Semiparametric Estimation of Diffusion Models with Applications in Finance

Dennis Kristensen

Ph.D. Thesis
Department of Economics,
London School of Economics and Political Science

2004

UMI Number: U190353

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U190353
Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

BRITISH LIBRARY OF POLITICAL AND ECONOMIC SCIENCE

THESES

F
8237

977456

Abstract

This thesis concerns semiparametric modelling and estimation of diffusion models, and the application of these in mathematical finance. Two general classes of semiparametric scalar diffusion models are proposed, and an estimator of the drift and the diffusion function based on discrete observations with a fixed time distance in between is defined. The asymptotic properties of the estimator is derived; in particular it is shown to be consistent and asymptotically normally. These semiparametric models can be applied to the pricing of financial derivatives. We assume that preliminary estimates of the drift and diffusion term are available, and give general conditions under which implied derivative prices calculated on the basis of the estimates will be consistent, and follow a normal distribution asymptotically. In particular, we verify these conditions for the proposed semiparametric estimator. The theoretical results are applied in an empirical study of a proxy of the Eurodollar short-term interest rate. We fit a semiparametric single-factor diffusion model to a data set of daily observations of the Eurodollar rate in the period 1973-1995. The resulting estimates of the drift and diffusion exhibit nonlinearities that standard parametric models cannot capture. We test the most flexible parametric single-factor model against the semiparametric alternative, and reject the model. Furthermore, it is demonstrated that the two competing models lead to significantly different bond prices.

Contents

1 Introduction **7**

2 Estimation in Diffusion Models **11**
- 2.1 Introduction 11
- 2.2 Diffusion Processes 12
- 2.3 Estimation in Scalar Diffusion Models 14
  - 2.3.1 The Parametric Model 14
  - 2.3.2 The Nonparametric Model 16
  - 2.3.3 Two Classes of Semiparametric Models 20
- 2.4 Conclusion 21

3 Term Structure Modelling with Diffusions **22**
- 3.1 Introduction 22
- 3.2 The Arbitrage-Free Term Structure 22
- 3.3 The Multi-Factor Model 25
- 3.4 The Heath-Jarrow-Morton Model 27
- 3.5 Conclusion 28

4 Estimation in Two Classes of Semiparametric Diffusion Models **29**
- 4.1 Introduction 29
- 4.2 Framework 32
- 4.3 The Nonparametric Estimator 37
- 4.4 The Semiparametric Estimator 42
  - 4.4.1 Class 1 43
  - 4.4.2 Class 2 49
- 4.5 Semiparametric Efficiency 51
- 4.6 Implementation 53
- 4.7 A Simulation Study 54
- 4.8 Conclusion 57

4.A Proofs 58
4.B Lemmas 64
4.B.1 Class 1 67
4.B.2 Class 2 84
4.B.3 Auxiliary Lemmas 89

5 Estimation of Partial Differential Equations 104
5.1 Introduction 104
5.2 Linear Parabolic Partial Differential Equations 107
5.2.1 Applications in Finance 111
5.2.2 Estimation of Diffusion Models 115
5.3 Estimation of Partial Differential Equations 116
5.3.1 A Parametric Estimator 120
5.3.2 A Semiparametric Estimator 123
5.3.3 A Nonparametric Estimator 126
5.4 Applications 129
5.5 Conclusion 132
5.A Proofs 133
5.B Auxiliary Lemmas 142
5.B.1 The Semiparametric Estimator 147
5.B.2 The Nonparametric Estimator 149

6 A Semiparametric Single-Factor Model of the Term Structure 155
6.1 Introduction 155
6.2 Single-Factor Term Structure Models 157
6.3 Estimation of the Risk Premium 159
6.4 The Data 161
6.5 Empirical Results 165
6.5.1 Estimation of the Single-Factor Model 165
6.5.2 Implied Bond and Derivative Prices 172
6.6 Conclusion 174

References 175

List of Figures

4.1 Normal probability plot of the estimators with $n = 500$. 56
4.2 Estimated drift function with 95% confidence bands, $n = 500$. 57

6.1 The Eurodollar spot rate in levels, 1973-1995. 164
6.2 The Eurodollar spot rate in differences, 1973-1995. 164
6.3 Kernel estimates of the stationary density, π, for the full sample and the subsample. 166
6.4 Sensitivity check: Semiparametric estimates of $\sigma^2(r;\theta)$ for different bandwidths. 169
6.5 Nonparametric estimate of μ with 95% confidence bands. 170
6.6 Comparison of nonparametric and parametric estimate of μ for the full sample, 1973-1995. 171
6.7 Comparison of nonparametric and parametric estimate of μ for the subsample, 1982-1995. 171

List of Tables

6.1 Parametric specifications of the spot rate. 159
6.2 Data descriptives. 165
6.3 Unit root test results. 165
6.4 Estimates of θ and β. 167
6.5 Sensitivity check of semiparametric estimates. 168
6.6 Estimate of the market price of risk, λ. 173
6.7 Implied bond prices of the semiparametric and parametric model. 173

Preface

This thesis has been submitted to the University of London in partial fulfillment of the requirements for the Ph.D. degree in Economics at the London School of Economics and Political Sciences (LSE).

Throughout my studies at the LSE, I have received invaluable encouragement and support from my supervisor, Oliver Linton, for which I am very grateful. Oliver has been a very inspiring guide into the world of non- and semiparametric econometrics. Anders Rahbek, Richard Blundell, and Xiaohong Chen have also been a great support during my postgraduate studies. Matthias Hagmann, Antonio Mele, Peter Robinson have read various parts of the manuscript. I am grateful for their comments and suggestions.

I would like to thank fellow students and researchers at the Department of Economics and the Financial Markets Group (FMG) at the LSE for making my four years there such a great experience. Part of the work presented here was done while I was visiting the Centre for Applied Microeconometrics (CAM) and the Department of Statistics and Operations Research (ASOR) at University of Copenhagen, and the Department of Economics at Yale University. Thanks to everyone at the respective places for making these enjoyable visits. Financial support by the Danish Research Agency, the FMG and the Department of Economics at the LSE is gratefully acknowledged.

Finally, a warm thanks to my family and friends for their support.

1

Introduction

Continuous time stochastic processes are widely used in dynamic models in economics and finance to describe phenomena evolving randomly over time, see e.g. Bergstrom (1990) and Duffie (1996). In finance, these have been used in mathematical models in areas as diverse as portfolio management, term structure modelling and asset pricing theory. In asset pricing theory these have been extensively used in the derivation of pricing formulae. Since the now famous option pricing models by Black and Scholes (1973) and Merton (1973, 1976), diffusion processes have played a vital role in this part of the finance literature. The main reason for the popularity of diffusion processes is that they enjoy a number of attractive properties which facilitates the theoretical analysis of the models. In particular, one has at one's disposal the powerful tools found in stochastic calculus. With these tools, assuming that the fundamental asset prices follows a diffusion process, one can derive closed form expressions of a contingent claim using hedging and no-arbitrage arguments.

These continuous-time models are often calibrated and tested using historical data. For a review of the empirical work using continuous-time models in finance, we refer to Sundaresan (2000). The implementation of the models very often involves an estimation step where the model is calibrated to the data set at hand. Most frequently, the economic variables of interest are observed at discrete points in time, e.g. daily, weekly or monthly observations, so-called discrete observations. As a first step, one needs to choose a statistical model for the diffusion process in question. The second step will then involve finding an appropriate estimator of the (possible infinite-dimensional) unknown parameters appearing in the model. The second step can become very involved due to the fact that analytical expressions of the conditional density, moments etc. of the discretely sampled process are not available.

In this thesis, we shall be concerned with semiparametric modelling and estimation of diffusion processes given discrete observations, and the application of such estimators in finance - in particular in the modelling of the term structure. My main contributions are found in Chapter 4-6, while Chapter 2-3 are introductory. Proofs of theorems and lemmas are found in the appendices situated at the end of each chapter. The outline of the thesis is as follows:

Chapter 2: Estimation in Diffusion Models. We first introduce the class of diffusion processes and present some of basic properties of these, and then gives an overview of the literature on the estimation of diffusion models. We distinguish between three different sampling schemes (continuous record, discrete sample with shrinking time distance, and discrete sample with fixed time distance), and three types of models (parametric, semiparametric and nonparametric). We pay particular attention to non- and semiparametric estimation methods which have proliferated in recent years. As part of the chapter, we therefore give a brief introduction to kernel- and series methods which are the two main approaches used in non- and semiparametric econometrics. The main conclusion of the chapter is that while for the two first sampling schemes estimators are derived in a fairly straightforward manner, the third one is more problematic.

Chapter 3: Term Structure Modelling with Diffusions. One particular area where diffusion processes are widely used is in the modelling of the term structure of interest rates. In this chapter, we give an overview of the most important models proposed in the literature, and the implications for bond and interest rate derivative pricing are discussed. The main emphasis is on the class of single-factor models, a simple yet flexible class of interest rate models. Most of the proposed specifications in the single-factor case are fully parametric models. None of these have proved to be very good at describing the observed interest rates. In Chapter 4 we therefore propose two classes of semiparametric models which can be used to model the term structure, and in Chapter 5 demonstrate what consequences the use of fitted versions of these will have for implied bond and interest rate derivative prices.

Chapter 4: Estimation in Two Classes of Semiparametric Diffusion Models. We here set up two general classes of semiparametric scalar diffusion models, and propose an estimator of both its nonparametric and parametric part given discrete observations with fixed sampling distance. The estimator of the parametric part is a maximum-likelihood-type, while the nonparametric part is estimated using kernel methods We derive the asymptotic distribution of the estimator under regularity conditions. This is followed by a discussion of the issue of semiparametric efficiency. We propose a 1-step Newton-Raphson estimator which should reach the effiency bound. A small simulation study examines the quality of the estimator in finite sample.

Chapter 5: Estimation of Partial Differential Equations. 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. In particular, we consider the estimator proposed in Chapter 4. We demonstrate how these results can be applied in the presented examples from the asset pricing theory.

Chapter 6: A Semiparametric Single-Factor Model of the Term Structure. This chapter is an empirical study where the results of Chapter 4 and 5 are employed in the modelling and estimation of a semiparametric single-factor interest rate model. We compare the fitted semiparametric model with standard fully parametric ones: First directly, by testing the fully parametric model against the semiparametric one. Secondly, we look at how much the bond prices predicted by the competing models differ; this yields an alternative measure of the performance of the models. The fitted semiparametric model picks up nonlinearities which the fully parametric model cannot capture. This leads to a rejection of the parametric model in favour of the semiparametric model in the direct comparison of the two fitted models. Moreover, the calculated bond prices implied by the two competing models are shown to be significantly different.

The chapters can be read independently of each other. This means however that some definitions, results etc. are repeated in the different chapters. I have tried to maintain the same notation throughout the thesis, but there may be some slight differences; this should hopefully not cause any confusion. Throughout the thesis, the following notation will be used:

$\mathbb{R}^q$ - the space of q-dimensional real vectors
$\mathbb{R}^{q\times p}$ - the space of $q \times p$ dimensional real matrices
$A^\top$ - the transpose of any matrix A
$\text{int}\,(A)$- the interior of any set A
C - a generic constant
$E\,[\cdot]$ - the expectations operator
$\text{var}\,(\cdot)$ - the variance operator
$\text{cov}\,(\cdot,\cdot)$ - the covariance operator
$1_A\,(\cdot)$ - the indicator function for any set A
$\sim$ - 'is distributed as'
$\to^P$ and $\to^d$- convergence in probability and distribution respectively
$\{X_t\}$ - short for $\{X_t|0 \leq t \leq T\}$ for some $0 < T \leq +\infty$. The value of T will in most cases be evident from the context; we shall specify T when deemed necessary.
$\{x_n\}$ - short for $\{x_n|1 \leq n \leq N\}$ for some $1 \leq N \leq +\infty$. The value of N will in most cases be evident from the context; we shall specify N when deemed necessary.
For $f : \mathbb{R} \times \Theta \mapsto \mathbb{R}$, $\Theta \subseteq \mathbb{R}^d$, we shall use $\partial^{ij}_{x,\theta} f\,(x;\theta)$ to denote $\partial^i\partial^j f\,(x;\theta)\,/\partial^i x\partial^j\theta$. In some cases, we shall write $f^{(i)}\,(x;\theta)$ for $\partial^i_x f\,(x;\theta)$, $\dot{f}\,(x;\theta)$ for $\partial_\theta f\,(x;\theta)$, and $\ddot{f}\,(x;\theta)$ for $\partial^2_\theta f\,(x;\theta)$.
For $f : \mathbb{R}^d \times \Theta \mapsto \mathbb{R}$, we shall use $\partial^\alpha_x f\,(x;\theta)$ to denote $\partial^{i_1}\cdots\partial^{i_d} f\,(x;\theta)\,/\partial^{i_1}x_1\cdots\partial^{i_d}x_d$ for any d-tuple of non-negative integers, $\alpha = (i_1, ..., i_d)$.
For $f : \mathbb{R} \times \mathcal{H} \mapsto \mathbb{R}$, $\mathcal{H}$ some function space equipped with a metric, we shall use $\nabla f\,(x;h_0)\,[dh]$ to denote the pathwise derivative w.r.t. h at the point $h_0 \in \mathcal{H}$ and in

the direction $dh \in \mathcal{H}$.

2

Estimation in Diffusion Models

2.1 Introduction

We here give a brief introduction to diffusion processes, present some of their basic properties, and then give a review of the literature on estimation of diffusion models. This literature spans a period of over thirty years and covers a wide range of different topics. We shall in the following try to give an overview of the main results with emphasis on non- and semiparametric estimation of discretely observed diffusions. We have chosen to classify the results presented here along two dimensions: Along the first, we have the type of model and along the second, the type of sampling scheme. We shall focus solely on time-homogenous, stationary scalar diffusions. No formal proofs are given, but references to the relevant studies are given.

Along the model dimension we shall differentiate between the fully parametric, semiparametric, and nonparametric case. Along the sampling scheme dimension, we assume that either of the three following samples is available: A continuous record of the process, a discrete sample with decreasing time distance between observations, and a discrete sample with fixed time distance between observations. A loss of information occurs as one moves out along either of these two dimensions: Inference becomes more difficult as one moves from the first towards the third sampling scheme since less of the process has been observed. Similarly, as one moves from a fully parametric specification towards a nonparametric specification, less initial information about the observed process is available. As we shall see, the convergence rate will potentially slow down as we move out along either of the two dimensions. As an additional problem, both from a technical and practical viewpoint, it is not possible to derive explicit identifying relations for a discretely observed diffusion process except for a few specific (parametric) models. In particular, explicit expressions of (conditional) moments, and the transition density are not available in general with these being required as input in e.g. GMM and MLE. This makes it difficult to set up an estimator and show it has the desired properties, and even if so the actual implementation will require either approximate or simulation-based methods.

The above problems associated with the estimation given discrete observations can be circumvented by assuming that the time distance between observations go to zero asymptotically. In this setting, the standard strategy is first to define an estimator of the continuously sampled process, and check that this has the desired properties. Then one can normally construct a suitably discretised version of this which can be used for a discrete sample. The discretisation error due to the discrete approximation will vanish asymptotically if the time distance between observations goes to zero sufficiently fast. This approach will unfortunately lead to an asymptotic bias when one assumes a fixed positive time distance between observations.

The chapter is organised as follows: First, in Section 2, we introduce the class of diffusion processes and some of its important properties. In Section 3, we then turn to the question of estimation where we first define the data generating process, and the three sampling schemes in question: Section 3.1 deals with the fully parametric case, Section 3.2 with the nonparametric model, while the semiparametric case is treated in Section 3.3. For each type of model, we differentiate between the three sampling schemes. Finally, we conclude in Section 4.

2.2 Diffusion Processes

We consider a process $\{X_t\}$ taking values in $\mathbb{R}^q$. The process is assumed to solve a stochastic differential equation (SDE) of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t \tag{2.1}$$

defined on a filtered probability space $(\Omega, \mathcal{F}, P)$ with associated filtration $\{\mathcal{F}_t\}$. The above formulae should be read as

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s,$$

where $\{W_t\}$ is a q-dimensional standard Brownian motion. The function $\mu : [0, \infty) \times \mathbb{R}^q \mapsto \mathbb{R}^q$ is normally called the *drift term* while $\sigma^2 : [0, \infty) \times \mathbb{R}^q \mapsto \mathbb{R}^{q\times q}$ is called the *diffusion term*. The drift and diffusion term can be interpreted as the instantaneous conditional mean and variance respectively,

$$\begin{aligned}\mu(t,x) &= \lim_{\Delta\to 0} E\left[X_{t+\Delta} - X_t | X_t = x\right],\\ \sigma^2(t,x) &= \lim_{\Delta\to 0} E\left[\left(X_{t+\Delta} - X_t\right)\left(X_{t+\Delta} - X_t\right)' | X_t = x\right].\end{aligned}$$

Further introduction to and treatment of stochastic differential equations and diffusions can be found in, amongst others, Karlin and Taylor (1981), Karatzas and Shreve (1991), Rogers and Williams (1994, 1996). We shall assume that a unique weak solution to (2.1) exists; see Karatzas and Shreve (1991, Section 5.2) for a formal definition of this concept. Sufficient conditions for this to hold can be found in the references given above. In most of the following, we shall only be concerned with time-homogenous SDE's,

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{2.2}$$

where the drift and diffusion term do not depend directly on the time parameter t.

The conditional distribution of $X_{t+\Delta}$ conditional on X_t is given by the transition density $p(t+\Delta,\cdot|t,x)$. For the time-homogeneous SDE, p satisfies $p(t+\Delta,\cdot|t,x) = p_\Delta(\cdot|x)$. Only in a few special cases is it possible to obtain an analytical expression of p. Under suitable conditions, cf. Karatzas and Shreve (1991, Section 5.5) and Meyn and Tweedie (1993), there exists an invariant density π for the time homogenous model.

In the univariate case ($q = 1$), one can furthermore derive an expression for π,

$$\pi(x) = \left[M\sigma^2(x)\,s(x)\right]^{-1} \tag{2.3}$$

where $s(x) = \exp[-2\int_{x^*}^{x} \mu(y)/\sigma^2(y)\,dy]$ is the scale function for some $x^* \in I$, where $I = (l,r)$, $-\infty \leq l < r \leq +\infty$, denotes the domain of the process, and $M > 0$ is a normalising factor. If the process is initialised with $X_0 \sim \pi$, we obtain a stationary and ergodic solution to (2.2). The distribution of the stationary solution $\{X_t\}$ is denoted P_π. Observe that the relation (2.3) can be inverted to express μ (σ^2) in terms of π and σ^2 (μ):

$$\mu(x) = \frac{1}{2\pi(x)}\frac{\partial}{\partial x}\left[\sigma^2(x)\pi(x)\right], \tag{2.4}$$

$$\sigma^2(x) = \frac{2}{\pi(x)}\int_l^x \mu(y)\,\pi(y)\,dy. \tag{2.5}$$

The class of diffusion processes proves to be closed to smooth transformations. For any twice differentiable function $f : \mathbb{R}^q \mapsto \mathbb{R}$, the process $Y_t = f(X_t)$ solves

$$dY_t = L_t f(X_t)\,dt + \partial_x f(X_t)\,\sigma(t,X_t)\,dW_t \tag{2.6}$$

where L_t is the so-called *infinitesimal generator* defined by

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

see Karatzas and Shreve (1991, p. 281). This is the celebrated *Itô's Lemma.* Taking conditional expectations in (2.6), one obtain for $t \leq T$,

$$E\left[f(X_T)\rceil X_t = x\right] = \int_t^T L_t E\left[f(X_s)\,|X_t = x\right] ds.$$

Differentiating w.r.t. t on both sides of the above equality, we obtain that the function $u(t,x) = E\left[f(X_T)\,|X_t = x\right]$ solves the partial differential equation

$$-\frac{\partial u}{\partial t} = L_t(u), \quad u(T,x) = f(x). \tag{2.7}$$

A formal proof of the above can be found in Karatzas and Shreve (1991, Theorem 5.7.6). This is the simplest version of the *Feynman-Kac formula.*

2.3 Estimation in Scalar Diffusion Models

We shall in the following discuss the estimation of the drift and diffusion function in a parametric, nonparametric and semiparametric framework. Throughout this section, the data generating process $\{X_t\}$ is assumed to be stationary and solve the univariate SDE

$$dX_t = \mu_0 (X_t) \, dt + \sigma_0 (X_t) \, dW_t, \tag{2.8}$$

where $\{W_t\}$ is univariate. We shall not discuss the estimation of multivariate or nonstationary models. Estimation methods for multivariate parametric diffusion models can be found in e.g. Aït-Sahalia (2003), Bibby and Sørensen (1995), Duffie and Singleton (1993), and Broze, Scaillet and Zakaïan (1998). Bandi and Moloche (2001) and Chen, Hansen and Scheinkman (2000b) consider nonparametric kernel and sieve estimators respectively for multivariate processes, the former allowing for nonstationarity. Aït-Sahalia (2002) gives a general result for the parametric MLE of nonstationary scalar diffusions. Bandi and Phillips (2003) and Nicolau (2004) develop kernel estimators for nonstationary scalar diffusions.

The three sampling schemes we consider are the following:

CS [Continuous sample]: We have observed $\{X_t | 0 \leq t \leq T\}$ for some $0 < T < +\infty$.

DS In-fill [Discrete sample, $\Delta \to 0$]: We have observed $\{X_{i\Delta} | 0 \leq i \leq n\}$ with $T = n\Delta$, and $\Delta \to 0$.

DS Fixed [Discrete sample, $\Delta > 0$ fixed]: We have observed $\{X_{i\Delta} | 0 \leq i \leq n\}$ with $T = n\Delta$, and $\Delta > 0$ fixed.

In the two discrete sample schemes, we assume for notational simplicity that the observations are equidistant; all of the following results also hold with Δ varying across observations. The asymptotics of the estimators considered in the following will in all three schemes be based on $T \to \infty$.[1] Kutoyants (2004) gives an in-depth treatment of the first case. A comprehensive overview of the literature on the estimation of diffusion models covering all three sampling schemes can be found in Prakasa Rao (1999).

2.3.1 The Parametric Model

We consider the following model,

$$dX_t = \mu (X_t; \theta) \, dt + \sigma (X_t; \theta) \, dW_t \tag{2.9}$$

[1] In the two first sampling schemes, consistent estimators of the diffusion function can be constructed for $T < \infty$ fixed. One cannot estimate the drift function consistently however in this case why we throughout consider the case $T \to \infty$.

where $\mu(\cdot;\theta)$ and $\sigma^2(\cdot;\theta)$ are known functions up to the parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ such that $\mu(\cdot;\theta_0) = \mu_0(\cdot)$ and $\sigma^2(\cdot;\theta_0) = \sigma_0^2(\cdot)$ for some $\theta_0 \in \Theta$. In the following, we present the MLE for each of the three sampling schemes in question.

CS. In this setting, the log-likelihood conditional on the initial value is given by

$$L_T^c(\theta) = \frac{1}{T}\int_0^T \frac{\mu(X_t;\theta)}{\sigma^2(X_t;\theta)} dX_t - \frac{1}{2T}\int_0^T \frac{\mu^2(X_t;\theta)}{\sigma^2(X_t;\theta)} dt, \tag{2.10}$$

see e.g. Kutoyants (2004, Theorem 1.12). Observe that we in fact are able to extract $\sigma_0^2(\cdot)$ in a deterministic manner since the quadratic variation of the process, $\{\langle X\rangle_t\}$, defined by

$$\langle X\rangle_T = \lim_{|t_{i+1}-t_i|\to 0} \sum_{i=1}^n \left(X_{t_{i+1}} - X_{t_i}\right)^2,$$

where $0 = t_0 < t_1 \cdots < t_n = T$, satisfies $\langle X\rangle_T = \int_0^T \sigma_0^2(X_t)\,dt$. So by differentiating $\langle X\rangle_t$ w.r.t. t we are able to estimate without error $\sigma_0^2(x)$ for any $x \in \{X_t | 0 \le t \le T\}$. We substitute the parametric version of the diffusion term for $\sigma_0^2(\cdot)$ in the log-likelihood function in (2.15), and obtain the MLE as

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \left\{ \frac{1}{T}\int_0^T \frac{\mu(X_t;\theta)}{\sigma_0^2(X_t)} dX_t - \frac{1}{2T}\int_0^T \frac{\mu^2(X_t;\theta)}{\sigma_0^2(X_t)} dt \right\}.$$

Under regularity conditions (see, for example, Kutoyants (2004, Theorem 2.8), the MLE satisfies $\sqrt{T}(\hat{\theta} - \theta_0) \to^d N\left(0, I_0^{-1}\right)$ with

$$I_0 = E_\pi\left[\left(\frac{\partial_\theta \mu(X_0;\theta_0)}{\sigma_0^2(X_0)}\right)^2\right]. \tag{2.11}$$

DS In-fill. First observe that in finite sample, we can no longer determine $\sigma_0^2(\cdot)$ so we now have to estimate this. We discretise $L_T^c(\theta)$ and obtain

$$\begin{aligned} \tilde{L}_n^c(\theta) &= \sum_{i=1}^n \frac{\mu(X_{i\Delta};\theta)}{\sigma^2(X_{i\Delta};\theta)}\left(X_{(i+1)\Delta} - X_{i\Delta}\right) - \frac{\Delta}{2}\sum_{i=1}^n \frac{\mu^2(X_{i\Delta};\theta)}{\sigma^2(X_{i\Delta};\theta)} \\ &= \sum_{i=1}^n \frac{\mu(X_{i\Delta};\theta)}{\sigma^2(X_{i\Delta};\theta)}\left(X_{(i+1)\Delta} - X_{i\Delta} - \frac{\Delta}{2}\mu(X_{i\Delta};\theta)\right). \end{aligned}$$

In the special case with $\theta = \left(\alpha, \sigma^2\right)$, $\mu(x;\theta) = \mu(X_{i\Delta};\alpha)$, and $\sigma^2(x;\theta) = \sigma^2$, Yoshida (1992) first defines a preliminary estimator of σ^2, $\hat{\sigma}^2 = T^{-1}\sum_{i=1}^n \left(X_{(i+1)\Delta} - X_{i\Delta}\right)^2$ and then use this to estimate α, $\hat{\alpha} = \arg\max_{\alpha\in A} \tilde{L}_n^c\left(\alpha, \hat{\sigma}^2\right)$. Under regularity conditions, Yoshida (1992) shows that

$$\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right) \to^d N\left(0, 2\sigma_0^2\right), \quad \sqrt{T}\left(\hat{\alpha} - \alpha_0\right) \to^d N\left(0, I_0^{-1}\right),$$

where I_0 defined as in (2.11), and the two estimators are asymptotically independent. Thus, $\hat{\alpha}$ inherits the properties of the MLE given a continuous sample. Also observe that while we now cannot estimate the diffusion term without error, its estimator converges with a faster rate than the one associated with the drift term. See also Dacunha-Castelle and Florens-Zmirou (1986) and Florens-Zmirou (1989).

DS Fixed. The above discrete time estimator $\tilde{\theta}$ will be biased if $\Delta > 0$ remains fixed since the discretisation error does not vanish, cf. Florens-Zmirou (1989). To avoid this type of bias, we need to use the actual transition density of the discretely sampled process. As noted earlier, the transition density cannot be written on analytical form however, except in a few simple cases. But it proves possible still to derive the properties of the (infeasible) MLE. Aït-Sahalia (2002) shows that under regularity conditions the estimator

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L_n(\theta),$$

where

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log p_\Delta (X_{i+1} | X_i; \theta), \tag{2.12}$$

is asymptotically normally distributed,

$$\sqrt{n}(\hat{\theta} - \theta_0) \to^d N\left(0, I_0^{-1}\right)$$

where $I_0 = E\left[\partial_{\theta\theta} \log p_\Delta (X_{i+1} | X_i; \theta_0)\right]$ is the information matrix. The MLE can be calculated using numerical methods. There is a number of different methods in the literature based on either approximations of p or simulations. Approximate methods have been developed in for example Aït-Sahalia (1999, 2002) and Lo (1988), while simulation-based methods can be found in for example Nicolau (2002) and Pedersen (1995).

2.3.2 The Nonparametric Model

In this section, we present a number of nonparametric estimators proposed in the literature. Nonparametric estimators is an alternative to standard parametric ones, imposing no parametric restrictions on the statistical model. This means that the risk of misspecification is smaller; on the other hand nonparametric estimators often suffer from a slower convergence rate compared to parametric ones. Kernel and sieve estimators are predominantly used in nonparametric statistics. In the following we give a quick introduction to these two types of estimation methods. For a general introduction to nonparametric methods in econometrics, we refer to Pagan and Ullah (1999). A good introduction to kernel and sieve-methods can be found in Silverman (1986) and Chen (2004) respectively. Applications of kernel estimators to stochastic processes can be found in Bosq (1998). Here, we shall mainly focus on kernel estimators.

Assume that we wish to estimate a density function without imposing any parametric assumptions on its form. One can in this case apply a standard histogram estimator for some given binwidth. The kernel density estimator can then be seen as a generalised version of this simple estimator. These are local estimators that estimate the density at a point of interest by smoothing the observations around this point. The basic ingredients in these type of estimators are a kernel $K : \mathbb{R} \mapsto \mathbb{R}$ and a bandwidth $h > 0$. The kernel is normally assumed to satisfy $\int K(x)\,dx = 1$, $\|K\|_2^2 = \int K^2(x)\,dx < \infty$ and $\int x^2 K(x)\,dx < \infty$ as a minimum. Standard densities are normally used as kernels, e.g. the Gaussian one. We write $K_h(x) = K(x/h)/h$ in the following.

Sieve or series estimators are of a more global nature. The idea is to assume that the function of interest belongs to a known (infinite-dimensional) function space. This is then approximated by a finite dimensional function space (the sieve space) which grows dense in the function space. A density can for example be expressed in terms of its Fourier coefficients. One may then estimate a finite number of the Fourier coefficients and then use these to estimate the density itself.

Density Estimation

We first set up a nonparametric estimator of the marginal density π.

CS. Here, we estimate the marginal density π by

$$\hat{\pi}(x) = \frac{1}{T}\int_0^T K_h(X_t - x)\,dt.$$

This estimator was first proposed by Banon (1978) and Nguyen (1979). Assuming that they exist, the derivatives of the density can be estimated by

$$\hat{\pi}^{(r)}(x) = \frac{1}{Th^r}\int_0^T K_h^{(r)}(X_t - x)\,dt, \quad r \geq 1.$$

Under regularity conditions,

$$\sqrt{Th^{2r+1}}(\hat{\pi}^{(r)}(x) - \pi_0^{(r)}(x)) \to^d N\left(0, \pi_0(x)\,||K^{(r)}||_2^2\right), \tag{2.13}$$

provided $Th^{2r+1} \to \infty$ and $Th^{2r+3} \to 0$. In certain cases, the super-optimal/parametric convergence rate, $\sqrt{T}$, can be obtained, see Bosq (1998). This is special to the case of continuous sampling.

DS In-fill/Fixed. We discretise the continuous sample estimator and obtain

$$\hat{\pi}^{(r)}(x) = \frac{1}{Th^r}\sum_{i=1}^n K_h^{(r)}(X_{i\Delta} - x)\,\Delta = \frac{1}{nh^r}\sum_{i=1}^n K_h^{(r)}(X_{i\Delta} - x).$$

We use this estimator for both of the discrete sample schemes. In the infill-case, $\hat{\pi}^{(r)}(x)$ satisfies (2.13), while in the fixed time distance case,

$$\sqrt{nh^{2r+1}}(\hat{\pi}^{(r)}(x) - \pi_0^{(r)}(x)) \to^d N\left(0, \pi_0(x)\,||K^{(r)}||_2^2\right),$$

provided $nh^{2r+1} \to \infty$ and $nh^{2r+3} \to 0$, cf. Robinson (1983). Observe that here the same estimator works both for the in-fill and fixed Δ case since we only wish to obtain information about the marginal distribution, but the asymptotic properties differ in the two sampling schemes.

Drift and Diffusion Estimation

We now turn to the question of estimating the drift and diffusion function nonparametrically.

CS. As in the parametric case, the diffusion term can here be determined without any uncertainty. So again we shall only be concerned with the estimation of the drift term. We here present two kernel estimators. Banon (1978) proposed to plug the known diffusion function into (2.4) together with kernel estimators of π and $\pi^{(1)}$. One could alternatively use the following kernel regression estimator,

$$\hat{\mu}(x) = \frac{\int_0^T K_h(X_t - x)\, dX_t}{\int_0^T K_h(X_t - x)\, dt}, \tag{2.14}$$

as suggested by Geman (1979). The discretised version of this is examined by Bandi and Phillips (2003), see below. Other studies of nonparametric drift estimation are found in Geman (1980) and Pham (1981).

DS In-fill. In this setting, we also need to estimate the diffusion term. Florens-Zmirou (1993), Jiang and Knight (1997) and Bandi and Phillips (2003) considered the following kernel estimator,

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K_h(X_{i\Delta} - x)\left(X_{(i+1)\Delta} - X_{i\Delta}\right)^2}{\sum_{i=1}^n K_h(X_{i\Delta} - x)\,\Delta}. \tag{2.15}$$

Under regularity conditions,

$$\sqrt{nh}(\hat{\sigma}^2(x) - \sigma_0^2(x)) \to^d N\left(0, \frac{4\|K\|_2^2 \sigma_0^4(x)}{\pi_0(x)}\right),$$

if $nh \to \infty$ and $nh^3 \to 0$. Having obtained this, we may now estimate $\mu(x)$ as before by plugging the kernel estimator of $\sigma^2(x)$ into (2.4) together with the kernel estimator of $\pi(x)$ and $\pi^{(1)}(x)$, cf. Jiang and Knight (1997). By the functional delta-method, this estimator, $\tilde{\mu}(x)$, satisfies

$$\sqrt{Th^3}(\tilde{\mu}(x) - \mu_0(x)) \to^d N\left(0, \frac{\|K^{(1)}\|_2^2 \sigma_0^2(x)}{4\pi_0(x)}\right),$$

given $Th^3 \to \infty$ and $Th^5 \to 0$. This estimator has convergence rate $\sqrt{Th^3}$ which is slower than the one of $\hat{\sigma}^2(x)$. In fact, one can consistently estimate $\sigma^2(x)$ given observations in the interval $[0, T]$ with $0 < T < \infty$ fixed, while $\mu(x)$ can only be estimated consistently as $T \to \infty$. In this sense, $\mu(\cdot)$ is harder to estimate than $\sigma^2(\cdot)$.

Bandi and Phillips (2003) constructed a discretised version of the alternative drift estimator proposed in (2.14),

$$\hat{\mu}(x) = \frac{\sum_{i=1}^n K_h(X_{i\Delta} - x)\left(X_{(i+1)\Delta} - X_{i\Delta}\right)}{\sum_{i=1}^n K_h(X_{i\Delta} - x)\,\Delta}.$$

It can be shown that

$$\sqrt{Th}(\hat{\mu}(x) - \mu_0(x)) \to^d N\left(0, \frac{\|K\|_2^2 \sigma_0^2(x)}{\pi_0(x)}\right),$$

as $Th \to \infty$ and $Th^3 \to 0$. Again, $\hat{\mu}(x)$ has a slower convergence rate than $\hat{\sigma}^2(x)$, but faster than $\tilde{\mu}(x)$.

The discretisation bias in finite sample has been analysed in Nicolau (2003). Stanton (1997) proposed alternative kernel estimators based on higher order approximations yielding a smaller discretisation bias. On the other hand, as demonstrated in Fan and Zhang (2003), the resulting asymptotic variance of Stanton's estimator increases.

DS Fixed. In this case, a completely different approach compared to the CD and DS In-fill case has been developed. The approach is based on the infinitesimal operator of the diffusion model, and the conditional expectations operator of the sampled process. First, as shown in Hansen et al (1998), on a suitable domain $\mathcal{D}$ the infinitesimal operator L has a discrete spectrum $\{\delta_j\}$ with associated eigenfunctions $\{\psi_j\}$; we have ordered the eigenvalues such that $0 = \delta_0 \leq \delta_1 \leq ....$ Defining the conditional expectations operator A by

$$A_\Delta(f)(x) = E[f(X_\Delta)|X_0 = x],$$

one is able to realise that A also has a discrete spectrum such that

$$A_\Delta(f)(x) = \sum_{i=0}^{\infty} \exp[-\Delta\delta_j] E_\pi[f(X_0)\psi_j(X_0)]\psi_j(x),$$

cf. Chen et al (2000a,b). Next, the eigenvalues and functions can be identified in the following manner: First observe that $\delta_0 = 0$ and $\psi_0(x) \equiv 1$. The following eigenvalues then satisfies

$$\exp[-\delta_j] = \sup_{\psi \in \mathcal{D}_j} E_\pi[\psi(X_\Delta)\psi(X_0)],$$

with the eigenfunction ψ_j being the solution to the above optimisation problem. Here,

$$\mathcal{D}_j = \left\{\psi \in \mathcal{D} | E_\pi[\psi(X_0)\psi_i(X_0)] = 0,\ i = 0, ..., j-1,\ E_\pi\left[\psi^2(X_0)\right] = 1\right\}.$$

So one can calculate the eigenvalues and -functions recursively. It can also be shown that for any eigenpair (δ_j, ψ_j), $j \geq 1$, the diffusion coefficient satisfies

$$\sigma_0^2(x) = \frac{-\delta_j \int_l^x \psi_j(y)\pi_0(y)\,dy}{\psi_j'(x)\pi_0(x)}. \tag{2.16}$$

Chen et al (2000a,b) suggest the following estimation procedure: First, by replacing the expectations in the above optimisation problem by the empirical counterpart and approximating the eigenfunction space using the method of sieves, estimators of the eigenvalues and -functions can be obtained. These are plugged into (2.16), yielding a nonparametric diffusion estimator. Finally, using the relationship (2.4), an estimator of the drift can be obtained. Darolles and Gourieroux (2001), Gobet et al (2002) give further results for this nonparametric estimator.

2.3.3 Two Classes of Semiparametric Models

As an intermediate step between the fully parametric and nonparametric setting, the class of semiparametric models are situated. This is a very large class of models. Here, we shall only consider the case where one either parameterises the drift or the diffusion term leaving the other term unspecified. This gives us the following two classes of semiparametric models:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t;\theta)\,dW_t, \tag{2.17}$$

or

$$dX_t = \mu(X_t;\theta)\,dt + \sigma(X_t)\,dW_t. \tag{2.18}$$

Since the two classes of models above are nested within the fully nonparametric model, one natural way to estimate any of the two models is as follows: First, obtain nonparametric estimators of both μ and σ^2. The unspecified part is then consistently estimated by the nonparametric estimator, while the parametric part can be estimated by choosing $\hat{\theta}$ as the value of $\theta \in \Theta$ that minimises some functional metric between the fully parametric form and the preliminary nonparametric estimator. For the model in (2.17), we may then define the estimator of θ as

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\sigma}^2(X_{i\Delta}) - \sigma^2(X_{i\Delta};\theta)\right]^2,$$

where $\hat{\sigma}^2(\cdot)$ is a preliminary nonparametric estimator of $\sigma^2(\cdot)$, while for the model in (2.18), we define

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\mu}(X_{i\Delta}) - \mu(X_{i\Delta};\theta)\right]^2,$$

where $\hat{\mu}(\cdot)$ is a preliminary nonparametric estimator of $\mu(\cdot)$. The squared distance metric could of course be substituted for alternative metrics.

CS. To the author's knowledge semiparametric estimators have not been considered for this sampling scheme.

DS In-fill. The strategy proposed above has been investigated in Bandi and Phillips (2000) using the kernel estimators proposed in Bandi and Phillips (2003) as preliminary estimators. They derive the asymptotic distribution of $\hat{\theta}$ and show that it is $\sqrt{n}$- and $\sqrt{T}$-asymptotically normally distributed for models in Class 1 and 2 respectively.

DS Fixed. Following the strategy of Bandi and Phillips (2000), one should be able to obtain similar results when substituting the kernel estimator of Bandi and Phillips (2003) with the sieve-estimator of Chen et al (2000a). The asymptotic distribution is not easy to obtain however in this case.

Aït-Sahalia (1996a) considered a special case of the class of models in (2.18). He assumed the following specification,

$$dX_t = \beta(\alpha - X_t)\,dt + \sigma(X_t)\,dW_t, \tag{2.19}$$

for which it holds that

$$E[X_{(i+1)\Delta}|X_{i\Delta}] = \alpha + e^{-\beta\Delta}\left(X_{i\Delta} - \alpha\right), \tag{2.20}$$

He then proposed to estimate $\theta = (\alpha, \beta)$ by generalised least squares yielding an estimator of the drift, $\hat{\mu}(x) = \hat{\beta}(x - \hat{\alpha})$. Next, substituting the parametric estimator of $\mu(x)$ and the kernel estimator of $\pi(x)$ into the relation (2.5), an estimator of $\sigma^2(x)$ is obtained. It is showed in Aït-Sahalia (1996a) that

$$\sqrt{nh}(\hat{\sigma}^2(x) - \sigma_0^2(x)) \to^d N\left(0, \frac{||K||_2^2\sigma_0^4(x)}{\pi_0(x)}\right).$$

The conditional mean expression (2.20) allows Aït-Sahalia (1996a) to estimate θ separately from the nonparametric part. But this expression is special to the model (2.19); in the general case where $\mu(x;\theta)$ is non-linear in x, one cannot derive a regression equation as the one above. In particular, the conditional mean (or any other conditional moment for that sake) will be a function not only of θ but also $\sigma^2(\cdot)$. Thus, in the general case other methods have to be employed. In Chapter 4, one such method is proposed which covers virtually any model in either of the two classes of semiparametric models.

2.4 Conclusion

We have in this chapter presented the main results in the estimation of diffusion models in a parametric, semiparametric and nonparametric setting respectively. It was observed that the case where a continuous record or a discrete sample with vanishing time distance between observations were available was relatively easy to deal with. However, estimation given a discrete sample with fixed time distance between observations created additional problems in all three types of models. In particular, it proved difficult to derive analytical expressions which allows one to identify the drift and diffusion term. And even if one has managed to derive such, one will have to rely on numerical approximations in order to implement the resulting estimator.

3

Term Structure Modelling with Diffusions

3.1 Introduction

The term structure enters as an input in many macroeconomic models. It is also required as an input in asset pricing models in general and interest rate derivative pricing models in particular. In the mathematical finance literature, the term structure is often modelled using diffusion processes. The use of these facilitates the theoretical analysis since one has at his disposal the whole machinery of stochastic calculus. In-depth treatment of the properties of this type of term structure models and can be found in for example Björk (1998, Chapter 15-17), Duffie (1996, Chapter 7). For a discussion of the modelling of the diffusion processes used to describe the term structure dynamics, we refer to Rogers (1995).

In this chapter, we give a brief introduction to the components entering term structure diffusion models, present some important results concerning bond and interest rate derivative pricing, and give a review of the various term structure diffusion models proposed in the literature. We put special emphasis on the class of so-called single-factor models. We shall not give any formal proofs of the results presented in this chapter, but merely refer to the relevant studies where these can be found.

We first set up the basic framework of a general term structure model in which one can price derivatives in Section 2. Assuming that the model is driven by a diffusion process, we present closed form expressions of bond and interest rate derivative prices. We then introduce the class of factor models in Section 3, while Section 4 deals with the class of so-called Heath-Jarrow-Morton models.

3.2 The Arbitrage-Free Term Structure

In this section, we introduce the framework we shall work within, and give some general results which are useful in the construction of term structure models and pricing of contingent claims.

We start out with some basic definitions. By a *zero-coupon bond* with maturity at time $T > 0$, we mean a financial security which pays the owner 1 unit of cash at time T; we shall also refer to such a security as a T-*bond.* We denote the price of a T-bond at time $t \leq T$ by $B_t(T)$. We assume that for any given $T > 0$, $\{B_t(T)\}$ follows a strictly positive, adapted process on the probability space $(P, \Omega, \mathcal{F})$ with an associated filtration $\{\mathcal{F}_t\}$. We then define (assuming the derivative $\partial B_t(T) / \partial T$ exists)

- The *yield to maturity*: $Y_t(T) = \log(B_t(T)) / (T - t)$.
- The *instantaneous forward rate*: $f_t(T) = \partial \log(B_t(T)) / \partial T$.
- The *short-term interest rate*: $r_t = f_t(t)$.

It is very much standard in the term structure literature to construct models in terms of either of the three variables introduced above. One can readily invert the first two of the above definitions and express any zero-coupon bond in terms of either the yield or the forward rate curve:

$$\begin{aligned} B_t(T) &= \exp\left[(T-t) Y_t(T)\right], \\ B_t(T) &= \exp\left[-\int_t^T f_t(s)\, ds\right]. \end{aligned} \tag{3.1}$$

In relation to the short-term interest rate one defines the so-called *money account*, β_t, given by

$$\beta_t = \exp\left[\int_0^t r_s ds\right],$$

or equivalently $d\beta_t = r_t \beta_t dt$ with $\beta_0 = 1$. Intuitively, β_t represents the amount of cash accumulated at time t if one starts with one unit cash at time zero and continually rolls over a bond with infinitesimal time to maturity. The asset β_t can be interpreted as a "locally risk free" asset since its infinitesimal rate of return, r_t, is known at time t.

Finally, we introduce a so-called *derivative* or *contingent claim.* This is a contract that pays the owner an adapted dividend stream $\{d_t\}$ until maturity $T > 0$ at which time he receives a pay-off X_T. The family of bond prices is the simplest example of a derivative where $d_t \equiv 0$ and $X_T \equiv 1$.

We say that the family of bond prices $\{B_t(T) | T > 0\}$ is *arbitrage-free* if

1. $B_T(T) = 1$ for any $T > 0$.
2. There exists a probability measure Q equivalent to P such that the process $Z_t(T) = B_t(T) / \beta_t$ is a martingale under Q,

$$E^Q\left[Z_t(T) | \mathcal{F}_s\right] = Z_s(T).$$

The probability measure Q is normally called the *risk-neutral measure*, while P (which is the measure under which we observe the prices) is denoted the *physical measure.* Assuming the existence of a risk neutral measure Q, the price at time $t \leq T$ of a claim satisfies

$$\Pi_t(T) = E^Q \left[\int_t^T d_s \exp\left[-\int_t^s r_u du \right] ds + X_T \exp\left[-\int_t^T r_u du \right] \middle| \mathcal{F}_t \right].$$

In particular, the price of a bond is given by

$$B_t(T) = E^Q \left[\exp\left[-\int_t^T r_s ds \right] \middle| \mathcal{F}_t \right].$$

Thus, under the additional assumption of the existence of a risk-neutral measure, we can also invert w.r.t. $\{r_t\}$. An important class of interest rate derivatives is where the dividend stream and the terminal pay-off both are functions of the short term interest rate, $d_t = d(t, r_t)$ and $X_T = c(r_T)$.

So in the term structure modelling, one is interested in constructing a measure Q such that 2. above is satisfied since this gives access to a closed form expression of any claim. For the specific model, one needs to establish the existence of Q and derive the dynamics of the variables of interest under this measure.

A leading case where one can establish the existence of Q is the one where $\{r_t\}$ is a diffusion process. Assume that $\{r_t\}$ solves a stochastic differential equation (SDE) of the form

$$dr_t = \mu_t dt + \sigma_t^\top dw_t, \tag{3.2}$$

where $\{\mu_t\}$ and $\{\sigma_t\}$ are $\mathbb{R}$- and $\mathbb{R}^q$-dimensional adapted processes respectively, while $\{w_t\}$ is a q-dimensional standard Brownian motion. Then for any adapted $\mathbb{R}^q$-valued process $\{\lambda_t\}$ such that the so-called Doleans exponential, $\{\mathcal{E}_t(\lambda \star W)\}$ defined by

$$\mathcal{E}_t(\lambda \star W) = E^P \left[\exp\left[\int_0^t \lambda_s dW_s - \frac{1}{2} \int_0^t \|\lambda_s\|^2 ds \right] \right],$$

is a P-martingale, there exists a unique risk-neutral measure Q under which

$$dr_t = \{\mu_t - \lambda_t^\top \sigma_t\} dt + \sigma_t^\top dW_t,$$

where $\{W_t\}$ is a q-dimensional standard Brownian motion under Q. Furthermore, the T-bond under P solves

$$dB_t(T) = B_t(T)\{r_t + \lambda_t^\top \sigma_t\} dt + B_t(T)\, s_t^\top(T)\, dW_t$$

while under Q,

$$dB_t(T) = B_t(T)\, r_t dt + B_t(T)\, s_t^\top(T)\, dW_t$$

for an adapted process $\{s_t(T)\}$ satisfying $s_t(T) = \sigma_t \partial B_t(T) / \partial r_t$. This means that under the physical measure the instantaneous returns from holding the bond differ from the short term interest rate r_t by $\lambda_t^\top s_t(T)$. Thus $\lambda_t^\top s_t(T)$ measures the *risk premium* for the T-bond, i.e. the rate of return over the risk free rate commanded by a T-bond. In particular, for $q = 1$, the quotient $\lambda_t = \lambda_t s_t(T) / s_t(T)$ can be interpreted as the risk premium per unit of volatility. This quotient, λ_t, is often termed the *market price of risk.* Observe that $\{\lambda_t\}$ does not depend on T such that all bonds will have the same market price of risk.

While we have been able to derive a closed form expression of any claim, it cannot be implemented before we have chosen the market price of risk process, $\{\lambda_t\}$. Often this is chosen such that the implied bond prices match the observed ones.

3.3 The Multi-Factor Model

Above we derived closed form expressions of any claim in a fairly general term structure model. This model is however so general that it cannot be calibrated nor implemented as it is. In the following, we shall further restrict the diffusion model in (3.2) to allow for actual calibration and implementation. We assume that a number of factors drive the short rate, and that these factors define a Markov process. This class of models are termed multi-factor models.

We assume that

$$r_t = R(F_t) \tag{3.3}$$

for some twice differentiable function $R : \mathbb{R}^q \mapsto \mathbb{R}$, and some q-dimensional process $\{F_t\}$ which solves a SDE of the form,

$$dF_t = \mu(t, F_t)\,dt + \sigma(t, F_t)\,dw_t, \tag{3.4}$$

under P where $\mu : [0, \infty) \times \mathbb{R}^q \mapsto \mathbb{R}^q$, $\sigma : [0, \infty) \times \mathbb{R}^q \mapsto \mathbb{R}^{q \times q}$, and $\{w_t\}$ is a q-dimensional Brownian motion. The variables in $\{F_t\}$ are normally referred to as the *factors.* These can either be chosen to be economically meaningful variables or some latent ones of unknown identity. By Itô's Lemma, we obtain that $\{r_t\}$ also solves a SDE. Thus, the multifactor model (3.3)-(3.4) is a special case of (3.2), and the pricing formulae in the previous section are valid for the multi-factor model.

A very popular class of factor models are the affine ones as proposed in Duffie and Kan (1996) and Duffee (2002). We assume that

$$r_t = \delta_0 + \delta^\top F_t,$$

and

$$dF_t = (a - BF_t)\,dt + \Sigma S_t^{1/2} dw_t$$

where B and Σ are $q \times q$-matrices, a is an q-vector, and S_t is an q-dimensional diagonal matrix with diagonal elements

$$[S_t]_{ii} = \alpha_i + \beta_i^\top F_t.$$

Finally, the market price for risk is an q-dimensional vector which is assumed to satisfy

$$\lambda_t = S_t d + S_t^- D F_t$$

where d is an q-dimensional vector, D an $q \times q$-matrix, and S_t^- is a diagonal matrix with

$$\left[S_t^-\right]_{ii} = \begin{cases} \left(\alpha_i + \beta_i^\top F_t\right)^{-1/2}, & \left(\alpha_i + \beta_i^\top F_t\right)^{-1/2} > 0 \\ 0, & \text{otherwise} \end{cases}.$$

This special structure ensures that the factor dynamics are affine both under the physical and risk-neutral measure. This in turns allows one to derive analytical expressions of the bond prices and various interest rate derivatives as demonstrated in Chacko & Das (2002).

Another special case within the class of factor-models is when $q = 1$. Assuming $F_t = r_t$ (such that $R(x) = x$), we obtain the class of so-called *single-factor models* where the short term interest rate is a Markov process,

$$dr_t = \mu(t, r_t)\, dt + \sigma(t, r_t)\, dw_t,$$

with $\{w_t\}$ being one-dimensional. Assume additionally that the risk premium process satisfies

$$\lambda_t = \lambda(t, r_t)$$

for some function λ. We then obtain that

$$dr_t = \mu^\lambda(t, r_t)\, dt + \sigma(t, r_t)\, dW_t, \quad \mu^\lambda(t, r) = \mu(t, r) - \lambda(t, r)\, \sigma(t, r)$$

under Q. Thus, $\{r_t\}$ is also a Markov process under Q, and we have that $\Pi_t(T) = u(t, r_t)$ for some function u. Using the Feynman-Kac formula, the valuation function u solves the following fundamental PDE,

$$\frac{\partial u}{\partial t} + \mu^\lambda(t, x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 u}{\partial x^2} - ru + d(t, r) = 0,$$

with terminal condition $u(T, r) = c(r)$.

In the following we present some of the specifications of μ and σ^2 in the single-factor model suggested in the literature. For a more detailed discussion of these and other models, we refer to Rogers (1995). The first model for the short term interest rate was proposed by Merton (1973). He suggested to model the short-term interest rate as a Brownian motion with drift,

$$dr_t = \mu dt + \sigma dw_t.$$

This model has the unfortunate property that with positive probability $r_t < 0$. It is furthermore non-stationary and exploding. Vasicek (1977) defined $\{r_t\}$ as an Ornstein-Uhlenbeck process,

$$dr_t = \beta(\alpha - r_t)\, dt + \sigma dw_t,$$

where $\alpha, \beta > 0$. There exists a stationary solution to this SDE, but again $r_t < 0$ with positive probability. Cox, Ingersoll and Ross (1985) (CIR) dealt with this problem, modelling $\{r_t\}$ as the solution to

$$dr_t = \beta(\alpha - r_t)\, dt + \sigma\sqrt{r_t} dw_t,$$

where $\alpha, \beta > 0$. Under suitable parameter restrictions, $\{r_t\}$ is a stationary process on the domain $I = \mathbb{R}_+$. Observe that all these three models belong to the affine class of factor models. More advanced parametric specifications have been proposed in the literature. For example,

$$\begin{aligned}
&\text{Aït-Sahalia(1996b):} && dr_t = \left\{\beta_0 + \beta_1 r_t + \beta_2 r_t^2 + \beta_3 r_t^{-1}\right\} dt + \sqrt{\sigma_0 + \sigma_1 r_t + \sigma_2 r_t^{\gamma}} dw_t, \\
&\text{Conley et al.(1997):} && dr_t = \left\{\beta_0 + \beta_1 r_t + \beta_2 r_t^2 + \beta_3 r_t^{-1}\right\} dt + \sigma r_t^{\gamma} dw_t, \\
&\text{Ahn and Gao(1999):} && dr_t = \left\{\beta_0 + \beta_1 r_t + \beta_2 r_t^2\right\} dt + \sqrt{\sigma_0 + \sigma_1 r_t + \sigma_2 r_t^3} dw_t.
\end{aligned}$$

In the financial industry, the above models are often generalised to allow for time-dependent parameters. This leads to the class of time-inhomogenous models. Ho and Lee (1986) proposed the first specific time-inhomogenous model, letting the parameter $\alpha = \alpha_t$ in the Merton-model to be time-dependent,

$$dr_t = \alpha_t dt + \sigma dw_t.$$

Similarly, many of the other time-homogenous models presented have been extended to allow for time-dependent parameters, see e.g. Hull and White (1990) for extended versions of the Vasicek- and CIR-model. The advantage of these models is that they can be calibrated on a daily basis to deliver a perfect fit of the current yield curve, something the corresponding time-homogenous models very often fail to do. This is very appealing from a practical point of view. On the other hand, these models say nothing about the dynamics of the time-varying parameters, and the they are therefore not very useful in predicting future yield curves which is needed in a bond and option pricing scenario.

The specification of $\{\lambda_t\}$ is still an open question. A relative simple specification has been favoured in the literature facilitating the calibration of the model and the calculation of the implied bond prices. In particular, a number of studies has chosen λ_t to be constant, as for example in Vasicek (1977) and Aït-Sahalia (1996a).

3.4 The Heath-Jarrow-Morton Model

Instead of modelling the short-term interest rate, Heath, Jarrow and Morton (1991), HJM henceforth, assumed that the forward rate with maturity at time T solved a SDE,

$$df_t(T) = \mu_t(T)\, dt + \sigma_t^{\top}(T)\, dw_t$$

under P, where $\{\mu_t(T)\}$ and $\{\sigma_t(T)\}$ are adapted processes taking values in $\mathbb{R}$ and $\mathbb{R}^q$ respectively while $\{w_t\}$ is a q-dimensional Brownian motion. A major advantage of this class of models over the factor models introduced in the previous section is that the HJM-model allows for a perfect fit of the current yield curve - this is simply chosen as the initial condition of the forward curve, $f_0(T)$. The bond prices can be recovered from the formula (3.1).

As before we examine under which conditions no arbitrage occur. Assume that there exists an adapted $\mathbb{R}^q$-valued process $\{\lambda_t\}$ such that the associated Doleans exponential is a P-martingale, and

$$\mu_t(T) = \sigma_t^\top(T)\,\sigma_t^*(T) - \sigma_t^\top(T)\,\lambda_t,$$

where

$$\sigma_t^*(T) = \int_t^T \sigma_t(s)\,ds.$$

Then there exists a unique risk-neutral measure Q. Under Q, the forward rate satisfies

$$df_t(T) = \sigma_t^\top(T)\,\sigma_t^*(T)\,dt + \sigma_t^\top(T)\,dW_t,$$

and the bond prices

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

where $\{W_t\}$ is a q-dimensional standard Brownian motion. An important point here is that under Q the forward rate and bond dynamics are characterised by the diffusion term $\sigma_t(T)$ alone - the drift term is of no importance. Thus, in order to price claims one only needs to specify and calibrate the volatility term.

A number of specific single-factor models can be shown to be a special case of the general HJM-model. For example, the Ho and Lee (1986) model can be written as a HJM-model This indicates that the HJM-setting is a more general way of describing the term structure.

The original HJM model suffers from stochastic singularity in the sense that the model implies deterministic relations between bonds of different maturities. This problem can be dealt with by extending the model to allow for a richer class of noise terms. Kennedy (1994, 1997) and Santa-Clara and Sornette (2000) are examples of this approach.

3.5 Conclusion

We have in this chapter presented a number of different term structure models which are based on diffusion processes. These have the advantage that the implied bond and interest rate derivative prices can be written on a closed form. While the finance theory is fully developed, it is still an open question which statistical model one should use when one takes the finance model to the data.

In the next chapter, we present two classes of semiparametric models which are very rich and highly flexible. These can be used in the modelling the term structure by a single-factor model. We also develop an estimator for any model in these two classes which allows one to calibrate the single-factor model using historical data of the short-term interest rate. Then in Chapter 5, we derive the asymptotic properties of the implied bond and derivative prices of the single-factor model based on the estimators in Chapter 4. In Chapter 6, we illustrate the use of the results obtained in Chapter 4 and 5 by fitting a specific semiparametric diffusion model to historical data.

4

Estimation in Two Classes of Semiparametric Diffusion Models

4.1 Introduction

Continuous time stochastic processes are widely used in dynamic models in economics and finance. In the past three decades since the groundbreaking work by Black and Scholes (1973), Merton (1973) stochastic processes have gained a major role in finance theory where they are used in the modelling of the dynamics of economic variables over time, for example interest rates, stock prices, and exchange rates; an overview of such models can be found in Duffie (1996). To a lesser extent these have also been used to model the dynamics of macroeconomic variables, see e.g. Bergstrom (1990). Unfortunately, economic theory has very little to say about the precise specification of the processes. As a consequence, a wide range of parametric models have been suggested in the literature, for example Black and Scholes (1973), Chan et al. (1992), Cox et al. (1985), Vasicek (1977), but it is not obvious that these models are able to deliver an adequate description of the observed process. This may lead to the use of a misspecified model that is not able to capture the true dynamics of the process in consideration. This again can have serious implications on the conclusions drawn from the model. Non- and semiparametric methods may help to detect and to some extent solve such problems, since these methods allow for a high degree of flexibility and should thereby better safeguard one against possible misspecification.

In this chapter, we consider a semiparametric approach to the modelling and estimation of scalar stochastic differential equations (SDE's) driven by a Brownian motion. Such processes are fully characterised by their drift and diffusion function, which we wish to model in a flexible manner. Two very general classes of models will be considered: In the first class, the drift is specified (up to an unknown parameter) while the diffusion term is left unspecified; in the second class it is the diffusion term that is parameterised while the drift term is not specified. We define an estimator for the drift and diffusion function for models in each of the two classes, and derive its asymptotic properties under regularity conditions. We also construct a simple test for parametric submodels against the semiparametric alternative. The main restriction we need to impose is that the diffusion processes in the two classes are strongly stationary since this property is used for identification of the unspecified term. This excludes for example time-inhomogenous processes, where the drift and diffusion functions are allowed to depend on time, since these are non-stationary by construction. The two classes are still very rich, and include a majority of the parametric homogeneous models proposed in the literature since these in most cases allow for stationary solutions. In particular, for any parameterisation of a stationary diffusion process, each of the two classes contains a semiparametric model which has this fully parametric model as a submodel.

