Nonparametric Modelling of Interest Rates

Dissertation

zur Erlangung des akademisches Grades eines

Doktors der Naturwissenschaften

(Dr.rer.nat.)

dem Fachbereich Mathematik der Universität Duisburg-Essen

vorgelegt im Oktober 2007

von

Evgeny Zoldin, geb. in Tscherkassy (die Ukraine)

Tag der mündlichen Prüfung: 26 Februar 2008
Gutachter: Prof. Dr. P. Laurie Davies
Prof. Dr. Walter Krämer

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The pioneering works of Black and Scholes (1973) and Merton (1973) opened new ways for economic research not only in the market of derivatives, but also in the four major financial markets of commodities, debt, equity, and foreign exchange. In this thesis we consider the financial assets of debt market, namely interest rate papers such as obligations, debt instruments, and bills. They are called discount papers because the holder receives from the issuer of the asset the nominal price he bought at a lower price, i.e. he bought the paper at a discount. There are termless assets, known as coupon papers, which do not have a maturity time, but at certain times a bonus is paid to the holders. In any case, the buyer of such paper gets some absolute profit for some period of time. The relation between profit value and buying price reduced to the period of one year gives us the discount rate or the yield of the financial paper. Because such papers, with minor reservations, allow a riskless capital investment, they are traded on secondary markets and valued above all by their yield. As well as the price of usual shares, the interest rate of discount papers mirrors the changes of the state of the market and consequently its value - the quotation - changes with a time. As an example of a discount paper, we consider 3-Month US Government Treasury Bills throughout this thesis. Figure 1.1 shows the development of its interest rate on daily trading on the secondary market for the period from January 1954 up to September 2006. Our main goal is modelling the behaviour of such assets, that is, to construct a stochastic model that generates the data, which "look like" the original data. At the same time, we do not claim that the original data were generated by such a model.

Though the lifetime of each issue is 3 months, the data represent some

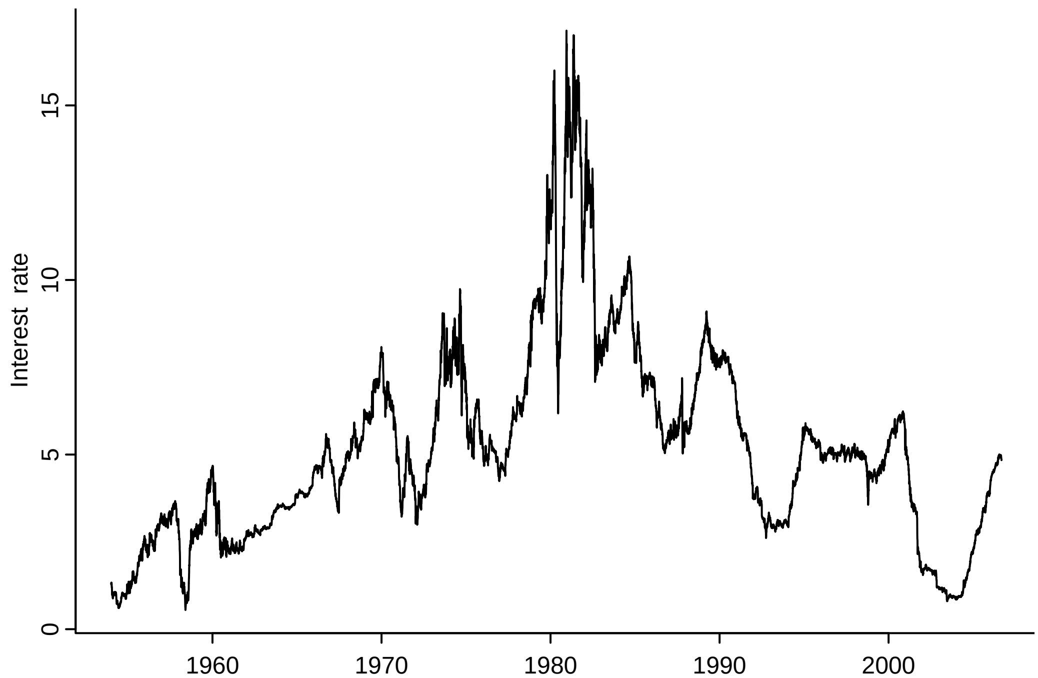

Fig. 1.1: 3-month US Government Bills for period from 04.Jan.1954 till 07.Sep.2006; Number of observations 13 159

common index of all issues present at that moment on the market. It stretches back over 50 years and can be considered as a single asset that was issued on 4th January 1954 and redeemed on 7th September 2006 for 100 dollars. The price of such imaginary paper can be retrospectively computed based on the daily interest rates. This price is displayed in Figure 1.2. The second curve is the price of another imaginary paper – the paper which has the same life period, the same issue, and redeem price, but a constant interest rate. This constant rate can serve in some sense as an ideal level for the 3-month bills for that time period.

The number of factors having an influence upon the price (interest rate) of assets is large and various – from the general economic situation up to every private participant in the market – and all of them cannot be exactly accounted for in the model. It is influenced by a sum of all small decisions which may be considered as noise whose intensity depends on the current value of the interest rate but not on the current time. Another factor is the deviation of the current rate from some true interest rate, which very

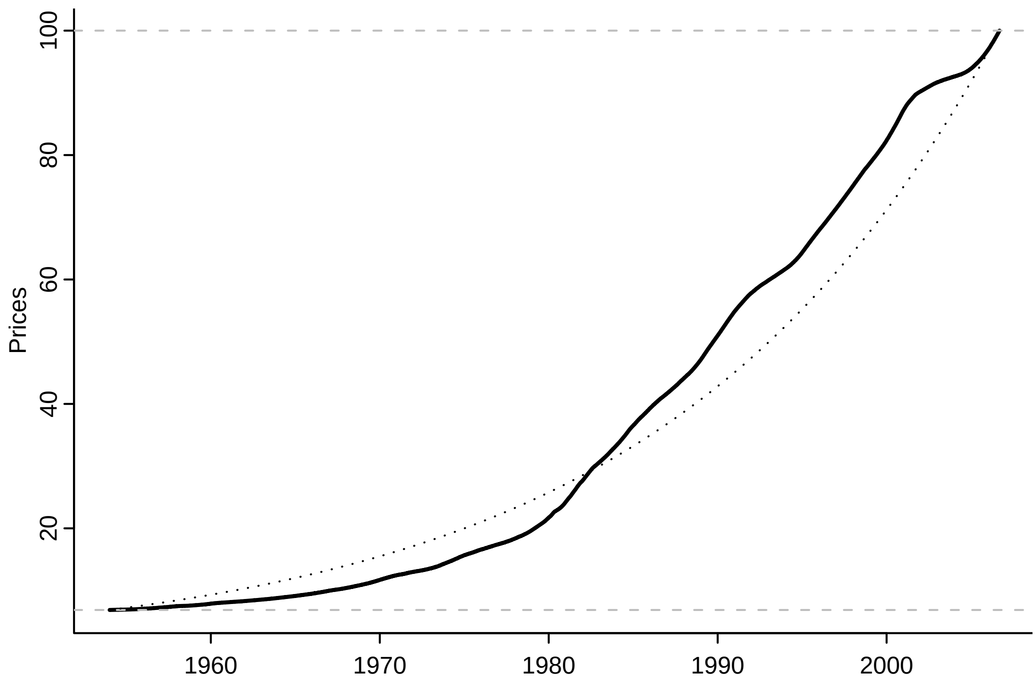

Fig. 1.2: Solid line – Price of “imaginary” 3-month Bills;
Dotted line – price of constant interest rate paper

approximately can be defined by the above mentioned constant rate. Such a "descriptive" model can be analytically described by a homogeneous diffusion process, i.e. the random process which satisfies the stochastic differential equation

$$dX(t) = b\big(X(t)\big)dt + \sigma\big(X(t)\big)dB(t), \quad X(0) \overset{a.s.}{=} x_0 \in \mathbb{R}, \tag{1.1}$$

where deterministic component $b\big(X(t)\big)dt$ corresponds to the second factor while the stochastic one $\sigma\big(X(t)\big)dB(t)$ represents the noise factor. Here the drift coefficient $b(x)$ and volatility (or diffusion) coefficient $\sigma(x)$ are real functions and $B(t)$ is standard Brownian motion. Often the drift and squared volatility functions are called instantaneous mean and variance respectively and the process $X(t)$ itself is called short interest rate process. Below we assume the process $X(t)$ is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in [0, T]$ with $T \approx 52.7$(years).

It seems reasonable, especially on commodity markets with strongly pronounced seasonality, to consider the drift and volatility to be dependent on the time t directly. But in this thesis we consider a stationary model. To complete the model specification we need to define the drift and volatility terms of the diffusion process. Many researchers use parametric models. Under parametric estimation the function is assumed to belong to some known family of functions parameterized by a parameter and main goal of the method is to determine the value of the parameter. In this thesis we use the nonparametric approach, that is, we make no assumptions in advance about the form of either the drift or volatility functions.

Chapter 2 provides some necessary facts of diffusion theory and discretization scheme that we use.

In Chapter 3 we give a short review of previous work on both parametric and nonparametric approaches in this area. Chapter 4 describes the whole scheme of the estimating and the modelling procedure.

The estimate of the volatility function is constructed in Chapter 5. The chapter also contains theoretical results on the consistency of the estimator and its convergence rate.

Similarly Chapter 6 describes the estimation of the drift coefficient again with theorems on the consistency of the estimator and its convergence rate.

In Chapter 7 we demonstrate the performance of our methods on a test model and its application to the real data, namely to historical data of 3-Month US Government Treasury Bills.

The proofs of theorems from Chapters 5 and 6 are collected in Chapter 8.

To this work it is enclosed a CD which contains source codes of the estimation procedures in C and R-package and a set of 2000 simulated samples. The detailed description is provided in "`readme.txt`" file in the root folder of the CD.

I am very thankful to Professor P.L. Davies for introducing to attractive research area of financial analysis, for his supervising throughout the whole period of my work on this thesis at the University of Duisburg-Essen, and for many productive ideas and remarks.

Also I would like to thank my colleagues Dr. M. Meise and C. Höhenrieder for their help at programmatic realisation of some algorithms, which are used in this work.

The financial support of the grant SFB 475 (Reduction of Complexity for Multivariate Data Structures) of the German Research Council (Deutsche Forschungsgemeinschaft, DFG) is much appreciated. This work is part of its section B1 (Capital Markets as Indicators of Economic Structural Change).

Chapter 2

Some facts on diffusion processes

In general to ensure the existence and uniqueness of a strong solution of the equation (1.1) the drift and volatility coefficients are subject to global Lipschitz and bounded growth conditions. That is, for all $x, y \in \mathbb{R}$ there exists some constant K such that

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leqslant K|x - y| \qquad \text{and}$$
$$b^2(x) + \sigma^2(x) \leqslant K(1 + x^2).$$

See, for example, Arnold (1974) or Gihman and Skorohod (1972). These conditions will not be used directly, but we suppose that a strong solution exists and unique.

Since $X(t)$ is an interest rate it is quite naturally to suppose $X(t)$ taking values on the half-line $[0, +\infty)$. For our goal we require some futher conditions. Namely, for the scale function

$$s(x) = \exp\left(-2\int_0^x \frac{b(y)}{\sigma^2(y)}dy\right) \tag{2.1}$$

we require

$$\int_0^{+\infty} s(x)dx = \infty, \tag{2.2}$$
$$\int_0^{+\infty} \frac{dx}{s(x)\sigma^2(x)} = C_X < \infty, \qquad \text{and}$$

$$\int_0^{+\infty} s(x)dx \int_x^{+\infty} \frac{dy}{s(y)\sigma^2(y)} < \infty. \tag{2.3}$$

It is known (see Gihman and Skorohod (1972)) that in the case of (2.2) the process $X(t)$ is ergodic with invariant density explicitly given by

$$\mu(x) = \frac{1}{s(x)\sigma^2(x)C_X}. \tag{2.4}$$

By Lemma 6.3 from Karlin and Taylor (1981, Chapter 15) from (2.2) follows

$$\int_0^{+\infty} s(x)dx \int_0^{x} \frac{dy}{s(y)\sigma^2(y)} = \infty,$$

and as proved by Mao (2006, Theorem 3.2) this together with the condition (2.3) implies uniform ergodicity of $X(t)$. On notion of various types of ergodicity and relations between the types we refer to Nummelin (1984), Chen (2001) and Chen (2002). We only remark that geometric and uniform ergodicity at Nummelin are called at Chen exponential and strong respectively.

Below we shall use the same notation $\mu(\cdot)$ for the measure of Lebesgue measurable sets $B \subseteq \mathbb{R}$ associated with the density $\mu(x)$ as

$$\mu(B) = \int_B \mu(x)dx$$

and this does not lead to ambiguity.

We work with daily trading prices and therefore we consider the sample data set as a realization of the process $X(t)$ on the interval $[0, T]$ observed at the discrete time moments jh, $j = 0, \ldots, N \approx T/h$, where T is the length of the trajectory $X(t)$ in years and the discretization step h corresponds to one trading day, i.e.

$$h \approx 0.004 = 1/250.$$

We transform the continuous time equation (1.1) into a discrete one using the classical Euler differential scheme

$$\begin{aligned} &X_0 = X(0) \qquad \text{and for } j = 0, \ldots, N-1 \\ &X_{j+1} \equiv X\big((j+1)h\big) = X(jh) + b\big(X(jh)\big)h + \sigma\big(X(jh)\big)\xi_j\sqrt{h} \end{aligned} \tag{2.5}$$

with ξ_j independent standard normal distributed random variables and X_0 is distributed with the invariant density μ. In contrast to standard discretization-based approach we do not assume that sampling interval h decreases if we add new data to the sample. That is, we add new observations to the end of the data row while the discretization step h remains fixed. Taking into account intraday data leads to study of high frequency properties of an interest rate, which are not a goal of this work (see Aït-Sahalia, 1996b).

Though the values domain of the process $X(t)$ can be unbounded we consider a bounded interval, say $[l, r]$, and will study the process $X(t)$ on that interval only. For simplicity using the scale transformation $(X(t)-l)/(r-l)$ one may think that the interval is $[0,1]$. On this interval we impose some additional restrictions on the diffusion coefficient σ and the invariant density μ. We assume that

> $\mu(x), \sigma(x) \geqslant \nu > 0$ for all $x \in [0,1]$ and
>
> $\sigma(x)$ is cdlg and on subintervals of continuity is Lipschitz with some constant K_σ.

Notice that the first condition together with (2.4) implies that both μ and σ are bounded on $[0,1]$. Also we require that the drift b is bounded on $[0,1]$.

For the ergodic process $X(t)$ a strong law of large numbers

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f\big(X(t)\big)dt \overset{a.s.}{=} \int_{\mathbb{R}} f(x)\mu(x)dx \tag{2.6}$$

holds for a wide class of measurable coefficients $f(\cdot)$ (Gihman and Skorohod, 1972, p.134). In particularly, we shall use such result for the functions $b(\cdot)$, $\sigma(\cdot)$ multiplied by the indicator function $\mathbb{1}_{\{\cdot\in[0,x]\}}$, $x \in [0,1]$.

Chapter 3

Review of previous work

There are two approaches to modelling of drift and volatility functions – parametric and nonparametric. In the parametric approach the family of functions is parameterized by some parameter θ and it is assumed that the "true" function belongs to this family. In the nonparametric case no restrictions on the form of the true function are imposed except regularity restrictions such as continuity, differentiability or monotonicity.

In the parametric case the equation (1.1) is considered in the form

$$dX(t) = b\big(X(t), \theta\big)dt + \sigma\big(X(t), \theta\big)dB(t), \quad \theta \in \Theta \subseteq \mathbb{R}^d.$$

Probably the best known models are those of Merton (1973) with $b(x, \theta) = \beta$ and $\sigma(x, \theta) = \sigma$, of Vasicek (1977) with $b(x, \theta) = \beta(\alpha - x)$ and $\sigma(x, \theta) = \sigma$, and of Cox, Ingersoll and Ross (1985) with $b(x, \theta) = \beta(\alpha - x)$ and $\sigma(x, \theta) = \sigma\sqrt{x}$. Here $\theta = (\alpha, \beta, \sigma)$. In these cases the conditional distributions of $X(t)$ are log-normal, normal and non-central chi-square respectively. And in the first case the process $X(t)$ is the geometric Brownian motion and in the second one the Ornstein-Uhlenbeck process. These models and several others are summarized in Table 3.1 taken from Aït-Sahalia (1996a). Below we refer the model of Cox, Ingersoll and Ross (1985) as CIR model.

An excellent survey of Sørensen (2004) introduces to parametric estimation techniques such as martingale and so-called simple estimating functions, analytical and numerical approximations of the likelihood function, Bayesian analysis and Markov Chain Monte-Carlo (MCMC) methods, indirect inference and the so-called efficient method of moments (EMM) and gives a list of corresponding references. A new estimation technique using transformation functions is developed by Kelly, Platen, and Sørensen (2004).

