Estimation of Partial Differential Equations with Applications in Finance[*]

Dennis Kristensen[†], London School of Economics

June 7, 2004

Abstract

Linear parabolic partial differential equations (PDE's) and diffusion models are closely linked through the celebrated Feynman-Kac representation of solutions to PDE's. In asset pricing theory, this leads to the representation of derivative prices as solutions to PDE's. We give a number of examples of this, including the pricing of bonds and interest rate derivatives. Very often derivative prices are calculated given preliminary estimates of the diffusion model for the underlying variable. We demonstrate that the derivative prices are consistent and asymptotically normally distributed under general conditions. We apply this result to three leading cases of preliminary estimators: Nonparametric, semiparametric and fully parametric ones. In all three cases, the asymptotic distribution of the solution is derived. Our general results have other applications in asset pricing theory and in the estimation of diffusion models; these are also discussed.

1   Introduction

Partial differential equations (PDE's) are used in fields as diverse as physics, biology, economics, and finance to model and analyse dynamic systems. One class of PDE's which has received particular attention are the linear parabolic ones (LPDE's). These make up a large class of PDE's which is of a sufficiently simple structure such that a thorough analysis of them is possible, see e.g. Friedman (1964) and Evans (1998) for an introduction and detailed analysis of their properties.

One area where LPDE's play an essential role is in asset pricing theory in general and in the pricing of financial derivatives in particular. The latter are securities whose pay-off is contingent of the value of an underlying variable, this for example being a stock price or an interest rate. The option pricing literature was revolutionised by the groundbreaking work of Black and Scholes (1973) and Merton (1973, 1976). Assuming that the underlying asset follows a geometric Brownian motion and that trading takes place in continuous time, they derived the price of an option as the solution to a LPDE using hedging and no-arbitrage arguments. This result has since then been generalised in various directions. In particular, the restriction that the fundamental asset price follows a geometric Brownian motion can be weakened to allow for basically any diffusion type process.

In the above framework, the option price is a functional of the so-called drift and diffusion term, these being functions characterising the diffusion process that the underlying asset is assumed to follow. Empirical applications of these option pricing formulae therefore almost always involve some sort of calibration of the drift and diffusion term. These calibrated terms can then substituted into the LPDE in place of the true but unknown ones, and the option price solved for. The calibration is often done by statistical estimation based on historical data. The implied option prices therefore inherit the statistical uncertainty associated with the estimated drift and diffusion term. It will be valuable to be able to measure how the estimation error (e.g. in terms of standard errors) in the drift and diffusion term affects the resulting option prices. This will allow one to evaluate the accuracy of the estimated prices. Moreover, such results can be used to construct a direct statistical test of the option price model by comparing the estimated prices with the observed ones.

We give general results for the asymptotic properties of the implied option prices given preliminary estimators of the drift and diffusion term. The implied/estimated price is obtained as the solution to a LPDE where the preliminary estimators have been plugged in. We shall here show that the estimated solution will be consistent when the preliminary estimators are. We also give general conditions under which the solution will be asymptotically normal distributed. In the option pricing framework, this means that the estimated prices are consistent if the drift and diffusion estimators for the underlying asset price diffusion are. Furthermore, we are able to calculate standard errors for the prices. We first state this result under fairly general conditions. We then verify these conditions for three specific types of preliminary estimators, a parametric, a semiparametric and a nonparametric one, and derive the asymptotic distribution in each case.

Similar results to the ones derived here can be found elsewhere in the literature. In the Black-Scholes model, the statistical properties of option prices given preliminary estimates of the diffusion term has been considered in a number of studies, see e.g. Boyle and Ananthanarayanan (1977) and Ncube and Satchell (1997). In a very general setting, Lo (1986) derived the asymptotic properties of

the implied option prices given preliminary parametric estimates of the drift and diffusion function. However, this was done under high-level conditions, and it was not verified that these actually hold. Furthermore, he was not able to give closed form expressions for the asymptotic distribution. Interest rate derivative pricing given kernel estimators of the short rate model was considered in Aït-Sahalia (1996a) and Jiang (1998). Our results extend these results to basically any asset pricing model which is driven by a finite number of state variables, and virtually any estimator of the drift and diffusion term in the model in question. In particular, our results include multi-factor interest rate models and stochastic volatility models. In the parametric case, we are able to derive an explicit expression of the asymptotic distribution which allows one to estimate this. In the general case, we are not able to do this; we are however still able to define a simple estimator of the asymptotic distribution which should be consistent.

Other applications of our general results are also available in the econometric analysis of diffusion models, e.g. GMM-type estimators [Bibby and Sørensen 1995, Duffie and Singleton 1993] and the estimation using observed option prices. We give a brief discussion of these applications.

Studies of solutions to (partial) differential equations given preliminary estimates of the driving coefficients are found elsewhere in the literature. Hausman and Newey (1995) consider a non-linear ODE and derive the asymptotic properties of an estimator of the solution when a preliminary estimator of the driving function is available. Vanhems (2003) deals with a similar problem where a nonlinear ODE depends on a conditional mean function. The conditional mean is then estimated by kernel methods, and the associated estimated solution is analysed. PDE's have also received some attention, in particular in the financial econometrics literature. In Aït-Sahalia (1996a), the estimation of interest rate derivative prices is treated given preliminary semiparametric estimators of the drift and diffusion function of the short-term interest rate. His analysis is based on a deterministic characterisation of the solution to the PDE as given in Friedman (1964), which he analyses using the functional delta method of Aït-Sahalia (1993). Jiang (1998) follows the same approach when analysing the properties of estimated option prices given fully nonparametric estimators of the drift and diffusion term. Finally, Chow et al (1999) also consider nonparametric estimation in the context of PDE's. But while we are concerned with the estimation of the solution given preliminary estimators of coefficients entering the PDE, they assume that the solution of the PDE has been observed with error, and then use this to estimate parameters entering the PDE.

A very nice feature of the class of LPDE's is the probabilistic interpretation which a solution to any PDE of this type can be given: Under weak regularity conditions the solution can be characterised as the conditional moment of a solution to an associated diffusion process. This is the celebrated Feynman-Kac Representation of solutions to LPDE's. This is exactly the link that allows one to translate the option price as the discounted expected value of the future price into the solution to a LPDE. Our analysis of the estimated solution is based on this stochastic representation as a conditional expectation involving a diffusion process. This approach has proved very fruitful in the analysis of various other problems related to this type of PDE's, see e.g. Freidlin (1985) for an exposition. So instead of directly working with the PDE of interest, we shall focus on a certain class of conditional moments of the associated stochastic differential equation (SDE) in terms of which the solution to the PDE can be expressed. One advantage of this approach is that while in the general case it is difficult to set up conditions for

the existence of a global solution to the PDE, the conditional moments of the SDE of interest will be well-defined under weak conditions. Another is that a closed form expression of the conditional moment is available which facilitates the statistical analysis of the estimator.

Once the general asymptotic result has been established, we apply it to three leading preliminary estimators: Fully parametric estimators of the drift and diffusion term (including MLE and GMM), semiparametric ones (see Aït-Sahalia 1996a and Chapter 4), and fully nonparametric ones (see Jiang and Knight 1997 and Bandi and Phillips 2003). In all three cases, we are able to derive the convergence rate and the asymptotic distribution of the solution. In particular, we demonstrate that even if non- and semiparametric preliminary estimators are used, the associated solution will converge with parametric rate. This appealing result follows from the higher level of regularity/smoothness of the solution to the PDE compared to the preliminary estimators. This is a well-known phenomenon found elsewhere in the literature on nonparametric estimation. One important consequence is that if the end goal of the econometric analysis of the asset price model is the pricing of derivatives, one will asymptotically in many cases be better off using non- and semiparametric estimators: These allow for a higher level of flexibility without slowing down the rate of convergence of the solution. Of course, if one has correctly specified a parametric model of the underlying SDE, a parametric estimator of the solution will in most cases enjoy higher efficiency and better finite sample properties than the nonparametric one. Moreover, inherent in nonparametric estimation is a problem of choosing some smoothing parameter; this problem, one does not face in a parametric setting.

The study is organised as follows. In the next section we first present the class of PDE's of interest and derive some useful properties of these; we then discuss various applications to finance and estimation of diffusions. In Section 3, a general result concerning consistency and asymptotic normality is first presented which is then applied to the aforementioned three types of estimators. These econometric results are then put into the framework of derivative pricing in Section 4, which also contains a discussion on the application of our results to GMM-type estimation of diffusion models and estimation based on observed option prices. Section 5 concludes. All proofs and lemmas have been relegated to appendix A and B respectively.

2   Linear Parabolic Partial Differential Equations

We shall in the following introduce the class of linear parabolic PDE's together with the concept of generalised solutions to these. We give conditions for these to be well-defined. The section ends with a presentation of the various applications of LPDE's to finance and estimation of diffusion models.

For any two functions $\mu : [0, \infty) \times \mathbb{R}^q \mapsto \mathbb{R}^q$ and $\sigma^2 : [0, \infty) \times \mathbb{R}^q \mapsto \mathbb{R}^{q \times q}$, we define the linear second order differential operator

$$L_t\left(u\right) = \sum_{i=1}^{q} \mu_i\left(t, x\right) \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{q} \sigma_{ij}^2\left(t, x\right) \frac{\partial^2 u}{\partial x_i \partial x_j}.$$

This is normally referred to as the infinitesimal generator, c.f. Karatzas and Shreve (1991, p. 281). For

$T > 0$, we shall then consider solutions $u : [0, T] \times \mathbb{R}^q \mapsto \mathbb{R}$ to the following Cauchy problem,

$$-\frac{\partial u}{\partial t} + au = L_t(u) + c, \tag{1}$$

$$u(T, x) = b(x), \tag{2}$$

for given functions $a : [0, T] \times \mathbb{R}^q \to [0, \infty)$, $b : \mathbb{R}^q \mapsto \mathbb{R}$ and $c : [0, T] \times \mathbb{R}^q \mapsto \mathbb{R}$.

Only in a few special cases is it possible to derive an explicit expression of the solution. This of course complicates the analysis of solutions to general PDE's, but one can get quite far by using implicit representations found in the literature. Friedman (1964) derives a deterministic expression of the solution; this is however very involved and appears difficult to work with. Instead, we shall here rely on the so-called Feynman-Kac Representation: This establishes a direct link between the solution to (1)-(2) and a conditional moment of the process $\{X_t\}$ solving a SDE,

$$\text{SDE}(\mu, \sigma) : dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \quad 0 \le t \le T, \tag{3}$$

with $\{W_t\}$ being a q-dimensional standard Brownian motion.[1] If a solution exists to (1)-(2), and certain growth conditions on c, b, and u are satisfied, we obtain that

$$u(t, x) = E_{t,x}\left[b(X_T)\exp\left[-\int_t^T a(u, X_u)\, du\right]\right] \tag{4}$$
$$+ E_{t,x}\left[\int_t^T c(s, X_s)\exp\left[-\int_t^s a(u, X_u)\, du\right] ds\right];$$

where $E_{t,x}[\cdot] = E[\cdot|X_t = x]$, see for example Karatzas and Shreve (1991, Theorem 5.7.6). We follow Freidlin (1985, p. 122) and call the Feyman-Kac representation of u the *generalised solution* to (1)-(2), since this may be well-defined even if no solution to the PDE exists. In our analysis of u we shall choose to work with this stochastic representation. The reason for this is that the solution can be written up in an explicit form in contrast to the deterministic approach.

The econometric problem which shall be considered here is the estimation of u given preliminary estimators of μ and σ^2. Initially, we do not make any assumptions about the nature of these estimators, but in most cases they arrive from historical observations of a process solving the SDE (3). Let μ_0 and σ_0^2 denote the true but unknown values of the drift and diffusion term, and $\{X_t^0\}$ the solution to $\text{SDE}(\mu_0, \sigma_0^2)$. Let u_0 denote the associated solution obtained from (4) with $\{X_t^0\}$ plugged in. Now, assume that $(\hat{\mu}, \hat{\sigma}^2)$ is a pair of estimators of (μ_0, σ_0^2). An obvious estimator of u is then obtained in the following manner: First, plug $(\hat{\mu}, \hat{\sigma}^2)$ into $\text{SDE}(\mu, \sigma)$ as given in (3). This yields an estimator of $\{X_t^0\}$ which we denote $\{\hat{X}_t\}$; this is then in turn plugged into (4), thereby obtaining an estimator of u_0 which we denote by $\hat{u}$. We are then interested in the asymptotic properties of $\hat{u}$, in particular we wish to give conditions for $\hat{u}(t, x)$ to be consistent and for

$$V_n^{-1/2}(t, x)(\hat{u}(t, x) - u_0(t, x)) \xrightarrow{d} N(0, 1)$$

to hold, for any $(t, x) \in [0, T] \times \mathbb{R}^q$, where $\{V_n(t, x)\}$ is some, possibly random, sequence. In the next section, we give precise conditions under which this result will hold.

---

[1] Here, we have implicitly assumed that $\sigma^2(t, x)$ is nonnegative definite such that the matrix square root, $\sigma(t, x)$, is well-defined.

Since the solutions in most cases cannot be written on an explicit form, numerical methods are normally used to solve the solution to the PDE (1)-(2). Hull (1997, Chapter 15) provides an overview of a number of numerical methods used in finance. The two most popular methods is the so-called finite-difference method and Monte Carlo methods. A thorough treatment of numerical solutions of PDE's using finite difference methods can be found in Ames (1992). Alternatively, the solution u can be obtained by the use of Monte Carlo methods; these are normally based on the Feynman-Kac representation. The Monte Carlo simulations can be done in the following manner: Let $\{X_s^{(i)} | t \leq s \leq T\}$, $i = 1, ..., N$, be N independent simulated paths of the SDE (3) with initial condition $X_t = x$. We then approximate $u(t, x)$ by

$$
\begin{aligned}
u^{(N)}(t, x) \;=\; & \frac{1}{N} \sum_{i=1}^{N} \left[b\left(X_T^{(i)} \right) \exp\left[-\int_t^T a\left(u, X_u^{(i)} \right) du \right] \right] \\
& + \frac{1}{N} \sum_{i=1}^{N} \left[\int_t^T c\left(s, X_s^{(i)} \right) \exp\left[-\int_t^s a\left(u, X_u^{(i)} \right) du \right] ds \right].
\end{aligned}
\tag{5}
$$

Let P^* denote the probability measure that we simulate under. Then $E^{P^*}\left[u^{(N)}(t, x) \right] = u(t, x)$, and, by the strong Law of Large Numbers, $u^{(N)}(t, x) \to^{P^*-\text{a.s.}} u(t, x)$ as $N \to \infty$. It is however not possible to obtain an exact continuous sample path of this type of stochastic processes; instead one often derives an approximate discrete time version of (3) from which one simulates. This approximate model can be chosen arbitrarily close to the actual one. For an overview of simulations of SDE's, we refer to Kloeden and Platen (1999).

We now wish to discuss the question of existence and uniqueness of the generalised solution and derive some of its properties. These will prove useful in the subsequent section when we deal with the econometric problem in question. Sufficient conditions for a solution to (1)-(2) can be found in Friedman (1964, Section I.4) and Evans (1998, Chapter 5). In the following, we construct a set of function pairs, $\mathcal{D}$, such that for any $(\mu, \sigma^2) \in \mathcal{D}$, the associated generalised solution u exists and is sufficiently well-behaved. This is done by restricting $\mathcal{D}$ in the following manner:

Definition The space $\mathcal{D}$ consists of all function pairs (μ, σ^2) where

1. μ and σ^2 are twice continuously differentiable in x such that:

   (a) There exists $K > 0$ such that

   $$
   \left\| \partial_x^\alpha \mu(t, x) \right\| \leq K(1 + \|x\|), \quad \left\| \partial_x^\alpha \sigma^2(t, x) \right\| \leq K(1 + \|x\|),
   $$

   for all $(t, x) \in [0, T] \times \mathbb{R}^q$ and $|\alpha| \leq 2$.

   (b) For all $N \geq 1$, there exists $K_N > 0$ such that

   $$
   \left\| \mu(t, x) - \mu(t, y) \right\| \leq K_N \|x - y\|, \quad \left\| \sigma^2(t, x) - \sigma^2(t, y) \right\| \leq K_N \|x - y\|,
   $$

   for all $t \in [0, T]$ and $\|x\|, \|y\| \leq N$.

2. There exists a constant $\underline{\sigma}^2 > 0$ such that $\sum_{i,j=1}^{q} \sigma_{ij}^2(t, x) y_i y_j \geq \underline{\sigma}^2 \|y\|^2$ for all $y \in \mathbb{R}^q$ and $(t, x) \in [0, T] \times \mathbb{R}^q$.

Observe that $\mathcal{D}$ is a well-defined function space. For any $\left(\mu, \sigma^2\right) \in \mathcal{D}$ and any initial condition, $X_0 = X^*$, which is independent of $\{W_t\}$ and satisfies $E\left[\|X^*\|^2\right] < \infty$, there exists an associated unique strong solution to (3), c.f. Friedman (1975, Theorem 5.2.2). Furthermore, if $E\left[\|X^*\|^{2p}\right] < \infty$, for some $p \geq 1$,

$$E\left[\|X_t\|^{2p}\right] \leq \left(1 + E\left[\|X^*\|^{2p}\right]\right) e^{C^* t} \tag{6}$$

for $0 \leq t \leq T$, where $C^* = C^*\left(K, p, T\right)$, c.f. Friedman (1975, Theorem 5.2.3). For $q = 1$, a weaker sufficient condition for existence and uniqueness is that μ and σ^2 are continuously differentiable and $\sigma^2\left(\cdot\right) > 0$, c.f. Karatzas and Shreve (1991, Theorem 5.5.15 and Corollary 5.3.23). The bound in (6) does not necessarily hold in this case however. For $q > 1$, weaker conditions for existence and uniqueness can be found in Meyn and Tweedie (1993). Most likely the results presented in the following hold for $\left(\mu, \sigma^2\right)$ situated in a larger function space, but for simplicity we shall restrict them to belong to $\mathcal{D}$. The existence and uniqueness results for $\{X_t\}$ hold without the differentiability conditions on μ and σ; these are used when we derive the asymptotic properties of $\hat{u}$.

In the following we consider a fixed pair $\left(\mu_0, \sigma_0^2\right) \in \mathcal{D}$, and denote the associated diffusion process by $\left\{X_t^0\right\}$. We also fix the initial condition of $\left\{X_t^0\right\}$ at some given random variable, X^*. First we define $L_p\left(X^*, [0,T] \times \mathbb{R}^q\right)$ as the space of functions $f : [0,T] \times \mathbb{R}^q \mapsto \mathbb{R}$ for which $E\left[\int_0^T \left|f\left(t, X_t^0\right)\right|^p dt\right] < \infty$. Next, we introduce a Sobolev-like space $W^{m,p} = W^{m,p}\left(X^*, [0,T] \times \mathbb{R}^q\right)$ for any $p \geq 1$ and $m \geq 0$. This is defined as the space of functions $f : [0,T] \times \mathbb{R}^q \mapsto \mathbb{R}$ which are m times continuously differentiable in their second argument and with $\partial_x^\alpha f \in L_p\left(X^*, [0,T] \times \mathbb{R}^q\right)$ for any $\alpha \in \{0, ..., k\}^q$ with $|\alpha| = \sum_{i=1}^q \alpha_i = k$, $0 \leq k \leq m$. We equip the space with the norm

$$\|f\|_{m,p} = \left(\sum_{|\alpha| \leq m} E\left[\int_0^T \left|\partial_x^\alpha f\left(t, X_t^0\right)\right|^p dt\right]\right)^{1/p}.$$

Observe that $W^{m,2}$ is a Hilbert space with inner product

$$\langle f, g\rangle_m = \sum_{|\alpha| \leq m} E\left[\int_0^T \partial_x^\alpha f\left(t, X_t^0\right) \partial_x^\alpha g\left(t, X_t^0\right) dt\right] \tag{7}$$

and that $W^{0,p} = L_p\left(X^*, [0,T] \times \mathbb{R}^q\right)$. Combining the above results, we observe that if (i) f has m derivatives in its second argument and these satisfies $\|\partial_x^\alpha f\left(t, x\right)\| \leq C\left(1 + \|x\|^r\right)$, $|\alpha| \leq m$, (ii) $\left(\mu_0, \sigma_0^2\right) \in \mathcal{D}$ and (iii) $E\left[\|X^*\|^{p^*}\right] < \infty$, then $f \in W^{m,p}$ with $p = p^*/r$. In particular, for any $\left(\mu, \sigma^2\right) \in \mathcal{D}$, $\left(\mu, \sigma^2\right) \in W^{2,p} \times W^{2,p}$ with $p \leq p^*$.

We impose the following conditions on the functions a, b, and c:

Condition 1 For some $r \geq 1$, $\left|\partial_x^\alpha a\left(t, x\right)\right| \leq C\left(1 + \|x\|^r\right)$, $\left|\partial_x^\alpha b\left(t, x\right)\right| \leq C\left(1 + \|x\|^r\right)$ and $\left|\partial_x^\alpha c\left(t, x\right)\right| \leq C\left(1 + \|x\|^r\right)$, $|\alpha| \leq 2$.

