

CREATES Research Paper 2007-1

Nonparametric Estimation and Misspecification Testing of Diffusion Models

Dennis Kristensen



School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark



Aarhus School of Business
University of Aarhus
Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF
COPENHAGEN



NONPARAMETRIC ESTIMATION AND MISSPECIFICATION TESTING OF DIFFUSION MODELS*

DENNIS KRISTENSEN[†]
COLUMBIA UNIVERSITY AND CREATES[‡]

January, 2007

Abstract

A nonparametric kernel estimator of the drift (diffusion) term in a diffusion model are developed given a preliminary parametric estimator of the diffusion (drift) term. Under regularity conditions, rates of convergence and asymptotic normality of the nonparametric estimators are established. We develop misspecification tests of parametric diffusion models based on the nonparametric estimators, and derive the asymptotic properties of the tests. We also propose a Markov Bootstrap method for the test statistics to improve on the finite-sample approximations. The finite sample properties of the estimators are examined in a simulation study.

*This paper is a revised version of a chapter of my PhD thesis at the LSE. I wish to thank my supervisor, Oliver Linton, for helpful advice and encouragement. Parts of this paper appeared in an earlier version titled "Estimation in Two Classes of Semiparametric Diffusion Models". Thanks to Danish Social Sciences Research Council (research grant no. 2114-04-0001) for financial support.

[†]E-mail: dk2313@columbia.edu. Dep. of Economics, International Affairs Building, MC 3308, 420 West 118th Street New York NY 10027.

[‡]Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation.

1 Introduction

Diffusion processes are widely used in the modelling of the dynamics of for example interest rates, stock prices, and exchange rates; an overview of such models can be found in Björk (2004). To a lesser extent these have also been used to model the dynamics of macroeconomic variables, see e.g. Bergstrom (1990). Unfortunately, economic theory has very little to say about the precise specification of the processes. As a consequence, a wide range of parametric models have been suggested in the literature, for example Ahn and Gao (1999), Chan et al. (1992), Cox et al. (1985), Vasicek (1977). By nature, a parametric specification imposes restrictions on the dynamics allowed for in the observed data. If the model is misspecified, this can have serious implications on the conclusions drawn from the fitted model. This motivates the use of non- and semiparametric estimation techniques which to a lesser degree suffer from these deficiencies.

In this study, we develop nonparametric kernel estimators of the drift and diffusion term given low-frequency observations from a univariate diffusion model. For each of the two estimators, we assume that a preliminary parametric estimator of the other term is available, and combine this with a kernel density estimator to obtain the nonparametric estimator of the drift or diffusion term. Our estimators are of interest in themselves in a semiparametric framework, but as we demonstrate they can also be used to construct goodness-of-fit tests of parametric models. The advantage of the proposed test statistics is that, in case of rejection, they allow the researcher to detect in which direction the parametric model departs from the semi-nonparametric alternative. Thus, the two proposed test statistics seem well-suited as a guide in the search for an appropriate parametric specification. We also propose a Bootstrap method to better approximate the finite sample distribution of the test statistics.

Our nonparametric estimators are based on the fact that the stationary density can be expressed in terms of the drift and diffusion term. Inverting this expression, one can write the drift (diffusion) term as a functional of the density and the diffusion (drift) term. This allows us to identify the drift (diffusion) term given a parameterisation of the diffusion (drift) together with a nonparametric estimator of the stationary density. This identification scheme originates from Wong (1964), and was further developed in Hansen and Scheinkman (1995), and Hansen et al. (1998). Using kernel methods to estimate the stationary density, we derive the asymptotic distribution of the two estimators under regularity conditions.

We then proceed to develop misspecification tests of fully parametric models based on the semi-nonparametric estimators. Assuming estimators of a fully parametric model are available, we define statistics based on L_2 distances between the parametric and nonparametric estimators of the drift and diffusion function. Under the hypothesis that the parametric specification is correct, we derive their asymptotic distribution. We do this for two cases: The first is the case where the bandwidth of the kernel estimator shrinks to zero at a suitable rate. Using standard U-statistics results, this leads to a Normal distribution asymptotically. In the second case, the bandwidth is kept fixed which in turn leads to an asymptotic distribution of an infinite sum of weighted χ^2 random variables.

These two sets of results indicate that in finite sample, the asymptotic distribution may be a poor proxy. This is also supported by simulation studies of kernel estimators in persistent diffusion processes, c.f. Pritsker (1998). To obtain a better approximation of the finite-sample distribution, we propose a Markov Bootstrap technique of the test statistics. It is shown that the proposed Bootstrap method is consistent.

As a byproduct of the asymptotic analysis of our test statistics, we generalize recent results by Fan (1998) and Fan and Ullah (1999) to L_2 distance metrics of nonlinear transformations of parametric versus nonparametric density estimates.

One can divide the literature on the non- and semiparametric estimation of diffusion processes into two categories: In the first category high frequency data is assumed to be available while in the second only low frequency data is assumed. In the former, the time distance between the discrete observations shrinks to zero, see e.g. Bandi and Phillips (2003,2005), Jiang and Knight (1997). Thus, for a fixed time distance, these estimators will in general suffer from a discretisation bias, c.f. Nicolau (2003). On the other hand, the asymptotics of these estimators do not rely on stationarity of the observed process.

In a fixed-time-distance framework, Aït-Sahalia (1996a) and Conley et al (1997) developed estimators of two specific semiparametric diffusion models and derived their asymptotic distributions. Kristensen (2006a) consider two general classes of semiparametric models for which estimators of the parametric component are developed. Nonparametric sieve estimators of diffusion models have been proposed in Chen et al. (2000a), Darolles and Gouriéroux (2001) and Gobet et al. (2004). Unfortunately, the asymptotic distributions of these sieve estimators are not known which in effect hamper their use in applied work. Nonparametric methods have also been employed to develop misspecification tests of parametric models. Aït-Sahalia (1996b) considered two types of goodness-of-fit test statistics where an L_2 -distance metric was used to compare the density implied by the parametric model with a nonparametric kernel density estimator. An alternative nonparametric goodness-of-fit test statistics were derived in Hong and Li (2004) where the data was transformed to uniformly distributed random variables. While all the above studies are more appropriate when only low-frequency data is available, they rely on a stationarity assumption of the observed diffusion process which may be questionable in some applications.

The remains of the paper is organized as follows: In Section 2, the nonparametric estimators of the drift and diffusion term are presented and their asymptotic properties derived. In Section 3, we propose a number of different test statistics for a parametric specification against the semi-nonparametric alternative and analyse their asymptotic behaviour. A bootstrap method for the test statistics is developed in Section 4. The finite-sample performance of the estimators are examined through a simulation study in Section 5. We conclude in Section 6. All proofs and lemmas have been relegated to Appendix A and B respectively.

The following notation will be used: Let $f * g(z) = \int_{\mathbb{R}} f(u) g(u+z) du$ denote the convolution of functions f and g . The derivative of order $s \geq 0$ of $f(x)$ is denoted $f^{(s)}(x)$.

2 Semi-Nonparametric Estimators of the Drift and Diffusion Term

Consider the continuous time process $\{X_t\} = \{X_t : t \geq 0\}$ which is the solution to a univariate time-homogenous diffusion model,

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t, \quad (1)$$

where $\{W_t\}$ is a standard Brownian motion. The domain of $\{X_t\}$ is denoted $I = (l, r)$ where $-\infty \leq l < r \leq \infty$. The functions $\mu : I \mapsto \mathbb{R}$ and $\sigma^2 : I \mapsto \mathbb{R}_+$ are the so-called drift and diffusion term respectively. Note that $\{X_t\}$ is a Markov process whose dynamics are fully characterised by its transition density $\{p_t\}$,

$$\int_A p_t(y|x) dy = P(X_{s+t} \in A | X_s = x), \quad s, t \geq 0,$$

for any Borel set $A \subseteq I$.

In a fully parametric framework, both the drift and diffusion would be specified up to some unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^k$, $\mu(\cdot; \theta)$ and $\sigma^2(\cdot; \theta)$. We are here interested in imposing fewer functional restriction on the drift and diffusion, and instead consider models which are situated in either (or both) of the two following classes:

Class 1: $\sigma^2(\cdot) = \sigma^2(\cdot; \alpha)$ for some parameter $\alpha \in \mathcal{A}$.

Class 2: $\mu(\cdot) = \mu(\cdot; \beta)$ for some parameter $\beta \in \mathcal{B}$.

If a model is situated exclusively in Class 1 (2), the drift (diffusion) term is unspecified, and the model is semiparametric. If a model is situated both in Class 1 and 2, both the drift and diffusion are specified, and the model is fully parametric with $\theta = (\alpha, \beta)$. We will assume that a preliminary estimator of either α or β is available, and use this to obtain a nonparametric estimator of either the drift or diffusion term for models in Class 1 and 2 respectively. We make no assumptions about where the preliminary estimator has arrived from, merely that it is sufficiently well-behaved.

The nonparametric estimators rely on the assumption of stationarity. Suppose that $\{X_t\}$ is strictly stationary and ergodic, in which case it has a stationary marginal density which we denote π satisfying $\int_A \pi(x) dx = P(X_t \in A)$, for any $t \geq 0$ and Borel-set $A \subseteq I$. It can be shown that the stationary density is given by

$$\pi(x) = \frac{M_{x^*}}{\sigma^2(x)} \exp \left[2 \int_{x^*}^x \frac{\mu(y)}{\sigma^2(y)} dy \right], \quad (2)$$

for some arbitrary point $x^* \in \text{int}I$, and normalization factor $M_{x^*} > 0$, c.f. Karlin and Taylor (1981, Section 15.6). It is possible to revert (2) in either of the two following ways,

$$\mu(x) = \frac{1}{2\pi(x)} \frac{\partial}{\partial x} [\sigma^2(x) \pi(x)], \quad (3)$$

$$\sigma^2(x) = \frac{2}{\pi(x)} \int_l^x \mu(y) \pi(y) dy. \quad (4)$$

From these expressions, we see that instead of specifying the drift and diffusion term, an alternative specification scheme would be to specify the marginal density together with either the drift or the diffusion term, an idea originating from Wong (1964); see Hansen and Scheinkman (1995), Hansen et al. (1998). This could be done in a fully parametric framework, but here we instead rely on a nonparametric estimator of π .

We now use the expressions in (3) and (4) to construct nonparametric estimators of the drift (in Class 1) and the diffusion (in Class 2) respectively. We assume that we have observed n observations from (1), $X_\Delta, X_{2\Delta}, \dots, X_{n\Delta}$, where $\Delta > 0$ is the fixed time distance between observations; without loss of generality, we set $\Delta \equiv 1$ in the following. Under stationarity, the following nonparametric kernel estimator of π is available:

$$\hat{\pi}(x) = \frac{1}{nh} \sum_{i=1}^n K_h(x - X_i), \quad (5)$$

where $K_h(z) = K(z/h)/h$, K is a kernel, and $h > 0$ is a bandwidth; see Silverman (1986) for an introduction to this estimator.

