximum likelihood methods as advocated by Chen, Wu and Yi (2009). In the two-step approach, the marginal distribution is first estimated by the empirical cdf, which in turn is plugged into the likelihood function of the model. This is then maximized with respect to the parameters of the UPD. We provide an asymptotic theory for both cases by importing results from Chen, Wu and Yi (2009) and Chen and Fan (2006), respectively. In particular, we provide primitive conditions for their high-level assumptions to hold in our diffusion setting. The resulting asymptotic theory shows $\sqrt{n}$-asymptotic normality of the parametric components. Given the estimates of parametric component, one can obtain semiparametric estimates of the drift and diffusion functions and we also provide an asymptotic theory for these.

Our modelling strategy has parametric ascendants: Bu et al. (2011), Eraker and Wang (2015) and Forman and Sørensen (2014) considered parametric transformations of UPDs for modelling short-term interest rates, variance risk premia and molecular dynamics, respectively. We here provide a more flexible class of models relative to theirs since we leave the transformation unspecified. At the same time, all the attractive properties of their models remain valid: The transition density of the observed process is induced by the UPD and so the estimation of copula-based diffusion models is computationally simple. Moreover, copula diffusion models can furthermore be easily employed in asset pricing applications since (conditional) moments are easily computed using the specification of the UPD. Finally, none of these papers fully addresses the identification issue and so our identification results are also helpful in their setting.

There are also similarities between our approach and the one pursued in Aït-Sahalia (1996a) and Kristensen (2010). They developed two classes of semiparametric diffusion models where either the drift or the diffusion term is specified parametrically and the remaining term is left unspecified. The remaining term is then recovered by using the triangular link between the marginal distribution, the drift and the diffusion terms that exist for stationary diffusions. In this way, the marginal distribution implicitly ties down the dynamics of the observed diffusion process. Unfortunately, it is very difficult to interpret the dynamic properties of the resulting semiparametric diffusion model. In contrast, in our setting, the UPD alone ties down the dynamics of the observed diffusion and so these are much better understood. The estimation of copula diffusions are also less computationally burdensome compared to the Pseudo Maximum Likelihood Estimator (PMLE) proposed in Kristensen (2010).

The remainder of this paper is organized as follows. Section 2 outlines our semiparametric modelling strategy. Section 3 investigates the identification issue of our model. In Section 4, we discuss the estimators of our model while Section 5 investigates their asymptotic properties. Section 6 presents a simulation study to examine the finite sample performance of our estimators. In Section 7, we consider a simple empirical application. Some concluding remarks are given in Section 8. All proofs and lemmas are collected in Appendices.

2 Copula-Based Diffusion Models

2.1 Framework

Consider a continuous-time process $Y = \{Y_t : t \geq 0\}$ with domain $\mathcal{Y} = (y_l, y_r)$, where $-\infty \leq y_l < y_r \leq +\infty$. We assume that Y satisfies

$$Y_t = V\left(X_t\right), \tag{2.1}$$

where $V : \mathcal{X} \mapsto \mathcal{Y}$ is a smooth monotonic univariate function and $X = \{X_t : t \geq 0\}$ solves the following parametric SDE:

$$dX_t = \mu_X\left(X_t; \theta\right) dt + \sigma_X\left(X_t; \theta\right) dW_t. \tag{2.2}$$

Here, $\mu_X\left(x; \theta\right)$ and $\sigma_X^2\left(x; \theta\right)$ are scalar functions that are known up to some unknown parameter vector $\theta \in \Theta$, where Θ is the parameter space, while W is a standard Brownian motion. We call X the underlying parametric diffusion (UPD) and let $\mathcal{X} = (x_l, x_r)$, $-\infty \leq x_l < x_r \leq +\infty$, denote its domain.

We call Y a *copula-based diffusion* since its dynamics are determined by the implied (dynamic) copula of the UPD X, as we will explain below. Given a discrete sample of Y, $Y_{i\Delta}$, $i = 0, 1, \ldots, n$, where $\Delta > 0$ denotes the time distance between observations, we are then interested in drawing inference regarding the parameter θ and the function V. Note here that we only observe Y while X remains unobserved since we leave V unspecified (unknown to us). For convenience, we collect the unknown component in the *structure* $\mathcal{S} \equiv (\theta, V)$.

The above class of models allows for added flexibility through the transformation V which we treat as a nonparametric object that we wish to estimate together with θ. By allowing for a broad nonparametric class of transformations V, our model is richer and more flexible compared to the fully parametric case with known or parametric specifications of V. In particular, as we shall see, any given member of the above class of models is able to completely match the marginal distribution of any given time series.

We will require that the underlying Markov process X sampled at $i\Delta$, $i = 1, 2, \ldots$, possesses a transition density $p_X\left(x|x_0; \theta\right)$,

$$\Pr\left(X_\Delta \in \mathcal{A} | X_0 = x_0\right) = \int_{\mathcal{A}} p_X\left(x|x_0; \theta\right) dx, \quad \mathcal{A} \subseteq \mathcal{X}. \tag{2.3}$$

Moreover, some of our results require X to be recurrent, a property which can be stated in terms of the so-called scale density and scale measure. These are defined as

$$s(x;\theta) := \exp\left\{-\int_{x^*}^{x} \frac{2\mu_X(z;\theta)}{\sigma_X^2(z;\theta)}dz\right\} \text{ and } S(x;\theta) := \int_{x^*}^{x} s(z;\theta)\,dz \tag{2.4}$$

for some $x^* \in \mathcal{X}$. We then impose the following:

Assumption 2.1. (i) $\mu_X(\cdot;\theta)$ and $\sigma_X^2(\cdot;\theta) > 0$ are twice continuously differentiable; (ii) the scale measure satisfies $S(x;\theta) \to -\infty\ (+\infty)$ as $x \to x_l\ (x_r)$; (iii) $\xi(\theta) = \int_{\mathcal{X}} \left\{\sigma_X^2(x;\theta)\,s(x;\theta)\right\}^{-1} dx < \infty$.

Assumption 2.2. The transformation V is strictly increasing with inverse $U = V^{-1}$, i.e., $y = V(x) \Leftrightarrow x = U(y)$, and is twice continuously differentiable.

Assumption 2.1(i) provides primitive conditions for a solution to eq. (2.2) to exist and for the transition density $p_X(x|x_0;\theta)$ to be well-defined, while Assumption 2.1(ii) implies that this solution is positive recurrent; see Bandi and Phillips (2003), Karatzas and Shreve (1991, Section 5.5) and McKean (1969, Section 5) for more details. Assumption 2.1(iii) strengthens the recurrence property to stationarity and ergodicity in which case the stationary marginal density of X takes the form

$$f_X(x;\theta) = \frac{\xi(\theta)}{\sigma_X^2(x;\theta)\,s(x;\theta)}, \tag{2.5}$$

where $\xi(\theta)$ was defined in Assumption 2.1(iii). However, stationarity will not be required for all our results to hold; in particular, some of our identification results and proposed estimators do not rely on stationarity. This is in contrast to the existing literature on dynamic copula models where stationarity is a maintained assumption.

Assumption 2.2 requires V to be strictly increasing; this is a testable restriction under the remaining assumptions introduced below which ensures identification: Suppose that indeed V is strictly decreasing; we then have $Y_t = \bar{V}\left(\bar{X}_t\right)$, where $\bar{V}(x) = V(-x)$ is increasing and $\bar{X}_t = -X_t$ has dynamics $p_X(-x|-x_0;\theta)$. Assuming that the chosen UPD satisfies $p_X(-x|-x_0;\theta) \neq p_X(x|x_0;\tilde{\theta})$ for $\theta \neq \tilde{\theta}$, we can test whether V indeed is decreasing or increasing.

The smoothness condition on V is imposed so that we can employ Ito's Lemma on the transformation to obtain that the continuous-time dynamics of Y can be written in terms of $\mathcal{S}$ as

$$dY_t = \mu_Y(Y_t;\mathcal{S})\,dt + \sigma_Y(Y_t;\mathcal{S})\,dW_t,$$

with

$$\mu_Y(y;\mathcal{S}) = \frac{\mu_X(U(y);\theta)}{U'(y)} - \frac{1}{2}\sigma_X^2(U(y);\theta)\frac{U''(y)}{U'(y)^3}, \tag{2.6}$$

$$\sigma_Y(y;\mathcal{S}) = \frac{\sigma_X(U(y);\theta)}{U'(y)}, \tag{2.7}$$

where we have used that, with $U'(y)$ and $U''(y)$ denoting the first two derivatives of $U(y)$, $V'(U(y)) = 1/U'(y)$ and $V''(U(y)) = -U''(y)/U'(y)^3$. In particular, Y is a Markov diffusion process. As can be seen from the above expressions, the dynamics of Y, as characterized by μ_Y and σ_Y^2, may appear quite complex with U potentially generating nonlinearities in both the drift and diffusion terms even if μ_X and σ_X^2 are linear. We demonstrate this feature in the subsequent subsection where we present examples of simple UPD's are able to generate non-linear shapes of μ_Y and σ_Y^2 via the non-linear transformation V. At the same time, if we transform Y by U we recover the dynamics of the UPD. As a consequence, the transition density of the discretely sampled process $Y_{i\Delta}$, $i = 0, 1, 2, ...$, can be expressed in terms of the one of X as

$$p_Y(y|y_0; \mathcal{S}) = U'(y)\, p_X(U(y)\,|U(y_0); \theta), \tag{2.8}$$

using standard results for densities of invertible transformations. By similar arguments, the stationary density of Y satisfies

$$f_Y(y; \mathcal{S}) = U'(y)\, f_X(U(y); \theta), \tag{2.9}$$

which shows that any choice for UPD is able to fully adapt to any given marginal density of Y due to the nonparametric nature of U.

The above expressions also highlights the following additional theoretical and practical advantages of our modelling strategy: First, for a given choice of U, we can easily compute $p_Y(y|y_0; \mathcal{S})$ and $f_Y(y; \mathcal{S})$ since computation of parametric transition densities and stationary densities of diffusion models is in general straightforward, even if they are not available on closed form. Second, Y inherits all its dynamic properties from X; and in the modelling of X, we can rely on a large literature on parametric modelling of diffusion models. Formally, we have the following straightforward results adopted from Forman and Sørensen (2014).

Proposition 2.1 *Suppose that Assumptions 2.1(i)–(ii) and 2.2 hold. Then the following results hold for the model (2.1)-(2.2):*

1. *If Assumption 2.1(iii) hold, then X is stationary and ergodic and so is Y.*
2. *The mixing coefficients of X and Y coincide.*
3. *If $E[|X_t|^{q_1}] < \infty$ and $|V(x)| \leq B(1 + |x|^{q_2})$ for some $B < \infty$ and $q_1, q_2 \geq 0$, then $E[|Y_t|^{q_1/q_2}] < \infty$.*
4. *If φ is an eigenfunction of X with corresponding eigenvalue ρ in the sense that $E[\varphi(X_1)\,|X_0] = \rho\varphi(X_0)$ then $\varphi \circ U$ is an eigenfunction of Y with corresponding eigenvalue ρ.*

The above theorem shows that, given knowledge (or estimates) of $\mathcal{S}$, the properties of Y in terms of mixing coefficients, moments, and eigenfunctions are well-understood since they are inherited from the specification of X. In addition, computations of conditional moments of Y can be done straightforwardly utilizing knowledge of the UPD. For example, for a given function G, the corresponding conditional moment can be computed as

$$E[G(Y_{t+s})\,|Y_t = y] = E[G_X(X_{t+s})\,|X_t = U(y)], \text{ where } G_X(x) := G(V(x)).$$

The right-hand side moment only involves X and so standard methods for computing moments of parametric diffusion models (e.g., Monte Carlo methods, solving partial differential equations, Fourier transforms) can be employed. This facilitates the use of our diffusion models in asset pricing where the price often takes the form of a conditional moment. We refer to Eraker and Wang (2015) for more details on asset pricing applications for our class of models; they take a fully parametric approach but all their arguments carry over to our setting.

The last result of the above theorem will prove useful for our identification arguments since these will rely on the fundamental nonparametric identification results derived in Hansen et al. (1998). Their results involve the spectrum of the observed diffusion process, and the last result of the theorem implies that the spectrum of Y is fully characterized by the spectrum of X together with the transformation. The eigenfunctions and their eigenvalues are also useful for evaluating long-run properties of Y. In our semiparametric approach, the eigenfunctions and corresponding eigenvalues of Y are easily computed from X and so we circumvent the problem of estimating these nonparametrically as done in, for example, Chen, Hansen and Scheinkman (2009) and Gobet et al. (2004).

2.2 Examples of UPDs

Our framework is quite flexible and in principle allows for any specification of the UPD for X. Many parametric models are available for that purpose, and we here present three specific examples from the literature on continuous-time interest rate modelling.

Example 1: Ornstein-Uhlenbeck (OU) model. The OU model (c.f. Vasicek, 1977) is given by

$$dX_t = \kappa \left(\alpha - X_t\right) dt + \sigma dW_t, \tag{2.10}$$

defined on the domain $\mathcal{X} = (-\infty, +\infty)$. The process is stationary if and only if $\kappa > 0$, in which case X mean-reverts to its unconditional mean α. The scale of X is controlled by σ. Its stationary and transition distributions are both normal, and the corresponding copula of the discretely sampled process is a Gaussian copula with correlation parameter $e^{-\kappa\Delta}$. For this particular model, the resulting drift and diffusion term of the observed process takes the form

$$\mu_Y \left(y; \mathcal{S}\right) = \frac{\kappa \left(\alpha - U\left(y\right)\right)}{U'\left(y\right)} - \frac{1}{2}\sigma^2 \frac{U''\left(y\right)}{U'\left(y\right)^3}, \quad \sigma_Y^2 \left(y; \mathcal{S}\right) = \frac{\sigma^2}{U'\left(y\right)^2}. \tag{2.11}$$

In Figure 2 (found in Section 6), we plot these two functions with U and θ fitted to the 7-day Eurodollar interest rate time series used in Aït-Sahalia (1996b). Observe that U generates non-linear behavior in μ_Y and σ_Y^2 despite the UPD being a linear Gaussian process.

Example 2: Cox-Ingersoll-Ross (CIR) model. The CIR process (c.f. Cox et al., 1985) is given by

$$dX_t = \kappa \left(\alpha - X_t\right) dt + \sigma \sqrt{X_t} dW_t. \tag{2.12}$$

The process has domain $\mathcal{X} = (0, +\infty)$ and is stationary if and only if $\kappa > 0$, $\alpha > 0$ and $2\kappa\alpha/\sigma^2 \geq 1$. Conditional on $X_{i\Delta}$, $X_{(i+1)\Delta}$ admits a non-central χ^2 distribution with fractional degrees of freedom while its stationary distribution is a Gamma distribution. To our best knowledge, the corresponding dynamic copula has not been analyzed before or used in empirical work. Figure 4 (in Section 6) displays μ_Y and σ_Y^2, with U and θ chosen in the same way as in Exampe 1. Compared to this example, the resulting drift and diffusion term of Y exhibit even stronger non-linearities.

Example 3: Nonlinear Drift Constant Elasticity Variance (NLDCEV) model. The NLDCEV specification (c.f. Conley et al., 1997) is given by

$$dX_t = \left(\sum_{i=-k}^{l} \alpha_i X_t^i \right) dt + \sigma X_t^{\beta} dW_t \tag{2.13}$$

with domain $\mathcal{X} = (0, +\infty)$. It is easily seen that when $\alpha_{-k} > 0$ and $\alpha_l < 0$ the drift term of the diffusion in (2.13) exhibits mean-reversions for large and small values of X. A popular choice for various studies in finance assumes that $k = 1$ and $l = 2$ or 3 (c.f. Aït-Sahalia, 1996b; Choi, 2009; Kristensen, 2010; Bu, Cheng and Hadri, 2017), in which case the drift has linear or zero mean-reversion in the middle part and much stronger mean-reversion for large and small values of X. Meanwhile, the CEV diffusion term is also consistent with most empirical findings of the shape of the diffusion term. It follows that since (2.13) is one of the most flexible parametric diffusions, diffusion processes that are unspecified transformations of (2.13) should represent a very flexible class of diffusion models. Similar to (2.12), the implied copula of the NLDCEV is new to the copula literature.

