Testing the parametric form of the volatility in continuous time diffusion models - an empirical process approach

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de
FAX: +49 234 3214 559

Mark Podolskij
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: podolski@cityweb.de

November 18, 2005

Abstract

In this paper we present two new tests for the parametric form of the variance function in diffusion processes $dX_t = b(t, X_t) + \sigma(t, X_t)dW_t$. Our approach is based on two stochastic processes of the integrated volatility. We prove weak convergence of these processes to centered processes whose conditional distributions given the process $(X_t)_{t\in[0,1]}$ are Gaussian. In the special case of testing for a constant volatility the limiting process is the standard Brownian bridge in both cases. As a consequence an asymptotic distribution free test (for the problem of testing for homoscedasticity) and bootstrap tests (for the problem of testing for a general parametric form) can easily be implemented. It is demonstrated that the new tests are more powerful with respect to Pitman alternatives than the currently available procedures for this problem. The asymptotic advantages of the new approach are also observed for realistic sample sizes in a simulation study, where the finite sample properties of a Kolmogorov-Smirnov test are investigated.

Keywords and Phrases: Specification tests, integrated volatility, bootstrap, heteroscedasticity, stable convergence, Brownian Bridge

1 Introduction

Modeling the dynamics of interest rates, stock prices exchange rates is an important problem in mathematical finance and since the seminar papers of Black and Scholes (1973) and Merton (1973) many theoretical models have been developed for this purpose. Most of these models are continuous time stochastic processes, because information arrives at financial markets in continuous time [see Merton (1990)]. A commonly used class of processes in mathematical finance for representing asset prices are Itô diffusions defined as a solution of the stochastic differential equation

$$dX_t = b\,(t, X_t)\,dt + \sigma\,(t, X_t)\,dW_t \tag{1.1}$$

where $(W_t)_t$ is a standard Brownian motion and b and σ denote the drift and volatility, respectively. Various models have been proposed in the literature for the different types of options [see e.g. Black and Scholes (1973), Vasicek (1977), Cox, Ingersoll and Ross (1985), Karatzas (1988), Constantinides (1992) or Duffie and Harrison (1993) among many others]. For a reasonable pricing of derivative assets in the context of such models a correct specification of the volatility is required and good estimates of this quantity are needed. For example, the pricing of European call options crucially depends on the functional form of the volatility [see Black and Scholes (1973)] and the same is true for other types of options [see e.g. Duffie and Harrison (1993) or Karatzas (1988) among many others].

A (correct) specification of a parametric form for the volatility has the advantage that the problem of its estimation is reduced to the determination of a low dimensional parameter. On the other hand a misspecification of drift or variance in the diffusion model (1.1) may lead to an inadequate data analysis and to serious errors in the pricing of derivative assets. Therefore several authors propose to check the postulated model by means of a goodness-of-fit test [see Ait Sahalia (1996), Corradi and White (1999), Dette and von Lieres und Wilkau (2003)]. Ait Sahalia (1996) assumes a time span approaching infinity as the sample size increases and considers the problem of testing a joint parametric specification of drift and variance, while in the other references a fixed time span is considered, where the discrete sampling interval approaches zero, and a parametric hypothesis regarding the volatility function is tested. This modeling might be more appropriate for high frequency data.

In the present paper we also consider the case of discretely observed data on a fixed time span, say $[0, 1]$, from the model (1.1) with increasing sample size. As pointed out by Corradi and White (1999) this model is appropriate for analyzing the pricing of European, American or Asian options. These authors consider the sum of the squared differences between a nonparametric and a parametric estimate of the variance function at a fixed number of points in the interval $[0, 1]$. Although this approach is attractive because of its simplicity, it has been argued by Dette and von Lieres und Wilkau (2003) that the results of the test may depend on the number and location of the points, where the parametric and nonparametric estimates are compared. Therefore these authors suggest a new test for the parametric form of the volatility in the diffusion model (1.1), which does not depend on the state x, i.e. $\sigma(t, X_t) = \sigma(t)$. The test is based on an L^2-distance between the volatility function in the model under the null hypothesis and alternative. This approach yields a consistent procedure against any (fixed) alternative, which can detect local alternatives converging to the null hypothesis at a rate $n^{-1/4}$. In the present paper an alternative test for the parametric form of the volatility function is proposed, which is based on an empirical process of the integrated volatility. Our motivation for considering functionals of stochastic processes as test statistics stems from the fact that tests of this type are more sensitive with respect to Pitman alternatives. Moreover the new tests are also applicable for testing parametric hypotheses on the volatility, which which depend on the state x.

In Section 2 we introduce some basic terminology and describe two kinds of parametric hypotheses for the volatility function. We also define two types of stochastic processes of the integrated volatility, which will be used for the construction of test statistics for these hypotheses. Section

3 contains our main results. We show convergence in probability of the stochastic processes to a random variable, which vanishes if and only if the null hypothesis is satisfied. Moreover, we also establish weak convergence of appropriately scaled processes of the integrated volatility to a centered process under the null hypothesis of a parametric form of the volatility. Consequently, the Kolmogorov-Smirnov and Cramér von Mises functional of these processes are natural test statistics. In general the limiting process is a complicated "function" of the data generating diffusion, but conditioned on the diffusion $(X_t)_{t\in[0,1]}$ it is a Gaussian process. In the problem of testing for homoscedasticity these tests are asymptotically distribution free and the limit distribution is given by a Brownian bridge. In Section 4 we study the finite sample properties of the proposed methodology and compare the new procedure with the currently available tests for the parametric form of the volatility function. For high frequency data the new tests yield a reliable approximation of the nominal level and substantial improvements with respect to power compared to the currently available procedures. Finally, all proofs and some auxiliary results are presented in an appendix.

2 Specification of a parametric form of the volatility

Let $(W_t)_{t\geq 0}$ denote a standard Brownian motion defined on an appropriate probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, P)$ with corresponding filtration $\mathcal{F}_t^W = \sigma(W_s, 0 \leq s \leq t)$ and assume that the drift and variance function in the stochastic differential equation (1.1)

$$\begin{aligned} b &: [0,1] \times \mathbb{R} \to \mathbb{R} \\ \sigma &: [0,1] \times \mathbb{R} \to \mathbb{R} \end{aligned}$$

are locally Lipschitz continuous, i.e. for every integer $M > 0$ there exists a constant K_M such that

$$|b\,(t,x) - b\,(t,y)| + |\sigma\,(t,x) - \sigma\,(t,y)| \leq K_M \; |x-y| \tag{2.1}$$

for all $t \in [0,1], x, y \in [-M, M]$, and there exists a constant K such that

$$|b\,(t,x)|^2 + |\sigma\,(t,x)|^2 \leq K^2\,(1+|x|^2) \tag{2.2}$$

for all $t \in [0,1], x \in \mathbb{R}$. It is well known that for an $\mathcal{F}_0$-measurable square integrable random variable ξ, which is independent of the Brownian motion $(W_t)_{t\in[0,1]}$, the assumptions (2.1) and (2.2) admit a unique strong solution $(X_t)_{t\in[0,1]}$ of the stochastic differential equation (1.1) with initial condition $X_0 = \xi$ which is adapted to the filtration $(\mathcal{F}_t)_{0\leq t\leq 1}$ [see e.g. Karatzas and Shreve (1991) p. 289]. The solution of the differential equation can be represented as

$$X_t = \xi + \int_0^t b\,(s, X_s)\,ds + \int_0^t \sigma\,(s, X_s)\,dW_s \quad \text{a.s.}, \tag{2.3}$$

where X_t is $\mathcal{F}_t$-measurable for all $t \in [0,1]$ and the paths $t \to X_t$ are almost surely continuous. In the literature various parametric functions have been proposed for different types of options [see e.g. Black and Scholes (1973), Vasicek (1977), Cox, Ingersoll and Ross (1985), Karatzas (1988), Constantinides (1992) or Duffie and Harrison (1993) among many others]. In principle the assumption on the volatility function in these models can be formulated in two ways that is

$$\bar{H}_0 : \sigma^2(t, X_t) = \sum_{j=1}^{d} \bar{\theta}_j \bar{\sigma}_j^2(t, X_t) \quad \forall\; t \in [0,1] \quad \text{(a.s.)}\;, \tag{2.4}$$

or

$$H_0 : \sigma(t, X_t) = \sum_{j=1}^{d} \theta_j \sigma_j(t, X_t) \quad \forall\, t \in [0,1] \quad \text{(a.s.)}\ , \tag{2.5}$$

where $\bar{\sigma}_1^2, \ldots, \bar{\sigma}_d^2$, respectively $\sigma_1, \ldots, \sigma_d$ are given and known volatility functions and $\bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_d)$, $\theta = (\theta_1, \ldots, \theta_d) \in \Theta \subset \mathbb{R}^d$ are unknown finite dimensional parameters. Throughout this paper we assume additionally that the drift and variance function satisfy a further Lipschitz condition of order $\gamma > \frac{1}{2}$, i.e.

$$\begin{aligned} &|b\,(t,x) - b\,(s,x)| + |\sigma(t,x) - \sigma(s,x)| \le L\,|t-s|^{\gamma} \\ &|b\,(t,x) - b\,(s,x)| + |\sigma_j(t,x) - \sigma_j(s,x)| \le L\,|t-s|^{\gamma}\ , \quad j = 1, \ldots, d, \end{aligned} \tag{2.6}$$

for all $s, t \in [0,1]$, $L > 0$. Note that the hypothesis (2.4) refers to the variance function σ^2 while the formulation (2.5) directly refers to the factor of the term dW_s in the stochastic differential equation (1.1). There exist in fact many models for prices of financial assets traded in continuous time, where both hypotheses are equivalent [see e.g. Vasicek (1977), Cox, Ingersoll and Ross (1985), Brennan and Schwartz (1979), Courtadon (1982), Chan, Karolyi, Longstaff and Sanders (1992)], but in general these hypotheses are not equivalent. A typical example for such a case is given by

$$\bar{H}_0 \ : \ \sigma^2(t, X_t) = \bar{\vartheta}_1 + \bar{\vartheta}_2 X_1^2, \quad \text{(a.s.)} \tag{2.7}$$

$$H_0 \ : \ \sigma(t, X_t) = \vartheta_1 + \vartheta_2 |X_t|, \quad \text{(a.s.)} \tag{2.8}$$

which is a slight generalization of the models considered in the cited references. We begin our discussion with the construction of a test statistic for the hypothesis H_0 in (2.5) and since the law of the process X depends only on σ^2 [see Revuz and Yor (1999) p. 293] we assume that the functions $\sigma, \sigma_1, \ldots, \sigma_d$ are strictly positive and linearly independent on every compact set $[0,1] \times [a,b]$, $a < b$. We assume additionally that $\sigma : [0,1] \times \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable such that for some constant $F > 0$

$$\sup_{s,t \in [0,1]} E[(\frac{\partial}{\partial y} \sigma(s, X_t))^4] < F, \ \sup_{s,t \in [0,1]} E[(\frac{\partial^2}{\partial y^2} \sigma(s, X_t))^4] < F \tag{2.9}$$

and

$$\sup_{s,t \in [0,1]} E[(\frac{\partial}{\partial y} \{\sigma_i(s, X_t)\sigma_j(s, X_t)\})^4] < F, \ \sup_{s,t \in [0,1]} E[(\frac{\partial^2}{\partial y^2} \{\sigma_i(s, X_t)\sigma_j(s, X_t)\})^4] < F \tag{2.10}$$

for all $1 \le i, j \le d$, where the differentiation in (2.9) and (2.10) is performed with respect to the second argument. Throughout this paper we assume that

$$E[|\xi|^4] < \infty, \tag{2.11}$$

and that the functions $\sigma_1, \ldots, \sigma_d$ in the linear hypothesis (2.5) satisfy the same assumptions (2.1), (2.2) and (2.9) as the volatility function σ. For the discussion of hypothesis (2.4) we need to replace σ by σ^2 in assumption (2.9), σ_i by $\bar{\sigma}_i^2$ in (2.10) and 4 by 8 in assumption (2.11).

