CREATES Research Paper 2010-43

Semi-Nonparametric Estimation and Misspecification Testing of Diffusion Models

Dennis Kristensen

Semi-Nonparametric Estimation and Misspecification Testing of Diffusion Models

Dennis Kristensen*
Columbia University and CREATES†

August 2010

Abstract

Novel transition-based misspecification tests of semiparametric and fully parametric univariate diffusion models based on the estimators developed in Kristensen (*Journal of Econometrics*, 2010) are proposed. It is demonstrated that transition-based tests in general lack power in detecting local departures from the null since they integrate out local features of the drift and volatility. As a solution to this, tests that directly compare drift and volatility estimators under the relevant null and alternative are also developed which exhibit better power against local alternatives.

Keywords: Diffusion process; kernel estimation; nonparametric; specification testing; semiparametric; transition density.

JEL-Classification: C12, C13, C14, C22.

*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

In this study, we develop semi-nonparametric estimators and misspecification tests of the so-called drift and diffusion functions in univariate diffusion models given low-frequency observations. The proposed estimators and tests provide the researcher with tools to investigate whether a given parametric specification of the drift and diffusion function is correct and allows him to test drift and diffusion specifications separately from each other. This is in contrast to existing methods found in the literature which simultaneously test correct specification of drift and diffusion terms.

Our estimation and testing procedure takes as starting point two classes of semiparametric diffusion models introduced in Kristensen (2010): In the first class, the drift term is known up to a finite-dimensional parameter while the diffusion term is left unspecified; in the second class, the diffusion term is on parametric form while the drift term is unknown. Kristensen (2010) develop estimators of the parametric component for a given model in either of the two classes. We demonstrate how the unspecified term in any of these semiparametric diffusion models can be estimated nonparametrically using kernel methods. These estimators are useful as guides in the search for a correct parametric specification since they provide information about the shape of the unspecified term. In addition, the estimators help us to develop novel misspecification tests of diffusion models.

We suggest two sets of tests: First, we propose tests for a given semiparametric diffusion model against a fully nonparametric alternative. Second, tests for a fully parametric model against either of its two semiparametric alternatives are developed. Our test statistics are chosen as weighted L_2-distances of the estimators of the so-called transition density obtained under null and alternative respectively. In addition, we also consider tests that directly compare drift or diffusion estimators. We explore the asymptotic properties of the tests both under null and alternative, and obtain a number of interesting results:

First, our transition-based test of a given semiparametric model against the fully nonparametric alternative is under the null first-order asymptotically equivalent to tests of fully parametric models as developed in Aït-Sahalia, Fan and Peng (2009) and Li and Tkacz (2006). This is due to the fact that estimators of the transition density under the semiparametric and parametric null respectively both converge with parametric rate, and as such the asymptotic distributions of our test statistics are completely driven by the fully nonparametric transition density estimator. The parametric rate of the semiparametric transition density estimator appears because computation of transition densities for low-frequency observations involves integration of both the drift and diffusion term (see e.g. Kristensen, 2008). This integration functions as an additional smoothing mechanism that speeds up convergence rate of the semiparametric estimator of the transition density even though it involves kernel estimators.

Second, our proposed transition-based test statistics of the fully parametric model against either of the two semiparametric alternatives converge with parametric rate under the null. This is non-standard within the class of tests based on L_2-distance measures of semi-nonparametric density estimators which in general converge with nonparametric rate. Instead our transition-based tests for the fully parametric null share similarities with the class semiparametric estimators and tests that exhibit parametric rate (see e.g. Andrews, 1994; Corradi and Swanson, 2005; Whang and Andrews, 1993).

Third, we study the power properties of the tests by considering their performance under contiguous alternatives. In particular, we show that, transition-based tests are not very suitable for detection of high-frequency departures in a diffusion framework. This is in contrast to density-based tests in standard, discrete-time setting. This maybe surprising result is due to the fact that local features of the drift and diffusion terms are integrated out in the computation of trasition densities and so local deviations get blurred out. It should be stressed that this problem is not special to our particular tests, but is shared by all other transition-based tests of diffusion models in the literature such as Aït-Sahalia et al (2009). As such our power analysis should be of general interest.

The lack of power against local alternatives leads us to propose two alternative tests of the parametric null against semiparametric alternatives based on direct comparison of drift and diffusion function estimators obtained under null and alternatives. We examine their asymptotic properties both under null and alternative: They converge with a slower rate than the transition-based tests, and thus are dominated by transition-based tests in terms of detecting global alternatives. On the other hand, the tests are better at detecting local deviations of drift and diffusion functions from the null, and so have better power against local alternatives. As such they complement our transition-based tests.

Finally, we conduct a higher-order analysis of the proposed tests under the null. This analysis demonstrates that first-order asymptotic distributions obtained under the null may be a poor proxy of their finite-sample distributions. We therefore propose a Markov bootstrap method that we hope will provide a better approximation of finite-sample distributions of the test statistics. This conjecture is supported by simulation results in Aït-Sahlia et al (2009) and Li and Tkacz (2006) who propose similar Bootstrap procedures for their tests.

The proposed tests and their theoretical analysis add to a growing literature on specification testing of diffusion models. This class of models is widely used in describing dynamics of asset pricing variables such as interest rates, stock prices, and exchange rates; see for example Björk (2004) for an overview. Since economic theory imposes little restrictions on asset price dynamics, statistical techniques are usually employed in the search for a correct specification. The literature on testing diffusion model specifications can roughly be divided up into two categories depending on whether high-frequency data is assumed available or not.

If high-frequency data is observed, simple nonparametric kernel-regression estimators of drift and diffusion terms can be used to test for correct specification (Bandi and Phillips, 2005; Corradi and White, 1999; Li, 2007; Negri and Nishiyama, 2009). In principle, these tests do not rely on stationarity which is an advantage over the approach taken here. On the other hand, asymptotic properties of estimators and associated tests do rely on the time distance between observations shrinking to zero; thus, estimators and tests will potentially be severely biased if only low-frequency data is available (see Nicolau, 2003).

To avoid the bias issues associated with high-frequency based tests, alternative tests based on fixed time distance between observations have been developed. Aït-Sahalia (1996b) propose to test for correct specification using a weighted L_2-distance to measure discrepancies between the marginal density under null and alternative. This class of tests was originally proposed in Bickel and Rosenthal (1973) in a cross-sectional setting; see also Fan (1994) and Gourieroux and Tenreiro (2001). Since the test of Aït-Sahalia (1996b) is only able to detect discrepancies in the marginal density, it is not consistent against all alternatives. This observation lead to the development of tests based on transition densities since these fully characterise diffusion models.

Our transition-based tests are most related to the ones developed in Aït-Sahalia et al (2009) and Li and Tkacz (2006) where fully nonparametric and parametric estimators of the transition density are compared. In a similar spirit, Hong and Li (2004) propose a test where transformed versions of the transition densities are compared, while Chen, Gao and Tang (2009) employ empirical likelihood techniques. These tests are all designed to examine the correct parametric specification of the drift and diffusion function jointly. In contrast, we are able to test the specification of each of the two functions characterising the model separately. Our local power analysis complements the one carried out in Aït-Sahalia et al (2009). They specify alternatives in terms of the transition densities and find that transition-based tests have the ability to detect local deviations form the null at a better rate than CvM type tests. However, given that the end goal is to test for the correct specification of drift and diffusion term, we instead specify our alternatives directly in terms of these. By doing so, we obtain some rather different power results for transition-based tests. In particular, we show that they are not able to detect local alternatives at a higher rate compared to CvM type tests. These seemingly contradictory results are due to the fact that Aït-Sahalia et al (2009) specify their alternatives in terms of the transition density while we focus on deviations in terms of underlying drift and diffusion functions. Since, as already noted above, the transition density involves integration over the drift and diffusion function, local features in these get smoothed out in the transition density and therefore not easily detected.

Our tests based on direct comparison of the drift and diffusion function estimates under null and alternative are related to the marginal density tests of Aït-Sahalia (1996b) and Huang (1997). However, our proposed tests involve non-trivial transformations of the marginal density and its derivatives and as such are able to detect different, more natural alternatives compared to their tests.

Instead of comparing transition densities, Kolmogorov-Smirnoff (KS) type tests have been proposed by Bhardwaj, Corradi and Swanson (2008) and Corradi and Swanson (2005) where estimators of the cumulative distribution functions (cdf's) are compared. This on one hand means that their tests converge with parametric rate under the null and as such are more powerful at detecting certain global alternatives compared to transition-based tests. On the other hand KS-type tests are known to have difficulties detecting local deviations from the null; a shortcoming that density-based tests do not suffer from (see e.g. Escanciano, 2009; Eubank and LaRiccia, 1992).

Finally, Kristensen (2010) proposes some specification tests which appear to be the only existing tests based on low-frequency data that allow for testing correct specifications of the drift and diffusion terms separately. However, Kristensen (2010) does not supply a complete asymptotic theory. Moreover, as with CvM and KS type tests, his proposed Hausmann-type tests of fully parametric models will in general have low power against local alternatives since they are based on only matching estimators of the parametric component obtained under null and under alternatives. In particular, his tests may not be consistent against all alternatives. In contrast, we base our tests on estimators of the nonparametric component under the alternative, and so expect them to enjoy better power properties.

The remains of the paper is organised as follows: In Section 2, we lay out the general framework of our analyis. The semi-nonparametric estimators of the drift and diffusion term are presented and their asymptotic properties derived in Section 3. In Section 4, we propose a number of different test statistics for a parametric specification against semi- and nonparametric alternatives and investigate their asymptotic behaviour. We discuss related tests in Section 5, while Bootstrap versions of the test statistics are developed in Section 6. The finite-sample performance of the estimators are examined through a simulation study in Section 7. We conclude in Section 8. All proofs have been relegated to the Appendix.

2 Framework

Consider the continuous time process $\{X_t\} = \{X_t : t \geq 0\}$ solving the following univariate Markov diffusion model,

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{1}$$

where $\{W_t\}$ is a standard Brownian motion. The domain of $\{X_t\}$ takes the form of an open interval $I = (l, r)$ 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. The dynamics of the process are described by the transition densities $p(y|x;t)$, $t \geq 0$, describing conditional distributions,

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

For diffusion models as given in eq. (1), the transition density can be expressed as the solution to the following partial differential equation (PDE) (see Friedman, 1976):

$$\frac{\partial p(y|x;t)}{\partial t} = \mathcal{A}\left[\mu,\sigma^2\right] p(y|x;t), \quad t>0, \ (x,y)\in I\times I, \tag{2}$$

with boundary condition $\lim_{t\to 0} p_t(y|x) = \delta(y-x)$. Here, $\mathcal{A}\left[\mu,\sigma^2\right]$ denotes the infinitesimal generator,

$$\mathcal{A}\left[\mu,\sigma^2\right] p(y|x;t) = \mu(x)\frac{\partial p(y|x;t)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 p(y|x;t)}{\partial x^2},$$

and $\delta(\cdot)$ Dirac's delta function. Thus, the drift and diffusion function fully characterise the transition density and we will write $p\left(y|x;t,\mu,\sigma^2\right)$ for the solution mapping that takes any drift and diffusion function into the corresponding transition density as given implicitly through the PDE in eq. (2).

We are interested in testing parametric specifications of the drift and diffusion function. We will throughout work under the maintained (nonparametric) hypothesis that $\{X_t\}$ is a Markov diffusion process,

$$H_{\mathrm{NP}}: \{X_t\} \text{ solves eq. (1) with } \sigma^2(\cdot) \text{ and } \mu(\cdot) \text{ unspecified.}$$

In the existing literature, tests have been developed for a fully parametric diffusion specification against this nonparametric alternative. The joint fully parametric hypothesis takes the form

$$H_{\mathrm{P}}: \sigma^2(\cdot) = \sigma^2(\cdot;\theta_{0,1}) \text{ and } \mu(\cdot) = \mu(\cdot;\theta_{0,2}) \text{ for some } (\theta_{0,1},\theta_{0,2})\in\Theta_1\times\Theta_2,$$

where $\Theta_k \subseteq \mathbb{R}^{d_k}$, $k=1,2$. Thus, under H_{P}, both drift and diffusion functions are known up to some finite-dimensional parameter. A plethora of tests of H_{P} vs. H_{NP} exist; see, for example, Aït-Sahalia et al (2009), Bhardwaj et al (2008), Chen et al (2009), Hong and Li (2004) and Li and Tkacz (2006). Most of these studies base their tests on comparison of estimators of (potentially transformed versions of) the transition density under the null and the alternative.

However, in case of rejection of H_{P}, such tests are not informative regarding whether misspecification of drift, diffusion or both are the cause of rejection. This motivates us to introduce the following two semiparametric hypotheses, which allow us to test for misspecification of the drift and diffusion term separately from each other:

$$H_{\mathrm{SP},1}: \sigma^2(\cdot) = \sigma^2(\cdot;\theta_{0,1}) \text{ for some } \theta_{0,1}\in\Theta_1, \tag{3}$$

and

$$H_{\mathrm{SP},2}: \mu(\cdot) = \mu(\cdot;\theta_{0,2}) \text{ for some } \theta_{0,2}\in\Theta_2. \tag{4}$$

If a model satisfy $H_{\mathrm{SP},1}$ ($H_{\mathrm{SP},2}$), the drift (diffusion) term is unspecified, and the model is semiparametric. Also note that if a model satisfies both $H_{\mathrm{SP},1}$ and $H_{\mathrm{SP},2}$, then both drift and diffusion are specified and the model is fully parametric. In particular, we have the following nesting of the hypotheses: $H_{\mathrm{P}} \subseteq H_{\mathrm{SP},k} \subseteq H_{\mathrm{NP}}$ for $k=1,2$.

In the next section, we first develop tests of each of the two semiparametric hypotheses, $H_{\mathrm{SP},1}$ and $H_{\mathrm{SP},2}$, against the nonparametric alternative. Secondly, we propose tests of H_{P} against each of the two semiparametric hypotheses. Together, the tests enable the econometrician to first test for the correct specification of, say, the drift term ($H_{\mathrm{SP},2}$ vs. H_{NP}), and then (if $H_{\mathrm{SP},2}$ is accepted) the correct specification of the diffusion term (H_{P} vs. $H_{\mathrm{SP},2}$).

In order to develop our tests, we first obtain estimators of the drift and diffusion functions under the two semiparametric hypotheses. The 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. This density satisfies $\int_A \pi(x)\,dx = P(X_t \in A)$, for any $t \geq 0$ and Borel set $A \subseteq I$, and can be written on the following form:

$$\pi(x) = \frac{M_{x^*}}{\sigma^2(x)} \exp\left[2\int_{x^*}^{x} \frac{\mu(y)}{\sigma^2(y)} dy\right], \tag{5}$$

for some some point $x^* \in \mathrm{int}I$, and normalisation factor $M_{x^*} > 0$, c.f. Karlin and Taylor (1981, Section 15.6). One can revert the expression in eq. (5) to obtain expressions of either drift or diffusion function:

$$\mu(x) = \frac{1}{2\pi(x)} \frac{\partial}{\partial x}\left[\sigma^2(x)\pi(x)\right], \tag{6}$$

$$\sigma^2(x) = \frac{2}{\pi(x)} \int_l^x \mu(y)\pi(y)\,dy. \tag{7}$$

From these expressions, we can identify the drift (diffusion) function from the diffusion (drift) term together with the marginal density; this point was already made in Wong (1964), and further pursued in Aït-Sahalia (1996a), Hansen and Scheinkman (1995), and Kristensen (2010). In particular, this allows us to identify the unspecified term under each of the two semiparametric hypotheses.

3 Semi-Nonparametric Estimators

We develop specific drift and diffusion estimators based on the identification scheme presented in the previous section: Suppose that we have $n+1$ observations available from eq. (1), $X_0, X_\Delta, X_{2\Delta}, ..., X_{n\Delta}$, where $\Delta > 0$ is the fixed time distance between observations; without loss of generality, we normalise time distance to $\Delta \equiv 1$ in the following. Under the relevant semiparametric hypothesis, $H_{\mathrm{SP},1}$ or $H_{\mathrm{SP},2}$, we assume that a preliminary estimator of the parametric component, θ_1 or θ_2, is available. We make no assumptions about where the preliminary estimators have arrived from, and merely require that they are sufficiently regular. One particular class of estimators are the pseudo-MLEs proposed in Kristensen (2010), but we do not restrict ourselves to these and the estimator of Aït-Sahalia (1996a) could also be used in the case of a linear drift specification.

Given estimators of the parametric components, we now just need to obtain an estimator of the marginal density, π. We here propose to use kernel methods to estimate it,

$$\hat{\pi}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \tag{8}$$

where $K_h(z) = K(z/h)/h$, K is a kernel, and $h > 0$ is a bandwidth; see Robinson (1983) for an introduction to kernel density estimators in a time series setting. We then combine estimators of the parametric component and the marginal density to obtain an estimator of the unspecified term.

First, consider $H_{\mathrm{SP},1}$: In this case, the diffusion term is parameterised and an estimator $\hat{\theta}_1$ is available together with the kernel estimator $\hat{\pi}$. We then estimate μ by substituting $\sigma^2(x;\hat{\theta}_1)$ and $\hat{\pi}$ into eq. (6):

$$\hat{\mu}(x) = \frac{1}{2\hat{\pi}(x)} \frac{\partial}{\partial x}\left[\sigma^2(x;\hat{\theta}_1)\hat{\pi}(x)\right]. \tag{9}$$

Under $H_{\mathrm{SP},2}$, we have a parametric estimator of the drift parameter, $\hat{\theta}_2$, which together with $\hat{\pi}$ can be used to estimate the diffusion term. Two alternative estimators present themselves: An obvious estimator would be to directly substitute $\mu(y;\hat{\theta}_2)$ and $\hat{\pi}$ into eq. (7), $\tilde{\sigma}^2(x) = \frac{2}{\hat{\pi}(x)}\int_l^x \mu(y;\hat{\theta}_2)\hat{\pi}(y)\,dy$. However, the integral $\int_l^x \mu(y)\pi(y)\,dy$ can be estimated without bias by a sample average, $\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{X_i\leq x\}\mu(X_i)\to^P \int_l^x \mu(y)\pi(y)\,dy$, where $\mathbb{I}\{\cdot\}$ is the indicator function. So we suggest to estimate $\sigma^2(x)$ by

$$\hat{\sigma}^2(x)=\frac{2}{\hat{\pi}(x)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{X_i\leq x\}\mu(X_i;\hat{\theta}_2). \tag{10}$$

To establish the asymptotic properties of the two estimators, we impose regularity conditions on the model:

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 $V(x)\to\infty$ as $|x|\to\infty$, and constants $b,c>0$ such that

$$\mu(x)V'(x)+\frac{1}{2}\sigma^2(x)V''(x)\leq -cV(x)+b. \tag{11}$$

A.2 The marginal density π is uniformly differentiable of order $m\geq 2$ with bounded derivatives, and satisfies $\int_I \pi(x)^{1-q}\,dx<\infty$ for some $q>0$. The conditional density $p(y|x)\equiv p(y|x;1)$ is uniformly differentiable of order m with $\sup_{x,y\in I}p(y|x)\pi(x)<\infty$.

A.3 The parametric drift and diffusion function satisfy:

1. $\theta_1\mapsto\sigma^2(x;\theta_1)$ is continuously differentiable satisfying $||\partial^{ij}_{x,\theta_1}\sigma^2(x;\theta_1)||\leq V(x)$, $i,j=0,1$.
2. $\theta_2\mapsto\mu(x;\theta_2)$ is continuously differentiable, satisfying $||\partial^{i}_{\theta_2}\mu(x;\theta_2)||\leq V(x)$, $i=0,1$.

A.4 For $k=1,2$: There exists $\theta_k^*\in\Theta_k$ and function $\psi_{\mathrm{SP},k}$ satisfying $E\left[\psi_{\mathrm{SP},k}(X_1|X_0)\right]=0$ and $E\left[||\psi_{\mathrm{SP},k}(X_1|X_0)||^{2+\delta}\right]<\infty$, such that $\sqrt{n}(\hat{\theta}_k-\theta_k^*)=\sum_{i=1}^{n}\psi_{\mathrm{SP},k}(X_i|X_{i-1})/n+o_P(1)$.

Assumption (A.1) is sufficient for a stationary and geometrically β-mixing solution to exist as shown in Meyn and Tweedie (1993); alternative mixing conditions for diffusion processes can be found in Chen, Hansen and Carrasco (2010) and Hansen and Scheinkman (1995). We will throughout assume that we have observed this solution. Some of the results stated in this section actually go through 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 Gourieroux and Tenreiro, 2001), we impose this restriction throughout for clarity. Most models found in the finance literature satisfy (A.1) under suitable restrictions on the parameters.

