DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Diffusion Copulas: Identification and Estimation

Ruijun Bu, Kaddour Hadri and Dennis Kristensen

CREATES Research Paper 2018-20

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Diffusion Copulas: Identification and Estimation

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

July 2018

Abstract

We propose a new semiparametric approach for modelling nonlinear univariate diffusions, where the observed processes are nonparametric transformations of underlying parametric diffusions (UPDs). 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. Semiparametric likelihood-based estimators of the UPD parameters are developed and we show that under regularity conditions both the parametric and nonparametric components converge with parametric rate towards Normal distributions. 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.

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 address: ruijunbu@liv.ac.uk.

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

[‡]University College London, CeMMaP, and Institute for Fiscal Studies. Correspondence to: Department of Economics, University College London,Gower Street, London WC1E 6BT, United Kingdom. E-mail address: d.kristensen@ucl.ac.uk (D. Kristensen)

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).

In a dynamic setting, Joe (1997) showed how bivariate parametric copulas could be used to model discrete-time Markov chains with flexible, nonparametric stationary marginal distributions. The resulting class of models are semiparametric but are relatively simple 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 do not have a clear interpretation in terms of the dynamics they imply when used to model Markov processes. 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 the theoretical analysis (such as establishing identification) and the practical implementation of such models.

We here propose a novel class of dynamic copulas that resolves these issues: We show how copulas can easily be generated using diffusion processes. The copulas have a clear interpretation in terms of dynamics since they are constructed from an underlying dynamic continuous-time process. Furthermore, conditions for identification of the parameters can be derived despite the fact that the copulas are implicit. Finally, the copulas can easily be computed using existing numerical methods for computing marginal and transition densities of diffusion processes.

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 our attention to the underlying process being a parametric Markov diffusion process which we refer to as underlying parametric diffusion (UPD). Within this class of dynamic copulas, the parametric copula is implied by the diffusion dynamics of the UPD and so has a clear interpretation in terms of dynamic properties. Moreover, a given parametric choice of the drift and diffusion functions for the UPD induces a copula for which we provide conditions for the parameters to be identified. The advantage of our modelling and identification strategy is two-fold: First, the dynamics of the resulting copula model are well-understood with the UPD tying down the dynamics of the observed process, while the nonparametric transformation only affects its marginal 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). Second, our identification results are transparent and provide simple-to-check conditions for the model parameters to be identified from data.

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. As a consequence, 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. This also spills over to computation and estimation with the Pseudo Maximum Likelihood Estimator (PMLE) proposed in Kristensen (2010) being computationally burdensome to implement.

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 and investigate their asymptotic properties. Section 6 presents a simulation study to examine the finite sample performance of our estimators. Some concluding remarks are given in Section 7. All the proofs are collected in Appendices.

2 Copula-Based Diffusion Models

2.1 Model and Its Properties

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

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

where $V : \mathcal{X} \mapsto \mathcal{Y}$ is a smooth monotonic univariate function and $X = \{X_t : t \geq 0\}$ is an underlying parametric diffusion (UPD) defined on $\mathcal{X} = (x_l, x_r)$, with $-\infty \leq x_l < x_r \leq +\infty$, which solves the following parametric SDE:

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

Here, $\mu_X(x; \theta)$ and $\sigma_X^2(x; \theta)$ 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. In most applications, θ is a finite-dimensional parameter but our identification results allows θ to be infinite-dimensional; for example, $\theta = (\mu_X, \sigma_X^2)$ in which case our model is fully nonparametric. Our proposed estimators and their asymptotic theories, however, restrict θ to be finite-dimensional.

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 observed at time points $t_i = i\Delta$, $i = 0, 1, \ldots, n$, 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 do not know V. For convenience, we collect the unknown component in the *structure* $\mathcal{S} \equiv (\theta, V)$.

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

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

Moreover, we shall, as a minimum, require X to be recurrent. To formally impose this property, we introduce

$$s(x; \theta) := \exp\left\{-2 \int_{x^*}^{x} \frac{\mu_X(z; \theta)}{\sigma_X^2(z; \theta)} dz\right\} \text{ and } S(x; \theta) := \int_{x^*}^{x} s(z; \theta)\, dz \tag{2.3}$$

denoting the so-called scale density and scale measure, respectively, where $x^* \in \mathcal{X}$ is some constant. We then impose the following:

Assumption 2.1. For all $\theta \in \Theta$: (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) $\int_{x_l}^{x_r} \sigma_X^{-2}(x; \theta)\, s(x; \theta)\, dx < \infty$.

Assumption 2.2. The transformation function V is strictly increasing with inverse function $U = V^{-1}$, 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) 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

$$f_X(x;\theta) = \frac{\xi(\theta)}{\sigma_X^2(x;\theta)\, s(x;\theta)}, \tag{2.4}$$

where $\xi(\theta)$ is a normalization constant ensuring that $f_X(x;\theta)$ integrates to one. This last part of Assumption 2.1 will not be required for all our results; 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 standard assumption.

Assumption 2.2 requires V to be strictly increasing; this is a testable assumption 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 $p_X(-x|-x_0;\theta) \neq p_X(x|x_0;\tilde{\theta})$ for $\theta \neq \tilde{\theta}$, existing (mis)specification tests (see Section 4.3) can then be used to detect whether V is increasing or not.

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

$$\mu_Y(y;\mathcal{S}) = \frac{\mu_X(U(y);\theta)}{U'(y)} - \frac{1}{2}\sigma_X^2(U(y);\theta)\frac{U''(y)}{U'(y)^3}, \tag{2.5}$$

$$\sigma_Y^2(y;\mathcal{S}) = \frac{\sigma_X^2(U(y);\theta)}{U'(y)^2}, \tag{2.6}$$

where $U'(y)$ and $U''(y)$ are the first two derivatives of $U(y)$. 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. At the same time, once we transform Y by U, we recover the dynamics of the UPD. In particular, the transition density of the discretely sampled process $Y_{i\Delta}$, $i = 0, 1, 2, \ldots$, can be expressed in terms of the one of X as

$$p_Y(y|y_0;\mathcal{S}) = U'(y)\, p_X(U(y)|U(y_0);\theta), \tag{2.7}$$

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

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

and so is also available on closed form. From a modelling and computational perspective, this simple link between the transition dynamics of Y and X is highly useful since the dynamics of parametric diffusion models are well-understood and the computation of parametric transition densities is in general straightforward, even if they are not available on closed form. Formally, we have the following results adopted from Forman and Sørensen (2014):

Theorem 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, so that X is stationary and ergodic, then so is Y.*

2. *The mixing coefficients of X and Y coincide.*

3. *If $E\left[|X_t|^{q_1}\right] < \infty$ and $|V(x)| \leq B\left(1+|x|^{q_2}\right)$ 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 $E\left[\varphi\left(X_1\right)|X_0\right] = \rho\varphi\left(X_0\right)$ 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

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

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 since it only requires the researcher to choose some parametric specification for X. Many parametric diffusion models are available for that purpose and any of these can in principle be employed. We here present three 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

$$dX_t = \kappa\left(\alpha - X_t\right)dt + \sigma dW_t, \tag{2.8}$$

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 $e^{-\kappa\Delta}$.

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

$$dX_t = \kappa\left(\alpha - X_t\right)dt + \sigma\sqrt{X_t}dW_t. \tag{2.9}$$

The process is stationary on $\mathcal{X} = (0, +\infty)$ 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.

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

$$dX_t = \left(\sum_{i=-k}^{l} \alpha_i X_t^i\right)dt + \sigma X_t^{\beta} dW_t. \tag{2.10}$$

It is easily seen that when $\alpha_{-k} > 0$ and $\alpha_l < 0$ the drift term of the diffusion in (2.10) 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 (2.10) is one of the most flexible parametric diffusions, diffusion processes that are unspecified transformations of (2.10) should represent a very flexible class of diffusion models. Similar to (2.9), the implied copula of the NLDCEV is most likely unestablished in 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 (2.4). 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, our modelling framework is 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. Now, in this literature, a given discrete-time, stationary scalar Markov process $Y = \{Y_i : i = 0, 1, \ldots, n\}$ is modelled through a bivariate parametric copula, say, $c_X(u_0, u; \theta)$, together with its stationary marginal cdf F_Y, i.e.,

$$p_Y(y|y_0; \theta, F_Y) = f_Y(y)\, c_X(F_Y(y_0), F_Y(y); \theta), \tag{2.11}$$

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

$$Y_i = F_Y^{-1}(X_i), \quad X_{i+1}|X_i = x_0 \sim c_X(x_0, \cdot; \theta),$$

so that Y_i is a transformation of an underlying Markov process $X_i \in [0,1]$ with a uniform marginal distribution and transition density $c(x_0, x; \theta)$. 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 given parametric specification of our continuous-time UPD and suppose that X is stationary with marginal cdf $F_X(x; \theta)$. In this case, we obtain the following link between the invariant marginal cdf of Y, $F_Y(y)$, and the transformation function can be expressed as

$$F_Y(y) = F_X(U(y); \theta) \Leftrightarrow U(y) = F_X^{-1}(F_Y(y); \theta). \tag{2.12}$$

That is, given knowledge of the parametric diffusion X, we can recover U from the marginal cdf of Y. Substituting (2.12) into (2.7), we see that p_Y can be expressed in the form of (2.11) where $c_X(u_0, u; \theta)$ is the density function of the (dynamic) copula[1] implied by the discretely sampled UPD X,

$$c_X(u_0, u; \theta) = \frac{p_X\left(F_X^{-1}(u; \theta)\,|F_X^{-1}(u_0; \theta); \theta\right)}{f_X\left(F_X^{-1}(u; \theta); \theta\right)}. \tag{2.13}$$

[1]The copula $C_X(u_0, u_1; \theta)$ for the discretely sampled UPD is defined as

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

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

Thus, our semiparametric class of diffusion models falls within the framework of copula-based Markov models.

