Testing Continuous Time Models in Financial Markets

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum politicarum
(dr. rer. pol.)
im Fach Statistik und Ökonometrie

eingereicht an der
Wirtschaftswissenschaftlichen Fakultãt
Humboldt-Universität zu Berlin

von
Herrn Dipl.-Math. Torsten Kleinow
geborem am 19.08.1972 in Potsdam

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Wirtschaftswissenschaftlichen Fakultãt:
Prof. Dr. M. Burda

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Priv.-Doz. Dr. Helmut Herwartz

eingereicht am: 22. Mai 2002
Tag der mündlichen Prüfung: 4. Juli 2002

Abstract

The aim of the thesis is to provide a wide range of statistical methods designed to test parametric assumptions about the evolution of continuous time processes in financial markets. The main focus is on the statistical methodology and the investigation of the properties of the proposed methods when applied to finite samples. The latter aspect is particularly important for empirical applications. All chapters include an empirical analysis of financial data using the developed methods.

Keywords:
Mathematical Finance, Statistics, Testing, Diffusion process

Zusammenfassung

Das Ziel der Dissertation ist die Entwicklung statistischer Testverfahren zur Überprüfung parametrischer Modelle für die Dynamik zeitstetiger Prozesse und die Anwendung der entwickelten Methoden auf Finanzmarktdaten. Besonderes Augenmerk wird auf die statistische Methodik und die Untersuchung der Testeigenschaften in endlichen Stichproben gelegt, da diese in empirischen Untersuchungen von entscheidener Bedeutung sind. Alle Kapitel der Dissertation umfassen eine empirische Analyse, in der die vorgestellten Tests auf Finanzmarktdaten angewandt werden.

Sclagwörter:
Finanzmathematik, Statistik, Testverfahren, Diffusionsprozess

Acknowledgement

The current version of the book has benefited considerably from the suggestions and comments of a number of colleagues and friends I would like to thank at this point.

First of all I would like to express my gratitude to my PhD advisor Professor Wolfgang Härdle who always supported the development of this work by a number of fruitful discussions. Without his support this book would not exist in its present form.

I also would like to thank my second advisor Dr. Helmut Herwartz for his valuable comments on the topic.

Main parts of the book mirror the joint work with a number of coauthors. Particularly I would like to thank Peter Hall, Alexander Korostelev, Eckhard Platen, Camillie Logeay, Song Xi Chen and Peter Schmidt.

I very much enjoyed the inspiring environment at the Institut für Statistik und Ökonometrie at Humboldt-Universität zu Berlin and I am grateful to my colleagues and all members of the Sonderforschungsbereich 373 for supporting me in writing this book and the papers that it is based on. Particularly I would like to thank Axel Werwatz and Rainer Schulz.

For the technical support, I would like to thank the members of the XploRe team, in particular, Sigbert Klinke for his help and the continuous development of XploRe and MD*Book.

I gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 373 "Quantifikation and Simulation ökonomischer Prozesse".

Berlin, May 2002 Torsten Kleinow

Contents

Frequently Used Notation viii

1 Introduction 1

2 Tests for the Hurst coefficient 6
2.1 Introduction . . . 6
2.2 Methodology and Theory . . . 8
2.2.1 R–S Analysis . . . 8
2.2.2 Approximating the Distribution of $\widehat{H}$ . . . 10
2.2.3 Confidence Regions and Hypothesis Testing . . . 12
2.2.4 Theoretical Properties . . . 13
2.3 Empirical Analysis . . . 17

3 Testing Diffusion Models 23
3.1 Model and Observations . . . 23
3.2 Hypotheses . . . 26
3.3 Tests Based On A Discrete Time Approximation . . . 26
3.3.1 Discrete Time Approximation . . . 27
3.3.2 Estimation of the Drift and Diffusion Coefficient . . . 28
3.3.3 Testing the Parametric Model . . . 30
3.3.4 Empirical Likelihood Tests . . . 32
3.3.5 Fixed Sample Properties . . . 45
3.4 Nonparametric Estimation of the Continuous Time Model . . . 49
3.4.1 Estimation of the marginal density . . . 49
3.4.2 Estimation of the diffusion coefficient . . . 51
3.4.3 Estimation of the drift . . . 56
3.4.4 Fixed Sample Properties . . . 56
3.5 Testing the Continuous Time Model . . . 60
3.5.1 Tests based on density comparison . . . 60
3.5.2 Directly testing the diffusion coefficient . . . 64
3.5.3 Empirical Likelihood Test About The Diffusion coefficient . . . 67

3.5.4 Testing Time-inhomogeneous Diffusion Coefficients 72
3.5.5 Fixed Sample Properties 73
3.6 Empirical Analysis 76
3.6.1 The analysis of the 7-day Eurodollar rate 76
3.6.2 German Stock Prices 77

4 A Semiparametric Model For A Stock Market Index 86
4.1 Introduction 86
4.2 Statistical Methodology for a Normalized Diffusion 89
4.2.1 Parametric Models 89
4.2.2 Kernel Smoothing 90
4.2.3 Estimation of Parameters 91
4.2.4 Testing the Parametric Model 94
4.3 Empirical Analysis of the S&P 500 94
4.4 Simulation Study 96

A Appendix 106
A.1 One-dimensional Diffusions 106
A.2 Proofs of Theorems and Lemmas of EL section 108
A.3 Estimation of Parameters of a Normalized Index 118

List of Figures

2.1 Illustration of the construction of the R part in the R/S analysis. . . 9
2.2 R–S plot for VW, $\widehat{H} = 0.606$ 17
2.3 Estimated density of $\widehat{H} - H$ for 400 simulated Brownian Motions with length 6900. The vertical lines determine the 0.05, 0.95 quantiles. 18
2.4 Bootstrap density of $\widehat{H} - H$ for the Volkswagen stock. The vertical lines determine the 0.05, 0.95 quantiles. 20

3.1 Scatterplots of the drift function of Y according to (3.50) with $c_T = 0.00$ and $c_T = 0.06$. 46
3.2 Power of the empirical likelihood test. The dotted lines indicate the 5% level 47
3.3 Empirical rejection level of the empirical likelihood test for the drift of an Ornstein Uhlenbeck process. ($\Delta = 1/50$ left figure, $\Delta = 1/250$ right figure) 49
3.4 Empirical rejection probabilities of EL test with $n = 50$ for nominal level $\alpha = 0.01$ and $\alpha = 0.05$. The degree of difference between H_0 and H_1 are $c_T = 2$ (red), $c_T = 1.5$ (yellow), $c_T = 1$ (green), $c_T = 0.5$ (blue), $c_T = 0$ (black). The parameters of the simulated Ornstein Uhlenbeck process are $a = -1$ and $\sigma = 0.5$. 50
3.5 Empirical rejection probabilities of EL test with $n = 250$ for nominal level $\alpha = 0.01$ and $\alpha = 0.05$. The degree of difference between H_0 and H_1 are $c_T = 2$ (red), $c_T = 1.5$ (yellow), $c_T = 1$ (green), $c_T = 0.5$ (blue), $c_T = 0$ (black). The parameters of the simulated Ornstein Uhlenbeck process are $a = -1$ and $\sigma = 0.5$. 51
3.6 Nonparametric estimates of σ^2 together with the empirical 90% confidence band. 59
3.7 Nonparametric estimates of m together with the empirical 90% confidence band. 80

3.8 Empirical level of $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ for different models and path length $nT = 1000$. The left column displays the results for the linear mean reverting drift and the right column corresponds to the Ahn-Gao drift. On the vertical axis the empirical level is displayed and the horizontal axis shows the degrees of freedom (k). The solid line is the level of $\mathcal{T}_T^{(n)}$, the dotted line is the level of $T_T^{(n)}$ and the thin vertical line is the nominal level 0.05. 81
3.9 Empirical level of $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ for the CKLS model with AG drift and path lengths $nT = 3000$ (upper plot) and $nT = 5000$ (lower plot). On the vertical axis the empirical level is displayed and the horizontal axis shows the degrees of freedom (k). The solid line is the level of $\mathcal{T}_T^{(n)}$, the dotted line is the level of $T_T^{(n)}$ and the thin vertical line is the nominal level of 0.05. 82
3.10 Empirical power of the EL test $\mathcal{T}_T^{(n)}$ the upper line corresponds to the CKLS model and the middle one to the square root model. The lower line represents the empirical level of the Vasicek model. The paths are simulated from the Vasicek model ($nT = 1000$, 1000 trajectories) . . 83
3.11 The 7-day Eurodollar rate. 83
3.12 The autocorrelation function of the Eurodollar spot rate up to $\tau = 500$ days. 84
3.13 The marginal density of the Eurodollar spot rate. 84
3.14 Parametric and nonparametric estimates of the diffusion (upper plot) and the drift function(lower plot). The green function is the parametric estimate of the Ahn-Gao model (Ahn-Gao drift, CKLS diffusion) and the red line corresponds to the parametric estimate of Square root process (LMR drift, CIR diffusion). The blue line in drift plot corresponds to the estimator in (3.23) and the black line is estimated with (3.68). 85

4.1 S&P 500 index S, average index $\bar{S}$ and normalized index X (lower plot). 88
4.2 The optimal flexible bandwidth $h_{opt}(t)$. 95
4.3 Nonparametric and parametric estimates of the drift $m(.)$ and squared diffusion coefficient $\sigma^2(.)$ with 90% confidence bands. 96
4.4 Nonparametric estimates for the drift and squared diffusion coefficients of a simulated geometric Ornstein-Uhlenbeck process, confidence bands and true functions. 99

List of Tables

2.1 Estimated Hurst coefficient of German stocks 21
2.2 Confidence regions for Hurst coefficients 22

3.1 Diffusion coefficient models used in the simulation study 57
3.2 Ahn-Gao model used in the simulation study 57
3.3 Drift functions used in the simulation study 73
3.4 Mean and variance of the two test statistics estimated from a sample of 1000 paths with length $nT = 1000$. 75
3.5 Values of the EL test statistic and estimated parameters for the 7-day Eurodollar rate. 77
3.6 Values of the EL test statistic and estimated parameters for the DAX and five German stocks. 78

4.1 Estimated values for γ and the estimated variance of X for different bandwidths h. 95
4.2 Estimated values for β for different fixed bandwidths h. 96
4.3 Estimated parameters. 97

Frequently Used Notation

$x \stackrel{\text{def}}{=} \ldots$ x is defined as ...
$\mathbb{R}$ real numbers
$\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\infty, \infty\}$
$A^\top$ transpose of matrix A
$X \sim D$ the random variable X has distribution D
$\mathrm{E}[X]$ expected value of random variable X
$Var(X)$ variance of random variable X
$\mathrm{Std}(X)$ standard deviation of random variable X
$\mathrm{Cov}(X, Y)$ covariance of two random variables X and Y
$\mathrm{U}[a, b]$ uniform distribution on $[a, b]$
$\mathrm{N}(\mu, \Sigma)$ normal distribution with expectation μ and covariance matrix Σ, a similar notation is used if Σ is the correlation matrix
cdf denotes the cumulative distribution function
pdf denotes the probability density function
$\mathrm{P}[A]$ or $\mathrm{P}(A)$ probability of a set A
$\boldsymbol{I}$ indicator function
$(F \circ G)(x) \stackrel{\text{def}}{=} F\{G(x)\}$ for functions F and G
$\mathcal{F}_t$ is the information set generated by all information available at time t

For two sequences of real numbers α_n and β_n:
$\alpha_n = \mathcal{O}(\beta_n)$ iff $\frac{\alpha_n}{\beta_n} \longrightarrow$ constant, as $n \longrightarrow \infty$
$\alpha_n = \mathcal{o}(\beta_n)$ iff $\frac{\alpha_n}{\beta_n} \longrightarrow 0$, as $n \longrightarrow \infty$
For two sequences of random variables A_n and B_n:
$A_n = \mathcal{O}_p(B_n)$ iff $\forall \varepsilon > 0\ \exists M,\ \exists N$ such that $\mathrm{P}[|A_n/B_n| > M] < \varepsilon,\ \forall n > N$.
$A_n = \mathcal{o}_p(B_n)$ iff $\forall \varepsilon > 0\ :\ \lim_{n\to\infty} \mathrm{P}[|A_n/B_n| > \varepsilon] = 0$.
For sequence of functions $\gamma_n(x)$ and a sequence of random numbers δ_n:
$\gamma_n(x) = \tilde{\mathcal{O}}_p(\delta_n)$ iff $\sup_{x\in S} |\gamma_(x)| = O_p(\delta_n)$
$\gamma_n(x) = \tilde{\mathcal{o}}_p(\delta_n)$ iff $\sup_{x\in S} |\gamma_n(x)| = o_p(\delta_n)$

Chapter 1

Introduction

Throughout this book we present statistical methods that test particular models for financial data. The data that we consider are generated by interest rate or asset price processes. The assets that we have in mind are stocks, exchange rates, index processes or any other kind of a risky security.

The evolution of the prices of these assets takes place in a continuous state space and in continuous time. If prices in financial markets do not vary continuously in time, they move and can be observed very frequently. In particular, the introduction of electronic trading systems, like XETRA, has rapidly increased the frequency of price fixings. On top of that, continuous time models have proofed their usefulness as approximations of reality and modern methods in mathematical finance rely on this kind of models. We therefore concentrate here on statistical methods developed for the quantitative analysis of financial data in continuous time.

Before we start with the presentation of the statistical methodology, we give a brief introduction into the theory of mathematical finance to motivate the remainder of the book.

To model the market that we consider here, we start with a spot interest rate process $\{r(t),\ t \in [0,T]\}$ and an asset price process $\{P(t),\ t \in [0,T]\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P}, \{\mathcal{F}(t),\ t \in [0,T]\})$ and adapted to the filtration $\{\mathcal{F}(t),\ t \in [0,T]\}$. Heuristically speaking, the adaption of r and P on $\{\mathcal{F}(t),\ t \in [0,T]\}$ means that the current values $r(t)$ and $P(t)$ are known at time t. In addition to r and P, a savings account $P_0(t)$ is given as the solution of the differential equation

$$dP_0(t) = r(t)P_0(t)dt \tag{1.1}$$

with initial condition $P(0) = 1$. P_0 can be interpreted as a bank deposit with an instantaneous return process r. The discounted asset price is then $\tilde{P}(t) = P(t)/P_0(t)$.

A trading strategy in such a financial market is a pair $(a_0(t), a_1(t))$ that represents the investments in the two assets P_0 and P and the corresponding portfolio process $V(t)$ is the current value of a portfolio according to the investments $(a_0(t), a_1(t))$, i.e. $V(t) = a_0(t)P_0(t) + a_1(t)P(t)$. We call a trading strategy self-financed if $V(t) = V(0) + \int_0^t a_0(t)dP_0(t) + \int_0^t a_1(t)dP(t)$. This means, that the value of the portfolio at time t is equal to the sum of the initial investment $V(0)$ and the gains earned from the investments up to time t.

A fundamental concept in the mathematical theory of financial markets is the no arbitrage condition. We say that an arbitrage opportunity exists, if there exists a self-financed trading strategy and a lower bound v such that for the corresponding portfolio process holds, (i) $V(T) - V(0) \geq 0$, (ii) $V(t) \geq v$ for all $t \in [0, T]$ and (iii) $\mathrm{P}[V(T) - V(0) > 0] > 0$. In this definition of arbitrage the second condition is of particular importance, since it prevents double-or-quits strategies.

The fundamental theorem of asset pricing states that in an arbitrage free market model exists a probability measure Q on $(\Omega, \mathcal{F})$ such that every discounted price process $\tilde{P}$ is a martingale under Q and Q is equivalent to P, i.e. P and Q have the same null sets. A consequence is, that the price process P need to be a semimartingale under the original measure P. Delbaen and Schachermayer (1994) proof that a arbitrage opportunity exists, if a price process P is not semimartingale. An important question in the quantitative analysis of financial data is therefore to check the assumption, that an observed process is a semimartingale.

In Chapter 2 we capture this problem by introducing a test about the Hurst coefficient H of a fractional Brownian motion (FBM). A FBM is an example for a stochastic process that is not a semimartingale except in the case of a Hurst coefficient H equal to 0.5, see Rogers (1997) for a formal proof. Hence a financial market model with a price process P that is assumed to be a FBM with $H \neq 0.5$ implies an arbitrage opportunity. Rogers (1997) also provides a direct construction of a trading strategy that produces arbitrage in this situation.

More precisely we test in Chapter 2 a null hypothesis about the Hurst coefficient of a smooth function of fractional Brownian Motion. Thus we do not restrict our class of models to one particular process, but include other processes that have the same fractal dimension as the FBM.

In addition to the presentation of estimation and testing procedure for the Hurst coefficient we derive the asymptotic distribution of the test and show how this distribution can be approximated by a bootstrap procedure. The chapter also includes an empirical analysis of German stocks.

In Chapter 3 we assume, that the considered market model is arbitrage free and thus we restrict the statistical methodology to processes that are semimartingales under the original probability measure P. Under this assumption we concentrate on the particular case where P is a diffusion process that is given as the solution of the stochastic differential equation

$$dP(t) = P(t)\Big\{\mu\{P(t)\}dt + \sigma\{P(t)\}dW(t)\Big\} \tag{1.2}$$

with an initial condition $P(0) = p$. $\{W(t),\ t \in [0,T]\}$ is a standard Brownian motion under P adapted to $\{\mathcal{F}(t),\ t \in [0,T]\}$. A solution of (1.2) is given by

$$P(t) = e^{X(t)}$$

with:

$$dX(t) = \left(\mu\{P(t)\} - \frac{1}{2}\sigma^2\{P(t)\}\right) dt + \sigma\{P(t)\} dW(t)\ .$$

Furthermore we assume that the spot rate r solves the stochastic differential equation

$$dr(t) = m_r\{r(t)\}dt + \sigma_r\{r(t)\}dW_0(t) \tag{1.3}$$

where W_0 is again a Brownian motion adapted to $\{\mathcal{F}(t),\ t \in [0,T]\}$ and W and W_0 are independent.

We define the risk premium as

$$\eta(t) \stackrel{\text{def}}{=} \frac{m\{P(t)\} - r(t)}{\sigma\{P(t)\}}$$

and for every $t \in [0,T]$ a probability measure Q_t via the Radon-Nikodym density

$$\frac{dQ_t}{d\mathrm{P}} = \exp\left\{\int_0^t \eta(s)dW(s) - \frac{1}{2}\int_0^t \eta^2(s)ds\right\}\ .$$

The equivalent martingale measure Q is then defined as a probability measure such that $Q(A) = Q_t(A)$ for all $A \in \mathcal{F}(t)$ and for all $t \in T$. From Girsanov's theorem, Karatzas and Shreve (1991), we have that

$$W_Q(t) \stackrel{\text{def}}{=} W(t) + \int_0^t \eta(s)ds$$

is a Brownian motion with respect to the equivalent martingal measure Q. Since the discounted asset price $\tilde{P}$ solves the SDE

$$d\tilde{P}(t) = \tilde{P}(t)\sigma\{\tilde{P}(t)\}dW_Q(t) \tag{1.4}$$

it is, under Q, an integral with respect to a Brownian motion and thus a martingale. We mention that the drift term m disappears from the definition of $\tilde{P}$ under the martingale measure.

One of the most important applications in mathematical finance is the pricing of contingent claims. In general, the price of any contingent claim that pays B at time T is given by the expectation of the discounted payoff under the equivalent martingale measure Q

$$H = \mathrm{E}_Q\left[\exp\left\{\int_t^T -r(t)dt\right\} B\middle|\mathcal{F}(t)\right] \tag{1.5}$$

In the particular case of the well known European call option that pays $\{P(T) - K\}^+$ in time T the option price $H = H\{P(t), T-t, K\}$ in time t can also be expressed as the solution of the partial differential equation

$$0 = rP\frac{\partial H}{\partial P} - rH - \frac{\partial H}{\partial t} + \frac{1}{2}(P\sigma)^2\frac{\partial^2 H}{\partial P^2} \tag{1.6}$$

with the boundary condition $H\{P, 0, K\} = \{p - K\}^+$.

From (1.3)-(1.6) it follows, that the parameters of interest for option pricing are the diffusion coefficients of P and r and the drift of r. The drift m of P influences the density process $dQ_t/d\mathrm{P}$, but under the martingale measure it disappears, and thus it does not influence derivative prices. Since these prices are expected discounted payoffs, the same argument does not hold for m_r, the drift of the spot rate process. For these reasons we propose in Chapter 3 methods that test parametric functional forms of the coefficients of P and r separately as well as methods that test for the whole dynamics of a diffusion process.

We divide the proposed statistical methods in Chapter 3 into two groups. The first group consists of methods that are based on an approximation of the continuous time process by a time series. This approximation gives the opportunity to apply the statistical tools developed for time series analysis. However, it turns out, that the potential of these methods is restricted when applied to diffusions. In particular asymptotic results are based on the assumption that T goes to infinity. For the second group of methods we directly apply properties of the continuous time diffusion. Since we observe the process only at discrete time points, we have two kinds of asymptotics here: (i) T tends to infinity and (ii) the time difference between two successive observations goes to zero. The latter makes the statistics of diffusions different from time series analysis, where fixed time intervals between successive observations are considered.

In general, the tests introduced in Chapter 3 compare nonparametrically estimated functions to parametric forms of these functions implied by the null hypothesis. The functions that are tested depend on the particular null hypothesis. We propose methods that compare the marginal density and the transition density of a process and thus test about the whole dynamics of the process. As described above the drift function does not influence the prices of derivatives and we therefore present quantitative methods that compare the estimated diffusion coefficient directly to its parametric form implied by the null hypothesis.

In the empirical study at the end of Chapter 3 we analyze the quantitative behavior of a spot interest rate process, namely the 7-day Eurodollar rate, German stocks and the German stock market index DAX. We find that, although the drift of r is important in mathematical finance, tests that include a drift specification will fail, since not enough data are available to produce reliable estimates for the drift term.

We therefore concentrate on the estimation and testing of the diffusion coefficient and find for all treated data, that the tested hypotheses are rejected.

The work is completed by a particular model of a stock market index that is proposed in Chapter 4. The model we consider treats a stock market index as a product of a smooth growth process and a normalized index. From a statistical point of view there arises the problem of nonparametric estimation of the conditional mean when the error terms are not independent and identically distributed, but are the observations of a diffusion process. In particular we consider the case of an Ornstein Uhlenbeck process.

Chapter 2

Semiparametric Bootstrap Approach to Hypothesis Tests and Confidence Intervals for the Hurst Coefficient

A major application of rescaled adjusted range analysis (R–S analysis) is the study of price fluctuations in financial markets. There, the value of the Hurst constant, H, in a time series may be interpreted as an indicator of the irregularity of the price of a commodity, currency or similar quantity. Interval estimation and hypothesis testing for H are central to comparative quantitative analysis. In this chapter we propose a new bootstrap, or Monte Carlo, approach to such problems. Traditional bootstrap methods in this context are based on fitting a process chosen from a wide but relatively conventional range of discrete time series models, including autoregressions, moving averages, autoregressive moving averages and many more. By way of contrast we suggest simulation using a single type of continuous-time process, with its fractal dimension. We provide theoretical justification for this method, and explore its numerical properties and statistical performance by application to real data on commodity prices and exchange rates.

2.1 Introduction

R–S analysis has its roots in early work of the British hydrologist H.E. Hurst, who investigated dependence properties of phenomena such as levels of the River Nile. The Hurst constant H, as the index of dependence is often called, always lies between 0 and 1, and equals $\frac{1}{2}$ for processes that have independent increments. Particular interest focuses on the hypothesis that $H > \frac{1}{2}$, indicating relatively long-range dependence.

For example, Hurst observed that $H = 0.91$ in the case of Nile data, indicating a strength of dependence that was well beyond what could be adequately explained assuming independent increments.

Today, a principal application of R–S analysis is to the study of fluctuations in financial markets, where the value of H is variously interpreted as an indicator of range of dependence, of irregularity and of nervousness. (Adler (1981) coined the word 'erraticism' to denote a quantitative measure of 'nervousness'.) To elucidate this point we note that the fractal dimension D of sample paths of a locally self-similar or self-affine random process increases monotonically with the irregularity of those paths; and that $D = 2 - H$, see e.g. Berry and Hannay (1978); Sayles and Thomas (1978); Adler (1981), Chapter 8; Mandelbrot et al. (1984); Hall et al. (1996). Therefore, a process with higher Hurst constant is more regular, or less erratic, or less 'nervous' then one with a lower value. For example, a time series of commodity prices that is characterised by a larger Hurst constant enjoys greater stability, over at least short periods of time; and trade in that commodity might be said to be subject to less nervousness. See for example Peters (1994).

As already mentioned in the introduction, the abscence of arbitrage is strongly related to the Hurst constant. A particular process with a Hurst coefficient different from 1/2 is the fractional Brownian motion (FBM) that is defined as a Gaussian process ζ with

$$\mathrm{P}(\zeta_0 = 0) = 1 \qquad \mathrm{E}(\zeta_t) = 0$$

and

$$\mathrm{E}(\zeta_{s+t} - \zeta_s)^2 = |t|^\alpha$$

for all s and t, where $\alpha = 2H \in (0, 2)$. Equivalently, ζ_t is defined to be that Gaussian process with zero mean and covariance

$$\gamma(s,t) \equiv \mathrm{Cov}(\zeta_s, \zeta_t) = \tfrac{1}{2}\left(|s|^\alpha + |t|^\alpha - |s-t|^\alpha\right). \tag{2.1}$$

See for example Beran (1994), p. 51ff and Peters (1994) p. 183ff. Rogers (1997) shows that the FBM is not a semimartingale for $H \neq 1/2$. On a heuristic level we have from (2.1) that $\zeta_t - \zeta_s$ is of order $|t-s|^H$ which means that

$$\sum_{j=1}^{2^n} |\zeta_{j2^{-n}} - \zeta_{(j-1)2^{-n}}|^p \approx (2^n)^{1-pH}.$$

It follows, that the order-p variation of ζ is infinite if $p < H^{-1}$ and zero if $p > H^{-1}$ which is consistent with the semimartingale property for $H = 1/2$ only. Note that in this case the FBM coincites with the standard Brownian motion.

For the above reasons and since the no arbitrage condition is essential in mathematical finance, point and interval estimation of the Hurst constant can be basic to quantitative descriptions of market fluctuations. And testing for significant differences between two Hurst constants, or between one constant and the value $\frac{1}{2}$, is fundamental to comparative quantitative analysis of market 'nervousness'. In this chapter we suggest bootstrap, or Monte Carlo, methods for constructing confidence intervals and hypothesis tests for Hurst indices.

Our methods are based on the estimator $\widehat{H}$ of H derived from R–S analysis, and involve simulating the sampled process using a time-adjusted version of fractional Brownian motion. We argue that, since the 'S' part of R–S analysis corrects for inhomogeneities in the data, it is unnecessary to reproduce them in the bootstrap algorithm.

This approach differs fundamentally from more traditional methods currently used for simulation, where the model is taken to be a relatively conventional discrete time series such as an autoregression, or moving average, or autoregressive moving average. See for example Peters (1994) Chapter 9. Instead, we suggest simulating a single type of continuous stochastic process, where the degree of irregularity is determined empirically through an estimator of H. We justify this approach through theoretical analysis, and assess its numerical and statistical properties using applications to real data on stock prices.

The idea of basing the bootstrap method on a continuous rather than a discrete stochastic process has been suggested before, but in the very different context of bootstrap methods for spatial samples of data on surface roughness, Davies and Hall (1998). There, the 'S' part of R–S analysis is usually omitted, since the observed process is generally scale-homogeneous. Such bootstrap methods are nonstandard, since they conform to neither the parametric nor nonparametric bootstrap approaches. They fall midway between the two, and might fairly be said to be semiparametric bootstrap methods.

2.2 Methodology and Theory

2.2.1 R–S Analysis

We observe a stochastic process X_t at time points $t \in \mathcal{I} = \{0, \ldots, N\}$. Let n be an integer that is small relative to N (asymptotically, as $N/n \to \infty$), and let A denote the integer part of N/n. Divide the 'interval' $\mathcal{I}$ into A consecutive 'subintervals', each of length n and with overlapping endpoints. In every subinterval correct the original datum X_t for location, using the mean slope of the process in the subinterval, obtaining $X_t - (t/n)\,(X_{an} - X_{(a-1)n})$ for all t with $(a-1)n \le t \le an$ and for all $a = 1, \ldots, A$. Over the a'th subinterval $\mathcal{I}_a = \{(a-1)n, (a-1)n+1, \ldots, an\}$, for $1 \le a \le A$, construct the smallest box (with sides parallel to the coordinate axes) such that the box contains all the fluctuations of $X_t - (t/n)\,(X_{an} - X_{(a-1)n})$ that occur within $\mathcal{I}_a$. Then, the height of the box equals

$$\begin{aligned} R_a &= \max_{(a-1)n \le t \le an} \left\{ X_t - \frac{t}{n}(X_{an} - X_{(a-1)n}) \right\} \\ &- \min_{(a-1)n \le t \le an} \left\{ X_t - \frac{t}{n}(X_{an} - X_{(a-1)n}) \right\} \end{aligned}$$

The construction of the boxes is illustrated in Figure 2.1.

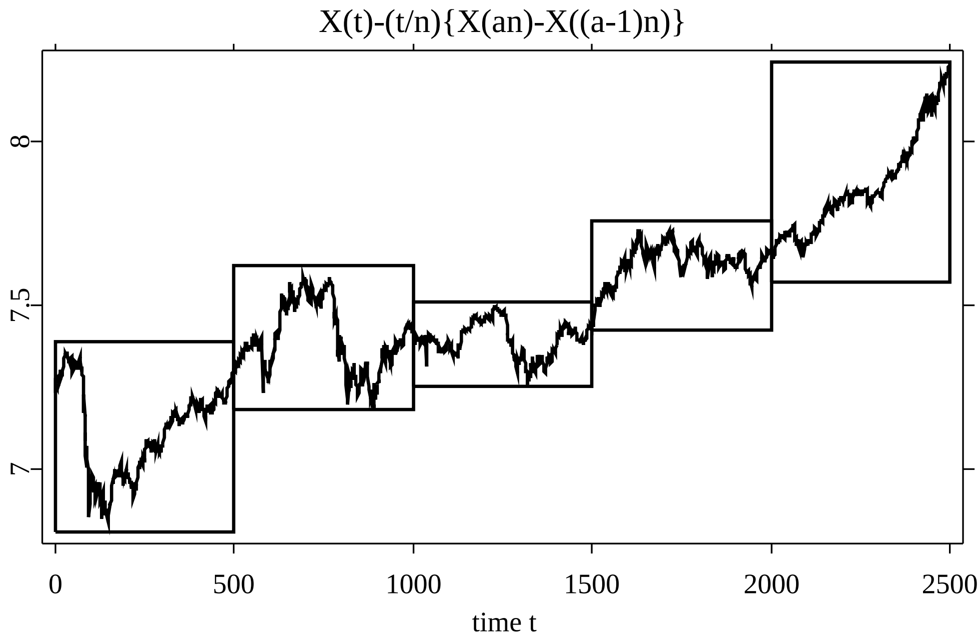


Figure 2.1: Illustration of the construction of the R part in the R/S analysis.

Let S_a denote the empirical standard error of the n variables $X_t - X_{t-1}$, for $(a-1)n + 1 \le t \le an$. If the process X is stationary then S_a varies little with a; in other cases, dividing R_a by S_a corrects for the main effects of scale inhomogeneity in both spatial and temporal domains.

The total area of the boxes, corrected for scale, is proportional in n to

$$\left(\frac{R}{S}\right)_n := A^{-1} \sum_{a=1}^{A} \frac{R_a}{S_a}\,. \tag{2.2}$$

The slope $\widehat{H}$ of the regression of $\log(R/S)_n$ on $\log n$, for k values of n, may be taken as an estimator of the Hurst constant H describing long-range dependence of the process X. See for example Beran (1994), Chapter 1 and Peters (1994) Chapters 4–6.

This R–S analysis, or 'rescaled adjusted range' analysis, dates from Hurst (1951). If the process X is stationary then correction for scale is not strictly necessary, and we may take each S_a to be the constant 1. In that case the R–S statistic $\widehat{H}$ is a version of the box-counting estimator that is widely used in physical science applications; see for example Carter et al. (1988), Sullivan and Hunt (1988) and Hunt (1990). The box-counting estimator is related to the capacity definition of fractal dimension Barnsley (1988), p. 172ff, and the R–S estimator may be interpreted in the same way. Statistical properties of the box-counting estimator have been discussed by Hall and Wood (1993).

A more detailed analysis, exploiting dependence among the errors in the regression of $\log(R/S)_n$ on $\log n$, may be undertaken in place of R–S analysis. See Kent and Wood (1997) for a version of this approach in the case where scale correction is unnecessary. However, as Kent and Wood show, the advantages of the approach tend to be asymptotic in character, and sample sizes may need to be extremely large before real improvements are obtained.

2.2.2 Approximating the Distribution of $\widehat{H}$

Depending on the value of H, and on the nature of the stochastic process X, the asymptotic distribution of $\widehat{H}$ (as $N \to \infty$, for fixed k) can be Normal or Rosenblatt; the latter was introduced by Taqqu (1875), following work of Rosenblatt (1961). (More concisely, in the Rosenblatt case the asymptotic distribution of $\widehat{H}$ is that of a finite linear form in correlated Rosenblatt-distributed random variables, but for simplicity we shall refer to this as a Rosenblatt distribution.) Indeed, the asymptotic distribution of $\widehat{H}$ can be Rosenblatt for $3/4 < H < 1$ and Normal for $0 < H \leq 3/4$; see Section 2.2.4. The Rosenblatt distribution that is relevant here is particularly complex, and its shape depends intimately on the unknown value of H. The distribution has not been tabulated.

If the value of k is large, i.e. the number of values of n for the linear regression is large then the Rosenblatt approximation becomes, by virtue of the central limit theorem, similar to the Normal approximation. However, the asymptotic variance is difficult to calculate. Moreover, it is known from work of Hall and Wood (1993) and Constantine and Hall (1994) that, due to long-range dependence, statistical performance of the estimator $\widehat{H}$ generally deteriorates for large k, and in fact optimal mean squared error properties are often achieved by keeping k fixed as N increases.

These considerations motivate Monte Carlo analysis, rather than more conventional asymptotic methods, in the range $3/4 < H < 1$. Even when H lies outside this interval there is much to be said for taking a Monte Carlo approach, however. Monte

Carlo simulation can be expected to capture many of the penultimate, second-order effects that describe departure of the distribution of $\widehat{H}$ from its asymptotic limit, so that even if the limiting distribution were known, the Monte Carlo approach would be expected to provide somewhat greater accuracy than the conventional asymptotic approximation. The second-order effects arise from finiteness of N, and from the fact that stochastic fluctuations of the scale correction in R–S analysis influence the true distribution of $\widehat{H}$ even though they do not affect the limit distribution.

A more familiar example of the same phenomenon is use of Student's t distribution to approximate the distribution of a Studentised ratio, even when the sampled distribution is not exactly Normally distributed. The Student's t approximation represents a 'penultimate' form of the Normal 'ultimate' limiting distribution. Even for data from a skew distribution the Student's t approach generally captures finite-sample properties better than the Normal approximation, despite the fact that it does not capture all second-order departures from Normality.

We shall show in Section 2.2.4 that in many cases the limiting distribution of $\widehat{H}$ depends only on H and a temporal scale factor. The spatial scale of the process X, and the process's potential heteroscedasticity and non-Gaussianity, do not feature in first-order asymptotic results. In large part this is a result of the 'S' component of R–S analysis. Therefore, the limiting distribution of $\widehat{H}$ is the same as it would be if X_t were ζ_t, where ζ is an elementary self-similar Gaussian process. The Gaussian process that we have in mind is the fractional Brownian motion, defined above.

We may simulate from a discrete approximation to ζ_t, say on the points $t_j = j/\nu$ for a large integer ν, by forming the $(2p\nu + 1) \times (2p\nu + 1)$ covariance matrix, M, of which the (i, j)'th element is $\gamma(t_i, t_j)$ for $-p\nu \le i, j \le p\nu$ (p an integer); and then using the spectral decomposition of M to generate Gaussian random $(2p\nu+1)$-vectors with this covariance. Alternatively, methods of Davies and Harte (1987), or those of Wood and Chan (1994) or of the many authors whose work is surveyed by Wood and Chan, may be employed.

Denote the original data set $\{X_1, \ldots, X_N\}$ by $\mathcal{X}$. Our bootstrap algorithm is as follows. Compute the estimator $\widehat{H}$, and in the steps below, take $\alpha = 2\widehat{H}$ when constructing the fractional Brownian motion ζ, conditional on $\mathcal{X}$. Let X_t^*, for $0 \le t \le N$, denote a realisation of the process ζ. Compute the corresponding value $\widehat{H}^*$ of $\widehat{H}$. Take the conditional distribution of $\widehat{H}^*$, given the data $\mathcal{X}$, to be a Monte Carlo approximation to the unconditional distribution of $\widehat{H}$; or alternatively, take the conditional distribution of $\widehat{H}^* - \widehat{H}$ to approximate the unconditional distribution of $\widehat{H} - H$. These approaches give rise respectively to the two percentile methods discussed in Section 2.2.3.

Some of the second-order properties that this approach does not capture may be addressed by fitting a smooth estimate of scale to the process ζ. For example, we might model the variance function $\sigma(t)^2 = Var(X_t)$, and thereby compute an estimator $\hat{\sigma}(\cdot)$ of $\sigma(\cdot)$; and simulate from the process $\hat{\sigma}(t)\,|t|^{-\alpha/2}\,\zeta_t$ rather than from ζ_t. In this case we should translate the time interval so as to avoid the origin.

2.2.3 Confidence Regions and Hypothesis Testing

Confidence intervals and hypothesis tests for H may be constructed using either of the two standard bootstrap percentile methods. For example, a nominal 95% confidence interval for H is given by $(\widehat{H}^{(1)}, \widehat{H}^{(2)})$, where $\widehat{H}^{(1)}$ and $\widehat{H}^{(2)}$ are defined by either $P(\widehat{H}^* \le \widehat{H}^{(1)}|\mathcal{X}) = P(\widehat{H}^* \ge \widehat{H}^{(2)}|\mathcal{X}) = 0.025$ or $P(\widehat{H}^* - \widehat{H} \le \widehat{H} - \widehat{H}^{(2)}|\mathcal{X}) = P(\widehat{H}^* - \widehat{H} \ge \widehat{H} - \widehat{H}^{(1)}|\mathcal{X}) = 0.025$. A test at the 5% level of the null hypothesis that $H = \frac{1}{2}$, corresponding to X being a random walk, is to reject the null if $(\widehat{H}^{(1)}, \widehat{H}^{(2)})$ does not contain the point $\frac{1}{2}$.

Given two independent samples from long-range dependent processes, leading to respective estimators $\widehat{H}_1$ and $\widehat{H}_2$ of Hurst constants, we may generate independent realisations from respective stochastic processes $\zeta^{(1)}$ and $\zeta^{(2)}$, and thereby compute a bootstrap approximation to the distribution of $\widehat{H}_1 - \widehat{H}_2$ or of $\widehat{H}_1 - \widehat{H}_2 - (H_1 - H_2)$. As before, this may be used as the basis of percentile-bootstrap confidence intervals and hypothesis tests for $H_1 - H_2$.