It is not always the case in our applications that the functions are differentiable as assumed here. We conjecture that the results also hold in this case. All the following results are derived under the implicitly maintained assumption that Condition 1 holds. The first result ensures that u exists and is well-defined for suitable choices of μ and σ^2:

Theorem 1 *For any $\left(\mu, \sigma^2\right) \in \mathcal{D}$, the associated generalised solution u exists. Furthermore, $\left|\partial_x^\alpha u\left(t, x\right)\right| \leq C\left(T\right)\left(1 + \|x\|^r\right)$ for $|\alpha| \leq 2$. In particular, $u \in W^{2,p}$ for any initial condition X^* with $E\left[\|X^*\|^{pr}\right] < \infty$.*

2.1 Applications in Finance

One particular area where PDE's of the linear parabolic type is widely used is in asset pricing theory in general and derivative pricing in particular. Derivatives are securities whose pay-off depends on some underlying variable, e.g. the price of a stock or an interest rate, with the most well-known example being options. Financial derivatives play an important role in the financial markets, and have consequently received great attention in the finance literature. Since the seminal work by Black-Scholes (1973) and Merton (1973), diffusion processes have played a prominent part in the asset pricing literature. Assuming that the fundamental asset prices solve an SDE, one is able to derive closed form solutions of derivative prices. In fact, one of the main results is that the price of the derivative is the solution to a PDE in the class considered here. Below, we give a brief overview of the various fields where our results can be applied. These examples illustrate the wide range of applications that parabolic PDE's have.

We first introduce the necessary notation. We fix the probability space $(P, \Omega, \mathcal{F})$ with an associated filtration $\{\mathcal{F}_t\}$. Here P denotes the physical measure under which we observe the processes introduced in the following.

Example 1: A General Asset Pricing Model. Consider a riskless asset $\{\beta_t\}$ given by

$$d\beta_t = r_t \beta_t dt,$$

for some adapted short-term interest rate process, $\{r_t\}$, see Chapter 2 for a discussion of these. We are also given N risky traded assets, each having an associated price process $\{S_t^{(i)}\}$, $i = 1, ..., N$. We assume that the process $\{S_t\}$, $S_t = (S_t^{(1)}, ..., S_t^{(N)})^\top$, solves a SDE,

$$dS_t = \mu_S\left(t, S_t\right) dt + \sigma_S\left(t, S_t\right) dw_t^S, \tag{8}$$

where $\left\{W_t^S\right\}$ is a N-dimensional standard Brownian motion. Each asset i has also an associated dividend stream $\{d_t^{(i)}\}$, $i = 1, ..., N$, which we collect in $\{d_t\}$, $d_t = (d_t^{(1)}, ..., d_t^{(N)})^\top$. Given the existence of an equivalent martingale measure, Q,[2] the price process then satisfies

$$S_t = E_t^Q\left[\exp\left[-\int_t^T r_s ds\right] S_T + \int_t^T \exp\left[-\int_t^s r_s du\right] d_s ds\right], \tag{9}$$

where $\{S_t\}$ has dynamics

$$dS_t = r_t S_t dt + \sigma_S\left(t, S_t\right) dW_t^S \tag{10}$$

under Q, see for example Duffie (1996, Chapter 6 and 8). Observe that μ_S does not enter the dynamics of $\{S_t\}$ under Q, and therefore has no influence on the option prices. Assume that $\{r_t\}$ and $\{d_t\}$ also solve SDE's under Q,

$$\begin{aligned} dr_t &= \mu_r\left(t, r_t\right) dt + \sigma_r\left(t, r_t\right) dW_t^r, \\ dd_t &= \mu_D\left(t, S_t\right) dt + \sigma_D\left(t, S_t\right) dW_t^D. \end{aligned}$$

---

[2] We shall not discuss conditions for the existence (and uniqueness) of Q, and merely assume its existence.

Then by defining

$$X_t = \left(S_t^\top, d_t^\top, r_t \right)^\top, \quad W_t = (W_t^S, W_t^D, W_t^r)^\top,$$

$$\mu(t,x) = \left(rS^\top, \mu_D^\top(t,S), \mu_r(t,r) \right)^\top, \quad \sigma(t,x) = \operatorname{diag}\left(\sigma_S(t,S), \sigma_D(t,S), \sigma_r(t,r) \right),$$

we observe that the pricing formula (9) takes the form of (4). More advanced models for the short term interest rate as presented below can without any problems be allowed for.

Example 1.1: The Black-Scholes Model. A special case of the above model is the (extended) Black-Scholes model where we have one risky asset ($N = 1$), say a stock, and a derivative on this stock. At time of maturity T, the derivative pays off $b(S_T)$. From (9), the following expression of the price of the derivative at time t, $\Pi_t(T)$, presents itself,

$$\Pi_t(T) = E_t^Q \left[\exp\left[-\int_t^T r_s ds \right] P_T(T) \right] = E_t^Q \left[\exp\left[-\int_t^T r_s ds \right] b(S_T) \right],$$

where $\{S_t\}$ solves (10) under Q. In the classic Black-Scholes model, it is assumed that the short-term interest rate is constant, $r_t \equiv r > 0$, and that $\{S_t\}$ is a geometric Brownian motion under P; that is,

$$dS_t = \mu S_t dt + \sigma S_t dw_t^S.$$

We then consider a call-option where the pay-off function is $b(x) = \max\{x - K, 0\}$ with K being the strike price.[3] In this case, the above conditional expectation can be shown to satisfy

$$\Pi_t(T) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution and

$$d_1 = \frac{\log(S/K) + \left(r + \sigma^2/2 \right)(T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}.$$

In the general case with more complex dynamics of $\{S_t\}$ and/or stochastic interest rates, an explicit expression for $\Pi_t(T)$ is not available. Instead, one has to rely on numerical methods to calculate the actual prices as discussed earlier.

Example 1.2: Stochastic Volatility Models. The classic Black-Scholes model is not able to match observed option prices very well. To deal with this empirical shortcoming stochastic volatility models were introduced, see e.g. Ghysels et al (1996) for a review. We still consider some stock price process $\{S_t\}$ but we now assume that

$$dS_t = \mu_S(S_t)\, dt + \sigma_S(S_t, v_t)\, dw_t^S, \tag{11}$$

where $\{v_t\}$ is non-traded/unobserved process solving

$$dv_t = \mu_v(v_t)\, dt + \sigma_v(v_t)\, dw_t^v \tag{12}$$

and, for simplicity, $\{w_t^S\}$ and $\{w_t^v\}$ are mutually independent standard Brownian motion.[4] This is an extension of the classic Black-Scholes model where $\{v_t\}$ can be interpreted as a stochastic volatility term. In this setting, (9) is still valid but now

$$dS_t = rS_t dt + \sigma_S(S_t, v_t)\, dW_t^S,$$

---

[3] Observe that b is not differentiable here. We conjecture that the results still hold for this case however.

[4] We can also allow for S_t entering the SDE for v_t, and also that v_t enters the drift function μ_S.

$$dv_t = \{\mu_v(v_t) - \lambda(S_t, v_t)\} dt + \sigma_v(v_t) dW_t^v,$$

under Q. Observe that the drift term of $\{v_t\}$ under Q includes the term $\lambda(t, S_t, v_t)$ which can be interpreted as the market price for volatility risk. A simple specification of (11)-(12) is found in Heston (1993) where

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dw_t^S, \tag{13}$$

$$dv_t = \beta(\alpha - v_t) dt + \sigma\sqrt{v_t} dw_t^v, \tag{14}$$

and $\lambda(S, v) = \lambda v$. In this case, the PDE can be solved explicitly; this is not possible in the general case though. To see that this model also can be accommodated for in our framework, define $X_t = (S_t, v_t)^\top$. This process then solves the SDE (3) with $\mu(x) = (rS, \mu_v(v) - \lambda(S, v))^\top$, $\sigma(x) = \text{diag}(\sigma_S(S, v), \sigma_v(v))$, and $W_t = (W_t^S, W_t^v)^\top$.

Example 2: Factor Models for the Term Structure. We assume that the short-term interest rate process, $\{r_t\}$, is a Markov process solving

$$dr_t = \mu(t, r_t) dt + \sigma(t, r_t) dw_t \tag{15}$$

under P. We are then able to derive the term structure of bonds. Following for example Björk (1998, Chapter 16), one may show that

$$dr_t = \{\mu(t, r_t) - \lambda(t, r_t) \sigma(t, r_t)\} dt + \sigma(t, r_t) dW_t \tag{16}$$

under Q for some process $\{\lambda_t\}$ which is often termed "the market price for risk". Now consider an interest rate derivative with associated dividend stream $d_t = d(t, r_t)$ and terminal pay-off $g(r_T)$. The price at any time t is then given as

$$\Pi_t(T) = E_t^Q \left[\exp\left[-\int_t^T r_s ds \right] g(r_T) + \int_t^T \exp\left[-\int_t^s r_s du \right] d(s, r_s) ds \right]. \tag{17}$$

A leading example of an interest rate derivative is a zero-coupon bond, characterised by $b(r_T) = 1$ and $d(t, r) = 0$.

The above model is a special case of the general multifactor models where the yield curve is driven by multiple factors. That is, the interest rate is given by $r_t = R(F_t)$ for some twice differentiable function $R : \mathbb{R}^q \mapsto \mathbb{R}$, and some q-dimensional diffusion process $\{F_t\}$. By Itô's Lemma, we then obtain that $\{r_t\}$ is also a diffusion process and the formula in (17) remains valid. Observe that the short term model above is a single-factor model ($q = 1$, $R(x) = x$ and $F_t = r_t$).

A class of factor models which has received particular attention is the affine one. In this setting the functions $F(x)$, $\mu(x)$, and $\sigma(x)\sigma(x)^\top$ all are assumed to be affine in x. These restrictions highly facilitates the analysis since it is possible to derive explicit expressions of bond prices. See for example Duffie and Kan (1996) and Duffee (2002).

Once the zero-coupon bond prices have been recovered, one can start to price coupon-bearing bonds, bond options and other derivatives with a bond as the underlying variable, e.g. yield options, swaps, caps, floors and futures. See Hull (1997, Chapter 16) for an introduction to these. Bond and interest rate derivative prices for any factor model can be put on the form of (4): Define $X_t = F_t$,

$a(t, x) = R(x)$, $b(x) = g(R(x))$, $c(t, x) = d(t, F(x))$; we then easily see that (17) takes the desired form.

Example 3: The Heath-Jarrow-Morton Model. In the Heath-Jarrow-Morton (1992) framework, the forward rate structure is modelled instead of the short rate. Let $f_t(T)$ denote the instantaneous forward rate with maturity T contracted at time t. This is defined as

$$f_t(T) = \frac{\partial \log B_t(T)}{\partial T}$$

where $B_t(T)$ is the price of a zero-coupon bond with maturity date T. One can reversely write $B_t(T) = \exp\left[-\int_t^T f_t(s)\, ds\right]$. In particular, the short rate satisfies $r_t = f_t(t)$. We assume the following dynamics of $f_t(T)$ under Q,

$$df_t(T) = \mu_t(T)\, dt + \sigma_t(T)\, dW_t,$$

where $\{W_t\}$ is a q-dimensional Brownian motion, while $\{\mu_t(T)\}$ and $\{\sigma_t(T)\}$ are adapted stochastic processes. The assumption of no-arbitrage implies that

$$\mu_t(T) = \sigma_t(T) \int_t^T \sigma_t(s)^\top\, ds,$$

c.f. Björk (1998, p. 269). Furthermore, the bond prices have the following dynamics,

$$dB_t(T) = r_t B_t(T)\, dt + \sigma_t^*(T) B_t(T)\, dW_t,$$

where $\sigma_t^*(T) = \int_t^T \sigma_s(T)\, ds$. Assuming that $\mu_t(T) = \mu(t, X_t; T)$ and $\sigma_t(T) = \sigma(t, X_t; T)$ for some finite-dimensional vector of state-variables X_t, the above pricing formula takes the form of (4).

2.2 Estimation of Diffusion Models

The type of partial differential equations in consideration here also appear in other areas. In the following, we give a brief discussion of their applications in the estimation of diffusions. The literature on the estimation of diffusion models is very large and still growing. One particular branch of this literature is concerned with estimation given discrete observations of the process, e.g. daily, weekly or monthly observations. This is the most realistic setting but also the least tractable; in particular the natural estimator, the MLE, proves to be difficult to implement. A large number of alternative estimators have been proposed as a result. But the asymptotic properties of these have either only been conjectured at or derived under high-level conditions. The results derived in the next section enable us to validate these high-level conditions. In the following, we shall present a number of estimation methods and discuss what is needed for the estimator to have the desired asymptotic properties. We shall only discuss these issues in a parametric framework, but it should be clear that our main results also are applicable in a non- and semiparametric setting.

We assume that we have discrete observations from the following SDE,

$$dX_t = \mu(X_t; \theta)\, dt + \sigma(X_t; \theta)\, dW_t, \tag{18}$$

for some unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^d$. In the following we discuss the estimation of θ.

Example 4: Estimation via Conditional Means. Since in many cases the transition density is of unknown form, the model is often estimated using estimating equations. In particular, one often use regression models of the form

$$b\left(X_{i\Delta}\right) = B\left(X_{(i-1)\Delta}; \theta\right) + \varepsilon_i$$

where $B\left(x; \theta\right) = E_\theta\left[b\left(X_\Delta\right) | X_0 = x\right]$ is the conditional mean where we write $E_\theta\left[\cdot\right]$ to indicate the dependence of the conditional mean on θ. Using this type of equations leads to GMM-type estimators as considered in, amongst others, Bibby and Sørensen (1995), Chacko & Viceira (2003), Duffie and Singleton (1993), Carrasco, Chernov, Florens & Ghysels (2002), Singleton (2001), Sørensen (1997). In order to derive the asymptotics of this type of estimators, we need to show that $B\left(x; \theta\right)$ is smooth and differentiable in θ. However, as noted earlier, an analytical expression of $B\left(x; \theta\right)$ often cannot be derived and is calculated using either simulations or approximate methods. One easily realise that $B\left(x; \theta\right) = u\left(0, x; \theta\right)$ where u solves the LPDE

$$-\frac{\partial u}{\partial t} = L_t\left(u; \theta\right), \quad u\left(\Delta, x; \theta\right) = b\left(x\right).$$

One example is the estimator proposed in Bibby and Sørensen (1995). We define the so-called estimating function,

$$G_n\left(\theta\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(X_{i\Delta} | X_{(i-1)\Delta}; \theta\right), \quad g\left(y|x; \theta\right) = \alpha\left(x; \theta\right)^\top \left\{b\left(y\right) - B\left(x; \theta\right)\right\}$$

where $b: \mathbb{R}^q \mapsto \mathbb{R}^m$ and B are given above and $\alpha: \mathbb{R}^q \times \Theta \mapsto \mathbb{R}^{m \times d}$ is a weighting function. The estimator is then chosen as the root, $G_n(\hat{\theta}) = 0$. An obvious choice is $b_1\left(x\right) = x$ and $b_2\left(x\right) = x^2$.

Example 5: Estimation via Observed Option Prices. Another application is in the estimation of the parameter θ using observed derivative prices. We here present the estimation method using the extended Black-Scholes model in Example 1.1 with constant interest rates, $r_t = r > 0$, but the idea can easily be adapted to other, more general models. For simplicity, we assume that we have observed over time a series of prices for a specific option with pay-off function g and fixed time to maturity $T > 0$. So no cross-sectional dimension is included. Let $\{P_i\}$ denote the observed option prices and $\{X_i\}$ the observed stock price. Assuming that the option prices have been observed with errors (due to market imperfections, observation errors etc.), we have the following regression model,

$$P_i = \Pi\left(X_i; \theta\right) + \varepsilon_i, \quad \Pi\left(x; \theta\right) = e^{-rT} E_\theta^Q\left[b\left(X_T\right) | X_0 = x\right].$$

The parameter θ may then be estimated by e.g. nonlinear least squares (assuming it is identified). Again, for the estimator to be consistent and asymptotically normal distributed one normally has to check that $\theta \mapsto \Pi\left(x; \theta\right)$ is continuous and differentiable. In the next section, we give regularity conditions which ensures this.

3 Estimation of Partial Differential Equations

In this section, we shall assume that preliminary estimators $\left(\hat{\mu}, \hat{\sigma}^2\right)$ are available, and then give conditions for the associated solution $\hat{u}$ to be consistent and asymptotically normal distributed.

We introduce the operator $\Gamma : \mathcal{D} \mapsto \mathcal{U}$ defined by

$$u(t,x) = \Gamma\left(\mu, \sigma^2\right)(t,x),$$

where u is the solution to (1)-(2) with $\left(\mu, \sigma^2\right)$ plugged in. We assume that we have obtained estimators, $\left(\hat{\mu}, \hat{\sigma}^2\right)$, of the true drift and diffusion term, $\left(\mu_0, \sigma_0^2\right)$. Given the definition of Γ, the true but unknown solution to the PDE is given by

$$u_0 = \Gamma\left(\mu_0, \sigma_0^2\right)$$

which we then estimate by

$$\hat{u} = \Gamma\left(\hat{\mu}, \hat{\sigma}^2\right).$$

By an extension of Slutsky's theorem from the Euclidean case to function spaces, the asymptotic properties of $\hat{u}$ will then follow from the ones of $\left(\hat{\mu}, \hat{\sigma}^2\right)$ given that Γ is sufficiently smooth. Roughly speaking, $\hat{u}$ will be consistent if $\left(\hat{\mu}, \hat{\sigma}^2\right)$ is so and Γ is continuous, while the asymptotic distribution will be induced by the one of $\left(\hat{\mu}, \hat{\sigma}^2\right)$ given that Γ is (pathwise) differentiable w.r.t. $\left(\mu, \sigma^2\right)$. To extend Slutsky's Theorem to hold on function spaces we need to ensure that $\mathcal{D}$ and $\mathcal{U}$ can be equipped with suitable norms. For now assume this is the case and let $\|\cdot\|_{\mathcal{D}}$ and $\|\cdot\|_{\mathcal{U}}$ denote the norms on $\mathcal{D}$ and $\mathcal{U}$ respectively. We then assume that our preliminary estimators satisfy $\left(\hat{\mu}, \hat{\sigma}^2\right) \in \mathcal{D}$ with $\left\|\left(\hat{\mu}, \hat{\sigma}^2\right) - \left(\mu_0, \sigma_0^2\right)\right\|_{\mathcal{D}} \rightarrow^P 0$. Consistency of $\hat{u} = \Gamma\left(\hat{\mu}, \hat{\sigma}^2\right)$ will now follow by continuity of Γ since this implies $\|\hat{u} - u_0\|_{\mathcal{U}} = \left\|\Gamma\left(\hat{\mu}, \hat{\sigma}^2\right) - \Gamma\left(\mu_0, \sigma_0^2\right)\right\|_{\mathcal{U}} \rightarrow^P 0$. Assume that the pathwise derivative of Γ w.r.t. μ and σ^2 at $\left(\mu_0, \sigma_0^2\right)$ exists. We denote these $\nabla_1 \Gamma[d\mu]$ and $\nabla_2 \Gamma\left[d\sigma^2\right]$ respectively and define $\nabla \Gamma\left[d\mu, d\sigma^2\right] = \nabla_1 \Gamma[d\mu] + \nabla_2 \Gamma\left[d\sigma^2\right]$. Assuming that

$$\left\|\Gamma\left(\hat{\mu}, \hat{\sigma}^2\right) - \Gamma\left(\mu_0, \sigma_0^2\right) - \nabla\Gamma\left[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2\right]\right\|_{\mathcal{U}} \leq C\left(\|\hat{\mu} - \mu_0\|_{\mathcal{D}}^2 + \left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{\mathcal{D}}^2\right),$$

$\nabla\Gamma$ will drive the asymptotic distribution under suitable conditions.

The approach outlined above has been widely used in the literature when working with functionals of nonparametric estimators. General result concerning the asymptotics of Γ when the preliminary estimator is a kernel estimator can be found in Aït-Sahalia (1993). Examples of applications of this approach to specific estimation problems can be found in Aït-Sahalia (1996a), Hausman and Newey (1995), Jiang (1998) and Vanhems (2003).

All subsequent results will be derived under the following additional condition which implicitly will be assumed throughout the remains of the chapter together with Condition 1:

Condition 2 $\left(\mu_0, \sigma_0^2\right) \in \mathcal{D}$

We first show that the functional $\Gamma : \mathcal{D} \mapsto \mathcal{U}$ is continuous.

Theorem 2 *For any* $\left(\mu, \sigma^2\right) \in \mathcal{D}$,

$$\left|\Gamma\left(\mu, \sigma^2\right)(t,x) - \Gamma\left(\mu_0, \sigma_0^2\right)(t,x)\right| \leq CT\left(1 + \|x\|^q\right)\left\{\|\hat{\mu} - \mu_0\|_{0,4} + \left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{0,4}\right\}$$

for $X^* = x$. *In particular,*

$$\|\hat{u} - u_0\|_{0,1} \leq CT\left(1 + E\left[\|X^*\|^r\right]\right)\left\{\|\hat{\mu} - \mu_0\|_{0,4} + \left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{0,4}\right\},$$

for $r \leq p^*$.