The existence of $m\geq 2$ derivatives of π assumed in (A.2) combined with the use of an mth order kernel as given in (B.1) below allow us to control the bias of the kernel density estimator and its first derivative. The smoothness of π as measured by its number of derivatives, m, determines how much the bias can be reduced with. The condition that π is m times differentiable is satisfied if μ and σ^2 are $m-1$ and m times differentiable respectively, c.f. eq. (5).

The tail condition imposed on π in (A.2) is used to obtain uniform convergence results for the semiparametric drift and diffusion estimators when analysing the associated semiparametric estimator of the transition density (see Lemma 2 in Section 4). The parameter $q>0$ measures the thickness of the tails of the marginal distribution, and is used to control the asymptotic impact of trimming introduced in the next section. The conditions on the transition density in (A.2) together with (A.1) allow us to bound the variance of $\hat{\pi}$, and will also become useful when analysing nonparametric estimators of the transition density in Section 4.

Assumption (A.3) in conjunction with (A.1) implies that the following two moments exist: $E[||\partial^i_{\theta_2}\mu(X_0;\theta_2)||]<\infty$ and $E[||\partial^{ij}_{x,\theta_1}\sigma^2(X_0;\theta_1)||]<\infty$. These are used when demonstrating uniform convergence of the semi-nonparametric estimators.

Assumption (A.4) imposes restrictions on the estimator $\hat{\theta}_k$ obtained under $H_{\mathrm{SP},k}$, $k=1,2$. The assumption is formulated so it holds both under $H_{\mathrm{SP},1}$ and $H_{\mathrm{SP},2}$ respectively, and the nonparametric alternative. Under the relevant null, there exists a parameter value $\theta_{0,k}$ such that either eq. (3) or (4) hold. It is then implicitly assumed that $\theta^*_k=\theta_{0,k}$ such that $\sigma^2(x;\theta^*_1)=\sigma^2(x)$ and $\mu(x;\theta^*_2)=\mu(x)$ respectively. If the relevant semiparametric null is false, no parameter value exists such that either eq. (3) or (4) hold. As such θ^*_k is a pseudo-true value in the sense that $\sigma^2(x;\theta^*_1)\neq\sigma^2(x)$ and $\mu(x;\theta^*_2)\neq\mu(x)$ respectively, and θ^*_k is just some parameter value that the estimator converges towards. Under the null, Assumption (A.4) is satisfied in great generality for most well-behaved estimators: For the fully parametric MLE's, Aït-Sahalia (2002) gives conditions for (A.4) to hold, while Kristensen (2010) give conditions under which semiparametric pseudo MLE's satisfy the conditions. Under the alternatives, we expect that (A.4) will still hold under great generality by employing the arguments similar to those in White (1982). For some of our results, the conditions imposed on the parametric estimators in (A.4) can be weakened to the requirement that they merely converge at a faster rate than the kernel estimator. However, for simplicity we maintain the stronger assumptions of (A.4) throughout.

Finally, we restrict the class of kernel functions to belong to the following family:

B.1 The kernel K is differentiable, and there exists constants $C,\eta>0$ such that

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

where $K^{(i)}(z)$ denotes the ith derivative. Furthemore, $\int_{\mathbb{R}}K(z)\,dz=1$, $\int_{\mathbb{R}}z^jK(z)\,dz=0$, $1\leq j\leq m-1$, and $\int_{\mathbb{R}}|z|^mK(z)\,dz<\infty$.

This class includes most standard kernels including the Gaussian and Uniform kernel. We are now able to state pointwise convergence results for the estimators of the unspecified term under the two semiparametric nulls:

Theorem 1 *Assume that (A.1)-(A.4) and (B.1) hold. Then for any point x in the interior of I:*

1. *Under $H_{\mathrm{SP},1}$: As $nh^3\to\infty$ and $nh^{3+2m}\to0$,*

$$\sqrt{nh^3}(\hat{\mu}(x)-\mu(x))\xrightarrow{d}N(0,V_\mu(x)),$$

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

2. *Under $H_{\mathrm{SP},2}$: As $nh\to\infty$, and $nh^{1+2m}\to0$,*

$$\sqrt{nh}(\hat{\sigma}^2(x)-\sigma^2(x))\xrightarrow{d}N(0,V_{\sigma^2}(x)),$$

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

The above result allows the researcher to plot the two estimators together with pointwise confidence bands. The pointwise asymptotic variances for $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ can be estimated by:

$$\hat{V}_\mu(x)=\frac{\sigma^4(x;\hat{\theta}_1)}{4\hat{\pi}(x)}\int_{\mathbb{R}}K^{(1)}(z)^2\,dz,\quad \hat{V}_{\sigma^2}(x)=\frac{\hat{\sigma}^4(x)}{\hat{\pi}(x)}\int_{\mathbb{R}}K(z)^2\,dz. \tag{12}$$

One can easily show, as is standard for kernel-based estimators, that both semi-nonparametric estimators are asymptotically independent across distinct points. This facilitates inference, for example when constructing pointwise confidence bands.

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 which is a well-established fact in the literature: Gobet, Hoffmann and Reiß (2004) show that the optimal convergence rate of the nonparametric estimation of the drift is slower than for the diffusion given low-frequency observations, and coin the nonparametric estimation of μ as an "ill-posed problem". Similarly, Bandi and Phillips (2003) demonstrate that with high-frequency observations of 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 $\Delta \to 0$ and $n\Delta \to \infty$.

4 Goodness-of-Fit Testing

We here develop tests of correct specifications of the drift and/or diffusion function. Our main focus will be on tests based on the transition density of the Markov process $\{X_t\}$, where a given null is tested against a given alternative by comparing estimators of the transition density obtained under the null and the alternative respectively. However, motivated by a power analysis of the proposed transition-based tests, we will also develop tests that directly compare drift and diffusion estimators under null and alternative. The two following subsections develop tests of the semiparametric and fully parametric hypotheses respectively and examine their properties.

4.1 Semiparametric Specification Tests

We consider testing either $H_{\mathrm{SP},1}$ or $H_{\mathrm{SP},2}$ against H_{NP}. In order to present our tests, we first introduce some additional notation: Recall that we have normalised the time distance between observations to $\Delta = 1$, such that $p(y|x) := p(y|x;1)$ is the transition density of the observed Markov chain, X_i, $i = 1, ..., n$. Let $f(y,x) = p(y|x)\pi(x)$ denote the corresponding joint density of (X_i, X_{i-1}). Under either of the two semiparametric hypotheseses, restrictions are imposed on the drift and diffusion term. Using eqs. (6)-(7), we define the restricted drift and diffusion terms under the respective nulls as

$$\mu_{\mathrm{SP},1}(x) = \frac{1}{2\pi(x)} \frac{\partial}{\partial x} \left[\sigma^2(x;\theta_1^*)\pi(x) \right], \quad \sigma^2_{\mathrm{SP},1}(x) = \sigma^2(x;\theta_1^*), \tag{13}$$

$$\mu_{\mathrm{SP},2}(x) = \mu(x;\theta_2^*), \quad \sigma^2_{\mathrm{SP},2}(x) = \frac{2}{\pi(x)} \int_l^x \mu(x;\theta_2^*)\pi(y)\,dy. \tag{14}$$

We let $p_{\mathrm{SP},k}(y|x;\theta_k) := p_{\mathrm{SP},k}(y|x;1,\theta_k)$ denote the transition density corresponding to the restricted drift and diffusion functions under $H_{\mathrm{SP},k}$, $k = 1,2$ at $t = \Delta = 1$. It can for example be represented as the solution (at $t = 1$) to the PDE in eq. (2) with the restricted drift and diffusion functions plugged in. When evaluated at the (pseudo-)true parameter value we simply write $p_{\mathrm{SP},k}(y|x) = p_{\mathrm{SP},k}(y|x;\theta_k^*)$.

Under the nonparametric hypothesis, H_{NP}, the drift and diffusion functions are left completely unspecified, and so we propose to estimate the unrestricted transition density, $p(y|x)$, under the alternative using standard kernel methods. A standard kernel estimator of the transition density for the observed data is

$$\hat{p}_{\mathrm{NP}}(y|x) = \frac{\hat{f}_{\mathrm{NP}}(y,x)}{\hat{\pi}_{\mathrm{NP}}(x)},$$

where, for some bandwidth $h_{\mathrm{NP}} > 0$,

$$\hat{f}_{\mathrm{NP}}(y,x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_{\mathrm{NP}}}(X_i - y)\, K_{h_{\mathrm{NP}}}(X_{i-1} - x), \quad \hat{\pi}_{\mathrm{NP}}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_{\mathrm{NP}}}(X_{i-1} - x).$$

Note that two different bandwidths are now being employed: Under the semiparametric null, we use the bandwidth h in the estimation of the univariate marginal density, while under the alternative h_{NP} is used to obtain a nonparametric estimator of the bivariate transition density.

Next, we obtain an estimator of the transition density under either of the two semiparametric hypotheseses, $p_{\mathrm{SP},k}(y|x)$. In both cases, we have drift and diffusion estimators available as developed in the previous section. These could in principle be used to obtain an estimator of $p_{\mathrm{SP},k}(y|x)$ by plugging them into the PDE in eq. (2) and then solving w.r.t. $p(y|x;t)$ (at $t = 1$). However, to establish theoretical properties of the resulting semiparametric estimator of the transition density, we have to modify the drift and diffusion estimators proposed in the previous section to control their tail behaviour. We first introduce a class of trimming functions $\tau_a(z)$:

B.2 The trimming function $\tau_a : \mathbb{R} \mapsto [0,1]$, $a > 0$, satisfies $\tau_a(z) = 1$ for $z \geq a$ and $\tau_a(z) = 0$ for $z \leq a/2$.

A simple way of constructing $\tau_a(z)$ is to choose a cdf F with support $[0,1]$, and define $\tau_a(z) = F((2z-a)/a)$ which then in great generality will satisfy (B.2); see also Andrews (1995, p. 572).

Given the trimming function, we redefine the estimators under the two semiparametric hypotheses, where we now use subscripts to differentiate between the two nulls,

$$\hat{\mu}_{\mathrm{SP},1}(x) = \frac{\hat{\tau}_a(x)}{2\hat{\pi}(x)}\frac{\partial}{\partial x}\left[\sigma^2(x;\hat{\theta}_1)\hat{\pi}(x)\right], \quad \hat{\sigma}^2_{\mathrm{SP},1}(x) = \hat{\tau}_a(x)\,\sigma^2(x;\hat{\theta}_1) + \underline{\sigma}^2(1 - \hat{\tau}_a(x)), \quad (15)$$

$$\hat{\mu}_{\mathrm{SP},2}(x) = \hat{\tau}_a(x)\,\mu(x;\hat{\theta}_2), \quad \hat{\sigma}^2_{\mathrm{SP},2}(x) = \frac{2\hat{\tau}_a(x)}{\hat{\pi}(x)}\int_l^x \mu(y;\hat{\theta}_2)\hat{\pi}(y)\,dy + \underline{\sigma}^2(1 - \hat{\tau}_a(x)), \quad (16)$$

where $\hat{\tau}_a(x) := \tau_a(\hat{\pi}(x))$, $a = a_n > 0$ is a trimming sequence, and $\underline{\sigma}^2 > 0$ a constant. The inclusion of the additional term $\underline{\sigma}^2(1 - \hat{\tau}_a(x))$ in the diffusion estimator guarantees that it is strictly positive for all $x \in I$ for n sufficiently large. The motivation for the trimming is two-fold: First, by combining results of Andrews (1995) and Kristensen (2009), the trimming of the semi-nonparametric component is used to show that $\hat{\mu}_{\mathrm{SP},1}(x) \to^P \tau_a(\pi(x))\,\mu_{\mathrm{SP},1}(x)$ and $\hat{\sigma}^2_{\mathrm{SP},2}(x) \to^P \tau_a(\pi(x))\,\sigma^2_{\mathrm{SP},2}(x)$ uniformly over $x \in I$, $k = 1,2$, c.f. Lemma 9. We will then let $a \to 0$ at a suitable rate such that asymptotically the trimming has no first-order effect asymptotically, $\tau_a(\pi(x))\,\mu_{\mathrm{SP},1}(x) \approx \mu_{\mathrm{SP},1}(x)$ and $\tau_a(\pi(x))\,\sigma^2_{\mathrm{SP},2}(x) \approx \sigma^2_{\mathrm{SP},2}(x)$; see, for example, Ai (1997) and Robinson (1988) for similar applications of trimming. Second, the trimming of the parametric component is introduced to ensure that the associated transition density exists: Due to trimming, $\hat{\mu}_{\mathrm{SP},k}$ and $\hat{\sigma}^2_{\mathrm{SP},k}$ are bounded and $\hat{\sigma}^2_{\mathrm{SP},k} > 0$, and we can therefore apply standard results to ensure that the associated diffusion process has a well-defined transition density; see, for example, Friedman (1976).

Given the above re-defined semiparametric drift and diffusion estimators, we define our estimator of the corresponding transition density, $\hat{p}_{\mathrm{SP},k}(y|x)$, as the solution to the following PDE at $t = 1$,

$$\frac{\partial \hat{p}_{\mathrm{SP},k}(y|x;t)}{\partial t} = \mathcal{A}\left[\hat{\mu}_{\mathrm{SP},k}, \hat{\sigma}^2_{\mathrm{SP},k}\right]\hat{p}_{\mathrm{SP},k}(y|x;t), \quad t > 0, \ (x,y) \in I \times I. \quad (17)$$

While the theoretical analysis of the estimator will rely on the above representation, its actual computation can be done using numerical techniques as developed in, amongst others, Aït-Sahalia (2002) and Kristensen and Shin (2008); see also Kristensen (2010, Section 5).

Given the non- and semiparametric estimates, we propose to test $H_{\text{SP},k}$ using the following statistic,

$$T_{\text{SP},k} = \int_I \int_I \left[\hat{p}_{\text{SP},k}(y|x) - \hat{p}_{\text{NP}}(y|x)\right]^2 w\left(y,x\right) dydx, \tag{18}$$

for some weighting function w. Similar test statistics have been considered in Aït-Sahalia et al (2009) and Li and Tkacz (2006) but in a different context, namely that of testing fully parametric models against a nonparametric alternative. By appropriate choice of w, the tests can be interpreted as second-order approximations of the generalised likelihood-ratio tests, c.f. Aït-Sahalia et al (2009, p. 1105) and Fan, Zhang and Zhang (2001).

Other transition-based distance measures could be used: For example, measures based on the Kuhlback-Leibler divergence (Robinson, 1991), the empirical likelihood (Chen et al, 2009), or integral transforms (Hong and Li, 2004). We focus on $T_{\text{SP},k}$, but conjecture that theoretical results for other distance measures could be derived by following the same proof strategy as used here for $T_{\text{SP},k}$.

As a first step towards establishing asymptotic properties of $T_{\text{SP},k}$, we investigate the properties of $\hat{p}_{\text{SP},k}(y|x)$. The analysis will rely on the representation of $\hat{p}_{\text{SP},k}(y|x)$ as the solution to eq. (17) at $t = 1$. To ensure that the solution exists (asymptotically) and is sufficiently regular, we impose the following assumption on the transition density:

A.5 The transition density under $H_{\text{SP},k}$, $p_{\text{SP},k}\left(y|x;t,\theta\right)$ for $t > 0$, exists as a solution to eq. (2) and satisfies $\left|\partial_x^i p_{\text{SP},k}\left(y|x;t,\theta\right)\right| \leq \gamma\left(y|x;t\right)$, $(t,x,y,\theta) \in (0,\Delta] \times I^2 \times \Theta$, $i = 0,1,2$, where

$$\gamma\left(y|x;t\right) = c_1 \frac{|y|^{\lambda_1} + |x|^{\lambda_1}}{t^{\alpha_1}} \exp\left[-c_2 \frac{|y|^{\lambda_2} + |x|^{\lambda_2}}{t^{\alpha_2}}\right] \tag{19}$$

for constants $c_j, \alpha_j, \lambda_j > 0$, $j = 1,2$.

The above assumption is high level. It would be preferable to give more primitive conditions in terms of the underlying drift and diffusion functions for the above regularity conditions to hold. However, to our knowledge, the only known sufficient conditions for existence of a solution to eq. (2) are overly restrictive and, for example, require that drift and diffusion functions are bounded, c.f. Friedman (1976). Such boundedness restrictions are violated by most standard models used in the literature, and rule out that the process is mixing, c.f. Chen, Hansen and Carrasco (2010). We therefore impose the high level conditions in (A.5) instead, which is similar to the conditions imposed in Kristensen (2010). We conjecture that the assumption could be replaced by alternative conditions such as the ones in Aït-Sahalia (2002).

Finally, with $m \geq 2$ and $q > 0$ given in (A.2), we impose the following conditions on the bandwidth and trimming parameter to ensure that $\hat{\mu}_{\text{SP},k}\left(x\right)$ and $\hat{\sigma}^2_{\text{SP},k}\left(x\right)$ converge sufficiently fast:

H.1 $\sqrt{n}ha^6/\log\left(n\right) \to \infty$, $\sqrt{n}h^3a^4/\log\left(n\right) \to \infty$, $n^{1/4}h^m a^{-3} \to 0$, $\sqrt{n}h^m a^{-1} \to 0$, and $\sqrt{n}a^{q/2} \to 0$.

H.2 $\sqrt{n}ha^4/\log\left(n\right) \to \infty$, $n^{1/4}h^m a^{-2} \to 0$, $\sqrt{n}h^m a^{-1} \to 0$, and $\sqrt{n}a^{q/2} \to 0$.

Depending on whether we work under $H_{\text{SP},1}$ or $H_{\text{SP},2}$, we will impose (H.1) or (H.2) respectively. The conditions involve both h and a and impose restrictions on how fast they jointly can go to zero. They are used to control higher-order bias and variance terms appearing in $\hat{p}_{\text{SP},k}(y|x)$; in particular, they ensure that the kernel-based estimators of the relevant semi-nonparametric component under the null converges with rate $o_P\left(n^{-1/4}\right)$ uniformly over $\{x : \pi\left(x\right) \geq a\}$, and that the trimming has no first-order impact on the semiparametric transition density estimator.

Utilising arguments developed in Kristensen (2008, 2010), we are now able to establish the following asymptotic expansion of the transition density estimator:

Lemma 2 *For $k \in \{1,2\}$: Assume that (A.1)-(A.5), (B.1)-(B.2) and (H.k) hold. Then,*

$$\hat{p}_{\mathrm{SP},k}(y|x) = p_{\mathrm{SP},k}(y|x) + \frac{1}{n}\sum_{i=1}^{n} D_{k,i}(y|x) + o_P\left(1/\sqrt{n}\right), \quad k = 1,2,$$

uniformly over (y,x) in any compact set of $I \times I$. Here, $D_{k,i}(y|x) = D_k(X_i, X_{i-1}; y, x)$ is given in eq. (35). In particular, $E[D_{k,i}(y|x)] = 0$ and $E\left[D_{k,i}^2(y|x)\right] < \infty$ for all (x,y).

From the above lemma, we see that $\hat{p}_{\mathrm{SP},k}(y|x)$ is $\sqrt{n}$-consistent. This holds despite the fact that nonparametric kernel estimators are employed as inputs in the computation of $\hat{p}_{\mathrm{SP},k}(y|x)$. The reason for this maybe surprising result can be found in the representation of $\hat{p}_{\mathrm{SP},k}(y|x)$ as a solution to a PDE: As such, the computation of $\hat{p}_{\mathrm{SP},k}(y|x)$ involves integrating over the drift and diffusion estimator which in turn speeds up the convergence rate; for more details, we refer to Kristensen (2008, 2010). An important consequence of the above lemma is that $\hat{p}_{\mathrm{SP},k}(y|x)$ converges at a faster rate than $\hat{p}_{\mathrm{NP}}(y|x)$, so we can exchange $\hat{p}_{\mathrm{SP},k}(y|x)$ for the unknown density in the derivation of the asymptotic properties of $T_{\mathrm{SP},k}$. Finally, we note that the theorem holds both under $H_{\mathrm{SP},k}$ and the alternative: Under the null $p_{\mathrm{SP},k}(y|x)$ equals the true data generating transition density while under the alternative it corresponds to the drift or diffusion restriction evaluated at the pseudo-true value.