Only a few studies in the existing literature have considered semiparametric diffusion models. Aït-Sahalia (1996a) proposes a semiparametric model with a linear parameterisation of the drift, while leaving the diffusion term unspecified. Conley et al. (1997) on the other hand suggest to use a simple parametric form for the diffusion term, while either applying a global series expansion or a locally linear approximation of the drift term. The model of Aït-Sahalia (1996a) belongs to the first class of models considered here, while the Conley et al. (1997) model is situated in the second one. These two models are quite general, but one may still want to allow for other, more flexible, specifications of either the drift or the diffusion term than the two proposed by the aforementioned authors. This is made possible with the two classes of semiparametric models proposed here, which allows for virtually any reasonable parameterisation of either the drift or diffusion term. In Bandi and Phillips (2000), least squares estimators for any parameterisation of either the drift or the diffusion term is proposed; see also Florens-Zmirou (1989) and Genon-Catalot (1990). Their results however depend on the time distance between observations shrinking to zero, the so-called infill assumption, while ours hold for a fixed time distance. We restrict our attention to diffusions driven by a Brownian motion.

The semiparametric models under consideration here can be very useful as an intermediate step in model building, moving from an initial nonparametric model towards a parsimonious fully parametric one. There is a large literature on fully nonparametric estimation of the drift and diffusion function. Most of the proposed estimators are based on kernel methods, making use of the characterisation of the drift and diffusion function as the instantaneous conditional mean and variance respectively. Assuming that the time distance between observations shrinks to zero as the number of observations goes to infinity, standard kernel regression methods can be used to consistently estimate the drift and diffusion term. This approach is pursued by, for example, Bandi and Phillips (2003), Jiang and Knight (1997), and Stanton (1997). These estimators are however prone to a discretization bias if the process is in fact observed at fixed time instants, cf. Nicolau (2003). Chen et al. (2000a), Darolles and Gouriéroux (2001) and Gobet et al. (2003) derive nonparametric estimators for univariate diffusion models by the method of sieves, allowing for a fixed time distance between observations. Their approach is based on the so-called infinitesimal operator of the diffusion model, which uniquely identifies the model. They decompose the operator into its eigenfunctions and demonstrate that from these one may recover the drift and diffusion term. Estimators of the eigenfunctions are then constructed, and thereby also estimators of the drift and diffusion function.

Our estimation method is based on the assumption that the sampled diffusion process is stationary and ergodic, thereby ensuring that an invariant density of the process exists. By using the Kolmogorov forward equation, the density can be expressed in terms of the drift and diffusion term. Inverting this expression, one can write the drift (diffusion) term as a functional of the density and the diffusion (drift) term. This allows us to uniquely identify the drift (diffusion) term given a parameterisation of the diffusion (drift) together with a nonparametric estimator of the invariant density. This idea originates from Wong (1964), and was further developed in Hansen and Scheinkman (1995), and Hansen et al. (1998). Aït-Sahalia (1996a) made use of the same link to estimate his semiparametric diffusion model. Due to the higher level of generality, our estimator becomes more involved than the one in Aït-Sahalia (1996a) though. There, a closed form estimator for the parametric part is derived, not depending on the nonparametric part. Unfortunately, in the general case it does not appear as if one can separate the estimation of the parametric part from the nonparametric one when given discrete observations. Instead, our estimator is obtained in the following three steps: First, we obtain a nonparametric estimator of the marginal density. Then the parametric part is estimated using the log-transition density of the diffusion process with the marginal density estimator plugged in as a nuisance parameter. Finally, the nonparametric part is estimated as a functional of the nonparametric density estimator and the parametric estimator.

The benefits from using the log-transition density to estimate the parameter are twofold: First, it is more likely that the parameter is identified since the transition density gives a full description of the probability structure of the sampled process.[1] Second, assuming that the nonparametric part is known, estimation of the parametric part by the log-transition density yields the efficient MLE. One would expect the semiparametric estimator to be close to the (infeasible) fully parametric MLE, and thereby enjoy a high level of efficiency.

Since it is not possible to directly evaluate the transition density, we propose either to use approximate (e.g. Aït-Sahalia, 2002) or simulation-based methods (see e.g. Durham and Gallant, 2002) in order to implement the estimator. The estimator obtained from these methods will enjoy the same properties as the actual, but infeasible one, under suitable conditions. The finite sample properties of the estimator using approximate likelihood is investigated in a small simulation study. Here, we will see that even for moderate sample sizes, our estimator performs well, and that the approximate method does a good job.

Under regularity conditions, we derive the asymptotic properties of the estimator, showing that the parametric part is $\sqrt{n}$-consistent, while the nonparametric part has a slower convergence rate. Also, the estimator is shown to follow a normal distribution asymptotically. The asymptotics of the estimator are based on discrete observations with a fixed time distance in between. This is in contrast to the papers on nonparametric kernel estimation of the drift and diffusion cited above, and is a desirable property since a continuous time record of observations may not be available in practice. High frequency (so-called tick-by-tick) data of, e.g., stock prices and exchange rates are now widely available. One

[1] A related problem is the so-called aliasing-problem where discretely sampled stochastic processes are indistinguishable, c.f. Phillips (1973). Hansen and Scheinkman (1995, p. 786) show however that the aliasing problem does not exist for reversible Markov processes.

could argue that these present a (nearly) continuous record, but the data often suffers from various market microstructure effects, see for example Dunis and Zhou (1998). One may therefore be willing to sacrifice some of the available observations to avoid having to deal with such effects, and only use observations of lower frequency (e.g. daily) when estimating the diffusion model.

The rest of the chapter is organised as follows: In Section 2, we set up the framework and give an informal introduction to the proposed estimation procedure. In Section 3, theoretical results concerning the nonparametric part of the estimator are given. The asymptotics of the parametric part of the estimator is derived in Section 4. We discuss the efficiency of the parametric part in Section 5, and propose a 1-step adjustment which should make it reach the semiparametric efficiency bound. The implementation of the estimator is discussed in Section 6, and the results of the simulation study is presented in Section 7. We conclude in Section 8. All proofs and lemmas are collected into the appendices.

Throughout the text, $g^{(\lambda)}(x;\theta)$ denotes the λth derivative w.r.t. x of a function $g:\mathbb{R}\times\Theta\mapsto\mathbb{R}$ with $g^{(0)}\equiv g$, while $\dot{g}(x;\theta)$ and $\ddot{g}(x;\theta)$ denote the first and second derivative w.r.t. θ. At times we shall however also denote derivatives by $\partial^{ij}_{x,\theta}g(x;\theta)\equiv\partial^i\partial^j g(x;\theta)/\partial^i x\partial^j\theta$. We shall write $\|g\|_\infty=\sup_{x\in I}|g(x)|$ and $\|g\|_2=(\int_I|g(x)|^2dx)^{1/2}$ for any function with domain $I\subseteq\mathbb{R}$.

4.2 Framework

Let $\{X_t\}=\{X_t:t\geq 0\}$ be the stochastic process solving the following homogenous SDE,

$$dX_t=\mu(X_t)\,dt+\sigma(X_t)\,dW_t, \tag{4.1}$$

where $\{W_t\}$ is a standard Brownian motion. The domain of $\{X_t\}$ is denoted $I=(l,r)$ where $-\infty\leq l<r\leq\infty$. We define the scale density $s(x)=\exp\left[-2\int_{x^*}^{x}\mu(y)/\sigma^2(y)\,dy\right]$, for some x^* in the interior of I. Sufficient conditions for strong stationarity are (S1) $\int_l^{x^*}s(x)\,dx=-\infty$, $\int_{x^*}^{r}s(x)\,dx=+\infty$, and (S2) $1/M\equiv\int_l^r\left[s(x)\sigma^2(x)\right]^{-1}dx<\infty$, cf. Karlin and Taylor (1981, Section 15.6) and Karatzas and Shreve (1991, Section 5.5). Under these conditions, $\{X_t\}$ is stationary and ergodic with an invariant measure π, $\pi(A)=\int_I P(X_t\in A|X_0=x)\,d\pi(x)$ for any Borel-set A, which has a density given by[2]

$$\pi(x)=\frac{M}{s(x)\sigma^2(x)}=\frac{M}{\sigma^2(x)}\exp\left[2\int_{x^*}^{x}\frac{\mu(y)}{\sigma^2(y)}dy\right]. \tag{4.2}$$

In a parametric framework, models for the above diffusion process is normally constructed by specifying the drift term, μ, and the diffusion term, σ^2, up to an unknown parameter vector $\theta\in\Theta$ where $\Theta\subseteq\mathbb{R}^d$ is a finite-dimensional parameter space. We see from (4.2) that one then implicitly also specifies the stationary density. It is possible to

[2] We here use π to denote both the measure and the density.

revert (4.2) in either of the two following ways,

$$\mu(x) = \frac{1}{2\pi(x)}\frac{\partial}{\partial x}\left[\sigma^2(x)\pi(x)\right], \tag{4.3}$$

$$\sigma^2(x) = \frac{2}{\pi(x)}\int_l^x \mu(y)\pi(y)\,dy. \tag{4.4}$$

So an alternative specification scheme would be to specify the marginal density together with either the drift or the diffusion term, an idea originating from Wong (1964); see also Cobb et al. (1983), Hansen and Scheinkman (1995), Hansen et al. (1998). This could be done in a fully parametric framework, but here we only specify either the drift or the diffusion term and then rely on a nonparametric estimator of π. For example, we may parameterise the diffusion term, and then plug this into (4.3) together with a nonparametric estimator of π. We thereby obtain a semiparametric estimator of μ, by which we mean that it depends both on a parameter, θ, and a function, π. These considerations lead us to suggest the following two semiparametric classes of diffusion models:

Class 1:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t;\theta)\,dW_t, \tag{4.5}$$

with $\mu(\cdot)$ unknown and $\sigma^2(\cdot;\theta)$ known up to the parameter θ.

Class 2:

$$dX_t = \mu(X_t;\theta)\,dt + \sigma(X_t)\,dW_t, \tag{4.6}$$

with $\mu(\cdot;\theta)$ known up to the parameter θ and $\sigma^2(\cdot)$ unknown.

Here and in the following, μ_0, σ_0^2 and π_0 will denote the true drift, diffusion and invariant density respectively associated with the data-generating process. To discuss the estimation of the two classes of models, let us as an example consider a model from Class 1. In this case, we are given a parameterisation of the diffusion term, $\sigma^2(\cdot;\theta)$, which we plug into the RHS of (4.3) together with a density π,

$$\mu(x;\theta,\pi) = \frac{1}{2\pi(x)}\frac{\partial}{\partial x}\left[\sigma^2(x;\theta)\pi(x)\right]. \tag{4.7}$$

To obtain an estimator of θ we then make use of the transition density p of $\{X_t\}$, which is characterised by $P(X_{t+\Delta} \in A | X_t = x) = \int_A p(y|x)\,dy$ for any Borel-set A. Since $\{X_t\}$ is completely characterised by μ and σ, p is a functional of these two, $p(y|x) = p(y|x;\mu(\cdot),\sigma(\cdot))$. In the following section, a precise expression of p as a functional of μ and σ is derived by utilising results of Dacunha-Castelle and Florens-Zmirou (1986). By plugging in $\sigma(x;\theta)$ and $\mu(x;\theta,\pi)$, a semiparametric version of the transition density, $p(y|x;\theta,\pi) = p(y|x;\mu(\cdot;\theta,\pi),\sigma(\cdot;\theta))$, now appears. This version of the transition density will be employed to perform MLE-like estimation of θ given a nonparametric estimator of π. Let $X_0, X_\Delta, X_{2\Delta}..., X_{n\Delta}$ be $n+1$ observations obtained from (4.5), where $\Delta > 0$ is the fixed time distance between observations; without loss of generality, we set $\Delta \equiv 1$ in the following.[3] The following nonparametric kernel estimator of the rth derivative, $\pi_0^{(r)}$ (assuming that it exists), is then available,

$$\hat{\pi}^{(r)}(x) = \frac{1}{nh_r^{r+1}} \sum_{i=1}^{n} K^{(r)}\left(\frac{x - X_i}{h_r}\right), \quad r \geq 1, \tag{4.8}$$

for a kernel K and a bandwidth h_r; see Silverman (1986) for an introduction to these concepts.[4] Note that we use potentially different bandwidths to estimate each derivative. Under regularity conditions, including $h_r = h_{r,n} \to 0$ and $nh_r^{2r+1} \to \infty$, $\hat{\pi}^{(r)}(x) \to^P \pi_0^{(r)}(x)$ as $n \to \infty$. We plug $\hat{\pi}$ and $\hat{\pi}^{(1)}$ into (4.7), yielding $\hat{\mu}(x;\theta) = \mu(x;\theta,\hat{\pi})$, which in turn is plugged into the transition density. We then propose to estimate θ by

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L_n(\theta, \hat{\mu}(\cdot;\theta)) \tag{4.9}$$

where

$$L_n(\theta,\mu) = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i|X_{i-1};\mu,\sigma(\cdot;\theta)). \tag{4.10}$$

Once $\hat{\theta}$ has been found, the obvious pointwise estimator of $\sigma^2(x)$ is $\sigma^2(x;\hat{\theta})$ while $\mu(x)$ is estimated by plugging $\hat{\theta}$ and $\hat{\pi}$ into (4.7) yielding $\hat{\mu}(x) = \mu(x;\hat{\theta},\hat{\pi})$. The above procedure is also applicable for models from Class 2, only this time we are given a full parameterisation of $\mu(\cdot) = \mu(\cdot;\theta)$, which can be substituted into (4.4) together with a nonparametric estimator of π, thereby obtaining a semiparametric estimator of $\sigma^2(\cdot) = \sigma^2(\cdot;\theta,\pi)$.

The dependence of the nonparametric estimators, $\hat{\mu}(x)$ in Class 1 and $\hat{\sigma}^2(x)$ in Class 2, on the smoothing parameter h (and a trimming parameter introduced later) chosen by the user is an undesirable feature, which they share with many other non- and semiparametric estimators. The sensitivity of the estimators towards h can be high, and one therefore has to be careful when choosing the bandwidth. Too small values of h can give imprecise estimates, while a too large choice can induce bias. Rules of thumb are often applied for the bandwidth choice, but data driven methods such as cross-validation may lead to better performance. In our framework, such methods are not readily available however. A further discussion of these and related issues can be found in Section 6.

The estimation procedure described above belongs to a general class of semiparametric estimation problems, where an estimator of a finite-dimensional parameter θ is obtained with the help of a preliminary estimator of an infinite-dimensional nuisance parameter (here, π). General treatments of the asymptotic properties of such profiled/concentrated semiparametric estimators can be found in e.g. Andrews (1994), Chen et al. (2003), Newey and McFadden (1994, Section 8). The estimation of the finite-dimensional parameter is performed by what we may call semiparametric MLE. There is a large literature on non- and semiparametric MLE,[5] but there the infinite-dimensional parameter is estimated to-

[3] To simplify the exposition, equidistant observations over time are assumed; our results can be extended to allow for varying time distances between observations.

[4] For notational convenience, we here leave out the first observation in the definition of $\hat{\pi}$. This will have no consequences for the asymptotic properties.

[5] See for example Murhpy and Van der Vaart (2000) and the references therein.

gether with the finite-dimensional one, while here we make use of a preliminary estimator of the former. This makes our asymptotic theory somewhat different from that strand of the literature. Instead our estimator fits nicely into a general class of semiparametric two-step estimators: In the first step a function is (nonparametrically) estimated, while in the second step this is used to obtain an estimator of a finite-dimensional parameter. So in this setting the function estimated in the first step can be seen as a nuisance parameter. In our case, the function in question is the invariant density. Chen et al. (2003) and Newey and McFadden (1994, Section 8) give general conditions for consistency and asymptotic normality for such profiled semiparametric estimators. Unfortunately, the problem at hand here cannot directly be dealt with in the framework of those two studies since we have to introduce trimming of our nonparametric estimators. We therefore have to modify their conditions in order to establish our theoretical results; Ai (1997) and Robinson (1988) contain related applications of trimming in a semiparametric framework. Furthermore, the transition density takes a very complicated form, and a careful analysis of it as a function of the drift and diffusion function is required in order to derive the asymptotic properties. In particular, the derivation of the asymptotic distribution of the parametric part is very cumbersome, and we are unable to give an explicit expression for the resulting asymptotic variance. We are however able to set up a consistent estimator of it. Finally, Chen et al. (2003) and Newey and McFadden (1994) only give conditions for i.i.d. data, while our observations are dependent. In order to handle this additional complication, we have to assume that our process is not only stationary, but weakly dependent (in fact, we assume it is β-mixing), and restrict the decay rate of the mixing-coefficients in a suitable manner. This should be seen as a technical assumption however used to facilitate our analysis rather than a necessary property needed for the results to carry through.

There are certain obstacles with the implementation of the proposed estimator since the transition density p for general specifications of μ and σ^2 cannot be written in an explicit form, thereby not allowing for direct evaluation. We resolve this problem by relying on either approximate methods (see e.g. Lo 1988, Aït-Sahalia 2002) or simulation-based methods (see e.g. Durham and Gallant 2002, Elerian et al 2001, Hurn et al 2003, Pedersen 1995). Applying such methods in the implementation of our estimator will have an asymptotically negligible effect on $\hat{\theta}$ if the order of approximation is allowed to increase with the sample size at a fast enough rate.

We remark that other criterion functions than $\log p$ could be used to estimate θ. In Aït-Sahalia (1996a) for example OLS is used; it is not clear however if this idea can be adapted to more general cases or only works for his specific choice of parameterisation. There is a variety of other estimating procedures in the literature for diffusion models, see e.g. Duffie and Singleton (1993), Hansen and Scheinkman (1995), Gallant and Long (1997), Gallant and Tauchen (1996), Gouriéroux et al. (1993), Sørensen (1997), but the log-likelihood approach is the most natural choice, and one would expect that this would yield a near-optimal estimator.

There is also room for different estimators of π_0. Our theoretical results are based on the use of the above kernel estimator of π_0, but can be substituted with alternative estimators such as series or spline estimators, cf. Stone (1990), as long as one is able to show uniform consistency with a sufficiently high convergence rate for this.

Observe that if π_0 was known, we would be in a fully parametric framework and $\hat{\theta}$ would be the maximum-likelihood estimator (MLE), which under regularity conditions would enjoy full efficiency. But since we have not fully specified our model, the asymptotic variance of $\hat{\theta}$ may not reach the Cramer-Rao bound. One would however expect that the asymptotic properties of $\hat{\theta}$ are closely related to the fully parametric MLE. As we shall see, the asymptotic distribution of $\hat{\theta}$ in fact equals that of the fully parametric MLE plus an additional term entering the variance; this is due to the fact that we use an estimator of π_0 instead of the unknown density itself. This is related to the issue of semiparametric efficiency, see Newey (1990) and Severini and Tripathi (2001) for overviews. It could be of interest to derive the efficiency bound for the semiparametric models of this paper, and see whether our estimator reaches it. This is non-trivial though. Most of the existing literature on semiparametric efficiency is concerned with i.i.d. data, while we work with a Markov process. Moreover, the analysis of the transition density as a functional of π is not easy, and will require a lot of additional work. In Section 5, we give a brief discussion of these issues, and propose a 1-step adjustment to our semiparametric estimator which we conjecture will reach the semiparametric efficiency bound. A rigorous treatment of the efficiency bound and the 1-step adjustment is left for future research.

As stressed earlier, we here restrict our attention to stationary diffusion processes. The above identification scheme can however be extended to a wider class of processes satisfying (S1), but not necessarily (S2). In this case, the invariant density π exists, but is not necessarily integrable, allowing for $\int_I \pi(x)\,dx = +\infty$. The density will still satisfy (4.2) (leaving out M), such that the relation given in (4.3) remains valid, while for (4.4) to hold one has to require $\lim_{x\to l} \pi(x)\,\sigma^2(x) = 0$.[6] In the groundbreaking work by Bandi and Phillips (2003), it is demonstrated that for this extended class of "weakly" non-stationary (so-called recurrent) processes, μ and σ^2 can be consistently estimated by kernel methods as $\Delta \to 0$; for related results, we refer to Karlsen and Tjøstheim (2001) and Park and Phillips (1998). However, it is not clear what the asymptotic behaviour of the estimators proposed above will be when (S2) does not hold; this will be investigated in future research.

Our estimation procedure cannot readily be extended to general multivariate diffusion models, since the link between the invariant density, the drift and the diffusion term utilised here does not necessarily hold in higher dimensions. If one is ready to restrict the attention to the class of multivariate models satisfying this relation,[7] the proposed estimation procedure should still work. But it would suffer from the well-known curse of dimensionality of nonparametric estimators. Moreover, the transition density in the general multivariate case is even more difficult to analyse than in the univariate one, so the task of establishing theoretical results for the parametric part of the estimator in a multivariate setting will be a rather difficult one.

[6] This will automatically be satisfied under (S2).

[7] This restriction is for example imposed by Chen et al (2000b) in their nonparametric study of multivariate diffusion models.

4.3 The Nonparametric Estimator

In this section, we show that the nonparametric estimators of μ and σ^2 proposed in the previous section will be pointwise consistent and asymptotically normally distributed for any given $\sqrt{n}$-consistent estimator of θ. So we here assume the existence of such an estimator. In the next section we show that the estimator of θ proposed in the previous section is indeed $\sqrt{n}$-consistent. We also give uniform convergence rates and define a simple test statistic allowing one to test any parametric submodel against the semiparametric alternative.

In Class 1, we simply plug in the initial estimators of the marginal density π and the parameter θ, yielding

$$\hat{\mu}(x) = \frac{1}{2\hat{\pi}(x)} \frac{\partial}{\partial x} \left[\sigma^2(x;\hat{\theta}) \hat{\pi}(x) \right], \tag{4.11}$$

where $\hat{\pi}$ is the kernel estimator in (4.8) and $\hat{\theta}$ is the estimator of θ. For Class 2, we observe that by the Law of Large Numbers (LLN) for stationary and ergodic sequences,

$$\frac{1}{n} \sum_{i=1}^{n} 1_{(l,x)}(X_i) \mu(X_i;\theta) \to^P \int_l^x \pi_0(y) \mu(y;\theta) \, dy,$$

for any $(x,\theta) \in I \times \Theta$ given the moment exists. We then define[8]

$$\hat{\sigma}^2(x) = \frac{2}{\hat{\pi}(x)} \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,x)}(X_i) \mu(X_i;\hat{\theta}).$$

As noted earlier, we have to assume that $\{X_t\}$ is stationary and ergodic in order to be able to identify the unspecified term. In fact, we require it to be geometrically β-mixing. The results stated in this section will actually hold under weaker mixing conditions. But since in the next section we need β-mixing in order to employ U-statistics results for dependent sequences (see Serfling, 1980; Arcones, 1995), we impose this restriction throughout for clarity. Similar conditions have been imposed elsewhere in the nonparametric literature to control the dependence structure, for example in Aït-Sahalia (1996a) and Robinson (1989). The following assumption (A0) is sufficient for $\{X_t\}$ to be well-defined, stationary and geometrically β-mixing. In particular, (A0) implies (S1)-(S2) given in the previous section.

A0 (i) The drift $\mu_0(\cdot)$ and diffusion $\sigma_0^2(\cdot) > 0$ are continuously differentiable, and (ii) there exists a function $V : \mathbb{R} \mapsto \mathbb{R}_+$ satisfying $V(x) \geq |x|^{\bar{q}}$ as $x \to l^+$ and r^- with $\bar{q} > 1$, and constants $b, c > 0$ such that

$$\mu_0(x) V'(x) + \frac{1}{2}\sigma_0^2(x) V''(x) \leq -cV(x) + b. \tag{4.12}$$

Under (A0), (i), there exists a unique solution to (4.1), cf. Karatzas and Shreve (1991, Theorem 5.5.15 and Corollary 5.3.23). The condition given in (4.12) is a so-called drift

[8] An alternative estimator would be $\hat{\sigma}^2(x) = 2\int_l^x \hat{\pi}(y) \mu(y;\hat{\theta}) dy / \hat{\pi}(x)$. The advantage of this is that it is continuous and differentiable. On the other hand, $\int_l^x \hat{\pi}(y) \mu(y;\theta) dy$ is a biased estimator of $\int_l^x \pi_0(y) \mu(y;\theta) dy$.

criterion, known from the ergodic theory for Markov chains. The function V is a norm-like function, and under (4.12), there exists $\rho \in (0,1)$ such that $E\left[V\left(X_{\Delta}\right)|X_0 = x\right] \leq \rho V\left(x\right) + b$, ensuring that the process is mean-reverting. This condition not only implies that the process is β-mixing with exponentially decaying mixing-coefficients, but also that $E_{\pi}\left[\left|X_0\right|^{\bar{q}}\right] < \infty$, where $E_{\pi}\left[\cdot\right]$ denotes the expectations operator w.r.t. the stationary measure of X. (A0) is based on results by Meyn and Tweedie (1992); alternative conditions for mixing of diffusion processes can be found in Chen et al. (1999), Hansen and Scheinkman (1995) and Veretennikov (1997); see also Karatzas and Shreve (1991, Section 5.5). Most parametric model found in the literature can be shown to satisfy (A0): Continuity and differentiability of μ and σ^2 are normally satisfied, and with $V\left(x\right) = x^q$, $q > 1$, the second condition becomes

$$q\mu\left(x\right)x + \frac{q\left(q-1\right)}{2}\sigma^2\left(x\right) \leq -cx^2, \tag{4.13}$$

as $|x| \to \infty$ (assuming $I = \mathbb{R}$). If for example $\mu\left(x\right) = \beta\left(\alpha - x\right)$, the condition becomes

$$\begin{aligned} \sigma^2\left(x\right) &\leq c_1 x^2 + c_2, \quad |x| \to \infty, \\ c_1 &\equiv \frac{2\left(q\beta - c_3\right)}{q\left(q-1\right)}, \; c_2 \equiv \frac{-2q\beta\alpha}{q\left(q-1\right)} \end{aligned}$$

with $0 < c_3 < q\beta$ where we require $\beta > 0$. This condition is satisfied for all the models with linear drift quoted in Table 6.1 when restricting the parameters in a suitable manner. Similarly, one may show that remaining models quoted in the table satisfy (4.13) under suitable parameter restrictions.

In some cases, one might want to have precise expressions of the convergence rate. This is for example the case in the next section where the convergence must take place at a sufficiently high rate. To speed up the convergence, we employ so-called higher order kernels in the estimation of $\pi^{(i)}$, allowing us to control the bias. We define the following class $\mathcal{K}\left(\omega,\lambda\right)$ of kernels first proposed by Parzen (1962) where $\omega, \lambda \geq 1$ are integers:

$\underline{\mathcal{K}\left(\omega,\lambda\right)}$ The kernel K satisfies $\int_{\mathbb{R}} K\left(x\right)dx = 1$; $\int_{\mathbb{R}} x^i K\left(x\right)dx = 0$, for $0 \leq i \leq \omega - 1$; $\int_{\mathbb{R}} |x|^{\omega} \left|K\left(x\right)\right| dx < \infty$; $K^{(i)}\left(x\right) \to 0$, $|x| \to \infty$, $0 \leq i \leq \lambda - 1$;
$\sup_x \left|K^{(i)}\left(x\right)\right| \max\left(|x|, 1\right) < \infty$, $0 \leq i \leq \lambda + 1$; $K^{(i)}$ is absolutely integrable with a Fourier transform Ψ_i satisfying $\int_{\mathbb{R}} \left(1 + |x|\right) \sup_{b \geq 1} \left|\Psi_i\left(bx\right)\right| dx < \infty$, $0 \leq i \leq \lambda$.

A discussion of the construction of specific kernels satisfying these conditions can be found in Bierens (1987). Using a kernel from this class makes it possible to reduce the bias of $\hat{\pi}$ and its derivatives, and thereby obtain a faster rate of convergence of $\hat{\pi}^{(i)}$, $0 \leq i \leq \lambda$. The smoothness of π_0 as measured by ω determines how much the bias can be reduced with. For the results stated in this section concerning the pointwise distribution and uniform consistency of the nonparametric estimators, bias reduction is not needed, and standard kernels can be used. But higher order kernels of the nonparametric estimators are useful when we need to get exact rates of convergence. This becomes relevant when proving $\sqrt{n}$-consistency of θ, see e.g. Robinson (1988) for an early application of higher order kernels to semiparametric estimation. Andrews (1995) gives uniform convergence rates of the density estimator and its derivatives using this type of kernels under fairly general conditions. We apply his results here even though the convergence rates stated there are not optimal. Masry (1996) obtains optimal convergence rates but only considers convergence on compact sets, while we wish to allow for a non-compact domain I. Similarly, Bosq (1998) establishes uniform consistency with a near optimal convergence rate on the whole of $\mathbb{R}$ for Markov processes, but estimators of the derivatives of the density are not considered. One could extend their results to hold on the whole of $\mathbb{R}$ and for density derivatives, but this is not the focus of this paper and we shall simply apply the results of Andrews (1995) here. The pointwise asymptotic distribution of $\hat{\pi}^{(r)}$ has been established in a number of papers, see e.g. Robinson (1983). Given the consistency and the asymptotic distribution of $\hat{\pi}^{(r)}$, the asymptotic properties of the two nonparametric estimators can now be derived using standard delta-methods.

One might also wish to have uniform convergence of the nonparametric estimators. This is for example needed in the next section when dealing with the asymptotics of $\hat{\theta}$. We wish to show uniform consistency of the nonparametric estimators in the supremum-norm. However, since the estimators and the limits themselves potentially are unbounded functions, this is not readily possible. To circumvent this problem, we control the tail behaviour of the estimator by trimming, ensuring that the nonparametric estimator equals zero outside a compact, but growing set. We define a sequence of sets $\hat{A} = \hat{A}_n$ by

$$\hat{A} = \{x|\hat{\pi}(x) \geq a\} \tag{4.14}$$

for some sequence $a = a_n \to 0$. We then show uniform convergence on the increasing set $\hat{A}$.

In addition to (A0), we impose the following assumptions:

A1 The true density, π_0, is ω times continuously differentiable on I with bounded derivatives.

A2 $\sqrt{n}(\hat{\theta} - \theta_0) = O_P(1)$.

The condition that π_0 is ω times continuously differentiable is satisfied if μ_0 and σ_0^2 are ω times continuously differentiable, cf. (4.2). Observe that all the models in Table 6.1 have infinitely differentiable drift and diffusion term so this is not a strong restriction for standard models. Since the rate of convergence of $\hat{\pi}$ and its derivatives is slower than $\sqrt{n}$, the asymptotic distribution of $\hat{\theta}$ will not have any effect on the ones of $\hat{\mu}$ and $\hat{\sigma}^2$. In particular, the efficiency of $\hat{\theta}$ is not important in this context. Condition (A2) can be weakened to allow for slower convergence rate of $\hat{\theta}$, as long as it is faster than $\sqrt{nh^3}$ ($\sqrt{nh}$) when estimating μ_0 (σ_0^2). If this is not the case, the asymptotic distribution of $\hat{\theta}$ will influence the one of the nonparametric estimator.

In the following let $\{x_i\}_{i=1}^N$ be a set of distinct points in the domain I, $x_i \neq x_j$ for $i \neq j$.

Theorem 1 (Class 1) *Assume that $K \in \mathcal{K}(\omega, 1)$, and (A0)-(A2) hold with $\omega \geq 3$; $\theta \mapsto \sigma^2(x;\theta)$ is continuously differentiable satisfying $||\partial_{x,\theta}^{ij}\sigma^2(x;\theta)|| \leq C(1 + |x|^{\bar{q}})$, $i,j = 0,1$; and $h_i \to 0$, and $nh_i^{2i+1} \to \infty$, $i = 0,1$. Then the nonparametric estimator of the drift is* *pointwise consistent and asymptotically normally distributed,*

$$\sqrt{nh_1^3}\left\{\hat{\mu}(x_1)-\mu_0(x_1)\right\}_{i=1}^N \xrightarrow{d} N(0,V_\mu),$$

where $V_\mu = \operatorname{diag}(\{V_\mu(x_i)\}_{i=1}^N)$ *is a diagonal matrix and* $V_\mu(x) = \frac{1}{4}||K^{(1)}||_2^2\sigma_0^4(x)/\pi_0(x)$. *Moreover,*

$$\sup_{x\in\hat{A}}|\hat{\mu}(x)-\mu_0(x)| = \sum_{i=0}^{1}\left\{O_P(n^{-1/2}a^{i-3}h_i^{-1-i})+O_P(a^{i-3}h_i^{\omega-i})\right\}.$$

Theorem 2 (Class 2) *Assume that* $K\in\mathcal{K}(\omega,0)$, *and (A0)-(A2) hold with* $\omega\geq 2$; $\theta\mapsto\mu(x;\theta)$ *is continuously differentiable, satisfying* $||\partial_\theta^i\mu(x;\theta)||\leq C\left(1+|x|^{\bar{q}/2}\right)$, $i=0,1$; $h_0\to 0$ *and* $nh_0\to\infty$. *Then the nonparametric estimator of the diffusion term is pointwise consistent and asymptotically normally distributed,*

$$\sqrt{nh_0}\left\{\hat{\sigma}^2(x_i)-\sigma_0^2(x_i)\right\}_{i=1}^N \xrightarrow{d} N(0,V_\sigma),$$

where $V_\sigma = \operatorname{diag}(\{V_\sigma(x_i)\}_{i=1}^N)$ *is a diagonal matrix with* $V_\sigma(x) = ||K||_2^2\sigma_0^4(x)/\pi_0(x)$. *Moreover,*

$$\sup_{x\in\hat{A}}\left|\hat{\sigma}^2(x)-\sigma_0^2(x)\right| = O_P(n^{-1/2}a^{-2}h_0^{-1})+O_P\left(a^{-2}h_0^{\omega}\right).$$

Pointwise estimators of the asymptotic variance for $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ respectively can be constructed as

$$\hat{V}_\mu(x) = \frac{1}{4}[\textstyle\int K'(y)^2\,dy]\sigma^4(x;\hat{\theta})/\hat{\pi}(x), \quad \hat{V}_\sigma(x) = [\textstyle\int K(y)^2\,dy]\hat{\sigma}^4(x)/\hat{\pi}(x). \qquad (4.15)$$

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

The pointwise rate of convergence of $\hat{\mu}$ might at first appear surprisingly slow given the interpretation of μ as the (instantaneous) conditional mean. In a standard nonparametric regression model, one is able to estimate the conditional mean with rate $\sqrt{nh}$, but observe that in our case μ is not only a functional of π alone but also of its derivative $\pi^{(1)}$ with the nonparametric estimator $\hat{\pi}^{(1)}$ having slower convergence rate than $\hat{\pi}$, $\sqrt{nh^3}$ relative to $\sqrt{nh}$. In contrast, we obtain the standard rate of convergence as found in kernel regressions for $\hat{\sigma}^2$. This owes to the fact that $\hat{\sigma}^2$ is only a function of $\hat{\pi}$ and not any of its derivatives. Thus, the drift is more difficult to estimate than the diffusion term in a nonparametric setting. This observation has been made elsewhere in the literature. Gobet et al. (2003) report similar results for their sieve-estimator, and coin the nonparametric estimation of μ given discrete observations as an "ill-posed problem". Similarly, Bandi and Phillips (2003) demonstrate that for a stationary diffusion, it is only possible to estimate $\mu(x)$

nonparametrically with $\sqrt{n\Delta h}$-rate, while $\sigma^2(x)$ can be estimated at the faster rate$\sqrt{nh}$ as $\Delta \to 0$ and $n\Delta \to \infty$.

The first part of the result stated in Theorem 2 has already been obtained by Aït-Sahalia (1996a) for the special case $\mu(x;\theta) = \beta(\alpha - x)$. So we here extend his result to hold for a more general class of semiparametric diffusion models.

Next, we set up a simple test for a parametric diffusion submodel against our semiparametric alternative. We start out with Class 1, for which we consider a parametric specification of the drift, $\mu(\cdot;\beta)$ for $\beta \in \mathcal{B} \subseteq \mathbb{R}^d$. We then wish to test the following nested hypothesis

$$H_{10} : \mu_0(\cdot) = \mu(\cdot;\beta_0) \text{ for some } \beta_0 \in \mathcal{B}$$

against the nonparametric alternative. Under the null the model is fully specified and the parameters (θ, β) can be estimated using standard methods, with the obvious one being MLE. Under regularity conditions, this will yield $\sqrt{n}$-consistent estimators of (θ, β). We base the our test statistic on the pointwise difference between the nonparametric and parametric estimate. For similar test procedures for conditional means, see Gozalo (1995, 1997) and Härdle and Mammen (1993). Under H_{10}, $\mu(x;\hat{\beta}) - \mu_0(x) = O_P(n^{-1/2})$, when a $\sqrt{n}$-consistent estimator $\hat{\beta}$ is available[9], and smoothness conditions are imposed on $\beta \mapsto \mu(x;\beta)$, such that

$$\sqrt{nh^3}\frac{\hat{\mu}(x) - \mu(x;\hat{\beta})}{\hat{V}_\mu^{1/2}(x)} = \frac{\hat{V}_\mu^{1/2}(x)}{V_\mu^{1/2}(x)}\sqrt{nh^3}\frac{\hat{\mu}(x) - \mu_0(x)}{V_\mu^{1/2}(x)} + o_P(1) \xrightarrow{d} N(0,1),$$

where $\hat{\mu}$ is the nonparametric estimator, while $V_\mu(x)$ and $\hat{V}_\mu(x)$ are given in Theorem 1 and Remark 1 respectively. Due to the asymptotic independence between $\hat{\mu}(x)$ an $\hat{\mu}(y)$ for any $x \neq y$, we are able to derive the distribution of the sum of squared differences across any given set of distinct points.

Theorem 3 (Class 1) *Assume that the conditions of Theorem 1 holds, and $\hat{\beta} - \beta_0 = O_P(n^{-1/2})$ under H_{10} with $\beta \mapsto \mu(x;\beta)$ being continuously differentiable. Then under H_{10},*

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

where $\hat{V}_\mu(x)$ is given in (4.15).

The above strategy can be applied to construct a test statistic in Class 2 for any hypothesis of the form

$$H_{20} : \sigma_0^2(\cdot) = \sigma^2(\cdot;\beta_0) \text{ for some } \beta_0 \in \mathcal{B}.$$

Theorem 4 (Class 2) *Assume that the conditions of Theorem 2 holds, and $\hat{\beta} - \beta_0 = O_P(n^{-1/2})$ under H_{20} with $\beta \mapsto \sigma^2(x;\beta)$ being continuously differentiable. Then under*

[9] One obvious estimator would be $\hat{\beta} = \arg\min_\beta n^{-1}\sum_{i=1}^{n}[\hat{\mu}(x) - \mu(x;\beta)]^2$, where $\hat{\mu}$ is the nonparametric estimator, c.f. Bandi & Phillips (1998).

H_{20},

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

where $\hat{V}_\sigma(x)$ *is given in (4.15).*

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

Instead of relying on the asymptotic distribution as an approximation of the finite-sample properties of T_n, it may be worthwhile to use bootstrapping since nonparametric goodness-of-fit tests appear to exhibit significant differences between nominal and true size in finite samples, see e.g. Fan (1994, 1995). It should be possible to show consistency of the bootstrap in our case by following her arguments.[10]

4.4 The Semiparametric Estimator

In this section we construct an estimator for θ and derive its asymptotic properties in each of the two classes of models. This is done along the lines proposed in Section 2, using the log-transition density to define our criterion function. For each class, we show that $\hat{\theta}$ is consistent, and converges weakly towards a normal distribution with $\sqrt{n}$-rate. The $\sqrt{n}$-consistency of $\hat{\theta}$ will then in turn imply that the results for the nonparametric estimators of μ and σ^2 stated in Theorem 1-4 are valid. The results stated in this section are established under the assumption that the domain $I = \mathbb{R}$. We conjecture that our results also hold for other domains by using arguments similar to those in Aït-Sahalia (2002). Allowing for such will however further complicate the proofs, since we need to give specific treatment to the boundary behaviour of $\{X_t\}$; in particular, the transition density will depend on the specified domain.

Drawing upon results of Dacunha-Castelle and Florens-Zmirou (1986), we are able to obtain an expression for $\log p$ (cf. Lemma 30) as a functional of μ and σ^2. This characterisation was also utilised by Aït-Sahalia (2002) in his derivation of an approximation of the likelihood-function. The log-density takes the following form,

$$\log p\left(x\,|x_0;\mu,\sigma^2\right) \propto -\frac{1}{4}\log\left[\sigma^2(x)\,\sigma^2(x_0)\right] - \left(\int_{x_0}^{x}\sigma^{-1}(w)\,dw\right)^2/2 + \log\left(E_B\left[\psi\left(x|x_0\right)\right]\right), \tag{4.16}$$

where

$$\psi(x\,|x_0) = \exp\left[\Delta\int_0^1 \lambda_Y\left(Z_t(x|x_0)\right)dt\right], \tag{4.17}$$

$$Z_t(x|x_0) = \gamma^{-1}\left(t\gamma(x) + (1-t)\,\gamma(x_0) + B_t\right), \tag{4.18}$$

[10] Fan (1994, 1995) only consider the i.i.d. bootstrap; in our setting a different bootstrap method have to be used, for example Horowitz (2003).

$$\lambda_Y(z) = -\frac{1}{2}\left[\mu_Y^2(z) + \partial_x \mu_Y(z)\,\sigma(z)\right], \tag{4.19}$$

$$\mu_Y(z) = \frac{\mu(z)}{\sigma(z)} - \frac{1}{4}\frac{\partial_x \sigma^2(z)}{\sigma(z)}, \tag{4.20}$$

$$\gamma(z) = \int \sigma(z)^{-1}\,dz. \tag{4.21}$$

and $\{B_t | 0 \leq t \leq 1\}$ is a standard Brownian Bridge with associated expectations operator $E_B[\cdot]$.[11] As can be seen, the function ψ depends on (μ, σ^2) in a fairly complicated way, so the analysis of $\log p$ as a functional of these is not straight forward. The analysis is further complicated by the presence of the Brownian Bridge in the expression of ψ.

To derive the asymptotic distribution, we need to modify the nonparametric estimators introduced in Section 2. As part of our proofs, we need to ensure that the nonparametric estimator converges uniformly towards a specified limit. In the previous section, we introduced a trimming set to ensure this. This technical device is widely used to establish theoretical results for semiparametric estimators, see for example Ai (1997) and Robinson (1988). However, the associated trimming function, $1\{\hat{\pi}(x) \geq a\}$ is discontinuous and nondifferentiable. So for technical reasons, we here follow the idea of Andrews (1995) and instead introduce a general trimming function, T, which is assumed to be continuous and differentiable. We assume that the function satisfies

$$T(x;\pi,a) = \begin{cases} 1, & \pi(x) \geq a \\ 0, & \pi(x) \leq a/2 \end{cases}, \tag{4.22}$$

for some sequence $a = a_n \to 0$.

$\underline{\mathcal{T}(\omega)}$ The function $T(x;\pi,a)$ (i) satisfies (4.22), (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.

The differentiability of T is assumed out of technical convenience; this property simplifies parts of our proofs. The speed with which a goes to zero will be restricted, so that the trimming has no effect on the asymptotics. Observe that the simplest choice of trimming function, $\hat{T} = 1\{\hat{\pi}(x) \geq a\}$, is not a member of $\mathcal{T}(\omega)$ for $\omega \geq 1$. One way of constructing a member of $\mathcal{T}(\omega)$ is to choose a cumulative density function, F, with support $[1/2, 1]$. Thus, $F(x) = 1$, $x \geq 1$, and $F(x) = 0$, $x \leq 1/2$, such that $T(x;\pi,a) = F(\pi(x)/a)$ satisfies (4.22). Under suitable conditions on F (and the density π), T belongs to $\mathcal{T}(\omega)$.

In the following we shall write $\hat{T}(x;a) = T(x;\hat{\pi},a)$ and $T_0(x;a) = T(x;\pi_0,a)$.

4.4.1 Class 1

In this subsection, we derive the asymptotic properties of $\hat{\theta}$ in Class 1. First, we redefine the nonparametric estimator of μ as follows:

$$\hat{\mu}(\cdot;\theta) = \frac{\hat{T}(x;a)}{2\hat{\pi}(x)}\frac{\partial}{\partial x}\left[\sigma^2(x;\hat{\theta})\hat{\pi}(x)\right].$$

[11] See Karatzas and Shreve (1991, p. 358-360) for a definition of the Brownian Bridge.

The use of the trimming function enables us to show that $||\hat{\mu}^{(i)}(\cdot;\theta) - \hat{\mu}_0^{(i)}(\cdot;\theta)||_\infty \to^P 0$, $i \geq 0$, where

$$\hat{\mu}_0(x;\theta) = \hat{T}(x;a)\,\mu_0(x;\theta), \quad \mu_0(x;\theta) = \frac{1}{2\pi_0(x)}\frac{\partial}{\partial x}\left[\sigma^2(x;\theta)\pi_0(x)\right].$$

We then propose the following estimator

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L_n(\theta, \hat{\mu}(\cdot;\theta)), \tag{4.23}$$

where

$$L_n(\theta,\mu) = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i|X_{i-1};\theta,\mu), \tag{4.24}$$

and $p(x|x_0;\theta,\mu) = p\left(x|x_0;\mu,\sigma^2(\cdot;\theta)\right)$ with $p\left(x|x_0;\mu,\sigma^2\right)$ given in (4.16). We observe that ψ only depends on μ and $\mu^{(1)}$, so when showing consistency of $\hat{\theta}$, we only need to show $||\hat{\mu}^{(i)} - \hat{\mu}_0^{(i)}||_\infty \to^P 0$, $i = 0, 1$. However, due to θ appearing in $Z_t(x|x_0;\theta)$, $\partial_\theta \log p$ depends on $\mu^{(i)}$, $i = 0,1,2$, and $\partial_\theta^2 \log p$ on $\mu^{(i)}$, $i = 0,1,2,3$. So in order to derive the asymptotic distribution of $\hat{\theta}$, we have to ensure that $||\hat{\mu}^{(i)} - \hat{\mu}_0^{(i)}||_\infty \to^P 0$, $i = 0,1,2,3$, and that convergence takes place with rate $n^{1/4}$.

We are now ready to set up the conditions, which we will work under.

C1.1 (A0)-(A1) holds with $\omega \geq 6$, the kernel $K \in \mathcal{K}(\omega, 4)$, and the trimming function $\hat{T} \in \mathcal{T}(4)$.

C1.2 The diffusion function $x \mapsto \sigma(x;\theta)$ is six times continuously differentiable for any $\theta \in \Theta$; $\theta \mapsto \sigma^2(x;\theta)$ is three times continuously differentiable for any $x \in \mathbb{R}$; $\underline{\sigma}^2 \leq \sigma^2(x;\theta)$, and $||\partial_{x,\theta}^{ij}\sigma^2(x;\theta)|| \leq \bar{\sigma}^2$, $0 \leq i \leq 4$ and $0 \leq j \leq 2$.

C1.3 (i) The drift function satisfies $||\partial_x^i \partial_\theta^j \mu_0(x;\theta)|| \leq C(1 + |x|^q)$, $0 \leq i \leq 6$ and $0 \leq j \leq 2$, with $4q + 2 + \delta \leq \bar{q}$ for some $\delta > 0$ where $\bar{q}$ given in (A0); (ii) $-C\left(1 + |z|^{\bar{q}}\right) \leq \lambda_Y(z;\theta,\mu_0(\cdot;\theta)) \leq \bar{\lambda}_Y$ uniformly in (x,θ).

C1.4 The parameter space $\Theta \subseteq \mathbb{R}^d$ is compact.

C1.5 The moment $L(\theta,\mu_0(\cdot;\theta)) = E_\pi\left[\log p(X_1|X_0;\theta,\mu_0(\cdot;\theta))\right]$ has a unique maximum at $\theta_0 \in \Theta$, such that

$$H(\theta_0,\mu_0) \equiv -E_\pi\left[\frac{\partial^2}{\partial\theta^2}\log p(X_1|X_0;\theta_0,\mu_0(\cdot;\theta_0))\right] \tag{4.25}$$

is positive definite.

C1.6a (i) $na^{2(i-4)}h_i^{2(1+i)} \to \infty$, (ii) $a^{i-4}h_i^{\omega-i} \to 0$,$\lambda = 0,1,2$, as $a, h_i \to 0$, for $i = 0,1,2$.

C1.6b (i) $na^{4(i-4)}h_i^{4(1+i)} \to \infty$ (ii) $na^{4(i-4)}h_i^{4(\omega-i)} \to 0$, (iii) $na^{4(4-i)}h_i^{1+i} \to 0$, and (iv) $a^{-2}h_i^{2(\omega-i)} \to 0$, for $0 \leq i \leq 3$; (v) $na^{-4}h_4^{10} \to \infty$, and (vi) $a^{-2}h_4^{\omega-4} \to 0$; (vii) $nP_{\pi\times B}(a/2 \leq \pi_0(Z_t) \leq a) \to 0$; (viii) $nP_\pi(a/2 \leq \pi_0(X_0) \leq a) \to 0$.

C1.7 The density $\bar{p}(z)$ given in (4.50) satisfies $\bar{p}^{(i)} = O(\pi_0^{(i)})$, $0 \leq i \leq 4$.

The smoothness criteria on π_0 in (C1.1) are used to ensure that the transition density is well-defined, and to decrease the bias from the kernel estimation. As discussed earlier, a high degree of smoothness together with the use of higher-order kernels will reduce the bias of the kernel estimator, cf. Lemma 31.

The smoothness assumptions on $\sigma^2(x;\theta)$ in (C1.2) is needed for the first and second derivative of $\log p$ w.r.t. θ to be well-defined. The boundedness conditions on $\sigma^2(x;\theta)$ and its derivatives are very restrictive. These bounds are primarily used to establish suitable bounds for the various terms entering $\log p$, in particular $E_B[\psi(x, x_0;\theta,\mu)]$. We conjecture that it should be possible to obtain these under weaker assumptions on σ^2, but this will complicate the proofs further. In practice the boundedness assumption should not be a problem, since one can always choose a parameterisation such that $\sigma^2(x;\theta)$ is constant outside a compact set, which can be chosen arbitrarily large.

The conditions (C1.1)-(C1.3) guarantee that the transition density exists (cf. Aït-Sahalia (2002), Proposition 2). There is some tension between (A0) and (C1.3), which both impose growth conditions on the drift function. In a fully parametric framework, (C1.4)-(C1.5) are standard assumptions when deriving the asymptotic properties of the MLE. In fact, if π_0 was known, the MLE of θ is consistent and asymptotically normally distributed under (C1.1)-(C1.5), cf. Aït-Sahalia (2002, Proposition 3 and Theorem 2).

Condition (C1.6a) and (C1.6b) restrict the choice of bandwidths and the trimming sequences to ensure that (a) $||\hat{\mu}^{(i)}(\cdot;\theta) - \hat{\mu}_0^{(i)}(\cdot;\theta)||_\infty \to^P 0$ and (b) $E_\pi[|\hat{\mu}_0^{(i)}(X_0;\theta) - \mu_0^{(i)}(X_0;\theta)|] \to 0$, together with derivatives w.r.t. θ, at a sufficiently fast rate. To prove consistency, we merely have to show that the convergence takes place, while $\sqrt{n}$-asymptotic normality requires that the convergence takes place with rate $n^{1/4}$. The first convergence creates a tension between a and h as they go to zero. Two bias and variance terms have to be controlled for: The one incurred from using $\hat{\pi}$ instead of π_0 in the estimation, which goes to zero as $h \to 0$, the other is caused by the trimming since the trimmed version of the score function may not equal zero for $a > 0$. One then has to balance the two effects to obtain consistency and asymptotic normality. For (b) to hold one needs that $a \to 0$, while for it to happen with rate $n^{1/4}$ we need to impose (C1.6b), (vii). The condition (C1.6b) is more restrictive than (C1.6a). When showing $\sqrt{n}$-asymptotic normality, further restrictions on the set of permissible bandwidth and trimming sequences are required since we now have to ensure that the two biases in (a) and (b) go to zero with a faster rate. The trimming bias can normally be avoided by introducing the trimming in such a way that the conditional expectations of $\partial_\theta \log p$ equals zero for any a as done in Ai (1997) and Robinson (1988). This does not appear to feasible here though since μ enters p in a complicated manner.

The measure $P_{\pi\times B}$ appearing in (C1.6b) is the product of the probability measures associated with the Brownian Bridge and $\{X_i\}$ respectively. The condition (C1.6b), (viii) will rely on the tail-thickness of the density π_0. To see this, we write

$$\begin{aligned} P_\pi\left(a/2 \le \pi_0\left(X_0\right) \le a\right) &= \int 1\left\{a/2 \le \pi_0\left(z\right) \le a\right\} \pi_0\left(z\right) dz \\ &\le \int 1\left\{\pi_0\left(z\right) \le a\right\} \pi_0\left(z\right) dz \\ &= \int 1\left\{\pi_0\left(z\right) \le a\right\} \pi_0^{\varepsilon}\left(z\right) \pi_0^{1-\varepsilon}\left(z\right) dz \\ &\le a^{\varepsilon} \int \pi_0^{1-\varepsilon}\left(z\right) dz \end{aligned}$$

for some $\varepsilon \in (0,1)$ such that $\int \pi_0^{1-\varepsilon}(z)\, dz < \infty$. The closer ε is to one, the thinner tails the density π_0 will have. So the tail-thickness of π_0 determines the rate with which a is allowed to go to zero.

The density $\bar{p}$ in (4.50) introduced in (C1.7) is implicitly given by $\int f(z)\bar{p}(z)\, dz = \int_0^1 E_{\pi\times B}\left[f\left(Z_t\left(X_1|X_0\right)\right)\right] dt$. The restriction imposed on $\bar{p}$ is used to ensure that the asymptotic variance of $\hat{\theta}$ is finite. It appears difficult to come up with primitive conditions for this to hold since $\bar{p}$ takes a very complex form.

In order to show consistency of the parametric part of our estimator, we basically have to demonstrate that the log-likelihood function is continuous w.r.t. μ in probability and that $\hat{\mu} \to^P \mu_0$ in a normed function space. Once this is established, standard consistency results for parametric estimators can be applied to $L_n\left(\theta, \mu_0\left(\cdot;\theta\right)\right)$, see e.g. Newey and McFadden (1994, Theorem 2.1) and Chen et al. (2003, Theorem 1). The following theorem establishes this result:

Theorem 5 *Assume that (C1.1)-(C1.6a) hold. Then* $\hat{\theta} \to^P \theta_0$.

Next, we show that $\hat{\theta}$ converges weakly towards a normal distribution with $\sqrt{n}$-rate. General conditions for this to hold are given in Andrews (1994), Chen et al. (2003, Theorem 2) and Newey and McFadden (1994). The results in Andrews (1994) only apply to the case where the initial nonparametric estimator does not influence the asymptotic variance however, which is not the case here. The other two studies deal with this situation, and we here follow their strategy.

The first and second derivative of $L_n\left(\theta, \mu\left(\cdot;\theta\right)\right)$ w.r.t. θ is denoted by

$$\begin{aligned} S_n\left(\theta,\mu\right) &= \frac{1}{n}\sum_{i=1}^{n} s\left(X_i|X_{i-1};\theta,\mu\right), \quad s\left(x|x_0;\theta,\mu\right) = \frac{\partial \log p\left(x|x_0;\theta,\mu\left(\cdot;\theta\right)\right)}{\partial\theta}; \\ H_n\left(\theta,\mu\right) &= \frac{1}{n}\sum_{i=1}^{n} h\left(X_i|X_{i-1};\theta,\mu\right), \quad h\left(x|x_0;\theta,\mu\right) = -\frac{\partial^2 \log p\left(x|x_0;\theta,\mu\left(\cdot;\theta\right)\right)}{\partial\theta\partial\theta'}. \end{aligned}$$

Expressions for s and h can be found in (4.31) and (4.37) respectively. We also introduce the pathwise derivative of s w.r.t. μ at (θ,μ) in the direction $d\mu$ which we denote $\nabla s\left(x|x_0;\theta,\mu\right)[d\mu]$ (see e.g. Bickel et al. (1993), Appendix 5 for an introduction to this concept), the corresponding sample version,

$$\nabla S_n\left(\theta_0,\hat{\mu}_0\right)[d\mu] = \frac{1}{n}\sum_{i=1}^{n} \nabla s\left(X_i|X_{i-1};\theta_0,\hat{\mu}_0\right)[d\mu].$$

and its moment, $\nabla S(\theta_0, \hat{\mu}_0)[d\mu] = E_\pi[\nabla s(X_i|X_{i-1};\theta_0, \hat{\mu}_0)[d\mu]]$. By Lemma 14, the pathwise derivative is well-defined, and satisfies

$$S_n(\theta_0, \hat{\mu}) - S_n(\theta_0, \hat{\mu}_0) - \nabla S_n(\theta_0, \hat{\mu}_0)[\hat{\mu} - \hat{\mu}_0] = o_P(n^{-1/2}).$$

Using standard U-statistics results for weakly dependent sequences, $\nabla S_n(\theta_0, \hat{\mu}_0)[\hat{\mu} - \hat{\mu}_0] = \nabla S(\theta_0, \mu_0)[\hat{\mu} - \hat{\mu}_0] + o_P(n^{-1/2})$ (cf. Lemma 16 and 17). Finally, $\nabla S(\theta_0, \mu_0)[\hat{\mu} - \hat{\mu}_0]$ can be written as a normed sum plus a remainder term with the latter being asymptotically negligible,

$$\nabla S(\theta_0, \mu_0)[\hat{\mu} - \hat{\mu}_0] = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i) + o_P(n^{-1/2}), \tag{4.26}$$

where $E_\pi[\delta(X_0)] = 0$ and $E_\pi\left[\|\delta(X_0)\|^2\right] < \infty$ (cf. Lemma 18). These results combined with the fact that $H_n(\hat{\theta}, \hat{\mu})$ converges towards $H(\theta_0, \mu_0)$ (cf. Lemma 19) proves the following result:

Theorem 6 *Assume that (C1.1)-(C1.7) hold, and that $\theta_0 \in \operatorname{int}(\Theta)$. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, H_0^{-1}(H_0 + V_0) H_0^{-1}\right),$$

where $H_0 = H(\theta_0, \mu_0)$, and $V_0 = \Omega_0 + 2\sum_{i=1}^{\infty} \Omega_i$ with $\Omega_i = E_\pi\left[\delta(X_0)\delta(X_i)^\top\right]$.

The extra term, V_0, in the variance expression is an adjustment term due to the use of $\hat{\pi}$ instead of π_0 in the estimation. If π_0 was known, $V_0 \equiv 0$, and the asymptotic variance expression would collapse to the standard inverse information matrix, H_0^{-1}. Instead, we here experience an increase in the asymptotic variance. The derivation of (4.26) is based on the Riesz Representation Theorem, and we therefore are not able to supply a closed form expression for δ. We are however able to show that it has mean zero and finite variance. Furthermore, it is possible to derive a consistent estimator of it, following the same strategy as in Newey (1994a). This estimator can in turn be used to obtain an estimator of the asymptotic variance by using the so-called HAC variance estimators, see e.g. Robinson and Velasco (1997). Here, we present an estimator based on the idea of Newey and West (1987).

Theorem 7 *Assume that (C1.1)-(C1.7) hold and $E_\pi[\|\delta(X_0)\|^{4+\delta}] < \infty$. Then consistent estimators of H_0 and V_0 respectively are given by $\hat{H}_n = H_n(\hat{\theta}, \hat{\mu})$ and*

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

where $w_{M,i} = 1 - [i/(M+1)]$, $\hat{\Omega}_i = n^{-1}\sum_{j=i}^{n} \hat{\omega}_j \hat{\omega}_{j-i}^\top$, $\hat{\omega}_j = \hat{\delta}_j - n^{-1}\sum_{k=1}^{n} \hat{\delta}_k$,

$$\hat{\delta}_j = \frac{1}{n}\sum_{k=1}^{n} \left.\frac{\partial s(X_k|X_{k-1};\hat{\theta}, \mu(\cdot;\hat{\theta}, \hat{\pi} + \alpha K_h(\cdot - X_j)))}{\partial\alpha}\right|_{\alpha=0},$$

and $M \to \infty$, $M/n^{1/8} \to 0$.

Observe that in the parametric framework of Newey and West (1987), it is required that $M_n/n^{1/4} \to 0$. We have to require that $M_n \to \infty$ at a slower rate due to the presence of the nonparametric part here, only exhibiting $n^{1/4}$-convergence rate. One advantage of the above variance estimator is its simple implementation; one can evaluate the variance estimator by numerical differentiation of $\log p\left(x|x_0; \theta, \mu\left(\cdot; \theta, \pi + \alpha K_h\left(\cdot - y\right)\right)\right)$ w.r.t. θ and α, instead of deriving the analytical derivatives (on the other hand, these may lead to superior numerical estimates). Since $E_\pi\left[\delta\left(X_0\right)\right] = 0$, one could leave out the average appearing in the expression for $\hat{\omega}_j$, but in finite sample this adjustment may improve on the performance of the estimator. An alternative to the variance estimator suggested here would be to construct one by either bootstrapping or subsampling, which should improve on the finite sample approximation; Hall (1992) and Politis, Romano and Wolf (1999) respectively provide in-depth treatment of these two methods. The recent work by Horowitz (2003), where a bootstrap method for Markov chains is suggested based on a kernel estimator of the transition density, is very well-suited for our framework. Since the sampled observations of the process $\{X_t\}$ indeed is a Markov chain, and we have here obtained a semiparametric estimator of the transition density, one should be able to adapt the results of Horowitz (2003) to our setting. Chen et al. (2003, Theorem B) give conditions for consistency of the bootstrap for a general class of semiparametric estimators. This is done under the assumption of i.i.d. observations, but combining their approach with Horowitz's results should yield the desired result for our estimator. The verification of this claim is out of the scope of this paper however.