Let us now consider a model from Class 1: In this case, the diffusion term is parameterised and an estimator $\hat{\alpha}$ is available together with the kernel estimator $\hat{\pi}$. We then estimate μ by substituting $\sigma^2(x; \hat{\alpha})$ and $\hat{\pi}$ into Eq. (3):

$$\hat{\mu}(x) = \frac{1}{2\hat{\pi}(x)} \frac{\partial}{\partial x} [\sigma^2(x; \hat{\alpha}) \hat{\pi}(x)]. \quad (6)$$

In Class 2, two alternative estimators present themselves: The obvious thing would be to directly substituting $\mu(y; \hat{\beta})$ and $\hat{\pi}$ into Eq. (4),

$$\hat{\sigma}^2(x) = \frac{2}{\hat{\pi}(x)} \int_l^x \mu(y; \hat{\beta}) \hat{\pi}(y) dy.$$

However, the integral $\int_l^x \mu(y) \pi(y) dy$ can be estimated without bias by a sample average,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\} \mu(X_i) \rightarrow^P \int_l^x \mu(y) \pi(y) dy,$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. So we suggest to estimate $\sigma^2(x)$ by

$$\hat{\sigma}^2(x) = \frac{2}{\hat{\pi}(x)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\} \mu(X_i; \hat{\beta}). \quad (7)$$

To establish the asymptotic properties of these two estimators, we impose the following assumptions:

- A.1** (i) The drift $\mu(\cdot)$ and diffusion $\sigma^2(\cdot) > 0$ are continuously differentiable.
(ii) there exists a twice continuously differentiable function $V : \mathbb{R} \mapsto \mathbb{R}_+$ with $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$, and constants $b, c > 0$ such that

$$\mu(x) V'(x) + \frac{1}{2} \sigma^2(x) V''(x) \leq -cV(x) + b. \quad (8)$$

A.2 The marginal density π is uniformly differentiable of order $m + 1 \geq 4$ with bounded derivatives. The conditional density $p(y|x) \equiv p_1(y|x)$ is uniformly fourth-order differentiable with

$$\sup_{x,y \in I} p(y|x) \pi(x) < \infty$$

A.3 1. $\beta \mapsto \mu(x; \beta)$ is continuously differentiable, satisfying $\|\partial_\beta^i \mu(x; \beta)\| \leq V(x)$, $i = 0, 1$.
 2. $\alpha \mapsto \sigma^2(x; \alpha)$ is continuously differentiable satisfying $\|\partial_{x,\alpha}^{ij} \sigma^2(x; \alpha)\| \leq V(x)$, $i, j = 0, 1$.

A.4 There exists functions ψ_k with $E[\psi_k(X_1|X_0)] = 0$ and $E[\|\psi_k(X_1|X_0)\|^{2+\delta}] < \infty$, $k = 1, 2$, such that:

1. $\hat{\alpha} = \alpha_0 + \sum_{i=1}^n \psi_1(X_i|X_{i-1})/n + o_P(1/\sqrt{n})$.
2. $\hat{\beta} = \beta_0 + \sum_{i=1}^n \psi_2(X_i|X_{i-1})/n + o_P(1/\sqrt{n})$.

B.1 The kernel K is differentiable with

$$\left| K^{(i)}(z) \right| \leq C |z|^{-\eta}, \quad \left| K^{(i)}(z) - K^{(i)}(z') \right| \leq C |z - z'|, \quad i = 0, 1$$

and

$$\int_{\mathbb{R}} z^j K(z) dz = \begin{cases} 1, & j = 0, \\ 0, & j = 1, \dots, m-1, \\ < \infty, & j = m. \end{cases}$$

As noted earlier, we have to assume that $\{X_t\}$ is stationary and ergodic in order to be able to identify the unspecified term. In fact, Assumption (A.1) is sufficient for $\{X_t\}$ to be well-defined, stationary and geometrically β -mixing. (A.1) is based on results by Meyn and Tweedie (1993); alternative mixing conditions for diffusion processes can be found in Hansen and Scheinkman (1995) and Veretennikov (1997); see also Karatzas and Shreve (1991, Section 5.5). We here require it to be geometrically β -mixing even though some of the results stated in this section will actually hold under weaker mixing conditions. But since in the next section we need β -mixing of geometric order to employ U-statistics results for dependent sequences (see Yoshihara, 1976; Fan and Li, 1999), we impose this restriction throughout for clarity. Many models found in the finance literature satisfy (A.1) under suitable restrictions on the parameters, including all the models cited in the Introduction.

The existence of higher order derivatives of π assumed in (A.2) combined with the use of a higher order kernel as given in (B.1) reduces the bias of the kernel estimator and its first derivative. The smoothness of π as measured by its number of derivatives, $m + 1$, determines how much the bias can be reduced with. The condition that π is $m + 1$ times differentiable is satisfied if μ and σ^2 are m and $m + 1$ times differentiable respectively, c.f. Eq. (2). The conditions on the conditional density p are used to control the variance component of the kernel estimator.

Assumption (A.3) ensures that $E \left[\partial_{\beta}^i \mu(X_0; \beta) \right] < \infty$ and $E \left[\partial_{x, \alpha}^{ij} \sigma^2(X_0; \alpha) \right] < \infty$. These moments are used when demonstrating uniform convergence of the nonparametric estimators.

The conditions on the preliminary estimators given in (A.4) is only needed in Theorem 5. All other results can be established under the weaker assumption that they converge at a faster rate than the kernel estimator. The condition is satisfied under great generality for most well-behaved estimators. For the fully parametric MLE, Aït-Sahalia (2002) gives conditions for (A.4) to hold, while Kristensen (2006a) give conditions under which a class of semiparametric estimators satisfy the condition.

One might also wish to have uniform convergence of the semi-nonparametric estimators. However, since the estimators and the limits themselves potentially are unbounded functions, this is not readily possible. To circumvent this problem, we control the tail behaviour of the estimator by trimming, ensuring that the nonparametric estimator equals zero outside a compact, but growing set. We define a sequence of trimming sets $\hat{A} = \hat{A}_n$ by

$$\hat{A} = \{x \in I | \hat{\pi}(x) \geq a\} \quad (9)$$

for some sequence $a = a_n \rightarrow 0$. We then show uniform convergence on the increasing set $\hat{A}$ by combining results of Hansen (2006) and Andrews (1995).

In the following let $\{x_i\}_{i=1}^N$ be a *fixed set* of distinct points in the domain I , $x_i \neq x_j$ for $i \neq j$. We then give pointwise and uniform results for each of the two estimators:

Theorem 1 (Class 1) *Assume that (A.1)-(A.2) (A.3.1)-(A.4.1) and (B.1) hold. Then:*

1. As $h \rightarrow 0$, $nh^3 \rightarrow \infty$ and $nh^{3+m} \rightarrow 0$:

$$\sqrt{nh^3} \{\hat{\mu}(x_i) - \mu_0(x_i)\}_{i=1}^N \xrightarrow{d} N\left(0, \text{diag}(\{V_{\mu}(x_i)\}_{i=1}^N)\right),$$

where

$$V_{\mu}(x) = \frac{\sigma^4(x)}{4\pi(x)} \int_{\mathbb{R}} K^{(1)}(z)^2 dz.$$

- 2.

$$\sup_{x \in \hat{A}} |\hat{\mu}(x) - \mu(x)| = \sum_{s=0}^1 \left\{ O_P \left(\sqrt{\log(n)} a^{-1-s} n^{-1/2} h^{-(1+2s)/2} \right) + O_P(a^{-1-s} h^m) \right\}.$$

Theorem 2 (Class 2) *Assume that (A.1)-(A.3), (A.4.2)-(A.5.2) and (B.1) hold. Then:*

1. As $h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^{1+m} \rightarrow 0$:

$$\sqrt{nh} \{\hat{\sigma}^2(x_i) - \sigma^2(x_i)\}_{i=1}^N \xrightarrow{d} N\left(0, \text{diag}(\{V_{\sigma}(x_i)\}_{i=1}^N)\right),$$

where

$$V_{\sigma}(x) = \frac{\sigma^4(x)}{\pi(x)} \int_{\mathbb{R}} K(z)^2 dz.$$

2.

$$\sup_{x \in \hat{A}} |\hat{\sigma}^2(x) - \sigma^2(x)| = O_P\left(\sqrt{\log(n)} a^{-2} n^{-1/2} h^{-1/2}\right) + O_P(a^{-2} h^m).$$

The pointwise asymptotic variances for $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ respectively can be estimated by:

$$\hat{V}_{\mu}(x) = \frac{\sigma^4(x; \hat{\alpha})}{4\hat{\pi}(x)} \int_{\mathbb{R}} K^{(1)}(z)^2 dz, \quad \hat{V}_{\sigma}(x) = \frac{\hat{\sigma}^4(x)}{\hat{\pi}(x)} \int_{\mathbb{R}} K(z)^2 dz. \quad (10)$$

We here only stated results for the estimation of μ and σ^2 but one can easily derive similar results for the estimators of the derivatives of μ and σ^2 . Observe that both nonparametric estimators are asymptotically independent across the points $\{x_i\}_{i=1}^N$. This is a well-known property of kernel-estimators, cf. Robinson (1983), which facilitates global inference, for example when constructing pointwise confidence bands, and testing hypotheses (see Section 3).

The first part of the result stated in Theorem 2 has already been obtained by Aït-Sahalia (1996a) for the semiparametric model where $\mu(x; \beta) = \beta_1(\beta_2 - x)$ and σ^2 was left unspecified. We here have extended his result to two general classes of semiparametric diffusion models.

The rate of convergence of $\hat{\mu}$ is slower than the one of $\hat{\sigma}^2$. This owes to the fact that $\hat{\mu}$ depends on both $\hat{\pi}$ and its first derivative, $\hat{\pi}^{(1)}$, while $\hat{\sigma}^2$ is only a function of $\hat{\pi}$. The density derivative has slower weak convergence rate than $\hat{\pi}$, $\sqrt{nh^3}$ relative to $\sqrt{nh}$, which the drift estimator inherits. Thus, the drift is more difficult to estimate than the diffusion term in our setting. This observation has been made elsewhere in the literature. Gobet et al. (2003) show that the optimal convergence rate of the nonparametric estimation of the drift is lower than for the diffusion, and coin the nonparametric estimation of μ given discrete observations as an "ill-posed problem". Similarly, Bandi and Phillips (2003) demonstrate that for a stationary diffusion, it is only possible to estimate $\mu(x)$ nonparametrically with $\sqrt{n\Delta h}$ -rate, while $\sigma^2(x)$ can be estimated at the faster rate $\sqrt{nh}$ as $\Delta \rightarrow 0$ and $n\Delta \rightarrow \infty$.

3 Goodness-of-Fit Testing

In this section, we propose a number of different specification tests for a parametric diffusion model against semiparametric alternatives. We consider the fully parametric hypothesis,

$$H_0 : \sigma^2(\cdot) = \sigma^2(\cdot; \alpha) \text{ and } \mu(\cdot) = \mu(\cdot; \beta) \text{ for some } (\alpha, \beta) \in \mathcal{A} \times \mathcal{B}.$$

We then wish to test H_0 against each of the two following semiparametric alternatives,

$$H_{A,1} : \sigma^2(\cdot) = \sigma^2(\cdot; \alpha) \text{ for some } \alpha \in \mathcal{A} \text{ and } \mu(\cdot) \neq \mu(\cdot; \beta) \text{ for all } \beta \in \mathcal{B},$$

and

$$H_{A,2} : \sigma^2(\cdot) \neq \sigma^2(\cdot; \alpha) \text{ for all } \alpha \in \mathcal{A} \text{ and } \mu(\cdot) = \mu(\cdot; \beta) \text{ for some } \beta \in \mathcal{B}.$$

Note that the situation $\sigma^2(\cdot) \neq \sigma^2(\cdot; \alpha)$ for all $\alpha \in \mathcal{A}$ and $\mu(\cdot) \neq \mu(\cdot; \beta)$ for all $\beta \in \mathcal{B}$ is not included in the two alternatives.

Under the null, the model is fully specified and the parameter vector $\theta = (\alpha, \beta)$ can be estimated using standard parametric estimation methods with the obvious one being MLE, see e.g. Aït-Sahalia (2002). Alternatively, given a preliminary parametric estimator of either α or β , one can obtain an estimator of the remaining parameter by matching our nonparametric estimator with the parametric specification, c.f. Bandi & Phillips (2005). For example in Class 1, given $\hat{\alpha}$, we can first obtain the nonparametric estimator $\hat{\mu}(x)$ and then estimate β by $\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} n^{-1} \sum_{i=1}^n [\hat{\mu}(x_i) - \mu(x_i; \beta)]^2$; similarly for Class 2.

Under each of the two alternatives, we use the relevant nonparametric estimator developed in the previous section. The preliminary parametric component, $\hat{\alpha}$ or $\hat{\beta}$, can either arrive from the fully parametric submodel such that $\hat{\alpha} = \tilde{\alpha}$ or $\hat{\beta} = \tilde{\beta}$, or be separate semiparametric estimates. In the latter case, the general semiparametric estimation method in Kristensen (2006a) can be employed. Under the null, it will obviously make no difference with regard to the asymptotics whether the preliminary parametric estimators used in the nonparametric estimators arrive from the fully parametric model or a semiparametric one. Under the alternative however, we expect that it will be quite influential whether the fully parametric or semiparametric estimator is used.