To derive the asymptotic properties of the test statistics, we impose the following restriction on the weighting function:

B.3 The weighting function $w : I \times I \mapsto \mathbb{R}_+$ is continuous with compact support.

The assumption of a fixed, compact support of w is made in order to control the tail behaviour of the estimators of transition densities. This assumption is fairly standard and is, for example, also imposed in Aït-Sahalia et al (2009); a similar restriction is imposed in Li and Tkacz (2006) who assume compact support of the joint density $f(y,x)$. Under (B.2), $T_{\mathrm{SP},k}$ can only detect departures from $H_{\mathrm{SP},k}$ that reveal themselves in the density within the support of w. However, under suitable regularity conditions on the tail behaviour of w, the drift and the diffusion, one should be able to allow for weighting functions with unbounded support, see e.g. Kristensen (2010) and Li and Tkacz (2006). This would lead to more complicated proofs however, and we therefore maintain (B.3) for simplicity.

In the following, let $(f * g)(z) = \int_{\mathbb{R}} f(u)\, g(u+z)\, du$ denote the convolution of any two functions f and g. We then have the following results for the asymptotic properties of the two tests:

Theorem 3 *For $k \in \{1,2\}$: Assume that (A.1)-(A.5), (B.1))-(B.3), (H.k) and $H_{\mathrm{SP},k}$ hold.*

(i) If $\lambda_n^2 := nh_{\mathrm{NP}}^{2m+2} \to \lambda^2 < \infty$, the following expansion holds:

$$nh_{\mathrm{NP}}\{T_{\mathrm{SP},k} - m_{\mathrm{SP}}\} = v_{\mathrm{SP}} U_n + \sqrt{h_{\mathrm{NP}}}\bar{v}_{\mathrm{SP}}\bar{U}_n + \lambda_n\left\{\sigma_v V_n + \bar{\sigma}_v \bar{V}_n\right\} + nh_{\mathrm{NP}}^{2m+1} B_k + O_P\left(\frac{\log(n)^2}{nh_{\mathrm{NP}}^3}\right)$$

where (U_n, V_n) and $\left(\bar{U}_n, \bar{V}_n\right)$ both converge towards bivariate standard normal distributions,

$$\begin{aligned} m_{\mathrm{SP}} &= \frac{1}{nh_{\mathrm{NP}}^2}\left[\int_{\mathbb{R}} K^2(z)\, dz\right]^2 \times \int_{\mathbb{R}^2} \frac{p(y|x)}{\pi(x)} w(y,x)\, dydx, \\ &\quad + \frac{1}{nh_{\mathrm{NP}}}\int_{\mathbb{R}} K^2(z)\, dz \times \int_{\mathbb{R}^2} \frac{p(y|x)}{\pi^2(x)} w(y,x)\, dydx, \end{aligned}$$

$$v_{\mathrm{SP}}^2 = 2\left[\int_{\mathbb{R}} (K*K)^2(z)\, dz\right]^2 \times \int_{I\times I} p^2(y|x) w(y,x)\, dydx,$$

and the parameters B_k, $\bar{v}_{\mathrm{SP}}^2$, σ_v^2 and $\bar{\sigma}_v^2$ are given in the proof.

(ii) In particular, if $nh_{\mathrm{NP}}^3/\log(n)^2 \to \infty$ and $nh_{\mathrm{NP}}^{2m+1} \to 0$,

$$nh_{\mathrm{NP}} \frac{T_{\mathrm{SP},k} - m_{\mathrm{SP}}}{v_{\mathrm{SP}}} \to^d N(0,1).$$

The first part of the theorem states an asymptotic expansion of $T_{\mathrm{SP},k}$, $k = 1, 2$, under weak restrictions on the bandwidth. The limiting distribution is in this general case quite involved and not easily evaluated. One could adjust the proposed test statistics by following the ideas of Bickel and Rosenblatt (1973) and Fan (1994) in order to remove the higher-order terms $\lambda_n \left\{\sigma_v V_n + \bar{\sigma}_v \bar{V}_n\right\}$ and $O_P\left(nh_{\mathrm{NP}}^{2m+1}\right)$. This however would have consequences for the resulting tests' power properties, c.f. Fan (1994).

Under additional restrictions on the bandwidth h_{NP}, we obtain a standard normal distribution of the tests which is similar to the results reported in Aït-Sahalia et al (2009, Theorems 1-2), and Li and Tkacz (2006, Theorem 1). In particular, as in these studies, the asymptotic distribution is entirely determined by the nonparametric estimator, $\hat{p}_{\mathrm{NP}}(y|x)$, since the estimator of the transition density under the null converges with parametric rate. This is the reason for that the asymptotic expansions in the first part are the same for both tests. It should also be noted that the asymptotic distribution in the second part is not affected by the dependence structure in data and is identical to the one found when data is i.i.d., see e.g. Fan (1994).

In order for the resulting test in the second part to become operational, consistent estimates of m_{SP} and v_{SP}^2 have to be obtained. This can easily be done by substituting the unknown quantities entering these for their estimates (either under the null or the alternative); see e.g. Li and Tkacz (2006, p. 867).

Next, we investigate the power of the proposed tests. To this end, we introduce the following two sequences of contiguous alternatives:

$$H_{\mathrm{SP},1}^c : \mu_n(x) = \frac{1}{2\pi(x)} \frac{\partial}{\partial x}\left[\sigma_n^2(x)\pi(x)\right], \quad \sigma_n^2(x) = \sigma^2(x;\theta_1^*) + g_n(x),$$

and

$$H_{\mathrm{SP},2}^c : \mu_n(x) = \mu(x;\theta_2^*) + g_n(x), \quad \sigma_n^2(x) = \frac{2}{\pi(x)} \int_l^x \mu_n(y)\pi(y)\,dy.$$

Here, $g_n : I \mapsto \mathbb{R}$ is a sequence of functions, which we will throughout restrict to be continuously differentiable with compact support uniformly in n. These alternatives posit that the diffusion model is stationary such that the unspecified term can be identified by eqs. (6) and (7) respectively, but that the parametric component is misspecified with $g_n(x)$ describing the degree of misspecification. In particular, if $g_n(x) = 0$ the null is true.

Under great generality, the estimator under the null, $\hat{\theta}_k$, will still satisfy Assumption (A.4) under $H_{\mathrm{SP},k}^c$. However, the data generating process (DGP) is now drifting, and as a consequence the pseudo-true parameter will in general also drift. At a first glance, the above alternatives therefore looks a bit odd since $\theta_k^* = \theta_{n,k}^*$ is non-constant, and a more natural specification of, for example, $H_{\mathrm{SP},1}^c$ would seem to be $\tilde{H}_{\mathrm{SP},1}^c : \sigma_n^2(x) = \sigma^2(x;\theta_{0,1}) + f_n(x)$, for some fixed value $\theta_{0,1}$. However, we here follow the existing literature and focus on $H_{\mathrm{SP},1}^c$ for the following reasons[1]: First, note that under $\tilde{H}_{\mathrm{SP},1}^c$ $\theta_{n,1}^* \to \theta_{0,1}$ in great generality as $f_n(x) \to 0$. Thus, $\tilde{H}_{\mathrm{SP},1}^c$ is included in $H_{\mathrm{SP},1}^c$ since we can always choose $g_n(x) := f_n(x) + \left\{\sigma^2(x;\theta_{0,1}) - \sigma^2\left(x;\theta_{n,1}^*\right)\right\}$. This also shows that $g_n(x)$ captures the over all impact of employing the misspecified model: It captures the explicit difference, $f_n(x)$, between the DGP and the parametric restriction, but also the impact that this difference has on the parametric estimator of the diffusion term, $\sigma^2(x;\theta_{0,1}) - \sigma^2\left(x;\theta_{n,1}^*\right)$. As a result, by working with $H_{\mathrm{SP},1}^c$ the notation and theoretical analysis become less burdensome since we do not explicitly have to specify how $\theta_{n,1}^*$ drifts towards $\theta_{0,1}$.

[1]See also Härdle and Mammen, (1993) for a detailed discussion in a regression setting.

It should be stressed that the above alternatives are different from the ones considered in, for example, Fan (1994) and Aït-Sahalia et al (2009) who specify alternatives in terms of the corresponding (transition) density. Since our focus is on testing for the correct specification of the drift and diffusion function, our alternatives seem to be the more natural ones though.

Since the proposed tests are based on transition densities, we first obtain an expression of the sequences transition densities corresponding to the above two contiguous alternatives. Let $p_n(y|x) = p(y|x; 1, \mu_n, \sigma_n^2)$ denote the sequence of transition densities corresponding to either of the two sequences of alternatives. Utilising that $p_n(y|x)$ and $p_{\mathrm{SP},k}(y|x)$ both solve a PDE on the form given in eq. (2), we obtain the following relationship between the two (see Proof of Theorem 4):

$$p_n(y|x) = p_{\mathrm{SP},k}(y|x) + \gamma_{\mathrm{SP},k}^{(n)}(y|x) + O(R_{\mathrm{SP},k}), \tag{20}$$

where

$$\gamma_{\mathrm{SP},1}^{(n)}(y|x) = \frac{1}{2}\int_I g_n'(w)\frac{\pi'(w)}{\pi(w)}\bar{p}_{\mu,1}(y,x,w)\,dw + \int_I g_n(w)\,\bar{p}_{\sigma^2,1}(y,x,w)\,dw, \tag{21}$$

$$\gamma_{\mathrm{SP},2}^{(n)}(y|x) = \int_I g_n(w)\,\bar{p}_{\mu,2}(y,x,w)\,dwdt + 2\int_I\int_l^w g_n(u)\,\pi(u)\,du\frac{1}{\pi(w)}\bar{p}_{\sigma^2,2}(y,x,w)\,dw, \tag{22}$$

and $R_{\mathrm{SP},1}$ and $R_{\mathrm{SP},2}$ are remainder terms given by

$$R_{\mathrm{SP},1} = \sup_{x\in I}|g_n(x)|^2 + \sup_{x\in I}\left|g_n'(x)\right|^2, \quad R_{\mathrm{SP},2} = \sup_{x\in I}|g_n(x)|^2.$$

Here, we have defined

$$\bar{p}_{\mu,k}(y,x,w) := \int_0^1 \frac{\partial p_{\mathrm{SP},k}(y|w;t)}{\partial w}p_{\mathrm{SP},k}(w|x;t)\,dt, \quad \bar{p}_{\sigma^2,k}(y,x,w) := \int_0^1 \frac{\partial^2 p_{\mathrm{SP},k}(y|w;t)}{\partial w^2}p_{\mathrm{SP},k}(w|x;t)\,dt \tag{23}$$

The above expressions of the deviations in terms of densities, $\gamma_{\mathrm{SP},k}^{(n)}(y|x)$, involves integrating over the deviation $g_n(x)$ appearing in the drift and diffusion term. This is due to the fact that any given diffusion transition density implicitly integrates over the underlying drift and diffusion terms as noted earlier.

Next, using arguments similar to those of Gourieroux and Tenreiro (2001, Proof of Theorem 3), we obtain the following theorem:

Theorem 4 *For $k \in \{1,2\}$: Assume that (A.1)-(A.5), (B.1)-(B.3), (H.k) and $H_{\mathrm{SP},k}^c$ hold. Then, as $nh_{\mathrm{NP}}^3/\log(n)^2 \to \infty$ and $nh_{\mathrm{NP}}^{2m+1} \to 0$„*

$$nh_{\mathrm{NP}}\{T_{\mathrm{SP},k} - m_{\mathrm{SP}}\} = v_{\mathrm{SP}}U_{n,1} + nh_{\mathrm{NP}}\int_I\int_I \gamma_{\mathrm{SP},k}^{(n)}(y|x)^2 w(y,x)\,dydx + O_P\left(R_{\mathrm{SP},k}^2\right) + o_P(1). \tag{24}$$

The above expression of the test statistic under contiguous alternatives corresponds to the ones found in Aït-Sahalia et al (2009, Theorem 3) and Fan (1994, Theorem 3.6), except that the deviation from the null, $\gamma_{\mathrm{SP},k}^{(n)}(y|x)$, here takes a more complicated form since it is expressed in terms of the underlying deviations from the hypothesised drift and diffusion function.

We now proceed to consider different specifications of the deviation from the null, $g_n(x)$. First, we first consider so-called Pitman alternatives on the form $g_n(x) = a_n g(x)$ for a sequence $a_n \to 0$ and a fixed function $g(w)$. These are global alternatives for which the deviations in terms of the transition density takes the form

$$\gamma_{\mathrm{SP},1}^{(n)}(y|x) = \frac{a_n}{2}\int_I g'(w)\frac{\pi'(w)}{\pi(w)}\bar{p}_{\mu,1}(y,x,w)\,dw + a_n\int_0^1\int_I g(w)\,\bar{p}_{\sigma^2,1}(y,x,w)\,dw,$$

and

$$\gamma_{\mathrm{SP},2}^{(n)}(y|x) = a_n \int_I g(w)\,\bar{p}_{\mu,2}(y,x,w)\,dw + 2a_n \int_I \int_l^w g(u)\,\pi(u)\,du \frac{1}{\pi(w)} \bar{p}_{\sigma^2,2}(y,x,w)\,dw.$$

Plugging these expressions into eq. (24), we easily see that both tests can in general detect global alternatives for which $\lim_{n\to\infty} nh_{\mathrm{NP}}a_n^2 > 0$. In particular, they can detect alternatives that vanish with rate $a_n = O\left(n^{-2/5}\right)$ when the bandwidth is chosen to vanish with rate $h_{\mathrm{NP}} = O\left(n^{-1/5}\right)$. Do note however that for particular directions (choices of $g(w)$), $\gamma_{\mathrm{SP},1}^{(n)}(y|x)$ and $\gamma_{\mathrm{SP},2}^{(n)}(y|x)$ are zero and so the transition-based tests cannot detect such alternatives. This shows that our tests are less powerful than CvM and KS type tests which can detect alternatives at parametric rate. The above results are in accordance with the analysis of kernel-based specification tests where the alternatives are directly expressed in terms of the density of interest, c.f. Aït-Sahalia et al (2009) and Fan (1994).

To investigate whether the above mentioned drawback of our test relative to cdf-based tests is peculiar to Pitman alternatives, we consider "local" deviations on the form $g_n(x) = a_n g((x-x_0)/b_n)$ as originally proposed in Rosenblatt (1975). We here introduce an additional sequence of $b_n \to 0$ and some $x_0 \in I$. For this class of drift and diffusion alternatives, the corresponding deviations in terms of the transition densities satisfy

$$\begin{aligned}
\gamma_{\mathrm{SP},1}^{(n)}(y|x) &= \frac{a_n}{2} \int_I \frac{1}{b_n} g'\left(\frac{w-x_0}{b_n}\right) \frac{\pi'(w)}{\pi(w)} \bar{p}_{\mu,1}(y,x,w)\,dw + a \int_I g\left(\frac{w-x_0}{b_n}\right) \bar{p}_{\sigma^2,1}(y,x,w)\,dw \\
&= \frac{a_n}{2} \frac{\pi'(x_0)}{\pi(x_0)} \bar{p}_{\mu,1}(y,x,x_0) \times \int g'(z)\,dz + a_n b_n \bar{p}_{\sigma^2,1}(y,x,x_0) \times \int g(z)\,dz \\
&\quad + o(a_n) + o(a_n b_n),
\end{aligned}$$

and, similarly,

$$\begin{aligned}
\gamma_{\mathrm{SP},2}^{(n)}(y|x) &= a_n b_n \bar{p}_{\mu,2}(y,x,x_0) \times \int g(z)\,dz + 2a_n b_n \pi(x_0) \int_{x_0}^r \frac{1}{\pi(w)} \bar{p}_{\sigma^2,2}(y,x,w)\,dw \times \int g(z)\,dz \\
&\quad + o(a_n) + o(a_n b_n).
\end{aligned}$$

By plugging the above expressions into eq. (24), we find that our tests are only able to detect local alternatives for which $\lim_{n\to\infty} nh_{\mathrm{NP}}a_n^2 b_n^2 > 0$. Moreover, alternatives which satisfy $\int g(z)\,dz = \int g'(z)\,dz = 0$ cannot be detected by $T_{\mathrm{SP},1}$ for fixed a_n and b_n, while $T_{\mathrm{SP},2}$ cannot detect alternatives satisfying $\int g(z)\,dz = 0$. This shows that our transition-based test statistics also have problems detecting high-frequency/local features in the drift and diffusion function. In particular, the above rates are the same as the one for KS and CvM type tests.

The above power results are not particular to our setting, and apply more generally. In particular, the transition-based tests of fully parametric models proposed in Aït-Sahalia et al (2009) also suffer from poor power against local alternatives: In Section 5, we revisit one of their tests and show that this also suffers from low power against alternatives in terms of the drift and diffusion term.

Our findings for local ("high-frequency") alternatives are somewhat analogous to the negative results reported for tests based on cumulative distribution functions (cdf's) such as the KS and CvM tests: High-frequency departures, as formulated in terms of the density, cannot be detected by such tests since the departures are integrated out in the computation of the cdf's, see e.g. Escanciano (2009) and Eubank and LaRiccia (1992). However, such tests are on the other hand more powerful at detecting global Pitman alternatives compared to tests based on transition densities such as ours, since the former can detect alternatives at parametric rate.

In conclusion, it appears as if tests based on L_2-distance measures of transition densities may not be very appropriate for detection of alternatives in terms of the underlying drift and diffusion functionst. One way to detect deviating features of these functions would be to obtain estimates of these under null and alternative and compare those directly instead of through the corresponding transition densities. Suppose that we have at our disposal fully nonparametric estimators of the drift and diffusion term, say $\hat{\mu}_{\mathrm{NP}}(x)$ and $\hat{\sigma}^2_{\mathrm{NP}}(x)$ such as the sieve estimators developed in Chen, Hansen and Scheinkman (2010) and Gobet et al (2004). Two natural classes of test statistics would then be

$$\bar{T}_{\mathrm{SP},1} = \int_I [\hat{\sigma}^2_{\mathrm{NP}}(x) - \hat{\sigma}^2_{\mathrm{SP},1}(x)]^2 \bar{w}(x)\, dx, \quad \bar{T}_{\mathrm{SP},2} = \int_I [\hat{\mu}_{\mathrm{NP}}(x) - \hat{\mu}_{\mathrm{SP},2}(x)]^2 \bar{w}(x)\, dx, \tag{25}$$

for some weighting function $\bar{w}(x)$. We expect that such tests would have better power properties against local alternatives. This conjecturre is supported by the theoretical results found in the next section where we demonstrate that in testing of fully parametric models against semiparametric alternatives this class of tests indeed have better local power properties. It would be of interest to develop similar results for the above two test statistics, but this is hampered by the fact that existing fully nonparametric estimators are quite complicated to analyze theoretically. We therefore leave this for future research.

4.2 Parametric Specification Tests

In this section, we develop tests of the fully parametric hypothesis, H_{P}, against either of the two semiparametric ones, $H_{\mathrm{SP},1}$ or $H_{\mathrm{SP},2}$. We will consider two types of tests: The first is similar in spirit to the tests considered in the previous section and based on an indirect comparison of the null and alternative through the corresponding transition density estimates. The second will directly compare the drift and diffusion estimates obtained under null and alternative. As we shall see, these two classes of tests have radically different asymptotic behaviour.

First, we introduce our transition-based tests: Under the alternative, we have the semiparametric estimate, $\hat{p}_{\mathrm{SP},k}(y|x)$, while under the null we assume an estimator of the parameters, $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2)$, is available. Under the null, the model is fully specified and the estimator $\tilde{\theta}$ could arrive from a range of standard parametric estimation methods such as maximum-likelihood (Aït-Sahalia, 2002; Kristensen and Shin, 2008) and method of moments (Bibby, Jacobsen and Sørensen, 2009; Hansen and Scheinkman, 1995). Associated with the fully parametric family of diffusion models under H_{P}, there exists a family of transition densities; this can be obtained by, for example, plugging the parametric drift and diffusion specification into eq. (2). We denote this family $p_{\mathrm{P}}(y|x;t,\theta) = p_{\mathrm{P}}\left(y|x;t,\mu(\cdot;\theta), \sigma^2(\cdot;\theta)\right)$, and we will again suppress the dependence on t when evaluated at $t = 1$. The estimated transition density under the null is then given by $\hat{p}_{\mathrm{P}}(y|x) := p_{\mathrm{P}}(y|x;\tilde{\theta})$. As with the semiparametric transition density estimator, $\hat{p}_{\mathrm{P}}(y|x)$ can in general not be written on closed form and numerical approximations have to be employed (Aït-Sahalia, 2002; Kristensen and Shin, 2008).