These techniques, being based on the percentile bootstrap, lack the pivotalness that bootstrap methods for confidence procedures should ideally enjoy. However, they have asymptotically correct levels, as N increases. Moreover, even when the statistic $\widehat{H}$ admits a Normal asymptotic distribution we lack a simple, computable variance estimator with which to correct for scale. And when the limiting distribution is Rosenblatt, rather than Normal, scale corrections are not sufficient to produce pivotalness, since the shape of the Rosenblatt distribution depends on the unknown Hurst constant through more than simply scale. For these reasons we argue that the percentile-t bootstrap, often suggested in simpler problems as a pivotal method for constructing confidence intervals and hypothesis tests with relatively accurate levels (see for example Hall (1992), p. 14f; Efron and Tibshirani (1993), p. 158f; Shao and Tu (1995), p. 94f; Davison and Hinkley (1997), p. 29f) is not appropriate in the present setting.

Instead, level accuracy may be enhanced by using the double bootstrap (see for example Hall (1992), p. 20ff; Efron and Tibshirani (1993), p. 263ff; Shao and Tu (1995), p. 155ff; Davison and Hinkley (1997), p. 103ff). However, the accuracy typically achived by double-bootstrap procedures cannot be expected to generalize in the present case, since our Gaussian model based on the fractional Brownian motion does not necessarily reflect all second-order features of the distribution of the sampled stochastic process X. It seems difficult to improve on this situation without introducing relatively complex high-order models for X.

2.2.4 Theoretical Properties

We will now formally prove the theoretical properties of $\hat{H}$ foreshadowed in 2.2.3. Suppose the data X_t, $t \in \mathcal{I}$, are generated as $X_t = g(Y_{\epsilon t}, t)$, where

(a) g is a smooth bivariate function,

(b) Y is a Gaussian process whose sample paths have fractal dimension $D = 2 - H$, and

(c) ϵ denotes a small positive constant.

The function g represents a possibly nonlinear transformation of Y, implying in particular that the observed process X is not necessarily Gaussian. Importantly, it allows a wide range of different types of inhomogeneity. By taking ϵ small we ensure that even if t_1 is moderately distant from t_2, X_{t_1} can be strongly correlated with X_{t_2}. This confers long-range dependence on the observed process. There is no difficulty in extending our results to the case where X is a function of a vector of Gaussian processes, say $X_t = g(Y_{\epsilon t}^{(1)}, \ldots, Y_{\epsilon t}^{(k)}, t)$. Here the Hurst index that prevails equals 2 minus the fractal dimension of sample paths of the process $Y^{(j)}$ that has the roughest sample paths. It is also possible to incorporate a smooth, monotone, nonlinear transformation of the time variable t. However, the simpler setting prescribed by condition (a) conveys the important characteristics of these more complex models.

We claim that, under models of the type characterised by (a)–(c), $\widehat{H}$ is consistent for H and has an asymptotic distribution that is either Normal or of the type introduced by Rosenblatt (1961). To formulate this assertion as a mathematical theorem we first elaborate on (a)–(c) with the following assumptions:

(A) the derivatives

$$g_{j_1 j_2}(y,t) = (\partial/\partial y)^{j_1}\, (\partial/\partial t)^{j_2}\, g(y,t)$$

are bounded for each $j_1, j_2 \geq 0$, and g_{10} does not vanish;

(B) the Gaussian process Y satisfies $E(Y_t) \equiv 0$, and for constants $c > 0$, $\alpha = 2H \in (\frac{1}{2}, 2)$ and $\beta > \min(\frac{1}{2}, 2-\alpha)$, $E(Y_{s+t} - Y_s)^2 = c|t|^\alpha + \mathcal{O}(|t|^{\alpha+\beta})$, uniformly in $s \in \mathcal{J} = [0,1]$, as $t \to 0$; and

(C) $\epsilon = 1/N \to 0$,

THEOREM 2.1 *We define* $\widehat{H}$ *by regression of* $\log(R/S)_n$ *on* $\log n$, *i.e.*

$$\log\left(\frac{R}{S}\right)_n = \widehat{H} \log n + C\ ,$$

for a fixed number, k, *of values* $\ell_1 m, \ldots, \ell_k m$ *of* n, *where* $\ell_1, \ldots, \ell_k$ *are fixed and* $m = m(\epsilon) \to \infty$ *as* $\epsilon \to 0$, *in such a manner that* $m^{-1} + m\epsilon = \mathcal{O}(\epsilon^a)$ *for some* $a > 0$.

Define $\xi = m\epsilon$ *and*

$$t_\xi = \begin{cases} \xi^{2(1-H)} \text{ if } 3/4 < H < 1 \\ \left(\xi \log \xi^{-1}\right)^{1/2} \text{ if } H = 3/4 \\ \xi^{1/2} \text{ if } 0 < H < 3/4 \,, \end{cases}$$

which converges to 0 as $\epsilon \to 0$*. Then, we claim that* $\widehat{H} - H$ *may be expressed as* $t_\xi Z_\xi$*, where* Z_ξ *has a proper limiting distribution as* $\epsilon \to 0$*.*

The regularity conditions may be relaxed in many circumstances. For example, the restriction in (B) that $\alpha > \frac{1}{2}$ may be dropped if $g(y,t) \equiv y$, and also in some other cases. The boundedness condition on derivatives of g may also be relaxed.

Crucially, the limiting distribution of $\widehat{H}$ depends only on H and $\ell_1, \ldots, \ell_k$; it does not depend on g or on the scale constant, c, appearing in the first-order approximation of the covariance. The main effects of scale and heteroscedasticity, entering through g and c, have cancelled due to rescaling by the terms S_a in (2.2), see the proof of Theorem 2.1. The limiting distribution is Normal when $0 < H \leq 3/4$, and a finite linear combination of correlated Rosenblatt distributions when $3/4 < H < 1$. Outline proofs of all these assertions are given in the appendix.

The results are foreshadowed by those of Hall and Wood (1993) for box-counting estimators, of which $\widehat{H}$ may be regarded as a scale-corrected version. We do not give the form of the limits, since it is complex (particularly in the Rosenblatt case), but it is of the type discussed by Hall and Wood (1993), p. 252. The relationships between statistical properties of a Gaussian process (e.g. Y), and of a smooth function of that process (e.g. X), have been addressed by Hall and Roy (1994).

The fact that the limiting distribution depends only on H and $\ell_1, \ldots, \ell_k$ justifies the bootstrap methods suggested in Section 2.2.2. Specifically, since the bootstrap algorithm preserves the way in which H and $\ell_1, \ldots, \ell_k$ contribute to the limiting distribution, and since $\widehat{H} \to H$ at a rate that is polynomial in ξ (indeed, at rate t_ξ), then the bootstrap produces confidence intervals and hypothesis tests that have asymptotically correct coverage. The fractional Brownian motion ζ, used as the basis for our simulations, is just one of many that could have been employed, satisfying condition (B) above.

Note particularly that we keep k fixed as ϵ decreases. If our regularity conditions were to allow $k = k(\epsilon)$ to diverge then the Rosenblatt limit would change to Normal, but as discussed by Constantine and Hall (1994), this would generally be at the expense of increased mean squared error of $\widehat{H}$.

PROOF of Theorem 2.1:

Put $Z_t = g(Y_t, t)$ and let $\mathcal{J} = [0,1]$. From the Taylor formula we have for any integer

B'

$$
\begin{aligned}
Z_{t_1} \;&=\; \sum_{j_1=0}^{B'} \sum_{j_2=0}^{B'} \frac{1}{j_1! j_2!} \,(Y_{t_1} - Y_{t_2})^{j_1}\,(t_1 - t_2)^{j_2}\, g_{j_1 j_2}(Y_{t_2}, t_2) \\
&\quad + \mathcal{O}_p\big(|t_1 - t_2|^{B'}\big) + \mathcal{O}_p\big(|Y_{t_1} - Y_{t_2}|^{B'}\big)\,.
\end{aligned}
$$

Given $B > 0$, we choose B' so large that $B'\alpha > 2B$. Then we get with assumption (B), uniformly in $t_1, t_2 \in \mathcal{J}$,

$$
Z_{t_1} = \sum_{j_1=0}^{B'} \sum_{j_2=0}^{B'} \frac{1}{j_1! j_2!} \,(Y_{t_1} - Y_{t_2})^{j_1}\,(t_1 - t_2)^{j_2}\, g_{j_1 j_2}(Y_{t_2}, t_2) + \mathcal{O}_p\big(|t_1 - t_2|^{B}\big)\,. \tag{2.3}
$$

This formula provides the opportunity to develop Taylor expansions of quantities such as R_a/S_a. It turns out that only the first term in such expansions contributes to asymptotic results. Nevertheless, higher-order Taylor-expansion terms should be included since, prior to correction for their means and analysis of their size, they are potential first-order contributors to limit theory for $(R/S)_n$. In our work the contributions of these high-order terms will be denoted by $Q_1, Q_2, \ldots$. For the sake of simplicity we ignore the mean correction in the definition of S_a.

Let $\mathcal{T} \subseteq \mathcal{J}$ denote a set of $n+1$ equally-spaced points $t_0 < \ldots < t_n$ within an interval of width $\delta = n\epsilon$, and write $S_{\mathcal{T}}$ and $U_{\mathcal{T}}$ for the empirical standard errors of the 'samples' $\{Z_{t_i} - Z_{t_{i-1}},\, 1 \le i \le n\}$ and $\{Y_{t_i} - Y_{t_{i-1}},\, 1 \le i \le n\}$, respectively. Then by (2.3), for all $\eta > 0$,

$$
\begin{aligned}
S_{\mathcal{T}}^2 \;&=\; g_{10}(Y_{t_2}, t_2)^2\, U_{\mathcal{T}}^2 + Q_1 + \mathcal{O}_p\big(\epsilon^{(\alpha/2)+B-\eta}\big)\,, &(2.4)\\
R_{\mathcal{T}} \;&\stackrel{\text{def}}{=}\; \max_{t\in\mathcal{T}} Z_t - \min_{t\in\mathcal{T}} Z_t \\
&= s\,|g_{10}(Y_{t_2}, t_2)|\,\big(Y_{T_{\mathcal{T}}} - Y_{T'_{\mathcal{T}}}\big) + Q_2 + \mathcal{O}_p\big(\delta^{B}\big)\,, &(2.5)
\end{aligned}
$$

where $T_{\mathcal{T}} = \operatorname{argmax}_{t\in\mathcal{T}} Z_t$, $T'_{\mathcal{T}} = \operatorname{argmin}_{t\in\mathcal{T}} Z_t$, and s denotes the sign of g_{10}. Hence, for all $\eta > 0$,

$$
\frac{R_{\mathcal{T}}}{S_{\mathcal{T}}} = \frac{s}{U_{\mathcal{T}}}\,\big(Y_{T_{\mathcal{T}}} - Y_{T'_{\mathcal{T}}}\big) + Q_3 + \mathcal{O}_p\big(\delta^{\alpha/2}\epsilon^{B-(\alpha/2)-\eta} + \delta^{B}\,\epsilon^{-(\alpha/2)-\eta}\big)\,, \tag{2.6}
$$

where Q_3 represents a series of ratios of terms, of the form $V/U_{\mathcal{T}}$, in Taylor expansions (in this sense, each summand is like the first term on the right-hand side of (2.6)), and the $\mathcal{O}_p(\cdot)$ remainder is of the stated order uniformly in $\mathcal{T}$. Note particularly that in forming the leading ratio in (2.6) the contribution $g_{10}(Y_{t_2}, t_2)$ has cancelled from the leading terms in (2.4) and (2.5), and likewise the effect of the constant c (see condition (B) in Section 2.2.4) may be seen to cancel. This results from the scaling aspect of R–S analysis, and explains why the process ζ from which we simulate when applying the bootstrap does not need to reflect either the properties of g or the value of c.

We deal with each ratio, V/W where $W = U_{\mathcal{T}}$, by expressing it as

$$\frac{V^2}{W^2} = \frac{(v + \Delta_V)^2}{w^2} \frac{1}{1 + \Delta_W}$$

where $\Delta_V = V - v$, $\Delta_W = (W^2 - w^2)/w^2$, $v = E(V)$ and $w^2 = E(W^2)$. With the power series expansion of $(1 + x)^{-1/2}$ we get

$$\frac{V}{W} = w^{-1}(v + \Delta_V)\,(1 + \tfrac{1}{2}\Delta_W + \frac{3}{8}\Delta_W^2 + \ldots)\,.$$

For purposes of exposition we shall confine attention to the three main terms in such an expansion, i.e. to $(v/w) + (\Delta_V/w) + \frac{1}{2}v(\Delta_W/w)$, in the case $V = Y_{T_{\mathcal{T}}} - Y_{T'_{\mathcal{T}}}$ and $W = U_{\mathcal{T}}$. (Without loss of generality, $s = 1$.) Other terms may be treated similarly, although the argument is lengthy.

Let Δ_{Va}, Δ_{Wa}, v_a and w_a denote versions of Δ_V, Δ_W, v and w when $\mathcal{T} = \mathcal{I}_a$, the latter defined in Section 2.2.1. Note that, by condition (B), $w_a = w^0\{1 + \mathcal{O}(\xi^\beta)\}$ uniformly in a, where w^0 does not depend on a or n. Since $\beta > \min(\frac{1}{2}, 2 - \alpha)$ (see condition (B)) then $\xi^\beta = o(t_\xi)$. Arguing thus it may be proved that A^{-1} times the sum over $1 \leq a \leq A$ of v_a/w_a equals $C\delta^{\alpha/2}(w^0)^{-1}\{1 + o(t_\xi)\}$, where $C > 0$ is a constant not depending on n.

Put $u = A^{-1}\delta^{-\alpha/2}w^0$, and let $S_\xi(n)$ equal u times the sum over $1 \leq a \leq A$ of the term Δ_{Va}/w_a. Methods of Hall and Wood (1993) may be used to show that the variance of $S_\xi(n)$ is asymptotically equal to a constant multiple of t_ξ^2, and that for the k values of n being considered, the variables $S_\xi(n)/t_\xi$ have a joint asymptotic distribution which is k-variate Normal when $0 < H \leq 3/4$, and k-variate Rosenblatt (Rosenblatt (1961); Taqqu (1875)) when $3/4 < H < 1$.

By considering properties of the variogram estimator of fractal dimension, methods of Constantine and Hall (1994) may be employed to prove that u times the sum over a of $v_a\Delta_{Wa}/w_a$ equals $o_p(t_\xi)$. (Here it is critical that m diverge to infinity.) If B is sufficiently large then u times the sum over a of the $\mathcal{O}_p(\cdot)$ remainder at (2.6) also equals $o_p(t_\xi)$, and similar methods may be applied to terms represented by Q_3 in the Taylor expansion. (The high-order contributions to bias of $\widehat{H}$ include terms of order ξ^α, but since we assumed $\alpha > \frac{1}{2}$ then this equals $o(t_\xi)$.) Arguing thus we may ultimately show that

$$(R/S)_n = C\delta^{\alpha/2}(w^0)^{-1}\{1 + S_\xi(n) + o_p(t_\xi)\}\,.$$

Hence, $\log(R/S)_n$ equals a quantity which does not depend on n and which goes into the intercept term in the regression, plus $(\alpha/2)\log n + S_\xi(n) + o_p(t_\xi)$. The result asserted in section 2.2.3 follows from this property. □

2.3 Empirical Analysis

To justify statistical methods based on the semimartingal property of asset prices in the following chapters, we test the null hypothesis $H_0 : H = 0.5$ for a variety of asset price processes. Thus, the aim of this section is to obtain an estimate $\widehat{H}$ of the Hurst coefficient H and to construct hypothesis tests and confidence intervals for H for the logarithm of the price process of certain German stocks.

Denote the logarithm of the price process of a stock (or index) by $\{X_t : 0 \leq t \leq T\}$. To estimate the Hurst coefficient H we apply R–S analysis, as described in Section 2.2.1, to N discrete observations $\{X_n : n = 1, \ldots, N\}$ of $\{X_t\}$ at times $t_1 \leq t_2 \leq \ldots \leq t_N$,

For the empirical study we used 6900 observations ($N = 6900$) of 24 German blue chip stocks obtained form the Datastream/Primark's database from 8th of January 1973 to the 18th of June 1999. The blue chips are included in the DAX, an index comprising 30 German stocks. We analysed Datastream performance indices instead of prices in order to avoid jumps in the respective time series due to dividend payments or rights issues. The obtained Hurst coefficients are shown in Table 2.1. Figure 2.2 shows the R–S plot for the price process of the stock of Volkswagen. The R–S plot also includes a line with slope 0.5, which correspond to Brownian motion. As one can

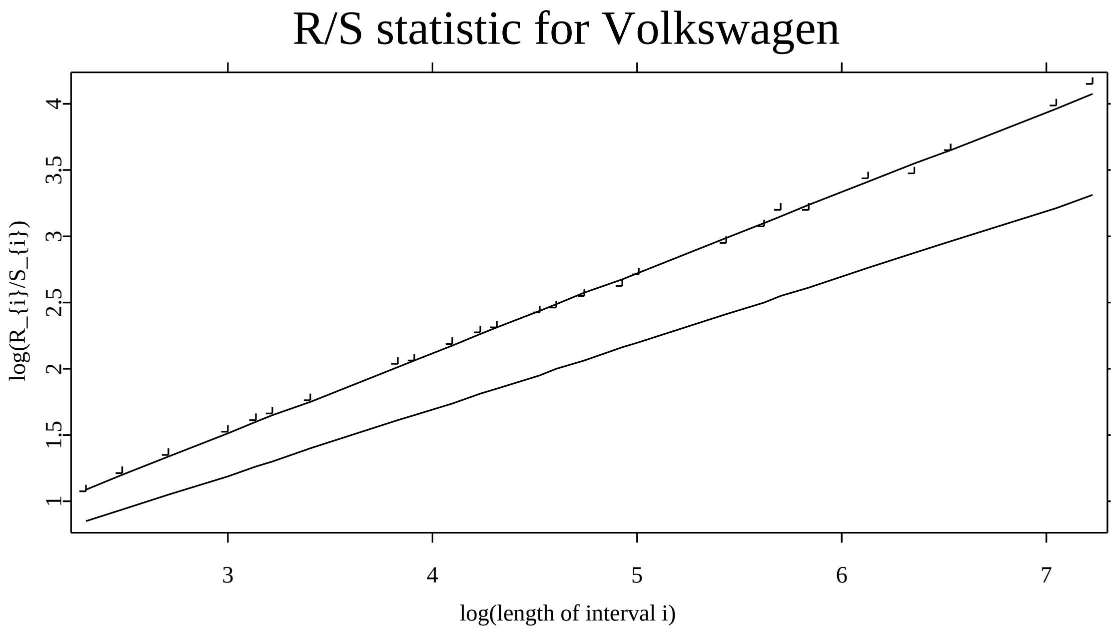


Figure 2.2: R–S plot for VW, $\widehat{H} = 0.606$

see, the R–S line has a different slope then it would have if the underlying process corresponded to a Brownian Motion.

In the first step of our empirical analysis we tested whether the Hurst coefficient of an asset was significantly different from 0.5 or not. A significant difference from 0.5 would indicate that X_t did not follow a Brownian Motion. In order to test the null hypothesis that $H = 0.5$, against the alternative $H \neq 0.5$, i.e.

$$h_0: \; H = 0.5 \qquad h_1: \; H \neq 0.5\,,$$

we approximated the distribution of $\widehat{H} - H$ conditional on the null hypothesis, and calculated the p-values, $P\{|\widehat{H} - E\widehat{H}| > |H_{\text{observed}} - E\widehat{H}| \mid h_0\}$, of the estimated $\widehat{H}$. For this approximation the bootstrap algorithm described in Section 2 was used. For $H = 0.5$ the fractional Brownian Motion coincided with usual Brownian Motion, which we simulated as a random walk. An estimate of the conditional density of $\widehat{H}^* - \widehat{H}$, computed from 400 simulated random walks of length 6900, is shown in figure 2.3. Table 2.1 shows the p-values for the estimated Hurst coefficient of the

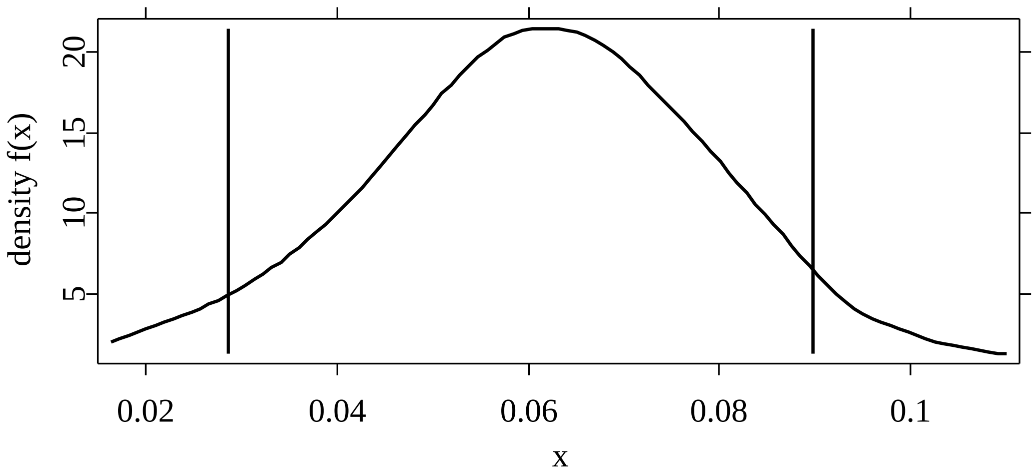


Figure 2.3: Estimated density of $\widehat{H} - H$ for 400 simulated Brownian Motions with length 6900. The vertical lines determine the 0.05, 0.95 quantiles.

stocks.

Our analysis suggests that the difference between the estimated Hurst index of the prices of BMW, Daimler, Mannesmann, Preussag, Siemens and Volkswagen, and the value the Hurst index would take if the stochastic process describing prices were Brownian motion, is so great that it cannot be adequately explained by stochastic fluctuations.

We studied the assets for which the estimated Hurst coefficient H was significantly different from 0.5. For our further analysis we assumed that the logarithm of the price processes are self similar with stationary increments, i.e.

$$c^{-H}(X_{ct})_{t\in\mathbb{R}} =_d (X_t)_{t\in\mathbb{R}} \quad \text{for all} \quad c > 0 \tag{2.7}$$

and for any $k \geq 1$ and any time points $t_1, \ldots, t_k$,

$$(X(t_1), \ldots, X(t_k)) =_d (X(t_1 + c), \ldots, X(t_k + c)) \quad \text{for all} \quad c \in \mathbb{R} \tag{2.8}$$

Here, $Y =_d Z$ means that Y and Z have the same distribution. These assumptions are often made in literature on financial market analysis. A well known model is the Multifractal Model of Asset Returns (MMAR) introduced by Calvet et al. (1997). In this model the logarithms of prices are assumed to follow a fractional Brownian Motion, i.e.

$$X(t) - X(0) = B_H(\theta(t)),$$

where $\theta(t)$ is a multifractal process with continuous, non-decreasing paths and stationary increments.

Under assumptions 2.7 and 2.8 the autocorrelation function $\rho(k) = E[\{X(t) - EX(t)\}\{X(t+k) - EX(t+k)\}]$ of $X(t)$ is approximately of the form ck^{2H-2}. More precisely, the following holds, Beran (1994):

$$\frac{\rho(k)}{H(2H-1)k^{2H-2}} \longrightarrow 1 \qquad 0 < H < 1,\ H \neq \frac{1}{2}, \qquad k \longrightarrow \infty.$$

This means that for $\widehat{H} > 0.5$, X_t has long memory. Stocks where long memory was detected are displayed in bold face in table 2.1.

The second step of our analysis was construction of confidence intervals. For this purpose we approximated the distribution of $\widehat{H} - H$ by that of $\widehat{H}^* - \widehat{H}$, where $\widehat{H}^*$ denotes the estimated value of the Hurst coefficient of simulated fractional Brownian Motions with coefficient $\widehat{\alpha} = 2\widehat{H}$. That is, we computed the conditional (on $X(t)$) distribution of the bootstrap form of $\widehat{H}^* - \widehat{H}$, as an approximation to the unconditional distribution of $\widehat{H} - H$. We applied the bootstrap method described in Section 2.2. To simulate fractional Brownian motion we used methods from Section 2.2.2 with $p = 1$ as well as the algorithm described in Beran (1994), p. 216. The latter is based on the finite Fourier transform of the autocovariance function of fractional Gaussian noise. Both methods lead to similar results.

The bootstrap densities for the different Hurst values of the assets which have significantly larger Hurst coefficient than a Brownian Motion were approximately the same except for the mean value. For this reason we calculated only the density of $\widehat{H}^* - \widehat{H}$ of the Volkswagen stock. It is shown in figure 2.4. The confidence intervals for the other assets were obtained by correcting this density for the different estimated Hurst coefficient. Table 2.2 shows the resulting confidence regions.

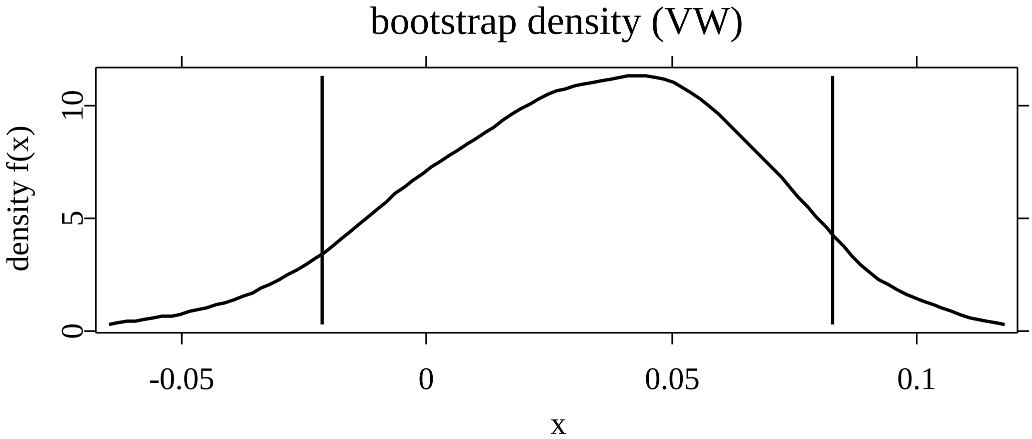


Figure 2.4: Bootstrap density of $\widehat{H} - H$ for the Volkswagen stock. The vertical lines determine the 0.05, 0.95 quantiles.

asset	$\widehat{H}$	p-value
Allianz	0.5642	0.6
BASF	0.5390	0.24
Bayer	0.5288	0.073
BMW	**0.5851**	**0.05**
Commerzbank	0.5536	0.88
Dt. Bank	0.5743	0.22
Daimler	**0.5859**	**0.05**
Degussa Hüls	0.5629	0.68
Dresdner Bank	0.5625	0.7
Hoechst	0.5420	0.37
HypoVereinsbank	0.5533	0.86
Karstadt	0.5552	0.95
Lufthansa	0.5584	0.89
Linde	0.5583	0.90
MAN	0.5605	0.79
Mannesmann	**0.5856**	**0.05**
Münchner Rück NA	0.5589	0.88
Preussag	**0.5884**	**0.035**
RWE	0.5398	0.29
Schering	0.5772	0.17
Siemens	**0.6007**	**0**
ThyssenKrupp	0.5794	0.13
Veba	0.5426	0.38
Volkswagen	**0.6049**	**0**

Table 2.1: Estimated Hurst coefficient of German stocks

asset	0.9 confidence region	0.95 confidence region
BMW	$[0.475, 0.579]$	$[0.466, 0.594]$
Daimler	$[0.476, 0.581]$	$[0.467, 0.596]$
Mannesmann	$[0.476, 0.580]$	$[0.467, 0.596]$
Preussag	$[0.481, 0.585]$	$[0.472, 0.601]$
Siemens	$[0.506, 0.610]$	$[0.497, 0.626]$
Volkswagen	$[0.514, 0.619]$	$[0.505, 0.634]$

Table 2.2: Confidence regions for Hurst coefficients

Chapter 3

Testing Diffusion Models

We will now assume, that the observed processes are semimartingales and thus do not contradict the no arbitrage condition, see Chapter 1. In particular we concentrate on diffusion processes. These processes are Markovian semimartingales and have almost surely continuous paths.

The chapter is organized as follows. We introduce the model and the available observations in Section 3.1. The null hypotheses are given in Section 3.2. In Section 3.3 we introduce testing procedures that are based on a discrete approximation of the continuous time process X by a time series. Section 3.4 captures nonparametric estimation methods for the marginal density, the drift and the diffusion coefficient of the continuous time model. Finally we introduce in Section 3.5 different tests about the dynamics of X based on the proposed estimators.

3.1 Model and Observations

Formally, we assume that the log price process of an underlying or an interest rate process is an one-dimensional diffusion $\{X(t),\ t \in [0,T]\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P}, \{\mathcal{F}_t\}_{t\in[0,T]})$. X is given as a strong solution of the stochastic differential equation

$$dX(t) = m\{X(t)\}dt + \sigma\{X(t)\}dW(t) \qquad t > 0 \tag{3.1}$$

where m and σ are smooth function, such that a unique strong solution of (3.1) exists and $\{W(t),\ t \in [0,T]\}$ is a standard Brownian Motion adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$. Furthermore we assume that σ^2 has continuous derivatives up to the second order. Conditions for the existence and uniqueness of a solution of (3.1) are given in Appendix A.1.

The dynamics of the process X are fully specified by the functional form of the drift coefficient m and the diffusion coefficient σ. The aim of this chapter is to test parametric models for both functions.

We mention, that both coefficient functions, σ^2 and m, are assumed to depend only on the current state $X(t)$ of X. When we discuss tests about a parametric form of σ^2 we will show, how these tests can be extended to the case, when σ also depends on time t. Basically this extension is realized via a transformation of X. However, for the drift function such a transformation does not exist. Other arbitrage free models to capture the dynamics of financial time series have been proposed in the literature. The basic concepts are stochastic volatility models and stochastic delay equations. In the first kind of models, an additional random process, the volatility process, influences the diffusion coefficient and/or the drift of X. This additional source of randomness yields an incomplete market model, where option prices are no longer unique. In the second approach, non Markovian semimartingales are used, i.e. both functions, the drift m and the diffusion coefficient σ depend on the path history of X, Hobson and Rogers (1998).

In order to make the proposed tests applicable to observed data, all tests and estimation methods are based on discrete observations of X. Thus, we assume that the diffusion process X is observed at equidistant time points

$$0, 1/n, \ldots, ([Tn]-1)/n, [Tn]/n$$

where $[a]$ denotes the integer part of a. From this discretization scheme we see that two kinds of asymptotics results coexist in the statistics of continuous time processes. The first kind is based on $n \to \infty$. This means that the number of observations per unit of time tends to infinity and due to the assumption of equidistant observations, the time distance between two observations is shrinking to 0. The second kind of asymptotics is $T \to \infty$, i.e. the time horizon, until which observations are available, is tending to infinity. Heuristically speaking, $n \to \infty$ is used when we derive asymptotic results about the diffusion coefficient σ^2. The second kind of asymptotics, $T \to \infty$ is used when we derive estimators or tests for the drift coefficient m or the marginal density of X. The results of this chapter rely on one or on both kinds of asymptotics. Thus we will specify the assumptions about T and n for every particular method.

Before we introduce statistical methods for X we give a few definitions.

DEFINITION 3.1 *We call a stochastic process X stationary, if and only if, the distribution of $(X(t_1), \ldots, X(t_k)$ is the same as the distribution of $(X(t_1+s), \ldots, X(t_k+s)$ for any $s > 0$, and $t_1 \geq 0, \ldots, t_k \geq 0$ and any $k = 1, 2, \ldots$.*

DEFINITION 3.2 *The process X given as the solution of (3.1) is α-mixing, if*

$$\alpha(u) \stackrel{\text{def}}{=} \sup_{\mathcal{A} \in \mathcal{F}_t; \mathcal{B} \in \mathcal{F}_{t+u}^{\infty}} |P(AB) - P(A)P(B)| \to 0$$

for $u \to \infty$. Here $\mathcal{F}_t^{\infty}$ denotes the σ-algebra generated by $\{(X_u), u \geq t\}$. We call the process X geometrically α-mixing if

$$\alpha(u) \leq a\rho^u$$

for some $a > 0$ and $\rho \in [0, 1)$.

For an introduction into α-mixing processes, see Bosq (1998) or Billingsley (1968).

Since all estimators and tests are restricted to the range where X is observed, we introduce the definition of the local time of X up to time $t \in [0, T]$.

DEFINITION 3.3 *For a diffusion X we define*

- *occupation measure ν_t: $\nu_t(B, \omega) \stackrel{\text{def}}{=} \int_0^t \boldsymbol{I}_B\{X(u, \omega)\}du$*
- *Local Time: $L_t(., \omega) \stackrel{\text{def}}{=} \frac{d\nu_t}{d\lambda}$ for $P-a.e.$ $\omega \in \Omega$ where λ is the Lebesque measure on $\mathbb{R}$.*

This definition is given in Bosq (1998). Heuristically speaking, the occupation measure $\nu_t(B)$ measures the time, that the process X spends in the set B up to time t and the local time $L_t(x)$ measures the time, that X spends in a neighborhood of x.

Using the local time we can now restrict the range on which we estimate parameters of X and test particular models. For this reason we use the notation

$$I_X \stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid L_T(x) > \varepsilon_L\} \tag{3.2}$$

for an arbitrary $\varepsilon_L > 0$. In the particular case of an ergodic process X, see Appendix A.1, we have from the ergodic theorem that the marginal density f of X is given by

$$f(x) = \lim_{t \to \infty} \frac{1}{t} L_t(x) \tag{3.3}$$

Hence, may rewrite the above notation as

$$I_X = \{x \in \mathbb{R} \mid f(x) > \varepsilon_f\} \tag{3.4}$$

where ε_f is a positive constant. For a stationary α-mixing process both expressions of I_X are asymptotically equivalent for $T \to \infty$. However, when we test the diffusion coefficient we do not assume that X is stationary and thus we use (3.2) as the definition of I_X. On a heuristic level we may think of I_X as the set of all points $x \in \mathbb{R}$ that the process X has visited up to time T. We will restrict all estimators and tests on this set, since we are only interested in the behavior of m and σ at points, where observations are available.

3.2 Hypotheses

As already mentioned in the introduction, option prices and risk measures depend on the particular dynamics of X. These dynamics are specified by a functional form of the drift function m and the diffusion coefficient σ in (3.1). In mathematical finance it is often assumed, that these functional forms are known up to a finite dimensional parameter θ, i.e. $m(x) = m(\theta, x)$ and $\sigma(x) = \sigma(\theta, x)$ with a usually unknown parameter θ.

It is the aim of this chapter to introduce goodness-of-fit tests, to verify the parametric models. Hence, we will test two null hypotheses.

Null hypothesis about the drift

$$\begin{aligned} H_0(m) \; : \; & \exists \theta_0 \in \Theta \; : \; \text{for every } t \in [0,T] : \\ & m\{X(t)\} = m\{\theta_0, X(t)\} \qquad \text{P-a.s.} \end{aligned} \tag{3.5}$$

Null hypothesis about the squared diffusion coefficient

$$\begin{aligned} H_0(\sigma^2) \; : \; & \exists \theta_0 \in \Theta \; : \; \text{for every } t \in [0,T] : \\ & \sigma^2\{X(t)\} = \sigma^2\{\theta_0, X(t)\} \qquad \text{P-a.s.} \end{aligned} \tag{3.6}$$

We mention, that the particular choice of the parameter θ is not part of the null hypotheses. For instance, the null hypothesis about σ^2 is true, if there is any parameter, such that the deviation between the true diffusion coefficient and the diffusion coefficient function as implied by the null hypothesis and the particular choice of the parameter, can be explained by random fluctuations.

Since m and σ fully specify the dynamics of the Markovian process X, the two hypotheses about the coefficients of X can be replaced by hypothesis about its transition density. Such an approach is outlined by Hong and Li (2002). However, in many applications a particular form for the coefficients need to be tested and a closed form for the transition density is not known. In addition, tests about the transition density require estimation methods for the drift function or for the parameters of it. Since these methods usually rely on the ergodic theorem, the class of possible models is restricted to stationary processes.

3.3 Tests Based On A Discrete Time Approximation

In this section we develop tests that are based on a time series approximation of X. The results are mainly based on the work of Härdle et al. (2001), Chen et al.

(2001) and Chen et al. (2002). We start with a discrete time approximation X^Δ of X and then introduce statistical estimation and testing procedures for X^Δ. The proposed test statistics directly compare nonparametric estimates of the drift m and the diffusion coefficient σ^2 to their parametric forms as implied by the null hypotheses. The first test is based on the L_∞ norm, while the second test, the Empirical Likelihood test, is equivalent to a studentized L_2 distance between the nonparametric and the parametric estimates of the drift.

3.3.1 Discrete Time Approximation

From Theorem 10.2.2., its proof and Remark 10.2.3. in Kloeden and Platen (1999) we know, that under a few technical assumptions on the functions m and σ^2, the Euler approximation

$$X^\Delta(t) = X^\Delta(0) + \int_0^t m\big\{X^\Delta(t_{i_s})\big\}ds \; + \int_0^t \sigma\big\{X^\Delta(t_{i_s})\big\}dW(s) \qquad t \in [0,T] \quad (3.7)$$

with $t_{i_s} = \max\{t_i \,,\, t_i \leq s\}$ and $t_i = i/n$ for $i = 0, \dots, [nT]$, converges in a mean square sense to X as $\Delta = 1/n \to 0$, i.e.,

$$\lim_{n\to\infty} \mathrm{E}(\sup_{0\le t\le T} |X^\Delta(t) - X(t)|^2) = 0,\; T > 0. \quad (3.8)$$

From now on, we assume that a discrete time approximation X^Δ exists in the form of (3.7) and that property (3.8) is satisfied. For the purposes of this section, $\Delta = 1/n$ will always be considered small enough that one can substitute X by X^Δ in our interpretation of the observed data. The increments of the Euler approximation (3.7) and so the observed data will have the form

$$X^\Delta(t_{i+1}) - X^\Delta(t_i) = m\big\{X^\Delta(t_i)\big\}\Delta \; + \sigma\big\{X^\Delta(t_i)\big\}\big\{W(t_{i+1}) - W(t_i)\big\} \quad (3.9)$$

for $i = 0, 1, \dots [nT]-1$. The observations $\{X^\Delta(t_i)\},\ i = 0, 1, \dots [nT]-1$ form a time series. As long as the step size Δ is small enough the concrete choice of Δ does not matter since all the relevant information about the model is contained in the drift m and diffusion coefficient σ.