$b(x)$	$\sigma(x)$	Reference
β	σ	Merton (1973)
$\beta(\alpha - x)$	σ	Vasicek (1977)
0	$\sigma x^{3/2}$	Cox (1975) Cox, Ingersoll and Ross (1980)
0	σx	Dothan (1978)
$\beta x\big(\alpha - \ln(x)\big)$	σx	Brennan and Schwartz (1979)
$\beta(\alpha - x)$	σx	Courtadon (1982)
$\alpha x^{\delta-1} + \beta x$	$\sigma x^{\delta/2}$	Marsch and Rosenfeld (1983)
$\beta(\alpha - x)$	$\sigma x^{1/2}$	Cox, Ingersoll and Ross (1985) Brown and Dybvig (1986) Gibbons and Ramaswamy (1993)
$\beta(\alpha - x)$	σx^{λ}	Chan et al. (1992)
$\alpha + \beta x + \gamma x^2$	$\sigma + \gamma x$	Constantinides (1992)
$\beta(\alpha - x)$	$\sqrt{\sigma + \gamma x}$	Duffie and Kan (1993)

Table 3.1: Some parametric models of the short-term interest rate process

As stated above we consider the asymptotic scheme with a fixed sampling period h. Other models of financial activity consider the situation when the length h tends to zero and this leads to the consideration of continuously observed diffusion models. The estimation of such continuous-time processes is well studied; see, for example, Lipster and Shiryaev (2001) or Kutoyants (1984).

The arrival of the computer era has stimulated and considerably accelerated the development of nonparametric methods, which require lengthy calculations. Over the last two decades the literature concerning nonparametric estimation of diffusion models has become very large and it continues to grow. The recent comprehensive overview by Fan (2005a) and comments by Phillips and Yu (2005), Sørensen (2005), Mykland and Zhang (2005), and Fan (2005b) cover a variety of nonparametric techniques in financial econometrics and demonstrate their applications to various aspects of both time-homogeneous and time-dependent diffusion models: drift and volatility terms, transition and state price densities. Additional references are Spokoiny (2000) who proposed a locally linear smoother with a data-driven bandwidth for nonparametric estimation of drift term and Rei (2006) who constructed an estimator of the volatility function for the embedded Markov chain and spectral properties of its Markov transition operator.

The asymptotic scheme with $h \to 0$ is studied, for instance, using kernel estimation methods by Florens-Zmirou (1993), Jacod (2000), wavelet methods by Genon-Catalot, Laredo and Picard (1992), Hoffmann (1999), Honoré (1997) or Kalman filter by Shoji (2002), Shoji (2004). A selective review made by Cai and Hong (2003) is concerned with nonparametric estimation and nonparametric testing of parametric continuous-time diffusion models both time-homogeneous, time-dependent and jump. See also references therein.

In Kloeden et al. (1996) the Euler discretization scheme (2.5) is called strong order 1/2 approximation because it has convergence rate $\sqrt{h}$. By the Itô-Taylor expansion one can obtain the higher-order approximation, for instance a strong order-one approximation for time-homogeneous model is given by

$$X_{j+1} = X_j + b\big(X_j\big)h + \frac{1}{2}\left\{\sigma\left(X_j + b\big(X_j\big)h + \sigma\big(X_j\big)\xi_j\sqrt{h}\right) + \sigma\big(X_j\big)\right\}\xi_j\sqrt{h}.$$

See Kloeden et al. (1996), relation (3.14). By simulating monthly test data from the latter and the Euler discretization schemes Fan (2005a) notices that difference between different scheme simulations is negligible and agreeing with Stanton (1997) he concludes that “as long as data are sampled monthly or more frequently, the errors introduces by using the Euler approximation are very small for stochastic dynamics that are similar to the CIR model”.

Applying nonparametric methods for study of US Treasury Bills data Aït-Sahalia (1996b) and Stanton (1997) concluded that the drift coefficient is nonlinear. However later this statement has been called into question in works of Pritsker (1998) and Chapman and Pearson (2000), where the drawbacks of kernel estimation methods of Aït-Sahalia (1996b) and Stanton (1997) are considered. Chapman and Pearson (2000) perform a Monte Carlo study of finite sample properties of estimators of Aït-Sahalia (1996b) and Stanton (1997) applying its to simulated paths of CIR model. They found that the typical estimated drift function displays nonlinearity, even though the true drift is linear. Commenting on Aït-Sahalia's (1996b) estimator Pritsker (1998) says that "to achieve the estimator precision implied by the asymptotic distribution with 22 years of daily data actually requires more than 2750 years of data" while Chapman and Pearson (2000) conclude that there is no definitive answer to the question whether the short rate drift actually nonlinear. For some additional discussion we refer to Cai and Hong (2003). Though this work also concerns with estimation of the drift term, we neither insist on its linearity or nonlinearity, because as it said above our main goal is to obtain a model that can produce data like to original ones.

Chapter 4

Estimation scheme

For estimation of the volatility and drift functions we propose two statistical Tukey procedures (Tukey 1993) both are based on a data decomposition, which following Tukey can be written as

$$\text{DATA} = \text{SIGNAL} + \text{NOISE},$$

where the SIGNAL is assumed to be simple and the NOISE is complex.

Since for the diffusion process (1.1) and for our discretization scheme (2.5) we have

$$\mathbb{E}(dX(t))^2 = (dt)^2\mathbb{E}\big(b(X(t))\big)^2 + \sigma^2(X(t))dt \quad \text{and}$$
$$\mathbb{E}(X_{j+1} - X_j)^2 = h^2\mathbb{E}\big(b(X_j)\big)^2 + \sigma^2(X_j)h$$

we can say that for small values of h the squared increments $(X_{j+1} - X_j)^2$ of the sequence $\{X_j\}$ are dominated by the volatility term

$$(X_{j+1} - X_j)^2 = \sigma^2(X_j)\xi_j^2 h + 2h^{3/2}b(X_j)\sigma(X_j)\xi_j + h^2 b^2(X_j),$$

that is,

$$(X_{j+1} - X_j)^2 \approx \sigma^2(X_j)\xi_j^2 h. \tag{4.1}$$

Using this relation the whole scheme of specifying the drift and volatility coefficients looks as follows:

- The first step is to associate the SIGNAL with the volatility function σ and estimate it using the squared increments $(X_{j+1} - X_j)^2$ associated with the DATA. For this we adopt the method developed by Davies (2006) and based on properties of the χ^2-distribution.

- In the second step we associate the SIGNAL with the drift term b and estimate it using the linear increments $X_{j+1} - X_j$ associated with the DATA and the already specified volatility σ. Here we apply the adoption of the taut string method developed by Davies and Kovac (2001).

Both methods were originally developed for the situations where the functions associated with the SIGNAL depend on the parameter t directly. In our framework that functions, i.e. σ and b depend on the state of the process $X(t)$. The simplicity of the SIGNAL is expressed in the fact that both methods delivery a piecewise constant estimator. And additionally the first method based on the properties of chi-squared distribution and yields the estimator with minimal count of constancy intervals, while the second one - the taut string method - results the minimal number of local extreme values of the estimator.

The performance of the proposed methods is demonstrated on the original as well on test data. As a test model we use the following diffusion process:

$$dX(t) = -1.5X(t)dt + 0.5\big(1 + \big|\sin(\pi X(t))\big|\big)dW(t). \tag{4.2}$$

In Figure 4.1 is shown a sample path generated from this model with the same precision of two digits after the decimal point and the same discretization step as the original data.

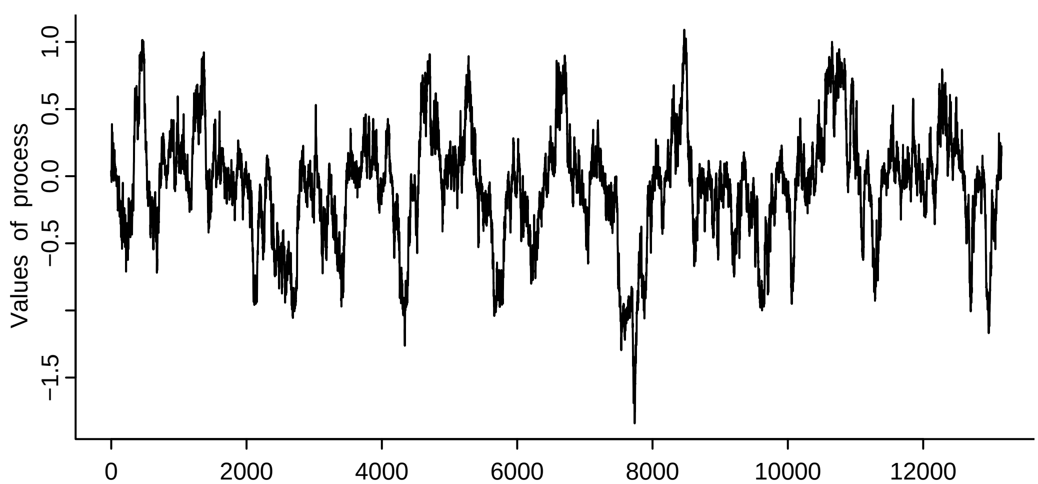

Fig. 4.1: A sample path of the model (4.2); Consisting of 13 159 observations

Chapter 5

Estimation of the volatility term

By $\{X_{(j)}\}$ we denote non-decreasing permutation of the sequence $\{X_i\}$ and order the equal X_i's chronologically. If i is those of indexes that corresponds to (j), then we put $\Delta X_{(j)} = X_{i+1} - X_i$. The relation (4.1) leads us to consider the decomposition

$$R_j = \sigma_j Z_j, \qquad j = 1, \dots, N,$$

where we put

$$R_j^2 = \frac{\left(\Delta X_{(j)}\right)^2}{h}, \qquad \sigma_j = \sigma(X_{(j)}), \tag{5.1}$$

and $\{Z_j\}$ is the respectively renumbered sequence $\{\xi_j\}$. Here R_j's represent DATA, σ_j's – SIGNAL and Z_j's – NOISE. From the last equality we have

$$\sum_{j \in I} \frac{R_j^2}{\sigma_j^2} = \sum_{j \in I} Z_j^2 \overset{d}{=} \chi^2(|I|),$$

where $|I| = \sharp\{j \mid j \in I\}$. If we denote by $qu\,(\alpha, k)$ an α-quantile of $\chi^2(k)$ random variable then for $\alpha \in [0, 1]$

$$\mathbb{P}\left\{ qu\left(\frac{1-\alpha}{2}, |I|\right) \leqslant \sum_{j \in I} Z_j^2 \leqslant qu\left(\frac{1+\alpha}{2}, |I|\right) \right\} = \alpha.$$

We are interested not in one interval I only, but in all intervals from some family of intervals $\mathcal{I}$ simultaneously and therefore we must let α depend on

N:

$$\mathbb{P}\left\{qu\left(\frac{1-\alpha_N}{2}, |I|\right) \leqslant \sum_{j\in I}\frac{R_j^2}{\sigma_j^2} \leqslant qu\left(\frac{1+\alpha_N}{2}, |I|\right), \quad I \in \mathcal{I}\right\} = \alpha.$$

For a given α the values of α_N can only be determined by simulations. Following Davies (2006) we choose

$$\alpha_N = 1 - \frac{2}{N^{1.15}\sqrt{4.6\pi \log N}} \tag{5.2}$$

what correponds to the choice $\sqrt{2.3\log(N)}$ for the threshold in Davies and Kovac (2001). Such choice of α_N is a good approximation for $\alpha = 0.7$ and values of N from 5000 to 15000. As we are looking for a piecewise constant estimator of the volatility function $\sigma(x)$ the family $\mathcal{I} = \mathcal{I}(N)$ is such partition $I_1, \ldots, I_m$ of interval $1, \ldots, N$ with minimal m that

$$\sigma_l^2(I) \leqslant \hat{\sigma}^2(I_k) \leqslant \sigma_u^2(I), \qquad I \subseteq I_k,\ k = 1, \ldots, m, \tag{5.3}$$

where the lower σ_l^2 and the upper σ_u^2 bounds for the value $\hat{\sigma}(I_k)$ of the estimator of the function σ on the interval I_k are defined as follows:

$$\sigma_l^2(I) = \max_{i,j\in I, i\leqslant j}\left\{\frac{\sum_{i\leqslant l\leqslant j} R_l^2}{qu\left(\frac{1+\alpha_N}{2}, j-i+1\right)}\right\}$$

$$\sigma_u^2(I) = \min_{i,j\in I, i\leqslant j}\left\{\frac{\sum_{i\leqslant l\leqslant j} R_l^2}{qu\left(\frac{1-\alpha_N}{2}, j-i+1\right)}\right\},$$

and the value of the estimator itself is

$$\hat{\sigma}(I_k) := \sqrt{\frac{1}{|I_k|}\sum_{j\in I_k} R_j^2}.$$

The whole details of the construction of such partition can be found in Davies (2006). If we replace the inequalities (5.3) by some weaker conditions

$$\sigma_l^2(I_k) \leqslant \sigma^2(I_k) \leqslant \sigma_u^2(I_k), \qquad k = 1, \ldots, m$$

then generally we obtain the smaller number m of intervals in the partition. For the detailed description of this procedure we refer to Höhenrieder (2007).

The result of the application of the second estimation procedure to data generated from the test model (4.2) is plotted on Figure 5.1 together with the mean absolute increments

$$\frac{1}{\sharp\{i \,|\, X_i = X_{(j)}\}h} \sum_{i \,|\, X_i = X_{(j)}} |X_{i+1} - X_i|, \qquad j = 1, \dots \tag{5.4}$$

and the true volatility function $0.5\big(1 + \big|\sin(\pi x)\big|\big)$. We can see that the accordance of the true and estimated volatility functions on the interval of the value range of X_i's from 1% till 99% is very good.

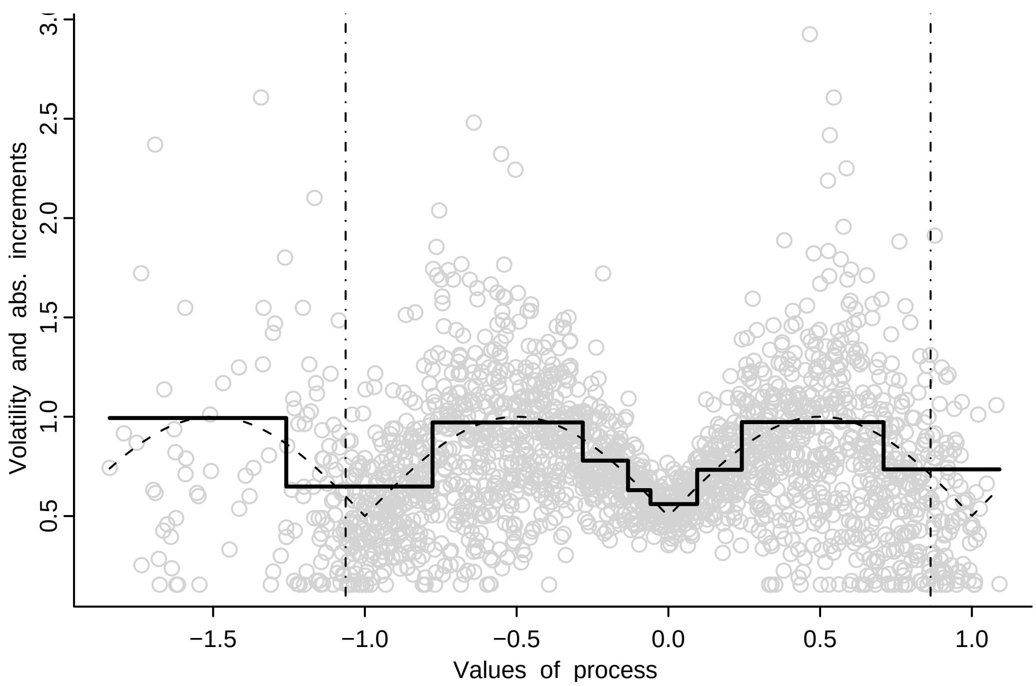

Fig. 5.1: Dashed line - true volatility function $0.5\big(1 + \big|\sin(\pi x)\big|\big)$;
Solid line - the piecewise constant estimation of the volatility;
Gray points - the mean absolute increments defined in (5.4);
Vertical lines – 1% and 99% quantiles of values range interval of X_i's

For every N and any subinterval I of the integer interval $1, \dots, N$ one can define a corresponding "continuous" interval $\tilde{I}$ as that subinterval $[x, y)$ of $[0, 1]$ which covers all values $\{X_{(j)},\, j \in I\}$. If the interval $\tilde{I}$ includes $X_{(1)}$ (i.e. $1 \in I$) than we put $x = 0$ otherwise x is defined as a half-sum of the smallest of $X_{(j)}$'s from $\tilde{I}$ and its previous value in the ordered sequence $\{X_{(j)}\}$. Similarly, if $\tilde{I}$ includes $X_{(N)}$ than $y = 1$ and we suppose the right end of $\tilde{I}$ is included otherwise y is a half-sum of the largest of $X_{(j)}$'s from $\tilde{I}$ and its next value from the sequence $\{X_{(j)}\}$. Thus the partition $\mathcal{I}(N)$ introduced above defines a set of "discrete" intervals of indexes $\{I\}$ and at the same time partitions the interval $[0, 1]$ and defines a set $\tilde{\mathcal{I}}(N)$ of "continuous" intervals $\{\tilde{I}\}$. For a point $x \in [0, 1]$ we will denote by

$\tilde{I}(x, N)$ – the constancy interval of the estimator $\hat{\sigma}(\cdot)$ which contains the point x and by

$I(x, N)$ – the "discrete" interval that corresponds to $\tilde{I}(x, N)$. (5.5)

For the formulation of the main result of this section we need some additional notations. We put

$$\varphi(N) = \frac{1}{1 - \alpha_N} \tag{5.6}$$

and notice that for the default choice of α_N (5.2) and in general for

$$\alpha_N \sim 1 - \frac{1}{N^\gamma}$$

with any $\gamma > 1$ we have

$$\varphi(N) \sim N^\gamma \qquad \text{and} \qquad \log \varphi(N) \sim \log N.$$

The next theorem is the main result of this section. It states uniform consistence and convergence rate of the estimator $\hat{\sigma}$.