Given $\hat{p}_{\mathrm{P}}(y|x)$ and $\hat{p}_{\mathrm{SP},k}(y|x)$, we then propose to test H_{P} against $H_{\mathrm{SP},k}$ by:

$$T_{\mathrm{P},k} = \int_I \int_I [\hat{p}_{\mathrm{P}}(y|x) - \hat{p}_{\mathrm{SP},k}(y|x)]^2\, w(y,x)\, dydx, \quad k = 1, 2.$$

To analyse the asymptotic properties of these two tests, we impose the following assumptions on the parametric model and its estimators:

A.6 The estimator $\tilde{\theta}$ satisfies $\tilde{\theta} = \theta^{**} + \sum_{i=1}^n \psi_{\mathrm{P}}(X_i|X_{i-1})/n + o_P(1/\sqrt{n})$ with $E[\psi_{\mathrm{P}}(X_1|X_0)] = 0$ and $E[||\psi_{\mathrm{P}}(X_1|X_0)||^{2+\delta}] < \infty$ for some $\delta > 0$.

A.7 The transition density under H_{P}, $p_{\mathrm{P}}(y|x;\theta)$, and its first two derivatives w.r.t. θ exist, and they are all continuous w.r.t. (y, x) for all θ.

As with the estimators under the semiparametric nulls, (A.6) allows for misspecification and will only assume that θ^{**} is equal to the true value when working under H_{P}. We will in general suppress dependence on θ when evaluated at $\theta = \theta^{**}$. Sufficient conditions for the above assumption to hold for the MLE can be found in Aït-Sahalia (2002) and for GMM-type estimators in Bibby et al (2009). As we shall see, to derive the asymptotic distribution of $T_{P,k}$ under the null, it is critical that the estimators of the parametric components are $\sqrt{n}$-asymptotically normally distributed. This is in contrast to the semiparametric tests, $T_{\mathrm{SP},k}$, where we only need that they converge at a sufficiently fast rate.

Theorem 5 *For $k \in \{1,2\}$: Assume that (A.1)-(A.7), (B.1)-(B.3), and (H.k) hold. Then under H_{P}:*

$$nT_{\mathrm{P},k} \to^d \int\int Z_k^2(y,x)\, w(y,x)\, dydx,$$

for $k = 1,2$, where $Z_k(y,x)$ is a Gaussian process with covariance kernel

$$\bar{\Sigma}\left((x,y),(x',y')\right) = \Sigma_0\left((x,y),(x',y')\right) + \sum_{i=1}^{\infty} \Sigma_i\left((x,y),(x',y')\right),$$

$$\Sigma_i\left((x,y),(x',y')\right) = E\left[\left\{\frac{\partial p_{\mathrm{P}}(y|x;\theta)}{\partial\theta'}\psi_{\mathrm{P},0} - D_{k,0}(y|x)\right\}\left\{\frac{\partial p_{\mathrm{P}}(y'|x';\theta)}{\partial\theta'}\psi_{\mathrm{P},i} - D_{k,i}(y'|x')\right\}\right],$$

and $\psi_{\mathrm{P},i} := \psi_{\mathrm{P}}(X_i|X_{i-1})$.

The above test statistic has the interesting property that it converges with parametric rate even though it involves nonparametric kernel estimators. This is due to the fact that the transition density under the semiparametric alternative, $\hat{p}_{\mathrm{SP},k}(y|x)$, converges with parametric rate. Moreover, the limiting distributions depend on the asymptotics of the underlying parametric estimators. Both these features are in contrast to the ones of the semiparametric transition-based tests. Instead, the asymptotic behaviour of $T_{\mathrm{P},k}$, $k = 1,2$, is similar to those of omnibus tests such as the KS and CvM test; see, for example, Bhardwaj et al (2008, Theorem 3) and Escanciano (2009).

These omnibus-type features of the tests in particular means that they are able to detect "global" alternatives with parametric rate; on the other hand, due to the integration involved when computing the transition densities, $T_{\mathrm{P},k}$ cannot detect local (or high-frequency) departures. To see this, consider the following two contiguous alternatives:

$$H_{\mathrm{P},1}^c : \mu_n(x) = \mu(x;\theta_2^{**}) + g_n(x), \quad \sigma_n^2(x) = \sigma^2(x;\theta_1^{**}) \tag{26}$$

and

$$H_{\mathrm{P},2}^c : \mu_n(x) = \mu(x;\theta_2^{**}), \quad \sigma_n^2(x) = \sigma^2(x;\theta_1^{**}) + g_n(x). \tag{27}$$

Here, $H_{\mathrm{P},k}^c$ will be used to examine the power properties of $T_{\mathrm{P},k}$. Note that under $H_{\mathrm{P},1}^c$ the diffusion function is correctly specified and as such it is a constant sequence; this is to ensure that the maintained assumption, $H_{\mathrm{SP},1}$ is correct. Similarly with $H_{\mathrm{P},2}^c$. As with the semiparametric pseudo-true values, θ_1^{**} and θ_2^{**} will in general be drifting under $H_{\mathrm{P},2}^c$ and $H_{\mathrm{P},1}^c$ respectively, and the discussion following the introduction of $H_{\mathrm{SP},1}^c$ and $H_{\mathrm{SP},2}^c$ also applies here.

As before, we let $p_n(y|x) = p_n(y|x;\mu_n,\sigma_n^2)$ denote the data generating transition density, where μ_n and σ_n^2 are given either by $H_{\mathrm{P},1}^c$ or $H_{\mathrm{P},2}^c$. We then obtain under $H_{\mathrm{P},k}^c$:

$$p_{\mathrm{P}}(y|x) = p_n(y|x) + \gamma_{\mathrm{P},k}^{(n)}(y|x) + O(R_{\mathrm{P}}), \tag{28}$$

where $R_{\mathrm{P}} = \sup_{x\in I}|g_n(x)|^2$,

$$\gamma_{\mathrm{P},1}^{(n)}(y|x) = \frac{1}{2}\int_I g_n(w)\,\bar{p}_\mu(y,x,w)\,dw, \quad \gamma_{\mathrm{P},2}^{(n)}(y|x) = \int_I g_n(w)\,\bar{p}_{\sigma^2}(y,x,w)\,dw,$$

and $\bar{p}_\mu(y,x,w)$ and $\bar{p}_{\sigma^2}(y,x,w)$ are defined as in the previous section, except that $p_{\mathrm{P}}(y|x)$ replaces $p_{\mathrm{SP},k}(y|x)$. The expression in eq. (28) can then in turn be used to derive asymptotic expansions of the tests under contiguous alternatives:

Theorem 6 *For $k \in \{1,2\}$: Assume that (A.1)-(A.7) and (B.1)-(B.3), and (H.k) hold. Then under $H^c_{P,k}$,*

$$\begin{aligned} nT_{\mathrm{P},k} &= \int\int Z^2_{n,k}(y,x)\,w(y,x)\,dydx + n\int_I\int_I \gamma^{(n)}_{\mathrm{P},k}(y|x)^2\,w(y,x)\,dydx \\ &\quad +2\sqrt{n}\int_I\int_I Z_{n,k}(x,y)\,\gamma^{(n)}_{\mathrm{P},k}(y|x)\,w(y,x)\,dydx + O_P\left(nR^2_{\mathrm{P}}\right) + O\left(R_{\mathrm{P}}/\sqrt{n}\right), \end{aligned}$$

where $Z_{n,k} \to^d Z_k$ on the support of w.

From the expression in Theorem 6, it is easily seen that $T_{\mathrm{P},k}$ can detect global alternatives on the form $g_n(x) = a_n g(x)$ for which $\lim_{n\to\infty} na^2_n > 0$. Thus, it can detect global alternatives vanishing at parametric rate, $a_n = O\left(n^{-1/2}\right)$. On the other hand, local alternatives on the form $g_n(x) = a_n g((x - x_0)/b_n)$ are not as easily detected. For this class of alternatives, we obtain

$$\gamma^{(n)}_{\mathrm{P},1}(y|x) = a_n\int_I g\left(\frac{x - x_0}{b_n}\right)\bar{p}_\mu(y,x,w)\,dw = a_n b_n \bar{p}_\mu(y,x,x_0) \times \int_I g(z)\,dz,$$

and similarly for $\gamma^{(n)}_{\mathrm{P},2}(y|x)$. Thus, deviations can only be detected if $\int_I g(z)\,dz \neq 0$ and $\lim_{n\to\infty} na^2_n b^2_n > 0$. This is akin to the semiparametric tests, and the discussion of these also applies here. In conclusion, the transition-based tests may not be suitable when the interest lies in detecting local, "high-frequency" departures in the drift and diffusion function from the null.

The problem with the transition-based tests lies in the fact that they integrate out the deviations appearing in the drift and/or diffusion function. We therefore introduce two alternative test statistics that directly compare the fully parametric and semiparametric estimators of the drift and diffusion function. Define

$$\bar{T}_{\mathrm{P},1} = \int_I [\mu(x;\tilde{\theta}_1) - \hat{\mu}_{\mathrm{SP},1}(x)]^2\bar{w}(x)\,dx, \quad \bar{T}_{\mathrm{P},2} = \int_I [\sigma^2(x;\tilde{\theta}_2) - \hat{\sigma}^2_{\mathrm{SP},2}(x)]^2\bar{w}(x)\,dx, \tag{29}$$

for some weighting function $\bar{w} : I \mapsto \mathbb{R}_+$. These tests are similar to the ones proposed in eq. We will assume that $\bar{w}$ has compact support which in particular implies that trimming of the semi-nonparametric estimators is not required; thus, we may use the ones given in eqs. (9)-(10) instead of eqs. (15)-(16).

Here, $\bar{T}_{\mathrm{P},k}$ tests H_{P} against $H_{\mathrm{SP},k}$, $k = 1, 2$. The intuition behind these two alternative test statistics is similar to the one for $T_{\mathrm{P},1}$ and $T_{\mathrm{P},2}$, but instead of measuring deviations from the null in terms of the transition densities we now directly measure discrepancies appearing in the drift or diffusion functions. To get a better understanding of what $\bar{T}_{\mathrm{P},k}$ is actually testing, it is worth noting that under the null $\bar{T}_{\mathrm{P},1} \approx I_0(\omega_{10}) + I_1(\omega_{11})$ and $\bar{T}_{\mathrm{P},2} \approx I_0(\omega_2)$, where

$$I_k(\omega) = \int_I [\hat{\pi}^{(k)}(x) - \pi^{(k)}(x)]^2\omega(x)\,dx, \tag{30}$$

for $k = 0, 1$, and ω_{10}, ω_{11} and ω_2 are appropriately chosen weighting functions (see the proof of Theorem 7 below for details). This highlights that $\bar{T}_{\mathrm{P},1}$ and $\bar{T}_{\mathrm{P},2}$ to a large extent are testing the correct specification of the marginal density as implied by the parametric specification under H_{P} against its nonparametric alternative. As such the tests are similar to the ones proposed in Aït-Sahalia (1996b) and Huang (1997).

This could also seem to indicate that one could instead use $I_k(\omega)$, $k = 0, 1$, to test H_{P} against $H_{\mathrm{SP},1}$ and $H_{\mathrm{SP},2}$. However, observe that $\bar{T}_{\mathrm{P},1}$ and $\bar{T}_{\mathrm{P},2}$ involve nontrivial transformations of the marginal density and therefore test different directions of departure from the null with special emphasis on the correct specification of the drift and diffusion respectively. In particular, when one specifies deviations from the null in terms of the drift and diffusion terms, then $I_0(\omega)$ and

$I_1(\omega)$ will distort some of the local features in the drift and diffusion term; see the discussion following Theorem 8 below for more details.

We also note that $\bar{T}_{\mathrm{P},2}$ shares some similarities with the specification tests proposed in Corradi and White (1999) and Li (2007). These two studies are only concerned with testing the correct specification of the diffusion term, and propose to test a given specification of σ^2 using $\bar{T}_{\mathrm{P},2}$ as given in Eq. (29) except that they employ the nonparametric estimator of $\sigma^2(\cdot)$ proposed in Florens-Zmirou (1989); see also Bandi and Phillips (2003). The advantage of the estimator of Florens-Zmirou (1989) is that it does not require as input a preliminary estimator of the drift function (as ours do). On the other hand, the estimator of Florens-Zmirou (1989) requires high-frequency observations and is only consistent as time distance between observations shrinks to zero, $\Delta \to 0$, sufficiently fast as $n \to \infty$ (c.f. Nicolau, 2003). So for low frequency data, the tests of Corradi and White (1999) and Li (2007) will be biased, and will not have a well-defined asymptotic distribution under the null.

The theorem is shown under the following regularity condition on the weighting function:

B.4 The weighting function $\bar{w} : I \mapsto \mathbb{R}_+$ is continuous and has compact support.

The discussion that followed Assumption B.3 also applies here. We are now able to derive the following result concerning the asymptotic distributions of the tests under the null:

Theorem 7 *Assume (A.1)-(A.6), (B.1) and (B.4) hold. Then under H_{P}:*

(i) As $nh^{m+5} \to 0$, $nh^{4m+5/2} \to 0$ and $nh^{1/2}/\log(n)^2 \to \infty$,

$$nh^{5/2}\frac{\bar{T}_{\mathrm{P},1} - \bar{m}_{\mathrm{P},1}}{\bar{v}_{\mathrm{P},1}} \to^d N(0,1),$$

where

$$\bar{m}_{\mathrm{P},1} = \frac{1}{4nh^3}\int_{\mathbb{R}} K'(z)^2\,dz \times \int_I \frac{\sigma^4(x)\,\bar{w}(x)}{\pi(x)}dx + \frac{1}{4nh}\int_{\mathbb{R}} K^2(z)\,dz \times \int_I \frac{\sigma^4(x)\,\bar{w}^2(x)\,\pi'(x)^2}{\pi^4(x)}dx,$$

$$\bar{v}^2_{\mathrm{P},1} = \frac{1}{8}\int_{\mathbb{R}} \left(K' * K'\right)^2(z)\,dz \times \int \pi^2(x)\sigma^8(x)\,\bar{w}^2(x)\,dx.$$

(ii) As $nh^{2m+1} \to 0$, $nh^{4m+1/2} \to 0$, and $nh^{3/2}/\log(n)^2 \to \infty$,

$$nh^{1/2}\frac{\bar{T}_{\mathrm{P},2} - \bar{m}_{\mathrm{P},2}}{\bar{v}_{\mathrm{P},2}} \to^d N(0,1),$$

where

$$\bar{m}_{\mathrm{P},2} = \frac{4}{nh}\int_{\mathbb{R}} K^2(z)\,dz \times \int_I \frac{\sigma^4(x)\,\bar{w}(x)}{\pi(x)}dx,$$

$$\bar{v}^2_{\mathrm{P},2} = 32\int_{\mathbb{R}} (K * K)^2(z)\,dz \times \int_I \frac{\sigma^8(x)\,\bar{w}^2(x)}{\pi^2(x)}dx$$

Consistent estimates of $\bar{m}_{\mathrm{P},k}$ and $\bar{v}^2_{\mathrm{P},k}$ can be obtained by substituting the unknown quantities entering these, that is, $\sigma^2(x)$ and $\pi(x)$, for their estimates. As part of the proof of Theorem 7, we derive asymptotic expansions of the two test statistics similar to those stated for the semi-parametric test statistics in Theorem 3. These expansions include additional higher-order terms which vanish under the restrictions imposed on the bandwidth in Theorem 7.

In contrast to the transition-based tests, $T_{\mathrm{P},1}$ and $T_{\mathrm{P},2}$, the above alternative tests converge with nonparametric rates and have standard normal distributions. This owes to the fact that in $\bar{T}_{\mathrm{P},1}$ and $\bar{T}_{\mathrm{P},2}$, the semi-nonparametric estimators, $\hat{\mu}_{\mathrm{SP},1}(x)$ and $\hat{\sigma}^2_{\mathrm{SP},2}(x)$, are not integrated over, and as such the asymptotic properties are similar to other kernel-based test statistics, c.f. Theorems 1 and 3.

One could consider a number of modified versions of the above test statistics by following the ideas of Kristensen (2007) and replace the parametric estimators of the drift (in $\bar{T}_{\mathrm{P},1}$) or diffusion (in $\bar{T}_{\mathrm{P},2}$) with kernel smoothed versions. As shown in that study, this removes some of the higher-order terms in the asymptotic expansions of the resulting test statistics such that weaker restrictions on allowable bandwidth sequences are needed. However, as demonstrated in Fan (1994), these modifications alter the power properties of the tests.

We now examine the power properties of the tests to add further insight to their (asymptotic) performance. We do this by revisiting the sequence of alternatives specified in eqs. (26)-(27).

Theorem 8 *Assume (A.1)-(A.6), (B.1) and (B.4) hold. Then:*

(i) Under $H^c_{\mathrm{P},1}$, as $nh^3 \to \infty$, and $nh^{3/2+2m} \to 0$:

$$nh^{5/2}\left\{\bar{T}_{\mathrm{P},1} - \bar{m}_{\mathrm{P},1}\right\} = \bar{v}_{\mathrm{P},1}U_{n,1} + nh^{5/2}\int_I g_n^2(x)\,\bar{w}(x)\,dx + o_P(1),$$

where $U_{n,1} \to^d N(0,1)$.

(ii) Under $H^c_{\mathrm{P},2}$, as $nh \to \infty$, and $nh^{1/2+2m} \to 0$,

$$nh^{1/2}\left\{\bar{T}_{\mathrm{P},2} - \bar{m}_{\mathrm{P},2}\right\} = \bar{v}_{\mathrm{P},2}U_{n,2} + nh^{1/2}\int_I g_n^2(x)\,\bar{w}(x)\,dx + o_P(1),$$

where $U_{n,2} \to^d N(0,1)$.

The above expressions reveal that $\bar{T}_{\mathrm{P},1}$ and $\bar{T}_{\mathrm{P},2}$ can only detect global alternatives on the form $g_n(x) = a_n g(x)$ for which $\lim_{n\to\infty} nh^{5/2}a_n^2 > 0$ and $\lim_{n\to\infty} nh^{1/2}a_n^2 > 0$ respectively. Thus, they are less powerful than $T_{\mathrm{P},k}$, $k = 1, 2$, in this regard. However, they are better at detecting local deviations from the null: For alternatives on the form $g_n(x) = a_n g((x - x_0)/b_n)$, we obtain

$$\int_I g_n^2(x)\,\bar{w}(x)\,dydx = a_n^2\int_I g^2\left(\frac{x - x_0}{b_n}\right)\bar{w}(x)\,dx = a_n^2 b_n \bar{w}(x_0)\int_I g^2(z)\,dz.$$

Thus, the tests can detect alternatives for which $\int_I g(z)\,dz = 0$, and the rates at which they can detect alternatives are $\lim_{n\to\infty} nh^{5/2}a_n^2 b_n > 0$ and $\lim_{n\to\infty} nh^{1/2}a_n^2 b_n > 0$ respectively. For suitable choices of h, high-frequency alternatives can therefore be detected by $\bar{T}_{\mathrm{P},1}$ and $\bar{T}_{\mathrm{P},2}$ at a better rate compared to $T_{\mathrm{P},1}$ and $T_{\mathrm{P},2}$; see Rosenblatt (1975) and Ghosh and Huang (1991) for related results.

The results of Theorems 6 and 8 are comparable to the ones found in the literature on testing for correct specifications of distributions using either nonparametric kernel density estimators or cumulative density function estimators (see e.g. Eubank and LaRiccia, 1992). In conclusion, depending on the type of alternatives of interest, one should either employ $T_{P,k}$ or $\bar{T}_{P,k}$, $k = 1, 2$.

5 Related Tests

We here briefly discuss the misspecification tests proposed by Aït-Sahalia (1996b) and Aït-Sahalia et al (2009) in relation to our tests.