13

We now derive an expression for the pathwise derivative of u w.r.t. (μ, σ^2) at (μ_0, σ_0^2) in the direction $(d\mu, d\sigma^2) = (\mu - \mu_0, \sigma^2 - \sigma_0^2)$. Let $\{\nabla_1 X_t\}$ and $\{\nabla_2 X_t\}$ be given as in in (53) and (54) respectively with $\nabla_i X_0 = 0$, $i = 1, 2$. From Lemma 12, $\{\nabla_1 X_t\}$ and $\{\nabla_2 X_t\}$ are the pathwise derivatives w.r.t. μ and σ^2 respectively in the L_2-sense. The pathwise derivative of X_t at (μ_0, σ_0^2) in the direction $(d\mu, d\sigma^2)$ in the L_2-sense is then given by

$$\nabla X_t = \nabla_1 X_t + \nabla_2 X_t.$$

The chain rule now implies that the pathwise derivative of Γ at (μ_0, σ_0^2) in the direction $(d\mu, d\sigma^2)$ is given by

$$\nabla \Gamma \left[d\mu, d\sigma^2\right] (t, x) \tag{19}$$

$$= E_{t,x} \left[b_x \left(X_T^0 \right) \nabla X_T \exp \left[- \int_t^T a \left(u, X_u^0 \right) du \right] \right] \tag{20}$$

$$- E_{t,x} \left[b \left(X_T^0 \right) \int_t^T a_x \left(u, X_s^0 \right) \nabla X_s ds \exp \left[- \int_t^T a \left(s, X_s^0 \right) ds \right] \right]$$

$$+ E_{t,x} \left[\int_t^T c_x \left(s, X_s^0 \right) \nabla X_s \exp \left[- \int_t^s a \left(u, X_u^0 \right) du \right] ds \right]$$

$$- E_{t,x} \left[\int_t^T c \left(s, X_s^0 \right) \exp \left[- \int_t^s a \left(u, X_u^0 \right) du \right] \left(\int_t^s a_x \left(u, X_u^0 \right) \nabla X_u du \right) ds \right].$$

By construction, $\{\nabla X_t\}$ and thereby $\nabla \Gamma$ is linear in $(d\mu, d\sigma^2)$. An alternative representation of $\nabla \Gamma$ is as the solution, v, to the following LPDE,

$$-\frac{\partial v}{\partial t} + av = L_t (v) + \bar{c} \left[d\mu, d\sigma^2\right] \tag{21}$$

$$v(T, x) = 0 \tag{22}$$

where

$$\bar{c} \left[d\mu, d\sigma^2\right] = \sum_{i=1}^q d\mu_i \frac{\partial u_0}{\partial x_i} + \frac{1}{2} \sum_{i=1}^q d\sigma_{ij}^2 \frac{\partial^2 u_0}{\partial x_i \partial x_j},$$

and $u_0 = \Gamma \left(\mu_0, \sigma_0^2 \right)$. The generalised solution of (21)-(22) is given as

$$\nabla \Gamma \left[d\mu, d\sigma^2\right] (t, x) = E_{t,x} \left[\int_t^T \bar{c} \left[d\mu, d\sigma^2\right] \left(s, X_s^0 \right) \exp \left[- \int_t^s a \left(u, X_u^0 \right) du \right] ds \right] \tag{23}$$

The following theorem shows that $\nabla \Gamma$ also has the desired properties discussed earlier:

Theorem 3 *For any* $(\mu, \sigma^2) \in \mathcal{D}$, $\nabla \Gamma \left[\mu - \mu_0, \sigma^2 - \sigma_0^2 \right]$ *is well-defined and satisfies*

$$\left| \hat{u} (t, x) - u_0 (t, x) - \nabla \Gamma \left[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2 \right] (t, x) \right| \leq b(x, T) \left\| \hat{\mu} - \mu_0 \right\|_{1,4}^2 + \left\| \hat{\sigma}^2 - \sigma_0^2 \right\|_{1,4}^2$$

and

$$\left| \nabla \Gamma \left[\mu, \sigma^2 \right] (t, x) \right| \leq b(x, T) \left(\left\| \mu \right\|_{0,2}^2 + \left\| \sigma^2 \right\|_{0,2}^2 \right),$$

with $X^* = x$, *where* $b(x, T) = CTe^{CT} \left[1 + \|x\|^{2q} \right]$.

Having obtained these two basic results, we are now ready to discuss the asymptotics of $\hat{u}$. As a first step, we obtain from Theorem 2 that $\hat{u}$ is consistent in the $\|\cdot\|_{0,2}$-norm if $\hat{\mu}$ and $\hat{\sigma}^2$ are in the $\|\cdot\|_{0,4}$-norm. This also gives a first lower bound on the convergence rate of $\hat{u}$. From the second of the two above theorems, we have that the pointwise convergence rate of $\hat{u}$ is determined by those of $\hat{\mu}$ and $\hat{\sigma}^2$ in the squared $\|\cdot\|_{1,4}$-norm together with the behaviour of the pathwise derivative $\nabla\Gamma$. If $\hat{\mu}$ and $\hat{\sigma}^2$ are sufficiently well-behaved, the asymptotic distribution of $\hat{u}$ will be determined by $\nabla\Gamma\left[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2\right]$. The following theorem states high-level conditions under which $\hat{u}(t, x)$ is asymptotically normally distributed.

Theorem 4 (Master Theorem) *Assume that $\left(\hat{\mu}, \hat{\sigma}^2\right) \in \mathcal{D}$, and $\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1)$ and $\left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{0,4} = o_P(1)$. Then $\|\hat{u}(t, x) - u_0(t, x)\|_{0,1} \to^P 0$ for any X^* with $E\left[\|X^*\|^r\right] < \infty$.*

If furthermore there exists a (possible random) sequence $\{A_n\}$ such that

1. *$\|\hat{\mu} - \mu_0\|_{1,4} = o_P\left(A_n^{-1/2}\right)$ and $\left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{1,4} = o_P\left(A_n^{-1/2}\right)$;*

2. *$A_n \nabla\Gamma\left[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2\right] \xrightarrow{d} N(0, V(t, x))$;*

Then,

$$A_n\left(\hat{u}(t, x) - u_0(t, x)\right) \xrightarrow{d} N(0, V(t, x)).$$

This theorem is very general, and not very useful per se. In order to apply it on specific estimators, one has to verify that 1. and 2. are satisfied. The first condition is normally fairly easy to check since this is merely a question of $\hat{\mu}$ and $\hat{\sigma}^2$ converging sufficiently fast in the norm $\|\cdot\|_{0,2}$. The verification of the second condition on the other hand requires more work since the precise form of $\nabla\Gamma$ is complicated. In the parametric case, it proves to be easy to check the second condition given sufficient smoothness conditions on μ and σ^2, and we are able to give an explicit expression of the variance term, c.f. Theorem 5. In the non- and semiparametric case, the following trick will be used: We observe that $\mathcal{D}$ is a linear subspace of $\mathcal{H} \equiv W^{0,2} \times W^{0,2}$ and that $\mathcal{H}$ is a Hilbert space equipped with the inner product $\langle\cdot, \cdot\rangle = \langle\cdot, \cdot\rangle_0$ as defined in (7). So the completion of $\mathcal{D}$, $\bar{\mathcal{D}}$, can be considered as a Hilbert space in its own right. Furthermore, $\nabla\Gamma$ is a continuous, linear operator on $\bar{\mathcal{D}}$, c.f. Theorem 3. We then apply Riesz Representation Theorem on $\nabla\Gamma$: There exists $d^* = (d_1^*, d_2^*) \in \bar{\mathcal{D}}$ such that

$$\nabla\Gamma\left[\mu, \sigma^2\right](t, x) = \langle\mu, d_1^*\rangle + \langle\sigma^2, d_2^*\rangle, \tag{24}$$

where $\langle\cdot, \cdot\rangle$ is given by

$$\langle f, g\rangle = E_{t,x}\left[\int_t^T f\left(u, X_u^0\right) g\left(u, X_u^0\right) du\right] = \int_{\mathbb{R}^d}\int_t^T p(u, y | t, x) f(u, y) g(u, y) \, du \, dy$$

and $p(u, y | t, x)$ denotes the conditional density of X_u^0 conditional on $X_t^0 = x$. This representation of $\nabla\Gamma$ is much easier to work with, and one can normally verify that each of the integrals converges in distribution when one plugs in $\hat{\mu}$ and $\hat{\sigma}^2$. In the case where $\mu(t, x) = \mu(x)$ and $\sigma^2(t, x) = \sigma^2(x)$, we can use the following, more simple inner product,

$$\langle f, g\rangle = \int_{\mathbb{R}^d} q_{T-t}(y | x) f(y) g(y) \, dy,$$

where $q_t(y|x) = \int_0^t p_u(y|x)\,du$ and $p_t(y|x) = p(u, y|u - t, x)$ is the homogeneous transition density.

Unfortunately, this approach does not supply us with the precise form of the asymptotic variance since the Riesz Representation Theorem does not tell us the precise form of $d^* \in \bar{\mathcal{D}}$ - only that such exists. A special case where the explicit form of d^* can be derived is when $a(t, x) = a$ is constant. Under this assumption, we obtain from (23) that

$$
\begin{aligned}
\nabla\Gamma\left[d\mu, d\sigma^2\right](t, x) &= e^{at} \sum_{i=1}^q E_{t,x}\left[\int_t^T d\mu_i\left(s, X_s^0\right) d_{1,i}^*\left(s, X_s^0\right) ds\right] \\
&\quad + \frac{1}{2} e^{at} \sum_{i,j=1}^q E_{t,x}\left[\int_t^T d\sigma_{ij}^2\left(s, X_s^0\right) d_{2,ij}^*\left(s, X_s^0\right) ds\right]
\end{aligned}
\tag{25}
$$

where

$$
d_{1,i}^*(t, x) = \frac{\partial u_0(t, x)}{\partial x_i} e^{-at}, \quad d_{2,ij}^*(t, x) = \frac{\partial^2 u_0(t, x)}{\partial x_i \partial x_j} e^{-at}.
$$

But even if the precise form of the variance is unknown, we shall demonstrate that it is possible to construct an estimator of it. Otherwise, one can apply bootstrap methods to estimate the distribution. The latter has the advantage of giving a better approximation of the finite-sample distribution, c.f. Hall (1992).

We shall now apply the above Master Theorem on three specific estimators of μ and σ^2, and derive the asymptotic properties of the associated estimated solution for each of these. In all three cases, the estimated solution will be $\sqrt{n}$-consistent, despite the fact that the preliminary estimators may have slower than $\sqrt{n}$-convergence rate. This is a well-known result from nonparametric estimation theory. While differentiation makes a problem more ill-posed/less regular, integration works as a regularization of the problem. The increased regularity of the problem in turn increases the convergence rate. A simple example of this is nonparametric density estimation: The optimal rate of convergence in the minimax sense of the nonparametric density estimator is $n^{2/(q+4)}$, while the optimal rate of the cumulative density estimator is $\sqrt{n}$.

The $\sqrt{n}$-convergence rate of estimators of solutions to a class of ordinary differential equations was established by Hausman and Newey (1995) and Vanhems (2003), and similar results were obtained for solutions to LPDE's for specific kernel estimators, c.f. Aït-Sahalia (1996a) and Jiang (1998). The result stated in Theorem 4 confirms this: For $\hat{u}(t, x)$ to be asymptotically normally distributed, we require that the preliminary estimators converge with $n^{1/4}$-rate, while $\nabla\Gamma\left[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2\right]$ converges with $\sqrt{n}$-rate. The latter will hold in great generality.

The three estimators we shall consider are all based on discrete observations of the underlying diffusion process with drift term μ_0 and diffusion term σ_0. In the following, we shall denote the sampled process by $\{x_t\}$, and the driving Brownian motion by $\{w_t\}$ such that

$$
dx_t = \mu_0(t, x_t)\,dt + \sigma_0(t, x_t)\,dw_t.
\tag{26}
$$

This is done in order not to confuse the sampled process with $\{X_t\}$ entering the expression of the generalised solution. We may and will choose the probability measures Q and P which $\{X_t\}$ and $\{x_t\}$ respectively operates under to be mutually independent.

3.1 A Parametric Estimator

We assume that $\mu_0(t,x) = \mu(t,x;\theta_0)$ and $\sigma_0^2(t,x) = \sigma^2(t,x;\theta_0)$ for some known parameterisation where $\theta_0 \in \Theta \subseteq \mathbb{R}^d$ is the true, unknown parameter, and that a preliminary estimator $\hat{\theta}$ is available. The estimator $\hat{\theta}$ could arrive from various estimation methods, the leading example being that it is based on discrete observations, $\{x_{i\Delta}\}$, of the process $\{x_t\}$. In this setting θ can be estimated by for example MLE (Pedersen 1995, Elerian et al 2001, Aït-Sahalia 2002) or GMM (Bibby and Sørensen 1995, Duffie and Singleton 1993). We do not have to restrict the observed process to be stationary; it may potentially be non-stationary and the estimator converging with a random convergence rate.

We then wish to derive the asymptotic properties of $\hat{u}$ associated with $\hat{\mu}(t,x) = \mu(t,x;\hat{\theta})$ and $\hat{\sigma}^2(t,x) = \sigma^2(t,x;\hat{\theta})$. This will be done under the following set of regularity conditions:

P.1 For any $\theta \in \Theta$: $\left(\mu(\cdot,\cdot;\theta), \sigma^2(\cdot,\cdot;\theta) \right) \in \mathcal{D}$.

P.2 $\partial_x^i \mu(t,x;\theta)$ and $\partial_x^i \sigma^2(t,x;\theta)$ are continuously differentiable w.r.t. θ such that

$$\left\| \partial_x^i \dot{\mu}(t,x;\theta) \right\| \leq C(1+\|x\|), \quad \partial_x^i \dot{\sigma}^2(t,x;\theta) \leq C(1+\|x\|),$$

for $i = 0, 1$.

P.3 $\partial_x \dot{\mu}(t,x;\theta)$ and $\partial_x \dot{\sigma}^2(t,x;\theta)$ are bounded.

P.4 The preliminary estimator $\hat{\theta}$ satisfies $V_n^{-1/2}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0,I)$ where $\theta_0 \in \text{int}\Theta$ and $\{V_n\}$ is a (possibly random) matrix-sequence which is positive definite and $\|V_n\| \to 0$ P-a.s.

We apply standard Taylor-expansions to obtain the desired result. First, it holds that $\hat{\mu}(t,x) = \mu(t,x;\hat{\theta})$ satisfies

$$E_{t,x}\left[\int_t^T \left[\left\| \partial_x^i \hat{\mu}(u,X_u^0) - \partial_x^i \mu_0(u,X_u^0) \right\|^4 \right] du \right]$$
$$= E_{t,x}\left[\int_t^T \left[\left\| \partial_x^i \dot{\mu}(u,X_u^0;\bar{\theta})(\hat{\theta} - \theta_0) \right\|^4 \right] du \right]$$
$$\leq C\left(1 + \|x\|^4\right) \|\hat{\theta} - \theta_0\|^4,$$

for $i = 0, 1$, and similarly that $\hat{\sigma}^2(t,x) = \sigma^2(t,x;\hat{\theta})$ satisfies

$$E_{t,x}\left[\int_t^T \left[\left\| \partial_x^i \hat{\sigma}^2(u,X_u^0) - \partial_x^i \sigma_0^2(u,X_u^0) \right\|^4 \right] du \right] \leq C\left(1 + \|x\|^4\right) \|\hat{\theta} - \theta_0\|^4,$$

for $i = 0, 1$. Thus, by Theorem 4, for any $0 \leq t \leq T$ and $x \in \mathbb{R}^q$,

$$|\hat{u}(t,x) - u_0(t,x) - \nabla\Gamma(t,x)| \leq C\left(1 + \|x\|^4\right) \|\hat{\theta} - \theta_0\|^2 = O_P(\|V_n\|)$$

The asymptotic distribution is then determined by $\nabla\Gamma(t,x)$ which in the parametric setting takes a

fairly simple form. We define

$$
\begin{aligned}
\dot{\Gamma}_0(t,x) &= E_{t,x}\left[b_x\left(X_T^0\right)\dot{X}_T^0\exp\left[-\int_t^T a\left(u,X_u^0\right)du\right]\right] \\
&\quad -E_{t,x}\left[b\left(X_T^0\right)\exp\left[-\int_t^T a\left(s,X_s^0\right)ds\right]\left(\int_t^T a_x\left(u,X_s^0\right)\nabla X_s^0 ds\right)\right] \\
&\quad +E_{t,x}\left[\int_t^T c_x\left(s,X_s^0\right)\dot{X}_s^0\exp\left[-\int_t^s a\left(u,X_u^0\right)du\right]ds\right] \\
&\quad -E_{t,x}\left[\int_t^T c\left(s,X_s^0\right)\exp\left[-\int_t^s a\left(u,X_u^0\right)du\right]\left(\int_t^s a_x\left(u,X_u^0\right)\dot{X}_u du\right)ds\right].
\end{aligned}
\tag{27}
$$

where $\{\dot{X}_t^0\}$ is the solution to the SDE

$$
\dot{X}_t^0 = \left\{\dot{\mu}\left(t,X_t^0;\theta\right)+\mu^{(1)}\left(t,X_t^0;\theta\right)\dot{X}_t^\theta\right\}dt+\left\{\dot{\sigma}\left(t,X_t^0;\theta\right)+\sigma^{(1)}\left(t,X_t^0;\theta\right)\dot{X}_t^\theta\right\}dW_t,
$$

with $\dot{X}_0^0 = 0$. It is then easily shown, using the same arguments as in the proof of Lemma 12, that

$$
E_{x,s}\left[||\nabla X_t\left[\hat{\mu}-\mu_0,\hat{\sigma}^2-\sigma_0^2\right]-\dot{X}_t^0(\hat{\theta}-\theta_0)||\right]\leq C\left(x\right)||\hat{\theta}-\theta_0||^2
$$

implying

$$
\nabla\Gamma\left[\hat{\mu}-\mu_0,\hat{\sigma}^2-\sigma_0^2\right](t,x)=\dot{\Gamma}_0\left(t,x\right)\left(\hat{\theta}-\theta_0\right)+O_P\left(||V_n||\right)
$$

We have now proved the following theorem:

Theorem 5 (Parametric Estimator) *Under (P.1)-(P.4), the parametric estimator $\hat{u}$ is consistent and satisfies*

$$
\hat{u}\left(t,x\right)-u_0\left(t,x\right)\xrightarrow{d}N\left(0,\dot{\Gamma}_0\left(t,x\right)^\top V_n\dot{\Gamma}_0\left(t,x\right)\right),
$$

where $\dot{\Gamma}_0\left(t,x\right)$ is given in (27).

So in this setting a closed form expression of the asymptotic variance is available.

Remark. An alternative characterization of $\dot{\Gamma}_0$ is as solution, v, to the LPDE given in (21)-(22) with $\bar{c}$ given by

$$
\bar{c}=\sum_{i=1}^q\dot{\mu}_i\frac{\partial u_0}{\partial x_i}+\frac{1}{2}\sum_{i=1}^q\dot{\sigma}_{ij}^2\frac{\partial^2 u_0}{\partial x_i\partial x_j}.
\tag{28}
$$

It readily follows from Lemma 9, that a consistent estimator of $\dot{\Gamma}_0\left(t,x\right)$ is obtained by substituting $(X^0,\dot{X}^0)$ by $(\hat{X},\widehat{\dot{X}})$ in (27), where the latter solves the SDE-system associated with the estimated drift and diffusion term,

$$
\begin{aligned}
d\hat{X}_t &= \mu(t,\hat{X}_t;\hat{\theta})dt+\sigma(t,\hat{X}_t;\hat{\theta})dW_t, \\
d\widehat{\dot{X}}_t &= \{\dot{\mu}(t,\hat{X}_t;\hat{\theta})+\mu^{(1)}(t,\hat{X}_t;\hat{\theta})\widehat{\dot{X}}_t\}dt+\{\dot{\sigma}(\hat{X}_t;\hat{\theta})+\sigma^{(1)}(\hat{X}_t;\hat{\theta})\widehat{\dot{X}}_t\}dW_t,
\end{aligned}
$$

where $\hat{X}_t = x$, and $\widehat{\dot{X}}_0 = 0$. Alternatively, one can obtain an estimator of $\dot{\Gamma}_0$ by solving (21)-(22) with $\bar{c}$ given in (28)and with u_0, $\dot{\mu}$ and $\dot{\sigma}^2$ substituted for $\hat{u}$, $\partial_\theta\hat{\mu}$ and $\partial_\theta\hat{\sigma}^2$ respectively.

3.2 A Semiparametric Estimator

In this section we consider the case where semiparametric estimators of the drift and diffusion term are available. We introduce the following two classes of scalar ($q = 1$) diffusion models:

Class 1 $dx_t = \mu(x_t)\, dt + \sigma(x_t; \theta)\, dw_t$ where $\theta \in \Theta \subseteq \mathbb{R}^d$ and $\mu(\cdot)$ is unspecified.