Theorem 5.1

1. *Suppose the volatility function σ is continuous and satisfies*

$$\inf_{x,y \in [x_1, x_2)} \frac{|\sigma(x) - \sigma(y)|}{|x - y|} > 0. \tag{5.7}$$

on an interval $[x_1, x_2) \subset [0, 1]$. Then for any $\varepsilon > 0$ there exists some positive constant A such that

$$\lim_{N \to \infty} \mathbb{P}\left\{ \sup_{x \in [x_1 + \varepsilon, x_2 - \varepsilon)} \left| \sigma^2(x) - \hat{\sigma}^2(x) \right| \leqslant A \max\left(\frac{(\log \varphi(N))^{2/3}}{N^{1/3}}, \frac{(\log \varphi(N))^{1/2}}{\varphi^{1/4}(N)} \right) \right\} = 1. \tag{5.8}$$

2. *If the function* σ *is constant on an interval* $[x_1, x_2) \subset [0,1]$ *then for any* $\varepsilon > 0$ *there exists some positive constant* A *such that*

$$\lim_{N\to\infty} \mathbb{P}\left\{ \sup_{x\in[x_1+\varepsilon, x_2-\varepsilon)} \left|\sigma^2(x) - \hat{\sigma}^2(x)\right| \leqslant \frac{A\sqrt{\log\varphi(N)}}{\min\left(N^{1/2}, \varphi^{1/4}(N)\right)} \right\} = 1. \quad (5.9)$$

Chapter 6

Estimation of the drift term

For a given function $y(\cdot)$ on $[0,1]$ we consider its integral

$$y(x) = \int_0^x y(v)dv$$

and for $C_{\mathcal{T}} > 0$ define the taut string $s\big(\cdot, C_{\mathcal{T}}\big)$ as a function with the smallest length

$$\int_0^1 \sqrt{1 + \Big(s^{(1)}\big(x, C_{\mathcal{T}}\big)\Big)^2}\, dx,$$

which lies in the Kolmogorov tube

$$\mathcal{T}\big(y, C_{\mathcal{T}}\big) = \{f(\cdot) \,|\, y(x) - C_{\mathcal{T}} \leqslant f(x) \leqslant y(x) + C_{\mathcal{T}},\, x \in [0,1]\}$$

and satisfies $s(0) = y(0)$ and $s(1) = y(1)$. It is clear that the taut string $s\big(\cdot, C_{\mathcal{T}}\big)$ is a piecewise linear function and its derivate $s\big(\cdot, C_{\mathcal{T}}\big)$ is piecewise constant and has the minimal modality amongst all functions which lie in the tube $T\big(y, C_{\mathcal{T}}\big)$ and satisfy the edge conditions. For other properties of taut strings we refer to Davies and Kovac (2001) and Barlow et al. (1972). An example of the taut string is showed in Figure 6.1.

6.1 Rounding problem

As mentioned above for the estimation of the drift function we adopt the taut string method developed by Davies and Kovac (2001). They considered a nonparametric regression model

$$Y(x_i) = b(x_i) + \sigma\xi_i, \qquad i = 1, \ldots, M \tag{6.1}$$

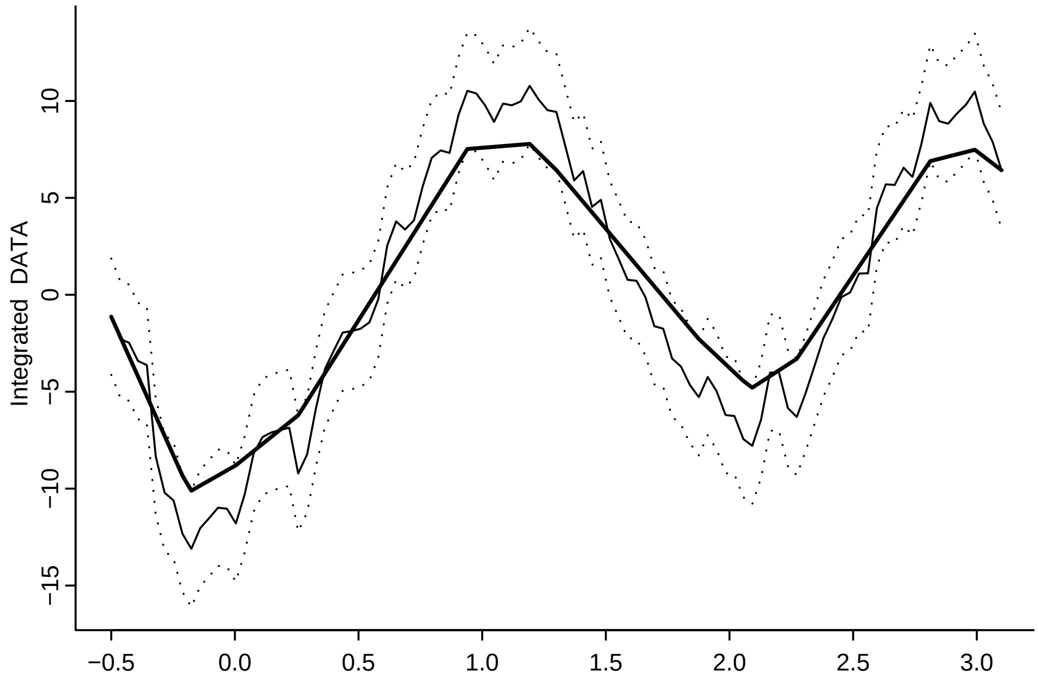

Fig. 6.1: Solid line – the integrated DATA, e.g. function $y(x)$;
Dashed line – the $C_{\mathcal{T}}$-width Kolmogorov tube;
Solid bold line – the taut string;

assumig that the all x_i's are different and strictly ordered. But it is not the case of the interest rate homogeneous diffusion models, where original data are only available rounded up to the second digit after the decimal point and therefore at every point x_i can be observed more than one response $Y_j(x_i)$. Therefore instead of the model (6.1) we consider

$$Y(x_i, j) = b(x_i) + \sigma(x_i)\xi_{ij}, \qquad j = 1, \dots, m_i, \quad i = 1, \dots, M \tag{6.2}$$

with the weigths $m_i = \sharp\{j \,|\, x_j = x_i\}$. There are two approaches to construction of the taut string and computing of the multiresolution coefficients [22, 1.5]. In the first one using mean responses we alter the last model as

$$Y_i' := \sum_{j \,|\, x_j = x_i} \frac{Y(x_i, j)}{m_i} = b(x_i) + \sigma(x_i)\xi_i', \qquad i = 1, \dots, M,$$

where

$$\xi_i' = \frac{1}{m_i} \sum_{j \mid x_j = x_i} \xi_{ij} \sim \mathcal{N}\left(0, \frac{1}{\sqrt{m_i}}\right)$$

and the values $\sqrt{m_i}\xi_i' \sim \mathcal{N}(0,1)$ are used as residuals r_i for the check of the multiresolution conditions [22, 1.6]. The number of observations used for the threshold in [22, 1.6] is the number of all different x_j's, i.e. M. In the second approach we use total responses and have the model

$$Y_i'' := \sum_{j \mid x_j = x_i} Y(x_i, j) = m_i b(x_i) + \sigma(x_i)\xi_i'', \qquad i = 1, \ldots, M,$$

where

$$\xi_i'' = \sum_{j \mid x_j = x_i} \xi_{ij} \sim \mathcal{N}\left(0, m_i\right).$$

Then for the multiresolution coefficients [22, 1.5] we take the residuals

$$r_i = \frac{Y_i'' - m_i b(x_i)}{\sigma(x_i)} \equiv \xi_i''$$

and compute the coefficients as

$$w_{k,l} = \frac{1}{\sqrt{\sum_{k<i\leqslant l} m_i}} \sum_{k<i\leqslant l} m_i r_i,$$

and the number of observations for the threshold value is the number of all x_j's, i.e. the sum of all weights $\sum_{1\leqslant i\leqslant M} m_i$.

The original interest rate data sets have more observations at the middle of the values interval and less at the edges of it. Both approaches work well on the test models with similar properties even with relatively weak signal. For the test model

$$Y_j = -\frac{x_j}{4} + \xi_j, \qquad \xi_j \sim \mathcal{N}(0,1), \quad j = 1, \ldots, N$$

with 1001 different x_j's from $[-1,1]$, with weights proportional to the normal density, and $N = 24151$ the results of the approaches are plotted on figure 6.2.

In the context of diffussion processes the weights m_i are proportional to the invariant density function and because of the strong law of large numbers for ergodic processes (2.6) and also relations (6.8), (6.9) and (6.11) we prefer in our framework the second approach. The exact constructions follow.

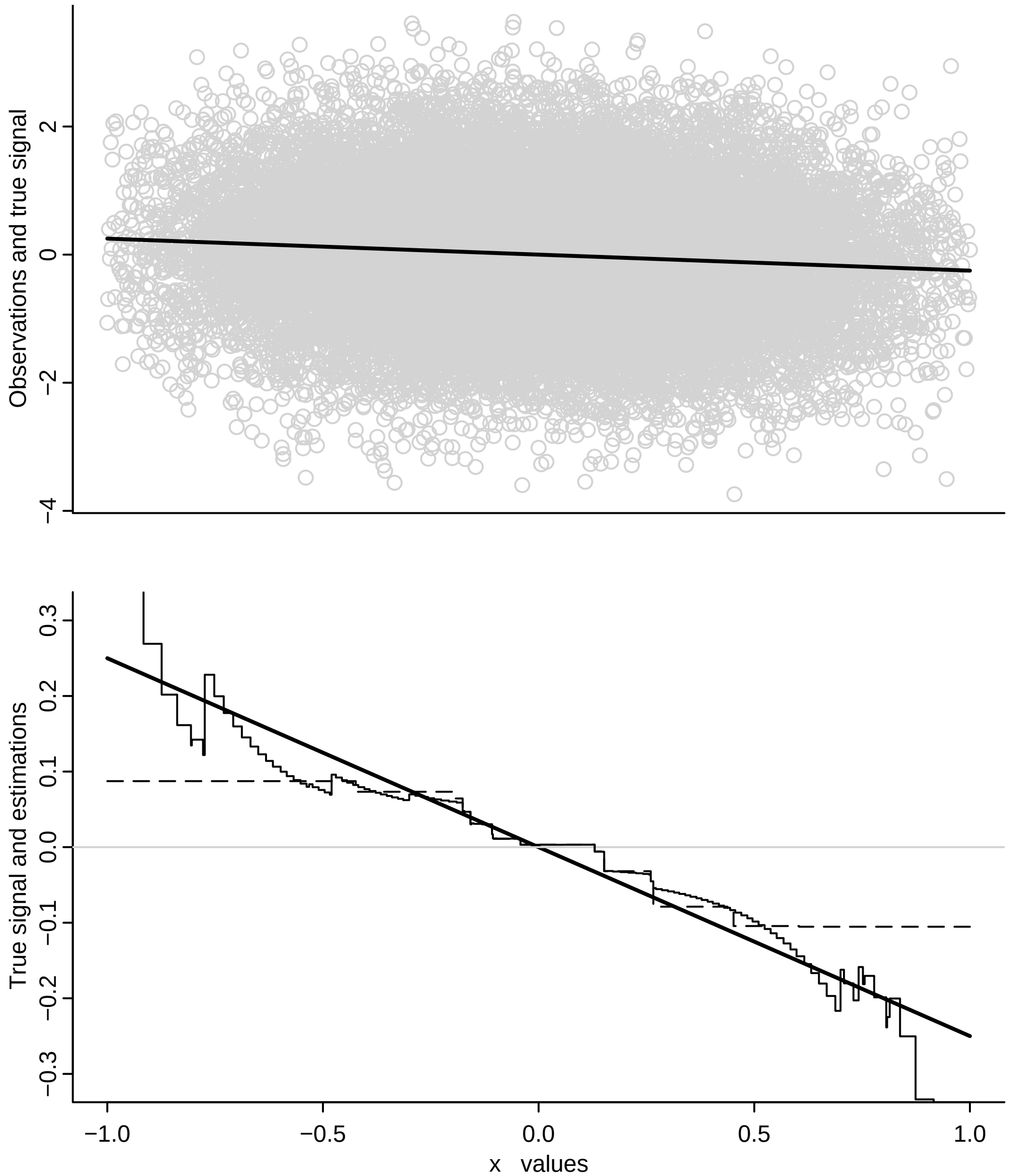

Fig. 6.2: Two approaches to nonparametric regression with weights;
The upper panel: Gray points – observations; Black line - the signal;
The lower panel: Solid line – the signal;
Dashed line – the estimation using mean responses;
Thin line – the estimation using total responses;

6.2 Construction of the estimator

In the context of Tukey's data decomposition we identify the increments $X_{j+1} - X_j$ with the DATA, the term $b(X_j)h$ with the SIGNAL, and the term $\sigma(X_j)\xi_j\sqrt{h}$ where the volatility σ is already known with the NOISE. The corresponding integrated process is

$$y_N(x) = \frac{1}{Nh} \sum_{0 \leqslant j < N} \mathbb{1}_{\{X_j \in [0,x)\}}(X_{j+1} - X_j), \tag{6.3}$$

where $\mathbb{1}_A$ is indicator of event A. Also we define the integrated SIGNAL and NOISE respectively

$$f_N(x) = \frac{1}{N} \sum_{0 \leqslant j < N} \mathbb{1}_{\{X_j \in [0,x)\}} b(X_j), \tag{6.4}$$

$$s_N(x) = \frac{1}{N\sqrt{h}} \sum_{0 \leqslant j < N} \mathbb{1}_{\{X_j \in [0,x)\}} \sigma(X_j)\xi_j, \tag{6.5}$$

and additionally

$$\mu_N(x) = \frac{1}{N} \sum_{0 \leqslant j < N} \mathbb{1}_{\{X_j \in [0,x)\}}. \tag{6.6}$$

If we put

$$f(x) = \mu(x)b(x) \tag{6.7}$$

then for $0 \leqslant x_1 \leqslant x_2 \leqslant 1$ from (2.6) it follows

$$\lim_{N\to\infty} \big(f_N(x_2) - f_N(x_1)\big) = \lim_{N\to\infty} \frac{1}{N} \sum_{0 \leqslant j < N} b(X_j) \mathbb{1}_{\{X_j \in [x_1,x_2)\}} \overset{a.s.}{=} \int_{x_1}^{x_2} f(x)dx \tag{6.8}$$

and analogously

$$\lim_{N\to\infty} \big(\mu_N(x_2) - \mu_N(x_1)\big) \overset{a.s.}{=} \int_{x_1}^{x_2} \mu(x)dx. \tag{6.9}$$

The convergence is uniform in $x_1, x_2 \in [0,1]$ and its rate is stated below. Now for some $C_{\mathcal{T}} > 0$ we denote the taut string that lies in the Kolmogorov tube $\mathcal{T}\big(y_N, C_{\mathcal{T}}/\sqrt{N}\big)$ by $s_N(x, C_{\mathcal{T}})$ and let the function $s_N(x) \equiv s_N(x, C_{\mathcal{T}})$ be the right derivative of $s_N(x, C_{\mathcal{T}})$ for $0 \leqslant x < 1$ and left derivative for $x = 1$. We require also some notation for the local extreme points of the functions

f and s_N. Let the function f have $K(f)$ local extremes p_i^e on the interval $(0,1)$. Denote by $\{p_i^l, p_i^r\}$, $i=1,\dots,K_N(C_{\mathcal{T}})$ such pairs of successive knots x_{j_i}, $i=1,\dots,K_N(C_{\mathcal{T}})$ of the taut string $s_N(\cdot)$ for which $s_N(\cdot)$ attains its local extremes and let m_i be the midpoint of the interval (p_i^l, p_i^r). It is worth pointing out that the values $K_N(C_{\mathcal{T}})$, p_i^l, p_i^r, m_i are random as properties of the random taut string while $K(f)$ and p_i^e are not. The following statement demonstrates that the estimator $s_N(\cdot)$ is consistent in the number of local extremes and its locations.

Theorem 6.1 *If the function f has a continuous first derivative on $[0,1]$ and $f^{(1)}(x)=0$ only for $x=p_i^e$, $i=1,\dots,K(f)$ then for all $\varepsilon>0$*

$$\lim_{C_{\mathcal{T}}\to\infty}\lim_{N\to\infty}\mathbb{P}\Bigg(\Big\{K_N(C_{\mathcal{T}})=K(f)\Big\}\bigcap\Big\{\max\nolimits_{1\leqslant i\leqslant K(f)}(p_i^r-p_i^l)\leqslant\varepsilon\Big\}\bigcap$$
$$\bigcap\Big\{\max\nolimits_{1\leqslant i\leqslant K(f)}|m_i-p_i^e|\leqslant\varepsilon\Big\}\Bigg)=1. \quad (6.10)$$

Let $\mathcal{Y}_N=[0,1]\setminus\bigcup_{i=1}^K[p_i^l,p_i^u]$ and $\mathcal{Y}_N^e=\bigcup_{i=1}^K[p_i^l,p_i^u]$. The next theorem substantiates the use of the derivative s_N of the taut string as an approximation of the function f and gives the measure of closeness between them.