For testing the hypothesis (2.5) we consider the following stochastic process

$$M_t := \int_0^t \Big\{ \sigma(s, X_s) - \sum_{j=1}^d \theta_j^{min} \sigma_j(s, X_s) \Big\} \, ds \;, \tag{2.12}$$

where the vector $\theta^{min} = (\theta_1^{min}, \dots, \theta_d^{min})^T$ is defined by

$$\theta^{min} := \text{argmin}_{\theta \in \mathbb{R}^d} \int_0^1 \Big\{ \sigma(s, X_s) - \sum_{j=1}^d \theta_j \sigma_j(s, X_s) \Big\}^2 \, ds \;. \tag{2.13}$$

Note that the null hypothesis in (2.5) is satisfied if and only if

$$M_t = 0 \qquad \forall \; t \in [0,1] \;\; a.s. \tag{2.14}$$

and that we use an L^2-distance to determine the best approximation of σ by linear combination of the functions $\sigma_1, \dots, \sigma_d$. This choice is mainly motivated by the sake of transparency and other distances as the L^1-distance could be used as well with an additional amount of technical difficulties.

From standard Hilbert space theory [see Achieser (1956)] we obtain

$$\theta^{min} = D^{-1} C \;, \tag{2.15}$$

where the matrix $D = (D_{ij})_{1 \le i,j \le d}$ and the vector $C = (C_1, \dots, C_d)^T$ are defined by

$$D_{ij} \;:=\; \langle \sigma_i, \sigma_j \rangle_2, \tag{2.16}$$

$$C_i \;:=\; \langle \sigma, \sigma_i \rangle_2, \tag{2.17}$$

and $\langle \cdot, \cdot \rangle_2$ denotes the standard inner product for functions $f, g : [0,1] \times \mathbb{R} \to \mathbb{R}$, that is

$$\langle f, g \rangle_2 = \int_0^1 f(t, X_t) g(t, X_t) dt \tag{2.18}$$

(here and throughout this paper we assume that the integrals exist almost surely, whenever they appear in the text). Throughout this paper we also assume the existence of the expectation

$$E[\det(D)^{-\rho}] < \infty \tag{2.19}$$

for some $\rho > 0$. The quantities in (2.16) and (2.17) can easily be estimated by

$$\hat{D}_{ij} \;:=\; \frac{1}{n} \sum_{k=1}^n \sigma_i(\frac{k}{n}, X_{\frac{k}{n}}) \sigma_j(\frac{k}{n}, X_{\frac{k}{n}}) \xrightarrow{P} D_{ij}, \tag{2.20}$$

$$\hat{C}_i \;:=\; \mu_1^{-1} n^{-\frac{1}{2}} \sum_{k=1}^n \sigma_i(\frac{k-1}{n}, X_{\frac{k-1}{n}}) |X_{\frac{k}{n}} - X_{\frac{k-1}{n}}| \xrightarrow{P} C_i \;, \tag{2.21}$$

where the symbol $\xrightarrow{P}$ means convergence in probability and μ_1 is the first absolute moment of a standard normal distribution. With the notation

$$\hat{D} \;=\; (\hat{D}_{ij})_{1 \le i,j \le d}, \tag{2.22}$$

$$\hat{C} \;=\; (\hat{C}_1, \dots, \hat{C}_d)^T \tag{2.23}$$

we obtain

$$\hat\theta^{min} := \hat D^{-1}\hat C \tag{2.24}$$

as estimate for the random variable θ^{min}. We finally introduce the stochastic process

$$\hat M_t := \hat B_t^0 - \hat B_t^T \hat D^{-1}\hat C \ , \tag{2.25}$$

as estimate of the process $(M_t)_{t\in[0,1]}$ defined in (2.12), where the quantities $\hat B_t^0$ and $\hat B_t = (\hat B_t^1, \dots, \hat B_t^d)^T$ are given by

$$\hat B_t^0 \ := \ \mu_1^{-1} n^{-\frac{1}{2}} \sum_{k=1}^{\lfloor nt \rfloor} |X_{\frac{k}{n}} - X_{\frac{k-1}{n}}|, \tag{2.26}$$

$$\hat B_t^i \ := \ \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \sigma_i(\frac{k}{n}, X_{\frac{k}{n}}), \qquad 1 \le i \le d, \tag{2.27}$$

respectively. Note that $M_t = 0$ (a.s.) for all $t \in [0,1]$ if and only if the null hypothesis H_0 is satisfied. Consequently, it is intuitively clear that the rejection of the null hypothesis for large values of

$$\sup_{t\in[0,1]} |\hat M_t| \ , \quad \int_0^1 |\hat M_t| dt \ \text{ or } \quad \int_0^1 |\hat M_t|^2 dt$$

yields a consistent test for the above problem.

Before we make these arguments more rigorous we briefly present the corresponding testing procedure for the hypothesis (2.4). In this case the analogue of the stochastic process M_t is defined by

$$N_t := \int_0^t \Big\{ \sigma^2(s, X_s) - \sum_{j=1}^d \bar\theta_j^{min} \bar\sigma_j^2(s, X_s) \Big\} \ ds \ , \tag{2.28}$$

where the random variable $\bar\theta^{min} = (\bar\theta_1^{min}, \dots, \bar\theta_d^{min})^T$ is given by

$$\bar\theta^{min} := \operatorname{argmin}_{\bar\theta\in\mathbb{R}^d} \int_0^1 \Big\{ \sigma^2(s, X_s) - \sum_{j=1}^d \bar\theta_j \bar\sigma_j^2(s, X_s) \Big\}^2 \ ds \ . \tag{2.29}$$

The nonobservable stochastic process can easily be estimated from the available data by

$$\hat N_t := \bar B_t^0 - \bar B_t^T \bar D^{-1} \bar C \ , \tag{2.30}$$

where

$$\bar D_{ij} \ := \ \frac{1}{n} \sum_{k=1}^n \bar\sigma_i^2(\frac{k}{n}, X_{\frac{k}{n}}) \bar\sigma_j^2(\frac{k}{n}, X_{\frac{k}{n}}), \quad i, j = 1, \dots, d \tag{2.31}$$

$$\bar C_i \ := \ \sum_{k=2}^n \bar\sigma_i^2(\frac{k-1}{n}, X_{\frac{k-1}{n}}) |X_{\frac{k}{n}} - X_{\frac{k-1}{n}}|^2, \quad i = 1, \dots, d \tag{2.32}$$

and the quantities $\bar{B}_t^0$ and $\bar{B}_t = (\bar{B}_t^1, \dots \bar{B}_t^d)^T$ are given by

$$\bar{B}_t^0 := \sum_{k=1}^{\lfloor nt \rfloor} |X_{\frac{k}{n}} - X_{\frac{k-1}{n}}|^2, \tag{2.33}$$

$$\bar{B}_t^i := \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \bar{\sigma}_i^2(\frac{k}{n}, X_{\frac{k}{n}}), \qquad 1 \leq i \leq d. \tag{2.34}$$

In the following section we investigate the stochastic properties of the processes

$$(\sqrt{n}(\hat{M}_t - M_t))_{t\in[0,1]}$$

and

$$(\sqrt{n}(\hat{N}_t - N_t))_{t\in[0,1]}.$$

In particular, we will prove weak convergence of these processes to centered processes, which are conditioned on the process $(X_t)_{t\in[0,1]}$ Gaussian processes. This is the basic result for the application of these processes in the problem of testing for the parametric form of the volatility in a continuous time diffusion model. The reason for considering both processes is twofold. On the one hand the weak convergence of the process $\sqrt{n}(\hat{M}_t - M_t)$ can be established under weaker assumptions on the model (1.1). On the other hand the statistic $\sqrt{n}(\hat{N}_t - N_t)$ can easily be extended to vector-valued diffusions [see Remark 3.7].

3 Main results

For the sake of brevity we mainly restrict ourselves to a detailed derivation of the stochastic properties of the process $\hat{M}_t$. The corresponding results for the process $\hat{N}_t$ can be obtained by similar arguments and the main statements are given at the end of this section for the sake of completeness. We begin our discussion with two auxiliary results regarding the estimators $\hat{D}$ and $\hat{B}_t^i$ defined in (2.22) and (2.27), which are also of own interest. Throughout this paper μ_i denotes the ith absolute moment of a standard normal distribution ($i = 1, 2$). Our first results clarify the order of difference between the empirical quantities $\hat{C}_i$, $\hat{B}_t^i$, $\hat{D}_{ij}$ and their theoretical counterparts C_i, B_t^i, D_{ij}, respectively.

Lemma 3.1. *If the assumptions stated in Section 2 are satisfied we have*

$$\hat{B}_t^i - \int_0^t \sigma_i(s, X_s)ds = o_p(n^{-\frac{1}{2}}) \qquad 1 \leq i \leq d$$

$$\hat{D} - D = o_p(n^{-\frac{1}{2}})$$

Throughout this paper the symbol

$$X_n \xrightarrow{\mathcal{D}_{st}} X$$

means that the sequence of random variables converges stably in law. Recall that a sequence of d-dimensional random variables $(X_n)_{n\in\mathbb{N}}$ converges stably in law with limit X, defined on an appropriate extension $(\Omega', \mathcal{F}', P')$ of a probability space $(\Omega, \mathcal{F}, P)$, if and only if for any $\mathcal{F}$-measurable and bounded random variable Y and any bounded and continuous function g the convergence

$$\lim_{n\to\infty} E[Yg(X_n)] \ = \ E[Yg(X)]$$

holds. This is obviously a slightly stronger mode of convergence than convergence in law [see Renyi (1963), Aldous and Eagleson (1978) for more details on stable convergence]. The following Lemma shows that the random variables $\hat{B}_t^0$ and $\hat{C}_i$ defined in (2.26) and (2.23) converge stably in law if they are appropriately normalized.