For the following we introduce the notation

$$\begin{aligned} X_i &\stackrel{\text{def}}{=} X^\Delta(t_i)\,, \qquad \mathcal{X} \stackrel{\text{def}}{=} (X_0 \dots, X_{[nT]-1}) && (3.10)\\ \varepsilon_i &\stackrel{\text{def}}{=} W(t_{i+1}) - W(t_i)\,, \qquad \varepsilon \stackrel{\text{def}}{=} (\varepsilon_0, \dots, \varepsilon_{[nT]-1}) \\ Y_i &\stackrel{\text{def}}{=} X_{i+1} - X_i = m\big(X_i\big)\Delta \; + \sigma\big(X_i\big)\sqrt{\Delta}\varepsilon_i\,, \end{aligned}$$

$$\mathcal{Y} \stackrel{\text{def}}{=} (Y_0, \dots, Y_{[nT]-1}) \quad (3.11)$$

with independent standard Gaussian random variables

$$\varepsilon_i = \frac{W(t_{i+1}) - W(t_i)}{\sqrt{\Delta}} \sim \mathrm{N}(0,1).$$

The same discretization scheme is applied by Hoffmann (1999). He uses this method to develop an adaptive nonparametric estimation procedure for σ^2 using wavelets.

We remark, that we assume here that the observed data are not generated by the continuous time model (3.1), but the discrete time process X^Δ as given in (3.9). We argue, that for statistical methods the difference between the two models is asymptotically negligible and thus all empirical results for the discrete time model are also valid for the corresponding continuous time process. In Section 3.4 we redefine X_i since we will not apply a discrete time approximation there and interpret the observations as realizations of the continuous time process X.

3.3.2 Estimation of the Drift and Diffusion Coefficient

We will now introduce methods to estimate the drift and diffusion coefficient of X nonparametrically. For the applied methods it is necessary, that the following condition is fulfilled, which we assume for the remainder of the section.

(DT1) X, as given by (3.1) is stationary and geometrically α-mixing, and thus X is ergodic.

Genon-Catalot et al. (2000) show, that both $\{X_i, i = 0, \dots, [nt]\}$ and $\{(X_{i+1} - X_i)^2, i = 0, \dots, [nt]-1\}$ are α-mixing if the continuous time process X is α-mixing, see Definition 3.2.

Let K be a bounded probability density function with a compact support on $[-1,1]$ that satisfies the moment conditions:

$$\mu_1(K) \stackrel{\text{def}}{=} \int uK(u)du = 0, \qquad \mu_2(K) \stackrel{\text{def}}{=} \int u^2K(u)du < \infty\,. \tag{3.12}$$

Let h be a positive smoothing bandwidth which will be used to smooth $(\mathcal{X}, \mathcal{Y})$ given in (3.10) and (3.11). In the following we use the notation

$$K_h(u) \stackrel{\text{def}}{=} \frac{1}{h} K\left(\frac{u}{h}\right)\,. \tag{3.13}$$

The nonparametric estimator considered is the local linear estimator, Fan and Gijbels (1996) and Härdle and Tsybakov (1997). To justify the use of this method, the time series X_i has to be ergodic and to meet some technical conditions, see Härdle and Tsybakov (1997). As mentioned above, the ergodicity of X_i follows from the ergodicity of X, which we assume here. The square root process and the Ornstein-Uhlenbeck process are examples, where the discrete time approximation (3.7) satisfies the technical conditions. Mainly, these conditions are, that the marginal distribution has a bounded, strictly positive density function and that the drift function and the diffusion coefficient are twice continuously differentiable.

The local linear estimator of the drift m at $x \in I_X$ is obtained by

$$\hat{m}_h(x) = \frac{1}{\Delta}\,\hat{\beta}_0(x) \tag{3.14}$$

with

$$\begin{aligned}
\hat{\beta}(x) &= \begin{pmatrix} \hat{\beta}_0(x) \\ \hat{\beta}_1(x) \end{pmatrix} \\
&= \operatorname{argmin}_{b_0,b_1} \left(\sum_{i=0}^{[nT]-1} \Big\{ Y_i - b_0 - b_1(X_i - x) \Big\}^2 K_h(x - X_i) \right).
\end{aligned}$$

The bandwidth $h > 0$ is chosen with respect to the Silvermans rule of thumb, see Härdle (1990).

We apply for the squared diffusion function $\sigma^2(x)$ a two-step estimation. First we compute from the above drift function estimator (3.14) the values $\hat{m}_h\{X_i\}$. In the second step we use the squared diffusion function estimator

$$\hat{\sigma}_h^2(x) = \frac{1}{\Delta}\,\hat{\delta}_0(x) \tag{3.15}$$

with

$$\begin{aligned}
\hat{\delta}(x) &= \begin{pmatrix} \hat{\delta}_0(x) \\ \hat{\delta}_1(x) \end{pmatrix} \\
&= \operatorname{argmin} \left(\sum_{i=0}^{[nT]-1} \Big\{ \big(Y_i - \Delta \hat{m}_h\{X_i\}\big)^2 \right. \\
&\qquad \left. - \delta_0 - \delta_1(X_i - x) \Big\}^2 K_h(x - X_i) \right)
\end{aligned}$$

and bandwidth $h > 0$.

The asymptotic properties of local polynomial estimates are studied in Fan and Gijbels (1996) and Härdle et al. (1999). Under some smoothness conditions with bandwidth $h = k_0/n^{1/5}$ for a constant $k_0 > 0$, the results applied to our case provide the following formulas on the asymptotic normality

$$n^{2/5}\,\{\hat{m}_h(x) - m(x)\} \stackrel{\mathcal{D}}{\rightarrow} N\left(\frac{k_0^2}{2}\mu_2(K)\,\Delta\, m''(x)\,,\; \frac{\|K\|_2^2\,\Delta\,\sigma^2(x)}{k_0 f(x)} \right) \tag{3.16}$$

$$n^{2/5}\left\{\hat{\sigma}_h^2(x)-\sigma^2(x)\right\} \xrightarrow{\mathcal{D}} N\left(\frac{k_0^2}{2}\mu_2(K)\left(\Delta\left(\sigma^2(x)\right)''+2(\Delta m'(x))^2\right), \frac{2\Delta^2\sigma^4(x)\|K\|_2^2}{k_0 f(x)}\right). \tag{3.17}$$

Here $\mu_2(K)$ is the second moment of the kernel K and

$$\|K\|_2 \stackrel{\text{def}}{=} \left(\int_{[-1,1]} K^2(x)dx\right)^{1/2}$$

is its L_2 norm. Furthermore $f(x)$ denotes the stationary density of X as given in Section 3.3.1.

3.3.3 Testing the Parametric Model

We construct tests to compare the nonparametric estimates introduced in 3.3.2 for $m\{.\}$ and $\sigma^2\{.\}$ to parametric forms implied by the null hypotheses. The nonparametric estimators we apply, are the local polynomial estimators in (3.14) and (3.15). The Nadaraya-Watson estimator is not used here, but will be applied in 3.3.4 when we incorporate the Empirical Likelihood methodology.

The tests in this section are based on pointwise confidence bands which we build with the bootstrap method. The idea is to bootstrap the original discrete time series and estimate each time the drift and squared diffusion coefficients nonparametrically as described in Section 3.3.2. With these estimates one can then construct pointwise confidence bands for the two functions.

We choose the bootstrap method because it leads to better coverage probabilities than, for instance, a Gaussian approximation. In Neumann and Kreiss (1998) it was shown for a time series similar to (3.7) that the coverage probability is of order $\mathcal{O}([nt]^{-q})$ for some $q > 0$, where $[nt]$ is the number of observations. A Gaussian approximation, see Hall (1985), leads to a coverage probability of order $\mathcal{O}(1/\ln([nt]))$.

Since the residuals ε_i of the local linear regression

$$Y_i = m(X_i) + \varepsilon_i$$

are not identically distributed, a naive resampling method will not mimic the true distribution of ε_i, see Härdle and Mammen (1993). Instead of naive resampling they suggest to apply a wild bootstrap procedure. For the wild bootstrap we calculate the residuals

$$\hat{\varepsilon}_i \stackrel{\text{def}}{=} Y_i - \hat{m}_h(X_i)$$

and use one observations for each residual ε_i to estimate the conditional distribution of $Y_i - m(X_i)$ given X_i. More precisely Härdle and Mammen (1993) define a two-point distribution $\hat{F}_i$ with expectation zero and with second and third moments given by $\hat{\varepsilon}_i^2$ and $\hat{\varepsilon}_i^3$ respectively. With this distribution we are now able to construct independent replications $\varepsilon^* \sim \hat{F}_i$ and use them to simulate new replications of the time series X_i^*.

The confidence bands for the nonparametric estimators are constructed in the following way, Härdle et al. (2001)

1. Choose a bandwidth g, which is larger than the optimal h in order to have oversmoothing. Estimate then nonparametrically $m(.)$ and $\sigma^2(.)$ and obtain the residual estimated errors:

$$\hat{\varepsilon}_i = \frac{Y_i - \Delta \hat{m}_g\{X_i\}}{\sqrt{\Delta}\hat{\sigma}_g\{X_i\}}.$$

   Since we make the assumption that the ε_i has zero-mean, we subtract the sample mean of $\hat{\varepsilon}_i$.

2. Replicate N times the series of the $(\hat{\varepsilon}_i)$ with wild bootstrap obtaining $(\varepsilon_i^{*,n})$ for $n = 1, \ldots, N$ and build N new bootstrapped series $(X^{*,n})$:

$$X_1^{*,n} = X_1$$

$$X_{i+1}^{*,n} - X_i^{*,n} = \Delta \hat{m}_g(X_i^{*,n}) + \sqrt{\Delta}\hat{\sigma}_g(X_i^{*,n})\varepsilon_i^{*,n}.$$

   Estimate again $m(z)$ and $\sigma^2(z)$ for each of the N bootstrapped series with bandwidth h.

3. Build the statistics:

$$T_m^* = \sup_z \frac{|\hat{m}_h^{*,n}(z) - \hat{m}_h(z)|}{\hat{\sigma}_h^{*,n}(z)}$$

   and

$$T_\sigma^* = \sup_z |(\hat{\sigma}^2)_h^{*,n}(z) - \hat{\sigma}_h^2(z)|$$

   .

4. Form the $(1-\alpha)$ confidence bands CB

$$CB(m(.)) = [\hat{m}_h(z) - \hat{\sigma}_h(z)t_{m,1-\alpha}\ ,\ \hat{m}_h(z) + \hat{\sigma}_h(z)t_{m,1-\alpha}]$$

   and

$$CB(\sigma^2(.)) = \left[\hat{\sigma}_h^2(z) - t_{\sigma,1-\alpha}\ ,\ \hat{\sigma}_h^2(z) + t_{\sigma,1-\alpha}\right]$$

   where $t_{m,1-\alpha}$ and $t_{\sigma,1-\alpha}$ denote the empirical $1-\alpha$-quantile of T_m^* and T_σ^*, respectively.

The asymptotic results for the $(1-\alpha)$ confidence bands $CB(m)$ and $(CB(\sigma^2))$, that is

$$P\left\{m(z) \in CB(m)\right\} \rightarrow 1-\alpha$$

and

$$P\left\{\sigma^2(z) \in CB(\sigma^2)\right\} \rightarrow 1-\alpha$$

respectively, are proven in Franke et al. (1998).

3.3.4 Empirical Likelihood Tests

The second test in this section is about a parametric form of the drift m of the time series $\mathcal{X}$ given in (3.7). We apply here the Empirical Likelihood (EL) methodology since it internally studentizes the test statistic and also captures features of its empirical distribution. In the context of goodness-of-fit tests for time series models, the EL concept was first introduced by Chen et al. (2001) and was applied to a discrete time approximation of a diffusion by Chen et al. (2002). We first give a short introduction into the Empirical Likelihood concept for independent and identically distributed data and then expand the results to the case of time series observations.

Introduction into Empirical Likelihood

Let us now as in Owen (1988) and Owen (1990) introduce the empirical likelihood (EL) concept. For a detailed discussion of EL tests and confidence bands we refer to Owen (2001). Suppose a sample $(U_1, \ldots, U_N)$ of independent identically distributed random variables in $\mathbb{R}^1$ according to a probability law with unknown distribution function F and unknown density f. For an observation $(u_1, \ldots, u_N)$ of $(U_1, \ldots, U_N)$ the likelihood function is given by

$$\bar{L}(f) = \prod_{i=1}^{N} f(u_i) \tag{3.18}$$

The empirical density calculated from the observations $(u_1, \ldots, u_N)$ is

$$f_N(u) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{I}\{u_i = u\} \tag{3.19}$$

where $\boldsymbol{I}$ denotes the indicator function. It is easy to see that f_N maximizes $\bar{L}(f)$ in the class of all probability density functions.

The objective of the empirical likelihood concept is the construction of tests and confidence intervals for a parameter $\theta = \theta(F)$ of the distribution of U_i. To keep things simple we illustrate the empirical likelihood method for the expectation $\mathrm{E}[U_i]$.

The null hypothesis is $\mathrm{E}[U_i] = \theta$. We can test this assumption based on the empirical likelihood ratio

$$R(F) \stackrel{\text{def}}{=} \frac{\bar{L}\{f(\theta)\}}{\bar{L}(f_N)} \tag{3.20}$$

where $f(\theta)$ maximizes $\bar{L}(f)$ subject to

$$\int U_i dF = \theta. \tag{3.21}$$

On a heuristic level we can reject the null hypothesis "under the true distribution F, U has expectation θ" if the ratio $R(F)$ is small relative to 1, i.e. the test rejects if $R(F) < r$ for a certain level $r \in (0, 1)$. More precisely, Owen (1990) proves the following theorem.

THEOREM 3.1 *Let $(U_1, \ldots, U_N)$ be independent and identically distributed one-dimensional random variables with expectation θ and variance σ^2. For a positive $r < 1$ let*

$$C_{r,N} = \left\{ \int U_i dF \;\middle|\; F \ll F_N, R(F) \geq r \right\}$$

be the set of all possible expectations of U with respect to distributions F dominated by F_N $(F \ll F_N)$. Then it follows

$$\lim_{N\to\infty} P[\theta \in C_{r,N}] = P[\chi^2 \leq -2 \log r] \tag{3.22}$$

where χ^2 is a χ^2-distributed random variable with one degree of freedom.

For the log EL ratio

$$L^R \stackrel{\text{def}}{=} -2 \log \left\{ \max_{\{F|F\ll F_N, \int u dF = \theta\}} R(F) \right\} = -2 \log \left\{ \max_{\{F|F\ll F_N, \int u dF = \theta\}} \frac{\bar{L}\{f(\theta)\}}{\bar{L}(f_N)} \right\}$$

follows directly from Theorem 3.1

$$\begin{aligned} \lim_{N\to\infty} \mathrm{P}\left[L^R \leq r \;\middle|\; \mathrm{E}U_i = \theta\right] &= \lim_{N\to\infty} \mathrm{P}\left[\max_{\{F|F\ll F_N, \int u dF = \theta\}} R(F) \geq -\frac{1}{2} e^r \;\middle|\; \mathrm{E}U_i = \theta \right] \\ &= \lim_{N\to\infty} \mathrm{P}[\theta \in C_{-(1/2)\exp r, N}] \\ &= \mathrm{P}[\chi^2 \leq r] \end{aligned}$$

This result suggests therefore to use the log-EL ratio L^R as the basic element of a test about a parametric hypothesis for the drift function of a time series or a diffusion process.

Empirical Likelihood Test About The Drift

We will now expand the results in 3.3.4 to the nonparametric drift estimation of the time series X_i, $i = 0, \ldots, [nT]$ based on $[nT]$ observations up to time T. An application of the EL methodology to the testing of the squared diffusion coefficient is given in 3.5.3. In a time series context an EL test about a parametric model of the drift of a time series is proposed by Chen et al. (2001) and in a diffusion context by Chen et al. (2002). We will follow the results of Chen et al. (2001). For the sake of simplicity we assume that $I_X = [0, 1]$. The asymptotic results we are going to show rely of the assumption that T tends to infinity, i.e. the length of the time interval where observations are available is increasing. We do not make any assumptions about n except, that n is large enough, such that the approximation of X by the time series $X_i, i = 0, \ldots, [nT] - 1$ as introduced in 3.3.1 is valid. The nonparametric estimator we apply here is the Nadaraya-Watson estimator

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} Y_i K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \,. \tag{3.23}$$

Let

$$\tilde{m}_{\hat{\theta}}(x) = \frac{\sum K_h(x - X_i) m_{\hat{\theta}}(X_i)}{\sum_{i=1}^{n} K_h(x - X_i)}$$

be the smoothed parametric model. The test statistic we are going to consider is based on the difference between $\tilde{m}_{\hat{\theta}}$ and $\hat{m}$, rather than directly between $\hat{m}$ and $m_{\hat{\theta}}$, in order to avoid the issue of bias associated with the nonparametric fit.

The local linear estimator, as given in 3.14, can be used to replace the Nadaraya-Watson estimator in estimating m. The local linear estimator is known for its attractive bias properties. However, as we compare $\hat{m}$ with $\tilde{m}_{\hat{\theta}}$ in formulating the goodness-of-fit test, the possible bias associated with the NW estimator is not an issue here. In addition, the NW estimator has a simpler analytic form. Extensions of the results to the local linear estimator based test can be derived in a similar fashion, although the proofs will be more involved.

Hypotheses about the drift

$$\begin{aligned} H_0(m) \;&:\; \exists \theta_0 \in \Theta \;:\; \text{for every } t \in [0, T] :\; m\{X(t)\} = m\{\theta_0, X(t)\} \qquad \text{P-a.s.} \\ H_1(m) \;&:\; \forall\, \theta \in \Theta \;:\; \text{for every } t \in [0, T] : \\ &\qquad |m\{X(t)\} = m\{\theta, X(t)\}| \geq c_T \Delta_T(X(t)) \qquad \text{P-a.s.} \end{aligned}$$

Δ_T, the local shift in the alternative, is a sequence of bounded functions and c_T is the order of difference between H_0 and H_1. This choice of the alternative ensures that the power of the proposed test depends on the number of observations n. In particular we will assume, that c_T tends to 0 with $n \to \infty$. This means, that the tests can better distinguish between the null hypothesis and the alternative when the number of observations is large, i.e. the power of the test depends on the number of observations.

Before we start with the derivation of the test statistic for the goodness-of-fit test for the drift m, we introduce the following set of assumptions in addition to (DT1).

(DT2) The kernel K is Lipschitz continuous in $[-1,1]$, that is $|K(t_1) - K(t_2)| \leq C||t_1 - t_2||$ where $||\cdot||$ is the Euclidean norm, and $h = \mathcal{O}\{T^{-1/5}\}$;

(DT3) f, m and σ^2 have continuous derivatives up to the second order in S.

(DT4) $\hat{\theta}$ is a parametric estimator of θ within the family of the parametric model, and

$$\sup_{x \in S} |m_{\hat{\theta}}(x) - m_\theta(x)| = \mathcal{O}_p(T^{-1/2}).$$

(DT5) $\Delta_T(x)$, the local shift in the alternative H_1, is uniformly bounded with respect to x and T, and $c_T = T^{-1/2}h^{-1/4}$ which is the order of the difference between H_0 and H_1.

(DT6) $E\{\exp(a_0|Y_1 - m(X_1)|)\} < \infty$ for some $a_0 > 0$; The conditional density of X given Y $f_{X|Y} \leq A_1 < \infty$, and the joint conditional density of (X_1, X_l) given (Y_1, Y_l) is bounded for all $l > 1$.

Assumptions (DT2) and (DT3) are standard in nonparametric curve estimation and are satisfied for example for bandwidths selected by cross validation, whereas (DT4) and (DT5) are common in nonparametric goodness-of-fit tests. It can be seen from the proof of Lemma 3.1 and the proof of Theorem 3.4 that the geometric α-mixing condition, assumption (DT1), can be weakened to $\alpha(u) \leq Cu^{-s(d)}$ where $s(d) > 2$ and is a monotone function of d. It is technically convenient to assume geometric the α-mixing.

We will now introduce the empirical likelihood concept for the testing problem that we consider here. For an arbitrary $x \in [0,1]$ and any function μ we have

$$\mathrm{E}\left[K\left(\frac{x - X_i}{h}\right)\{Y_i - \mu(x)\} \,\middle|\, \mathrm{E}[Y_i|X_i = x] = \mu(x) \right] \stackrel{h \to 0}{\longrightarrow} 0 \,. \tag{3.24}$$

Using this relationship we apply the EL methodology as introduced above to develop a test about $\mu(x)$ for an arbitrary $x \in I_X$ and then extend this test to global goodness-of-fit test. . Let $p_i(x)$ be nonnegative numbers representing a density for

$$K\left(\frac{x - X_i}{h}\right)\{Y_i - \mu(x)\} \qquad i = 0, \ldots, [nT] - 1$$

The empirical likelihood for $\mu(x)$ is

$$\mathcal{L}\{\mu(x)\} \stackrel{\text{def}}{=} \max \prod_{i=0}^{[nT]-1} p_i(x) \tag{3.25}$$

subject to

$$\sum_{i=0}^{[nT]-1} p_i(x) = 1 \text{ and } \sum_{i=0}^{[nT]-1} p_i(x) K\left(\frac{x - X_i}{h}\right) \{Y_i - \mu(x)\} = 0\ . \tag{3.26}$$

The second condition reflects (3.24).

Following Owen (2001), we find the maximum by introducing Lagrange multipliers and maximizing the Lagrangian function

$$\begin{aligned} \mathcal{H}(p, \lambda_1, \lambda_2) &= \sum_{i=0}^{[nT]-1} \log p_i(x) \qquad (3.27) \\ &\quad -\lambda_1 \sum_{i=0}^{[nT]-1} p_i(x) K\left(\frac{x - X_i}{h}\right) \{Y_i - \mu(x)\} - \lambda_2 \left\{\sum_{i=0}^{[nT]-1} p_i(x) - 1\right\} \end{aligned}$$

where λ_1 and λ_2 depend on x. The first order conditions are the equations in (3.26) and

$$\frac{\partial \mathcal{H}(p, \lambda_1, \lambda_2)}{\partial p_i(x)} = \frac{1}{p_i(x)} - \lambda_1 K\left(\frac{x - X_i}{h}\right) \{Y_i - \mu(x)\} - \lambda_2 = 0$$

for all $i = 0, \ldots, [nT] - 1$. With the normalization $\lambda_2 = n$ and $\lambda = \lambda_1/\lambda_2$ we obtain as a solution to (3.25) the optimal weights

$$p_i(x) = n^{-1} \left[1 + \lambda(x) K\left(\frac{x - X_i}{h}\right) \{Y_i - \mu(x)\}\right]^{-1} \tag{3.28}$$

where $\lambda(x)$ is the root of

$$\sum_{i=0}^{n} \frac{K\left(\frac{x-X_i}{h}\right) \{Y_i - \mu(x)\}}{1 + \lambda(x) K\left(\frac{x-X_i}{h}\right) \{Y_i - \mu(x)\}} = 0. \tag{3.29}$$

The maximum empirical likelihood is achieved at $p_i(x) = [nT]^{-1}$ corresponding to the nonparametric curve estimate $\mu(x) = \hat{m}(x)$. For a parameter estimate $\hat{\theta}$ we get the maximum empirical likelihood for the smoothed parametric model $\mathcal{L}\{\tilde{m}_{\hat{\theta}}(x)\}$. The log-EL ratio is

$$\ell\{\tilde{m}_{\hat{\theta}}(x)\} \stackrel{\text{def}}{=} -2 \log \frac{\mathcal{L}\{\tilde{m}_{\hat{\theta}}(x)\}}{\mathcal{L}\{\hat{m}(x)\}} = -2 \log[\mathcal{L}\{\tilde{m}_{\hat{\theta}}(x)\}[nT]^{[nT]}].$$

To study properties of the empirical likelihood based test statistic we need to evaluate $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ at an arbitrary x first, which requires the following theorem on $\lambda(x)$ that was first proven in Chen et al. (2001).

THEOREM 3.2 *Under the assumptions (DT1)-(DT6), we have for $\mu(x) = \tilde{m}_{\hat{\theta}}(x)$:*

$$\sup_{x \in I_X} |\lambda(x)| = \mathcal{O}_p\{([nT]h)^{-1/2} \log(nT)\}.$$

To prepare the proof we introduce the notation

$$\bar{U}_j(x) = ([nT]h)^{-1} \sum_{i=0}^{[nT]-1} \left[K\left(\frac{x - X_i}{h}\right) \{Y_i - \tilde{m}_{\hat{\theta}}(x)\} \right]^j \tag{3.30}$$

for $j = 1, 2, \ldots$ and give the following three Lemmas proven in Chen et al. (2001), see Appendix A.2.

LEMMA 3.1

$$\sup_{x \in I_X} |\bar{U}_1(x)| = \mathcal{O}_p\{([nT]h)^{-1/2} \log(nT)\}$$

LEMMA 3.2

$$P\{\inf_{x \in I_X} \bar{U}_2(x) \geq d_0\} = 1 \quad \textit{for a positive } d_0 > 0$$

LEMMA 3.3

$$\max_{1 \leq j \leq n} \sup_{x \in I_X} \left| K\left(\frac{x - X_i}{h}\right) \{Y_i - \tilde{m}_{\hat{\theta}}(x)\} \right| = \mathcal{O}_p\{([nT]h)^{1/2} \log^{-1}(nT)\}$$

Using these Lemmas we will now proof Theorem 3.2.

PROOF of Theorem 3.2:

From the definition of $\mathcal{H}$ in (3.27) and from the deviation of $p_i(x)$ we have that $p_i(x)$ is positive for all x and i. Following Owen (1990) and with the notation

$$\varepsilon_i(x) \stackrel{\text{def}}{=} K\left(\frac{x - X_i}{h}\right) \{Y_i - \tilde{m}_{\hat{\theta}}(x)\}$$

we get from (3.29)

$$\begin{aligned} 0 &= \left| \sum_{i=0}^{[nT]-1} \frac{\varepsilon_i(x)}{1 + \lambda(x)\varepsilon_i(x)} \right| \\ &= \left| \sum_{i=0}^{[nT]-1} \varepsilon_i(x) - \lambda(x) \sum_{i=0}^{[nT]-1} \frac{\varepsilon_i^2(x)}{1 + \lambda(x)\varepsilon_i(x)} \right| &(3.31) \\ &\geq |\lambda(x)| \sum_{i=0}^{[nT]-1} \frac{\varepsilon_i(x)}{1 + \lambda(x)\varepsilon_i(x)} - \left| \sum_{i=0}^{[nT]-1} \varepsilon_i(x) \right| . &(3.32) \end{aligned}$$

From this inequality and the definition of $\bar{U}_j(x)$ in (3.30) follows

$$0 \geq |\lambda(x)|\bar{U}_2(x)\left\{1+|\lambda(x)|\max_{0\leq i\leq [nT]-1}|\varepsilon_i(x)|\right\}^{-1} - |\bar{U}_1(x)|\,.$$

Taking the supremum over all $x \in [0,1]$ we have from Lemma 3.1

$$\sup_{x\in[0,1]}|\lambda(x)|\bar{U}_2(x)\left\{1+|\lambda(x)|\max_{1\leq 0\leq [nT]-1}|\varepsilon_i(x)|\right\}^{-1} = \mathcal{O}_p\{([nT]h)^{-1/2}\log(nT)\}$$

and the proof is completed by applying Lemma 3.2 and Lemma 3.3. □

An application of the power series expansion of $1/(1-.)$ applied to (3.29) and Theorem 3.2 yields

$$\sum_{i=0}^{[nT]-1} K\left(\frac{x-X_i}{h}\right)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\}\left[\sum_{j=0}^{\infty}\{-\lambda(x)\}^j K^j\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}^j\right] = 0.$$

Inverting the above expansion, we have

$$\lambda(x) = \bar{U}_2^{-1}(x)\bar{U}_1(x) + \tilde{o}_p\{([nT]h)^{-1}\log^2(nT)\}. \tag{3.33}$$

From (3.28), Theorem 3.2 and the Taylor expansion of $\log(1+.)$ we get

$$\begin{aligned}
\ell\{\tilde{m}_{\hat{\theta}}(x)\} &= -2\log[\mathcal{L}\{\tilde{m}_{\hat{\theta}}(x)\}[nT]^{[nT]}] \\
&= 2\sum_{i=0}^{[nT]-1}\log[1+\lambda(x)K\left(\frac{x-X_i}{h}\right)\{Y_j-\tilde{m}_{\hat{\theta}}(x)\}] \\
&= 2[nT]h\lambda(x)\bar{U}_1 - [nT]h\lambda^2(x)\bar{U}_2 + \tilde{o}_p\{([nT]h)^{-1/2}\log^3([nT])\}\,.
\end{aligned} \tag{3.34}$$

Inserting (3.33) in (3.34) yields

$$\ell\{\tilde{m}_{\hat{\theta}}(x)\} = [nT]h\bar{U}_2^{-1}(x)\bar{U}_1^2(x) + \tilde{o}_p\{([nT]h)^{-1/2}\log^3([nT])\}. \tag{3.35}$$

For any $x \in [0,1]$, let

$$v(x;h) = h\int_0^1 K_h^2(x-y)dy \quad \text{and} \quad b(x;h) = h\int_0^1 K_h(x-y)dy$$

be the variance and the bias coefficient functions associated with the NW estimator, respectively, see Wand and Jones (1995), Härdle et al. (2000). Let

$$S_{I,h} = \{x \in [0,1] |\min(|x-1|,|x|) > h\}.$$

For $h \to 0$, $S_{I,h}$ converges to the set of interior points in $I_X = [0,1]$. If $x \in S_{I,h}$, we have $v(x;h) \stackrel{\text{def}}{=} \int K^2(x)dx$ and $b(x;h) = 1$. Define

$$V(x;h) = \frac{v(x;h)\sigma^2(x)}{f(x)b^2(x;h)}.$$

Clearly, $V(x;h)/([nT]h)$ is the asymptotic variance of $\hat{m}(x)$ when $[nT]h \to \infty$, which is one of the conditions we assumed, Wand and Jones (1995) p. 125.

From assumption (DT4) we have

$$\begin{aligned}
\bar{U}_1(x) &= [nT]^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\} \\
&= [nT]^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i)\{Y_i - m_\theta(X_i)\} + \tilde{\mathcal{O}}_p(n^{-1/2}) \\
&= \hat{f}(x)\{\hat{m}(x) - \tilde{m}_\theta(x)\} + \tilde{\mathcal{O}}_p([nT]^{-1/2})
\end{aligned}$$

and with Theorem 2.2 in Bosq (1998)

$$\bar{U}_1(x) = f(x)b(x;h)\{\hat{m}(x) - \tilde{m}_\theta(x)\} + \tilde{\mathcal{O}}_p\{[nT]^{-1/2} + ([nT]h)^{-1}\log^2([nT])\}. \tag{3.36}$$

for any $x \in S_{I,h}$. (A.18) and (3.35) mean that

$$\begin{aligned}
\ell\{\tilde{m}_{\hat{\theta}}(x)\} &= ([nT]h)\bar{U}_2^{-1}\bar{U}_1^2 + \tilde{o}_p\{([nT]h)^{-1/2}\log^3([nT])\} \\
&= ([nT]h)\frac{f^2(x)b^2(x;h)}{f(x)v(x;h)\sigma^2(x)}\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2 + \tilde{\mathcal{O}}\{([nT]h)^{-1}h\log^2([nT])\} \\
&= ([nT]h)V^{-1}(x;h)\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2 + \tilde{\mathcal{O}}\{([nT]h)^{-1}h\log^2([nT])\} \\
&= Var(\hat{m}(x))^{-1}\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2 + \tilde{\mathcal{O}}\{([nT]h)^{-1}h\log^2([nT])\}\,. \qquad (3.37)
\end{aligned}$$

Therefore, $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ is asymptotically equivalent to a studentized L_2 distance between $\tilde{m}_{\hat{\theta}}(x)$ and $\hat{m}(x)$. It is this property that leads us to use $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ as the basic building block in the construction of a global test statistic for distinction between $\tilde{m}_{\hat{\theta}}$ and $\hat{m}$ in the next section. The use of the empirical likelihood as a distance measure and its comparison with other distance measures have been discussed in Owen (1991) and Baggerly (1998).

To extend the empirical likelihood ratio statistic to a global measure of goodness-of-fit, we choose k_T-equally spaced lattice points $t_1, t_2, \cdots, t_{k_T}$ in $I_X = [0,1]$ where $t_1 = 0$, $t_{k_T} = 1$ and $t_i \le t_j$ for $1 \le i < j \le k_T$. We let $k_T \to \infty$ and $k_T/T \to 0$ as $T \to \infty$. This essentially divides $[0,1]$ into k_T small nonoverlapping intervals of size $(k_T)^{-1}$. A simple choice is to let $k_T = [1/(2h)]$. Then we have with assumption (DT2), $k_T = C_1 T^{1/5} \to \infty$ and $k_T/T = C_2 T^{-4/5} \to 0$ as $T \to \infty$. This choice as justified later ensures asymptotic independence among $\ell\{\tilde{m}_{\hat{\theta}}(t_j)\}$ at different points t_j. Bins of different size can be adopted to suit situations where there are areas of low design density. This corresponds to the use of different bandwidth values in adaptive kernel smoothing. The main results of this chapter are not affected by un-equal bins. For the purpose of easy presentation, we consider bins of equal size.

As $\ell\{\tilde{m}_{\hat{\theta}}(t_j)\}$ measures the goodness-of-fit at a fixed t_j, an empirical likelihood based statistic that measures the global goodness-of-fit is defined as

$$\ell_n(\tilde{m}_{\hat{\theta}}) \stackrel{\text{def}}{=} \sum_{j=1}^{k_T} \ell\{\tilde{m}_{\hat{\theta}}(t_j)\}.$$

The following theorem was first proven by Chen et al. (2001), see Appendix A.2,

THEOREM 3.3 *Under the assumptions (DT1) - (DT6),*

$$k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) = ([nT]h)\int \frac{\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2}{V(x)}dx + \mathcal{O}_p\{k_T^{-1}\log^2([nT]) + h\log^2([nT])\} \tag{3.38}$$

where $V(x) \stackrel{\text{def}}{=} \lim_{h\to 0} V(x,h)$.

Härdle and Mammen (1993) proposed the L_2 distance

$$T_n = nh^{1/2}\int\{\hat{m}(x) - \tilde{m}_{\hat{\theta}}(x)\}^2\pi(x)dx$$

as a measure of goodness-of-fit where $\pi(x)$ is a given weight function. Theorem 3.3 indicates that the leading term of $k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}})$ is $h^{1/2}T_n$ with $\pi(x) = V^{-1}(x)$. The differences between the two test statistics are (a) the empirical likelihood test statistic automatically studentizes via its internal algorithm conducted at the background, so that there is no need to explicitly estimate $V(x)$; (b) the empirical likelihood statistic also captures other features such as skewness and kurtosis exhibited in the data without using the bootstrap resampling which involves more technical details when data are dependent. If we choose $k_T = [1/(2h)]$ as prescribed, then the remainder term in (3.38) becomes $\mathcal{O}_p\{h\log^2([nT])\}$.

We will now discuss the asymptotic distribution of the test statistic $\ell_n(\tilde{m}_{\hat{\theta}})$. The proof of Theorem 3.4 was given by Chen et al. (2001), see Appendix A.2.

THEOREM 3.4 *Suppose assumptions (DT1) - (DT6), then*

$$k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) \xrightarrow{\mathcal{L}} \int_0^1 \mathcal{N}^2(s)ds \text{ for } T \to \infty$$

where $\mathcal{N}$ is a Gaussian process on $[0,1]$, i.e. $\mathcal{N}(s)$ is normal for every s, with mean

$$E\{\mathcal{N}(s)\} = h^{1/4}\Delta_T(s)/\sqrt{V(s)}$$

and covariance

$$\Omega(s,t) = Cov\{\mathcal{N}(s),\mathcal{N}(t)\} = \sqrt{\frac{f(s)\sigma^2(s)}{f(t)\sigma^2(t)}}\frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}}$$

where

$$W_0^{(2)}(s,t) = \int_0^1 h^{-1}K\{(s-y)/h\}K\{(t-y)/h\}dy. \tag{3.39}$$

As K is a compact kernel on $[-1,1]$, when both s and t are in S_I (the interior part of $[0,1]$), we get from (3.39) with $u=(s-y)/h$

$$\begin{aligned} W_0^{(2)}(s,t) &= \int_{\frac{s-1}{h}}^{\frac{s}{h}} K(u)K\{u-(s-t)/h\}du \\ &= \int_{-\infty}^{\infty} K(u)K\{u-(s-t)/h\}du \\ &= K^{(2)}\left(\frac{s-t}{h}\right) \end{aligned} \tag{3.40}$$

where $K^{(2)}$ is the convolution of K, i.e.

$$K^{(2)}(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} K(x-u)K(u)du\ .$$

The compactness of K also means that $W_0^{(2)}(s,t)=0$ if $|s-t|>2h$ which implies $\Omega(s,t)=0$ if $|s-t|>2h$. Hence $\mathcal{N}(s)$ and $\mathcal{N}(t)$ are independent if $|s-t|>2h$. As

$$f(s)\sigma^2(s) = f(s)\sigma^2(t) + \mathcal{O}(h)$$

when $|s-t|\le 2h$, we get

$$\Omega(s,t) = \frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}} + \mathcal{O}(h), \tag{3.41}$$

So, the leading order of the covariance function is free of σ^2 and f, i.e. $\Omega(s,t)$ is completely known.

Let

$$\mathcal{N}_0(s) = \mathcal{N}(s) - \frac{h^{1/4}\Delta_T(s)}{\sqrt{V(s)}}. \tag{3.42}$$

Then $\mathcal{N}_0(s)$ is a normal process with zero mean and covariance Ω. The boundedness of K implies $W_0^{(2)}$ being bounded, and hence $\int_0^1 \Omega(t,t)dt < \infty$. We will now study the expectation and variance of $\int_0^1 \mathcal{N}^2(s)ds$. Let $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 \stackrel{\text{def}}{=} \int_0^1 \mathcal{N}^2(s)ds$ where

$$\begin{aligned}
\mathcal{T}_1 &= \int_0^1 \mathcal{N}_0^2(s)ds, \\
\mathcal{T}_2 &= 2h^{1/4}\int_0^1 V^{-1/2}(s)\Delta_T(s)\mathcal{N}_0(s)ds \quad \text{and} \\
\mathcal{T}_3 &= h^{1/2}\int_0^1 V^{-1}(s)\Delta_T^2(s)ds.
\end{aligned}$$

Before studying the properties of $\mathcal{T}_1$ and $\mathcal{T}_2$ we proof the following lemma.