We make the following assumption about the estimator $\tilde{\theta}$ obtained under H_0 :

A.5 Under H_0 , the estimator $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta})$ satisfies $\tilde{\theta} = \theta_0 + \sum_{i=1}^n \tilde{\psi}(X_i|X_{i-1})/n + o_P(1/\sqrt{n})$ with $E[\tilde{\psi}(X_1|X_0)] = 0$ and $E[||\tilde{\psi}(X_1|X_0)||^{2+\delta}] < \infty$.

Some of the following results (Theorem 3-4) hold under the weaker assumption that $\tilde{\theta}$ has a faster rate of convergence relative to the nonparametric estimators. The $\sqrt{n}$ -asymptotic normality which is imposed in (A.5) is only needed to establish Theorem 5.

The proposed test statistics are based on the square differences between the nonparametric and parametric estimates of the drift and diffusion term. We start by considering test statistics which only compare the semi-nonparametric and fully parametric estimator across a fixed number $1 \leq N < \infty$ of points in the domain, $\{x_i\}_{i=1}^N$, as given in the previous section. For similar test procedures for conditional means, see e.g. Gozalo (1997). Since we are considering a fixed number of distinct points, Theorem 1 and Theorem 2 together with the delta method and the continuous mapping theorem yield the following result:

Theorem 3 Under (A.1)-(A.5) and (B.1): As $h \rightarrow 0$, $nh^3 \rightarrow \infty$ and $nh^{3+m} \rightarrow 0$,

$$T_{1,n} = nh^3 \sum_{i=1}^N \left[\frac{\mu(x_i; \tilde{\beta}) - \hat{\mu}(x_i)}{\sqrt{\hat{V}_\mu(x_i)}} \right]^2 \xrightarrow{d} \chi^2(N),$$

while, $h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^{1+m} \rightarrow 0$,

$$T_{2,n} = nh \sum_{i=1}^N \left[\frac{\sigma^2(x_i; \tilde{\alpha}) - \hat{\sigma}^2(x_i)}{\sqrt{\hat{V}_\sigma(x_i)}} \right]^2 \xrightarrow{d} \chi^2(N),$$

where $\hat{V}_\mu(x)$ and $\hat{V}_\sigma(x)$ are given in (10) and (10) respectively.

The actual choice of N and $\{x_i\}_{i=1}^N$ is not obvious. Gozalo (1997) proposes to perform a random selection of points over I . Also, he shows that the number of points N used in the test statistic for μ (σ^2) can grow with n as long as it does so at a rate slower than $\sqrt{nh^3}$ ($\sqrt{nh}$). Still, the test statistics will be sensitive to the actual choice of $\{x_i\}_{i=1}^N$ which is a less attractive feature of the two test statistics.

Given this defect, we now develop test statistics that evaluate the square difference between the semi-nonparametric estimator and the parametric over the whole domain I . We do this by considering,

$$\begin{aligned} \bar{T}_{1,n} &= \int_I [\mu(x; \tilde{\beta}) - \hat{\mu}(x)]^2 \omega(x) dx, \\ \bar{T}_{2,n} &= \int_I [\sigma^2(x; \tilde{\beta}) - \hat{\sigma}^2(x)]^2 \omega(x) dx, \end{aligned}$$

for some weighting function $\omega : I \mapsto \mathbb{R}_+$.

Assume for the moment that α is known. Then,

$$\bar{T}_{1,n} = \int_I [G_1(x, \hat{\pi}(x)) - G_1(x, \tilde{\pi}(x))]^2 \omega(x) dx,$$

where $\tilde{\pi}(x) = \pi(x; \hat{\beta})$ and

$$G_1(x, \pi(x)) = \frac{1}{2\pi(x)} \frac{\partial}{\partial x} [\sigma^2(x) \pi(x)].$$

Similar expression holds for $\bar{T}_{2,n}$, if β was known. This highlights that $\bar{T}_{1,n}$ and $\bar{T}_{2,n}$ are in fact testing the parametric specification of the marginal density, $\pi(\cdot) = \pi(\cdot; \theta)$ for some $\theta \in \Theta$, against the nonparametric alternative, $\pi(\cdot) \neq \pi(\cdot; \theta)$ for all $\theta \in \Theta$. One could therefore test the hypothesis H_0 against both semiparametric alternatives by considering $\int_I [\hat{\pi}(x) - \tilde{\pi}(x)]^2 \omega(x) dx$. This test statistic has been examined in Aït-Sahalia (1996b), Fan (1994, 1995, 1998) and Fan and Ullah (1999). However, observe that $\bar{T}_{1,n}$ and $\bar{T}_{2,n}$ involve nontrivial transformations of the marginal density and therefore test different directions of departure from the null. In particular, if for example $\bar{T}_{1,n}$ rejects the null while $\bar{T}_{2,n}$ accepts it, the diffusion term appears to be correctly specified while it would be necessary to modify the parametric specification of the drift term. This is in contrast to $\int_I [\hat{\pi}(x) - \tilde{\pi}(x)]^2 \omega(x) dx$ where a rejection does not give any guidance to a more appropriate specification of the parametric model.

The following theorem gives asymptotic results of the two test statistics as $h \rightarrow 0$. The proof strategy is to linearize the test statistics in terms of π such that for appropriate weighting function w_1 and w_2 , $\int \left[\hat{\pi}^{(k)}(x) - \pi^{(k)}(x) \right]^2 w_k(x) dx$, $k = 1, 2$, will drive the asymptotic distributions of the two test statistics. We may then use same arguments as in Fan (1994) and Fan and Ullah (1999) to obtain the asymptotic distributions. The theorem is shown under the following regularity condition on the weighting function:

B.2 The weighting function $\omega : I \mapsto \mathbb{R}_+$ has compact support.

The assumption of a fixed, compact support of ω is made in order to control the tail behaviour of the estimators of the drift and diffusion term. Under suitable regularity conditions on the tail behaviour of ω and the drift and diffusion, one should be able to allow for infinite support by introducing a trimmed version, e.g. $\hat{\omega}(x) = \mathbb{I}\{x \in \hat{A}\}\omega(x)$ where $\hat{A}$ is the trimming set introduced in the previous section; in the limit $\hat{\omega}$ would have same support as ω . This would lead to more technical proofs however, and we therefore maintain (B.2) throughout for simplicity.

Theorem 4 Under (A.1)-(A.5) and (B.1)-(B.2):

1. As $h \rightarrow 0$, $nh^3 \rightarrow \infty$, and $nh^{3/2+2m} \rightarrow 0$,

$$nh^{5/2} (\bar{T}_{1,n} - c_1 / (nh^3)) \rightarrow^d N(0, \nu_1^2),$$

where

$$\begin{aligned} c_1 &= \frac{1}{4} \left\{ \int_{\mathbb{R}} K^{(1)}(z)^2 dz \right\} \left\{ \int_I \frac{\sigma^4(x)}{\pi(x)} \omega(x) dx \right\}, \\ \nu_1^2 &= \frac{1}{8} \left\{ \int_{\mathbb{R}} [K^{(1)} * K^{(1)}]^2(z) \frac{\sigma^4(z)}{\pi^2(z)} \omega(z) dz \right\} \left\{ \int_I \sigma^4(x) \omega(x) dx \right\}. \end{aligned}$$

2. As $h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^{1/2+2m} \rightarrow 0$,

$$nh^{1/2} (\bar{T}_{2,n} - c_2 / (nh)) \rightarrow^d N(0, \nu_2^2),$$

where

$$\begin{aligned} c_2 &= \left\{ \int_{\mathbb{R}} K(z)^2 dz \right\} \left\{ \int_I \frac{\sigma^4(x)}{\pi(x)} \omega(x) dx \right\}, \\ \nu_2^2 &= 2 \left\{ \int_{\mathbb{R}} [K * K]^2(z) \frac{\sigma^4(z)}{\pi^2(z)} \omega(z) dz \right\} \left\{ \int_I \sigma^4(x) \omega(x) dx \right\}, \end{aligned}$$

Consistent estimates of $c_k(n)$ and ν_k^2 can be obtained by substituting the unknown quantities entering these, that is, $\sigma^2(z)$ and $\pi(z)$, for their estimates.

Next, we consider a slight modification of the test statistics,

$$\begin{aligned}\tilde{T}_{1,n} &= \int_I [\tilde{\mu}(x) - \hat{\mu}(x)]^2 \omega(x) dx, \\ \tilde{T}_{2,n} &= \int_I [\tilde{\sigma}^2(x) - \hat{\sigma}^2(x)]^2 \omega(x) dx,\end{aligned}$$

where $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ are the same semi-nonparametric estimators as before, while

$$\tilde{\mu}(x) = \frac{1}{2(K_h * \tilde{\pi})(x)} \frac{\partial}{\partial x} [\sigma^2(x; \tilde{\alpha})(K_h * \tilde{\pi})(x)], \quad (11)$$

$$\tilde{\sigma}^2(x) = \frac{2}{(K_h * \tilde{\pi})(x)} \int_l^x \mu(y; \tilde{\beta})(K_h * \tilde{\pi})(y) dy, \quad (12)$$

$$\tilde{\pi}(x) = \frac{M_{x^*}(\tilde{\theta})}{\sigma^2(x; \tilde{\alpha})} \exp \left[2 \int_{x^*}^x \frac{\mu(y; \tilde{\beta})}{\sigma^2(y; \tilde{\alpha})} dy \right]. \quad (13)$$

So we here transform the parametric density estimator $\tilde{\pi}$ by $K * \tilde{\pi}$, thereby removing the bias incurred from the nonparametric kernel density estimation. The same strategy as in Fan (1994, Proof of Theorem 4.1) can be used to show that $\tilde{T}_{k,n}$ has the same asymptotic distribution as $\bar{T}_{k,n}$, $k = 1, 2$, but without the requirement that $nh^{3/2+2m} \rightarrow 0$ and $nh^{1/2+2m} \rightarrow 0$ respectively. Furthermore, we can derive the asymptotics of these test statistics for fixed $h > 0$; this is based on an extension of the results found in Fan (1998).

Theorem 5 *Under (A.1)-(A.5) and (B.1)-(B.2), for any fixed $h > 0$ and $k = 1, 2$,*

$$n\tilde{T}_{k,n} \rightarrow^d \int_{\mathbb{R}^d} Z_k^2(x, h) \omega(x) dx, \quad n \rightarrow \infty,$$

where $Z_k(\cdot, h) \sim N(0, \Sigma_k(\cdot, h))$ and $\Sigma_k(\cdot, h)$ satisfies

$$\begin{aligned} \int_{\mathbb{R}^d} \Sigma_k(x, h) a(x) \omega(x) dx &= \text{Var} \left(\left\langle \bar{\psi}_0^k(h), a \right\rangle \right) + 2 \sum_{i=1}^{\infty} \text{Cov} \left(\left\langle \bar{\psi}_0^k, a \right\rangle, \left\langle \bar{\psi}_i^k(h), a \right\rangle \right), \\ \langle a, b \rangle &= \int_{\mathbb{R}^d} a(x) b(x) \omega(x) dx, \end{aligned}$$

with the random functions $\bar{\psi}_i^k(\cdot, h)$, $k = 1, 2$ being given in Eq. (22) and (23) respectively.

This theorem highlights that the asymptotic distribution given in Theorem 4 may deliver a poor finite sample approximation. This may be further distorted by the need to estimate the unknown quantities entering the asymptotic distribution and the dependence in data, see e.g. Pritsker (1998). We therefore in the next section propose a Markov bootstrap method to obtain a better approximation of the finite sample distribution of the test statistics.