Lemma 3.2. *If the assumptions stated in Section 2 are satisfied we have for any $t_1, \dots, t_k \in [0,1]$*

$$\sqrt{n}\begin{pmatrix} \hat{B}_{t_1}^0 - <\sigma, 1>_2^{t_1} \\ \vdots \\ \hat{B}_{t_k}^0 - <\sigma, 1>_2^{t_k} \\ \hat{C}_1 - C_1 \\ \vdots \\ \hat{C}_d - C_d \end{pmatrix} \ \xrightarrow{\mathcal{D}_{st}} \ \mu_1^{-1}\sqrt{\mu_2-\mu_1^2}\int_0^1 \Sigma_{t_1,\dots t_k}^{\frac{1}{2}}(s, X_s)\ dW_s' \ ,$$

where W' denotes a $(d+k)$-dimensional Brownian motion, which is independent of the σ-field $\mathcal{F}$, and the matrix $\Sigma_{t_1,\dots t_k}(s, X_s)$ is defined by

$$\Sigma_{t_1,\dots t_k}(s,X_s) = \begin{pmatrix} v_{11}(s,X_s) & \cdots & v_{1k}(s,X_s) & w_{11}(s,X_s) & \cdots & w_{1d}(s,X_s) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{k1}(s,X_s) & \cdots & v_{kk}(s,X_s) & w_{k1}(s,X_s) & \cdots & w_{kd}(s,X_s) \\ w_{11}(s,X_s) & \cdots & w_{k1}(s,X_s) & s_{11}(s,X_s) & \cdots & s_{1d}(s,X_s) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w_{1d}(s,X_s) & \cdots & w_{kd}(s,X_s) & s_{d1}(s,X_s) & \cdots & s_{dd}(s,X_s) \end{pmatrix} \tag{3.1}$$

with

$$\begin{aligned} v_{ij}(s,X_s) &= \sigma^2(s,X_s) 1_{[0,t_i\wedge t_j)}(s) \qquad 1\le i\le k \\ w_{ij}(s,X_s) &= \sigma_j(s,X_s)\sigma^2(s,X_s)1_{[0,t_i)}(s) \qquad 1\le i\le k\ ,\ 1\le j\le d \\ s_{ij}(s,X_s) &= \sigma_i(s,X_s)\sigma_j(s,X_s)\sigma^2(s,X_s) \qquad 1\le i,j\le d \end{aligned}$$

and

$$<\sigma, 1>_2^t = \int_0^t \sigma(s,X_s)ds.$$

Note that the matrix $\Sigma_{t_1,\dots t_k}(s,X_s)$ defined in (3.1) can be represented as

$$\Sigma_{t_1,\dots,t_k}^{1/2}(s,X_s) = \frac{g_{t_1,\dots,t_k}(s,X_s)g_{t_1,\dots,t_k}(s,X_s)^T}{\sqrt{g_{t_1,\dots,t_k}(s,X_s)^T g_{t_1,\dots,t_k}(s,X_s)}}, \tag{3.2}$$

where the vector $g_{t_1,\dots,t_k}(s, X_s)$ is defined by

$$g_{t_1,\dots,t_k}(s, X_s) = (\sigma(s, X_s)I_{[0,t_1)}(s), \dots, \sigma(s, X_s)I_{[0,t_k)}(s), \sigma_1(s, X_s)\sigma(s, X_s), \dots, \sigma_d(s, X_s)\sigma(s, X_s))^T. \tag{3.3}$$

Now we state one of our main results. For this purpose we define the process

$$A_n(t) = \sqrt{n}(\hat{M}_t - M_t) \quad (t \in [0, 1]) \tag{3.4}$$

and obtain the following result.

Theorem 3.3. *If the assumptions given in Section 2 are satisfied, then the process $(A_n(t))_{t\in[0,1]}$ defined in (3.4) converges weakly on $D[0, 1]$ to a process $(A(t))_{t\in[0,1]}$, which is Gaussian conditioned on the σ-field $\mathcal{F}$. Moreover, the finite dimensional conditional distributions of the limiting process $(A(t_1), \dots A(t_k))^T$ are uniquely determined by the conditional covariance matrix*

$$\mu_1^{-2}(\mu_2 - \mu_1^2) \;\; V \int_0^1 \Sigma_{t_1,\dots t_k}(s, X_s) \; ds \; V^T \; , \tag{3.5}$$

where the $k \times (d + k)$-dimensional matrix V is defined by

$$V = (I_k|\tilde{V}) \qquad \tilde{V} = -\begin{pmatrix} B_{t_1}^T D^{-1} \\ \vdots \\ B_{t_k}^T D^{-1} \end{pmatrix}, \tag{3.6}$$

and $I_k \in \mathbb{R}^{k\times k}$ denotes the identity matrix.

Note that the identity $M_t \equiv 0$ holds (a.s.) for all $t \in [0, 1]$ if and only if the null hypothesis in (2.5) is satisfied, and consequently the null hypothesis is rejected for large values of a functional of the process $(\sqrt{n}\hat{M}_t)_{t\in[0,1]}$. For example, in the case of the Kolmogorov-Smirnov statistic

$$K_n = \sqrt{n} \sup_{t\in[0,1]} |\hat{M}_t|, \tag{3.7}$$

it follows from Theorem 3.3 that (under H_0)

$$K_n \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |A(t)|,$$

where the symbol $\xrightarrow{\mathcal{D}}$ denotes the weak convergence and the process $(A(t))_{t\in[0,1]}$ is defined in Theorem 3.3. In general, even under the null hypothesis H_0, the distribution of this process is rarely available and depends on the full process $(X_t)_{t\in[0,1]}$. However, conditioned on the process $(X_t)_{t\in[0,1]}$ the process $(A(t))_{t\in[0,1]}$ is Gaussian. Moreover, in the important case of testing for a constant volatility, i.e. $d = 1, \sigma_1(t, X_t) = 1$, the limit distribution of the process $(A_n(t))_{t\in[0,1]}$ is surprisingly simple.

Corollary 3.4. *Assume that the assumptions stated in Section 2 are satisfied and that the hypothesis $H_0 : \sigma(t, X_t) = \sigma$ for some $\sigma > 0$ has to be tested (that is $d = 1, \sigma_1(t, X_t) = 1$ in (2.5)). Under the null hypothesis the process $(A_n(t))_{t\in[0,1]}$ converges weakly on $D[0, 1]$ to the process*

$$(\mu_1^{-1}\sqrt{\mu_2 - \mu_1^2}\sigma B_t)_{t\in[0,1]} \ ,$$

where B_t denotes a Brownian bridge.

We now briefly consider the corresponding results for testing the hypothesis (2.4) based on the stochastic process $\hat{N}_t$ defined in (2.30). For this recall the definition of N_t in (2.28) and define

$$\bar{A}_n(t) = \sqrt{n}(\hat{N}_t - N_t). \tag{3.8}$$

The following result is proved by similar arguments as presented for the proof of Theorem 3.3 in the Appendix.

Theorem 3.5. *If the assumptions given in Section 2 are satisfied, then the process $(\bar{A}_n(t))_{t\in[0,1]}$ in (3.8) converges weakly on $D[0, 1]$ to a process $(\bar{A}(t))_{t\in[0,1]}$, which is Gaussian conditioned on the σ-field $\mathcal{F}$. Moreover, the finite dimensional conditional distributions of the limiting process $(\bar{A}(t_1), \ldots \bar{A}(t_k))^T$ are uniquely determined by the conditional covariance matrix*

$$2 \ \bar{V} \int_0^1 \bar{\Sigma}_{t_1,\ldots t_k}(s, X_s) \ ds \ \bar{V}^T \ , \tag{3.9}$$

where the $k \times (d + k)$-dimensional matrix $\bar{V}$ is defined by

$$\bar{V} = (I_k|\tilde{U}) \qquad \tilde{U} = -\begin{pmatrix} \bar{B}_{t_1}^T \bar{D}^{-1} \\ \vdots \\ \bar{B}_{t_k}^T \bar{D}^{-1} \end{pmatrix},$$

and the matrix $\bar{\Sigma}_{t_1,\ldots t_k}$ is given by

$$\bar{\Sigma}_{t_1,\ldots t_k}(s, X_s) = \begin{pmatrix} \bar{v}_{11}(s, X_s) & \cdots & \bar{v}_{1k}(s, X_s) & \bar{w}_{11}(s, X_s) & \cdots & \bar{w}_{1d}(s, X_s) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \bar{v}_{k1}(s, X_s) & \cdots & \bar{v}_{kk}(s, X_s) & \bar{w}_{k1}(s, X_s) & \cdots & \bar{w}_{kd}(s, X_s) \\ \bar{w}_{11}(s, X_s) & \cdots & \bar{w}_{k1}(s, X_s) & \bar{s}_{11}(s, X_s) & \cdots & \bar{s}_{1d}(s, X_s) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \bar{w}_{1d}(s, X_s) & \cdots & \bar{w}_{kd}(s, X_s) & \bar{s}_{d1}(s, X_s) & \cdots & \bar{s}_{dd}(s, X_s) \end{pmatrix} \tag{3.10}$$

with

$$\begin{aligned} \bar{v}_{ij}(s, X_s) &= \sigma^4(s, X_s)1_{[0,t_i\wedge t_j)}(s) \qquad 1 \le i \le k \\ \bar{w}_{ij}(s, X_s) &= \bar{\sigma}_j^2(s, X_s)\sigma^4(s, X_s)1_{[0,t_i)}(s) \qquad 1 \le i \le k \ , \ 1 \le j \le d \\ \bar{s}_{ij}(s, X_s) &= \bar{\sigma}_i^2(s, X_s)\bar{\sigma}_j^2(s, X_s)\sigma^4(s, X_s) \qquad 1 \le i, j \le d \end{aligned}$$

and

$$\bar{B}_t = \Big(\int_0^t \bar{\sigma}_1^2(s, X_s)ds, \ldots, \int_0^t \bar{\sigma}_d^2(s, X_s)ds \Big)^T.$$

For a construction of a test for the hypothesis (2.4) we calculate a Kolmogorov-Smirnov statistic and reject the null hypothesis for large values. We conclude this section with an investigation of the stochastic properties of the tests with respect to local alternatives. For the sake of brevity we restrict ourselves to the problem of testing for homoscedasticity, that is

$$H_0 : \sigma(t, X_t) = \sigma \quad \text{a.s.} \tag{3.11}$$

for some $\sigma > 0$ and local alternatives of the form

$$H_1^{(n)} : \sigma(t, X_t) = \sigma + \gamma_n h(t, X_t), \tag{3.12}$$

where h is a positive function and γ_n is a positive sequence converging to 0 at an appropriate rate. The problem of testing more general hypotheses can be treated exactly in the same way. The consideration of the null hypothesis of homoscedasticity additionally allows a comparison of the two approaches based on Theorem 3.3 and 3.5, because in the special case $d = 1$ the hypotheses (2.4) and (2.5) are in fact equivalent. Note that the process corresponding to the hypothesis $H_0 : \sigma(t, X_t) = \sigma$ is given by

$$\hat{M}_t = \mu_1^{-1} n^{-\frac{1}{2}} \Big(\sum_{k=2}^{\lfloor nt \rfloor} |X_{\frac{k}{n}} - X_{\frac{k-1}{n}}| - \frac{\lfloor nt \rfloor}{n} \sum_{k=2}^{n} |X_{\frac{k}{n}} - X_{\frac{k-1}{n}}| \Big). \tag{3.13}$$

Similarly, if the process defined by (2.30) is used we have (in the case $d = 1, \sigma_1(s, X_s) = 1$)

$$\hat{N}_t = \sum_{k=2}^{\lfloor nt \rfloor} (X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^2 - \frac{\lfloor nt \rfloor}{n} \sum_{k=2}^{n} (X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^2. \tag{3.14}$$

We finally also introduce the statistic proposed by Dette and von Lieres und Wilkau (2003) for the hypothesis (3.11), that is

$$\hat{G} = \sqrt{n} \Big\{ \frac{n}{3} \sum_{k=2}^{n} (X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^4 - \Big(\sum_{k=2}^{n} (X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^2 \Big)^2 \Big\} \tag{3.15}$$

The following results specify the asymptotic behaviour of the processes $\hat{M}_t, \hat{N}_t$ and the statistic $\hat{G}$ under local alternatives of the form (3.12).