Class 2 $dx_t = \mu(x_t; \theta)\, dt + \sigma(x_t)\, dw_t$ where $\theta \in \Theta \subseteq \mathbb{R}^d$ and $\sigma^2(\cdot)$ is unspecified.

Observe that the SDE's in both classes are assumed to be time-homogenous in which case the transition density satisfies $p(t, y|s, x) = p_{t-s}(y|x)$.

We assume that an estimator $\hat{\theta}$ is available satisfying

$$\hat{\theta} = \theta_0 + \frac{1}{n} \sum_{i=1}^{n} \psi(x_i; x_{i-1}) + o_P(n^{-1/2}), \tag{29}$$

for some influence function ψ with $E[\psi(x_i; x_{i-1})] = 0$ and $E[\|\psi(x_i; x_{i-1})\|^{2+\delta}] < \infty$ for some $\delta > 0$. One such estimator was derived in Kristensen (2004a). Assuming stationarity of $\{x_t\}$, there exists a stationary density π_0 satisfying

$$\pi_0(x) = \frac{M}{\sigma^2(x)} \exp\left[2 \int_{x^*}^{x} \frac{\mu_0(y)}{\sigma_0^2(y)} dy \right], \tag{30}$$

for some $x^* \in I$ and a normalising factor $M > 0$, see for example Karlin and Taylor (1981, Section 15.6). It is possible to revert (30) in either of the two following ways,

$$\mu_0(x) = \frac{1}{2\pi_0(x)} \frac{\partial}{\partial x} \left[\sigma_0^2(x) \pi_0(x) \right], \tag{31}$$

$$\sigma_0^2(x) = \frac{2}{\pi_0(x)} \int_l^x \mu_0(y) \pi_0(y)\, dy. \tag{32}$$

We estimate $\pi_0^{(m)}$ by the kernel estimator $\hat{\pi}^{(m)}$ given by

$$\hat{\pi}^{(m)}(x) = \frac{1}{nh^{m+1}} \sum_{i=1}^{n} K^{(m)}\left(\frac{x - x_i}{h} \right), \quad r \geq 0, \tag{33}$$

where for a kernel K and a bandwidth h; see Silverman (1986) for an introduction to these concepts. Given $\hat{\theta}$ and $\hat{\pi}^{(m)}$, $m = 0, 1$, we may then estimate the drift and diffusion term in the following manner. For a model in Class 1, we estimate $\mu_0(x)$ by $\hat{\mu}(x; \hat{\theta})$ where

$$\hat{\mu}(x; \theta) = \frac{1}{2} \partial_x \sigma^2(x; \theta) + \frac{1}{2} \sigma^2(x; \theta) \frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)}, \tag{34}$$

and $\sigma_0^2(x)$ by $\sigma^2(x; \hat{\theta})$. For any model in Class 2, we estimate $\mu_0(x)$ by $\mu(x; \hat{\theta})$ and $\sigma_0^2(x)$ by $\hat{\sigma}^2(x; \hat{\theta})$ where

$$\hat{\sigma}^2(x; \theta) = \frac{1}{2\hat{\pi}(x)} \frac{1}{n} \sum_{i=1}^{n} 1_{(l,x)}(x_i)\, \mu(x; \theta). \tag{35}$$

See Chapter 4 for more details on these estimators.

In Class 1, given consistency of $\hat{\theta}$, $\sigma^2(x; \hat{\theta})$ is a pointwise consistent estimator of $\sigma^2(x)$ given smoothness conditions of $\sigma^2(x; \theta)$ w.r.t. θ. This in turn yields consistency of $\hat{\mu}(x; \hat{\theta})$ in (34) by the delta method. Similarly for Class 2. We note that in both cases the convergence rate of the nonparametric part is slower than $\sqrt{n}$.

In the following, we will derive the asymptotics of $\hat{u}$ in each of the two classes. In order to do this we need to establish consistency of the two nonparametric estimators in the function norms $\|\cdot\|_{0,4}$ and $\|\cdot\|_{1,4}$. For this to hold, we need to introduce trimming in order to control the tailbehaviour of $\hat{\pi}$ since this appears in the denominator of both estimators. To this end, we introduce a trimming function, $\hat{T}$, which we require satisfies

$$T(x; \pi, a) = \left\{ \begin{array}{ll} 1, & \pi(x) \geq a \\ 0, & \pi(x) \leq a/2 \end{array} \right. , \tag{36}$$

for a positive sequence $a = a$ such that $a \to 0$. We impose further regularity conditions on the trimming function:

$\underline{\mathcal{T}(\omega)}$ The function $T(x; \pi, a)$ (i) satisfies (36), (ii) is ω times continuously differentiable in x with $\partial_x^i T(x; \pi, a)$ bounded, $i = 0, ..., \omega$, and (iii) continuously differentiable in a with $a\partial_a T(x; \pi, a)$ bounded.

In the following we shall write $\hat{T}(x; a) = T(x; \hat{\pi}, a)$ and $T_0(x; a) = T(x; \pi_0, a)$. Given $\hat{T}$, we redefine $\hat{\mu}$ in Class 1 as

$$\hat{\mu}(x) = \left\{ \frac{1}{2}\partial_x \sigma^2\left(x; \hat{\theta}\right) + \frac{1}{2}\sigma^2\left(x; \hat{\theta}\right) \frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} \right\} \hat{T}(x; a). \tag{37}$$

Similarly, we redefine $\hat{\sigma}^2$ in Class 2 as

$$\hat{\sigma}^2(x) = \frac{n^{-1}\sum_{i=1}^n 1_{(l,x)}(x_i)\hat{\mu}(x_i)}{2\hat{\pi}(x)}\hat{T}(x; a). \tag{38}$$

In order to establish sufficiently fast convergence of the nonparametric part in the appropriate functional norm, we introduce the following class $\mathcal{K}(\omega, \lambda)$ of higher-order, bias-reducing kernels, first proposed by Parzen (1962):

$\underline{\mathcal{K}(\omega, \lambda)}$ The kernel K satisfies $\int_{\mathbb{R}} K(x)\,dx = 1$; $\int_{\mathbb{R}} x^i K(x)\,dx = 0$, for $0 \leq i \leq \omega-1$; $\int_{\mathbb{R}} |x|^\omega |K(x)|\,dx < \infty$; $K^{(i)}(x) \to 0$ as $|x| \to \infty$, $0 \leq i \leq \lambda - 1$;

$\sup_{x \in \mathbb{R}} |K^{(i)}(x)| \max(|x|, 1) < \infty$, $0 \leq i \leq \lambda + 1$; $K^{(i)}$ is absolutely integrable with Fourier transform Ψ_i satisfying $\int_{\mathbb{R}} (1 + |x|)\sup_{b \geq 1} |\Psi_i(bx)|\,dx < \infty$, $0 \leq i \leq \lambda$.

We first derive the asymptotics for models in Class1. We assume the following:

SP.0 The sequence $\{x_i\}$ is stationary and β-mixing with geometrically decreasing mixing coefficients.

SP.1 The marginal density π_0 is ω times continuously differentiable with bounded derivatives.

SP.2 $\left(\mu_0, \sigma_0^2\right) \in \mathcal{D}$.

SP.3 The estimator $\hat{\theta}$ satisfies (29) with $\theta_0 \in \text{int}\Theta$.

SP.4 The transition density p_t exists for any $t \geq 0$ such that the mapping $y \mapsto p_t(y|x)$ is bounded, and continuously differentiable with bounded first derivative.

SP1.A The kernel $K \in \mathcal{K}(\omega, 2)$ and the trimming function $T \in \mathcal{T}(2)$. The bandwidth h and the trimming parameter a satisfies $n^{-1/2}a^{k-2}h^{-2-k} \to 0$ and $a^{k-2}h^{\omega-k} \to 0$, $k = 0, 1$, as $a, h \to 0$.

SP1.B The bandwidth h and the trimming parameter a satisfies

    1. $n^{-1/4}a^{k-2}h^{-1-k} \to 0$ and $n^{1/4}a^{k-2}h^{\omega-k} \to 0$, $k = 0, 1, 2$.

    2. $\int_t^T P_{t,x}\left(a/2 \leq \pi_0\left(X_s^0\right) \leq a\right) ds = o(n^{-4})$.

Theorem 6 (Class 1) *Assume that (SP.0)-(SP.3) and (SP1.A) hold and $\omega \geq 3$. Then $\|\hat{u} - u_0\|_{0,1} = o_P(1)$. If additionally (SP.4) and (SP1.B) hold and $\omega \geq 4$, then*

$$\sqrt{n}\left(\hat{u}(t, x) - u_0(t, x)\right) \xrightarrow{d} N\left(0, V(t, x)\right),$$

where

$$V(t, x) = \operatorname{var}\left(\nu(x_1|x_0; t, x)\right) + 2\sum_{i=1}^{\infty} \operatorname{cov}\left(\nu(x_1|x_0; t, x), \nu(x_{i+1}|x_i; t, x)\right)$$

and $\{\nu(x_i|x_{i-1}; t, x)\}$ is given in (46).

Next, we derive the asymptotics in Class 2. This is done under very much the same assumptions as the ones assumed for Class 1. Only do we need to slightly change the conditions on the bandwidth and trimming parameter:

SP2.A The kernel $K \in \mathcal{K}(\omega, 1)$ and the trimming function $T \in \mathcal{T}(1)$. The bandwidth h and the trimming parameter a satisfies $n^{-1/2}a^{-1}h^{-1} \to 0$ and $a^{-1}h^{\omega} \to 0$ as $a, h \to 0$.

SP2.B The bandwidth h and the trimming parameter a satisfies

    1. $n^{-1/4}a^{k-2}h^{-1-k} \to 0$ and $n^{1/4}a^{k-2}h^{\omega-k} \to 0$, $k = 0, 1$.

    2. $a^8 \int_t^T P_{t,x}\left(a/2 \leq \pi_0\left(X_s^0\right) \leq a\right) ds = o(n^{-4})$.

Theorem 7 (Class 2) *Assume that (SP.0)-(SP.3) and (SP2.A) hold and $\omega \geq 2$. Then $\|\hat{u} - u_0\|_{0,1} = o_P(1)$. If additionally (SP.4) and (SP2.B) hold and $\omega \geq 3$, then*

$$\sqrt{n}\left(\hat{u}(t, x) - u_0(t, x)\right) \xrightarrow{d} N\left(0, V(t, x)\right),$$

where

$$V(t, x) = \operatorname{var}\left(\nu(x_i|x_{i-1}; t, x)\right) + 2\sum_{i=1}^{\infty} \operatorname{cov}\left(\nu(x_1|x_0; t, x), \nu(x_{i+1}|x_i; t, x)\right)$$

and $\nu_i(t, x)$ is given in (49).

Sufficient conditions for (SP.0) to hold can be found in Kristensen (2004a). (SP.1) holds if μ_0 and σ_0^2 both are ω times continuously differentiable. Aït-Sahalia (2002) gives sufficient conditions for (SP.4) to hold

For both classes of estimators, the asymptotic variance $V(t,x)$ is of unknown form. One can use bootstrap methods to obtain an approximation of the distribution of the estimator. Alternatively, one can use an idea originating from Newey (1994a) to estimate the variance using the pathwise derivative. We only present the variance estimator for Class 1; the Class 2 case is dealt with similarly. We define

$$\hat{V}(t,x) = \hat{\Omega}_0(t,x) + \sum_{i=1}^{M} w_{M,i}(\hat{\Omega}_i(t,x) + \hat{\Omega}_i^{\top}(t,x)),$$

where $w_{M,i} = 1 - [i/(M+1)]$, $\hat{\Omega}_i(t,x) = n^{-1} \sum_{j=i}^{n} \hat{\omega}_j(t,x) \hat{\omega}_{j-i}^{\top}(t,x)$, $\hat{\omega}_j(t,x) = \hat{\nu}_j^{(1)}(t,x) + \hat{\nu}_j^{(2)}(t,x)$, and

$$
\begin{aligned}
\hat{\nu}_j^{(1)}(t,x) &= \left. \frac{\partial \Gamma(\mu(\cdot; \hat{\theta}, \hat{\pi} + \alpha K_h(\cdot - x_j)), \sigma^2(\cdot; \hat{\theta}),)(t,x)}{\partial \alpha} \right|_{\alpha=0}, \\
\hat{\nu}_j^{(2)}(t,x) &= \left. \frac{\partial \Gamma(\mu(\cdot; \theta, \hat{\pi}), \sigma^2(\cdot; \theta),)(t,x)}{\partial \theta} \right|_{\theta=\hat{\theta}} \psi(x_j | x_{j-1}).
\end{aligned}
$$

The two functions can be calculated using numerical derivatives. This estimator should be consistent as $M \to \infty$ and $M/n^{1/8} \to 0$. We will not give a formal proof of this, and instead refer to Section 4.4.

3.3 A Nonparametric Estimator

In this section we shall consider fully nonparametric kernel estimators of μ and σ^2 in the univariate case, $q = 1$. Such estimators have been considered in a series of papers, see e.g. Florens-Zmirou (1993), Jiang and Knight (1997), Stanton (1997), Bandi and Moloche (2001), Bandi and Phillips (2003). All these papers consider a sampling scheme where the time distance between observations $\Delta = \Delta_n \to 0$, as the number of observations $n \to \infty$; this is the so-called in-fill assumption. This enables one to reconstruct the full sample path in any compact interval in the limit, and thereby extract enough information about the infinitesimal conditional variance, σ^2, for it to be estimable. However, to construct an estimator of the infinitesimal mean, μ, it is necessary also to require that the length of the time interval in which the process is observed, $\bar{T} \to \infty$; this is the so-called long-span assumption. Bandi and Phillips (2003) obtain pointwise consistency and mixed asymptotic normality of the drift and diffusion estimators only assuming recurrence of the process thereby allowing for certain forms of non-stationarity. We apply the estimators proposed by Bandi and Phillips (2003). But it appears to be difficult to work under their general assumption of recurrence since the convergence rates of the estimators in the general case is path-dependent. This in particular makes it difficult to show consistency in a functional norm. So for simplicity, we restrict our attention to diffusion processes having a stationary marginal density π.

We assume that we have observed $\{x_i\}$, $x_i = x_{i\Delta}$, in the interval $[0, \bar{T}]$ where $\bar{T} = n\Delta \to \infty$. We shall assume that $\{x_t\}$ takes values on the interval $I \subseteq \mathbb{R}$, and that the process is stationary and mixing. As we shall see, the nonparametric estimator of μ used here only has $\sqrt{\bar{T}h} = \sqrt{n\Delta h}$-convergence rate, while the nonparametric estimator of σ^2 exhibits faster $\sqrt{nh}$-convergence rate. This in turn will mean

that the drift estimator will be the dominating term when deriving the asymptotics of $\hat{u}$. In particular, the convergence rate of $\hat{u}$ is $\sqrt{T}$ and not $\sqrt{n}$.

Before we define our estimators, we first introduce $m(x) \equiv \mu(x)\pi(x)$ and $s(x) \equiv \sigma^2(x)\pi(x)$ such that

$$\mu(x) = \frac{m(x)}{\pi(x)}, \quad \sigma^2(x) = \frac{s(x)}{\pi(x)}.$$

We then construct kernel estimators of π, m, and s,

$$\hat{\pi}(x) = n^{-1}\sum_{i=1}^{n} K_h(x_i - x),$$

$$\hat{m}(x) = n^{-1}\sum_{i=1}^{n} K_h(x_i - x)\frac{x_{i+1} - x_i}{\Delta},$$

$$\hat{s}(x) = n^{-1}\sum_{i=1}^{n} K_h(x_i - x)\frac{(x_{i+1} - x_i)^2}{\Delta}.$$

As in the previous section, we need to control the tailbehaviour of $\hat{\pi}$. So we introduce trimmed versions of our estimators,

$$\hat{\mu}(x) = \hat{T}(x;a)\frac{\hat{m}(x)}{\hat{\pi}(x)}, \quad \hat{\sigma}^2(x) = \hat{T}(x;a)\frac{\hat{s}(x)}{\hat{\pi}(x)}.$$

The basic conditions are almost the same as the ones assumed for the semiparametric estimators:

NP.0 $\{x_t\}$ is stationary and β-mixing with geometrically decreasing mixing coefficients.

NP.1 The kernel $K \in \mathcal{K}(\omega, 1)$ and the trimming function $T \in \mathcal{T}(1)$.

NP.2 The marginal density π_0 is ω times continuously differentiable with bounded derivatives.

NP.3 $\left(\mu_0, \sigma_0^2\right) \in \mathcal{D}$.

NP.4 The transition density p_t exists for any $t \geq 0$ such that the mapping $y \mapsto p_t(y|x)$ is bounded, and continuously differentiable with bounded derivative.

NP.5A The bandwidth h and the trimming parameter a satisfies $\bar{T}^{-1/2}a^{-2}h^{-1} \to 0$, $a^{-2}h^{\omega} \to 0$, and $\bar{T}^{1/2+\delta}\sqrt{\log_2(\bar{T})}\Delta^{3/4}\log(\Delta^{-1})^{1/4}a^{-2}h^{-2} \to 0$, as $a, h \to 0$.

NP.5B The bandwidth h and the trimming parameter a satisfies

1. $\bar{T}^{3/4+\delta}\sqrt{\log_2(\bar{T})}\Delta^{3/4}\log(\Delta^{-1})^{1/4}a^{k-2}h^{-2-k} \to 0$, $\bar{T}^{-1/4}a^{k-2}h^{-1-k} \to 0$ and $\bar{T}^{1/4}a^{k-2}h^{\omega-k} \to 0$, $k = 0, 1$.

2. $\int_t^T P_{t,x}\left(a/2 \leq \pi_0\left(X_s^0\right) \leq a\right)ds = o(n^{-4})$.

Applying results from Bosq (1998), we are able to show that $\hat{\pi}$, $\hat{m}$, and $\hat{s}$ are uniformly consistent on I, and also supply convergence rates. Given these, it is then an easy task to show that the nonparametric estimators of μ and σ^2 converge in the $\|\cdot\|_{0,4}$-norm. This shows consistency. We are able to strengthen this $\|\hat{\mu} - \mu_0\|_{1,4} = o_P\left(\bar{T}^{-1/4}\right)$ and $\left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{1,4} = o_P\left(\bar{T}^{-1/4}\right)$. The pathwise derivative consists of two

parts, the first part being a functional of $\hat{\mu}$, $\nabla_1\Gamma$, and the second a functional of $\hat{\sigma}^2$, $\nabla_2\Gamma$. It can now be shown that $\nabla_1\Gamma\left[\hat{\mu}-\mu_0\right]$ converges towards a normal distribution with speed $\sqrt{T}$, while $\nabla_2\Gamma\left[\hat{\sigma}^2-\sigma_0^2\right]$ does so with speed $\sqrt{n}$. Thus, the first term dominates the second one, implying that $\nabla_1\Gamma\left[\hat{\mu}-\mu_0\right]$ drives the asymptotic distribution.

Theorem 8 (Nonparametric) *Assume that (NP.0)-(NP.3) together with (NP.5A) hold with $\omega\geq 2$. Then the nonparametric estimator $\hat{u}$ is consistent. If additionally (NP4) and (NP5.B) hold then*

$$\sqrt{\overline{T}}\left(\hat{u}\left(t,x\right)-u_0\left(t,x\right)\right)\xrightarrow{d}N\left(0,V\left(t,x\right)\right),$$

where

$$V\left(t,x\right)=E\left[\sigma_0^2\left(x_s\right)\frac{d_\mu^{*2}\left(x_s\right)}{\pi_0^2\left(x_s\right)}\left(\int_t^T p_u\left(x_s|x\right)du\right)^2\right],$$

with d_1^* given in (24).

We propose to estimate the variance by $\hat{V}\left(t,x\right)$ as given in the previous section, only we redefine $\hat{\nu}_j^{(1)}\left(t,x\right)$ and $\hat{\nu}_j^{(2)}\left(t,x\right)$ as

$$
\begin{aligned}
\hat{\nu}_j^{(1)}\left(t,x\right) &= \left.\frac{\partial\Gamma(\mu(\cdot;\hat{\pi}+\alpha K_h\left(\cdot-x_j\right),\hat{m}),\sigma^2(\cdot;\hat{\pi},\hat{s},)\,)\left(t,x\right)}{\partial\alpha}\right|_{\alpha=0}, \\
\hat{\nu}_j^{(2)}\left(t,x\right) &= \left.\frac{\partial\Gamma(\mu(\cdot;\hat{\pi},\hat{m}+\alpha\left(x_{j+1}-x_j\right)K_h\left(\cdot-x_j\right)/\Delta),\sigma^2(\cdot;\hat{\pi},\hat{s},)\,)\left(t,x\right)}{\partial\alpha}\right|_{\alpha=0}.
\end{aligned}
$$

The two functions can be calculated using numerical methods. This estimator should be consistent as $M\to\infty$ and $M/T^{1/8}\to 0$. We will not give a formal proof of this.