Having obtained the estimator $\hat{\theta}$ in either of the two classes, one could now be interested in testing hypotheses concerning the parametric part, e.g. $H_0 : \theta = \theta_0$ for some given $\theta_0 \in \text{int}\Theta$. An obvious choice of test statistic for this hypothesis would appear to be the likelihood ratio,

$$T_n \equiv n\left[L_n(\hat{\theta}, \hat{\mu}(\cdot; \hat{\theta})) - L_n\left(\theta_0, \hat{\mu}\left(\cdot; \theta_0\right)\right)\right],$$

A general treatment of the semiparametric likelihood ratio test can be found in Murphy and Van der Vaart (1997) who show that under regularity conditions the likelihood-ratio converges towards a $\chi^2\left(p\right)$-distribution, where p is the dimension of θ. This is however not valid in our case. This owes to the fact here a preliminary estimator of the nonparametric part is used, while in Murphy and Van der Vaart (1997) the nonparametric part is estimated together with θ. This has strong implications for the asymptotic distribution of T_n. Instead one may use that $n\left(\hat{\theta} - \theta_0\right) H_0^{-1}\left(H_0 + V_0\right) H_0^{-1}\left(\hat{\theta} - \theta_0\right) \to^d \chi^2\left(p\right)$, or apply a GMM-type test statistic based on the score function, $S_n\left(\theta, \mu\right)$.

Given the $\sqrt{n}$-consistency of $\hat{\theta}$ as established above, Theorem 1 now establishes consistency and asymptotic normality of the nonparametric estimator of the drift, $\hat{\mu}\left(x\right) = \hat{\mu}(x; \hat{\theta})$. Note that the bandwidths used to estimate $\hat{\mu}$ should not be chosen to satisfy (C1.6). The bandwidth restrictions there were tailored to ensure a sufficiently fast convergence rate of both $\hat{\mu}$ and its first two derivatives while taking into account the trimming; this is not needed to prove pointwise asymptotic normality of $\hat{\mu}$.

4.4.2 Class 2

Here, we derive theoretical results for the estimator of θ for models in Class 2. Since our conditions, strategy of proof and results are very much the same as for Class 1, we will not give any thorough discussions of these, and instead refer to the previous section.

As in the previous section, we need to trim our estimator of $\sigma_0^2(x;\theta)$ to control for the tailbehaviour. Define

$$\hat{\sigma}^2(x) = \hat{T}(x;a)\frac{2}{\hat{\pi}(x)}\frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta) + (1-\hat{T}(x;a))\underline{\sigma}^2,$$

with $\hat{T}$ defined earlier. Observe that we here make sure that $\hat{\sigma}^2(x) \geq \underline{\sigma}^2$ for some lower bound $\underline{\sigma}^2 > 0$; this is needed since $\sigma^2(x)$ enters as a denominator in p. We furthermore define

$$\hat{\sigma}_0^2(x;\theta) = \hat{T}(y;a)\,\sigma_0^2(x;\theta) + (1-\hat{T}(x;a))\underline{\sigma}^2, \quad \sigma_0^2(x;\theta) = \frac{2}{\pi_0(x)}\int_l^x \pi_0(y)\,\mu(y;\theta)dy.$$

We are now able to establish $\left\|\hat{\sigma}^2(\cdot;\theta) - \hat{\sigma}_0^2(\cdot;\theta)\right\|_\infty \to^P 0$. In the estimation of the derivatives of σ^2 w.r.t. x, we cannot simply differentiate $\hat{\sigma}^2(x)$ because of the indicator function. Instead, we define

$$\partial_x\hat{\sigma}^2(x;\theta) \equiv \hat{T}(x;a)\left\{2\mu(x;\theta) - \frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}^2(x)}\frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta)\right\},$$

and similarly for higher order derivatives w.r.t. x.

We write $p\left(x|x_0;\theta,\sigma^2\right) = p\left(x|x_0;\mu(\cdot;\theta),\sigma^2\right)$ with $p\left(x|x_0;\mu,\sigma^2\right)$ as given in (4.16), and define our estimator as

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L_n\left(\theta,\hat{\sigma}^2(\cdot;\theta)\right),$$

$$L_n\left(\theta,\sigma^2\right) = \frac{1}{n}\sum_{i=1}^{n}\log p\left(X_i|X_{i-1};\theta,\sigma^2\right),$$

For the kernel estimator, we again use a higher order kernel of order $\omega \geq 5$. The following assumptions are imposed:

C2.1 (A0)-(A1) holds with $\omega \geq 6$, the kernel $K \in \mathcal{K}(\omega,4)$, and the trimming function $\hat{T} \in \mathcal{T}(4)$.

C2.2 (i) $\underline{\sigma}^2 \leq \sigma_0^2(x;\theta)$, and (ii) $||\partial_x^i\partial_\theta^j\sigma_0^2(x;\theta)|| \leq \bar{\sigma}^2$, $0 \leq i \leq 6$ and $0 \leq j \leq 2$.

C2.3 (i) The drift function satisfies $||\partial_x^i\partial_\theta^j\mu_0(x;\theta)|| \leq C(1+|x|^q)$, $0 \leq i \leq 6$ and $0 \leq j \leq 2$, with $4q+2+\delta \leq \bar{q}$ for some $\delta > 0$ where $\bar{q}$ given in (A0); (ii) $-C\left(1+|z|^{\bar{q}}\right) \leq \lambda_Y(z;\theta,\mu_0(\cdot;\theta)) \leq \bar{\lambda}_Y$ uniformly in (x,θ).

C2.4 The parameter space $\Theta \subseteq \mathbb{R}^d$ is compact.

C2.5 $L(\theta,\sigma_0(\cdot;\theta)) = E_\pi[\log p(X_1,X_0;\theta,\sigma_0(\cdot;\theta))]$ has a unique maximum at θ_0, and

$$H\left(\theta_0,\sigma_0^2\right) \equiv E_\pi\left[\frac{\partial^2}{\partial\theta\partial\theta'}\log p\left(X_1|X_0;\theta_0,\sigma_0^2(\cdot;\theta_0)\right)\right] \tag{4.27}$$

is non-singular.

C2.6a (i) $na^{2(i-5)}h_i^{2(1+i)} \to \infty$, (ii) $a^{i-5}h_i^{\omega-i} \to 0$ as a and $h_i \to 0$, $i = 0, 1, 2$.

C2.6b (i) $na^{4(i-5)}h_i^{4(1+i)} \to \infty$ (ii) $na^{4(i-5)}h_i^{4(\omega-i)} \to 0$, (iii) $na^{4(5-i)}h_i^{1+i} \to \infty$, (iv) $a^{-1}h_i^{\omega-i} \to 0$, $0 \le i \le 3$; (v) $na^{-6}h_4^{10} \to \infty$, (vi) $a^{-3}h_4^{\omega-4} \to 0$;

(vii) $nP_{\pi\times B}(a/2 \le \pi_0(Z_t) \le a) \to 0$; (viii) $nP_\pi(a/2 \le \pi_0(X_0) \le a) \to 0$.

C2.7 The density $\bar{p}(z)$ given in (4.50) satisfies $\bar{p}^{(i)} = O(\pi_0^{(i)})$, $0 \le i \le 4$.

The conditions are essentially the same as the ones imposed on the models in Class 1. Note that we here assume that the $\sigma_0^2(x;\theta)$ is bounded from below by $\underline{\sigma}^2$ which is known. The assumption that $\underline{\sigma}^2$ is known is used to simplify our proofs. One could allow for an unknown bound by introducing another trimming parameter $\underline{\sigma}_n^2 \to 0$. The proofs of consistency and asymptotic normality now proceed as for Class 1. First, the estimator is shown to be consistent:

Theorem 8 *Under (C2.1)-(C2.6a), $\hat{\theta} \to^P \theta_0$.*

We introduce the score $s\left(x|x_0;\theta,\sigma^2\right) = \partial_\theta \log p\left(x|x_0;\theta,\sigma^2(\cdot;\theta)\right)$, and the pathwise derivative of the score $s\left(x|x_0;\theta,\sigma^2\right)$ w.r.t. σ^2 in the direction $d\sigma^2$, $\nabla s\left(x|x_0;\theta,\sigma^2\right)\left[d\sigma^2\right]$ which will be used in the derivation of the asymptotic distribution. Using the same notation as in the previous section, we have that $\nabla S_n\left(\theta_0,\hat{\sigma}^2\right)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right] = \nabla S\left(\theta_0,\sigma_0^2\right)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right] + o_P\left(n^{-1/2}\right)$, cf. Lemmas 25 and 26. Furthermore, $\nabla S\left(\theta_0,\sigma_0^2\right)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right]$ can be written as a sum and a remainder term with the latter being asymptotically negligible (Lemma 27),

$$\nabla S\left(\theta_0,\sigma_0^2\right).\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right] = \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + o_P\left(1/\sqrt{n}\right).$$

It should be noted, that the function δ here is not identical to the δ-function appearing in Class 1. Finally, the Hessian $h\left(x|x_0;\theta,\sigma^2\right) = \partial_{\theta\theta}^2 \log p\left(x|x_0;\theta,\sigma^2(\cdot;\theta)\right)$ satisfies $H_n(\hat{\theta},\hat{\sigma}^2) \to^P H\left(\theta_0,\sigma_0^2\right)$, cf. Lemma 28. We are able to conclude:

Theorem 9 *Assume that (C2.1)-(C2.7) hold and that $\theta_0 \in \operatorname{int}(\Theta)$. Then the conclusions of Theorem 6 hold for Class 2 with $H_0 = H\left(\theta_0,\sigma_0^2\right)$.*

We also obtain a consistent estimator of the asymptotic variance:

Theorem 10 *Assume that (C2.1)-(C2.7) hold and $E_\pi[\|\delta(X_0)\|^{4+\delta}] < \infty$. Then consistent estimators of H_0 and V_0 respectively are given by $\hat{H}_n = H_n(\hat{\theta},\hat{\sigma}^2)$ and $\hat{V}_n$ given as in Theorem 7 with*

$$\hat{\delta}_j = \frac{1}{n}\sum_{k=1}^{n}\left.\frac{\partial s(X_k|X_{k-1};\hat{\theta},\sigma^2(\cdot;\hat{\theta},\hat{\pi}+\alpha K_h(\cdot - X_j)))}{\partial\alpha}\right|_{\alpha=0},$$

and $M \to \infty$, $M/n^{1/8} \to 0$.

Having obtained $\sqrt{n}$-consistency of $\hat{\theta}$, Theorem 2 establishes pointwise consistency and asymptotic normality of the nonparametric estimator of $\sigma_0^2(x)$.

4.5 Semiparametric Efficiency

As observed in the previous section, our semiparametric estimator is not adaptive in the general case since $V_0 > 0$. A natural question to ask is whether it at least reaches the semiparametric efficiency bound. We are unfortunately not able to give a rigorous answer to this, but due to the nature of our estimator we conjecture this is not the case. We furthermore propose a one-step adjustment to our estimator $\hat{\theta}$ which we conjecture will reach the bound in any circumstance.

For a semiparametric model, Stein (1956) defined the semiparametric efficiency bound as the "least favourable" parametric subproblem of the original semiparametric problem. The Fisher information of the semiparametric problem is obviously no greater than the information of any parametric subproblem. The semiparametric efficiency bound is then defined as the lower bound of the information of all parametric subproblems.

In our setting, the nonparametric part is π, so we therefore consider any smooth parameterisation of π for which the associated Fisher information may be derived. In the following assume for simplicity that θ is one-dimensional, and consider a smooth parameterisation of π, $\theta \mapsto \pi_\theta$, with $\pi_{\theta_0} = \pi_0$. The Fisher information of this subproblem is then given by

$$I(\dot{\pi}) = E_\pi \left[s_0 (X_1 | X_0) [\dot{\pi}]^2 \right]$$

where $s_0 [\dot{\pi}] = s(\theta_0, \pi_0) [\dot{\pi}]$, $s(\theta, \pi) [\dot{\pi}] = \partial_\theta \log p(y|x; \theta, \pi_0) + \nabla_\pi \log p(y|x; \theta, \pi_0) [\dot{\pi}]$, with $\nabla_\pi \log p$ denoting the pathwise derivative of the log-density w.r.t. π at (θ, π), and $\dot{\pi} = \partial_\theta \pi_\theta|_{\theta=\theta_0}$ is the tangent vector of the curve $\theta \mapsto \pi_\theta$ at θ_0. The space of tangent vectors/nuisance scores is given by $\mathcal{S} = \left\{ \dot{\pi} \in L_2(I) \mid \int_I \dot{\pi}(z)\, dz = 0 \right\}$, and we denote the closure of $\mathcal{S}$ by $\bar{\mathcal{S}}$. A tangent vector $\dot{\pi}^* \in \bar{\mathcal{S}}$ is then called the least favourable direction if

$$I(\dot{\pi}^*) = \inf_{\dot{\pi} \in \bar{\mathcal{S}}} I(\dot{\pi}),$$

and $I^{-1}(\dot{\pi}^*)$ is the semiparametric efficiency bound. The associated score function $s_0^* = s_0 [\dot{\pi}^*]$ is called the efficient score function, and any parameterisation $\theta \mapsto \pi_\theta$ which satisfies $\partial_\theta \pi_\theta|_{\theta=\theta_0} = \dot{\pi}^*$ is called a least favourable model. Observe that $\nabla_\pi \log p_0 [\dot{\pi}^*]$ is the projection of $-\partial_\theta \log p_0$ onto the space $\left\{ \nabla_\pi \log p_0 [\dot{\pi}] \,|\, \dot{\pi} \in \bar{\mathcal{S}} \right\}$. This characterisation was utilised in Severini and Tripathi (1999) to calculate the efficiency bound in a number of semiparametric problems.

But it appears problematic to find $\dot{\pi}^*$ in our case, since $s[\dot{\pi}]$ takes a form which is very difficult to analyse, cf. (4.31). And even if we were able to find $\dot{\pi}^*$, $I^{-1}(\dot{\pi}^*)$ would be difficult to compare to the variance of $\hat{\theta}$ derived in the previous section since we have no closed form expression for V_0. Instead, we construct a one-step estimator which is designed to reach the efficiency bound. Given the estimator proposed in the previous section, we perform a one-step Newton-Raphson iteration using an estimate of the efficient score. The resulting estimator will be semiparametric efficient. This procedure is very much a generalisation of the one-step Newton-Raphson estimator found in the fully parametric literature: An initial $\sqrt{n}$-consistent estimator is adjusted by the estimated score function making the resulting estimator efficient. This procedure has also been used in the semiparametric literature, see for example Drost and Klaassen and Werker (1997).

The main problem is to obtain an estimator of the efficient score. Here, we rely on the literature on semiparametric profile estimation. As mentioned earlier, a number of studies have developed a general theory for semiparametric profile likelihood estimators; see for example Wong and Severini (1991), Severini and Wong (1992) and Murphy and Van der Vaart (1997, 2000). A very nice property of these estimators is that, under regularity conditions, they reach the semiparametric efficiency bound. In the following, we first introduce the semiparametric profile estimator for our specific problem, and then define a one-step estimator of θ based on the profile likelihood which is computationally less demanding than the actual profile estimator. : There exists $\theta \mapsto \pi_\theta$ satisfying $\dot{\pi}_\theta = \dot{\pi}^*$;

The profile likelihood estimator is defined as

$$\tilde{\theta} = \arg\max_{\theta\in\Theta} L_n\left(\theta, \hat{\pi}_\theta\right),$$

where $\hat{\pi}_\theta = \arg\max_{\pi\in\Pi} L_n\left(\theta, \pi\right)$, and $\Pi \subseteq \left\{\pi \geq 0 | \int_I \pi\left(x\right) dx = 1\right\}$ is a subspace of all densities. Intuitively, this estimator should perform better than our estimator, $\hat{\theta}$. The latter is based on a fixed initial estimator $\hat{\pi}$, while the profile estimator relies on an estimator $\hat{\pi}_\theta$ which adjusts to θ. The profile estimator should then reach the semiparametric efficiency bound. General conditions for this can be found in Murphy and Van der Vaart (2000). These are: (i) there exists a least favourable model, (ii) $\hat{\pi}_{\theta_n} \to^P \pi_0$, and (iii)

$$E_\pi\left[s\left(\theta_0, \hat{\pi}_{\theta_n}\right)\left[\widehat{\dot{\pi}}_{\theta_n}\right]\right] = o_P\left(\|\theta_n - \theta_0\|\right) + o_P\left(n^{-1/2}\right),$$

for any any random sequence $\theta_n \to^P \theta_0$. If we start with (ii), since Π is an infinite-dimensional space, this condition is not easily verified.[12] One solution to this problem is to apply the method of sieves :[13] For each $n \geq 1$, let Π_n be a finite-dimensional space of densities with support I such that the sequence $\{\Pi_n\}$ grows dense in Π as $n \to \infty$. We then redefine $\hat{\pi}_\theta$ as $\hat{\pi}_\theta = \arg\max_{\pi\in\Pi_n} L_n\left(\theta, \pi\right)$. Under regularity conditions, $\hat{\pi}_\theta \to^P \pi_\theta$ for any given $\theta \in \Theta$, where

$$\pi_\theta = \arg\max_{\pi\in\Pi} E_\pi\left[\log p\left(\theta, \pi\right)\right]; \tag{4.28}$$

see for example Chen and Shen (1998). Sufficient conditions for (ii) to hold is then (a) $\theta \mapsto \pi_\theta$ is a continuous mapping and (b) $\sup_{\theta\in\Theta} \left\|\hat{\pi}_\theta - \pi_\theta\right\|_\infty \to^P 0$. These will hold if $\{\hat{\pi}_\theta | \theta \in \Theta\}$ is stochastically equicontinuous, cf. Newey (1991). The curve defined in (4.28) is moreover a natural candidate for the least favourable model in (i). Condition (iii) is a smoothness condition on the score function.

While we expect $\tilde{\theta}$ to reach the efficiency bound, it is much more computationally burdensome than $\hat{\theta}$ since at each given value of θ we have to perform a high-dimensional optimisation routine over Π_n in order to obtain $\hat{\pi}_\theta$. A computationally attractive alterna-

[12] The estimator $\hat{\pi}_\theta$ might not even be well-defined and, even if it is, very difficult to compute.
[13] See Chen (2004) for an overview of this method.

tive to $\tilde{\theta}$ is the following one-step adjustment estimator,

$$\check{\theta} = \hat{\theta} + H_n^{-1}(\hat{\theta}, \hat{\pi}_{\hat{\theta}}) S_n(\hat{\theta}, \hat{\pi}_{\hat{\theta}}),$$

where $\hat{\pi}_{\theta}$ is defined as before and $\hat{\theta}$ is the estimator considered in the previous section. The adjustment term is basically a Newton-Raphson iteration. Under the regularity conditions in Murphy and Van der Vaart (2000), $S_n(\hat{\theta}, \hat{\pi}_{\hat{\theta}}) = n^{-1} \sum_{i=1}^{n} s_0^* (X_i|X_{i-1}) + o_P(n^{-1/2})$. Given this, it should be possible to show that $\check{\theta}$ has the desired asymptotic properties, see Bickel et al (1993, Section 7.8). We shall not pursue this any further here, and leave the proof of this conjecture to future research.

4.6 Implementation

In this section we discuss the implementation of the estimator. As mentioned earlier, the transition density p does not in general have a closed form expression, and so one can not directly evaluate it. Instead, a number of different suggestions for how to either approximate or simulate it have been proposed in the literature. Lo (1988) observes that p solves a linear 2nd order partial differential equation, and suggests the application of numerical methods to solve it and thereby obtain p. A closed-form approximation of p can be found in Aït-Sahalia (2002), derived by using Edgeworth-expansion-type arguments. Durham and Gallant (2002), Elerian et al. (2001), Hurn et al (2003), Nicolau (2002) and Pedersen (1995) all consider simulation-based maximum-likelihood. Either of the above methods can be applied to our estimator. As mentioned in the previous section, in the implementation of the above mentioned methods evaluation of γ is not required, except for the method of Nicolau (2002).

An important part of the estimator is the choice of bandwidths and trimming parameter. An obvious way of choosing the bandwidths would be cross-validation methods, see Härdle et al. (1990); other options are rule-of-thumb and plug-in methods, see Silverman (1986) for a discussion of these and related methods. Most existing methods however are designed to minimise the mean square error, while the conditions imposed on the set of bandwidths when deriving asymptotic normality of $\hat{\theta}$ require them to be of a different order. So the above methods do not appear to be directly applicable in our case. This is demonstrated in Härdle et al. (1992) where results for the optimal bandwidth choice for the average derivative estimator is derived; it is shown that the optimal bandwidths used in the semiparametric estimation are not equivalent to the ones minimising the mean squared error. Powell and Stoker (1996) extend their results to other semiparametric problems. Data-driven methods to obtain bandwidths in semiparametric estimation are yet to be derived however. It is outside the scope of this study to construct a bandwidth choice method tailored to our application of the kernel estimator. Newey (1994a) suggests that in practice a good method would be to start with the standard cross-validated choice of bandwidth, and then decrease it until $\hat{\theta}$ does not change too much. Another rule-of-thumb method is the following: For $||\hat{\pi}^{(i)} - \pi_0^{(i)}||_\infty = o_P(n^{-1/4})$ to hold, we require that $n^{\frac{1}{4(1+i)}} h_i \to \infty$ and $n^{\frac{1}{4(\omega-i)}} h_i \to 0$. Restricting h_i to $h_i = cn^{-q_i}$, these restrictions can be written as $n^{\frac{1}{4(1+i)}-q_i} \to \infty$, and $n^{\frac{1}{4(\omega-i)}-q_i} \to 0$. Thus, we require that $\frac{1}{4(\omega-i)} < q_i < \frac{1}{4(1+i)}$, which holds if $\omega > 2i+1$. If this is satisfied, the optimal choice is $h_i = cn^{-\frac{1}{4(1+i)}} \log(n)$. Using some data-driven method minimising the MSE, we obtain $h_i^* = O\left(n^{-1/(2i+5)}\right)$. One way of choosing h_i in an application is then as

$$h_i = h_0^* n^{\frac{1}{5}-\frac{1}{4(1+i)}} \log(n) = h_0^* n^{\frac{4i-1}{20(1+i)}} \log(n),$$

or alternatively $h_i = h_i^* n^{\frac{1}{2i+5}-\frac{1}{4(1+i)}} \log(n)$.

Various studies suggest that the dependence structure of the available data will affect the performance of the kernel estimators in finite samples. In particular, strong dependence will deteriorate the finite sample performance. Hall et al. (1995) give theoretical results concerning robustness of the cross-validation procedure towards dependence, while Pritsker (1998), who reconsidered the work of Aït-Sahalia (1996b), demonstrates that for the data set used there, the asymptotic distribution of the marginal density estimator provided an unsatisfactory approximation of the finite-sample distribution. It appeared that the major problem was the strong dependence between the observations, which may slow down the convergence of the kernel density estimator; in such cases the use of the asymptotic distribution is not appropriate for tests and confidence bands.[14] The same problem is reported by Chapman and Pearson (2000) who also give evidence of potential boundary problems of the kernel estimates.

Another potential problem with the performance of our estimator is that kernel estimators of density derivatives appear to be systematically biased in finite sample. Stoker (1993) give theoretical evidence of a systematic bias towards zero of these, and suggests a method for correcting for this bias in weighted average derivative estimators; his results are generalised by Newey et al. (1992), see also Newey et al (2004). These results indicate that great care should be taken when choosing the bandwidths, and that the use of the asymptotic distribution to approximate the finite sample distribution of non- and semi-parametric estimators may not be a terribly good idea. In the worst case, the estimator of the drift and diffusion term may be heavily biased. This lends support to the application of bootstrap methods when conducting inference.

4.7 A Simulation Study

In this section we present results from a small simulation study. The simulation study demonstrates that the estimator performs well for moderate sample sizes, suggesting that the concerns put forward in the previous section may not be so relevant.

The estimator is implemented using the approximation of p suggested in Aït-Sahalia (2002). Let $p^{(J)}$ denote the approximation where $J \geq 1$ is an integer; as $J \to \infty$, $p^{(J)}(x\,|x_0) \to p(x\,|x_0)$ uniformly over (x, x_0) on any compact set under regularity condi-

[14] In a different context, Fan (1994) argues that rather the problems are caused by big differences between true and nominal sizes of the test in finite sample.

tions on μ and σ^2, cf. Aït-Sahalia (2002, Theorem 1). The Jth approximation requires the evaluation of the J first derivatives of μ and σ^2.

We choose a model in Class 1, so in order to implement the procedure for a given data set, we perform the following three-step procedure:

1. Obtain $\hat{\pi}^{(i)}$ for some kernel K and bandwidth h_i, $0 \leq i \leq J+1$.

2. Obtain $\hat{\theta}$ using the approximate MLE method of order$J$ with $\hat{\pi}^{(i)}$, $0 \leq i \leq J+1$, plugged in.

3. Calculate $\hat{\mu}(x) = \mu(x; \hat{\theta}, \hat{\pi})$.

We have 4 parameters ,which have to be chosen to run the above procedure: The kernel K, the bandwidth h, the trimming parameter a, and the approximation order J. One would expect that there would be a trade off between the size of J and the estimation of π: As J goes to infinity, the approximate likelihood approaches the true one; on the other hand, in the actual implementation a large value of J requires a large number of derivatives of π to be estimated. However, in the simulation study it was found that the estimator was very stable towards the choice of J. It appears that the higher order terms of the approximation (and thereby the estimation of higher order derivatives) are not terribly important.[15] On the other hand, one has to be careful with the choice of the bandwidths for $\hat{\pi}^{(i)}$, $0 \leq i \leq 3$; we found that the estimator of θ was relatively sensitive towards choice of h_i, $0 \leq i \leq 3$. The trimming parameter was chosen such that only observations between the 2.5th and 97.5th the empirical percentile were included in the estimation of θ; the full data set was used in the preliminary estimation of π and its derivatives however. We tried out other percentiles in the range 0-5 and 95-100 respectively without any significant changes in the results. We also tried out various kernels, finding that the performance of the estimator appears to be very robust towards the choice of kernel. In particular, in practice higher order kernels did provide any significant improvement on the performance of the estimator.

The model we simulate from is the so-called CIR- model suggested by Cox et al. (1985),

$$\begin{aligned} dX_t &= \mu(X_t)\,dt + \theta\sqrt{X_t}dW_t, \\ \mu(x) &= 0.5\,(0.08 - x)\,, \end{aligned}$$

with $\theta = \sqrt{0.02} = 0.1414$. The specification of μ and σ^2 ensures that the data generating process is stationary. In the estimation, $\mu(\cdot)$ and θ are the unknown parameters of interest. We set the time distance between observations to $\Delta = 1/12$. The advantage of this model is that the transition density is known so we can perform actual MLE when we allow ourselves to use the information that $\mu(x) = 0.5\,(0.08 - x)$. This allows us to compare the semiparametric and actual MLE. We simulate n observations of the process using the standard Euler scheme. For each data set we then go through the steps 1.-3. given above. We employ the second rule-of-thumb method suggested in the previous section to choose

[15] But this may be specific to the model considered here.

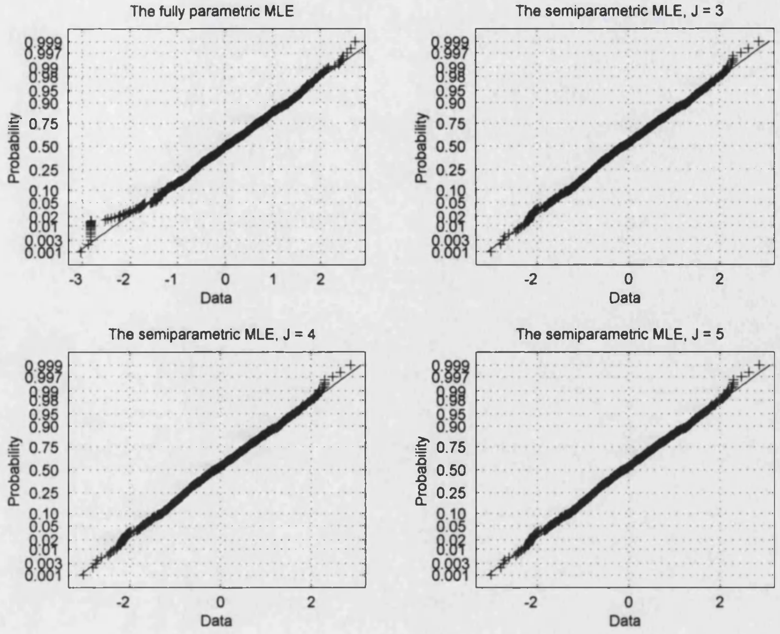

FIGURE 4.1. Normal probability plot of the estimators with $n = 500$.

the bandwidths for each data set. The results reported below could probably be improved upon by using data-driven bandwidth selection procedures, but our rule-of-thumb method seems to do a good job.

We simulate 1000 data sets, where each data set consists of $n = 500$ observations, and we choose $J = 3, 4, 5$. We consider the semiparametric estimator $\hat{\theta}_{500}^{(J)}$ for each of the 3 choices of J and also the actual fully parametric MLE, $\hat{\theta}_{0,500}$. In Figure 4.1, we have for each of the estimators made a QQ-plot of its empirical distribution against a $N\left(\theta, \hat{s}^2\right)$ distribution where $\hat{s}^2$ is the estimator's empirical variance. As can be seen, there are some slight problems with $\hat{\theta}_{0,500}$ in the left tail, which owes to numerical problems in the optimisation procedure. But the semiparametric estimator performs remarkably well with the choice of J having a negligible effect on the performance. The estimators are close to being unbiased: $\hat{\theta}_{0,500}$ has empirical mean 0.1416 and std. 0.050 while $\hat{\theta}_{500}^{(J)}$ has empirical mean 0.1410 (0.0047), 0.1412 (0.0049) and 0.1413 (0.0049) for $J = 3, 4, 5$. This indicates that for this specific model, the adjustment term δ is small.[16] Next, we report on the nonparametric part, μ. In Figure 4.2, the estimated drift for $J = 5$ is plotted together with the actual drift. As can bee seen, the estimator is biased but the true drift lies within its 95% confidence bands in the major part of the domain. However, in the left tail there appears to be problems. As expected, the estimator of the nonparametric part does not perform as well as the parametric part due to the slower convergence rate of the former.

[16] This may be due to the aforementioned numerical problems experienced with the MLE though, since these increase its empirical variance.

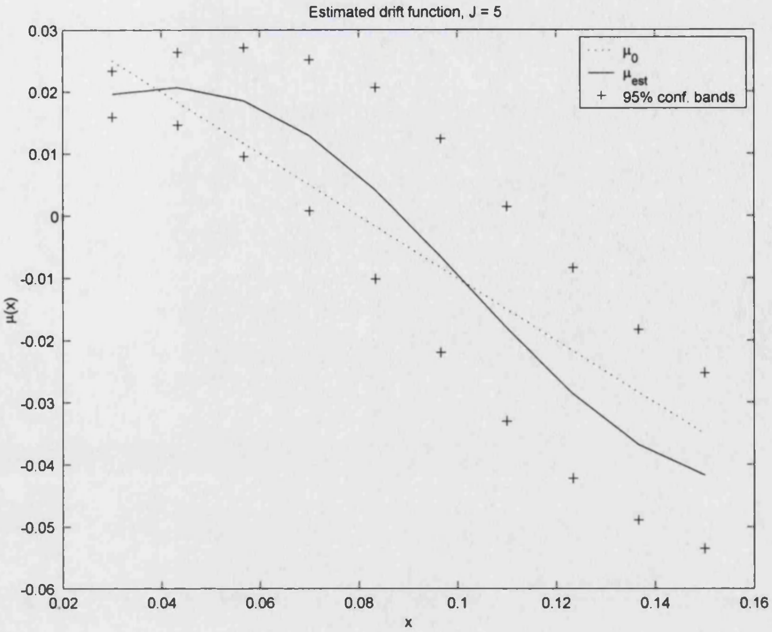

FIGURE 4.2. Estimated drift function with 95% confidence bands, $n = 500$.

4.8 Conclusion

We have considered two broad classes of diffusion models where either the drift or the diffusion term of the model is left unspecified while the other is specified up to a finite-dimensional parameter. Under the assumption of stationarity, estimators of both the parametric and nonparametric part were proposed, and their asymptotic properties were derived. We suggested that in the practical implementation of the estimator, approximate or simulation-based methods should be applied. Under suitable conditions these will have asymptotically negligible effects on the performance of the estimator. A small simulation study was carried out which supported the theoretical results. An approximation of the transition density was implemented, and we found that the choice of the approximation order had a small impact on the performance of the estimator, while the choice of bandwidths appeared to be important.

Various issues and extensions related to this study could be of interest to investigate in future research. As observed earlier, one may wish to allow for weak non-stationarity of the processes. Another important extension would be to consider multivariate diffusion models. It could also be of interest to consider other types of semiparametric continuous-time models. Here, we have restricted the noise process driving the SDE to be a Brownian motion, while allowing for an unspecified drift or diffusion term. One could take the alternative approach of specifying the two while leaving the noise process unspecified. This approach is pursued by Werker et al. (2000) where an extended version of the Vasicek (1977) model is considered. Finally, the issue of semiparametric efficiency was only discussed heuristically here; rigorous results in this area for the two classes of diffusion models are yet to be derived.

4.A Proofs

Proof of Theorem 1. We have

$$\begin{aligned}\hat{\mu}(x)-\mu_0(x) &= \frac{1}{2}\sigma^2(x;\theta_0)\left[\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)}-\frac{\pi_0^{(1)}(x)}{\pi_0(x)}\right] \\ &+\frac{1}{2}\left[\partial_x\sigma^2(x;\hat{\theta})-\partial_x\sigma^2(x;\theta_0)\right]+\frac{\hat{\pi}^{(1)}(x)}{2\hat{\pi}(x)}\left[\sigma^2(x;\hat{\theta})-\sigma^2(x;\theta_0)\right],\end{aligned}$$

where

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

for some $\bar{\theta}_i\in[\theta_0,\hat{\theta}]$, $i=0,1$, while

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

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

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

where $V_\pi=\text{diag}(\{V_\pi(x_i)\}_{i=1}^N)$ with $V_\pi(x)=\pi_0(x)\,||K^{(1)}||_2^2$, while the two remainder terms are $o_P(1)$, c.f. Lemma 31. The first part of the theorem now follows from Slutsky's Theorem. The uniform convergence result is obtained by combining the proof of Lemma 33 with 31. ■

Proof of Theorem 2. By Lemma 37 and arguments similar to the ones of the previous proof,

$$\hat{\sigma}^2(x)-\sigma_0^2(x)=2\int_l^x\mu(y;\theta_0)\pi_0(y)\,dy[\frac{1}{\hat{\pi}(x)}-\frac{1}{\pi_0(x)}]+O_P(n^{-1/2}),$$

where

$$\frac{1}{\hat{\pi}(x)}-\frac{1}{\pi_0(x)}=-\frac{1}{\pi_0^2(x)}\left[\hat{\pi}(x)-\pi_0(x)\right]+\frac{[\hat{\pi}(x)-\pi_0(x)]^2}{4\left(\lambda\hat{\pi}(x)+(1-\lambda)\pi_0(x)\right)^3},$$

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

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

where $V_\pi=\text{diag}(\{V_\pi(x_i)\}_{i=1}^N)$ with $V_\pi(x)=\pi_0(x)\,||K||_2^2$, while

$$\hat{\pi}(x)-\pi_0(x)=O_P(n^{-1/2}h^{-1})+O_P(h^\omega).$$

Slutsky's Theorem now gives the claimed asymptotic distribution. The uniform convergence result is established by combining Lemma 38 and 31. ■

Proof of Theorem 5. We are allowed to disregard the term $\sqrt{\pi(x)/\pi(x_0)}$ appearing in (4.56), since this does not depend on θ. Thus,

$$\hat{\theta} = \arg\min_{\theta\in\Theta} Q_n\left(\theta, \hat{\mu}\left(\cdot;\theta\right)\right)$$

where $Q_n(\theta,\mu) = \frac{1}{n}\sum_{i=1}^{n} q\left(X_i|X_{i-1};\theta,\mu\right)$ and

$$q\left(x|x_0;\theta,\mu\right) = -\frac{1}{4}\log\left(\sigma^2\left(x;\theta\right)\sigma^2\left(x_0;\theta\right)\right) - \frac{1}{2\Delta}\left(\int_{x_0}^{x}\sigma\left(w;\theta\right)^{-1}dw\right)^2 + \log E_B\left[\psi\left(x|x_0;\theta,\mu\right)\right].$$

We wish to show that 1) $\sup_{\theta\in\Theta}|Q_n(\theta,\hat{\mu}) - Q_n(\theta,\hat{\mu}_0)| \to^P 0$; 2) $\sup_{\theta\in\Theta}|Q_n(\theta,\hat{\mu}_0) - Q_n(\theta,\mu_0)| \to^P 0$; 3) $\sup_{\theta\in\Theta}||Q_n(\theta,\mu_0) - Q(\theta,\mu_0)|$, where $Q(\theta,\mu) = E_\pi\left[q\left(X_1|X_0;\theta,\mu\right)\right]$; and 4) $\theta \mapsto Q(\theta,\mu_0)$ is continuous with a unique maximum at θ_0.

To prove 1), write

$$\begin{aligned} q\left(x|x_0;\theta,\hat{\mu}\right) - q\left(x|x_0;\theta,\hat{\mu}_0\right) &= \log E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right] - \log E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right] \\ &= \log\left(\frac{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right]}{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right]}\right). \end{aligned}$$

Using that $(x-1)/x \le \log(x) \le x-1$, we see that

$$\begin{aligned} \frac{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right] - E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right]}{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right]} &\le \log\left(\frac{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right]}{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right]}\right) \\ &\le \frac{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right] - E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right]}{E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right]}. \end{aligned}$$

By Jensen's inequality and a 2nd order Taylor expansion of the exponential-function, we obtain

$$\begin{aligned} &\left|E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}\right)\right] - E_B\left[\psi\left(x|x_0;\theta,\hat{\mu}_0\right)\right]\right| \\ \le\; & E_B\left[\exp\left[2\Delta\int_0^1 a\lambda_Y\left(Z_t\left(x|x_0;\theta\right);\theta,\hat{\mu}\right) + (1-a)\,\lambda_Y\left(Z_t\left(x|x_0;\theta\right);\theta,\hat{\mu}_0\right)dt\right]\right. \\ &\left.\times\int_0^1\left|\lambda_Y\left(Z_t\left(x|x_0;\theta\right);\theta,\hat{\mu}\right) - \lambda_Y\left(Z_t\left(x|x_0;\theta\right);\theta,\hat{\mu}_0\right)\right|dt\right], \end{aligned}$$

where, by Lemma 34,

$$\begin{aligned} &\int_0^1\left|\lambda_Y\left(Z_t;\theta;\theta,\hat{\mu}\right) - \lambda_Y\left(Z_t;\theta;\theta,\hat{\mu}_0\right)\right|dt \qquad (4.29) \\ \le\; & \|\lambda_Y\left(\cdot;\theta,\hat{\mu}\right) - \lambda_Y\left(\cdot;\theta,\hat{\mu}_0\right)\|_\infty \le C\sum_{i=0}^{1}\|\hat{\mu}^{(i)} - \hat{\mu}_0^{(i)}\|_\infty. \end{aligned}$$

Using Jensen's inequality and (4.29) once more,

$$\begin{aligned}
&\frac{E_B[\exp[\Delta\int_0^1 a\lambda_Y(Z_t;\theta,\hat{\mu})+(1-a)\lambda_Y(Z_t;\theta,\hat{\mu}_0)\,dt]]}{E_B[\psi(x|x_0;\theta,\hat{\mu})]}\\
\leq\ & E_B[\exp[\Delta(1-a)\int_0^1|\lambda_Y(Z_t;\theta,\hat{\mu})-\lambda_Y(Z_t;\theta,\hat{\mu}_0)|\,dt]]\\
\leq\ & \exp\left[C\sum_{i=0}^{1}||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty\right].
\end{aligned}$$

In total,

$$\frac{|E_B[\psi(x|x_0;\theta,\hat{\mu})]-E_B[\psi(x|x_0;\theta,\hat{\mu}_0)]|}{E_B[\psi(x|x_0;\theta,\widehat{\mu})]}\leq C\exp\left[C\sum_{i=0}^{1}||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty\right]\sum_{i=0}^{1}||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty, \tag{4.30}$$

uniformly in θ. The above bound also holds with $\hat{\mu}$ and $\hat{\mu}_0$ interchanged. Claim 1) now follows from Lemma 33 and 31 together with the assumptions on the bandwidth and trimming parameter in (C1.6a).

To prove Claim 2), write $q(x|x_0;\theta,a)=q(x|x_0;\theta,\hat{T}(a)\,\mu_0)$. We then make a Taylor expansion, $q(x|x_0;\theta,a)=q(x|x_0;\theta,0)+\partial_a q(x|x_0;\theta,\bar{a})\,a$, $\bar{a}\in[0,a]$, and claim that $|\partial_a q(x|x_0;\theta,a)|\leq b(x|x_0)\,E_B[\int_0^1|\partial_a\hat{T}(Z_t;a)|^2dt]^{1/2}$ uniformly in $\theta\in\Theta$, where b does not depend on (a,θ) and $E_\pi[b(X_1|X_0)]<\infty$. This will yield 2) since

$$\begin{aligned}
\bar{a}E_{\pi\times B}\left[\int_0^1|\partial_a\hat{T}(Z_t;\bar{a})|^2dt\right]^{1/2} &\leq CP_{\pi\times B}(a/2\leq\hat{\pi}(Z_t)\leq a)^{1/2}\\
&\leq CP_{\pi\times B}(a/4\leq\pi_0(Z_t)\leq 2a)^{1/2}\\
&\to 0,
\end{aligned}$$

where the 2nd inequality holds for n sufficiently large. We have

$$|\partial_a q(x|x_0;\theta,a)|\leq\frac{E_B[|\partial_a\psi(x|x_0;\theta,a)|]}{E_B[\psi(x|x_0;\theta,a)]}\leq E_B\left[\Delta\int_0^1|\partial_a\lambda_Y(Z_t;\theta,a)|\,dt\right],$$

where $\partial_a\lambda_Y(Z_t;\theta,a)$ is given in (4.33) with $\nabla\mu(\cdot)=\partial_a\hat{T}(\cdot;a)\,\mu_0(\cdot;\theta)$ and $\nabla\sigma^2=0$. Using Lemma 34 together with (C1.2)-(C1.3),

$$\begin{aligned}
|\partial_a\lambda_Y(z;\theta,a)| &\leq \left(\frac{|\partial_x\sigma^2(z;\theta)\,\mu_0(z;\theta)|}{2\sigma^2(z;\theta)}+\frac{\mu_0^2(z;\theta)}{\sigma^2(z;\theta)}+\frac{|\mu_0^{(1)}(z;\theta)\,|}{2}\right)|\partial_a\hat{T}(z;a)\,|\\
&\leq C(1+|z|^q)\,|\partial_a\hat{T}(z;a)\,|,
\end{aligned}$$

such that by Lemma 36 $b(x|x_0)=C(1+|x|^q+|x_0|^q)$ will satisfy the desired bound. By (A0), b has a first moment.

Finally, 3) and 4) follow from standard fully parametric uniform LLN, see for example Tauchen (1985, Lemma 1): Observe that (i) Θ is compact; (ii) $\theta\mapsto q(x|x_0;\theta,\mu_0)$ is continuous; and (iii) $|q(x|x_0;\theta,\mu_0)|\leq C(1+|x|^q+|x_0|^q)$ with q given in (C1.3). The last claim follows from the fact that, by (C1.3) and Lemma 34,

$$-C\log(E_B[\exp[-C\int_0^1|B_t|^q\,dt]])\,(1+|x|^q+|x_0|^q)\leq\log\left(E_B\left[\psi\left(x|x_0;\theta,\mu_0\right)\right]\right)\leq\bar{\lambda}_Y$$

where $E_B[\exp[-C\int_0^1|B_t|^q\,dt]]<\infty$. We have now shown that the conditions of Newey and McFadden (1994, Theorem 2.1) are satisfied, and thereby that $\hat{\theta}$ is consistent. ∎

Proof of Theorem 6. Define $s_0\left(x|x_0\right)\equiv s\left(x|x_0;\theta_0,\mu_0\right)$. Lemma 11-19 then establishes that

$$\sqrt{n}(\hat{\theta}-\theta_0)=\left(H_0^{-1}+o_P\left(1\right)\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{s_0\left(X_i|X_{i-1}\right)+\delta\left(X_{i-1}\right)\right\}+o_P\left(1\right).$$

Using a CLT for mixing sequences, see e.g. Doukhan et al. (1994), we are able to conclude that $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}N\left(0,H_0^{-1}\Sigma_\infty H_0^{-1}\right)$, where $\Sigma_\infty=\Sigma_0+2\sum_{i=1}^{\infty}\Sigma_i$ and

$$\Sigma_i=E_\pi\left[\left\{s_0\left(X_1|X_0\right)+\delta\left(X_0\right)\right\}\left\{s_0\left(X_{i+1}|X_i\right)+\delta\left(X_i\right)\right\}'\right].$$

The moments Σ_∞ and H_0 are well-defined by Lemma 11, 18, and 19. Using that the process $\{X_t\}$ is a time reversible stationary Markov process, c.f. Hansen and Scheinkman (1995), together with the fact that is a martingale difference, it holds for any $i\geq 1$ that

$$\begin{aligned}E_\pi\left[s_0\left(X_1|X_0\right)s_0\left(X_{i+1}|X_i\right)^\top|X_i,X_1,X_0\right]&=s_0\left(X_1|X_0\right)E_\pi\left[s_0\left(X_{i+1}|X_i\right)^\top|X_i\right]\\&=s_0\left(X_1|X_0\right)\times 0,\end{aligned}$$

$$\begin{aligned}E_\pi\left[s_0\left(X_1|X_0\right)\delta\left(X_i\right)^\top\right]&=E_\pi\left[s_0\left(X_{i-1}|X_i\right)\delta\left(X_0\right)^\top\right]\\&=E_\pi\left[E_\pi\left[s_0\left(X_{i-1}|X_i\right)|X_i\right]\delta\left(X_0\right)^\top\right]\\&=E_\pi\left[E_\pi\left[s_0\left(X_i|X_{i-1}\right)|X_{i-1}\right]\delta\left(X_0\right)^\top\right]\\&=E_\pi\left[0\times\delta\left(X_0\right)^\top\right],\end{aligned}$$

and similar for the second cross-term. ∎

Proof of Theorem 7. Define $\tilde{\delta}_i=n^{-1}\sum_{k=1}^{n}\nabla_\pi s\left(X_k|X_{k-1};\theta_0,\hat{\mu}_0\right)\left[K_h\left(\cdot-X_i\right)\right]$ and $\bar{\delta}_i=E_\pi\left[\nabla_\pi s\left(X_1|X_0;\theta_0,\hat{\mu}_0\right)\left[K_h\left(\cdot-X_i\right)\right]\right]$, and observe that by definition of the pathwise derivative of s w.r.t. π, $\hat{\delta}_i=n^{-1}\sum_{k=1}^{n}\nabla_\pi s\left(X_k|X_{k-1};\hat{\theta},\hat{\mu}\right)\left[K_h\left(\cdot-X_i\right)\right]$. The first part of the proof then follows the one of Newey (1994, Lemma 5.5): By Lemma 13,

$$\begin{aligned}||\hat{\delta}_i-\tilde{\delta}_i||&\leq C\sum_{i,j=0}^{2}||\nabla\partial_{x,\theta}^{ij}\mu\left[K_h\left(\cdot-X_i\right)\right]||\\&\quad\times\left\{\sum_{i,j=0}^{2}||\partial_{x,\theta}^{ij}\hat{\mu}-\partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty+aE_B\left[\int_0^1|\partial_a\hat{T}\left(Z_t;a\right)|^2dt\right]^{1/2}+||\hat{\theta}-\theta_0||\right\},\end{aligned}$$

where $||\nabla_{x,\theta}^{ij}\mu\left[K_h\left(\cdot-X_i\right)\right]||_\infty\leq\sum_{k=0}^{i+1}a^{-1-k}h_k^{-1-k}$. It is then easily checked that under (C1.6b) $n^{-1}\sum_{i=1}^{n}||\hat{\delta}_i-\tilde{\delta}_i||^2=o_P\left(n^{-1/2}\right)$. From the proof of Lemma 16 it follows that

$n^{-1}\sum_{i=1}^n ||\tilde{\delta}_i - \bar{\delta}_i||^2 = o\left(n^{-1/2}\right)$, while by Lemma 18, $n^{-1}\sum_{i=1}^n ||\bar{\delta}_i - \delta_i||^2 = o\left(n^{-1/2}\right)$ with $\delta_i = \delta\left(X_{i\Delta}\right)$. In total, $n^{-1}\sum_{i=1}^n ||\hat{\delta}_i - \delta_i||^2 = o_P\left(n^{-1/2}\right)$. Assume that $d = \dim(\theta) = 1$ (otherwise consider $a'\hat{\delta}_i\hat{\delta}_j a$ for any d-dim. vector a), and obtain

$$\begin{aligned}
|\hat{\Omega}_i - \Omega_i| &\leq \frac{1}{n}\sum_{i=1}^n |\hat{\delta}_i\hat{\delta}_j - \delta_i\delta_j| \\
&\leq \frac{1}{n}\sum_{i=1}^n \left\{\delta_i|\hat{\delta}_j - \delta_j| + |\delta_j(\hat{\delta}_i - \delta_i)| + |(\hat{\delta}_j - \delta_j)(\hat{\delta}_i - \delta_j)|\right\} \\
&\leq 2\left(\frac{1}{n}\sum_{i=1}^n \delta_i^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n |\hat{\delta}_j - \delta_j|^2\right)^{1/2} + \frac{2}{n}\sum_{i=1}^n |\hat{\delta}_i - \delta_i|^2 \\
&= o_P(n^{-1/4}).
\end{aligned}$$

Next, very much copying the arguments of Newey and West (1987, Proof of Theorem 2), it then follows that $\hat{V} \to^P V_0$ under our conditions. Finally, $\hat{H} \to^P H_0$ by Lemma 19. ■

Proof of Theorem 8. As we did in Class 1, we modify our criterion function, and define

$$\hat{\theta} = \arg\min_{\theta\in\Theta} Q_n\left(\theta, \hat{\sigma}^2(\cdot;\theta)\right),$$

where $Q_n\left(\theta,\sigma^2\right) = \frac{1}{n}\sum_{i=1}^n q\left(X_i|X_{i-1};\theta,\sigma^2\right)$ and

$$\begin{aligned}
q\left(x|x_0;\theta,\sigma^2\right) &= -\frac{1}{4}\log\left[\sigma^2(x)\right] - \frac{1}{4}\log\left[\sigma^2(x_0)\right] \\
&\quad -\frac{1}{2\Delta}\left(\int_{x_0}^x \sigma^{-1}(w)\,dw\right)^2 + \log\left(E_B\left[\psi\left(x|x_0;\theta,\sigma^2\right)\right]\right).
\end{aligned}$$

We now follow the same three steps as in proof of Theorem 5, and therefore do not give all details. First,

$$\begin{aligned}
q\left(x|x_0;\theta,\hat{\sigma}^2\right) - q\left(x|x_0;\theta,\hat{\sigma}_0^2\right) &= -\frac{1}{4}\log\left(\frac{\hat{\sigma}^2(x;\theta)}{\hat{\sigma}_0^2(x;\theta)}\right) - \frac{1}{4}\log\left(\frac{\hat{\sigma}^2(x_0;\theta)}{\hat{\sigma}_0^2(x_0;\theta)}\right) \\
&\quad -\frac{1}{2\Delta}\left(\int_{x_0}^{x} \hat{\sigma}(w;\theta)^{-1}\,dw\right)^2 + \frac{1}{2\Delta}\left(\int_{x_0}^{x} \hat{\sigma}_0(w;\theta)^{-1}\,dw\right)^2 \\
&\quad +\log E_B\left[\psi\left(x|x_0;\theta,\hat{\sigma}^2\right)\right] - \log E_B\left[\psi\left(x|x_0;\theta,\hat{\sigma}_0^2\right)\right].
\end{aligned}$$

We have
$$\frac{\hat{\sigma}_0^2(z;\theta) - \hat{\sigma}^2(z;\theta)}{\hat{\sigma}^2(z;\theta)} \leq \log\left(\frac{\hat{\sigma}^2(z;\theta)}{\hat{\sigma}_0^2(z;\theta)}\right) \leq \frac{\hat{\sigma}^2(z;\theta) - \hat{\sigma}_0^2(z;\theta)}{\hat{\sigma}_0^2(z;\theta)}$$

where $\hat{\sigma}^2(z;\theta), \hat{\sigma}_0^2(z;\theta) \geq \underline{\sigma}^2$. Thus,

$$\left|\log\left(\frac{\hat{\sigma}^2(z;\theta)}{\hat{\sigma}_0^2(zx;\theta)}\right)\right| \leq \frac{\left\|\hat{\sigma}^2(\cdot;\theta) - \hat{\sigma}_0^2(\cdot;\theta)\right\|_\infty}{\underline{\sigma}^2}$$

With $z = x$ and x_0, this establishes the desired bound for the first two terms. The third term satisfies

$$\begin{aligned}
&\left(\int_{x_0}^{x} \hat{\sigma}(x;\theta)^{-1} dw\right)^2 - \left(\int_{x_0}^{x} \hat{\sigma}_0(x;\theta)^{-1} dw\right)^2 \\
= &\left(\int_{x_0}^{x} \hat{\sigma}(x;\theta)^{-1} - \hat{\sigma}_0(x;\theta)^{-1} dw\right)\left(\int_{x_0}^{x} \hat{\sigma}(x;\theta)^{-1} + \hat{\sigma}_0(x;\theta)^{-1} dw\right),
\end{aligned}$$

where

$$\begin{aligned}
\int_{x_0}^{x} \left|\hat{\sigma}(w;\theta)^{-1} - \hat{\sigma}_0(w;\theta)^{-1}\right| dw &\leq \underline{\sigma}^{-3} \int_{x_0}^{x} \left|\hat{\sigma}^2(w;\theta) - \hat{\sigma}_0^2(w;\theta)\right| dw \\
&\leq \underline{\sigma}^{-1}(|x| + |x_0|) \left\|\hat{\sigma}^2(\cdot;\theta) - \hat{\sigma}_0^2(\cdot;\theta)\right\|_\infty \\
\int_{x_0}^{x} \hat{\sigma}(w;\theta)^{-1} + \hat{\sigma}_0(w;\theta)^{-1} dw &\leq 2\underline{\sigma}^{-1}(|x| + |x_0|),
\end{aligned}$$

Define $\hat{Z}_t = Z_t(\hat{\sigma}^2)$ and $\hat{Z}_{0t} = Z_t(\hat{\sigma}_0)$. We then have

$$\begin{aligned}
\left|\log\left(\frac{E_B[\psi(x|x_0;\theta,\hat{\sigma}^2)]}{E_B[\psi(x|x_0;\theta,\hat{\sigma}_0^2)]}\right)\right| &\leq E_B[\exp[\Delta \int_0^1 |\lambda_Y(\hat{Z}_t;\theta,\hat{\sigma}^2) - \lambda_Y(\hat{Z}_{0t};\theta,\hat{\sigma}_0^2)|dt]] \\
&\quad \times E_B[\int_0^1 |\lambda_Y(\hat{Z}_t;\theta,\hat{\sigma}^2) - \lambda_Y(\hat{Z}_{0t};\theta,\hat{\sigma}_0^2)|dt],
\end{aligned}$$

where, using a Taylor expansion together with Lemma 39 and 42,

$$\begin{aligned}
|\lambda_Y\left(\hat{Z}_t;\theta,\hat{\sigma}_0^2\right) - \lambda_Y\left(\hat{Z}_{0t};\theta,\hat{\sigma}_0^2\right)| &\leq |\lambda_Y^{(1)}(w\hat{Z}_t + (1-w)\hat{Z}_{0t};\theta,\hat{\sigma}_0^2)||\hat{Z} - \hat{Z}_{0t}| \\
&\leq C\left(1 + a^{-3}\right)\left\|\hat{\sigma}^2 - \hat{\sigma}_0^2\right\|_\infty.
\end{aligned}$$

This together with Lemma 40, implies that

$$\left|Q_n(\theta,\hat{\sigma}^2) - Q_n(\theta,\hat{\sigma}_0^2)\right| \leq C(1 + (|x| + |x_0|))\, a^{-3} \sum_{i=0}^{2} \left\|\partial_x^i \hat{\sigma}^2 - \partial_x^i \hat{\sigma}_0^2\right\|_\infty$$

where, by Lemma 38 and 31 together with (C2.6a), $a^{-3}\left\|\partial_x^i \hat{\sigma}^2 - \partial_x^i \hat{\sigma}_0^2\right\|_\infty = o_P(1)$, $0 \leq i \leq 2$, and $E_\pi[|X_0|] < \infty$. Next, define $q(x|x_0;\theta,a) = q(x|x_0;\theta,\hat{T}(\cdot;a)\sigma_0^2 + (1-\hat{T}(\cdot;a))\underline{\sigma}^2)$. We obtain

$$\begin{aligned}
\partial_a q(x|x_0;\theta,a) = &-\frac{\partial_a \hat{\sigma}_0^2(x;\theta)}{4\hat{\sigma}_0^2(x;\theta)} - \frac{\partial_a \hat{\sigma}_0^2(x_0;\theta)}{4\hat{\sigma}_0^2(x_0;\theta)} + \frac{E_B[\partial_a \psi(x|x_0;\theta,a)]}{E_B[\psi(x|x_0;\theta,a)]} \\
&-\frac{1}{2}\int_{x_0}^{x} 1/\hat{\sigma}_0(w;\theta)\, dw \left(\int_{x_0}^{x} \frac{\partial_a \hat{\sigma}_0^2(w;\theta)}{\hat{\sigma}_0(w;\theta)} dw\right)
\end{aligned}$$

where $\partial_a \hat{\sigma}_0^2 = \partial_a \hat{T}(\cdot;a)\left(\sigma_0^2 - \underline{\sigma}^2\right)$, $\partial_a \partial_x^i \hat{\sigma}_0^2 = \partial_a \hat{T}(\cdot;a)\, \partial_x^i \sigma_0^2$. Thus,

$$\begin{aligned}
|\partial_a q(x|x_0;\theta,a)| &\leq \frac{1}{4\underline{\sigma}^2}\left(\left|\partial_a \hat{\sigma}_0^2(x;\theta)\right| + \left|\partial_a \hat{\sigma}_0^2(x_0;\theta)\right|\right) \\
&\quad + \frac{|x|+|x_0|}{2\underline{\sigma}^2}\left(\int_{x_0}^{x} \partial_a \hat{\sigma}_0^2(w;\theta)\,dw\right) \\
&\quad + E_B\left[\Delta \int_0^1 |\partial_a \lambda_Y(\hat{Z}_{0t};\theta,a)| + |\lambda_Y^{(1)}(\hat{Z}_{0t};\theta,a)||\partial_a \hat{Z}_{0t}|dt\right],
\end{aligned}$$

where $|\partial_a \hat{Z}_{0t}| \leq C\left(1+|x|+|x_0|+|B_t|\right)|\partial_a \hat{T}(Z_t;a)|$ and

$$\begin{aligned}
|\partial_a \lambda_Y(z;\theta,a)| &\leq \left|\frac{\mu^2(z;\theta)}{2\underline{\sigma}^2} - \frac{\mu(z;\theta)\,\partial_x^2 \sigma_0^2(z;\theta)}{2} + \frac{\left[\partial_x \sigma_0^2(z;\theta)\right]^2}{32\underline{\sigma}^2}\right| |\partial_a \hat{T}(z;a)| \\
&\quad + \left|\frac{\mu(z;\theta)\,\partial_x \sigma_0^2(z;\theta)}{2\underline{\sigma}^2} - \frac{\left[\partial_x \sigma_0^2(z;\theta)\right]^2}{16\underline{\sigma}^2} + \frac{\partial_x \sigma_0^2(z;\theta)}{8}\right| |\partial_a \hat{T}(z;a)| \\
&\leq C\left(1+\mu^2(z;\theta)+|\mu(z;\theta)|\right)\left|\partial_a \hat{T}(z;a)\right|,
\end{aligned}$$

$$|\lambda_Y^{(1)}(z;\theta,a)| \leq C(1+|\mu(z;\theta)|(|\mu^{(1)}(z;\theta)|+1)+\mu^2(z;\theta)+|\mu^{(1)}(z;\theta)|+|\mu^{(2)}(z;\theta)|).$$

We obtain by Lemma 42 and (C.1.3) that

$$\begin{aligned}
|\partial_a q(x|x_0;\theta,a)| &\leq b(x|x_0)\{|\partial_a \hat{T}(x;a)| + |\partial_a \hat{T}(x;a)| \\
&\quad + \int_{x_0}^{x} |\partial_a \hat{T}(w;a)|dw + E_B[|\int_0^1 |\partial_a \hat{T}(Z_t;a)\,dt|^2]^{1/2}\},
\end{aligned}$$

with $b(x|x_0) = C(1+|x|^{2q}+|x_0|^{2q})$, and conclude $|Q_n(\hat{\theta},\hat{\sigma}_0^2) - Q_n(\hat{\theta},\sigma_0^2)| \to^P 0$ by the properties of $\hat{T}$. Finally, $\sup_{\theta\in\Theta}\left|Q_n(\theta,\sigma_0^2) - Q(\theta,\sigma_0^2)\right| \to^P 0$ where $Q(\theta,\sigma_0^2) = E_\pi\left[q\left(X_1|X_0;\theta,\sigma_0^2\right)\right]$ since: (i) Θ is compact; (ii) $\theta \mapsto q\left(x|x_0;\theta,\sigma_0^2(\cdot;\theta)\right)$ is continuous; (iii) by (C2.3) and Lemma 39, $\left|q\left(x|x_0;\theta,\sigma_0^2\right)\right| \leq C(1+|x|^{2q}+|x_0|^{2q})$. ■

Proof of Theorem 9. This follows the same steps as the proof of Theorem 6, now only using Lemma 20-28. ■

Proof of Theorem 10. The claim is proved in the same fashion as Theorem 7, this time using Lemma 22, 25, 27 and 28. ■

4.B Lemmas

In this section we first derive expressions of the score and various derivatives of it, and then present and show a number of lemmas used in the proofs of the previous section. The following two subsections contain lemmas used in the proofs of the theorems of Class 1 and 2 respectively, while the third one contains auxiliary lemmas.