LEMMA 3.4 *Let X, Y be standard normal random variables with covariance $Cov(X,Y) = \rho$, i.e.*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{3.43}$$

Then we have:

$$Cov(X^2, Y^2) = 2\rho^2$$

PROOF :
Define $Z \sim \mathrm{N}(0,1)$ independent of X and $X' \stackrel{\text{def}}{=} \rho X + \sqrt{1-\rho^2}Z$. Then we get:

$$\begin{pmatrix} X \\ X' \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

$$\mathrm{Cov}(X^2, Y^2) = \mathrm{Cov}(X^2, X'^2) = 2\rho^2$$

□

From some basic results on stochastic integrals, Lemma 3.4 and (3.41) it follows,

$$\begin{aligned} \mathrm{E}(\mathcal{T}_1) &= \int_0^1 \Omega(s,s)ds = 1 \quad \text{and} \\ Var(\mathcal{T}_1) &= \mathrm{E}[\mathcal{T}_1^2] - 1 \qquad (3.44) \\ &= \int_0^1\int_0^1 \mathrm{E}\left[N_0^2(s)N_0^2(t)\right] dsdt - 1 \qquad (3.45) \\ &= 2\int_0^1\int_0^1 \Omega^2(s,t)dsdt \\ &= 2\int_0^1\int_0^1 \{W_0^{(2)}(s,t)\}^2\{W_0^{(2)}(s,s)W_0^{(2)}(t,t)\}^{-1}dsdt\,\{1+\mathcal{O}(h^2)\} \end{aligned}$$

From (3.40) and the fact that the size of the region $[0,1]\setminus S_{I,h}$ is $\mathcal{O}(h)$, we have

$$\begin{aligned} &\int_0^1\int_0^1 \{W_0^{(2)}(s,t)\}^2\{W_0^{(2)}(s,s)W_0^{(2)}(t,t)\}^{-1}dsdt \\ =\;& \{K^{(2)}(0)\}^{-2}\int_0^1\int_0^1 [K^{(2)}\{(s-t)/h\}]^2 dsdt\,\{1+o(1)\} \\ =\;& hK^{(4)}(0)\{K^{(2)}(0)\}^{-2} + o(h). \end{aligned}$$

Therefore,

$$Var(\mathcal{T}_1) = 2hK^{(4)}(0)\{K^{(2)}(0)\}^{-2} + o(h^2).$$

It is obvious that $\mathrm{E}(\mathcal{T}_2) = 0$ and

$$Var(\mathcal{T}_2) = 4h^{1/2}\int\int V^{-1/2}(s)\Delta_T(s)\Omega(s,t)V^{-1/2}(t)\Delta_T(t)dsdt.$$

As Δ_T and V^{-1} are bounded in $[0,1]$, there exists a constant C_1 such that

$$Var(\mathcal{T}_2) \le C_1 h^{1/2}\int\int \Omega(s,t)dsdt.$$

Furthermore we know from the discussion above,

$$\begin{aligned} \int\int \Omega(s,t)dsdt &= \int\int \frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}}dsdt + \mathcal{O}(h) \\ &= \int\int_{t-2h}^{t+2h} \frac{W_0^{(2)}(s,t)}{K^{(2)}(0)}dsdt + \mathcal{O}(h) \\ &\le 4\frac{1}{K^{(2)}(0)}C_1'h + C_1''h \end{aligned}$$

with other constants C'_1 and C''_1, and thus, there exists a constant C_2, such that

$$Var(\mathcal{T}_2) \leq C_2 h^{\frac{3}{2}}.$$

As $\mathcal{T}_3$ is non-random, we have

$$\mathrm{E}(\mathcal{T}) = 1 + h^{1/2}\int_0^1 V^{-1}(s)\Delta_T^2(s)ds \quad \text{and} \tag{3.46}$$

$$Var\{\mathcal{T}\} = 2hK^{(4)}(0)\{K^{(2)}(0)\}^{-2} + o(h) \tag{3.47}$$

(3.46) and (3.47) together with Theorem 3.4 give the asymptotic expectation and variance of the test statistic $k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}})$.

We now turn our interest to the derivation of the asymptotic distribution of $k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}})$. We do this by discretizing $\int_0^1 \mathcal{N}^2(s)ds$ as $(k_T)^{-1}\sum_{j=1}^{k_T}\mathcal{N}^2(t_j)$ where $\{t_j\}_{j=1}^{k_T}$ are the mid-points of the original bins in formulating $\ell_n(\tilde{m}_{\hat{\theta}})$. If we choose $k_T = [(2h)^{-1}]$ such that $|t_{j+1} - t_j| \geq 2h$ for all j, then $\{\mathcal{N}(t_j)\}$ are independent and each $\mathcal{N}(t_j) \sim \mathrm{N}(h^{1/4}\Delta_T(t_j)/\sqrt{V(t_j)}, 1)$. This means that under the alternative H_1

$$\sum_{j=1}^{k_T}\mathcal{N}^2(t_j) \sim \chi^2_{k_T}(\gamma_{k_T}),$$

a non-central χ^2 random variable with k_T degrees of freedom and the non-central component $\gamma_{k_T} = h^{1/4}\{\sum_{j=1}^{k_T}\Delta_T^2(t_j)/V(t_j)\}^{1/2}$. Under H_0,

$$\sum_{j=1}^{k_T}\mathcal{N}^2(t_j) \sim \chi^2_{k_T}$$

is χ^2-distributed with k_T degrees of freedom. This leads to a χ^2 test with significance level α which rejects H_0 if $\ell_n(\tilde{m}_{\hat{\theta}}) > \chi^2_{k_T,\alpha}$ where $\chi^2_{k_T,\alpha}$ is the $(1-\alpha)$-quantile of $\chi^2_{k_T}$. The asymptotic power of the χ^2 test is $\mathrm{P}\{\chi^2_{k_T}(\gamma_{k_T}) > \chi^2_{k_T,\alpha}\}$, which is sensitive to alternative hypotheses differing from H_0 in all directions.

We may also establish the asymptotic normality of $(k_T)^{-1}\sum_{i=0}^{k_T}\mathcal{N}^2(t_j)$ by applying the central limit theorem for a triangular array, which together with (3.46) and (3.47) means that

$$k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) \xrightarrow{\mathcal{L}} N\left(1 + h^{1/2}\int \Delta_T^2(s)V^{-1}(s)ds, 2hK^{(4)}(0)\{K^{(2)}(0)\}^{-2}\right).$$

A test for H_0 with an asymptotic significance level α is to reject H_0 if

$$k_T^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) > 1 + z_\alpha\{K^{(2)}(0)\}^{-1}\sqrt{2hK^{(4)}(0)} \tag{3.48}$$

where $\mathrm{P}(Z > z_\alpha) = \alpha$ and $Z \sim \mathrm{N}(0,1)$. The asymptotic power of this test is

$$1 - \Phi\left\{ z_\alpha - \frac{K^{(2)}(0) \int \Delta_T^2(s) V^{-1}(s) ds}{\sqrt{2K^{(4)}(0)}} \right\} \tag{3.49}$$

where Φ denotes the Gaussian distribution function.

We see from the above that the binning based on the bandwidth value h provides a key role in the derivation of the asymptotic distributions. However, the binning discretizes the null hypothesis and unavoidably leads to some loss of power as shown in the simulation reported in the next section. From the point of view of retaining power, we would like to have the size of the bins smaller than that prescribed by the smoothing bandwidth in order to increase the resolution of the discretized null hypothesis to the original H_0. However, this will create dependence between the empirical likelihood evaluated at neighbouring bins and make the above asymptotic distributions invalid. One possibility is to evaluate the distribution of $\int_0^1 \mathcal{N}_0^2(s) ds$ by using the approach of Wood and Chan (1994) simulating the normal process $\mathcal{N}^2(s)$ under H_0. However, this is not our focus here and hence is not considered in this chapter.

3.3.5 Fixed Sample Properties

We investigate the finite sample properties of the EL testing procedure in two simulation studies. In our first simulation we consider the time series model

$$Y_i = 2Y_{i-1}/(1 + Y_{i-1}^2) + c_T \sin(Y_{i-1}) + \sigma(Y_{i-1})\eta_i \tag{3.50}$$

for $i = 0, \ldots, T$ where $\{\eta_i\}$ are independent and identically distributed uniform random variables in $[-1, 1]$, η_i is independent of $X_i = Y_{i-1}$ for each i, and $\sigma(x) = \exp(-x^2/4)$. Note that the mean and the variance functions are both bounded which ensures the series is asymptotically stationary. To get rid of the impact of the initial conditions, we pre-run the series 100 times with $Y_{-100} = 0$. Figure 3.1 shows typical plots of (y_{i-1}, y_i) for $c_T = 0$ and $c_T = 0.06$. It appears from the figure, that there is no obvious difference between the two drift functions. However, the EL test is able two distinguish between the two cases ($c_T = 0$ and $c_T = 0.06$) as can be seen in Figure 3.2.

For the simulation study the sample sizes considered for each trajectory are $T =$ 500 and 1000 and c_T, the degree of difference between H_0 and H_1, takes value of 0, 0.03 and 0.06. As the simulation shows that the two empirical likelihood tests have very similar power performance, we will report the results for the test based on the χ^2 distribution only. To gauge the effect of the smoothing bandwidth h on the power, ten levels of h are used for each simulated sample to formulate the test statistic.

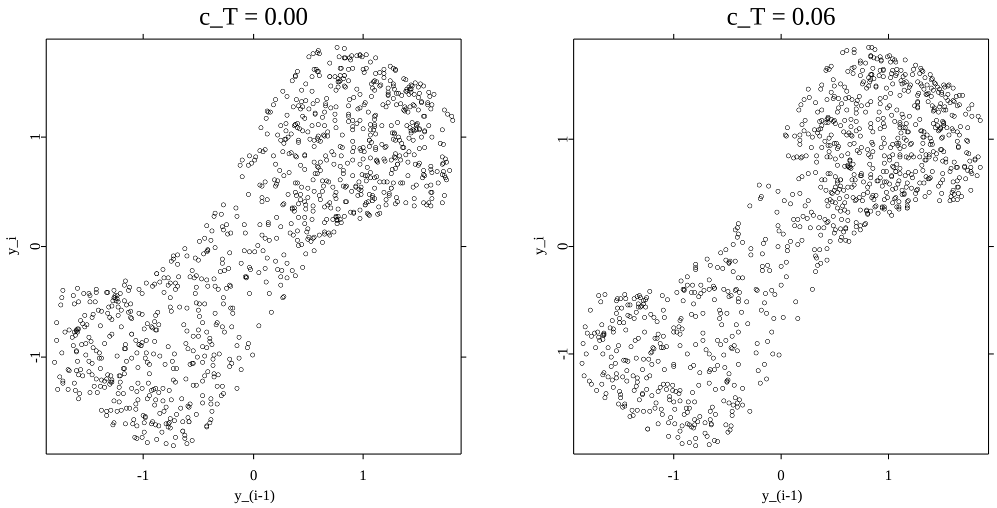


Figure 3.1: Scatterplots of the drift function of Y according to (3.50) with $c_T = 0.00$ and $c_T = 0.06$.

Figure 3.2 presents the power of the empirical likelihood test based on 5000 simulation with a nominal 5% level of significance. We notice that when $c_T = 0$ the simulated significance level of the test is very close to the nominal level for large range of h values which is especially the case for the larger sample size $T = 1000$. When c_T increases, for each fixed h the power increases as the distance between the null and the alternative hypotheses becomes larger. For each fixed c_T, there is a general trend of decreasing power when h increases. This is due to the discretization of H_0 by binning as discussed at the end of the previous section. We also notice that the power curves for $c_T = 0.06$ are a little erratic although they maintain the same trend as in the case of $c_T = 0.03$. This may be due to the fact that when the difference between H_0 and H_1 is large, the difference between the nonparametric and the parametric fits becomes larger and the test procedure becomes more sensitive to the bandwidths.

In our second simulation study we consider an Ornstein-Uhlenbeck process X fluctuating about 0 that satisfies the stochastic differential equation

$$dX(t) = m(a, X(t))dt + \sigma dW(t)$$

where W is a standard Brownian Motion and $m(a, x) = ax$ under the null hypothesis. The speed of adjustment parameter a has to be negative to ensure stationarity. To apply the empirical likelihood test we construct the time series X and Y as in Section

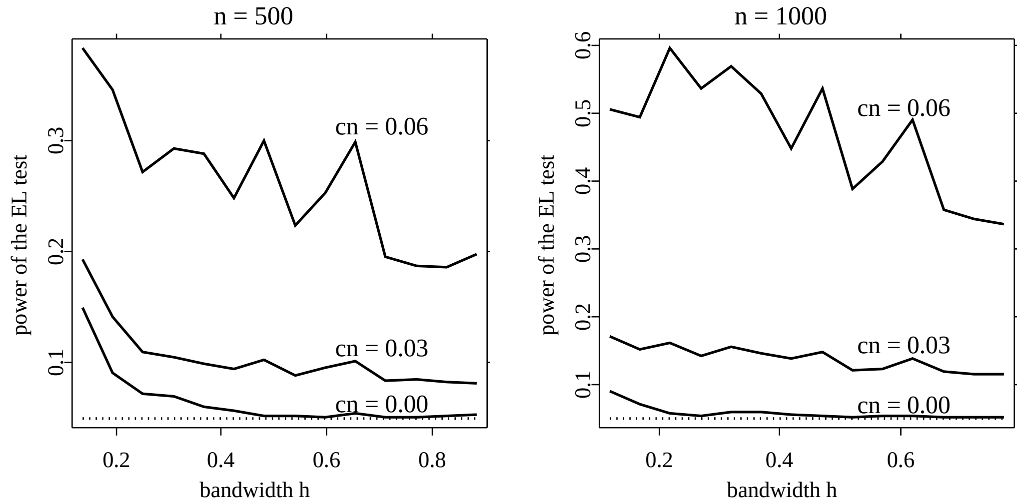


Figure 3.2: Power of the empirical likelihood test. The dotted lines indicate the 5% level

3.3.1, i.e.

$$\begin{aligned} X_i^\Delta &= X^\Delta(t_i) \,, \qquad X^\Delta = (X_0^\Delta, \ldots, X_{[nT]-1}^\Delta) \\ \varepsilon_i &= W(t_{i+1}) - W(t_i) \,, \qquad \varepsilon = (\varepsilon_0, \ldots, \varepsilon_{[nT]-1}) \\ Y_i^\Delta &= X_{i+1}^\Delta - X_i^\Delta = aX_i^\Delta \Delta \, + \sigma\varepsilon_i \,, \qquad Y^\Delta = (Y_0^\Delta, \ldots, Y_{[nT]-1}^\Delta) \end{aligned} \tag{3.51}$$

where we use again the notation $X(t)$ to denote the continuous time process and the subscript X_i to denote discrete time observations.

For $a = -1$ and $\sigma = 0.5$ we investigate the power of the test for $nT = 3000$ observations in a simulation study. The hypotheses for the test are chosen as in 3.3.4.

$$H_0(m) \;:\; \exists a \in [-\infty, 0) \;:\; \text{for every } t \in [0,T] : \; m\{X(t)\} = aX(t) \qquad \text{P-a.s.}$$

against the alternative

$$\begin{aligned} H_1(m) \;:\; \forall\, a \in [-\infty, 0) \;:\; & \text{ for every } t \in [0,T] : \\ & m\{X(t)\} = aX(t) + c_T\Delta_T(X(t)) \qquad \text{P-a.s.} \end{aligned}$$

with $\Delta_T(x) = \sin(\pi x)$. We simulate trajectories of five processes: $c_T = 0$ ($H_0(m)$), $c_T = 0.5$, 1, 1.5 and $c_T = 2$. For each c_T we simulate 1000 paths according to the Milstein scheme

$$X(t+\delta) = X(t) + m(\theta, X(t))\delta + \sigma(\theta, X(t))\sqrt{\delta}\varepsilon(t) + \frac{1}{2}\sigma^2(\theta, X(t))\delta\{\varepsilon(t)^2 - 1\} \tag{3.52}$$

for $\delta > 0$ and a sequence of independent standard normally distributed random variables $\varepsilon(t)$, see Kloeden and Platen (1999). We then apply to every simulated path the EL test about the hypothesis $H_0(m)$.

To understand the influence of the two kinds of asymptotics we make two simulation studies with the same sample size $nT = 3000$, one with $n = 50$, i.e. for $nT = 3000$ observations we have $T = 60$, and one with $n = 250$, i.e. $T = 12$. This means, that in the first simulation ($n = 50$) the approximation of X by the corresponding discrete time process X^Δ with $\Delta = 1/50$ is not as good as the approximation of X in the second simulation with $\Delta = 1/250$. On the other hand the time interval $[0, T]$ in which observations are available is much larger in the first simulation than in the second. This means that the mixing of the time series $X^{1/50}$ is stronger then the mixing of $X^{1/250}$. This effect also appears in the autocorrelation of X_i^Δ. It is well known, that the autocorrelation of an Ornstein-Uhlenbeck process is $\mathrm{Corr}(X(t), X(t+\tau)) = \exp\{a\tau\}$. Thus, the discrete process X_i^Δ has under the null hypothesis an autocorrelation of $\mathrm{Corr}(X_i^\Delta, X_{i+k}^\Delta) = \exp\{(a/n)k\}$, i.e. the autocorrelation for $n = 250$ is significantly higher than the autocorrelation for $n = 50$ when k is fixed. The reason is, that we observe the original process X at different times. We can think of $X_i^{1/50}$ as a rescaled version of $X_i^{1/250}$, where rescaling is done in time, for instance $X_1^{1/50} = X_5^{1/250}$. The same autocorrelation structure with a constant Δ can be constructed by choosing $a = -0.2$ for the second process. Thus a change in Δ corresponds to a change of a.

In the empirical analysis, Δ is not given, and thus the question of choosing Δ and verifying the assumptions arises. In particular assumption (DT1), i.e. the assumption, that X is stationary and α-mixing is critical in a fixed sample environment. Even if (DT1) is satisfied, the time interval $[0, T]$ might be too small to rely on the ergodic theorem. The reason could be, that Δ is to small or that the parameters of X generate strong dependencies between observations that are far away from each other. In the empirical analysis of the spot rate (7-day Eurodollar rate) we come again to this point.

The empirical rejection levels of the null hypothesis are shown in Figure 3.3 for nominal levels $\alpha = 0.01$ and $\alpha = 0.05$. It appears from Figure 3.3 that the nominal level of the EL test is better hold, when the approximation of the true continuous time process by a time series is better, i.e. there are more observations n of X in one unit of time.

The power of the test for different values of c_T is also investigated. The result is shown in Figure 3.4 and 3.5. The figures show, that the power of the test increases with T. The reason is, that the intervals, in which the dependencies are strong are shorter for smaller values of n and thus there are more observations that are "almost independent", i.e. have an autocorrelation close to 0. The figures also demonstrate, that the test decision of the EL test strongly depends on the order of difference c_T

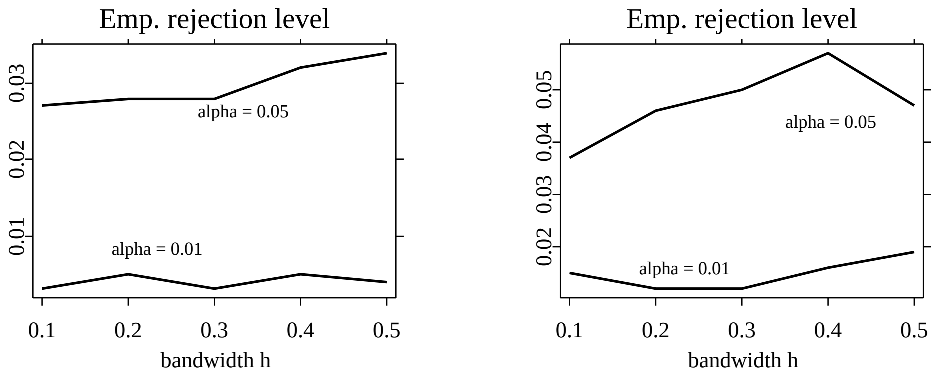


Figure 3.3: Empirical rejection level of the empirical likelihood test for the drift of an Ornstein Uhlenbeck process. ($\Delta = 1/50$ left figure, $\Delta = 1/250$ right figure)

between the H_0 and H_1. We also find, that the test depends on the chosen bandwidth that is used to nonparametrically estimate the drift. But this dependency seems to have a minor impact on the test decision.

3.4 Nonparametric Estimation of the Continuous Time Model

We will now study nonparametric estimators for the marginal density, the drift function and the diffusion coefficient of the process X given as the solution of (3.1) without approximating the true continuous time model by a discrete one. Therefore, we change the meaning of X_i. In this section we denote by X_i the value of X at i/n, i.e. $X_i \stackrel{\text{def}}{=} X(i/n)$ for $i = 0, \ldots, [nT]$.

For the nonparametric estimator based on the discrete sample of X we apply a Lipschitz continuous kernel with support on $[-1, 1]$, that satisfies the moment conditions in (3.12), i.e. K has finite first and second moments. As in Section 3.3 we use the notation $K_h(u) = 1/hK(u/h)$.

3.4.1 Estimation of the marginal density

To identify the marginal density we assume that X as given by (3.1) is stationary and α-mixing, see Definition 3.1 and 3.2. The estimator of the marginal density $f(x)$ at a state x of a diffusion process X is related to the local time of X as shown in (3.3). Using this relationship, we will estimate the marginal density by estimating the local time $L_t(x)$ and than rescale it by time. This means, that we need two kinds

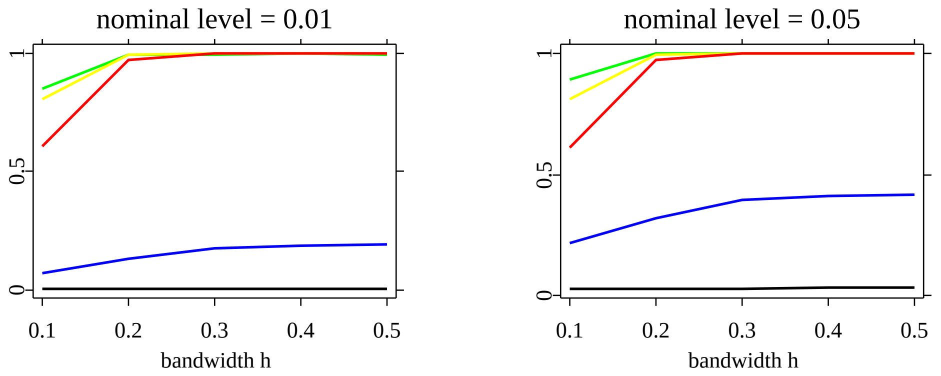


Figure 3.4: Empirical rejection probabilities of EL test with $n = 50$ for nominal level $\alpha = 0.01$ and $\alpha = 0.05$. The degree of difference between H_0 and H_1 are $c_T = 2$ (red), $c_T = 1.5$ (yellow), $c_T = 1$ (green), $c_T = 0.5$ (blue), $c_T = 0$ (black). The parameters of the simulated Ornstein Uhlenbeck process are $a = -1$ and $\sigma = 0.5$.

of asymptotics in this section. To estimate the local time, we have to assume, that n tends to infinity, i.e. that the time between two successive observations X_i and X_{i+1} is vanishing. On the other hand, relation (3.3) is based on the ergodic theorem and thus we assume, that T tends to infinity.

From the definition of L, Definition 3.3, it follows

$$L_t(x) = \lim_{h \to 0} \int_0^t K_h(X(u) - x)du \tag{3.53}$$

for every $t \in [0, T]$. A nonparametric estimator $L_T^{(n)}(x)$ of $L_T(x)$ based on the observations $X_0, \ldots, X_{[Tn]}$ is then given by an approximation of the integral in (3.53), i.e.

$$L_T^{(n)}(x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=0}^{[Tn]} K_{h_n}(X_i - x) \tag{3.54}$$

where h_n is a bandwidth satisfying the following assumption

(CT1) $h = n^\alpha$ for some α with $-1 < \alpha < -1/3$

Florens-Zmirou (1993) proves that $L_T^{(n)}(x)$ converges in the L^2 sense to $L_T(x)$, i.e. if $nh_n^4 \longrightarrow 0$ then

$$\mathrm{E}\left[\left\{L_T^{(n)}(x) - L_T(x)\right\}^2\right] \longrightarrow 0. \tag{3.55}$$

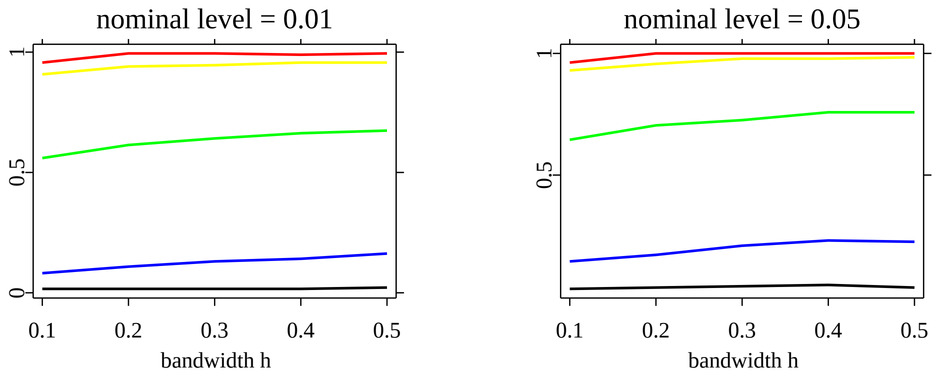


Figure 3.5: Empirical rejection probabilities of EL test with $n = 250$ for nominal level $\alpha = 0.01$ and $\alpha = 0.05$. The degree of difference between H_0 and H_1 are $c_T = 2$ (red), $c_T = 1.5$ (yellow), $c_T = 1$ (green), $c_T = 0.5$ (blue), $c_T = 0$ (black). The parameters of the simulated Ornstein Uhlenbeck process are $a = -1$ and $\sigma = 0.5$.

Applying (3.3) we define a nonparametric density estimator $f_T^{(n)}$ based on $[nT]$ discrete observations of X up to time T by $f_T^{(n)} \stackrel{\text{def}}{=} (1/T)L_T^{(n)}$. With (3.55) we have an analog result for $f_T^{(n)}$, i.e.

$$\mathrm{E}\left[\left\{f_T^{(n)}(x) - f(x)\right\}^2\right] \longrightarrow 0. \tag{3.56}$$

3.4.2 Estimation of the diffusion coefficient

Local Time Estimator

In contrast to the previous paragraph about the estimation of the marginal density, we do not need to assume that X is stationary or α-mixing for the nonparametric estimation of σ^2. The reason is, that the estimation of σ^2 is based on the approximation of the local time of X. For the same reason, the results about the estimation of σ^2 are asymptotic results for $n \to \infty$, but not for $T \to \infty$.

In Definition 3.3 we have already defined the local time of a diffusion X. A different definition of the semimartingale local time is given by Karatzas and Shreve (1991). They define the semimartingale local time $\Lambda_t(x)$ as a random field such that for every Borel-measurable function $k : \mathbb{R} \to [0, \infty)$ the following equation holds among others:

$$\int_0^t k\{X(u,\omega)\}\sigma^2(X(u,\omega))du = 2\int_{-\infty}^{\infty} k(a)\Lambda_t(a,\omega)da \tag{3.57}$$

for $\mathrm{P}-a.e.\ \omega \in \Omega$. The relationship between L and Λ is

$$\sigma^2(x) = \frac{2\Lambda_t(x)}{L_t(x)} \tag{3.58}$$

which follows immediately from equation (3.57).

Using equation (3.57) we get

$$2\Lambda_t(x) = \lim_{h\to 0} \int_0^t K_h \{X(u) - x\}\, d\langle X\rangle_u$$

where $\langle X\rangle_t$ denotes the quadratic variation of X up to time t. We define our estimator of $\Lambda_T(x)$ by approximating the integral,

$$2\Lambda_T^{(n)}(x) \overset{\text{def}}{=} \sum_{i=0}^{[Tn]-1} K_{h_n}(X_i - x)\{X_{i+1} - X_i\}^2 . \tag{3.59}$$

Combining (3.54) and (3.59) yields a nonparametric estimator $S_T^{(n)}(x)$ for $\sigma^2(x)$,

$$S_T^{(n)}(x) \quad = \quad \frac{n\sum_{i=0}^{[Tn]-1} K\left(\frac{X_i - x}{h_n}\right)\{X_{i+1} - X_i\}^2}{\sum_{i=0}^{[Tn]} K\left(\frac{X_i - x}{h_n}\right)} \tag{3.60}$$

which was first given by Florens-Zmirou (1993).

To achieve a test statistic for a particular set of points $x_1, \ldots, x_k$ we will now investigate the joint asymptotic distribution of $S_t^{(n)}(x_1), \ldots, S_t^{(n)}(x_k)$.

PROPOSITION 3.1 *Given k points $x_1, \ldots, x_k \in I_X$ and under the assumption $nh_n^3 \longrightarrow 0$ the random vector*

$$Z_T^{(n)} = (Z_T^{(n)}(x_1), \ldots, Z_T^{(n)}(x_k))^\top \qquad T \in (0, \infty)$$

with

$$Z_T^{(n)}(x_l) = \sqrt{nhL_T^{(n)}(x_l)}\left(\frac{S_T^{(n)}(x_l)}{\sigma^2(x_l)} - 1\right) \qquad l = 1, \ldots, k \tag{3.61}$$

converges in distribution to a random vector Z where Z has a joint standard normal distribution.

PROOF :

We proof the result only for $T = 1$. From Theorem 1 in Florens-Zmirou (1993) and

Theorem 1 in Jiang and Knight (1997) we know that $Z_1^{(n)}(x_l)$ converges in distribution to $Z_l \sim \mathrm{N}(0,1)$ for every $l = 1, \ldots, k$. We introduce the notation

$$\begin{aligned} a &= \min\{|x_{l_1} - x_{l_2}|;\ l_1, l_2 = 1, \ldots, k;\ l_1 \neq l_2\} \\ n_0 &= \min\{n | a > 2h_n\}\ . \end{aligned} \tag{3.62}$$

Following Florens-Zmirou (1993) we define

$$m_{i+1}(x_l) \stackrel{\text{def}}{=} \sqrt{\frac{n}{h_n}} K\left(\frac{X_i - x_l}{h_n}\right) \left[\{X_{i+1} - X_i\}^2 - \frac{\sigma^2(x_l)}{n}\right]$$

$$M_t^{(n)}(x_l) = \sum_{i=0}^{[nt]-1} m_{i+1}(x_l) \text{ for } t \in [0,1]$$

and get

$$Z_1^{(n)}(x_l) = \frac{M_1^{(n)}(x_l)}{\sigma^2(x_l)\sqrt{L_1^{(n)}(x_l)}}.$$

From (3.55) we have that

$$\tilde{Z}_n(x_l) \stackrel{\text{def}}{=} \frac{M_1^{(n)}(x_l)}{\sigma^2(x_l)\sqrt{L_1(x_l)}}.$$

also converges in distribution to a standard normal random variable.

For k arbitrary numbers u_l we define

$$C_l \stackrel{\text{def}}{=} \frac{u_l}{\sigma^2(x_l)\sqrt{L_t(x_l)}}$$

and

$$\tilde{Z}_n^{(u)} = \sum_{l=1}^{k} u_l \tilde{Z}_n(x_l) = \sum_{i=0}^{[nt]-1} \sum_{l=1}^{k} C_l m_{i+1}(x_l)$$

Then we have from Lemma 2 in Florens-Zmirou (1993) for every $n > n_0$ and with the notation $\mathrm{E}^{i,n}[.] = \mathrm{E}[.|\mathcal{F}_{i/n}]$ that

$$\begin{aligned} \sum_{i=0}^{[nt]-1} \mathrm{E}^{i,n}\left[\left(\sum_{l=1}^{k} C_l m_{i+1}(x_l)\right)^2\right] &= \sum_{l=1}^{k} C_l^2 \sum_{i=0}^{[nt]-1} \mathrm{E}^{i,n}[m_{i+1}^2(x_l)] \\ &= \sum_{l=1}^{k} C_l^2 \sigma^4(x_l) L_t(x_l) = \sum_{l=1}^{k} u_l^2 \end{aligned}$$

and
$$\sum_{i=0}^{[nt]-1} \mathrm{E}^{i,n} \left| \sum_{l=1}^{k} C_l m_{i+1}(x_l) \right|^3 \leq \sum_{l=1}^{k} |C_l|^3 \sum_{i=0}^{[nt]-1} \mathrm{E}^{i,n} |m_{i+1}(x_l)|^3 .$$
Applying proposition 5 in Florens-Zmirou (1993) we have that
$$\tilde{Z}_n^{(u)} \longrightarrow \tilde{Z}^{(u)}$$
in distribution, where $\tilde{Z}^{(u)}$ is normal distributed with zero expectation and variance $\sum_{l=1}^{k} u_l^2$. Since $L_t^{(n)}(x_l)$ converges in the L_2 sense to $L_t(x_l)$ the same follows for
$$\sum_{l=1}^{k} u_l Z_1^{(n)}(x_l) .$$

The convergence in distribution of $Z_t^{(n)}$ to a joint normal distribution follows with the Cramér-Wold device, Billingsley (1968). □

Other nonparametric estimators of the diffusion coefficient are proposed in the literature. We introduce two methods here, but do not investigate their properties is detail. For the test about a parametric form of σ^2 we will apply the estimator $S_T^{(n)}$.

Tanaka-Meyer-Formula

The first estimator applies the Tanaka-Meyer formula
$$\Lambda_t(x) = (X(t) - x)^- - (X(0) - x)^- + \int_0^t \boldsymbol{I}_{(-\infty,x]}\{X(s)\} dX(s) \; . \tag{3.63}$$
Kutoyants (1998) proposes a discrete approximation of the integral in (3.63) to estimate $\Lambda_t(x)$. He used this estimator to get a nonparametric estimator of the density of X. We remark that one could use also the Tanaka-Meyer formula for $(X(t) - x)^+$ or the formula for $|X(t) - x|$ without changing the results in principle.

Kolmogorov Forward Equation

We define the transition density $f(x, t|y, s)$ of X as the Radon-Nikodyn derivative of the transition law, i.e. $f(x, t|y, s)$ is the unique function that solves
$$\mathrm{P}[X(t) \in B | X(s) = y] = \int_B f(x, t|y, s) dx \tag{3.64}$$
for all Borel sets $B \in \mathbb{R}$.

To get an estimator for σ^2, Aït-Sahalia (1996) applies the Kolmogorov forward equation, Karlin and Taylor (1981), for ergodic diffusions

$$\begin{aligned}\frac{\partial f\{X(u),u|X(t),t\}}{\partial s} &= \frac{1}{2}\frac{\partial^2}{\partial X^2(u)}\left[\sigma^2\{X(u)\}f\{X(u),u|X(t),t\}\right] \\ &\quad -\frac{\partial}{\partial X(u)}\left[m\{X(u)\}f\{X(u),u|X(t),t\}\right]\end{aligned} \tag{3.65}$$

with $u = t + s$. By the Markov property of X the marginal density f is given by

$$f\{X(t+s)\} = \int_{-\infty}^{\infty} f\{X(t+s), t+s|X(t),t\} f\{X(t)\} dX(t) \; .$$

Since X is stationary it follows that $f\{X(t+s)\} = f\{X(t)\}$ and thus the derivative of $f\{X(t+s)\}$ with respect to s is equal to zero. Multiplying both sides of (3.65) with $f\{X(t)\}$ and integration with respect to $X(t)$ yields the ordinary differential equation

$$0 = -\frac{\partial}{\partial x}\left\{m(x)f(x)\right\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left\{\sigma^2(x)f(x)\right\} \tag{3.66}$$

and integrating twice with respect to x yields

$$\sigma^2(x) = \frac{2}{f(x)}\int_{-\infty}^{x} m(u)f(u)du \; .$$

Another way to derive the same estimator is to use the Tanaka-Meyer formula (3.63). Taking the expectation on both sides of (3.63) gives

$$\begin{aligned}\mathrm{E}\Lambda_t(x) &= \mathrm{E}[(X(t)-x)^-] - \mathrm{E}[(X(0)-x)^-] \\ &\quad +\mathrm{E}\left[\int_0^t \boldsymbol{I}_{(-\infty,x]}(X(s))\{m(X(s))ds + \sigma(X(s))dW(s)\}\right] \\ &= t\int_{-\infty}^{x} m(u)f(u)du \; .\end{aligned}$$

Choosing $\hat{\Lambda}_t(x) \stackrel{\text{def}}{=} \mathrm{E}\Lambda_t(x)$ and inserting it into (3.58) yields the same result as above.

Stanton (1997) proposes a nonparametric method that separately estimates the drift m and the diffusion coefficient σ. $S_t^{(n)}(x)$ coincides with one of the estimators given there. As we are interested in a smooth estimator for σ^2 and do not want to use any information about the drift coefficient m, we will apply the estimator in (3.60).

3.4.3 Estimation of the drift

Jiang and Knight (1997) apply the Kolmogorov forward equation (3.65) to develop a nonparametric estimator for the drift m. Integrating (3.66) with respect to x yields for the drift function

$$m(x) = \frac{1}{2f(x)} \frac{\partial}{\partial x}\{\sigma^2(x) f(x)\} = \frac{1}{2}\left\{\frac{\partial}{\partial x}\sigma^2(x) + \sigma^2(x)\frac{f'(x)}{f(x)}\right\} \tag{3.67}$$

Using (3.67) we define a nonparametric estimator of m by replacing the diffusion coefficient σ^2, the marginal density f and its derivative f' by their nonparametric estimators.

Estimators for σ^2 and f are already given in the previous sections. From (3.54) and the definition of $f_T^{(n)}(x)$, a natural way to estimate the derivative of the marginal density is

$$\hat{f}'(x) \stackrel{\text{def}}{=} \frac{\partial}{\partial x} f_T^{(n)}(x) = \frac{1}{nth_n^2} \sum_{i=0}^{[nt]} K'\left(\frac{X_i - x}{h_n}\right),$$

Wand and Jones (1995).

Jiang and Knight (1997) show, that the estimator $m_T^{(n)}(x)$ given by

$$m_T^{(n)}(x) \stackrel{\text{def}}{=} \frac{1}{2}\left\{\frac{\partial}{\partial x} S_T^{(n)}(x) + S_T^{(n)}(x)\frac{\hat{f}'(x)}{f_T^{(n)}(x)}\right\} \tag{3.68}$$

is pointwise consistent for $m(x)$. Under additional technical assumptions, Jiang and Knight (1997) show, that $S_T^{(n)}$ is differentiable.

3.4.4 Fixed Sample Properties

To investigate the performance of the proposed estimators we apply them to simulated diffusion processes.

The general stochastic differential equation, that the simulated process X follows is

$$dX(t) = m\{\theta, X(t)\}dt + \sigma\{\theta, X(t)\}dW(t) \qquad t > 0 \tag{3.69}$$

where θ is a parameter vector. To get discrete observations of X we use a Milstein scheme as in (3.52).

In the empirical analysis we test parametric models for the diffusion coefficient of the spot rate and various stock price processes. For this reason we will investigate the fixed sample properties of the nonparametric estimators applied to these models. A summary of the investigated models is given in Table 3.1. The parameters are

Name	$\sigma(x)$	θ_3
constant (VC)	θ_3	0.013
square root (CIR)	$\theta_3\sqrt{x}$	0.066
Chan, Karolyi, Longstaff, Sanders (CKLS)	$\theta_3 x^{1.5}$	1.2

Table 3.1: Diffusion coefficient models used in the simulation study

chosen accordingly to the estimated values in Ahn and Gao (1999). They estimated the parameters of different models applied to the one month US treasury bill rate. Note, that they did not estimate the parameters for all combinations of drift and diffusion coefficients that we use here. However, the parameters in our simulation study generate trajectories, that are positive and in the range of about 0.02 - 0.2 and thus might be a good choice to simulate interest rate processes.