Theorem 3.6. *Consider local alternatives of the form (3.12).*

(a) If the assumptions of Theorem 3.3 are satisfied, $\gamma_n = n^{-1/2}$, then the processes $(\sqrt{n}\hat{M}_t)_{t \in [0,1]}$ defined in (3.13) converges weakly on $D[0, 1]$ to the process $(\mu_1^{-1}\sqrt{\mu_2 - \mu_1^2}\sigma B_t + R_t)_{t \in [0,1]}$, where B_t denotes a Brownian bridge, the process R_t is given by

$$R_t = \Big(\int_0^t h(s, X_s)ds - t \int_0^1 h(s, X_s)ds \Big), \tag{3.16}$$

and the processes $(B_t)_{t \in [0,1]}$ and $(R_t)_{t \in [0,1]}$ are stochastically independent.

(b) *If the assumptions of Theorem 3.5 are satisfied,* $\gamma_n = n^{-1/2}$, *then the process* $(\sqrt{n}\hat{N}_t)_{t\in[0,1]}$ *defined by (3.14) converges weakly on* $D[0,1]$ *to the process* $(\sqrt{2}\sigma^2 B_t + 2\sigma R_t)_{t\in[0,1]}$, *where* B_t *denotes a Brownian bridge, the process* R_t *is given in (3.16) and the processes* $(B_t)_{t\in[0,1]}$ *and* $(R_t)_{t\in[0,1]}$ *are stochastically independent.*

(c) *If the assumptions of Theorem 3.5 are satisfied,* $\gamma_n = n^{-1/4}$, *then it follows for the random variable* $\hat{G}$ *defined in (3.15)*

$$\sqrt{n}\hat{G} \xrightarrow{\mathcal{D}} Z + 4\sigma^2\Big(\int_0^1 h^2(s,X_s)ds - \Big(\int_0^1 h(s,X_s)ds\Big)^2\Big) ,$$

where the random variables $Z \sim \mathcal{N}(0, \frac{8}{3}\sigma^8)$ *and* $(\int_0^1 h^2(s,X_s)ds - (\int_0^1 h(s,X_s)ds)^2)$ *are stochastically independent*

Note that it follows from Theorem 3.6 that goodness-of-fit tests based on the processes (3.13) and (3.14) are more powerful with respect to Pitman alternatives of the form (3.12) than the test which rejects the null hypothesis (3.11) for large values of the statistic $\hat{G}$. Moreover, Theorem 3.6 also shows that there will be no substantial differences between the tests based on the stochastic processe $\hat{M}_t$ and $\hat{N}_t$ with respect to power for local alternatives of the form (3.12) (besides that the asymptotic theory for the latter requires slightly stronger assumptions). We finally note again that a similar statement can be shown for the general hypotheses (2.4) and (2.5).

Remark 3.7. It is worthwhile to mention that the process $(N_t)_{t\in[0,1]}$ can easily be generalized to p-dimensional diffusions. For this assume that the drift function b in (1.1) is a p-dimensional vector, the volatiltiy is a $p \times q$ matrix, and that the underlying Brownian motion is q-dimensional. For functions $f, g : [0,1] \times \mathbb{R}^p \to \mathbb{R}^{p\times p}$ we define the (random) inner product

$$\langle f, g\rangle_2 = \int_0^1 \text{trace } (f(s,X_s)g(s,X_s)^T)ds$$

and denote by

$$\bar{\theta}^{min} := \text{argmin}_{\bar{\theta}\in\mathbb{R}^d} \langle \sigma\sigma^T - \sum_{j=1}^d \bar{\theta}_j\bar{\sigma}_j\bar{\sigma}_j^T, \sigma\sigma^T - \sum_{j=1}^d \bar{\theta}_j\bar{\sigma}_j\bar{\sigma}_j^T\rangle_2$$

Note that $\bar{\theta}^{min}$ can be written as

$$\bar{\theta}^{min} = \bar{D}^{-1}\bar{C} ,$$

where the elements of the matrix $\bar{D} = (\bar{D}_{ij})_{1\le i,j\le d}$ and the vector $\bar{C} = (\bar{C}_1, \dots, \bar{C}_d)^T$ are defined as

$$\bar{D}_{ij} := \langle \bar{\sigma}_i\bar{\sigma}_i^T, \bar{\sigma}_j\bar{\sigma}_j^T\rangle_2, \quad (3.17)$$

$$\bar{C}_i := \langle \sigma\sigma^T, \bar{\sigma}_i\bar{\sigma}_i^T\rangle_2, \quad (3.18)$$

Finally we define the $p \times p$ process

$$N_t := \int_0^t \left\{ \sigma(s, X_s)\sigma(s, X_s)^T - \sum_{j=1}^{d} \bar{\theta}_j^{min} \bar{\sigma}_j(s, X_s)\bar{\sigma}_j(s, X_s)^T \right\} ds \ , \tag{3.19}$$

then it is easy to see that the null hypothesis $\sigma\sigma^T = \sum_{j=1}^d \bar{\theta}_j \bar{\sigma}_j \bar{\sigma}_j^T$ is valid if and only if $M_t \equiv 0$ $\forall\ t \in [0, 1]$ (a.s.). This process is now estimated in an obvious way. For example, the first term in (3.19) can be approximated by the data by

$$\sum_{i=1}^{[nt]} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})(X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^T \xrightarrow{P} \int_0^t \sigma(s, X_s)\sigma(s, X_s)^T ds,$$

and the other terms are treated similarly. Consequently under appropriate assumptions on the drift b the volatility σ and the functions $\bar{\sigma}_1, \ldots \bar{\sigma}_d$ an analogue of Theorem 3.5 is available for the vector-valued diffusions.

4 Finite sample properties

In this section we investigate the finite sample properties of Kolmogorov-Smirnov tests based on the processes $(\hat{M}_t)_{t\in[0,1]}$ and $(\hat{N}_t)_{t\in[0,1]}$. We also compare these tests with the test, which was recently proposed by Dette, Podolskij and Vetter (2005) for the hypotheses of the form (2.4). We begin with a study of the quality of approximation by a Brownian bridge in the case of testing for homoscedasticity. In the second part of this section we briefly investigate the performance of a parametric bootstrap procedure for the problem of testing more general hypotheses and present an example analyzing exchange rate data. Here and throughout this section all reported results are based on 1000 simulation runs.

4.1 Testing for homoscedasticity

Recall from Corollary 3.4 that under the null hypothesis

$$M^{(n)} := \sqrt{n} \sup_{t\in[0,1]} |\frac{\hat{M}_t}{\hat{\beta}}| \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |B_t| \ , \tag{4.1}$$

where $(B_t)_{t\in[0,1]}$ denotes a Brownian bridge and $\hat{\beta}$ is given by

$$\hat{\beta} = \mu_1^{-2} \sqrt{\mu_2 - \mu_1^2} n^{-1/2} \sum_{i=2}^{n} |X_{\frac{i}{n}} - X_{\frac{i-1}{n}}|.$$

Similarly, it follows from Theorem 3.5 that

$$N^{(n)} := \sqrt{n} \sup_{t\in[0,1)} |\frac{\hat{N}_t}{\hat{\gamma}}| \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |B_t| \ , \tag{4.2}$$

where the process $(\hat{N}_t)_{t\in[0,1]}$ is defined in (3.14) and $\hat{\gamma}$ is given by

$$\hat{\gamma} = \sqrt{2}\sum_{i=1}^{n}(X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2.$$

The null hypothesis (3.11) of a constant volatility in the stochastic differential equation is now rejected if $M^{(n)}$ or $N^{(n)}$ exceed the corresponding quantile of the distribution of the maximum of a Brownian bridge on the interval $[0,1]$. In Table 4.1 we show the approximation of the nominal level of these tests for sample sizes $n = 100, 200, 500$. The data was generated according to the diffusion model (1.1) with $\sigma = 1$ and various drift functions $b(t,x)$.

	$M^{(n)}$						$N^{(n)}$					
n	100		200		500		100		200		500	
$b \backslash \alpha$	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0	.044	.098	.045	.099	0.046	0.093	.047	.082	.041	.077	.041	.098
2	.045	.077	.050	.099	0.044	0.106	.032	.078	.034	.078	.048	.092
x	.045	.095	.044	.081	0.042	0.091	.033	.089	.033	.081	.041	.082
$2-x$	.045	.086	.039	.092	0.053	0.092	.041	.075	.039	.069	.051	.090
tx	.042	.081	.050	.100	0.042	0.092	.037	.068	.042	.089	.051	.089

Table 4.1. *Approximation of the nominal level of the tests, which reject the null hypothesis of homoscedasticity for large values of the statistics $M^{(n)}$ and $N^{(n)}$. The critical values are obtained by the asymptotic law (4.1) and (4.2), respectively.*

We observe a reasonable approximation of the nominal level in most cases. The statistic $M^{(n)}$ usually yields a more precise approximation of the nominal level than the statistic $N^{(n)}$, which turns out to be slightly conservative for small sample sizes. We now investigate the power of both tests in the problem of testing for homoscedasticity. For the sake of comparison we consider the same scenario as in Dette, von Lieres und Wilkau (2003) who proposed the test based on the statistic $\hat{G}$ in (3.15) for the problem of checking homoscedasticity. Following these authors we chose the volatitlity functions

$$\sigma(t,x) = 1 + x; 1 + \sin(5x); 1 + xe^t; 1 + x\sin(5t); 1 + xt.$$

In Table 4.2 we present the corresponding rejection probabilities for the sample sizes $n = 100, 200$ and 500. The results are directly comparable with the results in Table 3 of Dette and von Lieres und Wilkau (2003) for the corresponding test based on the statistic (3.15). From Thoereom 3.6 we expect some improvement in local power with respect to Pitman alternatives by the new procedure and these theoretical advantages are impressively reflected in our simulation study. We observe a substantial increase in power for the new tests. In most cases the improvement is at least approximately 15% and there are many cases, for which the power of the new test with 200 observations already exceeds the power of the test of Dette, von Lieres und Wilkau (2003) for 500 observations.