Theorem 6.2 *Let the function f satisfy the conditions of Theorem 6.1 and additionally have a bounded second derivative $f^{(2)}$ which differs from zero at the local extremes of f. Then*

1. $\lim\limits_{C_{\mathcal{T}}\to\infty}\lim\limits_{N\to\infty}\mathbb{P}\left(p_i^e\in[p_i^l,p_i^u),\ i=1,...,K\right)=1.$
2. *For any $\varepsilon>0$ and all $i=1,\dots,K$*

$$\lim_{C_{\mathcal{T}}\to\infty}\lim_{N\to\infty}\mathbb{P}\left(\left|\frac{N^{1/6}|f^{(2)}(p_i^e)|^{1/3}(p_i^u-p_i^l)}{(24C_{\mathcal{T}})^{1/3}}-1\right|\leqslant\varepsilon\right)=1.$$

3. *Let x_j be a knot of $s_N(\cdot)$ such that x_j and x_{j+1} are either both on the upper $bu_N(\cdot):=y_N(\cdot)+C_{\mathcal{T}}/\sqrt{N}$ or both on the lower $bl_N(\cdot):=y_N(\cdot)-C_{\mathcal{T}}/\sqrt{N}$ bound of the tube $\mathcal{T}\big(y_N,C_{\mathcal{T}}/\sqrt{N}\big)$. Then for some $A>0$*

$$\lim_{C_{\mathcal{T}}\to\infty}\lim_{N\to\infty}\mathbb{P}\left(\max_j(x_{j+1}-x_j)|f^{(1)}(x_j)|^{2/3}\leqslant A\left(\frac{\log N}{N}\right)^{1/3}\right)=1.$$

4. There exists $A > 0$ such that

$$\lim_{C_{\mathcal{T}}\to\infty} \lim_{N\to\infty} \mathbb{P}\left(\sup_{x\in\mathcal{Y}_N} \frac{|f(x) - s_N(x)|}{|f^{(1)}(x)|^{1/3}} \leqslant A \left(\frac{\log N}{N} \right)^{1/3} \right) = 1.$$

5. There exists $A > 0$ such that

$$\lim_{C_{\mathcal{T}}\to\infty} \lim_{N\to\infty} \mathbb{P}\left(\sup_{x\in\mathcal{Y}_N^e} \frac{|f(x) - s_N(x)|}{|f^{(2)}(x)|^{1/3}} \leqslant A C_{\mathcal{T}}^{2/3} N^{-1/3} \right) = 1.$$

Relying on this theorem we consider the functions s_N as approximations of the function f and because of the definitions (6.3), (6.4), and (6.5) we have

$$y_N(x) = f_N(x) + \varepsilon_N(x)$$

and thus we expect that the differences

$$r_N(x) := y_N(x) - s_N(x),$$

which we call integrated residuals, approximate the integrated NOISE $\varepsilon_N(x)$. We now study the behaviour of $\varepsilon_N(x)$ for large N which will provide an appropriate choice of the constant $C_{\mathcal{T}}$.

Let $W(x) = W(\omega, x)$, $\omega \in \Omega$, $x \in [0, 1]$ be a random process with independent increments and

$$\begin{aligned} W(0) &\overset{a.s.}{=} 0, \\ W(x_2) - W(x_1) &\overset{d}{=} \mathcal{N}\left(0, \frac{1}{h}\int_{x_1}^{x_2} \sigma^2(x)\mu(x)dx\right), \quad 0 \leqslant x_1 \leqslant x_2 \leqslant 1 \end{aligned} \qquad (6.11)$$

Theorem 6.3 *On* $\mathrm{D}\big([0,1]\big)$ *the following weak convergence takes place:*

$$\sqrt{N}\varepsilon_N(\cdot) \Rightarrow W(\cdot) \text{ as } N \to \infty. \qquad (6.12)$$

Now we consider the set of all different ordered values $\{v_i\}$, $i = 0, \ldots, M$ taken by the process $X(t)$ and using the last theorem we have that for large N

$$\sqrt{N}\Big(\varepsilon_N(v_{i+1}) - \varepsilon_N(v_i)\Big) \sim \mathcal{N}\left(0, \frac{1}{h}\int_{v_i}^{v_{i+1}} \sigma^2(x)\mu(x)dx\right), \quad 0 \leqslant i < M.$$

We can say that the integrated residuals $\mathcal{R}_N(x)$ approximate the integrated NOISE $\mathcal{W}_N(x)$ if the sequence of random variables

$$Y_i := \sqrt{N}\Big(\mathcal{R}_N(v_{i+1}) - \mathcal{R}_N(v_i)\Big)$$

looks like a sequence of centered Gaussian random values with the corresponding variances. In order to check it we define the coefficients

$$w(t_1, t_2) = \sqrt{\frac{h}{\int_{t_1}^{t_2} \sigma^2(v)\mu(v)dv}} \sum_{i \,|\, t_1 \leqslant v_i < t_2} Y_i, \quad 0 \leqslant t_1 \leqslant t_2 \leqslant 1$$

and use the multiresolution conditions introduced by Davies and Kovac (2001). Namely we check whether the inequalities

$$|w(t_1, t_2)| \leqslant \sqrt{\tau \log N} \tag{6.13}$$

hold for all $0 \leqslant t_1 \leqslant t_2 \leqslant 1$ and for some $\tau > 0$. The idea of that bound is based on the limit relation for some $\tau > 0$

$$\lim_{n\to\infty} \mathbb{P}\left\{\max_{1\leqslant k\leqslant n} |Z_i| \leqslant \sqrt{\tau \log n}\right\} = 1 \tag{6.14}$$

where Z_i's are random variables with a common subgaussian distribution. Relations between the multiresolution coefficients $w(t_1, t_2)$ and the bound in (6.13) is given by the next theorem.

Theorem 6.4 *For each constant $A > 0$ there exists $\tau > 0$ such that the following hold:*

1. *For any multiresolution coefficient $w(t_{1,N}, t_{2,N})$ such that $t_{2,N} - t_{1,N} \leqslant A(\log N/N)^{1/3}$ the inequality (6.13) is true for all sufficiently large N.*

2. *For any multiresolution coefficient $w(x_j, x_{j+1})$ where x_j's are knots of the taut string $\mathcal{S}_N$ and the interval $[x_j, x_{j+1}]$ is neither of the extreme intervals $[p_i^l, p_i^r]$ the inequality (6.13) is true for all sufficiently large N.*

3. *For any multiresolution coefficient $w(t_{1,N}, t_{2,N})$ such that*

$$\liminf_{N\to\infty} \frac{t_{2,N} - t_{1,N}}{\log N} > 0$$

the inequality (6.13) is true for all sufficiently large N.

4. For any extreme interval $[p_i^l, p_i^r]$ there exists a subinterval $[p'_{1,N}, p'_{2,N}]$ such that for the multiresolution coefficient $w(p'_{1,N}, p'_{2,N})$ the inequality (6.13) eventually does not hold whatever the value of τ.

If in (6.14) the Z_i's are standard normal random variables then the relation (6.14) holds for any $\tau > 2$. But for our check of the multiresolution conditions (6.13) we choose $\tau = 2.5$ since it seems to give better results.

The checking of the multiresolution conditions requires an identification of the invariant density μ. For this goal we use the method developed by Davies and Kovac (2004), which also results a piecewise constant function. Figure 6.3 shows the estimation of the invariant density together with the true density function and the data histogram of our test model.

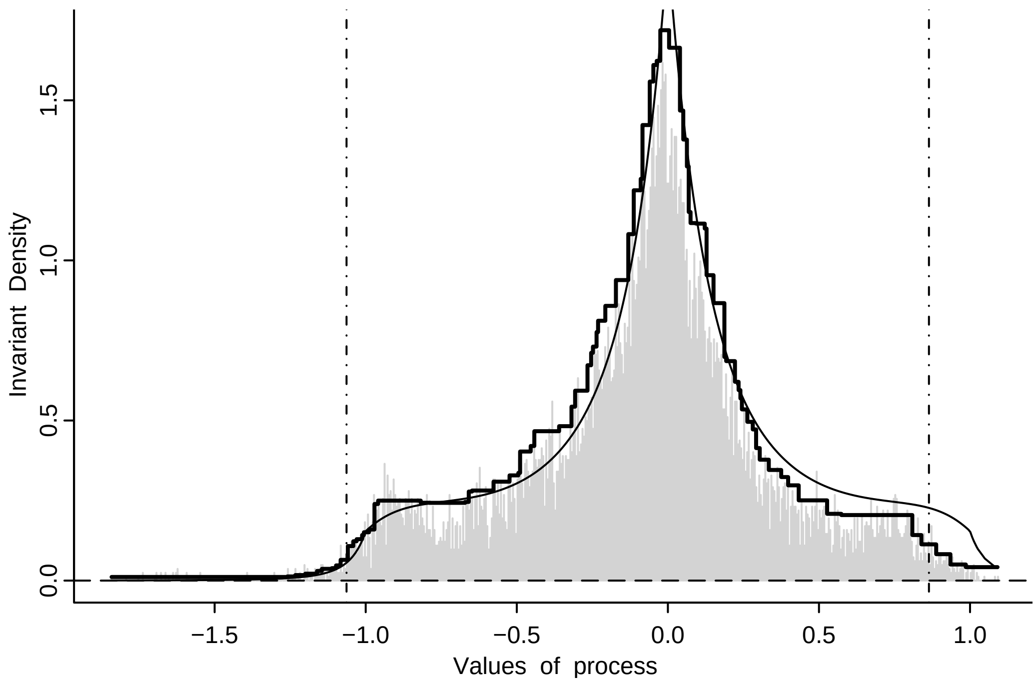

Fig. 6.3: Solid thin line - true invariant density of the process (4.2);
Solid bold line - the piecewise constant estimation of the invariant density;
Gray vertical lines - the data histogram;
Dashed vertical lines – 1% and 99% quantiles of the values range interval of X_i's

We now describe the algorithm of the choice of the tube width $C_{\mathcal{T}}$. Initially we take it so that the taut string s_N has one linear segment only and check the multiresolution conditions (6.13). If all of the inequalities hold then we stop the process, otherwise we reduce the width $C_{\mathcal{T}}$ and perform the next iteration.

For the better data accordance the final taut string can be altered in such a manner that its knots will be placed not on the borders of the supremum tube $\mathcal{T}$ but on the integrated data. This is demonstrated by Figure 6.4 where the final taut string for our test process (4.2), the corresponding integrated increments and the supremum tube are displayed. But such changes can shift the taut string out from the tube. In this case either the additional corresponding knots have to be incorporated or the changes should not be applied.

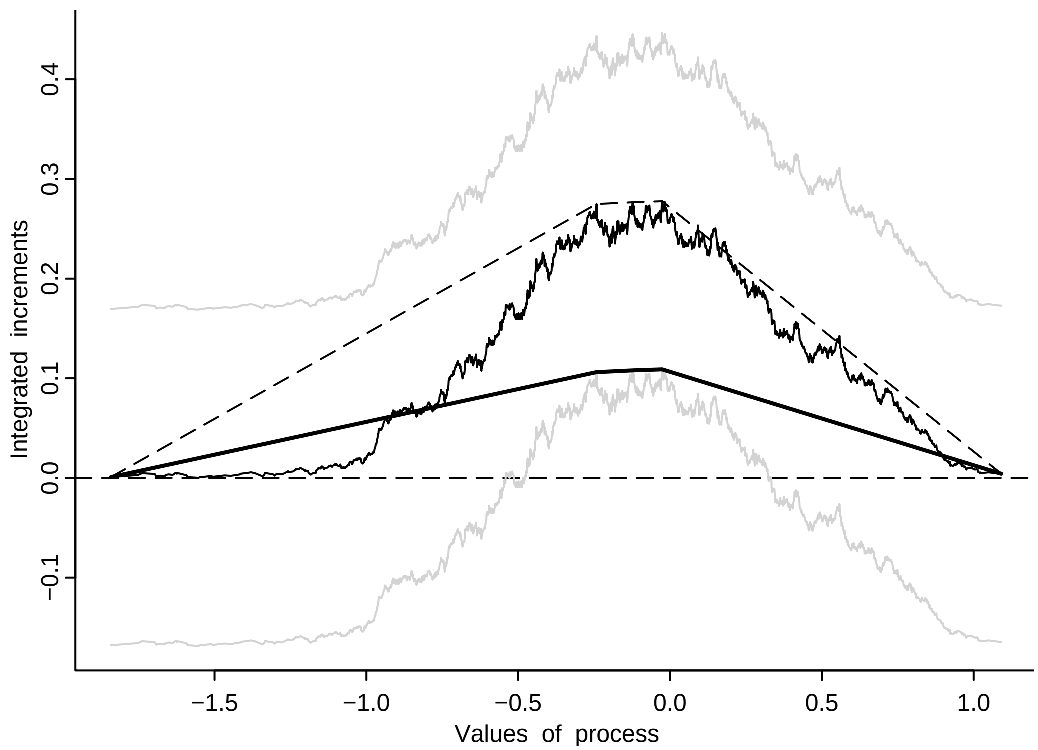

Fig. 6.4: Solid thin line - the integrated increments of a sample path of the process (4.2);
Dashed line - the altered taut string;
Solid bold line - the final taut string;
Gray lines - the bounds of the tube $\mathcal{T}$;

Another possibility to improve the data accordance is to squeeze the tube $\mathcal{T}$ between the iterations inversely to the data density - stronger at the edges and weaker in the middle of the data range interval. We apply such a squeezing both for the test model (Figure 6.5) and the original data (Figure 7.2).

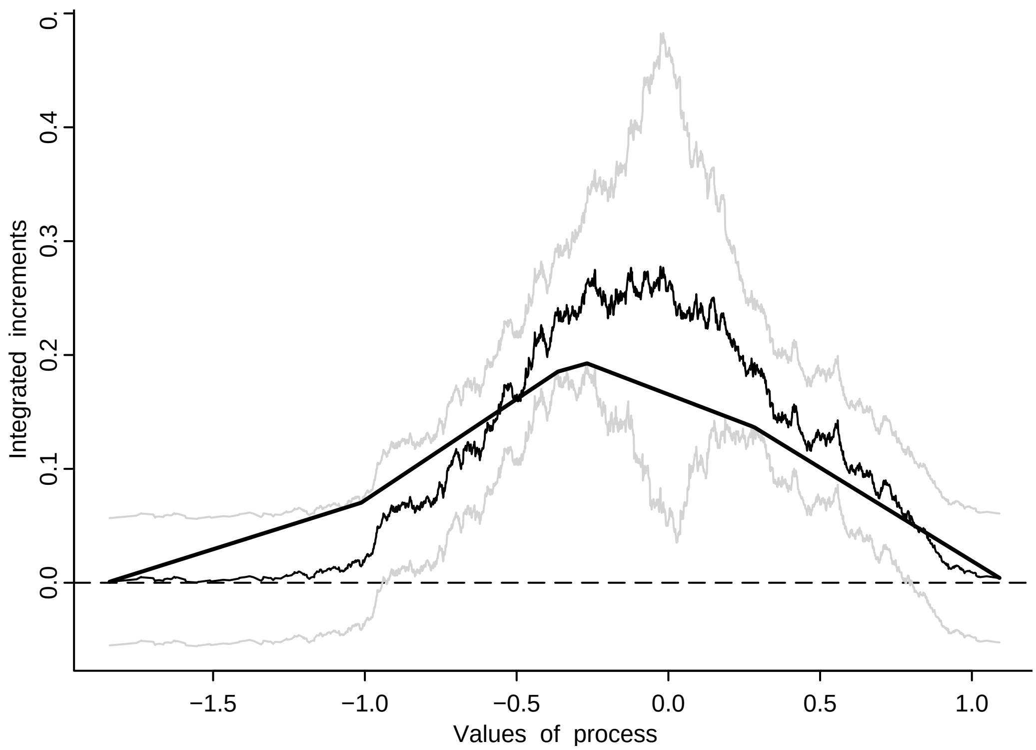

Fig. 6.5: Solid thin line - the integrated increments of a sample path of the process (4.2);
Solid bold line - the final taut string;
Gray lines - the bounds of the tube $\mathcal{T}$;

As the derivative $s_N(x)$ of the taut string aprroximates the function $f(x)$, i.e. the product of the drift term and the invariant density function, in order to obtain the estimation of the drift we must divide the derivative s_N by the density μ. For the test model the derivative of the final taut string is presented in Figure 6.6 while the drift term estimation in Figure 6.7. We note that the drift estimator well agrees with the true drift function.