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.

In contrast, 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, while at the same time we are able to derive precise conditions under which the parameters of the copula are identified from the data. Second, given that our copulas are induced by specifying the drift and diffusion functions of X, the time series dependence structure is much more easily inferred from our model. Third, the resulting model can be interpreted in a continuous-time context and so for asset pricing applications existing results for continuous-time models can be employed. 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

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

From this expression, we see that f_Y affects the resulting dynamics of Y.

3 Identification

Suppose that a particular parametric diffusion X written as in (2.2) (e.g., OU, CIR, NLDCEV) has been chosen. Given a discrete sample of Y, $Y_{i\Delta}$, $i = 0, 1, ..., n$, we then wish to estimate θ together with V. To this end, we first have to show that these are actually identified from data.

Formally, recall that $\mathcal{S} = (\theta, V)$ and let the model consist of all the structures that satisfy, as a minimum, Assumptions 2.1(i)–(ii) and 2.2. According to (2.7), each structure implies a conditional density $p_Y(y|y_0;\mathcal{S})$ of the observables. 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 conditional distribution of the observables, i.e. if for almost every (y, y_0) we have

$$p_Y(y|y_0;\mathcal{S}) = p_Y(y|y_0;\tilde{\mathcal{S}}).$$

The structure $\mathcal{S}$ is said to be identified within the model if $\mathcal{S} \sim \tilde{\mathcal{S}}$ implies $\mathcal{S} = \tilde{\mathcal{S}}$. In terms of the copula induced by the UPD as given in eq. (2.13), $\mathcal{S}$ is identified if $c_X(u_0, u; \theta)$ identifies θ. That is, $c_X(u_0, u; \theta) = c_X(u_0, u; \tilde{\theta})$ for almost every (u_0, u) if and only if $\tilde{\theta} \neq \theta$. However, in our setting, c_X is a complicated functional of p_X and F_X and without suitable *normalizations*, the identification condition in terms of c_X will generally not be satisfied. To see this, let X be the data-generating UPD with transition density $p_X(x|x_0; \theta)$, and suppose that there exists an increasing transformation $T(x)$ so that $\tilde{X} = T(X)$ has transition density $p_X(x|x_0; \tilde{\theta})$ for some $\tilde{\theta}$. Then $\mathcal{S} \sim \tilde{\mathcal{S}}$ where $\tilde{\mathcal{S}} = (\tilde{\theta}, \tilde{V})$ with $\tilde{V}(\tilde{x}) = V\left(T^{-1}(\tilde{x})\right)$. To illustrate this issue, let us return to Example 1:

Example 1 (continued). The UPD X is here characterized by $\theta = (\alpha, \kappa, \sigma)$. By Ito's Lemma, for any $a \in \mathbb{R}$ and $b > 0$, $\tilde{X} = a + bX$ is also an OU process on the form

$$d\tilde{X}_t = b\kappa(\alpha - X_t)\,dt + b\sigma dW_t = \kappa(b\alpha + a - (a + bX_t))\,dt + b\sigma dW_t = \kappa(\tilde{\alpha} - \tilde{X}_t)dt + \tilde{\sigma} dW_t,$$

where $\tilde{\alpha} = b\alpha + a$ and $\tilde{\sigma} = b\sigma$. Thus, $\mathcal{S} \sim \tilde{\mathcal{S}}$ where $\tilde{\mathcal{S}} = (\tilde{\theta}, \tilde{V})$ with $\tilde{\theta} = (\tilde{\alpha}, \kappa, \tilde{\sigma})$ and $\tilde{V}(\tilde{x}) = V((\tilde{x} - a)/b)$, and so we can only identify κ while we have to fix α and σ. This can also be seen in terms of the implied copula of the OU model: c_X is a Gaussian copula density which only depends on the parameter κ.

In the above example, determining which parameters that are identified from the dynamic copula is easy since the joint distribution of (X_Δ, X_0) and their stationary marginal distribution were known on closed form. However, in general, the marginal distributions of stationary Markov chains are not available on closed form. This in turn implies that the dynamic copula is not available on closed form and it is not obvious which parameters are identified. We believe this is a big part of the reason for the fact that in the existing literature on copula-based Markov models in discrete time, copulas have been imported from the existing literature on modelling cross-sectional dependence.

Identification of diffusion copula models is facilitated by three particular features of diffusion processes: First, the stationary distributions of Markov diffusion processes are available on closed form, c.f. eq. (2.4). Second, the process Y is itself a Markov diffusion process with its drift and diffusion terms satisfying eqs. (2.5)-(2.6). Third, under regularity conditions, the drift and diffusion terms of a discretely sampled process are nonparametrically identified (c.f. Hansen et al., 1998). The last property entails that, given observations of Y, we can treat μ_Y and σ_Y^2 as known in our identification argument.

These three properties will in the following be used to develop three different identification schemes: In the first one, we propose to fix the stationary distribution f_X so it is known, and choose a parametric form for either μ_X or σ_X^2. In the second one, we fix the diffusion function σ_X^2 so it is known and choose a parametric form for $\mu_X(x)$. Finally, one can alternatively fix the drift function μ_X to be zero and choose a parametric form for $\sigma_X^2(x)$. In each of the three cases, easy-to-verify conditions for identification can be derived. We analyze each scheme in turn in the following three subsections and show how they can be applied to standard UPDs as presented in

Examples 1-3. In particular, we demonstrate how any UPD, where both the drift and diffusion terms are specified up to unknown parameters, can via transformations be brought onto a form that satisfies each of the three identification schemes. Moreover, the second and third identification schemes work without imposing stationarity since the normalization does not involve any stationary distribution. This is in contrast to the existing literature on copula-based Markov models.

We wish to keep the analysis at a general level and provide formal identification results for general UPDs, not just specific models, thereby allowing for added flexibility in choosing the underlying diffusion model. We will therefore work under the following somewhat abstract identification condition:

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

This assumption is similar in spirit to existing papers on non- and semiparametric identification where it is routinely assumed that certain characteristics of the distribution of the observed process are nonparametrically identified from data. However, the above assumption is not completely innocuous and does impose some additional regularity conditions on the Data Generating Process (DGP). Below, we therefore provide primitive conditions for this assumption to hold.

The first set of primitive conditions can be found in Hansen et al. (1998) who showed that μ_Y and σ_Y^2 are identified from the discretely sampled process if Y is stationary and its infinitesimal operator has a discrete spectrum. Due to Theorem 2.1(4), we know 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:

Theorem 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) hold 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.

3.1 First Identification Scheme

As noted earlier, in the literature on estimation of discrete-time copula-based Markov models, identification is resolved by normalizing the underlying process X so that its stationary marginal distribution is uniform. We here generalize this idea and start out with a chosen marginal density f_X. Given this density, we show how to generate two classes of UPD that are stationary with marginal distribution f_X, in which case identification of the remaining parameters is guaranteed. While this normalization is difficult to impose in a discrete-time setting, it is straightforward to do so when X is a diffusion due to eq. (2.4) that ties together f_X, μ_X and σ_X^2. Specifically, one can choose a known (fixed) density $f_X(x)$ that describes the stationary distribution of X together with a parametric specification for either the drift or the diffusion function. We can then rearrange eq. (2.4) to back out the remaining term:

$$\mu_X(x;\theta) = \frac{1}{2f_X(x)}\frac{\partial}{\partial x}\left[\sigma_X^2(x;\theta) f_X(x)\right], \text{ or} \tag{3.1}$$

$$\sigma_X^2(x;\theta) = \frac{2}{f_X(x)}\int_{x_l}^{x} \mu_X(z;\theta) f_X(z)\,dz. \tag{3.2}$$

In the case where the researcher has a UPD specified in terms of a parametric drift term $\mu_X(x;\theta)$ and diffusion term $\sigma_X^2(x;\theta)$ in mind, one could here choose $f_X(x) = f_X(x;\bar{\theta})$ for some $\bar{\theta} \in \Theta$ chosen by the researcher.

Whether one generates the model through eq. (3.1) or (3.2), the resulting UPD is guaranteed to have stationary density $f_X(x)$ and the parameters entering the parametric diffusion (drift) term are identified under the following weak conditions:

Assumption 3.2 (i) The marginal density $f_X(x)$ is known (and so does not depend on θ) and three times differentiable; (ii) either the drift and the diffusion function of the UPD is specified parametrically while the remaining term satisfies eq. (3.1) or (3.2), respectively, so that Assumption 2.1(i)-(ii) are satisfied; (iii) there exists no $\tilde{\theta} \neq \theta$ such that $\mu_X(x;\tilde{\theta}) = \mu_X(x;\theta)$ or $\sigma_X^2(x;\tilde{\theta}) = \sigma_X^2(x;\theta)$ for all $x \in \mathcal{X}$.