	$M^{(n)}$						$N^{(n)}$					
n	100		200		500		100		200		500	
σ/α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$1+x$	.857	.888	.920	.944	.972	.976	.830	.863	.929	.949	.969	.978
$1+\sin(5x)$	.1.000	1.000	1.000	1.000	1.000	1.000	.998	.999	1.000	1.000	1.000	1.000
$1+xe^t$	.972	.981	.990	.993	0.999	0.999	.947	.968	.989	.996	0.998	0.998
$1+x\sin(5t)$	.781	.843	.890	.911	.962	.970	.776	.824	.882	.914	.961	.974
$1+tx$	.744	.797	.866	.891	.941	.955	.743	.780	.851	.878	.950	.987

Table 4.2. *Rejection probabilities of the tests, which reject the null hypothesis of homoscedasticity for large values of thes tatistics $M^{(n)}$ and $N^{(n)}$. The critical values are obtained by the asymptotic law (4.1) and (4.2), respectively.*

4.2 Testing for the parametric form of the volatility

As pointed out previously, for a general null hypothesis the asymptotic distribution of the processes depends on the underlying diffusion $(X_t)_{t\in[0,1]}$ and cannot be used for the calculation of critical values (except in the problem of testing for homoscedasticity). However, conditional on $(X_t)_{t\in[0,1]}$ the limiting processes in Theorem 3.3 and 3.5 are Gaussian and this suggests that the parametric bootstrap can be used to obtain critical values. In this paragraph we will investigate the finite sample performance of this approach. We explain this procedure for the process $(\hat{M}_t)_{t\in[0,1]}$, the corresponding bootstrap test for the process $(\hat{N}_t)_{t\in[0,1]}$ is obtained similarly. In a first step the least squares estimator $\hat{\theta}^{\min} = (\hat{\theta}_1^{\min}, \ldots, \hat{\theta}_d^{\min})^T$ defined in (2.24) is determined. Then the process $\sqrt{n}\hat{M}_t$ is standardized by an estimate of the conditional variance

$$s_t^2 = \mu_1^{-2}(\mu_2 - \mu_1^2)(1, -B_t^T D^{-1}) \int_0^1 \Sigma_t(s, X_s)ds \; (1, -B_t^T D^{-1})^T$$

For the corresponding estimate, say $\hat{s}_t^2$, the random variables B_t und D are replaced by their empirical counterparts defined in Section 3, and the random elements in the matrix $\Sigma_t(s, X_s)$ defined in (3.1) are replaced by the statistics

$$\sum_{k=1}^{[nt]} (X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^2 \xrightarrow{P} \int_0^t \sigma^2(s, X_s)ds$$

$$\sum_{k=1}^{[nt]} \sigma_i^2(\frac{k-1}{n}, X_{\frac{k-1}{n}})(X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^2 \xrightarrow{P} \int_0^t \sigma_i^2(s, X_s)\sigma^2(s, X_s)ds$$

$$\sum_{k=1}^{n} \sigma_j^2(\frac{k-1}{n}, X_{\frac{k-1}{n}})\sigma_i^2(\frac{k-1}{n}, X_{\frac{k-1}{n}})(X_{\frac{k}{n}} - X_{\frac{k-1}{n}})^2 \xrightarrow{P} \int_0^1 \sigma_j^2(s, X_s)\sigma_i^2(s, X_s)\sigma^2(s, X_s)ds$$

This yields the (standardized) Kolmogorov-Smirnov statistic

$$Z_n = \sup_{t\in[0,1]} \Big| \frac{\sqrt{n}\hat{M}_t}{\hat{s}_t} \Big| \tag{4.3}$$

In a second step data $X_{i/n}^{*(j)}$ $(i = 1, \ldots, n;\ j = 1, \ldots B)$ from the stochastic differential equation (1.1) with $b(t, x) \equiv 0$ and $\sigma(t, x) = \sum_{j=1}^{d} \hat{\theta}_j^{\min} \sigma_j(t, x)$ are generated [note that this choice corresponds to the null hypothesis (2.5)]. These data are used to calculate the bootstrap analogues

$$Z_n^{*(1)}, \ldots, Z_n^{*(B)}$$

of the statistic Z_n defined in (4.3). Finally the value of the statistic Z_n is compared with the corresponding quantiles of the bootstrap distribution.

We have investigated the performance of this resampling procedure for the problem of testing various linear hypotheses, where the volatility function depends on the variable x. The sample sizes are again $n = 100, 200, 500$ and $B = 500$ bootstrap replications are used for the calculation of the critical values. In particular we compare the two procedures based on $(\hat{M}_t)_{t\in[0,1]}$ and $(\hat{N}_t)_{t\in[0,1]}$ for testing the hypothesis

$$\text{(4.4)} \qquad \begin{aligned} \bar{H}_0 &: \sigma^2(t, x) = \bar{\theta} x^2 \\ H_0 &: \sigma(t, x) = \theta x \end{aligned}$$

In Table 4.3 we display the simulated level of the parametric bootstrap tests for various drift functions. We observe a better approximation of the nominal level by the test based on the process $(N_t)_{t\in[0,1]}$, in particular for small sample sizes. The Kolmogorov-Smirnov test based on the process $(\hat{M}_t)_{t\in[0,1]}$ yields a reliable approximation of the nominal level for sample sizes larger than 200, while the statistic based on the process $\hat{N}_t$ can already be recommended for $n = 100$. The results for sample size $n = 200$, 500 demonstrate that for high frequency data as considered in this paper both tests yield a reliable approximation of the nominal level. In Table 4.4 we show the simulated rejection probabilities for the case $b(t, x) = 2 - x$ and the alternatives

$$\sigma^2(t, x) = 1 + x^2, 1, 5|x|^{3/2}, 5|x|, (1 + x)^2$$

	$\hat{M}_t$						$\hat{N}_t$					
n	100		200		500		100		200		500	
b/α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
0	.125	.195	.081	.135	.074	.133	.052	.110	.050	.099	.061	.114
2	.076	.126	.066	.107	.048	.096	.084	.132	.079	.124	.069	.117
x	.094	.148	.071	.128	.048	.100	.069	.129	.057	.117	.054	.100
$2-x$	.082	.133	.065	.112	.063	.117	.048	.088	.043	.101	.043	.097
xt	.103	.166	.068	.130	.062	.116	.049	.103	.046	.099	.063	.105

Table 4.3. *Simulated level of the bootstrap test for the hypothesis (4.4) based on the standardized Kolmogorov-Smirnov functional of the processes* $(\hat{M}_t)_{t\in[0,1]}$ *and* $(\hat{N}_t)_{t\in[0,1]}$.

	$\hat{M}_t$						$\hat{N}_t$					
n	100		200		500		100		200		500	
σ^2/α	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$1+x^2$	.516	.587	.652	.720	.831	.885	.352	.467	.502	.627	.752	.828
1	.809	.862	.933	.955	.996	.998	.739	.838	.917	.960	.995	.997
$5\|x\|^{3/2}$	.371	.516	.511	.638	.743	.838	.252	.310	.388	.534	.485	.598
$5\|x\|$	.917	.882	.954	.970	.994	.997	.439	.551	.731	.858	.898	.949
$(1+x)^2$	.749	.815	.874	.920	.960	.976	.387	.500	.537	.751	.883	.934

Table 4.4. *Simulated rejection probabilities of the bootstrap test for the hypothesis (4.4) based on the standardized Kolmogorov-Smirnov functional of the processes* $(\hat{M}_t)_{t\in[0,1]}$ *and* $(\hat{N}_t)_{t\in[0,1]}$.

Note that the Kolmogorov-Smirnov test based on the process $(\hat{M}_t)_{t\in[0,1]}$ is substantially more powerful than the test based on the process $(\hat{N}_t)_{t\in[0,1]}$ which uses the squared differences. This superiority is partially bought by a worse approximation of the nominal level for smaller sample sizes [see the results for $n = 100$ and $n = 200$ in Table 4.3]. However, in the case $b(t,x) = 2 - x$ and $n = 200, 500$ both tests keep approximately their level, but the test based on $(\hat{M}_t)_{t\in[0,1]}$ is still substantially more powerful. Thus for high frequency data we recommend the application of the Kolmogorov-Smirinov test based on the process $(\hat{M}_t)_{t\in[0,1]}$.

It is also of interest to compare the power of the new tests with a bootstrap test, which was recently proposed by Dette, Podolskij and Vetter (2005) and is based on an estimate of the L^2-distance

$$M^2 = \min_{\theta_1,\dots,\theta_d} \int_0^1 \Big\{\sigma^2(t,X_t) - \sum_{j=1}^d \theta_j \sigma_j^2(t,X_t)\Big\}^2 dt.$$

Because this test yield a rather accurate approximation of the nominal level [see Table 1 in this reference] we mainly consider the Kolmogorov-Smirnov test based on the process $(\hat{N}_t)_{t\in[0,1]}$ in our comparison. The results in the right part of Table 4.4 are directly comparable with the results displayed in Tabel 4 of Dette, Podolskij and Vetter (2005). We observe that in most cases the new Kolmogorov-Smirnov test yields a substantial improvement with respect to power. For the sample size $n = 100$ the procedure is more powerful for detecting the alternatives $\sigma^2(t,x) = 1;\ 1 + x^2$ and less powerful for the alternative $\sigma^2(t,x) = 5|x|^{3/2}$. For the remaining two alternatives the new test yields slightly better results. One the other hand the asymptotic advantages of the Kolmogorov-Smirnov test become more visible for larger sample sizes ($n = 200$, $n = 500$), where it outperforms the test based on the L^2-distance in all cases. For example, the power of the test of Dette, Podolskij and Vetter (2005) with $n = 500$ observations is already achieved by the Kolmogorov-Smirnov test with $n = 200$ observations. Except for the alternative $\sigma^2(t,x) = 5|x|^{3/2}$ the power of the new test is substantially larger.

We finally note again that the power of the Kolmogorov-Smirnov test based on the process $(\hat{M}_t)_{t\in[0,1]}$ is even larger than the power obtained for $(\hat{N}_t)_{t\in[0,1]}$. Thus for high frequeny data the new tests are a substantial improvement of the currently available procedure for testing the parametric form of the diffusion coefficient in a stochastic differential equation.

4.3 Data Example

In this paragraph we apply the test based on the process $(N_t)_{t\in[0,1]}$ to tick-by-tick data. As a specific example we consider the log returns of the excange rate between the EUR and the US dollar in 2004. The data were available for 10 weeks between February and April 2004 and approximately 710 log returns were recorded per week. A typical picture for the 4th and 8th week is depicted in Figure 4.1.