As noted in the previous section, the two tests $\bar{T}_{\mathrm{P},1}$ and $\bar{T}_{\mathrm{P},2}$ are somewhat similar to the one proposed in Aït-Sahalia (1996b) which is on the form $I_0(m)$ as given in eq. (30). This test was originally proposed to test H_{P} against H_{NP}, but as noted above it seems more suitable for testing the parametric hypothesis against either $H_{\mathrm{SP},1}$ or $H_{\mathrm{SP},2}$. To see how our tests perform relative to this one, we analyse the power properties of a test based on $I_0(m)$: Consider again the contiguous alternative $H^c_{\mathrm{P},1}$. Using eq. (5), we obtain the following marginal density implied by the null,

$$\pi_{\mathrm{P}}(x) = \frac{M_{x^*}}{\sigma^2_{\mathrm{P}}(x)} \exp\left[2\int_{x^*}^{x} \frac{\mu_{\mathrm{P}}(y)}{\sigma^2_{\mathrm{P}}(y)} dy\right],$$

while the sequence of contiguous densities are given by

$$\pi_n(x) = \frac{M_{x^*}}{\sigma^2_{\mathrm{P}}(x)} \exp\left[2\int_{x^*}^{x} \frac{\mu_{\mathrm{P}}(y) + g_n(x)}{\sigma^2_{\mathrm{P}}(y)} dy\right] = \pi_{\mathrm{P}}(x)\exp\left[2\int_{x^*}^{x} \frac{g_n(x)}{\sigma^2_{\mathrm{P}}(y)} dy\right].$$

Thus, by using the same arguments as in the proof of Theorem 8, we obtain under $H^c_{\mathrm{P},1}$ that

$$nh^{1/2}\{I_0(m) - c_0\} = v_0 U_n + nh^{1/2}\int_I \left\{\exp\left[2\int_{x^*}^{x} \frac{g_n(x)}{\sigma^2_{\mathrm{P}}(y)} dy\right] - 1\right\} \pi_{\mathrm{P}}(x) m(x) dx + o_P(1),$$

for suitably defined parameters c_0 and v_0, and where $U_n \to^d N(0,1)$. This shows that $I_0(m)$ is not tailored to detect the deviation, $g_n(x)$. In particular, $g_n(x)$ is integrated over twice which has as consequence that $I_0(m)$ will suffer from similar issues as the transition-based tests. In contrast, $\bar{T}_{P,1}$ is designed to directly capture any deviations between $\mu_{\mathrm{P}}(x)$ and $\mu(y)$, c.f. Theorem 8(i). A similar analysis can be carried out under $H^c_{\mathrm{P},2}$.

Finally, to demonstrate that the reported poor power against both global and local deviations is a general problem for transition-based tests, we revisit one of the tests developed in Aït-Sahalia et al (2009). They propose to test H_{P} against H_{NP} by

$$T_{\mathrm{P}} = \int_I \int_I [\hat{p}_{\mathrm{P}}(y|x) - \hat{p}_{\mathrm{NP}}(y|x)]^2 w(y,x) dydx.$$

This test has the same asymptotic distribution as $T_{\mathrm{SP},k}$, $k = 1,2$, under the null of H_{P}. To investigate its power properties, we consider the following contiguous alternative,

$$H^c_{\mathrm{P}} : \mu_n(x) = \mu(x;\theta^{**}_2) + f_n(x), \quad \sigma^2_n(x) = \sigma^2(x;\theta^{**}_1) + g_n(x),$$

for two sequences $f_n(x)$ and $g_n(x)$ that measure deviations in the drift and diffusion term respectively. By following the same arguments as employed previously, it is easily shown that

$$p_{\mathrm{P}}(y|x) = p_n(y|x) + \gamma^{(n)}_{\mathrm{P}}(y|x) + O(R_{\mathrm{P}}),$$

where $R_{\mathrm{P}} = \sup_{x\in I}|f_n(x)|^2 + \sup_{x\in I}|g_n(x)|^2$ and

$$\gamma^{(n)}_{\mathrm{P}}(y|x) = \int_I f_n(w)\bar{p}_{\mu}(y,x,w) dw + \int_I g_n(w)\bar{p}_{\sigma^2}(y,x,w) dw.$$

By using the same arguments as in the proof of Theorem 3, we now obtain that

$$nh_{\mathrm{NP}}\{T_{\mathrm{P}} - m_{\mathrm{SP}}\} = v_{\mathrm{SP}}U_{n,1} + nh_{\mathrm{NP}}\int_I\int_I \gamma^{(n)}_{\mathrm{P}}(y|x)^2 w(y,x) dydx + O_P\left(R^2_{\mathrm{P}}\right) + o_P(1).$$

From this expression, we see that T_{P} cannot detect Pitman alternatives on the form $f_n(x) = a_n f(x)$ and $g_n(x) = a_n g(x)$ for which $\int_I f(w)\bar{p}_{\mu}(y,x,w) dw = 0$ and $\int_I g(w)\bar{p}_{\sigma^2}(y,x,w) dw = 0$. Moreover, for local alternatives on the form $g_n(x) = a_n f((x-x_0)/b_n)$ and $g_n(x) = a_n g((x-x_0)/b_n)$,

$$\gamma^{(n)}_{\mathrm{P}}(y|x) \simeq a_n b_n\left\{\bar{p}_{\mu}(y,x,x_0) \times \int_I f(z) dz + \bar{p}_{\sigma^2}(y,x,x_0) \times \int_I g(z) dz\right\},$$

and so T_{P} can only detect local alternatives for which $\lim_{n\to\infty} nh_{\mathrm{NP}}a^2_n b^2_n > 0$. Moreover, alternatives which satisfy $\int_I f(z) dz = \int_I g(z) dz$ are not detectable.

Aït-Sahalia et al (2009) also conduct a power analysis of T_{P} and conclude that it can detect local alternatives without any of the aformentioned problems. This seeming contradiction between our results and the ones of Aït-Sahalia et al (2009) are due to different formulations of alternatives: While Aït-Sahalia et al (2009) express their alternatives in terms of the transition density, we formulate them directly in terms of the underlying drift and diffusion functions.

In conclusion, departures from the drift and diffusion functions imposed under the null are in general not easily detected by transition-based tests since the deviations are smoothed out when the drift and diffusion functions are plugged into the transition density.

6 Markov Bootstrap Tests

The asymptotic distributions of the proposed test statistics derived in the previous section ignore several higher-order terms that will affect the finite-sample distributions: First, all asymptotic distributions, except the ones of $T_{\mathrm{P},1}$ and $T_{\mathrm{P},2}$, do not involve estimation errors due to unknown parametric components and additional covariance terms due to dependence in data. Second, they all are based on first-order linearisations of the test statistics and thereby ignore second-order terms. Third, various bias terms due to the kernel smoothing are not present. Fourth, in the implementation, we need to estimate unknown quantities entering the asymptotic distributions, which adds additional estimation errors to the tests.

In finite samples, the distributions will clearly depend on these additional components, and as such one could fear that the asymptotic distribution stated in the theorems may deliver a poor finite sample approximations. We therefore propose Markov bootstrap versions of the tests which are expected to perform better than the ones relying on approximations based on the asymptotic distribution. The simulation studies in Aït-Sahlia et al (2009) and Li and Tszask (2006) of Bootstrap versions of their nonparametric tests support this conjecture.

In the Markov bootstrap versions of the tests, we draw a new sample from the transition density under the relevant null, and use this sample to approximate the relevant distributions. The proposed bootstrap is similar to the one proposed by Fan (1995) in a cross-sectional setting and Li and Tszask (2006) in a time series setting. We also note that our proposal shares some similarities with the Markov bootstrap procedures examined in Horowitz (2003) and Andrews (2005) but in different settings, while Bhardwaj et al (2008) and Corradi and Swanson (2005) propose to use a block bootstrap in conjunction with their specification tests for diffusion models.

Let in the following T_n denote any one of the test statistics developed in the previous section, and $\hat{p}_0\left(y|x\right)$ and $\hat{\pi}_0$ denote the transition density and stationary density estimated under the relevant null ($H_{\mathrm{SP},1}$, $H_{\mathrm{SP},2}$ or H_{P}). The proposed bootstrap then proceeds as follows:

Step 1 Draw $X_0^* \sim \hat{\pi}_0$, and recursively $X_i^* \sim \hat{p}_0(\cdot|X_{i-1}^*)$, $i = 1, ..., n$.

Step 2 Replace the data $\{X_i\}_{i=1}^n$ with the bootstrap sample $\{X_i^*\}_{i=1}^n$ in the computation of estimators and test statistics; we denote the resulting test statistic T_n^*.

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

The initialisation in Step 1 could be exchanged for $X_0^* = X_0$ since we have a geometrically ergodic Markov chain. Since $\hat{p}_0(y|x)$ in general is not available on closed form, we propose to draw from it by utilising an Euler discretization scheme (see e.g. Corradi and Swanson, 2005; Gourieroux, Monfort and Renault, 1993). This will involve an additional error, but this can be controlled for by choosing a sufficiently small time step.

By relying on arguments similar to those in Bhardwaj et al (2008), Corradi and Swanson (2005) and Li and Tszask (2006), one should be able to show that the proposed Bootstrap versions of the parametric tests are consistent under suitable conditions. It should be noted though that in order to show consistency of the Bootstrap versions of the semiparametric tests, we first need to ensure that the bootstrap sample as generated by $\hat{p}_{\mathrm{SP},k}(y|x)$ is stationary and β-mixing. To this end, we need to further modify the semiparametric estimator of the drift functions, to ensure mean reversion.

One could potentially also use the Markov bootstrap to construct confidence bands for the semiparametric estimators.

7 Finite-Sample Performance of Estimators

We here examine how the semi-nonparametric estimators perform in finite samples. We choose as data generating models the CKLS model of Cha, Karolyi, Longstaff and Sanders (1992),

$$dX_t = \{\beta_1 + \beta_2 X_t\}\, dt + \sqrt{\alpha_1 X_t^{\alpha_2}} dW_t, \tag{CKLS}$$

and a restricted version of the model proposed in Aït-Sahalia (1996b),

$$dX_t = \left\{\beta_1 + \beta_2 X_t + \beta_3 X_t^2 + \beta_4 X_t^{-1}\right\} dt + \sqrt{\alpha_1 X_t^{\alpha_2}} dW_t. \tag{AS}$$

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 Aït-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 either CKLS or AS is the data generating process respectively: CKLS 1: $\mu(x)$ unknown and $\sigma^2(x) = \alpha_1 x^{\alpha_2}$; CKLS 2: $\mu(x) = \beta_1 + \beta_2 x$ and $\sigma^2(x)$ unknown; AS 1: $\mu(x)$ unknown and $\sigma^2(x) = \alpha_1 x^{\alpha_2}$; and AS 2: $\mu(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^{-1}$ and $\sigma^2(x)$ unknown. The parameters of the semiparametric models are estimated using the method proposed in Kristensen (2010). 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 (2008). This is implemented by simulating $N = 100$ values for each observation, using the Euler scheme with a step length of $\delta = \Delta/10$ (see Kristensen, 2010, for more details)

We first investigate the behaviour of the semi-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 fully parametric and semi-nonparametric drift estimates are plotted for the parameters (i) and (ii) respectively. For (i), Figure 1 shows that the semi-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 reported in Figure 2 where the semi-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. Interestingly, the shape of the mean of the semi-nonparametric diffusion estimator in Figure 4 is very similar to the one reported in Aït-Sahalia (1996a).

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 (2010, Section 6). The semi-nonparametric drift estimator performs fairly well, and has more or less the same level of precision as the parametric one. The performance of the semi-nonparametric diffusion estimator is not quite so good though.

8 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. However, if one is willing to restrict attention to multivariate models which does satisfy this relation, the proposed estimation and testing procedures 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,

$$\mu_i(x) = \frac{1}{2\pi(x)} \sum_{j=1}^{d} \frac{\partial}{\partial x_j} \left[\sigma_{ij}^2(x)\pi(x)\right]. \tag{31}$$

This restriction is for example imposed by Chen, Hansen and Scheinkman (2010) 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.

As revealed in the power analysis in Section 3.1 and 3.2, a more suitable class of tests for testing a fully nonparametric diffusion alternative against either the semiparametric or fully parametric nulls would be ones proposed in eq. (25). The analysis of these would be a useful addition to the one conducted here.

9 Acknowledgements

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, and Jianqing Fan for stimulating discussions. Parts of this paper appeared in an earlier version titled "Estimation in Two Classes of Semiparametric Diffusion Models". Financial Support from the Danish Social Sciences Research Council (through CREATES) and the National Science Foundation (grant no. SES-0961596) is gratefully acknowledged. Parts of this research was conducted while the author visited Princeton University and University of Copenhagen whose hospitality is gratefully acknowledged.

References

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.

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

Andrews, D.W.K., 1994, Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43-72.

Andrews, D.W.K., 1995, Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560-596.

Andrews, D.W.K., 2005, Higher-order improvements of the parametric bootstrap for Markov processes, in D.W.K. Andrews and J.H. Stock (Eds.): *Identification and Inference for Econometric Models: A Festschrift in Honor of T.J. Rothenberg.* Cambridge University Press, Cambridge.

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

Bandi, F.M. and P.C.B. Phillips, 2005, A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics* 137, 354-395.

Bhardwaj, G., V. Corradi and N.R. Swanson, 2005, A simulation-based specification test for diffusion processes. *Journal of Business & Economic Statistics* 26, 176-193.

Bibby, B.M., M. Jacobsen and M. Sørensen, 2009, Estimating functions for discretely sampled diffusion-type models, in Aït-Sahalia, Y. and Hansen, L.P. (Eds.): *Handbook of Financial Econometrics*, Vol. 1. Elsevier, Amsterdam.

Bickel, P. J. and M. Rosenblatt, 1973, On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071-1095.

Björk, T.,2004, *Arbitrage theory in continuous time, 2nd Edition.* Oxford University Press, Oxford.

Chan, K.C., , G.A. Karolyi, F.A. Longstaff and A.B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209-1227.

Chen, S. X., J. Gao and C. Y. Tang, 2009, A test for model specification of diffusion processes. *Annals of Statistics* 36, 167-198.

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

Chen, X., L.P. Hansen and J. Scheinkman, 2010, Nonlinear principal components and long run implications. Forthcoming in *Annals of Statistics.*

Corradi, V. and N.R. Swanson, 2005, Bootstrap specification tests for diffusion processes. *Journal of Econometrics* 124, 117-148.

Corradi, V. and H. White, 1999, Specification tests for the variance of a diffusion. *Journal of Time Series Analysis* 20, 253-270.

Doukhan, P., P. Massart and E. Rio, 1994, The central limit theorem for strongly mixing processes. *Annales de l'Institut Henri Poincaré, Section B*, 30, 63-82.

Doukhan, P., P. Massart and E. Rio, 1995, Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincaré, Section B*, 31, 393-427.

Escanciano, J.C., 2009, On the lack of power of omnibus specification tests. *Econometric Theory* 25, 162-194.

Eubank, R.L. and V.N. LaRiccia, 1992, Asymptotic comparison of Cramér-Von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Annals of Statistics* 20, 2071-2086.

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, J., C. Zhang and J. Zhang, 2001, Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153–193.

Florens-Zmirou, D., 1993, On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790-804.

Friedman, A., 1976, *Stochastic differential equations and applications, vol. 1.* Academic Press, New York.

Ghosh, B.K. and W.-M. Huang, 1991, The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Annals of Statistics* 19, 999-1009.

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

Gouriéroux, C., A. Monfort, E. Renault, 1993, Indirect inference. *Journal of Applied Econometrics* 8, S85-S118.

Gouriéroux, C. and C. Tenreiro, 2001, Local power properties of kernel based goodness of fit tests. *Journal of Multivariate Analysis* 78, 161-190.

Hansen, L.P. and J.A. Scheinkman, 1995, Back to the future: Generating moment implications for continuous time Markov processes. *Econometrica* 63, 767-804.

Härdle, W. and E. Mammen, 1993, Comparing nonparametric versus parametric regression fits. *Annals of Statististics* 21, 1926-1947.

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

Horowitz, J.L., 2003, Bootstrap methods for Markov processes. *Econometrica* 71, 1049–1082.

Huang, L.-S., 1997, Testing goodness-of-fit based on a roughness measure. *Journal of the American Statistical Association* 92, 1399-1402.

Karlin, S. and H.M. Taylor, 1981, *A second course in stochastic processes.* Academic Press, New York.

Kristensen, D., 2007, Nonparametric estimation and misspecification testing of diffusion models. CREATES Research Papers 2007-1, University of Aarhus.

Kristensen, D., 2008, Estimation of partial differential equations with applications in finance. *Journal of Econometrics* 144, 392-408.

Kristensen, D., 2009, Uniform convergence rates of kernel estimators with heterogeneous, dependent data. *Econometric Theory* 25, 1433-1445.

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

Kristensen, D. and Y. Shin, 2008, Estimation of dynamic models with nonparametric simulated maximum likelihood. CREATES Research Papers 2008-58, University of Aarhus.

Li, F., 2007, Testing the parametric specification of the diffusion function in a diffusion process. *Econometric Theory* 23, 221-250.

Li, F., and Tkacz, G., 2006, A consistent bootstrap test for conditional density functions with time-Series data. *Journal of Econometrics* 133, 863–886.

Meyn, S.P. and R.L. Tweedie, 1993, Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability* 25, 518-548.

Negri, I. and Nishiyama, Y., 2009, Goodness of fit test for ergodic diffusion processes. *Annals of the Institute of Statistical Mathematics* 61, 919-928.

Nicolau, J., 2003, Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory* 19, 754-777.

Robinson, P.M., 1983, Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–297.

Robinson, P.M., 1988, Root-n-consistent semiparametric regression. *Econometrica* 56, 931-954.

Robinson, P.M., 1991, Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58, 437-453.

Rosenblatt, M., 1975, A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3, 1-14.

Whang, Y.-J. and D.W.K. Andrews, 1993, Tests of specification for parametric and semiparametric models. *Journal of Econometrics* 57, 277-318.

White, H., 1982, Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25.

Wong, E., 1964, The construction of a class of stationary Markoff processes, in: R. Bellman (Ed.) *Sixteenth Symposium in Applied Mathematics - Stochastic Processes in Mathematical Physics and Engineering*, pp. 264-276. American Mathematical Society, Providence.

A Proofs

Proof of Theorem 1. To show the first part of the theorem, write

$$\begin{aligned}
\hat{\mu}(x)-\mu(x) &= \frac{1}{2}\sigma^2(x;\theta)\left[\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)}-\frac{\pi^{(1)}(x)}{\pi(x)}\right]+\frac{1}{2}\left[\frac{\partial\sigma^2(x;\hat{\theta})}{\partial x}-\frac{\partial\sigma^2(x;\theta)}{\partial x}\right] \\
&\quad+\frac{\hat{\pi}^{(1)}(x)}{2\hat{\pi}(x)}\left[\sigma^2(x;\hat{\theta})-\sigma^2(x;\theta_0)\right] \\
&=: A_1(x)+A_2(x)+A_3(x).
\end{aligned}$$

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

$$\frac{\partial^i\sigma^2(x;\hat{\theta})}{\partial x^i}-\frac{\partial^i\sigma^2(x;\theta_0)}{\partial x^i}=\frac{\partial^{i+1}\sigma^2\left(x;\bar{\theta}_i\right)}{\partial x^i\partial\theta'}(\hat{\theta}-\theta)=O_P\left(1/\sqrt{n}\right),$$

for some $\bar{\theta}_1\in[\theta_1,\hat{\theta}_1]$, $i=0,1$. We expand $A_1(x)$ in terms of $\pi^{(i)}(x)$, $i=0,1$:

$$\begin{aligned}
\sqrt{nh^3}A_1(x) &= \frac{\sigma^2(x;\theta)}{2\pi_0(x)}\sqrt{nh^3}[\hat{\pi}^{(1)}(x)-\pi_0^{(1)}(x)]-\frac{\pi_0^{(1)}(x)\,\sigma^2(x;\theta)}{2\pi_0^2(x)}\sqrt{nh^3}\left[\hat{\pi}(x)-\pi_0(x)\right] \\
&\quad+\sqrt{nh^3}O\left(|\hat{\pi}^{(1)}(x)-\pi_0^{(1)}(x)|^2+|\hat{\pi}(x)-\pi_0(x)|^2\right).
\end{aligned}$$

Using standard methods for kernel estimators, see Robinson (1983), we obtain, as $nh^{1+2i}\to\infty$ and $nh^{1+2(i+m)}\to 0$,

$$\sqrt{nh^{1+2i}}(\hat{\pi}^{(i)}(x)-\pi_0^{(i)}(x))\xrightarrow{d}N\left(0,V_i(x)\right),\quad i=0,1, \tag{32}$$

where $V_0(x)=\pi(x)\int K^2(z)\,dz$ and $V_1(x)=\pi(x)\int K^{(1)}(z)^2\,dz$, while the two remainder terms in $A_1(x)$ are $o_P(1)$. The weak convergence result in the first part of the theorem now follows from Slutsky's Theorem.