Assumption 3.2 implicitly imposes enough smoothness conditions on f_X so that the implied drift or diffusion term, as given by eq. (3.1) or (3.2), respectively, satisfies the smoothness conditions in Assumption 2.1 and the implied diffusion process X is stationary and ergodic. Moreover, it requires the implied drift and diffusion terms to be parameterized so that θ can be identified from observations of X. The following theorem shows that θ is in fact identified together with V from discretely observed Y:

Theorem 3.2 *Under Assumptions 2.1-2.2 and 3.1-3.2, $\mathcal{S}$ is identified.*

Example 1 (continued). The OU model has discrete spectrum (see Kessler and Sørensen, 1999). If we choose $f_X(x)$ as the standard normal distribution and the drift as $\mu_X(x;\theta) = -\kappa x$, the OU model with $\sigma_X^2(x;\theta) = 2\kappa$ is then obtained from eq. (3.2). Alternatively, with $f_X(x)$ as the standard normal distribution and choosing $\sigma_X^2(x;\theta) = 2\kappa$ leads to $\mu_X(x;\theta) = -\kappa x$ as the solution to eq. (3.1). In both cases $\theta = \kappa$ is identified together with the transformation function V by Theorem 3.2.

Example 2 (continued). The CIR model has discrete spectrum (see Kessler and Sørensen, 1999). Now, choosing $f_X(x)$ as the standard Gamma distribution with a fixed shape parameter α and choosing $\sigma_X^2(x;\theta) = 2\kappa x$ leads to $\mu_X(x;\theta) = \kappa(\alpha - x)$ as the solution to eq. (3.1). Reversely, we can first choose $\mu_X(x;\theta) = \kappa(\alpha - x)$ then eq. (3.2) delivers $\sigma_X^2(x;\theta) = 2\kappa x$. In either case, Assumption 3.2 is satisfied and identification of $\theta = (\kappa, \alpha)$ and V is ensured.

In Example 1, with f_X chosen as the standard normal, it is important to note that $-\kappa x$ is the only choice of $\mu_X(x;\theta)$ such that the diffusion X with $\sigma_X^2(x;\theta)$ satisfying eq. (3.2) is a OU process. For any drift other than $-\kappa x$, the resulting diffusion function will not be constant and so X will not be an OU process. But X still has the standard normal marginal distribution by construction and the identification result remains true as long as the chosen drift $\mu_X(x;\theta)$ satisfies Assumption 3.2(iii).

3.2 Second identification scheme

Our second identification strategy takes as starting point a fixed (and so parameter independent) diffusion function, $\sigma_X^2(x)$, together with a parametric drift function, $\mu_X(x;\theta)$. Again, if the researcher initially has a parametric UPD in mind where both the drift and diffusion terms are parametric, the second identification scheme would seem to require that all parameters in the diffusion term are fixed. However, this can be circumvented in the following way:

Let $\mu_X(x;\theta)$ and $\sigma_X^2(x;\theta)$ be a given specification of the UPD X: We then apply the so-called Lamperti transform of univariate diffusion processes to X,

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

for some $x^* \in \mathcal{X}$. Here, $\gamma(x;\theta)$ is the Lamperti transform. The resulting process is a unit diffusion process,

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

where

$$\mu_{\bar{X}}(\bar{x};\theta) = \frac{\mu_X\left(\gamma^{-1}(\bar{x};\theta);\theta\right)}{\sigma_X\left(\gamma^{-1}(\bar{x};\theta);\theta\right)} - \frac{1}{2}\frac{\partial\sigma_X}{\partial x}\left(\gamma^{-1}(\bar{x};\theta);\theta\right). \tag{3.3}$$

In particular, after transformation, any given UPD where both the drift and diffusion terms are parameterized, can be written on the form of the second identification scheme. Also note that since $\bar{X}$ is a invertible transformation of X, the dynamic properties of the two processes are identical. Thus, working with $\bar{X}$ or X as UPD will not make a difference in our modelling framework.

For a given UPD with parametric drift and known diffusion term, we then impose either of the following two restrictions:

Assumption 3.3. Either of the following hold: (i) The diffusion term $\sigma_X^2(x)$ is fixed (known) and there exists no $\tilde{\theta} \neq \theta$ such that $\mu_X(x;\tilde{\theta}) = \mu_X(x;\theta)$ for all $x \in \mathcal{X}$, and for some known $x_0 \in \mathcal{X}$ and $y_0 \in \mathcal{Y}$, $U(y_0) = x_0$; or (ii) there exists no $\eta \neq 0$ and $\tilde{\theta} \neq \theta$ such that $\mu_{\bar{X}}(x;\tilde{\theta}) = \mu_{\bar{X}}(x+\eta;\theta)$ for all $x \in \bar{\mathcal{X}}$.

Assumption 3.3 contains two alternative identification conditions: Condition (i) imposes a normalization condition on $U(y)$ and then requires the drift function to be specified in such a way that two different parameter values do not lead to the same drift function. The second identification condition imposes the normalization and identification condition on the transformed drift function. When verifying Assumption 3.3(ii) for the transformed unit diffusion $\bar{X}$ defined above, it will generally require us to fix some of the parameters that enter $\mu_X(x;\theta)$ and $\sigma_X^2(x;\theta)$ of the original process X, c.f. examples below.

Theorem 3.3 *Under Assumptions 2.1-2.2, 3.1 and 3.3, $\mathcal{S}$ is identified.*

Example 1 (continued). The Lamperti transform of the OU process in (2.8) 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(ii) to be satisfied; one such is $\alpha/\sigma = 0$ leading to the following identified model,

$$d\bar{X}_t = -\kappa\bar{X}_t dt + dW_t. \tag{3.4}$$

Example 2 (continued). The Lamperti transform of the CIR diffusion in (2.9) is given by

$$d\bar{X}_t = \left[\kappa\left(\frac{2}{\bar{X}_t}\frac{\alpha}{\sigma^2} - \frac{\bar{X}_t}{2}\right) - \frac{1}{2\bar{X}_t}\right]dt + dW_t, \tag{3.5}$$

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 (2.10) takes the form

$$d\bar{X}_t = \left[\sum_{i=-k}^{l} \alpha_i^* \bar{X}_t^{\frac{i-\beta}{1-\beta}} - \frac{\beta}{2(1-\beta)} \bar{X}_t^{-1} \right] dt + dW_t, \tag{3.6}$$

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 (2.8) and (2.9) are special cases of (2.10), both (3.4) and (3.5) are special cases of (3.6) with suitable parameter restrictions.

3.3 Third identification scheme

Our final identification strategy takes as input a parameter independent drift function, $\mu_X(x)$, together with a parametric diffusion function, $\sigma_X^2(x;\theta)$. Similar to our second scheme, if the initial UPD involves parametric specifications of both components, we can arrive at this situation through transformation: For given $\mu_X(x;\theta)$ and $\sigma_X^2(x;\theta)$ we transform X by its scale measure, $\bar{X}_t := S(X_t;\theta)$, 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

$$\sigma_{\bar{X}}^2(\bar{x};\theta) = s^2\left(S^{-1}(\bar{x};\theta);\theta\right)\sigma^2\left(S^{-1}(\bar{x};\theta);\theta\right). \tag{3.7}$$

We impose the following standard identifying assumption on the UPD (where in case of the model first having been scale-transformed, the assumption should be applied to the diffusion term given in eq. (3.7)):

Assumption 3.4. The drift term $\mu_X(x) = 0$ and one of the two following conditions hold: Either (i) there exists no $\tilde{\theta} \neq \theta$ such that $\sigma_X^2(x;\tilde{\theta}) = \sigma_X^2(x;\theta)$ for all $x \in \mathcal{X}$, while $U(y_0) = x_0$ and $U'(y_1) = x_1$ for known $y_0, y_1 \in \mathcal{Y}$ and $x_0, x_1 \in \mathcal{X}$; or (ii) there exists no $\eta_1 \neq 1$, $\eta_2 \neq 0$ and $\tilde{\theta} \neq \theta$ such that $\sigma_X^2(x;\tilde{\theta}) = \sigma_X^2(\eta_1 x + \eta_2;\theta)/\eta_1^2$ for all $x \in \mathcal{X}$.

In comparison to Assumption 3.3, we note two important differences: First, while Assumption 3.3 only restricts the diffusion term to be fixed (but potentially non-constant), we here restrict μ_X to be zero. Fortunately, as pointed out above, any UPD with non-zero drift can be easily transformed into a diffusion with zero drift and so this issue is minor. Second, note that 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 normalization of the process. That is, any location-scale transformation of X still leads to a zero-drift. Therefore, for the third scheme to work we need both a scale and location normalization.

Theorem 3.4 *Under Assumptions 2.1-2.2, 3.1 and 3.4, $\mathcal{S}$ is identified.*

4 Statistical Inference

4.1 Estimation

Suppose that we have specified a UPD X so that the parameter vector θ is identified by, for example, using any of the three identification schemes presented in the previous section. We then propose two alternative semiparametric estimators for θ. The first takes the form of a two-step Pseudo Maximum Likelihood Estimator (PMLE). The second is a semiparametric sieve-based ML estimator (SMLE), where θ and V are jointly estimated. To motivate the two estimators, suppose that U is known, in which case the MLE of θ is given by

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

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

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

If U is unknown, the above estimator is not feasible and we instead have to estimate it together with θ.