4 A Parametric Bootstrap

We here develop a parametric bootstrap for the test statistics proposed in the previous section along the same lines as in Fan (1995, 1998). We propose to draw a new sample from the parametric transition density $p(y|x; \tilde{\theta})$ and use these to approximate the distributions. In the following, T_n denotes either one of the test statistics developed in the previous description.

Step 1 Draw $X_1^* \sim \pi(\cdot; \tilde{\theta})$, and recursively $X_i^* \sim p(\cdot | X_{i-1}^*; \tilde{\theta})$, $i = 2, \dots, n$.

Step 2 Replace the data $\{X_i\}_{i=1}^n$ with the bootstrap sample $\{X_i^*\}_{i=1}^n$ in the estimation of the parametric components and the relevant test statistic; we denote these $\hat{\alpha}^*$, $\hat{\beta}^*$, $\tilde{\theta}^*$ and T_n^* respectively.

Step 3 Repeat Step 1-2 $B \geq 1$ times, each new sample being independent of the previous ones, yielding $T_{n,1}^*, \dots, T_{n,B}^*$. Use the empirical distribution of these to estimate the bootstrap one.

The initialization of the Bootstrap sample could be exchanged for $X_1^* = X_1$ since we are working with a geometrically ergodic Markov chain. In order to show that the proposed Bootstrap method is consistent, we impose an additional assumption, which is a modified version of the ones stated in Fan (1995, 1998):

$$\mathbf{A.6} \quad \hat{\alpha}^* = \hat{\alpha} + \frac{1}{n} \sum_{i=1}^n \psi_1(X_i | X_{i-1}) + o_{P^*}(1/\sqrt{n}), \quad \hat{\beta}^* = \hat{\beta} + \frac{1}{n} \sum_{i=1}^n \psi_2(X_i | X_{i-1}) + o_{P^*}(1/\sqrt{n}), \quad \tilde{\theta}^* = \tilde{\theta} + \frac{1}{n} \sum_{i=1}^n \tilde{\psi}(X_i | X_{i-1}) + o_{P^*}(1/\sqrt{n}).$$

Proposition 6 *Under (A.1)-(A.6) and (B.1)-(B.2), the distribution of $T_{k,n}^*$, $\bar{T}_{k,n}^*$ and $\tilde{T}_{k,n}^*$ converge almost surely to the same limit as the distribution of $T_{k,n}$, $\bar{T}_{k,n}$ and $\tilde{T}_{k,n}$ respectively, $k = 1, 2$.*

We will not give a proof of this result since it proceeds along the exact same lines as in Fan (1995, 1998).

5 A Simulation Study

We here examine how the nonparametric estimators perform in finite sample. For an application of the estimators to interest rates, we refer to Kristensen (2006b). We choose as data generating models the CKLS model and a submodel of the model proposed in Aït-Sahalia (1996b),

$$dX_t = \{\beta_1 + \beta_2 X_t\} dt + \sqrt{\alpha_1 X_t^{\alpha_2}} dW_t, \quad (\text{CKLS})$$

$$dX_t = \{\beta_1 + \beta_2 X_t + \beta_3 X_t^2 + \beta_4 X_t^{-1}\} dt + \sqrt{\alpha_1 X_t^{\alpha_2}} dW_t. \quad (\text{AS})$$

The data-generating parameters are chosen to match the estimates obtained when fitting the model by MLE to the Eurodollar interest rate data considered in Aït-Sahalia (1996a,b). The parameter estimates satisfy the β -mixing conditions found

in Ait-Sahalia (1996b) such that (A.1) holds. We measure time in years and set the time distance to $\Delta = 1/252$, thereby effectively ignoring holidays and weekends, and consider two sample sizes, $n = 2500, 5000$.

For each sample, we estimate the two following semiparametric models when CKLS and AS is the data generating process respectively,

$$\text{CKLS 1:} \quad dX_t = \mu(X_t) dt + \sqrt{\alpha_1 X_t^{\alpha_2}} dW_t, \quad (14)$$

$$\text{CKLS 2:} \quad dX_t = \{\beta_1 + \beta_2 X_t\} dt + \sigma(X_t) dW_t, \quad (15)$$

$$\text{AS 1:} \quad dX_t = \mu(X_t) dt + \sqrt{\alpha_1 X_t^{\alpha_2}} dW_t, \quad (16)$$

$$\text{AS 2:} \quad dX_t = \{\beta_1 + \beta_2 X_t + \beta_3 X_t^2 + \beta_4 X_t^{-1}\} dt + \sigma(X_t) dW_t, \quad (17)$$

The parameters of the semiparametric models are estimated using the MLE method proposed in Kristensen (2006a). Once the parametric component has been estimated, we calculate $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ for models in Class 1 and 2 respectively. We also estimate the fully parametric models (CKLS)-(AS) by MLE which allows us to compare the semiparametric and parametric estimates. In order to evaluate the likelihood in both the parametric and semiparametric case, we employ the simulated likelihood method of Kristensen and Shin (2006); see also Kristensen (2006a, Section 5). This is implemented by simulating $N = 100$ values for each observation, using the Euler scheme with a step length of $\delta = \Delta/10$.

We first investigate the behaviour of the nonparametric estimators for the CKLS model. We consider two sets of data generating parameter values, (i) $\alpha = (1.8207, 2.6217)$, $\beta = (0.0344, -0.2921)$ and (ii) $\alpha = (0.1547, 1.7079)$, $\beta = (0.0271, -0.4455)$. These are estimates from the Eurodollar data set using (i) the full sample 1973-1995 and (ii) the subsample 1982-1995. The first parameter set generates high volatility and low mean reversion while the second one generates just the opposite behaviour. In Figure 1-2, pointwise means and confidence bands of the the fully parametric and nonparametric drift estimates are plotted for the parameters (i) and (ii) respectively. For (i), Figure 1 shows that the nonparametric drift estimator performs well in the range $x \in [0.03, 0.12]$ while it is rather imprecise in tails. This is probably a consequence of that the process rarely visits outside this interval and that the strong persistence makes the nonparametric density estimator more biased. This is confirmed by the performance in Figure 2 where the nonparametric drift estimator becomes more precise in the tails with increased mean reversion. In Figure 3-4, the diffusion estimators are plotted. For both choices of parameter values, the estimator is very imprecise out in the right tail of the support. Moreover, a decrease in the volatility seemingly leads to a further deterioration of the performance.

Next, we examine the behaviour of the AS model. We do this with the parameters fitted to the full sample. In Figure 5 and 6 respectively, the drift and diffusion estimators are plotted. The parametric drift estimator is not very precise which owes to the fact that the drift parameters in the AS model are difficult to pin down, see also Kristensen (2006a, Section 5). In comparison, the nonparametric drift estimator

performs fairly well, and has more or less the same level of precision as the parametric one. The performance of the nonparametric diffusion estimator is not quite so good though.

6 Concluding Remarks

Extensions of our results to multivariate diffusion models would be of interest. However, our identification scheme cannot readily be extended to general multivariate diffusion models, since the link between the invariant density, the drift and the diffusion term utilised here does not necessarily hold in higher dimensions. If one is ready to restrict the attention to the class of multivariate models which does satisfy this relation, the proposed estimation procedure should still work. For example, one may consider the class of d -dimensional diffusions with drift $\mu : \mathbb{R}^d \mapsto \mathbb{R}^d$ and diffusion $\sigma^2 : \mathbb{R}^d \mapsto \mathbb{R}^{d \times d}$, where the following relationship holds between the drift and diffusion,

$$\mu_i(x) = \frac{1}{2\pi(x)} \sum_{j=1}^d \frac{\partial}{\partial x_j} [\sigma_{ij}^2(x) \pi(x)]. \quad (18)$$

This restriction is for example imposed by Chen et al (2000b) in their nonparametric study of multivariate diffusion models. Again, $\pi(x)$ can be estimated by kernel density methods which together with a parametric specification for σ^2 will lead to the same type of estimators considered here.

References

- Ahn, D.-H. & B. Gao (1999) A Parametric Nonlinear Model of Term Structure Dynamics. *Review of Financial Studies* 12, 721-762.
- Aït-Sahalia, Y. (1996a) Nonparametric Pricing of Interest Rate Derivative Securities. *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.
- Andrews, D.W.K. (1995) Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory* 11, 560-596.
- Bandi, F.M. & P.C.B. Phillips (2003) Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* 71, 241-283.
- Bandi, F.M. & P.C.B. Phillips (2005) A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions. Forthcoming, *Journal of Econometrics*.
- Bergstrom, A.R. (1990) *Continuous Time Econometric Modelling*. Oxford: Oxford University Press.
- Björk, T. (2004) *Arbitrage Theory in Continuous Time, 2nd Edition*. Oxford: Oxford University Press.
- Chan, K.C., Karolyi, G.A., Longstaff, F.A. & Sanders, A.B. (1992) An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance* 47, 1209-1227.
- Chen, X., L.P. Hansen & J. Scheinkman (2000a) Shape-preserving Estimation of Diffusions. Working paper, University of Chicago.
- Chen, X., L.P. Hansen & J. Scheinkman (2000b) Principal Components and the Long Run. Working paper, University of Chicago.
- Conley, T., L.P. Hansen, E. Luttmer & J. Scheinkman (1997) Short-term Interest Rates as Subordinated Diffusions. *Review of Financial Studies* 10, 525-577.
- Cox, J.C., J.E. Ingersoll & S. Ross (1985) A Theory of the Term Structure of Interest Rates. *Econometrica* 53, 373-384.
- Darolles, S. & C. Gouriéroux (2001) Truncated Dynamics and Estimation of Diffusion Equations. *Journal of Econometrics* 102, 1-22.
- Doukhan, P., P. Massart & E. Rio (1994) The Central Limit Theorem for Strongly Mixing Processes. *Ann. Inst. H. Poincaré* 30, 63-82.

- Fan, Y. (1994) Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method. *Econometric Theory* 10, 316-356.
- Fan, Y. (1995) Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test. *Econometric Reviews* 14, 367-382.
- Fan, Y. (1998) Goodness-of-Fit Tests Based on Kernel Density Estimators with Fixed Smoothing Parameters. *Econometric Theory* 14, 604-621.
- Fan, Y. & Q. Li (1999) Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing. *Journal of Nonparametric Statistics* 10, 245-271.
- Fan, Y. & A. Ullah (1999) On Goodness-of-Fit Tests for Weakly Dependent Processes Using Kernel Method. *Journal of Nonparametric Statistics* 11, 337-360.
- Gobet, E., M. Hoffmann & M. Reiß (2004) Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data. *Annals of Statistics* 32, 2223-2253.
- Gozalo, P.L. (1997) Nonparametric Bootstrap Analysis with Applications to Demographic Effects in Demand Functions. *Journal of Econometrics* 81, 357-393.
- Hansen, B. (2006) Uniform Convergence Rates of Kernel Estimation with Dependent Data. Working paper, University of Wisconsin.
- Hansen, L.P. & J.A. Scheinkman (1995) Back to the Future: Generating Moment Implications for Continuous Time Markov Processes. *Econometrica* 63, 767-804.
- Hansen, L.P., J.A. Scheinkman & N. Touzi (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. & J. Knight (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.
- Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*, 2nd Edition. New York: Springer Verlag.
- Karlin, S. & H.M. Taylor (1981) *A Second Course in Stochastic Processes*. Academic Press.
- Kristensen, D. (2006a) Semiparametric Maximum-Likelihood Estimation of Diffusion Models. Working paper, Columbia University.
- Kristensen, D. (2006b) Modelling the Short-Term Interest Rate: A Semiparametric Approach. Working paper, Columbia University.