Ahn and Gao (1999) also estimate a model introduced by Duffie and Kan (1996), where the diffusion coefficient is given by $\sigma(x) = \sqrt{\theta_3 + \theta_4 x}$. Since this model is not consistent for values of $X(t)$ smaller than $-\theta_3/\theta_4$ we will not use it in our simulation study.

To study the performance of the drift function estimator given in 3.4.3 and the influence of the unknown drift $m(x)$ on the estimates of σ^2, we combine each of the three diffusion coefficients with a drift function proposed by Ahn and Gao (1999), i.e. we simulate paths from three diffusion models. The function is given in Table 3.2 along with the parameter values estimated by Ahn and Gao (1999). For the reason of empirical relevance we use the given parameter values in our simulation.

Name	$m(x)$	θ_1	θ_2
Ahn-Gao model (AG)	$\theta_1(\theta_2 - x)x$	3.4	0.08

Table 3.2: Ahn-Gao model used in the simulation study

For every model we simulate 1000 paths of length $nT = 2500$. To simulate the trajectories we apply the Milstein scheme (3.52) with $\delta = 1/10$. This means, that we calculate 10 realizations per day but sample the data daily. Since the parameter values given in Tables 3.1 and 3.2 are annual values, we choose $n = 250$ (250 trading days per year), i.e. we have $T = 10$ years of observations.

We start with the diffusion coefficient estimation. The estimator we apply is the local time estimator given in (3.60). The mean of the estimated functions is shown in Figure 3.6 along with the 90% empirical confidence bands (green) and the true function (red). We find in the figures that the means of the estimates for the three diffusion coefficient functions are close to the corresponding true functions and we therefore conclude, that the estimators are unbiased in this situation.

In the first figure, constant σ^2, we find, that the confidence bands are small for states x of the process near 0.08. The reason is, that the level of mean reversion of the first model is 0.08 and therefore realizations of the process close to 0.08 occur more frequently than realizations in other regions of the state space. The middle and lower plot show that the width of the confidence bands increases with the level of X. The reason is that σ^2 is increasing in x and the variance of the estimator $S_T^{(n)}$ depends on the level of the true function σ^2 as it can be seen from Proposition 3.1.

With the knowledge about a nonparametric estimate of σ^2 we are now able to estimate the drift coefficient as in 3.4.3. The results for the Ahn-Gao drift function estimated from the three models described above are shown in Figure 3.7. The plots show the mean of the estimates (black), the true drift function (red) and the empirical 90% confidence band (green). As for the estimation of σ^2 the mean of the estimated drift is close to the true drift in all three models. However, it seems, that the estimator under-estimates the drift for values of x larger than 0.08 (the level of mean-reversion) and over-estimates the drift for $x < 0.08$.

Even if the bias of the drift estimator in this situation is negligible, 0 is contained in all confidence bands. Heuristically this means, that a drift function that is constantly 0 seems to be very likely in the situation here. Since the used parameter values correspond to estimated values by Ahn and Gao (1999) for an interest rate process, we doubt that for these kind of processes one can significantly distinguish between a zero drift and a mean reverting or quadratic drift. The reason is not only, that the drift is close to zero, but that the confidence bands are quite large. In particular 0 is included. The large confidence bands can be explained by the relative small sample size T, i.e. the number of years ($T = 10$) in which we have observations is too small to get a reliable drift estimate and to distinguish between a non zero and a zero drift. The situation would change when we use different parameters. In particular, when we increase the speed of adjustment parameter θ_1, we increase the influence of the drift on the instantaneous behavior of the process. In the empirical analysis we come again to that point.

It also appears from Figure 3.7 that the diffusion coefficient does have an influence on the preciseness of the drift estimation. In particular, diffusion coefficients, that are increasing in x, like the square root and the CKLS model, produce non constant confidence bands for the drift estimation. Heuristically speaking, the drift estimate becomes more imprecise the larger the state of the process, and thus the larger the diffusion coefficient function. The reason is, that the impact of the drift on the increments $X_{i+1} - X_i$ of the process becomes smaller when the diffusion coefficient gets larger. The diffusion coefficient can be interpreted as the instantaneous variance of

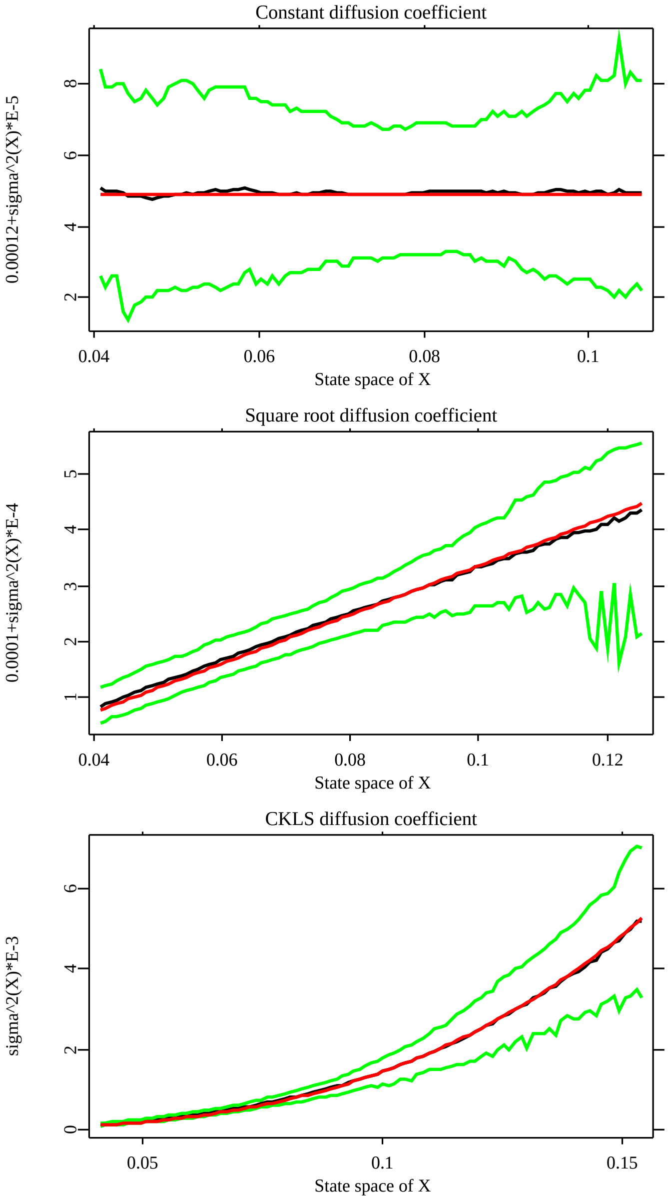


Figure 3.6: Nonparametric estimates of σ^2 together with the empirical 90% confidence band.

these increments. And the larger this variance is, the more difficult is the estimation of the instantaneous expectation $m(x)$. This phenomena can be compared to a univariate mean estimation, the higher the variance of the underlying random variable the more imprecise is the mean estimator.

3.5 Testing the Continuous Time Model

We now introduce test statistics to test the hypothesis in Section 3.2. The considered tests can be divided into two groups. The first group is based on the comparison of a nonparametric density estimate and a parametric density estimate as implied by the null hypotheses. The density that is used can be the marginal density as well as the transition density. The tests in the second group directly compare nonparametric estimates of the drift and/or the diffusion coefficient with their parametric forms implied by the null hypothesis.

Another feature of the proposed tests is the norm, that is used to compare two functions. We will use the L_∞-norm (sup-norm) and the L_2-norm. The use of the L_∞-norm has the advantage, that pointwise confidence intervals can be build around the nonparametric estimate. On the other hand, test statistics based on the L_2-norm have a known asymptotic distribution, a χ^2-distribution. We will here introduce tests, based on both norms. We start with tests based on density comparison.

3.5.1 Tests based on density comparison

We introduce two tests based on the comparison of the marginal density and the transition density respectively. Since both densities depend on the drift term m of X, both tests require, that the diffusion X is stationary and α-mixing and that T tends to infinity. On the other hand the densities are influenced by σ^2 and thus we also require that n tends to infinity.

Comparing The Marginal Densities

A test that compares the marginal density f of X as implied by the null hypothesis with the nonparametric estimate of f was introduced by Aït-Sahalia (1996). He applied this test to the 7-days Eurodollar spot rate. However, the test is of general interest when parametric models for diffusion processes are investigated.

For a diffusion process X that solves (3.1) it follows from the Kolmogorov forward equation (3.65), that the marginal density f of X is given by

$$f(x) = \frac{\xi}{\sigma^2(x)} \exp\{\int_{x_0}^{x} \frac{2m(z)}{\sigma^2(z)} dz\} \tag{3.70}$$

where $x_0 \in I_X$ is an arbitrary number and ξ is a normalization constant to ensure, that $f(x)$ integrates to 1, Karlin and Taylor (1981). Using the mapping between m, σ^2 and f given by (3.70), we can test the null hypotheses $H_0(m)$ and $H_0(\sigma^2)$ by testing a null hypothesis about f, i.e. we test the hypothesis

$$H_0(f) : \exists \theta_0 \in \Theta \ : \ \text{for every } t \in [0,T] : \ f\{X(t)\} = f\{\theta_0, X(t)\} \qquad \text{P-a.s.}$$

against a purely nonparametric alternative.

A corresponding test statistic is then

$$\frac{1}{k}\sum_{l=1}^{k}\left(f(\hat{\theta}, x_l) - f_T^{(n)}(x_l)\right)^2 f_T^{(n)}(x_l)$$

where x_l, $l = 1, \ldots, k$ are equidistant grid points in I_X and $\hat{\theta}$ is a square root consistent estimator for θ.

Aït-Sahalia (1996) used a different test statistic, i.e.

$$M \stackrel{\text{def}}{=} nb_n \min_{\theta \in \Theta} \frac{1}{n}\sum_{i=0}^{n-1}\left(f(\theta, X_i) - f_T^{(n)}(X_i)\right)^2$$

where b_n is the bandwidth used for the nonparametric estimation of f and nb_n is a normalizing constant. To calculate $\hat{M}$ he first estimates θ by

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} \frac{1}{n}\sum_{i=0}^{n-1}\left(f(\theta, X_i) - f_T^{(n)}(X_i)\right)^2$$

and then inserts the estimator into the test statistic M, i.e.

$$\hat{M} \stackrel{\text{def}}{=} nb_n \frac{1}{n}\sum_{i=0}^{n-1}\left(f(\hat{\theta}, X_i) - f_T^{(n)}(X_i)\right)^2 .$$

However, we can use any square root consistent estimator for θ without changing the asymptotic properties of $\hat{M}$.

Aït-Sahalia (1996) shows, that the asymptotic distribution of $\hat{M}$ is normal, i.e.

$$b_n^{-1/2}\{\hat{M} - E_M\} \longrightarrow^d \text{N}(0, V_M)$$

with

$$\begin{aligned} E_M &= \left(\int_{-\infty}^{\infty} K^2(x)dx\right)\left(\int_{I_X} f^2(x)dx\right) \\ V_M &= \left(\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} K8u)K(u+x)du\right\}^2 dx\right)\left(\int_{I_X} f^4(x)dx\right) \end{aligned}$$

which can be estimated by

$$\hat{E}_M = \left(\int_{-\infty}^{\infty} K^2(x)dx\right)\left(\sum_{i=0}^{[nt]} f_T^{(n)}(X_i)\right)$$

$$\hat{V}_M = \left(\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} K8u)K(u+x)du\right\}^2 dx\right)\left(\sum_{i=0}^{[nt]}\left\{f_T^{(n)}(x)\right\}^3\right).$$

Since the proposed test statistic $\hat{M}$ is based on properties of the marginal density of X, it cannot distinguish between processes with the same marginal distributions but different dependency structures. Thus an unique identification of a pair of drift and diffusion coefficients is not possible. To overcome this problem, Hong and Li (2002) propose a test based on the comparison of the transition densities.

Comparing The Transition Densities

Since the process X given in (3.1) is a Markov process, all information about it's distribution is contained in it's transition density as defined in (3.64) and the initial distribution of X_0. Thus, there is a unique mapping between a pair of drift and diffusion coefficient $\{m, \sigma\}$ and the corresponding transition density $f(x,t|y,s)$. The hypotheses in Section 3.2 can therefore be transformed into a single hypothesis about the transition density.

$$\begin{aligned} H_0(f) \ : \ \exists\theta_0 \in \Theta \ : \ & \text{for every } s,t \in [0,T] : \\ & f\{X(t),t|X(s),s\} = f\{X(t),t|\theta_0,X(s),s\} \qquad \text{P-a.s.} \end{aligned} \tag{3.71}$$

where the alternative is again purely nonparametric.

A natural way to test a parametric model of the transition density is to compare a parametrically estimated density with it's nonparametric estimate. Since the convergence rate of the nonparametric estimator is even slower than the rate for the marginal density, Hong and Li (2002) suggest to apply the probability integral transformation to get a transformed series of data

$$Z_i \stackrel{\text{def}}{=} \int_{-\infty}^{X_i} f\{x, i/n \mid \theta, X_{i-1}, (i-1)/n\}dx$$

for $i = 0, \ldots, [nT]-1$ and $X_i = X(i/n)$ as in 3.3.1. Under $H_0(f)$ there exist a θ such that the series $\{Z_i, i = 0, \ldots, [nT]-1\}$ is independent and identically distributed U[0, 1], Rosenblatt (1952) and Diebold et al. (1998). The test proposed by Hong and Li (2002) is based on the comparison of the two-dimensional density $g(z_1, z_2)$

of (Z_i, Z_{i+1}) and the two-dimensional uniform density 1. The application of the probability transformation in this situation has two major advantages. Under the null hypothesis there is no dependency in the data and the two-dimensional density as implied by $H_0(f)$ is a constant, which means that there is no asymptotic bias for the nonparametric estimator under $H_0(f)$.

The bivariate density estimator applied to Z is

$$g_T^{(n)}(z_1, z_2) \stackrel{\text{def}}{=} ([nT] - j)^{-1} \sum_{i=j}^{[nT]-1} K_{h_n}(z_1 - \hat{Z}_i) K_{h_n}(z_2 - \hat{Z}_{i-j})$$

where $\hat{Z}$ is the transformed series implied by the null hypothesis and the parameter estimator $\hat{\theta}$. We remark, that Hong and Li (2002) applied a different estimator in order to reduce the boundary bias and the relatively large variance of $g_T^{(n)}$ in the boundary region. However, we do not apply any correction for the boundary bias in all other tests proposed here. In order to make the tests comparable we do not apply the bias correction for $g_T^{(n)}$ as in Hong and Li (2002).

The test decision is based on L_2 distance between $g_T^{(n)}$ and 1, the joint density of the bivariate U$[0,1]$ distribution, i.e.

$$\hat{M}(j) \stackrel{\text{def}}{=} \int_0^1 \int_0^1 [g_T^{(n)}(z_1, z_2) - 1]^2 dz_1 dz_2 \; .$$

We do not use any weighting function to weight the differences, since under $H_0(f)$ the observations of Z are uniformly distributed over the integration range. The test statistic is a centered and scaled version of $\hat{M}(j)$,

$$\hat{Q}(j) \stackrel{\text{def}}{=} \left[([nt] - j) h \hat{M}(j) - A_h \right] / V^{1/2}$$

with

$$\begin{aligned}
A_h &\stackrel{\text{def}}{=} (h^{-1} - 2) \int_{-1}^{1} K^2(u) du + 2 \int_0^1 \int_{-1}^{b} k(u, b) du db, \\
k(u, b) &\stackrel{\text{def}}{=} \frac{1}{k(u)} \int_{-1}^{b} K(v) dv \\
V &\stackrel{\text{def}}{=} 2 \left[\int_{-1}^{1} \left\{ \int_{-1}^{1} K(u + v) K(v) dv \right\}^2 \right]^2 .
\end{aligned}$$

Hong and Li (2002) show that $\hat{Q}(j)$ is under the null hypothesis asymptotically standard normally distributed for all j and the the covariance between $\hat{Q}(i)$ and $\hat{Q}(j)$ for $i \neq j$ converges in probability to zero.

3.5.2 Directly testing the diffusion coefficient

We now derive a direct test for the diffusion coefficient that does not incorporate any assumptions about the drift. The proposed test compares a nonparametric estimate of σ^2 to the smoothed parametric function that is implied by the null hypothesis. The nonparametric estimator we use, is $S_T^{(n)}$ given in (3.60). Since this estimator is asymptotically, for $n \to \infty$, independent of m, the test is asymptotically independent of m too. Hence we do not have to estimate the drift or parameters of it and therefore do not require that X is stationary or that T tends to infinity.

The hypotheses about the squared diffusion coefficient for the test are:

$$\begin{aligned} H_0(\sigma^2) \;&:\; \exists \theta_0 \in \Theta \;:\; \text{for every } t \in [0,T] :\; \sigma^2\{X(t)\} = \sigma^2\{\theta_0, X(t)\} \qquad \text{P-a.s.} \\ H_1(\sigma^2) \;&:\; \forall\, \theta \in \Theta \;:\; \text{for every } t \in [0,T] : \\ &\qquad |\sigma^2\{X(t)\} = \sigma^2\{\theta, X(t)\}| \geq c_n \Delta_n(X(t)) \qquad \text{P-a.s.} \end{aligned}$$

As for the EL test about the drift in Section 3.3.4 Δ_n, the local shift in the alternative, is a sequence of bounded functions and c_n is the order of difference between H_0 and the alternative. The difference here is that both, Δ_n and c_n do not depend on T but on n. With a similar argument as in Section 3.3.4 the construction of H_1 ensures that the power of the proposed test depends on the number of observations per unit of time n but does not depend on T.

To formalize the test, we make the following assumptions about σ and θ.

(CT2) The following holds for σ^2:

$$|\sigma^2(\theta, x) - \sigma^2(\theta_0, x)| \leq D(x)\|\theta - \theta_0\| \qquad \forall\, x \in I_X$$

where $D(x)$ is a constant depending on x.

(CT3) $\hat{\theta}$ is a square root consistent parametric estimator of θ within the family of the parametric model, i.e. $\|\hat{\theta} - \theta\| = \mathcal{O}_p(n^{-1/2})$.

It is well known that for fixed n and h_n, $S_T^{(n)}(x)$ is a biased estimator of $\sigma^2(x)$. Thus we will not compare it directly with $\sigma^2(\hat{\theta}, x)$ but with its smoothed version

$$\tilde{\sigma}^2(x) \stackrel{\text{def}}{=} \frac{\sum_{i=0}^{[nt]} K_{h_n}(X_i - x)\sigma^2(X_i)}{\sum_{i=0}^{[nt]} K_{h_n}(X_i - x)}. \tag{3.72}$$

We remark that Proposition 3.1 is also valid, if we replace $\sigma^2(x)$ in (3.61) by $\tilde{\sigma}^2(\hat{\theta}, x)$.

For an arbitrary point $x \in I_X$ we introduce the test statistic

$$\begin{aligned} T_T^{(n)}(x) &= \sqrt{nh_n L_T^{(n)}(x_l)} \left(\frac{S_T^{(n)}(x)}{\tilde{\sigma}^2(\hat{\theta}, x)} - 1 \right) \\ &= \sqrt{nh_n L_T^{(n)}(x_l)} \left(\frac{S_T^{(n)}(x)}{\tilde{\sigma}^2(x)} \frac{\tilde{\sigma}^2(x)}{\tilde{\sigma}^2(\hat{\theta}, x)} - 1 \right) \\ &= \sqrt{nh_n L_T^{(n)}(x_l)} \left(\frac{S_T^{(n)}(x)}{\tilde{\sigma}^2(x)} - 1 \right) + R_T^{(n)}(x) \end{aligned} \tag{3.73}$$

From Proposition 3.1 we know that $T_T^{(n)}(x)$ converges in distribution to $Z + R$, where Z is standard normally distributed and R is the limit of

$$R_T^{(n)}(x) = \sqrt{nh_n L_T^{(n)}(x)} \frac{\tilde{\sigma}^2(x) - \tilde{\sigma}^2(\hat{\theta}, x)}{\tilde{\sigma}^2(\hat{\theta}, x)} \frac{S_T^{(n)}(x)}{\tilde{\sigma}^2(x)}$$

for $n \to \infty$. Proposition 1 and 3 in Florens-Zmirou (1993) imply that $S_T^{(n)}(x)/\tilde{\sigma}^2(x)$ converges to 1 in the L^2 sense if $nh^4 \ll$ tends to 0. Under H_0 assumption (CT2) and (CT3) imply

$$\frac{\tilde{\sigma}^2(x) - \tilde{\sigma}^2(\hat{\theta}, x)}{\tilde{\sigma}^2(\hat{\theta}, x)} = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)$$

and it follows with Proposition 1 in Florens-Zmirou (1993) that

$$R_T^{(n)}(x) = \sqrt{L_T(x)/T}\mathcal{O}_p(\sqrt{h_n})\left\{1 + \mathcal{O}_p(nh_n^4)\right\} = \mathcal{O}_p(\sqrt{h_n}).$$

To study the properties of the test statistic $T_T^{(n)}(x)$ under the alternative H_1, we make the following assumption about c_n and Δ_n.

(CT4) $\Delta_n(x)$ is bounded with respect to n and x. and $c_n = 1/\sqrt{nh_n}$.

With assumption (CT4) we have under the alternative

$$\frac{\tilde{\sigma}^2(x) - \tilde{\sigma}^2(\hat{\theta}, x)}{\tilde{\sigma}^2(\hat{\theta}, x)} \longrightarrow \frac{\sigma^2(x) - \sigma^2(\hat{\theta}, x)}{\sigma^2(\hat{\theta}, x)} \longrightarrow \frac{c_n \Delta_n(x)}{\sigma^2(\theta_0, x)}$$

and thus $R_T^{(n)}(x) = \Delta_n(x)/\sigma^2(\theta_0, x)$

To get a global goodness-of-fit test we choose k arbitrary points $x_1, \ldots, x_k \in I_X$ and built the test statistic

$$T_T^{(n)} = \sum_{l=1}^{k} \{T_T^{(n)}(x_l)\}^2 \,. \tag{3.74}$$

We now study the asymptotic distribution of $T_T^{(n)}$ under the null hypothesis.

PROPOSITION 3.2 *If H_0 holds and nh_n^3 tends to zero, we have for every k and every set of points $x_1, \ldots, x_k \in I_X$ with $x_i \neq x_j$ for $i \neq j$ that $T_T^{(n)}$ converges in distribution to a χ^2-distributed random variable with k degrees of freedom.*

PROOF :
With (3.62) we have for every $n > n_0$ and for every $i \neq j$ that

$$\mathrm{Cov}\{T_T^{(n)}(x_i), T_T^{(n)}(x_j)\} = 0$$

and from Proposition 3.1 it follows that $T_T^{(n)}$ is asymptotically χ^2-distributed with k degrees of freedom. □

With a similar proof we obtain that $T_T^{(n)}$ converges under the alternative to a non-central χ^2-distributed random variable with k degrees of freedom and non-centrality parameter $\sum_{l=1}^k \Delta_n^2(x_l)/\sigma^4(\theta_0, x_l)$.

We remark that the proposed test statistic is asymptotically equivalent to the L_2 distance between $S_T^{(n)}$ and $\tilde{\sigma}^2(\hat{\theta}, .)$. In a nonparametric regression context Härdle and Mammen (1993) propose a L_2 test statistic

$$T_{HM} \stackrel{\mathrm{def}}{=} nh_n^{\frac{1}{2}} \int \{S_T^{(n)}(x) - \tilde{\sigma}^2(\hat{\theta}, x)\}^2 \pi(x) dx$$

with a certain weight function $\pi(x)$. For a fixed bandwidth h_n and with $k_n = 1/(2h_n)$, $x_l = h_n + 2h_n(l-1)$ for $l = 1, \ldots, k_n$ we get that

$$\frac{1}{k_n} T_T^{(n)} = \frac{1}{k_n} nh_n \sum_{l=1}^{k_n} \frac{L_T^{(n)}(x_l)}{\tilde{\sigma}^4(\hat{\theta}, x)} \{S_T^{(n)}(x_l) - \tilde{\sigma}^2(\hat{\theta}, x_l)\}^2$$

is the Riemann approximation of $\sqrt{h_n} T_{HM}$ with the weight function $\pi(x) = L_T^{(n)}(x)/\tilde{\sigma}^4(\hat{\theta}, x)$.

It appears in the simulation study in Section 3.5.5 that a test based on $T_t^{(n)}$ is too conservative for all considered models. The reason could be, that $\pi(x)$ does not reflect features of the empirical distribution of $S_T^{(n)}(x_l) - \tilde{\sigma}^2(\hat{\theta}, x_l)$. One way to improve the test is to change the weighting function $\pi(x)$.

Hoffmann (1999) applies another approach to estimate the diffusion coefficient, that is based on time discretization. From Itô's formula we get with an appropriate function μ depending on m and σ

$$\begin{aligned} \{X_{i+1} - X_i\}^2 &= \int_{i/n}^{(i+1)/n} \sigma^2\{X(u)\}du \\ &\quad + \int_{i/n}^{(i+1)/n} \mu\{X(u)\}dW(u) + \mathcal{O}(n^{-2}) \\ &\approx \sigma^2(X_i)\frac{1}{n} + \mu(X_i)\sqrt{\frac{1}{n}}w_i \end{aligned} \tag{3.75}$$

where $w_i \sim \mathrm{N}(0,1)$ for all $i = 0, \ldots, [Tn]-1$. Hoffmann (1999) applied this approach to develop an adaptive nonparametric estimation procedure for σ using wavelets. Note, that a Nadaraya-Watson estimator for $\sigma^2(x)$ applied in (3.75) coincides with $S_T^{(n)}(x)$.

The approximation in (3.75) suggests to modify the test statistic in a way, that captures the heteroscedasticity of the error terms $\mu(X_i)\sqrt{1/n}w_i$. For this reason we will now propose a test statistic based on the empirical likelihood concept.

3.5.3 Empirical Likelihood Test About The Diffusion coefficient

The main advantage of Empirical Likelihood methods is their ability to studentize internally and to correct test statistics and confidence intervals for empirical properties of the data. This is the reason, why we introduce a test about σ^2 based on the EL methodology. We follow the results in 3.3.4 to derive the test statistic.

For the sake of simplicity we study the test of σ based on the observations up to time $T = 1$. The general case follows directly.

With the notation

$$\eta_i^{(x)}\{s\} \stackrel{\text{def}}{=} K\left(\frac{X_i - x}{h_n}\right)\left[n\left\{X_{i+1} - X_i\right\}^2 - s(x)\right] \qquad i = 0, \ldots, n-1$$

for a positive function s with support I_X, we get from the definition of $S_1^{(n)}(x)$ for any $x \in I_X$

$$S_1^{(n)}(x) - \tilde{\sigma}^2(\hat{\theta}, x) = \frac{1}{nh_n L_1^{(n)}(x)} \sum_{i=0}^{n-1} \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}$$

and might rewrite $T_1^{(n)}(x)$ in the following way

$$T_1^{(n)}(x) = \frac{1}{\tilde{\sigma}^2(\hat{\theta}, x)\sqrt{L_1^{(n)}(x)}} \sqrt{\frac{n}{h_n}} \sum_{i=0}^{n-1} \frac{1}{n}\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\} \tag{3.76}$$

The first part of (3.76) is a factor to standardize the variance of $T_1^{(n)}(x)$. The second part is a mean over $\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta},.)\}$ that gives equal weight $1/n$ to every i.

To introduce the EL concept we now replace $T_1^{(n)}(x)$ by a similar statistic which gives different weights to each i.

$$\tilde{T}_1^{(n)}(x) = \frac{1}{\tilde{\sigma}^2(\hat{\theta},x)\sqrt{L_1^{(n)}(x)}}\sqrt{\frac{n}{h_n}}\sum_{i=0}^{n-1} p_i \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta},.)\} \tag{3.77}$$

with $\sum_{i=0}^{n-1} p_i = 1$. For a fixed point x we follow Chen et al. (2001) to derive an EL test statistic.

The empirical likelihood $\mathcal{L}$ for $s(x)$ is defined by

$$\mathcal{L}\{s(x)\} \stackrel{\text{def}}{=} \max \prod_{i=0}^{n-1} p_i(x) \tag{3.78}$$

subject to

$$\sum_{i=0}^{n-1} p_i(x) = 1 \;\text{ and }\; \sum_{i=0}^{n-1} p_i(x)\eta_i^{(x)}\{s\} = 0\,. \tag{3.79}$$

The second condition reflects, that under the null hypothesis $\mathrm{E}[\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta},.)\}]$ converges to 0 for $n \to \infty$ and $h_n \to 0$. The test is based on the EL ratio $\mathcal{L}\{\tilde{\sigma}^2(\hat{\theta},x)\}/\mathcal{L}\{S_1^{(n)}(x)\}$, which should be close to 1 if the null hypothesis is true. To formalize this idea and to derive a test statistic we study the properties of $\mathcal{L}\{s(x)\}$.

As in Section 3.3.4, we find the maximum of $\mathcal{L}\{s(x)\}$ by introducing Lagrange multipliers and maximizing the Lagrangian function

$$\mathcal{H}(p,\lambda_1,\lambda_2) \;=\; \sum_{i=0}^{n-1}\log p_i(x) - \lambda_1\sum_{i=0}^{n-1} p_i(x)\eta_i^{(x)}\{s\} - \lambda_2\left\{\sum_{i=0}^{n-1} p_i(x) - 1\right\}$$

The first order conditions are the equations in (3.79) and

$$\frac{\partial\mathcal{H}(p,\lambda_1,\lambda_2)}{\partial p_i(x)} = \frac{1}{p_i(x)} - \lambda_1\eta_i^{(x)}\{s\} - \lambda_2 = 0$$

for all $i = 0,\ldots,n-1$. We obtain the optimal weights as in Section 3.3.4

$$p_i(x) = n^{-1}\left[1 + \lambda(x)\eta_i^{(x)}\{s\}\right]^{-1} \tag{3.80}$$

where $\lambda(x)$ is the root of

$$\sum_{i=0}^{n-1}\frac{\eta_i^{(x)}\{s\}}{1+\lambda(x)\eta_i^{(x)}\{s\}} = 0. \tag{3.81}$$

Again, the maximum empirical likelihood is achieved at $p_i(x) = n^{-1}$ corresponding to $s(x) = S_1^{(n)}(x)$ and for a parameter estimate $\hat{\theta}$ we get the maximum empirical likelihood for the smoothed parametric model $\mathcal{L}\{\tilde{\sigma}^2(\hat{\theta}, x)\}$. The log-EL ratio is

$$\ell\{\tilde{\sigma}^2(\hat{\theta}, x)\} \stackrel{\text{def}}{=} -2 \log \frac{\mathcal{L}\{\tilde{\sigma}^2(\hat{\theta}, x)\}}{\mathcal{L}\{S_1^{(n)}(x)\}} = -2 \log[\mathcal{L}\{\tilde{\sigma}^2(\hat{\theta}, x)\} n^n].$$

As in the discrete time case for the EL test about the drift m, we have to show, that the Lagrange multipliers $\lambda(x)$ are tending to 0 uniformly in I_X.

THEOREM 3.5 *For $s(x) = \tilde{\sigma}^2(\hat{\theta}, x)$ and under the assumptions (CT1) - (CT4) we have*

$$\sup_{x \in I_X} |\lambda(x)| = \mathcal{O}_p\{(nh_n)^{-1/2}\}.$$

For the following we redefine $\bar{U}_j(x)$

$$\bar{U}_j(x) = \frac{1}{nh_n} \sum_{i=0}^{n-1} [\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}]^j \ . \tag{3.82}$$

PROOF :
With a similar argument as in the proof of Theorem 3.2 we can show

$$\begin{aligned} 0 &= \left| \sum_{i=0}^{n-1} \frac{\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}}{1 + \lambda(x)\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}} \right| \\ &\geq |\lambda(x)| \sum_{i=0}^{n-1} \frac{[\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}]^2}{1 + \lambda(x)\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}} - \left| \sum_{i=0}^{n-1} \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\} \right| \ . \end{aligned} \tag{3.83}$$

From this inequality and the definition of $\bar{U}_j(x)$ in (3.82) follows

$$0 \geq |\lambda(x)|\bar{U}_2(x) \left\{ 1 + |\lambda(x)| \max_{1 \leq j \leq n} \left| \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\} \right| \right\}^{-1} - |\bar{U}_1(x)| \ . \tag{3.84}$$

With the notation $\varepsilon_i = n(X_{i+1} - X_i)^2 - \sigma^2(X_i)$ we split $\bar{U}_1(x)$ into three parts

$$\begin{aligned} \bar{U}_1(x) &= I_1 + I_2 + I_3 \\ I_1 &= L_T^{(n)}[\tilde{\sigma}^2(\theta, x) - \tilde{\sigma}^2(\hat{\theta}, x)] \\ I_2 &= 1/n \sum_{i=0}^{n-1} K_{h_n}(x - X_i)\varepsilon_i \\ I_3 &= 1/nc_n \sum_{i=0}^{n-1} K_{h_n}(x - X_i)\Delta_n(X_i) \end{aligned}$$

Assumption (CT2) and (CT3) yield $I_1 = \mathcal{O}(n^{-1/2})$ and assumption (CT4) yields $I_3 = \mathcal{O}(c_n)$.

$$I_2 = L_T^{(n)}(x)\{S_T^{(n)}(x) - \sigma^2(x)\} + L_T^{(n)}(x)\{\sigma^2(x) - \tilde{\sigma}^2(x)\}$$

Since σ^2 is assumed to be continuous and $S_T^{(n)}$ is a consistent estimator for σ^2 with order nh_n^4, see Florens-Zmirou (1993), I_2 is of order $o_p(nh^4)$. Thus the leading term is I_3 and $\bar{U}_1(x)$ is of order $o_P(1/\sqrt{nh_n})$.

We now show that there exists a constant such that

$$\mathrm{P}[\inf_{x\in I_X} \bar{U}_2(x) > d_0] = 1 \ . \tag{3.85}$$

Similar to the decomposition of $\bar{U}_1(x)$ it can be shown, that it is sufficient to show that

$$\mathrm{P}\left[\inf_{x\in I_X} 1/(nh_n)\sum_{i=0}^{n-1} K^2\left(\frac{X_i - x}{h_n}\right)\varepsilon_i^2 > d_0\right] = 1 \ .$$

$$\begin{aligned}
\frac{1}{nh_n}\sum_{i=0}^{n-1} K^2\left(\frac{X_i - x}{h_n}\right)\varepsilon_i^2 \ &\geq\ \frac{1}{nh_n}\sum_{i=0}^{n-1} \boldsymbol{I}\left\{K^2\left(\frac{X_i - x}{h_n}\right) > C_1\right\} \\
&\qquad \times K^2\left(\frac{X_i - x}{h_n}\right)\varepsilon_i^2 \\
&\geq\ nC_1\sum_{i=0}^{n-1}\left(\frac{\varepsilon_i}{n}\right)^2 \\
&\sim\ nC_1\sum_{i=0}^{n-1}\mu^2(X_i)w_i^2 \\
&\geq\ C_2\sum_{i=0}^{n-1} w_i^2
\end{aligned}$$

where we use an approximation similar to (3.75). w_i are standard normal and μ is given in (3.75).

To complete the proof of the theorem, it remains to show that

$$\max_{1\leq j\leq n}\sup_{x\in I_X}\left|\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}\right| = o_p\{(nh_n)^{1/2}\} \ . \tag{3.86}$$

Let $v_i = \sup_{x\in I_X}|\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}|$. Since K, σ^2 and Δ_n are bounded in I_X, we have with a similar argument as above, that $v_i \leq C_1|\varepsilon_i|$. From the Chebyshev inequality follows

$$\begin{aligned}
\mathrm{P}\left(v_i > (nh_n)^{1/2}\right) \ &\leq\ \mathrm{P}\left(|\epsilon_i| \geq C_2(nh_n)^{1/2}\right) \\
&\leq\ C_3\mathrm{E}|\varepsilon_i|(nh_n)^{-1/2}
\end{aligned}$$

With the approximation (3.75) we have that $\mathrm{E}|(1/n)\varepsilon_i| = \mathcal{O}_P(n^{-2})$ and thus

$$\mathrm{P}\left(v_i > (nh_n)^{1/2}\right) \leq C_4 (nh_n)^{-1/2} 1/n$$

Thus, we have with assumption (CT1) that $\sum_{n=1}^{\infty} \mathrm{P}\left(v_i > (nh_n)^{1/2}\right) < \infty$. According to the Borel-Cantelli lemma, $v_i > (nh_n)^{1/2}$ finitely often with probability 1. This means that $\max_{1\leq i \leq n} v_i > (nh_n)^{1/2}$ finitely often, which proofs (3.86).

To finish the proof of Theorem 3.5 we take the supremum over all $x \in I_X$ in (3.84)

$$\sup_{x\in I_X} |\lambda(x)| \bar{U}_2(x) \left\{1 + |\lambda(x)| \max_{1\leq j\leq n} \left|\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}\right|\right\}^{-1} = \mathcal{O}_p\{(nh_n)^{-1/2}\}$$

which completes the proof together with (3.85) and (3.86).

□

As in Section 3.3.4, we apply the power series expansion of $1/(1-.)$ to (3.81). Theorem 3.5 then yields

$$\sum_{i=0}^{n-1} \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}) \left[\sum_{j=0}^{\infty} (-\lambda(x))^j (\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\})^j\right] = 0. \tag{3.87}$$

and we have from (3.82), Theorem 3.5 and (3.87)

$$\lambda(x) = \bar{U}_2^{-1}(x)\bar{U}_1(x) + \tilde{o}_p\{(nh_n)^{-1}\}. \tag{3.88}$$

From (3.80), Theorem 3.5 and the Taylor expansion of $\log(1+.)$ we get in a similar way as in (3.34)

$$\begin{aligned} \ell\{\tilde{\sigma}^2(\hat{\theta}, x)\} &= -2\log[\mathcal{L}\{\tilde{\sigma}^2(\hat{\theta}, x)\} n^n] \\ &= 2nh_n\lambda(x)\bar{U}_1 - nh_n\lambda^2(x)\bar{U}_2 + \tilde{o}_p\{(nh_n)^{-3/2}\} \end{aligned} \tag{3.89}$$

Inserting (3.88) in (3.89) yields

$$\ell\{\tilde{\sigma}^2(\hat{\theta}, x)\} = nh_n\bar{U}_2^{-1}(x)\bar{U}_1^2(x) + \tilde{o}_p\{(nh_n)^{-3/2}\}$$

and with the definition of U_1 and U_2 we approximate $\ell\{\tilde{\sigma}^2(\hat{\theta}, x)\}$ by

$$\ell\{\tilde{\sigma}^2(\hat{\theta}, x)\} \approx \frac{\left(\sum_{i=0}^{n-1} \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}\right)^2}{\sum_{i=0}^{n} (\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\})^2}$$

and for the general case $T \neq 1$ we have

$$\ell_t\{\tilde{\sigma}^2(\hat{\theta}, x)\} \approx \frac{\left(\sum_{i=0}^{[Tn]-1} \eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\}\right)^2}{\sum_{i=0}^{[Tn]-1} (\eta_i^{(x)}\{\tilde{\sigma}^2(\hat{\theta}, .)\})^2}. \tag{3.90}$$

For k points $x_1, \ldots, x_k$ we define the global EL goodness-of-fit test statistic $\mathcal{T}_1^{(n)}$ as in Chen et al. (2001),

$$\mathcal{T}_1^{(n)} \stackrel{\text{def}}{=} \sum_{l=1}^{k} \ell\{\tilde{\sigma}^2(\hat{\theta}, x_l)\}$$

and for $T \neq 1$

$$\mathcal{T}_T^{(n)} \stackrel{\text{def}}{=} \sum_{l=1}^{k} \ell_T\{\tilde{\sigma}^2(\hat{\theta}, x_l)\} \; . \tag{3.91}$$

As in Chen et al. (2001) we can show, that the asymptotic distribution of $\mathcal{T}_t^{(n)}$ under the null hypothesis is again a χ^2-distribution with k degrees of freedom and that $1/k\mathcal{T}_t^{(n)}$ is asymptotically equivalent to a L_2 distance between $S_T^{(n)}$ and $\tilde{\sigma}^2(\hat{\theta}, x)$. This means, that both test statistics, $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$, are asymptotically equivalent. However, the simulation study shows, that the ability of the EL test statistic to internally use features of the empirical distribution of $S_T^{(n)} - \tilde{\sigma}^2(\hat{\theta}, x)$ results in a smaller empirical level and thus produces more reliable results.