Examples 1-2 are attractive from a computational standpoint since the corresponding transition densities are available on closed-form thereby facilitating their implementation. But this comes at the cost of the dynamics being somewhat simple. The NLDCEV model implies more complex and richer dynamics but on the other hand its transition density is not available on closed form. However, the marginal pdf of the NLDCEV process, as well as more general specifications, can be evaluated in closed form by (2.5). Moreover, closed-from approximations of the transition density of the NLDCEV model developed by, for example, Aït-Sahalia (2002) and Li (2013) can be employed. Alternatively, simulated versions of the transition density can be computed using the techniques developed in, for example, Kristensen and Shin (2012) and Bladt and Sørensen (2014). In either case, an approximate version of the exact likelihood can be easily computed, thereby allowing for simple estimation of even quite complex underlying UPDs.

2.3 Related Literature

As already noted in the introduction, copula-based diffusions are related to the class of so-called *discrete-time* copula-based Markov models; see, for example, Chen and Fan (2006) and references therein. To map the notation and ideas of this literature into our continuous-time setting, we set the sampling time distance $\Delta = 1$ in the remaining part of this section.

Let us first introduce copula-based Markov models where a given discrete-time, stationary scalar Markov process $Y = \{Y_i : i = 0, 1, \ldots, n\}$ is modelled through a bivariate parametric copula density[1], say, $c_X(u_0, u; \theta)$, together with its stationary marginal cdf F_Y, i.e., so that Y's transition density satisfies

$$p_Y(y|y_0; \theta, F_Y) = f_Y(y)\, c_X(F_Y(y_0), F_Y(y); \theta), \tag{2.14}$$

where $f_Y(y) = F_Y'(y)$. An alternative representation of this model is

$$Y_i = F_Y^{-1}\left(\bar{X}_i\right), \quad \bar{X}_{i+1}|\bar{X}_i = x_0 \sim c_X(x_0, \cdot; \theta), \tag{2.15}$$

so that Y_i is a transformation of an underlying Markov process $\bar{X}_i \in [0,1]$; the latter having a uniform marginal distribution and transition density $c_X(x_0, x; \theta)$. Thus, if $c_X(x_0, x; \theta)$ is induced by an underlying Markov diffusion transition density, the corresponding copula-based Markov model falls within our framework.

Reversely, consider a copula-based diffusion and suppose that the UPD X is stationary with marginal cdf $F_X(x; \theta)$. By definition of Y, its marginal cdf satisfies

$$F_Y(y) = F_X(U(y); \theta) \Leftrightarrow U(y) = F_X^{-1}(F_Y(y); \theta). \tag{2.16}$$

Substituting the last expression for U into (2.8), we see that p_Y can be expressed in the form of (2.14) where $c_X(u_0, u; \theta)$ is the density function of the (dynamic) copula implied by the discretely sampled UPD X,

$$c_X(u_0, u; \theta) = \frac{p_X\left(F_X^{-1}(u; \theta) | F_X^{-1}(u_0; \theta); \theta\right)}{f_X\left(F_X^{-1}(u; \theta); \theta\right)}. \tag{2.17}$$

Thus, any discretely sampled stationary copula-based diffusion satisfies (2.15) with $\bar{X}_i = F_X(X_i)$.

However, the literature on copula-based Markov models focus on discrete-time models with standard copula specifications derived from bivariate distributions in an i.i.d. setting. Using copulas that are originally derived in an i.i.d. setting complicates the interpretation of the dynamics of the resulting Markov model, and conditions for the model to be mixing, for example, can be quite complicated to derive; see, e.g., Beare (2010) and Chen, Wu and Yi (2009). This also implies that very few standard copulas can be interpreted as diffusion processes; to our knowledge, the only one is the Gaussian copula which corresponds to the OU process in Example 1.

The reader may now wonder why we do not simply generate dynamic copulas by first deriving the transition density $p_X(x|x_0; \theta)$ for a given discrete-time Markov model and then obtain the corresponding Markov copula through eq. (2.17)? The reason is that for most discrete-time Markov models the stationary distribution $F_X(x; \theta)$ is not known on closed form. Thus, first of all, $F_X^{-1}(u; \theta)$ and thereby also c_X have be approximated numerically. Second, since c_X is now not available on closed form, the analysis of which parameters one can identify from the resulting copula model becomes very challenging. And identification in copula-based Markov models is a non-trivial problem: Generally, for a given parametric Markov model, not all parameters are identified from the corresponding copula as given in (2.17) and some of them have to be normalized.

[1]The copula $C_X(u_0, u_1; \theta)$ for a given Markov process is defined as

$$C_X(u_0, u_1; \theta) = \Pr\left(X_0 \leq F_X^{-1}(u_0; \theta), X_1 \leq F_X^{-1}(u_1; \theta)\right).$$

The corresponding copula density is then given by $c_X(u_0, u_1; \theta) = \partial^2 C_X(u_0, u_1; \theta) / (\partial u_0 \partial u_1)$.

We here directly generate copulas through an underlying continuous-time diffusion model for X. This resolves the aforementioned drawbacks of existing copula-based Markov models: First, we are able to generate highly flexible copulas so far not considered in the literature. Second, given that our copulas are induced by specifying the drift and diffusion functions of X, the time series properties are much more easily inferred from our model, c.f. Theorem 2.1. Third, by Ito's Lemma, eqs. (2.6)-(2.7) provide us with explicit expressions linking the drift and diffusion terms of the observed diffusion process Y to the UPD through the transformation V; this will allow us to derive necessary and sufficient conditions for identification in the following. Fourth, in terms of estimation, the stationary distribution of a given diffusion model has an explicit form, c.f. eq. (2.5), which allows us to develop computationally simple estimators of copula diffusion models. Finally, some of our identification results will not require stationarity and so expands the scope for using copula-type models in time series analysis.

Our modelling strategy is also related to the ideas of Aït-Sahalia (1996a) and Kristensen (2010, 2011) where F_Y is left unspecified while either the drift, μ_Y, or the diffusion term, σ_Y^2, is specified parametrically. As an example, consider the former case where $\sigma_Y^2(y;\theta)$ is known up to the parameter θ. Given knowledge of the marginal density f_Y (or a nonparametric estimator of it), the diffusion term can then be recovered as a functional of f_Y and μ_Y as

$$\mu_Y(y; f_Y, \theta) = \frac{1}{2f_Y(y)} \frac{\partial}{\partial y} \left[\sigma_Y^2(y;\theta) f_Y(y) \right].$$

So in their setting f_Y pins down the resulting dynamics of Y in a rather opaque manner.

3 Identification

Suppose that a particular specification of the UPD as given in (2.2) has been chosen. Given the discrete sample of Y, the goal is to obtain consistent estimates of θ together with V. To this end, we first have to show that these are actually identified from data. In order to do so, we need to be precise about which primitives we can identify from data. Given the primitives, we then wish to recover (θ, V). In the cross-sectional literature, one normally take as given the distribution of data and then establish a mapping between this and the structural parameters. In our setting, we are able to learn about the transition density of our data, p_Y, from the population and so it would be natural to use this as primitive from which we wish to recover (θ, V). However, the mapping from p_Y to (θ, V) is not available on closed form in general in our setting and so this identification strategy appears highly complicated. Instead we will take as primitives the drift, μ_Y, and diffusion term, σ_Y^2, of Y and then show identification of (θ, V) from these. This identification argument relies on us being able to identify μ_Y and σ_Y^2 in the first place, which we formally assume here:

Assumption 3.1 The drift, μ_Y, and the diffusion, σ_Y^2, are nonparametrically identified from the discretely sampled process Y.

The above assumption is not completely innocuous and does impose some additional regularity conditions on the Data Generating Process (DGP). We therefore first provide sufficient conditions under which Assumption 3.1 holds. The first set of conditions are due to Hansen et al. (1998) who showed that Assumption 3.1 is satisfied if Y is stationary and its infinitesimal operator has a discrete spectrum. Theorem 2.1(4) is helpful in this regard since it informs us that the spectrum of Y can be recovered from the one of X. In particular, if X is stationary with a discrete spectrum, then Y will have the same properties. Since the dynamics of X is known to us, the properties of its spectrum are in principle known to us and so this condition can be verified a priori. The second set of primitive conditions come from Bandi and Phillips (2003): They show that as $\Delta \to 0$ and $n\Delta \to \infty$, the drift and diffusion functions of a recurrent Markov diffusion process are identified. This last result holds without stationarity, but on the other hand requires high-frequency observations.

In order to formally state the above two results, we need some additional notation. Recall that the infinitesimal operator, denoted L_X, of a given UPD X is defined as

$$L_{X,\theta} g(x) := \mu_X(x;\theta) g'(x) + \frac{1}{2}\sigma_X^2(x;\theta) g''(x),$$

for any twice differentiable function $g(x)$. We follow Hansen et al. (1998) and restrict the domain of L_X to the following set of functions:

$$\mathcal{D}(L_{X,\theta}) = \left\{ g \in L_2(f_X) : g' \text{ is a.c., } L_{X,\theta} g \in L_2(f_X) \text{ and } \lim_{x \downarrow x_l} \frac{g'(x)}{s(x)} = \lim_{x \uparrow x_u} \frac{g'(x)}{s(x)} = 0 \right\}.$$

where a.c. stands for absolutely continuous. The spectrum of $L_{X,\theta}$ is then the set of solution pairs (φ, ρ), with $\varphi \in \mathcal{D}(L_{X,\theta})$ and $\rho \geq 0$, to the following eigenvalue problem, $L_{X,\theta}\varphi = -\rho\varphi$. We refer to Hansen et al. (1998) and Kessler and Sørensen (1999) for a further discussion and results regarding the spectrum of L_X. The following result then holds:

Proposition 3.1 *Suppose that Assumption 2.1(i)-(ii) is satisfied. Then Assumption 3.1 holds under either of the following two sets of conditions:*

1. *Assumption 2.1(iii) holds and $L_{X,\theta}$ has a discrete spectrum where θ is the data-generating parameter value.*
2. $\Delta \to 0$ *and* $n\Delta \to \infty$.

Importantly, the above result shows that Assumption 3.1 can be verified without imposing stationarity. Unfortunately, this requires high-frequency information ($\Delta \to 0$). To our knowledge, there exists no results for low-frequency ($\Delta > 0$ fixed) identification of the drift and diffusion terms of scalar diffusion processes under non-stationarity. But by inspection of the arguments of Hansen et al. (1998) one can verify that at least the diffusion component is nonparametrically identified from low-frequency information without stationarity.

We are now ready to analyze the identification problem. Recall that $\mathcal{S} = (\theta, V)$ contains the objects of interest and let our model consist of all the structures that satisfy, as a minimum, Assumptions 2.1(i)–(ii) and 2.2. According to (2.6)-(2.7), each structure implies a drift and diffusion term of the observed process. We shall say that two structures $\mathcal{S} = (\theta, V)$ and $\tilde{\mathcal{S}} = (\tilde{\theta}, \tilde{V})$ are *observationally equivalent*, a property which we denote by $\mathcal{S} \sim \tilde{\mathcal{S}}$, if they imply the same drift and diffusion of Y, i.e.

$$\forall y \in \mathcal{Y} : \mu_Y (y; \mathcal{S}) = \mu_Y \left(y; \tilde{\mathcal{S}}\right) \text{ and } \sigma_Y (y; \mathcal{S}) = \sigma_Y \left(y; \tilde{\mathcal{S}}\right). \tag{3.1}$$

The structure $\mathcal{S}$ is then said to be identified within the model if $\mathcal{S} \sim \tilde{\mathcal{S}}$ implies $\mathcal{S} = \tilde{\mathcal{S}}$. In our setting, without suitable *normalizations* on the parameters of the UPD, identification will generally fail. To see this, observe that any given structure $\mathcal{S}$ is observationally equivalent to the following process: Choose any one-to-one transformation $T : \mathcal{X} \mapsto \mathcal{X}$, and rewrite the DGP implied by $\mathcal{S}$ as

$$Y_t = \tilde{V}\left(\tilde{X}_t\right), \quad \tilde{V}(x) = V(T(x)), \tag{3.2}$$

where $\tilde{X}_t = T^{-1}(X_t)$ solves

$$d\tilde{X}_t = \mu_{T^{-1}(X)}\left(\tilde{X}_t; \theta\right) dt + \sigma_{T^{-1}(X)}\left(\tilde{X}_t; \theta\right) dW_t, \tag{3.3}$$

with

$$\mu_{T^{-1}(X)}(x;\theta) = \frac{\mu_X(T(x);\theta)}{\partial T(x)/(\partial x)} - \frac{1}{2}\sigma_X^2(T(x);\theta)\frac{\partial^2 T(x)/\left(\partial x^2\right)}{\partial T(x)/(\partial x)^3}, \tag{3.4}$$

$$\sigma_{T^{-1}(X)}(x;\theta) = \frac{\sigma_X(T(x);\theta)}{\partial T(x)/(\partial x)}. \tag{3.5}$$

Suppose now that there exists $\tilde{\theta}$ so that $\mu_{T^{-1}(X)}(x;\theta) = \mu_X\left(x;\tilde{\theta}\right)$ and $\sigma_{T^{-1}(X)}(x;\theta) = \sigma_X\left(x;\tilde{\theta}\right)$. Then the alternative representation (3.2)-(3.3) is a member of our model with structure $\tilde{\mathcal{S}} = (\tilde{\theta}, \tilde{V})$ which is observationally equivalent to $\mathcal{S} = (V, \theta)$. The following result provides a complete characterizations of the class of observationally equivalent structures for a given model:

Theorem 3.2 *Suppose that Assumptions 3.1 is satisfied. For any two structures $\mathcal{S} = (V, \theta)$ and $\tilde{\mathcal{S}} = (\tilde{V}, \tilde{\theta})$ satisfying Assumptions 2.1(i) and 2.2, the following hold: $\mathcal{S} \sim \tilde{\mathcal{S}}$ if and only if there exists one-to-one transformation $T : \mathcal{X} \mapsto \mathcal{X}$ so that*

$$\tilde{V}(x) = V(T(x)) \tag{3.6}$$

and, with $\mu_{T^{-1}(X)}\left(x;\tilde{\theta}\right)$ and $\sigma_{T^{-1}(X)}\left(x;\tilde{\theta}\right)$ given in eqs. (3.4)-(3.5),

$$\text{(i) } \mu_{T^{-1}(X)}(x;\tilde{\theta}) = \mu_X(x;\theta) \text{ and (ii) } \sigma_{T^{-1}(X)}(x;\tilde{\theta}) = \sigma_X(x;\theta). \tag{3.7}$$

In particular, the data-generating structure is identified if and only if there exists no one-to-one transformation T such that (3.7) holds for $\theta \neq \tilde{\theta}$.

Note that the above theorem does not require stationarity since it is only concerned with the mapping $\mathcal{S} \mapsto (\mu_Y(\cdot;\mathcal{S}), \sigma_Y(\cdot;\mathcal{S}))$ which is well-defined irrespectively of whether data is stationary. The first part of the theorem provides a exact characterization of when any two structures are equivalent, namely if there exists a transformation T so that (3.6)-(3.7) hold. The second part comes as a natural consequence of the first part: If there exists no such transformation, then the data-generating structure must be identified.

Unfortunately, the above result may not always be useful in practice since it requires us to search over all possible one-to-one transformations T and for each of these verify that there exists no $\theta \neq \tilde{\theta}$ for which eq. (3.7) holds. In some cases, it proves useful to first normalize the UPD suitably and then verify eq. (3.7) in the normalized version. First note that for any one-to-one transformation $\bar{T}(\cdot;\theta) : \mathcal{X} \mapsto \bar{\mathcal{X}}$, an equivalent representation of the model is

$$Y_t = V\left(\bar{X}_t\right),$$

where the "normalised" UPD $\bar{X}_t := \bar{T}^{-1}(X_t;\theta) \in \bar{\mathcal{X}}$ solves

$$d\bar{X}_t = \mu_{\bar{X}}\left(\bar{X}_t;\theta\right)dt + \sigma_{\bar{X}}\left(\bar{X}_t;\theta\right)dW_t,$$

with

$$\mu_{\bar{X}}(\bar{x};\theta) = \frac{\mu_X\left(\bar{T}(\bar{x};\theta);\theta\right)}{\partial\bar{T}(\bar{x};\theta)/(\partial\bar{x})} - \frac{1}{2}\sigma_X^2\left(\bar{T}(\bar{x};\theta);\theta\right)\frac{\partial^2\bar{T}(\bar{x};\theta)/\left(\partial\bar{x}^2\right)}{\partial\bar{T}(\bar{x};\theta)/(\partial\bar{x})^3}, \tag{3.8}$$

$$\sigma_{\bar{X}}(\bar{x};\theta) = \frac{\sigma_X\left(\bar{T}(\bar{x};\theta);\theta\right)}{\partial\bar{T}(\bar{x};\theta)/(\partial\bar{x})}. \tag{3.9}$$

Given that the above representation is observationally equivalent to the original model, we can still employ Theorem 3.2 but with $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$ replacing μ_X and σ_X. Verifying the identification conditions stated in the second part of the theorem for the normalised versions will in some situations be easier by judicious choice of $\bar{T}$.