The PMLE relies on Y being stationary in which case U satisfies $U\left(y\right) = F_X^{-1}\left(F_Y\left(y\right); \theta\right)$. The unknown cdf F_Y can be estimated by the empirical cdf defined as

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

where $\mathbb{I}\left\{\cdot\right\}$ denotes the indicator function. One could alternatively use the following kernel smoothed empirical cdf,

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

where $\mathcal{K}_h\left(y\right) = \mathcal{K}\left(y/h\right)$ with $\mathcal{K}\left(y\right) = \int_{-\infty}^{y} K\left(z\right) dz$, K being a kernel (e.g., the standard normal density), and $h > 0$ being a bandwidth going to zero at a certain rate as $n \to \infty$. Replacing F_Y in the expression of U with either $\tilde{F}_Y$ or $\hat{F}_Y$, we obtain the following two estimators of U,

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

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 $\hat{U}(y;\theta)$ 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

$$U'(y) = \frac{f_Y(y)}{f_X(U(y);\theta)}, \tag{4.2}$$

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,

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

where Θ is the parameter space and

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

is an adjusted version of the log-likelihood function $L_n(\theta, U)$ where we have subtracted the parameter-invariant term $\sum_{i=1}^{n} \log f_Y(Y_{i\Delta})$. One can easily check, by rewriting the above in terms of the implied copula of X, that this estimator is equivalent to the one analyzed in Chen and Fan (2006). For added transparency, we here, however, maintain the above expression since in our framework the model is formulated in terms of the transition density of X and the transformation U.

The second estimator replaces the inverse of the unknown transformation function, $U(y)$ by a sieve approximation $U_m(y) \in \mathcal{U}_m$ where $\mathcal{U}_m$ is a finite-dimensional function space reflecting the properties of U, $m = 1, 2, ...$ We here require, as a minimum, that $\mathcal{U}_m$ is restricted to differentiable and strictly increasing functions. For a given choice of $\mathcal{U}_m$, we obtain the following semiparametric sieve maximum-likelihood estimator,

$$(\hat{\theta}_{\mathrm{SMLE}}, \hat{U}_m) = \arg\max_{\theta\in\Theta, U_m\in\mathcal{U}_m} L_n(\theta, U_m).$$

The above SMLE shares similarities with the one proposed by Chen, Wu and Yi (2009) for the estimation of copula-based Markov models. However, while they estimate θ and F_Y, we here estimate θ and U. Under stationarity, there is a one-to-one correspondence between U and F_Y, but note that this involves evaluating $F_X^{-1}(x;\theta)$ which in general is not available on closed form in our setting: The cdf for general diffusion models is itself not available on closed-form and its evaluation involves numerical integration or Monte Carlo methods. This in turn also implies that its inverse has to be computed using numerical methods. By reparameterizing the model in terms of U, we avoid these issues. Moreover, the above SMLE does not require Y to be stationary and so broadens the scope for modelling and estimation of copula-type dynamic models. On the other hand, again in comparison with the PMLE, the numerical implementation of this SMLE involves joint maximization over both θ and U_M, which is a harder numerical problem. In terms of statistical efficiency, $\hat{\theta}_{\mathrm{SMLE}}$ will in general reach the semiparametric efficiency bound under stationarity, while the PMLE is inefficient.

Once an estimator for θ has been obtained, we can estimate the drift and diffusion terms of Y using the expressions given in (2.5) and (2.6) 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 $\mathcal{U}_m$ is restricted to twice continuously differentiable functions. For the PMLE, estimating these derivatives take some more work: The first derivative is given in eq. (4.2) while the second one is given by

$$U''(y) = \frac{f_Y'(y)}{f_X(U(y);\theta)} - \frac{f_X'(U(y);\theta)\, f_Y(y)^2}{f_X(U(y);\theta)^3}, \tag{4.3}$$

where $f_X'(x;\theta)$ and $f_Y'(y)$ are the first derivatives of $f_X(x;\theta)$ and $f_Y(y)$, respectively. Since $\tilde{F}_Y(y)$ is not differentiable, we replace this with $\hat{F}_Y(y)$, leading to the following three-step estimators of the drift and diffusion functions

$$\hat{\mu}_Y(y) = \frac{\mu_X(\hat{U}(y);\hat{\theta}_{\text{PMLE}})}{\hat{U}'(y)} - \frac{1}{2}\sigma_X^2(\hat{U}(y);\hat{\theta}_{\text{PMLE}})\frac{\hat{U}''(y)}{\hat{U}'(y)^3}, \tag{4.4}$$

$$\hat{\sigma}_Y^2(y) = \frac{\sigma_X^2(\hat{U}(y);\hat{\theta}_{\text{PMLE}})}{\hat{U}'(y)^2}, \tag{4.5}$$

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

4.2 Asymptotic Theory

4.2.1 Estimation of Parametric Component

We here establish an asymptotic theory for the proposed estimators. 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 this assumption to hold. Instead of referring to these different sets of conditions, we simply maintain Assumption 4.1 to save space in the following. One implication of Assumption 4.1 is that $p_Y(y|y_0;\mathcal{S}) \neq p_Y(y|y_0;\mathcal{S}_0)$ which in turn implies that $E[\log p_Y(Y_\Delta|Y_0;\mathcal{S})] < E[\log p_Y(Y_\Delta|Y_0;\mathcal{S}_0)]$ 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

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

(ii) With $s(x;\theta)$ and $S(x;\theta)$ defined in (2.3),

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

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

We here abstain the precise 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. One can show that the OU and the CIR models, for example, satisfy these conditions. The following result now follows from the general theory of Chen and Fan (2006) and Chen, Wu and Yi (2009), respectively:

Theorem 4.1 *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 the second moment of the efficient score function for ψ defined in Chen, Wu and Yi (2009).

4.2.2 Estimation of Drift and Diffusion Functions

Once the transformation function V has been estimated together with the UPD parameters, we may estimate the drift and diffusion terms of Y, μ_Y and σ_Y^2. These estimates provide the researchers with a better understanding of the dynamic properties of Y such as the level of mean-reversion and volatility. They can furthermore be used for specification testing along the lines of Kristensen (2011). We here focus on the kernel-based estimators of μ_Y and σ_Y^2 given in eqs. (4.4)-(4.5) and impose 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 our functional estimators of μ_Y and σ_Y^2 are well defined. Assumption 4.5 ensures that the estimator of the UPD parameter converges to the truth sufficiently fast, so that the asymptotic properties of $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ are determined by the properties of the kernel estimators. Clearly, this implies that our asymptotic results will be applicable to not only the PMLE and the SMLE derived above but also to any other $\sqrt{n}$-consistent estimators. Assumption 4.6 regulates the kernel functions and allow for most standard kernels such as the Gaussian and the Uniform kernels. We are now able to state pointwise convergence results for our kernel-based estimators of μ_Y and σ_Y^2 using the standard functional delta-method for kernel estimators.

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

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

where

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

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

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

where

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

The above theorem is based on the standard results for kernel density estimators as found in Robinson (1983) which requires Y to be stationary and strongly mixing with a sufficiently smooth marginal density f_Y. Assumption 2.1-2.2 and 4.2(i) guarantee that these are satisfied.

4.3 Specification Testing

Our semiparametric diffusion model relaxes some restrictions imposed by a fully parametric diffusion model, but it still imposes parametric restrictions on the UPD or equivalently on the dynamic copula implied by the UPD. From an empirical modelling point of view, it would be desirable to be able to test the semiparametric specification against the nonparametric alternative using discrete sample. Assuming that Y is Markovian and we are interested in testing whether its dynamics is described by a given UPD, we are effectively jointly testing the assumption that (i) Y is a diffusion process (which can be pretested using the tests developed in Florens et al., 1998 and Kanaya, 2008) and (ii) the UPD X in terms of its implied copula is correctly specified. Two classes of semiparametric specification tests may be considered for our model:

The first class of tests are often referred to as the transition-based specification tests. This class of tests compare the distance between the semiparametric estimator of the transition density or transition distribution under the null with the corresponding fully nonparametric estimator under the alternative. Typically, either the Kolmogorov-Smirnov type or Cramér-Von Mises test is entertained. The test will reject the null hypothesis if the sample statistic is too large compared to the critical value implied by the limiting distribution of the test statistic. Transition-based specification tests include Hong and Li (2005), Aït-Sahalia et al. (2009), Kristensen (2010, 2011), etc. Interested readers are referred to these papers for more detailed discussions on this subject.

The second class of tests focus more specifically on the underlying parametric copula of the semiparametric model. Under the maintained assumption that the process is Markovian and its marginal distribution is estimated nonparametrically, the main task is to test whether the parametric copula function is correctly specified. Consequently, our semiparametric specification test is closely related to the goodness-of-fit tests for parametric copulas. This type of tests usually examine the distance between the estimator of the parametric copula function and the nonparametric empirical copula for a given sample. Similar to the transition-based tests, both the sup-type and the L_2-type distance measures are often considered. However, since the majority of such test statistics are non-pivotal, their critical values typically need to be simulated from computationally intensive bootstrap procedures. Interested readers can refer to Fermanian (2005) and Genest and Rémillard (2008) for excellent surveys on existing literature.

5 Triangular Models

We here extend the modelling framework to a multivariate setting with a triangular structure. For simplicity, we here focus on the bivariate case, but all arguments can be generalized.