Most of the lemmas concerns the score, its pathwise derivative w.r.t. μ (in Class 1) or σ^2 (in Class 2), and the Hessian. We derive expressions of these below which can be used in both Class 1 and 2. In the following, let $p(x|x_0;\theta) = p\left(x|x_0;\mu(\cdot;\theta),\sigma^2(\cdot;\theta)\right)$. First we differentiate the $\log p$ w.r.t. θ to obtain the score s. This yields

$$s(x|x_0) = -\frac{1}{4}\frac{\dot{\sigma}^2(x)}{\sigma^2(x)} - \frac{1}{4}\frac{\dot{\sigma}^2(x_0)}{\sigma^2(x_0)} - \frac{1}{2}\left(\int_{x_0}^{x} 1/\sigma(w)\,dw\right)\left(\int_{x_0}^{x} \frac{\dot{\sigma}^2(w)}{\sigma(w)}dw\right) + \frac{E_B[\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]}, \tag{4.31}$$

where

$$\dot{\psi}(x|x_0) = \psi(x|x_0;\theta)\,\Delta\int_0^1 \dot{\lambda}_Y(Z_t) + \lambda_Y^{(1)}(Z_t)\,\dot{Z}_t dt,$$

$$\lambda_Y(z) = -\frac{\mu^2(z)}{2\sigma^2(z)} + \frac{\mu(z)\,\partial_x\sigma^2(z)}{2\sigma^2(z)} - \frac{\mu^{(1)}(z)}{2} - \frac{\left[\partial_x\sigma^2(z)\right]^2}{32\sigma^2(z)} + \frac{\partial_x\sigma^2(z)}{8}, \tag{4.32}$$

$$\begin{aligned} \dot{\lambda}_Y(z) &= \left[\frac{\partial_x\sigma^2(z)}{2\sigma^2(z)} - \frac{\mu(z)}{\sigma^2(z)}\right]\dot{\mu}(z) - \frac{\dot{\mu}^{(1)}(z)}{2} \\ &\quad + \left[\frac{\mu^2(z)}{2\sigma^4(z)} - \frac{\mu(z)\,\partial_x\sigma^2(z)}{2\sigma^2(z)} + \frac{\left[\partial_x\sigma^2(z)\right]^2}{32\sigma^4(z)}\right]\dot{\sigma}^2(z) \\ &\quad + \left[\frac{\mu(z)}{2\sigma^2(z)} - \frac{\partial_x\sigma^2(z)}{16\sigma^2(z)} + \frac{1}{8}\right]\partial_x\dot{\sigma}^2(z), \end{aligned} \tag{4.33}$$

$$\begin{aligned} \lambda_Y^{(1)}(z) &= -\frac{\mu(z)\,\mu^{(1)}(z)}{\sigma^2(z)} + \frac{\mu^2(z)\,\partial_x\sigma^2(z)}{2\sigma^4(z)} + \frac{\mu^{(1)}(z)\,\partial_x\sigma^2(z)}{2\sigma^2(z)} \\ &\quad + \frac{\mu(z)\,\partial_x^2\sigma^2(z)}{2\sigma^2(z)} - \frac{\mu(z)\left[\partial_x\sigma^2(z)\right]^2}{2\sigma^4(z)} - \frac{\mu^{(2)}(z)}{2} \\ &\quad - \frac{\partial_x\sigma^2(z)\,\partial_x^2\sigma^2(z)}{16\sigma^2(z)} + \frac{\left[\partial_x\sigma^2(z)\right]^3}{32\sigma^4(z)} + \frac{\partial_x^2\sigma^2(z)}{8}, \end{aligned} \tag{4.34}$$

$$\dot{Z}_t = \dot{\gamma}^{-1}(\gamma(Z_t)) + \sigma(\gamma(Z_t))^{-1}\left\{t\dot{\gamma}(x) + (1-t)\,\dot{\gamma}(x_0)\right\}, \tag{4.35}$$

and

$$\dot{\gamma}(z) = -\int \frac{\dot{\sigma}^2(z)}{2\sigma^3(z)}dz, \quad \dot{\gamma}^{-1}(z) = -\sigma\left(\gamma^{-1}(z)\right)\dot{\gamma}\left(\gamma^{-1}(z)\right).$$

We have here assumed that

$$\frac{\partial}{\partial\theta}E_B\left[\psi(x|x_0;\theta,\mu)\right] = E_B\left[\dot{\psi}(x|x_0;\theta,\mu)\right]. \tag{4.36}$$

We demonstrate in the two subsections that this calculation is valid for models in Class 1 and 2 respectively.

Next, we derive the expression of the pathwise derivative of s w.r.t. μ and σ^2 in the direction $\left(\nabla\mu, \nabla\sigma^2\right)$. We denote this ∇s. We get

$$\begin{aligned}\nabla s(x|x_0) &= \frac{1}{4}\frac{\dot{\sigma}^2(x)\nabla\sigma^2(x)}{\sigma^4(x)} - \frac{1}{4}\frac{\nabla\dot{\sigma}^2(x)}{\sigma^2(x)} + \frac{1}{4}\frac{\dot{\sigma}^2(x_0)\nabla\sigma^2(x_0)}{\sigma^4(x_0)} \\ &\quad -\frac{1}{4}\frac{\nabla\dot{\sigma}^2(x_0)}{\sigma^2(x_0)} - \frac{1}{4}\int_{x_0}^{x}\frac{\nabla\sigma^2(w)}{\sigma^3(w)}dw\int_{x_0}^{x}\frac{\dot{\sigma}^2(w)}{\sigma(w)}dw \\ &\quad -\frac{1}{2}\int_{x_0}^{x}\sigma^{-1}(w)\,dw\int_{x_0}^{x}\frac{\nabla\dot{\sigma}^2(w)}{\sigma(w)} - \frac{\dot{\sigma}^2(w)\nabla\sigma^2(w)}{2\sigma^3(w)}dw \\ &\quad +\frac{E_B[\nabla\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]} - \frac{E_B[\dot{\psi}(x|x_0)]E_B[\nabla\psi(x|x_0)]}{E_B[\psi(x|x_0)]^2}.\end{aligned} \tag{4.37}$$

where

$$\nabla\psi(x|x_0) = \psi(x|x_0)\,\Delta\int_0^1\nabla\lambda_Y(Z_t) + \lambda_Y^{(1)}(Z_t)\nabla Z_t dt,$$

$$\begin{aligned}\nabla\dot{\psi}(x|x_0) &= \nabla\psi(x|x_0)\,\Delta\left\{\int_0^1\dot{\lambda}_Y(Z_t) + \lambda_Y^{(1)}(Z_t)\dot{Z}_t + \nabla\dot{\lambda}_Y(Z_t)\right. \\ &\quad \left.+\dot{\lambda}_Y^{(1)}(Z_t)\nabla Z_t + \nabla\lambda_Y^{(1)}(Z_t)\dot{Z}_t + \lambda_Y^{(2)}(Z_t)\nabla Z_t\dot{Z}_t dt\right\},\end{aligned}$$

and

$$\begin{aligned}\nabla\lambda_Y(z) &= D_\mu(z)\nabla\mu(z) - \frac{\nabla\mu^{(1)}(z)}{2} + D_{\sigma^2}(z)\nabla\sigma^2(z) \\ &\quad +D_{\partial_x\sigma^2}(z)\nabla\partial_x\sigma^2(z) + \frac{\nabla\partial_x^2\sigma(z)}{8},\end{aligned} \tag{4.38}$$

with

$$D_\mu(z) = \frac{\partial_x\sigma^2(z)}{2\sigma^2(z)} - \frac{\mu(z)}{\sigma^2(z)}, \tag{4.39}$$

$$D_{\sigma^2}(z) = \frac{\mu^2(z)}{2\sigma^4(z)} - \frac{\mu(z)\partial_x\sigma^2(z)}{2\sigma^2(z)} + \frac{\left[\partial_x\sigma^2(z)\right]^2}{32\sigma^4(z)}, \tag{4.40}$$

$$D_{\partial_x\sigma^2}(z) = \frac{\mu(z)}{2\sigma^2(z)} - \frac{\partial_x\sigma^2(z)}{16\sigma^2(z)}. \tag{4.41}$$

Then,

$$\begin{aligned}\nabla\dot{\lambda}_Y(z) &= \dot{D}_\mu(z)\nabla\mu(z) + D_\mu(z)\nabla\dot{\mu}(z) - \frac{\nabla\dot{\mu}^{(1)}(z)}{2} \\ &\quad +\dot{D}_{\sigma^2}(z)\nabla\sigma^2(z) + D_{\sigma^2}(z)\nabla\dot{\sigma}^2(z) \\ &\quad +\dot{D}_{\partial_x\sigma^2}(z)\nabla\partial_x\sigma^2(z) + D_{\partial_x\sigma^2}(z)\nabla\partial_x\dot{\sigma}^2(z) + \frac{\nabla\partial_x^2\dot{\sigma}^2(z)}{8},\end{aligned} \tag{4.42}$$

$$\begin{aligned}\nabla\lambda_Y^{(1)}(z) &= D_\mu^{(1)}(z)\nabla\mu(z)+D_\mu(z)\nabla\mu^{(1)}(z)-\frac{\nabla\mu^{(2)}(z)}{2} \\ &\quad +D_{\sigma^2}^{(1)}(z)\nabla\sigma^2(z)+\left[D_{\sigma^2}(z)+D_{\partial_x\sigma^2}^{(1)}(z)\right]\nabla\partial_x\sigma^2(z) \\ &\quad +D_{\partial_x\sigma^2}(z)\nabla\partial_x^2\sigma^2(z)+\frac{\nabla\partial_x^3\sigma^2(z)}{8}.\end{aligned} \tag{4.43}$$

Finally,

$$\nabla Z_t=\nabla\gamma^{-1}(\gamma(Z_t))+\sigma(\gamma(Z_t))^{-1}\{t\nabla\gamma(x)+(1-t)\nabla\gamma(x_0)\}, \tag{4.44}$$

$$\begin{aligned}\nabla\dot{Z}_t &= \nabla\dot{\gamma}^{-1}(\gamma(Z_t))+\sigma^{-1}(\gamma(Z_t))\{t\nabla\gamma(x)+(1-t)\nabla\gamma(x_0)\} \\ &\quad -\frac{\nabla\sigma^2(\gamma(Z_t))}{2\sigma^3(\gamma(Z_t))}\{t\dot{\gamma}(x)+(1-t)\dot{\gamma}(x_0)\} \\ &\quad -\frac{\partial_x\sigma^2(\gamma(Z_t))}{2\sigma^3(\gamma(Z_t))}\{t\dot{\gamma}(x)+(1-t)\dot{\gamma}(x_0)\}\{t\nabla\gamma(x)+(1-t)\nabla\gamma(x_0)\},\end{aligned} \tag{4.45}$$

and

$$\begin{aligned}\nabla\gamma(z) &= -\int\frac{\nabla\sigma^2(z)}{2\sigma^3(z)}dz, \quad \nabla\dot{\gamma}(z)=\int\frac{3\dot{\sigma}^2(z)\nabla\sigma^2(z)}{4\sigma^3(z)}-\frac{\nabla\dot{\sigma}^2(z)}{2\sigma^3(z)}dz, \\ \nabla\gamma^{-1}(z) &= -\sigma\left(\gamma^{-1}(z)\right)\nabla\gamma\left(\gamma^{-1}(z)\right), \\ \nabla\dot{\gamma}^{-1}(z) &= \frac{\dot{\sigma}^2\left(\gamma^{-1}(z)\right)}{2\sigma\left(\gamma^{-1}(z)\right)}\nabla\gamma\left(\gamma^{-1}(z)\right)-\sigma\left(\gamma^{-1}(z)\right)\nabla\dot{\gamma}\left(\gamma^{-1}(z)\right).\end{aligned}$$

The Hessian, h, is now obtained with $\nabla\mu\equiv\dot{\mu}$, $\nabla\dot{\mu}\equiv\ddot{\mu}$, $\nabla\sigma^2\equiv\dot{\sigma}^2$, and $\nabla\dot{\sigma}^2\equiv\ddot{\sigma}^2$. In Class 1, the pathwise derivative of s w.r.t. $\partial_{x,\theta}^{ij}\mu$, $i,j=0,1$, in the direction $d\mu$ can now be obtained by choosing $\nabla\mu\equiv d\mu$, $\nabla\dot{\mu}=d\dot{\mu}$, $\nabla\mu^{(1)}\equiv d\mu^{(1)}$, $\nabla\dot{\mu}^{(1)}\equiv d\dot{\mu}^{(1)}$ and $\nabla\partial_{x,\theta}^{ij}\sigma^2\equiv 0$, $i,j=0,1$. In Class 2, the pathwise derivative of s w.r.t. $\partial_{x,\theta}^{ij}\sigma^2$ in the direction $d\sigma^2$, $i,j=0,1$, is given above with $\nabla\sigma^2\equiv d\sigma^2$, $\nabla\dot{\sigma}^2=d\dot{\sigma}^2$, $\nabla\partial_x\sigma^2\equiv d\partial_x\sigma^2$, $\nabla\partial_x\dot{\sigma}^2\equiv d\partial_x\dot{\sigma}^2$ and $\nabla\partial_{x,\theta}^{ij}\mu\equiv 0$, $i,j=0,1$.

4.B.1 Class 1

The Score

An expression for the score, $s(x|x_0;\theta,\mu)$, was derived in (4.31). A sufficient condition for (4.36) to hold is that $\psi(x|x_0;\theta,\mu)$ is bounded by an integrable function uniformly in θ; this is the case for $\mu=\mu_0$, and $\hat{\mu}_0$ by assumption (C1.3).

Lemma 11 *Under (C1.1)-(C1.7), $S_n(\theta_0,\hat{\mu}_0)=S_n(\theta_0,\mu_0)+o_P\left(n^{-1/2}\right)$, where for some $\delta>0$, $E_\pi[\|s\left(X_1|X_0;\theta_0;\sigma_0^2\right)\|^{2+\delta}]<\infty$.*

Proof. In the following we suppress the dependence on $\theta=\theta_0$ and write f_a to indicate the dependence of any function f on the trimming parameter a. We have

$$S_n(\theta_0,\hat{\mu}_0)=S_n(\theta_0,\mu_0)+\left\{n^{-1}\sum_{i=1}^{n}\frac{\partial s_{\bar{a}}(X_i|X_{i-1})}{\partial a}\right\}a,$$

where $\bar{a} \in [0, a]$ and $s_a(x|x_0) = s(x|x_0; \theta_0, \hat{T}(\cdot; a)\, \mu_0(\cdot; \theta))$. We claim that the last term is $o_P(n^{-1/2})$. The derivative $\partial_a s_a(X_i|X_{i-1}) = \nabla s_a(X_i|X_{i-1})$ where ∇s is given in (4.37) with $\nabla \partial^{ij}_{x,\theta}\mu \equiv \partial_a \hat{T}(a)\, \partial^{ij}_{x,\theta}\mu_0$, and $\nabla \partial^{ij}_{x,\theta}\sigma^2 \equiv 0$. We have

$$\|\partial_a s_a(x|x)\| \leq \frac{E_B[\|\partial_a \dot{\psi}_a(x|x_0)\|]}{E_B[\psi_a(x|x_0)]} - \frac{E_B[\|\dot{\psi}_a(x|x_0)\|] E_B[\|\partial_a \psi_a(x|x_0)\|]}{E_B[\psi_a(x|x_0)]^2}$$

where

$$\begin{aligned}\frac{E_B[\|\partial_a \dot{\psi}_a(x|x_0)\|]}{E_B[\psi_a(x|x_0)]} &\leq \Delta^2 E_B\left[\|\int_0^1 \partial_a \lambda_{Y,a}(Z_t)\, dt \int_0^1 \dot{\lambda}_{Y,a}(Z_t) + \lambda^{(1)}_{Y,a}(Z_t)\, \dot{Z}_t dt\|\right] \\ &\quad + \Delta E_B\left[\|\int_0^1 \partial_a \dot{\lambda}_{Y,a}(Z_t) + \partial_a \lambda^{(1)}_{Y,a}(Z_t)\, \dot{Z}_t dt\|\right]\end{aligned}$$

$$\frac{E_B[\|\dot{\psi}_a(x|x_0)\|]}{E_B[\psi_a(x|x_0)]} \leq \Delta E_B\left[\int_0^1 \|\dot{\lambda}_{Y,a}(Z_t)\| + |\lambda^{(1)}_{Y,a}(Z_t)|\|\dot{Z}_t\| dt\right],$$

and

$$\frac{E_B[|\partial_a \psi_a(x|x_0)|]}{E_B[\psi_a(x|x_0)]} \leq \Delta E_B\left[\int_0^1 |\partial_a \lambda_{Y,a}(Z_t)|\, dt\right].$$

Using the bounds in Lemma 34, together with

$$|\partial_a \lambda_{Y,a}(z)| \leq \left|\partial_a \hat{T}(z; a)\right| (|D_\mu(z)|\, |\mu_0(z)| + \frac{1}{2}|\mu_0^{(1)}(z)|),$$

$$\partial_a \lambda^{(1)}_{Y,a}(z) \leq \left|\partial_a \hat{T}(z; a)\right| (|D^{(1)}_\mu(z)|\, |\mu_0(z)| + |D_\mu(z)|\, |\mu_0^{(1)}(z)| + \frac{1}{2}|\mu_0^{(2)}(z)|),$$

and (C1.3), we see that

$$\|\partial_a s_a(x|x_0)\| \leq a E_B[\int_0^1 |\partial_a \hat{T}(Z_t; a)\, |dt] \times b(x|x_0),$$

where the function $b(x|x_0) = C(1 + |x|^{2q+1} + |x_0|^{2q+1})$ has $(2+\delta)$th moment by (C2.3) and (A0). Thus,

$$E_\pi[\|\partial_a s_a(X_1|X_0)\|] \leq E_\pi\left[b^2(X_1|X_0)\right]^{1/2} \times E_{\pi \times B}[\int_0^1 |\partial_a \hat{T}(Z_t; a)\, |^2 dt]^{1/2}.$$

The first part of the lemma now follows from (C1.6b). By the same arguments as above,

$$\begin{aligned}\|s_0(x|x_0)\| &\leq \frac{E_B[\|\dot{\psi}_0(x|x_0)\|]}{E_B[\psi_0(x|x_0)]} \leq \Delta E_B\left[\int_0^1 \|\dot{\lambda}_{Y,0}(Z_t)\| + |\lambda^{(1)}_{Y,0}(Z_t)|\|\dot{Z}_t\| dt\right] \\ &\leq b(x|x_0).\end{aligned}$$

Thus, $E_\pi\left[\|s_0(x|x_0)\|^{2+\delta}\right] < \infty$. ■

The Pathwise Derivative of the Score

The pathwise derivative of s w.r.t. $\partial^{ij}_{x,\theta}\mu$, $i, j = 0, 1$ in the direction $\nabla \partial^{ij}_{x,\theta}\mu$ is given in (4.37) with $\nabla \partial^{ij}_{x,\theta}\sigma^2 \equiv 0$, $i, j = 0, 1$.

Lemma 12 *Assume that (C1.1)-(C1.5) hold. Then for any direction* $\nabla\mu$,

$$\|\nabla s(\theta_0,\mu_0)[\nabla\mu]\| \le b(x|x_0)\sum_{i,j=0}^{2} E_B\left[\int_0^1 ||\partial_{x,\theta}^{i,j}\nabla\mu(Z_t)||^2 dt\right]^{1/2}, \tag{4.46}$$

where $E_\pi[b^{2+\delta}(X_1|X_0)] < \infty$, *for some* $\delta > 0$.

Proof. In the following we suppress the dependence on θ_0, μ_0 and $\nabla\mu$. Using the bounds in Lemma 35, we obtain

$$\begin{aligned}
&\frac{E_B[||\nabla\dot{\psi}(x|x_0)||]}{E_B[\psi(x|x_0)]}\\
\le\ & \frac{E_B[\psi(x|x_0)\Delta^2\int_0^1|\nabla\lambda_Y(Z_t)|\,dt\int_0^1||\dot{\lambda}_Y(Z_t)||+|\lambda_Y^{(1)}(Z_t)|||\dot{Z}_t||dt]}{E_B[\psi(x|x_0)]}\\
&+\frac{E_B[\psi(x|x_0)\Delta\int_0^1||\nabla\dot{\lambda}_Y(Z_t)||+|\nabla\lambda_Y^{(1)}(Z_t)|||\dot{Z}_t||dt]}{E_B[\psi(x|x_0)]}\\
\le\ & \Delta^2 E_B\left[\int_0^1|\nabla\lambda_Y(Z_t)|\,dt\int_0^1||\dot{\lambda}_Y(Z_t)||+|\lambda_Y^{(1)}(Z_t)|||\dot{Z}_t||dt\right]\\
&+\Delta E_B\left[\int_0^1||\nabla\dot{\lambda}_Y(Z_t)||+|\nabla\lambda_Y^{(1)}(Z_t)|||\dot{Z}_t||dt\right]\\
\le\ & Cb_1(x|x_0)\sum_{i,j=0}^{2}E_B\left[\int_0^1||\nabla\partial_{x,\theta}^{ij}\mu(Z_t)||^2dt\right]^{1/2},
\end{aligned}$$

where

$$\begin{aligned}
b_1(x|x_0) &= E_B\left[\int_0^1\left(|D_\mu(Z_t)|^2+1\right)dt\int_0^1||\dot{\lambda}_Y(Z_t)||^2+|\lambda_Y^{(1)}(Z_t)|^2||\dot{Z}_t||^2dt\right]^{1/2}\\
&\quad+\sum_{i=0}^{1}E_B\left[\int_0^1||\partial_\theta^i\dot{D}_\mu(Z_t)||^2\right]^{1/2}+E_B\left[\int_0^1||\dot{Z}_t||^2dt\right]^{1/2}\\
&\quad+\sum_{i=0}^{1}E_B\left[\int_0^1\left|\partial_x^iD_\mu(Z_t)\right|^2||\dot{Z}_t||^2dt\right]^{1/2}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{E_B[||\dot{\psi}(x|x_0)||]}{E_B[\psi(x|x_0)]}\frac{E_B[|\nabla\psi(x|x_0)|]}{E_B[\psi(x|x_0)]}\\
\le\ & E_B\left[\Delta\int_0^1||\dot{\lambda}_Y(Z_t)||+|\lambda_Y^{(1)}(Z_t)|||\dot{Z}_t||dt\right]E_B\left[\Delta\int_0^1\nabla\lambda_Y(Z_t)\,dt\right]\\
\le\ & b_2(x|x_0)\sum_{i,j=0}^{2}E_B\left[\int_0^1||\nabla\partial_{x,\theta}^{ij}\mu^{(i)}(Z_t)||^2dt\right]^{1/2}
\end{aligned}$$

where

$$b_2(x|x_0) = E_B\left[\int_0^1\left(|D_\mu(Z_t)|^2+1\right)dt\int_0^1||\dot{\lambda}_Y(Z_t)||^2+|\lambda_Y^{(1)}(Z_t)|^2||\dot{Z}_t||^2dt\right]^{1/2}.$$

We then have to show that $E_\pi[b_i^{2+\delta}(X_1|X_0)] < \infty$, $i = 1, 2$. Using the bounds established in Lemma 34, 35 and 36 together with (C1.3), we see that this will hold if $E_\pi[|X_0|^{2q+2+\delta}] < \infty$; this is satisfied by (C1.3) and (A0). ■

Lemma 13 *Under (C1.1)-(C1.5), there exist a function b with $E_\pi\left[b^{2+\delta}(X_1, X_0)\right] < \infty$ such that for all $\theta \in \Theta$ and with $||\mu||_{2,\infty} \equiv \sum_{i,j=0}^{2} ||\partial_{x,\theta}^{ij}\mu||_\infty$,*

$$\begin{aligned}
&\|\nabla s(x|x_0;\theta,\hat{\mu})[d\mu] - \nabla s(x|x_0;\theta_0,\mu_0)[d\mu]\| && (4.47)\\
\leq\ & b(x|x_0)\,||d\mu||_{2,\infty}\left\{||\hat{\mu}-\hat{\mu}_0||_{2,\infty} + aE_B\left[\int_0^1 |\partial_a\hat{T}(Z_t;a)|^2 dt\right]^{1/2} + \|\theta-\theta_0\|\right\}
\end{aligned}$$

Proof. Let $\theta \in \Theta$ be given, and write $\nabla s_0(x|x_0) = \nabla s(x|x_0;\theta,\hat{\mu}_0)[d\mu]$, $\nabla s(x|x_0) = \nabla s(x|x_0;\theta,\hat{\mu})[d\mu]$ and similarly for any other function depending on $\hat{\mu}_0$ and $\hat{\mu}$ respectively. We write,

$$\begin{aligned}
&\nabla s(x|x_0) - \nabla s_0(x|x_0)\\
=\ & \left\{\frac{E_B[\nabla\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]} - \frac{E_B[\nabla\dot{\psi}_0(x|x_0)]}{E_B[\psi_0(x|x_0)]}\right\}\\
&+\left\{\frac{E_B[\dot{\psi}_0(x|x_0)]}{E_B[\psi_0(x|x_0)]}\frac{E_B[\nabla\psi_0(x|x_0)]}{E_B[\psi_0(x|x_0)]} - \frac{E_B[\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]}\frac{E_B[\nabla\psi(x|x_0)]}{E_B[\psi(x|x_0)]}\right\}\\
=\ & : A_1 + A_2.
\end{aligned}$$

The first term is can be written as

$$\begin{aligned}
A_1 &= \frac{E_B[\nabla\dot{\psi}_0(x|x_0)]E_B[\psi(x|x_0)-\psi_0(x|x_0)]}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]} + \frac{E_B[\nabla\dot{\psi}_0(x|x_0)] - E_B[\nabla\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]}\\
&= : A_{11} + A_{12},
\end{aligned}$$

Then, using the bound for $E_B[|\psi(x|x_0) - \psi_0(x|x_0)|]/E_B[\psi(x|x_0)]$ obtained in the proof of Lemma 20,

$$\begin{aligned}
\|A_{11}\| &\leq CE_B\left[\int_0^1 (|\nabla\lambda_{Y,0}(Z_t)|+1)\,dt\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)|| + |\lambda_{Y,0}^{(1)}(Z_t)|||\dot{Z}_t||dt\right]\\
&\quad\times\frac{E_B[|\psi(x|x_0)-\psi_0(x|x_0)|]}{E_B[\psi(x|x_0)]}\\
&\leq Cb_{11}(x|x_0)\sum_{i,j=0}^{2}\exp\left[||\partial_{x,\theta}^{i,j}\hat{\mu} - \partial_{x,\theta}^{i,j}\hat{\mu}_0||_\infty\right]\\
&\quad\times||\partial_{x,\theta}^{i,j}\hat{\mu} - \partial_{x,\theta}^{i,j}\hat{\mu}_0||_\infty\sum_{i,j=0}^{2}E_B\left[|\nabla\partial_{x,\theta}^{i,j}\mu(Z_t)|^2\right]^{1/2}
\end{aligned}$$

where

$$\begin{aligned}
b_{11}(x|x_0) &= E_B[\left(\int_0^1 |\mu_0(Z_t)|^2 + 1dt\right)\int_0^1 \left(|\mu_0(Z_t)|^2 + |\mu_0(Z_t)|^2 \|\dot{\mu}_0(Z_t)\|^2 + \|\dot{\mu}_0(Z_t)\|^2\right) \\
&\quad + |\hat{\mu}_0(Z_t)|^2 + |\mu_0(Z_t)|^2 |\mu_0(Z_t)|^2 + |\mu_0(Z_t)|^2 \|\dot{Z}_t\|^2 dt \\
&\quad + \int_0^1 \|\dot{\mu}_0(Z_t)\|^2 + |\mu_0(Z_t)|^2 + \left(|D^{(1)}_{\mu,2}(Z_t)|^2 + |D^{(1)}_{\mu,2}(Z_t)|^2 + 1\right)\|\dot{Z}_t\|^2 dt]^{1/2} \\
&\le C(\sum_{i,j=0}^{2} E_B[\int_0^1 \|\partial^{i,j}_{x,\theta}\hat{\mu}_0(Z_t)\|^6 dt]^{1/2} + 1)(1+|x|+|x_0|) \\
&\le C(\sum_{i,j=0}^{2} E_B[\int_0^1 \|\partial^{i,j}_{x,\theta}\mu_0(Z_t)\|^6 dt]^{1/2} + 1)(1+|x|+|x_0|).
\end{aligned}$$

And the other term satisfies

$$\begin{aligned}
A_{12} &= \Big\{E_B[\Delta^2 \int_0^1 \nabla\lambda_Y(Z_t)\,dt \int_0^1 \dot{\lambda}_Y(Z_t) + \lambda^{(1)}_Y(Z_t)\,\dot{Z}_t dt \\
&\quad -\Delta^2 \int_0^1 \nabla\lambda_{Y,0}(Z_t)\,dt \int_0^1 \dot{\lambda}_{Y,0}(Z_t) + \lambda^{(1)}_{Y,0}(Z_t)\,\dot{Z}_t dt]\Big\} \\
&\quad +E_B[\Delta \int_0^1 \left[\nabla\dot{\lambda}_Y(Z_t) - \nabla\dot{\lambda}_{Y,0}(Z_t)\right] + \left[\nabla\lambda^{(1)}_Y(Z_t) - \nabla\lambda^{(1)}_{Y,0}(Z_t)\right]\dot{Z}_t dt] \\
&\quad +\frac{\psi(x|x_0) - \psi_0(x|x_0)}{\psi(x|x_0)}\Big\{E_B[\Delta^2 \int_0^1 \nabla\lambda_{Y,0}(Z_t)\,dt \int_0^1 \dot{\lambda}_{Y,0}(Z_t) + \lambda^{(1)}_{Y,0}(Z_t)\,\dot{Z}_t dt \\
&\quad \Delta \int_0^1 \nabla\dot{\lambda}_{Y,0}(Z_t) + \nabla\lambda^{(1)}_{Y,0}(Z_t)\,\dot{Z}_t dt]\Big\} \\
&= : A_{12,1} + A_{12,2} + A_{12,3},
\end{aligned}$$

where, by Lemma 34 and 35,

$$\begin{aligned}
\|A_{12,1}\| &\le E_B\left[\int_0^1 |\nabla\lambda_Y(Z_t)|\,dt \int_0^1 \|\dot{\lambda}_Y(Z_t) - \dot{\lambda}_{Y,0}(Z_t)\| + |\lambda^{(1)}_Y(Z_t) - \lambda^{(1)}_{Y,0}(Z_t)|\|\|\dot{Z}_t\|dt\right] \\
&\quad +E_B\left[\int_0^1 \|\dot{\lambda}_{Y,0}(Z_t) + \lambda^{(1)}_{Y,0}(Z_t)\,\dot{Z}_t\|dt \int_0^1 |\nabla\lambda_Y(Z_t) - \nabla\lambda_{Y,0}(Z_t)|dt\right] \\
&\le C\sum_{i=0}^{1} E_B[\int_0^1 \left|\nabla\partial^i_x\mu(Z_t)\right|^2 dt]^{1/2} E_B[\int_0^1 |D_{\mu,1}(Z_t)|^2 + 1dt]^{1/2} \\
&\quad \times \sum_{i=0}^{1}\sum_{j=0}^{1} a^{-j-2}\|\partial^{ij}_{x,\theta}\hat{\mu} - \partial^{ij}_{x,\theta}\hat{\mu}_0\|_\infty \\
&+ \int_0^1 \|\dot{\lambda}_{Y,0}(Z_t)\| + |\lambda^{(1)}_{Y,0}(Z_t)|\|\|\dot{Z}_t\|dt \\
&\quad \times \sum_{i,j=0}^{2} E_B[\int_0^1 \|\nabla\partial^{i,j}_{x,\theta}\mu(Z_t)\|^2 dt]^{1/2} E_B[\int_0^1 |\partial^{i,j}_{x,\theta}D_\mu(Z_t) - \partial^{i,j}_{x,\theta}D_{\mu,0}(Z_t)|^2 dt]^{1/2} \\
&\le b_{12,1}(x|x_0)(\sum_{i,j=0}^{2} \|\partial^{i,j}_{x,\theta}\hat{\mu} - \partial^{i,j}_{x,\theta}\hat{\mu}_0\|)(\sum_{i,j=0}^{2} E_B\left[\|\nabla\partial^{ij}_{x,\theta}\mu(Z_t)\|^2\right]^{1/2}),
\end{aligned}$$

$$\begin{aligned}
\|A_{12,2}\| &\leq CE_B[\int_0^1 ||\nabla\dot{\lambda}_Y(Z_t) - \nabla\dot{\lambda}_{Y,0}(Z_t)|| + |\nabla\lambda_Y^{(1)}(Z_t) - \nabla\lambda_{Y,0}^{(1)}(Z_t)|||\dot{Z}_t||dt] \\
&\leq CE_B[\int_0^1 ||\dot{D}_\mu(Z_t) - \dot{D}_{\mu,0}(Z_t)||\,|\nabla\mu(Z_t)|\,dt] \\
&\quad + CE_B[\int_0^1 \|D_\mu(Z_t) - D_{\mu,0}(Z_t)\|\,||\nabla\dot{\mu}(Z_t)||dt] \\
&\quad + CE_B[\int_0^1 |D_\mu^{(1)}(Z_t) - D_{\mu,0}^{(1)}(Z_t)|\,|\nabla\mu(Z_t)|]dt \\
&\quad + CE_B[\int_0^1 D_\mu(Z_t) - D_{\mu,0}(Z_t)\,|\nabla\mu^{(1)}(Z_t)\,\left\|\dot{Z}_t\right\|dt] \\
&\leq Cb_{12,2}(x|x_0)\sum_{i,j=0}^{2}||\partial_{x,\theta}^{i,j}\hat{\mu} - \partial_{x,\theta}^{i,j}\hat{\mu}_0||_\infty \sum_{i,j=0}^{2} E_B[||\nabla\partial_{x,\theta}^{i,j}\mu(Z_t)||^2]^{1/2}
\end{aligned}$$

and

$$\begin{aligned}
\|A_{12,3}\| &\leq Cb_{12,3}(x|x_0)\sum_{i,j=0}^{2}\exp\left[C||\partial_{x,\theta}^{i,j}\hat{\mu} - \partial_{x,\theta}^{i,j}\hat{\mu}_0||_\infty\right]||\partial_{x,\theta}^{i,j}\hat{\mu} - \partial_{x,\theta}^{i,j}\hat{\mu}_0||_\infty \\
&\quad \times \sum_{i,j=0}^{2} E_B[||\nabla\partial_{x,\theta}^{i,j}\mu(Z_t)||^2]^{1/2},
\end{aligned}$$

Define $b_{12} = \sum_i b_{12,i}$, where $b_{12,2}(x|x_0) = C(1 + |x| + |x_0|)$,

$$\begin{aligned}
b_{12,1}(x|x_0) &= E_B[\int_0^1 |D_\mu(Z_t)|^2 + 1dt]^{1/2} + E_B[\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||^2 \\
&\quad + |\lambda_{Y,0}^{(1)}(Z_t)|^2||\dot{Z}_t||^2dt]^{1/2},
\end{aligned}$$

$$\begin{aligned}
b_{12,3}(x|x_0) &= E_B[(\int_0^1 |D_{\mu,0}(Z_t)|^2 + 1dt)(\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||^2 + |\lambda_{Y,0}^{(1)}(Z_t)|^2\left\|\dot{Z}_t\right\|^2 dt)]^{1/2} \\
&\quad + E_B[\int_0^1 ||\dot{D}_{\mu,0}(z)||^2 + (|D_{\mu,0}^{(1)}(z)|^2 + |D_{\mu,0}(z)|^2 + 1)||\dot{Z}_t||^2dt]^{1/2}.
\end{aligned}$$

We then turn to A_2, which satisfies

$$\begin{aligned}
\|A_2\| &\leq \frac{E_B[||\dot{\psi}_0(x|x_0)||]}{E_B[\psi_0(x|x_0)]}\left|\frac{E_B[\nabla\psi_0(x|x_0)]}{E_B[\psi_0(x|x_0)]} - \frac{E_B[\nabla\psi(x|x_0)]}{E_B[\psi(x|x_0)]}\right| \\
&\quad + \frac{E_B[|\nabla\psi(x|x_0)|]}{E_B[\psi(x|x_0)]}\left\|\frac{E_B[\dot{\psi}_0(x|x_0)]}{E_B[\psi_0(x|x_0)]} - \frac{E_B[\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]}\right\| \\
&\leq CE_B[\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)|| + |\lambda_{Y,0}^{(1)}(Z_t)|||\dot{Z}_t||dt]\left|\frac{E_B[\nabla\psi_0(x|x_0)]}{E_B[\psi_0(x|x_0)]} - \frac{E_B[\nabla\psi(x|x_0)]}{E_B[\psi(x|x_0)]}\right| \\
&\quad + CE_B[\int_0^1 |\nabla\lambda_Y(Z_t)|\,dt]\left\|\frac{E_B[\dot{\psi}_0(x|x_0)]}{E_B[\psi_0(x|x_0)]} - \frac{E_B[\dot{\psi}(x|x_0)]}{E_B[\psi(x|x_0)]}\right\|
\end{aligned}$$

where

$$
\begin{aligned}
&\frac{E_B\left[\nabla\psi_0\left(x|x_0\right)\right]}{E_B\left[\psi_0\left(x|x_0\right)\right]}-\frac{E_B\left[\nabla\psi\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]}\\
=&\ \frac{E_B\left[\nabla\psi_0\left(x|x_0\right)\right]E_B\left[\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]E_B\left[\psi_0\left(x|x_0\right)\right]}+\frac{E_B\left[\nabla\psi_0\left(x|x_0\right)\right]-E_B\left[\nabla\psi\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]}\\
=&\ :A_{21}+A_{22}
\end{aligned}
$$

$$
\|A_{21}\|\leq CE_B[\int_0^1\nabla\lambda_{Y,0}\left(Z_t\right)dt]\frac{E_B\left[\left|\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)\right|\right]}{E_B\left[\psi\left(x|x_0\right)\right]}
$$

and

$$
\begin{aligned}
A_{22}\ =&\ \Delta^2\int_0^1\nabla\lambda_Y\left(Z_t\right)-\nabla\lambda_{Y,0}\left(Z_t\right)dt+\Delta\int_0^1\left[\nabla\lambda_Y^{(1)}\left(Z_t\right)-\nabla\lambda_{Y,0}^{(1)}\left(Z_t\right)\right]\dot{Z}_tdt\\
&+\frac{\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)}{\psi\left(x|x_0\right)}\Delta^2\int_0^1\nabla\lambda_{Y,0}\left(Z_t\right)dt\\
=&\ :A_{22,1}+A_{22,2}+A_{22,3}.
\end{aligned}
$$

The three terms are treated just as we did for $A_{12,i}$, $i=1,2,3$, with resulting bounds of the same order. Finally, $E_B\left[\dot{\psi}\left(x|x_0\right)\right]/E_B\left[\psi\left(x|x_0\right)\right]$ and $E_B\left[\nabla\psi\left(x|x_0\right)\right]/E_B\left[\psi\left(x|x_0\right)\right]$ take the same form such that they satisfy the same bounds. Using the bounds in Lemma 34, 35 and 36, we get that $b\left(x|x_0\right)\equiv\sum_{ij}b_{ij}\leq C(1+|x|^{2q+1}+|x_0|^{2q+1})$ with q given in (C1.3).

Next,

$$
\nabla s(\theta,\hat{T}\left(\cdot;a\right)\mu_0)\left[d\mu\right]-\nabla s\left(\theta,\mu_0\right)\left[d\mu\right]=\nabla\partial_a s(\theta,\hat{T}\left(\cdot;\bar{a}\right)\mu_0)\left[d\mu\right]a
$$

where

$$
\begin{aligned}
&\partial_a s(\theta,\hat{T}\left(\cdot;a\right)\mu_0)\left[d\mu\right]\\
=&\ \frac{E_B[\nabla\partial_a\dot{\psi}\left(x|x_0\right)]}{E_B\left[\psi\left(x|x_0\right)\right]}-\frac{E_B[\nabla\dot{\psi}\left(x|x_0\right)]E_B\left[\partial_a\psi\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]^2}\\
&-\frac{E_B[\partial_a\dot{\psi}\left(x|x_0\right)]E_B\left[\nabla\psi\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]^2}-\frac{E_B[\dot{\psi}\left(x|x_0\right)]E_B\left[\nabla\partial_a\psi\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]^2}\\
&+\frac{2E_B[\dot{\psi}\left(x|x_0\right)]E_B\left[\nabla\psi\left(x|x_0\right)\right]E_B\left[\partial_a\psi\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right]^3},
\end{aligned}
$$

and, using arguments similar to the ones employed in the proof of Lemma 11, one will realise that the desired bound will hold with b given above. Finally, the inequality

$$
\|\nabla s\left(\theta,\mu_0\right)\left[d\mu\right]-\nabla s\left(\theta_0,\mu_0\right)\left[d\mu\right]\|\leq b\left(x|x_0\right)||d\mu||_{2,\infty}\|\theta-\theta_0\|
$$

is shown to hold in the same manner. ■

Lemma 14 *Under (C1.1)-(C1.6), there exists a function* b *with* $E_\pi\left[b\left(X_1|X_0\right)\right] < \infty$ *such that*

$$\left\|s\left(\theta_0,\hat{\mu}\right) - s\left(\theta_0,\hat{\mu}_0\right) - \nabla s\left(\theta_0,\hat{\mu}_0\right)\left[\hat{\mu}-\hat{\mu}_0\right]\right\| \leq b\left(x|x_0\right) \sum_{i,j=0}^{2} ||\partial_{x,\theta}^{ij}\hat{\mu} - \partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty.$$

Proof. In the following we suppress the dependence on $\theta = \theta_0$ and $\nabla\mu = \hat{\mu} - \hat{\mu}_0$, and write $s\left(x|x_0\right) = s\left(x|x_0;\theta_0,\hat{\mu}\right)$, $s_0\left(x|x_0\right) = s\left(x|x_0;\theta_0,\hat{\mu}_0\right)$, $\nabla s\left(x|x_0\right) = \nabla s\left(x|x_0;\theta_0,\hat{\mu}_0\right)\left[\hat{\mu}-\hat{\mu}_0\right]$. Similarly, for any other function. We have

$$\begin{aligned}
& s\left(x|x_0\right) - s_0\left(x|x_0\right) - \nabla s\left(x|x_0\right) \\
= \; & \frac{E_B[\dot{\psi}\left(x|x_0\right)]}{E_B\left[\psi\left(x|x_0\right)\right]} - \frac{E_B[\dot{\psi}_0\left(x|x_0\right)]}{E_B\left[\psi_0\left(x|x_0\right)\right]} - \frac{E_B[\nabla\dot{\psi}_0\left(x|x_0\right)]}{E_B\left[\psi_0\left(x|x_0\right)\right]} \\
& + \frac{E_B[\dot{\psi}_0\left(x|x_0\right)]}{E_B\left[\psi_0\left(x|x_0\right)\right]^2} E_B\left[\nabla\psi_0\left(x|x_0\right)\right] \\
= \; & \frac{E_B\left[\psi\left(x|x_0\right)\right] - E_B\left[\psi_0\left(x|x_0\right)\right]}{E_B\left[\psi\left(x|x_0\right)\right] E_B\left[\psi_0\left(x|x_0\right)\right]} \begin{pmatrix} 1 & \frac{E_B[\dot{\psi}_0(x|x_0)]}{E_B[\psi_0(x|x_0)]} \end{pmatrix} \\
& \times \begin{pmatrix} E_B[\dot{\psi}\left(x|x_0\right)] - E_B[\dot{\psi}_0\left(x|x_0\right)] \\ E_B\left[\psi\left(x|x_0\right)\right] - E_B\left[\psi_0\left(x|x_0\right)\right] \end{pmatrix} \\
& - \frac{E_B[\dot{\psi}_0\left(x|x_0\right)]}{E_B\left[\psi_0\left(x|x_0\right)\right]^2} \left\{E_B\left[\psi\left(x|x_0\right)\right] - E_B\left[\psi_0\left(x|x_0\right)\right] - E_B\left[\nabla\psi_0\left(x|x_0\right)\right]\right\} \\
& + \frac{1}{E_B\left[\psi_0\left(x|x_0\right)\right]} \left\{E_B[\dot{\psi}\left(x|x_0\right)] - E_B[\dot{\psi}_0\left(x|x_0\right)] - E_B[\nabla\dot{\psi}_0\left(x|x_0\right)]\right\} \\
= \; & A_1 + A_2 + A_3,
\end{aligned}$$

and wish to show that

$$\|A_k\| \leq b_k\left(x|x_0\right) \times \left(\sum_{i,j=0}^{2} ||\partial_{x,\theta}^{ij}\hat{\mu} - \partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty^2\right),$$

where $E_\pi\left[b_k\left(X_1|X_0\right)\right] < \infty$, $k = 1, 2, 3$. First

$$\begin{aligned}
|A_1| \; \leq \; & \frac{E_B\left[\psi_0\left(x|x_0\right)\right] + ||E_B[\dot{\psi}_0\left(x|x_0\right)]||}{E_B\left[\psi\left(x|x_0\right)\right] E_B\left[\psi_0\left(x|x_0\right)\right]^2} \\
& \times \left(E_B[||\dot{\psi}\left(x|x_0\right) - \dot{\psi}_0\left(x|x_0\right)||^2] + E_B\left[\left\|\psi\left(x|x_0\right) - \psi_0\left(x|x_0\right)\right\|^2\right]\right)^2,
\end{aligned}$$

where, by (4.30), $|\psi\left(x|x_0\right) - \psi_0\left(x|x_0\right)| = O_P\left(\sum_{i=0}^{1} ||\hat{\mu}^{(i)} - \hat{\mu}_0^{(i)}||_\infty\right)$; we then show that

$$E_B[||\dot{\psi}\left(x|x_0\right) - \dot{\psi}_0\left(x|x_0\right)||] = O_P\left(\sum_{i,j=0}^{2} ||\partial_{x,\theta}^{ij}\hat{\mu} - \partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty\right). \tag{4.48}$$

$$
\begin{aligned}
& E_B[||\dot{\psi}(x|x_0) - \dot{\psi}_0(x|x_0)||] \\
\leq\ & E_B[\psi(x|x_0) \int_0^1 ||\dot{\lambda}_Y(Z_t) - \dot{\lambda}_{Y,0}(Z_t)||dt] \\
& + E_B\left[|\psi(x|x_0) - \psi_0(x|x_0)| \int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt\right] \\
\leq\ & C_1 E_B[\psi(x|x_0)] \left\{\sum_{i,j=0}^2 ||\partial_{x,\theta}^{ij}\hat{\mu} - \partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty\right\} \\
& + C_2 E_B\left[\exp\left[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt\right] \int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt\right] \\
& \times ||\hat{\mu}^{(1)} - \hat{\mu}_0^{(1)}||_\infty.
\end{aligned}
$$

From these inequalities, we see that $\|A_1\| \leq C\left(\sum_{i=1}^6 \|A_{1i}\|\right) \sum_{i,j=0}^2 ||\partial_{x,\theta}^{ij}\hat{\mu} - \partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty^2$, where

$$
\begin{aligned}
A_{11} &= \frac{E_B[\psi(x|x_0)]}{E_B[\psi_0(x|x_0)]}, \\
A_{12} &= \frac{E_B[\psi(x|x_0)]}{E_B[\psi_0(x|x_0)]} \frac{E_B[||\dot{\psi}(x|x_0)||]}{E_B[\psi_0(x|x_0)]}, \\
A_{13} &= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]}, \\
A_{14} &= \frac{E_B[||\dot{\psi}_0(x|x_0)||]E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]^2}, \\
A_{15} &= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt] \int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]}, \\
A_{16} &= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt] \int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]} \\
& \quad \times \frac{E_B[||\dot{\psi}(x|x_0)||]}{E_B[\psi_0(x|x_0)]}
\end{aligned}
$$

We show that $\|A_{1i}\| = O_P(1)$, $1 \leq i \leq 6$:

$$
A_{11} = \frac{E_B[\psi(x|x_0)]}{E_B[\psi_0(x|x_0)]} \leq E_B\left[\exp\left[\int_0^1 |\lambda_Y(Z_t;\hat{\mu}) - \lambda_{Y,0}(Z_t)|\,dt\right]\right] = O_P(1);
$$

$$
\begin{aligned}
A_{12} &\leq \frac{E_B[\psi(x|x_0)]\,E_B[||\dot{\psi}_0(x|x_0)||]}{E_B[\psi_0(x|x_0)]^2} = \hat{B}_{1,1} \times \frac{E_B[||\dot{\psi}_0(x|x_0)||]}{E_B[\psi_0(x|x_0)]} \\
&\leq O_P(1) \times E_B[\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt],
\end{aligned}
$$

where $E_{\pi\times B}[\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt] < \infty$;

$$\begin{aligned}
A_{13} &= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]} \\
&= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]]^2}{E_B[\psi(x|x_0)]^2}\frac{E_B[\psi(x|x_0)]}{E_B[\psi_0(x|x_0)]} \\
&= O_P(1)^2 \times O_P(1);
\end{aligned}$$

$$\begin{aligned}
A_{14} &= \frac{E_B[||\dot{\psi}_0(x|x_0)||]E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]^2} \\
&= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]]^2}{E_B[\psi(x|x_0)]^2}A_{12} \\
&= O_P(1)^2 \times A_{12};
\end{aligned}$$

$$\begin{aligned}
A_{15} &= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]} \\
&= E_B[\exp[\Delta(1-a)\int_0^1 |\lambda_Y(Z_t) - \lambda_{Y,0}(Z_t;)|\,dt\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt]^2 \\
&\quad \times A_{11} \\
&= O_P(1)^2 \times E_B[\int_0^1 ||\dot{\lambda}_Y(Z_t)||^2 dt];
\end{aligned}$$

$$\begin{aligned}
A_{1,6} &= \frac{E_B[\exp[\Delta \int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt]\int_0^1 ||\dot{\lambda}_{Y,0}(Z_t)||dt]^2}{E_B[\psi(x|x_0)]\,E_B[\psi_0(x|x_0)]} \\
&\quad \times\frac{E_B[||\dot{\psi}_0(x|x_0)||]}{E_B[\psi_0(x|x_0)]} \\
&= O_P(1) \times E_B[\int_0^1 ||\dot{\lambda}_Y(Z_t)||^3 dt]
\end{aligned}$$

Next, we consider A_2. First, by a 2nd order Taylor expansion of the exponential-function,

$$\begin{aligned}
&\psi(x|x_0) - \psi_0(x|x_0) - \psi_0(x|x_0)\left(\int_0^1 \lambda_Y(Z_t) - \lambda_{Y,0}(Z_t)\,dt\right) \\
&= \Delta\exp\left[\Delta\int_0^1 a\lambda_Y(Z_t) + (1-a)\lambda_{Y,0}(Z_t)\,dt\right]\left(\int_0^1 \lambda_Y(Z_t) - \lambda_{Y,0}(Z_t)\,dt\right)^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
& \left|\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)-\nabla\psi_0\left(x|x_0\right)\right| \\
\leq\ & \psi_0\left(x|x_0\right)\int_0^1\left|\lambda_Y\left(Z_t\right)-\lambda_{Y,0}\left(Z_t\right)-\nabla\lambda_{Y,0}\left(Z_t\right)\right|dt+ \\
& \Delta\exp\left[\Delta\int_0^1 a\lambda_Y\left(Z_t\right)+\left(1-a\right)\lambda_{Y,0}\left(Z_t\right)dt\right]\int_0^1\left|\lambda_Y\left(Z_t\right)-\lambda_{Y,0}\left(Z_t\right)\right|^2dt \\
\leq\ & C\psi_0\left(x|x_0\right)\|\hat{\mu}-\hat{\mu}_0\|_{1,\infty}^2 \\
& +C\exp\left[\Delta\int_0^1 a\lambda_Y\left(Z_t\right)+\left(1-a\right)\lambda_{Y,0}\left(Z_t\right)dt\right]\sum_{i=0}^1||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty^2
\end{aligned}$$

Plugging this inequality into A_2,

$$\begin{aligned}
\|A_2\| \ \leq\ & C\left(\textstyle\sum_{i=0}^1||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty^2\right)\frac{E_B[\dot{\psi}_0\left(x|x_0\right)]}{E_B\left[\psi_0\left(x|x_0\right)\right]} \\
& \times\left\{1+\frac{E_B[\exp[\Delta\int_0^1\int_0^1 a\lambda_Y\left(Z_t\right)+\left(1-a\right)\lambda_{Y,0}\left(Z_t\right)dt]]}{E_B\left[\psi_0\left(x|x_0\right)\right]}\right\}
\end{aligned}$$

where the second term is $O_P\left(1\right)$, and $E_\pi[E_B[\dot{\psi}_0\left(x|x_0\right)]/E_B\left[\psi_0\left(x|x_0\right)\right]]<\infty$, as shown earlier.

Finally, for A_3 we have

$$\begin{aligned}
& \dot{\psi}\left(x|x_0\right)-\dot{\psi}_0\left(x|x_{00}\right)-\nabla\dot{\psi}_0\left(x|x_0\right) \\
=\ & \left[\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)\right]\left[\Delta\int_0^1\dot{\lambda}_Y\left(Z_t\right)-\dot{\lambda}_{Y,0}\left(Z_t\right)dt\right] \\
& +\Delta\psi_0\left(x|x_0\right)\left[\int_0^1\dot{\lambda}_Y\left(Z_t\right)-\dot{\lambda}_{Y,0}\left(Z_t\right)-\nabla\dot{\lambda}_{Y,0}\left(Z_t\right)dt\right] \\
& +\int_0^1\dot{\lambda}_{Y,0}\left(Z_t\right)dt\left[\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)-\nabla\psi_0\left(x|x_0\right)\right]
\end{aligned}$$

We treat each of the terms separately: Using (4.30) together with Lemma 35,

$$\begin{aligned}
& \frac{E_B[(\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right))\Delta\int_0^1[\dot{\lambda}_Y\left(Z_t\right)-\dot{\lambda}_{Y,0}\left(Z_t\right)]dt]}{E_B\left[\psi_0\left(x|x_0\right)\right]} \\
\leq\ & C\frac{E_B[\exp[\Delta\int_0^1 a\lambda_Y\left(Z_t\right)+\left(1-a\right)\lambda_{Y,0}\left(Z_t\right)dt]]}{E_B\left[\psi_0\left(x|x_0\right)\right]}\textstyle\sum_{i,j=0}^2||\partial_{x,\theta}^{ij}\hat{\mu}-\partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty^2 \\
=\ & O_P\left(1\right)\times\textstyle\sum_{i,j=0}^2||\partial_{x,\theta}^{ij}\hat{\mu}-\partial_{x,\theta}^{ij}\hat{\mu}_0||_\infty^2;
\end{aligned}$$

we apply (C1.3) and Lemma 35 on the second term; and

$$\begin{aligned}
& \frac{E_B[|\psi\left(x|x_0\right)-\psi_0\left(x|x_0\right)-\nabla\psi_0\left(x|x_0\right)|\int_0^1||\dot{\lambda}_{Y,0}\left(Z_t\right)||dt]}{E_B\left[\psi_0\left(x|x_0\right)\right]} \\
\leq\ & O_P\left(1\right)\times E_B[\int_0^1||\dot{\lambda}_{Y,0}\left(Z_t\right)||dt]\times\textstyle\sum_{i=0}^1||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty^2.
\end{aligned}$$

■

We define

$$\nabla_{\pi} s(\mu)[d\pi] \equiv \nabla s(\mu)[\nabla_{\pi}\mu(x)[\hat{\pi}-\pi_0]], \tag{4.49}$$

where $\nabla_{\pi}\mu$ is given in Lemma 33. This function is the pathwise derivative w.r.t. π.

Lemma 15 *Under (C1.1)-(C1.5),*

$$\|\nabla s(x|x_0;\theta,\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]-\nabla_{\pi} s(x|x_0;\theta,\hat{\mu}_0)[\hat{\pi}-\pi_0]\| \leq b(x|x_0)\sum_{i=0}^{3} a^{i-5}\|\hat{\pi}^{(i)}-\pi_0^{(i)}\|_\infty^2$$

uniformly in $\theta \in \Theta$, where $\nabla_{\pi} s$ is given in (4.49) and $E_{\pi}[b(X_1|X_0)] < \infty$.

Proof. To see this, we make use of the linearity of ∇s and the function b given in Lemma 12 to write

$$\begin{aligned}
& \|\nabla s(x|x_0;\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]-\nabla_{\pi} s(x|x_0;\hat{\mu}_0)[\hat{\pi}-\pi_0]\| \\
= \; & \|\nabla s(x|x_0;\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0-\nabla\hat{\mu}_0[\hat{\pi}-\pi_0]]\| \\
\leq \; & b(x|x_0)\sum_{i,j=0}^{2} E_B\left[\int_0^1 \|\partial_{x,\theta}^{ij}[\hat{\mu}-\hat{\mu}_0-\nabla_{\pi}\hat{\mu}_0[\hat{\pi}-\pi_0]](Z_t)\|^2 dt\right]^{1/2} \\
& + b(x|x_0)\sum_{i,j=0}^{2} E_B\left[\int_0^1 \|\partial_{x,\theta}^{ij}[\hat{\mu}-\hat{\mu}_0-\nabla_{\pi}\hat{\mu}_0[\hat{\pi}-\pi_0]](Z_t)\|^2 dt\right]^{1/2},
\end{aligned}$$

where $E_{\pi}[b(X_1|X_0)] < \infty$. The result now follows from Lemma 33. ■

The Adjustment Term

In this section we show that the pathwise derivative of the score can be written as a normalised sum and a remainder term which can be ignored.

Lemma 16 *Under (C1.1)-(C1.6), $\nabla S_n(\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0] = \nabla S(\hat{\mu}_0)[\hat{\mu}_0-\hat{\mu}_0] + o_P(n^{-1/2})$.*

Proof. We split up in 4 terms,

$$\begin{aligned}
& \nabla S_n(\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]-\nabla S(\hat{\mu}_0)[\hat{\mu}_0-\hat{\mu}_0] \\
= \; & \{\nabla S_n(\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]-\nabla_{\pi} S_n(\hat{\mu}_0)[\hat{\pi}-\pi_0]\} \\
& + \{\nabla_{\pi} S_n(\hat{\mu}_0)[\hat{\pi}-E_{\pi}[\hat{\pi}]]-\nabla_{\pi} S(\hat{\mu}_0)[E_{\pi}[\hat{\pi}]-\pi_0]\} \\
& + \{\nabla S(\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]-\nabla_{\pi} S(\hat{\mu}_0)[\hat{\pi}-\pi_0]\} \\
& + \{\nabla_{\pi} S_n(\hat{\mu}_0)[E_{\pi}[\hat{\pi}]-\pi_0]-\nabla_{\pi} S(\hat{\mu}_0)[E_{\pi}[\hat{\pi}]-\pi_0]\},
\end{aligned}$$

where we have made use of that $d\pi \mapsto \nabla_{\pi} s(x|x_0;\hat{\mu}_0)[d\pi]$ is linear. The two first terms are of order $o_P(1/\sqrt{n})$ by Lemma 15 and 31 together with (C1.6b). To show that the third term is $o_P(1/\sqrt{n})$, we apply Lemma 32: Defining $Y_i = (X_i, X_{i-1})$, it is easily seen that $\{Y_i\}$ is a stationary and β-mixing sequence since $\{X_i\}$ is, and that the mixing coefficients, $\{\beta_{Y,i}\}$, of $\{Y_i\}$ satisfies $\beta_{Y,i} = \beta_{i-1}$ where $\{\beta_i\}$ are the mixing coefficients of $\{X_i\}$. Using the linearity of $\nabla_\pi s$ in $d\pi$ once more, we may write

$$\begin{aligned}\nabla_\pi S_n(\hat{\mu}_0)[\hat{\pi} - E_\pi[\hat{\pi}]] &= \nabla_\pi S_n(\hat{\mu}_0)[\hat{\pi}] - \nabla_\pi S_n(\hat{\mu}_0)[E_\pi[\hat{\pi}]],\\ \nabla_\pi S(\hat{\mu}_0)[\hat{\pi} - E_\pi[\hat{\pi}]] &= \nabla_\pi S(\hat{\mu}_0)[\hat{\pi}] - \nabla_\pi S(\hat{\mu}_0)[E_\pi[\hat{\pi}]].\end{aligned}$$

So by defining

$$m_n(Y_i, Y_j) = \nabla_\pi s(X_i | X_{i-1}; \hat{\mu}_0)[\underline{K}_h(\cdot - X_j)],$$

where $\underline{K}_h(x) = \left[h_0^{-1}K(x/h_0), h_1^{-2}K^{(1)}(x/h_1), h_2^{-3}K^{(2)}(x/h_2), h_3^{-4}K^{(3)}(x/h_3)\right]$, we obtain that

$$\nabla_\pi S_n(\hat{\mu}_0)[\hat{\pi} - E_\pi[\hat{\pi}]] - \nabla_\pi S(\hat{\mu}_0)[\hat{\pi} - E_\pi[\hat{\pi}]] = V_n - V_{1,n} - V_{2,n} + \bar{V}_n,$$

where V_n, $V_{1,n}$, $V_{2,n}$, and $\bar{V}_n$ are defined as in Lemma 32. We then check that the two conditions of this lemma are satisfied. The first condition follows by Lemma 12 and Hölder's inequality,

$$\begin{aligned}& E_\pi[\|m_n(Y_i, Y_j)\|^p]^{1/p}\\ =\ & E_\pi\left[\|\nabla_\pi s(X_i | X_{i-1}; \hat{\mu}_0)\left[\hat{T}(\cdot; a)\underline{K}_h(\cdot - X_j)\right]\|^p\right]^{1/p}\\ \le\ & E_\pi[b^p(X_i, X_{i-1})]^{1/p}\textstyle\sum_{i=0}^3 a^{-1-i}h_i^{-i}\\ \le\ & CE_\pi[b^p(X_i, X_{i-1})]^{1/p}\textstyle\sum_{i=0}^3 a^{-1-i}h_i^{-i},\end{aligned}$$

for any $p > 2$, where we have made use of the boundedness of K and its first three derivatives. By the restrictions put on the bandwidths, $a^{-1-i}h_i^{-i}n^{-1/2} \to 0$, $0 \le i \le 3$, while $E_\pi[b^p(X_i, X_{i-1})] < \infty$ with $p = 2 + \delta$. The second condition follows easily from the fact that $\beta_{Y,i} = \beta_{i-1}$ together with the geometric decay of $\{\beta_i\}$, c.f. Lemma 29.