Below, we present three particular normalising transformations that we have found useful in this regard. The chosen transformations allow us to provide easy-to-check conditions for a given UPD to be identified. For a given UPD, the researcher is free to apply either of the three identification schemes depending on which is the easier one to implement. The three schemes lead to different normalizations/parametrizations, but they all lead to models that are exactly identified (no over-identifying restrictions are imposed) and so are observationally equivalent: The resulting form of μ_Y and σ_Y will be identical irrespectively of which scheme is employed.

The three transformations that we consider also highlights three alternative modelling approaches: Instead of starting with a parametric UPD as found in the existing literature, such as Examples 1-3, one can alternatively build a UPD with unit diffusion ($\sigma_X = 1$), zero drift ($\mu_X = 0$) or known marginal distribution. As we shall see, either of these three modelling approaches are in principle as flexible as the standard approach where the researcher jointly specifies the drift and diffusion term.

3.1 First Scheme

In our first identification scheme, we choose to normalize X_t by the so-called Lamperti transform,

$$\bar{X}_t = \bar{T}^{-1}(X_t;\theta) := \gamma(X_t;\theta), \quad \gamma(x;\theta) = \int_{x^*}^{x}\frac{1}{\sigma_X(z;\theta)}dz,$$

for some $x^* \in \mathcal{X}$. The resulting process is a unit diffusion process,

$$d\bar{X}_t = \mu_{\bar{X}}\left(\bar{X}_t;\theta\right)dt + dW_t,$$

with domain $\bar{\mathcal{X}} = (\bar{x}_l, \bar{x}_r)$, where $\bar{x}_r = \lim_{x\to x_r^+}\gamma(x;\theta)$ and $\bar{x}_l = \lim_{x\to x_l^-}\gamma(x;\theta)$, and drift function

$$\mu_{\bar{X}}(\bar{x};\theta) = \frac{\mu_X\left(\gamma^{-1}(\bar{x};\theta);\theta\right)}{\sigma_X\left(\gamma^{-1}(\bar{x};\theta);\theta\right)} - \frac{1}{2}\frac{\partial\sigma_X}{\partial x}\left(\gamma^{-1}(\bar{x};\theta);\theta\right). \tag{3.10}$$

For the unit diffusion version of the UPD, the equivalence condition (3.7)(ii) becomes

$$1 = \sigma_{\bar{X}}(\bar{x};\theta) = \sigma_{T^{-1}(\bar{X})}\left(\bar{x};\tilde{\theta}\right) = \frac{1}{\partial T(\bar{x})/(\partial x)},$$

which can only hold if $T(\bar{x}) = \bar{x} + \eta$ for some constant $\eta \in \mathbb{R}$. Thus, we can restrict attention to this class of transformations and (3.7)(i) becomes:

Assumption 3.2. With $\mu_{\bar{X}}$ given in (3.10): There exists no $\eta \neq 0$ and $\tilde{\theta} \neq \theta$ such that $\mu_{\bar{X}}(\bar{x};\tilde{\theta}) = \mu_{\bar{X}}(\bar{x}+\eta;\theta)$ for all $\bar{x} \in \bar{\mathcal{X}}$.

Assumption 3.2 imposes a normalization condition on the transformed drift function to ensure identification. When verifying Assumption 3.2 for the transformed unit diffusion $\bar{X}$ defined above, we will generally need to fix some of the parameters that enter $\mu_X(x;\theta)$ and $\sigma_X^2(x;\theta)$ of the original process X, see below.

Corollary 3.3 *Under Assumptions 2.1(i), 2.2 and 3.1, $\mathcal{S}$ is identified if and only if Assumption 3.2 is satisfied.*

The above transformation result can be applied to standard parametric specifications when $\gamma(x;\theta)$ is available on closed-form. But it also highlights that in terms of modelling copula diffusions, we can without loss of generality build a model where we from the outset restrict $\sigma_X = 1$ and only model the drift term μ_X. For example, we could choose the following flexible polynomial drift model where we have already normalized the diffusion term:

$$dX_t = \left(\sum_{i=1}^{l}\alpha_i X_t^i\right)dt + dW_t, \tag{3.11}$$

where $\theta = (\alpha_1, ..., \alpha_l)$. Corollary 3.3 shows that this particular copula diffusion specification is identified without further restrictions on θ. Below we apply Corollary 3.3 to some of the standard parametric diffusions introduced earlier:

Example 1 (continued). The Lamperti transform of the OU process in (2.10) is given by

$$d\bar{X}_t = \kappa\left(\alpha/\sigma - \bar{X}_t\right)dt + dW_t.$$

Since α/σ is a location shift of $\bar{X}$, we need to normalize α/σ in order for the identification condition 3.3 to be satisfied; one such is $\alpha/\sigma = 0$ leading to the following identified model,

$$d\bar{X}_t = -\kappa \bar{X}_t dt + dW_t. \tag{3.12}$$

Example 2 (continued). The Lamperti transform of the CIR diffusion in (2.12) is given by

$$d\bar{X}_t = \left[\kappa\left(\frac{2}{\bar{X}_t}\frac{\alpha}{\sigma^2} - \frac{\bar{X}_t}{2}\right) - \frac{1}{2\bar{X}_t}\right]dt + dW_t, \tag{3.13}$$

which only depends on $\theta = (\kappa, \alpha^*)$ where $\alpha^* = \alpha/\sigma^2$. Note that the dimension of the parameter vector reduced from 3 to 2. Crucially, it also suggests that we can only identify α and σ^2 up to a ratio. Hence, normalization requires fixing either α, σ^2, or their ratio.

Example 3 (continued). It can be easily verified that the Lamperti transform of the NLDCEV diffusion in (2.13) takes the form

$$d\bar{X}_t = \left[\sum_{i=-k}^{l} \alpha_i^* \bar{X}_t^{\frac{i-\beta}{1-\beta}} - \frac{\beta}{2(1-\beta)}\bar{X}_t^{-1}\right]dt + dW_t, \tag{3.14}$$

where $\alpha_i^* := \alpha_i \sigma^{\frac{i-1}{1-\beta}} (1-\beta)^{\frac{i-\beta}{1-\beta}}$, $i = -k, ..., l$. Hence, the parameters $\theta = \left(\beta, \alpha_{-k}^*, ..., \alpha_{-l}^*\right)$ are identified and the number of parameters is reduced from $l+k+3$ to $l+k+2$. Note that just as (2.10) and (2.12) are special cases of (2.13), both (3.12) and (3.13) are special cases of (3.14).

3.2 Second Scheme

Our second identification strategy transforms X by its scale measure defined in eq. (2.4),

$$\bar{X}_t := S(X_t; \theta),$$

which brings the diffusion process onto its natural scale,

$$d\bar{X}_t = \sigma_{\bar{X}}\left(\bar{X}_t; \theta\right) dW_t,$$

where the drift is zero (and so known) while

$$\sigma_{\bar{X}}^2(\bar{x}; \theta) = s^2\left(S^{-1}(\bar{x}; \theta); \theta\right)\sigma^2\left(S^{-1}(\bar{x}; \theta); \theta\right). \tag{3.15}$$

Since the drift term is zero, the identification condition (3.7)(i) becomes

$$0 = -\frac{1}{2}\sigma_{\bar{X}}^2\left(T(\bar{x}); \tilde{\theta}\right)\frac{\partial^2 T(\bar{x}) / \left(\partial \bar{x}^2\right)}{\partial T(\bar{x}) / (\partial \bar{x})^3}, \tag{3.16}$$

which can only hold if $\partial^2 T(\bar{x}) / \left(\partial \bar{x}^2\right) = 0$. We can therefore restrict attention to linear transformations $T(\bar{x}) = \eta_1 \bar{x} + \eta_2$, for some constants $\eta_1, \eta_2 \in \mathbb{R}$, in which case (3.7)(ii) becomes:

Assumption 3.3. With $\sigma^2_{\bar{X}}$ given in (3.15): There exists no $\eta_1 \neq 1$, $\eta_2 \neq 0$ and $\tilde{\theta} \neq \theta$ such that $\sigma^2_{\bar{X}}(\bar{x};\tilde{\theta}) = \sigma^2_{\bar{X}}\left(\eta_1 \bar{x} + \eta_2; \theta\right)/\eta_1^2$ for all $\bar{x} \in \bar{\mathcal{X}}$.

In comparison to Assumption 3.2, we here have to impose two normalizations to ensure identification. The intuition for this is that setting the drift to zero does not act as a complete normalization of the process: Any additional scale transformation of $\bar{X}$ still leads to a zero-drift process. Therefore, for the third scheme to work we need both a scale and location normalization.

Theorem 3.4 *Under Assumptions 2.1(i)–(ii), 2.2 and 3.1, $\mathcal{S}$ is identified if and only if Assumption 3.3 is satisfied.*

Compared to the first identification scheme, it is noticeably harder to apply this one to existing parametric diffusion models since the inverse of the scale transform is usually not available in closed form. But, similar to the first identification scheme, the result shows that without loss of flexibility, we can focus on UPDs with zero drift and then model the diffusion term in a flexible manner, e.g.,

$$dX_t = \exp\left(\sum_{i=1}^{l-1} \beta_i X_t^i + \beta_l \left|X_t\right|^l\right) dW_t. \tag{3.17}$$

Corollary 3.4 shows that this UPD is identified together with V without any further parameter restrictions on $\theta = (\beta_1, ..., \beta_l)$.

3.3 Third scheme

Our third identification strategy transforms a given stationary UPD by its marginal cdf,

$$\bar{X}_t = F_X\left(X_t;\theta\right). \tag{3.18}$$

In this case, there is generally no simplification in terms of the drift and diffusion term, which take the form

$$\begin{aligned} \mu_{\bar{X}}\left(\bar{x};\theta\right) &= \mu_X\left(F_X^{-1}\left(\bar{x};\theta\right);\theta\right) f_X\left(F_X^{-1}\left(\bar{x};\theta\right);\theta\right) \\ &\quad + \frac{1}{2}\sigma_X^2\left(F_X^{-1}\left(\bar{x};\theta\right);\theta\right) f'_X\left(F_X^{-1}\left(\bar{x};\theta\right);\theta\right) \end{aligned} \tag{3.19}$$

and

$$\sigma_{\bar{X}}\left(\bar{x};\theta\right) = \sigma_X\left(F_X^{-1}\left(\bar{x};\theta\right);\theta\right) f_X\left(F_X^{-1}\left(\bar{x};\theta\right);\theta\right). \tag{3.20}$$

for $\bar{x} \in \bar{\mathcal{X}} = (0,1)$. But the marginal distribution is now known with $\bar{X}_t \sim U(0,1)$ and we can directly identify the transformation function by $U(y) = F_Y(y)$, c.f. eq. (2.16). The identification condition then takes the form:

Assumption 3.4. With $\mu_{\bar{X}}(\bar{x};\theta)$ and $\sigma_{\bar{X}}(\bar{x};\theta)$ given in eqs. (3.19)-(3.20), the following hold:

$$\forall \bar{x} \in (0,1) : \mu_{\bar{X}}(\bar{x};\theta) = \mu_{\bar{X}}\left(\bar{x};\tilde{\theta}\right) \text{ and } \sigma_{\bar{X}}(\bar{x};\theta) = \sigma_{\bar{X}}\left(\bar{x};\tilde{\theta}\right) \Leftrightarrow \theta = \tilde{\theta}.$$

Corollary 3.5 *Under Assumptions 2.1-2.2 and 3.1, $\mathcal{S}$ is identified if and only if Assumption 3.4 is satisfied.*

The above result is only useful for showing identification of a given UPD if $F^{-1}(\bar{x};\theta)$ is available on closed form. But similar to the previous identification schemes, it demonstrates we can restrict attention to diffusions with known marginal distributions in the model building phase. Specifically, one can choose a known density $f_X(x)$ that describes the stationary distribution of X together with a parametric specification for, say, the drift function. We can then rearrange eq. (2.5) to back out the diffusion term of the UPD:

$$\sigma_X^2(x;\theta) = \frac{2}{f_X(x)} \int_{x_l}^{x} \mu_X(z;\theta) f_X(z)\, dz. \tag{3.21}$$

If the drift is specified so that $\mu_X(\cdot;\theta) \neq \mu_X(\cdot;\tilde{\theta})$ for $\theta \neq \tilde{\theta}$, then Assumption 3.4 will be satisfied for this model. Alternatively, one could choose a parametric specification of the diffusion term and then derive the corresponding drift term of the UPD satisfying

$$\mu_X(x;\theta) = \frac{1}{2f_X(x)} \frac{\partial}{\partial x}\left[\sigma_X^2(x;\theta) f_X(x)\right].$$

The resulting copula diffusion model is identified as long as the chosen diffusion term satisfies $\sigma_X(\cdot;\theta) \neq \sigma_X(\cdot;\tilde{\theta})$ for $\theta \neq \tilde{\theta}$, then Assumption 3.4 will be satisfied for this model.

Below, we apply the third identification scheme to the OU and CIR model:

Example 1 (continued). The stationary distribution of (2.10) is $N\left(\alpha, v^2\right)$ with $v^2 = \sigma^2/2\kappa$ and so the marginal density and cdf takes the form $f_X(x;\theta) = \frac{1}{v}\phi\left(\frac{x-\alpha}{v}\right)$ and $F_X(x;\theta) = \Phi\left(\frac{x-\alpha}{v}\right)$, where ϕ and Φ denote the density and cdf of the $N(0,1)$ distribution. Applying the transformation (3.18) yields, after some tedious calculations,

$$d\bar{X}_t = -2\kappa\Phi^{-1}\left(\bar{X}_t\right)\phi\left(\Phi^{-1}\left(\bar{X}_t\right)\right)dt + \sqrt{2\kappa}\phi\left(\Phi^{-1}\left(\bar{X}_t\right)\right)dW_t,$$

which is independent of α and σ^2 and these therefore have to be fixed, leaving κ as the only free parameter. This is the same finding as with the first identification strategy.

Example 2 (continued). The stationary distribution of the CIR process is a Γ-distribution with scale parameter $\omega = 2\kappa/\sigma^2$ and shape parameter $\nu = 2\kappa\alpha/\sigma^2$. Thus, the marginal density and cdf can be written as

$$\begin{aligned} f_X(x;\theta) &= f_X(x;\omega,\nu) = \frac{\omega^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\omega x} \\ F_X(x;\theta) &= F_X(x;\omega,\nu) = \frac{1}{\Gamma(\nu)} \gamma(\nu,\omega x) \end{aligned}$$

where $\Gamma(\nu)$ is the gamma function and $\gamma(\nu, \omega x)$ is the lower incomplete gamma function. Applying the transformation (3.18) yields

$$\mu_{\bar{X}}(\bar{x};\theta) = \left[\kappa\left(\frac{\nu}{2\kappa} - \frac{\gamma^{-1}(\nu, \bar{x}\Gamma(\nu))}{2\kappa}\right) + \left(\frac{\nu-1}{2} - \frac{\gamma^{-1}(\nu, \bar{x}\Gamma(\nu))}{2}\right)\right] \frac{2\kappa}{\Gamma(\nu)} \gamma^{-1}(\nu, \bar{x}\Gamma(\nu))^{\nu-1} e^{-\gamma^{-1}(\nu, \bar{x}\Gamma(\nu))}$$

and

$$\sigma^2_{\bar{X}}(\bar{x};\theta) = 2\kappa\gamma^{-1}(\nu, \bar{x}\Gamma(\nu)) \left[\frac{1}{\Gamma(\nu)} \gamma^{-1}(\nu, \bar{x}\Gamma(\nu))^{\nu-1} e^{-\gamma^{-1}(\nu, \bar{x}\Gamma(\nu))}\right]^2.$$

Note that $\mu_{\bar{X}}(\bar{x};\theta)$ and $\sigma^2_{\bar{X}}(\bar{x};\theta)$ only depend on κ and ν, which means we can only identify α and σ^2 up to a ratio say $\alpha^* = \alpha/\sigma^2$. Hence, either α or σ^2 must be fixed, which is in accordance with what we found when applying the first identification strategy to the CIR. We could, for example, set $\sigma^2 = 2\kappa$ which leads to the following normalized CIR

$$dX_t = \kappa(\alpha - X_t)\,dt + \sqrt{2\kappa X_t}dW_t.$$

4 Estimation

In this section we develop two alternative semiparametric estimators of θ and V for a given specification of the UPD. The first takes the form of a two-step Pseudo Maximum Likelihood Estimator (PMLE). The second is a semiparametric sieve-based ML estimator (SMLE). We consider two different scenarios when developing estimators: In the first one (see Section 4.1), Y is observed at low frequency which we formally define as the case when $\Delta > 0$ is fixed as $n \to \infty$. In the second one (see Section 4.2), high-frequency data is available so that $\Delta \to 0$ as $n \to \infty$.