Bandi and Moloche (2001) generalise the above nonparametric estimators of μ and σ^2 to the multivariate case. In Jeffrey et al (2004), a kernel estimator of the volatility function in a class of Heath-Jarrow-Morton models is proposed. Series estimator of μ and σ^2 for a one-dimensional diffusion has been proposed by Chen et al (2000a, b). We conjecture that similar results to the one given above can be derived for these estimators. In particular, the curse of dimensionality will not be a problem in the estimation of u; the dimension of the underlying diffusion process $\{X_t\}$ will have no effect on the rate of convergence.

Jiang and Knight (1997) propose an alternative drift estimator that makes explicit use of the assumption of stationarity of the process. Jiang (1998) examines the estimation of solutions to PDE's when their nonparametric estimators are plugged in. He claims that the estimated solution, $\hat{u}$, converges with $\sqrt{n}$-rate. We believe there is a mistake in his proof since his drift estimator only converges with speed $\sqrt{\overline{T}h^3}$. Thus, the convergence of $\hat{u}$ should not be able to exceed $\sqrt{\overline{T}}$.

4   Applications

In this section, we return to the examples given in Section 2.1 and 2.2 and discuss how the results derived in the previous section can be applied to these.

We discussed in Section 2.1 how LPDE's can be used to characterise derivative prices. The results given in the previous section can now ensures that option prices calculated using preliminary estimated

models of the underlying variables are consistent and asymptotically normally distributed in great generality. In particular, this enables us to calculate standard errors of the estimated prices which gives us a measure of the statistical accuracy of the prices and allows us to test the individual asset pricing model.

Example 1 (continued). Under the physical measure P, assume that we have a parametric diffusion model of (8),

$$dS_t = \mu_S(t, S_t; \theta) dt + \sigma_S(t, S_t; \theta) dw_t^S, \tag{39}$$

the dividend stream is zero, $d_t = 0$, and the short-rate is assumed to be constant, $r_t = r > 0$. We assume that an estimator $\hat{\theta}$ of θ is available; this may have been obtained using historical observations of the stock prices, $S_{i\Delta}$, $i = 1, ..., n$, and applying MLE, GMM or some other method. Defining $x_i = S_{i\Delta}$, $\mu = \mu_S$, and $\sigma = \sigma_S$, Theorem 5 gives conditions under which any implied derivative price based on this estimator will be asymptotically normally distributed.

In the stochastic volatility model, we also have to obtain an estimator of the market price for volatility, λ. Assuming that this is a known function up to θ, $\lambda(S, v) = \lambda(S, v; \theta)$, the results carry through given smoothness conditions on λ of the same type as imposed on $\mu_S(t, S; \theta)$ and $\sigma_S^2(t, S; \theta)$. The estimator of λ may have been obtained using other data than historical observations of the stock price(s). This can also be accommodated for.

Example 2 (continued). We assume that we have observed the short rate at discrete points in time, $r_{i\Delta}$, $i = 1, ..., n$, and that it under P solves

$$dr_t = \mu(r_t) dt + \sigma(r_t; \theta) dw_t.$$

The semiparametric estimator considered in Section 3.2 is then used to estimate μ and σ^2. Taking the market price of risk, $\lambda(r)$, for given/known, Theorem 6 gives us the asymptotic distribution of any implied bond and interest rate derivative price.

We can also allow for an unknown market price for risk for which we have an estimator, $\hat{\lambda}$. Since $\lambda(r)$ enters the drift function linearly, one can easily accommodate for this in our proofs. Defining $\mu^\lambda = \mu - \lambda\sigma$, we first obtain

$$\left| \hat{u}(t, x) - u_0(t, x) - \nabla\Gamma[\hat{\mu}^\lambda - \mu_0^\lambda, \hat{\sigma}^2 - \sigma_0^2](t, x) \right| \le b(x, T) (||\hat{\mu}^\lambda - \mu_0^\lambda||_{1,4}^2 + ||\hat{\sigma}^2 - \sigma_0^2||_{1,4}^2),$$

where $\left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{1,4} = o_P(n^{-1/4})$ and

$$||\hat{\mu}^\lambda - \mu_0^\lambda||_{1,4} \le ||\hat{\mu} - \mu_0||_{1,4} + \hat{\lambda}||_{1,4}^2 \hat{\sigma} - \sigma_0||_{1,4} + ||\sigma_0||_{1,4} \hat{\lambda} - \lambda_0||_{1,4} = o_P(n^{-1/4}),$$

if $||\hat{\lambda} - \lambda_0||_{1,4} = o_P(n^{-1/4})$. Next, due to the linearity of $\nabla_1\Gamma$,

$$\begin{aligned} \nabla\Gamma[\hat{\mu}^\lambda - \mu_0^\lambda, \hat{\sigma}^2 - \sigma_0^2](t, x) &= \nabla_1\Gamma[\hat{\lambda}\hat{\sigma} - \lambda_0\sigma_0](t, x) + \nabla_1\Gamma[\hat{\mu} - \mu_0](t, x) \\ &\quad + \nabla_2\Gamma[\hat{\sigma}^2 - \sigma_0^2](t, x). \end{aligned}$$

The second and third term is treated in the proof of Theorem 6, while the first one requires a bit of work: Observe that

$$\hat{\lambda}\hat{\sigma} - \lambda_0\sigma_0 = \sigma_0(\hat{\lambda} - \lambda_0) + \frac{\lambda_0}{\sigma_0}(\hat{\sigma}^2 - \sigma_0^2) + O(||\hat{\sigma}^2 - \sigma_0^2||^2 + ||\hat{\lambda} - \lambda_0||^2),$$

25

such that

$$
\begin{aligned}
\nabla_1\Gamma[\hat{\lambda}\hat{\sigma} - \lambda_0\sigma_0](t,x) = &\ \nabla_1\Gamma[\sigma_0(\hat{\lambda} - \lambda_0)](t,x) + \nabla_1\Gamma\left[\lambda_0/\sigma_0\left(\hat{\sigma}^2 - \sigma_0^2\right)\right](t,x) \\
& + O\left(||\hat{\sigma}^2 - \sigma_0^2||_{0,2}^2 + ||\hat{\lambda} - \lambda_0||_{0,2}^2\right).
\end{aligned}
$$

The term $\nabla_1\Gamma[\lambda_0/\sigma_0(\hat{\sigma}^2 - \sigma_0^2)]$ can be treated as $\nabla_2\Gamma[\hat{\sigma}^2 - \sigma_0^2]$, while $\nabla_1\Gamma[\sigma_0(\hat{\lambda} - \lambda_0)]$, will converge in distribution in great generality.

Next we turn to the two examples given in Section 2.2 concerning the estimation of diffusion models. We check for each of the two examples that under regularity conditions the proposed estimator will be consistent and asymptotically normally distributed.

Example 4 (continued). We here give primitive conditions under which the estimator proposed by Bibby and Sørensen (1995) is consistent and asymptotically normally distributed. For simplicity, we only consider the univariate case $(q = 1)$ and assume that $b(x) = x$, while $\alpha(x;\theta) = \alpha(x)$ is parameter independent. We first set up a set of conditions:

C.4.1 There exists $\bar{p} \geq 2$ and constants $c_0, c_1 > 0$ such that

$$
2\mu(x;\theta)\|x\|^{2\bar{p}-1} + (\bar{p} - 1)\sigma_0^2(x)\|x\|^{2(\bar{p}-1)} \leq c_0 - c_1\|x\|^{2\bar{p}}.
$$

C.4.2 The drift and diffusion function, $\left(\mu(\cdot;\theta), \sigma^2(\cdot;\theta)\right)$, in (18) belongs to $\mathcal{D}$ for any $\theta \in \Theta$.

C.4.3 The function $\alpha : \mathbb{R} \mapsto \mathbb{R}$ satisfies $|\alpha(x)| \leq C\left(1 + \|x\|^{\bar{p}/2}\right)$.

C.4.4 The matrix $H(\theta) \equiv E_\theta[\dot{g}(X_\Delta|X_0;\theta)]$ is positive definite.

It now follows from Theorem 2 and 3 that $B(x;\theta)$ is continuously differentiable in θ, and by (6) $|B(x;\theta)| \leq C(K,\Delta)(1 + |x|)$, $\left|\dot{B}(x;\theta)\right| \leq C(K,\Delta)(1 + |x|^2)$. The first condition implies, c.f. Meyn and Tweedie (1993), that $\{X_t\}$ is stationary and ergodic (assuming it has been started at its invariant distribution) with $E_\pi[|X_0|^{2\bar{p}}] < \infty$. We have that

$$
|g(y|x;\theta)|^2 \leq C\left(1 + \|x\|^{\bar{p}}\right)\left(|y|^2 + |x|^2\right), \quad \|\dot{g}(y|x;\theta)\| \leq C\left(1 + \|x\|^{\bar{p}/2}\right)\left(|y|^2 + |x|^2\right)
$$

where

$$
E\left[\left(1 + \|X_0\|^{\bar{p}}\right)\left(|X_\Delta|^2 + |X_0|^2\right)\right] \leq CE\left[1 + \|X_0\|^{2\bar{p}}\right]^{1/2} E_\pi\left[\left(|X_\Delta|^4 + |X_0|^4\right)\right]^{1/2} < \infty.
$$

By Law of Large Numbers and a central limit theorem for martingales, we now obtain that

$$
\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow^d N\left(0, H^{-1}(\theta_0)V(\theta_0)H^{-1}(\theta_0)\right) \tag{40}
$$

where $V(\theta) = E_\theta\left[g(X_\Delta|X_0;\theta)g(X_\Delta|X_0;\theta)^\top\right]$.

Example 5 (continued). We here give conditions under which the least squares estimator of θ based on observe option prices is consistent and asymptotically normally distributed,

$$
\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} (P_i - \Pi(X_i, T;\theta))^2,
$$

where $X_i = X_{i\Delta}$. We assume that (C.4.1)-(C.4.2) hold and that

C.5.1 The pay-off function $g : \mathbb{R} \mapsto \mathbb{R}$ is continuously differentiable and satisfies $\left| \partial_x^i g(x) \right| \leq C \left(1 + \|x\|^{\bar{p}/2} \right)$, $i = 0, 1$.

C.5.2 The matrix $H(\theta) \equiv E_\theta \left[\dot{\Pi}(X_i, T; \theta) \dot{\Pi}(X_i, T; \theta)^\top \right]$ is positive definite.

C.5.3 The error sequence $\{\varepsilon_i\}$ is independent of $\{X_i\}$, i.i.d. and with $E[\varepsilon_i] = 0$, $\sigma_\varepsilon^2 = E[\varepsilon_i^2] < \infty$.

Under (C.4.1)-(C.4.2) and (C.5.1)-(C.5.2), we obtain by standard arguments that

$$\sqrt{n}(\hat{\theta} - \theta_0) \to^d N\left(0, \sigma_\varepsilon^2 H^{-1}(\theta_0) \right). \tag{41}$$

The assumptions in (C.5.3) on the errors can be weakened substantially.

5   Conclusion

We have investigated the properties of estimated solutions of a PDE given preliminary estimates of the driving coefficients of the PDE. We gave general conditions under which the estimated solution was consistent and asymptotically normally distributed, and checked that these were satisfied in three leading examples.

We demonstrated that these results have widespread use both in finance and econometrics. In particular, the results can be used when drawing inference on implied derivative prices given estimates of the dynamics of the underlying asset. Also in the literature on estimation of discretely observed diffusions are our results useful.

A   Proofs

Proof of Theorem 1. We have that

$$
\begin{aligned}
|u(t,x)| &\leq E_{t,x}\left[|b(X_T)|\right] + E_{t,x}\left[\int_t^T |c(s, X_s)|\, ds\right] \\
&\leq C\left(1 + E_{t,x}\left[|X_T|^r\right]\right) + E_{t,x}\left[\int_t^T C\left(1 + E_{t,x}\left[|X_s|^r\right]\right) ds\right] \\
&\leq C\left(1 + \|x\|^r\right),
\end{aligned}
$$

where we have used (6). We obtain that

$$
\begin{aligned}
\frac{\partial u(t,x)}{\partial x_i} &= E_{t,x}\left[\exp\left[-\int_t^T a(u, X_u)\, du\right]\left\{ b_x(X_T) Y_T^{(i)} - b(X_T) \int_t^T a_x(s, X_s) Y_s^{(i)} ds\right.\right. \\
&\quad \left.\left. + \int_t^T c_x(s, X_s) Y_s^{(i)} - c(s, X_s) \int_t^s a_x(u, X_u) Y_u^{(i)} du \right\}\right].
\end{aligned}
$$

where $\{Y_t^{(i)}\}$ is given in Lemma 11, such that $|\partial u(t,x)/\partial x_i|$ is bounded by

$$E_{t,x}\left[\|b_x(X_T)\|\,\|Y_T^{(i)}\| + |b(X_T)|\int_t^T \|a_x(s,X_s)\|\,\|Y_s^{(i)}\|ds\right]$$

$$+E_{t,x}\left[\int_t^T |c_x(s,X_s)|\,\|Y_s^{(i)}\| + \int_t^s \|a_x(u,X_u)\|\,\|Y_u^{(i)}\|duds\right]$$

$$\leq C\left(1+\left(\int_t^T E_{t,x}\left[\|X_s\|^{2r}\right]ds\right)^{1/2}\right)\left(1+\left(\int_t^T E_{t,x}\left[\|Y_s^{(i)}\|^2\right]ds\right)^{1/2}\right)$$

$$\leq C\left(1+\|x\|^q\right),$$

for any $(t,x)\in[0,T]\times\mathbb{R}^q$, where we have used (6). The expression of $\partial^2 u(t,x)/\partial x_i\partial x_j$ is not presented for brevity; one may show that $\left|\partial^2 u(t,x)/\partial x_i\partial x_j\right|\leq C\left(1+\|x\|^r\right)$, for any $(t,x)\in[0,T]\times\mathbb{R}^d$. This shows the first part of the theorem. We then easily realise that

$$\|u\|_{2,p}\leq C\left(1+\int_0^T E\left[\|X_t^0\|^{pr}\right]\right)\leq C\left(1+E\left[\|X^*\|^{pr}\right]\right)<\infty,$$

for $pr\leq p^*$. $\blacksquare$

Proof of Theorem 2. We see that

$$\left|E_{t,x}\left[b(X_T)\exp\left[-\int_t^T a(u,X_u)\,du\right]-b(X_T^0)\exp\left[-\int_t^T a(u,X_u^0)\,du\right]\right]\right|$$

$$\leq C\left(1+\|x\|^r\right)\left(\int_t^T E_{t,x}\left[\left|a(u,X_u)-a(u,X_u^0)\right|^2\right]du\right)^{1/2}+E_{t,x}\left[\left|b(X_T)-b(X_T^0)\right|\right]$$

where, by Lemma 10,

$$E_{t,x}\left[\|a(u,X_u)-a(u,X_u)\|^2\right] \leq C(u,x)\left(\int_t^u E_{t,x}\left[\|\mu(v,X_v^0)-\mu_0(v,X_v^0)\|^4\right]dv\right.$$

$$\left.+\int_t^u E_{t,x}\left[\|\sigma^2(v,X_v^0)-\sigma_0^2(v,X_v^0)\|^4\right]dv\right)^{1/2},$$

with $C(t,x)=Ct\left(1+\|x\|^r\right)$, and

$$E_{t,x}\left[\left|b(X_T)-b(X_T^0)\right|\right] \leq C(T,x)\left(\int_t^T E_{t,x}\left[\|\mu(u,X_u^0)-\mu_0(u,X_u^0)\|^2\right]du\right.$$

$$\left.+\int_t^T E_{t,x}\left[\|\sigma^2(u,X_u^0)-\sigma_0^2(u,X_u^0)\|^2\right]du\right)^{1/2}.$$

Similarly,

$$\left|E_{t,x}\left[\int_t^T c(s,X_s)\exp\left[-\int_t^s a(u,X_u)\,du\right]ds\right]\right.$$

$$\left.-E_{t,x}\left[\int_t^T c(s,X_s^0)\exp\left[-\int_t^s a(u,X_u^0)\,du\right]ds\right]\right|$$

$$\leq C(T,x)\left(\int_t^T E_{t,x}\left[\|\mu(u,X_u^0)-\mu_0(u,X_u^0)\|^4\right]du\right.$$

$$\left.+\int_t^T E_{t,x}\left[\|\sigma^2(u,X_u^0)-\sigma_0^2(u,X_u^0)\|^4\right]du\right)^{1/2}$$

28

We conclude that

$$|u(t,x) - u_0(t,x)| \leq C(T,x) \left(\int_t^T E_{t,x} \left[\| \mu(u, X_u^0) - \mu_0(u, X_u^0) \|^4 \right] du \right.$$
$$\left. + \int_t^T E_{t,x} \left[\| \sigma^2(u, X_u^0) - \sigma_0^2(u, X_u^0) \|^4 \right] du \right)^{1/4}$$

Taking expectations,

$$\|u - u_0\|_{0,1} \leq CT(1 + E[\|X^*\|^q]) \left\{ \|\mu - \mu_0\|_{0,4} + \|\sigma^2 - \sigma_0^2\|_{0,4} \right\}.$$

∎

Proof of Theorem 3. For any function f, we have

$$\left\| f(x) - f(x_0) - f^{(1)}(x_0)(x - x_0) \right\| \leq \left\| f^{(2)}(\lambda x + (1 - \lambda) x_0) \right\| \|x - x_0\|^2,$$

for some $\lambda \in [0,1]$. Thus,

$$\left| E_{t,x} \left[b(X_T) \exp \left[-\int_t^T a(u, X_u) du \right] \right] - E_{t,x} \left[b(X_T^0) \exp \left[-\int_t^T a(u, X_u^0) du \right] \right] \right.$$
$$- E_{t,x} \left[b_x(X_T^0) [X_T - X_T^0] \exp \left[-\int_t^T a(u, X_u^0) du \right] \right]$$
$$\left. + E_{t,x} \left[b(X_T^0) \exp \left[-\int_t^T a(s, X_s^0) ds \right] \left(\int_t^T a_x(u, X_s^0) [X_s - X_s^0] ds \right) \right] \right|$$
$$\leq CTe^{CT} \left[1 + \|x\|^{4r} \right]^{1/2} \left\{ E_{t,x} \left[\int_t^T \|X_s - X_s^0\|^4 ds + \right]^{1/2} + E_{t,x} \left[\|X_T - X_T^0\|^4 \right]^{1/2} \right\},$$

Also,

$$\left| E_{t,x} \left[b_x(X_T^0) \exp \left[-\int_t^T a(u, X_u^0) du \right] [X_T - X_T^0 - \nabla X_T] \right] \right|$$
$$\leq C(T,x) E_{t,x} \left[\|X_T - X_T^0 - \nabla X_T\|^2 \right]^{1/2}$$

and

$$\left\| E_{t,x} \left[b(X_T^0) \exp \left[-\int_t^T a(s, X_s^0) ds \right] \left(\int_t^T a_x(u, X_s^0) [X_s - X_s^0 - \nabla X_s] ds \right) \right] \right\|$$
$$\leq C(T,x) E_{t,x} \left[\left(\int_t^T \|X_s - X_s^0 - \nabla X_s\|^2 ds \right) \right]^{1/2}$$

Similarly,

$$\left| E_{t,x} \left[\int_t^T c(s, X_s) \exp \left[-\int_t^s a(u, X_u) du \right] ds \right] \right.$$
$$- E_{t,x} \left[\int_t^T c(s, X_s^0) \exp \left[-\int_t^s a(u, X_u^0) du \right] ds \right]$$
$$- E_{t,x} \left[\int_t^T c_x(s, X_s^0) [X_s - X_s^0] \exp \left[-\int_t^s a(u, X_u^0) du \right] ds \right]$$
$$\left. + E_{t,x} \left[\int_t^T c(s, X_s^0) \exp \left[-\int_t^s a(u, X_u^0) du \right] \left(\int_t^s a_x(u, X_u^0) [X_u - X_u^0] du \right) ds \right] \right|$$
$$\leq C(T,x) E_{t,x} \left[\int_t^T \|X_s - X_s^0\|^4 ds \right]^{1/2},$$

and

$$\left\| E_{t,x} \left[\int_t^T c_x \left(s, X_s^0 \right) \exp \left[- \int_t^s a \left(u, X_u^0 \right) du \right] \left[X_s - X_s^0 - \nabla X_s \right] ds \right] \right\|$$
$$\leq \quad C \left(T, x \right) E_{t,x} \left[\int_t^T \left\| X_{Ts} - X_s^0 - \nabla X_s \right\|^2 ds \right]^{1/2} .$$

Combining these results we obtain

$$|u \left(t, x \right) - u_0 \left(t, x \right) - \nabla \Gamma \left(t, x \right)|$$
$$\leq \quad C \left(T, x \right) \left\{ E_{t,x} \left[\int_t^T \left\| X_s - X_s^0 \right\|^4 ds + \right]^{1/2} + E_{t,x} \left[\left\| X_T - X_T^0 \right\|^4 \right]^{1/2} \right\}$$
$$+ C \left(T, x \right) E_{t,x} \left[\left(\int_t^T \left\| X_s - X_s^0 - \nabla X_s \right\|^2 ds \right) \right]^{1/2} .$$

An application of Lemma 9 and 12 then proves the claim.