3.5.4 Testing Time-inhomogeneous Diffusion Coefficients

To extend the proposed methodology to time-inhomogeneous coefficients, we now assume that the diffusion process X is given as the solution of

$$dX(t) = m\{X(t), t\}dt + \sigma\{X(t), t\}dW(t) \qquad t > 0$$

and we replace our null hypothesis about σ^2 by

$$\begin{aligned} H_0' \;\; : \;\; & \exists \theta_0 \in \Theta \; : \; \text{for every } t \in [0, T] : \\ & \sigma^2\{X(t), t\} = \sigma^2\{\theta_0, X(t), t\} \qquad \text{P-a.s. .} \end{aligned}$$

Furthermore we replace assumption (CT2) by

(CT2')

$$|\sigma^2(\theta, x, t) - \sigma^2(\theta_0, x, t)| \leq D(x, t)\|\theta - \theta_0\| \qquad \forall\, x \in I_X, \; \forall\, t \in [0, T]$$

where $D(x, t)$ is a constant depending on x and t and the set I_X is defined as in (3.2).

Applying Itô's formula to $g(x, t) \stackrel{\text{def}}{=} \int_0^x 1/\sigma(\hat{\theta}, z, t)dz$, Karatzas and Shreve (1991), we get for $Y(t) \stackrel{\text{def}}{=} g(X(t), t)$

$$dY(t) = m_Y\{X(t), t\}dt + \frac{\sigma(X(t), t)}{\sigma(\hat{\theta}, X(t), t)}dW(t) \qquad t > 0$$

where $m_Y(x,t)$ is given by

$$m_Y(x,t) = \frac{\partial}{\partial t} g(x,t) + \frac{\partial}{\partial x} g(x,t) m(x,t) + 0.5 \frac{\partial}{\partial x^2} g(x,t) \sigma^2(x, \hat{\theta}, s)$$

By replacing x by $g^{-1}(y)$ in the last equation, we get from the assumptions (CT2') and (CT3) under the null hypothesis a diffusion Y with constant diffusion coefficient equal to $1 + \mathcal{O}_p(n^{-1/2})$, for which 1 is a square root consistent estimator. Since the proposed tests do not depend on the drift, and the diffusion coefficient of Y is asymptotically independent of t, we are now in the situation described above.

3.5.5 Fixed Sample Properties

We investigate the finite sample properties of the two proposed tests about the diffusion coefficient by simulating various models and applying the test to the simulated data. We again simulate from the process given as the solution of (3.1). For the simulation we apply the Milstein scheme (3.52).

For the reason of empirical relevance we will investigate the fixed sample properties of the tests applied to the models that are used in the empirical analysis, compare 3.4.4 and Table 3.1.

In addition to the Ahn-Gao model, compare 3.4.4, we also apply a linear mean reverting drift in our simulation here. The used parameter values again correspond to the estimated values by Ahn and Gao (1999) for the one month US treasury bill rate. A summary of the drift functions used in the simulation is given in Table 3.3.

Name	$m(x)$	θ_1	θ_2
Linear mean reverting model (LMR)	$\theta_1(\theta_2 - x)$	0.13	0.08
Ahn-Gao model (AG)	$\theta_1(\theta_2 - x)x$	3.4	0.08

Table 3.3: Drift functions used in the simulation study

We simulate from every combination of the given diffusion coefficient and drift function, i.e. we simulate paths of 6 different processes. For every model we simulate 1000 paths of length $nT = 1000$, $nT = 3000$ and $nT = 5000$. For $nT = 1000, 3000$ we simulate 10 observations each day, but sample the process daily. For $nT = 5000$ we simulate 20 observations per day and sample the data daily. Since the parameter values given in Tables 3.1 and 3.3 are annual values, we choose $n = 250$ (250 trading days per year) and $T = 4, 12, 20$ years.

Both test statistics $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ depend on the choice of the degrees of freedom k, on the bandwidth h and on the points $x_1, \ldots, x_k$. For given degrees of freedom k we choose

$$h = 1/(2k) \quad \text{and} \quad x_l = h + 2h(l-1) \tag{3.92}$$

for $l = 1, \ldots, k$. This choice guarantees that the random variables $T_T^{(n)}(x_l)$ and $\ell\{\tilde{\sigma}^2(\hat{\theta}, x_l)\}$ are uncorrelated. The parameter estimates are obtained from the quadratic variation.

For $nT = 1000$ the empirical levels of both tests $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ are shown in Figure 3.8. The results indicate, that the empirical level of the EL test statistic $\mathcal{T}_T^{(n)}$ is close to the nominal level only for degrees of freedom between about 4 and 6 and the test based on $T_T^{(n)}$ is to liberal even for small degrees of freedom. This statement holds independently of the model that is tested. The nonlinearity of the drift seems to have almost no impact on the empirical level of the test.

Figure 3.9 shows the empirical level for the test about the CKLS diffusion coefficient when the length of the simulated paths is 3000 (upper plot) and 5000 (lower plot). For the simulation we used a nonlinear drift (the AG model). As we expected the empirical level is closer to the nominal level when the sample size is increasing. For the LMR drift and the other two diffusion functions we get similar pictures.

The simulations show, that the performance of the test strongly depends on the choice of k, the degrees of freedom of the asymptotic χ^2-distribution. If k is to large, the approximation of $T_T^{(n)}(x)$ and $\ell_T\{\tilde{\sigma}^2(\hat{\theta}, x)\}$ by normally distributed random variables fails and thus the test statistics $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ are not χ^2-distributed. In addition we see from Figure 3.8 that the empirical level of the test increases with k. The reason seems to be clear, the larger k the smaller is h_n. It is a well known feature of nonparametric estimators, that the variance of the estimator is decreasing in h_n. Thus a larger k, smaller h_n, yields a larger variance of $T_T^{(n)}(x)$ and $\ell_T\{\tilde{\sigma}^2(\hat{\theta}, x_l)\}$ and thus a larger expectation of the test statistics. For the 6 simulated models we report the estimated variance and mean of the test statistics in Table 3.4. It also appears from Figure 3.8 and Table 3.4 that the internal studentization of the EL test statistics reduces the variance of $\mathcal{T}_T^{(n)}$ and thus the empirical level of the test is closer to the nominal level than the empirical level of the $T_T^{(n)}$ test.

On the other hand, the comparison of the parametric function $\tilde{\sigma}^2(\hat{\theta}, .)$ and $S_T^{(n)}$ is done only at k points. This means, that the smaller k the less function values are used for the test decision. One way to solve this trade off, is to use overlapping intervals for the calculation of the smoother. But in this approach we lose the asymptotic independence of $T_T^{(n)}(x_l)$ and thus $T_T^{(n)}$ is not asymptotically χ^2-distributed. A similar argument holds for $\mathcal{T}_T^{(n)}$.

One possible solution to solve the problem of small sample sizes and to make the test more reliable in such situations is the use of a bootstrap approximation of

	k	Vasicek mean	Vasicek Var	Square Root mean	Square Root Var	CKLS mean	CKLS Var
			EL test statistic $\mathcal{T}_T^{(n)}$				
LMR	3	2.42	5.32	2.31	4.89	2.37	5.38
	7	6.87	16.84	6.75	16.25	7.08	19.06
	11	11.60	35.04	11.74	31.12	12.18	34.40
AG	3	2.43	5.49	2.34	5.11	2.25	4.88
	7	6.89	16.51	6.91	16.66	6.85	17.18
	11	11.81	34.30	12.05	37.71	11.93	34.31
			test statistic $T_T^{(n)}$				
LMR	3	2.82	7.54	2.61	6.75	2.82	7.44
	11	12.70	41.56	12.26	37.10	12.42	37.98
AG	3	2.83	7.45	2.63	6.72	2.66	6.71
	11	12.67	37.16	12.43	39.56	12.51	41.07

Table 3.4: Mean and variance of the two test statistics estimated from a sample of 1000 paths with length $nT = 1000$.

the asymptotic distribution. Using the bootstrap methodology we could construct the test statistics from small overlapping intervals $(x_l - h_n, x_l + h_n)$. One possible bootstrap approach that could be applied in this situation is the local bootstrap method introduced by Paparoditis and Politis (2000). It captures the dependency structure of the data. However, the application of bootstrap is behind the scope of this work.

To investigate the power of the EL test we simulate 1000 paths of the Vasicek model with linear drift ($nT = 1000$) and test the three diffusion coefficient models given in Table 3.1 with this data. The result is shown in Figure 3.10. It appears from that figure, that the power of the test for the square root model is smaller than that of the CKLS model. However, the difference of the empirical rejection level between the (true) Vasicek model and the square root model is significant. This means that the proposed test is able to distinguish these two models. An inclusion of the Ahn-Gao drift does not change the result in principle. Since the test based on $T_T^{(n)}$ does not hold its nominal level, we will not use it in our empirical analysis and we do not investigate it's power.

3.6 Empirical Analysis

We apply the estimation procedures and the Empirical Likelihood test about the diffusion coefficient proposed in this Chapter to different data including interest rates (7-day Eurodollar rate), asset prices and stock market index processes. We obtain all data sets from Thomson Financial Datastream. All data are sampled daily.

3.6.1 The analysis of the 7-day Eurodollar rate

We start with the analysis of the 7-day Eurodollar rate. The data we use are daily observations of the spot rate from 1975/01/02 to 2002/02/18. These are 7078 observations. The evolution of the process is shown in Figure 3.11.

To provide some intuition about the behavior of the Eurodollar spot rate, we start with some descriptive statistics. Figure 3.12 shows the autocorrelation function of the spot rate. It appears, that even after 500 trading days, the autocorrelation is still about 0.5 and thus the mixing coefficient of this process seems to be very small. As mentioned in the last chapter, the statistical methodology for the drift function relies on the ergodic theorem. Heuristically speaking the ergodic theorem states, that observations of a mixing process can be treated like independent observations, if the dependency between these observations tends to zero as the time between them gets larger.

In a next step we estimate the drift and diffusion coefficient. As mentioned above, the estimation of the drift might be incorrect, due to the week mixing of the interest rate. However, to provide a complete picture of the process we show in Figure 3.14 parametric and nonparametric estimates for both, the drift and the diffusion coefficient. To estimate the diffusion coefficient we apply the estimator in (3.60). For the drift coefficient the Nadaraya-Watson estimator (3.23) as well as the estimator that was introduced by Jiang and Knight (1997), see (3.68), are applied.

To test parametric hypothesis about the Eurodollar spot rate, we test models about its diffusion coefficient only. The reason is again the large autocorrelation that we observe in the data. All models in Table 3.1 are tested. The parameters estimated from $n = 250$ trading days per year are given in Table 3.5 along with the values of the EL test statistic. To remove the influences of economic crises or structural breaks in the series we also investigate two subsamples. The first subsample is the period from the beginning of 1983 to the end of the sample, 2002/02/18. We use this time interval to remove the impact of the debt crisis of the developing countries in the beginning of the eighties. The second subsample considered is the period from January 1988 up to the end of the sample excluding the October 1987 crash.

As it can be seen, the EL test rejects all models, independently of the chosen degrees of freedom of their asymptotic χ^2-distribution. This result indicates, that the deviations of the estimated parametric diffusion functions from the nonparametrically estimated one cannot be explained by random fluctuations. Since the drift is not used in the construction of the test, an inclusion of a certain drift function will not change the result. In particular we have seen in the simulation study, that the test is robust at least against quadratic drift functions. This means, despite the drift function of interest rate models is important for the valuation of continent claims, all spot rate models that use one of the tested diffusion coefficients can be rejected. This result coincides with the empirical findings by Hong and Li (2002).

	value of $\mathcal{T}_T^{(n)}$				$\hat{\theta}$
k	5	7	9	11	
0.05 critical values	11.070	14.067	16.919	19.675	
0.01 critical values	15.086	18.475	21.666	24.725	
1975/01/02 – 2002/02/18					
$\sigma(\theta, x) = \theta$	208.277	274.721	924.651	550.362	0.043
$\sigma(\theta, x) = \theta\sqrt{x}$	32.696	56.268	153.364	111.100	0.126
$\sigma(\theta, x) = \theta x^{1.5}$	71.004	260.461	192.945	195.044	1.779
1983/01/03 – 2002/02/18					
$\sigma(\theta, x) = \theta$	428.088	494.119	577.438	854.375	0.037
$\sigma(\theta, x) = \theta\sqrt{x}$	69.048	109.827	150.644	243.480	0.119
$\sigma(\theta, x) = \theta x^{1.5}$	133.680	95.752	329.994	271.199	1.887
1988/01/01 – 2002/02/18					
$\sigma(\theta, x) = \theta$	236.125	298.087	288.529	447.579	0.031
$\sigma(\theta, x) = \theta\sqrt{x}$	23.390	42.265	39.413	91.197	0.103
$\sigma(\theta, x) = \theta x^{1.5}$	25.349	40.783	98.820	125.468	1.849

Table 3.5: Values of the EL test statistic and estimated parameters for the 7-day Eurodollar rate.

3.6.2 German Stock Prices

The EL test is also applied to the German stock market index DAX and to the German stocks Allianz, Bayer, Deutsche Bank, RWE and VW. The data we use are daily observations of the assets from 01.07.1991 to 19.02.2002. These are 2778 observations. We apply the test not to the original data but to the log prices, $X(t) = \log P(t)$, where

$P(t)$ is the observed price of the asset at time t. The results of the EL test are given in Table 3.6.

		value of $\mathcal{T}_T^{(n)}$			$\hat{\theta}$
	k	3	7	11	
	0.05 critical values	7.815	14.067	19.675	
DAX	$\sigma(\theta, x) = \theta$	95.495	274.544	276.595	0.181
	$\sigma(\theta, x) = \theta\sqrt{x}$	78.983	332.413	367.068	0.073
	$\sigma(\theta, x) = \theta x^{1.5}$	33.779	239.898	253.275	0.009
Allianz	$\sigma(\theta, x) = \theta$	71.274	260.952	266.548	0.259
	$\sigma(\theta, x) = \theta\sqrt{x}$	59.163	335.127	336.712	0.126
	$\sigma(\theta, x) = \theta x^{1.5}$	21.158	210.860	220.641	0.024
Bayer	$\sigma(\theta, x) = \theta$	119.001	135.356	248.024	0.221
	$\sigma(\theta, x) = \theta\sqrt{x}$	145.471	130.443	255.342	0.138
	$\sigma(\theta, x) = \theta x^{1.5}$	18.203	55.027	117.207	0.041
Deutsche Bank	$\sigma(\theta, x) = \theta$	164.887	101.440	455.881	0.243
	$\sigma(\theta, x) = \theta\sqrt{x}$	232.280	105.342	569.002	0.145
	$\sigma(\theta, x) = \theta x^{1.5}$	122.952	55.802	421.737	0.036
RWE	$\sigma(\theta, x) = \theta$	120.130	172.589	289.700	0.229
	$\sigma(\theta, x) = \theta\sqrt{x}$	113.045	187.370	314.577	0.136
	$\sigma(\theta, x) = \theta x^{1.5}$	37.562	103.120	174.242	0.038
VW	$\sigma(\theta, x) = \theta$	43.655	199.624	196.352	0.290
	$\sigma(\theta, x) = \theta\sqrt{x}$	20.487	187.114	183.105	0.164
	$\sigma(\theta, x) = \theta x^{1.5}$	3.817	117.371	137.169	0.047

Table 3.6: Values of the EL test statistic and estimated parameters for the DAX and five German stocks.

As for the interest rate, all supposed models are rejected by the test, except the CKLS model is not rejected for the VW stock price process when $k = 3$.

The empirical results indicate that affine diffusion processes might not be appropriate to model financial time series, like interest rates or stock prices. A number of alternative models is proposed in the literature. Hobson and Rogers (1998) propose a complete model, i.e. without an additional source of randomness. They model price processes as the solution of a stochastic delay differential equation, where the diffusion and drift coefficients depend on the whole history of the process. Stochastic volatility models, where the diffusion coefficient depends on an additional non observable volatility process are another way to capture the dynamics observed in the market, Hofmann et al. (1992). As these models yield incomplete markets, derivative prices are not unique.

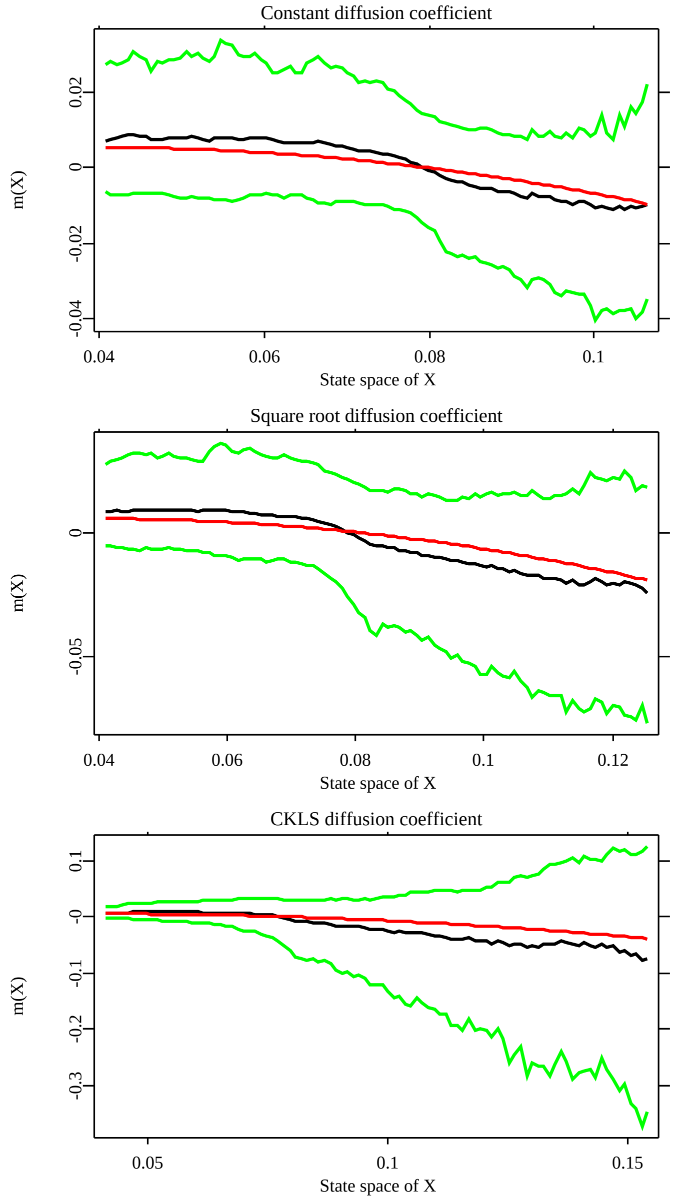


Figure 3.7: Nonparametric estimates of m together with the empirical 90% confidence band.

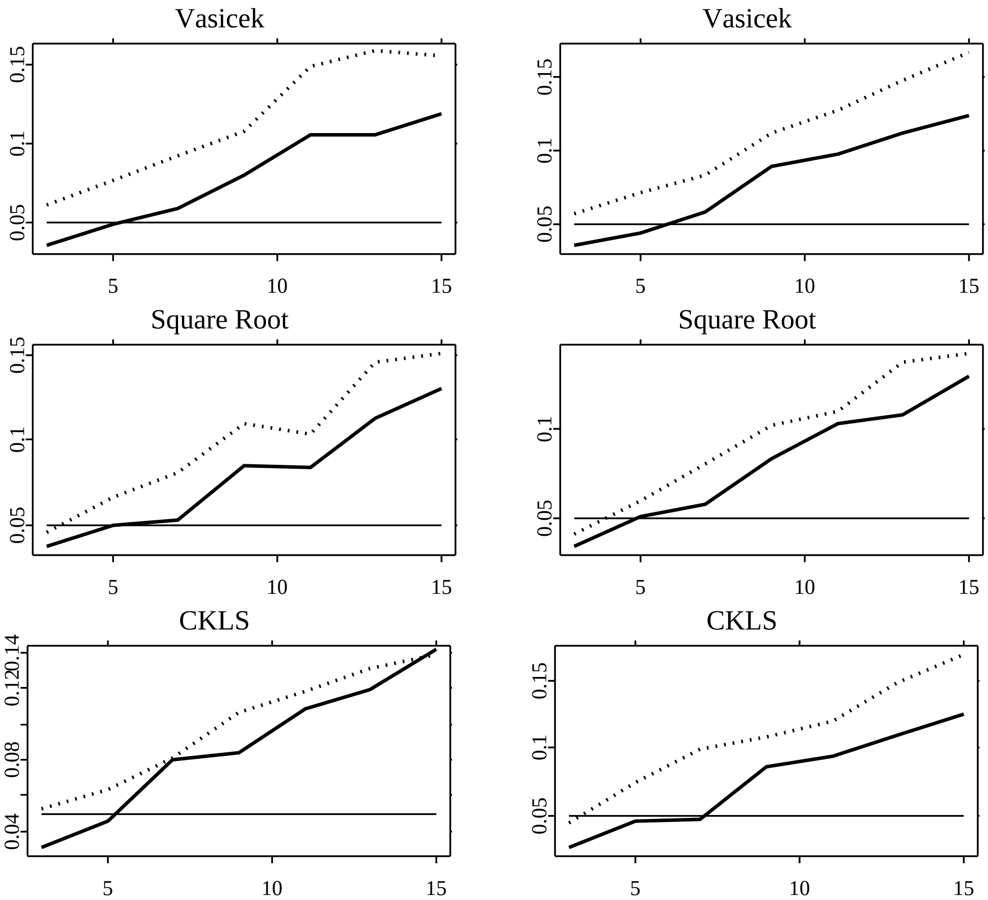


Figure 3.8: Empirical level of $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ for different models and path length $nT = 1000$. The left column displays the results for the linear mean reverting drift and the right column corresponds to the Ahn-Gao drift. On the vertical axis the empirical level is displayed and the horizontal axis shows the degrees of freedom (k). The solid line is the level of $\mathcal{T}_T^{(n)}$, the dotted line is the level of $T_T^{(n)}$ and the thin vertical line is the nominal level 0.05.

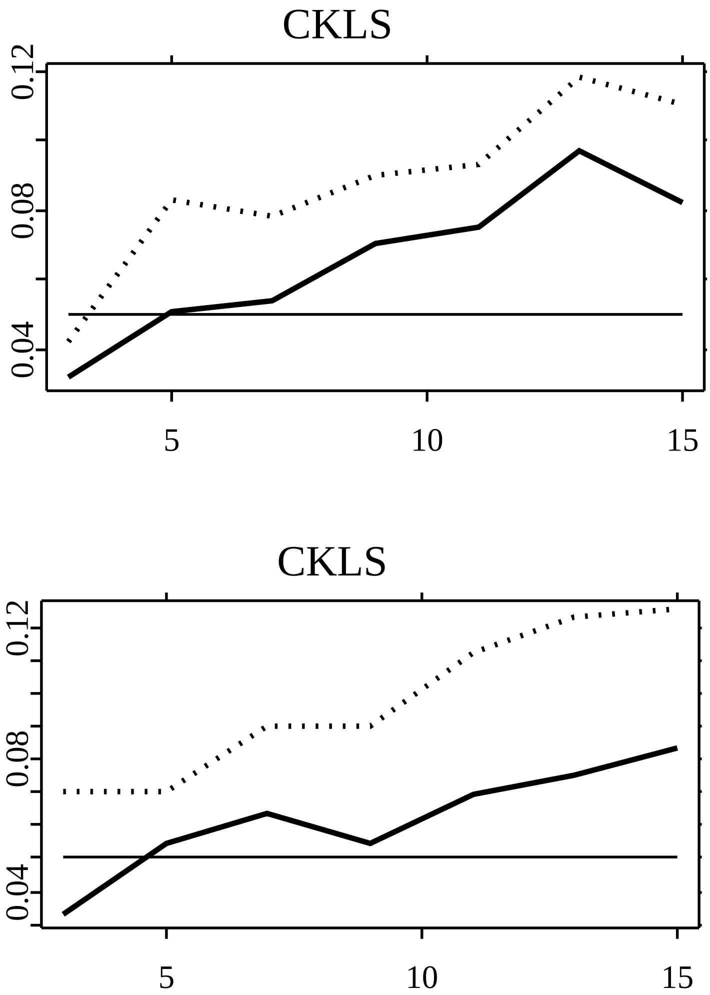


Figure 3.9: Empirical level of $T_T^{(n)}$ and $\mathcal{T}_T^{(n)}$ for the CKLS model with AG drift and path lengths $nT = 3000$ (upper plot) and $nT = 5000$ (lower plot). On the vertical axis the empirical level is displayed and the horizontal axis shows the degrees of freedom (k). The solid line is the level of $\mathcal{T}_T^{(n)}$, the dotted line is the level of $T_T^{(n)}$ and the thin vertical line is the nominal level of 0.05.

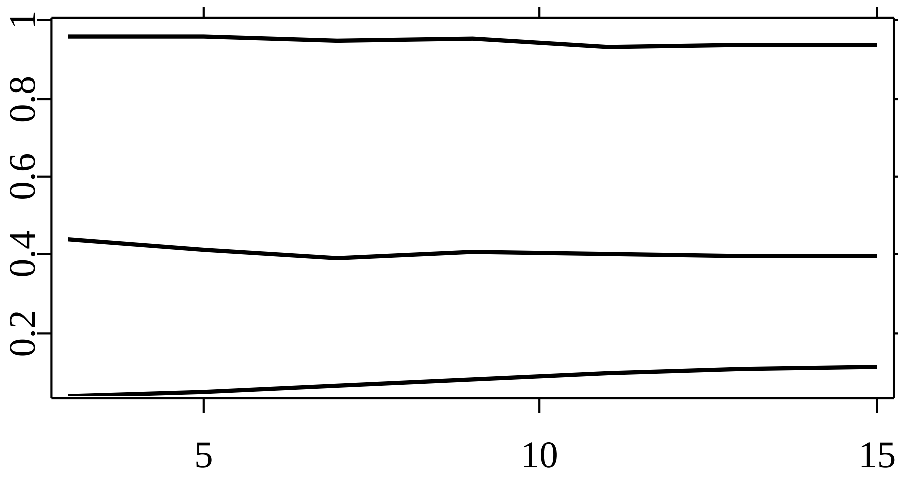


Figure 3.10: Empirical power of the EL test $\mathcal{T}_T^{(n)}$ the upper line corresponds to the CKLS model and the middle one to the square root model. The lower line represents the empirical level of the Vasicek model. The paths are simulated from the Vasicek model ($nT = 1000$, 1000 trajectories)

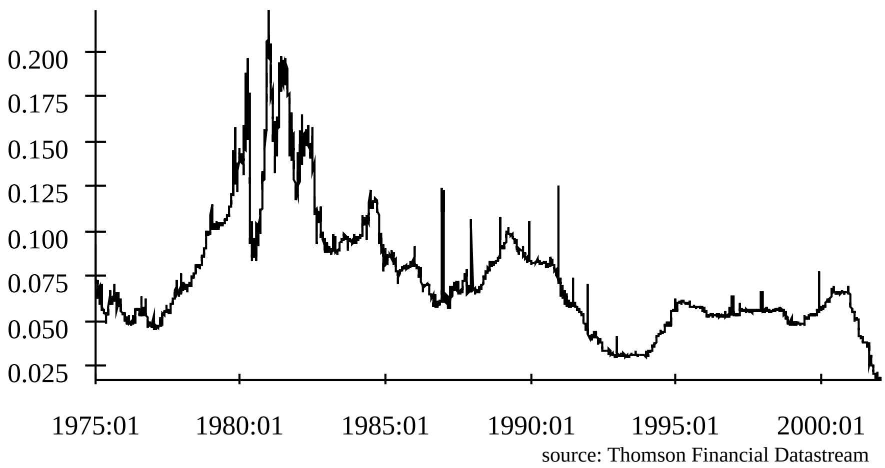


Figure 3.11: The 7-day Eurodollar rate.

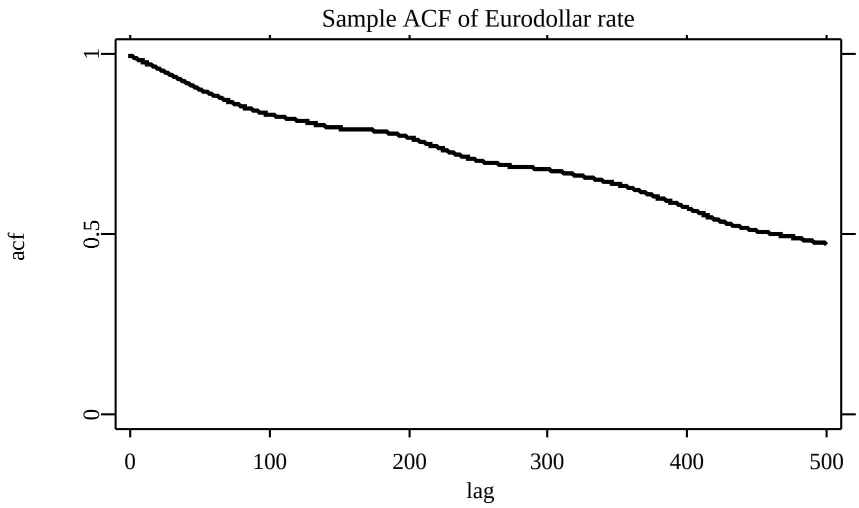


Figure 3.12: The autocorrelation function of the Eurodollar spot rate up to $\tau = 500$ days.

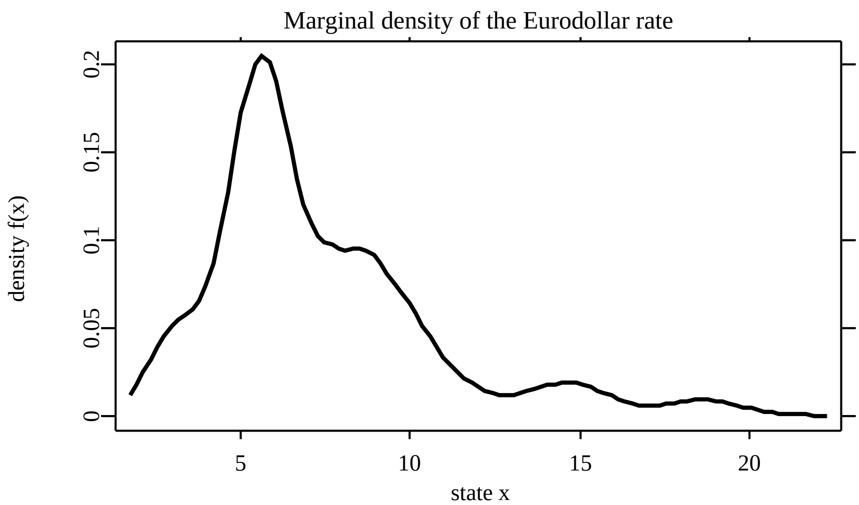


Figure 3.13: The marginal density of the Eurodollar spot rate.

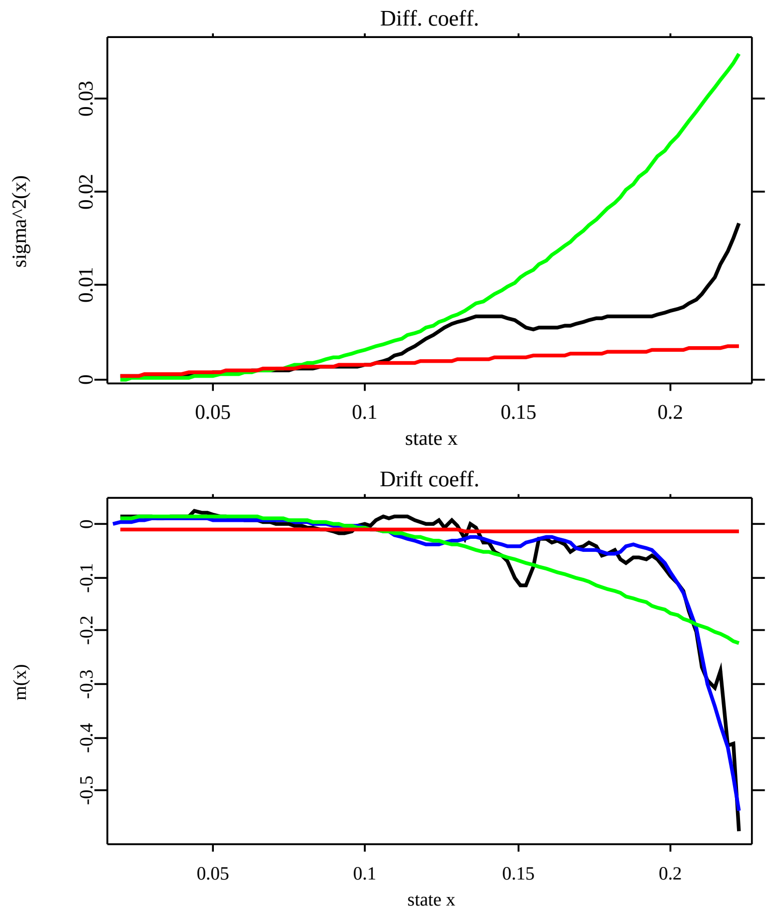


Figure 3.14: Parametric and nonparametric estimates of the diffusion (upper plot) and the drift function(lower plot). The green function is the parametric estimate of the Ahn-Gao model (Ahn-Gao drift, CKLS diffusion) and the red line corresponds to the parametric estimate of Square root process (LMR drift, CIR diffusion). The blue line in drift plot corresponds to the estimator in (3.23) and the black line is estimated with (3.68).

Chapter 4

A Semiparametric Model For A Stock Market Index

4.1 Introduction

We study here a case where a non-stationary diffusion process, an index, is observed. To be able to apply methods that rely on ergodicity we express the observed process as the product of an ergodic process and a smooth function of time. This smooth function is interpreted as average growth of the index. Due to the unknown impact of the average growth on the observed data, the ergodic part of our model is not directly observable. The proposed methodology combines recently developed non-parametric and parametric methods in order to estimate and probe the drift and diffusion coefficients of the ergodic process.

To illustrate our methodology we concentrate here on the empirical analysis of a particular stock market index, the S&P 500. The statistical analysis of stock prices, exchange rates etc. is similar but not in the focus of this chapter. We concentrate here on the case where an index is modeled by a scalar diffusion process.

The framework of Platen (2000) fully characterizes a financial market by the specification of the different denominations of the, so called, best benchmark portfolio. The stock index and the index benchmarked stock prices can be interpreted as denominations of the best benchmark portfolio. As a consequence, exchange prices are ratios of corresponding denominations of the best benchmark portfolio. Furthermore, this portfolio represents the optimal growth portfolio, see Karatzas and Shreve (1998). A well diversified market index, as the S&P 500, comes close to the optimal growth portfolio. For this reason, the inference for the index is also the first step in the statistical analysis of an exchange rate or stock price.

We assume that an appropriately normalized index process $X = \{X(t), t \geq 0\}$ can be interpreted as an ergodic process. Based on this assumption we focus on the inference of this normalized process X instead of the index $S = \{S(t), t \geq 0\}$ itself. This allows us to direct our attention towards the identification of the drift and diffusion coefficient functions of an ergodic diffusion. In Figure 4.1 we plot the S&P 500 index S with daily data from 1977 to 1997 together with an average index $\bar{S} = \{\bar{S}(t), t \geq 0\}$. Such an average index $\bar{S}$ can be obtained in different ways. For instance, it could be exogenously given by a function of economic and financial quantities, i.e. inflation rate, growth rate of the domestic product, interest rate, etc.. It could also be derived by a kernel smoothing procedure, with an appropriate bandwidth or filter length h. This is the choice which we will study in this chapter.

We construct the normalized index X by dividing the original index S by the above described average index $\bar{S}$, that is

$$X(t) = \frac{S(t)}{\bar{S}(t)} \tag{4.1}$$

for $t \geq 0$. The resulting normalized index X, derived via a kernel smoother, is shown in Figure 4.1. Its path resembles that of a stationary diffusion process. Note in the middle of our plot the sudden decline caused by the 1987 crash, which we do not remove from our sample.

We assume for the value of the index $S(t)$ at time t a representation of the form

$$S(t) = S(0)Z(t) \exp\left\{\int_0^t \eta(s)ds\right\} \tag{4.2}$$

for $t \geq 0$. Here $\eta(t)$ is interpreted as the deterministic, time dependent growth rate of the index at time t. Furthermore $Z(t)$ denotes the value of a positive ergodic diffusion process Z at time t, that means, Z solves the Itô stochastic differential equation (SDE)

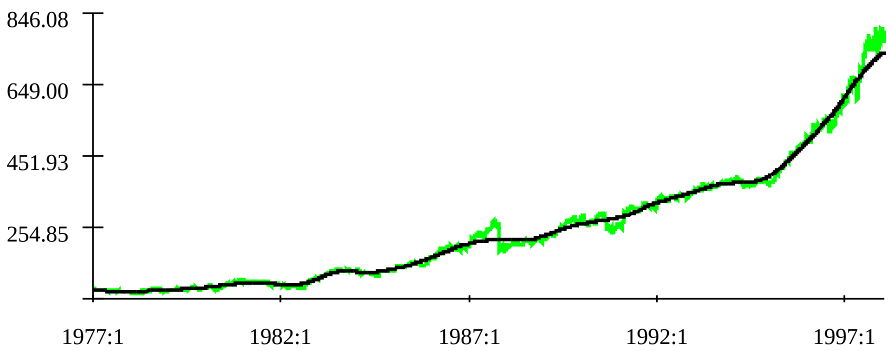
$$dZ(t) = m\{Z(t)\}dt + \sigma\{Z(t)\}dW(t) \tag{4.3}$$

for $t \geq 0$. Here $W = \{W(t), t \geq 0\}$ denotes a standard Wiener process and $m\{.\}$ and $\sigma\{.\}$ are the drift and diffusion coefficient functions. Due to the factor $S(0)$ we assume that Z is stable about 1, which models a mean reverting behavior. On the other hand, $Z(t)$ has to be positive for all $t \geq 0$.