4.1 Low-frequency estimators

To motivate the two estimators, suppose that U is known, in which case the MLE of θ is given by

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta\in\Theta} L_n(\theta, U),$$

where $L_n(\theta, U)$ is the log-likelihood of $\{Y_{i\Delta} : i = 0, 1, ..., n\}$,

$$L_n(\theta, U) = \frac{1}{n}\sum_{i=1}^{n}\left\{\log p_X\left(U(Y_{i\Delta})|U\left(Y_{(i-1)\Delta}\right);\theta\right) + \log U'(Y_{i\Delta})\right\}, \tag{4.1}$$

where p_X was is defined in eq. (2.3). If U is unknown, the above estimator is not feasible and we instead have to estimate it together with θ.

Our PMLE assumes Y is stationary in which case U satisfies eq. (2.16), where F_X is known up to θ while F_Y is unknown. The latter can be estimated by the empirical cdf defined as

$$\tilde{F}_Y(y) = \frac{1}{n+1}\sum_{i=0}^{n}\mathbb{I}\{Y_{i\Delta} \le y\},$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function, or alternatively by the following kernel smoothed empirical cdf,

$$\hat{F}_Y(y) = \frac{1}{n+1}\sum_{i=0}^{n} \mathcal{K}_h(Y_{i\Delta} - y), \tag{4.2}$$

where $\mathcal{K}_h(y) = \mathcal{K}(y/h)$ with $\mathcal{K}(y) = \int_{-\infty}^{y} K(z)\,dz$, K being a kernel (e.g., the standard normal density), and $h > 0$ a bandwidth. Replacing F_Y in eq. (2.16) with either $\tilde{F}_Y$ or $\hat{F}_Y$, we obtain the following two alternative estimators of U,

$$\tilde{U}(y;\theta) = F_X^{-1}(\tilde{F}_Y(y);\theta); \quad \hat{U}(y;\theta) = F_X^{-1}(\hat{F}_Y(y);\theta). \tag{4.3}$$

Since $\hat{F}_Y(y) = \tilde{F}_Y(y) + O(h^2)$, the above two estimators of U will be first-order asymptotically equivalent under appropriate bandwidth conditions. A natural way to estimate θ in our semiparametric framework would then be to substitute either $\hat{U}(y;\theta)$ or $\tilde{U}(y;\theta)$ into $L_n(\theta, U)$. However, in the latter case, this is not possible since $L_n(\theta, U)$ depends on U' and $\tilde{U}$ is not differentiable. However, note that

$$U'(y) = \frac{f_Y(y)}{f_X(U(y);\theta)}, \tag{4.4}$$

so that $\log U'(y) = \log f_Y(y) - \log f_X(U(y);\theta)$. Since the first term is parameter independent, it can be ignored and so we arrive at the following semiparametric PMLE,

$$\hat{\theta}_{\text{PMLE}} = \arg\max_{\theta\in\Theta} \bar{L}_n(\theta, \tilde{U}(\cdot;\theta)),$$

where Θ is the parameter space and

$$\bar{L}_n(\theta, U) = \frac{1}{n}\sum_{i=1}^{n} \left\{\log p_X\left(U(Y_{i\Delta})|U\left(Y_{(i-1)\Delta}\right);\theta\right) - \log f_X(U(Y_{i\Delta});\theta)\right\}$$

is $L_n(\theta, U) - \sum_{i=1}^{n} \log f_Y(Y_{i\Delta})/n$. One can easily check that, by rewriting the above in terms of the implied copula of X, this estimator is equivalent to the one analyzed in Chen and Fan (2006).

Our second proposal, the SMLE, replaces the unknown density function $f_Y(y)$ by a sieve approximation $f_{Y,m}(y) \in \mathcal{F}_m$ where $\mathcal{F}_m$ is a finite-dimensional function space reflecting the properties of f_Y, $m = 1, 2, \ldots$. For a given candidate density, we then compute

$$U(y; f_{Y,m}, \theta) = F_X^{-1}(F_{Y,m}(y);\theta)$$

where $F_{Y,m}(y) = \int_{y_l}^{y} f_{Y,m}(z)\,dz$. Substituting this into the likelihood function yields the following semiparametric sieve maximum-likelihood estimator,

$$(\hat{\theta}_{\text{SMLE}}, \hat{f}_{Y,m}) = \arg\max_{\theta\in\Theta, f_{Y,m}\in\mathcal{F}_m} L_n(\theta, U(\cdot; f_{Y,m}, \theta)). \tag{4.5}$$

The above SMLE is identical to the one proposed by Chen, Wu and Yi (2009) for the estimation of copula-based Markov models, except that while they estimate the parameters of a copula function, we estimate those of the drift and diffusion functions of the UPD. In comparison with the PMLE, the numerical implementation of the SMLE involves joint maximization over both θ and $\mathcal{F}_m$, which is a harder numerical problem and potentially more time-consuming. In terms of statistical efficiency, $\hat{\theta}_{\text{SMLE}}$ will in general reach the semiparametric efficiency bound under stationarity, while the PMLE is inefficient.

Both of the above estimators require us to evaluate $F_X^{-1}(x;\theta)$ which in general is not available on closed form and so has to be computed using numerical methods, e.g., numerical integration or Monte Carlo methods combined with a equation solver. For the SMLE, one can circumvent this issue by directly approximating U instead of f_Y: For a given finite-dimensional function space of one-to-one transformations $\mathcal{U}_m$, an alternative to the SMLE in (4.5) is $(\tilde{\theta}_{\text{SMLE}}, \tilde{U}_m) = \arg\max_{\theta\in\Theta, U_m\in\mathcal{U}_m} L_n(\theta, U_m)$. We expect this to be computationally more efficient compared to the density version above; the theoretical analysis of this alternative SMLE is left for future research.

Once an estimator for θ has been obtained, we can estimate the drift and diffusion terms of Y using the expressions given in (2.6) and (2.7) by replacing θ and U with their estimators. However, this involves estimating the first and second derivative of U. For the SMLE this is not an issue assuming that $\mathcal{F}_m$ is a differentiable function space. For the PMLE, since $\tilde{U}(y;\theta)$ is not differentiable, we instead use the kernel smoothed version $\hat{U}(y;\theta)$, leading to the following three-step estimators of the drift and diffusion functions

$$\hat{\mu}_Y(y) = \frac{\mu_X(\hat{U}(y);\hat{\theta}_{\text{PMLE}})}{\hat{U}'(y)} - \frac{1}{2}\sigma_X^2(\hat{U}(y);\hat{\theta}_{\text{PMLE}})\frac{\hat{U}''(y)}{\hat{U}'(y)^3}, \tag{4.6}$$

$$\hat{\sigma}_Y^2(y) = \frac{\sigma_X^2(\hat{U}(y);\hat{\theta}_{\text{PMLE}})}{\hat{U}'(y)^2}, \tag{4.7}$$

where $\hat{U}(y) = F_X^{-1}(\hat{F}_Y(y);\hat{\theta}_{\text{PMLE}})$.

4.2 High-frequency estimators

We now turn to the case where high-frequency data is available; this scenario is formally modelled as $\Delta \to 0$ as $n \to \infty$. The proposed estimators described in the previous section remains valid, but an alternative estimation method is available in this case since the exact density of the underlying UPD, p_X, is well-approximated by

$$\hat{p}_X(x|x_0;\theta) = \frac{1}{\sqrt{2\pi\Delta}}\sigma_X(x_0;\theta)\exp\left[-\frac{(x - x_0 - \mu_X(x_0;\theta)\Delta)^2}{2\sigma_X^2(x_0;\theta)\Delta}\right] \tag{4.8}$$

as $\Delta \to 0$, c.f. Kessler (1997). We then propose to estimate θ using either the two-step or sieve approach described in the previous section, except that we here replace $p_X(x|x_0;\theta)$ with its high-frequency approximation, $\hat{p}_X(x|x_0;\theta)$, in the definition of $L_n(\theta, U)$ and $\bar{L}_n(\theta, U)$. The advantage of doing so is computational in that $\hat{p}_X(x|x_0;\theta)$ is on closed form for any given UPD while $p_X(x|x_0;\theta)$ generally can only be evaluated using numerical methods as pointed out earlier.

For most standard UPD's, the parameters can be decomposed into $\theta = (\theta_1, \theta_2)$ so that $\mu_X(x_0;\theta_1) = \mu_X(x_0;\theta_1)$ and $\sigma_X(x_0;\theta) = \sigma_X(x_0;\theta_2)$ only depends on the first and second component, respectively. One could hope to be able to estimate θ_1 and θ_2 separately in this case. For known U, this is indeed possible. We could, for example, use least-squares methods similar to Kanaya and Kristensen (2018) where θ_1 and θ_2, respectively, are estimated by the minimizers of the following two least-squares objectives,

$$L_{n,\Delta}^{(\mu)}\left(\theta_1;U\right) = \sum_{i=1}^{n} w_i^{(\mu)}\left(U\left(Y_{i\Delta}\right) - U\left(Y_{(i-1)\Delta}\right) - \mu_X\left(U\left(Y_{(i-1)\Delta}\right);\theta_1\right)\Delta\right)^2, \tag{4.9}$$

$$L_{n,\Delta}^{(\sigma)}\left(\theta_2;U\right) = \sum_{i=1}^{n} w_i^{(\sigma)}\left(\left\{U\left(Y_{i\Delta}\right) - U\left(Y_{(i-1)\Delta}\right)\right\}^2 - \sigma_X^2\left(U\left(Y_{(i-1)\Delta}\right);\theta_2\right)\Delta\right)^2, \tag{4.10}$$

where $w_i^{(\mu)} = w^{(\mu)}\left(Y_{(i-1)\Delta}, Y_{i\Delta}\right)$ and $w_i^{(\sigma)} = w^{(\sigma)}\left(Y_{(i-1)\Delta}, Y_{i\Delta}\right)$ are weighting functions.

This approach, however, faces two complications in our setting: First, after applying any of the three normalizations presented in Section 3 in order to achieve identification, the resulting drift and diffusion of the UPD tend to share parameters. Second, U is unknown and has to be estimated together with θ. In the case of PMLE, $\tilde{U}\left(y;\theta\right)$ in eq. (4.3) generally depends on both θ_1 and θ_2 since $f_X\left(x;\theta\right)$ does. Thus, if we replace U by $\tilde{U}\left(y;\theta\right)$ in the above objectives, we cannot separately estimate θ_1 and θ_2. Similarly, the SMLE requires joint estimation of U together with θ in which case it would have to be re-estimated for each of the two objectives. In conclusion, these least-squares estimators are rarely useful in practice.

Another alternative approach, inspired by Bandi and Phillips (2007), see also Kristensen (2011), would be to first obtain non-parametric estimates of μ_Y and σ_Y^2 and then match these with the ones implied by the copula model,

$$Q_{n,\Delta}^{(\mu)}\left(\mathcal{S}\right) = \sum_{i=1}^{n} w_i^{(\mu)}\left(\hat{\mu}_Y\left(Y_{i\Delta}\right) - \mu_Y\left(Y_{i\Delta};\mathcal{S}\right)\right)^2, \quad Q_{n,\Delta}^{(\sigma)}\left(\mathcal{S}\right) = \sum_{i=1}^{n} w_i^{(\sigma)}\left(\hat{\sigma}_Y^2\left(Y_{i\Delta}\right) - \sigma_Y^2\left(Y_{i\Delta};\mathcal{S}\right)\right)^2,$$

where $\hat{\mu}_Y\left(\cdot\right)$ and $\hat{\sigma}_Y^2\left(\cdot\right)$ are the first-step nonparametric estimators; see Bandi and Phillips (2007) for their precise forms. This procedure suffers from the same issue as the least-squares one described in the previous paragraph. An additional complication is that it involves multiple smoothing parameters: First, $\hat{\mu}_Y\left(\cdot\right)$ and $\hat{\sigma}_Y^2\left(\cdot\right)$ depend on two bandwidths and converge with slow rates and, second, $\mu_Y\left(\cdot;\mathcal{S}\right)$ and $\sigma_Y^2\left(\cdot;\mathcal{S}\right)$ involve derivatives of U and so if we replace U by its kernel-smoothed estimator, $\hat{U}$, the two objective funtions will depend on the first and second order derivatives of the kernel density estimator of f_Y, which in turn depends on additional bandwidth. All together, these estimators will be complicated to implement due to the multiple bandwidths that the econometrician have to choose. Moreover, their asymptotic analysis and behaviour will be non-standard.

5 Asymptotic Theory

5.1 Low-frequency Estimation of Parametric Component

We here establish an asymptotic theory for the proposed estimators in the case of low-frequency data ($\Delta > 0$ fixed). In the theoretical analysis we shall work under the following high-level identification condition:

Assumption 4.1 $\mathcal{S}_0$ is identified.

The previous section provided three different sets of primitive conditions for Assumption 4.1 to hold in terms of $(\mu_Y(\cdot;\mathcal{S}), \sigma_Y(\cdot;\mathcal{S}))$. This combined with Assumption 3.1 then implies that the mapping $(\mu_Y(\cdot;\mathcal{S}), \sigma_Y(\cdot;\mathcal{S})) \mapsto p_Y(y|y_0;\mathcal{S})$ is injective so that different drift and diffusion terms lead to different transition densities. One implication of Assumptions 3.1 and 4.1 is $E[\log p_Y(Y_\Delta|Y_0;\mathcal{S})] < E[\log p_Y(Y_\Delta|Y_0;\mathcal{S}_0)]$ for any $\mathcal{S} \neq \mathcal{S}_0$, c.f. Newey and McFadden (1994, Lemma 2.2). This ensures that the SMLE identifies $\mathcal{S}_0$ in the limit. Regarding the PMLE, we note that it replaces U by $\hat{U}(y;\theta) = F_X^{-1}(\hat{F}_Y(y;\theta))$. By the LLN of stationary and ergodic sequences, $\hat{U}(y;\theta) \to^P U(y;\theta) = F_X^{-1}(F_Y(y;\theta))$, where, by the same arguments as before, $E[\log p_Y(Y_\Delta|Y_0;\theta, U(\cdot;\theta))] < E[\log p_Y(Y_\Delta|Y_0;\theta_0, U(\cdot;\theta_0))]$. Thus, the PMLE will also in the limit identify θ_0.

Next, we import conditions from Chen et al. (2010) guaranteeing, in conjunction with our own Assumptions 2.1-2.2, that the UPD X, and thereby Y, is stationary and β-mixing with mixing coefficients decaying at either polynomial rate (c.f. Corollary 5.5 in Chen et al., 2010) or geometric rate (c.f. Corollary 4.2 in Chen et al., 2010):

Assumption 4.2. (i) μ_X and σ_X^2 satisfies

$$\lim_{x\to x_r}\left\{\frac{\mu_X(x;\theta_0)}{\sigma_X(x;\theta_0)} - \frac{1}{2}\frac{\partial\sigma_X(x;\theta_0)}{\partial x}\right\} \leq 0, \qquad \lim_{x\to x_u}\left\{\frac{\mu_X(x;\theta_0)}{\sigma_X(x;\theta_0)} - \frac{1}{2}\frac{\partial\sigma_X(x;\theta_0)}{\partial x}\right\} \geq 0;$$

(ii) With $s(x;\theta)$ and $S(x;\theta)$ defined in (2.4),

$$\lim_{x\to x_r}\left\{\frac{s(x;\theta_0)\,\sigma_X(x;\theta_0)}{S(x;\theta_0)}\right\} > 0, \qquad \lim_{x\to x_u}\left\{\frac{s(x;\theta_0)\,\sigma_X(x;\theta_0)}{S(x;\theta_0)}\right\} < 0;$$

Assumption 4.2(ii) is a strengthening of Assumption 4.2(i). For the analysis of the PMLE, Assumption 4.2(i) suffices while we need the stronger Assumption 4.2(ii) to establish an asymptotic theory for the SMLE. As we mentioned before, it is not always straightforward to verify the required mixing conditions for copula-based (discrete-time) Markov models such as Chen and Fan (2006) and Chen, Wu and Yi (2009). In contrast, either sets of conditions stated in Assumption 4.2 can be easily verified by directly examining the drift and diffusion functions of the UPD X.