The following inequalities yield the second part of the theorem: For any $(d\mu, d\sigma) \in \bar{\mathcal{D}}$,

$$\left| E_{t,x} \left[b_x \left(X_T^0 \right) \nabla X_T \left[d\mu, d\sigma^2 \right] \exp \left[- \int_t^T a \left(u, X_u^0 \right) du \right] \right] \right|$$
$$\leq \quad E_{t,x} \left[\left\| b_x \left(X_T^0 \right) \right\|^2 \right]^{1/2} E_{t,x} \left[\left\| \nabla X_T \left[d\mu, d\sigma^2 \right] \right\|^2 \right]^{1/2} ,$$

$$\left| E_{t,x} \left[b \left(X_T^0 \right) \exp \left[- \int_t^T a \left(s, X_s^0 \right) ds \right] \left(\int_t^T a_x \left(u, X_s^0 \right) \nabla X_s ds \right) \right] \right|$$
$$\leq \quad E_{t,x} \left[\left\| b_x \left(X_T^0 \right) \right\|^2 \int_t^T a_x \left(u, X_s^0 \right)^2 ds \right]^{1/2} E_{t,x} \left[\int_t^T \left\| \nabla X_s \left[\mu, \sigma^2 \right] \right\|^2 ds \right]^{1/2} ,$$

$$\left| E_{t,x} \left[\int_t^T c_x \left(s, X_s^0 \right) \nabla X_s \exp \left[- \int_t^s a \left(u, X_u^0 \right) du \right] ds \right] \right|$$
$$\leq \quad E_{t,x} \left[\int_t^T \left\| c_x \left(s, X_s^0 \right) \right\|^2 ds \right]^{1/2} E_{t,x} \left[\int_t^T \left\| \nabla X_s \left[d\mu, d\sigma^2 \right] \right\|^2 ds \right]^{1/2} ,$$

and

$$\left| E_{t,x} \left[\int_t^T c \left(s, X_s^0 \right) \exp \left[- \int_t^s a \left(u, X_u^0 \right) du \right] \left(\int_t^s a_x \left(u, X_u^0 \right) \nabla X_u du \right) ds \right] \right|$$
$$\leq \quad E_{t,x} \left[\int_t^T \left\| c \left(s, X_s^0 \right) \right\|^2 \int_t^s a_x^2 \left(u, X_u^0 \right) du ds \right] E_{t,x} \left[\int_t^T \left\| \nabla X_s \left[d\mu, d\sigma^2 \right] \right\|^2 ds \right]^{1/2}$$

where

$$E_{t,x} \left[\left\| \nabla X_s \left[d\mu, d\sigma^2 \right] \right\|^2 \right] \quad \leq \quad \int_t^T E_{t,x} \left[\left\| d\mu \left(X_u^0 \right) \right\|^2 \right] du + \int_t^T E_{t,x} \left[\left\| d\sigma^2 \left(X_u^0 \right) \right\|^2 \right] du$$
$$= \quad \left\| \left(d\mu, d\sigma^2 \right) \right\|_{0,2}^2 .$$

for all $s \in [t, T]$, by Lemma 12. We conclude that $\left(d\mu, d\sigma^2 \right) \mapsto \nabla \Gamma \left[d\mu, d\sigma^2 \right]$ is a linear, continuous functional. ∎

Proof of Theorem 6. Under the (SP.0)-(SP.3) and (SP1.A), it follows from Lemma 13 that $\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1)$, and $\|\hat{\sigma} - \sigma_0\|_{0,4} = o_P(1)$. The consistency part now follows from the first part of Theorem 4. To prove asymptotic normality, we first observe from Lemma 13, that $\|\hat{\mu} - \mu_0\|_{1,4}^2 = o_P(n^{-1/2})$, and $\|\hat{\sigma}^2 - \sigma_0^2\|_{1,4}^2 = o_P(n^{-1/2})$. Thus, the asymptotic distribution of $\hat{u}$ is determined by $\nabla\Gamma$ as given in (19). We linearise μ and σ^2 w.r.t. π, $\pi^{(1)}$ and θ: Define $\nabla\mu_0 = \hat{T}\nabla_\pi\mu_0 + \nabla_\theta\mu_0$ where $\nabla_\pi\mu_0\left[d\pi, d\pi^{(1)}\right] = \nabla_0\mu_0\left[d\pi\right] + \nabla_1\mu_0\left[d\pi^{(1)}\right]$, and

$$\nabla_0\mu_0\left[d\pi\right] = -\frac{1}{2}\sigma_0^2(x)\frac{\pi_0^{(1)}(x)}{\pi_0^2(x)}d\pi(x), \quad \nabla_1\mu_0\left[d\pi^{(1)}\right] = \frac{1}{2}\sigma_0^2(x)\frac{d\pi^{(1)}(x)}{\pi_0(x)},$$

$$\nabla_\theta\mu_0\left[d\theta\right] = \dot{\mu}_0(x)\,d\theta, \quad \dot{\mu}_0(x) = \frac{1}{2}\partial_x\dot{\sigma}_0^2(x) + \frac{1}{2}\sigma_0^2(x)\frac{\pi_0^{(1)}(x)}{\pi_0(x)}$$

and $\nabla\sigma_0^2[d\theta] = \nabla_\theta\sigma_0^2[d\theta] = \dot{\sigma}_0^2(x)\,d\theta$, where $\dot{\sigma}_0^2(x) = \partial_\theta\sigma^2(x;\theta_0)$. We observe from Lemma 14 that

$$\|\hat{\mu}(\cdot;\hat{\theta}) - \mu_0 - \nabla\mu_0[\hat{\pi} - \pi_0, \hat{\theta} - \theta_0]\|_{0,2} = o_P(n^{-1/2}),$$

and

$$\|\sigma^2\left(\cdot;\hat{\theta}\right) - \sigma_0^2 - \nabla\sigma_0^2[\hat{\theta} - \theta_0]\|_{0,2} = o_P(n^{-1/2}),$$

under (SP1.B). Since $\nabla\Gamma$ is linear,

$$\nabla\Gamma(t,x)\left[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2\right] = \nabla\Gamma(t,x)\left[\hat{\mu} - \mu_0 - \nabla\mu_0, \hat{\sigma}^2 - \sigma_0^2 - \nabla\sigma_0^2\right]$$
$$+\nabla\Gamma(t,x)\left[\nabla\mu_0, \nabla\sigma_0^2\right],$$

where, by the continuity of $\nabla\Gamma$, c.f. Theorem 3,

$$\nabla\Gamma(t,x)\left[\hat{\mu} - \mu_0 - \nabla\mu_0, \hat{\sigma}^2 - \sigma_0^2 - \nabla\sigma_0^2\right]$$
$$\leq C\|\hat{\mu}(\cdot;\hat{\theta}) - \mu_0 - \nabla\mu_0[\hat{\pi} - \pi_0, \hat{\theta} - \theta_0]\|_{0,2}$$
$$+C\|\hat{\sigma}^2\left(X_s^0\right) - \sigma_0^2\left(X_s^0\right) - \dot{\sigma}^2\left(X_s^0\right)(\hat{\theta} - \theta_0)\|_{0,2}$$
$$= o_P(n^{-1/2}).$$

By Riesz's Representation Theorem,

$$\nabla\Gamma(t,x)\left[\nabla\mu_0, \nabla\sigma_0^2\right]$$
$$= \int_t^T E_{t,x}\left[d_1^*\left(X_u^0\right)\nabla_0\mu_0\left(X_u^0\right)\right]ds + \int_t^T E_{t,x}\left[d_1^*\left(X_u^0\right)\nabla_1\mu_0\left(X_u^0\right)\right]ds$$
$$+ \int_t^T E_{t,x}\left[d_1^*\left(X_u^0\right)\nabla_\theta\mu_0\left(X_u^0\right)\right]ds + \int_t^T E_{t,x}\left[d_1^*\left(X_u^0\right)\nabla\sigma_0^2\left(X_u^0\right)\right]ds$$

for some $d^* = (d_1^*, d_2^*) \in \bar{\mathcal{D}}$. Each of the four terms in the above expression will make up a part of the influence function for $\nabla\Gamma$: First,

$$\int_t^T E_{t,x}\left[d_1^*\left(X_u^0\right)\nabla_0\mu_0\left(X_u^0\right)[\hat{\pi}]\right]ds$$
$$= -\frac{1}{2n}\sum_{i=1}^n\int_t^T\int\hat{T}(y;a)\,d_1^*(y)\,\sigma_0^2(y)\frac{\pi_0^{(1)}(y)}{\pi_0^2(y)}p_{u-t}(y|x)\frac{1}{h}K\left(\frac{y-x_i}{h}\right)dydu$$
$$= \frac{1}{n}\sum_{i=1}^n\hat{T}(x_i;a)\,\nu_1(x_i;t,x) + O_P\left(a^{-1}h^{\omega-1}\right).$$

31

where

$$\nu_1(y; t, x) = -\frac{1}{2} d_1^*(y) \sigma_0^2(y) \frac{\pi_0^{(1)}(y)}{\pi_0^2(y)} \int_t^T p_{s-t}(y|x) \, ds. \tag{42}$$

The last equality follows from the fact that for any m times continuously differentiable function with $g^{(i)}$ bounded, $0 \le i \le m$,

$$
\begin{aligned}
\int g(z) K_h(z-y) \, dz &= \int g(y+hz) K(z) \, dz \\
&= \int \left\{ g(y) + g^{(1)}(y) hz + \ldots + g^{(m)}(\bar{z}) h^m z^m \right\} K(z) \, dz \\
&= g(y) + h^m \int g^m(\bar{z}) z^m K(z) \, dz,
\end{aligned}
$$

for some $\bar{z} \in [y, y+hz]$. The result is then obtained by applying this with $g = \hat{T}\nu_1$ and $m = \omega - 1$ such that $g^{(m)} = \sum_{i=0}^m \hat{T}^{(i)} \nu_1^{(m-i)}$. The boundedness condition holds since $\hat{T}^{(i)}(x; a) = 0$, $1 \le i \le \omega - 1$, for x outside a compact interval, while $\pi_0(y) \left| \nu_1^{(\omega-1)}(y; t, x) \right| = O(1)$ such that $\hat{T}(y; a) \left| \nu_1^{(\omega-1)}(y; t, x) \right| = O(a^{-1})$. By the same arguments,

$$\int_t^T E_{t,x}\left[\hat{T}\left(X_s^0; a\right) d_1\left(X_s^0\right) \nabla_1 \mu\left(X_s^0\right) \left[\hat{\pi}^{(1)}\right] \right] ds = \frac{1}{n} \sum_{i=1}^n \hat{T}(x_i; a) \nu_2(x_i; t, x) + O_P\left(a^{-1} h^{\omega-1}\right),$$

where

$$\nu_2(y; t, x) = -\frac{1}{2} \frac{\partial}{\partial y} \left[\frac{d_1(y) \sigma_0^2(y)}{\pi_0(y)} \left\{ \int_t^T p_s(y|x) \, ds \right\} \right]. \tag{43}$$

The last two terms are easily dealt with since

$$
\begin{aligned}
\int_t^T E_{t,x}\left[d_1\left(X_s^0\right) \dot{\mu}\left(X_s^0\right) (\hat{\theta} - \theta_0) \right] ds &= \frac{1}{n} \sum_{i=1}^n \nu_3(x_i|x_{i-1}; t, x) + o_P(n^{-1/2}), \\
E_{t,x}\left[d_2\left(X_s^0\right) \dot{\sigma}\left(X_s^0\right) (\hat{\theta} - \theta_0) \right] ds &= \frac{1}{n} \sum_{i=1}^n \nu_4(x_i|x_{i-1}; t, x) + o_P(n^{-1/2}),
\end{aligned}
$$

where

$$\nu^{(3)}(x_i|x_{i-1}; t, x) = \left\{ \int_t^T E_{t,x}\left[d_1\left(X_s^0\right) \dot{\mu}_0\left(X_s^0\right) \right] ds \right\} \psi(x_i|x_{i-1}), \tag{44}$$

$$\nu^{(4)}(x_i|x_{i-1}; t, x) = \left\{ \int_t^T E_{t,x}\left[d_2\left(X_s^0\right) \dot{\sigma}_0\left(X_s^0\right) \right] ds \right\} \psi(x_i|x_{i-1}). \tag{45}$$

All together, $\nabla \Gamma(t, x) = \frac{1}{n} \sum_{i=1}^n \nu(x_i|x_{i-1}; t, x) + o_P(n^{-1/2})$ with

$$\nu(x_i|x_{i-1}; t, x) = \sum_{k=1}^2 \nu_k(x_i; t, x) + \sum_{k=3}^4 \nu_k(x_i|x_{i-1}; t, x). \tag{46}$$

∎

Proof of Theorem 7. Under (SP2.A), $\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1)$, and $\|\hat{\sigma}^2 - \sigma_0^2\|_{0,4} = o_P(1)$ by Lemma 13 which gives us consistency. From Lemma 13, we obtain $\|\hat{\mu} - \mu_0\|_{1,4}^2 = o_P(n^{-1/2})$, and

$\|\hat{\sigma} - \sigma_0\|_{1,4}^2 = o_P(n^{-1/2})$ under (SP2.1)-(SP2.4) and (SP2.B). We linearise μ and σ w.r.t. π and θ: Define $\nabla\sigma_0^2[d\pi, d\theta] = \hat{T}\nabla_\pi\sigma_0^2[d\pi] + \nabla_\theta\sigma_0^2[d\theta]$, where

$$\nabla_\pi\sigma_0^2[d\pi] = -\frac{d\pi(x)}{2\pi_0^2(x)} \int_l^x \mu_0(y)\,\pi_0(y)\,dy,$$

$$\nabla\sigma_0^2[d\theta] = \dot{\sigma}_0^2(x)\,d\theta, \quad \dot{\sigma}_0^2(x) = \frac{1}{2\pi_0(x)}\int_l^x \dot{\mu}_0(y)\,d\pi(y)\,dy,$$

and $\nabla\mu_0[d\theta] = \nabla_\theta\mu_0[d\theta] = \dot{\mu}_0(x)\,d\theta$, where $\dot{\mu}_0(x) = \partial_\theta\mu(x;\theta_0)$. By standard Taylor expansions,

$$\|\mu(\cdot;\hat{\theta}) - \mu_0 - \nabla\mu_0[\hat{\theta} - \theta_0]\|_{1,4} = o_P(n^{-1/2}),$$

while by Lemma 14,

$$\|\hat{\sigma}^2 - \sigma_0^2 - \nabla\sigma_0^2[\hat{\pi} - \pi_0, \hat{\theta} - \theta_0]\|_{1,4} = o_P(n^{-1/2}),$$

under (SP2.B). Thus, by linearity and continuity of $\nabla\Gamma$,

$$\begin{aligned}
\nabla\Gamma(t,x)[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2] &= \int_t^T E_{t,x}[d_1^*(X_s^0)\,\dot{\mu}(X_s^0)]\,ds(\hat{\theta} - \theta_0) \\
&+ \int_t^T E_{t,x}[d_2^*(X_s^0)\,\dot{\sigma}(X_s^0)]\,ds(\hat{\theta} - \theta_0) \\
&+ \int_t^T E_{t,x}[d_2^*(X_s^0)\,\nabla_\pi\sigma^2(X_s^0)[\hat{\pi} - \pi_0]]\,ds \\
&+ o_P(n^{-1/2}),
\end{aligned}$$

where d^* is given by the Riesz Representation Theorem. Proceeding as in the proof of Theorem 6, we obtain

$$\int_t^T E_{t,x}[d_2^*(X_s^0)\,\nabla_\pi\sigma^2(X_s^0)[\hat{\pi}]]\,ds = \frac{1}{n}\sum_{i=1}^n \nu_1(x_i;t,x) + o_P(n^{-1/2})$$

where

$$\nu_2(y;t,x) = d_2^*(y)\frac{1}{2\pi_0^2(y)}\int_l^y \mu_0(y)\,\pi_0(y)\,dy\int_t^T p_s(y|x)\,ds. \tag{47}$$

The derivatives w.r.t. θ have the following influence function,

$$\nu_2(y|z;t,x) = \left\{\int_t^T E_{t,x}[d_1^*(X_s^0)\,\dot{\mu}(X_s^0)]\,ds + \int_t^T E_{t,x}[d_2^*(X_s^0)\,\dot{\sigma}(X_s^0)]\,ds\right\}\psi(z;y) \tag{48}$$

In total, $\nabla u(t,x) = \frac{1}{n}\sum_{i=1}^n \nu(x_i|x_{i-1};t,x) + o_P(n^{-1/2})$, where

$$\nu(x_i|x_{i-1};t,x) = \nu_1(x_i;t,x) + \nu_2(x_i|x_{i-1};t,x) \tag{49}$$

∎

Proof of Theorem 8. Under (NP.1)-(NP.4) and (NP.6a), $\hat{\mu}$ and $\hat{\sigma}^2$ are consistent on I in the $\|\cdot\|_{0,4}$-norm, c.f. Lemma 15. This proves consistency of $\hat{u}$.

To derive the asymptotic distribution of $\hat{u}$, we proceed as in the previous two proofs. Applying Lemma 16, $\|\hat{\mu} - \mu_0\|_{1,4}^2 = o_P(\bar{T}^{-1/2})$, and $\|\hat{\sigma}^2 - \sigma_0^2\|_{1,4}^2 = o_P(n^{-1/2})$. So as before, the asymptotic distribution of $\hat{u}$ is determined by $\nabla\Gamma[d\mu, d\sigma^2]$ as given in (19). We define $\nabla\mu_0(x)[dm, d\pi] =$

$\hat{T}\{\nabla_m\mu_0(x)[dm] + \nabla_\pi\mu_0(x)[d\pi]\}$ and $\nabla\sigma_0^2(x)[ds, d\pi] = \hat{T}\{\nabla_s\sigma_0^2(x)[d\tau] + \nabla_\pi\sigma_0^2(x)[d\pi]\}$ where

$$\nabla_m\mu_0(x)[dm] = \frac{1}{\pi_0(x)}dm(x), \quad \nabla_\pi\mu_0(x)[d\pi] = -\frac{m_0(x)}{\pi_0^2(x)}d\pi(x) \tag{50}$$

$$\nabla_s\sigma_0^2(x)[d\tau] = \frac{1}{\pi_0(x)}ds(x), \quad \nabla_\pi\sigma_0^2(x)[d\pi] = -\frac{s_0(x)}{\pi_0^2(x)}d\pi(x). \tag{51}$$

Applying Lemma 16 once more, it easily follows that

$$\|\hat{\mu} - \mu_0 - \nabla\mu_0[\hat{m} - m_0, \hat{\pi} - \pi_0]\|_{0,2} = o_P(\bar{T}^{-1/2}),$$
$$\|\hat{\sigma}^2 - \sigma_0^2 - \nabla\sigma_0^2[\hat{s} - s_0, \hat{\pi} - \pi_0]\|_{0,2} = o_P(n^{-1/2}).$$

Thus, by linearity and continuity of $\nabla\Gamma$ together with Riesz's Representation Theorem,

$$\nabla\Gamma(t,x)[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2]$$
$$= \int_t^T E_{t,x}\left[d_1^*(X_u^0)\nabla\hat{\mu}_0(X_u^0)[\hat{m} - m_0, \hat{\pi} - \pi_0]\right]du$$
$$+ \int_t^T E_{t,x}\left[d_2^*(X_u^0)\nabla\sigma_0^2(X_u^0)[\hat{s} - s_0, \hat{\pi} - \pi_0]\right]du + o_P(\bar{T}^{-1/2}).$$

We now show that this expression converges towards the claimed distribution:

$$\int_t^T E_{t,x}\left[\hat{T}(X_u^0; a)\,d_1^*(X_u^0)\,\nabla_m\hat{\mu}_0(X_u^0)[\hat{m}]\right]du$$
$$= n^{-1}\sum_{i=1}^n \hat{T}(x_i; a)\frac{x_{i+1} - x_i}{\Delta}d_1^*(x_i)\frac{1}{\pi_0(x_i)}\int_t^T p_u(x_i|x)\,du + o_P(\bar{T}^{-1/2});$$

$$\int_t^T E_{t,x}\left[\hat{T}(X_u^0; a)\,d_1^*(X_u^0)\,\nabla_\pi\hat{\mu}_0(X_u^0)[\hat{\pi}]\right]du$$
$$= -n^{-1}\sum_{i=1}^n \hat{T}(x_i; a)\,d_1^*(x_i)\frac{\mu_0(x_i)}{\pi_0(x_i)}\int_t^T p_u(x_i|x)\,du + o_P(\bar{T}^{-1/2});$$

$$\int_t^T E_{t,x}\left[\hat{T}(X_u^0; a)\,d_2^*(X_u^0)\,\nabla\hat{\sigma}_0^2(X_u^0)[\hat{s}]\right]du$$
$$= n^{-1}\sum_{i=1}^n \hat{T}(x_i; a)\frac{(x_{i+1} - x_i)^2}{\Delta}\frac{d_2^*(x_i)}{\pi_0(x_i)}\int_t^T p_u(x_i|x)\,du + o_P(n^{-1/2});$$

$$\int_t^T E_{t,x}\left[\hat{T}(X_u^0; a)\,d_2^*(X_u^0)\,\nabla\hat{\sigma}_0^2(X_u^0)[\hat{\pi}]\right]du$$
$$= -n^{-1}\sum_{i=1}^n \hat{T}(x_i; a)\frac{d_2^*(x_i)\,\sigma_0^2(x_i)}{\pi_0(x_i)}\int_t^T p_u(x_i|x)\,du + o_P\left(n^{-1/2}\right);$$

In both cases, the trimming is asymptotically negligible so in total,

$$\nabla\Gamma(t,x)[\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2] = n^{-1}\sum_{i=1}^n\left\{\frac{x_{i+1} - x_i}{\Delta} - \mu_0(x_i)\right\}\bar{d}_1^*(x_i)$$
$$+ n^{-1}\sum_{i=1}^n\left\{\frac{(x_{i+1} - x_i)^2}{\Delta} - \sigma_0^2(x_i)\right\}\bar{d}_2^*(x_i) + o_P(n^{-1/2}).$$

where

$$\bar{d}_i^*(y; t, x) = d_i^*(y) \frac{1}{\pi_0(y)} \int_t^T p_u(y|x) \, du.$$

Using same arguments as employed in the proof of Lemma 15,

$$\frac{\sqrt{\Delta}}{\sqrt{n}} \sum_{i=1}^n \bar{d}_1^*(x_i) \left\{ \frac{x_{i+1} - x_i}{\Delta} - \mu(x_i) \right\} = \frac{1}{\sqrt{\bar{T}}} \int_0^{\bar{T}} \bar{d}_1^*(x_s) \sigma(x_s) \, dW_s + o_P(1),$$

$$\frac{\sqrt{\Delta}}{\sqrt{n}} \sum_{i=1}^n \bar{d}_2^*(x_i) \left\{ \frac{(x_{i+1} - x_i)^2}{\Delta} - \sigma_0^2(x_i) \right\} = o_P(1).$$

Thus,

$$\sqrt{\bar{T}} \nabla \Gamma(t, x) [\hat{\mu} - \mu_0, \hat{\sigma}^2 - \sigma_0^2] = \frac{1}{\sqrt{\bar{T}}} \int_0^{\bar{T}} \sigma_0(x_s) \bar{d}_1^*(x_s; t, x) \, dW_s + o_P(1)$$

where the leading term weakly converges towards a normal distribution with mean zero and variance $V(x, t) = E\left[\sigma_0^2(x_0) \bar{d}_1^2(x_0; t, x)\right]$. ∎

B Auxiliary Lemmas

Lemma 9 *For any $(\mu, \sigma^2) \in \mathcal{D}$,*

$$E_s\left[\left\|X_t - X_t^0\right\|^2\right] \leq \left\|X_s - X_s^0\right\|^2 + 6 \int_s^T E_s\left[\left\|\mu\left(X_u^0\right) - \mu_0\left(X_u^0\right)\right\|^2\right] du$$

$$+ 6 \int_s^T E_s\left[\left\|\sigma^2\left(u, X_u^0\right) - \sigma_0^2\left(u, X_u^0\right)\right\|^2\right] du$$

for $s \leq t \leq T$.