Consider a bivariate process (S_t, Y_t) where the first component S_t, can be thought of as the log-price of an asset and Y_t as the stochastic volatility process. Suppose that the first component solves a fully parametric SDE,

$$dS_t = \mu_S (S_t, Y_t; \theta_S) \, dt + \sigma_S (S_t, Y_t; \theta_S) \, dW_{S,t}$$

where the drift and diffusion terms potentially depend on both components. The second component, Y_t, is modelled semiparametrically on the form of eqs. (2.1)-(2.2). The two Brownian motions, $W_{S,t}$ and W_t, may be correlated, thereby capturing leverage effects, and let $\rho = cov \, (dW_{S,t}, dW_t)$ denote the correlation coefficient. A particular example falling within the above general framework is the following:

$$dS_t = \mu dt + \sqrt{Y_t} dW_{S,t},$$

where Y_t is the stochastic volatility process assumed to follow our semiparametric diffusion process. As such, the bivariate model depends on (μ, ρ, θ, V), where (θ, V) is the structure of the semiparametric diffusion model defined in the Section 2. We here follow Aït-Sahalia and Kimmel (2007) and assume that a proxy for the volatility process is observed.

Given a sample of (S_t, Y_t) at discrete time points, identification of this bivariate model follows straightforwardly from our univariate results: First, given the triangular structure, we can identify the parameters entering the UPD for Y and the transformation function using only data of Y. Once the nonparametric component has been identified, we can rely on existing results for identification of fully parametric diffusion models to identify the remaining parameters entering the specification of S.

Estimation of the model can be done by the following simple extension of the semiparametric estimators developed in the univariate case: First, estimate V and θ appearing in eqs. (2.1)-(2.2) using only the observations of Y; this can be done through either the PMLE or the sieve MLE. This provides us with a $\sqrt{n}$-consistent estimator of V. Given V and θ (or an estimator of these), the bivariate model is fully parametric and all remaining parameters, μ and ρ, can be estimated using standard parametric methods such as the MLE of Aït-Sahalia and Kimmel (2007). By the same arguments as in Chen and Fan (2006), it follows that the estimator of μ and ρ will be $\sqrt{n}$-asymptotically normally distributed under great generality.

6 Monte Carlo Simulations

In this section, we investigate the finite sample performance of our semiparametric pseudo ML inference procedure in comparison with a similar multi-stage but fully parametric ML procedure by Monte Carlo simulations. The details of the latter procedure are explained in Section 6.2.

6.1 Data Generating Processes

In this simulation study, we consider two UPDs:

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

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

Both have closed-form transition density and also closed-form marginal density thereby facilitating the numerical implementation. Note that both processes have been normalized by letting $\sigma^2 = 2\kappa$. This particular normalization has the advantage that the marginal distributions of X in (6.1) and (6.2) are now invariant to the mean-reversion parameter κ. Hence, by varying the value of κ alone, 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 semiparametric estimators of θ, μ_Y and σ_Y^2 in relation to the fully parametric estimators.

Next, we specify the transformation of the DGP of Y. We do this by specifying a parametric marginal distribution with cdf $F_Y(y;\phi)$ which induces the transformation V as

$$Y_t = V\left(X_t; \phi \right) = F_Y^{-1}\left(F_X\left(X_t; \theta \right); \phi \right). \tag{6.3}$$

Given $F_Y(y;\phi)$, its corresponding pdf $f_Y(y;\phi)$, and $p_X(x|x_0;\theta)$, the transition density of the true DGP of Y then takes the form

$$p_Y\left(y|y_0; \theta, \phi \right) = f_Y\left(y; \phi \right) c_X\left(F_Y\left(y_0; \phi \right), F_Y\left(y; \phi \right); \theta \right). \tag{6.4}$$

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). The location-scale version of the SKST distribution Y has the following density function:

$$f_Y\left(y; \phi \right) = \begin{cases} \dfrac{bq}{v}\left(1 + \dfrac{1}{\tau - 2}\left(\dfrac{\frac{b}{v}\left(y - m \right) + a}{1 - \lambda} \right)^2 \right)^{-(\tau+1)/2} & \text{if} \quad y < m - av/b, \\ \dfrac{bq}{v}\left(1 + \dfrac{1}{\tau - 2}\left(\dfrac{\frac{b}{v}\left(y - m \right) + a}{1 + \lambda} \right)^2 \right)^{-(\tau+1)/2} & \text{if} \quad y \geq m - av/b, \end{cases} \tag{6.5}$$

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

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

Note that the location-scale SKST has four parameters, i.e., $\phi \in (m, v, \lambda, \tau)$ 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 in (6.5) and normalized UPD in (6.1) or (6.2) via (6.3) is referred to as the OU-SKST or the CIR-SKST diffusion model correspondingly.

6.2 Calibration and Estimation

The true data-generating parameters ϕ and θ are chosen to match the estimates obtained when the model is fitted to the classic data set of 5505 observations of daily ($\Delta = 1/252$) observations of 7-day Eurodollar rate used in Aït-Sahalia (1996b). The calibration is based on a 2-Stage MLE (2SMLE). In the first stage, the SKST distribution is fitted to the data (as if they are i.i.d) to obtain the MLE $\hat{\phi}$. We then plug the corresponding MLEs of $F_Y(y; \hat{\phi})$ and $f_Y(y; \hat{\phi})$ into (6.4) which is then maximized with respect to θ to obtain $\hat{\theta}$. The calibrated parameter values of the marginal SKST distribution are $\left(\hat{m}, \hat{v}, \hat{\lambda}, \hat{\tau}\right) = (0.0835, 0.0358, 0.5193, 25.3708)$, and those of the underlying OU and CIR diffusions are $\hat{\kappa} = 1.1376$ and $(\hat{\kappa}, \hat{\alpha}) = (0.7653, 1.1653)$, respectively.

Artificial samples of sizes 2202 and 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. We therefore 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 it creates data series with 1st-order autocorrelation coefficient ρ_1 equal to 0.9 approximately, which is a reasonably high persistent level without being excessively close to the unit root. Finally, 500 replications for each case are generated. The values of the true data-generating parameters of the underlying diffusion X and the approximate value of ρ_1 for each scenario can be found in Tables 1-3.

Two estimation methods are investigated in this simulation study. The first estimation method is the 2SMLE which is simply the method used for our calibration. This fully parametric estimator is included mainly as a benchmark to be compared with our semiparametric PMLE. Note that the only difference between the 2-stage ML method and our pseudo ML method is that the former estimates the marginal distribution F_Y by the parametric ML, while the latter estimates it by the nonparametric empirical cdf. The marginal density and its first derivatives, which are the two components for estimating the drift and diffusion functions, are all estimated by nonparametric kernel estimators using the Gaussian kernel, and the bandwidth is obtained by Silverman's Rule-of-thumb (c.f. Silverman, 1986).

The calibrated SKST distribution, the benchmark Normal distribution and the kernel estimator of the marginal distribution are plotted in Figure 1. The nonparametric kernel estimator does not suffer from misspecification, so we expect the kernel density to reflect the most important features of our data. We observe that this is indeed the case. On one hand, we note that the fitted kernel-based density (dashed line) is heavily skewed to the right compared to the fitted Normal distribution (dotted line). On the other hand, we observe that the kernel density has a right tail much thicker and a left tail substantially thinner than those of the fitted Normal distribution. In this regard, the SKST specification (solid line) does, as we expect, reflect the most vital distributional features and thus our choice is suitably justified.

[Figure 1]

6.3 Estimation Results

6.3.1 Parametric Component

The results for the OU-SKST case are presented in Table 1. The true value of κ calibrated from Aït-Sahalia (1996b) data is 1.1376. The persistence level implied by the OU-SKST process is very high, with an average ρ_1 (over our 500 replications) approximately equal to 0.9944. Three additional scenarios are generated by multiplying κ by factors of 5, 10 and 20, generating three decreasing persistent levels.

[Table 1]

The scaled bias and the scaled RMSE (defined as the ratios of the actual bias and the actual RMSE over the true parameter value, respectively) are reported in order to examine the performance of the two estimation methods. Overall, the results from the two estimation methods are generally comparable with the same magnitudes. Specifically, at sample size 5505 and in terms of the scaled bias, the fully parametric 2SMLE outperforms the PMLE in cases with $\kappa = 1.1376$, 5.6882 and 11.377, but not when $\kappa = 22.753$. This may suggest that when sample sizes are relatively large and the DGP is highly persistent, the fully parametric 2SMLE (benefiting from parametric structure) outperforms our semiparametric PMLE. However, when the DGP becomes more stationary (a situation which our model is designed for), our PMLE overtakes the 2SMLE slightly. This is not too surprising, since it is well understood that nonparametric density estimators do not perform well in small and moderate samples when data are persistent (see, for example, Pritsker, 1998). In terms of the scaled RMSE, basically a similar story takes place. Although the 2SMLE is better than the PMLE even when $\kappa = 22.753$, the discrepancy between the two methods becomes negligible when persistence reduces. At sample size 2202, the results are similar in terms of both the bias and the RMSE. However, we note that as sample size decreases, the difference between the two methods also becomes smaller. 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.