Finally, we impose the same conditions as used in the asymptotic analysis of the PMLE in Chen and Fan (2006) and Chen, Wu and Yi (2009), respectively, on the copula implied by the chosen UPD and the sieve density in the case of SMLE:

Assumption 4.3. (i) $c_X(u_0, u;\theta)$ defined in (2.17) satisfies the regularity conditions set out in Chen and Fan (2006, A1-A3, A4 or A4', A5-A6); (ii) $c_X(u_0, u;\theta)$ and the sieve space $\mathcal{F}_m$ satisfy Assumptions 3.1-3.4 and 4.1–4.7, respectively, in Chen, Wu and Yi (2009).

We here abstain from stating the precise, mostly technical, conditions and refer the interested reader to Chen and Fan (2006) and Chen, Wu and Yi (2009); broadly speaking their conditions translate into moment bounds and smoothness conditions on the log-transition density of the UPD. These conditions depend on the precise choice of the UPD and so will have to be verified on a case-by-case basis. In Appendix B, we verify the conditions for models in Examples 1–2.

The following result now follows from the general theory of Chen and Fan (2006) and Chen, Wu and Yi (2009), respectively:

Theorem 5.1 *Under Assumptions 2.1-2.2, 4.1, 4.2(i) and 4.3(i),*

$$\sqrt{n}(\hat{\theta}_{\text{PMLE}} - \theta_0) \to^d N\left(0, B^{-1}\Sigma B^{-1}\right),$$

where B and Σ are defined in Chen and Fan (2006, A1 and A_n^*).

Under Assumptions 2.1-2.2, 4.1, 4.2(ii) and 4.3(ii),

$$\sqrt{n}(\hat{\theta}_{\text{SMLE}} - \theta_0) \to^d N\left(0, \mathcal{I}_*^{-1}(\theta)\right),$$

where $\mathcal{I}_*$ is defined in Chen, Wu and Yi (2009).

Consistent estimators of the asymptotic variances, $B^{-1}\Sigma B^{-1}$ and $\mathcal{I}_*^{-1}(\theta)$, can be found in Chen and Fan (2006) and Chen, Wu and Yi (2009), respectively.

5.2 High-frequency Estimation of Parametric Component

Next, we discuss the asymptotic properties of the PMLE based on the high-frequency log-likelihood that takes as input $\hat{p}_X(x|x_0;\theta)$ defined in eq. (4.8); a complete analysis of the PMLE and SMLE in a high-frequency setting is left for future research. In the following, we let $T := n\Delta$ denote the sampling range, which will be assumed to diverge as $\Delta \to 0$.

The high-frequency PMLE is given by $\hat{\theta}_{\text{PMLE}} = \arg\max_{\theta\in\Theta} \hat{L}_n\left(\theta, \tilde{U}(\cdot;\theta)\right)$ where

$$\hat{L}_n(\theta, U) = \frac{1}{n}\sum_{i=1}^{n}\left\{\log \hat{p}_X\left(U(Y_{i\Delta})|U\left(Y_{(i-1)\Delta}\right);\theta\right) - \log f_X\left(U(Y_{i\Delta});\theta\right)\right\},$$

and $\tilde{U}(Y_{i\Delta};\theta)$ defined in (4.3). We first specialize the general result of Kanaya (2018, Theorem 2) by choosing $B = \psi = 1$ and $K_h(y) = I\{y \leq 0\}$ in his notation to obtain that under our Assumption 4.2,

$$\sup_{y\in\mathcal{Y}}\left|\tilde{F}_Y(y) - F_Y(y)\right| = O_P\left(\sqrt{\Delta}/\log\Delta\right) + O_P\left(\log T/\sqrt{T}\right), \tag{5.1}$$

where the two terms on the right-hand side correspond to discretization bias and sampling variance, respectively. By letting T grow sufficiently fast as $\Delta \to 0$, the first term can be ignored. Under regularity conditions on μ_X and σ_X so that $(y, y_0) \mapsto \hat{p}_X\left(F_X^{-1}(y;\theta)|F_Y^{-1}(y_0);\theta\right)/f_Y(y_0)$ satisfies Lipschitz conditions similar to the ones in Chen and Fan (2006), we then obtain

$$\sup_{\theta\in\theta}\left|\hat{L}_n\left(\theta, \tilde{U}(\cdot;\theta)\right) - \hat{L}_n(\theta, U(\cdot;\theta))\right| = O_P\left(\sqrt{\Delta}/\log\Delta\right) + O_P\left(\log T/\sqrt{T}\right),$$

where $U(y;\theta) = F_Y\left(F_X^{-1}(y;\theta)\right)$. Consistency of the PMLE now follows by extending the arguments of Kessler (1997) to allow for the presence of the parameter-dependent transformation $U(y;\theta)$. Next, to simplify our discussion of the asymptotic distribution of the PMLE, we consider two special cases:

First, suppose that suppose that, after suitable normalizations, $\sigma_X(x)$ is known and only $\mu_X(x;\theta)$ is parameter dependent. In this case, we expect that Kessler's results generalize so that $\hat{\theta}_{\text{PMLE}}$ will converge with $\sqrt{T}$-rate towards a Normal distribution, where the asymptotic variance will have to be adjusted to take into account the first-step estimation of $\hat{F}_Y$.

Next, consider the opposite scenario, $\mu_X(x_0)$ is known and only $\sigma_X(x_0;\theta)$ is parameter dependent. With U known, Kessler (1997) shows that $\hat{\theta}_{\text{PMLE}}$ converges with $\sqrt{n}$-rate towards a Normal distribution in this case. Note the faster convergence rate compared to the drift estimator. However, in our setting $U(y;\theta)$ is parameter dependent, and as a consequence this result appears to no longer apply: $U(y;\theta)$ enters $\hat{L}_n(\theta, U)$ in the same way that μ_X does and so the score of $\hat{L}_n(\theta, U(\cdot;\theta))$ will have a component on the same form as in the first case and so will converge with $\sqrt{T}$-rate instead of $\sqrt{n}$-rate. Moreover, the presence of the first-step estimator $\tilde{F}_Y(y)$, which also converge with $\sqrt{T}$-rate, will generate an additional variance term. In total, estimators of diffusion parameters appear not to enjoy "super" consistency in our setting due to the way that the unknown transformation U enters the likelihood.

5.3 Estimation of Drift and Diffusion Functions

We here analyze the asymptotic properties of the kernel-based estimators of μ_Y and σ_Y^2 given in eqs. (4.6)-(4.7). We only do so for the low-frequency case; the analysis of the high-frequency case should proceed in a similar fashion. Our analysis takes as starting point the following regularity conditions on the estimator of the parametric component and the kernel function:

Assumption 4.4. The transformation function V is four times continuously differentiable.

Assumption 4.5. The estimator $\hat{\theta}$ of the parameter of the UPD X is $\sqrt{n}$-consistent.

Assumption 4.6. The kernel K is differentiable, and there exists constants $D, \omega > 0$ such that

$$\left|K^{(i)}(z)\right| \leq D|z|^{-\omega}, \quad \left|K^{(i)}(z) - K^{(i)}(\tilde{z})\right| \leq D|z - \tilde{z}|, \quad i = 0, 1,$$

where $K^{(i)}(z)$ denotes the ith derivative of $K(z)$. Moreover, $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\kappa_2 = \int_{\mathbb{R}} z^2 K(z)\,dz < \infty$.

Assumption 4.4 ensures the existence of the 3rd and 4th derivatives of $U(y)$, which in turn ensure that relevant quantities entering the asymptotic distributions of $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ are well defined. Assumption 4.5 implies that the asymptotic properties of $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ are determined by the properties of the kernel density estimator alone. The proposed PMLE and SMLE satisfy this condition under our Assumptions 4.1-4.3, but other $\sqrt{n}$-consistent estimators are allowed for. Assumption 4.6 regulates the kernel functions and allow for most standard kernels such as the Gaussian and the Uniform kernels. Using the functional delta-method together with standard results for kernel density estimators, as found in Robinson (1983), we obtain:

Theorem 5.2 *Under Assumptions 2.1-2.2, 4.2(i), and 4.4-4.6, we have as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$,*

$$\sqrt{nh^3}\left\{\hat{\mu}_Y(y) - \mu_Y(y) - h^2 B_{\mu_Y}(y)\right\} \to^d N\left(0, V_{\mu_Y}(y)\right),$$

where

$$B_{\mu_Y}(y) = -\frac{\kappa_2 \sigma_Y^2(y) f_Y'''(y)}{4 f_Y(y)}, \quad V_{\mu_Y}(y) = \frac{\sigma_Y^4(y)}{4 f_Y(y)} \int_{\mathbb{R}} K'(z)^2 \, dz.$$

Also, as $n \to \infty$, $h \to 0$ and $nh \to \infty$, we have

$$\sqrt{nh}\{\hat{\sigma}_Y^2(y) - \sigma_Y^2(y) - h^2 B_{\sigma_Y^2}(y)\} \to^d N\left(0, V_{\sigma^2}(y)\right),$$

where

$$B_{\sigma_Y^2}(y) = -\frac{\kappa_2 \sigma_Y^2(y) f_Y''(y)}{f_Y(y)}, \quad V_{\sigma_Y^2}(y) = \frac{4\sigma_Y^4(y)}{f_Y(y)} \int_{\mathbb{R}} K(z)^2 \, dz.$$

We see that both estimators suffer from smoothing biases, $B_{\mu_Y}(y)$ and $B_{\sigma_Y^2}(y)$. If $h \to 0$ sufficiently fast, these biases will be negiglible. Also note that the convergence rates of the drift estimator is slower compared to the diffusion estimator. These features are similar to the asymptotic properties of the semi-nonparametric drift and diffusion estimators considered in Kristensen (2011).

6 Monte Carlo Simulations

In this section, we compare the finite sample performance of our low-frequency semiparametric PMLE with that of a fully parametric PMLE (described below) through Monte Carlo simulations.

6.1 Data Generating Processes

We consider the following normalized versions of the UPDs of Examples 1–2,

$$\text{OU} \; : \; dX_t = -\kappa X_t dt + \sqrt{2\kappa} dW_t, \quad \theta = \kappa, \tag{6.1}$$

$$\text{CIR} \; : \; dX_t = \kappa(\alpha - X_t) dt + \sqrt{2\kappa X_t} dW_t, \quad \theta = (\kappa, \alpha). \tag{6.2}$$

The chosen normalizations have the advantage that the marginal distributions of X are invariant to the mean-reversion parameter κ. Hence, by varying κ, we can change the persistence level of X (and thus Y) while keeping the marginal distributions fixed. In this way, we can examine the impact of persistence on the performance of the proposed estimators of θ, μ_Y and σ_Y^2.

Next, we specify the transformation of the DGP of Y. This is done by choosing marginal cdf $F_Y(y;\phi)$, where ϕ is a hyper parameter governing the shape of the cdf, which induces the transformation $V(X_t;\phi) = F_Y^{-1}(F_X(X_t;\theta);\phi)$. With $f_Y(y;\phi) = F_Y'(y;\phi)$, the transition density of the true DGP of Y then takes the form

$$p_Y(y|y_0;\theta,\phi) = f_Y(y;\phi)\, c_X(F_Y(y_0;\phi), F_Y(y;\phi);\theta). \tag{6.3}$$

We choose $F_Y(y;\phi)$ as a flexible distribution to reflect stylized features such as asymmetry and fat-tailedness of observed financial data. Specifically, we use the Skewed Student-t (SKST) Distribution of Hansen (1994) with density

$$f_Y(y;\phi) = \begin{cases} \dfrac{bq}{v}\left(1 + \dfrac{1}{\tau-2}\left(\dfrac{\frac{b}{v}(y-m)+a}{1-\lambda}\right)^2\right)^{-(\tau+1)/2} & \text{if} \quad y < m - av/b, \\ \dfrac{bq}{v}\left(1 + \dfrac{1}{\tau-2}\left(\dfrac{\frac{b}{v}(y-m)+a}{1+\lambda}\right)^2\right)^{-(\tau+1)/2} & \text{if} \quad y \geq m - av/b, \end{cases} \tag{6.4}$$

where $v > 0$, $2 < \tau < \infty$, $-1 < \lambda < 1$, $a = 4\lambda q\left(\dfrac{\tau-2}{\tau-1}\right)$, $b^2 = 1 + 3\lambda^2 - a^2$ and $q = \Gamma((\tau+1)/2)/\sqrt{\pi(\tau-2)\Gamma^2(\tau/2)}$. We collect the hyper parameters in $\phi = (m, v, \lambda, \tau)$ which has to be chosen in order to fully specify the DGP. While m and v are the unconditional mean and standard deviation of the distribution, λ controls the skewness and τ controls the degrees of freedom (hence the fat-tailedness) of the distribution. The 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; Bu, Fredj and Li, 2017).

The transformed diffusion Y generated by the SKST marginal distribution together with the normalized UPD in (6.1) or (6.2) is referred to as the OU-SKST or the CIR-SKST model, respectively. The true data-generating parameters ϕ and θ are chosen as estimates obtained from fitting the parametric versions of the two models to the 7-day Eurodollar interest rate time series used in Aït-Sahalia (1996b). The estimation is based on a fully parametric two-stage PMLE. In the first stage, the SKST distribution is fitted to the data (as if they are i.i.d) to obtain $\hat{\phi}$. We then substitute $F_Y(y;\hat{\phi})$ and $f_Y(y;\hat{\phi})$ into (6.3) which is then maximized with respect to θ to obtain $\hat{\theta}$ for each of the two UPD's. The calibrated parameter values of the marginal SKST distribution are $(\hat{m}, \hat{v}, \hat{\lambda}, \hat{\tau}) = (0.0835, 0.0358, 0.5193, 25.3708)$, and those of the underlying OU and CIR diffusions are $\hat{\kappa} = 1.1376$ and $(\hat{\kappa}, \hat{\alpha}) = (0.7653, 1.1653)$, respectively.

We compare the fitted SKST and Normal distributions with a nonparametric kernel estimate in Figure 1. We see that the SKST distribution does a reasonable job at capturing the marginal distribution found in data while the Normal one does not provide a very good fit.

[Figure 1]

Artificial samples of sizes $n = 2202$ and $n = 5505$, respectively, are then generated using $\phi = \hat{\phi}$ and $\theta = \hat{\theta}$ as our true data-generating parameters. For both OU-SKST and CIR-SKST, θ involves the mean-reversion parameter κ which controls the level of persistence. We create 3 additional scenarios by multiplying κ by factors of 5, 10, and 20 while keeping everything else unchanged. Collectively, we have a total of 8 cases corresponding to 2 sample sizes and 4 persistence levels. The maximum factor 20 is chosen because the implied 1st-order autocorrelation coefficient $\rho_1 \approx 0.9$, which is a reasonably high persistent level without being excessively close to the unit root. Finally, 500 replications for each case are generated.

6.2 Estimation Results

We compare our low-frequency PMLE of θ with the corresponding fully parametric PMLE (PPMLE) described above that we used for our calibration. Note that the only difference between the two estimators is that the former estimates the marginal distribution F_Y parametrically, while the latter estimates it nonparametrically.

The relative bias and RMSE (defined as the ratios of the actual bias and the actual RMSE over the true parameter value, respectively) of the estimators of the parametric components of the OU-SKST case are presented in Table 1. Overall, the results from the two estimation methods are generally comparable with the same magnitudes. The semiparametric PMLE tends to do better in terms of bias while the parametric PMLE dominates in terms of variance. However, as the level of persistence decreases, the two estimators' performance is close to identical.

[Table 1]

The results for the CIR-SKST case are presented in Table 2 and 3 which are qualitatively very similar to the ones for the OU-SKST. Overall, the performance of the PMLE is comparable with that of the PPMLE with very similar estimation errors. Moreover, the gap in the performance of the PMLE relative to the PPMLE appears to narrow when the true DGP gets less persistent.