Lemma 10 *For any $(\mu, \sigma^2) \in \mathcal{D}$,*

1. *For any integer $p \geq 1$, there exists constant $C(p)$ such that*

$$E_s\left[\left\|X_t - X_t^0\right\|^{2p}\right]$$

$$\leq \left\|X_s - X_s^0\right\|^{2p} + C(m) \int_s^T E_s\left[\left\|\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right\|^{2p}\right] du$$

$$+ C(m) \int_s^T E_s\left[\left\|\sigma\left(u, X_u^0\right) - \sigma_0\left(u, X_u^0\right)\right\|^{2p}\right] du$$

2. *If $f : [0, T] \times \mathbb{R}^q \mapsto \mathbb{R}$ satisfies $\|f(t, x) - f(t, y)\| \leq C(1 + \|x\|^p + \|y\|^p)(\|x - y\|)$ then*

$$E_{x,s}\left[\left\|f(t, X_t) - f\left(t, X_t^0\right)\right\|\right]$$

$$\leq C(t, x) \left(\int_s^t E_{x,s}\left[\left\|\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right\|^2\right] du \right.$$

$$\left. \int_s^T E_{x,s}\left[\left\|\sigma^2\left(u, X_u^0\right) - \sigma_0^2\left(u, X_u^0\right)\right\|^2\right] du \right)^{1/2},$$

where $C(t, x) = Ct(1 + \|x^p\|)$.

Lemma 11 *For any $\left(\mu, \sigma^2\right) \in \mathcal{D}$, the $\mathbb{R}^q$-valued diffusion process*

$$dY_t^{(i)} = \mu^{(1)}\left(t, X_t^{s,x}\right) Y_t^{(i)} dt + Y_t^{(i)} \sigma^{(1)}\left(t, X_t^{s,x}\right) dW_t, \quad Y_s^{(i)} = e_i, \tag{52}$$

where $e_i = \{e_{ij}\}$ with $e_{ij} = 0$ for $i \neq j$ and $e_{ii} = 1$, exists and $Y_t^{(i)} = \partial X_t^{s,x}/\partial x_i$ in the L_2-sense.
If furthermore μ and σ^2 are twice continuously differentiable and satisfy

$$\|\partial_x^\alpha \mu(t, x)\| + \|\partial_x^\alpha \sigma(t, x)\| \leq C\left(1 + \|x\|\right),$$

for $|\alpha| = 2$ then $Y_t^{(i,j)} = \partial^2 X_t^{s,x}/\partial x_i \partial x_j$ also exists in the L_2-sense.

Lemma 12 *For any $\left(\mu, \sigma^2\right) \in \mathcal{D}$:*

1. *the process $\{\nabla_1 X_t\}$ given by*

$$d\nabla_1 X_t = \left\{ d\mu\left(t, X_t^0\right) + \mu_0^{(1)}\left(t, X_t^0\right) \nabla_1 X_t \right\} dt + \sigma_0^{(1)}\left(t, X_t^0\right) \nabla_1 X_t dW_t, \tag{53}$$

   and

$$d\nabla_2 X_t = \mu_0^{(1)}\left(t, X_t^0\right) \nabla_2 X_t dt + \left\{ \frac{1}{2}\sigma_0^{-1}\left(t, X_t^0\right) d\sigma^2\left(t, X_t^0\right) + \sigma_0^{(1)}\left(t, X_t^0\right) \nabla_2 X_t \right\} dW_t, \tag{54}$$

   with $d\mu = \mu - \mu_0$, satisfies

$$E_s\left[\left\|X_t - X_t^0 - \nabla X_t\right\|^2\right]$$
$$\leq \quad \left\|X_s - X_s^0\right\|^2 + \sum_{|\alpha| \leq 1} \int_s^t E_s\left[\left\|\partial_x^\alpha \mu\left(u, X_u^0\right) - \partial_x^\alpha \mu_0\left(u, X_u^0\right)\right\|^4\right] du \tag{55}$$

   where $\{X_t\}$ solves $SDE\left(\mu, \sigma_0^2\right)$ and

$$E_s\left[\left\|\nabla X_t\right\|^2\right] \leq \int_s^t E_s\left[\left\|\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right\|^2\right] du. \tag{56}$$

2. *The process $\{\nabla_2 X_t\}$ given in (54) with $d\sigma^2 = \sigma^2 - \sigma_0^2$, satisfies*

$$E_s\left[\left\|X_t - X_t^0 - \nabla_2 X_t\right\|^2\right] \tag{57}$$
$$\leq \quad \left\|X_s - X_s^0\right\|^2 + \underline{\sigma}^{-2} \sum_{|\alpha| \leq 1} \int_s^t E_s\left[\left\|\partial_x^\alpha \sigma^2\left(u, X_u^0\right) - \partial_x^\alpha \sigma_0^2\left(u, X_u^0\right)\right\|^4\right] du$$

   where $\{X_t\}$ solves $SDE\left(\mu_0, \sigma^2\right)$ and

$$E_s\left[\left\|\nabla_2 X_t\right\|^2\right] \leq \underline{\sigma}^{-2} \int_s^t E_s\left[\left\|\sigma^2\left(u, X_u^0\right) - \sigma_0^2\left(u, X_u^0\right)\right\|^2\right] du. \tag{58}$$

B.1 The Semiparametric Estimator

Lemma 13 *Under (SP.0)-(SP.3) and (SP1.A) [(SP2.A)], the estimators for (μ, σ) in Class 1 [2] satisfy*

$$\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1), \quad \|\hat{\sigma}^2 - \sigma_0^2\|_{0,4} = o_P(1).$$

If additionally (SP1.B) [(SP2.B)] then

$$\|\hat{\mu} - \mu_0\|_{1,4} = o_P\left(n^{-1/4}\right), \quad \|\hat{\sigma}^2 - \sigma_0^2\|_{1,4} = o_P\left(n^{-1/4}\right).$$

Proof. In Class 1, the convergence of the diffusion estimator follows by the arguments used in the parametric case. For the drift estimator, by Kristensen (2004a),

$$\|\hat{\mu}^{(i)}(\cdot; \theta) - \hat{T}\mu_0^{(i)}(\cdot; \theta)\|_\infty \leq C \sum_{k=0}^{i+1} a^{k-2-i} \|\hat{\pi}^{(k)} - \pi_0^{(k)}\|_\infty,$$

and

$$\|\hat{\pi}^{(k)} - \pi_0^{(k)}\|_\infty = O_P\left(n^{-1/2}h^{-1-k}\right) + O_P\left(h^{\omega-k}\right). \tag{59}$$

Furthermore, $\hat{\mu}_0^{(i)}(x) - \mu_0^{(i)}(x) = a\partial_a\hat{T}(x; \bar{a})\mu_0^{(i)}(x)$, $\bar{a} \in [0, a]$, such that

$$\int_t^T E_{t,x}\left[\|\hat{\mu}_0^{(i)}(X_s^0) - \mu_0^{(i)}(X_s^0)\|^4\right] ds$$

$$\leq a^4 \left(\int_t^T E_{t,x}\left[|\partial_a\hat{T}(x; \bar{a})|^8\right] ds\right)^{1/2} \left(\int_t^T E_{t,x}\left[\|\mu_0^{(i)}(X_s^0)\|^8\right] ds\right)^{1/2},$$

where $\int_t^T E_{t,x}\left[\|\mu_0^{(i)}(X_s^0)\|^8\right] ds < \infty$. Under (SP1.A), we obtain that $a\left(\int_t^T E_{t,x}\left[|\partial_a\hat{T}(x; \bar{a})|^8\right] ds\right)^{1/8} = o_P(1)$. Under (SP1.B), it is $o_P\left(n^{-1/4}\right)$.

The result for $\hat{\sigma}^2$ follows along the same lines, again using the results obtained in Kristensen (2004a). ∎

Lemma 14 *Under (SP.0)-(SP.3) and (SP1.A-B), the estimators for (μ, σ) in Class 1 satisfy uniformly in $\theta \in \Theta$,*

$$\|\hat{\mu} - \mu_0 - \nabla_\pi\mu[\hat{\pi} - \pi_0] - \nabla_\theta\mu[\hat{\theta} - \theta_0]\|_{0,2} = o_P(n^{-1/2}),$$
$$\|\hat{\sigma}^2 - \sigma_0^2 - \nabla_\theta\sigma^2[\hat{\theta} - \theta_0]\|_{0,2} = o_P(n^{-1/2}).$$

Under (SP.0)-(SP.3) and (SP2.A-B), the estimators for (μ, σ) in Class 2 satisfy uniformly in $\theta \in \Theta$,

$$\|\hat{\mu} - \mu_0 - \nabla_\theta\mu[\hat{\theta} - \theta_0]\|_{1,4} = o_P(n^{-1/2}),$$
$$\|\hat{\sigma}^2 - \sigma_0^2 - \nabla_\pi\sigma^2[\hat{\pi} - \pi_0] - \nabla_\theta\sigma^2[\hat{\theta} - \theta_0]\|_{1,4} = o_P(n^{-1/2}).$$

Proof. The result for the diffusion estimator in Class 1 is proved by the same arguments as the ones applied in the parametric case. For the drift estimator we have

$$\hat{\mu}^{(i)}(x; \hat{\theta}) = \hat{\mu}^{(i)}(x; \theta_0) + \partial_\theta\hat{\mu}^{(i)}(x; \bar{\theta})(\hat{\theta} - \theta_0),$$

and, by the results of Kristensen (2004a),

$$||\hat{\mu}^{(i)}\left(\cdot;\theta_0\right) - \hat{\mu}_0^{(i)}\left(\cdot;\theta_0\right) - \nabla_\pi\hat{\mu}^{(i)}\left(\cdot;\theta_0\right)\left[\hat{\pi}-\pi_0\right]||_\infty \leq C\sum_{i=0}^{k+1}a_n^{i-k-3}||\hat{\pi}^{(i)}-\pi_0^{(i)}||_\infty^2,$$

where $\nabla_\pi\hat{\mu}^{(i)}$ is the pathwise derivative w.r.t. π. The convergence of the RHS is given in (59). Using the same arguments as in the previous proof,

$$\int_t^T E_{t,x}\left[|\hat{\mu}^{(i)}\left(X_s^0;\theta_0\right) - \mu^{(i)}\left(X_s^0;\theta_0\right)|^4\right]ds = o_P\left(n^{-2}\right),$$

and

$$\int_t^T E_{t,x}\left[|\nabla\hat{\mu}_0^{(i)}\left(x;\theta_0\right)\left[\hat{\pi}-\pi_0\right] - \nabla\mu^{(i)}\left(x;\theta_0\right)\left[\hat{\pi}-\pi_0\right]|^4\right]ds = o_P\left(n^{-2}\right),$$

under (SP1.B) for $i=0,1$. Also,

$$||\partial_\theta\hat{\mu}^{(i)}\left(x;\bar{\theta}\right) - \partial_\theta\mu^{(i)}\left(x;\theta_0\right)|| \leq ||\partial_\theta\hat{\mu}^{(i)}\left(x;\bar{\theta}\right) - \partial_\theta\mu^{(i)}\left(x;\bar{\theta}\right)|| + ||\partial_\theta^2\mu^{(i)}\left(x;\bar{\theta}\right)||||\hat{\theta}-\theta_0||$$

$$||\partial_x^i\hat{\sigma}^2 - \partial_x^i\sigma_0^2 - \partial_x^i\dot{\sigma}_0^2(\hat{\theta}-\theta_0)|| \leq ||\partial_x^i\ddot{\sigma}^2\left(x;\bar{\theta}\right)||\,||\hat{\theta}-\theta_0||^2 \leq ||\partial_x^i\bar{\ddot{\sigma}}^2\left(x\right)||||\hat{\theta}-\theta_0||^2,$$

Given the assumptions, we see that the terms on the the right hand side of the above inequalities are $o_P(n^{-1/2})$.

The results for Class 2 follow along the same lines applying results of Kristensen (2004a). ∎

B.2   The Nonparametric Estimator

Lemma 15 *Under (NP.0)-(NP.4),*

$$\begin{aligned}
\sup_{x\in I}\left|\hat{\pi}^{(k)}\left(x\right) - \pi_0^{(k)}\left(x\right)\right| &= O_P(A_{n,T}h^{-2k}) + O_P(\bar{T}^{-1/2}h^{-1-2k}) + O_P\left(h^{\omega-k}\right), \\
\sup_{x\in I}\left|\hat{m}^{(k)}\left(x\right) - m_0^{(k)}\left(x\right)\right| &= O_P(A_{n,T}h^{-2k}) + O_P(\bar{T}^{-1/2}h^{-1-2k}) + O_P\left(h^{\omega-k}\right), \\
\sup_{x\in I}\left|\hat{s}^{(k)}\left(x\right) - s_0^{(k)}\left(x\right)\right| &= O_P(A_{n,T}h^{-2k}) + O_P(\bar{T}^{-1/2}h^{-1-2k}) + O_P\left(h^{\omega-k}\right)
\end{aligned}$$

for $k=0,1$, where for some $\delta > 0$,

$$A_{n,T} = \bar{T}^{1/2+\delta}\sqrt{\log_2\left(\bar{T}\right)}h^{-2}\Delta^{3/4}\log\left(\Delta^{-1}\right)^{1/4}.$$

Proof. Define $\tilde{\pi}\left(x\right) = \bar{T}^{-1}\int_0^{\bar{T}}K_h\left(x_s - x\right)ds$. We then first show that $\sup_{x\in\mathbb{R}}|\hat{\pi}\left(x\right) - \tilde{\pi}\left(x\right)| = O_P\left(A_{n,T}\right)$. It holds that

$$\begin{aligned}
E\left[|\hat{\pi}\left(x\right) - \tilde{\pi}\left(x\right)|\right] &\leq \frac{1}{n}\sum_{i=1}^n\int_{(i-1)\Delta}^{i\Delta}E\left[|K_h\left(x_i-x\right) - K_h\left(x_s-x\right)|\right]ds \\
&\leq C\frac{\Delta\kappa_{n,\bar{T}}}{h^2}
\end{aligned}$$

where $\kappa_{n,\bar{T}} \equiv \max_i\sup_{s\in[i\Delta,(i+1)\Delta]}|x_s - x_i| = O_P(\Delta^{1/2}\sqrt{\log\left(\Delta^{-1}\right)})$ by Levy's modulus of continuity, c.f. Karatzas and Shreve (1991, Theorem 9.25), and where the RHS does not depend on x. We then

use the same idea as in Bosq (1998, p. 50): Define $B_n = \left\{ x : |x| \leq \bar{T}^\gamma \right\}$, for some $\gamma > 0$, and a covering of B_n,

$$B_{i,n} = \left\{ x : |x - x_{i,n}| \leq \frac{\bar{T}^\gamma}{M_{n,\bar{T}}} \right\}, \quad i = 1, ..., M_{n,\bar{T}}.$$

We have for $x \in B_{i,n}$,

$$|\hat{\pi}(x) - \tilde{\pi}(x)| \leq |\hat{\pi}(x) - \hat{\pi}(x_{i,n})| + |\tilde{\pi}(x) - \tilde{\pi}(x_{i,n})| + |\hat{\pi}(x_{i,n}) - \tilde{\pi}(x_{i,n})|,$$

where

$$|\hat{\pi}(x) - \hat{\pi}(x_{i,n})| \leq C \frac{\bar{T}^\gamma}{h^2 M_{n,\bar{T}}}, \quad |\tilde{\pi}(x) - \tilde{\pi}(x_{i,n})| \leq C \frac{\bar{T}^\gamma}{h^2 M_{n,\bar{T}}}.$$

Thus,

$$A_{n,T}^{-1} \sup_{|x| \leq \bar{T}^\gamma} |\hat{\pi}(x) - \tilde{\pi}(x)| \leq 2C \frac{A_{n,T}^{-1} \bar{T}^\gamma}{h^2 M_{n,\bar{T}}} + A_{n,T}^{-1} \max_{i=1,...,M_{n,\bar{T}}} |\hat{\pi}(x_{i,n}) - \tilde{\pi}(x_{i,n})|$$

with

$$
\begin{aligned}
P(A_{n,T}^{-1} \max_{i=1,...,M_{n,\bar{T}}} |\hat{\pi}(x_{i,n}) - \tilde{\pi}(x_{i,n})| > \varepsilon) &\leq \sum_{i=1}^{M_{n,\bar{T}}} P\left(A_{n,T}^{-1} |\hat{\pi}(x_{i,n}) - \tilde{\pi}(x_{i,n})| > \varepsilon \right) \\
&\leq CM_{n,\bar{T}} A_{n,T}^{-1} \frac{\Delta \kappa_{n,\bar{T}}}{\varepsilon h^2}.
\end{aligned}
$$

We choose $M_{n,\bar{T}} = \left[A_{n,\bar{T}}^{-1} \bar{T}^\gamma h^{-2} \log_2 (\bar{T}) \right] + 1$, and using arguments similar to the ones in Bosq (1998, p. 52), we then obtain $\sup_{x \in \mathbb{R}} |\hat{\pi}(x) - \tilde{\pi}(x)| = O_P(A_{n,T})$ if

$$A_{n,\bar{T}}^{-2} \bar{T}^{1+\delta} h^{-4} \log_2 (\bar{T}) \Delta \kappa_{n,\bar{T}} = O(1)$$

for some $\delta > 0$. Finally, applying the same arguments as Bosq (1998, Corollary 4.6), it holds

$$\sup_{x \in \mathbb{R}} |\tilde{\pi}(x) - \pi_0(x)| = O_P \left(\bar{T}^{-1/2} h^{-1} \right) + O_P (h^\omega).$$

The proof of that $\sup_{x \in \mathbb{R}} \left| \hat{\pi}^{(1)}(x) - \tilde{\pi}^{(1)}(x) \right| = o_P \left(\bar{T}^{-1/4} \right)$ is shown in a similar manner under the condition that $\bar{T}^{3+\delta} \log_2 (\bar{T})^2 h^{-12} \Delta^3 \log (\Delta^{-1}) \to 0$ for some $\delta > 0$, while

$$\sup_{x \in \mathbb{R}} |\tilde{\pi}^{(1)}(x) - \pi_0^{(1)}(x)| = O_P(\bar{T}^{-1/2} h^{-3}) + O_P \left(h^{\omega-1} \right).$$