[Table 2 and 3]

Overall, the performance of the PMLE is comparable with the fully parametric 2SMLE with very similar estimation errors. Moreover, the gap in the performance of the PMLE relative to the 2SMLE appears to narrow when we move in the following two directions: (i) when the true DGP gets less persistent; and (2) when sample size gets smaller[2].

6.3.2 Semiparametric Drift and Diffusion Functions

We now investigate the performance of the semiparametric functional estimators of the drift and diffusion terms of our data generating diffusion Y. To examine the quality of these functional estimators, we plot the mean and pointwise 95% confidence bands from the 500 estimates against the truth.

Figure 2 plots the estimated drift and diffusion functions from the 2-stage parametric method and our 2-stage semiparametric method where the true DGP is the OU-SKST process with $\kappa = 22.753$ and sample size 2202. First of all, it is important to point out that, as we can see, although the underlying OU process has a linear drift and a constant diffusion function, the transformed process does exhibit strong nonlinearities in both terms. In particular, to a large extent such nonlinearities closely resemble the nonlinearities depicted in, for example, Aït-Sahalia (1996b), Jiang and Knight (1997), and Stanton (1997). Secondly, the mean estimates from both estimation methods are fairly close to the truth, but the variability of the semiparametric estimator is noticeably larger than the parametric method, especially on the right side of the function. However, such difference in the variability is not surprising for the following two reasons. Firstly, as shown in Section 4.2.2, the drift function converges at a slow rate $\sqrt{nh^3}$ due to its dependence on the nonparametric estimator of the first derivative of the marginal density. Secondly, from Figure 1 we can see that the marginal distribution has a long right tail, which implies that the support extends far into the right, but observations of large realizations are relatively few. As a result, the right tail of the distribution is very difficult to estimate by the kernel estimator. The situation for the semiparametric diffusion estimator is better since it converges at a faster rate $\sqrt{nh}$, although the fully parametric method is still slightly better at this sample size. 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. However, it is very important to point

[2] Simulation results based on sample sizes 1101 and 11010 (not reported for space of economy) deliver similar qualitative conclusions.

out that although both the OU and CIR processes have linear drifts, because of their distinctive diffusion terms, the drift and diffusion terms of the transformed OU and those of the transformed CIR processes are quite different. In particular, the drift of the transformed CIR process has a much stronger nonlinear downward pull on the right end and the diffusion exhibits higher growth rate. These are exactly the featured nonlinearities found by, for example, Aït-Sahalia (1996b), Jiang and Knight (1997), and Stanton (1997). This suggests that our semiparametric transformed diffusion model can do a very good job in creating the documented nonlinearities while at the same time being analytically and numerically tractable. This is evidence that our semiparametric model may be an appealing alternative to the less tractable fully nonparametric models and the less flexible fully parametric models.

7 Conclusion

We develop a copula-based semiparametric approach for modelling stationary nonlinear univariate diffusions. We show that our model specifications potentially encompass very general parametric stationary diffusions as well as their time-invariant transformations. 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. Potential future work under this framework may include extensions to multivariate diffusions or even non-Markovian stochastic processes.

8 Acknowledgments

Kristensen acknowledges support from the European Research Council (starting grant no. ERC-2012-StG 312474), the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, and CREATES, which is funded by the Danish National Research Foundation (DNRF78).

References

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.

Aït-Sahalia, Y., Fan, J., Peng, H., 2009. Nonparametric transition-based tests for jump-diffusions. Journal of the American Statistical Association 104, 1102-1116.

Aït-Sahalia, Y., R. Kimmel., 2007. Maximum likelihood estimation of stochastic volatility models. Journal of Financial Economics 83, 413-452.

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, 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.

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

Fermanian, J.D., 2005. Goodness-of-fit tests for copulas. Journal of Multivariate Analysis 95, 119-152.

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

Florens, J.P., Renault, E., Touzi, N., 1998. Testing embeddability by stationary reversible continuous-time Markov processes. Econometric Theory 14, 744-769.

Genest, C., Rémillard, B., 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44, 1096-1127.

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.

Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with application to spot interest rates. Review of Financial Studies 18, 37-84.

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.

Kanaya, S., 2008. Non-parametric specification testing for continuous-time Markov processes: Do the processes follow diffusions? Manuscript. Oxford University.

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

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.

Pritsker, M., 1998. Nonparametric density estimation of tests of continuous time interest rate models. Review of Financial Studies 11, 449-487.

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. $F_Y(y)$ is identified from data. This together with knowledge of $F_X(x)$, as chosen by us, allows us to identify $U(y)$ from eq. (2.12). We can therefore treat $X_{i\Delta} = U(Y_{i\Delta})$, $i = 1, 2, ...$, as directly observed and so μ_X and σ_X^2 are identified due to Assumption 3.1. Assumption 3.2 then ensures that θ is identified. ■

Proof of Theorem 3.3. First, recall that μ_Y and σ_Y^2 are identified from the discrete sampled Y under Assumption 3.1 and so can be treated as known. Next, observe that by Ito's Lemma,

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

$$\sigma_Y^2(y) = \frac{\sigma_X^2(U(y);\theta)}{U'(y)^2}. \tag{A.2}$$

We now establish identification under Assumption 3.3(i): From (A.2), we obtain the following separable Ordinary Differential Equation (ODE):

$$\frac{\sigma_X(U(y))}{U'(y)} = \sigma_Y(y),$$

with solution

$$S_X(U(y)) = S_Y(y) + k \Leftrightarrow U(y) = S_X^{-1}(S_Y(y) + k),$$

for some constant $k \in \mathbb{R}$, where $S_X(x) = \int 1/\sigma_X(x)\,dx$ and $S_Y(y) = \int 1/\sigma_Y(y)\,dy$; note here that $S_X(x)$ is invertible since $\sigma_X(x) > 0$. The normalization imposed on U fixes k,

$$x_0 = S_X^{-1}(S_Y(y_0) + k) \Longleftrightarrow k = S_X(x_0) - S_Y(y_0).$$

This in turn implies that $U(y)$ is identified. Next, observe that

$$\mu_X(U(y);\theta) = U'(y)\left\{\mu_Y(y) + \frac{1}{2}\sigma_X^2(U(y))\frac{U''(y)}{U'(y)^3}\right\},$$

where the right-hand side is known. This together with the assumption that $\mu_X(x;\theta) \neq \mu_X(x;\tilde{\theta})$ for any $\theta \neq \tilde{\theta}$ imply that θ is identified.

Next, we establish identification under Assumption 3.3(ii): Since $\sigma_{\bar{X}} = 1$, we have $\sigma_Y^2(y) = U'(y)^{-2}$. Hence, given $\sigma_Y^2(y)$ we identify $U'(y)$ and so also $U''(y)$. Meanwhile, $U'(y)$ identifies $U(y)$ up to some constant η, i.e., $U(y) = U_0(y) + \eta$, where U_0 is known. Moreover, given $\mu_Y(y)$, $U'(y)$, $U''(y)$ and $U_0(y)$, $\mu_Y(y) = \mu_{\bar{X}}(U(y);\theta)/U'(y) + \frac{1}{2}U''(y)U'(y)$ identifies $\mu_{\bar{X}}(U(y);\theta) = \mu_{\bar{X}}(U_0(y) + \eta;\theta)$. The normalization imposed on $\mu_{\bar{X}}$ together with the requirement that θ is uniquely identified from $\mu_{\bar{X}}$ ensure identification of η and θ, and thereby $U(y)$. ■

Proof of Theorem 3.4. By the same arguments used to obtain eqs. (A.1)-(A.2),

$$\mu_Y(y) = \frac{\mu_X(U(y))}{U'(y)} - \frac{1}{2}\sigma_Y^2(y)\frac{U''(y)}{U'(y)} = \frac{\mu_X(U(y)) - \frac{1}{2}\sigma_Y^2(y)U''(y)}{U'(y)}$$

which can be rewritten as

$$\mu_Y(y)\,U'(y) + \frac{1}{2}\sigma_Y^2(y)\,U''(y) = \mu_X(U(y)).$$

Since $\mu_X(x) = 0$, we obtain

$$-2\frac{\mu_Y(y)}{\sigma_Y^2(y)} = \frac{U''(y)}{U'(y)} = \frac{\partial \log U'(y)}{\partial y}.$$

That is, $\log U'(y) = k_1 - 2\int_{y_1}^{y} \frac{\mu_Y(z)}{\sigma_Y^2(z)}dz$ and so $U'(y) = \exp\left(k_1 - 2\int_{y_1}^{y} \frac{\mu_Y(z)}{\sigma_Y^2(z)}dz\right)$ which in turn implies

$$U(y) = k_0 + \int_{y_0}^{y} \exp\left(k_1 - 2\int_{y_1}^{v} \frac{\mu_Y(z)}{\sigma_Y^2(z)}dz\right)dv.$$

The two normalizations, $U'(y_1) = x_1$ and $U(y_0) = x_0$, imply that $k_1 = \log x_1$ and $k_0 = x_0$. In particular, $U(y)$ is identified.