[Table 2 and 3]

Next, we investigate the performance of the semiparametric estimators of μ_Y and σ_Y^2 in eqs. (4.6)-(4.7) relative to their fully parametric estimators. In Figure 2, we plot their pointwise means and 95% confidence bands from the 500 estimates against the truth for the OU-SKST process with $\kappa = 22.753$ and sample size 2202. First, it is worth noting that μ_Y and σ_Y^2 exhibit strong nonlinearities that closely resemble the nonlinearities depicted in, for example, Aït-Sahalia (1996b), Jiang and Knight (1997), and Stanton (1997). Second, the mean estimates from both estimation methods are fairly close to the truth, but the variability of the semiparametric estimators is noticeably larger than the parametric ones, especially in the right end of the range. This is not surprising: Firstly, as shown in Theorem 5.2, $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ converges at slower than $\sqrt{n}$-rate due to the use of kernel estimators of f_Y. From Figure 1, we can see that f_Y has a long right tail which is difficult to estimate by the kernel estimator in small and moderate samples. Figure 3 presents the same estimators at sample size 5505. At this larger sample size, the bias is even smaller for both methods and the variability of these estimates are also reduced significantly. Overall, although the parametric method obviously has the advantage due to its parametric structure, our semiparametric method also provides fairly satisfactory estimation results.

[Figure 2 and 3]

The drift and diffusion estimators from the two methods where the true DGP is the CIR-SKST process with $\kappa = 15.307$ and the two sample sizes are presented in Figure 4 and 5, respectively. Almost identical qualitative conclusions can be reached.

[Figure 4 and 5]

7 Empirical Application

7.1 Data

As an empirical illustration, we here model the time series dynamics of the CBOE Volatility Index data using copula diffusion models. The data consists of the daily VIX index from January 2, 1990 to July 19, 2019 (7445 observations). It is displayed and summarized in Figure 6 and Table 4, respectively. The time series plot shows a clear pattern of mean reversion, and Augmented Dickey-Fuller tests with reasonable lags all rejected the unit root hypothesis at 5% significance level, which justifies the use of stationary diffusion models. The mean and the standard deviation is of VIX is 19.21 and 7.76, respectively. Meanwhile, the skewness and the kurtosis are 2.12 and 10.85, respectively, suggesting that the stationary distribution deviates quite substantially from normality. This is more formally confirmed by the highly significant Jarque-Bera test statistic with a negligible p-value.

[Figure 6 and Table 4]

7.2 Models

We focus on whether two well known parametric transformed diffusion models proposed for modelling VIX are supported by the data against their semiparametric alternatives. The two parametric models are the transformed-OU model of Detemple and Osakwe (2000) (DO) and the transformed-CIR model of Eraker and Wang (2015) (EW). Specifically, the DO model is the exponential transform of the OU process, which can be written as

$$Y_t = \exp(X_t), \qquad dX_t = \kappa(\alpha - X_t)\,dt + \sigma dW_t$$

and the EW model is a parameter-dependent transformation of the CIR process, which is given by

$$Y_t = \frac{1}{X+\delta} + \varrho, \qquad dX_t = \kappa(\alpha - X_t)\,dt + \sigma\sqrt{X_t}dW_t$$

Meanwhile, the two semiparametric models we consider are the same two models considered in our simulations, namely, the nonparametrically transformed OU and CIR models, which we denote as

NPTOU and NPTCIR, respectively. Their associated normalized UPD processes are given in (6.1) and (6.2).

Importantly, we maintain the assumption that the VIX is a Markov diffusion process. In particular, we rule out jumps and stochastic volatility (SV) in the VIX which is inconsistent with the empirical findings of, e.g., Kaeck and Alexander (2013). However, their models are fully parametric and so impose much stronger functional form restrictions on the drift and diffusion component compared to our semiparametric approach. Specifically, jumps and SV components are often used to capture extremal events (fat tails). It is possible that these components are needed in explaining the VIX dynamics due to the restrictive drift and diffusion specifications they consider. Our semiparametric approach allows for more flexibility in this respect and so can be seen as a competing approach to capturing the same features in data. An interesting research topic would be to develop tools that allow for formal statistical comparison of our class of models against these alternative ones.

7.3 Results

For each of the two UPDs, we examine whether the parametric specification of the transformation is supported by the data. We do this by testing each of the parametric models against the semiparametric alternative where the transformation is left unspecified. We do so by computing a pseudo Likelihood Ratio (pseudo-LR) test statistic defined as the difference between the pseudo log-likelihood (pseudo-LL) of the semiparametric model and the log-likelihood (LL) of the parametric model. Since the model under the alternative is semiparametric and estimated by pseudo-ML, the pseudo-LL test statistic will not follow a χ^2-distribution. We therefore resort to a parametric bootstrap procedure: For each of the two pseudo-LR test, we simulate 1000 new time series from the parametric model using as data-generating parameter values the MLEs obtained from the original sample. For each of the 1000 new data sets, of the same size as the original one, we estimate both the parametric model and the semiparametric model and compute the corresponding pseudo-LR statistic. Finally, we use the 95th and 99th quantiles from the simulated distribution of the pseudo-LR statistic as our 5% and 1% bootstrap critical values, respectively.

The pseudo-LL is computed using the log-likelihood given in (4.1) with $U(y)$ and $\log U'(y;\theta)$ replaced by $\tilde{U}'(y;\theta)$ given in (4.3) and $\log \tilde{U}'(y;\theta) = \log \hat{f}_Y(y) - \log f_X\left(\tilde{U}(y);\theta\right)$, respectively. Here, $\hat{f}_Y(y)$ is the kernel density estimator which requires us choosing a bandwidth. There is a lack of consensus on the right procedure for choosing bandwidths for kernel estimators using dependent data. We therefore considered a sequence of bandwidths constructed by multiplying the Silverman's rule of thumb bandwidth, denoted as h_S, by a factor k between 0.75 and 1.75 on a small grid. Visual inspection of these density estimates revealed that with k is around 1.5, the resulting density appears to be the most satisfactory in terms of smoothness and the revelation of distributional features of the data. For this reason, we report our inferential results based on the relatively optimal bandwidth $1.5h_S = 2.0730$ below. However, our conclusions remain unchanged for any bandwidth within the aforementioned range.

Our estimation and testing results are reported in Table 5. The upper panel of the table presents the parameter estimates for the models together with their standard errors in the parentheses underneath. For the two semiparametric models, these were computed using the estimators proposed in Chen and Fan (2006). Recall that due to normalization, only κ is estimated for the NPTOU model and only κ and α for the NPTCIR model. In addition, while κ has the same interpretation (i.e. rate of mean reversion) and scale in all four models, α has different scales in the two transformed CIR models. For both the transformed OU and the transformed CIR classes of models, we can see that the PMLEs of the mean-reversion parameter $\hat{\kappa}$ are slightly lower than their corresponding MLE estimates. The same difference applies to their standard errors. This shows that parametric (mis-)specification of the stationary distribution does have a quite significant impact on the estimation of the dynamic parameters.

[Table 5]

The lower panel presents the LL values and the our pseudo-LR test results. We can see that the EW model has a much higher LL (-1.1585) than the DO model (-1.1724), suggesting much better goodness of fit to the data by the former. This is not entirely surprising because the EW model is more flexible both in terms of the UPD and the transformation function compared to the DO model. Meanwhile, the NPTCIR model has a higher pseudo-LL than the NPTOU. Since they have identical stationary distributions, such a difference is solely due to the additional flexibility of the UPD of the former. Most importantly, we see that when the underlying diffusions are the same, models with nonparametric transformation have much higher LLs than those with parametric transformations. More specifically, the resulting pseudo-LR between the NPTOU model and the DO model is 290.7263, and that between the NPTCIR model and the EW model is 40.8606. This proves that the exponential transformation of the DO model is too restrictive, and that while the transformation function of the EW model is more flexible, it is still rather restrictive relatively to our nonparametric alternative.

To formally assess the significance of the observed differences, we present the empirical 5% and 1% critical values and the corresponding p-values of our pseudo-LR tests, obtained from our bootstrap procedure described above. For both tests, we observe that those critical values are all negative and the p-values are both exactly zero. This means that the original pseudo-LRs of 290.7263 and 40.8606 are not only far greater than their corresponding empirical critical values but also greater than any of the bootstrap pseudo-LRs when the parametric model under the null hypothesis is true. This suggests that when either the DO model or the EW model is the true model, the corresponding NPTOU model or the NPTCIR model is unlikely to produce a higher LL value than the parametric model itself. This is fairly strong evidence that the parametric assumptions made by the DO and the EW models are not supported by our data and our nonparametrically transformed models are strongly favored.

The reason for the rejection of the two parametric models can be found in the implied stationary densities of the two models which we plot in Figure 7 together with the kernel density estimator. As can be seen from this figure, the parametric specifications are unable to capture the middle range of the empirical distribution of VIX; in contrast, the two semiparametric alternatives are constructed so that they match the empirical distribution exactly.

[Figure 7]

8 Conclusion

We propose a novel semiparametric approach for modelling stationary nonlinear univariate diffusions. The class of models can be thought of as Markov copula models where the copula is implied by the UPD model. Primitive conditions for the identification of the UPD parameters together with the unknown transformations from discrete samples are provided. We derive the asymptotic properties for our semiparametric likelihood-based estimators of the UPD parameters and kernel-based drift and diffusion estimators. Our simulation results suggest that our semiparametric method performs well in finite sample compared to the fully parametric method, and our relatively simple application shows that the parametric assumptions on the transformation function of the well known DO model and EW model are rejected by the data against our nonparametric alternatives. Potential future work under this framework may include extensions to multivariate diffusions and jump-diffusions.

References

Ahn, D.-H., Gao, B., 1999. A Parametric nonlinear 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.

Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation Approach. Econometrica 70, 223-262.

Bandi, F.M., 2002. Short-term interest rate dynamics: A spatial approach. Journal of Financial Economics 65, 73-110.

Bandi, F.M., Phillips, P.C.B., 2003, Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241-283.

Beare, B.K., 2010. Copulas and temporal dependence. Econometrica 78, 395-410.

Bladt, M., Sørensen, M., 2014. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Bernoulli 20, 645-675.

Bu, R., Cheng, J., Hadri, K., 2017. Specification analysis in regime-switching continuous-time diffusion models for market volatility. Studies in Nonlinear Dynamics and Econometrics 21(1), 65-80.

Bu, R., Fredj, J., Li, Y., 2017. An empirical comparison of transformed diffusion models for VIX and VIX futures. Journal of International Financial Markets, Institutions and Money 46, 116-127.

Bu, R., Giet, L., Hadri, K., 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.

Chen, X., Fan, Y., 2006. Estimation of copula-based semiparametric time series models. Journal of Econometrics 130, 307-335.

Chen, X., Hansen, L.P., Scheinkman, J., 2009. Nonlinear principal components and long run implications of multivariate diffusions. Annals of Statistics 37, 4279-4312.

Chen, X., Hansen, L.P., Carrasco, M., 2010. Nonlinearity and temporal dependence. Journal of Econometrics 155, 155-169.

Chen, X., Wu, W.B., Yi, Y., 2009. Efficient estimation of copula-based semiparametric Markov models. Annals of Statistics 37, 4214-4253.

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., Hansen, L., Luttmer, E., Scheinkman, J., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525-577.

Cox, J., Ingersoll, J., 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.

Eraker, B., Wang, J., 2015. A non-linear dynamic model of the variance risk premium, Journal of Econometrics 187, 547-556.

Forman, J.L., Sørensen, M., 2014. A transformation approach to modelling multi-modal diffusions. Journal of Statistical Planning and Inference 146, 56-69.

Gobet, E., Hoffmann, M., Reiß, M., 2004. Nonparametric estimation of scalar diffusions based on low frequency data. Annals of Statistics 32, 2223-2253.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705-730.

Hansen, L.P., Scheinkman, J., Touzi, N., 1998. Spectral methods for identifying scalar diffusions. Journal of Econometrics 86, 1-32.

Jiang, G., Knight, J., 1997. A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. Econometric Theory 13, 615-645.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Jondeau, E., Rockinger, M., 2006. The copula-GARCH model of conditional dependencies - an international stock application. Journal of International Money and Finance 25, 827-853.

Kaeck, A., Alexander, C., 2013. Continuous-time VIX dynamics: On the role of stochastic volatility of volatility. International Review of Financial Analysis 28, 46-56.

Kanaya, S., Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach. Econometric Theory 33, 874-914.

Kanaya, S., Kristensen, D., 2016. Estimation of stochastic volatility models by nonparametric filtering. Econometric Theory 32, 861-916.

Karatzas, I., Shreve, S., 1991. Brownian Motion and Stochastic Calculus, 2nd ed. Springer-Verlag, New York.

Kessler, M., 1997, Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics 24, 211–229.

Kessler, M., Sørensen, M., 1999. Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli 5, 299-314.

Kristensen, D., 2010. Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. Journal of Econometrics 156, 239-259.

Kristensen, D., 2011. Semi-nonparametric estimation and misspecification testing of diffusion models. Journal of Econometrics 164, 382-403.

Kristensen, D., Shin, Y., 2012. Estimation of dynamic models with nonparametric simulated maximum likelihood. Journal of Econometrics 167, 76-94.

Li, C., 2013. Maximum-likelihood estimation for diffusion processes via closed-form density expansions, Annals of Statistics 41, 1350-1380.

McKean, H.P., 1969. Stochastic Integrals. Academic Press. New York.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam (chapter 36).

Patton, A., 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometrics 2, 130-168.

Robinson, P., 1983. Nonparametric estimators for time series. Journal of Time Series Analysis 4, 185-207.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. Chapman and Hall, London.

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.

A Proofs

Proof of Theorem 3.2. From eqs. (3.2)-(3.5), it is obvious that (3.6)-(3.7) imply $\mathcal{S} \sim \tilde{\mathcal{S}}$. Now, suppose that $\mathcal{S} \sim \tilde{\mathcal{S}}$; this implies that $\mu_Y(y;\mathcal{S}) = \mu_Y\left(y;\tilde{\mathcal{S}}\right)$ and $\sigma_Y^2(y;\mathcal{S}) = \sigma_Y^2\left(y;\tilde{\mathcal{S}}\right)$, where μ_Y and σ_Y^2 are given in eqs. (2.6)-(2.7). That is, for all $y \in \mathcal{Y}$,

$$\begin{aligned}
\frac{\mu_X(U(y);\theta)}{U'(y)} - \frac{1}{2}\sigma_X^2(U(y);\theta)\frac{U''(y)}{U'(y)^3} &= \frac{\mu_X\left(\tilde{U}(y);\tilde{\theta}\right)}{\tilde{U}'(y)} - \frac{1}{2}\sigma_X^2\left(\tilde{U}(y);\tilde{\theta}\right)\frac{\tilde{U}''(y)}{\tilde{U}'(y)^3}, \\
\frac{\sigma_X(U(y);\theta)}{U'(y)} &= \frac{\sigma_X\left(\tilde{U}(y);\tilde{\theta}\right)}{\tilde{U}'(y)}.
\end{aligned}$$

Since V is one-to-one we can set $y = V(x)$ in the above to obtain the following for all $x \in \mathcal{X}$,

$$\begin{aligned}
&\frac{\mu_X(U(V(x));\theta)}{U'(V(x))} - \frac{1}{2}\sigma_X^2(U(V(x));\theta)\frac{U''(V(x))}{U'(V(x))^3} \qquad \text{(A.1)} \\
&= \frac{\mu_X\left(\tilde{U}(V(x));\tilde{\theta}\right)}{\tilde{U}'(V(x))} - \frac{1}{2}\sigma_X^2\left(\tilde{U}(V(x));\tilde{\theta}\right)\frac{\tilde{U}''(V(x))}{\tilde{U}'(V(x))^3}, \\
\frac{\sigma_X(U(V(x));\theta)}{U'(V(x))} &= \frac{\sigma_X\left(\tilde{U}(V(x));\tilde{\theta}\right)}{\tilde{U}'(V(x))}. \qquad \text{(A.2)}
\end{aligned}$$

Define $T(x) = \tilde{U}(V(x)) \Leftrightarrow T^{-1}(x) = U\left(\tilde{V}(x)\right)$, and observe that

$$U(V(x)) = x, \quad U'(V(x))V'(x) = 1, \quad \frac{\partial T(x)}{\partial x} = \tilde{U}'(V(x))V'(x).$$

Eq. (A.2) combined with the above implies (3.7)(ii),

$$\sigma_X(x;\theta) = \frac{\sigma_X(U(V(x));\theta)}{U'(V(x))V'(x)} = \frac{\sigma_X\left(\tilde{U}(V(x));\tilde{\theta}\right)}{\tilde{U}'(V(x))V'(x)} = \frac{\sigma_X\left(T(x);\tilde{\theta}\right)}{\partial T(x)/(\partial x)} = \sigma_{T^{-1}(X)}\left(x;\tilde{\theta}\right). \qquad \text{(A.3)}$$