The remaining two claims are shown along the same lines, see e.g. Bandi and Phillips (2003). We briefly sketch the proofs. First we show that $\hat{m}(x) - \tilde{m}(x) = o_P(\bar{T}^{-1/4})$, where $\tilde{m}(x)$ is given by $\tilde{m}(x) = \bar{T}^{-1} \int_0^{\bar{T}} K_h(x_s - x) \mu(x_s) ds$. We have

$$\frac{x_{i+1} - x_i}{\Delta} - \mu(x_i) = \Delta^{-1} \int_{i\Delta}^{(i+1)\Delta} \{\mu(x_s) - \mu(x_i)\} ds + \Delta^{-1} \int_{i\Delta}^{(i+1)\Delta} \sigma(x_s) dW_s, \tag{60}$$

such that

$$
\begin{aligned}
\frac{1}{n\Delta} \sum_{i=1}^n K_h(x_i - x) \int_{i\Delta}^{(i+1)\Delta} |\mu(x_s) - \mu(x_i)| ds &\leq \frac{\kappa_{n,\bar{T}}}{n} \sum_{i=1}^n K_h(x_i - x) \left| \mu^{(1}(x_i + o_P(1)) \right| \\
&= O_P \left(\frac{\kappa_{n,\bar{T}}}{h} \right),
\end{aligned}
$$

39

and

$$\left| n^{-1} \sum_{i=1}^{n} K_h \left(x_i - x \right) \mu \left(x_i \right) - T^{-1} \int_0^{\bar{T}} K_h \left(x_s - x \right) \mu \left(x_s \right) ds \right|$$

$$\leq \quad \bar{K} \frac{\kappa_{n,\bar{T}}}{h^2} \bar{T}^{-1} \int_0^{\bar{T}} \left| \mu \left(x_s \right) \right| ds + \bar{K} \frac{\kappa_{n,\bar{T}}}{h} \bar{T}^{-1} \int_0^{\bar{T}} \left| \mu^{(1} \left(x_s \right) \right| ds$$

$$= \quad O_P \left(\frac{\kappa_{n,\bar{T}}}{h^2} \right).$$

where the bound does not depend on x. Next,

$$E \left[\Delta^{-1} n^{-1} \sum_{i=1}^{n} \int_{i\Delta}^{(i+1)\Delta} \sigma \left(x_s \right) \left[K_h \left(x_i - x \right) - K_h \left(x_s - x \right) \right] dw_s \right] = 0,$$

and

$$E[| \Delta^{-1} n^{-1} \sum_{i=1}^{n} \int_{i\Delta}^{(i+1)\Delta} \sigma \left(x_s \right) \left[K_h \left(x_i - x \right) - K_h \left(x_s - x \right) \right] dw_s |^2] \leq C \frac{\kappa_{n,\bar{T}}^2}{h^4} \bar{T}^{-1} E \left[\sigma^2 \left(x_s \right) \right]$$

The process $S_{\bar{T}} \left(x \right) = \bar{T}^{-1} \int_0^{\bar{T}} K_h \left(\frac{x_s - x}{h} \right) \sigma \left(x_s \right) dw_s$ has mean zero and variance

$$\text{var} \left[S_{\bar{T}} \left(x \right) \right] = \left(\bar{T} h \right)^{-1} \left(\sigma^2 \left(x \right) \pi \left(x \right) \int K^2 \left(z \right) dz + o \left(1 \right) \right).$$

Hence, $S_{\bar{T}} \left(x \right) = O_P(1/\sqrt{\bar{T} h})$. We may now extend this to uniform convergence. Finally, using same arguments as in Bosq (1998, Section 4.3.1), we obtain that

$$\sup_{x \in \mathbb{R}} \left| \tilde{m} \left(x \right) - m_0 \left(x \right) \right| = O_P \left(\bar{T}^{-1/2} h^{-1/2} \right) + O_P \left(h^{\omega} \right).$$

To prove $\sup_{x \in \mathbb{R}} |\hat{s} \left(x \right) - s_0 \left(x \right)| = o_P(\bar{T}^{-1/4})$, we first apply Itô's Lemma on (60) to obtain

$$\frac{\left(x_{i+1} - x_i \right)^2}{\Delta} - \sigma^2 \left(x_i \right) = \frac{2}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \mu \left(x_s \right) \left(x_s - x_i \right) ds + \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \sigma^2 \left(x_s \right) - \sigma^2 \left(x_i \right) ds$$

$$+ \frac{2}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \sigma \left(x_s \right) \left(x_s - x_i \right) dw_s,$$

and by using similar arguments as before,

$$\frac{2}{\Delta n} \sum_{i=1}^{n} K_h \left(x_i - x \right) \int_{i\Delta}^{(i+1)\Delta} \mu \left(x_s \right) \left(x_s - x_i \right) ds = O_P \left(\frac{\kappa_{n,\bar{T}}^2}{h^2} \right),$$

$$\frac{2}{\Delta n} \sum_{i=1}^{n} K_h \left(x_i - x \right) \int_{i\Delta}^{(i+1)\Delta} \sigma^2 \left(x_s \right) - \sigma^2 \left(x_i \right) ds = O_P \left(\frac{\kappa_{n,\bar{T}}}{h^2} \right),$$

$$n^{-1} \sum_{i=1}^{n} K_h \left(x_i - x \right) \sigma^2 \left(x_i \right) - T^{-1} \int_0^{\bar{T}} K_h \left(x_s - x \right) \sigma^2 \left(x_s \right) ds = O_P \left(\frac{\kappa_{n,\bar{T}}}{h^2} \right),$$

and

$$T^{-1} \int_0^{\bar{T}} K_h \left(x_s - x \right) \sigma^2 \left(x_s \right) ds - \sigma^2 \left(x \right) = O_P(\bar{T}^{-1/2} h^{-1/2}) + O_P \left(h^{\omega} \right).$$

40

The remaining variance term, $S_n(x) = n^{-1}\Delta^{-1}\sum_{i=1}^{n} s_{i,n}(x)$ with

$$s_{i,n}(x) = 2K_h(x_i - x)\int_{i\Delta}^{(i+1)\Delta}\sigma(x_s)(x_s - x_i)\,dw_s$$

defines a martingale, and we obtain

$$\operatorname{var}[S_n(x)] = \frac{\sigma^4(x)\pi(x) + o(1)}{nh} + O\left(\frac{\Delta^{1/2}\sqrt{\log(\Delta^{-1})}}{h^2}\right).$$

Hence, $S_n(x) = O_P\left(1/\sqrt{nh}\right)$. $\blacksquare$

Lemma 16   1. *Under (NP.0)-(NP.5A),*

$$\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1), \quad \|\hat{\sigma}^2 - \sigma_0^2\|_{0,4} = o_P(1).$$

2. *Under (NP.0)-(NP.5B),*

$$\|\hat{\mu} - \mu_0\|_{1,4} = o_P(\bar{T}^{-1/4}), \quad \|\hat{\sigma}^2 - \sigma_0^2\|_{1,4} = o_P(n^{-1/4}),$$

$$\|\hat{\mu} - \mu_0 - \nabla\mu\|_{0,2} = o_P(\bar{T}^{-1/4}), \quad \|\hat{\sigma}^2 - \sigma_0^2 - \nabla\sigma^2\|_{0,2} = o_P(n^{-1/4}),$$

   where

$$\partial_x\hat{\mu} = \hat{T}(x;a)\left\{\frac{\hat{m}^{(1)}}{\hat{\pi}} - \frac{\hat{m}}{\hat{\pi}}\frac{\hat{\pi}^{(1)}}{\hat{\pi}}\right\}, \quad \partial_x\hat{\sigma}^2 = \hat{T}(x;a)\left\{\frac{\hat{s}^{(1)}}{\hat{\pi}} - \frac{\hat{s}}{\hat{\pi}}\frac{\hat{\pi}^{(1)}}{\hat{\pi}}\right\}.$$

Proof. We have

$$E_{t,x}\left[\int_t^T\left|\hat{\mu}^{(i)}(X_s^0) - \mu_0^{(i)}(X_s^0)\right|^4 ds\right]^{1/4} \leq \sup_{x\in\mathbb{R}}\hat{T}(x;a)\left|\hat{\mu}^{(i)}(x) - \mu_0^{(i)}(x)\right|$$

$$+ aE\left[\int_t^T|\partial_a\hat{T}(X_s^0);a|^4\left|\mu_0(X_s^0)\right|^4 ds\right]^{1/4},$$

where

$$E_{t,x}\left[\int_t^T|\partial_a\hat{T}(X_s^0);a|^4|\mu_0^{(i)}(X_s^0)|^4 ds\right]^{1/4}$$

$$\leq E_{t,x}\left[\int_t^T|\partial_a\hat{T}(X_s^0);a|^8 ds\right]^{1/8} E_{t,x}\left[\int_t^T|\mu_0^{(i)}(X_s^0)|^8 ds\right]^{1/8}.$$

It holds that

$$\sup_{x\in\mathbb{R}}\hat{T}(x;a)\left|\hat{\mu}(x) - \mu_0(x)\right| \leq \sup_{x\in\mathbb{R}}\hat{T}(x;a)\left|\frac{\hat{m}(x)}{\hat{\pi}(x)} - \frac{m_0(x)}{\pi_0(x)}\right|$$

$$\leq a^{-1}\|\hat{m} - m_0\|_\infty + a^{-2}\|\hat{\pi} - \pi_0\|_\infty,$$

and

$$\sup_{x\in\mathbb{R}}\hat{T}(x;a)\left|\hat{\mu}^{(1)}(x) - \mu_0^{(1)}(x)\right|$$

$$\leq a^{-2}\|\hat{m} - m_0\|_\infty + a^{-1}\|\hat{m}^{(1)} - m_0^{(1)}\|_\infty + a^{-3}\|\hat{\pi} - \pi_0\|_\infty + a^{-2}\|\hat{\pi}^{(1)} - \pi_0^{(1)}\|_\infty,$$

Using the rates of convergence established in Lemma 15, we obtain $\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1)$ and $\|\hat{\mu} - \mu_0\|_{1,4} = o_P(\bar{T}^{-1/4})$ under (NP5.A) and (NP5.B) respectively.

With $\nabla \mu$ given in (50) it holds that

$$\mu(x) - \mu_0(x) - \nabla\mu_0(x)[m - m_0, \pi - \pi_0]$$
$$= \frac{\pi(x) - \pi_0(x)}{\pi(x)\pi_0(x)}\left[[m(x) - m_0(x)] - \frac{m_0(x)}{\pi_0(x)}[\pi(x) - \pi_0(x)]\right]$$

such that

$$|\hat{\mu}(x) - \hat{\mu}_0(x) - \nabla\hat{\mu}_0(x)[\hat{m} - m_0, \hat{\pi} - \pi_0]|$$
$$\leq \hat{T}(x;a)\frac{1 + |m_0(x)|}{\hat{\pi}(x)\pi_0(x)}\left\{|\hat{m}(x) - m_0(x)|^2 + |\hat{\pi}(x) - \pi_0(x)|^2\right\}.$$

Thus,

$$\int_0^t \left|\hat{\mu}(X_s^0) - \hat{\mu}_0(X_s^0) - \nabla\hat{\mu}_0(X_s^0)[\hat{m} - m_0, \hat{\pi} - \pi_0]\right|^2 ds$$
$$\leq C\left(\int_0^t \hat{T}(X_s^0;a)\left(\frac{1 + |m_0(X_s^0)|}{\hat{\pi}(X_s^0)\pi_0(X_s^0)}\right)^2 ds\right)^{1/2} \times$$
$$\left(\int_0^t \left|\hat{m}(X_s^0) - m_0(X_s^0)\right|^2 + \left|\hat{\pi}(X_s^0) - \pi_0(X_s^0)\right|^4 ds\right)^{1/2}$$

where

$$\int_0^t \hat{T}(X_s^0;a)\left(\frac{1 + |m_0(X_s^0)|}{\hat{\pi}(X_s^0)\pi_0(X_s^0)}\right)^2 ds \leq a^{-2}\int_0^t 1 + |\mu_0(X_s^0)|^2 ds.$$

∎

References

Aït-Sahalia, Y. (1993) The Delta Method for Nonparametric Kernel Functionals. Working paper, Princeton.

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

Aït-Sahalia, Y. (2002) Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica* 70, 223-262.

Aït-Sahalia, Y. & J. Duarte (2003) Nonparametric Option Pricing under Shape Restrictions. *Journal of Econometrics* 116, 9-47.

Ames, W.F. (1992) *Numerical Methods for Partial Differential Equation*s. Academic Press, San Diego.

Andrews, D.W.K. (1995) Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory* 11, 560-596.

Bandi, F.M. & G. Moloche (2001) On the Functional Estimation of Multivariate Diffusion Processes. Working paper, MIT.

Bandi, F.M. & P.C.B. Phillips (2000) Accelerated Asymptotics for Diffusion Model Estimation. Working paper, GSB, University of Chicago.

Bandi, F.M. & P.C.B. Phillips (2003) Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* 71, 241-283.

Bibby, B. and M. Sørensen (1995) Martingale Estimating Functions for Discretely Observed Diffusion Processes. *Bernoulli* 1, 17-39.

Bierens, H.J. (1987) Kernel Estimators of Regression Functions. In T.F. Bewley (ed.) *Advances in Econometrics: Fifth World Congress* Vol. 1, 99-144. Cambridge: Cambridge University Press.

Björk, T. (1998) *Arbitrage Theory in Continuous Time*. Oxford: Oxford University Press.

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

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes*. New York: Springer-Verlag.

Boyle, P.P. and A.L. Ananthanarayanan (1977) The Impact of Variance Estimation in Option Valuation Models. *Journal of Financial Economics* 5, 375-388.

Broze, L., O. Scaillet & J.-M. Zakaïan (1998) Quasi-Indirect Inference for Diffusion Processes. *Econometric Theory* 14, 161-186.

Carrasco, M., M. Chernov, J.-P. Florens & E. Ghysels (2002) Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions. Working paper, University of Rochester.

Chacko, G. & L. Viceira (2003) Spectral GMM Estimation of Continuous-Time Processes. *Journal of Econometrics* 116, 259-292.

Chen, X., L.P. Hansen & J. Scheinkman (2000a) Shape-preserving Estimation of Diffusions. Working paper, University of Chicago.

Chen, X., L.P. Hansen & J. Scheinkman (2000b) Principal Components and the Long Run. Working paper, University of Chicago.

Chow, P.-L., I.A. Ibragimov & R.Z. Khasminskii (1999) Statistical Approach to Some Ill-Posed Problems for Linear Partial Differential Equations. *Probability Theory and Related Fields* 113, 421-441.

Duffee, G.R. (2002) Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* 57, 405-443.

Duffie, D. (1996) *Dynamic Asset Pricing Theory*, 3rd edition. New Jersey: Princeton University Press.

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

Duffie, D. & K.J. Singleton (1993) Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica* 61, 929-952.

Elerian, O., S. Chib, S. & N. Shephard (2001) Likelihood Inference for Discretely Observed Non-linear Diffusions. *Econometrica* 69, 959-993.

Evans, L.C. (1998) *Partial Differential Equations*. Providence: American Mathematical Society.

Freidlin, M. (1985) *Functional Integration and Partial Differential Equations*. New Jersey: Princeton University Press.

Friedman, A. (1964) *Partial Differential Equations of Parabolic Type*. New Jersey: Prentice-Hall.

Friedman, A. (1975) *Stochastic Differential Equations and Application*s, vol. 1. New York: Academic Press.

Garcia, R., E. Ghysels and E. Renault (2004) The Econometrics of Option Pricing. Working paper 2004s-04, CIRANO. Forthcoming in *Handbook of Financial Econometrics*.

Gysels, E., A. Harvey & E. Renault (1996) Stochastic Volatility. In *Handbook of Statistics, Vol. 14* (eds. Maddala, G. & C. Rao). Elsevier Science B.V.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P., S.N. Lahiri & J. Polzehl (1995) On Bandwidth Choice in Nonparametric Regression with Both Short- and Long-Range Dependent Errors. *Annals of Statistics* 23, 1921-1936.

Hausman, J. & W. Newey (1995) Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss. *Econometrica* 63, 1445-1476.

Härdle, W., J.D. Hart, J.S. Marron & A.B. Tsybakov (1992) Bandwidth Choice for Average Derivative Estimation. *Journal of the American Statistical Association* 87, 218-233.

Härdle, W., J.S. Marron & M.P. Wand (1990) Bandwidth Choice for Density Derivatives. *Journal of the Royal Statistical Society*, Series B, 52, 223-232.

Heath, D.C., R.A. Jarrow & A. Morton (1992) Bond Pricing and the Term Structure of Interest Rates: A New methodology for Contingent Claims Valuation. *Econometrica* 60, 77-105.

Heston, S.L. (1993) A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6, 327-343

Ho, T.S.Y. & S.-B. Lee (1986) Term Structure Movements and Pricing Interest Rate Contingent Claims. *Journal of Finance* 41, 1011-1029.

Horowitz, J. (2003) Bootstrap Methods for Markov Processes. *Econometrica* 71, 1049-1082.

Hull, J.C. (1997) *Options, Futures and Other Derivatives*. New Jersey: Prentice-Hall.

Hull, J.C. & A.D. White (1993) One-Factor Interest-Rate Models and the Valuation of Interest-Rate Derivative Securities. *Journal of Financial and Quantitative Analysis* 28, 235-254.

Jeffrey, A., D. Kristensen, O. Linton, T. Nguyen & P.C.B. Phillips (2004) Nonparametric Estimation of a Multifactor Heath-Jarrow-Morton Model: An Integrated Approach. Forthcoming in *Journal of Financial Econometrics.*

Jiang, G (1998) Nonparametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Prices of Derivative Securities. *Journal of Financial and Quantitative Analysis* 33, 465-497.

Jiang, G. & J. Knight (1997) A Nonparametric Approach to the Estimation of Diffusion Processes - with an Application to a Short-term Interest Rate Model. *Econometric Theory* 13, 615-645.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus,* 2nd Edition. New York: Springer Verlag.

Karlin, S. & H.M. Taylor (1981) *A Second Course in Stochastic Processes*. Academic Press.

Kloeden, P.E. & E. Platen (1999) *Numerical Solution of Stochastic Differential Equations* (2nd ed.). Heidelberg: Springer-Verlag.

Kristensen, D. (2004a) Estimation in Two Classes of Semiparametric Diffusion Models. Working paper, LSE.

Kristensen, D. (2004b) A Semiparametric Single-Factor Model for the Term Structure of Interest Rates. Working paper, LSE.

Lo, A. (1986) Statistical Tests of Contingent-Claims Asset-Pricing Models. *Journal of Financial Economics* 17, 143-173.

Lo, A. (1988) Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data. *Econometric Theory* 4, 231-247.

Masry, E. (1996) Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates. *Journal of Time Series Analysis* 17, 571-599.

Merton, R.C. (1976) Option Pricing when Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics* 3, 125-144.

Meyn, S.P. & R.L. Tweedie (1993) Stability of Markovian processes III: Foster-Lyapunov Criteria for Continuous-time Processes. *Advances in Applied Probability* 25, 518-548.

Ncube, M. & S. Satchell (1997) The Statistical Properties of the Black-Scholes Option Price. *Mathematical Finance* 7, 287-305

Newey, W.K. (1994a) Kernel Estimation of Partial Means and a General Variance Estimator. *Econometric Theory* 10, 233-253.

Newey, W.K., F. Hsieh & J. Robins (1992) Bias Corrected Semiparametric Estimation. Working paper, MIT.

Newey, W.K., F. Hsieh & J. Robins (2004) Twicing Kernels and a Small Bias Property of Semiparametric Estimators. *Econometrica* 72, 947-962.

Newey, W.K. & K.D. West (1987) A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703-708.

Parzen, E. (1962) On Estimation of a Probability Density Function and Mode. *Annals of Mathematical Statistics* 33, 1065-1076.

Pedersen, A.R. (1995) A New Approach to Maximum-Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics* 22, 55-71.

Powell, J.L. & T.M. Stoker (1996) Optimal Bandwidth Choice for Density-Weighted Averages. *Journal of Econometrics* 75, 291-316.

Robinson, P.M. & C. Velasco (1997) Autocorrelation-Robust Inference. In *Robust Inference* (eds. G.S. Maddala & C.R. Rao), 267-298. Amsterdam: North-Holland.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Singleton, K. (2001) Estimation of Affine Pricing Models Using the Empirical Characteristic Function. *Journal of Econometrics* 102, 111-141.

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

Stoker, T.M. (1993) Smoothing Bias in Density Derivative Estimation. *Journal of the American Statistical Association* 88, 855-863.

Sundaresan, S.M. (2000) Continuous-Time Methods in Finance: A Review and an Assessment. *Journal of Finance* 55, 1569-1622.

Vanhems, A. (2003) Nonparametric Study of Solutions of Differential Equations. Working paper, ESC Toulouse.

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