- Kristensen, D. & Y. Shin (2006) Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood. Working paper, University of Wisconsin.
- Meyn, S.P. & R.L. Tweedie (1993) Stability of Markovian processes III: Foster-Lyapunov Criteria for Continuous-time Processes. *Advances in Applied Probability* 25, 518-548.
- Nicolau, J. (2003) Bias Reduction in Nonparametric Diffusion Coefficient Estimation. *Econometric Theory* 19, 754-777.
- Politis, D.N. & J.P. Romano (1994) Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap. *Statistica Sinica* 4, 461-476.
- Pritsker, M.G. (1998) Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models. *Review of Financial Studies* 11, 449-487.
- Robinson, P.M. (1983) Nonparametric Estimators for Time Series. *Journal of Time Series Analysis* 4, 185-297.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- Vasicek, O. (1977) An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5, 177-188.
- Veretennikov, A.Y. (1997) On Polynomial Mixing Bounds for Stochastic Differential Equations. *Stochastic Processes and their Applications* 70, 115-127.
- Wong, E. (1964) The Construction of a Class of Stationary Markoff Processes. *Sixteenth Symposium in Applied Mathematics - Stoch. Proc. in Math. Physics and Engineering* (ed. R. Bellman), 264-276. Providence: American Mathematical Society.
- Yoshihara, K. (1976) Limiting Behaviour of U-Statistics for Stationary, Absolutely Regular Processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 35, 237-252.

A Proofs

Proof of Theorem 1. We first show pointwise weak convergence: For any given $x \in I$,

$$\begin{aligned}\hat{\mu}(x) - \mu(x) &= \frac{1}{2}\sigma^2(x; \theta_0) \left[\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)} \right] \\ &\quad + \frac{1}{2} \left[\partial_x \sigma^2(x; \hat{\theta}) - \partial_x \sigma^2(x; \theta_0) \right] + \frac{\hat{\pi}^{(1)}(x)}{2\hat{\pi}(x)} \left[\sigma^2(x; \hat{\theta}) - \sigma^2(x; \theta_0) \right] \\ &=: I_1(x) + I_2(x) + I_3(x).\end{aligned}$$

We have $I_k(x) = O_P(1/\sqrt{n})$, $k = 2, 3$, since, by (A.4),

$$\partial_x^i \sigma^2(x; \hat{\theta}) - \partial_x^i \sigma^2(x; \theta_0) = \partial_x^i \dot{\sigma}^2(x; \bar{\theta}_i) (\hat{\theta} - \theta_0) = O_P(1/\sqrt{n}),$$

for some $\bar{\theta}_i \in [\theta_0, \hat{\theta}]$, $i = 0, 1$. For the first term,

$$\begin{aligned}\sqrt{nh^3} \frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)} &= \frac{1}{\pi_0(x)} \sqrt{nh^3} [\hat{\pi}^{(1)}(x) - \pi_0^{(1)}(x)] \\ &\quad - \frac{\pi_0^{(1)}(x)}{\pi_0^2(x)} \sqrt{nh^3} [\hat{\pi}(x) - \pi_0(x)] \\ &\quad + \sqrt{nh^3} O(|\hat{\pi}^{(1)}(x) - \pi_0^{(1)}(x)|^2 + |\hat{\pi}(x) - \pi_0(x)|^2).\end{aligned}$$

Using standard methods for kernel estimators, see Robinson (1983), we obtain

$$\sqrt{nh^3} \{\hat{\pi}^{(1)}(x_i) - \pi_0^{(1)}(x_i)\}_{i=1}^N \xrightarrow{d} N\left(0, \text{diag}(\{V_\pi(x_i)\}_{i=1}^N)\right),$$

where $V_\pi(x) = \pi_0(x) \int K^{(1)}(z) dz$, while the two remainder terms are $o_P(1)$. The first part of the theorem now follows from Slutsky's Theorem.

To prove the uniform convergence result we introduce some additional trimming sets. Define

$$A(\varepsilon) = A_1(\varepsilon) \cap A_2(\varepsilon) \tag{19}$$

where

$$A_1(\varepsilon) = \{x | \hat{\pi}(x) \geq \varepsilon a\}, \quad A_2(\varepsilon) = \{x | \pi(x) \geq \varepsilon a\}, \tag{20}$$

for any $\varepsilon > 0$. As shown in Andrews (1994, p. 588), $A(\varepsilon) \supseteq A_1(2\varepsilon)$ with probability 1 as $n \rightarrow \infty$ under our conditions. Since

$$\sup_{x \in A_1(1)} |\mu(x; \hat{\pi}) - \mu(x; \pi_0)| \leq \sup_{x \in A(1/2)} |\mu(x; \hat{\pi}) - \mu(x; \pi_0)|,$$

we establish convergence uniformly over $A(1/2)$. We have

$$\sup_{x \in A(1/2)} |I_2(x)| \leq \frac{1}{2a} \left\{ \sup_{x \in A(1/2)} \pi_0(x) \|\partial_x \dot{\sigma}^2(x; \bar{\theta}_i)\| \right\} \|\hat{\theta} - \theta_0\| = O_P(a^{-1} n^{-1/2}),$$

$$\sup_{x \in A(1/2)} |I_3(x)| \leq \frac{C}{a^2} \sup_{x \in A(1/2)} \{ \pi_0(x) \|\dot{\sigma}^2(x; \bar{\theta}_i)\| \} \|\hat{\theta} - \theta_0\| = O_P(a^{-2}n^{-1/2}),$$

and

$$\sup_{x \in A(1/2)} |I_1(x)| \leq \frac{1}{2a} \left\{ \sup_{x \in A(1/2)} \pi_0(x) \sigma^2(x; \theta_0) \right\} \sup_{x \in A(1/2)} \left[\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)} \right]$$

where

$$\begin{aligned} & \sup_{x \in A(\varepsilon)} \left| \frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi^{(1)}(x)}{\pi(x)} \right| \\ & \leq \sup_{x \in A(\varepsilon)} \left\{ \hat{\pi}(x)^{-1} |\hat{\pi}^{(1)}(x) - \pi^{(1)}(x)| \right\} + \sup_{x \in A(\varepsilon)} |\pi^{(1)}(x)| \left| \frac{1}{\hat{\pi}(x)} - \frac{1}{\pi(x)} \right| \\ & \leq Ca^{-1} \|\hat{\pi}^{(1)} - \pi^{(1)}\|_\infty + Ca^{-2} \|\hat{\pi} - \pi\|_\infty. \end{aligned}$$

The result now follows from Lemma 7 with $s = 0, 1$. ■

Proof of Theorem 2. We have

$$\begin{aligned} \hat{\sigma}^2(x) - \sigma^2(x) &= \frac{2}{\hat{\pi}(x)} \frac{1}{n} \sum_{i=1}^n \left\{ \mu(X_i; \hat{\theta}) - \mu(X_i; \theta_0) \right\} \mathbb{I}\{X_i \leq x\} \\ &+ \frac{2}{\hat{\pi}(x)} \frac{1}{n} \sum_{i=1}^n \left\{ \mu(X_i; \theta_0) \mathbb{I}\{X_i \leq x\} - \int_l^x \mu(y; \theta_0) \pi(y) dy \right\} \\ &+ 2 \int_l^x \mu(y; \theta_0) \pi(y) dy \left\{ \frac{1}{\hat{\pi}(x)} - \frac{1}{\pi(x)} \right\} \\ &= : I_1(x) + I_2(x) + I_3(x), \end{aligned}$$

where $I_2(x) = O_P(n^{-1/2})$ by the CLT for mixing processes, c.f. Doukhan et al (1994), and

$$I_1(x) = \frac{2}{\hat{\pi}(x)} \left\{ \frac{1}{n} \sum_{i=1}^n \dot{\mu}(X_i; \bar{\theta}) \mathbb{I}\{X_i \leq x\} \right\} (\hat{\theta} - \theta_0) = O_P(n^{-1/2}).$$

For the third term, first note that

$$\frac{1}{\hat{\pi}(x)} - \frac{1}{\pi(x)} = -\frac{1}{\pi^2(x)} [\hat{\pi}(x) - \pi(x)] + \frac{[\hat{\pi}(x) - \pi(x)]^2}{4(\lambda \hat{\pi}(x) + (1-\lambda)\pi(x))^3},$$

for some $\lambda \in [0, 1]$. Using standard results for kernel estimators, see Robinson (1983), we obtain

$$\sqrt{nh} \{ \hat{\pi}(x_i) - \pi(x_i) \}_{i=1}^N \xrightarrow{d} N \left(0, \text{diag}(\{V_\pi(x_i)\}_{i=1}^N) \right),$$

where $V_\pi(x) = \pi(x) \int K(z)^2 dz$, while

$$\hat{\pi}(x) - \pi(x) = O_P(n^{-1/2}h^{-1}) + O_P(h^m).$$

Slutsky's Theorem now gives the claimed asymptotic distribution.

Recall the definition of $A(\varepsilon)$ in (19), and the results associated with this trimming set. Then

$$\begin{aligned} \sup_{x \in A(1/2)} |I_1(x)| &\leq \frac{2}{a} \left\{ \frac{1}{n} \sum_{i=1}^n \dot{\mu}(X_i; \bar{\theta}) \right\} \|\hat{\theta} - \theta_0\| = O_P(a^{-1}n^{-1/2}), \\ \sup_{x \in A(1/2)} |I_2(x)| &\leq \frac{2}{a} \frac{1}{n} \sum_{i=1}^n \left\{ \mu(X_i; \theta_0) \mathbb{I}\{X_i \leq x\} - \int_l^x \mu(y; \theta_0) \pi(y) dy \right\} = O_P(a^{-1}n^{-1/2}), \\ \sup_{x \in A(1/2)} |I_3(x)| &\leq 2 \int_l^r |\mu(y; \theta_0)| \pi(y) dy \left\{ \frac{1}{\hat{\pi}(x)} - \frac{1}{\pi(x)} \right\} \leq O(a^{-2}) \sup_{x \in A(1/2)} |\hat{\pi}(x) - \pi(x)|. \end{aligned}$$

The result now follows from Lemma 7 with $s = 0$. ■

Proof of Theorem 3. First consider $T_{1,n}$: Under H_0 , $\mu(x_i; \hat{\beta}) - \mu_0(x_i) = \dot{\mu}(x_i; \bar{\beta})(\hat{\beta} - \beta) = O_P(n^{-1/2})$ such that

$$\sqrt{nh^3} \frac{\hat{\mu}(x_i) - \mu(x_i; \hat{\beta})}{\hat{V}_\mu^{1/2}(x_i)} = \frac{\hat{V}_\mu^{1/2}(x_i)}{V_\mu^{1/2}(x_i)} \sqrt{nh^3} \frac{\hat{\mu}(x_i) - \mu_0(x_i)}{V_\mu^{1/2}(x_i)} + o_P(1) \xrightarrow{d} Z_i,$$

for $i = 1, \dots, N$, where $\{Z_i\}_{i=1}^n$ are i.i.d. standard Normal, c.f. Theorem 1. The result now follows by the continuous mapping theorem. The asymptotic distribution of $T_{2,n}$ is derived in a similar fashion. ■

Proof of Theorem 4. Define $\hat{\gamma} = (\hat{\pi}, \hat{\pi}^{(1)})$. We then have

$$nh^{5/2} \bar{T}_{1,n} = nh^{5/2} \int_I [G_1(x, \hat{\gamma}(x); \hat{\alpha}) - G_1(x, \tilde{\gamma}(x); \tilde{\alpha})]^2 \omega(x) dx,$$

where

$$G_1(x, y; \alpha) = \frac{\partial \sigma^2(x; \alpha)}{\partial x} + \frac{\sigma^2(x; \alpha) y_2}{2y_1}. \quad (21)$$

The result now follows from Lemma 10.