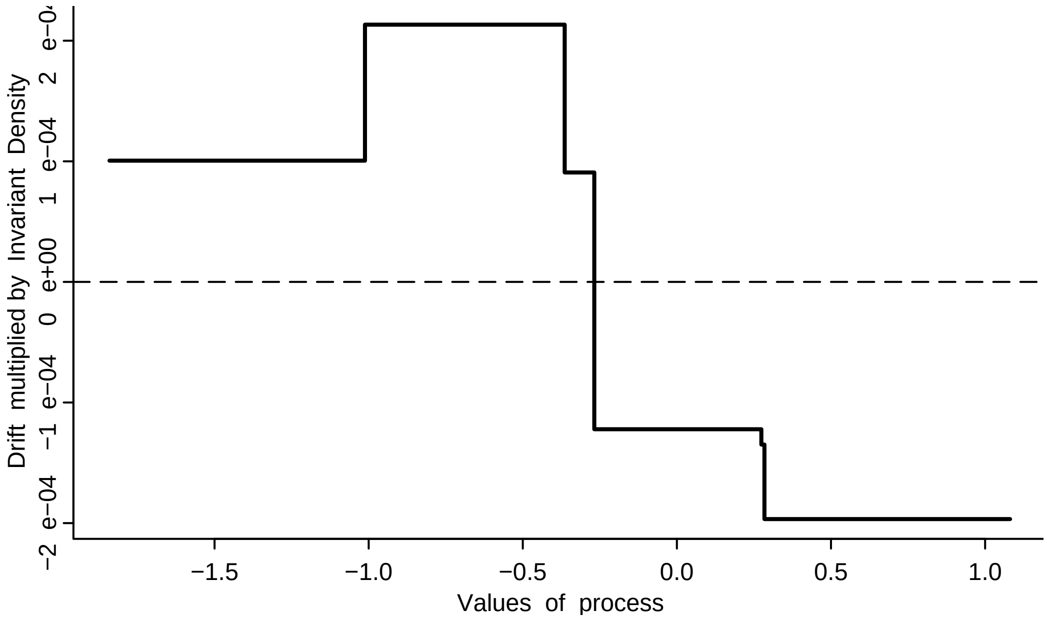

Fig. 6.6: The derivative of the final taut string for the process (4.2);

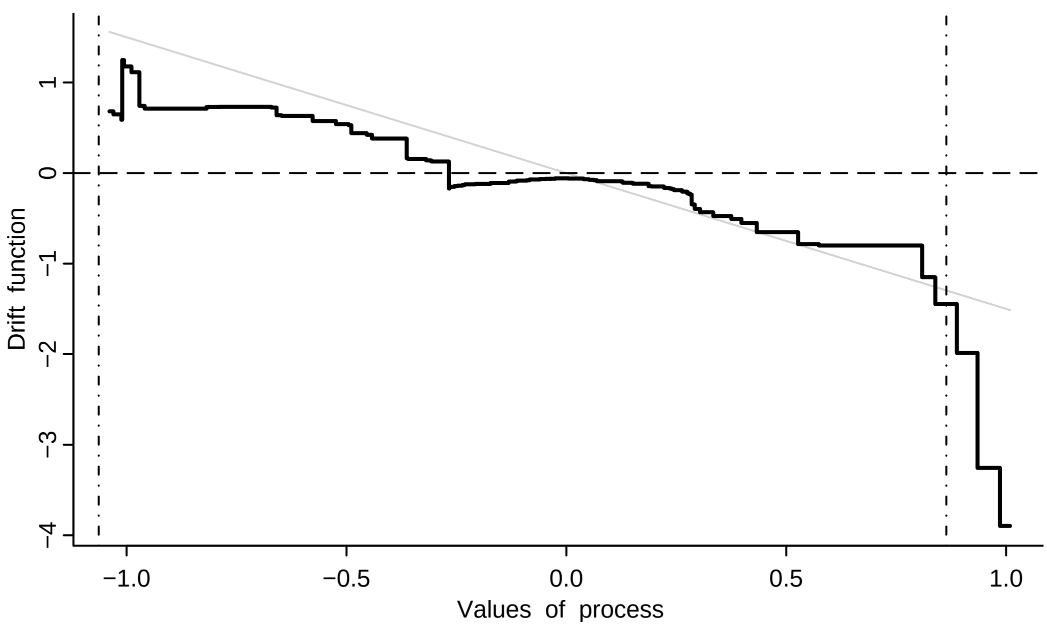

Fig. 6.7: Gray line - true drift function $-1.5x$;
Solid bold line - the piecewise constant estimation of the drift term;
Vertical lines – 1% and 99% quantiles of the values range interval of X_i's;

Chapter 7

Applying the methods to the real data

In this section we plot the results of applying of the estimation methods to the real data of 3-Month US Government Tresuary Bills.

Figure 7.1 shows the absolute values of the interest rate increments and the estimation of the volatility function.

Figure 7.2 represents the final taut string for the estimation of the drift term and the next Figure 7.2 shows the derivative of the taut string, i.e. the estimation of the product of the invariant density and the drift functions. The piecewise constant estimation of the invariant density function is plotted in Figure 7.3 while the final estimation of the drift in Figure 7.5.

As the volatility and drift function are estimated and the model is completely specified we can simulate a dynamic of the interest rate and compare it with the original data. Figure 7.6 shows the first 8 successive simulated trajectories which are plotted together with the original one. All 2000 simulations in JPEG format can be found on the CD enclosed to this thesis.

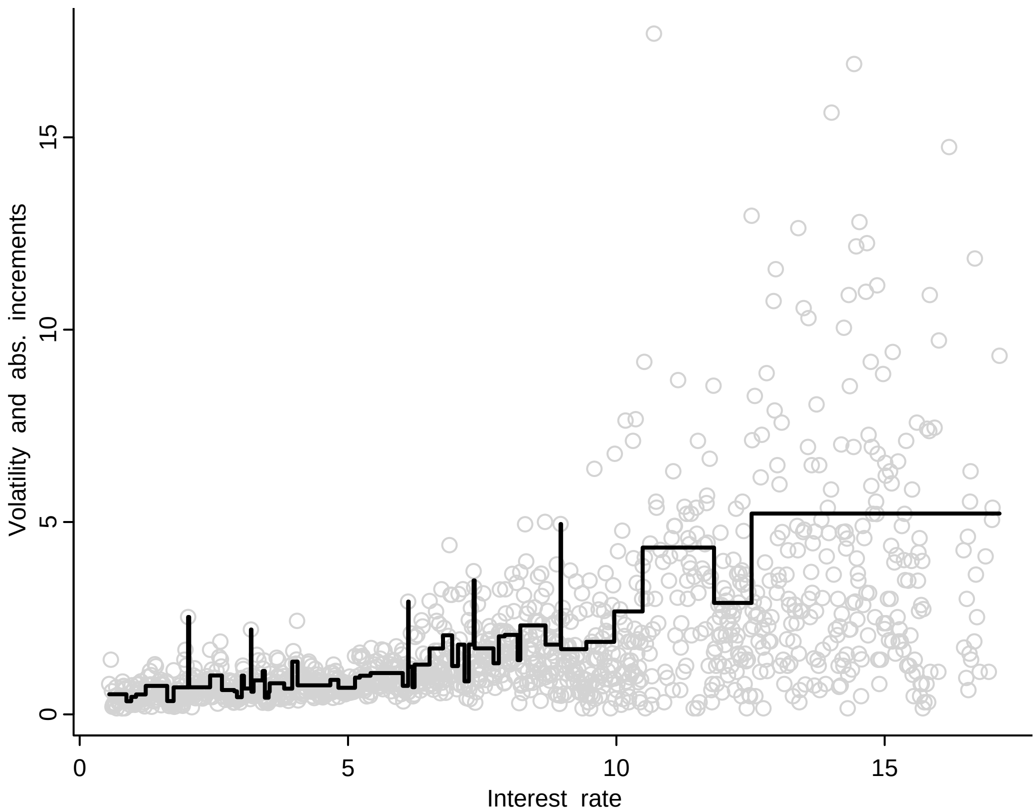

Fig. 7.1: 3-Month US Government Tresuary Bills
Solid line - the piecewise constant estimation of the volatility;
Gray points - the mean absolute increments defined in (5.4);

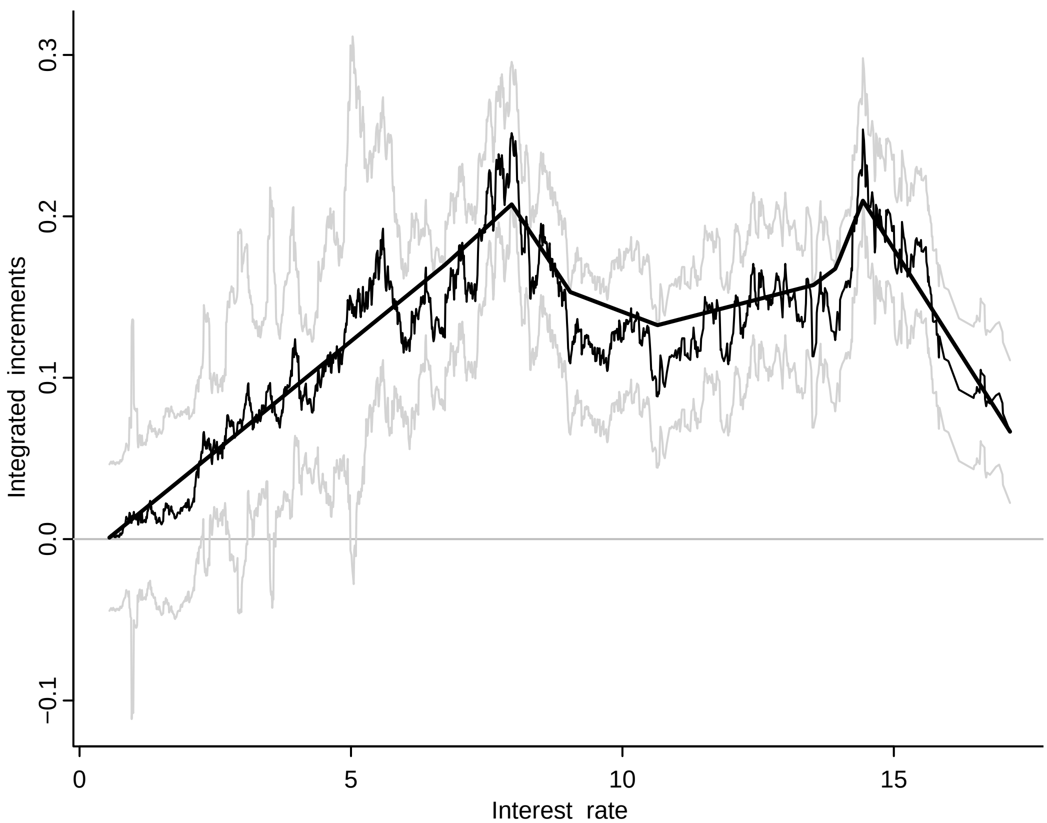

Fig. 7.2: 3-Month US Government Tresuary Bills
Solid think line - the cumulative integrated increments;
Solid bold line - the final taut string;
Gray lines - the bounds of the tube $\mathcal{T}$;

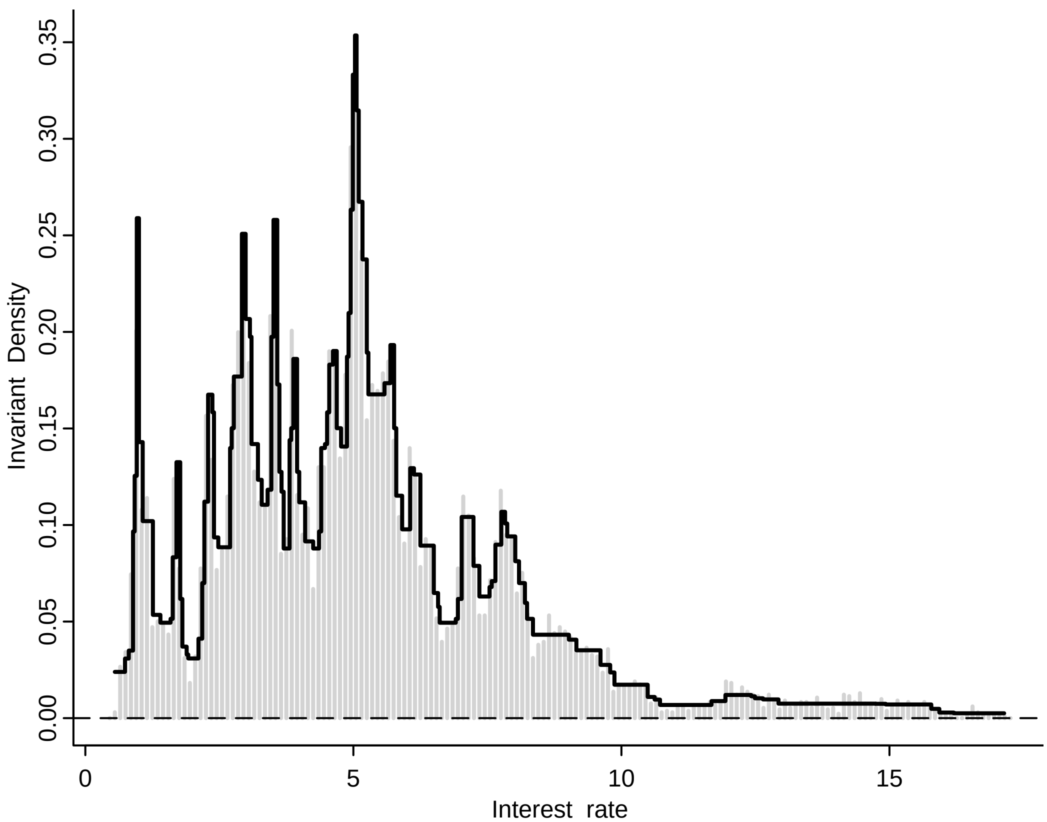

Fig. 7.3: 3-Month US Government Tresuary Bills
Solid bold line - the piecewise constant estimation of the invariant density;
Gray vertical lines - the interest rate data histogram;

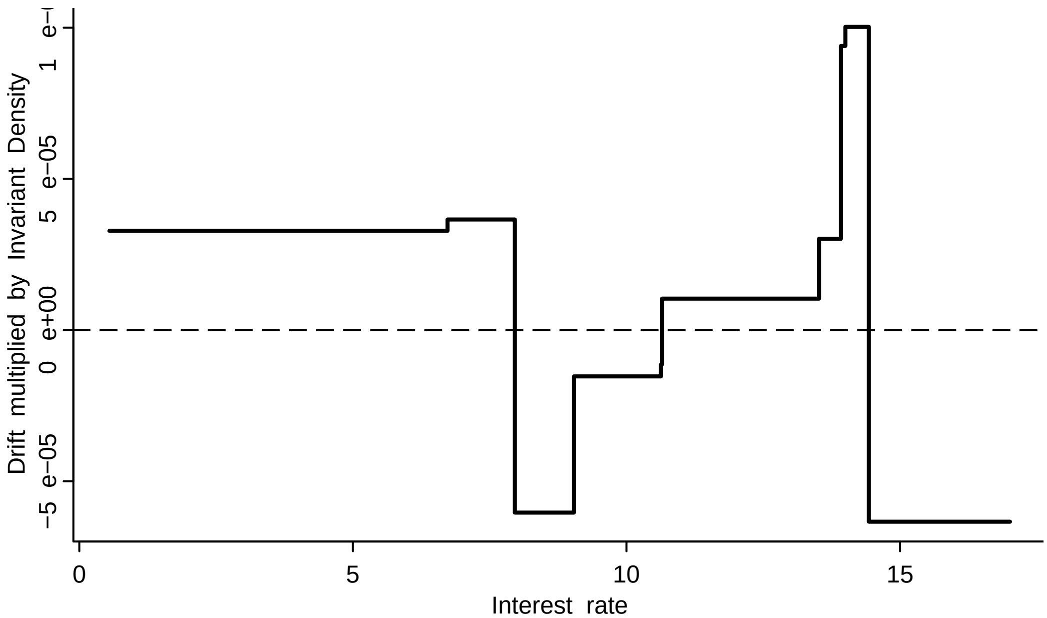

Fig. 7.4: 3-Month US Government Tresuary Bills
The derivative of the final taut string for the drift term estimation;

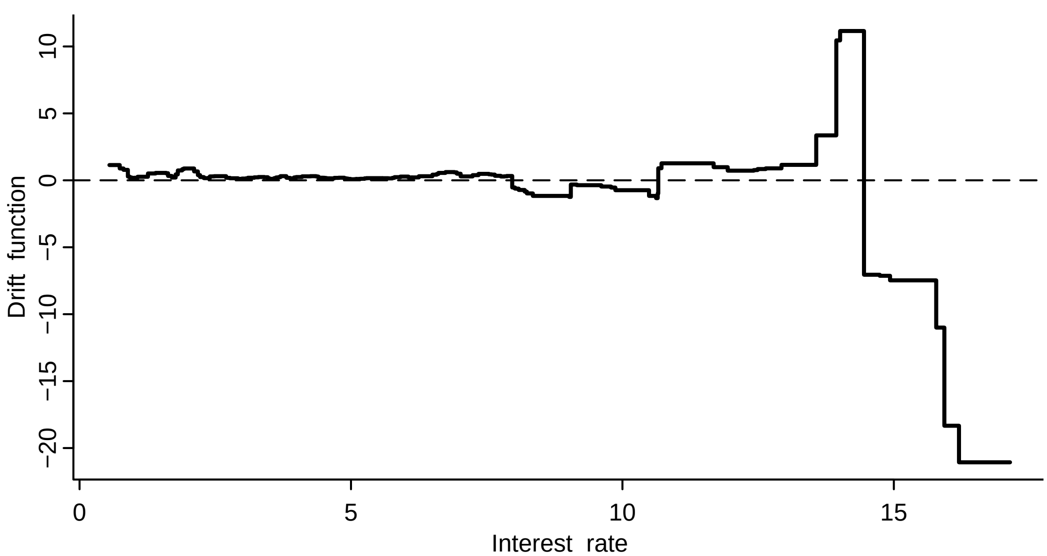

Fig. 7.5: 3-Month US Government Tresuary Bills
Solid bold line - the piecewise constant estimation of the drift term;