Next, we show that $\nabla_\pi S_n(\hat{\mu}_0)[E_\pi[\hat{\pi}] - \pi_0] = \nabla_\pi S(\hat{\mu}_0)[E_\pi[\hat{\pi}] - \pi_0] + o_P(n^{-1/2})$. By Lemma 12,

$$\begin{aligned}\|\nabla_\pi s(x | x_0\ \hat{\mu}_0)[E_\pi[\hat{\pi}] - \pi_0]\| &= \|\nabla s(x | x_0\ \hat{\mu}_0)[\nabla\hat{\mu}_0[E_\pi[\hat{\pi}] - \pi_0]]\|\\ &\le b(x|x_0)\sum_{i=0}^3 a^{-1-i}\|E_\pi[\hat{\pi}^{(i)}] - \pi_0^{(i)}\|_\infty,\end{aligned}$$

where $\sqrt{n}a^{-1-i}\|E_\pi[\hat{\pi}^{(i)}] - \pi_0^{(i)}\|_\infty = O\left(\sqrt{n}a^{-1-i}h_i^{\omega-i}\right) \to 0$, $0 \le i \le 3$, by (C1.6b). ∎

Lemma 17 *Under (C1.1)-(C1.6),* $\nabla S(\hat{\mu}_0)[\hat{\mu} - \hat{\mu}_0] = \nabla S(\mu_0)[\hat{\mu} - \hat{\mu}_0] + o_P(n^{-1/2})$.

Proof. By Lemma 13 and Hölder's inequality,

$$\begin{aligned}& \sqrt{n}\|\nabla S(\hat{\mu}_0)[\hat{\mu} - \hat{\mu}_0] - \nabla S(\mu_0)[\hat{\mu} - \hat{\mu}_0]\|\\ \le\ & E_\pi\left[b^2(X_1|X_0)\right]^{1/2} \times a\sqrt{n}E_{\pi\times B}[\int_0^1 |\partial_a\hat{T}(Z_t; a)|^2 dt]^{1/2},\end{aligned}$$

where the last term goes to zero by (C1.6b). ∎

Lemma 18 *Under (C1.1)-(C1.7), there exists a function δ with $E_\pi[\delta(X_0)] = 0$ and $E_\pi\left[\|\delta(X_0)\|^{2+\varepsilon}\right] < \infty$ for some $\varepsilon > 0$ such that*

$$\nabla S(\mu_0)[\hat{\mu} - \hat{\mu}_0] = \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + o_P(n^{-1/2}).$$

Proof. We suppress any dependence on μ_0 and θ_0. We introduce the Hilbert space $\mathcal{H}_k$ given by

$$\mathcal{H}_k = \{f : \mathbb{R} \mapsto \mathbb{R} | E_{\pi\times B}[\int_0^1 f^{(i)}(Z_t)^4\, dt] < \infty,\ 0 \leq i \leq k+1\},$$

where $Z_t = Z_t(X_1|X_0)$, which we equip with the inner product

$$\langle f, g\rangle = E_{\pi\times B}[\int_0^1 f(Z_t)\, g(Z_t)\, dt].$$

We first show that there exists a density $\bar{p}$ such that for any function $f \in \mathcal{H}_k$,

$$E_{\pi\times B}[\int_0^1 f(Z_t)\, dt] = \int_{\mathbb{R}} f(z)\,\bar{p}(z)\, dz.$$

Observe that

$$E_{\pi\times B}[\int_0^1 f(Z_t)\, dt] = \int_0^1\int_{\mathbb{R}^3} f(Z(t,x,x_0,b))\, p(x|x_0)\,\pi_0(x_0)\, p_B(t,b)\, dx dx_0 db dt.$$

where

$$Z(t,x,x_0,b) = \gamma_0^{-1}\left(t\gamma_0(x) + (1-t)\gamma_0(x_0) + \sqrt{\Delta}b\right),$$

and

$$p_B(t,b) = \phi\left(\frac{b}{\sqrt{t(1-t)}}\right) = \frac{1}{\sqrt{2\pi t(1-t)}}\exp\left[\frac{-b^2}{2t(1-t)}\right].$$

is the marginal density associated with the Brownian Bridge at time t. We now make the transformation

$$(t,x,x_0,b) \mapsto (t,x,x_0,z) = F(t,x,x_0,b) \equiv (t,x,x_0,Z(t,x,x_0,b))$$

with inverse

$$(t,x,x_0,b) = F^{-1}(t,x,x_0,z) = \left(t,x,x_0,Z^{-1}(t,x,x_0,z)\right)$$

where

$$Z^{-1}(t,x,x_0,z) = \frac{\gamma_0(z) - t\gamma_0(x) - (1-t)\gamma_0(x_0)}{\sqrt{\Delta}}.$$

We derive the derivative of F^{-1},

$$DF^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{\partial Z^{-1}(t,x,x_0,z)}{\partial x} & \frac{\partial Z^{-1}(t,x,x_0,z)}{\partial x_0} & \frac{\partial Z^{-1}(t,x,x_0,z)}{\partial t} & \frac{\partial Z^{-1}(t,x,x_0,z)}{\partial z} \end{bmatrix},$$

where $\partial Z^{-1}(t,x,x_0,z)/\partial z = 1/\left(\sqrt{\Delta}\sigma_0(z)\right)$. So the determinant of DF^{-1} is $\left|DF^{-1}\right| = 1/\left(\sqrt{\Delta}\sigma(z)\right)$, and the claim holds with $\bar{p}$ given by

$$\bar{p}(z) = \frac{1}{\sqrt{\Delta}\sigma_0(z)} \int_{\mathbb{R}^2} p(x|x_0)\,\pi_0(x_0)\left\{\int_0^1 p_B\left(t, Z^{-1}(t,x,x_0,z)\right)dt\right\}dx\,dx_0. \tag{4.50}$$

We shall also make use of the following result: By repeated use of integration by parts,

$$\begin{aligned}\int_{\mathbb{R}} f(z)\,\pi^{(i)}(z)\,dz &= f(z)\,\pi^{(i-1)}(z)\Big]_{z=-\infty}^{z=\infty} - \int_{\mathbb{R}} f^{(1)}(z)\,\pi^{(i-1)}(z)\,dz \\ = \quad -\int_{\mathbb{R}} f^{(1)}(z)\,\pi^{(i-1)}(z)\,dz &= \cdots = (-1)^i \int_{\mathbb{R}} f^{(i)}(z)\,\pi(z)\,dz,\end{aligned} \tag{4.51}$$

for any function f satisfying $f^{(j)}(z)\,\pi^{(i-j)}(z)\big]_{z=-\infty}^{z=\infty} = 0,\ 0 \le j \le i$.

We now write

$$\nabla S(\mu_0)[\mu] = \sum_{i=0}^{2} \nabla_{i0} S(\mu_0)\left[\partial_x^i \mu\right] + \sum_{i=0}^{1} \nabla_{i1} S(\mu_0)\left[\partial_x^i \dot{\mu}\right],$$

where $\nabla_{ij} S(\mu_0)[\partial_{x,\theta}^{ij}\mu]$ is the pathwise derivative w.r.t. $\partial_{x,\theta}^{ij}\mu$. By Lemma 12, it follows that $\partial_{x,\theta}^{ij}\mu \mapsto \nabla_{ij} S(\mu_0)[\partial_{x,\theta}^{ij}\mu]$ is a bounded linear functional on $\mathcal{H}$. By Riesz's Representation Theorem, it therefore holds that there exists a function $d = (d_{i,j})$ with $d_{i,j} \in \mathcal{H}_i$ such that

$$\begin{aligned}\nabla S(\mu_0)[\nabla\mu] \quad &= \quad \sum_{i=0}^{2} E_{\pi\times B}\left[\int_0^1 \nabla\mu^{(i)}(Z_t;\theta_0)\,d_{i,0}(Z_t)\,dt\right] \\ &\quad + \sum_{i=0}^{1} E_{\pi\times B}\left[\int_0^1 \nabla\dot{\mu}^{(i)}(Z_t;\theta_0)\,d_{i,1}(Z_t)\,dt\right],\end{aligned} \tag{4.52}$$

for any $\partial_{x,\theta}^{ij}\nabla\mu \in \mathcal{H}_i$. For $0 \le i \le 2$, $0 \le j \le 1$, since $\partial_{x,\theta}^{ij}\hat{\mu}$ and $\partial_{x,\theta}^{ij}\hat{\mu}_0$ and the first $i+1$ first derivatives are bounded, they belong to $\mathcal{H}_i$ for any $n \ge 1$; moreover, for $0 \le k \le i+1$,

$$\begin{aligned}\int_0^1 E_{\pi\times B}\left[\|\partial_{x,\theta}^{i+k,j}\mu_0(Z_t)\|^4\right] \quad &\le \quad C\left(1 + \int_0^1 E_{\pi\times B}\left[|Z_t|^{4q}\right]\right) \\ &\le \quad C\left(1 + E_\pi\left[|X_0|^{4q}\right] + \int_0^1 E_B\left[|B_t|^{4q}\right]dt\right) < \infty,\end{aligned}$$

by (C1.3), (4.66) and (A0); so $\partial_{\theta,x}^{ij}\mu_0 \in \mathcal{H}_i$. Each term in (4.52) is now shown to be on the desired form. Starting out with the first term,

$$\nabla\hat{\mu}_0(x)[d\pi] = \hat{T}(x;a)\left[D_{0,0}(x)\,d\pi(x) + D_{0,1}(x)\,d\pi^{(1)}(x)\right]$$

where $D_{0,0}$ and $D_{0,1}$ are given in (4.64).By Lemma 33 and (C1.6b),

$$
\begin{aligned}
& E_{\pi\times B}\left[\int_0^1 \left[\hat{\mu}(Z_t)-\hat{\mu}_0(Z_t)\right] d_{0,0}(Z_t)\,dt\right] \\
= & -\frac{1}{2}\int_{\mathbb{R}} \hat{T}(z;a)\left[\hat{\pi}(z)-\pi_0(z)\right]\frac{\sigma^2(z)\,\pi_0^{(1)}(z)}{\pi_0^2(z)} d_{0,0}(z)\,\bar{p}(z)\,dz \\
& +\frac{1}{2}\int_{\mathbb{R}} \hat{T}(z;a)\left[\hat{\pi}^{(1)}(z)-\pi_0^{(1)}(z)\right]\frac{\sigma^2(z)}{\pi_0(z)} d_{0,0}(z)\,\bar{p}(z)\,dz + o_P(n^{-1/2}) \\
= & \int_{\mathbb{R}} \hat{T}(z;a)\,\nu_0(z)\,\hat{\pi}(z)\,dz + \int_{\mathbb{R}} \hat{T}(z;a)\,\nu_1(z)\,\hat{\pi}^{(1)}(z)\,dz + o_P(n^{-1/2}),
\end{aligned}
$$

where

$$
\nu_0(z) = -\frac{1}{2}\frac{\sigma_0^2(z)\,\pi_0^{(1)}(z)}{\pi_0^2(z)} d_{0,0}(z)\,\bar{p}(z), \quad \nu_1(z) = \frac{1}{2}\frac{\sigma_0^2(z)}{\pi_0(z)} d_{0,0}(z)\,\bar{p}(z). \tag{4.53}
$$

Observe that for any $n \geq 1$, there exists $R > 0$ such that $\partial_x \hat{T}(z;a) = 0$, $|z| > R$. Thus, by (4.51),

$$
\int_{\mathbb{R}} \hat{T}(z;a)\,\nu_1(z)\,\hat{\pi}^{(1)}(z)\,dz = -\int_{\mathbb{R}} \hat{T}^{(1)}(z;a)\,\nu_1(z)\,\hat{\pi}(z)\,dz - \int_{\mathbb{R}} \hat{T}(z;a)\,\nu_1^{(1)}(z)\,\hat{\pi}(z)\,dz,
$$

where we claim that $\int_{\mathbb{R}} \hat{T}^{(1)}(z;a)\,\nu_1(z)\,\hat{\pi}(z)\,dz = o_P(n^{-1/2})$; this will be shown later. But first we define $\delta_{0,0}(z) = \nu_0(z) - \nu_1^{(1)}(z) = 1/(2\pi_0(z))\,\partial_x\left[\sigma_0^2(z)\,d_{1,0}(z)\,\bar{p}(z)\right]$, and claim that

$$
\begin{aligned}
\int_{\mathbb{R}} \hat{T}(z;a)\,\delta_1(z)\,\hat{\pi}(z)\,dz &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} \hat{T}(z;a)\,\delta_1(z)\,K_h(z-X_i)\,dz \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{T}(X_i;a)\,\delta_1(X_i) + o_P(n^{-1/2}),
\end{aligned} \tag{4.54}
$$

where $E_\pi[\delta_{0,0}(X_0)] = 0$ and $E_\pi[\|\delta_{0,0}(X_0)\|^{2+\epsilon}] < \infty$. It is easily seen that $E_\pi[\delta_{0,0}(X_0)] = 0$, while

$$
\begin{aligned}
E_\pi[\|\delta_{0,0}(X_0)\|^{2+\epsilon}] &= \frac{1}{2^{2+\epsilon}}\int_{\mathbb{R}}\frac{1}{\pi_0^{1+\epsilon}(z)}\left\|\partial_z\left[\sigma_0^2(z)\,d_{0,0}(z)\,\bar{p}(z)\right]\right\|^{2+\epsilon} dz \\
&\leq C\int_{\mathbb{R}}\frac{1}{\pi_0^{1+\epsilon}(z)}\left\{\|d_{0,0}(z)\|^{2+\epsilon} + \|\partial_z d_{0,0}(z)\|^{2+\epsilon}\right\}\bar{p}^{2+\epsilon}(z)\,dz \\
&\quad + C\int_{\mathbb{R}}\frac{1}{\pi_0^{1+\epsilon}(z)}\|d_{0,0}(z)\|^{2+\epsilon}\,|\partial_z\bar{p}(z)|^{2+\epsilon}\,dz
\end{aligned}
$$

where we have made use of (C1.2). By (C1.7), $\int_{\mathbb{R}}\{d_{0,0}^{2+\epsilon}(z)+\partial_z d_{0,0}^{2+\epsilon}(z)\}\bar{p}(z)\,dz$ bounds the first integral, while the second is bounded by

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{|\partial_z \pi_0(z)|^{1+\epsilon}}{\pi_0^{1+\epsilon}(z)} \|d_{0,0}(z)\|^{2+\epsilon} |\partial_z \bar{p}(z)|\,dz \\
\leq\ & C\int_{\mathbb{R}} |\mu_0(z)|^{1+\epsilon} \|d_{0,0}(z)\|^{2+\epsilon} |\partial_z \bar{p}(z)|\,dz \\
=\ & C\int_{\mathbb{R}} \left[|\mu_0(z)|^{\epsilon} |\partial_z \mu_0(z)| \|d_{0,0}(z)\|^{2+\epsilon} + |\mu_0(z)|^{1+\epsilon} \|d_{0,0}(z)\|^{\epsilon} \|\partial_z d_{0,0}(z)\|\right] \bar{p}(z)\,dz
\end{aligned}$$

where the last equality follows by (4.51). Since $d_{0,0}, \mu_0 \in \mathcal{H}_1$ this proves the claim. Defining

$$w_n(x) = \int_{\mathbb{R}} \hat{T}(z;a)\,\delta_{0,0}(z)\,K_h(z-x)\,dz - \hat{T}(x;a)\,\delta_{0,0}(x),$$

the last equality (4.54) will hold if $E_\pi[w_n(X_0)] = o\left(n^{-1/2}\right)$ and $\mathrm{var}[n^{-1/2}\sum_{i=1}^n w_n(X_i)] = o(1)$. First observe that

$$\begin{aligned}
w_n(x) &= h_0^{\omega}\int_{\mathbb{R}} \pi_0^{(\omega)}(z^*)\,\hat{T}(z;a)\,\delta_{0,0}(z)\,dz = h_0^{\omega}\int_{\mathbb{R}} \pi_0^{(\omega)}(z^*)\,\frac{\hat{T}(z;a)}{\pi_0(z)}\pi_0(z)\,\delta_{0,0}(z)\,dz \\
&= O\left(h_0^{\omega}a^{-1}\right)
\end{aligned}$$

uniformly over $x \in \mathbb{R}$. This implies that $\sqrt{n}E_\pi[w_n(X_1)] = O\left(\sqrt{n}h_0^{\omega}a^{-1}\right) = o(1)$ by (C1.6b), while the variance term is bounded by

$$\mathrm{var}(w_n(X_1)) + 2\sum_{i=1}^{n-1}\mathrm{cov}(w_n(X_1), w_n(X_{i+1})) \leq E_\pi\left[\|w_n(X_1)\|^2\right]\left\{1+2\sum_{i=1}^{n-1}\beta_i^{1/2}\right\}$$

where, by dominated convergence and (C1.6b), $E_\pi[\|w_n(X_1)\|^2] = O\left(h_0^{2\omega}a^{-2}\right) = o(1)$.

Treating the remaining four terms of (4.52) in a similar fashion, we obtain the claimed result with $\delta = \sum_{i=0}^{2}\sum_{j=0}^{1}\delta_{i,j}$ if $n^{-1/2}\sum_{i=1}^n\{1-\hat{T}(X_i;a)\}\delta(X_i) = o_P(1)$. But $1 - \hat{T}(x;a) = a\partial_a\hat{T}(x;\bar{a})$ for some $\bar{a} \in [0,a]$ where

$$E_\pi[\|\partial_a\hat{T}(X_i;\bar{a})\,\delta(X_i)\,\|] \leq E_\pi[|\partial_a\hat{T}(X_i;a)\,|^2]^{1/2} \times E_\pi[\|\delta(X_i)\,\|^2]^{1/2},$$

and the result now follows from (C1.6b).

Finally, we have to show that $\int_{\mathbb{R}}\hat{T}^{(1)}(z;a)\,\nu_1(z)\,\hat{\pi}(z)\,dz = o_P(n^{-1/2})$. Going through the same steps as before, we obtain

$$\int_{\mathbb{R}}\hat{T}^{(1)}(z;a)\,\nu_1(z)\,\hat{\pi}(z)\,dz = \sum_{i=1}^n\hat{T}^{(1)}(X_i;a)\,\nu_1(X_i) + o_P(n^{-1/2}),$$

where

$$E_\pi[|\hat{T}^{(1)}(X_i;a)\,|\,|\nu_1(X_i)|] \leq E_\pi[|\hat{T}^{(1)}(X_i;a)\,|^2]E_\pi\left[\|\nu_1(X_i)\|^2\right]^{1/2} = o_P(n^{-1/2})$$

by (C1.6b) since $E_\pi[\|\nu_1(X_i)\|^2] < \infty$, and $E_\pi[|\hat{T}^{(1)}(X_i;a)\,|^2] \leq CP_\pi(a/2 \leq \pi_0(X_i) \leq a)$.
■

The Hessian

Lemma 19 *Under (C1.1)-(C1.7),* $\sup_{\theta\in\Theta}\|H_n(\theta,\hat{\mu})-H(\theta,\mu_0)\|\to^P 0$.

Proof. The Hessian, $h=\partial_\theta s$, is given in (4.12) with $\nabla\partial_x^i\mu=\partial_x^i\hat{\mu}$, $\nabla\partial_x^i\dot{\mu}=\partial_x^i\ddot{\mu}$, $\nabla\partial_x^i\sigma^2=\partial_x^i\dot{\sigma}^2$, and $\nabla\partial_x^i\dot{\sigma}^2=\partial_x^i\ddot{\sigma}^2$. We claim that

$$\begin{aligned}\|h(x|x_0;\theta,\hat{\mu})-h(x|x_0;\theta,\hat{\mu}_0)\| &\leq b(x|x_0)\sum_{i,j=0}^{3}\|\partial_{x,\theta}^{ij}\hat{\mu}-\partial_{x,\theta}^{ij}\hat{\mu}_0\|_\infty,\\ \|h(x|x_0;\theta,\hat{\mu}_0)-h(x|x_0;\theta,\mu_0)\| &\leq b(x|x_0)\,aE_B\left[\int_0^1|\partial_a\hat{T}(Z_t;a)|dt\right],\\ \|h(x|x_0;\theta,\mu_0)\| &\leq b(x|x_0),\end{aligned}$$

where $E_\pi[b(X_1|X_0)]<\infty$. This will give the desired result, c.f. the proof of Theorem 5. The inequalities are established in the same fashion as in the proof of Lemma 13. ■

4.B.2 Class 2

The Score

The expression for the score given in (4.31) is also valid for Class 2, since $\psi(x|x_0;\theta,\sigma^2)$ is bounded for $\sigma^2=\hat{\sigma}_0^2$ and σ_0^2 by (C2.3).

Lemma 20 *Under (C2.1)-(C2.7),*

$$S_n(\theta_0;\hat{\sigma}_0^2)=S_n(\theta_0;\sigma_0^2)+o_P(n^{-1/2}),$$

where $E_\pi[\|s(X_1|X_0;\theta_0;\sigma_0^2)\|^{2+\delta}]<\infty$.

Proof. Using the same notation and strategy as in the proof of Lemma 11, we obtain

$$\begin{aligned}&\|\partial_a s(x|x_0;a)\|\\ \leq\ & C\left(1+|x|^{3q+1}+|x_0|^{3q+1}\right)\\ &\times\left(|\partial_a\hat{T}(x;a)|+|\partial_a\hat{T}(x_0;a)|+\int_{x_0}^{x}|\partial_a\hat{T}(w;a)|dw+E_B\left[\int_0^1|\partial_a\hat{T}(Z_t;a)|dt\right]\right),\end{aligned}$$

by Lemma 40 and 42. The proof now proceeds as the one of Lemma 11, and we conclude that (C2.6b) ensures that the result holds. ■

The Pathwise Derivative of the Score

The pathwise derivative of s w.r.t. $\partial_{x,\theta}^{ij}\sigma^2$, $i,j=0,1$, w.r.t. $\nabla\sigma^2$is given in (4.37) with $\nabla\partial_{x,\theta}^{ij}\mu\equiv 0$, $i,j=0,1$.

Lemma 21 *Assume that (C2.1)-(C2.6) hold. Then*

$$\begin{aligned}\left\|\nabla s\left(x|x_0;\theta,\sigma_0^2\right)\left[\nabla\sigma^2\right]\right\| &\leq C\sum_{j=0}^{1}\left(\|\partial_\theta^j\nabla\sigma^2(x)\|+\|\partial_\theta^j\nabla\sigma^2(x_0)\|\right)\\ &+b(x|x_0)\sum_{j=0}^{1}\int_{\min\{x,x_0\}}^{\max\{x,x_0\}}|\partial_\theta^j\nabla\sigma^2(w)|dw\\ &+b(x|x_0)\sum_{i=0}^{2}\sum_{j=0}^{1}E_B\left[\int_0^1\|\partial_x^i\partial_\theta^j\nabla\sigma^2(Z_t)\|^2dt\right]^{1/2},\end{aligned}$$

where $E_\pi\left[b^{2+\delta}(X_1|X_0)\right]<\infty$.

Proof. Using the bounds in (C2.2),

$$\begin{aligned}\|\nabla s(x|x_0)\| &\leq C\left(\nabla\sigma^2(x)+\nabla\dot\sigma^2(x)+\nabla\sigma^2(x_0)+\nabla\dot\sigma^2(x_0)\right)\\ &+C(|x|+|x_0|)\left(\int_{x_0}^{x}\nabla\sigma^2(w)\,dw+\int_{x_0}^{x}\nabla\dot\sigma^2(w)\,dw\right)\\ &+\frac{E_B[\|\nabla\dot\psi(x|x_0)\|]}{E_B[\psi(x|x_0)]}-\frac{E_B[\|\dot\psi(x|x_0)\|]E_B[\|\nabla\psi(x|x_0)\|]}{E_B[\psi(x|x_0)]^2}.\end{aligned}\tag{4.55}$$

We obtain

$$\begin{aligned}&\frac{|\nabla\psi(x|x_0)|}{\psi(x|x_0)}\\ \leq\;&\Delta\left(\int_0^1\sum_{i=0}^{2}|D_i(Z_t)||\nabla\partial_x^i\sigma^{2(i)}(Z_t)|+|\lambda_Y^{(1)}(Z_t)||\nabla Z_t|\,dt\right)\\ \leq\;&\Delta\int_0^1\sum_{i=0}^{2}|D_i(Z_t)||\nabla\partial_x^i\sigma^{2(i)}(Z_t)|dt+C\int_0^1|\lambda_Y^{(1)}(Z_t)|\int_0^{Z_t}\left|\nabla\sigma^2(w)\right|dwdt\\ &+C\left(\int_0^x\left|\nabla\sigma^2(w)\right|dw+\int_0^{x_0}\left|\nabla\sigma^2(w)\right|dw\right)\int_0^1|\lambda_Y^{(1)}(Z_t)|dt\\ =\;&\left(\sum_{i=0}^{2}\int_0^1|D_i(Z_t)|^2\,dt\right)^{1/2}\left(\sum_{i=0}^{2}\int_0^1\left|\nabla\sigma^2(Z_t)\right|^2dt\right)^{1/2}\\ &+\int_0^1|\lambda_Y^{(1)}(Z_t)|dt\left(\int_0^x\left|\nabla\sigma^2(w)\right|dw\right).\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\|\nabla\dot\psi(x|x_0)\|}{\psi(x|x_0)} &\leq \frac{\|\nabla\psi(x|x_0)\|}{\psi(x|x_0)}\Delta\int_0^1\|\dot\lambda_Y(Z_t)\|+|\lambda_Y^{(1)}(Z_t)|\|\dot Z_t\|dt\\ &+\Delta\int_0^1\|\nabla\dot\lambda_Y(Z_t)\|+|\nabla\lambda_Y^{(1)}(Z_t)|\|\dot Z_t\|+|\lambda_Y^{(1)}(Z_t)|\|\nabla\dot Z_t\|dt\end{aligned}$$

with bounds for $\nabla\dot\lambda_Y$, $\nabla\lambda_Y^{(1)}$, and $\left\|\nabla\dot Z_t\right\|$ given in Lemma 41 and 42. Collecting the various terms, we establish the desired bound with $E_\pi[b^{2+\delta}(X_1|X_0)]<\infty$ because of (C2.3) and (A0). ■

Lemma 22 *Under (C2.1)-(C2.5), there exist a function b with $E_\pi\left[b^{2+\delta}(X_1,X_0)\right]<\infty$ such that for all $\theta\in\Theta$ and with $||\sigma^2||_{3,1,\infty}\equiv\sum_{i=0}^{3}\sum_{j=0}^{1}||\partial_{x,\theta}^{ij}\sigma^2||_\infty$,*

$$\begin{aligned}&\left\|\nabla s\left(x|x_0;\theta,\hat\sigma^2\right)\left[d\sigma^2\right]-\nabla s\left(x|x_0;\theta_0,\sigma_0^2\right)\left[d\sigma^2\right]\right\|\\ \le\;& b\left(x|x_0\right)||d\sigma^2||_{3,1,\infty}\left\{||\hat\sigma^2-\sigma_0^2||_{3,1,\infty}+\|\theta-\theta_0\|+a\hat A\right\},\end{aligned}$$

where

$$\hat A=|\partial_a\hat T(x;a)|+|\partial_a\hat T(x;a)|+\int_{\min\{x,x_0\}}^{\max\{x,x_0\}}|\partial_a\hat T(w;a)|dw+E_B\left[\int_0^1|\partial_a\hat T(Z_t;a)|^2dt\right]^{1/2}.$$

Proof. Let $\theta\in\Theta$ be given, and write $\nabla s_0(x|x_0)=\nabla s\left(x|x_0;\theta,\hat\sigma_0^2\right)[d\mu]$, $\nabla s(x|x_0)=\nabla s\left(x|x_0;\theta,\hat\sigma^2\right)\left[d\sigma^2\right]$ and similarly for any other function depending on $\hat\sigma_0^2$ and $\hat\sigma^2$ respectively. It holds that

$$\begin{aligned}&\nabla s\left(x|x_0;\theta,\hat\sigma^2\right)\left[d\sigma^2\right]-\nabla s\left(x|x_0;\theta_0,\hat\sigma_0^2\right)\left[d\sigma^2\right]\\ =\;&\frac14 d\sigma^2(x)\left[\frac{\partial_\theta\hat\sigma^2(x)}{\hat\sigma^4(x)}-\frac{\partial_\theta\hat\sigma_0^2(x)}{\hat\sigma_0^4(x)}\right]-\frac14 d\dot\sigma^2(x)\left[\frac{1}{\hat\sigma^2(x)}-\frac{1}{\hat\sigma_0^2(x)}\right]\\ &+\frac14 d\sigma^2(x_0)\left[\frac{\partial_\theta\hat\sigma^2(x_0)}{\hat\sigma^4(x_0)}-\frac{\partial_\theta\hat\sigma_0^2(x_0)}{\hat\sigma_0^4(x_0)}\right]-\frac14 d\dot\sigma^2(x_0)\left[\frac{1}{\hat\sigma^2(x_0)}-\frac{1}{\hat\sigma_0^2(x_0)}\right]\\ &-\frac12\left(\int_{x_0}^{x}\sigma^{-1}(w)\,dw\right)\int_{x_0}^{x}\nabla\dot\sigma^2(w)\left[\frac{1}{\hat\sigma^2(w)}-\frac{1}{\hat\sigma_0^2(w)}\right]dw\\ &+\frac12\left(\int_{x_0}^{x}\sigma^{-1}(w)\,dw\right)\int_{x_0}^{x}\frac{\nabla\sigma^2(w)}{2}\left[\frac{\partial_\theta\hat\sigma^2(w)}{2\hat\sigma^3(w)}+\frac{\partial_\theta\hat\sigma_0^2(w)}{2\hat\sigma_0^3(w)}\right]dw\\ &+\frac12\int_{x_0}^{x}\left[\frac{1}{\hat\sigma^2(w)}-\frac{1}{\hat\sigma_0^2(w)}\right]dw\int_{x_0}^{x}\frac{\nabla\dot\sigma^2(w)}{\hat\sigma_0(w)}-\frac{\partial_\theta\hat\sigma_0^2(w)\nabla\sigma^2(w)}{2\hat\sigma_0^3(w)}dw\\ &+\left[\frac{E_B[\nabla\dot\psi(x|x_0)\left[d\sigma^2\right]]}{E_B\left[\psi(x|x_0)\right]}-\frac{E_B[\nabla\dot\psi_0(x|x_0)\left[d\sigma^2\right]]}{E_B\left[\psi_0(x|x_0)\right]}\right]\\ &-\left[\frac{E_B[\dot\psi(x|x_0)]E_B\left[\nabla\psi(x|x_0)\left[d\sigma^2\right]\right]}{E_B\left[\psi(x|x_0)\right]^2}-\frac{E_B[\dot\psi_0(x|x_0)]E_B\left[\nabla\psi_0(x|x_0)\left[d\sigma^2\right]\right]}{E_B\left[\psi_0(x|x_0)\right]^2}\right].\end{aligned}$$

Thus,

$$\begin{aligned}&\left\|\nabla s\left(x|x_0;\theta,\hat\sigma^2\right)\left[d\sigma^2\right]-\nabla s\left(x|x_0;\theta_0,\hat\sigma_0^2\right)\left[d\sigma^2\right]\right\|\\ \le\;& b\left(x|x_0\right)\left\|d\sigma^2\right\|_{3,1,\infty}||\hat\sigma^2-\sigma_0^2||_{3,1,\infty}\\ &+\left\|\frac{E_B[\nabla\dot\psi(x|x_0)\left[d\sigma^2\right]]}{E_B\left[\psi(x|x_0)\right]}-\frac{E_B[\nabla\dot\psi_0(x|x_0)\left[d\sigma^2\right]]}{E_B\left[\psi_0(x|x_0)\right]}\right\|\\ &+\left\|\frac{E_B[\dot\psi(x|x_0)]E_B\left[\nabla\psi(x|x_0)\left[d\sigma^2\right]\right]}{E_B\left[\psi(x|x_0)\right]^2}-\frac{E_B[\dot\psi_0(x|x_0)]E_B\left[\nabla\psi_0(x|x_0)\left[d\sigma^2\right]\right]}{E_B\left[\psi_0(x|x_0)\right]^2}\right\|,\end{aligned}$$

where $b(x|x_0)=C(1+|x|^{2q+1}+|x_0|^{2q+1})$. The last two terms are treated in the same way as in the proof of Lemma 13. Using the same notation and strategy as in that proof, we obtain

$$\begin{aligned}
\|A_{11}\| &\leq CE_B\left[\int_0^1 \left(|\nabla\lambda_{Y,0}(Z_t)|+1\right)dt\int_0^1 \|\dot{\lambda}_{Y,0}(Z_t)\|+|\lambda^{(1)}_{Y,0}(Z_t)|\|\|\dot{Z}_t\||dt\right] \\
&\quad\times\frac{E_B\left[|\psi(x|x_0)-\psi_0(x|x_0)|\right]}{E_B\left[\psi(x|x_0)\right]} \\
&\leq Cb(x|x_0)\,C\sum_{k=0}^{3}\sum_{l=0}^{1}a^{-3-i}\left\|\partial^{kl}_{x,\theta}\hat{\sigma}^2-\partial^{kl}_{x,\theta}\hat{\sigma}_0^2\right\|_\infty\sum_{i,j=0}^{2}E_B\left[|\nabla\partial^{i,j}_{x,\theta}\sigma^2(Z_t)|^2\right]^{1/2}
\end{aligned}$$

by Lemma 40, 41 and 42, and similarly for the other terms. Next,

$$\nabla s\left(\theta,\hat{\sigma}_0^2\right)\left[d\sigma^2\right]-\nabla s\left(\theta,\sigma_0^2\right)\left[d\sigma^2\right]=\nabla\partial_a s\left(\theta,\hat{\sigma}_0^2\right)\left[d\sigma^2\right]|_{a=\bar{a}}a$$

where we claim that

$$\nabla\partial_a s(\theta,\hat{\sigma}_0^2)\left[d\sigma^2\right]=\nabla\partial_a s(\theta,\hat{T}(\cdot;\bar{a})\,\sigma_0^2+(1-\hat{T}(\cdot;\bar{a}))\underline{\sigma}^2)\left[d\sigma^2\right]$$

satisfies $a\left\|\nabla\partial_a s(\theta,\hat{\sigma}_0^2)\left[d\sigma^2\right]\right\|\leq b(x|x_0)\,\hat{A}_n$. We see that

$$\begin{aligned}
&\nabla\partial_a s(\theta,\hat{T}(\cdot;\bar{a})\,\sigma_0^2+(1-\hat{T}(\cdot;\bar{a}))\underline{\sigma}^2)\left[d\sigma^2\right] \\
=\;&\frac{d\sigma^2(x)}{4}\left[\frac{\partial_{\theta,a}\hat{\sigma}_0^2(x)}{\hat{\sigma}_0^4(x)}-\frac{\partial_\theta\hat{\sigma}_0^2(x)\,\partial_a\hat{\sigma}_0^2(x)}{2\hat{\sigma}_0^6(x)}\right] \\
&\frac{d\sigma^2(x_0)}{4}\left[\frac{\partial_{\theta,a}\hat{\sigma}_0^2(x_0)}{\hat{\sigma}_0^4(x_0)}-\frac{\dot{\sigma}^2(x)\,\partial_a\hat{\sigma}_0^2(x_0)}{2\hat{\sigma}_0^6(x_0)}\right] \\
&+\frac{3}{8}\int_{x_0}^{x}\frac{d\sigma^2(w)\,\partial_a^2\hat{\sigma}_0^2(w)}{\hat{\sigma}_0^5(w)}dw\int_{x_0}^{x}\frac{\partial_\theta\hat{\sigma}_0^2(w)}{\hat{\sigma}_0(w)}dw \\
&+\frac{1}{4}\int_{x_0}^{x}\frac{d\sigma^2(w)}{\hat{\sigma}_0^3(w)}dw\int_{x_0}^{x}\frac{\partial_\theta\hat{\sigma}_0^2(w)\,\partial_a^2\hat{\sigma}_0^2(w)}{2\hat{\sigma}_0^3(w)}-\frac{\partial_{\theta,a}\hat{\sigma}_0^2(w)}{\hat{\sigma}_0(w)}dw \\
&+\frac{1}{4}\int_{x_0}^{x}\frac{d\sigma^2(w)}{\hat{\sigma}_0^3(w)}dw\int_{x_0}^{x}\frac{\partial_\theta\hat{\sigma}_0^2(w)\,\partial_a\hat{\sigma}_0^2(w)}{2\hat{\sigma}_0^3(w)}-\frac{\partial_{\theta,a}\hat{\sigma}_0^2(w)}{\hat{\sigma}_0(w)}dw \\
&+\frac{1}{2}\int_{x_0}^{x}\frac{\partial_a\hat{\sigma}_0^2(w)}{\hat{\sigma}_0^3(w)}dw\int_{x_0}^{x}\frac{d\dot{\sigma}^2(w)}{\sigma(w)}-\frac{\dot{\sigma}^2(w)\,d\sigma^2(w)}{2\sigma^3(w)}dw \\
&-\frac{1}{2}\int_{x_0}^{x}\frac{1}{\sigma(w)}dw\int_{x_0}^{x}\frac{d\dot{\sigma}^2(w)\,\partial_a\hat{\sigma}_0^2(w)}{\hat{\sigma}_0^4(w)}+\frac{\partial_\theta\hat{\sigma}_0^2(w)\,d\sigma^2(w)\,\partial_a\hat{\sigma}_0^2(w)}{2\hat{\sigma}_0^5(w)} \\
&+\frac{1}{2}\int_{x_0}^{x}\frac{1}{\sigma(w)}dw\int_{x_0}^{x}\frac{\partial_{\theta,a}\hat{\sigma}_0(w)\,d\sigma^2(w)}{2\hat{\sigma}_0^3(w)}dw \\
&\frac{E_B[\nabla\partial_a\dot{\psi}(x|x_0)\left[d\sigma^2\right]]}{E_B\left[\psi(x|x_0)\right]}-\frac{E_B[\nabla\dot{\psi}(x|x_0)\left[d\sigma^2\right]]E_B\left[\partial_a\psi(x|x_0)\right]}{E_B\left[\psi(x|x_0)\right]^2} \\
&-\frac{E_B[\partial_a\dot{\psi}(x|x_0)]E_B\left[\nabla\psi(x|x_0)\left[d\sigma^2\right]\right]}{E_B\left[\psi(x|x_0)\right]^2}-\frac{E_B[\dot{\psi}(x|x_0)]E_B\left[\nabla\partial_a\psi(x|x_0)\left[d\sigma^2\right]\right]}{E_B\left[\psi(x|x_0)\right]^2} \\
&+\frac{2E_B[\dot{\psi}(x|x_0)]E_B\left[\nabla\psi(x|x_0)\left[d\sigma^2\right]\right]E_B\left[\partial_a\psi(x|x_0)\right]}{E_B\left[\psi(x|x_0)\right]^3}.
\end{aligned}$$

Using arguments similar to the ones employed in the proof of Lemma 20, one will realise that the desired bound will hold with $b(x|x_0)=C(1+|x|^{2q+1}+|x_0|^{2q+1})$. Finally, the inequality

$$\|\nabla s(\theta,\mu_0)[d\mu]-\nabla s(\theta_0,\mu_0)[d\mu]\|\leq b(x|x_0)\,\|d\mu\|_{2,\infty}\,\|\theta-\theta_0\|$$

is shown to hold in the same manner. ■

Lemma 23 *Under (C2.1)-(C2.6),*

$$\left\|S_n\left(\theta_0,\hat\sigma^2\right)-S_n\left(\theta_0,\hat\sigma_0^2\right)-\nabla S_n\left(\theta_0,\hat\sigma_0^2\right)\left[\hat\sigma^2-\hat\sigma_0^2\right]\right\|\leq O_P(1)\times\left\|\hat\sigma^2-\hat\sigma_0^2\right\|_\infty^2.$$

Proof. This follows along the same lines as the proof of Lemma 14, only here two extra terms have to be dealt with, $Z_t\left(\hat\sigma^2\right)-Z_t\left(\hat\sigma_0^2\right)-\nabla Z_t\left[\hat\sigma^2-\hat\sigma_0^2\right]$ and $\dot Z_t\left(\hat\sigma^2\right)-\dot Z_t\left(\hat\sigma_0^2\right)-\nabla\dot Z_t\left[\hat\sigma^2-\hat\sigma_0^2\right]$. This is done by applying the results of Lemma 42. ■

We define $\nabla_\pi s\left(x|x_0;\theta,\sigma^2\right)[d\pi]=\nabla s\left(x|x_0;\theta,\sigma^2\right)\left[\nabla\sigma^2[d\pi]\right]$, and obtain

Lemma 24 *Under (C2.1)-(C2.6), there exists b with $E_\pi\left[b\left(X_1|X_0\right)\right]<\infty$ such that*

$$\left\|\nabla s\left(x|x_0;\theta_0,\hat\sigma_0^2\right)\left[\hat\sigma^2-\hat\sigma_0^2\right]-\nabla_\pi s\left(x|x_0;\theta_0,\hat\sigma_0^2\right)\left[\hat\pi-\pi_0\right]\right\|\leq b(x|x_0)\sum_{i=0}^{3}a^{i-5}\|\hat\pi^{(i)}-\pi_0^{(i)}\|_\infty^2.$$

Proof. This follows from Lemma 21 together with 38 and 31. ■

The Adjustment Term

In this section we show that the pathwise derivative can be written as a normalised sum.

Lemma 25 *Under (C2.1)-(C2.7),*

$$\nabla S_n\left(\theta_0,\hat\sigma_0^2\right)\left[\hat\sigma^2-\sigma_0^2\right]=\nabla S\left(\theta_0,\hat\sigma_0^2\right)\left[\hat\sigma^2-\sigma_0^2\right]+o_P(n^{-1/2}).$$

Proof. The result is obtained by copying the arguments of the proof of Lemma 16, only now using Lemma 21 and 24. ■

Lemma 26 *Under (C2.1)-(C2.7),*

$$\nabla S\left(\theta_0,\hat\sigma_0^2\right)\left[\hat\sigma^2-\hat\sigma_0^2\right]=\nabla S\left(\theta_0,\sigma_0^2\right)\left[\hat\sigma^2-\hat\sigma_0^2\right]+o_P(n^{-1/2}).$$

Proof. This can be shown by the same line of arguments employed in the proof of Lemma 20. ■

Lemma 27 *Under (C2.1)-(C2.7), there exists a function δ with $E_\pi\left[\delta\left(X_0\right)\right]=0$ and $E_\pi\left[\left\|\delta\left(X_0\right)\right\|^{2+\epsilon}\right]<\infty$ such that*

$$\nabla S\left(\theta_0,\sigma_0^2\right)\left[\hat\sigma^2-\hat\sigma_0^2\right]=\frac{1}{n}\sum_{i=1}^{n}\delta\left(X_i\right)+o_P(n^{-1/2}).$$

Proof. We proceed as in the proof of Lemma 18, only now defining the Hilbert space $\mathcal{H}_k$ as

$$\mathcal{H}_k=\{f:\mathbb{R}\mapsto\mathbb{R}|f\text{ is }k\text{ times continuously differentiable and bounded}\},$$

since we have that, in contrast to μ, σ^2 is bounded. We equip $\mathcal{H}_k$ with the inner product

$$\begin{aligned}\langle f,g\rangle &= \int f(x)\,g(x)\,\pi_0(x)\,dx + E_{\pi\times B}\left[\int_0^1 f(Z_t)\,g(Z_t)\,dt\right]\\ &\quad+\int\int\left(\int_{\min\{x,x_0\}}^{\max\{x,x_0\}} f(w)\,g(w)\,dw\right)p_0(x|x_0)\,\pi_0(x)\,dx dx_0.\end{aligned}$$

■

The Hessian

Lemma 28 *Under (C2.1)-(C2.7),* $\sup_{\theta\in\Theta}\left\|H_n\left(\theta,\hat{\sigma}^2\right)-H\left(\theta,\sigma_0^2\right)\right\|\to^P 0$.

Proof. The Hessian, $h=\partial_\theta s$, is given in (4.37) with $\nabla\partial_x^i\mu=\partial_x^i\dot{\mu}$, $\nabla\partial_x^i\dot{\mu}=\partial_x^i\ddot{\mu}$, $\nabla\partial_x^i\sigma^2=\partial_x^i\dot{\sigma}^2$, and $\nabla\partial_x^i\dot{\sigma}^2=\partial_x^i\ddot{\sigma}^2$. Copy the arguments of Lemma 19, this time employing the inequalities in Lemma 40, 41, and 42 to obtain that

$$\left\|h\left(x|x_0;\theta,\hat{\sigma}^2\right)-h\left(x|x_0;\theta,\hat{\sigma}_0^2\right)\right\|\le b(x|x_0)\sum_{i,j=0}^{3}||\partial_{x,\theta}^{ij}\hat{\sigma}^2-\partial_{x,\theta}^{ij}\hat{\sigma}_0^2||_\infty,$$

for some function b with $E_\pi\left[b(X_1|X_0)\right]$By the same line of arguments employed in the proof of Theorem 8, we obtain

$$h\left(x|x_0;\theta,\hat{\sigma}_0^2\right)-h\left(x|x_0;\theta,\sigma_0^2\right)=\partial_a h\left(x|x_0;\theta,\hat{\sigma}_0^2\right)a$$

where, following the same procedure as in the proof of Lemma 22,

$$\begin{aligned}\left\|\partial_a h\left(x|x_0;\theta,\hat{\sigma}_0^2\right)\right\| &\le b(x|x_0)\{|\partial_a\hat{T}(x;a)|+|\partial_a\hat{T}(x;a)|\\ &\quad+\int_{\min\{x,x_0\}}^{\max\{x,x_0\}}|\partial_a\hat{T}(w;a)|dw+E_B[\int_0^1|\partial_a\hat{T}(Z_t;a)|^2dt]^{1/2}\}.\end{aligned}$$

■

4.B.3 Auxiliary Lemmas

Lemma 29 *Under (A0),* $\{X_i\}$ *is* β*-mixing with* $\beta_i\le c\exp(-\rho i)$, $\rho>0$, *and* $E_\pi\left[V(X_t)\right]<\infty$.

Proof. This follows from Meyn & Tweedie (1993, Theorem 6.1). ■

Lemma 30 *Under (C1.1)-(C1.3) or (C2.1)-(C2.3), the transition density* p *for* $\{X_t\}$ *with invariant density* π*, takes the form*

$$p(x|x_0)=\frac{1}{\sqrt{\sigma(x)\,\sigma(x_0)}\sqrt{2\pi\Delta}}\sqrt{\frac{\pi(x)}{\pi(x_0)}}\exp\left[-\left(\int_{x_0}^{x}\sigma(w)^{-1}dw\right)^2/2\Delta\right]E_B\left[\psi(x|x_0)\right]. \tag{4.56}$$

Proof. We have by Dacunha-Castelle and Florens-Zmirou (1986, Lemma 1),

$$p(x|x_0) = \frac{1}{\sigma(x)\sqrt{2\pi\Delta}} \exp\left[-(\gamma(x)-\gamma(x_0))^2/2\Delta + \int_{\gamma(x_0)}^{\gamma(x)} \mu_Y(w)\,dw\right] E_B[\psi(x,x_0)],$$

Using (4.21) and

$$\frac{1}{2}\left[\log\left(\sigma^2(w)\pi(w)\right)\right]_{x_0}^{x} = \int_{x_0}^{x} \frac{\mu(w)}{\sigma^2(w)}dw,$$

(c.f. (4.3)), we obtain that $\gamma(x) - \gamma(x_0) = \int_{x_0}^{x} \sigma(w)^{-1}\,dw$, and

$$\begin{aligned}
\int_{\gamma(x_0)}^{\gamma(x)} \mu_Y(w)\,dw &= \int_{x_0}^{x} \mu_Y(\gamma(w))\,\sigma(w)^{-1}\,dw \\
&= \int_{x_0}^{x} \left[\frac{\mu(w)}{\sigma(w)} - \frac{1}{2}\frac{\partial\sigma(w)}{\partial w}\right]\sigma(w)^{-1}\,dw \\
&= \int_{x_0}^{x} \left[\frac{\mu(w)}{\sigma^2(w)} - \frac{1}{2}\frac{\partial\log\sigma(w)}{\partial w}\right]dw \\
&= \frac{1}{2}\left[\log\left(\sigma^2(w)\pi(w)\right)\right]_{x_0}^{x} - \frac{1}{2}\left[\log(\sigma(w))\right]_{x_0}^{x} \\
&= \log(\sigma(x)/\sigma(x_0)) + \frac{1}{2}\log(\pi(x)/\pi(x_0)) - \frac{1}{2}\log(\sigma(x)/\sigma(x_0)) \\
&= \frac{1}{2}\log(\sigma(x)/\sigma(x_0)) + \frac{1}{2}\log(\pi(x)/\pi(x_0))
\end{aligned}$$

Plugging in these expressions,

$$\begin{aligned}
p(x|x_0) &= \frac{1}{\sigma(x)\sqrt{2\pi\Delta}} \exp\left[-\left(\int_{x_0}^{x}\sigma(w)^{-1}\,dw\right)^2/2\Delta\right]\sqrt{\frac{\sigma(x)}{\sigma(x_0)}}\sqrt{\frac{\pi(x)}{\pi(x_0)}}E_B[\psi(x|x_0)] \\
&= \frac{1}{\sqrt{\sigma(x)\sigma(x_0)}\sqrt{2\pi\Delta}}\sqrt{\frac{\pi(x)}{\pi(x_0)}}\exp\left[-\left(\int_{x_0}^{x}\sigma(w)^{-1}\,dw\right)^2/2\Delta\right]E_B[\psi(x|x_0)].
\end{aligned}$$

■

To prove that $\hat{\pi}$ and its derivatives converge uniformly, we make use of a general result by Andrews (1994, Theorem 1):

Lemma 31 *Assume that (A0)-(A1) holds and the kernel $K \in \mathcal{K}(\omega,\lambda)$. Then*

$$\sup_{x\in I}\left|\hat{\pi}^{(\lambda)}(x) - \pi_0^{(\lambda)}(x)\right| = O_P(n^{-1/2}h_\lambda^{-1-\lambda}) + O_P(h_\lambda^{\omega-\lambda}) \tag{4.57}$$

When working with the pathwise derivatives, we need a U-statistics result for dependent sequences. Let $\{Y_i\}$ be a stationary and β-mixing sequence with mixing coefficients $\{\beta_{Y,i}\}$ taking values in $\mathbb{R}^p$. Consider the sequence

$$V_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i=1}^{n} m_n(Y_i, Y_j)$$

where $(m_n)_{n\geq 1}$ is a sequence of Borel measurable functions, $m_n : \mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}^q$. We introduce the following projections,

$$\begin{aligned} V_{1,n} &= \frac{1}{n}\sum_{i=1}^{n} m_{1,n}(Y_i), \quad m_{1,n}(y_1) = \int m_n(y_1, y_2)\, dP(y_2), \\ V_{2,n} &= \frac{1}{n}\sum_{j=1}^{n} m_{2,n}(Y_j), \quad m_{2,n}(y_1) = \int m_n(y_1, y_2)\, dP(y_2), \\ \overline{V}_n &= \int m_n(y_1, y_2)\, dP(y_1, y_2), \end{aligned}$$

where P is the distribution of $\{Y_i\}$ on $\mathbb{R}^p$. The following projection theorem is a standard results from the theory of V-statistics, see for example Arcones (1995), Robinson (1989) and Yoshihara (1976).

Lemma 32 *Assume that for some $p > 2$,*

1. $M_n \equiv \sup_{i,j\geq 1} E\left[\|m_n(Y_i, Y_j)\|^p\right]^{1/p} = o(\sqrt{n})$.

2. $\sum_{k=1}^{\infty} \beta_{Y,k} k^{2/(p-2)} < \infty$.

Then,

$$V_n - V_{1,n} - V_{2,n} + \overline{V}_n = o_P(n^{-1/2}). \tag{4.58}$$

We deliver a proof for completeness based on Arcones (1995).

Proof. We first define the U-statistics, $U_n = 1/n^2 \sum_{1\leq i\neq j\leq n} h_n(Y_i, Y_j)$, where

$$h_n(y_1, y_2) = m_n(y_1, y_2) + m_n(y_2, y_1) - 2\overline{V}_n,$$

and its projection, $U_{1,n} = 1/n \sum_{i=1}^{n} h_{1,n}(Y_i)$, where

$$h_{1,n}(y_1) = \int h_n(y_1, y_2)\, d\pi_Y(y_2) = m_{1,n}(y_1) + m_{2,n}(y_1) - 2\overline{V}_n.$$

Note that $h_n(y_1, y_2)$ is symmetric and that

$$V_n - \overline{V}_n = \frac{1}{n^2}\sum_{i=1}^{n} m_n(Y_i, Y_i) + \frac{n-1}{2n} U_n,$$

and $V_{1,n} + V_{2,n} - 2\overline{V}_n = U_{1,n}$. Since

$$\frac{1}{n^2}\sum_{i=1}^{n} E\left[\|m_n(Y_i, Y_i)\|\right] \leq \frac{1}{n^2}\sum_{i=1}^{n} E\left[\|m_n(Y_i, Y_i)\|^p\right]^{1/p} \leq \frac{1}{n} M_n = o\left(1/\sqrt{n}\right),$$

we have that $n^{-2}\sum_{i=1}^{n} m_n(Y_i, Y_i) = o_P\left(n^{-1/2}\right)$. So we show that $U_n - U_{1,n} = o_P\left(n^{-1/2}\right)$. The proof of this follows along the same lines as the proof of Arcones (1995, Theorem 1). He works with a function h which is independent of n but we observe that his Lemma 2 and 3 hold for n-dependent h functions as well.

Let $(\xi_i)_{i\geq 1}$ be a sequence of i.i.d. variables with the same distribution as Y_0. Arcones (1995, Lemma 2) implies that for any $n \geq 1$, $1 \leq i \leq n$, $1 \leq r \leq n-i$ and $a > 0$,

$$\left|E\left[\|h_n(Y_i, Y_{i+r})\|^p 1_{\{|m_n|\leq a\}}\right] - E\left[\|h_n(Y_i, \xi_2)\|^p 1_{\{|m_n|\leq a\}}\right]\right| \leq \beta_{Y,r} a^p,$$

which again implies that

$$E\left[|h_n(Y_i, \xi_2)|^p\right] \leq 2M_n^p \tag{4.59}$$

for any $n \geq 1$ and $1 \leq i \leq n$. Define the function

$$\tilde{h}_n(y_1, y_2) = h_n(y_1, y_2) - h_{1,n}(y_1) - h_{1,n}(y_2) + h_n(y_1, y_2).$$

Observe that $\tilde{h}_n(y_1, y_2)$ is also symmetric and, by (4.59), $\sup_{1\leq i\leq n} E[|\tilde{h}_n(Y_i, Y_j)|^p] < 8M_n^p$. We may then apply Arcones (1995, Lemma 3) to $\tilde{h}_n$ to obtain

$$E[n^{-3/2}\|\sum_{1\leq i\neq j\leq n} \tilde{h}_n(Y_i, Y_j)\|] \leq C(8n^{-1}M_n^2 + 8n^{-1}M_n^2 \sum_{j=1}^{n-1} j\beta_{Y,j}^{2/(p-2)}). \tag{4.60}$$

From our assumptions, $\sum_{j=k}^{2k} j^{2/(p-2)}\beta_{Y,j} \to 0$ such that $\beta_{Y,2k}k^{p/(p-2)} \to 0$ since

$$\sum_{j=k}^{2k} j^{2/(p-2)}\beta_{Y,j} \geq \sum_{j=k}^{2k} j^{2/(p-2)}\beta_{Y,2k} = \beta_{Y,2k} O\left(k^{p/(p-2)}\right).$$

We conclude that (4.60) converges towards zero. Hence,

$$\sqrt{n}(U_n - U_{1,n}) = \frac{1}{2n^{3/2}} \sum_{1\leq i\neq j\leq n} \tilde{h}_n(Y_i, Y_j) = o_P(1).$$

■

Class 1

Lemma 33 *Under (C1.1)-(C1.6a),*

$$\|\partial_\theta^j \hat{\mu}^{(k)}(\cdot;\theta) - \partial_\theta^j \hat{\mu}_0^{(k)}(\cdot;\theta)\|_\infty \leq C\sum_{i=0}^{k+1} a^{i-k-2}\|\hat{\pi}^{(i)} - \pi_0^{(i)}\|_\infty,$$

$$\|\partial_\theta^j \partial_x^k \hat{\mu}(\cdot;\theta) - \partial_\theta^j \partial_x^k \hat{\mu}_0(\cdot;\theta) - \partial_\theta^j \partial_x^k \nabla_\pi \hat{\mu}_0(\cdot;\theta)[\hat{\pi} - \pi_0]\|_\infty \leq C\sum_{i=0}^{k+1} a^{i-k-3}\|\hat{\pi}^{(i)} - \pi_0^{(i)}\|_\infty^2$$

uniformly over $\theta \in \Theta$, where $\partial_\theta^j \partial_x^k \nabla_\pi \hat{\mu}_0(\cdot;\theta)[d\pi]$ is given in (4.63)-(4.65).

Proof. We fix $\theta \in \Theta$, and suppress it throughout the proof. We introduce some trimming sets. Define

$$A(\varepsilon) = A_1(\varepsilon) \cap A_2(\varepsilon) \tag{4.61}$$

where

$$A_1(\varepsilon) = \{x|\hat{\pi}(x) \geq \varepsilon a\}, \quad A_2(\varepsilon) = \{x|\pi_0(x) \geq \varepsilon a\}, \tag{4.62}$$

for any $\varepsilon > 0$. As shown in Andrews (1994, p. 588), $A(\varepsilon) \supseteq A_1(2\varepsilon)$ with probability 1 as $n \to \infty$ under (C1.1) and (C1.6a). Assume that $\sup_{x\in A(\varepsilon)} |\mu(x;\hat{\pi}) - \mu(x;\pi_0)| \to^P 0$. Then, by (T0),

$$\|\hat{\mu} - \hat{\mu}_0\|_\infty \leq \sup_{x\in A_1(1)} |\mu(x;\hat{\pi}) - \mu(x;\pi_0)| \leq \sup_{x\in A(1/2)} |\mu(x;\hat{\pi}) - \mu(x;\pi_0)| \to^P 0.$$

The same argument works for derivatives w.r.t. x and θ. So in the following we establish convergence uniformly over $A(\varepsilon)$.