Since $n^{-1} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\} \mu(X_i; \hat{\beta}) = \int_l^x \mu(z; \beta) \pi(z) dz + O_P(n^{-1/2})$, we have

$$nh^{1/2} \bar{T}_{2,n} = nh^{1/2} \int_I [G_2(x, \hat{\pi}(x); \hat{\beta}) - G_2(x, \tilde{\pi}(x); \tilde{\beta})]^2 \omega(x) dx + o_P(1)$$

where

$$G_2(x, y; \beta) = \frac{2}{y} \int_l^x \mu(z; \beta) \pi(z) dz.$$

Again, Lemma 10 supplies us with the desired result. ■

Proof of Theorem 5. The result for $\tilde{T}_{1,n}$ follows from an application of Lemma 8 on G_1 defined in (21) such that

$$\begin{aligned}\Psi_i^1(x, h) &= \frac{\sigma^4(x) K^* \pi^{(1)}(x)^2}{4K^* \pi(x)^2} \frac{1}{h^2} \left\{ K\left(\frac{X_i - x}{h}\right) - E\left[K\left(\frac{X_i - x}{h}\right)\right] \right\} \\ &+ \frac{\sigma^4(x)}{4K^* \pi(x)} \frac{1}{h^2} \left\{ K^{(1)}\left(\frac{X_i - x}{h}\right) - E\left[K^{(1)}\left(\frac{X_i - x}{h}\right)\right] \right\} \\ &+ \frac{\sigma^4(x)}{4K^* \pi(x)} \frac{\partial [K^* \pi^{(1)}](x; \theta_0)}{\partial \theta} \begin{bmatrix} \psi_3(X_i | X_{i-1}) \\ \psi_4(X_i | X_{i-1}) \end{bmatrix} \\ &+ \frac{1}{2\pi(x)} \frac{\partial}{\partial x} \left[\frac{\partial \sigma^2(x; \alpha)}{\partial \alpha'} \pi(x) \right] \left\{ \psi_1(X_i | X_{i-1}) + \tilde{\psi}_1(X_i | X_{i-1}) \right\}.\end{aligned}\quad (22)$$

For $\tilde{T}_{2,n}$, we need to take into account the presence of $n^{-1} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\} \mu(X_i; \hat{\beta})$. Lemma 8 can easily be extended to allow for this yielding the claimed result with

$$\begin{aligned}\Psi_i^2(x, h) &= \frac{\sigma^4(x)}{K^* \pi(x)} \frac{1}{h} \left\{ K\left(\frac{X_i - x}{h}\right) - E\left[K\left(\frac{X_i - x}{h}\right)\right] \right\} \\ &+ \frac{2}{K^* \pi(x)} \left\{ \mathbb{I}\{X_i \leq x\} \mu(X_i) - \int_l^x \mu(y) \pi(y) dy \right\} \\ &+ \frac{\sigma^4(x)}{K^* \pi(x)} \frac{\partial [K^* \pi](x; \theta_0)}{\partial \theta'} \begin{bmatrix} \tilde{\psi}_1(X_i | X_{i-1}) \\ \tilde{\psi}_2(X_i | X_{i-1}) \end{bmatrix} \\ &+ \frac{2}{K^* \pi(x)} \int_l^x \frac{\partial \mu(y; \beta_0)}{\partial \beta'} \pi(y) dy \left\{ \psi_2(X_i | X_{i-1}) + \tilde{\psi}_2(X_i | X_{i-1}) \right\}.\end{aligned}\quad (23)$$

■

B Lemmas

We state the needed lemmas for a more general class of processes. We shall work under the following set of assumptions:

- C.1** $\{X_t\}$, $X_t \in \mathbb{R}^d$, is stationary and β -mixing with mixing coefficients satisfying $\beta_t = O(\rho^t)$ for some $0 < \rho < 1$.
- C.2** The marginal density $f : \mathbb{R}^d \mapsto \mathbb{R}$ has $s + m \geq 2$, derivatives which are bounded and uniformly continuous. For $t \geq 1$, the density of (X_0, X_t) , $f_t : \mathbb{R}^{2d} \mapsto \mathbb{R}$ is uniformly fourth-order differentiable and $\sup_{t,x,y} f_t(x, y) < \infty$.
- C.3** The kernel $K : \mathbb{R}^d \mapsto \mathbb{R}$ has s derivatives which satisfy

$$|D^\alpha K(z)| \leq C |z|^{-\eta}, \quad |D^\alpha K(z) - D^\alpha K(z')| \leq C |z - z'|,$$

for $|\alpha| = s$, and

$$\int z_1^{\alpha_1} \cdots z_d^{\alpha_d} K(z) dz = \begin{cases} 1, & |\alpha| = 0 \\ 0 & |\alpha| = 1, \dots, m-1, \\ C & |\alpha| = m \end{cases},$$

The first lemma give uniform convergence rate of the density derivative estimator:

Lemma 7 Under (C.1), (C.6)-(C.7), with $h = cn^{-\gamma}$, $0 < \gamma < 1$,

$$\sup_{x \in I} \left| D^\alpha \hat{f}(x) - D^\alpha f(x) \right| = O_P \left(\sqrt{\log(n)} n^{-1/2} h^{-(d+2s)/2} \right) + O_P(h^m),$$

for $|\alpha| = s \leq m$.

Proof. Using standard techniques, we obtain under (C.6)-(C.7) that for some $\bar{h} \in [0, h]$,

$$E[D^\alpha \hat{f}(x)] - D^\alpha f(x) = \frac{h^m}{m!} \sum_{|a|=m} \int D^{\alpha+|a|} f(x + z\bar{h}) z_1^{a_1} \cdots z_d^{a_d} K(z) dz = O(h^m)$$

uniformly in $x \in I$. Next, we define

$$\hat{G}(x) = \frac{1}{nh} \sum_{i=1}^n G\left(\frac{x - X_i}{h}\right), \quad G(z) = D^\alpha K(z).$$

It is easily checked that our choice of G satisfies Assumption 1 in Hansen (2005) under (C.6). We then obtain from Hansen (2005, Proof of Theorem 3) that

$$E \left[(\hat{G}(x) - E[\hat{G}(x)])^2 \right] = O_P \left(\log(n) n^{-1} h^{-d} \right)$$

uniformly in $x \in I$. Thus,

$$\begin{aligned} E \left[(D^\alpha \hat{f}(x) - D^\alpha f(x))^2 \right] &= \left\{ E[D^\alpha \hat{f}(x)] - D^\alpha f(x) \right\}^2 + h^{-2s} E \left[(\hat{G}(x) - E[\hat{G}(x)])^2 \right] \\ &= O(h^{2m}) + O_P \left(\log(n) n^{-1} h^{-(d+2s)} \right). \end{aligned}$$

■

Next, we give asymptotic results for statistics of the following form

$$\tilde{T}_n = \int_{\mathbb{R}^d} \left[G(x, \hat{\gamma}(x); \hat{\theta}) - G(x, K^* \tilde{\gamma}(x); \tilde{\theta}) \right]^2 \omega(x) dx,$$

and

$$\bar{T}_n = \int_{\mathbb{R}^d} \left[G(x, \hat{\gamma}(x); \hat{\theta}) - G(x, \tilde{\gamma}(x); \tilde{\theta}) \right]^2 \omega(x) dx,$$

where $\hat{\gamma}(x) = (\hat{\gamma}_1(x), \dots, \hat{\gamma}_k(x))$, $\tilde{\gamma}(x) = (\tilde{\gamma}_1(x), \dots, \tilde{\gamma}_k(x))$, and

$$\begin{aligned} \hat{\gamma}_i(x) &= \frac{1}{nh^d} \sum_{i=1}^n D^{\alpha_i} K\left(\frac{x - X_i}{h}\right), \\ \tilde{\gamma}_i(x) &= D^{\alpha_i} \tilde{f}(u), \quad \gamma_i(x) = D^{\alpha_i} f(u) \\ \tilde{f}(x) &= f(x; \tilde{\theta}) \text{ for some estimator } \tilde{\theta}. \end{aligned}$$

The functions $\omega : \mathbb{R}^d \mapsto \mathbb{R}_+$ and $G : \mathbb{R}^d \times \mathbb{R}^m \times \Theta \mapsto \mathbb{R}$ are known functions, and $\hat{\theta}, \tilde{\theta} \in \Theta \subseteq \mathbb{R}^l$ are parametric estimators. We shall work under the null,

$$H_0 : \exists \theta_0 \in \Theta : f(\cdot) = f(\cdot; \theta_0),$$

where equality holds in the almost sure sense w.r.t. the Lebesgue measure.

We derive two sets of asymptotic results. The first concerns the case where $h > 0$ is fixed, while the second the case where $h \rightarrow 0$ as $n \rightarrow \infty$. We make the following assumptions in addition to (C.1)-(C.3):

C.4 The parametric density $f(x; \theta)$ and its first two derivatives w.r.t. θ are bounded. The function $D^\alpha f(x; \theta)$ and its first two derivatives w.r.t. θ are bounded by a function $B(x)$ satisfying $\int_{\mathbb{R}^d} B(x) dx < \infty$.

C.5 $\hat{\theta} = \theta_0 + \frac{1}{n} \sum_{i=1}^n \psi(X_i|X_{i-1}) + o_P(1/\sqrt{n})$ and $\tilde{\theta} = \theta_0 + \frac{1}{n} \sum_{i=1}^n \tilde{\psi}(X_i|X_{i-1}) + o_P(1/\sqrt{n})$, where $E[\psi(X_1|X_0)] = E[\tilde{\psi}(X_1|X_0)] = 0$, $E[|\psi(X_1|X_0)|^{2+\delta}] < \infty$ and $E[|\tilde{\psi}(X_1|X_0)|^{2+\delta}] < \infty$.

C.6 The function $G(x, y; \theta)$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^k$, $\theta \in \Theta$, is twice continuously differentiable in its 2nd and 3rd argument

C.7 $\omega(x)$ has bounded support.

We obtain the following result for the case where the bandwidth $h > 0$ is fixed as $n \rightarrow \infty$:

Lemma 8 Under (C.1)-(C.7), for any fixed $h > 0$,

$$n\tilde{T}_n \rightarrow^d \int_{\mathbb{R}^d} Z_h^2(x) \omega(x) dx,$$

where $Z_h \sim N(0, \Sigma_h)$ and Σ_h satisfies for any function $a \in L_2(\omega) = \{g | \int_{\mathbb{R}^d} g^2(x) \omega(x) dx < \infty\}$,

$$\langle \Sigma_h, a \rangle = \text{Var}(\langle \Psi_{0,h}, a \rangle) + 2 \sum_{i=1}^{\infty} \text{Cov}(\langle \Psi_{0,h}, a \rangle, \langle \Psi_{i,h}, a \rangle),$$

$$\langle a, b \rangle = \int_{\mathbb{R}^d} a(x) b(x) \omega(x) dx,$$

with

$$\begin{aligned} \Psi_{i,h}(x) &= \sum_{i=1}^k \frac{\partial G(x, K^* \gamma(x); \theta_0)}{\partial y_i} \frac{1}{h^{d+|\alpha_i|}} \left\{ D^{\alpha_i} K\left(\frac{X_i - x}{h}\right) - E\left[D^{\alpha_i} K\left(\frac{X_i - x}{h}\right)\right] \right\} \\ &+ \frac{\partial G(x, K^* \gamma(x); \theta_0)}{\partial y} \frac{\partial [K^* \gamma(\cdot; \theta)](x)}{\partial \theta'} \psi(X_i|X_{i-1}) \\ &+ \frac{\partial G(x, K^* \gamma(x); \theta_0)}{\partial \theta'} \left\{ \psi(X_i|X_{i-1}) + \tilde{\psi}(X_i|X_{i-1}) \right\}. \end{aligned}$$

Proof. By a 2nd order Taylor expansion,

$$\begin{aligned}
G(x, y; \theta) - G(x, y_0; \theta_0) &= \frac{\partial G(x, y_0; \theta_0)}{\partial y} [y - y_0] + \frac{\partial G(x, y_0; \theta_0)}{\partial \alpha} [\theta - \theta_0] \\
&+ \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 G(x, \bar{y}; \bar{\theta})}{\partial y_i \partial y_j} [y_i - y_{0,i}] [y_j - y_{0,j}] \\
&+ \frac{1}{2} \sum_{i,j=1}^l \frac{\partial^2 G(x, \bar{y}; \bar{\theta})}{\partial \theta_i \partial \theta_j} [\theta_i - \theta_{0,i}] [\theta_j - \theta_{0,j}] \\
&+ \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^l \frac{\partial^2 G(x, \bar{y}; \bar{\theta})}{\partial y_i \partial \theta_j} [y_i - y_{0,i}] [\theta_j - \theta_{0,j}]
\end{aligned}$$

for some $\bar{y}$ between y and y_0 and $\bar{\theta}$ between θ and θ_0 . We then apply this to both $G(x, \hat{\gamma}(x); \theta_0)$ and $G(x, K^* \tilde{\gamma}(x); \theta_0)$, and collect all the first order terms in $Z_n(x) = \sum_{i=1}^n \Psi_i(x) / \sqrt{n}$ where $\Psi_i(x)$ is given in the lemma.