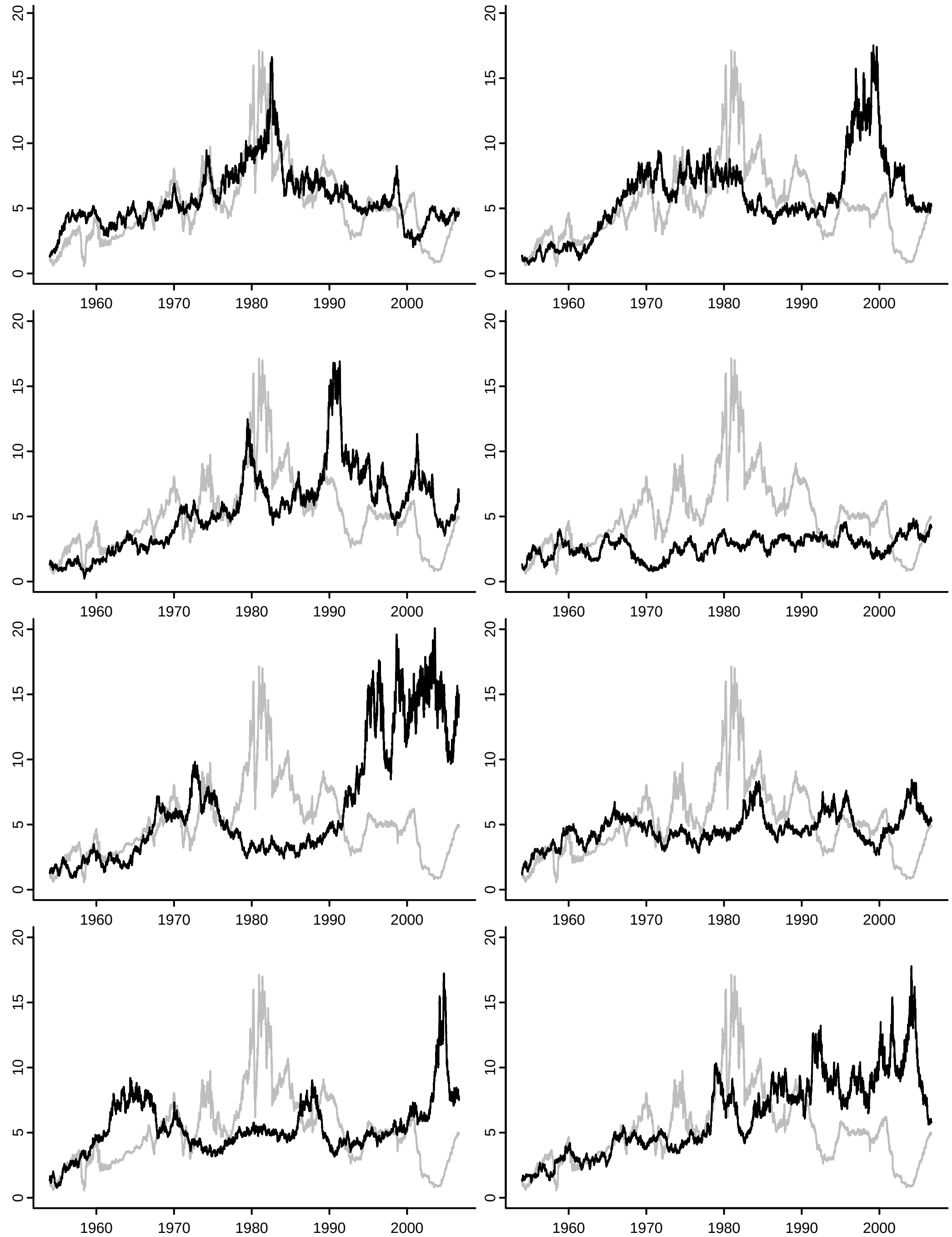

Fig. 7.6: First 8 successive simulations of 3-Month US Tresuary Bills;
Black line - the simulated path; Gray line - the original data;

Chapter 8

Proofs

8.1 Some auxiliary limit theorems and inequalities

Further on for any intervals of type $(x, y]$ or $[y, x)$, $x, y \in [0, 1]$ we suppose that the end x is included if $x = 0$ in the first case and $x = 1$ in the second one.

The proofs of the main results - Theorems 5.1, 6.2 and 6.4 - essentially relies on Theorem 6.3 for the drift term and on weak convergence of the process

$$\zeta_N(x) = \frac{1}{\sqrt{N}} \sum_{0 \leqslant j < N} \mathbb{1}_{\{X_j \in [0,x]\}} \sigma^2(X_j)(\xi_j^2 - 1)$$

to some Gaussian process, say $V(x)$, with independent increments and

$$\begin{aligned} V(0) &\overset{a.s.}{=} 0, \\ V(x_2) - V(x_1) &\overset{d}{=} \mathcal{N}\left(0, 2\int_{x_1}^{x_2} \sigma^4(x)\mu(x)dx\right), \quad 0 \leqslant x_1 \leqslant x_2 \leqslant 1. \end{aligned} \tag{8.1}$$

Therefore in the first place we proof (6.12) and show that

$$\zeta_N(\cdot) \Rightarrow V(\cdot) \text{ as } N \to \infty. \tag{8.2}$$

For this we will show that finite-dimensional distributions of $\sqrt{N}\varepsilon_N(\cdot)$ and $\zeta_N(\cdot)$ converge to those of the processes $V(\cdot)$ and $W(\cdot)$ respectively and after that we establish the tightness of the sequences $\sqrt{N}\varepsilon_N(\cdot)$ and $\zeta_N(\cdot)$. We carry out the proof for the case of the process $\xi_N(\cdot)$, while for another case it requires just some minor changes.

Let $\{\eta_j\}_{j\geqslant 0}$ be a sequence of mutually independent and independent from X_j and ξ_j standard normal random variables. Below we use bold symbols for elements of the D-dimensional real space ($D \geqslant 1$).

Lemma 8.1 *For $D \geqslant 1$ and a set of disjoint intervals $\tilde{I}_d = (x_d^l, x_d^r] \subseteq [0,1]$, $d = 1, \ldots, D$ we consider a sequence of random vectors $\{\boldsymbol{\psi}_j\}_{j\geqslant 0}$ with the components defined as*

$$\psi_j \equiv \psi_j(\tilde{I}_d) = \sqrt{2}\eta_j \mathbb{1}_{\{X_j \in \tilde{I}_d\}} \sigma^2(X_j),\ d = 1, \ldots, D. \tag{8.3}$$

Then a sequence of the sums

$$\boldsymbol{\Psi}_N = \frac{1}{\sqrt{N}} \sum_{0\leqslant j<N} \boldsymbol{\psi}_j \tag{8.4}$$

converges in distribution to a normal vector $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, R)$ with a zero expectation $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^D$ and a covariance matrix $R = (r_{db})_{d,b=1}^D$ with elements

$$r_{db} = \begin{cases} 2\displaystyle\int_{\tilde{I}_d} \sigma^4(v)\mu(v)dv, & \text{if} \quad d = b, \\ 0, & \text{if} \quad d \neq b. \end{cases}$$

Proof. Let $\mathbf{t} = (t_1, \ldots, t_D) \in \mathbb{R}^D$, $\mathbf{1}_j = \left(\mathbb{1}_{\{X_j \in \tilde{I}_1\}}, \ldots, \mathbb{1}_{\{X_j \in \tilde{I}_D\}}\right)$ and $< \cdot, \cdot >$ denotes an inner product in $\mathbb{R}^D$. By ergodicity (2.6) and because the intervals $\tilde{I}_d$'s are disjoint we have

$$\frac{1}{N}\sum_{0\leqslant j<N} < \mathbf{t}, \sigma^2(X_j)\mathbf{1}_j >^2 = \frac{1}{N}\sum_{0\leqslant j<N}\left(\sum_{1\leqslant d\leqslant D} t_d \sigma^2(X_j)\mathbb{1}_{\{X_j\in\tilde{I}_d\}}\right)^2 =$$

$$\frac{1}{N}\sum_{0\leqslant j<N}\sum_{1\leqslant d\leqslant D}\left(t_d\sigma^2(X_j)\mathbb{1}_{\{X_j\in\tilde{I}_d\}}\right)^2 = \sum_{1\leqslant d\leqslant D} t_d^2 \frac{\sum_{0\leqslant j<N}\sigma^4(X_j)\mathbb{1}_{\{X_j\in\tilde{I}_d\}}}{N} \xrightarrow[N\to\infty]{\text{a.s.}}$$

$$\sum_{1\leqslant d\leqslant D} t_d^2 \int_{\tilde{I}_d}\sigma^4(x)\mu(x)dx = \sum_{1\leqslant d\leqslant D}\frac{t_d^2 r_{dd}}{2} = \frac{< \mathbf{t}^2, \mathbf{r} >}{2},$$

where $\mathbf{t}^2 = (t_1^2, \ldots, t_D^2)$ and $\mathbf{r} = (r_{11}, \ldots, r_{DD})$. From here using conditional expectation for the characteristic function of $\boldsymbol{\Psi}_N$ we obtain

$$\mathbb{E}\exp\left(i < \mathbf{t}, \boldsymbol{\Psi}_N >\right) = \mathbb{E}\mathbb{E}\left(\exp\left(i < \mathbf{t}, \boldsymbol{\Psi}_N >\right) \Big| X_0, \ldots, X_N\right) =$$
$$\mathbb{E}\mathbb{E}\left(\exp\left(\frac{i}{\sqrt{N}} \sum_{0 \leqslant j < N} \sum_{1 \leqslant d \leqslant D} t_d \sqrt{2}\sigma^2(X_j) \mathbb{1}_{\{X_j \in \tilde{I}_d\}} \eta_j\right) \Big| X_0, \ldots, X_N\right) =$$
$$\mathbb{E}\mathbb{E}\left(\prod_{0 \leqslant j < N} \exp\left(i\sqrt{\frac{2}{N}} < \mathbf{t}, \sigma^2(X_j) \mathbf{1}_j > \eta_j\right) \Big| X_0, \ldots, X_N\right) = (\star),$$

and since η_j's are mutually independent and independent from X_j' we can continue the latter as

$$(\star) = \mathbb{E}\left(\prod_{0 \leqslant j < N} \exp\left(-\frac{\left(\sqrt{\frac{2}{N}} < \mathbf{t}, \sigma^2(X_j)\mathbf{1}_j >\right)^2}{2}\right)\right) =$$
$$\mathbb{E}\exp\left(-\frac{\sum_{0 \leqslant j < N} < \mathbf{t}, \sigma^2(X_j)\mathbf{1}_j >^2}{N}\right) \xrightarrow[N \to \infty]{} \exp\left(-\frac{< \mathbf{t}^2, \mathbf{r} >}{2}\right). \qquad \square$$

Lemma 8.2 *Finite-dimensional distributions of the process* $\xi_N(x), \quad x \in [0, 1]$, *converge to those of* $V(x)$ *defined in (8.1).*

Proof. Let $0 \leqslant x_1^l < x_1^r \leqslant \ldots \leqslant x_D^l < x_D^r \leqslant 1$ be fixed and denote $I_d = (x_d^l, x_d^r]$, $d = 1, \ldots, D$. Define

$$Y_j = Y_j(\tilde{I}) = (\xi_j^2 - 1)\mathbb{1}_{\{X_j \in \tilde{I}\}} \sigma^2(X_j), \tag{8.5}$$

introduce vectors

$$\mathbf{Y}_j = \left(Y_j(\tilde{I}_1), \ldots, Y_j(\tilde{I}_D)\right),$$

and recall the definition (8.3) of $\boldsymbol{\psi}_j$, $j = 0, \ldots, N-1$.

It is sufficient to show that for any twice differentiable function $g(\cdot)$ on $\mathbb{R}^D$ with bounded all second derivatives

$$\mathbb{E}g(\mathbf{S}_N) - \mathbb{E}g(\boldsymbol{\Psi}_N) \to 0 \quad \text{as } N \to \infty,$$

where $\mathbf{\Psi}_N$ defined in (8.4) and

$$\mathbf{S}_N = \frac{1}{\sqrt{N}} \sum_{0 \leqslant j < N} \mathbf{Y}_j.$$

We shall follow the line of the proof of Lindeberg's theorem from Billingsley (1968). If $\mathbf{x} \in \mathbb{R}^D$ then we denote

$$g_d^{(1)}(\mathbf{x}) = \frac{\partial g(\mathbf{x})}{\partial x_d} \qquad \text{and} \qquad g_{d,b}^{(2)}(\mathbf{x}) = \frac{\partial^2 g(\mathbf{x})}{\partial x_d \partial x_b}, \quad 1 \leqslant d, b \leqslant D.$$

Using a Taylor expansion we have

$$\rho(\mathbf{y}) = \left| g(\mathbf{x}+\mathbf{y}) - g(\mathbf{x}) - \sum_{1 \leqslant d \leqslant D} g_d^{(1)}(\mathbf{x}) y_d - \frac{1}{2} \sum_{1 \leqslant d,b \leqslant D} g_{d,b}^{(2)}(\mathbf{x}) y_d y_b \right|$$

and

$$\rho(\mathbf{y}) \leqslant C(|y_1|^3 + \ldots + |y_D|^3) \tag{8.6}$$

where the constant C only depends on the function $g(\cdot)$. Now we define a sequence of sums

$$\mathbf{R}_{N,j} = \frac{1}{\sqrt{N}} \sum_{0 \leqslant i < j} \mathbf{Y}_i + \frac{1}{\sqrt{N}} \sum_{j < i < N} \boldsymbol{\psi}.$$

Notice

$$\begin{aligned}
&\mathbf{S}_N = \mathbf{R}_{N,N-1} + \frac{1}{\sqrt{N}} \mathbf{Y}_{N-1}, \\
&\mathbf{\Psi}_N = \mathbf{R}_{N,0} + \frac{1}{\sqrt{N}} \boldsymbol{\psi}_0, \\
&\mathbf{R}_{N,j} + \frac{1}{\sqrt{N}} \mathbf{Y}_j = \mathbf{R}_{N,j+1} + \frac{1}{\sqrt{N}} \boldsymbol{\psi}_{j+1}, \quad j = 0, \ldots, N-1
\end{aligned}$$

and the sums over empty set of indexes we assign zero. Thus

$$\mathbb{E}g(\mathbf{S}_N) - \mathbb{E}g(\boldsymbol{\eta}) = \sum_{0 \leqslant j < N} \mathbb{E} \left(g \left(\mathbf{R}_{N,j} + \frac{1}{\sqrt{N}} \mathbf{Y}_j \right) - g \left(\mathbf{R}_{N,j} + \frac{1}{\sqrt{N}} \boldsymbol{\psi}_j \right) \right)$$

or using the definition of $\rho(\cdot)$

$$\Big|\mathbb{E}g(\mathbf{S}_N) - \mathbb{E}g(\boldsymbol{\eta})\Big| \leqslant$$
$$\sum_{0\leqslant j<N}\sum_{1\leqslant d\leqslant D}\left|\mathbb{E}\left(g_d^{(1)}(\mathbf{R}_{N,j})\frac{\sigma^2(X_j)\mathbb{1}_{\{X_j\in\tilde{I}_d\}}}{\sqrt{N}}\left((\xi_j^2-1)-\sqrt{2}\eta_j\right)\right)\right| +$$
$$\sum_{0\leqslant j<N}\sum_{1\leqslant d,b\leqslant D}\left|\mathbb{E}\left(g_d^{(2)}(\mathbf{R}_{N,j})\frac{\sigma^4(X_j)\mathbb{1}_{\{X_j\in\tilde{I}_d\}}}{\sqrt{N}}\left((\xi_j^2-1)^2-2\eta_j^2\right)\right)\right| +$$
$$\sum_{0\leqslant j<N}\mathbb{E}\left(\rho\left(\frac{1}{\sqrt{N}}\mathbf{Y}_j\right)+\rho\left(\frac{1}{\sqrt{N}}\boldsymbol{\psi}_j\right)\right).$$

Because for every j both ξ_j and η_j are independent from $\mathbf{R}_{N,j}$ and X_j and because

$$\mathbb{E}\left((\xi_j^2-1)-\sqrt{2}\eta_j\right) = \mathbb{E}\Big((\xi_j^2-1)^2-2\eta_j^2\Big) = 0,$$

the first two summands on the right-hand side of the last inequality are equal zero. It remains to show that the third term tends to zero as N increases. From (8.6) we can write

$$\sum_{0\leqslant j<N}\mathbb{E}\left(\rho\left(\frac{1}{\sqrt{N}}\mathbf{Y}_j\right)+\rho\left(\frac{1}{\sqrt{N}}\boldsymbol{\psi}_j\right)\right) \leqslant$$
$$C\sum_{0\leqslant j<N}\sum_{1\leqslant d\leqslant D}\mathbb{E}\left(\frac{1}{N^{3/2}}\left(|\xi_j^2-1|^3+|\sqrt{2}\eta_j|^3\right)\sigma^6(X_j)\mathbb{1}_{\{X_j\in\tilde{I}_d\}}\right) =$$
$$\frac{C}{N^{3/2}}\sum_{0\leqslant j<N}\left(\mathbb{E}\left(|\xi_j^2-1|^3+|\sqrt{2}\eta_j|^3\right)\sum_{1\leqslant d\leqslant D}\mathbb{E}\sigma^6(X_j)\mathbb{1}_{\{X_j\in I_d\}}\right) =$$
$$\frac{CC_1}{N^{3/2}}\sum_{0\leqslant j<N}\mathbb{E}\sigma^6(X_j)\mathbb{1}_{X_j\in\cup I_d} \leqslant \frac{NCC_1\max_{x\in[0,1]}\sigma^6(x)}{N^{3/2}} \to 0$$

as $N\to\infty$ and the proof is complete. □

It is worth to point out that though the vectors $\mathbf{Y}_j$ (and consequently $\mathbf{S}_N$) have non-correlated but not independent components the sequence of their sums $\mathbf{S}_N/\sqrt{N}$ converge to a vector whose components are non-correlated too and by normality even independent.