Next, divide through with $V'(x)$ in (A.1) and rearrange to obtain

$$\begin{aligned}
\mu_X(x;\theta) &= \frac{\mu_X\left(T(x);\tilde{\theta}\right)}{\partial T(x)/(\partial x)} + \frac{1}{2}\left\{\sigma_X^2(x;\theta)\frac{U''(V(x))}{U'(V(x))^3V'(x)} - \sigma_X^2\left(T^{-1}(x);\tilde{\theta}\right)\frac{\tilde{U}''(V(x))}{\tilde{U}'(V(x))^3V'(x)}\right\} \\
&= \frac{\mu_X\left(T(x);\tilde{\theta}\right)}{\partial T(x)/(\partial x)} + \frac{1}{2}\sigma_X^2\left(T(x);\tilde{\theta}\right)\left\{\frac{1}{\tilde{U}'(V(x))^2V'(x)^3}\frac{U''(V(x))}{U'(V(x))^3} - \frac{\tilde{U}''(V(x))}{\tilde{U}'(V(x))^3V'(x)}\right\}
\end{aligned}$$

where the second equality uses (A.3). Eq. (3.7)(i) now follows since

$$\begin{aligned}
&\frac{1}{\tilde{U}'(V(x))^2V'(x)^3}\frac{U''(V(x))}{U'(V(x))^3} - \frac{\tilde{U}''(V(x))}{\tilde{U}'(V(x))^3V'(x)} \\
&= \frac{1}{\tilde{U}'(V(x))^3V'(x)^3}\left[\frac{\tilde{U}'(V(x))U''(V(x))}{U'(V(x))^3} - \tilde{U}''(V(x))V'(x)^2\right] \\
&= \frac{-1}{\tilde{U}'(V(x))^3V'(x)^3}\left[\tilde{U}'(V(x))V''(x) + \tilde{U}''(V(x))V'(x)^2\right] \\
&= -\frac{\partial^2 T(x)/\left(\partial x^2\right)}{\partial T(x)/(\partial x)^3}.
\end{aligned}$$

■

Proof of Theorem 5.1. We first note that the PMLE takes the same form as the one analyzed in Chen and Fan (2006) with the general copula considered in their work satisfying eq. (2.17). The desired result will follow if we can verify that the conditions stated in their proof are satisfied by our assumptions: First, by Assumptions 2.1, the discrete sample $\{X_{i\Delta} : i = 0, 1, \ldots, n\}$ generated by the UPD X is first-order Markovian and with marginal density $f_X(x;\theta)$ and transition density $p_X(x|x_0;\theta)$. Hence, the copula density $c_X(u_0, u; \theta)$ in (2.17) implied by X is absolutely continuous with respect to the Lebesgue measure on $[0,1]^2$ due to its continuity in $F_X(x;\theta)$, $f_X(x;\theta)$ and $p_X(x|x_0;\theta)$. Moreover, the implied copula is neither the Fréchet-Hoeffding upper or lower bound due to Assumption 2.1, i.e., $\sigma_X^2(x;\theta) > 0$ for all $x \in \mathcal{X}$. Thus, Chen and Fan (2006, Assumption 1) is satisfied. Second, our Assumption 4.2(i) ensures that X is β-mixing with polynomial decay rate. Third, by Theorem 2.1, Y is mixing with the same mixing properties as X and so satisfies Chen and Fan (2006, Assumption 1). The remaining conditions are met by Assumption 4.3(i).

For the analysis of the proposed sieve MLE, we note that it takes the same form as the one analyzed in Chen, Wu and Yi (2009) and so their results carry over to our setting. Their Assumption M and assumption of β-mixing property are satisfied by Y under our Assumptions 2.1, 2.2, and 4.2(ii) together with our Theorem 2.1. The remaining conditions are met by Assumption 4.3(ii). ■

Proof of Theorem 5.2. Similar to the proof strategy employed in Lemma C.1, we define

$$\tilde{\mu}_Y(y) = \frac{\mu_X(U(y);\theta)}{U'(y)} - \frac{1}{2}\sigma_X^2(U(y);\theta)\frac{\hat{U}''(y)}{U'(y)^3}, \quad \tilde{\sigma}_Y^2(y) = \frac{\sigma_X^2(U(y);\theta)}{\hat{U}'(y)^2},$$

and, with $f_Y^{(i)}$ denoting the ith derivative of f_Y and similar for other functions, arrive at

$$\begin{aligned}
&\sqrt{nh^3}\left\{\hat{\mu}_Y(y) - \mu_Y(y) - \frac{1}{2}h^2\kappa_2\frac{f_Y^{(3)}(y)}{f_X(U(y);\theta)}\left[-\frac{\sigma_X^2(U(y);\theta)}{2U'(y)^3}\right]\right\} \\
= &\sqrt{nh^3}\left\{\tilde{\mu}_Y(y) - \mu_Y(y) - \frac{1}{2}h^2\kappa_2\frac{f_Y^{(3)}(y)}{f_X(U(y);\theta)}\left[-\frac{\sigma_X^2(U(y);\theta)}{2U'(y)^3}\right]\right\} + o_p(1) \\
= &-\frac{\sigma_X^2(U(y);\theta)}{2U'(y)^3}\sqrt{nh^3}\left\{\hat{U}^{(2)}(y) - U^{(2)}(y) - \frac{1}{2}h^2\kappa_2\frac{f_Y^{(3)}(y)}{f_X(U(y);\theta)}\right\} + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{nh}\left\{\hat{\sigma}_Y^2(y) - \sigma_Y^2(y) - \frac{1}{2}h^2\kappa_2\frac{f_Y^{(2)}(y)}{f_X(U(y);\theta)}\left[-\frac{2\sigma_X^2(U(y);\theta)}{U'(y)^3}\right]\right\} \\
= &\sqrt{nh}\left\{\tilde{\sigma}_Y^2(y) - \sigma_Y^2(y) - \frac{1}{2}h^2\kappa_2\frac{f_Y^{(2)}(y)}{f_X(U(y);\theta)}\left[-\frac{2\sigma_X^2(U(y);\theta)}{U'(y)^3}\right]\right\} + o_p(1) \\
= &-\frac{2\sigma_X^2(U(y);\theta)}{U'(y)^3}\sqrt{nh}\left\{\hat{U}'(y) - U'(y) - \frac{1}{2}h^2\kappa_2\frac{f_Y^{(2)}(y)}{f_X(U(y);\theta)}\right\} + o_p(1).
\end{aligned}$$

These together with (C.1) and (C.2) of Lemma C.1 and Slutsky's Theorem complete the proof. ■

B Verification of conditions for OU and CIR model

We here verify the technical conditions of Chen and Fan (2006) for the normalized versions of the OU and CIR model given in eqs. (6.1) and (6.2), respectively. For both examples, we will require that $U(y;\theta)$, as defined in eq. (2.16), and its first and second-order derivatives w.r.t θ are polynomially bounded in y. This imposes growth restrictions on the transformation function and is used to easily verify various moment conditions in the following. Also note that the criterion $l(U_{i-1},U_i;\theta)$ in Chen and Fan (2006) takes the form $l(U_{i-1},U_i;\theta) := \log p_X\left(U(Y_{i\Delta};\theta);U\left(Y_{(i-1)\Delta};\theta\right);\theta\right) - \log f_X\left(U(Y_{i\Delta};\theta);\theta\right)$, where $U_i = F_Y(Y_{i\Delta})$, in our notation.

B.1 OU model

Assumption 4.2: It is easily seen that $\left\{\frac{\mu_X(x;\theta_0)}{\sigma_X(x;\theta_0)} - \frac{1}{2}\frac{\partial\sigma_X(x;\theta_0)}{\partial x}\right\} = -\sqrt{\frac{\kappa}{2}}x$ and $\frac{s(x;\theta_0)\sigma_X(x;\theta_0)}{S(x;\theta_0)} = \exp\left(\frac{x^2}{2}\right)/\int_{x^*}^{x}\exp\left(\frac{z^2}{2}\right)dz$. Assumption 4.2 is verified by taking the relevant limits.

Assumption 4.3: The implied copula of the normalized OU process is Gaussian, for which Assumption 4.3(i) and 4.3(ii) are satisfied as discussed in Chen and Fan (2006) and Chen, Wu, and Yi (2009), respectively.

B.2 CIR model

Assumption 4.2: We obtain $\left\{\frac{\mu_X(x;\theta_0)}{\sigma_X(x;\theta_0)} - \frac{1}{2}\frac{\partial\sigma_X(x;\theta_0)}{\partial x}\right\} = \frac{(2\alpha-1)}{2}\sqrt{\frac{\kappa}{2x}} - \sqrt{\frac{\kappa}{4x}}$ and $\frac{s(x;\theta_0)\sigma_X(x;\theta_0)}{S(x;\theta_0)} = \frac{\exp\{x\}}{x^\alpha}\sqrt{2\kappa}\sqrt{x}/\int_{x^*}^{x}\frac{\exp\{z\}}{z^\alpha}dz$ and the assumption is verified by taking relevant limits.

Assumption 4.3. First observe that

$$p_X(x|x_0;\theta) = \exp\left[c_0(\theta) - c(\theta)\left(x + e^{-\kappa\Delta}x_0\right)\right]\frac{x_0}{x}I_{\alpha-1}\left(2c^2(\theta)\sqrt{xx_0}\right),$$

where $I_q(\cdot)$ is the so-called modified Bessel function of the first kind and of order q and $c_0(\theta,\Delta) > 0$ and $c(\theta,\Delta) > 0$ are analytic functions. Moreover, f_X is here the density of a gamma distribution and so all polynomial moments of X exist. Since U is assumed to be polynomially bounded, this implies that all polynomial moments of Y also exist. All smoothness conditions imposed in Chen and Fan (2006) are trivially satisfied since $p_X(x|x_0;\theta)$ and $U(y;\theta)$ are twice continuously differentiable w.r.t their arguments and so will not be discussed any further. Similarly, we have already shown that Y is geometrically mixing. It remains to verify the moment conditions and the identifying restrictions imposed in C1-C.5 in Proposition 4.2 and A2-A6 in Chen and Fan (2006).

C1 is satisfied if we restrict $\theta = (\alpha,\kappa)$ to be situated in a compact set on $\mathbb{R}_+^2$ that contains the true value. Observe that

$$\log p_X(x|x_0;\theta) = c_0(\theta) - c(\theta)\left(x + e^{-\kappa\Delta}x_0\right) + \log\left(\frac{x_0}{x}\right) + \log I_{\alpha-1}\left(2c^2(\theta)\sqrt{xx_0}\right).$$

Thus,

$$
\begin{aligned}
s_{\theta}\left(x|x_0;\theta\right) \quad : \quad &= \frac{\partial \log p_X\left(x;x_0;\theta\right)}{\partial\theta} \\
&= \dot{c}_0\left(\theta\right) - \dot{c}\left(\theta\right)\left(x + e^{-\kappa\Delta}x_0\right) + c\left(\theta\right)\Delta e^{-\kappa\Delta}x_0 + \frac{I'_{\alpha-1}\left(2c^2\left(\theta\right)\sqrt{xx_0}\right)4c\left(\theta\right)\sqrt{xx_0}\dot{c}\left(\theta\right)}{I_{\alpha-1}\left(2c^2\left(\theta\right)\sqrt{xx_0}\right)} \\
&\quad + \begin{bmatrix} \frac{\dot{I}_{\alpha-1}\left(2c^2(\theta)\sqrt{xx_0}\right)}{I_{\alpha-1}\left(2c^2(\theta)\sqrt{xx_0}\right)} \\ 0 \end{bmatrix},
\end{aligned}
$$

where $\dot{c}_0\left(\theta\right) = \partial c_0\left(\theta\right)/\left(\partial\theta\right)$ and similar for other functions, $I'_{\alpha-1}\left(x\right) = \partial I_{\alpha-1}\left(x\right)/\left(\partial x\right)$, and $\dot{I}_{\alpha-1}\left(x\right) = \partial I_{\alpha-1}\left(x\right)/\left(\partial\alpha\right)$. It is easily verified that $\left|I'_{\alpha-1}\left(x\right)/I_{\alpha-1}\left(x\right)\right|$ and $\left|\left|I'_{\alpha-1}\left(x\right)/I_{\alpha-1}\left(x\right)\right|\right|$ are both bounded by a polynomial in x. Thus, $\left\|s_X\left(x|x_0;\theta\right)\right\|$ is bounded by a polynomial uniformly in $\theta \in \Theta$. The expressions of $s_x\left(x|x_0;\theta\right) := \partial \log p_X\left(x;x_0;\theta\right)/\left(\partial x\right)$ and $s_{x_0}\left(x|x_0;\theta\right) := \partial \log p_X\left(x;x_0;\theta\right)/\left(\partial x_0\right)$ are on a similar form and also polynomially bounded. Now, observe that

$$
\begin{aligned}
l_{\theta}\left(U_{i-1},U_i;\theta\right) \quad : \quad &= \frac{\partial l\left(U_{i-1},U_i;\theta\right)}{\partial\theta} \\
&= s_{\theta}\left(U\left(Y_{i\Delta};\theta\right)|U\left(Y_{(i-1)\Delta};\theta\right);\theta\right) \\
&\quad + s_x\left(U\left(Y_{i\Delta};\theta\right)|U\left(Y_{(i-1)\Delta};\theta\right);\theta\right)\dot{U}\left(Y_{i\Delta};\theta\right) \\
&\quad + s_{x_0}\left(U\left(Y_{i\Delta};\theta\right)|U\left(Y_{(i-1)\Delta};\theta\right);\theta\right)\dot{U}\left(Y_{(i-1)\Delta};\theta\right) \\
&\quad - \frac{\partial \log f_X\left(U\left(Y_{i\Delta};\theta\right);\theta\right)}{\partial\theta}.
\end{aligned}
$$

Given that the model is correctly specified and identified, it follows by standard arguments for MLE that $E\left[l_{\theta}\left(U_i,U_{i-1};\theta\right)\right] = 0$ if and only if θ equals the true value.

C4. From the above expression of $l_{\theta}\left(U_i,U_{i-1};\theta\right)$ together with our assumption on $U\left(y;\theta\right)$, it is easily checked that it is bounded by a polynomial in $\left(Y_{i\Delta},Y_{(i-1)\Delta}\right)$ uniformly in $\theta \in \Theta$. It now follows that $E\left[\sup_{\theta}\left\|l_{\theta}\left(U_i,U_{i-1};\theta\right)\right\|^p\right] < \infty$ for any $p \geq 1$.

C5.

$$
l_{\theta,1}\left(U_{i-1},U_i;\theta\right) = \frac{\partial l_{\theta}\left(U_{i-1},U_i;\theta\right)}{\partial U_{i-1}}, \quad l_{\theta,2}\left(U_{i-1},U_i;\theta\right) = \frac{\partial l_{\theta}\left(U_{i-1},U_i;\theta\right)}{\partial U_i}
$$

are again bounded by polynomials in $\left(Y_{i\Delta},Y_{(i-1)\Delta}\right)$ and so have all relevant moments.

A1(ii)-(iii). With $W_{1,i}$ and $W_{2,i}$ defined in (4.2)-(4.3) in Chen and Fan (2006) and

$$
l_{\theta,\theta}\left(U_{i-1},U_i;\theta\right) = \frac{\partial^2 l\left(U_{i-1},U_i;\theta\right)}{\partial\theta\partial\theta'},
$$

$$
\lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{l_{\theta}\left(U_{i-1},U_i;\theta\right) + W_{1,i} + W_{2,i}\right\}\right),
$$

and $E\left[l_{\theta,\theta}\left(U_{i-1},U_i;\theta\right)\right]$ to have full rank. We have been unable to verify these two conditions due to the complex form of the score and hessian of the CIR model.

A4. Observe that $\left|W_{1,i}\right| \leq E\left[\left|U_{i-1}\right|\left\|l_{\theta,1}\left(U_{i-1},U_i;\theta\right)\right\|\right] < \infty$ and similar for $W_{2,i}$. Thus, both have all relevant moments.

A5-A6 have already been verified above.