To show the second part of the theorem, write

$$\begin{aligned}
\hat{\sigma}^2(x)-\sigma^2(x) &= 2\int_l^x\mu(y;\theta_2)\pi(y)\,dy\left\{\frac{1}{\hat{\pi}(x)}-\frac{1}{\pi(x)}\right\} \\
&\quad+\frac{2}{\hat{\pi}(x)}\frac{1}{n}\sum_{i=1}^n\left\{\mu(X_i;\hat{\theta}_2)-\mu(X_i;\theta_2)\right\}\mathbb{I}\left\{X_i\leq x\right\} \\
&\quad+\frac{2}{\hat{\pi}(x)}\frac{1}{n}\sum_{i=1}^n\left\{\mu(X_i;\theta_2)\mathbb{I}\left\{X_i\leq x\right\}-\int_l^x\mu(y;\theta_2)\pi(y)\,dy\right\} \\
&=: B_1(x)+B_2(x)+B_3(x),
\end{aligned}$$

where $B_3(x)=O_P(1/\sqrt{n})$ by the CLT for mixing processes, c.f. Doukhan et al (1994), and

$$B_3(x)=\frac{2}{\hat{\pi}(x)}\left\{\frac{1}{n}\sum_{i=1}^n\frac{\partial\mu(X_i;\bar{\theta}_2)}{\partial\theta'}\mathbb{I}\left\{X_i\leq x\right\}\right\}(\hat{\theta}_2-\theta_2)=O_P\left(n^{-1/2}\right),$$

for some $\bar{\theta}_2\in[\theta_2,\hat{\theta}_2]$. Regarding $B_1(x)$, first note that

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

for some $\lambda\in[0,1]$. Using standard results for kernel estimators, see Robinson (1983), the second term on the left hand side is $O_P(h^{2m})+O_P(1/(nh))$. The weak convergence result now follows from eq. (32) combined with Slutsky's Theorem. ■

Proof of Lemma 2. Define $\bar{\mu}_{\mathrm{SP},k}(x;\theta_k) = \tau_a(\pi(x))\,\mu_{\mathrm{SP},k}(x;\theta_k)$ and $\bar{\sigma}^2_{\mathrm{SP},k}(x;\theta_k) = \tau_a(\pi(x))\,\sigma^2_{\mathrm{SP},k}(x;\theta_k) + \underline{\sigma}^2(1-\tau_a(\pi(x)))$, and let $\bar{p}_{\mathrm{SP},k}(y|x;\theta_k)$ denote the transition density corresponding to these trimmed versions. In the following we suppress their dependence on the parameter when evaluated at the true value. We employ Kristensen (2010, Lemma 5) in conjunction with the uniform convergence results in Lemma 9 to obtain that

$$\hat{p}_{\mathrm{SP},k}(y|x) = \bar{p}_{\mathrm{SP},k}(y|x) + \nabla\bar{p}(y|x)\left[\hat{\mu}_{\mathrm{SP},k} - \bar{\mu}_{\mathrm{SP},k}, \hat{\sigma}^2_{\mathrm{SP},k} - \bar{\sigma}^2_{\mathrm{SP},k}\right] + o_P\left(1/\sqrt{n}\right),$$

under the conditions imposed on the bandwidth in (H.k), $k=1,2$, where $\nabla p(y|x)\left[d\mu, d\sigma^2\right]$ is the pathwise derivative of $\bar{p}_{\mathrm{SP},k}(y|x)$ w.r.t. the drift and diffusion function in the direction $\left(d\mu, d\sigma^2\right)$. It is the solution (at $t=1$) to the following PDE,

$$\frac{\partial\nabla\bar{p}(y|x;t)}{\partial t} = \mathcal{A}\left[\mu_{\mathrm{SP},k}, \sigma^2_{\mathrm{SP},k}\right]\nabla\bar{p}(y|x;t) + \mathcal{A}\left[d\mu, d\sigma^2\right]\bar{p}_{\mathrm{SP},k}(y|x;t), \tag{33}$$

with $\nabla\bar{p}(y|x;0)\left[d\mu, d\sigma^2\right] = 0$. The solution at $t=1$ can be represented as:

$$\begin{aligned}\nabla\bar{p}(y|x)\left[d\mu, d\sigma^2\right] &= \int_0^1\int_I d\mu(w)\frac{\partial\bar{p}_{\mathrm{SP},k}(y|w;t)}{\partial w}\bar{p}_{\mathrm{SP},k}(w|x;t)\,dwdt \\ &\quad + \int_0^1\int_I d\sigma^2(w)\frac{\partial^2\bar{p}_{\mathrm{SP},k}(y|w;t)}{\partial w^2}\bar{p}_{\mathrm{SP},k}(w|x;t)\,dwdt.\end{aligned} \tag{34}$$

Using Kristensen (2010, Lemma 5), it follows that $\partial^i\bar{p}_{\mathrm{SP},k}(y|x)/\partial x^i = \partial^i p_{\mathrm{SP},k}(y|x)/\partial x^i + O(a^q)$, $i=0,1,2$, where $q>0$ is given in Assumption (A.3). Thus, as $\sqrt{n}a^q \to 0$, $\partial^i\bar{p}_{\mathrm{SP},k}(y|x)/\partial x^i = \partial^i p_{\mathrm{SP},k}(y|x)/\partial x^i + o(1/\sqrt{n})$, which in turn implies that

$$\nabla\bar{p}(y|x;0)\left[d\mu, d\sigma^2\right] = \nabla p(y|x)\left[d\mu, d\sigma^2\right] + o\left(1/\sqrt{n}\right),$$

where $\nabla p(y|x)\left[d\mu, d\sigma^2\right]$ is the pathwise derivative of the untrimmed transition density, $p_{\mathrm{SP},k}(y|x)$. This pathwise derivative has the same representation as $\nabla\bar{p}(y|x)\left[d\mu, d\sigma^2\right]$ given in eq. (34), but with $p_{\mathrm{SP},k}(y|x;t)$ replacing $\bar{p}_{\mathrm{SP},k}(y|x;t)$ on the right hand side.

We now analyze the two integrals appearing in the representation of $\nabla p(y|x)\left[d\mu, d\sigma^2\right]$ with $\left(d\mu, d\sigma^2\right) = (\hat{\mu}_{\mathrm{SP},k} - \bar{\mu}_{\mathrm{SP},k}, \hat{\sigma}^2_{\mathrm{SP},k} - \bar{\sigma}^2_{\mathrm{SP},k})$ for the two classes of semiparametric estimators. First consider the estimators under $H_{\mathrm{SP},1}$: Proceeding as in Kristensen (2010, Proof of Theorem 2), under the conditions imposed on bandwidth and the trimming sequence,

$$\int_0^1\int_I\left[\hat{\mu}_{\mathrm{SP},1}(w) - \bar{\mu}_{\mathrm{SP},1}(w)\right]\frac{\partial p_{\mathrm{SP},1}(y|w;t)}{\partial w}p_{\mathrm{SP},1}(w|x;t)\,dt = \frac{1}{n}\sum_{i=1}^n D_{1,1}(X_i,y,x) + o_P\left(1/\sqrt{n}\right),$$

while, by the mean-value theorem,

$$\begin{aligned}&\int_0^1\int_I\left[\hat{\sigma}^2_{\mathrm{SP},1}(w) - \bar{\sigma}^2_{\mathrm{SP},1}(w)\right]\frac{\partial^2 p_{\mathrm{SP},1}(y|w;t)}{\partial w^2}p_{\mathrm{SP},1}(w|x;t)\,dwdt \\ &= \frac{1}{n}\sum_{i=1}^n D_{1,2}(X_i, X_{i-1}, y, x) + o_P\left(1/\sqrt{n}\right),\end{aligned}$$

where

$$D_{1,1}(z_1,y,x) = -\frac{1}{2\pi_0(z_1)}\frac{\partial}{\partial z_1}\left[\sigma^2_{\mathrm{SP},1}(z_1)\int_0^1\frac{\partial p_{\mathrm{SP},1}(y|z_1;t)}{\partial z_1}p_{\mathrm{SP},1}(z_1|x;t)\,dt\right],$$

$$D_{1,2}(z_1,z_2,y,x) = \psi_{\mathrm{SP},1}(z_1|z_2)\int_0^1\int_I\frac{\partial\sigma^2_{\mathrm{SP},1}(w;\theta)}{\partial\theta}\frac{\partial^2 p_{\mathrm{SP},1}(y|w;t)}{\partial w^2}p_{\mathrm{SP},1}(w|x;t)\,dwdt.$$

Next, consider the estimators under $H_{\mathrm{SP},2}$: Again, proceeding as Kristensen (2010, Proof of Theorem 2), we obtain under the conditions imposed on the bandwidth and the trimming sequence that

$$\int_0^1 \int_I \left[\hat{\sigma}^2_{\mathrm{SP},2}(w) - \sigma^2_{\mathrm{SP},2}(w)\right] \frac{\partial^2 p_{\mathrm{SP},2}(y|w;t)}{\partial w^2} p_{\mathrm{SP},2}(w|x;t)\,dwdt = \frac{1}{n}\sum_{i=1}^n D_{2,1}(X_i,y,x) + o_P\left(1/\sqrt{n}\right),$$

while, by the mean-value theorem,

$$\begin{aligned}
&\int_0^1 \int_I \left[\hat{\mu}_{\mathrm{SP},2}(w) - \mu_{\mathrm{SP},2}(w)\right] \frac{\partial p_{\mathrm{SP},2}(y|w;t)}{\partial w} p_{\mathrm{SP},2}(w|x;t)\,dwdt \\
= &\ \frac{1}{n}\sum_{i=1}^n D_{2,2}(X_i, X_{i-1}, y, x) + o_P\left(1/\sqrt{n}\right),
\end{aligned}$$

where

$$\begin{aligned}
D_{2,1}(z_1,y,x) &= 2\mu_{\mathrm{SP},2}(z_1)\int_0^1 \int_l^{z_1} \frac{1}{\pi(w)} \frac{\partial^2 p_{\mathrm{SP},2}(y|w;t)}{\partial w^2} p_{\mathrm{SP},2}(w|x;t)\,dwdt \\
&\quad -2\frac{\sigma^2_{\mathrm{SP},2}(z_1)}{\pi^2(z_1)}\int_0^1 \frac{\partial^2 p_{\mathrm{SP},2}(y|z_1;t)}{\partial z_1^2} p_{\mathrm{SP},2}(z_1|x;t)\,dt,
\end{aligned}$$

$$D_{2,2}(z_1,z_2,y,x) = \psi_{\mathrm{SP},2}(z_1|z_2)\int_0^1 \int_I \frac{\partial \mu_{\mathrm{SP},2}(w;\theta)}{\partial \theta} \frac{\partial p_{\mathrm{SP},2}(y|w;t)}{\partial w} p_{\mathrm{SP},2}(w|x;t)\,dwdt.$$

The claimed result now holds with

$$D_{k,i}(y|x) = D_{k,1}(X_i,y,x) + D_{k,2}(X_i,X_{i-1},y,x), \quad k=1,2. \tag{35}$$

■

Proof of Theorem 3. First note that we can replace $\hat{p}_{\mathrm{SP},k}(y|x)$, $k=1,2$, by $p_{\mathrm{SP},k}(y|x) = p(y|x)$ in the following since it converges with $\sqrt{n}$-rate, c.f. Lemma 2, and we now proceed to analyze

$$T_{\mathrm{SP}} := \int_I \int_I \left[p(y|x) - \hat{p}_{\mathrm{NP}}(y|x)\right]^2 w(y,x)\,dydx.$$

By a Taylor expansion in terms of f and π,

$$T_{\mathrm{SP}} = \int_I \int_I \left[\frac{f(y|x)}{\pi_0(x)} - \frac{\hat{f}_{\mathrm{NP}}(y,x)}{\hat{\pi}_{\mathrm{NP}}(x)}\right]^2 w(y,x)\,dydx = I + \bar{I} + R,$$

where

$$I := \int_I \int_I \left[f(y,x) - \hat{f}_{\mathrm{NP}}(y,x)\right]^2 m(y,x)\,dydx, \quad \bar{I} := \int_I \int_I \left[\pi(x) - \hat{\pi}_{\mathrm{NP}}(x)\right]^2 \bar{m}(x)\,dx,$$

with $m(y,x) := w(y,x)/\pi^2(x)$ and $\bar{m}(x) := \int_I f(y,x)w(y,x)\,dy/\pi^4(x)$, and

$$R = O_P\left(\sup_{x,y\in I} |f(y,x) - \hat{f}_{\mathrm{NP}}(y,x)|^4\right) + O_P\left(\sup_{x\in I} |\pi(x) - \hat{\pi}_{\mathrm{NP}}(x)|^4\right).$$

Under the requirement that $\lambda^2 = \lim_{n\to\infty} nh_{\mathrm{NP}}^{2m+2} < \infty$, it follows from Gourieroux and Tenreiro (2001, Theorem 4.1) that

$$nh_{\mathrm{NP}}\{I - \mu - B\} = v_{\mathrm{SP}}U_n + \sqrt{n}h_{\mathrm{NP}}^{m+1}\sigma_v V_n + o_P\left(\sqrt{n}h_{\mathrm{NP}}^{m+1}\right) + o_P(1), \tag{36}$$

$$nh_{\mathrm{NP}}^{1/2}\{\bar{I} - \bar{\mu} - \bar{B}\} = \bar{v}_{\mathrm{SP}}\bar{U}_n + \sqrt{n}h_{\mathrm{NP}}^{m+1/2}\bar{\sigma}_v\bar{V}_n + o_P\left(\sqrt{n}h_{\mathrm{NP}}^{m+1/2}\right) + o_P(1) \tag{37}$$

where

$$
\begin{aligned}
\mu &: = \frac{1}{nh_{\text{NP}}^{2}}\left[\int_{\mathbb{R}} K^{2}(z)\, dz\right]^{2} \times \int_{I\times I} f(y,x) m_{1}(y,x)\, dy dx, \\
\bar{\mu} &: = \frac{1}{nh_{\text{NP}}} \int_{\mathbb{R}} K^{2}(z)\, dz \times \int_{I} \pi(x) m_{2}(x)\, dx,
\end{aligned}
$$

$$
\begin{aligned}
B &: = \int_{I\times I}\left[\int_{I\times I} \frac{1}{h_{\text{NP}}^{2}} K\left(\frac{u_{1}-y}{h_{\text{NP}}}\right) K\left(\frac{u_{1}-y}{h_{\text{NP}}}\right) f(u_{1},u_{2}) du_{1} du_{2} - f(y,x)\right]^{2} m(y,x)\, dy dx, \\
\bar{B} &: = \int_{I}\left[\int_{I} \frac{1}{h_{\text{NP}}} K\left(\frac{u-x}{h_{\text{NP}}}\right) \pi(u) du - \pi(x)\right]^{2} \bar{m}(x)\, dx,
\end{aligned}
$$

$$
\begin{aligned}
v_{\text{SP}}^{2} &= 2\left[\int_{\mathbb{R}} (K*K)^{2}(z)\, dz\right]^{2} \times \int_{I\times I} f^{2}(y,x) m^{2}(y,x)\, dy dx, \\
\bar{v}_{\text{SP}}^{2} &= 2\int_{\mathbb{R}} (K*K)^{2}(z)\, dz \times \int_{I} \pi^{2}(x) \bar{m}^{2}(x)\, dx,
\end{aligned}
$$

and (U_n, V_n) and $(\bar{U}_n, \bar{V}_n)$ both converge towards a bivariate standard Normal distribution.

Due to the smoothness conditions imposed on $p(y|x)$ and $\pi(x)$ and K being an mth order kernel,

$$
\begin{aligned}
B &= \int_{I\times I}\left[\int_{\mathbb{R}^{2}} K(z_{1}) K(z_{2}) \left[f(y+z_{1}h_{\text{NP}}, x+z_{2}h_{\text{NP}}) - f(y,x)\right] dz_{1} dz_{2}\right]^{2} m(y,x)\, dy dx \\
&= \int_{I\times I}\left[\sum_{i,j\leq m} h_{\text{NP}}^{i+j} \frac{\partial^{i+j} f(y,x)}{\partial x^{i} \partial y^{j}} \int_{\mathbb{R}^{2}} K(z_{1}) K(z_{2}) z_{1}^{i} z_{2}^{j} dz_{1} dz_{2} + o(h_{\text{NP}}^{m})\right]^{2} m(y,x)\, dy dx \\
&= h_{\text{NP}}^{2m} \times \int_{I\times I}\left[\frac{\partial^{2m} f(y,x)}{\partial x^{m} \partial y^{m}}\right]^{2} m(y,x)\, dy dx \times \left[\int_{\mathbb{R}} K(z) z^{m} dz\right]^{4} + o\left(h_{\text{NP}}^{2m}\right) \\
&= : h_{\text{NP}}^{2m} b + o\left(h_{\text{NP}}^{2m}\right)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
\bar{B} &= h_{\text{NP}}^{2m} \times \int_{I\times I}\left[\frac{\partial^{m} \pi(x)}{\partial x^{m}}\right]^{2} \bar{m}(x)\, dx \times \left[\int_{\mathbb{R}} K(z) z^{m} dz\right]^{2} + o\left(h_{\text{NP}}^{2m}\right) \\
&= : h_{\text{NP}}^{2m} \bar{b} + o\left(h_{\text{NP}}^{2m}\right).
\end{aligned}
$$

Finally, applying Kristensen (2009, Theorem 1) together with standard arguments for the bias components of the kernel density estimators,

$$
nh_{\text{NP}} R = O_{P}\left(nh_{\text{NP}}^{4m+1}\right) + O_{P}\left(\frac{\log(n)^{2}}{nh_{\text{NP}}^{3}}\right).
$$

In total,

$$
\begin{aligned}
& nh_{\text{NP}}\left\{T_{\text{SP}} - \mu - \bar{\mu}\right\} \\
= \; & nh_{\text{NP}}\left\{I - \mu - B\right\} + nh_{\text{NP}}\left\{\bar{I} - \bar{\mu} - \bar{B}\right\} + nh_{\text{NP}}\left(B + \bar{B}\right) + nh_{\text{NP}} R \\
= \; & v_{\text{SP}} U_{n} + \sqrt{h_{\text{NP}}}\, \bar{v}_{\text{SP}} \bar{U}_{n} + \lambda_{n}\left\{\sigma_{v} V_{n} + \bar{\sigma}_{v} \bar{V}_{n}\right\} + nh_{\text{NP}}^{2m+1}\left(b + \bar{b}\right) + O_{P}\left(\frac{\log(n)^{2}}{nh_{\text{NP}}^{3}}\right)
\end{aligned}
$$

where λ_n^2 is defined in the theorem; this proves the first part of the theorem. The second part is a direct consequence of this representation. ■

Proof of Theorem 4. Consider the contiguous alternative $H_{SP,k}^c$ $(k = 1, 2)$: Since the pseudo-true parameter values are drifting, note that the restricted drift and diffusion terms implied by the null as given in eqs. (13) and (14) respectively are also drifting. We therefore have that $\sigma_{\mathrm{SP},k}^2(x) = \sigma_{\mathrm{SP},k,n}^2(x)$ and $\mu_{\mathrm{SP},k}(x) = \mu_{\mathrm{SP},k.n}(x)$ are sequences, and so in turn is the associated transition density, $p_{\mathrm{SP},k}(y|x) = p_{\mathrm{SP},k,n}(y|x)$. We suppress their dependence on n in the following for notational ease.

Define the deviations from the null hypothesis, $d\mu_n(x) = \mu_n(x) - \mu_{\mathrm{SP},k}(x)$ and $d\sigma_n^2(w) = \sigma_n^2(x) - \sigma_{\mathrm{SP},k}^2(x)$, where $\mu_{\mathrm{SP},k}(x)$ and $\sigma_{\mathrm{SP},k}^2(x)$ are the drift and diffusion term. By Kristensen (2010, Lemma 5),

$$\left|p_n(y|x) - p_{\mathrm{SP},k}(y|x) - \gamma_{\mathrm{SP},k}^{(n)}(y|x)\right| \leq CB_n(x),$$

where $\gamma_{\mathrm{SP},k}^{(n)}(y|x) = \nabla p_n(y|x)[d\mu_n, d\sigma_n]$ is the pathwise derivative w.r.t. the drift and diffusion term, c.f. Proof of Lemma 2, and

$$B_n(x) = \int_I \left(|d\mu_n(w)|^2 + \left|d\sigma_n^2(w)\right|^2\right) \bar{\gamma}(w|x)\, dw.$$

Next, by the same arguments as in the Proof of Lemma 2,

$$\gamma_{\mathrm{SP},k}^{(n)}(y|x) = \int_I d\mu_n(w)\, \bar{p}_{\mu,k}(y, x, w)\, dw + \int_I d\sigma_n^2(w)\, \bar{p}_{\sigma^2,k}(y, x, w)\, dw,$$

where $\bar{p}_{\mu,k}(y, x, w)$ and $\bar{p}_{\sigma^2,k}(y, x, w)$ are defined in eq. (23).