The space $L_2(\omega)$ defined in the lemma is a Hilbert space. Since ω has bounded support and $\Psi_i(x)$ is continuous a.s., we easily see that $Z_n \in L_2(\omega)$ a.s. Then, by Politis and Romano (1994, Theorem 2.3), $Z_n \rightarrow^d Z$ in $L_2(\omega)$ where Z has been defined in the lemma. We claim that

$$n \int_{\mathbb{R}^d} \left[G(x, \hat{\gamma}(x); \hat{\theta}) - G(x, K^* \tilde{\gamma}(x); \tilde{\theta}) \right]^2 \omega(x) dx = \int_{\mathbb{R}^d} Z_n^2(x) \omega(x) dx + o_P(1).$$

If this holds, then the result follows by the continuous mapping theorem since $g \mapsto \int_{\mathbb{R}^d} g^2(x) \omega(x) dx$ is a continuous functional on $L_2(\omega)$.

The claim will hold if we can show that all squared higher order terms in the 2nd order Taylor expansions are $o_P(1)$. We do this only for the first one; the remaining ones follow by the same argument:

$$\begin{aligned}
& n \int_{\mathbb{R}^d} \left\| \frac{\partial G(x, D^\alpha \bar{f}(x); \bar{\theta})}{\partial y_i \partial y'_j} \right\|^2 \|\hat{\gamma}(x) - K^* \gamma(x)\|^4 \omega(x) dx \\
&= \frac{1}{n} \int_{\mathbb{R}^d} \left\| \frac{\partial G(x, D^\alpha \bar{f}(x); \bar{\theta})}{\partial y_i \partial y'_j} \right\|^2 \|\tilde{Z}_n(x)\|^4 \omega(x) dx \\
&\leq \frac{B}{n} \int_{\mathbb{R}^d} \|\tilde{Z}_n(x)\|^4 \omega(x) dx \\
&= \frac{1}{n} \times O_P(1),
\end{aligned}$$

since, again using Politis and Romano (1994, Theorem 2.3), $\tilde{Z}_n = (\tilde{Z}_{n,1}, \dots, \tilde{Z}_{n,k})'$ given by

$$\tilde{Z}_{n,j}(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ D^{\alpha_j} K \left(\frac{X_i - x}{h} \right) - E \left[D^{\alpha_j} K \left(\frac{X_i - x}{h} \right) \right] \right\}$$

weakly converges in $L_2(\omega)$ towards $\tilde{Z}$ where $\tilde{Z}$ is a Gaussian process. ■

Next, we consider the asymptotics as $h \rightarrow 0$. We first consider the case where the parametric density is smoothed.

Lemma 9 *Under (C.1)-(C.6), for any bounded function $w : \mathbb{R}^d \mapsto \mathbb{R}_+$ and $|\alpha| = s \geq 0$,*

$$nh^{d/2+2s} \left\{ \int_{\mathbb{R}^d} \left[D^\alpha \hat{f}(x) - K^* D^\alpha \tilde{f}(x) \right]^2 w(x) dx - c_\alpha / (nh^{d+2s}) \right\} \rightarrow^d N(0, \sigma_\alpha^2),$$

as $nh^{d+2s} \rightarrow \infty$, where

$$\begin{aligned} c_\alpha &= \left\{ \int_{\mathbb{R}^d} D^\alpha K(z)^2 dz \right\} \left\{ \int_{\mathbb{R}^d} f(x) w(x) dx \right\}, \\ \sigma_\alpha^2 &= 2 \left\{ \int_{\mathbb{R}^d} [D^\alpha K * D^\alpha K]^2(z) w(z) dz \right\} \left\{ \int_{\mathbb{R}^d} f^2(x) w(x) dx \right\}. \end{aligned}$$

If furthermore $nh^{d/2+2s+2m} \rightarrow 0$,

$$nh^{d/2+2s} \left\{ \int_{\mathbb{R}^d} \left[D^\alpha \hat{f}(x) - D^\alpha \tilde{f}(x) \right]^2 w(x) dx - c_\alpha / (nh^{d+2s}) \right\} \rightarrow^d N(0, \sigma_\alpha^2).$$

Proof. We show this by an extension of Fan and Ullah (1999, Theorem 4.1), applying the U-statistics result in Fan and Li (1994). We have that

$$\begin{aligned} & \int_{\mathbb{R}^d} \left[D^\alpha \hat{f}(x) - K^* D^\alpha \tilde{f}(x) \right]^2 w(x) dx \\ &= \int_{\mathbb{R}^d} \left[D^\alpha \hat{f}(x) - K^* D^\alpha f(x) \right]^2 w(x) dx \\ & \quad + \int_{\mathbb{R}^d} \left[K^* D^\alpha f(x) - K^* D^\alpha \tilde{f}(x) \right]^2 w(x) dx \\ & \quad + 2 \int_{\mathbb{R}^d} \left[D^\alpha \hat{f}(x) - K^* D^\alpha f(x) \right] \left[K^* D^\alpha f(x) - K^* D^\alpha \tilde{f}(x) \right] w(x) dx \\ &= : I_{n,1} + I_{n,2} + I_{n,3}. \end{aligned}$$

First,

$$I_{n,1} = \frac{2}{n^2 h^{2(d+s)}} \sum_{1 \leq i < j \leq n} H_n(X_i, X_j) + \frac{1}{n^2 h^{2(d+s)}} \sum_{i=1}^n G_n(X_i),$$

where

$$\begin{aligned} H_n(u, v) &= \int_{\mathbb{R}^d} \{ \bar{K}_n(x, u) - E[\bar{K}_n(x, X_1)] \} \{ \bar{K}_n(x, v) - E[\bar{K}_n(x, X_1)] \} w(x) dx, \\ G_n(u) &= \int_{\mathbb{R}^d} \{ \bar{K}_n(x, u) - E[\bar{K}_n(x, X_1)] \}^2 w(x) dx, \end{aligned}$$

and $\bar{K}_n(x, y) = D^\alpha K((x - y)/h)$. We easily see that $H_n(u, v)$ is a symmetric function with $E[H_n(u, X_1)] = 0$, and we can therefore apply Fan and Li (1994, Theorem

2.1) on $U_n := \sum_{1 \leq i < j \leq n} H_n(X_i, X_j)$. Observe that $H_n(X_i, X_j)$ takes the same form as the function $\bar{H}_n(X_i, X_j)$ in Fan and Ullah (1999, Eq. (17)) except that our kernel $\bar{K}_n(z)$ is given as $D^\alpha K_n(z/h)$ and we have a weighting function $w(x)$. One can follow the steps in Fan and Ullah (1999, Proof of Theorem 4.1) to show that under (C.1)-(C.7), $\sqrt{2}U_n/(n\sigma_n) \rightarrow^d N(0, 1)$ as $nh^\gamma \rightarrow 0$ for some $\gamma > d + 2s$, where

$$\begin{aligned}
\sigma_n^2 & : = E \left[H_n^2(\tilde{X}_1, \tilde{X}_2) \right] \\
& \sim \int \int E \left[\bar{K}_n(x, \tilde{X}_1) \bar{K}_n(y, \tilde{X}_1) \right]^2 w(x) w(y) dx dy \\
& \sim \int \int \left\{ \int \bar{K}_n(x, z) \bar{K}_n(y, z) f(z) dz \right\}^2 w(x) w(y) dx dy \\
& \sim h^{3d} \int \int \left\{ \int D^\alpha K(u) D^\alpha K(u+v) f(x-uh) du \right\}^2 w(x) w(v) dx dv \\
& \sim h^{3d} \int \int \left\{ \int D^\alpha K(u) D^\alpha K(u+v) du \right\}^2 f(x)^2 w(x) w(v) dx dv \\
& \sim h^{3d} \sigma_\alpha^2.
\end{aligned}$$

Thus,

$$nh^{d/2+2s} \left\{ \frac{2}{n^2 h^{2(d+s)}} \sum_{1 \leq i < j \leq n} H_n(X_i, X_j) \right\} \rightarrow^d N(0, \sigma_\alpha^2).$$

For the second term, one can show, again following the arguments in Fan and Ullah (1999, Proof of Theorem 4.1), that

$$\begin{aligned}
& \frac{1}{n^2 h^{2(d+s)}} \sum_{i=1}^n G_n(X_i) \\
& = \frac{1}{n^2 h^{d+2s}} \sum_{i=1}^n \int_{\mathbb{R}^d} \frac{1}{h^d} \{ \bar{K}_n(x, X_i) - E[\bar{K}_n(x, X_1)] \}^2 w(x) dx \\
& = \frac{1}{n h^{d+2s}} \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \frac{1}{h^d} D^\alpha K\left(\frac{x-u}{h}\right)^2 f(u) du \right\} w(x) dx + o_P\left(\frac{1}{n h^{d+2s}}\right) \\
& = \frac{1}{n h^{d+2s}} \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} D^\alpha K(z)^2 f(x+zh) dz \right\} w(x) dx + o_P\left(\frac{1}{n h^{d+2s}}\right) \\
& = \frac{1}{n h^{d+2s}} c_\alpha + o_P\left(\frac{1}{n h^{d+2s}}\right).
\end{aligned}$$

Following the same arguments as in Fan and Ullah (1999, Proof of Theorem 4.1), one can show that $nh^{d/2+2s}I_{n,k} = o_P(1)$, $k = 2, 3$. This yields the first result.

To show the second result, we have to account for an additional bias term, $I_{n,4}$, due to the non-centering of $D^\alpha \tilde{f}(x)$. This is given by

$$I_{n,4} := \int_{\mathbb{R}^d} \left[K^* D^\alpha \tilde{f}(x) - D^\alpha \tilde{f}(x) \right]^2 w(x) dx = O_P(h^{2m}).$$

The additional requirement ensures that this vanishes asymptotically. ■

Lemma 10 Assume (C.1)-(C.7) hold and $|\alpha_k| = s \geq 0$ while $|\alpha_i| < s$, $i = 1, \dots, k-1$. Then,

$$nh^{d/2+2s} \left\{ \tilde{T}_n - c_{\alpha_k} / \left(nh^{d+2s} \right) \right\} \rightarrow^d N(0, \sigma_{\alpha_k}^2),$$

as $nh^{d+2s} \rightarrow \infty$, where c_α and σ_α^2 are given in Lemma 9 with

$$w(x) = \left[\frac{\partial G(x, K^* \gamma(x); \alpha_0)}{\partial y_m} \right]^2 \omega(x).$$

If furthermore $nh^{d+2s+2m} \rightarrow 0$,

$$nh^{d/2+2s} \left\{ \bar{T}_n - c_{\alpha_m} / \left(nh^{d+2s} \right) \right\} \rightarrow^d N(0, \sigma_{\alpha_m}^2),$$

where now

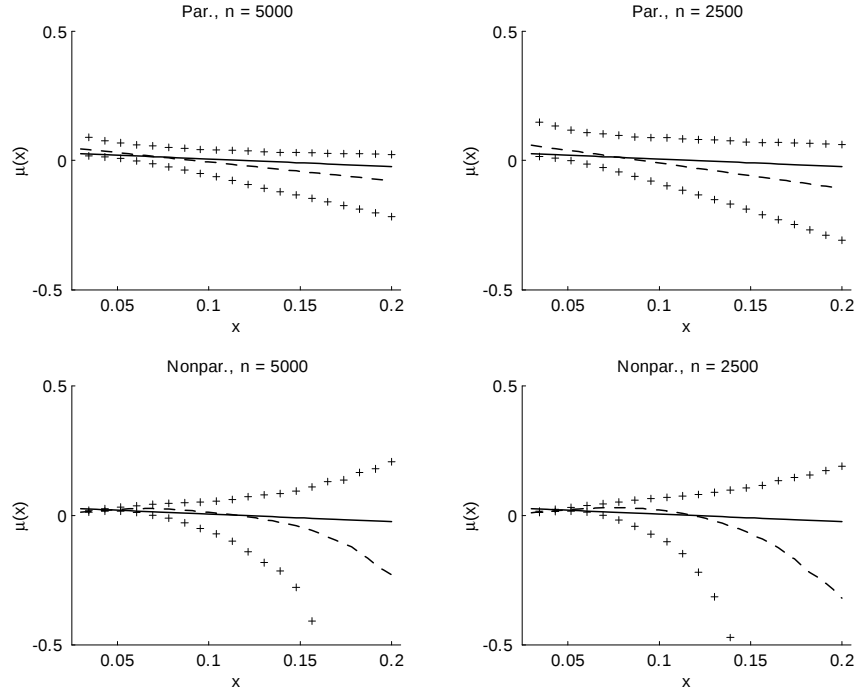
$$w(x) = \left[\frac{\partial G(x, \gamma(x); \alpha_0)}{\partial y_m} \right]^2 \omega(x).$$