We observe that $\mu(x;\pi) = \frac{1}{2}\partial_x\sigma^2(x) + \frac{1}{2}\sigma^2(x)\pi^{(1)}(x)/\pi(x)$. Thus, by (C1.2),

$$|\mu(x;\hat{\pi}) - \mu(x;\pi_0)| \leq \frac{1}{2}\bar{\sigma}^2 \left|\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)}\right|,$$

where

$$\begin{aligned}
&\sup_{x\in A(\varepsilon)} \left|\frac{\hat{\pi}^{(1)(x)}}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)}\right| \\
\leq\ & \sup_{x\in A(\varepsilon)} \left\{\hat{\pi}(x)^{-1} |\hat{\pi}^{(1)}(x) - \pi_0^{(1)}(x)|\right\} + \sup_{x\in A(\varepsilon)} |\pi_0^{(1)}(x)| \left|\frac{1}{\hat{\pi}(x)} - \frac{1}{\pi_{0(x)}}\right| \\
\leq\ & Ca^{-1}\|\hat{\pi}^{(1)} - \pi_0^{(1)}\|_\infty + Ca^{-2}\|\hat{\pi} - \pi_0\|_\infty .
\end{aligned}$$

Next,

$$\mu^{(1)}(x;\pi) = \frac{1}{2}\partial_x^2\sigma^2(x) + \frac{1}{2}\partial_x\sigma^2(x)\frac{\pi^{(1)}(x)}{\pi(x)} + \frac{1}{2}\sigma^2(x)\left(\frac{\pi^{(2)}(x)}{\pi(x)} - \frac{\pi^{(1)}(x)^2}{\pi(x)^2}\right),$$

such that, by (C1.2),

$$\begin{aligned}
|\mu^{(1)}(x;\hat{\pi}) - \mu_0^{(1)}(x;\pi_0)| \ \leq\ & \bar{\sigma}_1^2\left|\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)}\right| + \frac{1}{2}\bar{\sigma}^2\left|\frac{\hat{\pi}^{(2)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(2)}(x)}{\pi_0(x)}\right| \\
&+\frac{1}{2}\bar{\sigma}^2\left|\frac{\hat{\pi}^{(1)}(x)^2}{\hat{\pi}(x)^2} - \frac{\pi_0^{(1)}(x)^2}{\pi_0(x)^2}\right|,
\end{aligned}$$

where

$$\begin{aligned}
\sup_{x\in A(\varepsilon)}\left|\frac{\hat{\pi}^{(2)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(2)}(x)}{\pi_0(x)}\right| \ &\leq\ a^{-1}\|\hat{\pi}^{(2)} - \pi_0^{(2)}\|_\infty + Ca^{-2}\|\hat{\pi} - \pi_0\|_\infty, \\
\sup_{x\in A(\varepsilon)}\left|\frac{\hat{\pi}^{(1)}(x)^2}{\hat{\pi}(x)^2} - \frac{\pi_0^{(1)}(x)^2}{\pi_0(x)^2}\right| \ &\leq\ \sup_{x\in A(\varepsilon)}\left|\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} + \frac{\pi_0^{(1)}(x)}{\pi_0(x)}\right| + \sup_{x\in A(\varepsilon)}\left|\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0(x)}\right| \\
&\leq\ a^{-2}\|\hat{\pi}^{(2)} - \pi_0^{(2)}\|_\infty + Ca^{-3}\|\hat{\pi} - \pi_0\|_\infty .
\end{aligned}$$

The results for $\mu^{(i)}$, $i = 2, 3, 4$, follow in the same manner. The derivatives w.r.t. θ satisfy same bounds due to the bounds imposed on $\partial_{x,\theta}^{i,j}\sigma^2$, $0 \leq i \leq 4$ and $0 \leq j \leq 2$.

Next, we turn to the pathwise derivatives. We define

$$\nabla\hat{\mu}_0(x)[d\pi] = \hat{T}(x;a)\left\{D_{0,0}(x)\,d\pi(x) + D_{0,1}(x)\,d\pi^{(1)}(x)\right\} \tag{4.63}$$

where

$$D_{0,0}(x) = -\frac{1}{2}\sigma^2(x)\frac{\pi_0^{(1)}(x)}{\pi_0^2(x)}, \quad D_{0,1}(x) = \frac{1}{2}\sigma^2(x)\frac{1}{\pi_0(x)}, \tag{4.64}$$

and inductively for $i \geq 1$,

$$\nabla\partial_x^i\hat{\mu}_0(x)[d\pi] = \hat{T}(x;a)\sum_{j=0}^{i+1} D_{i,j}(x)\,d\pi^{(j)}(x).$$

where

$$\begin{aligned} D_{i,0}(x) &= D_{i-1,0}^{(1)}(x), \\ D_{i,j}(x) &= D_{i-1,j-1}(x) + D_{i-1,j}^{(1)}(x), \quad 1 \leq j \leq i, \\ D_{i,i+1}(x) &= \frac{1}{2}\sigma^2(x)\frac{1}{\pi_0(x)}. \end{aligned} \tag{4.65}$$

We show that the assertion holds for $\nabla\mu_0^{(i)}$, $i = 0, 1$. The remaining ones follow along the same lines. Observe that by a standard 2nd order Taylor expansion of the function $(f, g) \mapsto f/g$,

$$\begin{aligned} &\frac{f}{g} - \frac{f_0}{g_0} - \frac{f - f_0}{g_0} + \frac{f_0(g - g_0)}{g_0^2} \\ = \;& \frac{(f - f_0)(g - g_0)}{(\lambda g + (1-\lambda)g_0)^2} - 2\frac{\lambda g + (1-\lambda)g_0}{(\lambda g + (1-\lambda)g_0)^3}(g - g_0)^2. \end{aligned}$$

Thus,

$$\begin{aligned} |\hat{\mu} - \hat{\mu}_0 - \nabla\hat{\mu}_0[\hat{\pi} - \pi_0]| &\leq \frac{1}{2}\bar{\sigma}^2\left|\frac{\hat{\pi}^{(1)}}{\hat{\pi}} - \frac{\pi_0^{(1)}}{\pi_0} + \frac{\pi_0^{(1)}}{\pi_0^2}(\hat{\pi} - \pi_0) - \frac{1}{\pi_0}(\hat{\pi}^{(1)} - \pi_0^{(1)})\right| \\ &\leq C\left\{a^{-3}||\hat{\pi} - \pi_0||_\infty^2 + a^{-2}||\hat{\pi}^{(1)} - \pi_0^{(1)}||_\infty^2\right\}. \end{aligned}$$

In a similar fashion, Taylor expanding the function $(f, g) \mapsto f^2/g^2$,

$$\begin{aligned} &\left|\hat{\mu}^{(1)} - \hat{\mu}_0^{(1)} - \nabla\hat{\mu}_0^{(1)}[\hat{\pi} - \pi_0]\right| \\ \leq\; & \bar{\sigma}_1^2\left|\frac{\hat{\pi}^{(1)}}{\hat{\pi}} - \frac{\pi^{(1)}}{\pi} + \frac{\pi_0^{(1)}}{\pi_0^2}(\hat{\pi} - \pi_0) - \frac{1}{\pi_0}(\hat{\pi}^{(1)} - \pi_0^{(1)})\right| \\ & + \frac{1}{2}\bar{\sigma}^2\left|\frac{\hat{\pi}^{(2)}}{\hat{\pi}} - \frac{\pi^{(2)}}{\pi} + \frac{\pi_0^{(2)}}{\pi_0^2}(\hat{\pi} - \pi_0) - \frac{1}{\pi_0}(\hat{\pi}^{(2)} - \pi_0^{(2)})\right| \\ & + \frac{1}{2}\bar{\sigma}^2\left|\frac{\hat{\pi}^{(1)2}}{\hat{\pi}^2} - \frac{\pi^{(1)2}}{\pi^2} - 2\frac{\pi_0^{(1)}}{\pi_0^2}(\hat{\pi}^{(1)} - \pi_0^{(1)}) + 2\frac{\pi_0^{(1)2}}{\pi_0^3}(\hat{\pi} - \pi_0)\right| \\ \leq\; & C\sum_{i=0}^{2} a^{-4+i}||\hat{\pi}^{(i)} - \pi_0^{(i)}||_\infty^2. \end{aligned}$$

The results for the derivatives w.r.t. θ are shown in the same fashion. ■

Lemma 34 *Under (C1.1)-(C1.6a),*

$$\left\|\partial_{x,\theta}^{kl}\lambda_Y(\cdot;\theta,\hat{\mu})-\partial_{x,\theta}^{kl}\lambda_Y(\cdot;\theta,\hat{\mu}_0)\right\|_\infty\leq C\sum_{i,j=0}^{k+1}a^{i-k-1}||\partial_{x,\theta}^{ij}\hat{\mu}-\partial_{x,\theta}^{ij}\hat{\mu}_0^{(j)}||_\infty,$$

for $0\leq k,l\leq 2$. *Moreover,*

$$\begin{aligned}|\lambda_Y(z;\theta,\mu)| &\leq C\left(1+|\mu(z;\theta)|^2+|\mu^{(1)}(z;\theta)|\right),\\ |\lambda_Y^{(1)}(z;\theta,\mu)| &\leq C\left(1+|\mu(z;\theta)|^2+|\mu(z)|\,\mu^{(1)}(z;\theta)|+|\mu^{(2)}(z;\theta)|\right),\\ ||\dot{\lambda}_Y(z;\theta,\mu)|| &\leq C\left(1+|\mu(z;\theta)|^2+|\mu(z)|\,|\dot{\mu}(z;\theta)|+||\dot{\mu}^{(1)}(z;\theta)||\right).\end{aligned}$$

Proof. We suppress the dependence on $\theta\in\Theta$ which we fix. From (4.32),

$$\|\lambda_Y(\cdot;\hat{\mu})-\lambda_Y(\cdot;\hat{\mu}_0)\|_\infty\leq\frac{1}{2\underline{\sigma}}\left\|\hat{\mu}^2-\hat{\mu}_0^2\right\|_\infty+\frac{1}{2}||\hat{\mu}^{(1)}-\hat{\mu}_0^{(1)}||_\infty\leq\frac{\overline{\sigma}^2}{2\underline{\sigma}}\sum_{i=0}^{1}a^{-1+i}||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty,$$

since $||\hat{\mu}^2-\hat{\mu}_0^2||\leq||\hat{\mu}+\hat{\mu}_0^2||_\infty|\,|\hat{\mu}-\hat{\mu}_0|\,|_\infty\leq Ca^{-1}||\hat{\mu}-\hat{\mu}_0||_\infty$. By similar calculations, the remaining inequalities follow.

Next,

$$|\lambda_Y(z)|\leq\frac{|\mu(z)|^2}{2\underline{\sigma}^2}+\frac{|\mu(z)|\,\bar{\sigma}_1^2}{2\underline{\sigma}^2}+\frac{|\mu^{(1)}(z)|}{2}-\frac{\bar{\sigma}_1^4}{32\underline{\sigma}^2}+\frac{\bar{\sigma}_1^2}{8}\leq C(1+|\mu(z)|^2+|\mu^{(1)}(z)|),$$

and similarly for the two other bounds. ■

Lemma 35 *Under (C1.1)-(C1.4),*

1. *The function D_μ in (4.39) satisfies*

$$||\partial_x^k\partial_\theta^l D_\mu(z;\theta,\mu)||\leq C\sum_{i=0}^{k}\sum_{j=0}^{l}(1+||\partial_x^i\partial_\theta^j\mu(z;\theta)||)$$

and

$$||\partial_x^k\partial_\theta^l D_\mu(z;\theta,\mu_1)-\partial_x^k\partial_\theta^l D_\mu(z;\theta,\mu_2)||\leq C\sum_{i=0}^{k}\sum_{j=0}^{l}||\partial_x^i\partial_\theta^j\mu_1-\partial_x^i\partial_\theta^j\mu_2||_\infty.$$

2. *There exists a constant $C>0$, such that*

$$||\lambda_Y(z;\theta,\hat{\mu})-\lambda_Y(z;\theta,\hat{\mu}_0)-\nabla\lambda_Y(z;\theta,\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]||\leq C\,\|\hat{\mu}-\hat{\mu}_0\|_\infty^2,$$

$$||\lambda_Y^{(1)}(z;\theta,\hat{\mu})-\lambda_Y^{(1)}(z;\theta,\hat{\mu}_0)-\nabla\lambda_Y^{(1)}(z;\theta,\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]||\leq C\sum_{i=0}^{1}||\hat{\mu}^{(i)}-\hat{\mu}_0^{(i)}||_\infty^2$$

$$||\dot{\lambda}_Y(z;\theta,\hat{\mu})-\dot{\lambda}_Y(z;\theta,\hat{\mu}_0)-\nabla\dot{\lambda}_Y(z;\theta,\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0]||\leq C\sum_{i=0}^{1}\left\|\partial_\theta^i\hat{\mu}-\partial_\theta^i\hat{\mu}_0\right\|_\infty^2.$$

Proof. Given the bounds on σ^2 in (C1.2), the first part of the theorem easily follows. The first claim in the second part follows from

$$\begin{aligned}\lambda_Y(\hat{\mu}) - \lambda_Y(\hat{\mu}_0) - \nabla\lambda_Y(\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0] &= -\frac{1}{2\underline{\sigma}^2}\left[\mu^2 - \hat{\mu}_0^2 - 2\hat{\mu}_0(\hat{\mu}-\hat{\mu}_0)\right] \\ &= -\frac{1}{2\underline{\sigma}^2}(\mu-\hat{\mu}_0)^2.\end{aligned}$$

For the second claim, we write

$$\begin{aligned}&\lambda_Y^{(1)}(\hat{\mu}) - \lambda_Y^{(1)}(\hat{\mu}_0) - \nabla\lambda_Y^{(1)}(\hat{\mu}_0)[\hat{\mu}-\hat{\mu}_0] \\ =\ & -\frac{1}{\sigma^2}\left(\mu\mu^{(1)} - \mu_0\mu_0^{(1)} - \mu_0(\mu^{(1)}-\mu_0^{(1)}) - \mu_0^{(1)}(\mu-\mu_0)\right) \\ &+\frac{\partial_x\sigma^2}{2\underline{\sigma}^4}\left(\mu^2-\hat{\mu}_0^2-2\hat{\mu}_0(\mu-\hat{\mu}_0)\right) \\ =\ & -\frac{1}{2\underline{\sigma}^2}(\mu-\hat{\mu}_0)(\mu^{(1)}-\hat{\mu}_0^{(1)}) + \frac{\partial_x\sigma^2}{2\underline{\sigma}^4}(\mu-\hat{\mu}_0)^2,\end{aligned}$$

where $||\mu-\hat{\mu}_0||_\infty||\mu^{(1)}-\hat{\mu}_0^{(1)}||_\infty \le ||\mu-\hat{\mu}_0||_\infty^2 + ||\mu^{(1)}-\hat{\mu}_0^{(1)}||_\infty^2$. We obtain the third inequality by similar arguments. ■

Lemma 36 *Under (C1.1)-(C1.4),* $\left|\partial_\theta^i Z_t(x|x_0;\theta)\right| \le C(1+|x|+|x_0|+|B_t|)$, $i=0,1,2$.

Proof. First, $|\gamma(x;\theta)| \le C(1+|x|)$. For example, for $x>0$,

$$\begin{aligned}\gamma(x;\theta) &= \gamma(0;\theta) + \int_0^x \frac{1}{\sigma(w;\theta)}dw \le \gamma(0;\theta) + \int_0^x \underline{\sigma}^{-1}dw \le \gamma(0;\theta) + \underline{\sigma}^{-1}x, \\ \gamma(x;\theta) &\ge \gamma(0;\theta) + \bar{\sigma}^{-1}x,\end{aligned}$$

where $\sup_{\theta\in\Theta}|\gamma(0;\theta)| < \infty$. This implies that for any $t\in[0,1]$,

$$\begin{aligned}|Z_t| &= |\gamma^{-1}(t\gamma(x) + (1-t)\gamma(x_0) + \sqrt{\Delta}B_t)| \\ &\le C(1+t|\gamma(x)| + (1-t)|\gamma(x_0)| + \sqrt{\Delta}|B_t|) \\ &\le C(1+|x|+|x_0|+|B_t|).\end{aligned} \tag{4.66}$$

The claimed bounds for $\dot{Z}_t$ and $\ddot{Z}_t$ are established in a similar fashion. ■

Class 2

Lemma 37 *Assume that: (i) (A0) holds: (ii)* $\theta \mapsto \mu(x;\theta)$ *is* $k+1$ *times continuously differentiable, satisfying* $||\partial_\theta^i\mu(x;\theta)|| \le C|x|^{\bar{q}/(2+\delta)}$ *for* $0\le i\le k+1$ *and some* $\delta>0$. *Then*

$$\sup_{(x,\theta)\in\mathbb{R}\times\Theta}\left|\frac{1}{n}\sum_{i=1}^n 1_{(l,x)}(X_i)\,\partial_\theta^k\mu(X_i;\theta) - \int_l^x \pi_0(y)\,\partial_\theta^k\mu(y;\theta)\,dy\right| = O_P(n^{-1/2}).$$

Proof. We have that (i) $\{X_i\}$ is stationary and absolutely regular/β-mixing with geometrically decreasing mixing coefficients, and (ii) $\left|1_{(l,x)}(z)\,\partial_\theta^k\mu(z;\theta)\right| \le b(z) \equiv C\left(1+|z|^{\bar{q}/(2+\delta)}\right)$

with $E_\pi\left[b^{2+\delta}(X_0)\right]<\infty$. Finally, we claim that (iii) the ε-entropy with bracketing of $\mathcal{G}$ for the $L_q(\pi_0)$-metric, $H_{B,q}(\varepsilon,\mathcal{G},\pi_0)$, where $\mathcal{G}=\left\{g|g(z)=1_{(l,x)}(z)\,\partial_\theta^k\mu(z;\theta),\ (x,\theta)\in I\times\Theta\right\}$, satisfies $H_{B,q}(\varepsilon,\mathcal{G},\pi)\leq C\varepsilon^{-\rho}$ for some $\rho<1/2$ and $q>2$. Applying Doukhan, Massart and Rio (1995), (i)-(iii) yield the result. To prove the last claim, (iii), define $\mathcal{H}=\left\{h|h(z)=1_{(l,x)}(z),\ x\in I\right\}$, such that $\mathcal{G}=\left\{g|g(z;\theta,h)=h(z)\,\partial_\theta^k\mu(z;\theta),\ (\theta,h)\in\Theta\times\mathcal{H}\right\}$. It holds that for any $(\theta,h),(\theta',h')\in\Theta\times\mathcal{H}$,

$$
\begin{aligned}
\left|g(z;\theta,h)-g\left(z;\theta',h'\right)\right| &\leq \left|\partial_\theta^k\mu(z;\theta)\right|\left\|h-h'\right\|_q+\left\|h'\right\|_q\left|\partial_\theta^k\mu(z;\theta)-\partial_\theta^k\mu\left(z;\theta'\right)\right| \\
&= \left|\mu(z;\theta)\right|\left\|h-h'\right\|_q+\left\|h'\right\|_q\left|\partial_\theta^{k+1}\mu\left(z;\bar{\theta}\right)\right|\left\|\theta-\theta'\right\| \\
&\leq b(z)\left(\left\|h-h'\right\|_q+\left\|\theta-\theta'\right\|\right).
\end{aligned}
$$

By the same arguments as in the proof of Chen et al (2003, Theorem 3), it now follows that

$$
H_{B,q}(\varepsilon,\mathcal{G},\pi)\leq H(C\varepsilon^{1/s},\Theta,\|\cdot\|)+H_q(C\varepsilon^{1/s},\mathcal{H},\pi_0),
$$

for any $s\in(0,1]$, where $H(\varepsilon,\Theta,\|\cdot\|)$ and $H_q(\varepsilon,\mathcal{H},\pi_0)$ are the ε-entropies of Θ (for the Euclidean norm) and $\mathcal{H}$ (for the $L_q(\pi_0)$-metric) respectively. By van de Geer (2000, Lemma 2.5), $H(\varepsilon,\Theta,\|\cdot\|)\leq d\log\left(4C\varepsilon^{-1}+1\right)$ while, by van de Geer (2000, Theorem 3.11 and Example 3.7.4a), $H_q(\varepsilon,\mathcal{H},\pi_0)\leq\log\left(C\varepsilon^{-q}\right)$. This proves (iii). ■

Lemma 38 *Under (C2.1)-(C2.4), the following holds uniformly over* $(x,\theta)\in\mathbb{R}\times\Theta$, $0\leq i\leq 4$, *and* $0\leq j\leq 2$:

$$
||\partial_x^i\partial_\theta^j\hat{\sigma}^2(x;\theta)-\partial_x^i\partial_\theta^j\hat{\sigma}_0^2(x;\theta)||\leq O_P(a^{-1-i}n^{-1/2})+C\sum_{k=0}^{i}a^{-2-i+k}||\hat{\pi}^{(k)}-\pi_0^{(k)}||_\infty,
$$

$$
\begin{aligned}
&||\partial_x^i\partial_\theta^j\hat{\sigma}^2(x;\theta_0)-\partial_x^i\partial_\theta^j\hat{\sigma}_0^2(x;\theta_0)-\partial_x^i\partial_\theta^j\nabla_\pi\hat{\sigma}_0^2(x;\theta_0)[\hat{\pi}-\pi_0]|| \\
\leq\ & O_P(1)\times\sum_{k=0}^{i}a^{-3-i+k}||\hat{\pi}^{(k)}-\pi_0^{(k)}||_\infty^2+O_P(n^{-1/2})\times\sum_{k=0}^{i}a^{-2-i+k}||\hat{\pi}^{(k)}-\pi_0^{(k)}||_\infty.
\end{aligned}
$$

Proof. In the following we fix $\theta\in\Theta$ and suppress any dependence on it. Recall the definition of $A(\varepsilon)$ in (4.61), and the results associated with this trimming set. Then

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\hat{\sigma}^2(x)-\hat{\sigma}_0^2(x)\right| \\
\leq\ & \sup_{x\in A(\varepsilon)}\left\{\hat{\pi}(x)^{-1}\left|\frac{1}{n}\sum_{i=1}^{n}1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta)-\int_{-\infty}^{x}\pi_0(y)\,\mu(y;\theta)\,dy\right|\right\} \\
&+2\sup_{x\in A(\varepsilon)}\left\{\int_{-\infty}^{x}|\mu(y)|\,\pi_0(y)\,dy\left|\frac{1}{\hat{\pi}(x)}-\frac{1}{\pi_0(x)}\right|\right\} \\
\leq\ & O_P(a^{-1}n^{-1/2})+Ca^{-2}\left\|\hat{\pi}-\pi_0\right\|_\infty.
\end{aligned}
$$

Next,

$$
\begin{aligned}
& \left|\partial_x \hat{\sigma}^2(x) - \partial_x \hat{\sigma}_0^2(x)\right| \\
\leq\ & 2\hat{T}(x;a)\left(\left|\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)}\right|\left|\hat{\sigma}^2(x) - \hat{\sigma}_0^2(x)\right| + \hat{\sigma}_0^2(x)\left|\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}^2(x)} - \frac{\pi_0^{(1)}(x)}{\pi_0^2(x)}\right|\right) \\
\leq\ & C\left\{O_P(a^{-2}n^{-1/2}) + a^{-3}\left\|\hat{\pi} - \pi_0\right\|_\infty + a^{-2}\|\hat{\pi}^{(1)} - \pi_0^{(1)}\|_\infty\right\},
\end{aligned}
$$

The remaining bounds follow along the same lines.

We now turn to the pathwise derivatives of $\sigma^2(x)$ w.r.t. π. Define

$$
\begin{aligned}
& \nabla_\pi \hat{\sigma}_0^2(x)[\nabla\pi] \\
=\ & \hat{T}(x;a)\left\{\frac{-\nabla\pi(x)}{\pi_0(x)}\sigma_0^2(x) + \frac{2}{\pi_0(x)}\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta) - \sigma_0^2(x)\right\}
\end{aligned}
$$

such that

$$
\begin{aligned}
& \left|\hat{\sigma}^2(x) - \hat{\sigma}_0^2(x) - \partial_x^i \nabla_\pi \hat{\sigma}_0^2(x)[\hat{\pi} - \pi_0]\right| \\
\leq\ & \left|\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta)\right|\left|\frac{1}{\hat{\pi}(x)} - \frac{1}{\pi_0(x)} + \frac{\hat{\pi}(x) - \pi_0(x)}{\pi_0^2(x)}\right| \\
& + \frac{|\hat{\pi}(x) - \pi_0(x)|}{\pi_0^2(x)}\left|\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta) - \int_{-\infty}^x \pi_0(y)\,\mu(y;\theta)\,dy\right| \\
\leq\ & O_P(1) \times \left|\frac{(\hat{\pi}(x) - \pi_0(x))^2}{\pi_0^3(x) + \hat{\pi}^3(x)}\right| + O_P(n^{-1/2}) \times \frac{|\hat{\pi}(x) - \pi_0(x)|}{\pi_0^2(x)}.
\end{aligned}
$$

Similarly, with

$$
\begin{aligned}
\nabla_\pi \partial_x \hat{\sigma}^2(x;\theta) \equiv\ & \hat{T}(x;a)\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta)\left\{\frac{4\pi_0^{(1)}(x)}{\pi_0(x)^3}d\pi(x) - \frac{2d\pi^{(1)}(x)}{\pi_0(x)^2}\right\} \\
& -\hat{T}(x;a)\frac{\pi_0^{(1)}(x)}{\pi_0^2(x)}\frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x)}(X_i)\,\mu(X_i;\theta) + \partial_x \hat{\sigma}_0^2(x;\theta)
\end{aligned}
$$

we obtain

$$\begin{aligned}
& \left|\partial_x\hat{\sigma}^2(x) - \partial_x\hat{\sigma}_0^2(x) - \nabla_\pi\partial_x\hat{\sigma}_0^2(x)[\hat{\pi} - \pi_0]\right| \\
\leq\ & O_P(1) \times \hat{T}(x;a)\left|\frac{2\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)^2} - \frac{2\pi_0^{(1)}(x)}{\pi_0(x)^2}\right| \\
& + O_P(1) \times \hat{T}(x;a)\left|\frac{4\pi_0^{(1)}(x)}{\pi_0(x)^3}(\hat{\pi}(x) - \pi_0(x))\right| \\
& + O_P(1) \times \hat{T}(x;a)\left|\frac{2\left(\hat{\pi}^{(1)}(x) - \pi_0^{(1)}(x)\right)}{\pi_0(x)^2}\right| \\
& + O_P\left(n^{-1/2}\right) \times \hat{T}(x;a)\left\{\frac{4\left|\pi_0^{(1)}(x)\right|}{\pi_0(x)^3}|\hat{\pi}(x) - \pi_0(x)| + \frac{2\left(\hat{\pi}^{(1)}(x) - \pi_0^{(1)}(x)\right)}{\pi_0(x)^2}\right\} \\
\leq\ & O_P(1) \times \sum_{k=0}^{1} a^{-4+k}||\hat{\pi}^{(k)} - \pi_0^{(k)}||_\infty^2 + O_P\left(n^{-1/2}\right) \times \sum_{k=0}^{1} a^{-3+k}||\hat{\pi}^{(k)} - \pi_0^{(k)}||_\infty.
\end{aligned}$$

By similar arguments the pathwise derivatives of $\partial_x^2\hat{\sigma}^2(x)$ and $\partial_x^3\hat{\sigma}^2(x)$ are shown to satisfy the claimed inequalities. Once these have been established, the inequalities involving the pathwise derivatives of $\partial_x^i\partial_\theta^j\hat{\sigma}^2(x)$ are easily proven. ■

Lemma 39 *Under (C2.1)-(C2.4), it holds that (i)* $-(1 + C|z|^q) \leq \lambda_Y(z;\theta,\sigma_0^2) \leq \bar{\lambda}_Y$ *and (ii)* $|\lambda_Y^{(1)}(z;\theta,\hat{\sigma}_0^2)| \leq C + \left|\mu^{(2)}(z;\theta)\right|$.

Proof. This follows from (C2.2)-(C2.3). ■

Lemma 40 *Under (C2.1)-(C2.4) and (C2.6a),*

$$\left|\partial_{x,\theta}^{ij}\lambda_Y\left(z;\theta,\hat{\sigma}^2\right) - \partial_{x,\theta}^{ij}\lambda_Y\left(z;\theta,\hat{\sigma}_0^2\right)\right| \leq C\sum_{k=0}^{i+2}\sum_{l=0}^{j} a^{-3-i}||\partial_{x,\theta}^{kl}\hat{\sigma}^2 - \partial_{x,\theta}^{kl}\hat{\sigma}_0^2||_\infty$$

for $0 \leq i,j \leq 2$.

Proof. In the following we suppress the dependence on θ. We have

$$\lambda_Y(z) = -\frac{\mu^2(z)}{2\sigma^2(z)} + \frac{\mu(z)\partial_x\sigma^2(z)}{2\sigma^2(z)} - \frac{\mu^{(1)}(z)}{2} - \frac{\left[\partial_x\sigma^2(z)\right]^2}{32\sigma^2(z)} + \frac{\partial_x\sigma^2(z)}{8}, \tag{4.67}$$

such that

$$\begin{aligned}
\left|\lambda_Y\left(z;\hat{\sigma}^2\right) - \lambda_Y\left(z;\hat{\sigma}_0^2\right)\right| \leq\ & \left(\frac{\mu^2(z)}{2} + \frac{\mu(z)\partial_x\hat{\sigma}_0^2(z)}{2} + \frac{\left[\partial_x\hat{\sigma}_0^2(z)\right]^2}{32}\right)\left|\frac{1}{\hat{\sigma}^2(z)} - \frac{1}{\hat{\sigma}_0^2(z)}\right| \\
& + \left(\frac{\mu(z)}{2\hat{\sigma}^2(z)} + \frac{1}{8}\right)\left|\partial_x\hat{\sigma}^2(z) - \partial_x\hat{\sigma}_0^2(z)\right| \\
& + \frac{1}{32\hat{\sigma}^2(z)}\left|\left[\partial_x\hat{\sigma}^2(z)\right]^2 - \left[\partial_x\hat{\sigma}_0^2(z)\right]^2\right|
\end{aligned}$$

where, by Lemma 38,

$$\begin{aligned}\left|\frac{1}{\hat{\sigma}^2(z)}-\frac{1}{\hat{\sigma}_0^2(z)}\right| &\le \frac{\left|\hat{\sigma}^2(z)-\hat{\sigma}_0^2(z)\right|}{\left[\lambda\hat{\sigma}^2(z)+(1-\lambda)\hat{\sigma}_0^2(z)\right]^2}\le \underline{\sigma}^{-4}\left|\hat{\sigma}^2(x)-\hat{\sigma}_0^2(x)\right|,\\ \left|\left[\partial_x\hat{\sigma}^2(z)\right]^2-\left[\partial_x\hat{\sigma}_0^2(z)\right]^2\right| &\le \left|\partial_x\hat{\sigma}^2(z)+\partial_x\hat{\sigma}_0^2(z)\right|\left|\partial_x\hat{\sigma}^2(z)-\partial_x\hat{\sigma}_0^2(z)\right|\\ &\le Ca^{-1}\left|\partial_x\hat{\sigma}^2(z)-\partial_x\hat{\sigma}_0^2(z)\right|,\end{aligned}$$

and for $z\in A(1)$,

$$\frac{\mu^2(z)}{2}+\frac{\mu(z)\,\partial_x\hat{\sigma}_0^2(z)}{2}+\frac{\left[\partial_x\hat{\sigma}_0^2(z)\right]^2}{32}\le Ca^{-1}\left(1+a^{-1}\right),$$

$$\frac{\mu(z)}{2\hat{\sigma}^2(z)}+\frac{1}{8}\le C\left(1+a^{-1}\right),\quad \frac{1}{32\hat{\sigma}^2(z)}\le C.$$

This shows the first inequality. By the same arguments the claims follow for $\partial_\theta^i\lambda_Y(x;\theta,\pi)$ and $\partial_x^i\lambda_Y(x;\theta,\pi)$, $i=1,2$. ■

Lemma 41 *Under (C2.1)-(C2.4) and (C2.6a),*

$$\|\partial_x^i\partial_\theta^j D_{\sigma^2}\left(z;\theta,\sigma^2\right)\|\le C(1+\|\partial_x^i\partial_\theta^j\mu(z;\theta)\|),\quad \|\partial_x^i\partial_\theta^j D_{\partial_x\sigma^2}(z)\left(z;\theta,\sigma^2\right)\|\le C,$$

and

$$\left|\lambda_Y\left(z;\hat{\sigma}^2\right)-\lambda_Y\left(z;\hat{\sigma}_0^2\right)-\nabla\lambda_Y\left(z;\sigma_0^2\right)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right]\right|\le C\sum_{i=0}^{2}\left\|\partial_x^i\hat{\sigma}^2-\partial_x^i\hat{\sigma}_0^2\right\|_\infty^2,$$

$$\left|\lambda_Y^{(1)}\left(z;\hat{\sigma}^2\right)-\lambda_Y^{(1)}\left(z;\hat{\sigma}_0^2\right)-\nabla\lambda_Y^{(1)}\left(z;\hat{\sigma}_0^2\right)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right]\right|\le C\sum_{i=0}^{3}\left\|\partial_x^i\hat{\sigma}^2-\partial_x^i\hat{\sigma}_0^2\right\|_\infty^2,$$

$$\|\dot{\lambda}_Y\left(Z_t;\hat{\sigma}^2\right)-\dot{\lambda}_Y\left(Z_t;\hat{\sigma}_0^2\right)-\nabla\dot{\lambda}_Y\left(Z_t;\hat{\sigma}_0^2\right)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right]\|\le C\sum_{i=0}^{2}\sum_{j=0}^{1}\|\partial_{x\theta}^{ij}\hat{\sigma}^2-\partial_{x\theta}^{ij}\hat{\sigma}_0^2\|_\infty^2$$

Proof. Define

$$\nabla\lambda_Y\left(\hat{\sigma}_0^2\right)\left[\nabla\sigma^2\right]=D_{\sigma^2}(z)\nabla\sigma^2(z)+D_{\partial_x\sigma^2}(z)\nabla\partial_x\sigma^2(z)+\frac{\nabla\partial_x^2\sigma(z)}{8}.$$

By similar calculations as the ones in the proof of Lemma 35, we obtain the first inequality. The remaining ones follow along the same lines. ■

Define

$$Z_t\left(\sigma^2\right)=\gamma^{-1}\left((1-t)\gamma\left(x_0;\sigma^2\right)+t\gamma\left(x;\sigma^2\right)+\sqrt{\Delta}B_t;\sigma^2\right)$$

where $\gamma\left(x;\sigma^2\right)=\int\sigma(x)^{-1}dx$; we write $\hat{Z}_t=\hat{Z}_t\left(\theta,\hat{\sigma}^2\right)$, $\hat{Z}_{0t}=\hat{Z}_t\left(\theta,\hat{\sigma}_0^2\right)$, and $Z_{0t}(a)=Z_t\left(\hat{\sigma}_0^2(\cdot;\theta,a)\right)$.

Lemma 42 *Under (C2.1)-(C2.4) and (C2.6a), for all $(x,x_0,\theta)\in\mathbb{R}^2\times\Theta$,*

1. $|\hat{Z}_t(\theta)|,|\hat{Z}_{0t}(\theta)|,|Z_{0t}(\theta)|\le C(|x|+|x_0|+|B_t|)$.

2.

$$\begin{aligned}\left|\hat{Z}_t(\theta)-\hat{Z}_{0t}(\theta)\right| &\le C\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty,\\ \left\|\partial_\theta^k Z_t(\theta)-\partial_\theta^k Z_{0t}(\theta)\right\| &\le b(x|x_0)\sum_{i=0}^k\left\|\partial_\theta^i\hat{\sigma}^2-\partial_\theta^i\hat{\sigma}_0^2\right\|_\infty,\end{aligned}$$

where $E_\pi\left[b\left(X_1|X_0\right)\right]<\infty$.

3. *The pathwise derivatives of* $\hat{Z}_{0t}$, *and* $\partial_\theta\hat{Z}_{0t}$ *w.r.t.* σ^2 *exist and satisfy,*

$$\begin{aligned}\left|\nabla\hat{Z}_{0t}\left[d\sigma^2\right]\right| &\le C\int\left|d\sigma^2(w;\theta)\right|dw|_{w=x}+\int\left|d\sigma^2(w;\theta)\right|dw|_{w=x_0},\\ \left\|\nabla\partial_\theta\hat{Z}_{0t}\left[d\sigma^2\right]\right\| &\le C\int_0^{\max\{|x|,|x_0|,|\hat{Z}_{0t}|\}}\sum_{i=0}^1\left\|d\partial_\theta^i\sigma^2(w;\theta)\right\|dw,\end{aligned}$$

$$\left|\hat{Z}_t(\theta_0)-\hat{Z}_{0t}(\theta_0)-\nabla\hat{Z}_{0t}(\theta_0)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right]\right|\le b(x|x_0)\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty^2,$$

$$\left\|\partial_\theta\hat{Z}_t(\theta_0)-\partial_\theta\hat{Z}_{0t}(\theta_0)-\nabla\partial_\theta\hat{Z}_{0t}(\theta_0)\left[\hat{\sigma}^2-\hat{\sigma}_0^2\right]\right\|\le b(x|x_0)\sum_{i=0}^2\left\|\partial_\theta^i\hat{\sigma}^2-\partial_\theta^i\hat{\sigma}_0^2\right\|_\infty^2,$$

where $E_\pi\left[b\left(X_1|X_0\right)\right]<\infty$.

Proof. With $\tilde{B}_t=(1-t)\hat{\gamma}(x_0)+t\hat{\gamma}(x)+\sqrt{\Delta}B_t$,

$$\begin{aligned}\left|\hat{Z}_t-\hat{Z}_{0t}\right| &\le \left|\hat{\gamma}^{-1}(\tilde{B}_t)-\hat{\gamma}^{-1}(\tilde{B}_{0t})\right|+\left|\hat{\gamma}^{-1}(\tilde{B}_{0t})-\hat{\gamma}_0^{-1}(\tilde{B}_{0t})\right|\\ &\le \bar{\sigma}\left|\tilde{B}_t-\tilde{B}_{0t}\right|+\frac{\bar{\sigma}}{\underline{\sigma}^3}\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty,\end{aligned}$$

where

$$\left|\tilde{B}_t-\tilde{B}_{0t}\right|\le\left|\hat{\gamma}(x)-\hat{\gamma}_0(x)\right|+\left|\hat{\gamma}(x_0)-\hat{\gamma}(x_0)\right|\le C\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty.$$

Similarly for $\partial_\theta\hat{Z}_t$ given in (4.35): With $b=\hat{\gamma}^{-1}(y)$ and $b_0=\hat{\gamma}_0^{-1}(y)$,

$$\begin{aligned}\left\|\partial_\theta\left(\hat{\gamma}^{-1}\right)(y)-\partial_\theta\left(\hat{\gamma}_0^{-1}\right)(y)\right\| &\le \hat{\sigma}(b)\left\|\partial_\theta\hat{\gamma}(b)-\partial_\theta\hat{\gamma}_0(b)\right\|\\ &\quad+\hat{\sigma}(b)\left\|\partial_\theta\hat{\gamma}_0(b)-\partial_\theta\hat{\gamma}_0(b_0)\right\|\\ &\quad+\left\|\partial_\theta\hat{\gamma}_0(b_0)\right\|\left|\hat{\sigma}(b)-\hat{\sigma}_0(b)\right|\\ &\quad+\left\|\partial_\theta\hat{\gamma}_0(b_0)\right\|\left|\hat{\sigma}_0(b)-\sigma_0(b_0)\right|\\ &\le C\left(|b_0|\sum_{i=0}^1\left\|\partial_\theta^i\hat{\sigma}^2-\partial_\theta^i\hat{\sigma}_0^2\right\|_\infty+|b-b_0|\right)\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\partial_\theta\left(\hat{\gamma}^{-1}\right)(\tilde{B}_t)-\partial_\theta\left(\hat{\gamma}_0^{-1}\right)(\tilde{B}_{0t})\right\| &\le C\left(|\tilde{B}_{0t}|\sum_{i=0}^1\left\|\partial_\theta^i\hat{\sigma}^2-\partial_\theta^i\hat{\sigma}_0^2\right\|_\infty+\left|\tilde{B}_t-\tilde{B}_{0t}\right|\right)\\ &\le C\left(|x|+|x_0|\right)\sum_{i=0}^1\left\|\partial_\theta^i\hat{\sigma}^2-\partial_\theta^i\hat{\sigma}_0^2\right\|_\infty\end{aligned}$$

while

$$\begin{aligned}
& \left\|\hat{\sigma}\left(\hat{Z}_t\right)\left[(1-t)\,\partial_\theta\hat{\gamma}\,(x_0)+t\partial_\theta\hat{\gamma}\,(x)\right]-\hat{\sigma}_0(\hat{Z}_{0t})[(1-t)\,\partial_\theta\hat{\gamma}_0\,(x_0)+t\partial_\theta\hat{\gamma}_0\,(x)]\right\| \\
\leq\ & \bar{\sigma}\left(\|\partial_\theta\hat{\gamma}\,(x_0)-\partial_\theta\hat{\gamma}_0\,(x_0)\|+\|\partial_\theta\hat{\gamma}\,(x)-\partial_\theta\hat{\gamma}_0\,(x)\|\right) \\
& +\left(\|\partial_\theta\hat{\gamma}_0\,(x_0)\|+\|\partial_\theta\hat{\gamma}_0\,(x)\|\right)\left(|\hat{\sigma}^2(\hat{Z}_t)-\hat{\sigma}^2(\hat{Z}_{0t})|+\left|\hat{\sigma}^2(\hat{Z}_{0t})-\hat{\sigma}_0^2(\hat{Z}_{0t})\right|\right) \\
\leq\ & C\,(|x|+|x_0|)\sum_{i=0}^{1}\left\|\partial_\theta^i\hat{\sigma}^2-\partial_\theta^i\hat{\sigma}_0^2\right\|_\infty .
\end{aligned}$$

The last inequality in 1. is proved along the same lines.

For $\nabla\hat{Z}_{0t}$ given in (4.44), we have

$$\left|\nabla\hat{Z}_{0t}\left[d\sigma^2\right]\right|\leq\frac{\bar{\sigma}}{\underline{\sigma}^2}\left(\int_0^x d\sigma^2\,(w)\,dw+\int_0^{x_0}d\sigma^2\,(w)\,dw+\int_0^{\hat{Z}_{0t}}d\sigma^2\,(w)\,dw\right).$$

Observe

$$\hat{\gamma}\,(x)-\gamma_0\,(x)-\nabla\hat{\gamma}_0\,(x)\,[\hat{\sigma}-\hat{\sigma}_0]=\int_0^x\hat{\sigma}\,(w)^{-1}-\hat{\sigma}_0\,(w)^{-1}+\frac{[\hat{\sigma}-\hat{\sigma}_0]\,(w)}{\hat{\sigma}_0^2\,(w)}dw$$

where the RHS is bounded by

$$\int_0^x\frac{|\hat{\sigma}\,(w)-\hat{\sigma}_0\,(w)|^2}{|\lambda\hat{\sigma}\,(w)-(1-\lambda)\,\hat{\sigma}_0\,(w)|^3}dw\leq\underline{\sigma}^{-3}\left\|\hat{\sigma}\ {}^{-2}\ \hat{\sigma}_0^2\right\|_\infty^2 .$$

By a Taylor expansion of $\hat{\gamma}^{-1}$ with $x=\hat{\gamma}_0^{-1}\,(y)$ and $z=\hat{\gamma}^{-1}\,(\lambda\hat{\gamma}_0\,(x)+(1-\lambda)\,\hat{\gamma}\,(x))$,

$$\begin{aligned}
\left|\hat{\gamma}^{-1}\,(y)-\hat{\gamma}_0^{-1}\,(y)-\nabla\left(\hat{\gamma}_0^{-1}\right)(y)\right| &= \left|\hat{\sigma}^{(1)}\,(z)\right|\hat{\sigma}\,(z)\,|\hat{\gamma}_0\,(x)-\hat{\gamma}\,(x)|^2 \\
&\quad +\hat{\sigma}_0\left(\hat{\gamma}_0^{-1}\,(y)\right)|\hat{\gamma}_0\,(x)-\hat{\gamma}\,(x)-\nabla\hat{\gamma}_0\,(x)| \qquad (4.68)\\
&\leq C\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty^2 .
\end{aligned}$$

In total,

$$\begin{aligned}
|\hat{Z}_t-\hat{Z}_{0t}-\nabla\hat{Z}_{0t}| \leq\ & |\hat{\gamma}^{-1}(\tilde{B}_t)-\hat{\gamma}^{-1}(\tilde{B}_{0t})-\hat{\sigma}(\hat{Z}_{0t})(\tilde{B}_t-\tilde{B}_{0t})| \\
& +|\hat{\gamma}^{-1}(\tilde{B}_{0t})-\hat{\gamma}_0^{-1}(\tilde{B}_{0t})-\nabla\left(\hat{\gamma}_0^{-1}\right)(\tilde{B}_{0t})| \\
& +\hat{\sigma}_0(\hat{Z}_{0t})|\tilde{B}_t-\tilde{B}_{0t}-(1-t)\,\nabla\hat{\gamma}_0\,(x_0)-t\nabla\hat{\gamma}_0\,(x)\,|,
\end{aligned}$$

where

$$\begin{aligned}
\left|\hat{\gamma}^{-1}(\tilde{B}_t)-\hat{\gamma}^{-1}(\tilde{B}_{0t})-\hat{\sigma}(\hat{Z}_{0t})(\tilde{B}_t-\tilde{B}_{0t})\right| &\leq \left|\hat{\sigma}^{(1)}(\hat{Z}_{0t})\hat{\sigma}(\hat{Z}_{0t})\right||\tilde{B}_t-\tilde{B}_{0t}| \\
&\leq Ca^{-1}\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty^2 ,
\end{aligned}$$

and

$$\begin{aligned}
\left|\hat{\gamma}^{-1}(\tilde{B}_{0t})-\hat{\gamma}_0^{-1}(\tilde{B}_{0t})-\nabla\hat{\gamma}_0^{-1}(\tilde{B}_{0t})\right| &\leq C\left(|\hat{Z}_{0t}|+1\right)\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty^2 , \\
|\tilde{B}_t-\tilde{B}_{0t}-(1-t)\,\nabla\hat{\gamma}_0\,(x_0)-t\nabla\hat{\gamma}_0\,(x)\,| &\leq C\,(|x|+|x_0|)\left\|\hat{\sigma}^2-\hat{\sigma}_0^2\right\|_\infty^2 .
\end{aligned}$$

We may obtain the last set of inequalities by similar but more lengthy arguments. ■

5

Estimation of Partial Differential Equations

5.1 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.

In this chapter, 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 chapter 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.

5.2 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

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

This is normally referred to as the infinitesimal generator, cf. 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,

$$-\frac{\partial u}{\partial t} + au = L_t(u) + c, \qquad (5.1)$$

$$u(T,x) = b(x), \qquad (5.2)$$

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 (5.1)-(5.2) and a conditional moment of the process $\{X_t\}$ solving a SDE,

$$\mathrm{SDE}(\mu,\sigma) : dX_t = \mu(t,X_t)\,dt + \sigma(t,X_t)\,dW_t, \quad 0 \le t \le T, \qquad (5.3)$$

with $\{W_t\}$ being a q-dimensional standard Brownian motion.[1] If a solution exists to (5.1)-(5.2), and certain growth conditions on c, b, and u are satisfied, we obtain that

$$\begin{aligned} u(t,x) &= E_{t,x}\left[b(X_T)\exp\left[-\int_t^T a(s,X_s)\,ds\right]\right] \\ &\quad + E_{t,x}\left[\int_t^T c(s,X_s)\exp\left[-\int_t^s a(v,X_v)\,dv\right]ds\right]; \end{aligned} \tag{5.4}$$

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 (5.1)-(5.2), 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 (5.3). 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 SDE(μ_0, σ_0^2). Let u_0 denote the associated solution obtained from (5.4) 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 SDE(μ, σ) as given in (5.3). This yields an estimator of $\{X_t^0\}$ which we denote $\{\hat{X}_t\}$; this is then in turn plugged into (5.4), 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

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

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.

To avoid confusion in the following, we wish to emphasize that we are here working with two probability measures: The first is the probability measure under which we take expectations in (5.4) when calculating $\hat{u}$; the second is the one w.r.t. which the estimators are measurable. The former measure can be chosen by the researcher, and we shall here choose it to be independent of the latter. In effect, we are working with a product measure. So eventhough $\hat{u}$ is calculated as a conditional expectations under this first measure, it is still random since the functions $\hat{\mu}, \hat{\sigma}^2$ which are plugged into these are independent of the first measure.

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 (5.1)-(5.2). 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

[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.

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 (5.3) with initial condition $X_t = x$. We then approximate $u(t, x)$ by

$$\begin{aligned} u^{(N)}(t,x) &= \frac{1}{N}\sum_{i=1}^{N}\left[b\left(X_T^{(i)}\right)\exp\left[-\int_t^T a\left(v, X_v^{(i)}\right)dv\right]\right] \\ &+\frac{1}{N}\sum_{i=1}^{N}\left[\int_t^T c\left(s, X_s^{(i)}\right)\exp\left[-\int_t^s a\left(v, X_v^{(i)}\right)dv\right]ds\right]. \end{aligned} \tag{5.5}$$

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 (5.3) 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 (5.1)-(5.2) 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
   $$\|\partial_x^\alpha \mu(t,x)\| \leq K(1+\|x\|), \quad \left\|\partial_x^\alpha \sigma^2(t,x)\right\| \leq K(1+\|x\|),$$
   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
   $$\|\mu(t,x) - \mu(t,y)\| \leq K_N \|x-y\|, \quad \left\|\sigma^2(t,x) - \sigma^2(t,y)\right\| \leq K_N \|x-y\|,$$
   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 $(\mu, \sigma^2) \in \mathcal{D}$ 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 (5.3), cf. Friedman (1975, Theorem 5.2.2). Furthermore, if $E\left[\|X^*\|^{2p}\right] < \infty$, for some $p \geq 1$,

$$E\left[\|X_t\|^{2p}\right] \leq \left(1 + E\left[\|X^*\|^{2p}\right]\right) e^{C^* t} \tag{5.6}$$

for $0 \leq t \leq T$, where $C^* = C^*(K, p, T)$, cf. 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 $\sigma^2(\cdot) > 0$, cf. Karatzas and Shreve (1991, Theorem 5.5.15 and Corollary 5.3.23). The bound in (5.6) 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 (μ, σ^2) 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 $(\mu_0, \sigma_0^2) \in \mathcal{D}$, and denote the associated diffusion process by $\{X_t^0\}$. We also fix the initial condition of $\{X_t^0\}$ at some given random variable, X^*. First we define $L_p(X^*, [0,T] \times \mathbb{R}^q)$ 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 $W^{m,p} = W^{m,p}(X^*, [0,T] \times \mathbb{R}^q)$ 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 $\partial_x^\alpha f \in L_p(X^*, [0,T] \times \mathbb{R}^q)$ 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

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

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

$$\langle f, g \rangle_m = \sum_{|\alpha| \leq m} E\left[\int_0^T \partial_x^\alpha f\left(t, X_t^0\right) \partial_x^\alpha g\left(t, X_t^0\right) dt\right] \tag{5.7}$$

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

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

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

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

5.2.1 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,

$$dS_t = \mu_S(t, S_t)\, dt + \sigma_S(t, S_t)\, dw_t^S, \tag{5.8}$$

where $\{W_t^S\}$ 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

$$S_t = E_t^Q\left[\exp\left[-\int_t^T r_s ds\right] S_T + \int_t^T \exp\left[-\int_t^s r_v dv\right] d_s ds\right], \tag{5.9}$$

where $\{S_t\}$ has dynamics

$$dS_t = r_t S_t dt + \sigma_S(t, S_t)\, dW_t^S \tag{5.10}$$

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

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

that $\{r_t\}$ and $\{d_t\}$ also solve SDE's under Q,

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

Then by defining

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

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

we observe that the pricing formula (5.9) takes the form of (5.4). 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 (5.9), 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 (5.10) 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

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

consider some stock price process $\{S_t\}$ but we now assume that

$$dS_t = \mu_S(S_t)\,dt + \sigma_S(S_t, v_t)\,dw_t^S, \tag{5.11}$$

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

$$dv_t = \mu_v(v_t)\,dt + \sigma_v(v_t)\,dw_t^v \tag{5.12}$$

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, (5.9) is still valid but now

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

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

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 (5.11)-(5.12) is found in Heston (1993) where

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dw_t^S, \tag{5.13}$$

$$dv_t = \beta(\alpha - v_t)\,dt + \sigma\sqrt{v_t}\,dw_t^v, \tag{5.14}$$

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 (5.3) with $\mu(x) = (rS, \mu_v(v) - \lambda(S, v))^\top$, $\sigma(x) = \operatorname{diag}(\sigma_S(S, v), \sigma_v(v))$, 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

$$dr_t = \mu(t, r_t)\,dt + \sigma(t, r_t)\,dw_t \tag{5.15}$$

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

$$dr_t = \{\mu(t, r_t) - \lambda(t, r_t)\,\sigma(t, r_t)\}\,dt + \sigma(t, r_t)\,dW_t \tag{5.16}$$

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

$$\Pi_t(T) = E_t^Q\left[\exp\left[-\int_t^T r_s ds\right] g(r_T) + \int_t^T \exp\left[-\int_t^s r_s du\right] d(s, r_s)\,ds\right]. \tag{5.17}$$

[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.

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 (5.17) 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 (5.4): 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 (5.17) 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

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

cf. 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 (5.4).

5.2.2 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,

$$dX_t = \mu(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dW_t, \tag{5.18}$$

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

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

where $B(x;\theta) = E_\theta[b(X_\Delta)\,|X_0 = x]$ is the conditional mean where we write $E_\theta[\cdot]$ 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 $B(x;\theta)$ is smooth and differentiable in θ. However, as noted earlier, an analytical expression of $B(x;\theta)$ often cannot be derived and is calculated using either simulations or approximate methods. One easily realise that $B(x;\theta) = u(0,x;\theta)$ where u solves the LPDE

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

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

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

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 $b_1(x) = x$ and $b_2(x) = x^2$.

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,

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

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 $\theta \mapsto \Pi(x; \theta)$ is continuous and differentiable. In the next section, we give regularity conditions which ensures this.

5.3 Estimation of Partial Differential Equations

In this section, we shall assume that preliminary estimators $(\hat{\mu}, \hat{\sigma}^2)$ 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

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

where u is the solution to (5.1)-(5.2) with (μ, σ^2) plugged in. We assume that we have obtained estimators, $(\hat{\mu}, \hat{\sigma}^2)$, of the true drift and diffusion term, (μ_0, σ_0^2). Given the definition of Γ, the true but unknown solution to the PDE is given by

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

which we then estimate by

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

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 $(\hat{\mu}, \hat{\sigma}^2)$ given that Γ is sufficiently smooth. Roughly speaking, $\hat{u}$ will be consistent if $(\hat{\mu}, \hat{\sigma}^2)$ is so and Γ is continuous, while the asymptotic distribution will be induced by the one of $(\hat{\mu}, \hat{\sigma}^2)$ given that Γ is (pathwise) differentiable w.r.t. (μ, σ^2). 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 $(\hat{\mu}, \hat{\sigma}^2) \in \mathcal{D}$ with $\|(\hat{\mu}, \hat{\sigma}^2) - (\mu_0, \sigma_0^2)\|_{\mathcal{D}} \to^P 0$. Consistency of $\hat{u} = \Gamma(\hat{\mu}, \hat{\sigma}^2)$ will now follow by continuity of Γ since this implies $\|\hat{u} - u_0\|_{\mathcal{U}} = \|\Gamma(\hat{\mu}, \hat{\sigma}^2) - \Gamma(\mu_0, \sigma_0^2)\|_{\mathcal{U}} \to^P 0$. Assume that the pathwise derivative of Γ w.r.t. μ and σ^2 at (μ_0, σ_0^2) exists. We denote these $\nabla_1\Gamma[d\mu]$ and $\nabla_2\Gamma[d\sigma^2]$ respectively and define $\nabla\Gamma[d\mu, d\sigma^2] = \nabla_1\Gamma[d\mu] + \nabla_2\Gamma[d\sigma^2]$. Assuming that

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

$\nabla\Gamma$ 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 $(\mu_0, \sigma_0^2) \in \mathcal{D}$

We first show that the functional $\Gamma : \mathcal{D} \mapsto \mathcal{U}$ is continuous. This is stated in the following theorem:

Theorem 44 *For any* $(\mu, \sigma^2) \in \mathcal{D}$,

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

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

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

for $r \leq p^*$.

This basically shows that consistency of the preliminary estimators $\hat{\mu}$ and $\hat{\sigma}^2$ implies consistency of $\hat{u}$ as a continuous functional of these.

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 (5.53) and (5.54) respectively with $\nabla_i X_0 = 0$, $i = 1, 2$. From Lemma 54, $\{\nabla_1 X_t\}$ 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

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

Given the pathwise derivative of $\{X_t\}$, we are able to introduce the pathwise derivative of Γ. Making use of the chain rule, it should be clear that the pathwise derivative of Γ at

(μ_0, σ_0^2) in the direction $(d\mu, d\sigma^2)$ is given by

$$\begin{aligned}
& \nabla\Gamma\left[d\mu, d\sigma^2\right](t,x) \qquad (5.19)\\
= \; & E_{t,x}\left[b_x\left(X_T^0\right)\nabla X_T \exp\left[-\int_t^T a\left(s, X_s^0\right)ds\right]\right] \\
& -E_{t,x}\left[b\left(X_T^0\right)\int_t^T a_x\left(s, X_s^0\right)\nabla X_s ds \exp\left[-\int_t^T a\left(s, X_s^0\right)ds\right]\right] \\
& +E_{t,x}\left[\int_t^T c_x\left(s, X_s^0\right)\nabla X_s \exp\left[-\int_t^s a\left(v, X_v^0\right)dv\right]ds\right] \\
& -E_{t,x}\left[\int_t^T c\left(s, X_s^0\right)\exp\left[-\int_t^s a\left(v, X_v^0\right)dv\right]\left(\int_t^s a_x\left(v, X_v^0\right)\nabla X_v dv\right)ds\right].
\end{aligned}$$

This is formally shown in the Appendix. 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,

$$-\frac{\partial v}{\partial t} + av = L_t(v) + \bar{c}\left[d\mu, d\sigma^2\right] \qquad (5.20)$$

$$v(T,x) = 0 \qquad (5.21)$$

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 $u_0 = \Gamma(\mu_0, \sigma_0^2)$. The generalised solution of (5.20)-(5.21) is given as

$$\nabla\Gamma\left[d\mu, d\sigma^2\right](t,x) = E_{t,x}\left[\int_t^T \bar{c}\left[d\mu, d\sigma^2\right]\left(s, X_s^0\right)\exp\left[-\int_t^s a\left(u, X_u^0\right)du\right]ds\right] \qquad (5.22)$$

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

Theorem 45 *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

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

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 44 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 46 (Master Theorem) *Assume that* $(\hat{\mu}, \hat{\sigma}^2) \in \mathcal{D}$*, and* $\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1)$ *and* $\|\hat{\sigma}^2 - \sigma_0^2\|_{0,4} = o_P(1)$*. Then* $\|\hat{u}(t,x) - u_0(t,x)\|_{0,1} \to^P 0$ *for any* X^* *with* $E[\|X^*\|^r] < \infty$.

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

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

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 $\|\cdot\|_{0,4}$. 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, cf. Theorem 47. 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 (5.7). 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 $\bar{\mathcal{D}}$, cf. Theorem 45. We then apply Riesz Representation Theorem on $\nabla\Gamma$: There exists $d^* = (d_1^*, d_2^*) \in \bar{\mathcal{D}}$ such that

$$\nabla\Gamma\left[\mu, \sigma^2\right](t,x) = \langle\mu, d_1^*\rangle + \langle\sigma^2, d_2^*\rangle, \tag{5.23}$$

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

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

and $p(s,y|t,x)$ denotes the conditional density of X_s^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 (5.22) that

$$\begin{aligned} \nabla\Gamma\left[d\mu, d\sigma^2\right](t,x) &= e^{at}\sum_{i=1}^{q} E_{t,x}\left[\int_t^T d\mu_i\left(s, X_s^0\right) d_{1,i}^*\left(s, X_s^0\right) ds\right] \\ &+\frac{1}{2}e^{at}\sum_{i,j=1}^{q} E_{t,x}\left[\int_t^T d\sigma_{ij}^2\left(s, X_s^0\right) d_{2,ij}^*\left(s, X_s^0\right) ds\right] \end{aligned} \tag{5.24}$$

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, cf. 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, cf. Aït-Sahalia (1996a) and Jiang (1998). The result stated in Theorem 46 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

$$dx_t = \mu_0(t,x_t)\,dt + \sigma_0(t,x_t)\,dw_t. \tag{5.25}$$

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.

5.3.1 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 $V_n^{-1/2}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,I)$ 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.

The conditions are fairly weak, and are satisfied by a range of parametric diffusion models. The boundedness assumption in (P.3) is assumed for convenience and can be weakened to some polynomial bound. The conditions in (P.4) are satisfied for most well-behaved estimators. In particular, the MLE in both the stationary and nonstationary case satisfies (P.4) under weak conditions as can be found in Aït-Sahalia (2003).

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

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

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

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

for $i = 0, 1$. Thus, by Theorem 46, 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

$$\begin{aligned}\dot{\Gamma}_0(t,x) &= E_{t,x}\left[b_x\left(X_T^0\right)\dot{X}_T^0\exp\left[-\int_t^T a\left(s,X_s^0\right)ds\right]\right] \\ &\quad -E_{t,x}\left[b\left(X_T^0\right)\exp\left[-\int_t^T a\left(s,X_s^0\right)ds\right]\left(\int_t^T a_x\left(v,X_v^0\right)\dot{X}_v^0dv\right)\right] \\ &\quad +E_{t,x}\left[\int_t^T c_x\left(s,X_s^0\right)\dot{X}_s^0\exp\left[-\int_t^s a\left(v,X_v^0\right)dv\right]ds\right] \\ &\quad -E_{t,x}\left[\int_t^T c\left(s,X_s^0\right)\exp\left[-\int_t^s a\left(v,X_v^0\right)dv\right]\left(\int_t^s a_x\left(v,X_u^0\right)\dot{X}_v^0dv\right)ds\right].\end{aligned} \tag{5.26}$$

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 54, that

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

implying

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

We have now proved the following theorem:

Theorem 47 (Parametric Estimator) *Under (P.1)-(P.4), the parametric estimator $\hat{u}$ is consistent and satisfies*

$$\sqrt{\dot{\Gamma}_0(t,x)^\top V_n\dot{\Gamma}_0(t,x)}\,(\hat{u}(t,x)-u_0(t,x))\xrightarrow{d}N(0,1),$$

where $\dot{\Gamma}_0(t,x)$ is given in (5.26).