Consider first $H_{\mathrm{SP},1}^c$: In this case,

$$\begin{aligned} d\sigma_n^2(x) &= \left\{\sigma_n^2(x) - \sigma^2(x;\theta_0)\right\} + \left\{\sigma^2(x;\theta_0) - \sigma_{\mathrm{SP},1}^2(x)\right\} \\ &= g_n(x) + \dot{\sigma}^2(x;\theta_0)\,\xi_{1,n} + O\left(\left\|\xi_{1,n}\right\|^2\right) \end{aligned}$$

while

$$\begin{aligned} d\mu_n(x) &= \frac{1}{2\pi(x)} \frac{\partial}{\partial x}\left[\left(\sigma_n^2(x) - \sigma_{\mathrm{SP},1}^2(x)\right)\pi(x)\right] \\ &= \frac{1}{2\pi(x)} \frac{\partial}{\partial x}\left[g_n(x)\pi(x)\right] + \frac{\xi_{1,n}}{2\pi(x)} \frac{\partial}{\partial x}\left[\dot{\sigma}^2(x;\theta_0)\pi(x)\right] + O\left(\left\|\xi_{1,n}\right\|^2\right). \end{aligned}$$

Plugging these two expressions into $\gamma_{\mathrm{SP},1}^{(n)}(y|x)$ above we obtain the expression in eq. (21). Next, consider $H_{\mathrm{SP},2}^c$. Here, $d\mu_n(x) = \mu_n(x) - \mu_{\mathrm{SP},2}^2(x) = g_n(x)$, and

$$d\sigma_n^2(x) = \frac{2}{\pi_0(x)} \int_l^x d\mu_n(y)\pi(y)\, dy = \frac{2}{\pi_0(x)} \int_l^x g_n(y)\pi(y)\, dy.$$

Substituting those into $\gamma_{\mathrm{SP},2}^{(n)}(y|x)$, we obtain the expression in eq. (22).

We now proceed as in the Proof of Theorem 3:

$$T_{\mathrm{SP},k} = \int_I \int_I \left[p_{\mathrm{SP},k}(y|x) - \hat{p}_{\mathrm{NP}}(y|x)\right]^2 w(y, x)\, dydx = I_1 + I_2 + R,$$

where R is a higher-order remainder term,

$$I_1 := \int_I \int_I \left[f_{\mathrm{SP},k}(y, x) - \hat{f}_{\mathrm{NP}}(y, x)\right]^2 m_{n,1}(y, x)\, dydx, \quad I_2 := \int_I \int_I \left[\pi(x) - \hat{\pi}_{\mathrm{NP}}(x)\right]^2 m_{n,2}(x)\, dx,$$

with $m_{n,1}(y,x)$ and $m_{n,2}(x)$ given in the Proof of Theorem 3, and $f_{\mathrm{SP},k}(y,x) = p_{\mathrm{SP},k}(y|x)\pi(x)$ denoting the joint density under the null. Also let $f_n(y,x) = p_n(y|x)\pi(x)$ denote the sequence of joint densities under the alternative. Substituting the expression of $p_{\mathrm{SP},k}(y|x)$ given in eq. (20) into I_1 and ignoring the higher-order term $R^2_{\mathrm{SP},k}$,

$$\begin{aligned}
I_1 &= \int_I\int_I \left[f_{\mathrm{SP},k}(y,x) - \hat{f}_{\mathrm{NP}}(y,x)\right]^2 m_{n,1}(y,x)\,dydx \\
&= \int_I\int_I \left[f_n(y|x) - \hat{f}_{\mathrm{NP}}(y,x) - \gamma^{(n)}_{\mathrm{SP},k}(y|x)\pi(x)\right]^2 m_{n,1}(y,x)\,dydx \\
&= \int_I\int_I \left[f_n(y,x) - \hat{f}_{\mathrm{NP}}(y,x)\right]^2 m_{n,1}(y,x)\,dydx + \int_I\int_I \gamma^{(n)}_{\mathrm{SP},k}(y|x)^2\pi^2(x)\,m_{n,1}(y,x)\,dydx \\
&\quad +2\int_I\int_I \left[\hat{f}_{\mathrm{NP}}(y,x) - f_n(y,x)\right]\gamma^{(n)}_{\mathrm{SP},k}(y|x)\pi(x)\,m_{n,1}(y,x)\,dydx \\
&= : I_{11} + I_{12} + I_{13},
\end{aligned}$$

Due to the assumptions imposed on $g_n(x)$, we note that Assumptions (A.1)-(A.2) remains true for the diffusion model corresponding to (μ_n, σ_n^2). Thus, we can recycle the same arguments used in Proof of Theorem 3 to obtain $nh_{\mathrm{NP}}\{I_{11} - \mu_1\} = v_{\mathrm{SP}}U_1 + o_P(1)$, while, using the same arguments as in Gourieroux and Tenreiro (2001), $I_{13} = O_P(n^{-1/2})$. Since we consider alternatives where the marginal density remains correctly specified, the second term, I_2, still satisfies eq. (37). In total, as $nh^3_{\mathrm{NP}} \to \infty$ and $nh^{4m+1}_{\mathrm{NP}} \to 0$,

$$\begin{aligned}
nh_{\mathrm{NP}}\{T_{\mathrm{SP},k} - m_{\mathrm{SP}}\} &= nh_{\mathrm{NP}}\{I_{11} - \mu_1\} + nh_{\mathrm{NP}}I_{12} + nh_{\mathrm{NP}}I_{13} + nh_{\mathrm{NP}}\{I_2 - \mu_2\} + nh_{\mathrm{NP}}R \\
&= v_{\mathrm{SP}}U_1 + nh_{\mathrm{NP}}I_{12} + o_P(1).
\end{aligned}$$

■

Proof of Theorem 5. Under Assumption (A.6), the parametric estimator satisfies

$$\begin{aligned}
\hat{p}_{\mathrm{P}}(y|x) - p(y|x) &= \frac{\partial p_{\mathrm{P}}(y|x;\theta)}{\partial\theta'}(\tilde{\theta} - \theta) + O_P\left(||\tilde{\theta} - \theta||^2\right) \\
&= \frac{\partial p_{\mathrm{P}}(y|x;\theta)}{\partial\theta'}\frac{1}{n}\sum_{i=1}^n \psi_{P,i} + o_P\left(1/\sqrt{n}\right),
\end{aligned}$$

where $p_{\mathrm{P}}(y|x;\theta) = p(y|x)$ under the null, while Lemma 2 supplies us with an expansion of $\hat{p}_{\mathrm{SP},k}(y|x)$. Substituting these two expansions into $T_{\mathrm{P},k}$ yields:

$$\begin{aligned}
T_{\mathrm{P},k} &= \int_I\int_I \left[\hat{p}_{\mathrm{P}}(y|x) - \hat{p}_{\mathrm{SP},k}(y|x)\right]^2 w(y,x)\,dydx \\
&= \int_I\int_I \left[\{\hat{p}_{\mathrm{P}}(y|x) - p(y|x)\} - \{\hat{p}_{\mathrm{SP},k}(y|x) - p(y|x)\}\right]^2 w(y,x)\,dydx \\
&= \frac{1}{n}\int_I\int_I Z^2_{n,k}(x,y)\,w(y,x)\,dydx + o_P\left(\frac{1}{n}\right),
\end{aligned}$$

where $Z_{n,k}(x,y)$ is an empirical process,

$$Z_{n,k}(x,y) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\frac{\partial p(y|x;\theta)}{\partial\theta'}\psi_{P,i} - D_{k,i}(y|x)\right\}. \tag{38}$$

Let $\mathcal{C} \subseteq I \times I$ denote the (compact) support of $w(y,x)$. We then wish to show that $Z_{n,k}(x,y)$ weakly converges on $\mathcal{C}$ towards the stochastic process $Z_k(x,y)$ defined in the theorem. By Lemma 2, Assumption (A.5) and the CLT for stationary and mixing sequences (Doukhan et al, 1994), $Z_{n,k}(x,y) \to^d Z_k(x,y)$ for any given $(x,y) \in \mathcal{C}$. Appealing to standard arguments from empirical process theory, see e.g. Doukhan et al (1995), it follows by Lemma 2 and Assumption (A.6) that $Z_{n,k}(x,y)$ is stochastically equicontinuous. The result now follows by the Continuous Mapping Theorem. ■

Proof of Theorem 6. The representation of the sequence of transition densities given in eq. (28) follows by the same arguments as in the Proof of Theorem 4. We then obtain

$$\begin{aligned} T_{\mathrm{P},k} &= \int_I \int_I [\hat{p}_{\mathrm{P}}(y|x) - \hat{p}_{\mathrm{SP},k}(y|x)]^2 \, w(y,x) \, dydx \\ &= \int_I \int_I [\{\hat{p}_{\mathrm{P}}(y|x) - p_{\mathrm{P}}(y|x)\} - \{\hat{p}_{\mathrm{SP},k}(y|x) - p_n(y|x)\} + \{p_n(y|x) - p_{\mathrm{P}}(y|x)\}]^2 \, w(y,x) \, dydx \\ &= \int_I \int_I [\{\hat{p}_{\mathrm{P}}(y|x) - p_{\mathrm{P}}(y|x)\} - \{\hat{p}_{\mathrm{SP},k}(y|x) - p_n(y|x)\}]^2 \, w(y,x) \, dydx \\ &\quad + \int_I \int_I [p_n(y|x) - p_{\mathrm{P}}(y|x)]^2 \, w(y,x) \, dydx \\ &\quad + 2 \int_I \int_I [\{\hat{p}_{\mathrm{P}}(y|x) - p_{\mathrm{P}}(y|x)\} - \{\hat{p}_{\mathrm{SP},k}(y|x) - p_n(y|x)\}] \, [p_n(y|x) - p_{\mathrm{P}}(y|x)] \, w(y,x) \, dydx \\ &= : I_1 + I_2 + I_3 \end{aligned}$$

The first term, I_1, can be analyzed analogously to Proof of Theorem 5, while, by eq. (28),

$$I_2 = \int_I \int_I \gamma_{\mathrm{P},k}^{(n)}(y|x)^2 \, w(y,x) \, dydx + O\left(R_{\mathrm{P}}^2\right).$$

Finally, with $Z_{n,k}(x,y)$ defined in eq. (38), I_3 can be written as

$$I_3 = \frac{2}{\sqrt{n}} \int_I \int_I Z_{n,k}(x,y) \, \gamma_{\mathrm{P},k}^{(n)}(y|x) \, w(y,x) \, dydx + O_P\left(R_{\mathrm{P}}/\sqrt{n}\right).$$

■

Proof of Theorem 7. First consider $\bar{T}_{\mathrm{P},1}$: Since the drift estimator under the null and the diffusion estimator under the alternative both converge with parametric rate we may replace them with the true, unknown ones and redefine our estimators as:

$$\tilde{\mu}(x) = \frac{1}{2}\frac{\partial \sigma^2(x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\pi'(x)}{\pi(x)}, \quad \hat{\mu}(x) = \frac{1}{2}\frac{\partial \sigma^2(x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\hat{\pi}'(x)}{\hat{\pi}(x)}.$$

Next, by a Taylor expansion w.r.t. $\pi(x)$ and $\pi'(x)$,

$$\begin{aligned} \bar{T}_{\mathrm{P},1} &= \int_I [\hat{\pi}'(x) - \pi'(x)]^2 \omega_1(x) \, dx + \int_I [\hat{\pi}(x) - \pi(x)]^2 \omega_2(x) \, dx + R \\ &= : I_1 + I_2 + R, \end{aligned}$$

where $\omega_1(x) := \sigma^4(x) \bar{w}(x) / \left(4\pi^2(x)\right)$, $\omega_2(x) := \sigma^4(x) \bar{w}(x) \pi'(x)^2 / \left(4\pi^4(x)\right)$, and

$$R = O_P\left(\sup_{x,y\in I} |\hat{\pi}'(x) - \pi'(x)|^4\right) + O_P\left(\sup_{x\in I} |\hat{\pi}(x) - \pi(x)|^4\right).$$

First, consider I_1: With $\bar{\pi}'(x) = E\left[\hat{\pi}'(x)\right]$, write

$$\begin{aligned} I_1 &= \int_I [\bar{\pi}'(x) - \pi'(x)]_1^2 \omega_1(x) \, dx + 2\int_I [\hat{\pi}'(x) - \bar{\pi}'(x)][\bar{\pi}'(x) - \pi'(x)] \omega_1(x) \, dx \\ &\quad + \int_I [\hat{\pi}'(x) - \bar{\pi}'(x)]^2 \omega_1(x) \, dx \\ &= : I_{11} + I_{12} + I_{13}. \end{aligned}$$

Using that uniformly over $x \in I$,

$$\bar{\pi}'(x) = \pi'(x) + h^m \pi^{(m)}(x) + o(h^m), \tag{39}$$

the first term can be written as

$$I_{11} = h^{2m} \int_I \pi^{(m)}(x)^2 \omega_1(x)\, dx + o\left(h^{2m}\right).$$

The second term, again using eq. (39), satisfies

$$\begin{aligned}
I_{12} &= 2\frac{h^m}{nh^2} \sum_{i=1}^{n} \int_I \left\{ K'\left(\frac{x - X_i}{h}\right) - E\left[K'\left(\frac{x - X_0}{h}\right)\right]\right\} \left\{\pi^{(m)}(x) + o(1)\right\} \omega_1(x)\, dx \\
&= 2\frac{h^m}{n} \sum_{i=1}^{n} G_n(X_i) + o_P\left(\frac{h^m}{\sqrt{n}}\right),
\end{aligned}$$

where $G_n(u)$ is given by

$$\begin{aligned}
G_n(u) &= \frac{1}{h^2} \int_I \left\{ K'\left(\frac{x - u}{h}\right) - E\left[K'\left(\frac{x - X_0}{h}\right)\right]\right\} \pi^{(m)}(x)\, \omega_1(x)\, dx \\
&= \frac{1}{h} \int_I \left\{ K\left(\frac{x - u}{h}\right) - E\left[K\left(\frac{x - X_0}{h}\right)\right]\right\} \pi^{(m+1)}(x)\, \omega_1'(x)\, dx.
\end{aligned}$$

The third term can be rewritten as

$$\begin{aligned}
I_{13} &= \int_I \left[\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} \left\{ K'\left(\frac{x - X_i}{h}\right) - E\left[K'\left(\frac{x - X_i}{h}\right)\right]\right\}\right]^2 \omega_1(x)\, dx \\
&= \frac{1}{n^2 h^{5/2}} \sum_{i,j=1}^{n} H_n(X_i, X_j),
\end{aligned}$$

where

$$\begin{aligned}
H_n(u, v) &:= \frac{1}{h^{3/2}} \int_I \left\{ K'\left(\frac{x - u}{h}\right) - E\left[K'\left(\frac{x - X_0}{h}\right)\right]\right\} \\
&\quad \times \left\{ K'\left(\frac{x - v}{h}\right) - E\left[K'\left(\frac{x - X_0}{h}\right)\right]\right\} \omega_1(x)\, dx.
\end{aligned}$$

Combining these expressions and following the arguments Gourieroux and Tenreiro (2001, p. 182-184) (see also Huang, 1997), we then obtain

$$\begin{aligned}
I_1 &= \frac{1}{n^2 h^{5/2}} \sum_{i,j=1}^{n} \left\{ H_n(X_i, X_j) - E\left[H_n(X_i, X_j)\right]\right\} + \frac{2h^{m-1}}{n} \sum_{i=1}^{n} G_n(X_i) \\
&\quad + \frac{1}{nh^{5/2}} E\left[H_n(X_0, X_0)\right] + \frac{1}{n^2 h^{5/2}} \sum_{i \neq j}^{n} E\left[H_n(X_i, X_j)\right] \\
&\quad + h^{2m} \int_I \pi^{(m)}(x)^2 \omega_1(x)\, dx + o\left(h^{2m}\right) \\
&= \frac{1}{nh^{5/2}} \mathcal{H}_n + \frac{h^m}{\sqrt{n}} \mathcal{G}_n + \frac{1}{nh^{5/2}} E\left[H_n(X_0, X_0)\right] + h^{2m} \int_I \pi^{(m)}(x)^2 \omega_1(x)\, dx \\
&\quad + o_P\left(\frac{1}{nh^{5/2}}\right) + o_P\left(\frac{h^{m-1}}{\sqrt{n}}\right) + o_P\left(h^{2m}\right),
\end{aligned}$$

where

$$\mathcal{H}_n := \frac{2}{n}\sum_{i<j}\left\{H_n\left(X_i,X_j\right)-E\left[H_n\left(X_i,X_j\right)\right]\right\},\quad \text{and } \mathcal{G}_n := \frac{2}{\sqrt{n}}\sum_{i=1}^{n} G_n\left(X_i\right).$$

The mean component satisfies

$$\begin{aligned}
E\left[H_n\left(X_0,X_0\right)\right] &= \frac{1}{h^{3/2}}\int_I\left\{E\left[K'\left(\frac{x-X_0}{h}\right)^2\right]-E\left[K'\left(\frac{x-X_0}{h}\right)\right]^2\right\}\omega_1(x)\,dx \\
&= \frac{1}{h^{1/2}}\int_I\int_I K'(u)^2\,\omega_1(x+uh)\,\pi(x+uh)\,dudx+O\left(h^{1/2}\right) \\
&= \frac{1}{h^{1/2}}\int_{\mathbb{R}} K'(z)^2\,dz\times\int_I\omega_1(x)\,\pi(x)\,dx+o\left(\frac{1}{h^{1/2}}\right)+O\left(h^{1/2}\right),
\end{aligned}$$

such that $E\left[H_n\left(X_0,X_0\right)\right]/\left(nh^{5/2}\right)=\mu_1+o(1)$, where

$$\mu_1=\frac{1}{nh^3}\int_{\mathbb{R}}K'(z)^2\,dz\times\int_I\omega_1(x)\,\pi(x)\,dx=\frac{1}{4nh^3}\int_{\mathbb{R}}K'(z)^2\,dz\times\int_I\frac{\sigma^4(x)\,\bar{w}(x)}{\pi(x)}dx$$

We can now appeal to the arguments of Gourieroux and Tenreiro (2001, Proof of Theorem 3.2) to conclude that

$$I_1=\mu_1+\frac{1}{nh^{5/2}}\mathcal{H}_n+\frac{h^m}{\sqrt{n}}\mathcal{G}_n+O_P\left(h^{2m}\right)+o_P\left(\frac{1}{nh^{5/2}}\right)+o_P\left(\frac{h^m}{\sqrt{n}}\right).$$

Given that $\mathcal{H}_n$ and $\mathcal{G}_n$ converge towards a bivariate normal distribution with covariance zero and marginal variances v_{P}^2 and σ_{P}^2 (see Gourieroux and Tenreiro, 2001, Theorem 3.1 for their expressions), it follows that

$$nh^{5/2}\left\{I_1-\mu_1\right\}=v_{\mathrm{P},1}U_n+\sqrt{n}h^{m+5/2}\sigma_{\mathrm{P},1}V_n+O_P\left(nh^{2m+5/2}\right)+o_P\left(\sqrt{n}h^{m+5/2}\right).$$

Here, one can verify that

$$v_{\mathrm{P},1}^2=2\int_{\mathbb{R}}\left(K'*K'\right)^2(z)\,dz\times\int\pi^2(x)\omega_1^2(x)\,dx=\frac{1}{8}\int_{\mathbb{R}}\left(K'*K'\right)^2(z)\,dz\times\int\frac{\sigma^{16}(x)\,\bar{w}^2(x)}{\pi^2(x)}dx.$$

Next, by a direct application of Gourieroux and Tenreiro (2001, Theorem 3.2).

$$nh^{1/2}\left\{I_2-\mu_2\right\}=\bar{v}_{\mathrm{P},1}\bar{U}_n+\sqrt{n}h^{m+1/2}\bar{\sigma}_{\mathrm{P},1}\bar{V}_n+O_P\left(nh^{2m+1/2}\right)+o_P\left(\sqrt{n}h^{m+1/2}\right)\tag{40}$$

where $\left(\bar{U}_n,\bar{V}_n\right)$ converge in distribution towards a bivariate standard Normal distribution, and

$$\mu_2=\frac{1}{nh}\int_{\mathbb{R}}K^2(z)\,dz\times\int_I\pi(x)\omega_2(x)\,dx=\frac{1}{4nh}\int_{\mathbb{R}}K^2(z)\,dz\times\int_I\frac{\sigma^4(x)\,\bar{w}(x)\,\pi'(x)^2}{\pi^3(x)}dx,$$

and

$$\bar{v}_{\mathrm{P},1}^2=2\int_{\mathbb{R}}(K*K)^2(z)\,dz\times\int_I\pi^2(x)\omega_2^2(x)\,dx=\frac{1}{8}\int_{\mathbb{R}}(K*K)^2(z)\,dz\times\int_I\frac{\sigma^4(x)\,\bar{w}(x)\,\pi'(x)^4}{\pi^{14}(x)}dx$$

Finally, by Kristensen (2009, Theorem 1) and standard kernel bias evaluations,

$$nh^{5/2}R=O_P\left(nh^{4m+5/2}\right)+O_P\left(\frac{\log(n)^2}{nh^{1/2}}\right).$$

In total,

$$\begin{aligned} nh^{5/2}\left\{\bar{T}_{\mathrm{P},1}-\mu_1-\mu_2\right\} &= nh^{5/2}\left\{I_1-\mu_1\right\}+nh^{5/2}\left\{I_2-\mu_2\right\}+nh^{5/2}R \\ &= v_{\mathrm{P},1}U_n+h^2\bar{v}_{\mathrm{P},1}\bar{U}_n+\sqrt{n}h^{m+5/2}\left\{\sigma_{\mathrm{P},1}V_n+\bar{\sigma}_{\mathrm{P},1}\bar{V}_n\right\} \\ &\quad +o_P\left(\sqrt{n}h^{m+5/2}\right)+O_P\left(nh^{4m+5/2}\right)+O_P\left(\frac{\log\left(n\right)^2}{nh^{1/2}}\right). \end{aligned}$$

The first part of the theorem now follows under the conditions imposed on the bandwidth.