Proof of the weak convergence 8.2. The convergence of finite-dimensional distributions is established by Lemma 8.2.

In our situation such criterion as Theorem 15.6 from Billingsley (1968) is not suited for establish of tightness of the sequence $\xi_N(\cdot)$ and therefore we use another one proved by Genest, Ghoudi, and Remillard (1996), which generalises those of Billingsley. For this we need some additional notation.

Let T be a set of pairs (x, y) such that $0 \leqslant x \leqslant y \leqslant 1$ and $\mathbf{T}$ be a set of functions $\rho(x, y) : T \to [0, \infty)$ which decrease in the first argument and increase in the second one. Function $d \in \mathbf{T}$ is said to be "diagonally vanishing" if and only if for any $\gamma > 0$ one can find $\delta > 0$ such that

$$\sum_{1 \leqslant i \leqslant k} d(x_{i-1}, x_i) < \gamma$$

for all δ-coarse partition $0 = x_0 \leqslant x_1 \leqslant \ldots \leqslant x_k = 1$.

Example 1 from Genest, Ghoudi, and Remillard (1996) is adopted for our goal as following. Let finite-dimensional distributions of a sequence of centred Gaussian process $X_N(\cdot)$ on $[0, 1]$ converge to those of some another centred Gaussian process $X(\cdot)$. Assume that the restriction of $d_N(x, y) := \mathbb{E}\big(X_N(y) - X_N(x)\big)^2$ on T belongs to $\mathbf{T}$ and that the pth power of $d_\infty(x, y) := \mathbb{E}\big(X(y) - X(x)\big)^2$ is diagonally vanishing for some $p > 0$. If there exists $C \in (0, 1)$ such that

$$d_N\left(x, \frac{x+y}{2}\right) \leqslant C d_N(x, y),$$

then $X_N \Rightarrow X$.

For $0 \leqslant x_1 \leqslant x_2 \leqslant 1$ we check that the functions

$$d_\infty(x_1, x_2) = \mathbb{E}\big(V(x_2) - V(x_1)\big)^2 = 2\int_{x_1}^{x_2} \sigma^4(x)\mu(x)dx$$

and

$$d_N(x_1, x_2) = \mathbb{E}\big(\xi_N(x_2) - \xi_N(x_1)\big)^2 = \frac{2}{N} \sum_{0 \leqslant j < N} \mathbb{E}\mathbb{1}_{\{X_j \in (x_1, x_2]\}} \sigma^4(X_j)$$

satisfy the conditions of that example. Notice that because ξ_j is independent from ξ_i, X_i, $i<j$, and X_j's are stationary

$$d_N(x_1,x_2)=\mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{0\leqslant j<N}\mathbb{1}_{\{X_j\in(x_1,x_2]\}}\sigma^2(X_j)(\xi_j^2-1)\right)=$$

$$\frac{2}{N}\sum_{0\leqslant i<j<N}\mathbb{E}\left(\mathbb{1}_{\{X_j\in(x_1,x_2],\,X_i\in(x_1,x_2]\}}\sigma^2(X_j)\sigma^2(X_i)(\xi_i^2-1)\underbrace{\mathbb{E}(\xi_j^2-1)}_{=0}\right)+$$

$$\frac{1}{N}\sum_{0\leqslant j<N}\mathbb{E}\mathbb{1}_{\{X_j\in(x_1,x_2]\}}\sigma^4(X_j)\mathbb{E}(\xi_j^2-1)^2=$$

$$\frac{1}{N}\sum_{0\leqslant j<N}2\int_{x_1}^{x_2}\sigma^4(x)\mu(x)dx=d_\infty(x_1,x_2).$$

It is easy to see that $d_N(\cdot,\cdot)$ decreases in the first argument and increases in the second one. Denote $M_+=\max\big(\sigma^4(x)\mu(x)\big)$ and $M_-=\min\big(\sigma^4(x)\mu(x)\big)$, $x\in[0,1]$. Since the functions σ and μ bounded away from zero on $[0,1]$ then $M_->0$. If for any $\gamma>0$ we put $\delta=\gamma/(4M_+^2)$ then for all δ-coarse partition $0=x_0\leqslant\ldots\leqslant x_L=1$

$$\sum_{1\leqslant i\leqslant L}d_\infty^2(x_{i-1},x_i)=\sum_{1\leqslant i\leqslant L}\left(2\int_{x_{i-1}}^{x_i}\sigma^4(x)\mu(x)dx\right)^2\leqslant$$

$$\sum_{1\leqslant i\leqslant L}4M_+(x_i-x_{i-1})\int_{x_{i-1}}^{x_i}\sigma^4(x)\mu(x)dx\leqslant 4\delta M_+^2\sum_{1\leqslant i\leqslant L}(x_i-x_{i-1})=\gamma.$$

Thus the function d_∞^2 is diagonally vanishing. Further for any $0\leqslant x\leqslant y\leqslant 1$

$$2\int_{(x+y)/2}^{y}\sigma^4(x)\mu(x)dx\geqslant\int_{(x+y)/2}^{y}\sigma^4(x)\mu(x)dx+\int_{(x+y)/2}^{y}M_-dx\geqslant$$

$$\frac{M_-}{M_+}\int_{\frac{x+y}{2}}^{y}\sigma^4(x)\mu(x)dx+M_-\int_{x}^{\frac{x+y}{2}}\frac{\sigma^4(x)\mu(x)}{M_+}dx=\frac{M_-}{M_+}\int_x^y\sigma^4(x)\mu(x)dx,$$

and therefore

$$d_N(x,y)=d_N\left(x,\frac{x+y}{2}\right)+2\int_{(x+y)/2}^{y}\sigma^4(x)\mu(x)dx\geqslant$$

$$d_N\left(x,\frac{x+y}{2}\right)+\frac{M_-}{M_+}d_N(x,y)$$

or

$$d_N\left(x, \frac{x+y}{2}\right) \leqslant \frac{M_+ - M_-}{M_+} d_N(x,y) = C d_N(x,y),$$

where $C = (M_+ - M_-)/M_+ < 1$. □

For the random values $\mathbb{1}_{\{X_j \in (t_1,t_2]\}}\sigma(X_j)\xi_j$ the subgaussian condition

$$\mathbb{E}\exp\left(x\mathbb{1}_{\{X_j\in(t_1,t_2]\}}\sigma(X_j)\xi_j\right) \leqslant \exp\left(\frac{\sigma_+^2 x^2}{2}\right)$$

is fulfilled with $\sigma_+ = \max_{x\in[0,1]} \sigma(x)$ and it justifies the use of the multiresolution threshold inequality (6.13).

As for Lévy modulus of continuity of Brownian motion for the processes $V(x)$ and $W(x)$ with probability 1 holds

$$\lim_{\delta\to 0} \sup_{\substack{0\leqslant x, x\leqslant 1-\delta' \\ 0<\delta'\leqslant\delta}} \max\Big\{|V(x+\delta') - V(x)|, |W(x+\delta') - W(x)|\Big\} \leqslant C_{\mu,\sigma}\sqrt{-\delta\log\delta}$$

for some constant $C_{\mu,\sigma} > 0$ which depends on the functions σ, b and μ. We will use it for the prelimit processes as

$$\sup_{\substack{0\leqslant x, x+\delta'\leqslant 1 \\ 0<\delta'\leqslant\delta_N}} \sqrt{N}|\hat{\varepsilon}_N(x+\delta') - \hat{\varepsilon}_N(x)| \leqslant C_{\mu,\sigma}\sqrt{-\delta_N\log\delta_N}, \qquad \text{and} \tag{8.7}$$

$$\sup_{\substack{0\leqslant x, x+\delta'\leqslant 1 \\ 0<\delta'\leqslant\delta_N}} |\xi_N(x+\delta') - \xi_N(x)| \leqslant C_{\mu,\sigma}\sqrt{-\delta_N\log\delta_N}, \tag{8.8}$$

where δ_N is of the form $(\log N)^\alpha N^{-\beta}$ for some positive α and β and N is large enough.

The next theorem states the similar inequality for the process $f_N(x)$ defined in (6.4) and gives a convergence rate for uniform on $[0,1]$ ergodic theorem for empirical measure of the process $X(t)$.

Theorem 8.3 *Let function $g : \mathbb{R} \to \mathbb{R}$ be bounded on $[0,1]$. Then there exists constant $A > 0$ which depends on functions g and μ such that with probability 1 for all N large enough*

$$\sup_{\substack{0\leqslant x, x+\delta'\leqslant 1 \\ 0<\delta'\leqslant\delta_N}} \frac{\left|\sum_{0\leqslant j<N}\left(\mathbb{1}_{\{X_j\in[x,x+\delta')\}}g(X_j) - \int_x^{x+\delta'} g(z)\mu(dz)\right)\right|}{\sqrt{N}\sqrt{-\delta_N\log\delta_N}} \leqslant A,$$

where $\delta_N = (\log N)^\alpha N^{-\beta}$ with $\alpha \geqslant 0$ and $\beta > 0$.

Proof. Let $P^j(x,B) = \mathbb{P}\{X_j \in B \,|\, X_0 = x\}$, $B \subseteq \mathbb{R}$. For $x_1, x_2 \in [0,1]$, $x_1 \leqslant x_2$, we define

$$\bar{g}(x; x_1, x_2) \equiv \bar{g}(x) = \mathbb{1}_{\{x\in[x_1,x_2)\}} g(x) - \int_{x_1}^{x_2} g(y)\mu(dy), \quad x \in \mathbb{R},$$
$$P^j \bar{g}(x) = \int_R \bar{g}(y) P^j(x,dy), \quad \text{and}$$
$$\sigma_g^2(x_1,x_2) \equiv \sigma_g^2 = \int_{\mathbb{R}} \bar{g}^2(x)\mu(dx) + 2\int_{\mathbb{R}} \bar{g}(x) \sum_{1\leqslant j} P^j \bar{g}(x)\mu(dx).$$

By Theorem 6.15 from Nummelin (1984) for the uniform ergodic chain $\{X_j\}$ there exists $\rho < 1$ and some $C_1 > 0$ such that

$$\sup_{x\in\mathbb{R}} \sup_{A\subseteq\mathbb{R}\,|\,\mu(A)>0} |P^j(x,A) - \mu(A)| \leqslant C_1\rho^j.$$

Because $\int_{\mathbb{R}} \bar{g}(x)\mu(dx) = 0$ and $\bar{g}$ is bounded on $[0,1]$ this implies for σ_g

$$\left|\sum_{1\leqslant j} P^j g(x)\right| = \left|\sum_{1\leqslant j} \int_R \bar{g}(y)\left(P^j(x,dy) - \mu(dy)\right)\right| \leqslant$$
$$\sum_{1\leqslant j} \int_R |\bar{g}(y)| \left|P^j(x,dy) - \mu(dy)\right| \leqslant \sum_{1\leqslant j} \|\bar{g}\| \left|P^j(x,\mathbb{R}) - \mu(\mathbb{R})\right| \leqslant \frac{\|\bar{g}\| C_1 \rho}{1-\rho},$$

where $\|\bar{g}\| = \max_{x\in\mathbb{R}} |\bar{g}(x)| < +\infty$ since $g(x)$ is bounded on $[0,1]$. Thus

$$\sigma_g^2(x_1,x_2) \leqslant \int_{\mathbb{R}} \bar{g}^2(x)\mu(dx) + 2\int_{\mathbb{R}} |\bar{g}(x)| \left|\sum_{1\leqslant j} P^j g(x)\right| \mu(dx) \leqslant$$
$$\int_{x_1}^{x_2} g^2(x)\mu(dx) - \left(\int_{x_1}^{x_2} g(x)\mu(dx)\right)^2 +$$
$$\frac{\|\bar{g}\| C_1\rho}{1-\rho}\left(\int_{x_1}^{x_2} |g(x)|\mu(dx) + \left|\int_{x_1}^{x_2} g(x)\mu(dx)\right|\right) \leqslant C_2(x_2 - x_1) \quad (8.9)$$

for $C_2 = \max_{x\in[0,1]} \mu(x) g^2(x) + \max_{x\in[0,1]} \mu(x) g(x) \dfrac{2\|\bar{g}\| C_1\rho}{1-\rho}$.

The set of intervals $[x_1, x_2)$, $x_1, x_2 \in [0,1]$, is a Vapnik-Červonenkis class and because of boundedness of $g(x)$ on $[0,1]$ the family of functions

$\mathscr{F} = \{g(x)\mathbb{1}_{\{x\in[x_1,x_2)\}} \,|\, 0 \leqslant x_1 \leqslant x_2 \leqslant 1\}$ satisfies Pollard's entropy condition (Dudley (1978), Theorem 2.1, c), d))

$$\int_0^{+\infty} \sqrt{\log\left(\sup_Q D_2(\varepsilon, \mathscr{F}, Q)\right)} d\varepsilon < +\infty,$$

where the supremum is taken over all measures Q with finite support and the covering number $D_2(\varepsilon, \mathscr{F}, Q)$ is defined as

$$D_2(\varepsilon, \mathscr{F}, Q) = \sup\left\{ m \,\middle|\, \text{for some } f_1, \dots, f_m \in \mathscr{F},\ \int |f_i - f_j|^2 dQ > \varepsilon^2 \right\}.$$

The function $|\bar{g}(x; 0, 1)|$ is an envelope function for the family $\mathscr{F}$ and satisfies

$$\int_{\mathbb{R}} \bar{g}^2(x; 0, 1)\mu(dx) < +\infty,$$

and therefore by Theorem 4.3 of Chen (1999) the sums

$$\frac{1}{\sqrt{N}} \sum_{1\leqslant j<N} (|\bar{g}(X_j; 0, x)| - \mathbb{E}|\bar{g}(X_j; 0, x)|)$$

convergence in distribution to a normal distribution as N increases. Then by Theorem 4.9 of Tsai (1998) for the partial sums $S_N(x) = \sum_{0\leqslant j<N} \bar{g}(X_j; 0, x)$ holds

$$\frac{S_N(\cdot)}{\sqrt{N}} \Rightarrow R(\cdot) \quad \text{as } N \to \infty,$$

where $R(x)$, $x \in [0, 1]$, is a centered Gaussian process with

$$\mathbb{E}R(x)R(y) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E} \sum_{0\leqslant i,j<N} \bar{g}(X_i; 0, x)\bar{g}(X_j; 0, y), \quad x, y \in [0, 1].$$

The limit exists because by Cauchy's inequality

$$\frac{1}{N}\mathbb{E}\left| \sum_{0\leqslant i,j<N} \bar{g}(X_i; 0, x)\bar{g}(X_j; 0, y) \right| \leqslant$$
$$\frac{1}{N}\sqrt{\mathbb{E}\left(\sum_{0\leqslant i<N} \bar{g}(X_i; 0, x)\right)^2 \mathbb{E}\left(\sum_{0\leqslant j<N} \bar{g}(X_j; 0, y)\right)^2} = \sqrt{\frac{\mathbb{E}S_N^2(x)}{N}\frac{\mathbb{E}S_N^2(y)}{N}}$$

and by stationarity of $\{X_j\}$

$$\frac{1}{N}\mathbb{E}S_N^2(x) = \frac{1}{N}\sum_{0\leqslant i,j<N}\mathbb{E}\bar{g}(X_i;0,x)\bar{g}(X_j;0,x) =$$
$$\frac{1}{N}\sum_{0\leqslant j<N}\mathbb{E}\bar{g}^2(X_j;0,x) + \frac{2}{N}\sum_{1\leqslant j<N}(N-j)\mathbb{E}\bar{g}(X_0;0,x)\bar{g}(X_j;0,x) =$$
$$\int_{\mathbb{R}}\bar{g}^2(y;0,x)\mu(dy) + 2\sum_{1\leqslant j<N}\left(1-\frac{j}{N}\right)\int_{\mathbb{R}}\bar{g}(y;0,x)P^j\bar{g}(y;0,x)\mu(dy) \to$$
$$\to \sigma_g^2(0,x) \quad \text{as } N\to\infty,$$

where the last convergence follows from Kronecker's lemma. Moreover

$$\mathbb{E}(R(y)-R(x))^2 = \sigma_g^2(\min\{x,y\},\max\{x,y\}), \quad x,y\in[0,1].$$