Proof. We make the exact same expansions as in the proof of Lemma 8, and apply Lemma 9 on each linear term. We see that all terms associated with lower order derivatives α_i , $i = 1, \dots, m-1$, of the kernel estimator and the parametric components will disappear at a faster rate, and so the asymptotics will only be driven by the kernel estimator with the highest derivative, α_k . That is,

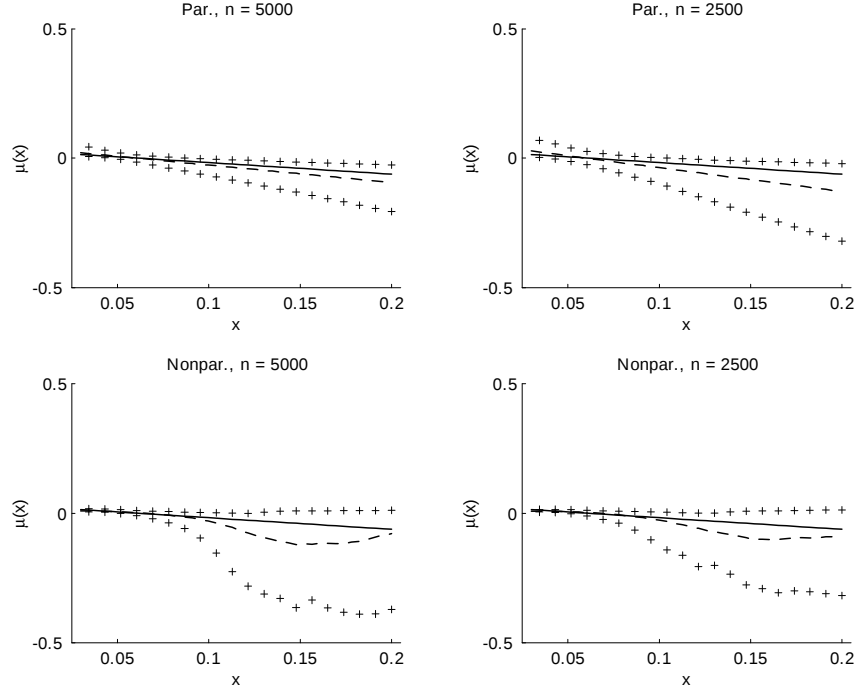
$$\begin{aligned} & nh^{d/2+2s} \left\{ \tilde{T}_n - c_{\alpha_k} / \left(nh^{d+2s} \right) \right\} \\ = & nh^{d/2+2s} \left\{ \int_{\mathbb{R}^d} \left[D^{\alpha_k} \hat{f}(x) - K^* D^{\alpha_k} \tilde{f}(x) \right]^2 w(x) dx - c_{\alpha_k} / \left(nh^{d+2s} \right) \right\} + o_P(1) \end{aligned}$$

with w given in the lemma, and similar for $\bar{T}_n$. ■

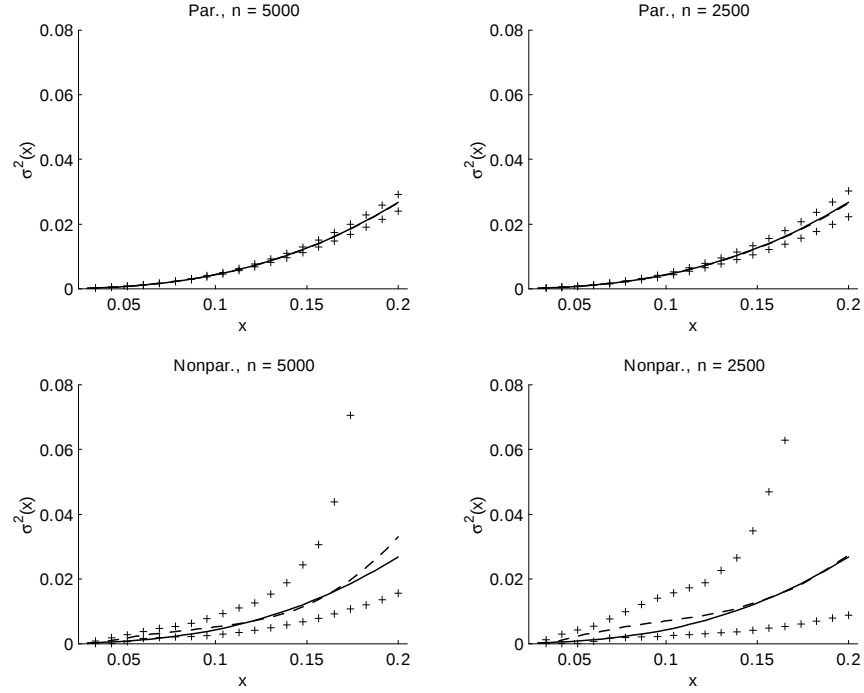
C Figures



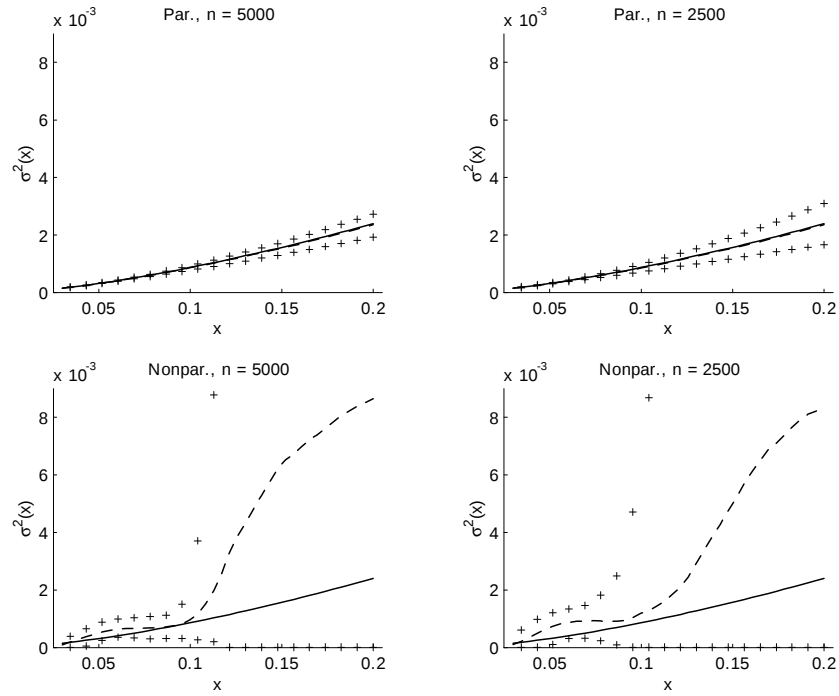
Estimates of $\mu(x)$ for the CKLS(i) model.



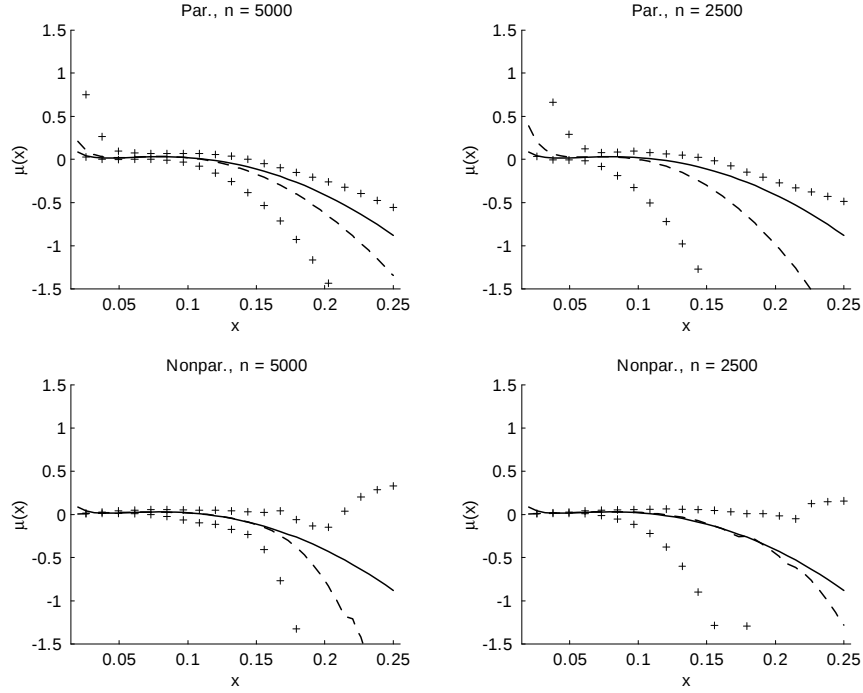
Estimates of $\mu(x)$ for the CKLS(ii) model.



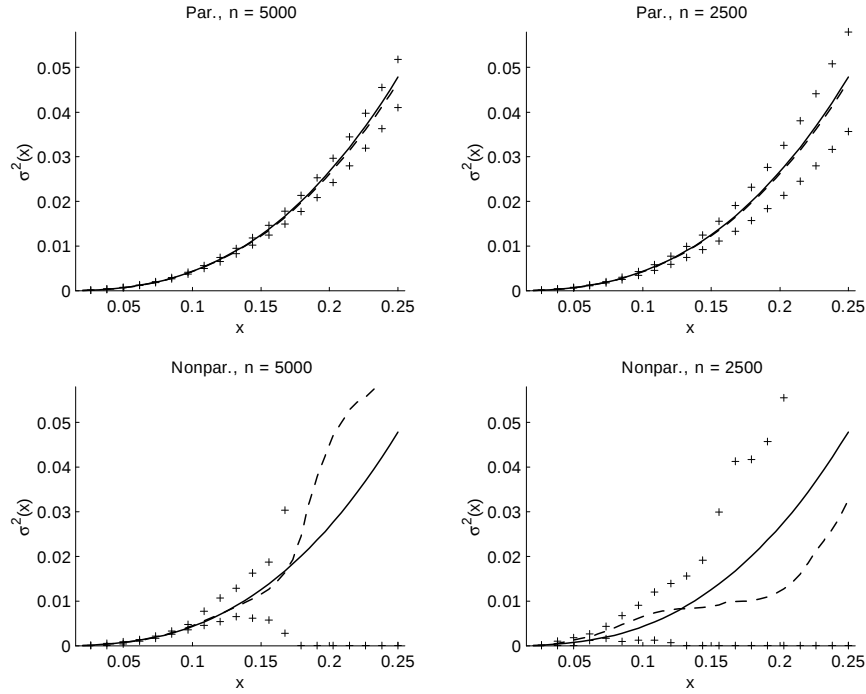
Estimates of $\sigma^2(x)$ for the CKLS(i) model.



Estimates of $\sigma^2(x)$ for the CKLS(ii) model.



Estimates of $\mu(x)$ for the AS model.



Estimates of $\sigma^2(x)$ for the AS model.

Research Papers 2007



2007-1 Dennis Kristensen: Nonparametric Estimation and Misspecification
Testing of Diffusion Models