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 (5.20)-(5.21) with $\bar{c}$ given by

$$\bar{c}=\sum_{i=1}^q\dot{\mu}_i\frac{\partial u_0}{\partial x_i}+\frac{1}{2}\sum_{i=1}^q\dot{\sigma}_{ij}^2\frac{\partial^2u_0}{\partial x_i\partial x_j}. \tag{5.27}$$

It readily follows from Lemma 51, that a consistent estimator of $\dot{\Gamma}_0(t,x)$ is obtained by substituting $(X^0,\dot{X}^0)$ by $(\hat{X},\widehat{\dot{X}})$ in (5.26), 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 $\hat{\dot{X}}_0 = 0$. Alternatively, one can obtain an estimator of $\dot{\Gamma}_0$ by solving (5.20)-(5.21) with $\bar{c}$ given in (5.27)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.

5.3.2 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

$$\hat{\theta} = \theta_0 + \frac{1}{n}\sum_{i=1}^{n} \psi(x_i; x_{i-1}) + o_P(n^{-1/2}), \tag{5.28}$$

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 Chapter 4. Assuming stationarity of $\{x_t\}$, there exists a stationary density π_0 satisfying

$$\pi_0(x) = \frac{M}{\sigma^2(x)} \exp\left[2\int_{x^*}^{x} \frac{\mu_0(y)}{\sigma_0^2(y)}dy\right], \tag{5.29}$$

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 (5.29) in either of the two following ways,

$$\mu_0(x) = \frac{1}{2\pi_0(x)}\frac{\partial}{\partial x}\left[\sigma_0^2(x)\,\pi_0(x)\right], \tag{5.30}$$

$$\sigma_0^2(x) = \frac{2}{\pi_0(x)}\int_l^x \mu_0(y)\,\pi_0(y)\,dy. \tag{5.31}$$

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

$$\hat{\pi}^{(m)}(x) = \frac{1}{nh^{m+1}}\sum_{i=1}^{n} K^{(m)}\left(\frac{x - x_i}{h}\right), \quad r \geq 0, \tag{5.32}$$

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

$$\hat{\mu}(x;\theta) = \frac{1}{2}\partial_x\sigma^2(x;\theta) + \frac{1}{2}\sigma^2(x;\theta)\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)}, \tag{5.33}$$

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

$$\hat{\sigma}^2(x;\theta) = \frac{1}{2\hat{\pi}(x)}\frac{1}{n}\sum_{i=1}^{n} 1_{(l,x)}(x_i)\,\mu(x;\theta). \tag{5.34}$$

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 (5.33) 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

$$T(x;\pi,a) = \begin{cases} 1, & \pi(x) \geq a \\ 0, & \pi(x) \leq a/2 \end{cases}, \tag{5.35}$$

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 (5.35), (ii) is $\omega \geq 0$ 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

$$\hat{\mu}(x) = \left\{\frac{1}{2}\partial_x\sigma^2\left(x;\hat{\theta}\right) + \frac{1}{2}\sigma^2\left(x;\hat{\theta}\right)\frac{\hat{\pi}^{(1)}(x)}{\hat{\pi}(x)}\right\}\hat{T}(x;a). \tag{5.36}$$

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

$$\hat{\sigma}^2(x) = \frac{n^{-1}\sum_{i=1}^{n} 1_{(l,x)}(x_i)\,\hat{\mu}(x_i)}{2\hat{\pi}(x)}\hat{T}(x;a). \tag{5.37}$$

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) where $\omega, \lambda \geq 1$ are integers:

$\underline{\mathcal{K}(\omega,\lambda)}$ 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 $|x| \to \infty$, $0 \leq i \leq \lambda - 1$;

$\sup_{x\in\mathbb{R}}\left|K^{(i)}(x)\right|\max(|x|,1) < \infty$, $0 \leq i \leq \lambda + 1$; $K^{(i)}$ 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 (5.28) with $\theta_0 \in \text{int}\Theta$.

SP.4 The transition density p_t exists for any $t \geq 0$ such that the mapping $y \mapsto p_t\left(y|x\right)$ 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 48 (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

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

and $\{\nu\left(x_i|x_{i-1};t,x\right)\}$ *is given in (5.45).*

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 49 (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

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

and $\{\nu(x_i|x_{i-1};t,x)\}$ is given in (5.48).

Sufficient conditions for (SP.0) to hold can be found in Chapter 4. (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; see also Section 4.4. 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

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

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.

5.3.3 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 $\sqrt{\bar{T}}$ 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,

$$\begin{aligned} \hat{\pi}(x) &= n^{-1}\sum_{i=1}^{n} K_h(x_i - x), \\ \hat{m}(x) &= n^{-1}\sum_{i=1}^{n} K_h(x_i - x)\frac{x_{i+1} - x_i}{\Delta}, \\ \hat{s}(x) &= n^{-1}\sum_{i=1}^{n} K_h(x_i - x)\frac{(x_{i+1} - x_i)^2}{\Delta}. \end{aligned}$$

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

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 $\bar{T}^{1/2+\delta}\sqrt{\log_2(\bar{T})}\Delta^{3/4}\log\left(\Delta^{-1}\right)^{1/4}a^{-2}h^{-2} \to 0$, as $a, h \to 0$.

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

1. $\bar{T}^{3/4+\delta}\sqrt{\log_2(\bar{T})}\Delta^{3/4}\log\left(\Delta^{-1}\right)^{1/4}a^{k-2}h^{-2-k} \to 0$, $\bar{T}^{-1/4}a^{k-2}h^{-1-k} \to 0$ 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 $\|\hat{\sigma}^2 - \sigma_0^2\|_{1,4} = o_P\left(\bar{T}^{-1/4}\right)$. 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 $\sqrt{\bar{T}}$, 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 50 (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*

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

where

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

with d_1^* given in (5.23).

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

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

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 $\sqrt{Th^3}$. Thus, the convergence of $\hat{u}$ should not be able to exceed $\sqrt{T}$.

5.4 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 (5.8),

$$dS_t = \mu_S (t, S_t; \theta)\, dt + \sigma_S (t, S_t; \theta)\, dw_t^S, \tag{5.38}$$

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 47 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 48 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

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

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

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

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 48, while the first one requires a bit of work: Observe that

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

such that

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

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, $(\mu(\cdot;\theta), \sigma^2(\cdot;\theta))$, in (5.18) 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 44 and 45 that $B(x;\theta)$ is continuously differentiable in θ, and by (5.6) $|B(x;\theta)| \leq C(K,\Delta)(1+|x|)$, $\left|\dot{B}(x;\theta)\right| \leq C(K,\Delta)(1+|x|^2)$. The first condition implies, cf. 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

$$\sqrt{n}(\hat{\theta}-\theta_0) \to^d N\left(0, H^{-1}(\theta_0)\, V(\theta_0)\, H^{-1}(\theta_0)\right) \tag{5.39}$$

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,

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

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 $E[\varepsilon_i] = 0$, $\sigma_\varepsilon^2 = E\left[\varepsilon_i^2\right] < \infty$.

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

$$\sqrt{n}(\hat{\theta}-\theta_0) \to^d N\left(0, \sigma_\varepsilon^2 H^{-1}(\theta_0)\right). \tag{5.40}$$

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

5.5 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 our results prove useful.

5.A Proofs

Proof of Theorem 43. 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 (5.6). 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 53, such that $|\partial u(t,x)/\partial x_i|$ is bounded by

$$\begin{aligned}&E_{t,x}\left[\|b_x(X_T)\|\,\|Y_T^{(i)}\|+|b(X_T)|\int_t^T\|a_x(s,X_s)\|\,\|Y_s^{(i)}\|ds\right]\\ &+E_{t,x}\left[\int_t^T|c_x(s,X_s)|\,\|Y_s^{(i)}\|+\int_t^s\|a_x(u,X_u)\|\,\|Y_u^{(i)}\|duds\right]\\ \leq\;& E_{t,x}\left[C\left(1+\|X_T\|^r\right)\|Y_T^{(i)}\|+C\left(1+\|X_T\|^r\right)\int_t^T C\left(1+\|X_u\|^r\right)\|Y_u^{(i)}\|du\right]\\ &+E_{t,x}\left[\int_t^T C\left(1+\|X_s\|^r\right)\|Y_s^{(i)}\|+C\left(1+\|X_s\|^r\right)\int_t^s C\left(1+\|X_u\|^r\right)\|Y_u^{(i)}\|duds\right]\\ \leq\;& C\left(1+\left(\int_t^T E_{t,x}\left[\|X_s\|^{2r}\right]ds\right)^{1/2}\right)\left(1+\left(\int_t^T E_{t,x}\left[\|Y_s^{(i)}\|^2\right]ds\right)^{1/2}\right)\\ \leq\;& C\left(1+\|x\|^q\right),\end{aligned}$$

for any $(t,x)\in[0,T]\times\mathbb{R}^q$, where we have used (5.6). 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 $pr\leq p^*$. ■

Proof of Theorem 44. We see that

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

where, by Lemma 52,

$$
\begin{aligned}
E_{t,x}\left[\left\|a\left(u,X_u\right)-a\left(u,X_u\right)\right\|^2\right] \ \leq\ & C\left(u,x\right)\left(\int_t^u E_{t,x}\left[\left\|\mu\left(v,X_v^0\right)-\mu_0\left(v,X_v^0\right)\right\|^4\right]dv\right. \\
&\left.+\int_t^u E_{t,x}\left[\left\|\sigma^2\left(v,X_v^0\right)-\sigma_0^2\left(v,X_v^0\right)\right\|^4\right]dv\right)^{1/2},
\end{aligned}
$$

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

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

Similarly,

$$
\begin{aligned}
&\left|E_{t,x}\left[\int_t^T c\left(s,X_s\right)\exp\left[-\int_t^s a\left(u,X_u\right)du\right]ds\right]\right. \\
&\left.-E_{t,x}\left[\int_t^T c\left(s,X_s^0\right)\exp\left[-\int_t^s a\left(u,X_u^0\right)du\right]ds\right]\right| \\
\leq\ & C\left(T,x\right)\left(\int_t^T E_{t,x}\left[\left\|\mu\left(u,X_u^0\right)-\mu_0\left(u,X_u^0\right)\right\|^4\right]du\right. \\
&\left.+\int_t^T E_{t,x}\left[\left\|\sigma^2\left(u,X_u^0\right)-\sigma_0^2\left(u,X_u^0\right)\right\|^4\right]du\right)^{1/2}
\end{aligned}
$$

We conclude that

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

Taking expectations,

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

■

Proof of Theorem 45. For any function f , we have

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

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

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

Also,

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

and

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

Similarly,

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

and

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

Combining these results we obtain

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

An application of Lemma 51 and 54 then proves the claim.

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

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

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

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

and

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

where

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

for all $s \in [t,T]$, by Lemma 54. 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 48. Under the (SP.0)-(SP.3) and (SP1.A), it follows from Lemma 55 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 46. To prove asymptotic normality, we first observe from Lemma 55, that $\|\hat{\mu}-\mu_0\|_{1,4}^2 = o_P(n^{-1/2})$, and $\left\|\hat{\sigma}^2-\sigma_0^2\right\|_{1,4}^2 = o_P(n^{-1/2})$. Thus, the asymptotic distribution of $\hat{u}$ is determined by $\nabla\Gamma$ as given in (5.19). 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 $\nabla_\pi\mu_0\left[d\pi,d\pi^{(1)}\right] = \nabla_0\mu_0[d\pi] + \nabla_1\mu_0\left[d\pi^{(1)}\right]$, and

$$\begin{aligned} \nabla_0\mu_0[d\pi] &= -\frac{1}{2}\sigma_0^2(x)\frac{\pi_0^{(1)}(x)}{\pi_0^2(x)}d\pi(x), \quad \nabla_1\mu_0\left[d\pi^{(1)}\right] = \frac{1}{2}\sigma_0^2(x)\frac{d\pi^{(1)}(x)}{\pi_0(x)}, \\ \nabla_\theta\mu_0[d\theta] &= \dot{\mu}_0(x)d\theta, \quad \dot{\mu}_0(x) = \frac{1}{2}\partial_x\dot{\sigma}_0^2(x) + \frac{1}{2}\sigma_0^2(x)\frac{\pi_0^{(1)}(x)}{\pi_0(x)} \end{aligned}$$

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 56 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,

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

where, by the continuity of $\nabla\Gamma$, c.f. Theorem 45,

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

By Riesz's Representation Theorem,

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

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,

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

where

$$\nu_1(y;t,x)=-\frac{1}{2}d_1^*(y)\,\sigma_0^2(y)\frac{\pi_0^{(1)}(y)}{\pi_0^2(y)}\int_t^T p_{s-t}(y|x)\,ds. \tag{5.41}$$

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

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

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 $\hat{T}^{(i)}(x; a) = 0$, $1 \leq i \leq \omega - 1$, 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,

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

where

$$\nu_2(y; t, x) = -\frac{1}{2} \frac{\partial}{\partial y} \left[\frac{d_1(y)\, \sigma_0^2(y)}{\pi_0(y)} \left\{ \int_t^T p_s(y|x)\, ds \right\} \right]. \tag{5.42}$$

The last two terms are easily dealt with since

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

where

$$\nu_3(x_i | x_{i-1}; t, x) = \left\{ \int_t^T E_{t,x}\left[d_1\left(X_s^0\right) \dot{\mu}_0\left(X_s^0\right)\right] ds \right\} \psi(x_i | x_{i-1}), \tag{5.43}$$

$$\nu_4(x_i | x_{i-1}; t, x) = \left\{ \int_t^T E_{t,x}\left[d_2\left(X_s^0\right) \dot{\sigma}_0\left(X_s^0\right)\right] ds \right\} \psi(x_i | x_{i-1}). \tag{5.44}$$

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

$$\nu(x_i | x_{i-1}; t, x) = \sum_{k=1}^{2} \nu_k(x_i; t, x) + \sum_{k=3}^{4} \nu_k(x_i | x_{i-1}; t, x). \tag{5.45}$$

■

Proof of Theorem 49. Under (SP2.A), $\|\hat{\mu} - \mu_0\|_{0,4} = o_P(1)$, and $\left\|\hat{\sigma}^2 - \sigma_0^2\right\|_{0,4} = o_P(1)$ by Lemma 55 which gives us consistency. From Lemma 55, 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

$$\begin{aligned} \nabla_\pi \sigma_0^2[d\pi] &= -\frac{d\pi(x)}{2\pi_0^2(x)} \int_l^x \mu_0(y)\, \pi_0(y)\, dy, \\ \nabla \sigma_0^2[d\theta] &= \dot{\sigma}_0^2(x)\, d\theta, \quad \dot{\sigma}_0^2(x) = \frac{1}{2\pi_0(x)} \int_l^x \dot{\mu}_0(y)\, d\pi(y)\, dy, \end{aligned}$$

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 56,

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

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

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

where d^* is given by the Riesz Representation Theorem. Proceeding as in the proof of Theorem 48, we obtain

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

where

$$\nu_2\left(y; t, x\right) = d_2^*\left(y\right)\frac{1}{2\pi_0^2\left(y\right)}\int_l^y \mu_0\left(y\right)\pi_0\left(y\right)dy\int_t^T p_s\left(y|x\right)ds. \tag{5.46}$$

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

$$\nu_2\left(y|z; t, x\right) = \left\{\int_t^T E_{t,x}\left[d_1^*\left(X_s^0\right)\dot{\mu}\left(X_s^0\right)\right] ds + \int_t^T E_{t,x}\left[d_2^*\left(X_s^0\right)\dot{\sigma}\left(X_s^0\right)\right] ds\right\}\psi\left(z; y\right) \tag{5.47}$$

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

$$\nu\left(x_i|x_{i-1}; t, x\right) = \nu_1\left(x_i; t, x\right) + \nu_2\left(x_i|x_{i-1}; t, x\right) \tag{5.48}$$

■

Proof of Theorem 50. 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 57. This proves consistency of $\hat{u}$.

To derive the asymptotic distribution of $\hat{u}$, we proceed as in the previous two proofs. Applying Lemma 58, $\|\hat{\mu} - \mu_0\|_{1,4}^2 = o_P\left(\bar{T}^{-1/2}\right)$, 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 $\nabla\Gamma\left[d\mu, d\sigma^2\right]$ as given in (5.19). We define $\nabla\mu_0\left(x\right)\left[dm, d\pi\right] = \hat{T}\left\{\nabla_m\mu_0\left(x\right)\left[dm\right] + \nabla_\pi\mu_0\left(x\right)\left[d\pi\right]\right\}$ and $\nabla\sigma_0^2\left(x\right)\left[ds, d\pi\right] = \hat{T}\left\{\nabla_s\sigma_0^2\left(x\right)\left[d\tau\right] + \nabla_\pi\sigma_0^2\left(x\right)\left[d\pi\right]\right\}$ where

$$\nabla_m\mu_0\left(x\right)\left[dm\right] = \frac{1}{\pi_0\left(x\right)}dm\left(x\right), \quad \nabla_\pi\mu_0\left(x\right)\left[d\pi\right] = -\frac{m_0\left(x\right)}{\pi_0^2\left(x\right)}d\pi\left(x\right) \tag{5.49}$$

$$\nabla_s\sigma_0^2\left(x\right)\left[d\tau\right] = \frac{1}{\pi_0\left(x\right)}ds\left(x\right), \quad \nabla_\pi\sigma_0^2\left(x\right)\left[d\pi\right] = -\frac{s_0\left(x\right)}{\pi_0^2\left(x\right)}d\pi\left(x\right). \tag{5.50}$$

Applying Lemma 58 once more, it easily follows that

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

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

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

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

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

$$\begin{aligned}&\int_t^T E_{t,x}\left[\hat{T}\left(X_u^0;a\right)d_1^*\left(X_u^0\right)\nabla_\pi\hat{\mu}_0\left(X_u^0\right)\left[\hat{\pi}\right]\right]du\\ =& -\int_t^T E_{t,x}\left[\hat{T}\left(X_u^0;a\right)d_1^*\left(X_u^0\right)\frac{m_0\left(X_u^0\right)}{\pi_0^2\left(X_u^0\right)}\hat{\pi}\left(X_u^0\right)\right]du\\ =& -n^{-1}\sum_{i=1}^n\int_t^T\int\hat{T}\left(y;a\right)d_1^*\left(y\right)\frac{m_0\left(y\right)}{\pi_0^2\left(y\right)}K_m\left(x_i-y\right)p_u\left(y|x\right)dydu\\ =& -n^{-1}\sum_{i=1}^n\hat{T}\left(x_i;a\right)d_1^*\left(x_i\right)\frac{m_0\left(x_i\right)}{\pi_0^2\left(x_i\right)}\int_t^T p_u\left(x_i|x\right)du+o_P(\bar{T}^{-1/2})\\ =& -n^{-1}\sum_{i=1}^n\hat{T}\left(x_i;a\right)d_1^*\left(x_i\right)\frac{\mu_0\left(x_i\right)}{\pi_0\left(x_i\right)}\int_t^T p_u\left(x_i|x\right)du+o_P(\bar{T}^{-1/2});\end{aligned}$$

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

$$\begin{aligned}
&\int_t^T E_{t,x}\left[\hat{T}\left(X_u^0;a\right)d_2^*\left(X_u^0\right)\nabla\hat{\sigma}_0^2\left(X_u^0\right)[\hat{\pi}]\right]du \\
&= -\int_t^T E_{t,x}\left[\hat{T}\left(X_u^0;a\right)d_2^*\left(X_u^0\right)\frac{s_0\left(X_u^0\right)}{\pi_0^2\left(X_u^0\right)}\hat{\pi}\left(X_u^0\right)\right]du \\
&= -n^{-1}\sum_{i=1}^n \hat{T}\left(x_i;a\right)\frac{d_2^*\left(x_i\right)\sigma_0^2\left(x_i\right)}{\pi_0\left(x_i\right)}\int_t^T p_u\left(x_i|x\right)du + o_P\left(n^{-1/2}\right);
\end{aligned}$$

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

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

where

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

Using same arguments as employed in the proof of Lemma 57,

$$\begin{aligned}
\frac{\sqrt{\Delta}}{\sqrt{n}}\sum_{i=1}^n \bar{d}_1^*\left(x_i\right)\left\{\frac{x_{i+1}-x_i}{\Delta}-\mu\left(x_i\right)\right\} &= \frac{1}{\sqrt{\bar{T}}}\int_0^{\bar{T}}\bar{d}_1^*\left(x_s\right)\sigma\left(x_s\right)dW_s + o_P\left(1\right), \\
\frac{\sqrt{\Delta}}{\sqrt{n}}\sum_{i=1}^n \bar{d}_2^*\left(x_i\right)\left\{\frac{\left(x_{i+1}-x_i\right)^2}{\Delta}-\sigma_0^2\left(x_i\right)\right\} &= o_P\left(1\right).
\end{aligned}$$

Thus,

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

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

5.B Auxiliary Lemmas

Lemma 51 *For any* $\left(\mu,\sigma^2\right)\in\mathcal{D}$,

$$\begin{aligned}
E_s\left[\left\|X_t-X_t^0\right\|^2\right] &\leq \left\|X_s-X_s^0\right\|^2 + 6\int_s^T E_s\left[\left\|\mu\left(X_u^0\right)-\mu_0\left(X_u^0\right)\right\|^2\right]du \\
&\quad + 6\int_s^T E_s\left[\left\|\sigma^2\left(u,X_u^0\right)-\sigma_0^2\left(u,X_u^0\right)\right\|^2\right]du
\end{aligned}$$

for $s\leq t\leq T$.

Proof. Since

$$\begin{aligned} X_t &= X_s + \int_s^t \mu(s, X_s)\, ds + \int_s^t \sigma(s, X_s)\, dW_s, \\ X_t^0 &= X_s^0 + \int_s^t \mu_0\left(s, X_s^0\right) ds + \int_s^t \sigma_0\left(s, X_s^0\right) dW_s \end{aligned}$$

We obtain

$$\begin{aligned} X_t - X_t^0 &= \eta_t + X_s - X_s^0 + \int_s^t \left[\mu(u, X_u) - \mu\left(u, X_u^0\right)\right] du \\ &\quad + \int_s^t \left[\sigma(u, X_u) - \sigma\left(u, X_u^0\right)\right] dW_u. \end{aligned} \tag{5.51}$$

where

$$\eta_t = \int_s^t \left[\mu\left(u, X_u^0\right) - \mu_0\left(s, X_s^0\right)\right] ds + \int_s^t \left[\sigma\left(u, X_u^0\right) - \sigma_0\left(u, X_u^0\right)\right] dW_u.$$

We introduce a truncation to obtain Lipschitz inequalities for μ and σ. Define

$$I_{n,t} = \begin{cases} 1, & |X_t|, \left|X_t^0\right| \leq n \text{ for } t \in [s, T] \\ 0, & \text{otherwise} \end{cases}$$

which is $\mathcal{F}_t$-measurable and satisfies $I_{n,t} = I_{n,s} I_{n,t}$ for $0 \leq s \leq t$. With $Y_{n,t} = I_{n,t}\left[X_t - X_t^0\right]$, we then get

$$Y_{n,t} = \eta_{n,t} + I_{n,t} \int_s^t I_{n,u}\left[\mu(u, X_u) - \mu\left(u, X_u^0\right)\right] du + I_{n,t} \int_s^t I_{n,u}\left[\sigma(u, X_u) - \sigma\left(u, X_u^0\right)\right] dW_u,$$

where

$$\eta_{n,t} = I_{n,t} \int_s^t I_{n,u}\left[\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right] du + I_{n,t} \int_s^t I_{n,u}\left[\sigma\left(u, X_u^0\right) - \sigma_0\left(u, X_u^0\right)\right] dW_u.$$

Since μ and σ are continuously differentiable, for every $n \geq 1$, there exists a $K_n > 0$ such that

$$\begin{aligned} \|\mu(u, x) - \mu(u, y)\|^2 &\leq K_n \|x - y\|^2, \\ \|\sigma(u, x) - \sigma(u, y)\|^2 &\leq K_n \|x - y\|^2, \end{aligned}$$

for $\|x\|, \|y\| \leq n$, and $s \leq u \leq T$. Thus,

$$\begin{aligned}
E_s |Y_{n,t}|^2 &\leq 3E_s |\eta_{n,t}|^2 + 3E_s \left[\|X_s - X_s^0\|^2 \right] + \\
&\quad 3E_s \left[\int_s^t I_{n,u} \|\mu(u, X_u) - \mu(u, X_u^0)\|^2 du \right] \\
&\quad +3E_s \left[\int_s^t I_{n,u} \|\sigma(u, X_u) - \sigma(u, X_u^0)\|^2 du \right] \\
&\leq 3E_s |\eta_{n,t}|^2 + 3\|X_s - X_s^0\|^2 + 3K_n E_s \left[\int_s^t I_{n,u} \|X_u - X_u^0\|^2 du \right] \\
&\quad +3K_n E_s \left[\int_s^t I_{n,u} \|X_u - X_u^0\|^2 du \right] \\
&\leq 3E_s |\eta_{n,t}|^2 + 3\|X_s - X_s^0\|^2 + 3tK_n \int_s^t E_s \left[\|Y_{n,u}\|^2 \right] du \\
&\quad +3K_n \int_s^t E_s \left[\|Y_{n,u}\|^2 \right] du \\
&\leq 3E_s |\eta_{n,t}|^2 + 3\|X_s - X_s^0\|^2 + 6(1+T) K_n \int_s^t E_s \left[\|Y_{n,u}\|^2 \right] du.
\end{aligned}$$

We also have that

$$E_s |\eta_{n,t}|^2 \leq 3 \int_s^T E_s \left[\|\mu(X_u^0) - \mu_0(X_u^0)\|^2 + \|\sigma(u, X_u^0) - \sigma_0(u, X_u^0)\|^2 \right] du$$

So with

$$\begin{aligned}
\delta &\equiv 3\|X_s - X_s^0\|^2 + 3\int_s^T E_s \left[\|\mu(X_u^0) - \mu_0(X_u^0)\|^2 \right] du \\
&\quad +3\int_s^T E_s \left[\|\sigma(u, X_u^0) - \sigma_0(u, X_u^0)\|^2 \right] du
\end{aligned}$$

we obtain

$$0 \leq E|Y_{n,t}|^2 \leq 2\delta + \beta_n \int_s^t E|Y_{n,s}|^2 ds$$

with $\beta_n = 6(1+T) K_n > 0$. By Karatzas and Shreve (1991, Problem 5.2.7),

$$\begin{aligned}
\sup_{s \leq t \leq T} E|Y_{n,t}|^2 &\leq 2\delta + \delta\beta_n \sup_{s \leq t \leq T} \int_s^t e^{-\beta_n(t-s)} ds \\
&\leq 2\delta \left[1 + \beta_n \int_0^T e^{-\beta_n(t-s)} ds \right] \\
&\leq 2\delta \left[1 - e^{-\beta_n T} \right] \\
&\leq 2\delta.
\end{aligned}$$

We conclude that $\sup_{s \leq t \leq T} E_s \left[\|X_t - X_t^0\|^2 \right] \leq \delta$, since this bound holds uniformly over $n \geq 1$. ∎

Lemma 52 *For any* $(\mu, \sigma^2) \in \mathcal{D}$,

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

$$\begin{aligned} & E_s\left[\left\|X_t - X_t^0\right\|^{2p}\right] \\ \leq \ & \left\|X_s - X_s^0\right\|^{2p} + C(m) \int_s^T E_s\left[\left\|\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right\|^{2p}\right] du \\ & + C(m) \int_s^T E_s\left[\left\|\sigma\left(u, X_u^0\right) - \sigma_0\left(u, X_u^0\right)\right\|^{2p}\right] du \end{aligned}$$

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*

$$\begin{aligned} & E_{x,s}\left[\left\|f(t, X_t) - f\left(t, X_t^0\right)\right\|\right] \\ \leq \ & C(t,x)\left(\int_s^t E_{x,s}\left[\left\|\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right\|^2\right] du\right. \\ & \left.\int_s^T E_{x,s}\left[\left\|\sigma^2\left(u, X_u^0\right) - \sigma_0^2\left(u, X_u^0\right)\right\|^2\right] du\right)^{1/2}, \end{aligned}$$

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

Proof. Apply Itô's Lemma with $f(x) = x^{2m}$ on the process $X_t - X_t^0$ as written in (5.51), and then proceed as in the proof of Lemma 51. This yields 1. The second result follows by combining the inequality that f satisfies with Lemma 51. ■

Lemma 53 *For any $(\mu, \sigma^2) \in \mathcal{D}$, the $\mathbb{R}^q$-valued diffusion process*

$$dY_t^{(i)} = \mu^{(1)}\left(t, X_t^{s,x}\right) Y_t^{(i)} dt + Y_t^{(i)} \sigma^{(1)}\left(t, X_t^{s,x}\right) dW_t, \quad Y_s^{(i)} = e_i, \tag{5.52}$$

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

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

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.

Proof. By assumption, $X_t^x = X_t^{s,x}$ and $X_t^{x+h} = X_t^{s,x+h}$ are well-defined unique solutions for any x and h. Define

$$\begin{aligned} Y_t^h &= h^{-1}\left(X_t^{x+h} - X_t^x\right) = 1 + h^{-1} \int_s^t \mu\left(u, X_u^{x+h}\right) - \mu\left(u, X_t^x\right) du \\ & + h^{-1} \int_s^t \sigma\left(u, X_u^{x+h}\right) - \sigma\left(u, X_t^x\right) dW_u. \end{aligned}$$

We can write

$$\begin{aligned}
& h^{-1}\int_s^t \mu\left(u, X_u^{x+h}\right) - \mu\left(u, X_t^x\right) du \\
= \; & h^{-1}\int_s^t \left[\int_0^1 \frac{\partial}{\partial\alpha}\mu\left(u, X_u^x + \alpha\left[X_u^{x+h} - X_u^x\right]\right) d\alpha\right] du \\
= \; & \int_s^t \left[\int_0^1 \mu_x\left(u, X_u^x + \alpha\left[X_u^{x+h} - X_u^x\right]\right) d\alpha\right] h^{-1}\left[X_u^{x+h} - X_u^x\right] du \\
= \; & \int_s^t \left[\int_0^1 \mu_x\left(u, X_u^x + \alpha\left[X_u^{x+h} - X_u^x\right]\right) d\alpha\right] Y_t^h du.
\end{aligned}$$

Similarly,

$$h^{-1}\int_s^t \sigma\left(u, X_u^{x+h}\right) - \sigma\left(u, X_t^x\right) dW_u = \int_s^t \left[\int_0^1 \sigma_x\left(u, X_u^x + \alpha\left[X_u^{x+h} - X_u^x\right]\right) d\alpha\right] Y_t^h dW_u.$$

By Lemma 51,

$$\begin{aligned}
E_{s,x}\left[\left\|Y_t - Y_t^h\right\|^2\right] \le \; & 6\int_s^t E\left[\left\|Y_u\left[\mu_x\left(u, X_u^x\right) - \left[\int_0^1 \mu_x\left(u, X_u^x + \alpha h Y_u\right) d\alpha\right]\right]\right\|^2 du\right] \\
& +6\int_s^t E\left[\left\|Y_u\left[\sigma_x\left(u, X_u^x\right) - \left[\int_0^1 \sigma_x\left(u, X_u^x + \alpha h Y_u\right) d\alpha\right]\right]\right\|^2 du\right]
\end{aligned}$$

where the two terms on the RHS go to zero as $h \to 0$.

The proof concerning the second derivative follows along the same lines. ■

Lemma 54 *For any $\left(\mu, \sigma^2\right) \in \mathcal{D}$:*

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

$$d\nabla_1 X_t = \left\{d\mu\left(t, X_t^0\right) + \mu_0^{(1)}\left(t, X_t^0\right)\nabla_1 X_t\right\} dt + \sigma_0^{(1)}\left(t, X_t^0\right)\nabla_1 X_t dW_t, \tag{5.53}$$

and

$$d\nabla_2 X_t = \mu_0^{(1)}\left(t, X_t^0\right)\nabla_2 X_t dt + \left\{\frac{1}{2}\sigma_0^{-1}\left(t, X_t^0\right) d\sigma^2\left(t, X_t^0\right) + \sigma_0^{(1)}\left(t, X_t^0\right)\nabla_2 X_t\right\} dW_t, \tag{5.54}$$

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

$$\begin{aligned}
& E_s\left[\left\|X_t - X_t^0 - \nabla X_t\right\|^2\right] \\
\le \; & \left\|X_s - X_s^0\right\|^2 + \sum_{|\alpha|\le 1}\int_s^t E_s\left[\left\|\partial_x^\alpha \mu\left(u, X_u^0\right) - \partial_x^\alpha \mu_0\left(u, X_u^0\right)\right\|^4\right] du
\end{aligned} \tag{5.55}$$

where $\{X_t\}$ solves SDE(μ, σ_0^2) and

$$E_s\left[\left\|\nabla X_t\right\|^2\right] \le \int_s^t E_s\left[\left\|\mu\left(u, X_u^0\right) - \mu_0\left(u, X_u^0\right)\right\|^2\right] du. \tag{5.56}$$

2. The process $\{\nabla_2 X_t\}$ given in (5.54) with $d\sigma^2 = \sigma^2 - \sigma_0^2$, satisfies

$$\begin{aligned} & E_s\left[\left\|X_t - X_t^0 - \nabla_2 X_t\right\|^2\right] && (5.57) \\ \leq\ & \left\|X_s - X_s^0\right\|^2 + \underline{\sigma}^{-2} \sum_{|\alpha|\leq 1} \int_s^t E_s\left[\left\|\partial_x^\alpha \sigma^2\left(u, X_u^0\right) - \partial_x^\alpha \sigma_0^2\left(u, X_u^0\right)\right\|^4\right] du \end{aligned}$$

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

$$E_s\left[\left\|\nabla_2 X_t\right\|^2\right] \leq \underline{\sigma}^{-2} \int_s^t E_s\left[\left\|\sigma^2\left(u, X_u^0\right) - \sigma_0^2\left(u, X_u^0\right)\right\|^2\right] du. \tag{5.58}$$

Proof. We only show the part of the theorem concerning $\{\nabla_1 X_t\}$. The proof of the second part follows along the same lines. We have

$$X_t - X_t^0 - \nabla_1 X_t = X_s - X_s^0 + \int_s^t \mu\left(u, X_u\right) - \mu\left(u, X_u^0\right) - \partial_x \mu_0\left(u, X_u^0\right) \nabla_1 X_u du$$

where

$$\begin{aligned} & E_s\left[\int_s^t \left\|\mu\left(u, X_u\right) - \mu\left(u, X_u^0\right) - \partial_x \mu_0\left(u, X_u^0\right) \nabla_1 X_u\right\|^2 du\right] \\ \leq\ & E_s\left[\int_s^t \left\|\mu\left(u, X_u^0\right) - \mu\left(u, X_u^0\right) - \partial_x \mu\left(u, X_u^0\right)\left(X_u - X_u^0\right)\right\|^2 du\right] \\ & + E_s\left[\int_s^t \left\|\partial_x \mu\left(u, X_u^0\right) - \partial_x \mu_0\left(u, X_u^0\right)\right\|^2 \left\|X_u - X_u^0\right\|^2 du\right] \\ & + E_s\left[\int_s^t \left\|\partial_x \mu_0\left(u, X_u^0\right)\right\|^2 \left\|X_u - X_u^0 - \nabla_1 X_u\right\|^2 du\right] \\ \leq\ & C E_s\left[\int_s^t \left\|X_u - X_u^0\right\|^4 du\right] \\ & + C\left(\int_s^T E_s\left[\left\|\partial_x \mu\left(u, X_u^0\right) - \partial_x \mu_0\left(u, X_u^0\right)\right\|^4\right] du\right)^{1/2} \left(E_s \int_s^t \left\|X_u - X_u^0\right\|^4 du\right)^{1/2} \\ & + C E_s\left[\int_s^t \left\|X_u - X_u^0 - \nabla_1 X_u\right\|^2 du\right], \end{aligned}$$

By applying Lemma 51 and collecting the resulting terms, we obtain the result by Karatzas and Shreve (1991, Problem 5.2.7).

The last inequality follows by an application of Lemma 51 on $\nabla_1 X_t[\mu - \mu_0] = \nabla_1 X_t[\mu] - \nabla_1 X_t[\mu_0, \sigma_0]$. ■

5.B.1 The Semiparametric Estimator

Lemma 55 *Under (SP.0)-(SP.3) and (SP1.A) [(SP2.A)], the estimators for (μ, σ) in Class 1 [2] satisfy*

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

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

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

Proof. In Class 1, the convergence of the diffusion estimator follows by the arguments used in the parametric case. For the drift estimator, by Lemma 33,

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

and

$$||\hat{\pi}^{(k)} - \pi_0^{(k)}||_\infty = O_P\left(n^{-1/2}h^{-1-k}\right) + O_P\left(h^{\omega-k}\right). \tag{5.59}$$

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

$$\begin{aligned}
&\int_t^T E_{t,x}\left[||\hat{\mu}_0^{(i)}(X_s^0) - \mu_0^{(i)}(X_s^0)||^4\right] ds \\
\leq\; & a^4\left(\int_t^T E_{t,x}\left[|\partial_a\hat{T}(x;\bar{a})|^8\right] ds\right)^{1/2}\left(\int_t^T E_{t,x}\left[||\mu_0^{(i)}(X_s^0)||^8\right] ds\right)^{1/2},
\end{aligned}$$

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 by using Lemma 38. ■

Lemma 56 *Under (SP.0)-(SP.3) and (SP1.A-B), the estimators for (μ,σ) in Class 1 satisfy uniformly in $\theta \in \Theta$,*

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

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

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

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 Lemma 33),

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

where $\nabla_\pi\hat{\mu}^{(i)}$ is given in (4.63)-(4.65). The convergence of the RHS is given in (5.59). 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

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

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

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

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

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, this time applying Lemma 38. ■

5.B.2 The Nonparametric Estimator

Lemma 57 *Under (NP.0)-(NP.4),*

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

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

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

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

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

where $\kappa_{n,\bar{T}}\equiv\max_i\sup_{s\in[i\Delta,(i+1)\Delta]}|x_s-x_i|=O_P(\Delta^{1/2}\sqrt{\log(\Delta^{-1})})$ 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 $B_n = \{x : |x| \leq \bar{T}^{\gamma}\}$, for some $\gamma > 0$, and a covering of B_n,

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

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

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

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 $\tilde{m}(x) = \bar{T}^{-1}\int_0^{\bar{T}} K_h(x_s - x)\,\mu(x_s)\,ds$. We have

$$\frac{x_{i+1} - x_i}{\Delta} - \mu(x_i) = \Delta^{-1}\int_{i\Delta}^{(i+1)\Delta} \{\mu(x_s) - \mu(x_i)\}\,ds + \Delta^{-1}\int_{i\Delta}^{(i+1)\Delta} \sigma(x_s)\,dW_s, \quad (5.60)$$

such that

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

and

$$\begin{aligned}
&\left|n^{-1}\sum_{i=1}^{n}K_h(x_i-x)\mu(x_i)-T^{-1}\int_0^{\bar{T}}K_h(x_s-x)\mu(x_s)\,ds\right| \\
=\ & n^{-1}\left|\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta}K_h(x_i-x)\mu(x_i)-K_h(x_s-x)\mu(x_s)\,ds\right| \\
\le\ & n^{-1}\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta}|\mu(x_s)|\,|K_h(x_i-x)-K_h(x_s-x)|\,ds \\
&+n^{-1}\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta}K_h(x_s-x)\left|\mu^{(1)}(x_s+o_P(1))\right||x_i-x_s|\,ds \\
=\ & \bar{K}\frac{\kappa_{n,\bar{T}}}{h^2}\bar{T}^{-1}\int_0^{\bar{T}}|\mu(x_s)|\,ds+\bar{K}\frac{\kappa_{n,\bar{T}}}{h}\bar{T}^{-1}\int_0^{\bar{T}}\left|\mu^{(1}(x_s)\right|ds \\
=\ & O_P\left(\frac{\kappa_{n,\bar{T}}}{h^2}\right).
\end{aligned}$$

where the bound does not depend on x. Next,

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

and

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

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

$$\begin{aligned}
\operatorname{var}\left[S_{\bar{T}}(x)\right] &= \left(\bar{T}h\right)^{-1}E\left[h^{-1}K^2\left(\frac{x_s-x}{h}\right)\sigma^2(x_s)\right] \\
&= \left(\bar{T}h\right)^{-1}\left(\sigma^2(x)\,\pi(x)\int K^2(z)\,dz+o(1)\right).
\end{aligned}$$

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

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

To prove $\sup_{x\in\mathbb{R}} |\hat{s}(x) - s_0(x)| = o_P(\bar{T}^{-1/4})$, we first apply Itô's Lemma on (5.60) to obtain

$$\begin{aligned}\frac{(x_{i+1}-x_i)^2}{\Delta} - \sigma^2(x_i) &= \frac{2}{\Delta}\int_{i\Delta}^{(i+1)\Delta} \mu(x_s)(x_s - x_i)\,ds + \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} \sigma^2(x_s) - \sigma^2(x_i)\,ds \\ &\quad + \frac{2}{\Delta}\int_{i\Delta}^{(i+1)\Delta} \sigma(x_s)(x_s - x_i)\,dw_s,\end{aligned}$$

and by using similar arguments as before,

$$\begin{aligned}\frac{2}{\Delta n}\sum_{i=1}^{n} K_h(x_i - x)\int_{i\Delta}^{(i+1)\Delta} \mu(x_s)(x_s - x_i)\,ds &= O_P\left(\frac{\kappa_{n,\bar{T}}^2}{h^2}\right), \\ \frac{2}{\Delta n}\sum_{i=1}^{n} K_h(x_i - x)\int_{i\Delta}^{(i+1)\Delta} \sigma^2(x_s) - \sigma^2(x_i)\,ds &= O_P\left(\frac{\kappa_{n,\bar{T}}}{h^2}\right), \\ n^{-1}\sum_{i=1}^{n} K_h(x_i - x)\sigma^2(x_i) - T^{-1}\int_0^{\bar{T}} K_h(x_s - x)\sigma^2(x_s)\,ds &= O_P\left(\frac{\kappa_{n,\bar{T}}}{h^2}\right),\end{aligned}$$

and

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

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

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

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

Lemma 58 *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

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

where

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

It holds that

$$\begin{aligned} \sup_{x\in\mathbb{R}} \hat{T}(x;a)\left|\hat{\mu}(x)-\mu_0(x)\right| &\leq \sup_{x\in\mathbb{R}} \hat{T}(x;a)\left|\frac{\hat{m}(x)}{\hat{\pi}(x)}-\frac{m_0(x)}{\pi_0(x)}\right| \\ &\leq a^{-1}\left\|\hat{m}-m_0\right\|_\infty + a^{-2}\left\|\hat{\pi}-\pi_0\right\|_\infty, \end{aligned}$$

and

$$\begin{aligned} & \sup_{x\in\mathbb{R}} \hat{T}(x;a)\left|\hat{\mu}^{(1)}(x)-\mu_0^{(1)}(x)\right| \\ \leq\ & \sup_{x\in\mathbb{R}} \hat{T}(\cdot;a)\left\{\left|\frac{\hat{m}^{(1)}}{\hat{\pi}}-\frac{m_0^{(1)}}{\pi_0}\right| + \left|\frac{\hat{m}}{\hat{\pi}}\frac{\hat{\pi}^{(1)}}{\hat{\pi}}-\frac{m_0}{\pi_0}\frac{\pi_0^{(1)}}{\pi_0}\right|\right\} \\ \leq\ & a^{-2}\left\|\hat{m}-m_0\right\|_\infty + a^{-1}\|\hat{m}^{(1)}-m_0^{(1)}\|_\infty + a^{-3}\left\|\hat{\pi}-\pi_0\right\|_\infty + a^{-2}\|\hat{\pi}^{(1)}-\pi_0^{(1)}\|_\infty, \end{aligned}$$

Using the rates of convergence established in Lemma 57, 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 (5.49) it holds that

$$\begin{aligned} & \mu(x)-\mu_0(x)-\nabla\mu_0(x)\left[m-m_0,\pi-\pi_0\right] \\ =\ & \frac{m(x)}{\pi(x)}-\frac{m_0(x)}{\pi_0(x)}-\frac{1}{\pi_0(x)}\left[m(x)-m_0(x)\right]+\frac{m_0(x)}{\pi_0^2(x)}\left[\pi(x)-\pi_0(x)\right] \\ =\ & \frac{\pi(x)-\pi_0(x)}{\pi(x)\pi_0(x)}\left[\left[m(x)-m_0(x)\right]-\frac{m_0(x)}{\pi_0(x)}\left[\pi(x)-\pi_0(x)\right]\right] \end{aligned}$$

such that

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

Thus,

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

where

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

■

6

A Semiparametric Single-Factor Model of the Term Structure

6.1 Introduction

The short-term interest rate is a state variable which enters in many different strands of economic and financial theory. It has strong implications for the pricing of fixed income securities and interest rate derivatives, e.g. bonds, options, futures, swaps. But it is also used in general asset pricing, and as an input in macroeconomic models, e.g. in the analysis of the business cycle. It is therefore of great interest to obtain a suitable model describing its dynamics. Diffusion processes are widely used for this purpose, which owes to the fact that continuous-time models greatly facilitate the theoretical analysis of financial markets. They prove particularly useful in derivative pricing since continuous-time arbitrage arguments then can be applied, allowing for a relative simple, and yet elegant, solution to the problem, see e.g. Duffie (1996). The theoretical option prices turn out to functionals of the underlying short-term interest rate, so in order to apply these, one has to (i) set up an appropriate model for the interest rate and (ii) calibrate this to the market of interest. There is a huge literature dealing with (i), ranging from relatively simple Markov models of the short-term rate (so-called single-factor models), over multi-factor models where the short-term rate is assumed to depend on several (potentially unobserved) factors, to the class of Heath-Jarrow-Morton (1992) [HJM] type models where a continuum of factors drives the yield curve.

It is however still an open question which of the many proposed models is the most adequate when calibrating it to interest rate data, see Rogers (1995) for a discussion of these issues. There is a large number of studies where different diffusion models of the term structure are implemented using historical interest rate data. These studies have mainly focused on parametric specifications of the diffusion model. But there appears to be no universal model which fits all interest rate data equally well. It is therefore still an open question which model one should choose given a specific interest rate data set. Given this problem, one may benefit from using non- or semiparametric methods since these allow for a degree of flexibility compared to the parametric case.

In this chapter, we focus on the class of single-factor models of the short-term interest rate. This class of term structure models assumes that the short-term interest rate is a Markov process solving a SDE, in which case this state variable drives the whole yield curve. Economic theory puts no restrictions on the drift and diffusion term, except that the resulting short-term interest rate should be positive. Thus, a model for the short-term interest rate can only be judged by how well it fits the available data. In such situations, instead of restricting oneself to a parametric model, it would be more appropriate to "let data speak for itself", which is exactly what non- and semiparametric models do. A number of different parametric models for the SDE has been proposed and tried out on historical data with varying degrees of success. Aït-Sahalia (1996b) fitted some of these parametric models to an interest rate data set, and then tested each model against a nonparametric alternative. The striking conclusion was that none of these could be accepted as the true model; more flexible models were needed. A number of other empirical studies have found similar evidence of nonlinearities both in the drift and diffusion term for this type of data, which the models of Aït-Sahalia (1996a) and Conley et al. (1997) cannot capture, see Ahn and Gao (1999), Bandi (2002), Jiang and Knight (1997, Stanton (1997), Tauchen (1995). The class of single-factor diffusion models are characterised by its drift and diffusion function which can be interpreted as the instantaneous mean and variance respectively. We propose a semiparametric diffusion model where we choose a very flexible parametric form for the diffusion term while leaving the drift term unspecified. The chosen parameterisation of the diffusion term is highly flexible, and the model encompasses most of the parametric models found in the literature. This model is very general; in particular, it includes most of the parametric models suggested in the literature as special cases. The model can be estimated using the general estimation procedure proposed in Chapter 4. The parametric part is estimated by a profiled version of the log-likelihood, while the drift term is estimated using kernel methods. Since the semiparametric model nests most of the parametric specifications as special cases, we are able to perform a specification test of each of these models against the semiparametric alternative.

We fit our model to a proxy of the short-term Eurodollar interest rate. Various parametric single-factor models have been fitted to this interest rate in a number of empirical studies, see for example Aït-Sahalia (1996b), Elerian et al (2001), Durham (2002). The conclusions drawn in the different studies are not conclusive, and it is not clear which model should be preferred. We reexamine the data set, fitting the proposed semiparametric diffusion model to it. We find nonlinearities in the drift function that even the most flexible parametric model cannot capture. The performance of the semiparametric model is then compared with the parametric model proposed in Aït-Sahalia (1996b); this is the most flexible parametric single-factor model found in the literature. The comparison is made along two lines: First we test the parametric diffusion models against the semiparametric alternative, using the test statistic proposed in Chapter 4. Second, we calculate a range of bond prices predicted by the competing models and see whether they are statistically significant; this is done using the results of Chapter 5. The second comparison is the most useful for practitioners if the end goal with the model is to price bonds and derivatives.

In Section 2, we give an overview over the various parametric single factor models proposed in the literature. In order to apply a single-factor model to price bonds and options the market price for risk has to be determined; in Section 3, we present a general estimation procedure to do this. The Eurodollar interest rate data set is presented in Section 4 along with a discussion of the various studies who have previously examined this. In Section 5, the estimation results for the semiparametric model and the implied bond and derivative prices are presented. We conclude in Section 6.

6.2 Single-Factor Term Structure Models

Single-factor models constitute a relatively simple class of models where the whole term structure is driven by one single state variable, the short-term interest rate. More advanced models such as multi-factor models, and HJM-models might be a more plausible way of describing the dynamics of the term structure, but this comes at the cost of a more difficult and computationally intensive implementation. Most of the applied studies of multi-factor and models only consider linear specifications in order to overcome the difficulties of estimating the model. Prominent examples are Brennan and Schwartz (1979), Chen and Scott (1992, 1993), Dai and Singleton (2000), Longstaff and Schwartz (1992). As an alternative, Ahn, Dittmar & A.R. Gallant (2002) consider a quadratic specification. In these type of models, while a large number of empirical studies have argued that at least two- or three-factor models are needed to fit the term structure properly, it is also found in that the short rate accounts for up to 90 percent of the variation in the data, see e.g. Litterman and Scheinkman (1991). Thus, it is of interest even within a multi-factor framework to find a suitable model for the short rate.

In the class of single-factor models we consider here, the short-term rate solves a time-homogenous stochastic differential equation (SDE) of the form

$$dr_t = \mu\left(r_t\right)dt + \sigma\left(r_t\right)dW_t, \tag{6.1}$$

where $\{W_t\}$ is a standard Brownian motion. We then wish to model the drift term, $\mu : \mathbb{R}_+ \mapsto \mathbb{R}$, and the diffusion term, $\sigma^2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$. A more flexible class of single-factor models can be constructed by using time-inhomogenous SDE's where μ and σ^2 are allowed to depend on time t. These are widely used in the financial industry, since these can be calibrated on a daily basis to deliver a perfect fit of the current yield curve, see e.g. Ho and Lee (1986), Hull and White (1990). But it is not evident that this leads to better out of sample performance and more correct pricing. In particular, these models do not specify the dynamics of the time varying coefficients. This and other arguments against time-inhomogenous models can be found in Dybvig (1997) and Backus et al. (1998).

Within the framework of single-factor models, the price of any bond or interest rate derivative can be shown to be a functional of μ and σ^2, see for example Chapter 3. So in order to be able to price such claims correctly, one needs to specify μ and σ^2 correctly. The traditional models normally assume a linear drift and diffusion term, but in the past decade a large body of empirical work has indicated that such specifications do not fit observed interest rates very well. Using a misspecified term structure model can have serious implications. For example, as observed in Chapman and Pearson (2001), "the existence and strength of nonlinear mean reversion have important implications for the likelihood of extreme interest rate changes and for the distribution of interest rate changes over long time horizon. As a result, they have significant implications for value-at risk calculations over long horizon and asset-liability management. Moreover, nonlinear mean reversion may also have implications for pricing long-term bonds and interest rate options". We may add, that the presence of nonlinearities in the diffusion term of course also will have important implications in the aforementioned applications.

A large part of the finance literature has focused on σ^2 as the crucial parameter of interest in derivative pricing, while to a certain extent neglecting the role of μ. In a Black-Scholes setting where the underlying variable is a traded asset this focus is correct since only σ^2 enters the derivative pricing formula. But in interest rate derivative pricing, the drift will also enter the formula and can have important effects on the prices. Moreover, in the calibration of the model, both the drift and diffusion term has to be specified correctly in order to avoid biased estimates. Given discrete observations, one can in most cases not separate the estimation of σ^2 from μ, these are invariably linked together in the estimation. So even if one has correctly specified σ^2, misspecification of μ will lead to a biased estimator of σ^2, which in turn will have implications for the pricing of derivatives.

In the past decade, a number of empirical studies have been directed towards finding an appropriate specification of μ and σ^2. Economic theory puts no restrictions on the drift and diffusion term, except that the resulting short-term interest rate should stay positive. Thus, a model for the short-term interest rate can only be judged by how well it fits the available data. In such situations, instead of restricting oneself to a parametric model, it would be more appropriate to "let data speak for itself", which is exactly what non- and semiparametric models do. A number of different parametric models for the SDE has been proposed and tried out on historical data with varying degrees of success. Aït-Sahalia (1996b) fitted some of these parametric models to an interest rate data set, and then tested each model against a nonparametric alternative. The striking conclusion was that none of these could be accepted as the true model; more flexible models were needed. A number of other empirical studies have found similar evidence of nonlinearities both in the drift and diffusion term for this type of data, which the models of Aït-Sahalia (1996a) and Conley et al. (1997) cannot capture, see Ahn and Gao (1999), Bandi (2002), Jiang and Knight (1997, Stanton (1997), Tauchen (1995). The importance of nonlinearities in the two terms should not be downplayed.

We here propose the following semiparametric model where the drift term is unspecified, while the diffusion term follows the flexible parameterisation proposed in Aït-Sahalia (1996b),

$$dr_t = \mu(r_t)\,dt + \sqrt{\sigma_0 + \sigma_1 r_t + \sigma_2 r_t^{\gamma}}dW_t. \tag{6.2}$$

Some of the most popular (stationary) models are quoted in Table 6.1. As can be seen, the semiparametric model encompasses most of these models. It is therefore possible to test each of the parametric alternative against the semiparametric alternative using the test statistic developed in Chapter 4.

TABLE 6.1. Parametric specifications of the spot rate.

$\mu(r)$	$\sigma^2(r)$	Reference
$\beta(\alpha - r)$	σ^2	Vasicek (1977)
$\beta r(\alpha - \log(r))$	$\sigma^2 r^2$	Brennan and Schwartz (1979)
$\beta(\alpha - r)$	$\sigma^2 r^2$	Courtadon (1982)
$\beta_1 r + \beta_2 r^{-(1+\gamma)}$	$\sigma^2 r^\gamma$	Marsh and Rosenfeld (1983)
$\beta(\alpha - r)$	$\sigma^2 r$	Cox et al. (1985)
$\beta(\alpha - r)$	$\sigma^2 r^\gamma$	Chan et al. (1992)
$\beta_0 + \beta_1 r + \beta_2 r^2$	$(\sigma_0 + \sigma_1 r)^2$	Constantinides (1992)
$\beta(\alpha - r)$	$\sigma_0^2 + \sigma_2^2 r$	Duffie and Kan (1996)
$\beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 r^{-1}$	$\sigma_0 + \sigma_1 r + \sigma_2 r^\gamma$	Aït-Sahalia (1996b)
$\beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 r^{-1}$	$\sigma^2 r^\gamma$	Tauchen (1995), Conley et al. (1997)
$\beta_0 + \beta_1 r + \beta_2 r^2$	$\sigma_0 + \sigma_1 r + \sigma_2 r^3$	Ahn and Gao (1999)
$\beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 r^{-1}$	$\sigma_0 + \sigma_1 r + \sigma_2 r^2 + \sigma_2 r^3$	Elerian et al (2001)

6.3 Estimation of the Risk Premium

In Chapter 3, we derived formulae for bond and option prices in a single-factor framework. An important ingredient in these was the risk premium process, $\{\lambda_t\}$. In order to apply our calibrated models to the pricing of such securities we therefore have to obtain an estimate of this process. In the following, we go through some of the methods suggested in the literature. For a more detailed treatment, we refer to Garcia, Ghysels and Renault (2004).

We start out with a general model for an option price Π. Assume that the price satisfies

$$\Pi = \Gamma(X, Z) + \varepsilon,$$

where X is a collection of random variables (including, for example, the current value of the underlying asset), Z a collection of deterministic characteristics associated with the option (time to maturity, type of pay-off function etc.), and ε is an error term (present due to e.g. pricing errors, failure of the theoretical model to perfectly match the observed data).

A branch of the empirical option pricing literature proposes to estimate the function Γ non- or semiparametrically, thereby not having to specify the dynamics of the underlying variable. Prominent examples of this approach is Aït-Sahalia and Lo (1998), Aït-Sahalia and Duarte (2003), Bondarenko (2003). This approach has the advantage of not imposing any restrictions on the dynamics of the underlying variable, and not making the estimation of any risk premium necessary. On the other hand, the precision of predicted option prices will suffer from the slower convergence of nonparametric estimators.

Restricting the dynamics of the term structure to be of diffusion type gives us additional information about the function. In the arbitrage-free framework of the single-factor model, it holds that

$$\Gamma(x, z) = \Gamma(r, g, t, T) = E^Q\left[g(r_T)\exp\left[-\int_t^T r_u du\right] | r_t = r\right],$$

cf. Chapter 3. Assuming that $\{r_t\}$ is time-homogenous under Q, the formula simplifies further to

$$\Gamma(r, g, \tau) = E^Q \left[g(r_\tau) \exp\left[-\int_0^\tau r_u du \right] | r_0 = r \right], \tag{6.3}$$

where $\tau = T - t$ is time to maturity. The dynamics of $\{r_t\}$ under Q depends on the risk premium λ,

$$dr_t = \{\mu(r_t) + \lambda(r_t)\sigma(r_t)\} dt + \sigma(r_t) dW_t. \tag{6.4}$$

Let $\{(\Pi_{ij}, r_i, g_j, \tau_j) | 1 \le i \le n,\ 1 \le j \le J\}$ be a collection of observed option prices together with the associated observed short rate and the characteristics. Taking μ and σ^2 for given (in our case, we will have preliminary estimates of these), the only unknown is λ. This yields the following regression model,

$$\Pi_{ij} = \Gamma(r_i, g_j, \tau_j; \lambda) + \varepsilon_{ij},$$

where the function Γ takes the form (6.3), and we assume that $E[\varepsilon_{ij}|r_i] = 0$. We may then estimate the unknown function λ by for example least squares. In practice this means that we choose λ such that the option prices implied by the single-factor model mimic the observed ones as closely as possible. It is still an open question whether λ can be identified in the above regression model. Observe that in fact the estimation problem here in some sense is the inverse of the one considered in Chapter 5. While there we had derived the asymptotic properties of the solution to the PDE given estimators of μ and σ^2, we here have observed solutions to the PDE (bond and derivative prices) from which we wish to extract an estimator of the one of the coefficients driving the PDE. The main problem here is then the inversion of the functional $u = \Gamma(\mu - \lambda\sigma, \sigma^2)$, with Γ given in Chapter 5, w.r.t. its first argument such that $\mu - \lambda\sigma = \Gamma^{-1}(u, \sigma^2)$. If the inverse of Γ is well-defined, identification of λ is ensured. Assuming an affine specification of μ, σ^2 and λ^2, a closed form expression of Γ is available, and one can in this setting show that the parameters entering λ are identifiable, cf. Duffie and Kan (1996). In the general case however, the function Γ is a complicated functional of λ which cannot be written on analytical form, and the identification problem is not easily resolved. To the author's knowledge, no general results concerning identification of λ exist. Here, we shall therefore simply assume that λ is identified. A nonparametric estimator of λ can be obtained by the method of sieves. Normally however, one assumes a parametric version, $\lambda(r) = \lambda(r; \theta)$, where θ is an unknown finite-dimensional parameter. For example, $\lambda(r) = \lambda$ in Vasicek (1977) and Aït-Sahalia (1996a), and $\lambda(r) = \lambda/\sigma\sqrt{r}$ in Cox-Ingersoll-Ross (1985). The general estimation procedure presented above is also applicable to multifactor models.