To make our parametric model specific we may choose for Z a square root process, that is positive and stable about an equilibrium reference level. The square root process is also known as the Cox-Ingersoll-Ross (CIR) process, see Cox et al. (1985). The functional form (4.2) that models the index is a special case of the minimal market model (MMM) proposed in Platen (2000).

Another parametric model arises if we choose

$$Z(t) = \exp\{U(t)\} \tag{4.4}$$



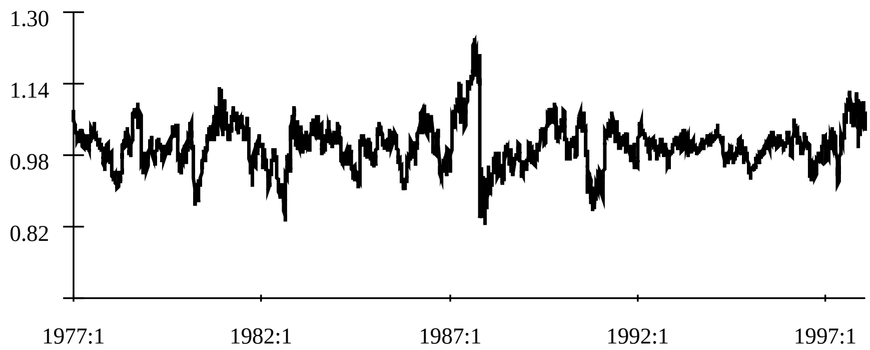


Figure 4.1: S&P 500 index S, average index $\bar{S}$ and normalized index X (lower plot).

with an Ornstein-Uhlenbeck process $U = \{U(t), t \geq 0\}$. This leads us to the exponential of an Ornstein-Uhlenbeck process as index model. Such a model has been used, for instance, in Föllmer and Schweizer (1993), Platen and Rebolledo (1996) and Fleming and Sheu (1999).

To compute the average index $\bar{S}$ in (4.1), we apply a kernel smoother to the logarithm of S and then calculate $\bar{S}$ as the exponential of this smoothed process. This removes the average deterministic growth in (4.2). For the analysis of the resulting normalized index we have to take into account that the residuals $\ln S - \ln \bar{S}$ are corrupted by the smoother. This will be shown in detail later on. It means, that the normalized index X is not a diffusion and in particular it does not equal the diffusion Z. For this reason we cannot directly apply estimation methods for discretely observed diffusions. From the statistical point of view we are faced with a nonparametric regression model with error terms that are not independent and identically distributed but are the discrete observations of a diffusion process. The analysis of these error terms and the clarification of their relationship to Z is a main task in this chapter.

We remark, that the index process S is itself a diffusion. When Z is specified according to (4.4) with an Ornstein-Uhlenbeck process U, Itô's formula yields the representation

$$dS(t) = \{\eta(t) - \beta\}S(t)dt + \gamma \exp\left\{\int_0^t \eta(s)ds\right\} dW(t) \tag{4.5}$$

for $t \geq 0$. The parameters β, γ and η cannot be easily separated in this representation. For this reason, we develop a statistical methodology for models that are based on the representation (4.2).

In Section 4.2.1 we introduce the parametric model for Z. The kernel smoothing and the computation of X is described in Section 4.2.2. The choice of the kernel and bandwidth and its influence on the average index is discussed in Section 4.2.3 together with the corresponding parameter estimation methods. In Section 4.2.4, a parametric model is tested versus a purely nonparametric alternative. This test is carried out by the bootstrap technique described in Section 3.3.3 and the Emprirical Likelihood methodology of Section 3.3.4. In Section 4.3 and 4.4 we apply the introduced methodology to S&P 500 data and also in a simulation study.

We remark that the proposed methodology applies directly to situations, where normalized data can be modeled by an ergodic diffusion process. Emphasis is here given to the case of an Ornstein-Uhlenbeck process, and results on the influence of the kernel smoother are included for this case.

4.2 Statistical Methodology for a Normalized Diffusion

4.2.1 Parametric Models

As discussed in the introduction, one can, in principle, use various parametric ergodic diffusion models. Let us mention two examples. Both of them have mean reverting drift coefficients. In the case where the squared diffusion coefficient has the form

$$\sigma^2(z) = \nu^2 z, \ \ z > 0, \tag{4.6}$$

with a positive constant ν, we obtain in (4.2) a square root process Z. Here we assume that Z satisfies the SDE

$$dZ(t) = \{\psi - \varphi\, Z(t)\}dt + \nu\sqrt{Z(t)}dW(t) \tag{4.7}$$

for $t \geq 0$ and with $\psi > \nu^2/2, \varphi > 0$. Note that a stationary and ergodic solution of (4.7) exists with the expected value $\mu_\infty = E[Z] = \psi/\varphi > 0$. Since the ratio $Z(t)/\mu_\infty$ is again a square root process and any constant term can be absorbed by $S(0)$ in (4.2), one can for simplicity assume that $\mu_\infty = E[Z] = 1$. This choice leads us to the SDE

$$dZ(t) = \varphi\{1 - Z(t)\}dt + \nu\sqrt{Z(t)}dW(t). \tag{4.8}$$

for $t \geq 0$.

We obtain a second example for an ergodic diffusion by defining Z as in (4.4), where U denotes the well-known Ornstein-Uhlenbeck process with

$$dU(t) = -\beta U(t)dt + \gamma dW(t). \tag{4.9}$$

for $t \geq 0$. Since U fluctuates about its reference level 0 and is ergodic, Z as given in (4.4) is an ergodic, positive diffusion process fluctuating about 1.

4.2.2 Kernel Smoothing

Denote by K_h a smoother with a kernel K and a bandwidth h. The smoothing of any process is denoted by a convolution operator $(*)$. As mentioned before, the normalized index $X(t)$ in (4.1) can be defined by the exponential of the difference of

$$L(t) = \ln\{Z(t)\} \tag{4.10}$$

and its smoother $(K_h * L)(t)$, that is:

$$X(t) = \exp\left\{ \ln S(t) - (K_h * \ln S)(t) \right\} = \exp\left\{ L(t) - (K_h * L)(t) \right\}. \tag{4.11}$$

Equation (4.11) holds if we neglect the difference between the accumulated deterministic growth rate $\int_0^t \eta(s)ds$ in (4.2) and its smoother, this means:

$$\int_0^t \eta(s)\, ds - \left(K_h * \int_0^{\cdot} \eta(s)\, ds \right)(t) \approx 0.$$

Here we arrive at a delicate point of our study. If we want to remove efficiently the deterministic growth rate in (4.2), then the value h should be chosen relatively small. Indeed, smaller values for h reduce the bias. On the other hand, the smaller the value of h is chosen, the more X is corrupted by $K_h * L$ in (4.11).

The smoother $K_h * L$ is differentiable for differentiable kernels K and thus of bounded variation. Due to the smoothing procedure $K_h * L$ involves also future information about L. Thus X is not a diffusion process. For this reason, we cannot treat $\ln X(t)$ in (4.11) as the logarithm of a square root process or as an Ornstein-Uhlenbeck process. A more detailed analysis of X has therefore to be performed. This is the objective of the next section. However note, in the case when $\bar{S}$ is obtained exogenously and not by a smoothing procedure, X might still be a diffusion.

4.2.3 Estimation of Parameters

In this section we assume that the only observations available are those of

$$\ln X(t) = L(t) - (K_h * L)(t) \tag{4.12}$$

in (4.11) and that L is the Ornstein-Uhlenbeck process U given in (4.9). The estimation problem that we now consider is that for the parameters β and γ in (4.9). In principle the value of γ can be restored from the quadratic variation of either $Z(t)$ or $L(t)$. For differentiable kernels K_h in (4.11), the process $(K_h * L)(t)$ is also differentiable. For this reason it holds that

$$\lim_{\Delta t \to 0} \sum_{i=1}^{n} \left\{ \ln X(i\Delta t) - \ln X(i\Delta t - \Delta t) \right\}^2 \stackrel{L^2}{=} \int_0^T d < L >_t \tag{4.13}$$

for $n = T/\Delta t$. Here $d < L >_t$ denotes the differential of the quadratic variation of the process L at time t. Empirical results confirm that the quadratic variation is not sensitive to the choice of h. For more details on that see Table 4.1. The following formula provides a stable estimate of γ^2 in the form

$$\begin{aligned} \hat{\gamma}^2 &= T^{-1} \sum_{i=1}^{n} \left\{ \ln X(i\Delta t) - \ln X(i\Delta t - \Delta t) \right\}^2 \\ &\approx T^{-1} \int_0^T d < L >_t . \end{aligned} \tag{4.14}$$

To estimate the speed of adjustment parameter β in (4.9) we could use the well-known form of the stationary variance of the Ornstein-Uhlenbeck process L. Along with (4.14) this would result in a first estimator of β with

$$\hat{\beta}_1 = \hat{\gamma}^2 / (2 \ Var[L]) . \tag{4.15}$$

Unfortunately, the substitution of $Var[L]$ by $Var[\ln X]$ makes $\hat{\beta}_1$ strongly dependent on h. Indeed, the variance

$$Var[\ln X] = Var[L - K_h * L] \tag{4.16}$$

increases as h grows, and only for very large values of h we can expect that $Var[\ln X] \approx Var[L]$.

It is not just the variance of the random process $\ln X$ that changes with h. Also its autocorrelation function depends on the bandwidth h. The correlation between the values of the process $\ln X$, distant by a constant time length $\tau > 0$, diminishes as h decreases. For this reason we propose a selection method for h based on the simultaneous estimation of β from the variance and from the autocorrelation function of the process $\ln X$. The idea is simple, if for each value of h there are two different estimates of the same parameter β, then the best choice of h is considered to be that, which brings these estimates as close as possible to each other.

The autocorrelation function $\rho^{(L)}(\tau)$ of the Ornstein-Uhlenbeck process L equals

$$\rho^{(L)}(\tau) = e^{-\beta\tau} \tag{4.17}$$

for $\tau > 0$. Thus, β represents the absolute value of the slope of this function at zero. Hence another estimate of β from the observations of L would be

$$\hat{\beta}_2 = \Big|\frac{\partial^+}{\partial\tau}\rho^{(L)}(\tau)\Big|\Big|_{\tau=0}, \tag{4.18}$$

where

$$\frac{\partial^+}{\partial\tau}\rho(\tau) = \lim_{s\to 0,\, s>0} \frac{\rho(\tau+s)-\rho(\tau)}{s}$$

for $\tau \geq 0$ denotes the right hand derivative of ρ with respect to τ.

Unfortunately, the estimator in (4.18) is not feasible since L is not observed. In Appendix A.3 we show for the process $\ln X$ that its stationary variance is asymptotically

$$Var[\ln X] = \frac{\gamma^2}{2\beta}\left(1 - \frac{c_K}{\beta h} + \mathcal{O}(h^{-2})\right) \text{ as } h \to \infty, \tag{4.19}$$

where the constant c_K depends on the kernel K. Furthermore, we prove in Appendix A.3 for the autocorrelation function $\rho_h^{(\ln X)}(\tau)$ of $\ln X$ the asymptotics

$$\begin{aligned}\rho_h^{(\ln X)}(\tau) &= \text{Corr}\left[\ln X(\tau); \ln X(0)\right] \\ &= \frac{e^{-\beta\tau} - \frac{c_K}{\beta h} + \mathcal{O}(\tau h^{-2})}{1 - \frac{c_K}{\beta h} + \mathcal{O}(h^{-2})}, \qquad \tau \geq 0\end{aligned} \tag{4.20}$$

as $h \to \infty$ with the same constant c_K as in (4.19). In Appendix A.3 this constant is calculated for the rectangle and the Epanechnikov kernels.

It follows from equation (4.19) that the first-order approximation of the stationary variance of $\ln X$ is

$$Var[\ln X] \approx \left\{1 - \frac{c_K}{\beta h}\right\} Var[L]. \tag{4.21}$$

By (4.20), the slope of the autocorrelation of $\ln X$ at zero is asymptotically

$$\Big|\frac{\partial^+}{\partial\tau}\rho_h^{(\ln X)}(\tau)\Big|\Big|_{\tau=0} \approx \frac{\beta}{1 - \frac{c_K}{\beta h}} \tag{4.22}$$

as $h \to \infty$. Thus this slope is steeper than that of ρ^L at $\tau = 0$.

The immediate consequence of (4.19) and (4.20) is that the formulas (4.15) and (4.18) for $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively, have to be modified if the process $\ln X$ rather than L is observed. In Appendix A.3 we show that the correct modification is provided by the expressions

$$\hat{\beta}_1(h) = \frac{\hat{\gamma}^2}{2\,Var[\ln X]} - \frac{c_K}{h} \tag{4.23}$$

and

$$\hat{\beta}_2(h) = \left|\frac{\partial^+}{\partial\tau}\rho_h^{(\ln X)}(\tau)\right|\Big|_{\tau=0} - \frac{c_K}{h}, \tag{4.24}$$

respectively. Finally, our method for the selection of h is based on the following balance equation

$$\hat{\beta}_1(h) = \hat{\beta}_2(h) \tag{4.25}$$

which equals both estimates.

After h is chosen, we need to restore the process L, which is needed in the remaining nonparametric and parametric analysis. From (4.11), proceeding formally, one arrives at the following iterative formula:

$$\begin{aligned} L = \ln X + K_h * L &= \ln X + K_h * (\ln X + K_h * L) \\ &= \dots \\ &= \ln X + K_h * \ln X + K_h * K_h * \ln X + \dots . \end{aligned} \tag{4.26}$$

The justification for the restoration formula (4.26) comes from the fact that if one neglects the boundary effects, the smoothing operator K_h is a contracting operator in L_2, as shown in Appendix A.3. In the practical application of (4.26), we rely on the fact that the smoother of the original process L is close to the smoother of $L - K_h * L$. In practice, only one or two convolutions are meaningful. After the restoration process is completed, the parameter β can be estimated directly from L by (4.15).

We were able to establish in this chapter the above correction terms for the Ornstein-Uhlenbeck process. One could, in principle, estimate parameters also under the assumption that X itself is a square root process or another ergodic diffusion. However, if the average index $\bar{S}$ is calculated via a smoothing procedure, a similar bandwidth selection method has to be developed. At that stage this is left for future research.

4.2.4 Testing the Parametric Model

We can now apply the methods introduced in Chapter 3 to test the parametric form of the normalized index process. The first step is derivation of the null hypotheses about the drift and diffusion coefficient of Z. To derive the null hypotheses in the case when Z is the exponential of an Ornstein Uhlenbeck process, we apply Itô's formula to $Z(t) = \exp\{U(t)\}$. Here U satisfies (4.9) and one obtains

$$\begin{aligned} dZ(t) &= d(\exp\{U(t)\}) \\ &= Z(t)\left\{-\beta \ln Z(t) + \frac{1}{2}\gamma^2\right\} dt + \gamma Z(t) dW(t) \end{aligned} \quad (4.27)$$

for $t \geq 0$. The null hypotheses of the tests are therefore

$$H_0(m) : m(z) = z\left\{-\beta \ln z + \frac{1}{2}\gamma^2\right\}$$

and

$$H_0(\sigma^2) : \sigma^2(z) = \gamma^2 z^2,$$

while the alternative is nonparametric.

4.3 Empirical Analysis of the S&P 500

We apply the methods introduced in Section 4.2 to daily observations of the S&P 500 index from 31.12.1976 to 31.12.1997 (5479 observations). The data are obtained from Thomson Financial Datastream.

For the kernel smoothing of S we choose the Epanechnikov kernel. The constant c_K that appears in the correction terms in (4.23) and (4.24) are known for this particular kernel, see Appendix A.3.

As already mentioned in Section 4.2.3 the estimates for the parameter γ calculated from formula (4.14) are small relative to 1 and do not change significantly with h. Table 4.1 shows the estimated values for different values of h. The variance of the process X is also shown in that table. The small variance and the fact that X is stable about 1 justifies to concentrate on the case of a geometric Ornstein-Uhlenbeck process defined by (4.4) and (4.9).

The next step in our analysis is the choice of h. Due to the long range of observations we apply a flexible bandwidth to the data. This flexible bandwidth was calculated by splitting the data in overlapping subintervals of different lengths and calculating an optimal fixed bandwidth for every subinterval. The bandwidth is chosen to be optimal with respect to the balance equation (4.25). To get a continuous optimal bandwidth function $h_{opt}(t)$ we interpolate the resulting values. The function

h	200	250	300	350	400
$Var(X)$	0.0018303	0.0023465	0.0029246	0.0035622	0.0042183
$\hat{\gamma}$	0.0090593	0.0090703	0.0090849	0.0090991	0.0091103

Table 4.1: Estimated values for γ and the estimated variance of X for different bandwidths h.

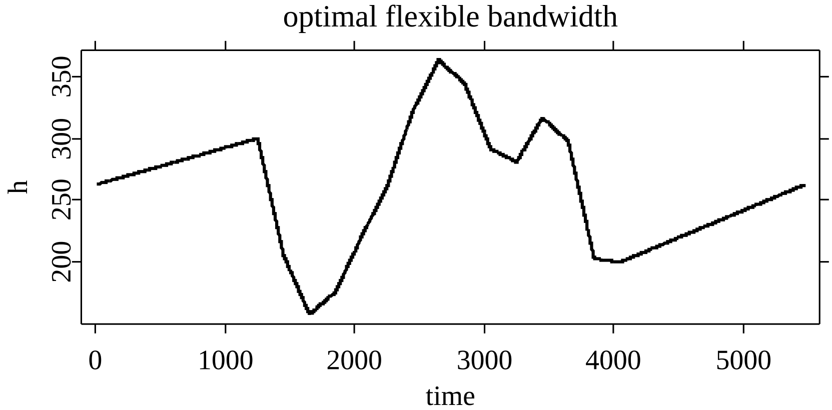


Figure 4.2: The optimal flexible bandwidth $h_{opt}(t)$.

$t \mapsto h_{opt}(t)$ is shown in Figure 4.2. The final values for β are $\hat{\beta}_1(h_{opt}) = 0.010352$ and $\hat{\beta}_2(h_{opt}) = 0.0089721$ and the ratio is $\hat{\beta}_1(h_{opt})/\hat{\beta}_1(h_{opt}) = 1.1538$. For fixed bandwidths in the range of h_{opt} this ratio is given in Table 4.2. All these ratios are larger than those for h_{opt} which justifies the use of the flexible bandwidth. The estimated value for γ is $\hat{\gamma}(h_{opt}) = 0.0091033$.

Now we are in the position to restore the path of the process Z and to estimate the parameters. We get the following estimates from the restored path

$$\hat{\beta}_1 = 0.01003, \qquad \hat{\beta}_2 = 0.0093294,$$
$$\hat{\beta}_1/\hat{\beta}_2 = 1.0751, \qquad \hat{\gamma} = 0.0092454.$$

To finish the empirical analysis we apply the test procedure described in Section 3.5 for the hypotheses in Section 4.2.4. Figure 4.3 shows the nonparametric estimates of the drift and squared diffusion coefficient of the restored process Z together with the 90% confidence bands. The almost straight lines show the parametric estimates with respect to the estimated values of the restored process Z. The vertical lines enclose the interval where 99% of the observed data reside.

h	200	225	250	275	300
$\hat{\beta}_1(h)$	0.017913	0.01569	0.013826	0.012197	0.010776
$\hat{\beta}_2(h)$	0.01328	0.012098	0.011092	0.0098626	0.0086047
$\hat{\beta}_1(h)/\hat{\beta}_2(h)$	1.3489	1.2969	1.2465	1.2367	1.2523

Table 4.2: Estimated values for β for different fixed bandwidths h.

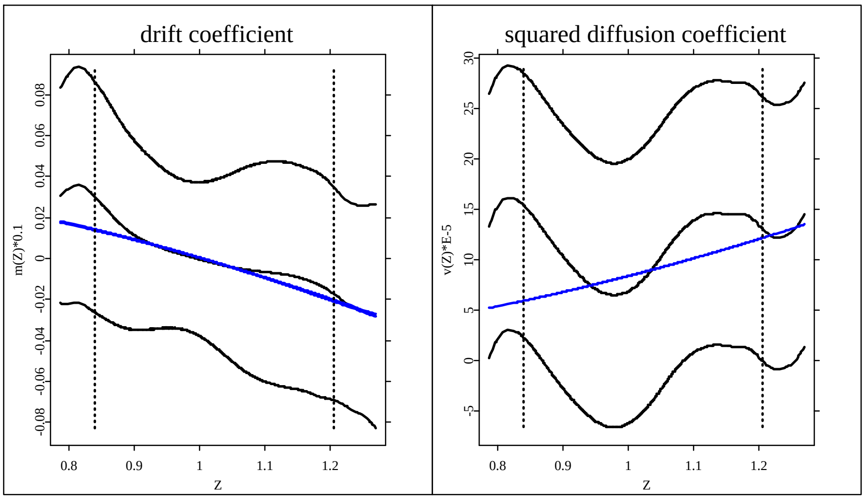


Figure 4.3: Nonparametric and parametric estimates of the drift $m(.)$ and squared diffusion coefficient $\sigma^2(.)$ with 90% confidence bands.

Both parametric functions are surely inside the confidence bands. Thus the null hypothesis of the geometric Ornstein–Uhlenbeck process cannot be rejected.

4.4 Simulation Study

We perform now a simulation study by applying the estimation methods introduced in Section 4.2.3 to simulated trajectories of the Ornstein-Uhlenbeck process U. The drift and diffusion parameters β and γ in (4.9) are estimated directly from the observations of U as well as from the residual of a kernel smoothing procedure.

It is well known that the transition probability of an Ornstein-Uhlenbeck process is normal with conditional mean

$$E[U_{t+\Delta}|U_t = u] = ue^{-\beta\Delta}$$

and conditional variance

$$Var(U_{t+\Delta}|U_t = u) = \frac{\gamma^2}{-2\beta}\left(e^{-2\beta\Delta} - 1\right).$$

Using this Gaussian transition probability we simulate 100 paths of the process U with time step size $\Delta = 1$. The true parameters are set to $\beta = 0.01$ and $\gamma = 0.01$, which correspond approximately to the empirical estimates for the S&P 500 index in Section 4.3.

For the analysis of the directly observed process U we apply three estimators for the speed of adjustment parameter β. Besides $\hat{\beta}_1$ and $\hat{\beta}_2$ introduced in (4.15) and (4.18), we use also the estimator

$$\hat{\beta}_3 = -\frac{1}{\Delta}\ln\frac{\sum_{i=1}^{n} U_{i-1}U_i}{\sum_{i=1}^{n} U_{i-1}^2}, \tag{4.28}$$

which is based on martingal estimating functions and was proposed in Bibby and Sørensen (1995). It is easy to see, that $\hat{\beta}_3$ is related to the autocorrelation function of U. For details about this estimator and the theory of martingal estimating functions we refer to Bibby and Sørensen (1995) and the references therein. The diffusion coefficient is estimated via the slope of the quadratic variation, similarly as in (4.14).

The first row of Table 4.3 shows the means of the corresponding estimated values. In the second row the variance of the estimates are shown. We emphasize that the results are based on a directly observed simulated diffusion.

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	γ	$\hat{\beta}_1/\hat{\beta}_2$	opt h
0.01070	0.01018	0.01028	0.00996		
4.589e-06	3.695e-06	4.541e-06	1.059e-08		
0.00964	0.00967		0.00996	0.99435	295.0
7.161e-06	6.258e-06		1.067e-08	0.00248	

Table 4.3: Estimated parameters.

Furthermore, we simulate the logarithm of the index in (4.2) $\ln S$ as the sum of a linear function and U. In a second step we calculate $\ln X$ as in (4.11) with the

Epanechnikov kernel, see Appendix A.3. We then estimate from the simulated data the parameters β and γ of U by the methods in (4.23), (4.24) and (4.14). This gives us an idea about the fixed sample behavior of these estimation methods when the residuals of a kernel smoothing procedure are observed instead of those of an Ornstein-Uhlenbeck process itself.

The estimated values calculated from the simulated trajectories of $\ln X$ are shown in the third and fourth row of Tables 4.3. The results clearly demonstrate that the correction terms in (4.23) and (4.24) are necessary to obtain reasonable estimates. In the situation considered here, the correction terms equal each other and have approximately the value $c_k/h \approx 0.0061$, see Appendix A.3. Since the correction terms for $\hat{\beta}_3$ are not considered, we do not report them in Table 4.3.

The table also shows the mean and the variance of the ratio $\hat{\beta}_1/\hat{\beta}_2$ used to select the bandwidth h, see (4.25). The mean of the selected bandwidth h, which brings this ratio as close as possible to one, is given in the last column.

The second part of the simulation study treats the bootstrap procedure. We apply the bootstrap methodology as introduced in Section 3.3.3 to a simulated path of an Ornstein-Uhlenbeck process U following the dynamics in (4.9) with parameters $\beta = 0.01$, $\gamma = 0.01$ and $\Delta = 1$. The values of the parameters are reasonable with respect to the empirical results for the S&P 500. The number of observations is 5000 and the number of the bootstrapped series for the confidence bands is 160. The two plots in Figure 4.4 show the nonparametric estimators for the drift and squared diffusion coefficient together with their 90% confidence bands constructed by the bootstrap procedure. The plots also show the true parametric functions for the drift and diffusion coefficient. The dotted vertical lines are the empirical 0.005 and 0.995 quantiles of the stationary distribution of $\exp(U)$. If we only consider the range between these quantities, i.e. the range where 99% of the data reside, then both of the parametric functions remain inside the confidence bands. This means, the null hypotheses $H_0(m)$ and $H_0(\sigma^2)$ as in Section 4.2.4 cannot be rejected for data in this range.

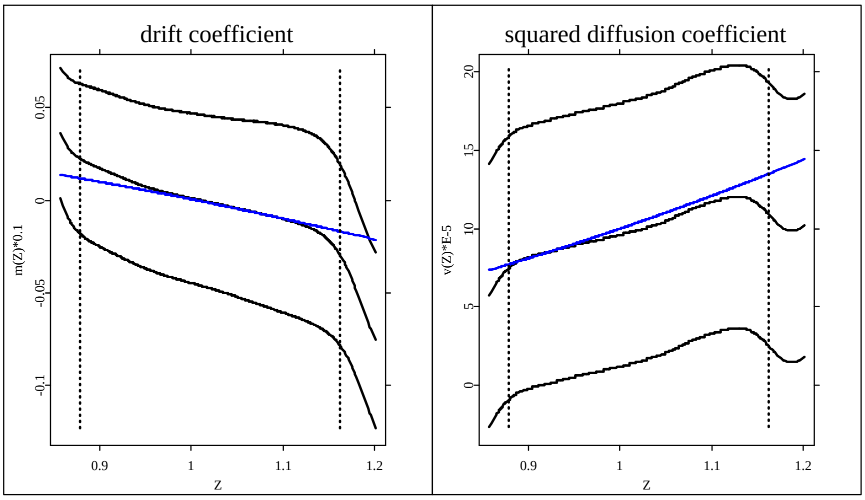


Figure 4.4: Nonparametric estimates for the drift and squared diffusion coefficients of a simulated geometric Ornstein-Uhlenbeck process, confidence bands and true functions.

Bibliography

Adler, R.J. (1981). *The Geometry of Random Fields.* New York: John Wiley & Sons.

Ahn, Dong-Hyun and Gao, Bin (1999). A Parametric Nonlinear Model of Term Structure Dynamics. *The Review of Financial Studies*, 12(4):721–762.

Aït-Sahalia, Yacine (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64(3):527–560.

Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika*, 85:535–547.

Barnsley, Michael (1988). *Fractals everywhere.* Boston, MA etc.: Academic Press, Inc.

Beran, J. (1994). *Statistics for Long-Memory Processes.* London: Chapman and Hall.

Berry, M.V. and Hannay, J.H. (1978). Topography of random surfaces. *Nature*, 273:573.

Bibby, Bo Martin and Sørensen, Michael (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1/2):17 – 40.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley & Sons.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.*, volume 110 of *Lecture Notes in Statistics.* Heidelberg: Springer-Verlag.

Calvet, L. and Fisher, A. and Mandelbrot, B.B. (1997). A Multifractal Model of Asset Returns. Cowles Foundation Discussion Paper # 1164.

Carter, P.H. and Cawley, R. and Mauldin, R.D. (1988). Mathematics of Dimension Measurements of Graphs of Functions. In Weitz, D., Sander, L., and Mandelbrot, B., editors, *Proc. Symb. Fractal Aspects of Materials, Disordered Systems.*, pages 183–186. Pittsburgh, PA: Materials Research Society.

Chan, K.C. and Karolyi, G.A. and Longstaff, F.A. and Sanders, A.B. (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance*, 47:1209–1227.

Chen, S. X. and Härdle, W. and Kleinow, T. (2001). An Empirical Likelihood Goodness-of-Fit Test for Time Series. Discussion paper 1/2001, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Chen, S. X. and Härdle, W. and Kleinow, T. (2002). An Empirical Likelihood Goodness-of-Fit Test for Diffusions. In Härdle, W., Kleinow, T., and Stahl, G., editors, *Applied Quantitative Finance.* New York: Springer Verlag.

Constantine, A.G. and Hall, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. Roy. Statist. Soc. Ser. B*, 56:97–113.

Cox, J.C. and Ingersoll, J.E. and Ross, S.A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 2(53):385–407.

Davies, R.B. and Harte, D.S. (1987). Tests for Hurst effect. *Biometrika*, 74:95–101.

Davies, S. and Hall, P. (1998). Fractal analysis of surface roughness using spatial data. *J. Roy. Statist. Soc. Ser. B*, to appear.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press.

Delbaen, Freddy and Schachermayer, Walter (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520.

Diebold, F.X. and Hunther, T. and Tay, A. (1998). Evaluating Density Forecasts with Application to Financial Risk Management. *International Economic Review*, 39:863–883.

Duffie, Darrell and Kan, Rui (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406.

Durrett, Richard (1991). *Stochastic Calculus.* New York: Springer-Verlag.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* London: Chapman and Hall.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications - Theory and Methodologies.* New York: Chapman and Hall.

Fleming, Wendell H. and Sheu, Shuenn-Jyi (1999). Optimal long term growth rate of expected utility of wealth. *Ann. Appl. Probab.*, 9(3):871–903.

Florens-Zmirou, Danielle (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, 30(4):790–804.

Föllmer, Hans and Schweizer, Martin (1993). A microeconomic approach to diffusion models for stock prices. *Mathematical Finance*, 3(1):1–23.

Franke, J. and Kreiss, J-P. and Mammen, E. and Neumann, M.H. (1998). Properties of The Nonparametric autoregressive bootstrap. Discussion paper, Sonderforschungsbereich 373, No. 54/98, Humboldt-Universität zu Berlin.

Genon-Catalot, V. and Jeantheau, T. and Larédo, C. (2000). Stochastic Volatility Models as Hidden Markov Models and Statistical Applications. *Bernoulli*, 6(6).

Hall, P. (1985). Resampling a Coverage Pattern. *Stoch. Proc. Appl.*, 20:231–246.

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

Hall, P. and Matthews, D. and Platen, E. (1996). Algorithms for analyzing nonstationary time series with fractal noise. *J. Computat. Graph. Statist.*, 5:351–364.

Hall, P. and Roy, R. (1994). On the relationship between fractal dimension and fractal index for stationary stochastic processes. *nn. Appl. Probab.*, 4:241–253.

Hall, Peter and Wood, Andrew (1993). On the performance of box-counting estimators of fractal dimension. *Biometrika*, 80(1):246–252.

Härdle, W (1990). *Applied Nonparametric Regression.* Number 19 in Econometric Society Monographs. Cambridge University Press.

Härdle, W. and Kleinow, T. and Korostelev, A. and Logeay, C. and Platen, E. (2001). Semiparametric Diffusion Estimation and Application to a Stock Market Index. Discussion Paper 24/2001, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Härdle, W. and Klinke, S. and Müller, M. (1999). *XploRe -The Statistical Computing Environment.* New York: Springer-Verlag.

Härdle, W. and Mammen, E. (1993). Comparing Nonparametric versus Parametric Regression Fits. *The Annals of Statistics*, 21:1926–1947.

Härdle, W. and Müller, M. and Sperlich, St. and Werwatz, A. (2000). Non- and Semiparametric Modelling. XploRe e-book, www.xplore-stat.de.

Härdle, W. and Tsybakov, A.B. (1997). Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression. *Journal of Econometrics*, 81:223–242.

Hobson, David G. and Rogers, L.C.G. (1998). Complete models with stochastic volatility. *Mathematical Finance*, 8(1):27–48.

Hoffmann, Marc (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, 79(1):135–163.

Hofmann, Norbert and Platen, Eckhard and Schweizer, Martin (1992). Option pricing under incompleteness and stochastic volatility. *Mathematical Finance*, 2(3):153–187.

Hong, Yongmiao and Li, Haitao (2002). Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates. Working paper, Cornell University.

Hunt, Fern (1990). Error analysis and convergence of capacity dimension algorithms. *SIAM J. Appl. Math.*, 50(1):307–321.

Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Engineers*, 116:770–799.

Jiang, G. and Knight, J. (1997). A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model. *Econometric Theory*, 13(5):615–645.

Karatzas, Ioannis and Shreve, Steven E. (1991). *Brownian Motion and Stochastic Calculus.* New York: Springer-Verlag.

Karatzas, Ioannis and Shreve, Steven E. (1998). *Methods of Mathematical Finance.*, volume 39 of *Applications of Mathematics, Stochastic Modelling and Applied Probability.* New York: Springer-Verlag.

Karlin, Samuel and Taylor, Howard M. (1981). *A second course in stochastic processes.* New York etc.: Academic Press, A Subsidiary of Harcourt Brace Jovanovich, Publishers. XVIII, 542 p. $ 35.00 .

Kent, John T. and Wood, Andrew T.A. (1997). Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. R. Stat. Soc., Ser. B*, 59(3):679–699.

Kloeden, Peter E. and Platen, Eckhard (1999). *Numerical Solution of Stochastic Differential Equations.*, volume 23 of *Applications of Mathematics.* Berlin, Heidelberg: Springer-Verlag.

Kutoyants, Yu.A. (1998). Efficient density estimation for ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, 1(2):131–155.

Mandelbrot, B.B. and Passoja, D.E. and Paullay, A.J. (1984). Fractal character of surfaces of metals. *Nature*, 308:721–722.

Neumann, M.H. and Kreiss, J-P. (1998). Regression-type inference in nonparametric autoregression. *The Annals of Statistics*, 26:1570–1613.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18:90–120.

Owen, A. (1991). Empirical likelihood for linear model. *The Annals of Statistics*, 19:1725–1747.

Owen, Art B. (2001). *Empirical Likelihood.*, volume 92 of *Monographs on Statistics and Applied Probability.* Chapman & Hall/CRC.

Paparoditis, Efstathios and Politis, Dimitris N. (2000). The local bootstrap for kernel estimators under general dependence conditions. *Annals of the Institute of Statistical Mathematics*, 52(1):139–159.

Peters, E.E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics.* New York: John Wiley & Sons.

Platen, Eckhard (2000). Risk Premia and Financial Modelling Without Measure Transformation. University of Technology Sydney, School of Finance & and Economics and Department of Mathematical Sciences.

Platen, Eckhard and Rebolledo, Rolando (1996). Principles for modelling financial markets. *Journal of Applied Probability*, 33(3):601–613.

Pollak, Moshe and Siegmund, David (1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, 72:267–280.

Rogers, L.C.G. (1997). Arbitrage with fractional Brownian motion. *Mathematical Finance*, 7(1):95–105.

Rogers, L. C. G. and Williams, David (1987). *Diffusions, Markov Processes, and Martingals, Volume 2: Ito calculus.* New York: John Wiley & Sons Ltd.

Rosenblatt, Murray (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23:470–472.

Rosenblatt, M. (1961). Independence and dependence. In Neyman, J., editor, *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, pages 411–433. Berkeley: University of California Press.

Sayles, R.S. And Thomas, T.R. (1978). Surface topography as a nonstationary random process. *Nature*, 271:431–434.

Shao, J. And Tu, D. (1995). *The Jackknife and Bootstrap.* New York: Springer-Verlag.

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

Sullivan, F. and Hunt, F. (1988). How to estimate capacity dimension. *Nuclear Physics B (Proc. Suppl.)*, pages 125–128.

Taqqu, M.S. (1875). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete*, 31:287–302.

van der Vaart, Aad and Wellner, Jon A. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. New York: Springer-Verlag.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Number 60 in Monographs in Statistics and Applied Probability. Chapman & Hall.

Wood, A. T. A. and Chan, G. (1994). Simulation of stationary Gaussian process in $[0,1]^d$. *J. Comp. Graph. Stat.*, 3:409–432.

Appendix A

Appendix

A.1 One-dimensional Diffusions

To ensure the existence of a solution of (3.1) we make the following assumptions

(X1) For all $\theta \in \Theta$ we assume, that their exists an interval (l, r) with $l, r \in \mathbb{R} \cup \{-\infty; \infty\}, l < r$ such that $m(.), \sigma(.), m(\theta, .), \sigma(\theta, .) \in C^1((l, r), \mathbb{R})$ and $\exists K > 0$ such that:

$$\begin{aligned} m^2(\theta, v) + \sigma^2(\theta, v) &\leq K(1 + v^2) \\ m^2(v) + \sigma^2(v) &\leq K(1 + v^2) \end{aligned}$$

and

$$\forall v \in (l, r) : \ \sigma(\theta, v) > 0 \ \sigma(, v) > 0$$

From this assumptions follows, that the martingale problem MP(m,σ^2) has a unique solution and thus a unique solution of (3.1) exists, Karatzas and Shreve (1991) and Durrett (1991).

(X2) For the scale function $\varphi(x)$ of σ_t^2

$$\varphi(x) \stackrel{\text{def}}{=} \int_c^x exp\left\{-2 \int_c^v \frac{m(u)}{\sigma^2(u)} du\right\} dv, \qquad x \in \mathbb{R}$$

holds for l and r in (X1):

$$\varphi(l+) = -\infty, \qquad \varphi(r-) = \infty$$

It follows, that the solution of (3.1) is recurrent and for the exit time S of X form (l, r)

$$S \stackrel{\text{def}}{=} \inf\{t \geq 0 : \ X(t) \notin (l, r)\}$$

holds $P[S = \infty] = 1$, Theorem 5.5.29 (Feller's Test for Explosion) and Proposition 5.5.22 in Karatzas and Shreve (1991).

(X3) For l and r holds

$$M \stackrel{\text{def}}{=} \int_l^r \frac{1}{\sigma^2(x)\varphi'(x)} dx < \infty$$

This conditions ensures, that X is stationary with marginal density

$$f(x) = \frac{1}{M} \frac{1}{\sigma^2 \varphi'(x)} \qquad \forall\, x \in (l, r)$$

if the distribution of the initial variable $X(0)$ has density f, Pollak and Siegmund (1985) and Karatzas and Shreve (1991). The ergodic theorem for diffusions, Rogers and Williams (1987), ensures the ergodic property of X.

We will now give three examples of diffusion processes, that we use in the simulation studies and the empirical analysis.

Example A.1 Ornstein-Uhlenbeck-Process:

$$dX(t) = \theta_1(\theta_2 - X(t))dt + \theta_3 dW(t)$$

If $X(0)$ has a normal distribution with expectation θ_2 and variance $\theta_3^2/(2\theta_1)$ than X is stationary and $l = -\infty$ and $r0\infty$.