We applied the proposed procedures to test the hypotheses $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1$, $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1|x|$, $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1 x^2$ and $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1 + \bar{\theta}_2 x^2$. The corresponding p-values are depicted in Table 4.5. The null hypothesis $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1$ is cleary rejected in all cases. For the hypotheses $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1|x|$ and $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1 x^2$ the results do not indiate a clear structure. In the remaining case $\bar{H}_0 : \sigma^2(t,x) = \bar{\theta}_1 + \bar{\theta}_2 x^2$ we observe relatively large p-values, which gives some evidence for the null hypothesis in all weeks under consideration. Further details of this evaluation are available from the authors.

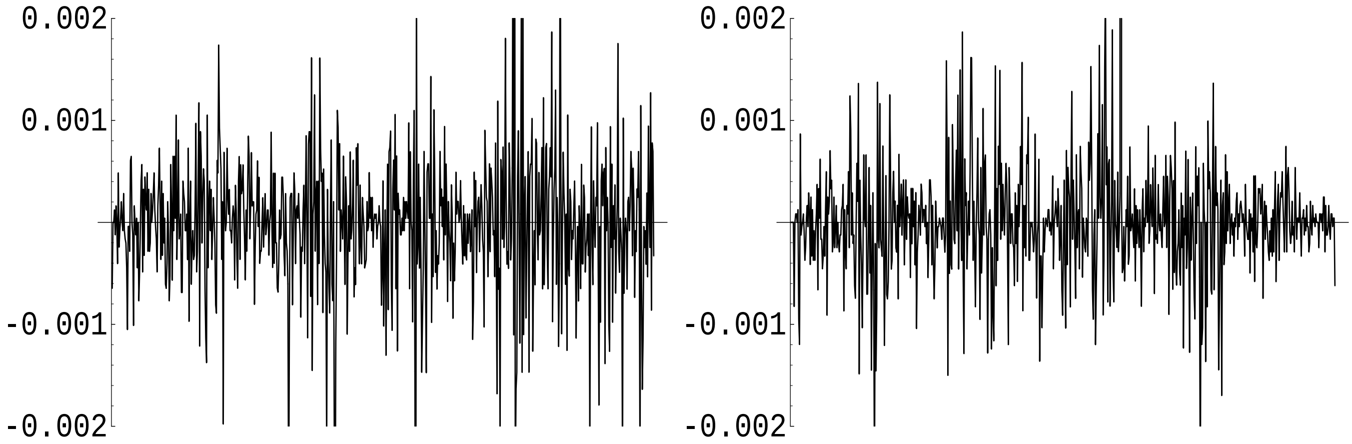

Figure 4.1. *Log returns of the EUR/USD exchange rate for two different weeks.*

week	1th	2th	3th	4th	5th	6th	7th	8th	9th	10th
n	714	714	713	714	714	714	708	714	718	710
$\sigma^2(t,x) = \theta_1$	0.000	0.026	0.000	0.002	0.002	0.000	0.004	0.000	0.001	0.010
$\sigma^2(t,x) = \theta_1\|x\|$	0.142	0.294	0.000	0.060	0.352	0.062	0.546	0.000	0.056	0.000
$\sigma^2(t,x) = \theta_1 x^2$	0.748	0.714	0.000	0.976	0.774	0.368	0.634	0.000	0.710	0.000
$\sigma^2(t,x) = \theta_1 + \theta_2 x^2$	0.880	0.996	0.886	0.994	0.978	0.986	0.968	0.974	0.966	0.988

Table 4.5. *p-values of the test based on the process $(N_t)_{t\in[0,1]}$ for various hypotheses on the volatility function. The table shows the results for ten weeks. The second row shows the number of the available data at each week.*

5 Appendix: Proofs

5.1 Proof of Lemma 3.1.

The proof of the following result is obtained along the lines of Dette, Podolskij and Vetter (2005) and therefore omitted. □

5.2 Proof of Lemma 3.2.

For the sake of brevity we restrict ourselves to a proof of asymptotic normality of the component

$$\sqrt{n}(\hat{C}_1 - C_1) \tag{4.1}$$

The general case is shown by exactly the same arguments using the results of Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2004). For a proof of the stable convergence of the statistic (4.1) we introduce the notation

$$\Delta_i^n X = X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \tag{4.2}$$

$$\beta_i^n = \sqrt{n}\sigma(\frac{i-1}{n}, X_{\frac{i-1}{n}})\Delta_i^n W \tag{4.3}$$

$$g(x) = |x| \tag{4.4}$$

$$\rho_x(f) = E[f(X)] \ , \text{ where } X \sim N(0, x^2) \tag{4.5}$$

$$\rho_{\sigma_s}(f) = \rho_{\sigma(s, X_s)}(f) \tag{4.6}$$

and decompose the proof in three parts.

(1) We prove the assertion

$$U^n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})[g(\beta_i^n) - \rho_{\sigma_{\frac{i-1}{n}}}(g)] \xrightarrow{\mathcal{D}_{st}} \nu \int_0^1 \sigma_1(s, X_s)\sigma(s, X_s) dW_s', \tag{4.7}$$

where $\nu = \sqrt{\mu_2 - \mu_1^2}$.

(2) We show the estimate

$$U^n - V^n \xrightarrow{P} 0, \tag{4.8}$$

where the random variable V_n is defined by

$$V^n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})[g(\sqrt{n}\Delta_i^n X) - E[g(\sqrt{n}\Delta_i^n X)|\mathcal{F}_{\frac{i-1}{n}}]] \tag{4.9}$$

(3) We prove the estimate

$$\sqrt{n}\mu_1(\hat{C}_1 - C_1) - V^n \xrightarrow{P} 0 \tag{4.10}$$

Recalling the definition of $\hat{C}_i$ in (2.21) and observing (4.1), (4.7) - (4.10) it follows

$$\sqrt{n}(\hat{C}_1 - C_1) \xrightarrow{\mathcal{D}_{st}} \mu_1^{-1}\sqrt{\mu_2 - \mu_1^2}\int_0^1 \sigma_1(s, X_s)\sigma(s, X_s)dW_s,$$

which proves the assertion of Lemma 3.2 for the second component.

Proof of (4.7). We introduce the random variable

$$\xi_i^n = \frac{1}{\sqrt{n}}\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})[g(\beta_i^n) - \rho_{\sigma_{\frac{i-1}{n}}}(g)] \tag{4.11}$$

and obtain the representation $U^n = \sum_{i=1}^n \xi_i^n$. Note that g is an even function and observe the identities

$$E[\xi_i^n|\mathcal{F}_{\frac{i-1}{n}}] = 0 \tag{4.12}$$

$$E[\xi_i^n \Delta_i^n W|\mathcal{F}_{\frac{i-1}{n}}] = 0 \tag{4.13}$$

$$E[|\xi_i^n|^2|\mathcal{F}_{\frac{i-1}{n}}] = n^{-1}(\mu_2 - \mu_1^2)\sigma_1^2(\frac{i-1}{n}, X_{\frac{i-1}{n}})\sigma^2(\frac{i-1}{n}, X_{\frac{i-1}{n}}) \tag{4.14}$$

Next, let N be any bounded martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, P)$, which is orthogonal to W (this means that the quadratic variation process $< M, N >_t$ is equal to 0). It follows from Barndorff-Nielsen et al. (2004)

$$E[\xi_i^n \Delta_i^n N|\mathcal{F}_{\frac{i-1}{n}}] = 0, \tag{4.15}$$

and finally Theorem IX 7.28 in Jacod and Shiryaev (2003) implies

$$U_n \xrightarrow{\mathcal{D}_{st}} \sqrt{\mu_2 - \mu_1^2}\int_0^1 \sigma_1(s, X_s)\sigma(s, X_s)dW'_s,$$

which proves (4.7).

Proof of (4.8). We consider the representation

$$V_n - U_n = \sum_{i=1}^n (\zeta_i^n - E[\zeta_i^n|\mathcal{F}_{\frac{i-1}{n}}]), \tag{4.16}$$

where the random variables ζ_i^n are given by

$$\zeta_i^n = \frac{1}{\sqrt{n}}\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})[g(\sqrt{n}\Delta_i^n X) - g(\beta_i^n)]\ , \tag{4.17}$$

and note that it is sufficient to prove

$$\sum_{i=1}^n E[|\zeta_i^n|^2] \to 0. \tag{4.18}$$

For a proof of (4.18) we calculate using Burkholder inequality, Lemma 6.2 in Appendix 6 and (2.11)

$$
\begin{aligned}
\sum_{i=1}^n E[|\zeta_i^n|^2] &= \frac{1}{n}\sum_{i=1}^n E\sigma_1^2(\frac{i-1}{n}, X_{\frac{i-1}{n}})[g(\sqrt{n}\Delta_i^n X) - g(\beta_i^n)]^2 \\
&\leq \sum_{i=1}^n E\sigma_1^2(\frac{i-1}{n}, X_{\frac{i-1}{n}})| \int_{\frac{i-1}{n}}^{\frac{i}{n}} b(s, X_s)\ ds + \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma(s, X_s) - \sigma(\frac{i-1}{n}, X_{\frac{i-1}{n}})\ dW_s|^2 \\
&\leq 2\sum_{i=1}^n E\{\sigma_1^4(\frac{i-1}{n}, X_{\frac{i-1}{n}})\}^{\frac{1}{2}} E\{| \int_{\frac{i-1}{n}}^{\frac{i}{n}} b(s, X_s)\ ds|^4\}^{\frac{1}{2}} \\
&+ E\{\sigma_1^4(\frac{i-1}{n}, X_{\frac{i-1}{n}})\}^{\frac{1}{2}} E\{| \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma(s, X_s) - \sigma(\frac{i-1}{n}, X_{\frac{i-1}{n}})\ dW_s|^4\}^{\frac{1}{2}} \\
&= o(1),
\end{aligned}
$$

which completes the proof of (4.8).