Some alternative estimation procedure can be found in the literature. Consider the yield of a zero-coupon bond with maturity at time T, $Y_t(T) = \log(B_t(T)) / (T - t)$ with $B_t(T)$ being the price of zero-coupon bond with maturity at time $T \ge t$. Vasicek (1977) observed that

$$\frac{\partial Y_t(T)}{\partial T}|_{T=t} = \frac{1}{2}[\mu(r_t) - \sigma(r_t)\lambda(r_t)],$$

and proposed to use observed yields to approximate the right hand side of the equation. Jiang (1998) derived an expression of the risk premium in terms of two different yields and the dynamics of those. By Itô's Lemma, $Y_t(T)$ solves an SDE,

$$dY_t(T) = \alpha(r_t, T)\,dt + \kappa(r_t, T)\,dW_t, \tag{6.5}$$

where the no-arbitrage assumption imposes restrictions on the drift and diffusion term. In fact, the following relation have to hold for any $T_1, T_2 > 0$,

$$\lambda(r_t) = \frac{\Delta Y_t(T_1, T_2) + \frac{1}{2}\left[\tau_1^2\kappa^2(r_t, T_1) - \tau_2^2\kappa^2(r_t, T_2) + \tau_2\alpha(r_t, T_2) - \tau_1\alpha(r_t, T_1)\right]}{\tau_2\kappa(r_t, T_2) - \tau_1\kappa(r_t, T_1)}, \tag{6.6}$$

where $\Delta Y_t(T_1, T_2) = Y_t(T_1) - \Delta Y_t(T_2)$ is the yield spread and $\tau_i = T_i - t$, $i = 1, 2$, is the time to maturity. Jiang (1998) then proposes to estimate λ by choosing two representative bonds, fit a diffusion model of the type (6.5) to their yields using nonparametric methods, and then plug these estimators into (6.6).

6.4 The Data

The data set consists of 5505 daily observations from June 1, 1973 to February 25, 1995 of the 7-day Eurodollar rate. Eurodollars are any dollar denominated deposit in commercial banks outside of the U.S. Eurodollar accounts are not transferable but banks can lend on the basis of the Eurodollar accounts they hold. The interest rate charged for Eurodollar loans is often based upon the London Interbank Offer Rate (LIBOR). The Eurodollar rate is considered the benchmark interest rate for corporate funding, and Eurodollar futures are by far the most actively traded interest-rate product. A more detailed account of the Eurodollar market is found in Burghardt (2003).

The 7-day rate should be a reasonable good proxy for the short-term interest rate.[1] We do not use a lower maturity since this might lead to various market micro structure problems. We shall not attempt to remove any seasonal effects, such as weekend effects, from the data, and simply treat Monday as the first day after Friday such that we have 252 observations per year. We measure time in days such that the time between observations, $\Delta = 1/252$.

A number of single-factor models have been fitted to this particular data set. Aït-Sahalia (1996a) fitted a semiparametric model with linear drift and unspecified diffusion term to the 7-days Eurodollar rate and found strong nonlinearities in the diffusion term. He compared his semiparametric model to the CIR and Vasicek-model (which both are nested within his semiparametric model), both in terms of the actual model fit but also the resulting bond and option prices. He concluded that the two parametric model were significantly different from his semiparametric specifications, both in terms of model estimates and implied prices.

Aït-Sahalia (1996b) reinvestigated the data set, setting up a nonparametric specification test which allowed him to test any parametric (stationary) diffusion model against a nonparametric alternative. He tested a number of parametric single-factor models; all

[1] Chapman et al (1999) find that using the 7-day rate as a proxy when fitting standard parametric short-term interest rate model does apparently not lead to significantly different implied bond prices.

of these were rejected by his test. As an answer to the failure of the standard models, Aït-Sahalia (1996b) then proposed the following highly flexible, parametric single-factor model,

$$dr_t = \{\beta_0 + \beta_1 r_t + \beta_2 r_t^2 + \beta_3 r_t^{-1}\} \, dt + \sqrt{\sigma_0 + \sigma_1 r_t + \sigma_2 r_t^{\gamma}} dW_t. \tag{6.7}$$

The drift term is parameterised such that the drift can be nearly zero in a large part of the domain, but still ensure mean reversion in the tails. The specification test accepted this new model as giving an adequate description of the data. The conclusions in Aït-Sahalia (1996b) have been questioned in a number of later studies however due to poor finite sample performance of his proposed test. Chapman and Pearson (2000) and Pritsker (1998) presented evidence of that the asymptotic distribution of the test statistic delivers a very poor approximation of the finite sample distribution in the presence of strong serial correlation. And this is exactly the case with the interest rate data set used in Aït-Sahalia (1996b). In particular, Pritsker (1998) demonstrated in a simulation study that the test statistic is prone to reject correctly specified model when using the asymptotic critical values.

Hong and Li (2002) and Thompson (2000) have suggested alternative specification tests which should exhibit improved finite sample properties compared to the one proposed by Aït-Sahalia (1996b). The main idea in both studies is to apply a transformation of the data which should decrease the serial correlation. Hong and Li (2002) applied their test to the same data set as used in Aït-Sahalia (1996b), and still rejected all the standard parametric models examined in Aït-Sahalia (1996b), but also the model in (6.7). They argued that the data exhibits strong non-Markovian behaviour, and that the class of single-factor (Markov) models is too restrictive in this sense. The application of the specification test in Thompson (2000) to the Eurodollar 7-days rate data set also lead to the rejection of all the parametric models, including the one in (6.7). His explanation for the failure of the parametric models differs from the one of Hong and Li (2002) however. He argued that the problem is that the driving noise process is misspecified. By using either jump- or gamma-processes instead of the Brownian motion, he was able to accept a relatively simple parametric model.

In Bandi (2002), the kernel estimators of Bandi and Phillips (2003) were employed to fit the model (6.1) nonparametrically to the Eurodollar data set. Nonlinearities were present in both the estimated drift and diffusion term. In particular, the drift estimate was nearly zero in a major part of the data domain, but exhibited mean-reversion in its right tail while the behaviour in the left tail was inconclusive. The kernel estimator is robust to departures from the stationarity assumption normally imposed in single-factor models. There is no clear-cut evidence of non-stationarity in the data however.

In a fully parametric framework, a number of studies have reexamined the model (6.7). Elerian et al (2001) fitted the model to the same Eurodollar data set in a Bayesian framework using simulated maximum-likelihood techniques. They found that the parameters σ_2 and γ in (6.7) were difficult to identify in the data, and instead proposed the following slightly different parameterisation,

$$dr_t = \{\beta_0 + \beta_1 r_t + \beta_2 r_t^2 + \beta_3 r_t^{-1}\} \, dt + \sqrt{\sigma_0 + \sigma_1 r_t + \sigma_2 r_t^2 + \sigma_3 r_t^3} dW_t. \tag{6.8}$$

While the parameter estimates were easier to pin down, this new model did not appear to fit the data very well, and they found that any significant mean reversion only appeared in the estimated drift when this was assumed in the prior.[2] Elerian et al (2001) concluded that a single-factor model was not a very appropriate description of the data, and conjectured that a stochastic volatility model of the type in Andersen and Lund (1997) would be needed. This conjecture was confirmed in Durham (2002) where the model (6.7) was compared with a stochastic volatility model using simulated maximum-likelihood methods. He found that the single-factor model fit the data poorly, while the stochastic volatility (2-factor) model on the other hand did a good job. Similarly, Hurn and Lindsay (2002) found that the parameters in both the drift and diffusion term of (6.7) were difficult to identify and suggested the use orthogonal polynomials to amend this problem. They estimated the model using discrete time approximations however which means that their estimates are very likely to suffer from discretisation bias.

The general conclusion to be drawn form the empirical studies seems to be that the Eurodollar short-term rate is difficult to model properly within a single-factor framework and that the use of multi-factor models improve on the fit. If one restricts attention to single-factor models, inconclusive results have been obtained about the degree of (non-linear) mean reversion. Some studies have found evidence of nonlinear mean-reversion while others have rejected this hypothesis. Moreover, the more advanced parametric models seems to suffer from poor identification of the parameters. In the following, we shall re-examine the single-factor models using the above semiparametric diffusion model as a starting point.

The raw data is plotted in levels and differences in Figure 6.1 and 6.2 respectively. Figure 6.1 shows that the data exhibits a very strong correlation over time as is usually found in interest rate data. Taking differences, we see in Figure 6.2 that the short-term interest rate behaves as heteroskedastic white noise, which indicates that it is close to being a random walk. One should also notice the significant different behaviour of the rate in the period 1979-1981. The significant break in the data set in this period is due to the so-called Fed-Experiment where the U.S. Federal Reserve targeted monetary aggregates instead of, as done before and after, interest rate levels. One may argue that the data from this period should be left out or a Markov-switching model should be used; the latter alternative is pursued in Ang and Bekaert (2002). We shall in the following try to fit a model both to the full sample and the subsample 1982-1995, the latter excluding the period of the Fed-Experiment.

In Table 6.2, descriptive statistics of the data set is presented for the full period and the subperiod 1982-1995. Notably is the very strong autocorrelation in levels, while the correlation in the differenced data is decreasing fairly quickly. This is the case for both the full sample and the subsample, but it is less pronounced in the full sample due to the Fed-Experiment. In a linear modelling framework, the persistent autocorrelation in the levels would lead one to conclude that the process is non-stationary; the resulting estimate of the drift term is insignificantly different from zero such that the interest rate would be

[2] Similar findings are reported in Jones (2003)

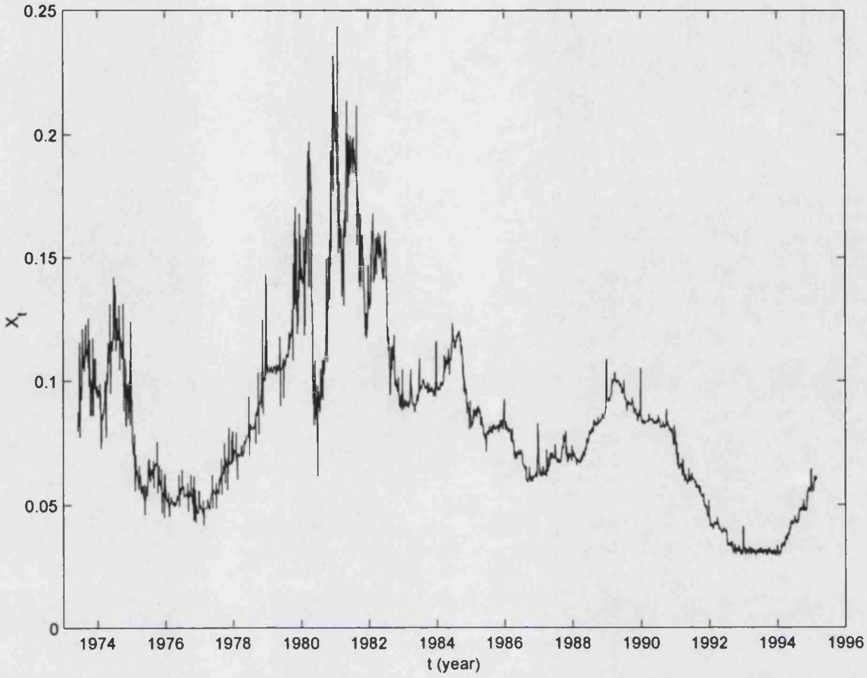

FIGURE 6.1. The Eurodollar spot rate in levels, 1973-1995.

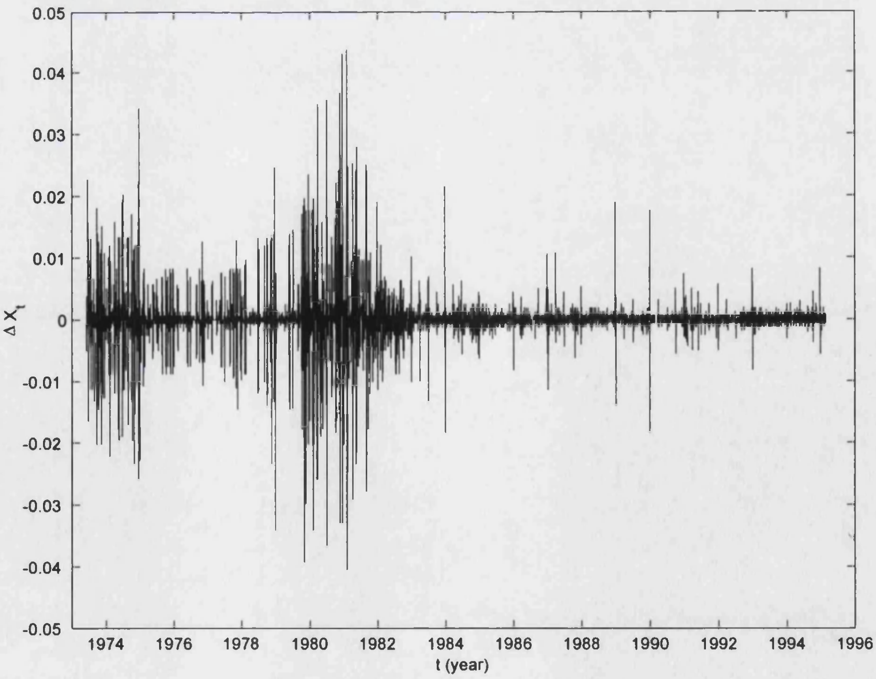

FIGURE 6.2. The Eurodollar spot rate in differences, 1973-1995.

deemed to be driven by a random walk and thereby being non-stationary. Moreover, all the descriptive statistics of the full sample are significantly different from the ones of the subsample. This confirms that the period of the Fed-Experiment lead to significantly different behaviour of the interest rate.

TABLE 6.2. Data descriptives.

		Mean	SD	Skewness	Kurtosis	ρ_1	ρ_2	ρ_3	ρ_4
1973-1995	r	0.0836	0.0359	0.9863	4.2022	0.9549	0.9093	0.8674	0.8329
	Δr	-3.453 10^{-6}	0.0041	-0.1238	30.8529	-0.2671	-0.0393	-0.0369	0.0319
1982-1995	r	0.0714	0.0240	-0.2448	2.1148	0.9802	0.9648	0.9458	0.9167
	Δr	-2.244 10^{-5}	0.0015	0.4253	54.8300	-0.0483	-0.0440	-0.0561	-0.0916

Notes: ρ_i denotes the correlation coefficient of order i. The reported autocorrelations for r are monthly while the autocorrelations for Δr are daily.

We now investigate further the seemingly nonstationary behaviour of the data in a linear framework. This is done by unit root tests: We set up a standard AR model, $\Delta r_i = \alpha_0 + (\phi - 1) r_{i-1} + \sum_{k=1}^{20} \alpha_k \Delta r_{i-k} + \varepsilon_i$, estimate the parameters by least squares, and then perform the Augmented Dickey-Fuller (ADF) test as outlined in Said and Dickey (1984). We also implement the $Z(t)$-test as proposed in Phillips (1987) using the model $r_i = \alpha_0 + \phi r_{i-1} + \varepsilon_i$; this should have less distortions in the presence of MA(1)-errors (see Phillips and Perron, 1988). The results are reported in Table 6.3 where we reject the hypothesis of a unit root for large negative values of the test statistic. The ADF test leads to non-conclusive results in the full sample with rejecting a unit root at a 10% level while accepting the hypothesis at a 5% level. The $Z(t)$ for the full sample on the other hand clearly rejects the hypothesis on a 5% level. In the subsample, both tests clearly accept the unit root hypothesis. So in a linear framework there is mixed evidence of non-stationarity in the full sample, while the subsample appear to be driven by a random walk. But if the drift is non-linear the above regression model is misspecified, and the estimation and test results invalid. A non-linear drift and diffusion term may lead to different results. As we shall see in the next section, the estimated drift and diffusion term generate processes with seemingly non-stationary behaviour in a major part of its domain with the process as a whole being stationary.

TABLE 6.3. Unit root test results.

		$\hat{\phi}$	Test Statistic	5% critical value	10% critical value
1973-1995	ADF test	0.9970	-2.60	-2.87	-2.59
	$Z(t)$ test	0.9935	-3.27	-2.87	-2.59
1982-1995	ADF test	0.9991	-1.47	-2.87	-2.59
	$Z(t)$ test	0.9963	-2.98	-2.87	-2.59

6.5 Empirical Results

6.5.1 Estimation of the Single-Factor Model

In Figure 6.3, we report the nonparametric kernel estimate of the marginal density as given in (4.8) for the data, using both the full sample and the subsample. A Gaussian kernel was used while the bandwidth was chosen by cross validation. The density estimate

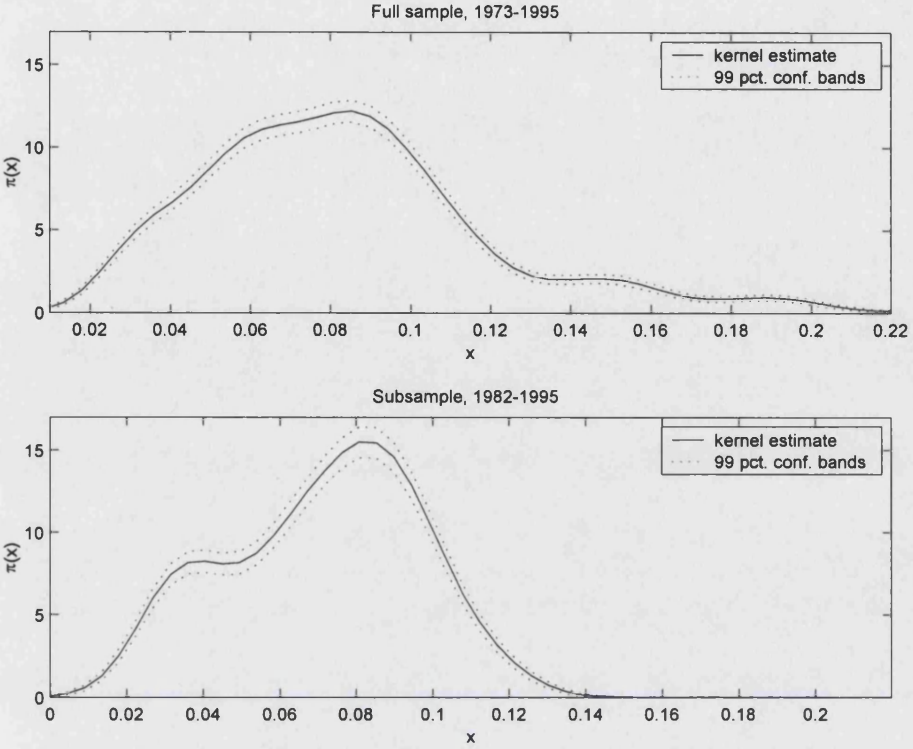

FIGURE 6.3. Kernel estimates of the stationary density, π, for the full sample and the subsample.

has a highly non-Gaussian shape for both samples. In the full sample, the density has a very long right tail due to the Fed-Experiment during which very high levels of interest rates were observed. Excluding these, one is left with the density in the second plot which is bimodal; a slightly smaller choice of bandwidth will give a trimodal shape of the density with the last mode being around 0.12. If $\{r_t\}$ is stationary, this indicates that contrary to what most of the models in Table 6.1 would suggest, the interest rate here evolves around not just one but two-three steady states. This is a strong indication of nonlinearities in the drift and the diffusion term.

We now report the estimates for our semiparametric model. As a benchmark, we also fit the model in (6.7) to the data set. Both models are estimated using the approximate log-likelihood suggested by Aït-Sahalia (2002) with order of approximation $M = 6$. For the semiparametric model, the bandwidths are chosen as described below, and we trim the data at the 1st and 99th empirical percentile. In the fully parametric model we had problems obtaining a precise estimate since the likelihood curve is relatively flat in the vicinity of the optimum. This was particularly a problem along the dimensions of σ_2 and γ, and was very pronounced when fitting the model to the subsample 1982-1995. By closer examination, we found that a wide range of parameter values generated very much the same shape of $\sigma^2(x;\theta)$ in the domain $x \in [0.02, 0.14]$ while for $x > 0.14$ different parameter choices lead to significantly different behaviour of the diffusion term. Since we do not have observations greater than 0.14 in the subsample, this is evidently a problem. However, this identification problem did not occur when estimating the semiparametric model; this is probably due to the fact that the diffusion parameters in the semiparametric estimation procedure now also enters the drift function, and this allows us to pin them down more precisely. This identification problem has also been reported in Elerian et al (2001) and Hurn and Lindsay (2002). The latter suggested to reparameterise both the drift and diffusion term using orthogonal polynomials to circumvent this problem; we plan to do this in a later version of this paper.

The resulting estimates of $\theta = (\sigma_0, \sigma_1, \sigma_2, \gamma)$ and (for the parametric model) $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ for both the full sample and the subsample are reported in Table 6.4 with associated standard errors. First, in both samples, we see that the parametric and semiparametric estimates are fairly close, but still significantly different at a 5%-level from each other. We also observe that the estimates of the fully parametric model satisfy the stationarity conditions imposed in Aït-Sahalia (1996b). The standard errors associated with σ_2 and γ are relatively large; most likely this owes to the aforementioned problem with the flat likelihood-curve. The estimates for the two different samples are not significantly different from each other; so for the fully parametric model, the Fed-Experiment does not influence the estimates.

TABLE 6.4. Estimates of θ and β.

	1973-1995		1982-1995	
	Parametric	Semiparametric	Parametric	Semiparametric
β_0	-0.3054	-	-0.3106	-
	(0.1590)	-	(0.1484)	-
β_1	5.5729	-	5.6579	-
	(1.2301)	-	(1.1872)	-
β_2	-30.9985	-	-30.4436	-
	(2.0307)	-	(3.6436)	-
β_3	0.0047	-	0.0053	-
	(0.0011)	-	(0.0019)	-
σ_0	-0.0006	0.0120	-0.0006	0.0110
	(0.0005)	(0.0009)	(0.0004)	(0.0011)
σ_1	0.0210	0.0213	0.0221	0.0123
	(0.0090)	(0.0108)	(0.0122)	(0.0134)
σ_2	30.8413	28.0874	32.5880	23.0874
	(7.8007)	(7.7745)	(8.7490)	(8.6923)
γ	4.2573	4.3019	4.2166	4.0019
	(1.5422)	(1.5361)	(2.7491)	(2.1923)

Notes: Standard errors are reported in parentheses. In the parametric model these were estimated by standard covariance estimation methods. In the semiparametric model the estimator proposed in Theorem 7 was used

The semiparametric estimates proved to be fairly robust over a range of bandwidth choices. An initial set of bandwidths was chosen by using standard cross-validation methods. We then generated a grid of bandwidth centered around this initial one. For each set of bandwidths, we obtained an estimate of θ. In Table 6.5, the results of this sensitivity check for the subsample are reported for seven different bandwidth sets. The bandwidths decrease as one moves from left to right in the table with the 7th bandwidth being the cross-validated one. Relatively large bandwidths choices, slightly bigger than the initial cross-validated ones, gave the most reliable estimates for our sample. This probably stems from the fact that the cross-validation procedure does not take into account the dependence in the data which is very strong in our case.

A plot of the semiparametric estimates of σ^2 for the first four set of bandwidths can be found in Figure 6.4. The estimates do not vary a great deal across the different bandwidths

TABLE 6.5. Sensitivity check of semiparametric estimates.

Bandwidth	1	2	3	4	5	6	7
σ_0	0.0112	0.0110	0.0113	0.0110	0.0104	0.0103	0.0109
σ_1	0.0129	0.0124	0.0127	0.0131	0.0120	0.0121	0.0116
σ_2	21.5643	24.1624	23.1004	22.3427	24.0175	23.3548	21.9449
γ	4.2699	4.2598	4.2978	4.2296	3.8650	3.7434	3.9873

Notes: Bandwidth 1 ($10^{-3}\times$): h_1= 10.09, h_2= 12.23, h_3= 13.89, h_4= 15.10, h_5= 17.22, h_6= 18.74, h_7= 19.97, h_8= 21.11.
Bandwidth 7 ($10^{-3}\times$): h_1= 5.17, h_2= 6.33, h_3= 6.97, h_4= 7.49, h_5= 7.85, h_6= 8.84, h_7= 9.60, h_8= 10.36.

which indicates that our estimation procedure is fairly robust. Similar results were obtained for the full sample.

Next, we report our nonparametric estimates of the drift function. Here, we use slightly higher bandwidths compared to the ones we used in the estimation of θ. In Figure 6.5, the nonparametric estimator of μ is plotted with pointwise 95%-confidence bands for both the full sample and the subsample. The confidence bands are calculated using Theorem 1. The range over which we plot the estimates was chosen as the one of the data from the subsample. We see that for both periods, the nonparametric estimate of μ is insignificantly different from zero in the range between 0.03 and 0.12; this is consistent with our earlier observation that locally the short-term interest rate behaves as a random walk. But for values of r less than 0.035 and greater than 0.12, $\hat{\mu}(r)$ is significantly different from zero implying a mean reverting effect; in the interval $[0.03, 0.12]$, $\{r_t\}$ is allowed to evolve basically as a random walk but if it leaves this part of its domain and takes on too small or great values it is pulled back again. So in a global sense $\{r_t\}$ can be seen as being stationary. While the drift estimates for the two periods have the same shape, they are still markedly different in the domain we have chosen here. In particular, in the full sample, there is strong mean reversion in the left tail while no significant such appears in the right tail. In contrast, the drift estimate in the smaller sample has relatively weak mean reversion in the left tail, but a strong one in the right tail. The "missing" mean reversion of the drift estimate in the full sample is however simply due to the fact that the domains within which we have observed the interest rate in the full sample and the subsample are not the same, and we are only able to compare the estimates within the domain of the subsample. Mean reversion for the full sample first appears further out in its right tail, cf. figure 6.6. In total however the two drift estimates are not significantly different from each other in the domain of the data in the subsample.

The results found here are compatible with other empirical studies of the short-term interest rate. Jiang and Knight (1997) and Bandi (2002) obtain nonparametric kernel estimates of μ that exhibit a similar behaviour. In a discrete time framework, parametric Markov switching AR-models have proved to be able to generate the same type of dynamics as the ones we have found here. For a recent application of this type of models to the short term interest rate see Ang and Bekaert (2002).

One should however be careful with the tailbehaviour of the nonparametric estimates, which may be an artifact of the use of kernel estimators. These are known for not being precise in the tails and outside of the support of the data as a combination of their local nature and the sparsity of data there. For the subsample, the data support is within

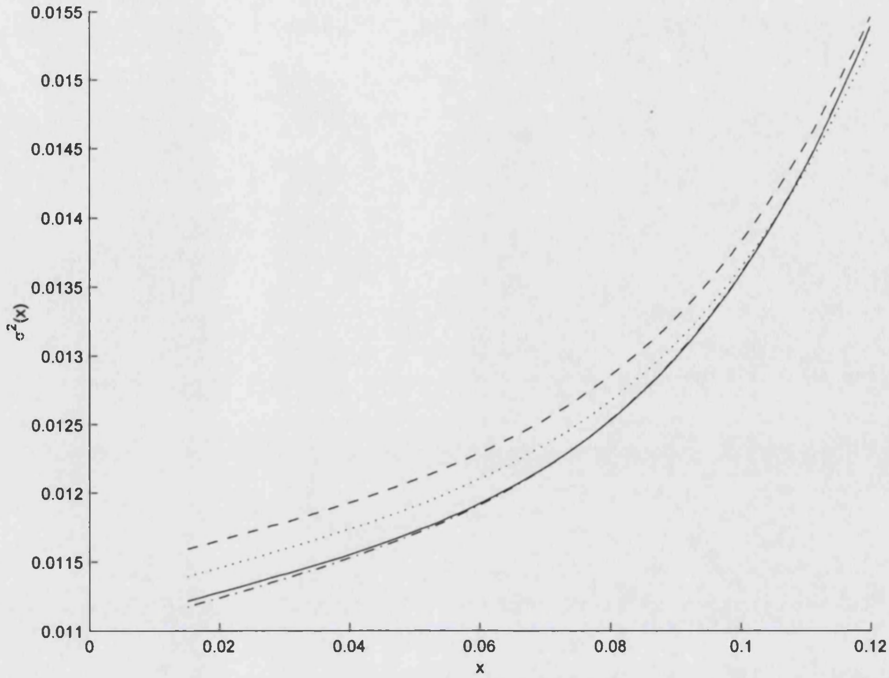

FIGURE 6.4. Sensitivity check: Semiparametric estimates of $\sigma^2(r;\theta)$ for different bandwidths.

$[0.03, 0.12]$ so one may question the quality of the local estimates of $\mu(x)$ outside of this interval, where it may be prone to a certain degree of erratic behaviour. Chapman and Pearson (2000) report in a simulation study that this is the case when using the nonparametric kernel estimator of Jiang and Knight (1997). Another issue related to the behaviour at the left tail is that the domain in our case is $I = (0, \infty)$. It is by now recognised that kernel densities estimates based on symmetric kernels may perform poorly near the boundary of the support, which may be another source of bias in our case near 0^+. Recall from the simulation study that $\hat{\mu}(x)$ did not perform well near the lower boundary for the CIR-model; one could suspect the same to be the case here. Observe however that in order for the short-term interest rate to remain positive, the drift has to be positive near zero. So at least our estimates have the right sign. Bandi (2002), when applying the nonparametric kernel estimator of Bandi and Phillips (2003), reports similar estimates of the drift function in the right tail of the data support. He however only reports estimates within the data support and is not able to conclude what the drift looks like close to 0^+. The tailbehaviour may also be a result of us imposing the assumption of stationarity on the process as argued by Jones (2003). In conclusion, the estimates for $x \notin [0.03, 0.12]$ should be interpreted with care. We plan to carry out further monte Carlo studies of these issues, and also apply bootstrapping when calculating the confidence bands.

In Figure 6.6, we compare the nonparametric estimator of the drift to the parametric estimator in the full sample. The nonparametric drift estimator is very wiggly in the right tail, and one might want to use a varying bandwidth to remove this effect. Comparing the parametric and nonparametric estimates, the most striking feature is the very strong mean reversion found in the nonparametric ones. We see that the parametric model mimics the nonparametric estimate pretty well in the major part of the domain, but has problems capturing the curvature of the drift term out in the tails. Similarly, when comparing the

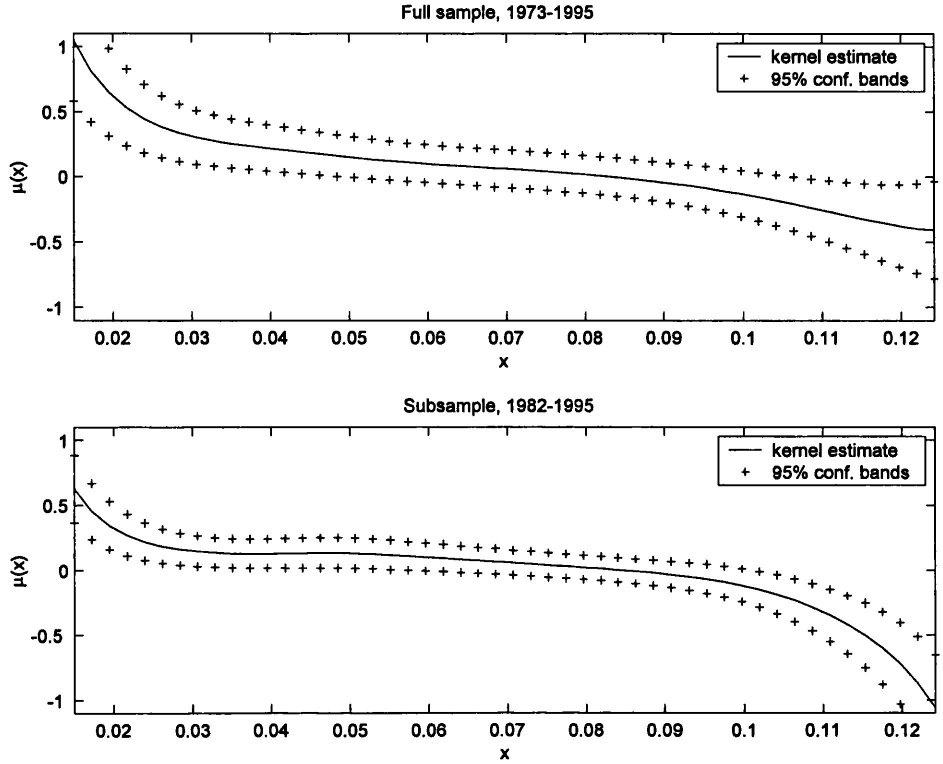

FIGURE 6.5. Nonparametric estimate of μ with 95% confidence bands.

two drift estimators for the subsample which can be found in Figure 6.7. Observe that in the subsample, the nonparametric drift estimate only crosses zero once (at $x = 0.08$, i.e. there is one steady-state for this model) while in the full sample, it reaches zero twice (at $x = 0.09$ and 0.19, i.e. there are two steady-states). This is caused by the Fed-Experiment. The parametric drift estimates are very close to zero, in particular in the subsample. This gives support to the earlier finding that the short rate may be modelled as a heteroschedastic random walk since the process $dr_t = \sigma(r_t)\, dW_t$ is the continuous-time equivalent of a such.

We test the parametric specification against the semiparametric alternative to see whether the latter supplies us with an adequate description of data. We do this using the test statistic proposed in Theorem 3 in Chapter 4. In both samples, we choose the number of points $d = 40$, and x_i as the $i/41 \times 100$th empirical percentile of the sample in question. The realised value of the test statistic is $T_n = 280.65$ and 552.96 for the full sample and the subsample respectively, while the critical value at a 1%-level is 63.6907. So we clearly reject the hypothesis that the parameterisation of the drift considered here is appropriate.[3] It could now be of interest to set up a parsimonious parametric model which was able to generate the same behaviour of the drift as our semiparametric estimate does. This is left for future work.

For both samples, the estimates for the parametric model are both qualitatively and quantitatively very different from the ones reported in Aït-Sahalia (1996b). In particu-

[3]The distribution of test statistics for parametric vs. nonparameteric alternatives are known to be poorly approximated by their asymptotic distribution. So it would probably be more appropriate to perform bootstrap here, see e.g. Fan (1994, 1995). This is left for future research.

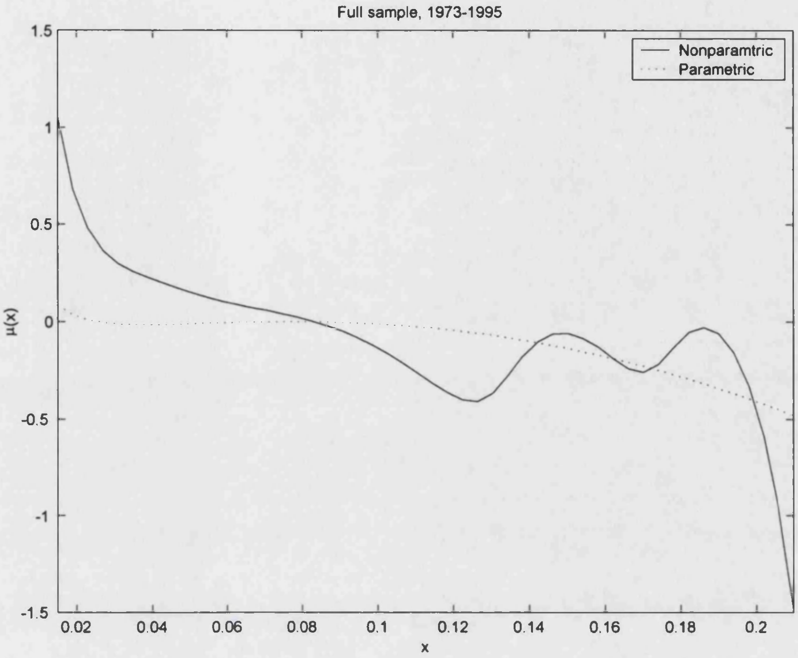

FIGURE 6.6. Comparison of nonparametric and parametric estimate of μ for the full sample, 1973-1995.

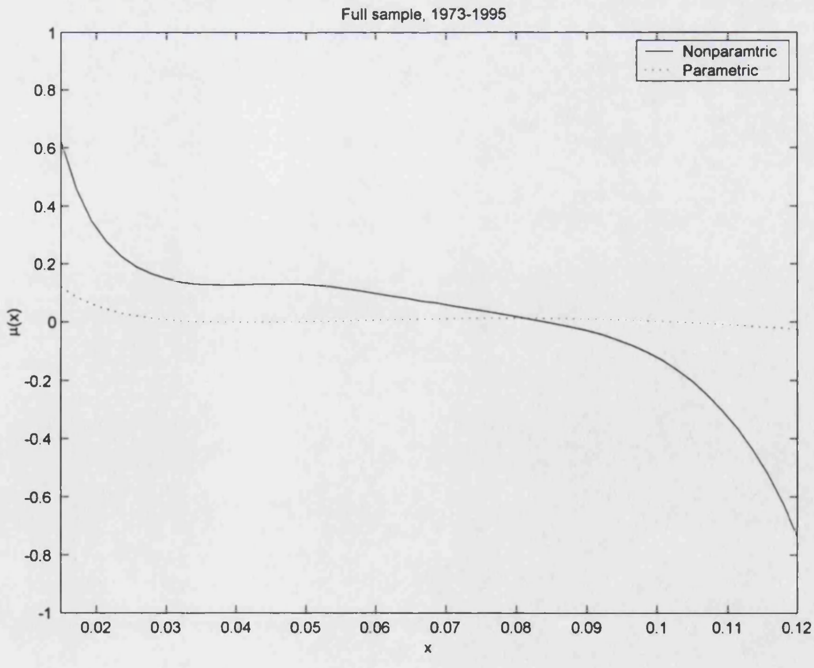

FIGURE 6.7. Comparison of nonparametric and parametric estimate of μ for the subsample, 1982-1995.

lar, while the estimated drift in Aït-Sahalia (1996b) predicts two turning points (steady-states), we here only have one. The second steady-state is partially reinforced in the semiparametric model fitted to the full sample. The estimated diffusion term in Aït-Sahalia (1996b) has a smile while ours is monotonously increasing. Hurn and Lindsay (2002) report qualitatively similar estimates to ours with their drift and diffusion term having the same shape as ours, but on a much smaller scale. Our estimates are fairly close to the ones obtained in Durham (2002) though.

6.5.2 Implied Bond and Derivative Prices

Given the calibrated models obtained in the previous section, we now wish to see what implications the competing models will have on bond and interest rate derivative prices. We compute bond prices implied by the competing models to see if they generate significantly different prices. This is only done for the models fitted using data in the period 1982-1995.

In order to do this, we first need to estimate the risk premium as discussed in Section 3. We use the general estimation procedure where the risk premium is chosen to minimise the squared difference between implied and observed bond prices. Here, we follow Aït-Sahalia (1996) and Vasicek (1977) amongst others and assume that the risk premium is constant over time. This facilitates the estimation since we then only have one parameter to optimise with respect to. We estimate this parameter, λ, using the least squares method presented in Section 3: Given the estimates of θ and the drift function and any value of λ, we calculate a set of implied bond prices and compare them to the observed prices. The data set used in the calibration consists of daily observations of 1-, 3- and 6-month Eurodollar bond prices, $(B_{i\Delta}(\tau_j))$, and the daily observations of the short-term interest rate, $r_{i\Delta}$, $i = 1, ..., n$, $j = 1, 2, 3$, where $\tau_1 = 1/12$, $\tau_1 = 1/4$ and $\tau_3 = 1/2$ are the times to maturities. We assume the following model for the observed bond prices,

$$B_{i\Delta}(\tau_j) = \Pi_B(r_i, \tau_j; \lambda) + \varepsilon_{ij}, \quad E^P[\varepsilon_{ij}] = 0, E^P\left[\varepsilon_{ij}^2\right] = \sigma_\varepsilon^2$$

where

$$\Pi_B(r, \tau; \lambda) = E^Q\left[\exp\left[-\int_0^\tau r_s ds\right] |r_0 = r\right]$$

and $\{r_t\}$ has dynamics (6.4) with $\lambda(r) = \lambda$. The estimate of λ is the obtained by running non-linear least squares on the above regression model. The (approximated) implied bond prices, $\Pi_B(r, \tau)$, are obtained by Monte Carlo simulation of the short term interest rate process under the risk-neutral measure.

Here, we use zero-coupon prices from the period December 1, 1994 to February 25, 1995 in our calibration of λ. This gives us $n = 154$ observations; we have actually a much larger data set of bond prices available which could be used, but the Monte Carlo simulations are fairly time-consuming so we here choose to only use a small proportion of this. In Table 6.6, the estimated market price of risk premiums are reported when one uses the semiparametric and parametric fit respectively. In Chapter 5, Section 4 we demonstrated that under weak regularity conditions,

$$\sqrt{n}(\hat{\lambda} - \lambda_0) \to^d N\left(0, \sigma_\varepsilon^2 H^{-1}(\lambda_0)\right),$$

where

$$H(\lambda) \equiv E_\theta \left[\dot{\Pi}_B (r_{i\Delta}; \lambda)^\top \dot{\Pi}_B (r_{i\Delta}; \lambda) \right], \quad \Pi_B (r; \lambda) = \{\Pi_B (r, \tau_j)\}_{j=1}^3 .$$

The standard errors were calculated using (5.26). They were not adjusted for the fact that the drift and diffusion term came from a preliminary estimation step. That is, reported standard errors of $\hat{\lambda}$ are calculated conditional on the estimated drift and diffusion term.

TABLE 6.6. Estimate of the market price of risk, λ.

	Semiparametric	Parametric
Estimate of λ	-8.5494 10^{-2}	-1.0596 10^{-1}
Standard Error	(2.0603 10^{-2})	(3.0651 10^{-2})

Notes: Standard errors are reported in parentheses. For both the parametric and semiparametric model (5.26) was used to calculate these.

We are now able to calculate bond prices under the risk neutral measure. In Table 6.7, we report the implied bond prices given either the semiparametric or the parametric fit of the short-term interest rate. We report prices for four different times to maturities, 0.5, 1, 5 and 10 years, and 4 different levels of the current short rate. In general as the maturity increases, the prices difference between the two competing models increase. For most maturities, the prices implied by the parametric model fall outside one or more standard deviations of the semiparametric prices. So we get significantly different prices when applying the semiparametric model compared to the parametric one.

TABLE 6.7. Implied bond prices of the semiparametric and parametric model.

	Short rate level			
Maturity (years)	0.04	0.06	0.08	0.10
0.5	97.2866	99.1478	96.1416	94.2130
	(0.1054)	(0.0868)	(0.0883)	(0.0697)
	97.1404	98.3368	96.5284	92.7327
1	94.9322	93.0204	91.0659	88.3761
	(0.1147)	(0.0906)	(0.0835)	(0.0657)
	92.2625	92.6393	91.2850	91.0595
5	65.6758	65.2582	62.6428	60.8854
	(0.1131)	(0.1078)	(0.0995)	(0.0656)
	65.3361	63.6363	62.7447	60.7810
10	41.7999	40.0005	39.1834	38.7795
	(0.0916)	(0.0803)	(0.0721)	(0.0681)
	42.9604	40.5908	40.0661	39.0452

Notes: (i) All prices correspond to a face value of the bond equal to $100. Each cell has three elements: The first and third are the implied prices of the semiparametric and parametric model respectively; the second the associated standard error of the semiparametric model.
(ii) The s.e.'s were calculated by using the estimator outlined in Section 5.3.2.

6.6 Conclusion

We have proposed a new semiparametric diffusion model for the short-term interest rate which contains most parametric models found in the term structure literature. We estimated the semiparametric model using a data set of daily observations of the 7-day Eurodollar rate in the period 1973-1995. For comparison, we also fitted the parametric model proposed by Aït-Sahalia (1996b) to the data, this being the most flexible parametric single-factor model found in the literature. Due to so-called Fed-Experiment in 1979-1981, there is a break in the data in this period compared to before and after. So we estimated the models using both the full sample and the subsample 1982-1995. Estimating the models using either of the two samples lead to markedly different results. The two sets of estimates were however not statistically different from each other. If one wishes to model the full period, one may gain from either including dummies or using a Markov-switching model.

Both the drift and diffusion term in the fitted semiparametric model exhibited a nonlinear behaviour which most of the parametric models in the literature cannot generate. The nonparametric estimate of the drift were significantly different from the one of the most flexible parametric model in both the full sample and the subsample. In both samples, this was due to a much stronger degree of mean reversion in the nonparametric estimates. The parametric form was not able to allow for a mean reversion in the tails of this type, and could therefore not mimic the shape of the nonparametric drift properly there. We tested the parametric model against the semiparametric alternative and rejected it in favour of the semiparametric one. The implications on the pricing of bonds were also examined. We found that the implied bond prices predicted by the two models were significantly different from each other. So not only do the models differ on a basic level, but they will also lead to different prices. This is particularly important for market practitioners.

As a next step, it would be interesting to develop further statistical tests to evaluate the performance of the model when using it for pricing derivatives. Such would be a useful tool in model selection and evaluation, helping the practitioner to choose a parsimonious model without misspecifications, and which at the same time performs well in a pricing scenario. Also, we plan to set up a fully parametric model which is consistent with the semiparametric estimates.

How to extend our approach to cover semiparametric multifactor models is not obvious since our estimator is not easily extended to multivariate diffusion models. If one is ready to restrict one's attention to a smaller class of multifactor models however, it should be possible to adapt the approach used here to this more general case, cf. Chen et al (2000b).

References

Ahn, D.-H., R.F. Dittmar & A.R. Gallant (2002) Quadratic Term Structure Models: Theory and Evidence. *Review of Financial Studies* 15, 243-288.

Ahn, D.-H. & B. Gao (1999) A Parametric Nonlinear Model of Term Structure Dynamics. *Review of Financial Studies* 12, 721-762.

Ai, C. (1997) A Semiparametric Maximum Likelihood Estimator. *Econometrica* 65, 933-963.

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. (1996b) Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies* 9, 385-426.

Aït-Sahalia, Y. (1999) Transition Densities for Interest Rates and Other Nonlinear Diffusions. *Journal of Finance* 54, 499-547.

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

Aït-Sahalia, Y. (2003) Closed-Form Likelihood Expansions for Multivariate Diffusions. Working paper, Princeton University.

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

Aït-Sahalia, Y. & A.W. Lo (1998) Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *Journal of Finance* 53, 499-547.

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

Andersen, T.G. & J. Lund (1997) Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate. *Journal of Econometrics* 77, 343-377.

Andrews, D.W.K. (1994) Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity. *Econometrica* 62, 43-72.

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

Ang, A. & G. Bekaert (2002) Short Rate Nonlinearities and Regime Switches. *Journal of Economic Dynamics and Control* 26, 1243-1276.

Arcones, M.A. (1995) On the Central Limit Theorem for U-Statistics under Absolute Regularity. *Statistics and Probability Letters* 24, 245-249.

Backus, D., Foresi, S. & Zin, S. (1998) Arbitrage Opportunities in Arbitrage-Free Models of Bond Pricing. *Journal of Business and Economic Statistics* 16.

Bandi, F.M. (2002) Short-Term Interest Rate Dynamics: A Spatial Approach. *Journal of Financial Economics* 65, 73-110.

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.

Banon, G. (1978) Nonparametric Identification for Diffusion Processes. *Siam Journal of Control and Optimization* 16, 380-395.

Bergstrom, A.R. (1990) *Continuous Time Econometric Modelling.* Oxford: Oxford University Press.

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

Bickel, P.J., C.A.J. Klaassen, Y. Ritov & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* The John Hopkins University Press.

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.

Bondarenko, O. (2003) Estimation of Risk-Neutral Densities Using Positive Convolution Approximations. *Journal of Econometrics* 116, 85-112.

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

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

Brennan, M. & E. Schwartz (1979) A Continuous Time Approach to the Pricing of Bonds. *Journal of Banking and Finance* 3, 133-155.

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

Burghardt, G. (2003) *The Eurodollar Futures and Options Handbook.* New York: McGraw-Hill.

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. & S. Das (2002) Pricing Interest Rate Derivatives: A General Approach. *Review of Financial Studies* 15,195-242.

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

Chan, K.C., Karolyi, G.A., Longstaff, F.A. & Sanders, A.B. (1992) An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance* 47, 1209-1227.

Chapman, D., J.B. Long Jr & N. Pearson (1999) Using Proxies for the Short Rate: When Are Three Months Like an Instant? *Review of Financial Studies* 12, 763-806.

Chapman, D. & N. Pearson (2000) Is the Short Rate Drift Actually Nonlinear? *Journal of Finance* 55, 355-388.

Chapman, D. & N. Pearson (2001) Recent Advances in Estimating Term-Structure Models. *Financial Analysts Journal* 57, 77-95.

Chen, R.R. & L. Scott (1992) Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross models of the Short-Term Interest Rates. *Review of Financial Studies* 5, 613-636

Chen, R.R. & L. Scott (1993) Maximum Likelihood Estimation for a Multifactor General Equilibrium Model of the Term Structure of Interest Rates *Journal of Fixed Income* 3, 14–31.

Chen, X. (2004) Semiparametric and Nonparametric Estimation via the Method of Sieves. Forthcoming in *Handbook of Econometrics* vol. 6 (eds. J.J. Heckman & E.E. Leamer). Amsterdam: North-Holland

Chen, X., L.P. Hansen & M. Carrasco (1999) Nonlinearity and Temporal Dependence. Working paper, University of Rochester.

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.

Chen, X., O. Linton & I. Van Keilegom (2003) Estimation of Semiparametric Models when the Criterion Function is not Smooth. *Econometrica* 71, 1591-1608.

Chen, X. & X. Shen (1998) Sieve Extremum Estimates for Weakly Dependent Data. *Econometrica* 66, 289-314

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.

Cobb, L., P. Koppstein & N.H. Chen (1983) Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family. *Journal of the American Statistical Association* 78, 124-130.

Conley, T., L.P. Hansen, E. Luttmer & J. Scheinkman (1997) Short-term Interest Rates as Subordinated Diffusions. *Review of Financial Studies* 10, 525-577.

Constantinides, G.M. (1992) A Theory of the Nominal Term Structure of Interest Rates. *Review of Financial Studies* 5, 531-552.

Courtadon, G. (1982) The Pricing of Options on Default Free Bonds. *Journal of Financial and Quantitative Analysis* 17, 531-552.

Cox, J.C., J.E. Ingersoll & S. Ross (1985) A Theory of the Term Structure of Interest Rates. *Econometrica* 53, 373-384.

Dacunha-Castelle, D. & D. Florens-Zmirou (1986) Estimation of the Coefficients of a Diffusion from Discrete Observations. *Stochastics* 19, 263-284.

Dai, Q. & K.J. Singleton (2000) Specification Analysis of Affine Term Structure Models. *Journal of Finance* 55, 1943-1978.

Darolles, S. & C. Gouriéroux (2001) Truncated Dynamics and Estimation of Diffusion Equations. *Journal of Econometrics* 102, 1-22.

Doukhan, P., P. Massart & E. Rio (1994) The Central Limit Theorem for Strongly Mixing Processes. *Annales de l'Institut Henri Poincare, Series B, Probability and Statistics* 30, 63-82.

Doukhan, P., P. Massart & E. Rio (1995) Invariance Principles for Absolutely Regular Processes. *Annales de l'Institut Henri Poincare, Series B, Probability and Statistics* 31, 393-427.

Drost, F.C., C.A.J. Klaassen & B.J.M. Werker (1997). Adaptive Estimation in Time-Series Models. *Annals of Statistics* 25, 786–817.

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.

Dunis, C. & B. Zhou (eds.) (1998) *Nonlinear Modelling of High Frequency Financial Time Series*. New York: John Wiley & Sons.

Durham, G.B. (2002) Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate. Working paper, University of Iowa. Forthcoming in *Journal of Financial Economics.*

Durham, G.B. & R.A. Gallant (2002) Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *The Journal of Business and Economic Statistics* 20, 297-316.

Dybvig, P. (1989) Bond and Bond Option Pricing Based on the Current Term Structure. In *Mathematics of Derivative Securities* (eds. M.A.H. Dempster & S. Pliska). Cambridge University Press.

Elerian, O., S. Chib & 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.

Fan, J. & C. Zhang (2003) A Reexamination of Diffusion Estimators with Applications to Financial Model Validation. *Journal of the American Statistical Association* 98, 118-134.

Fan, Y. (1994) Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method. *Econometric Theory* 10, 316-356.

Fan, Y. (1995) Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test. *Econometric Reviews* 14, 367-382.

Florens-Zmirou, D. (1989) Approximate Discrete Time Schemes for Statistics of Diffusion Processes. *Statistics* 20, 547-557.

Florens-Zmirou, D. (1993) On Estimating the Diffusion Coefficient from Discrete Observations. *Journal of Applied Probability* 30, 790-804.

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 Applications*, vol. 1. New York: Academic Press.

Gallant, A.R. & J.R. Long (1997) Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square. *Biometrika* 84, 125-141.

Gallant, A.R. & G. Tauchen (1996) Which Moments to Match? *Econometric Theory* 12, 657-681.

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

Geman, S. (1979) On a Common Sense Estimator for the Drift of a Diffusion. Working paper, Brown University.

Geman, S. (1980) An Application of the Method of Sieves: Functional Estimation for the Drift of a Diffusion. *Colloq. Math. Soc. Janos Bolyai* 6, Nonparametric Statistical Inference, Budapest.

Genon-Catalot, V. (1990) Maximum Contrast Estimation for Diffusion Processes from Discrete Observations. *Statistics* 21, 99-116.

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

Gobet, E., M. Hoffmann & M. Reiß (2003) Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data. Working paper, Universit́es de Paris 6 & 7 - CNRS (UMR 7599). Forthcoming in *Annals of Statistics.*

Gouriéroux, C., A. Monfort & E. Renault (1993) Indirect Inference. *Journal of Applied Econometrics* 8, 85-118

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.

Hansen, L.P. & J.A. Scheinkman (1995) Back to the Future: Generating Moment Implications for Continuous Time Markov Processes. *Econometrica* 63, 767-804.

Hansen, L.P., J.A. Scheinkman & N. Touzi (1998) Spectral Methods for Identifying Scalar Diffusions. *Journal of Econometrics* 86, 1-32.

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.

Hong, Y. & Li, H. (2002) Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates. Working paper, Cornell University.

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.

Hurn, A. S. & K. A. Lindsay (2002) On the Specification of the Drift and Diffusion Functions for Continuous-time Models of the Spot Interest Rate. *Oxford Bulletin of Economics and Statistics* 64, 547-564.

Hurn, A.S., K.A. Lindsay & V.L. Martin (2003) On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential Equation. *Journal of Time Series Analysis* 24, 45-63.

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.

Jones, C.S. (2003) Nonlinear Mean Reversion in the Short-Term Interest Rate. *Review of Financial Studies* 16, 793-843.

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.

Karlsen, H. & D. Tjøstheim (2001) Nonparametric Estimation in Null Recurrent Time Series. *Annals of Statistics* 29, 372-416.

Kennedy, D.P. (1994) The Term Structure of Interest Rates as a Gaussian Random Field. *Mathematical Finance* 4, 247-258.

Kennedy, D.P. (1997) Characterizing Gaussian Models of the Term Structure of Interest Rates. *Mathematical Finance* 7, 107-18.

Kessler, M. & A. Rahbek (2001) Identification and Inference for Cointegrated and Ergodic Gaussian Diffussions. Statistical Inference for Stochastic Processes, X, x-x.

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

Kutoyants, Y.A. (2004) *Statistical Inference for Ergodic Diffusion Processes.* London: Springer-Verlag.

Liptser, R.S & A.N. Shiryaev (2001). *Statistics of Random Processes, Vol. 1: General Theory.* New York: Springer-Verlag.

Litterman, R. & J. Scheinkman (1991) Common Factors Affecting Bond Returns. *Journal of Fixed Income*, 3, 54-61.

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.

Longstaff, F.A. & E.S. Schwartz (1992) Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model. *Journal of Finance* 47, 1259-1282.

Marsh, T.A. & E.R. Rosenfeld (1983) Stochastic Processes for Interest Rates and Equilibrium Bond Prices. *Journal of Finance* 38, 635-646.

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. (1973) Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science* 4, 141-183.

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.

Murphy, S.A. & A.W. van der Vaart (1997) Semiparametric Likelihood Ratio Inference. *Annals of Statistics* 25, 1471–1509.

Murphy, S.A. & A.W. van der Vaart (2000) On Profile Likelihood. *Journal of the American Statistical Association* 95, 449-485.

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

Newey, W.K. (1990) Semiparametric Efficiency Bounds. *Journal of Applied Econometrics* 5, 99-135.

Newey, W.K. (1991) Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica* 59, 1161-1167.

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

Newey, W.K. (1994b) The Asymptotic Variance of Semiparametric Estimators. *Econometrica* 62, 1349-1362.

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. & D. McFadden (1994) Large Sample Estimation and Hypothesis Testing. In *Handbook of Econometrics* vol. IV, ch. 36. Amsterdam: North-Holland.

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

Nguyen, H.T. (1979) Density Estimation in a Continuous-Time Markov Process. *Annals of Statistics* 7, 341-348.

Nicolau, J. (2002) A New Technique for Simulating the Likelihood of Stochastic Differential Equations. *Econometrics Journal* 5, 91-103.

Nicolau, J. (2003) Bias Reduction in Nonparametric Diffusion Coefficient Estimation. *Econometric Theory* 19, 754-777.

Nicolau, J. (2004) Nonparametric Estimation of Second Order Differential Equations. Working paper, Universidade Técnica de Lisboa.

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics.* Cambridge: Cambridge University Press.

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.

Pham, T.D. (1981) Nonparametric Estimation of the Drift Coefficient in the Diffusion Equation. *Math. Operatonsfosch. Statist., Ser. Statistics* 1, 61-73.

Phillips, P.C.B. (1973) The Problem of Identification in Finite Parameter Continuous-Time Models. *Journal of Econometrics* 4, 351-362.

Phillips, P.C.B. (1987) Time Series Regression with a Unit Root. *Econometrica* 55, 277-301.

Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion Paper no. 1181, Yale University.

Phillips, P.C.B. & P. Perron (1988) Testing for a Unit Root in Time Series Regression. *Biometrika* 75, 335-346.

Politis, D.N, J.P. Romano & M. Wolf (1999): *Subsampling.* New York: Springer-Verlag.

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

Prakasa Rao, B.L.S. (1999) *Statistical Inference for Diffusion Type Processes.* Oxford: Oxford University Press.

Pritsker, M.G. (1998) Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models. *Review of Financial Studies* 11, 449–487.

Revuz, D. & M. Yor (1998) *Continuous Martingales and Brownian Motion*, 3rd Edition. New York: Springer-Verlag.

Robinson, P.M. (1983) Nonparametric Estimators for Time Series. *Journal of Time Series Analysis* 4, 185–297.

Robinson, P.M. (1988) Root-N-Consistent Semiparametric Regression. *Econometrica* 56, 931-954.

Robinson, P.M. (1989) Hypothesis Testing in Nonparametric and Semiparametric Models for Econometric Time Series. *Review of Economic Studies* 56, 511-534.

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

Rogers, L. (1995) Which Model For Term-Structure of Interest Rates Should One Use?. In *Mathematical Finance* (eds. M.H.A. Davis, D. Duffie, W.H. Fleming and S.E. Shreve). New York: Springer-Verlag.

Rogers, L. & D. Williams (1994) *Diffusions, Markov Processes and Martingales, Vol. 1, Itô Calculus* (2nd Ed.). New York: John Wiley and Sons.

Rogers, L. & D. Williams (1996) *Diffusions, Markov Processes and Martingales, Vol. 2, Foundations* (2nd Ed.). New York: John Wiley and Sons.

Said, S.E. & D.A. Dickey (1984) Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. *Biometrika* 71, 599-607.

Santa-Clara, P. & D. Sornette (2001) The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks. *Review of Financial Studies* 14, 149-185.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons.

Severini, T.A. & G. Tripathi (2001) A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models. *Journal of Econometrics* 102, 23-66.

Severini, T. & W. Wong (1992) Generalized Profile Likelihood and Conditionally Parametric Models. *Annals of Statistics* 20, 1768-1802.

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.

Stein, C. (1956) Efficient Nonparametric Testing and Estimation. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 187-195. Berkeley: University of California Press

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

Stone, C.J. (1990) Large Sample Inference for Log-Spline Models. *Annals of Statistics* 18, 717-741.

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

Sørensen, M. (1997) Estimating Functions for Discretely Observed Diffusions: A Review. *Selected Proceedings of the Symposium on Estimating Functions*, 305-325 (eds. I.V. Basawa, V.P. Godambe and R.L. Taylor), IMS Lecture Notes 32. Hayward: Institute of Mathematical Statistics.

Tauchen, G.E. (1985) Diagnostic Testing and Evaluation of Maximum Likelihood Models. *Journal of Econometrics* 30, 415-443.

Tauchen, G.E. (1995) New Minimum Chi-Square Methods in Empirical Finance. In *Advances in Economics and Econometrics: Theory and Applications* (eds. K. Wallace & D. Kreps) vol. III, 279-317. Cambridge: Cambridge University Press.

Thompson, S.B. (2000) Specification Tests for Continuous Time Models. Working paper, Harvard University.

van de Geer, S. (2000) *Empirical Processes in M-Estimation.* Cambridge: Cambridge University Press.

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.

Veretennikov, A.Y. (1997) On Polynomial Mixing Bounds for Stochastic Differential Equations. *Stochastic Processes and their Applications* 70, 115-127.

Werker, B.J.M., M. Hallin & C. Koell (2000) Optimal Inference for Discretely Observed Semiparametric Ornstein-Uhlenbeck Processes. *Journal of Statistical Planning and Inference* 91, 323-340.

Wong, E. (1964) The Construction of a Class of Stationary Markoff Processes. *Sixteenth Symposium in Applied Mathematics - Stoch. Proc. in Math. Physics and Engineering* (ed. R. Bellman), 264-276. Providence: American Mathematical Society.

Wong, W.H. & T,A. Severini (1991) On Maximum Likelihood Estimation in Infinite Dimensional Parameter Space. *Annals of Statististics* 19, 603-632.

Yoshida, N. (1992) Estimation for Diffusion Processes from Discrete Observations. *Journal of Multivariate Analysis* 41, 220-242.

Yoshihara, K. (1976) Limiting Behaviour of U-Statistics for Stationary, Absolutely Regular Processes. *Zeitschrift für Wahrenscheinlichkeittheorie und verwandte Gebeite* 35, 237-252.