Example A.2 Square root process:

Cox et al. (1985) introduce the square root process as a model for interest rates. We call a diffusion X square root process, if X solves the SDE

$$dX(t) = \theta_1\{\theta_2 - X(t)\}dt + \theta_3\sqrt{X(t)}dW(t)\ . \tag{A.1}$$

X is stationary and positive if $2\theta_1\theta_2/\theta_3^2 - 1$ is positive.

Example A.3 Ahn-Gao model:

The Ahn-Gao process was introduced by Ahn and Gao (1999) to model interest rate processes. The diffusion coefficient is the Chan et al. (1992) (CKLS) coefficient but the drift term is not linear. We call a diffusion X an Ahn-Gao process, if X solves the following differential equation:

$$dX(t) = \theta_1\{\theta_2 - X(t)\}X(t)dt + \theta_3 X(t)^{1.5}dW(t)\ . \tag{A.2}$$

Ahn and Gao (1999) proof, that under the condition that θ_1, θ_2 and θ_3 are positive, a stationary strictly positive solution of (A.2) exists. They also provide a closed form for the marginal density, the transition density and the conditional and unconditional moments.

A.2 Proofs of Theorems and Lemmas of EL section

We give all proofs as in Chen et al. (2001) for time series.
PROOF of Lemma 3.1:
Define $\epsilon_i = Y_i - m(X_i)$ and write $\bar{U}_1(x) = I_1(x) + I_2(x) + I_3(x)$ where

$$I_1 = [nT]^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i)\{m_\theta(X_i) - \tilde{m}_{\hat{\theta}}(x)\} = \hat{f}(x)\{\tilde{m}_\theta(x) - \tilde{m}_{\hat{\theta}}(x)\},$$

$I_2 = [nT]^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i)\epsilon_i$ and $I_3 = [nT]^{-1} c_n \sum_{i=0}^{[nT]-1} K_h(x - X_i)\Delta_n(X_i)$. As

$$\sup_{x \in S} |[nT]^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i) - f(x)| \xrightarrow{a.s.} 0$$

as shown in Bosq (1998)p.49, condition (DT5) implies,

$$\sup_{x \in S} |I_1(x)| = \mathcal{O}_p([nT]^{-1/2}) \quad \text{and} \quad \sup_{x \in S} |I_3(x)| = \mathcal{O}_p(c_n). \tag{A.3}$$

Let $M_n = b_0 \log(n)$ for some positive constant b_0. Split $I_2(x)$ into two parts:

$$I_2^+(x) = n^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i)\epsilon_i I(|\epsilon_i| \geq M_n)$$

and

$$I_2^-(x) = n^{-1} \sum_{i=0}^{[nT]-1} K_h(x - X_i)\epsilon_i I(|\epsilon_i| < M_n).$$

As $\sup_{x \in S} |I_2^+(x)| \leq C([nT]h)^{-1} \sum_{i=0}^{[nT]-1} |\epsilon_i| I(|\epsilon_i| \geq M_n)$ for some $C > 0$, the Cauchy-Schwartz inequality implies that

$$E\left[\sup_{x \in S} |I_2^+(x) - E\{I_2^+(x)\}|\right] \leq 2C([nT]h)^{-1} \sum_{i=0}^{[nT]-1} \{E(|\epsilon_i|^2)\mathrm{P}(|\epsilon_i| \geq M_n)\}^{1/2}.$$

From the Chebyshev inequality and condition (DT6), for a positive constant η_0,

$$\begin{aligned} &\mathrm{P}\left[M_n^{-1}([nT]h)^{1/2} \sup_{x \in S} |I_2^+(x) - E\{I_2^+(x)\}| \geq \eta_0)\right] \\ &\quad \leq 2C\eta_0^{-1} n^{1/2} h^{-d/2} M_n^{-1} \exp\{-\tfrac{1}{2} a_0 b_0 \log(n)\}. \end{aligned}$$

By properly choosing b_0, the right hand sight converges to zero as $n \to \infty$. This means that

$$\sup_{x \in S} |I_2^+(x) - E\{I_2^+(x)\}| = o_p\{([nT]h)^{-1/2}\log(n)\}. \tag{A.4}$$

Let $\phi_i(x) = K(\frac{x-X_i}{h})\epsilon_i I(|\epsilon_i| < M_n)$, $Z_i(x) = \phi_i(x) - E\{\phi_i(x)\}$. Clearly, at each fixed x, $\{Z_i(x)\}$ has zero mean, is bounded by $b = C_1 M_n$ and geometrical the α-mixing. Put $\eta = (h/n)^{1/2} M_n \eta_0$. From Theorem 1.3 of Bosq (1998),

$$\begin{aligned} &\mathrm{P}\left[|I_2^-(x) - E\{I_2^-(x)\}| > ([nT]h)^{-1/2} M_n \eta_0\right] = \mathrm{P}\left(|\sum Z_i(x)| > n\eta\right) \\ \leq\ & 4\exp[-\eta^2 q/\{8v^2(q)\}] + 22(1 + 4C_1 M_n/\eta)^{1/2} q\alpha\{[n/(2q)]\} \end{aligned} \tag{A.5}$$

where $q = \eta_0 M_n^2 \sqrt{n} h^{-d/2}$, $p = n/q$, $v^2(q) = \frac{2}{p^2}\sigma^2(q) + \frac{b\eta}{2}$ and

$$\sigma^2(q) = \max_{0 \le j \le 2q-1} E\{\beta_1(p) Z_{[jp]+1}(x) + \sum_{i=2}^{p} Z_{[jp]+i}(x) + \beta_2(p) Z_{[jp]+1}(x)\}^2.$$

In the last equation $\beta_1(p) = [jp] + 1 - jp$ and $\beta_2(p) = (j+1)p - [(j+1)p]$. By the stationarity of $\{(X_i, Y_i)\}$,

$$\sigma^2(q) \le (p+2) E\{Z_1^2(x)\} + J \tag{A.6}$$

where

$$J = 2p \sum_{l=1}^{p} (1 - \frac{l-1}{p}) |\operatorname{Cov}\{Z_1(x), Z_{l+1}(x)\}| + 2|\operatorname{Cov}\{Z_1(x), Z_{p+1}(x)\}|.$$

Condition (DT6) implies that $E(|\epsilon|^\delta) < \infty$ for some $\delta > 2$. Using the Davydov's lemma,

$$\begin{aligned} |\operatorname{Cov}\{Z_1(x), Z_{l+1}(x)\}| &\le 2\delta(\delta-2)^{-1}\{E|K(\frac{x-X_i}{h})\epsilon_i|^\delta\}^{2/\delta} \alpha^{1-2/\delta}(l) \\ &\le Ch\alpha^{1-2/\delta}(l). \end{aligned} \tag{A.7}$$

Following the approach used in Fan and Gijbels (1996), we let $d_n \to \infty$ be a sequence of integers such that $d_n h \to 0$ and split J as

$$\begin{aligned} J_1 &= 2p \sum_{l=1}^{d_n - 1} \left(1 - \frac{l-1}{p}\right) \left|\operatorname{Cov}\{Z_1(x), Z_{l+1}(x)\}\right| \quad \text{and} \\ J_2 &= 2p \sum_{l=d_n}^{p} \left(1 - \frac{l-1}{p}\right) \left|\operatorname{Cov}\{Z_1(x), Z_{l+1}(x)\}\right| \\ &\quad + 2|\operatorname{Cov}\{Z_1(x), Z_{p+1}(x)\}|. \end{aligned} \tag{A.8}$$

As $\text{Cov}\{Z_1(x), Z_{l+1}(x)\}| \le Var\{Z_1(x)\} \le C$, $J_1 \le Cpd_n = o(ph)$. From (A.7) and condition (DT1), we have $J_2 = o(ph)$ as well. These imply that

$$J = J_1 + J_2 = o(ph) \tag{A.9}$$

and hence $\sigma^2(q) \le Cph$. The particular forms of q, b and η mean that $v^2(q) \le Cqh/n$ and

$$\exp[-\eta^2 q/\{8v^2(q)\}] \le \exp\left(-C_1 M_n^2 \eta_0^2\right). \tag{A.10}$$

The geometric the α-mixing condition implies:

$$(1 + 4Ch^d/\eta)^{1/2} q\alpha\{n/(2q)\} \le C_2([nT]h^{-d})^{-3/4} M_n^2 \rho^{\frac{1}{2}\eta_0} M_n^{-2}([nT]h)^{1/2}. \tag{A.11}$$

Combining (A.10), (A.11) with (A.5) and noticing that both (A.10) and (A.11) are free of x, we have

$$\begin{aligned} &\sup_{x\in S} \mathrm{P}[|I_2^-(x) - E\{I_2^-(x)\}| \ge ([nT]h)^{-1/2}\log(n)\eta_0] \qquad \text{(A.12)} \\ \le\ & \exp\left(-C_1 b_0^2 \log^2(n)\eta_0^2\right) + C_2 h^{-3d/4} n^{3/4} M_n^2 \rho^{\frac{1}{2}\eta_0} b_0^{-2} \log^{-2}(n)([nT]h)^{1/2}. \end{aligned}$$

Let $\{B_k\}_{k=1}^{v_n}$ be a set of equal volume disjoint hypercubes with centers $\{s_k\}_{k=1}^{v_n}$ such that $S = \bigcup_{k=1}^{v_n} B_k$, $v_n = [n^{t_0}]$ for some $t_0 > 0$ and $\sup_{x\in B_k} ||x - s_k|| \le cv_n^{-1}$. Based on this partition of S, and let $I_2^{-\star}(x) = I_2^{(}x) - E\{I_2^-(x)\}$

$$\sup_{x\in S} |I_2^-(x) - E\{I_2^-(x)\}| \le \max_{k=1,\dots,v_n} |I_2^{-\star}(s_k)| + \sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})|$$

where $k(x)$ being the index of the hypercube containing x. Note that

$$\begin{aligned} \mathrm{P}\{\max_{k=1,\dots,v_n} |I_2^{-\star}(s_k)| \ge ([nT]h)^{-1/2}\eta_0 M_n\} \ &\le\ n^{t_0} \sup_{x\in S} \mathrm{P}\{|I_2^-(x) - E\{I_2^{-1}(x)\}| \\ &\ge\ ([nT]h)^{-1/2} M_n \eta_0\}, \end{aligned}$$

By properly choosing b_0, (A.12) implies that

$$\max_{k=1,\dots,v_n} |I_2^{-\star}(s_k)| = o_p\{([nT]h)^{-1/2}\log(n)\}. \tag{A.13}$$

As K is Lipschitz continuous,

$$\sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| \le Ch^{-1}n^{-t_0}\left(n^{-1}\sum_{i=0}^{[nT]-1}|\epsilon_i| + E|\epsilon_i|\right).$$

Note that $n^{-1}\sum|\epsilon_i| \overset{w.s.}{\to} E|\epsilon_i|$, and $E|\epsilon_i| \le C$. We get with probability one

$$\sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| \le Ch^{-1}n^{-t_0}.$$

By choosing $t_0 > 3/\{2(d+1)\}$, we have

$$P\{\sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| \geq ([nT]h)^{-1/2}\log(n)\eta_0\} \to 0,$$

which means that

$$\sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| = o_p\{([nT]h)^{-1/2}\log(n)\}. \tag{A.14}$$

Clearly, (A.4), (A.13) and (A.14) complete the proof. □

PROOF of Lemma 3.2:
We need to do a few things before proving the lemma. Similar to the derivation of Lemma 3.1 and the proof of Theorem 2.2 of Bosq (1998), it can be shown that for any smooth function g in $\mathbb{R}^d$

$$\begin{aligned} &\sup_{x\in S} |n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)g(X_i) - f(x)v(x;h)g(x)| \\ =\ & \mathcal{O}_p\{([nT]h)^{-1/2}\log(n) + h\}, \end{aligned} \tag{A.15}$$

$$\begin{aligned} &\sup_{x\in S} |n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\epsilon_i^2 - f(x)v(x;h)\sigma^2(x)| \\ =\ & \mathcal{O}_p\{([nT]h)^{-1/2}\log(n) + h\} \end{aligned} \tag{A.16}$$

and

$$\sup_{x\in S} |n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\epsilon_i| = \mathcal{O}_p\{([nT]h)^{-1/2}\log(n)\} \tag{A.17}$$

where the h-order terms in the remainders are due to the bias associated with the kernel estimator. Note that

$$\bar{U}_2(x) = n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\{m_\theta(X_i) - \tilde{m}_{\hat{\theta}}(x) + \epsilon_i + c_n\Delta_n(X_i)\}^2 = \sum_{l=1}^{6} J_l(x)$$

where, from (A.15) to (A.17),

$$
\begin{aligned}
J_1(x) &= n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\{m_\theta(X_i)-\tilde{m}_{\hat\theta}(x)\}^2 = \tilde{O}_p\{n^{-1/2}+h\} \\
J_2(x) &= n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\epsilon_i^2 \to f(x)v(x;h)\sigma^2(x) + \tilde{O}_p\{([nT]h)^{-1/2}\log(n)+h\} \\
J_3(x) &= n^{-1}hc_n^2 \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\Delta_n^2(X_i) = \tilde{O}_p(c_n^2) \\
J_4(x) &= 2n^{-1}hc_n \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\{m_\theta(X_i)-\tilde{m}_{\hat\theta}(x)\}\Delta(X_i) = \tilde{O}_p\{c_n(n^{-1}+h)\} \\
J_5(x) &= 2n^{-1}h \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\{m_\theta(X_i)-\tilde{m}_{\hat\theta}(x)\}\epsilon_i = \tilde{o}_p(n^{-1/2}) \\
J_6(x) &= 2n^{-1}hc_n \sum_{i=0}^{[nT]-1} K_h^2(x-X_i)\epsilon_i\Delta_n(X_i) = \tilde{O}_p\{c_n([nT]h)^{-1/2}\log(n)\}.
\end{aligned}
$$

In summary of the above results, we have

$$\sup_{x\in S}|\bar{U}_2(x) - f(x)v(x;h)\sigma^2(x)| = \mathcal{O}_p(h). \tag{A.18}$$

As $f(x)v(x;h)\sigma^2(x)$ is uniformly bounded below,

$$\inf_{x\in S} f(x)v(x;h)\sigma^2(x) \ge d_0 \quad \text{for some } d_0 > 0. \tag{A.19}$$

Since

$$\inf_{x\in S}|U_2(x)| \ge -\sup_{x\in S}|U_2(x) - f(x)v(x;h)\sigma^2(x)| + \inf_{x\in S}|f(x)v(x;h)\sigma^2(x)|,$$

The proof is completed by (A.18) and (A.19). □

PROOF of Lemma 3.3:
Let $w_i = \sup_{x\in S}|K\left(\frac{x-X_i}{h}\right)\{Y_i - \tilde{m}_{\hat\theta}(x)\}|$. As K, m and Δ_n are bounded in S, $w_i \le C_1|\epsilon_i| + C_2$. From the Chebyshev inequality and Condition (DT6)

$$
\begin{aligned}
\mathrm{P}\left(w_i > ([nT]h)^{1/2}\{\log(n)\}^{-1}\right) &\le \mathrm{P}\left(|\epsilon_i| \ge C_3([nT]h)^{1/2}\{\log(n)\}^{-1}\right) \\
&\le C_4\exp\{-C_5([nT]h)^{1/2}\log^{-1}(n)\}
\end{aligned}
$$

Thus, $\sum_{n=1}^{\infty}\mathrm{P}\left(w_i > ([nT]h)^{1/2}\{\log(n)\}^{-1}\right) < \infty$. According to the Borel-Cantelli lemma, $w_i > ([nT]h)^{1/2}\{\log(n)\}^{-1}$ finitely often with probability 1. This means that $Z_n = \max_{1\le i\le n} w_i > ([nT]h)^{1/2}\{\log(n)\}^{-1}$ finitely often, which completes the proof. □

PROOF of Theorem 3.3:
From (3.36) and (A.18)

$$\begin{aligned}\bar{U}_2^{-1}\bar{U}_1^2 &= \left[n^{-1}\sum_{i=0}^{[nT]-1} W_h(x-X_i)\{\epsilon_i + c_n\Delta_n(X_i)\}\right]^2 \\ &\quad + \tilde{\mathcal{O}}_p\{([nT]h)^{-1}h\log^2(n)\} \qquad \text{(A.20)}\end{aligned}$$

where $W_h(x-X_i) = K_h(x-X_i)/\{f(x)v(x;h)\sigma^2(x)\}^{1/2}$. Note that $([nT]h^1)^{1/2}c_n = \mathcal{O}(h^{1/4})$. Let

$$\begin{aligned}A &= k_T^{-1}([nT]h)\sum_{j=1}^{k_T}\int_{B_j}\left[n^{-1}\sum_{i=0}^{[nT]-1} W_h(t_j-X_i)\{\epsilon_i+c_n\Delta_n(X_i)\}\right]^2 \\ &\quad -[n^{-1}\sum_{i=0}^{[nT]-1} W_h(t-X_i)\{\epsilon_i+c_n\Delta_n(X_i)\}]^2\Bigg]dt \\ &= k_T^{-1}\sum_{j=1}^{k_T}\int_{B_j} T_{1j}(t)T_{2j}(t)dt \qquad \text{(A.21)}\end{aligned}$$

where for $t \in B_j$

$$\begin{aligned}T_{1j}(t) &= n^{-1/2}\sum_{i=0}^{[nT]-1}\{W_h(t_j-X_i)-W_h(t-X_i)\}\{\epsilon_i+c_n\Delta_n(X_i)\}, \\ T_{2j}(t) &= n^{-1/2}\sum_{i=0}^{[nT]-1}\{W_h(t_j-X_i)+W_h(t-X_i)\}\{\epsilon_i+c_n\Delta_n(X_i)\}.\end{aligned}$$

Let $M_n = b_0\log(n)$ for a positive constant b_0 and $\omega_i = \epsilon_i + c_n\Delta_n(X_i)$. Define

$$\begin{aligned}T_{1j}^+(t) &= ([nT]h^{-d})^{-1/2}\sum_{i=0}^{[nT]-1}\{W_h(t_j-X_i)-W_h(t-X_i)\}\omega_i I(|\omega_i|>M_n), \\ T_{1j}^-(t) &= ([nT]h^{-d})^{-1/2}\sum_{i=0}^{[nT]-1}\{W_h(t_j-X_i)-W_h(t-X_i)\}\omega_i I(|\omega_i|\le M_n).\end{aligned}$$

Similar definitions apply for $T_{2j}^+(t)$ and $T_{2j}^-(t)$. It may be shown similar to the derivation of (A.4) that for $l = 1$ and 2

$$\max_{j=1,\dots,k_T}\sup_{t\in B_j}|T_{1j}^+(t) - E\{T_{1j}^+(t)\}| = o_p\{([nT]h)^{-1/2}\log(n)\}. \qquad \text{(A.22)}$$

Let $\phi_i(t) = h\{W_h(t_j - X_i) - W_h(t - X_i)\}\omega_i I(|\omega_i| < M_n)$ and $Z_i(t) = \phi_i(t) - E\{\phi_i(t)\}$. Then, for $u_n \to \infty$ (the exact order of u_n will be decided later)

$$\mathrm{P}\{|T_{1j}^{-}(t) - E\{T_{1j}^{-}(t)\} > u_n^{-1}\eta_0\} \;=\; \mathrm{P}\{|\sum_{i=0}^{[nT]-1} Z_i(t)| > n\eta\}$$

where $\eta = (h/n)^{1/2}u_n^{-1}\eta_0$. Note that $|Z_i(t)| \le CM_n k_T^{-1}h^{-1}$. Let $b = CM_n k_T^{-1}h^{-1}$ and $q = n^{1/2}M_n\eta_0 u_n^{-1}h^{-1}$. Similar to the derivation of (A.12) and employing again Theorem 1.3 of Bosq (1998), we have

$$\mathrm{P}\{|\sum_{i=0}^{[nT]-1} Z_i(t)| > n\eta\} \;\le\; 4\exp\{-\frac{\eta^2 q}{8v^2(q)}\} + 22(1 + b/\eta)^{1/2}q\alpha\left([n/(2q)]\right)$$

where $v^2(q) \le Cqh(nk_T)^{-1}$. The upper bound for $v^2(q)$ can be obtained using the same approach in deriving a similar bound for the same name quantity as given between (A.5) and (A.9). By choosing $u_n = b_1 k_T^{1/2}\log^{-1}(n)$ for some positive b_1,

$$\exp\{-\frac{\eta^2 q}{8v^2(q)}\} \le \exp\left(-C\frac{k_T\eta_0}{u_n^2}\right) = \exp\{-Cb_1\eta_0\log(n)\}$$

and

$$(1 + b/\eta)^{1/2}q\alpha\left([n/(2q)]\right) \le Cn^{3/4}M_n^{3/2}u_n^{-1/2}h^{-3/2-d/4}k_n^{-1/2}\rho^{M_n^{-1}\eta_0^{-1}n^{1/2}u_n h}.$$

As the right hand sides of the above two inequalities are free of t, we have

$$\begin{aligned}
&\sup_{t\in B_j}\mathrm{P}[|T_{1j}^{-}(t) - E\{T_{1j}^{-}(t)\}| \ge b_1 k_n^{-1/2}\log(n)\eta_0] \\
\le\;& \exp\{-Cb_1\eta_0\log(n)\} \\
&\quad + Cn^{3/4}M_n^{3/2}u_n^{-1/2}h^{-3/2-d/4}k_n^{-1/2}\rho^{M_n^{-1}\eta_0^{-1}n^{1/2}u_n h}. \qquad \text{(A.23)}
\end{aligned}$$

Let $\{B_{jl}\}_{l=1}^{v_j}$ be a partition of B_j of equal size hypercubes B_{jl} where v_j be an integer tending to ∞ as $n \to \infty$. Employing similar derivations to those in deriving (A.13) and (A.14) and utilizing (A.23), it can be shown that

$$\sup_{t\in[t_j,t_{j+1}]}|T_{1j}^{-}(t) - E\{T_{1j}^{-}(t)\}| = \mathcal{O}_p\{k_n^{-1/2}\log(n)\}. \qquad \text{(A.24)}$$

A similar derivation will show that

$$\sup_{t\in[t_j,t_{j+1}]}|T_{2j}^{-}(t) - E\{T_{2j}^{-}(t)| = \mathcal{O}_p\{k_n^{-1/2}\log(n)\}. \qquad \text{(A.25)}$$

From (A.22), (A.24) and (A.25) we have for $l = 1$ and 2

$$\sup_{t\in[t_j,t_{j+1}]} |T_{lj}(t)| = \mathcal{O}_p\{([nT]h)^{-1/2}\log(n) + k_n^{-1/2}\log(n)\}. \tag{A.26}$$

These together with (A.21) complete the proof. □

PROOF of Theorem 3.4:

We first derive the mean and the covariance of $\hat{m}(x) - \tilde{m}_\theta(x)$. We use $\tilde{O}()$ and $\tilde{o}()$ to denote quantities which are $\mathcal{O}()$ and $o()$ uniformly with respect to $x \in S$. It is noted that

$$\begin{aligned} & E\{\hat{m}(x) - \tilde{m}_\theta(x)\} \\ = \; & E\left[\frac{n^{-1}\sum_{i=0}^{[nT]-1} W_h(x - X_i)\{\epsilon_i + c_n\Delta_n(X_i)\}}{b(x;h)f(x)}\{1 + \frac{\hat{f}(x) - b(x;h)f(x)}{b(x;h)f(x)} + \dots\}\right] \\ = \; & c_n\Delta_n(x)\{1 + \tilde{O}(h)\} \end{aligned}$$

When x is in the interior of S, the above $\tilde{O}(h)$ term will be $\tilde{O}(h^2)$. This means that

$$E\left[([nT]h)^{1/2}V^{-1/2}(x)\{\hat{m}(x) - \tilde{m}_\theta(x)\}\right] = ([nT]h)^{1/2}c_n\Delta_n(x)V^{-1/2}(x)\{1 + \tilde{o}(1)\}. \tag{A.27}$$

Let $\omega_i = \epsilon_i + c_n\Delta_n(X_i)$. Then,

$$\begin{aligned} & V^{1/2}(s;h)V^{1/2}(t;h)\,\mathrm{Cov}\{\hat{m}(s) - \tilde{m}_\theta(s), \hat{m}(t) - \tilde{m}_\theta(t)\} \\ = \; & \mathrm{Cov}\{n^{-1}\sum_{i=0}^{[nT]-1} W_h(s - X_i)\omega_i, n^{-1}\sum_{i=0}^{[nT]-1} W_h(t - X_i)\omega_i\}\{1 + \tilde{o}(1)\} \\ = \; & \Big[n^{-1}\,\mathrm{Cov}\{W_h(s - X_1)\omega_1, W_h(t - X_1)\omega_1\} \\ & + n^{-1}\sum_{l=2}^{n}(1 - l/n)\,\mathrm{Cov}\{W_h(s - X_1)\omega_1, W_h(s - X_l)\omega_l\}\Big]\{1 + \tilde{o}(1)\} \end{aligned}$$

Standard derivations show

$$\mathrm{Cov}\{W_h(s - X_1)\omega_1, W_h(t - X_1)\omega_1\} = h^{-d}\sqrt{\frac{f(s)\sigma^2(s)}{f(t)\sigma^2(t)}}\frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}} + \tilde{o}(h^{-d}),$$

where $W_0^{(2)}$ is defined in (3.39) and $W_0^{(2)}(t,t) = v(t;h)$. Using the same arguments which establish (A.9) in the proof of Lemma 3.1 , we can show that

$$\sum_{l=2}^{n}(1 - l/n)\,\mathrm{Cov}\{W_h(s - X_1)\omega_1, W_h(s - X_l)\omega_l\} = \tilde{o}(h^{-d}).$$

Thus,

$$([nT]h)\,\mathrm{Cov}\left[\frac{\{\hat{m}(s)-\tilde{m}_{\hat{\theta}}(s)\}}{\sqrt{V(s)}},\frac{\{\hat{m}(t)-\tilde{m}_{\hat{\theta}}(t)\}}{\sqrt{V(t)}}\right]$$
$$=\sqrt{\frac{f(s)\sigma^2(s)}{f(t)\sigma^2(t)}}\frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}}\{1+\tilde{o}(1)\}. \tag{A.28}$$

Next we want to show that for k distinct $t_1, t_2, \cdots, t_k \in [0,1]^d$,

$$([nT]h)^{1/2}\left(\frac{\{\hat{m}(t_1)-\tilde{m}_{\theta}(t_1)\}}{V(t_1)},\cdots,\frac{\{\hat{m}(t_k)-\tilde{k}_{\theta}(t_k)\}}{V(t_k)}\right)\xrightarrow{\mathcal{L}} N_k(\mu_k,\Omega_k). \tag{A.29}$$

Here $N_k(\mu_k, \Omega_k)$ is a k-dimensional normal distribution with mean vector

$$\mu_k = ([nT]h)^{1/2}c_n\left(\Delta_n(t_1)f^{1/2}(t_1)V^{-1/2}(t_1),\cdots,\Delta_n(t_k)f^{1/2}(t_k)V^{-1/2}(t_k)\right)^T$$

and covariance matrix $\Omega_k = (\omega_{ij})_{k\times k}$, where

$$\omega_{ij}=\sqrt{\frac{f(t_i)\sigma^2(t_i)}{f(t_j)\sigma^2(t_j)}}\frac{W_0^{(2)}(t_i,t_j)}{\sqrt{W_0^{(2)}(t_i,t_i)W_0^{(2)}(t_j,t_j)}}.$$

From Theorem 3.4 of Bosq (1998), $V^{-1/2}(t_i)\{\hat{m}(t_i)-\tilde{m}_{\theta}(t_i)\}$ is asymptotically normally distributed at each t_i. Then (A.29) is obtained by applying the Cramér-Wold device.

From Theorem 1.5.4 of van der Vaart and Wellner (1996), we only need to show that $([nT]h)^{1/2}\hat{m}()/V^{-1/2}()$ is asymptotically tight in $C([0,1]^d)$. To simplify the presentation, we only prove the case for $d = 1$.

From Theorem 8.1 and Theorem 12.3 of Billingsley (1968), we need only to show that

$$([nT]h)^{1/2}V^{-1/2}(0)\{\hat{m}(0)-\tilde{m}_{\hat{\theta}}(0)\}\text{ is tight and} \tag{A.30}$$
$$P\{([nT]h)^{1/2}|V^{-1/2}(t_1)\{\hat{m}(t_1)-\tilde{m}_{\theta}(t_1)\}-V^{-1/2}(t_2)\{\hat{m}(t_2)-\tilde{m}_{\theta}(t_2)\}|>\eta\}$$
$$\le C(t_1-t_2)^{\alpha}/\eta_0^{\gamma}, \tag{A.31}$$

for any $\eta_0 > 0$, some $\gamma > 0$ and $\alpha > 1$.

As $V^{-1/2}(0)\{\hat{m}(0)-\tilde{m}_{\hat{\theta}}(0)\}$ has finite mean and variance, (A.30) is readily established from the Markov inequality. Note that

$$
\begin{aligned}
& ([nT]h)^{1/2}\Big[V^{-1/2}(t_1)\{\hat{m}(t_1)-\tilde{m}_\theta(t_1)\}-V^{-1/2}(t_2)\{\hat{m}(t_2)-\tilde{m}_\theta(t_2)\}\Big] \\
= \; & ([nT]h)^{1/2}n^{-1}\sum_{i=0}^{[nT]-1}\{W_h(t_1-X_i)-W_h(t_2-X_i)\}\{\epsilon_i+c_n\Delta(x_i)\}+o_p(1) \\
= \; & ([nT]h)^{1/2}n^{-1}\sum_{i=0}^{[nT]-1}\{W_h(t_1-X_i)-W_h(t_2-X_i)\}\epsilon_i+o_p(1)
\end{aligned}
$$

So, it is sufficient to prove for any $\eta > 0$,

$$
P\{h|\sum_{i=0}^{[nT]-1} Z_i| > \sqrt{nh}\eta_0\} \le C(t_1-t_2)^\alpha/\eta_0^\gamma. \tag{A.32}
$$

where $Z_i = h\{W_h(t_1-X_i)-W_{0h}(t_2-X_i)\}\epsilon_i$. Split Z_i into $Z_{i1} = Z_iI(|\epsilon_i| < M_n)$ and $Z_{i2} = Z_iI(|\epsilon_i| > M_n)$ where M_n is a larger number slowly tending to ∞. Clearly, $|Z_{i1}| \le b =: C|t_1-t_2|M_n/h$. Using again Theorem 1.3 of Bosq (1998),

$$
\begin{aligned}
& \mathrm{P}\{|\sum_{i=0}^{[nT]-1} Z_{i1}| > \tfrac{1}{2}([nT]h)^{1/2}\eta_0\} = \mathrm{P}\left(|\textstyle\sum_{i=0}^{[nT]-1} Z_{i1}| > n\eta\right) \\
\le \; & 4\exp\{-\frac{\eta^2 q}{8v^2(q)}\} + C(b/\eta)^{1/2}q\alpha\{[n/(2q)]\}
\end{aligned}
$$

where $q = n^{1/2}h^{-3/2}M_n\eta$ and $v^2(q) = Cqh|t_1-t_2|/n$. Thus,

$$
\exp\{-\frac{\eta^2 q}{8v^2(q)}\} \le \exp\left(-C\eta_0^2|t_1-t_2|^{-1}\right) \le C|t_1-t_2|^2\eta_0^{-2}
$$

and condition (DT1) implies that $(b/\eta)^{1/2}q\alpha\{[n/(2q)]\} \to 0$. Therefore,

$$
\mathrm{P}\{|\sum_{i=0}^{[nT]-1} Z_{i1}| > \tfrac{1}{2}([nT]h)^{1/2}\eta_0\} \le C|t_1-t_2|^2\eta_0^{-2}. \tag{A.33}
$$

Standard techniques, similar to those used in studying the properties of I_2^+ in the proof of Lemma 2, show that as $n \to \infty$

$$
\mathrm{P}\{|\sum_{i=0}^{[nT]-1} Z_{i2}| > \tfrac{1}{2}([nT]h)^{1/2}\eta_0\} \to 0.
$$

This and (A.33) prove (A.32), and complete the proof for the tightness.

□

A.3 Estimation of Parameters of a Normalized Index

Let $U(t)$ be the Ornstein-Uhlenbeck process satisfying (4.9). Introduce the autocovariance of the process $U - K_h * U$ as

$$\mathrm{Cov}(\tau) = \mathrm{Cov}[(U - K_h * U)(\tau);\ (U - K_h * U)(0)],\ 0 < \tau << h.$$

Let

$$\rho_h(\tau) = \frac{\mathrm{Cov}(\tau)}{\mathrm{Cov}(0)}$$

be the autocorrelation function of $U - K_h * U$.

Proposition 1. (i) *If $K(u)$ is the rectangle kernel, i.e., $K(u) = (1/2)\, \boldsymbol{I}(|u| \le 1)$, then as $h \to \infty$ we have that*

$$Cov(\tau) \;=\; \frac{\gamma^2}{2\beta}\Big(\ e^{-\beta\tau} - \frac{1}{\beta h} - \frac{1}{2\,\beta h}\,(\tau/h) + \mathcal{O}(h^{-2})\ \Big). \tag{A.34}$$

(ii) If $K(u) = (3/4)(1-u^2)\,\mathbb{I}(|u| \le 1)$ is the Epanechnikov kernel, then

$$\mathrm{Cov}(\tau) \;=\; \frac{\gamma^2}{2\beta}\Big(\ e^{-\beta\tau} - \frac{1.8}{\beta h} - \frac{1.5}{\beta h}\,(\tau/h)^2 + \mathcal{O}(h^{-2})\ \Big). \tag{A.35}$$

Proposition 2. *Under the assumptions of Proposition 1, up to the terms of the magnitude $\mathcal{O}(h^{-2})$, the following equation holds for β:*

$$\frac{\partial^+}{\partial\tau}\rho_h(\tau)\Big|_{\tau=0} = -\frac{\beta}{1 - c_K/(\beta h)}.$$

The solution of this equation is approximately

$$\hat\beta_2 = \Big|\frac{\partial^+}{\partial\tau}\rho_h(\tau)\Big|\Big|_{\tau=0} - \frac{c_K}{h} + \mathcal{O}(h^{-2})$$

where $c_K = 1$ if K is the rectangle kernel, and $c_K = 1.8$ if K is the Epanechnikov kernel.

PROOF of Proposition 1.:

(i) Integrate the both sides of (4.9) with the rectangle kernel. The integration results in

$$-\frac{1}{2\beta h}\big\{U(t+h) - U(t-h)\big\} = (K_h * U)(t) - \frac{\gamma}{2\beta h}\big\{W(t+h) - W(t-h)\big\},$$

or

$$(K_h * U)(t) = \frac{\gamma}{2\beta h}\big(W(t+h) - W(t-h)\big) - \frac{1}{2\beta h}\big(U(t+h) - U(t-h)\big). \quad \text{(A.36)}$$

The autocovariance function equals

$$\begin{aligned}\mathrm{Cov}(\tau) \;&=\; E[U(\tau)U(0)] - E[U(\tau)\,(K_h * U)(0)] \\ &\quad - E[U(0)\,(K_h * U)(\tau)] + E[(K_h * U)(\tau)\;(K_h * U)(0)]. \quad \text{(A.37)}\end{aligned}$$

The first term in the latter formula is $E[U(\tau)U(0)] = \{\gamma^2/(2\beta)\}e^{-\beta\tau}$. With the help of (A.36) and the explicit representation

$$U(t) = \int_{-\infty}^{t} \exp\{-\beta(t-s)\}\,\gamma\,dW(s),$$

one finds by direct calculation that each of the negative terms on the right-hand side of (A.37) contributes

$$-\frac{\gamma^2}{(2\beta^2\,h)} + \mathcal{O}(h^{-2}),$$

while the covariance of $K_h * U$ adds up to

$$\frac{\gamma^2(2h-\tau)}{(2\beta h)^2} + \mathcal{O}(h^{-2}).$$

Combining these results, we arrive at (A.34).

(ii) Integrating (4.9), we find as in (A.36) that

$$(K_h * U)(t) = -\frac{3}{2\beta h^3}\int_{t-h}^{t+h}(s-t)U(s)ds + \frac{\gamma}{\beta}\,(K_h * W)(t). \quad \text{(A.38)}$$

It is straightforward to verify that the variance of the first term on the right-hand side of (A.38) has the magnitude $\mathcal{O}(h^{-2})$ for h large. This term is negligible as compared to the second one. As in part (i), we obtain

$$\begin{aligned}\mathrm{Cov}(\tau) \;&=\; \frac{\gamma^2}{2\beta}e^{-\beta\tau} - 2\left(\frac{3\gamma^2}{4\beta^2 h} + \mathcal{O}(h^{-2})\right) + \frac{\gamma^2}{\beta^2}\,E\big[(K_h * W)(\tau)\,(K_h * W)(0)\big] \\ &=\; \frac{\gamma^2}{2\beta}e^{-\beta\tau} - \frac{3\gamma^2}{2\beta^2 h} + \frac{\gamma^2}{\beta^2}\left(\frac{3}{5} - \frac{3}{4}\left(\frac{\tau}{h}\right)^2\right) + \mathcal{O}(h^{-2}). \quad \text{(A.39)}\end{aligned}$$

This proves (A.35). □

PROOF of Proposition 2.:
The slope of the autocorrelation at zero follows from (A.34) and (A.35). Let

$$A = \left|\frac{\partial^+}{\partial\tau}\rho_h(\tau)\right|\Big|_{\tau=0},$$

then for $\hat{\beta}_2$ the quadratic equation $\hat{\beta}_2^2\, h - \hat{\beta}_2\, hA + c_K A = 0$ holds with root $\hat{\beta}_2 = A - c_K/h + \mathcal{O}(h^{-2})$. □

Proposition 3. *Let C_0 be a space of continuous function with finite support. Define K as the rectangle or Epanechnikov kernel. Then the operator K_h is a contracting operator on the space $L_2 \bigcap C_0$* with the L_2-norm.

PROOF of Proposition 3.:
The Fourier transformation for the rectangle kernel is $\tilde{K}(z) = (\sin z)/z$, and for the Epanechnikov kernel is $\tilde{K}(z) = 3(\sin z - z\cos z)/z^3$ with unique maximum value 1 at $z = 0$. Thus, for the n-th iterative convolution, $\|K^n\|_2 \to 0$ as $n \to \infty$. This confirms the result. □

Lebenslauf

Name:	Torsten Kleinow
10/1991 - 09/1998	Studium an der Humboldt-Universität zu Berlin in der Fachrichtung Mathematik
seit 10/1998	Wissenschaftlicher Mitarbeiter an der Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät Institut für Statistik und Ökonometrie Lehrstuhl Prof. W. Härdle

Selbständigkeitserklärung

Bei der Erstellung meiner Dissertation habe ich ausser der angegebenen Literatur keine weitere Literatur benutzt. Weiterhin ist mir bei der Erstellung dieser Arbeit ausschliesslich von den Personen Hilfe zuteil geworden, die in der Danksagung genannt werden.

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Torsten Kleinow
22. Mai 2002