Proof of (4.10). Obviously, the assertion (4.10) follows from the statements

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}}) E[g(\sqrt{n}\Delta_i^n X) - g(\beta_i^n)|\mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{P} 0 \tag{4.19}
$$

$$
\sqrt{n} \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma_1(s, X_s)\rho_{\sigma_s}(g) - \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})\rho_{\sigma_{\frac{i-1}{n}}}(g)\ ds \quad \xrightarrow{P} \quad 0 \tag{4.20}
$$

Note that $\sigma, \sigma_1 > 0$, which shows that $\rho_{\sigma_s}(g) = \mu_1\sigma_s$ and (4.20) follows along the lines of Dette, Podolskij and Vetter (2005). For a proof of (4.19) we define the set

$$
A_i^n := \{|\sqrt{n}\Delta_i^n X - \beta_i^n| > |\beta_i^n|\}, \tag{4.21}
$$

and obtain the decomposition

$$
g(\sqrt{n}\Delta_i^n X) - g(\beta_i^n) = R_{in}^1 + R_{in}^2 - R_{in}^3, \tag{4.22}
$$

with

$$
\begin{aligned}
R_{in}^1 &= g'(\beta_i^n)(\sqrt{n}\Delta_i^n X - \beta_i^n), &(4.23)\\
R_{in}^2 &= [g(\sqrt{n}\Delta_i^n X) - g(\beta_i^n)]1_{A_i^n}, &(4.24)\\
R_{in}^3 &= g'(\beta_i^n)(\sqrt{n}\Delta_i^n X - \beta_i^n)1_{A_i^n}, &(4.25)
\end{aligned}
$$

where 1_A denotes the indicator function of the set A. Note that the decomposition (4.22) follows from the fact that the random variables $\sqrt{n}\Delta_i^n X$ and β_i^n have the same sign if $(\sqrt{n}\Delta_i^n X, \beta_i^n)$ is an element of $(A_i^n)^c$ (here B^c denotes the complement of the set B). Note also that g' is defined on $\mathbb{R}\backslash\{0\}$ and that $\sigma > 0$. We now decompose R_{in}^1 as follows

$$
R_{in}^1 = R_{in}^{1.1} + R_{in}^{1.2}, \tag{4.26}
$$

where

$$
\begin{aligned}
R_{in}^{1.1} &:= \sqrt{n} g'(\beta_i^n)\Big[\frac{1}{n} b(\frac{i-1}{n}, X_{\frac{i-1}{n}}) + \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma'(\frac{i-1}{n}, X_{\frac{i-1}{n}})\Big(\int_{\frac{i-1}{n}}^{s} \sigma(\frac{i-1}{n}, X_{\frac{i-1}{n}}) dW_t\Big) dW_s\Big] \\
R_{in}^{1.2} &:= \sqrt{n} g'(\beta_i^n)\Big[\int_{\frac{i-1}{n}}^{\frac{i}{n}} b(s, X_s) - b(\frac{i-1}{n}, X_{\frac{i-1}{n}})\ ds + \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma(s, X_s) - \sigma(\frac{i-1}{n}, X_s)\ dW_s \\
&+ \int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma'(\frac{i-1}{n}, X_{\frac{i-1}{n}})\Big(\int_{\frac{i-1}{n}}^{s} b(t, X_t) dt + \int_{\frac{i-1}{n}}^{s} \sigma(t, X_t) - \sigma(\frac{i-1}{n}, X_{\frac{i-1}{n}})\ dW_t\Big)\ dW_s \\
&+ \frac{1}{2}\int_{\frac{i-1}{n}}^{\frac{i}{n}} \sigma''(\frac{i-1}{n}, \xi_i^n)(X_s - X_{\frac{i-1}{n}})^2\ dW_s\Big] \\
&= R_{in}^{1.2.1} + R_{in}^{1.2.2} + R_{in}^{1.2.3} + R_{in}^{1.2.4},
\end{aligned}
$$

the last line defines the random variables $R_{in}^{1.2.j}$ $(j = 1, 2, 3, 4)$ and $\xi_i^n = \vartheta_i^n X_{\frac{i-1}{n}} + (1 - \vartheta_i^n) X_s$ for some $\vartheta_i^n \in [0, 1]$. Here σ', σ'' denote the first and the second derivative with respect to the second variable, respectively. Because $R_{in}^{1.1}$ is an odd function of $\Delta_i^n W$ and $\Delta_i^n W$ is independent of the σ-field $\mathcal{F}_{\frac{i-1}{n}}$ we obtain

$$
E[R_{in}^{1.1} | \mathcal{F}_{\frac{i-1}{n}}] = 0. \tag{4.27}
$$

Several applications of the Cauchy-Schwartz inequality, Burkholder inequality, Lemma 6.2 in the Appendix and (2.11) now yield

$$
\begin{aligned}
\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}}) E[R_{in}^{1.2.1} | \mathcal{F}_{\frac{i-1}{n}}] &= O_p(n^{-1}) \\
\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}}) E[R_{in}^{1.2.2} | \mathcal{F}_{\frac{i-1}{n}}] &= o_p(n^{-\frac{1}{2}}) \\
\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}}) E[R_{in}^{1.2.3} | \mathcal{F}_{\frac{i-1}{n}}] &= O_p(n^{-1}) \\
\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}}) E[R_{in}^{1.2.4} | \mathcal{F}_{\frac{i-1}{n}}] &= O_p(n^{-1})
\end{aligned}
$$

(uniformly with respect to $i = 1, \ldots, n$), and we obtain

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}}) E[R_{in}^{1} | \mathcal{F}_{\frac{i-1}{n}}] \xrightarrow{P} 0. \tag{4.28}
$$

In order to derive similar estimates for R_{in}^2 and R_{in}^3 we note that it follows from Barndorff-Nielsen et al. (2004) that we may assume the existence of positive constants, say $C_1, C_2 > 0$ such that

$$
\begin{aligned}
C_1 &< |\sigma| < C_2, \\
C_1 &< |\sigma_j| < C_2.
\end{aligned} \tag{4.29}
$$

Recalling the definition of the set A_i^n in (4.21) and observing the estimate

$$
1_{A_i^n} \leq 1_{\{|\sqrt{n}\Delta_i^n X - \beta_i^n| \geq \varepsilon\}} + 1_{\{|\beta_i^n| < \varepsilon\}}\ , \tag{4.30}
$$

we obtain by using Burkholder inequality, Lemma 6.2 in the Appendix and (2.11)

$$E[1_{A_i^n}|\mathcal{F}_{\frac{i-1}{n}}] \leq \frac{E[|\sqrt{n}\Delta_i^n X - \beta_i^n|^2]}{\varepsilon^2} + K_1\varepsilon \leq K_2(\frac{1}{n\varepsilon^2} + \varepsilon) \ , \tag{4.31}$$

for some constants $K_1, K_2 > 0$. With the choice $\varepsilon = n^{-\frac{1}{3}}$ it therefore follows

$$E[1_{A_i^n}|\mathcal{F}_{\frac{i-1}{n}}] = O_p(n^{-\frac{1}{3}}). \tag{4.32}$$

A further application of the Cauchy-Schwartz inequality yields

$$\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})E[R_{in}^2|\mathcal{F}_{\frac{i-1}{n}}] = O_p(n^{-\frac{2}{3}}),$$

$$\sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})E[R_{in}^3|\mathcal{F}_{\frac{i-1}{n}}] = O_p(n^{-\frac{2}{3}}),$$

which implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})E[R_{in}^2|\mathcal{F}_{\frac{i-1}{n}}] \quad \stackrel{P}{\longrightarrow} \quad 0,$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_1(\frac{i-1}{n}, X_{\frac{i-1}{n}})E[R_{in}^3|\mathcal{F}_{\frac{i-1}{n}}] \quad \stackrel{P}{\longrightarrow} \quad 0.$$

The assertion (4.19) finally follows from (4.22), (4.28), which proves (4.10). By the arguments given at the beginning the proof of Lemma 3.2 is completed. □

5.3 Proof of Theorem 3.3.

Recall the definition of $\hat{B}_t^i$ $(i = 0, \ldots, d)$ in (2.26) and (2.27), then it follows from Lemma 3.2 that

$$\sqrt{n} \begin{pmatrix} \hat{B}_{t_1}^0 - < \sigma, 1 >_2^{t_1} \\ \vdots \\ \hat{B}_{t_k}^0 - < \sigma, 1 >_2^{t_k} \\ \hat{C}_1 - C_1 \\ \vdots \\ \hat{C}_d - C_d \end{pmatrix} \xrightarrow{\mathcal{D}_{st}} \mu_1^{-1}\sqrt{\mu_2 - \mu_1^2} \int_0^1 \Sigma_{t_1,\ldots t_k}^{\frac{1}{2}}(s, X_s)\, dW_s' \ , \tag{4.33}$$

where W' denotes a $(d+k)$-dimensional Brownian motion, which is independent of W and the matrix $\Sigma_{t_1,\ldots t_k}(s, X_s)$ is defined by (3.1). Now an application of the Delta-method for stable convergence [see the proof of Theorem 4 in Dette, Podolskij and Vetter (2005)] yields weak convergence of the finite dimensional distributions, that is

$$\sqrt{n}\Big(\hat{M}_{t_1} - M_{t_1}, \ldots, \hat{M}_{t_k} - M_{t_k}\Big)^T \xrightarrow{\mathcal{D}_{st}} \mu_1^{-1}\sqrt{\mu_2 - \mu_1^2} \ \ V \int_0^1 \Sigma_{t_1,\ldots t_k}^{\frac{1}{2}}(s, X_s)\, dW_s' \ , \tag{4.34}$$

where the $k \times (d+k)$ matrix V is defined by (3.6). We finally prove tightness of the sequence $\sqrt{n}(\hat{M}_t - M_t)$. For this we use the decomposition

$$\sqrt{n}\Big(\hat{M}_t - M_t\Big) = \sqrt{n}(\hat{B}_t^0 - B_t^0) + \sqrt{n}B_t^T D^{-1}(\hat{C} - C) + o_p(1), \tag{4.35}$$

which follows from the definition of the processes M_t and $\hat{M}_t$ in (2.12) and (2.25), respectively, and from Lemma 3.1. Tightness of the process $\sqrt{n}(\hat{B}_t^0 - B_t^0)$ follows from Barndorff-Nielsen et al. (2004). For the second term in (4.35) we note that in view of (4.29) it is sufficient to prove

$$\lim_{\delta\to 0}\ \limsup_{n\to\infty} P(\sqrt{n}\delta|D_{kl}^{-1}||\hat{C}_j - C_j| > \varepsilon) = 0 \tag{4.36}$$

for all $\varepsilon > 0$ and all $1 \le k, l, j \le d$ where D_{kl}^{-1} denote the element of the matrix D^{-1} in the position (k, ℓ). For this we use Markov's inequality and obtain

$$\begin{aligned} P(\sqrt{n}\delta|D_{kl}^{-1}||\hat{C}_j - C_j| > \varepsilon) &\le \frac{n^{\frac{\varrho}{4}}\delta^{\frac{\varrho}{2}}E[|D_{kl}^{-1}|^{\frac{\varrho}{2}}|\hat{C}_j - C_j|^{\frac{\varrho}{2}}]}{\varepsilon^{\frac{\varrho}{2}}} \\ &\le \frac{n^{\frac{\varrho}{4}}\delta^{\frac{\varrho}{2}}E[|D_{kl}^{-1}|^{\rho}]^{\frac{1}{2}}E[|\hat{C}_j - C_j|^{\rho}]^{\frac{1}{2}}}{\varepsilon^{\frac{\varrho}{2}}} \end{aligned}$$

Moreover, it follows from the proof of Lemma 3.2 that $E[n^{\frac{\varrho}{2}}|\hat{C}_j - C_j|^{\rho}] = O(1)$ and an application of Cramer's rule, (4.29) and (2.19) yield $E[|D_{kl}^{-1}|^{\rho}] < \infty$, which proves (4.36) and by (4.35) the tightness of the process $\sqrt{n}(\hat{M}_t - M_t)$. □