Next, consider $\bar{T}_{\mathrm{P},2}$: By the same arguments as before, we may set

$$\tilde{\sigma}^2\left(x\right)=\frac{2}{\pi\left(x\right)}\int_l^x\mu\left(y\right)\pi\left(y\right)dy,\quad \hat{\sigma}^2\left(x\right)=\frac{2}{\hat{\pi}\left(x\right)}\int_l^x\mu\left(y\right)\pi\left(y\right)dy$$

in the following. Thus, using that $\int_l^x\mu\left(y\right)\pi\left(y\right)dy=\pi\left(x\right)\sigma^2\left(x\right)$ together with the mean-value theorem,

$$\begin{aligned} \bar{T}_{\mathrm{P},2} &= \int_I[\tilde{\sigma}^2(x)-\hat{\sigma}^2\left(x\right)]^2\bar{w}\left(x\right)dx \\ &= 4\int_I\left[\frac{1}{\pi\left(x\right)}-\frac{1}{\hat{\pi}\left(x\right)}\right]^2\pi^2\left(x\right)\sigma^4\left(x\right)\bar{w}\left(x\right)dx \\ &= 4\int_I\left[\pi\left(x\right)-\hat{\pi}\left(x\right)\right]^2\frac{\sigma^4\left(x\right)\bar{w}\left(x\right)}{\pi^2\left(x\right)}dx+R, \end{aligned}$$

where

$$R=O_P\left(\sup_x\left|\pi\left(x\right)-\hat{\pi}\left(x\right)\right|^4\right)=O_P\left(h^{4m}\right)+O_P\left(\frac{\log\left(n\right)^2}{n^2h^2}\right).$$

From Gourieroux and Tenreiro (2001, Theorem 3.2) and the usual bias expressions, we now obtain

$$nh^{1/2}\left\{\bar{T}_{\mathrm{P},2}-\bar{m}_{\mathrm{P},2}\right\}=\bar{v}_{\mathrm{P},2}U_n+\sqrt{n}h^{m+1/2}\sigma_{\mathrm{P},2}V_n+O_P\left(nh^{2m+1/2}\right)+nh^{1/2}R_n,$$

where

$$\bar{m}_{\mathrm{P},2}=\frac{4}{nh}\int_{\mathbb{R}}K^2\left(z\right)dz\times\int_I\pi(x)\left\{\frac{\sigma^4\left(x\right)\bar{w}\left(x\right)}{\pi^2\left(x\right)}\right\}dx=\frac{4}{nh}\int_{\mathbb{R}}K^2\left(z\right)dz\times\int_I\frac{\sigma^4\left(x\right)\bar{w}\left(x\right)}{\pi(x)}dx,$$

and

$$\begin{aligned} \bar{v}_{\mathrm{P},2} &= 2\int_{\mathbb{R}}\left(K*K\right)^2\left(z\right)dz\times\int_I\pi^2(x)\left\{\frac{4\sigma^4\left(x\right)\bar{w}\left(x\right)}{\pi^2\left(x\right)}\right\}^2dx \\ &= 32\int_{\mathbb{R}}\left(K*K\right)^2\left(z\right)dz\times\int_I\frac{\sigma^8\left(x\right)\bar{w}^2\left(x\right)}{\pi^2\left(x\right)}dx \end{aligned}$$

■

Proof of Theorem 8. Consider first $\bar{T}_{\mathrm{P},1}$, where the drift under the null, $\mu_P(x)$, say, can be written as $\mu_{\mathrm{P}}(x)=\mu_n(x)+g_n\left(x\right)$. Then,

$$\bar{T}_{\mathrm{P},1}=\int_I[\hat{\mu}\left(x\right)-\mu_n(x)]^2\bar{w}\left(x\right)dx+\int_I g_n^2(x)\bar{w}\left(x\right)dx+2\int_I[\hat{\mu}\left(x\right)-\mu_n(x)]g_n(x)\bar{w}\left(x\right)dx.$$

The first term is treated as in the Proof of Theorem 7, while the third term is a higher-order term which can be ignored. Regarding $\bar{T}_{\mathrm{P},2}$, the diffusion term under the null can be written as $\sigma_{\mathrm{P}}^2(x)=\sigma_n^2(x)+g_n\left(x\right)$ such that

$$\bar{T}_{\mathrm{P},2}=\int_I[\hat{\sigma}^2(x)-\sigma_n^2(x)]^2\bar{w}\left(x\right)dx+\int_I g_n^2(x)\bar{w}\left(x\right)dx+2\int_I[\hat{\sigma}^2(x)-\sigma_n^2(x)]g_n(x)\bar{w}\left(x\right)dx,$$

and we proceed as with $\bar{T}_{\mathrm{P},1}$. ■

B Auxiliary Lemmas

Lemma 9 *Assume that (A.1)-(A.4) and (B.1)-(B.2) hold. Then:*

$$\sup_{x\in I}\left|\hat{\mu}_{\mathrm{SP},1}(x)-\tau_a(\pi(x))\,\mu_{\mathrm{SP},1}(x)\right|=\sum_{k=0}^{1}\left\{O_P\left(n^{-1/2}\sqrt{\log(n)}a^{-2+k}h^{-(1+2k)/2}\right)+O_P\left(a^{-2+k}h^m\right)\right\},$$

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

Proof. This follows along the same lines as Kristensen (2010, Proofs of Lemmas 9-10). ■

C Figures

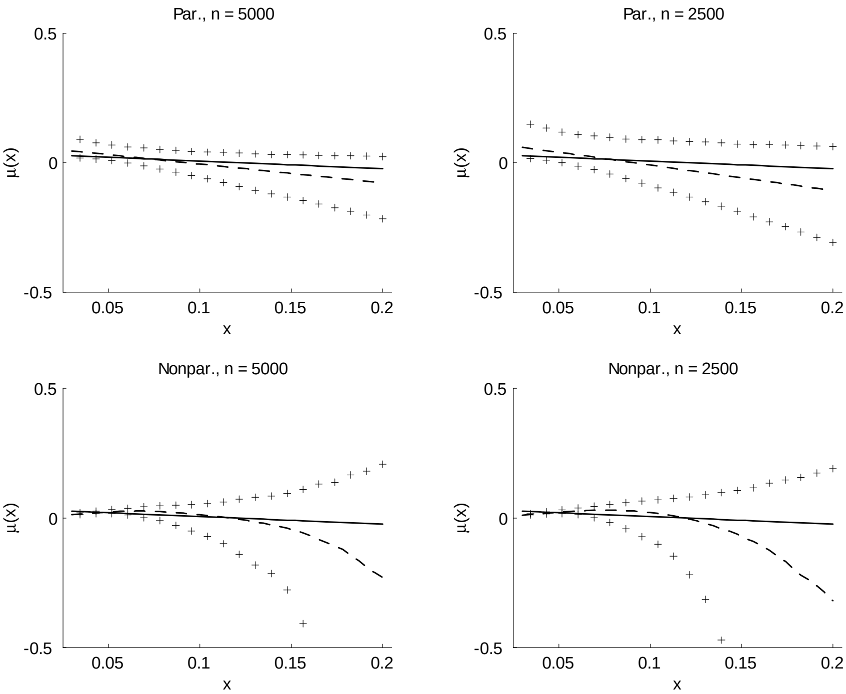

Figure 1: Estimates of $\mu(x)$ for the CKLS(i) model. Full line = true function, dotted line = mean of estimate, plusses = 95% confidence interval.

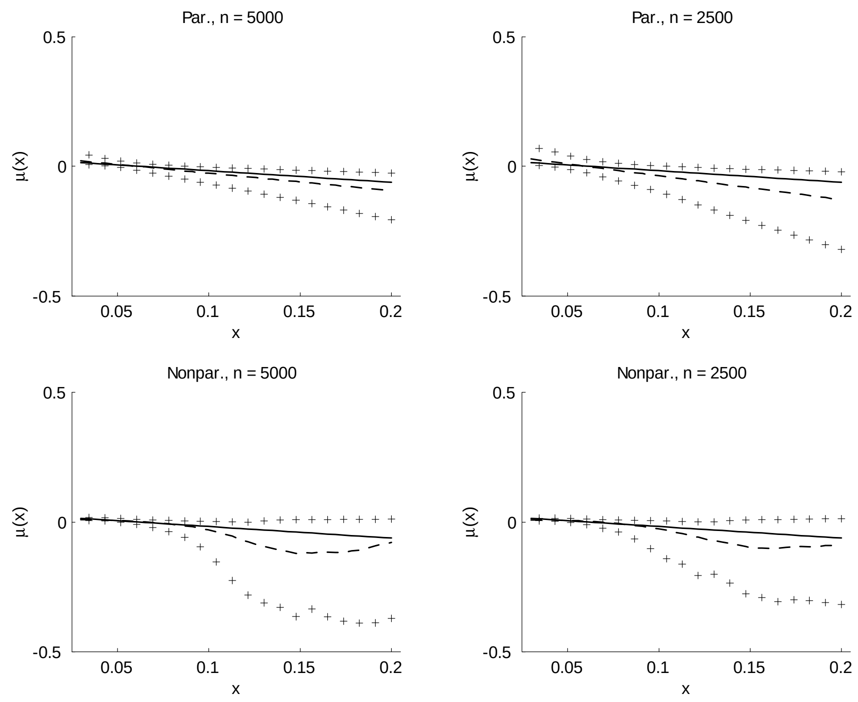

Figure 2: Estimates of $\mu(x)$ for the CKLS(ii) model. Full line = true function, dotted line = mean of estimate, plusses = 95% confidence interval.

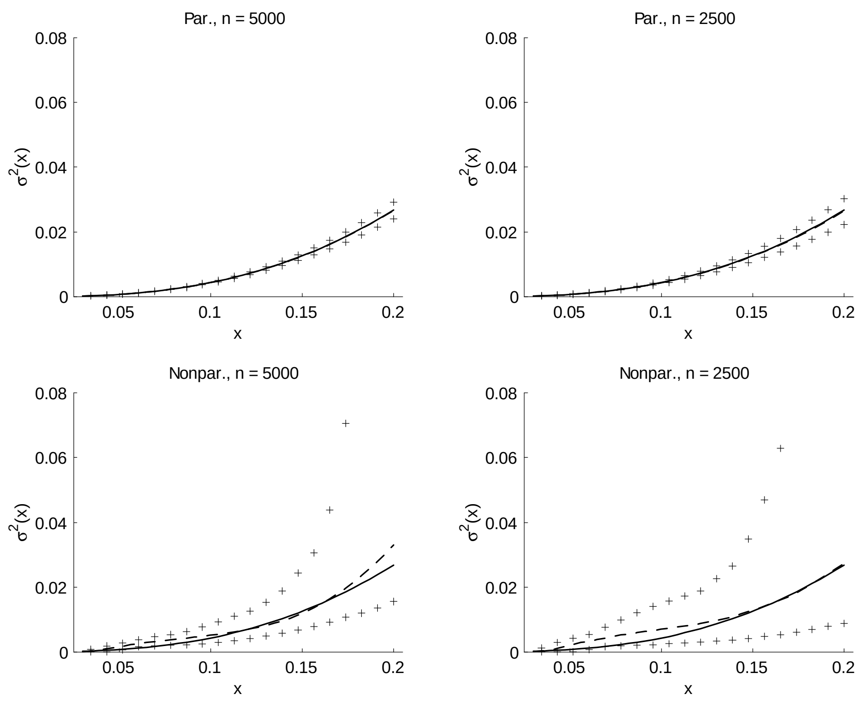

Figure 3: Estimates of $\sigma^2(x)$ for the CKLS(i) model. Full line = true function, dotted line = mean of estimate, plusses = 95% confidence interval.

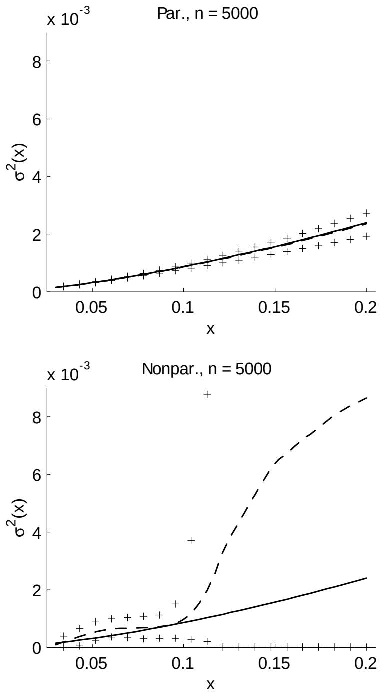

Figure 4: Estimates of $\sigma^2(x)$ for the CKLS(ii) model. Full line = true function, dotted line = mean of estimate, plusses = 95% confidence interval.

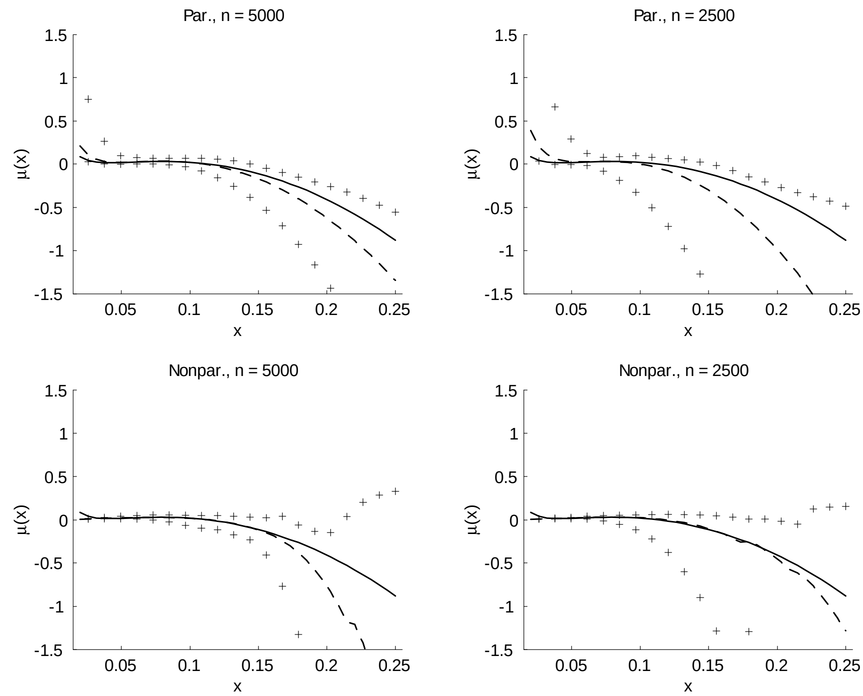

Figure 5: Estimates of $\mu(x)$ for the AS model. Full line = true function, dotted line = mean of estimate, plusses = 95% confidence interval.

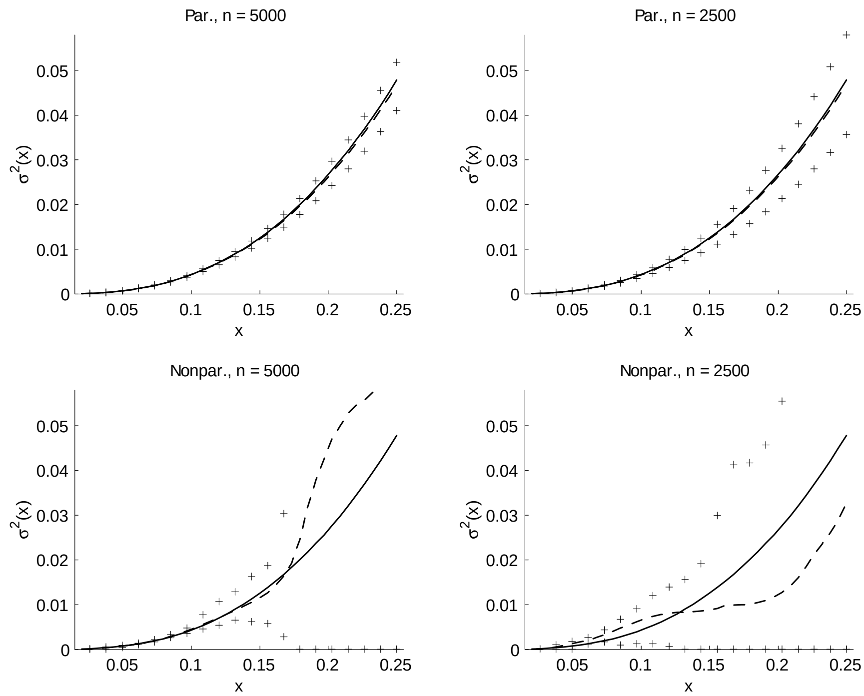

Figure 6: Estimates of $\sigma^2(x)$ for the AS model. Full line = true function, dotted line = mean of estimate, plusses = 95% confidence interval.

Research Papers 2010

CREATES

Center for Research in Econometric Analysis of Time Series

2010-29: Nikolaus Hautsch and Mark Podolskij: Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence

2010-30: Martin M. Andreasen: Non-linear DSGE Models and The Central Difference Kalman Filter

2010-31: Morten Ørregaard Nielsen and Per Frederiksen: Fully Modified Narrow-Band Least Squares Estimation of Weak Fractional Cointegration

2010-32: Mogens Bladt and Michael Sørensen: Simple simulation of diffusion bridges with application to likelihood inference for diffusions

2010-33: Fernando Baltazar-Larios and Michael Sørensen: Maximum likelihood estimation for integrated diffusion processes

2010-34: Leonidas Tsiaras: The Forecast Performance of Competing Implied Volatility Measures: The Case of Individual Stocks

2010-35: Leonidas Tsiaras: Dynamic Models of Exchange Rate Dependence Using Option Prices and Historical Returns

2010-36: Robinson Kruse and Rickard Sandberg: Linearity Testing in Time-Varying Smooth Transition Autoregressive Models under Unknown Degree of Persistency

2010-37: Tom Engsted and Thomas Q. Pedersen: The log-linear return approximation, bubbles, and predictability

2010-38: Thomas Q. Pedersen: Predictable return distributions

2010-39: Rasmus Tangsgaard Varneskov: The Role of Dynamic Specification in Forecasting Volatility in the Presence of Jumps and Noisy High-Frequency Data

2010-40: Antonis Papapantoleon and David Skovmand: Picard Approximation of Stochastic Differential Equations and Application to Libor Models

2010-41: Ole E. Barndorff–Nielsen, Fred Espen Benth and Almut E. D. Veraart: Modelling electricity forward markets by ambit fields

2010-42: Robinson Kruse and Philipp Sibbertsen: Long memory and changing persistence

2010-43: Dennis Kristensen: Semi-Nonparametric Estimation and Misspecification Testing of Diffusion Models