Next, we establish identification under Assumption 3.4(ii): From the above characterization of U, we easily see that, for some constant η_1,

$$U'(y) = \exp\left[-2\int^{y} \frac{\mu_Y(z)}{\sigma_Y^2(z)}dz\right]\eta_1 = U_0'(y)\,\eta_1,$$

where $U_0'(y)$ is known. This in turn implies that $U(y) = \eta_1 U_0(y) + \eta_2$ for some other constant η_2. Then, we note that

$$\begin{aligned} \sigma_X^2(U(y);\theta) &= \sigma_Y^2(y)\,U'(y)^2 = \sigma_Y^2(y)\,\eta_1^2 U_0'(y)^2 \\ \frac{\sigma_X^2(\eta_1 U_0(y) + \eta_2;\theta)}{\eta_1^2} &= \sigma_Y^2(y)\,U_0'(y)^2 \end{aligned}$$

The right hand side is known and so θ is identified under the normalization in Assumption 3.4(ii). ■

Proof of Theorem 4.1. 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. (2.13). 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 has absolutely continuous marginal distribution $F_X(x;\theta)$, marginal density $f_X(x;\theta)$ and transition density $p_X(x|x_0;\theta)$ with respect to the Lebesgue measure. Hence, the copula density $c_X(u_0, u;\theta)$ in (2.13) 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), except that the conditional density of Y in our case takes the form $p_Y(y|y_0;\theta,U) = U'(y)\,p_X(U(y)|U(y_0);\theta)$ while in their work it is given by $p\left(y|y_0;\theta,F_Y^{-1}\right) = \partial F_Y^{-1}(y)/(\partial y)\,c\left(F_Y^{-1}(y)|F_Y^{-1}(y_0);\theta\right)$. Relabelling our U and p_X as F_Y^{-1} and c, respectively, all their arguments and results carry over to our setting. Assumption M and the β-mixing property required by Chen, Wu and Yi (2009) 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 4.2. Similar to the proof strategy employed in Lemma B.1, we define

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

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 (B.1) and (B.2) of Lemma B.1 and Slutsky's Theorem complete the proof. ■

B Lemma

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

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

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

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

Proof. Let $\hat{f}_Y^{(i)}(y)$ 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 $nh^{1+2i} \to \infty$,

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

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

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

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

Meanwhile, from (4.3) we have

$$\hat{U}''(y) = \frac{\hat{f}_Y'(y)}{f_X(\hat{U}(y); \hat{\theta})} - \frac{f_X'(\hat{U}(y); \hat{\theta}) \hat{f}_Y(y)^2}{f_X(\hat{U}(y); \hat{\theta})^3}.$$

Define

$$\tilde{U}''(y) = \frac{\hat{f}_Y'(y)}{f_X(U(y); \theta_0)} - \frac{f_X'(U(y); \theta_0) f_Y(y)^2}{f_X(U(y); \theta_0)^3},$$

and a similar argument leads to

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

which together with (B.3) yield (B.2). ∎

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

		Bias/κ			
Sample Size		2202		5505	
True Parameter Value	ρ_1	2SMLE	PMLE	2SMLE	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	2SMLE	PMLE	2SMLE	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 Value		ρ_1	2SMLE	PMLE	2SMLE	PMLE
$\kappa = 0.7653$	$\alpha = 1.1653$	0.9921	0.9023	1.5269	0.4576	0.7717
$\kappa = 3.8267$	$\alpha = 1.1653$	0.9675	0.2358	0.3347	0.1194	0.1754
$\kappa = 7.6533$	$\alpha = 1.1653$	0.9399	0.1328	0.1816	0.0646	0.0853
$\kappa = 15.307$	$\alpha = 1.1653$	0.8917	0.0768	0.0928	0.0349	0.0398
			RMSE/κ			
Sample Size			2202		5505	
True Parameter Value		ρ_1	2SMLE	PMLE	2SMLE	PMLE
$\kappa = 0.7653$	$\alpha = 1.1653$	0.9921	1.2424	1.7509	0.6692	0.9231
$\kappa = 3.8267$	$\alpha = 1.1653$	0.9675	0.3881	0.4511	0.2363	0.2746
$\kappa = 7.6533$	$\alpha = 1.1653$	0.9399	0.2431	0.2771	0.1498	0.1672
$\kappa = 15.307$	$\alpha = 1.1653$	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 Value		ρ_1	2SMLE	PMLE	2SMLE	PMLE
$\kappa = 0.7653$	$\alpha = 1.1653$	0.9921	0.9458	1.0299	0.6192	0.8720
$\kappa = 3.8267$	$\alpha = 1.1653$	0.9675	0.4353	0.5171	0.1899	0.2554
$\kappa = 7.6533$	$\alpha = 1.1653$	0.9399	0.2633	0.3152	0.1033	0.1279
$\kappa = 15.307$	$\alpha = 1.1653$	0.8917	0.1302	0.1646	0.0663	0.0780
			RMSE/α			
Sample Size			2202		5505	
True Parameter Value		ρ_1	2SMLE	PMLE	2SMLE	PMLE
$\kappa = 0.7653$	$\alpha = 1.1653$	0.9921	1.5614	1.5784	1.1222	1.4309
$\kappa = 3.8267$	$\alpha = 1.1653$	0.9675	0.8443	0.9197	0.3867	0.4462
$\kappa = 7.6533$	$\alpha = 1.1653$	0.9399	0.5473	0.5695	0.2298	0.2558
$\kappa = 15.307$	$\alpha = 1.1653$	0.8917	0.2802	0.3139	0.1453	0.1684

Figure 1: Marginal Densities of the Eurodollar Rates

Solid = SKST Density, Dashed = Kernel Density, Dotted = Normal Density

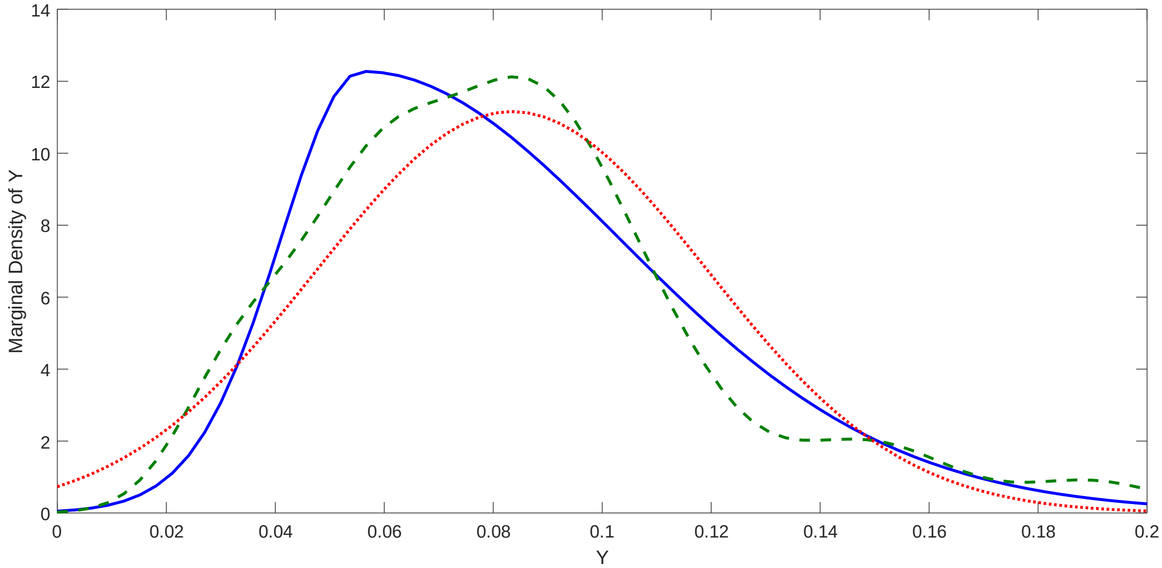


Figure 2: Estimated Drift and Diffusion for the OU-SKST Model ($T = 2202$)

Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

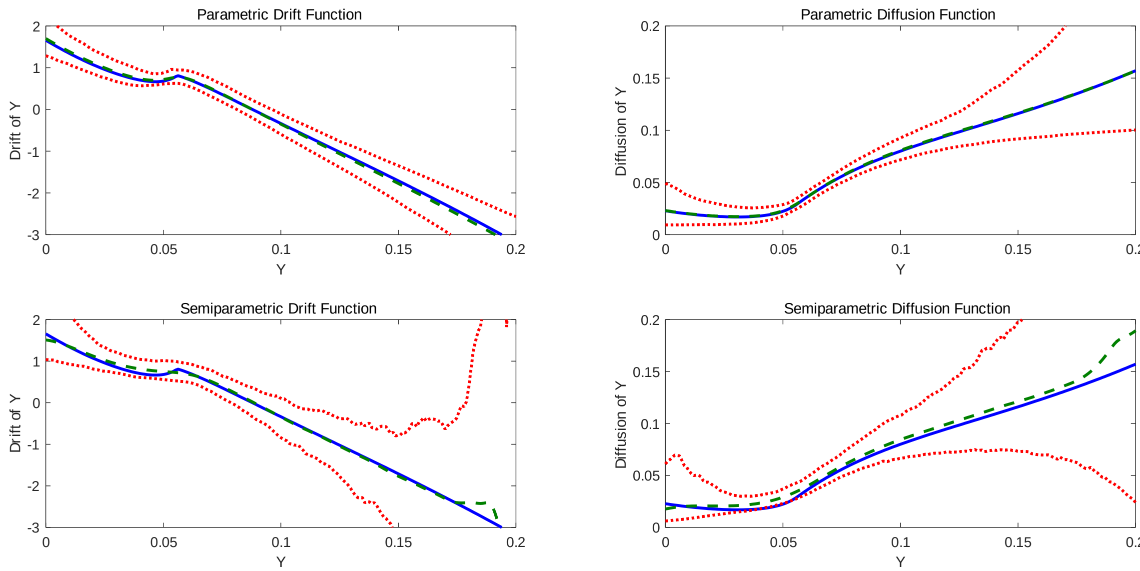


Figure 3: Estimated Drift and Diffusion for the OU-SKST Model ($T = 5505$)

Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

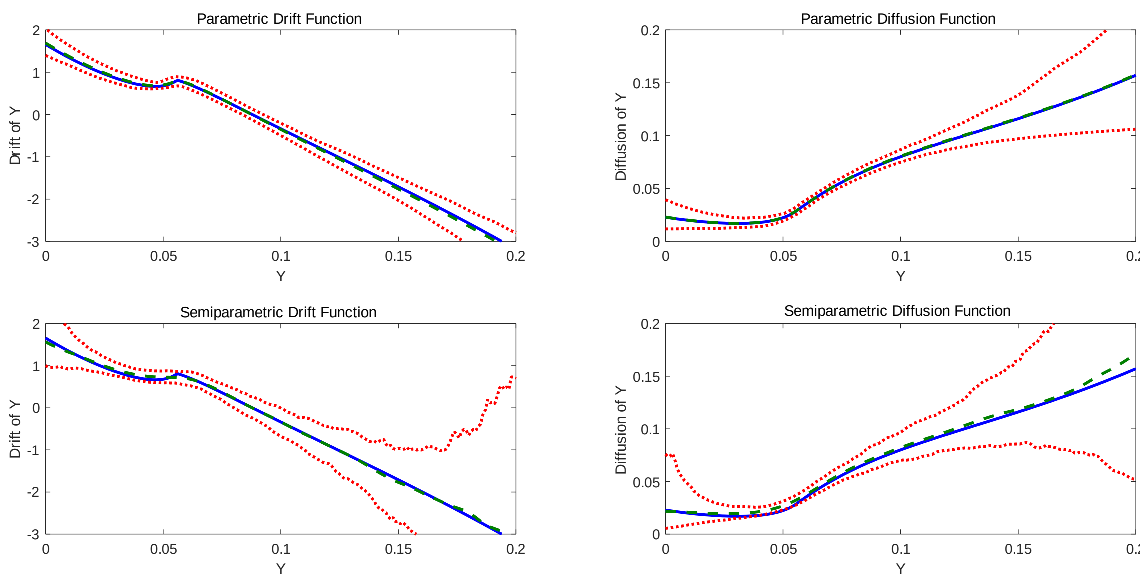


Figure 4: Estimated Drift and Diffusion for the CIR-SKST Model ($T = 2202$)

Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands

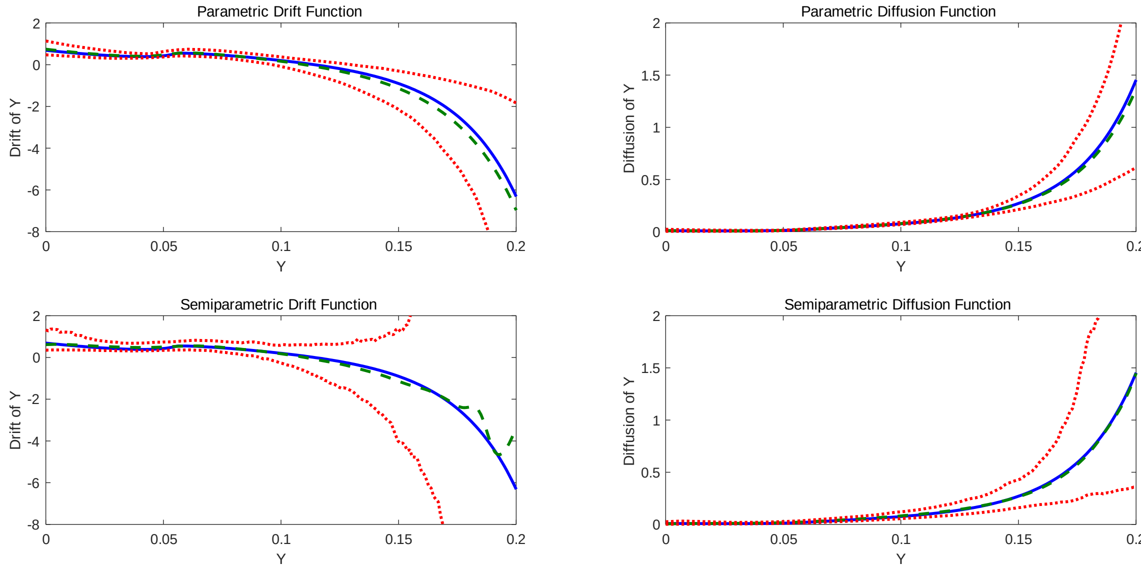
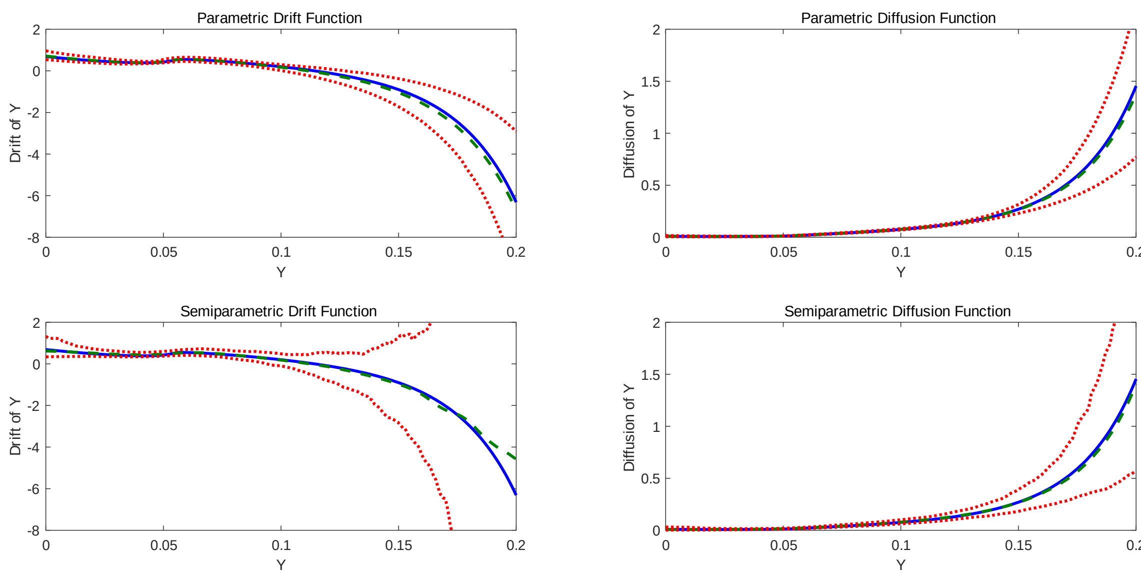


Figure 5: Estimated Drift and Diffusion for the CIR-SKST Model ($T = 5505$)

Solid = True Function, Dashed = Mean of Estimates, Dotted = 95% Confidence Bands



Research Papers 2018

CREATES
Center for Research in Econometric Analysis of Time Series

2018-04: Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Option Panels in Pure-Jump Settings

2018-05: Torben G. Andersen, Martin Thyrsgaard and Viktor Todorov: Time-Varying Periodicity in Intraday Volatility

2018-06: Niels Haldrup and Carsten P. T. Rosenskjold: A Parametric Factor Model of the Term Structure of Mortality

2018-07: Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Risk Premia Embedded in Index Options

2018-08: Torben G. Andersen, Nicola Fusari and Viktor Todorov: Short-Term Market Risks Implied by Weekly Options

2018-09: Torben G. Andersen and Rasmus T. Varneskov: Consistent Inference for Predictive Regressions in Persistent VAR Economies

2018-10: Isabel Casas, Xiuping Mao and Helena Veiga: Reexamining financial and economic predictability with new estimators of realized variance and variance risk premium

2018-11: Yunus Emre Ergemen and Carlos Velasco: Persistence Heterogeneity Testing in Panels with Interactive Fixed Effects

2018-12: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Economic Policy Uncertainty and Long-Run Stock Market Volatility and Correlation

2018-13: Emilio Zanetti Chini: Forecasting dynamically asymmetric fluctuations of the U.S. business cycle

2018-14: Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Models with Multiplicative Decomposition of Conditional Variances and Correlations

2018-15: Changli He, Jian Kang, Timo Teräsvirta and Shuhua Zhang: The Shifting Seasonal Mean Autoregressive Model and Seasonality in the Central England Monthly Temperature Series, 1772-2016

2018-16: Ulrich Hounyo and Rasmus T. Varneskov: Inference for Local Distributions at High Sampling Frequencies: A Bootstrap Approach

2018-17: Søren Johansen and Morten Ørregaard Nielsen: Nonstationary cointegration in the fractionally cointegrated VAR model

2018-18: Giorgio Mirone: Cross-sectional noise reduction and more efficient estimation of Integrated Variance

2018-19: Kim Christensen, Martin Thyrsgaard and Bezirgen Veliyev: The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing

2018-20: Ruijun Bu, Kaddour Hadri and Dennis Kristensen: Diffusion Copulas: Identification and Estimation