5.4 Proof of Corollary 3.4

Note that in the case $d = 1$ and $\sigma_1 = 1$ we have for the quantities in (2.27) and (2.20)

$$\hat{B}_t^1 = \lfloor nt \rfloor / n, \quad \hat{D} = 1,$$

and the statistic $\hat{M}_t$ in (2.25) reduces to (3.13). Moreover, under the null hypothesis of homoscedasticity it follows in the case $k = 2$ for the matrix V

$$V = \begin{pmatrix} 1 & 0 & -t_1 \\ 0 & 1 & -t_2 \end{pmatrix}.$$

Note that under the hypothesis $H_0 : \sigma(t, X_t) = \sigma$ we have $\theta^{\min} = \sigma$ and the matrix Σ_{t_1,t_2} in (3.1) can be calculated as

$$\Sigma_{t_1 t_2}(s, X_s) = \sigma^2 \begin{pmatrix} I_{[0,t_1)}(s) & I_{[0,t_1\wedge t_2)}(s) & I_{[0,t_1)}(s) \\ I_{[0,t_1\wedge t_2)}(s) & I_{[0,t_2)}(s) & I_{[0,t_2)}(s) \\ I_{[0,t_1)}(s) & I_{[0,t_2)}(s) & 1 \end{pmatrix}$$

Consequently, the limiting process is Gaussian and determined by its covariance kernel. From the representation

$$\int_0^1 \Sigma_{t_1,t_2}(s, X_s)ds = \sigma^2 \begin{pmatrix} t_1 & t_1 \wedge t_2 & t_1 \\ t_1 \wedge t_2 & t_2 & t_2 \\ t_1 & t_2 & 1 \end{pmatrix}$$

we obtain

$$V \int_0^1 \Sigma_{t_1,t_2}(s, X_s)ds\; V^T = \sigma^2 \begin{pmatrix} t_1(1-t_1) & t_1 \wedge t_2 - t_1 t_2 \\ t_1 \wedge t_2 - t_1 t_2 & t_2(1-t_2) \end{pmatrix}.$$

Therefore it follows from Theorem 3.3 that under the null hypothesis of homoscedasticity the process $\sqrt{n}\hat{M}_t$ converges in law on $D[0,1]$, that is

$$\sqrt{n}\hat{M}_t \Longrightarrow \mu_1^{-1}\sqrt{\mu_2 - \mu_1^2}\sigma B_t \;,$$

where B_t denotes a Brownian bridge with covariance kernel $k(t_1,t_2) = t_1 \wedge t_2 - t_1 t_2$, which completes the proof of Corollary 3.4. □

5.5 Proof of Theorem 3.6.

We will only prove part (b) of the theorem. All other cases are treated by similar arguments. Since the drift function b does not influence the limiting process we assume without loss of generality that $b = 0$. With the notation $X_t^{H_0} = \sigma W_t$ we obtain the decomposition

$$X_t = X_t^{H_0} + \gamma_n \int_0^t h(s, X_s)dW_s = X_t^{H_0} + \gamma_n X_t^{H_1} \tag{4.37}$$

where the last identity defines the process $X_t^{H_1}$. This yields

$$\hat{N}_t = \sum_{k=1}^{\lfloor nt \rfloor} |\Delta_i^n X|^2 - \frac{\lfloor nt \rfloor}{n} \sum_{k=1}^{n} |\Delta_i^n X|^2 = \hat{N}_t^{H_0} + S_n. \tag{4.38}$$

Here $\hat{N}_t^{H_0}$ denotes the process defined by (3.14), where the random variables $X_{\frac{k}{n}}$ have to be replaced by the corresponding quantities $X_{\frac{k}{n}}^{H_0}$ and the process S_n is defined by

$$S_n \;=\; P_{n1} + P_{n2} - N_{n1} - N_{n2}$$

with

$$\begin{aligned} P_{n1} &= 2\gamma_n \sum_{k=1}^{\lfloor nt \rfloor} \Delta_i^n X^{H_0} \Delta_i^n X^{H_1} \\ P_{n2} &= \gamma_n^2 \sum_{k=1}^{\lfloor nt \rfloor} |\Delta_i^n X^{H_1}|^2 \\ N_{n1} &= 2\frac{\lfloor nt \rfloor}{n}\gamma_n \sum_{k=1}^{n} \Delta_i^n X^{H_0} \Delta_i^n X^{H_1} \\ N_{n2} &= \frac{\lfloor nt \rfloor}{n}\gamma_n^2 \sum_{k=1}^{n} |\Delta_i^n X^{H_1}|^2 \end{aligned}$$

(here and in the following discussion the dependence of P_{nj} and N_{nj} $(j = 1, 2)$ on the index t will not be reflected by our notation). A straightforward calculation yields

$$\begin{aligned} P_{n1} &= 2\gamma_n \sigma \int_0^t h(s, X_s) ds + o_p(\gamma_n), \\ P_{n2} &= \gamma_n^2 \int_0^t h^2(s, X_s) ds + o_p(\gamma_n^2), \\ N_{n1} &= 2t\gamma_n \sigma \int_0^1 h(s, X_s) ds + o_p(\gamma_n), \\ N_{n2} &= t\gamma_n^2 \int_0^1 h^2(s, X_s) ds + o_p(\gamma_n^2), \end{aligned}$$

which gives (observing that $\gamma_n = n^{-1/2}$)

$$\sqrt{n}\hat{N}_t = \sqrt{n}\hat{N}_t^{H_0} + 2\sigma \left(\int_0^t h(s, X_s) ds - t \int_0^1 h(s, X_s) ds \right) + o_p(1)$$

The assertion now follows from Corollary 3.4. □

6 Appendix: two auxiliary results

Lemma 6.1 *If assumptions (2.1), (2.2) are satisfied and* $E[|X_0|^{2m}] < \infty$ *for some* $m \in \mathbb{N}$*, then the inequalities*

$$E[\sup_{0 \le s \le t} |X_s|^{2m}] \le C_{m,K}(1 + E[|X_0|^{2m}]) e^{C_{m,K} t} \quad \forall t \in [0, 1]$$

$$E[|X_t - X_s|^{2m}] \le C_{m,K}(1 + E[|X_0|^{2m}]) |t - s|^m \quad \forall s, t \in [0, 1]$$

hold for some $C_{m,K} > 0$.

Proof of Lemma 6.1 See Karatzas/Shreve (1991) p.306. □

Lemma 6.2 *If assumptions (2.1), (2.2), (2.6) are satisfied and* $E[|X_0|^{2m}] < \infty$ *for some* $m \in \mathbb{N}$ *then we have*

(A.1) $$\sup_{0 \le t \le 1} E[|b(t + h, X_{t+h}) - b(t, X_t)|^{2m}] = O(|h|^m) \quad (h \downarrow 0)$$

This statement remains true for functions $\sigma, \sigma_1, \ldots, \sigma_d$.

Proof of Lemma 6.2 We first observe that

$$\begin{aligned} E[|b(t + h, X_{t+h}) - b(t, X_t)|^{2m}] &\le E[(|b(t + h, X_{t+h}) - b(t, X_{t+h})| + |b(t, X_{t+h}) - b(t, X_t)|)^{2m}] \\ &\le 2^m E[|b(t + h, X_{t+h}) - b(t, X_{t+h})|^{2m} + |b(t, X_{t+h}) - b(t, X_t)|^{2m}] \end{aligned}$$

It follows from (2.1), (2.6) and Lemma 6.1

$$E[|b(t+h, X_{t+h}) - b(t, X_{t+h})|^{2m}] = O(|h|^m)$$

and

$$E[|b(t, X_{t+h}) - b(t, X_t)|^{2m}] = O(|h|^m)$$

Combining the above estimates we obtain the result. □

Acknowledgements. The work of the authors was supported by the Deutsche Forschungsgmeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen). The authors are also grateful to I. Gottschlich, who typed this paper with considerable technical expertise and to M. Vetter for computational assistance.

References

Achieser, N.J. (1956). Theory of approximation. Dover Publications Inc., New York

Ait-Sahalia, Y. (1996). Testing continuous time models of the spot interest rate. Rev. Financ. Stud. 9, 385-426.

Aldous, D.J., Eagleson, G.K. (1978). On mixing and stability of limit theorems. Ann. of Prob. 6(2), 325-331.

Barndorff-Nielsen, O.E. Graversen, S.E., Jacod, J., Podolskij, M., Shephard, N. (2004). A central limit theorem for realised power and bipower variations of continuous semimartingales. Preprint, Ruhr-Universität Bochum. Available at:
http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm

Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. J. Polit. Econ. 81, 637-654.

Brennan, M.J., Schwartz, E.S. (1979). A continuous time approach to the pricing of bonds. Journal of Banking and Finance 3, 133-155.

Chan, K.C., Karolyi, G.a., Kongstaff, F.a., Sanders, a.B. (1992). An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209-1227.

Constantinides, G.M. (1992). A theory of the nominal term structure of interest rates. Rev. Finan. Studies 5, 531-552.

Corradi, V., White, H. (1999). Specification tests for the variance of a diffusion. J. Time Series 20, 253-270.

Courtadon, C. (1982). The pricing of options on defautl-free bonds. Journal of Financing and Quantitative Analysis 17, 75-100.

Cox, J.C., Ingersoll, J.E., Ross, S.A. (1985). A theory of the term structure of interest rates. Econometrica 53, 385-407.

Dette, H., von Lieres und Wilkau, C. (2003). On a test for a parametric form of volatility in continuous time financial models. Financ. Stoch. 7, 363-384

Dette, H., Podolskij, M., Vetter, M. (2005). Estimation of integrated volatility in continuous time financial models with applications to goodness-of-fit testing. Preprint, Ruhr-Universität Bochum. To appear: Scandinavian Journal of Statistics. Available at:
http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm

Duffie, J.D., Harrison, J.M. (1993). Arbitrage pricing of Russian options and perpetual lookback options. Ann. Appl. Probab. 3, 641-651.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law, *Seminaire de Probabilites* **XXXI**, 232-246.

Jacod, J.,Shiryaev, A. N. (2003). Limit theorems for stochastic processes. 2nd ed. Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer.

Karatzas, I. (1988). On pricing of American options. Appl. Math. Optimization **17**, 37-60.

Karatzas, I., Shreve, S.E. (1991). Brownian Motion and Stochastic Calculus. New York: Springer.

Merton, R.C. (1973). Theory of rational option pricing. Bell Journal of Economics and Management Science 4, 141-183.

Merton, R.C. (1990). Continuous-Time Finance. Cambridge: Blackwell

Revuz, D., Yor M. (1999). Continuous Martingales and Brownian Motion. Berlin: Springer.

Vasicek, O. (1977). An equilibrium characterization of the term structure. J. Finan. Economics 5, 177-188.