C Lemma

Lemma C.1 *Under Assumptions 2.1-2.2, 4.2(i), and 4.4-4.6, we have as* $n \to \infty$*,* $h \to 0$*,* $nh \to \infty$*,*

$$\sqrt{nh}\left\{\hat{U}'(y) - U'(y) - \frac{1}{2}h^2\kappa_2 \frac{f_Y''(y)}{f_X(U(y);\theta_0)}\right\} \to^d N\left(0, \frac{U'(y)^2}{f_Y(y)}\int_{\mathbb{R}} K(z)^2\,dz\right), \tag{C.1}$$

and as $n \to \infty$*,* $h \to 0$*,* $nh^3 \to \infty$*,*

$$\sqrt{nh^3}\left\{\hat{U}''(y) - U''(y) - \frac{1}{2}h^2\kappa_2 \frac{f_Y'''(y)}{f_X(U(y);\theta_0)}\right\} \to^d N\left(0, \frac{U'(y)^2}{f_Y(y)}\int_{\mathbb{R}} K'(z)^2\,dz\right). \tag{C.2}$$

Proof. With $\hat{F}_Y(y)$ given in (4.2), let $\hat{f}_Y^{(i)}(y) = \hat{F}_Y^{(i+1)}(y)$, for $i = 1, 2$, be the ith derivative of the kernel marginal density estimator. Using standard methods for kernel estimators (c.f. Robinson, 1983), we obtain under the assumptions of the lemma that, as $n \to \infty, h \to 0$, and $nh^{1+2i} \to \infty$,

$$\sqrt{nh^{1+2i}}\left\{\hat{f}_Y^{(i)}(y) - f_Y^{(i)}(y) - \frac{1}{2}h^2\kappa_2 f_Y^{(i+2)}(y)\right\} \to^d N(0, V_i(y)) \tag{C.3}$$

where $V_i(y) = f_Y(y)\int_{\mathbb{R}} K^{(i)}(z)^2\,dz$. Assumptions 2.1 and 4.4 ensure that $f_Y(y)$ is sufficiently smooth so that $f_Y^{(2)}(y)$ and $f_Y^{(3)}(y)$ exist. Assumption 4.2(i) and 4.6 regulate the mixing property of Y and the kernel function, respectively, as required by Robinson (1983).

From (4.4) we have $\hat{U}'(y) = \hat{f}_Y(y)/f_X(\hat{U}(y);\hat{\theta})$. Now define $\hat{U}_0'(y) = \hat{f}_Y(y)/f_X(U(y);\theta_0)$ and note that Assumption 4.4 and 4.5 together with the delta-method imply $\hat{U}'(y) - \hat{U}_0'(y) = O_P(1/\sqrt{n}) = o_P(1/\sqrt{nh})$. It then follows that

$$\begin{aligned}
&\sqrt{nh}\left\{\hat{U}'(y) - U'(y) - \frac{1}{2}h^2\kappa_2 f_Y^{(2)}(y)\frac{1}{f_X(U(y);\theta_0)}\right\} \\
=\ &\sqrt{nh}\left\{o_P\left(1/\sqrt{nh}\right) + \hat{U}_0'(y) - U'(y) - \frac{1}{2}h^2\kappa_2 f_Y^{(2)}(y)\frac{1}{f_X(U(y);\theta_0)}\right\} \\
=\ &\frac{1}{f_X(U(y);\theta_0)}\sqrt{nh}\left\{\hat{f}_Y(y) - f_Y(y) - \frac{1}{2}h^2\kappa_2 f_Y^{(2)}(y)\right\} + o_P(1).
\end{aligned}$$

Using (C.3) and the same arguments as in Kristensen (2011, Proof of Theorem 1), we arrive at (C.1).

Next, observe that $U''(y) = \frac{f_Y'(y)}{f_X(U(y);\theta)} - \frac{f_X'(U(y);\theta)f_Y(y)^2}{f_X(U(y);\theta)^3}$ where $f_X'(x;\theta)$ and $f_Y'(y)$ are the first derivatives of $f_X(x;\theta)$ and $f_Y(y)$, respectively. Similarly, it is easily checked that $\hat{U}''(y) = \frac{\hat{f}_Y'(y)}{f_X(\hat{U}(y);\hat{\theta})} - \frac{f_X'(\hat{U}(y);\hat{\theta})\hat{f}_Y(y)^2}{f_X(\hat{U}(y);\hat{\theta})^3}$. Define $\hat{U}_0''(y) = \frac{\hat{f}_Y'(y)}{f_X(U(y);\theta_0)} - \frac{f_X'(U(y);\theta_0)f_Y(y)^2}{f_X(U(y);\theta_0)^3}$ and apply arguments similar to before to obtain

$$\begin{aligned}
&\sqrt{nh^3}\left\{\hat{U}''(y) - U''(y) - \frac{1}{2}h^2\kappa_2 f_Y^{(3)}(y)\frac{1}{f_X(U(y);\theta_0)}\right\} \\
=\ &\frac{1}{f_X(U(y);\theta_0)}\sqrt{nh^3}\left\{\hat{f}_Y'(y) - f_Y'(y) - \frac{1}{2}h^2\kappa_2 f_Y^{(3)}(y)\right\} + o_p(1)
\end{aligned}$$

which together with (C.3) yield (C.2). ■

D Tables and Figures

Table 1: Bias and RMSE of κ in the OU-SKST Model

		Bias/κ			
Sample Size		2202		5505	
True Parameter Value	ρ_1	PPMLE	PMLE	PPMLE	PMLE
$\kappa = 1.1376$	0.9944	0.6121	1.1379	0.2690	0.5054
$\kappa = 5.6882$	0.9758	0.1230	0.1987	0.0652	0.0939
$\kappa = 11.377$	0.9531	0.0656	0.0888	0.0400	0.0441
$\kappa = 22.753$	0.9093	0.0385	0.0383	0.0270	0.0210
		RMSE/κ			
Sample Size		2202		5505	
True Parameter Value	ρ_1	PPMLE	PMLE	PPMLE	PMLE
$\kappa = 1.1376$	0.9944	0.8603	1.2932	0.4476	0.6224
$\kappa = 5.6882$	0.9758	0.2420	0.2930	0.1454	0.1655
$\kappa = 11.377$	0.9531	0.1574	0.1730	0.0974	0.1044
$\kappa = 22.753$	0.9093	0.1059	0.1133	0.0668	0.0711

Table 2: Bias and RMSE of κ in the CIR-SKST Model

		Bias/κ			
Sample Size		2202		5505	
True Parameter Values	ρ_1	PPMLE	PMLE	PPMLE	PMLE
$(\kappa, \alpha) = (0.7653, 1.1653)$	0.9921	0.9023	1.5269	0.4576	0.7717
$(\kappa, \alpha) = (3.8267, 1.1653)$	0.9675	0.2358	0.3347	0.1194	0.1754
$(\kappa, \alpha) = (7.6533, 1.1653)$	0.9399	0.1328	0.1816	0.0646	0.0853
$(\kappa, \alpha) = (15.307, 1.1653)$	0.8917	0.0768	0.0928	0.0349	0.0398

		RMSE/κ			
Sample Size		2202		5505	
True Parameter Values	ρ_1	PPMLE	PMLE	PPMLE	PMLE
$(\kappa, \alpha) = (0.7653, 1.1653)$	0.9921	1.2424	1.7509	0.6692	0.9231
$(\kappa, \alpha) = (3.8267, 1.1653)$	0.9675	0.3881	0.4511	0.2363	0.2746
$(\kappa, \alpha) = (7.6533, 1.1653)$	0.9399	0.2431	0.2771	0.1498	0.1672
$(\kappa, \alpha) = (15.307, 1.1653)$	0.8917	0.1712	0.1847	0.1003	0.1068

Table 3: Bias and RMSE of α in the CIR-SKST Model

		Bias/α			
Sample Size		2202		5505	
True Parameter Values	ρ_1	PPMLE	PMLE	PPMLE	PMLE
$(\kappa, \alpha) = (0.7653, 1.1653)$	0.9921	0.9458	1.0299	0.6192	0.8720
$(\kappa, \alpha) = (3.8267, 1.1653)$	0.9675	0.4353	0.5171	0.1899	0.2554
$(\kappa, \alpha) = (7.6533, 1.1653)$	0.9399	0.2633	0.3152	0.1033	0.1279
$(\kappa, \alpha) = (15.307, 1.1653)$	0.8917	0.1302	0.1646	0.0663	0.0780

		RMSE/α			
Sample Size		2202		5505	
True Parameter Values	ρ_1	PPMLE	PMLE	PPMLE	PMLE
$(\kappa, \alpha) = (0.7653, 1.1653)$	0.9921	1.5614	1.5784	1.1222	1.4309
$(\kappa, \alpha) = (3.8267, 1.1653)$	0.9675	0.8443	0.9197	0.3867	0.4462
$(\kappa, \alpha) = (7.6533, 1.1653)$	0.9399	0.5473	0.5695	0.2298	0.2558
$(\kappa, \alpha) = (15.307, 1.1653)$	0.8917	0.2802	0.3139	0.1453	0.1684

Table 4: Descriptive Statistics of Daily VIX

Sample Period	January 2, 1990 - July 19, 2019
Sample Size	7445
Mean	19.21
Median	17.31
Std Dev.	7.76
Skewness	2.12
Kurtosis	10.85
Jarque-Bera Statistic	24669.26

Table 5: Model Estimation and Pseudo-LR Test Results

	Transformed OU		Transformed CIR	
	DO	NPTOU	EW	NPTCIR
$\hat{\kappa}$	4.4888	3.8191	4.0741	3.7541
	(0.5795)	(0.4525)	(0.5597)	(0.4257)
$\hat{\alpha}$	2.8890		0.0524	14.6916
	(0.0423)		(0.0032)	(8.8484)
$\hat{\sigma}^2$	1.0818		0.0695	
	(0.0179)		(0.0097)	
$\hat{\varrho}$			0.1916	
			(0.4827)	
$\hat{\delta}$			0.0072	
			(0.0029)	
$LL\left(10^4\right)$	-1.1724	-1.1579	-1.1585	-1.1565
LR		290.7263		40.8606
$CV_{0.05}$		-52.1521		-23.6766
$CV_{0.01}$		-30.5511		-10.9027
p-value		0.0000		0.0000

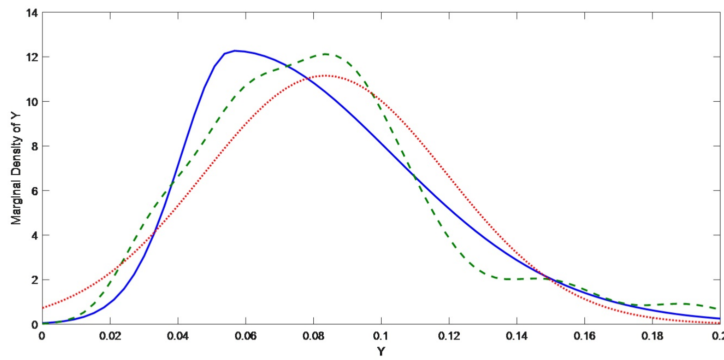

Figure 1: Marginal Densities of the Eurodollar Rates.
Solid = SKST Density, Dashed = Kernel Density, Dotted = Normal Density

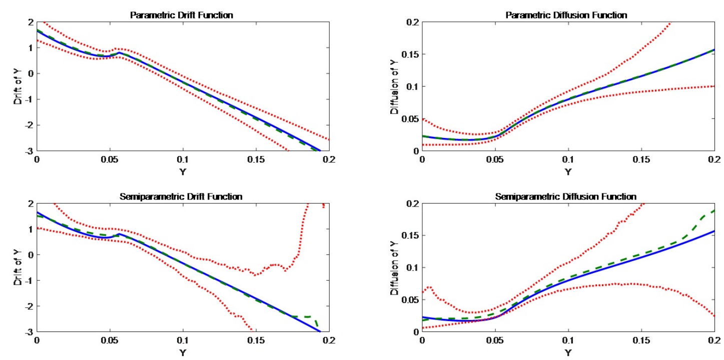

Figure 2: Estimated Drift and Diffusion for the OU-SKST Model ($T = 2202$) .
Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

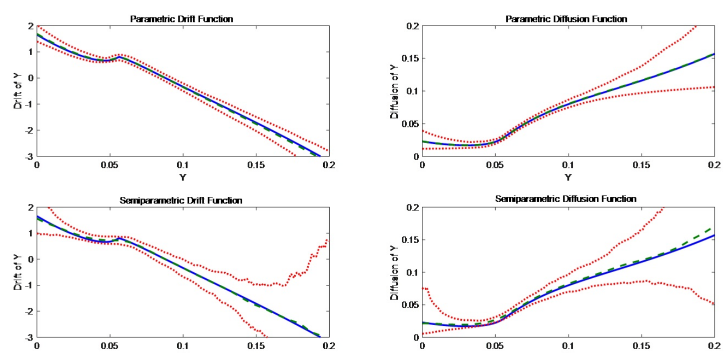

Figure 3: Estimated Drift and Diffusion for the OU-SKST Model ($T = 5505$).
Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

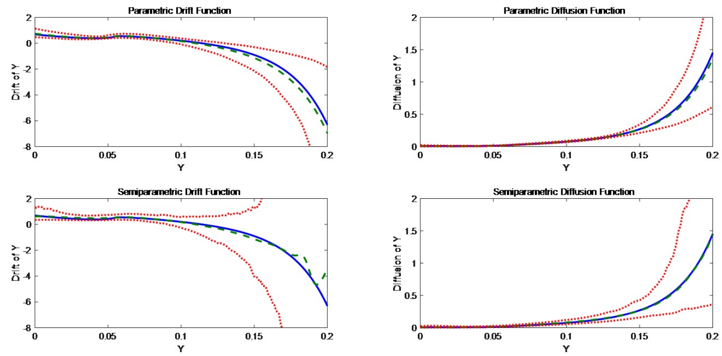

Figure 4: Estimated Drift and Diffusion for the CIR-SKST Model ($T = 2202$).
Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

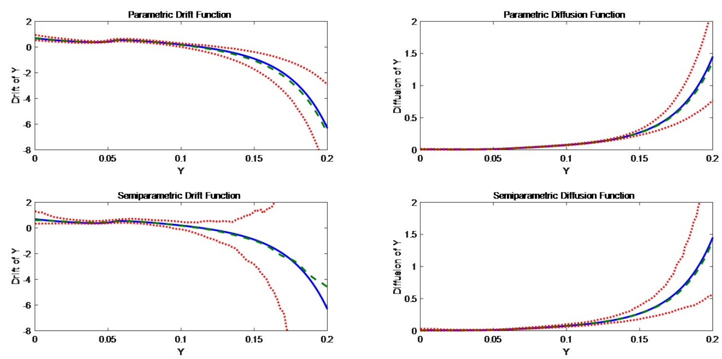

Figure 5: Estimated Drift and Diffusion for the CIR-SKST Model ($T = 5505$).
Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

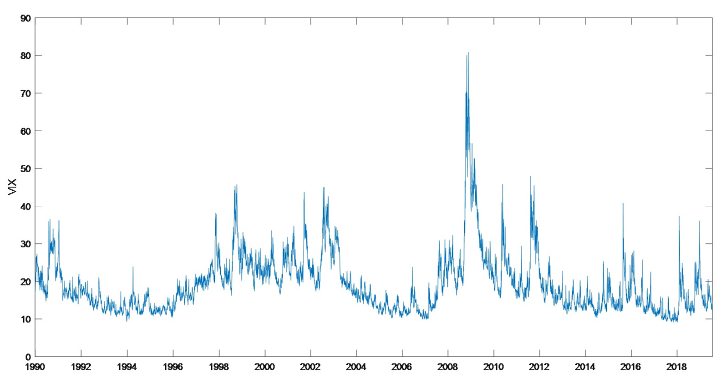

Figure 6: Time Series of Daily VIX

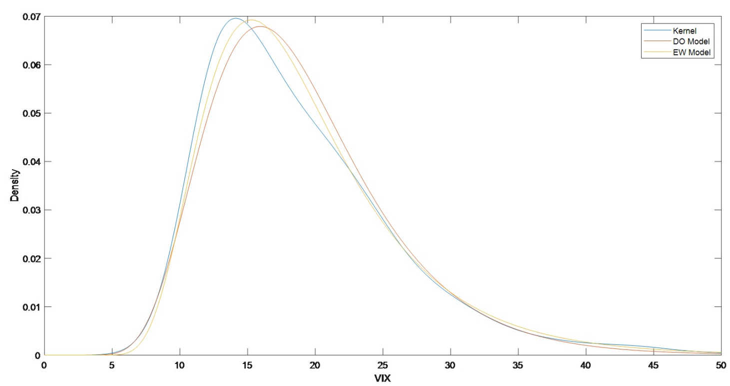

Figure 7: Estimated Marginal Densities of Daily VIX