Thus by (8.9)

$$\mathbb{E}(R(y)-R(x))^2 \leqslant C_2|x-y| \tag{8.10}$$

and from Lemma 2.1 of Markus and Shepp (1971) it follows that $R(x)$ is continuous. Then by Theorems 2.1 and 3.1 of Garsia, Rodemich, and Rumsey (1970) for $R(x)$ there exists a random variable $\eta(\omega)\geqslant 4$ with $\mathbb{E}\eta\leqslant 16\sqrt{2}$ such that for almost all ω we have

$$|R(y)-R(x)| \leqslant 16\int_0^{|y-x|}\sqrt{\log\left(\frac{\eta}{z^2}\right)}dr(z),$$

where

$$r^2(z) = \max_{\substack{|y-x|\leqslant z,\\ x,y\in[0,1]}}\mathbb{E}(R(y)-R(x))^2 \leqslant C_2 z, \quad z\in[0,1].$$

Using (8.10) and variable transformation $z=\eta^{1/2}\exp(-t^2/2)$ we estimate the last integral as

$$\int_0^{\delta}\sqrt{\log\left(\frac{\eta}{z^2}\right)}dr(z) \leqslant \sqrt{C_2}\eta^{1/4}\int_{\sqrt{\log(\eta/\delta^2)}}^{+\infty}t^2\exp(-t^2/4)dt \leqslant$$
$$2\sqrt{C_2}\delta^{1/2}\left(\sqrt{\log\left(\frac{\eta}{\delta^2}\right)}+\frac{2}{\sqrt{\log(\frac{\eta}{\delta^2})}}\right) \leqslant 2\sqrt{C_2\delta\log\left(\frac{\eta}{\delta^2}\right)}\left(1+\frac{2}{\sqrt{\log 4}}\right).$$

Therefore for $C_3 = 4C_2\left(1 + 2/\sqrt{\log 4}\right)^2$ almost surely

$$\sup_{\substack{|y-x|\leqslant\delta \\ x,y\in[0,1]}} |R(y) - R(x)| \leqslant \sqrt{C_3\delta\log\left(\frac{\eta}{\delta^2}\right)}$$

and because of continuity of supremum function for $\varepsilon > 0$, some $C_4 \geqslant 1$ and N large enough we obtain

$$\mathbb{P}\left\{\sup_{\substack{|y-x|\leqslant\delta \\ x,y\in[0,1]}} \frac{|S_N(y) - S_N(x)|}{\sqrt{N}} \geqslant \varepsilon\right\} \leqslant C_4\mathbb{P}\left\{\sup_{\substack{|y-x|\leqslant\delta \\ x,y\in[0,1]}} |R(y) - R(x)| \geqslant \varepsilon\right\} \leqslant$$
$$C_4\mathbb{P}\left\{C_3\delta\log\left(\frac{\eta}{\delta^2}\right) \geqslant \varepsilon^2\right\} \leqslant C_4\frac{\mathbb{E}\exp\left(\log\left(\frac{\eta}{\delta^2}\right)\right)}{\exp\left(\varepsilon^2/(C_3\delta)\right)} \leqslant \frac{16\sqrt{2}C_4}{\delta^2}\exp\left(-\frac{\varepsilon^2}{C_3\delta}\right).$$

Now we set $\varepsilon = \sqrt{\tau C_3\delta\log(1/\delta)}$ with some τ. This yields

$$\mathbb{P}\left\{\sup_{\substack{|y-x|\leqslant\delta \\ x,y\in[0,1]}} \frac{|S_N(y) - S_N(x)|}{\sqrt{N}} \geqslant \varepsilon\right\} \leqslant$$
$$\frac{16\sqrt{2}C_4}{\delta^2}\exp\left(-\tau\log\left(\frac{1}{\delta}\right)\right) = 16\sqrt{2}C_4\delta^{\tau-2}.$$

If $\delta = \delta_N \to 0$ then the choice $\tau \geqslant 2(1 - \log N/\log\delta_N)$ insures

$$\sum_{0\leqslant N}\mathbb{P}\left\{\sup_{\substack{|y-x|\leqslant\delta \\ x,y\in[0,1]}} \frac{|S_N(y) - S_N(x)|}{\sqrt{N}} \geqslant \sqrt{\beta C_3\delta\log(1/\delta)}\right\} \lesssim \sum_{0\leqslant N} N^{-2} < +\infty$$

and by Borel-Cantelli Lemma the assertion of the theorem follows. □

Lemma 8.4 *Let $\{\xi_j\}_{j\geqslant 1}$ be a sequence of independent standard normal random variables and $\{k_N\}$ an integer sequence satisfying*

$$\frac{k_N}{\log N} \to \infty \qquad \text{as} \qquad N \to \infty. \tag{8.11}$$

Then almost surely

$$\limsup_{N\to\infty}\max_{1\leqslant l\leqslant N-k_N+1}\sqrt{\frac{k_N}{\log N}}\left|\frac{1}{k_N}\sum_{l\leqslant j<l+k_N}(\xi_j^2 - 1)\right| \leqslant 2.$$

Proof. Since ξ_j^2 has $\chi^2(1)$ distribution the moment generating function of $\xi_j^2 - 1$ is

$$\mathbb{E}e^{t(\xi_j^2-1)} = \frac{e^{-t}}{\sqrt{1-2t}}, \quad t < \frac{1}{2}.$$

Because this function and the function $\exp(t^2/2)$ are continuous on $(0, 1/2)$ then there exists some $T_0 \in (0, 1/2)$ such that

$$\mathbb{E}e^{t(\xi_j^2-1)} \leqslant e^{2t^2} \qquad \text{for} \quad |t| \leqslant T_0.$$

And by Theorem 15 from Petrov (1975, p.52)

$$\mathbb{P}\left\{\left|\sum_{1\leqslant j\leqslant k_N}(\xi_j^2-1)\right| \geqslant x\right\} \leqslant 2\exp\left(-\frac{x^2}{2k_N}\right) \qquad \text{for } 0 \leqslant x \leqslant T_0 k_N.$$

Because of (8.11) for any $\alpha > 0$ and sufficiently large N

$$(2+\alpha)\sqrt{k_N \log N} \leqslant T_0 k_N.$$

Then for such α and N

$$\mathbb{P}\left\{\left|\sum_{1\leqslant j\leqslant k_N}(\xi_j^2-1)\right| \geqslant (2+\alpha)\sqrt{k_N\log N}\right\} \leqslant$$
$$\exp\left(-\frac{(2+\alpha)^2 k_N \log N}{2k_N}\right) = N^{-\frac{(2+\alpha)^2}{2}} \leqslant N^{-(2+\alpha)}$$

and thus

$$\sum_{N\geqslant 1}\mathbb{P}\left\{\max_{1\leqslant l\leqslant N-k_N+1}\sqrt{\frac{k_N}{\log N}}\left|\frac{1}{k_N}\sum_{l\leqslant j<l+k_N}(\xi_j^2-1)\right| \geqslant 2+\alpha\right\} \leqslant$$
$$\sum_{N\geqslant 1}(N-k_N+1)\mathbb{P}\left\{\left|\sum_{1\leqslant j\leqslant k_N}(\xi_j^2-1)\right| \geqslant (2+\alpha)\sqrt{k_N\log N}\right\}$$
$$\leqslant \sum_{N\leqslant 1}\frac{N-k_N+1}{N^{2+\alpha}} < \sum_{N\leqslant 1} N^{-(1+\alpha)} < \infty.$$

By Borel-Cantelli Lemma and because α is arbitrary the assertion of the lemma follows. $\square$

8.2 Some inequalities for χ^2-distribution

Below we will use Stirling's formula for the gamma function ([1, 6.1.37])

$$\frac{1}{\Gamma(k)} = \frac{e^k}{k^k}\sqrt{\frac{k}{2\pi}}\left(1 + O\left(\frac{1}{k}\right)\right). \tag{8.12}$$

Also we need to state some properties of the incomplete gamma functions.

Lemma 8.5 *For the incomplete gamma functions $\gamma(k,x)$ and $\Gamma(k,x)$ defined as*

$$\gamma(k+1,x) = \int_0^x t^k e^{-t} dt \quad \text{and} \quad \Gamma(k+1,x) = \int_x^\infty t^k e^{-t} dt, \quad x \geqslant 0$$

and for all integer k the following is true:

$$\frac{x^{k+1}e^{-x}}{2(x-k)} \leqslant \Gamma(k+1,x) \leqslant \frac{x^{k+1}e^{-x}}{x-k} \quad \text{for } x \geqslant k+\sqrt{k} \qquad \text{and} \tag{8.13}$$

$$\frac{x^{k+1}e^{-x}}{2(k-x)} \leqslant \gamma(k+1,x) \leqslant \frac{x^{k+1}e^{-x}}{k-x} \quad \text{for } 0 \leqslant x \leqslant k-\sqrt{k}. \tag{8.14}$$

Proof. Using the variable transformation $t = x(1+s)$ we obtain

$$\gamma(k+1,x) = x^{k+1}e^{-x}\int_{-1}^0 (1+s)^k e^{-xs} ds \quad \text{and}$$
$$\Gamma(k+1,x) = x^{k+1}e^{-x}\int_0^\infty (1+s)^k e^{-xs} ds.$$

The right-hand side inequalities follow from the obvious inequality $1+x \leqslant e^x$ and

$$\Gamma(k+1,x) \leqslant x^{k+1}e^{-x}\int_0^\infty e^{(k-x)s} ds = \frac{x^{k+1}e^{-x}}{x-k} \quad (\textit{here} \quad x > k) \quad \text{and}$$
$$\gamma(k+1,x) \leqslant x^{k+1}e^{-x}\int_{-1}^0 e^{(k-x)s} ds = \frac{x^{k+1}e^{-x}\left(1-e^{x-k}\right)}{k-x} \leqslant \frac{x^{k+1}e^{-x}}{k-x}$$
$$(\textit{here} \quad x < k).$$

For the left-hand side inequalities we denote $h(s) = (1+s)^k e^{-xs}$ and consider

$$h'(s) = h(s)\left(\frac{k}{1+s} - x\right) = \frac{h(s)}{1+s}(k - x(1+s)) \qquad \text{and}$$
$$h''(s) = h(s)\left[\left(\frac{k}{1+s} - x\right)^2 - \frac{k}{(1+s)^2}\right] = \frac{h(s)}{(1+s)^2}\left((k - x(1+s))^2 - k\right).$$

For the case $x \leqslant k - \sqrt{k}$ and $s \in [-1, 0]$ we have

$$(k - x(1+s))^2 - k \geqslant (k-x)^2 - k \geqslant (k - (k - \sqrt{k}))^2 - k = 0,$$

and for the case $x \geqslant k + \sqrt{k}$ and $s \geqslant 0$

$$(k - x(1+s))^2 - k \geqslant (k-x)^2 - k \geqslant (k - (k + \sqrt{k}))^2 - k = 0.$$

Thus for both cases the function $h(s)$ is convex and therefore $h(s) \geqslant \max\{0, h(0) + h'(0)s\} = \max\{0, 1 + (k-x)s\}$. It follows

$$\Gamma(k+1, x) \geqslant x^{k+1}e^{-x}\int_0^{\frac{1}{x-k}} 1 + (k-x)s\,ds = \frac{x^{k+1}e^{-x}}{2(x-k)} \text{ (here } x > k) \quad \text{and}$$
$$\gamma(k+1, x) \geqslant x^{k+1}e^{-x}\int_{\frac{1}{x-k}}^{0} 1 + (k-x)s\,ds = \frac{x^{k+1}e^{-x}}{2(k-x)} \quad \text{(here } x < k). \quad \square$$

Remark. For x and k such that $(x-k)/\sqrt{k}$ tends to infinity as k increases Tricomi (1950, p.140) showed that

$$\Gamma(k+1, x) = \frac{x^{k+1}e^{-x}}{x-k}\left(1 + O\left(\frac{k}{(x-k)^2}\right)\right). \tag{8.15}$$

The probability density function of $\chi^2(k)$ random value is

$$f_k(x) = \frac{\exp(-x/2)\,x^{k/2-1}}{2^{k/2}\Gamma(k/2)}. \tag{8.16}$$

With a use of the incomplete gamma functions $\gamma(k, x)$ and $\Gamma(k, x)$ we can write

$$\mathbb{P}\left\{\chi^2(k) \geqslant x\right\} = \frac{\Gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \qquad \text{and} \qquad \mathbb{P}\left\{\chi^2(k) \leqslant x\right\} = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \quad \text{for } x \geqslant 0. \tag{8.17}$$

If $x \geqslant k+\sqrt{2k}$ then using (8.12) and (8.13) for the first probability we obtain

$$\frac{e^{\frac{k-x}{2}}\left(\frac{x}{k}\right)^{\frac{k}{2}}\sqrt{k}\big(1+O(k^{-1})\big)}{2(x-k+2)\sqrt{\pi}} \leqslant \mathbb{P}\left\{\chi^2(k) \geqslant x\right\} \leqslant \frac{e^{\frac{k-x}{2}}\left(\frac{x}{k}\right)^{\frac{k}{2}}\sqrt{k}\big(1+O(k^{-1})\big)}{(x-k+2)\sqrt{\pi}}. \tag{8.18}$$

And similarly for $0 \leqslant x \leqslant k-2\sqrt{k}$ using (8.12) and (8.14) for the second probability

$$\frac{e^{\frac{k-x}{2}}\left(\frac{x}{k}\right)^{\frac{k}{2}}\sqrt{k}\big(1+O(k^{-1})\big)}{2(k-x-2)\sqrt{\pi}} \leqslant \mathbb{P}\left\{\chi^2(k) \leqslant x\right\} \leqslant \frac{e^{\frac{k-x}{2}}\left(\frac{x}{k}\right)^{\frac{k}{2}}\sqrt{k}\big(1+O(k^{-1})\big)}{(k-x-2)\sqrt{\pi}}. \tag{8.19}$$

Lemma 8.6 *If a sequence of integers $\{k_N\}_{N\geqslant 1}$ is bounded by some K then*

$$\lim_{N\to\infty}\frac{\log\varphi(N)}{qu\left(\frac{1+\alpha_N}{2},k_N\right)} = \frac{1}{2}, \tag{8.20}$$

and for some positive C_0, C_1 and all sufficiently large N

$$C_0\left(\varphi(N)\right)^{-2/K_-} \leqslant qu\left(\frac{1-\alpha_N}{2},k_N\right) \leqslant C_1\left(\varphi(N)\right)^{-2/K}, \tag{8.21}$$

where $K_- = \liminf k_N$.

Proof. Taking logarithm from terms of (8.18) with $k = k_N \leqslant K$ and $x = x(N) = qu\left(\frac{1+\alpha_N}{2},k_N\right)$ we obtain in consideration of $qu\left(\frac{1+\alpha_N}{2},k_N\right) \to \infty$

$$-\frac{x}{2}+\left(\frac{k}{2}-1\right)\log x + C_0 \leqslant -\log\big(2\varphi(N)\big) \leqslant -\frac{x}{2}+\left(\frac{k}{2}-1\right)\log x + C_1,$$

where constants C_0 and C_1 depend on K. Consequently

$$\frac{\log\varphi(N)}{qu\left(\frac{1+\alpha_N}{2},k_N\right)} \to \frac{1}{2} \qquad \text{as } N\to\infty.$$

Taking into account that $x = qu\left(\frac{1-\alpha_N}{2},k_N\right) \to 0$ from (8.19) with $k = k_N$ we can write

$$x^{k/2}\cdot C_0 \leqslant \frac{1}{2\varphi(N)} \leqslant x^{k/2}\cdot C_1,$$

and constants C_0 and C_1 depend on K. As corollary

$$\left(\frac{1}{2C_1\varphi(N)}\right)^{2/K_-} \leqslant qu\left(\frac{1-\alpha_N}{2},k_N\right) \leqslant \left(\frac{1}{2C_0\varphi(N)}\right)^{2/K}$$

for all N large enough. □

Lemma 8.7 *Let a sequence of integers* $\{k_N\}_{N\geqslant 1}$, $1 \leqslant k_N \leqslant N$ *tends to infinity and let function* $g(N)$ *be defined from by the equation*

$$qu\left(\frac{1+\alpha_N}{2}, k_N\right) = k_N(1+g(N)). \tag{8.22}$$

There exist some positive constants A_0, A_1 *such that for all sufficiently large* N

$$A_0\frac{\log\varphi(N)}{k_N} \leqslant g(N) \leqslant A_1 \max\left\{\sqrt{\frac{\log\varphi(N)}{k_N}}, \frac{\log\varphi(N)}{k_N}\right\}. \tag{8.23}$$

Moreover, if

$$\frac{\log\varphi(N)}{k_N} \to 0 \tag{8.24}$$

then

$$g(N) \sim 2\sqrt{\frac{\log\varphi(N)}{k_N}} \qquad as \quad N\to\infty. \tag{8.25}$$

Proof. Suppose that $g(N) \leqslant C/\sqrt{k_N}$ for some constant C. From the left-hand side inequality in (8.18) using Taylor expansion for logarithm we have

$$\mathbb{P}\Big\{\chi^2(k_N) \geqslant k_N(1+g(N))\Big\} \geqslant \mathbb{P}\Big\{\chi^2(k_N) \geqslant k_N + C\sqrt{k_N}\Big\} \geqslant$$

$$\exp\left(-\frac{C\sqrt{k_N}}{2}\right)\left(1+\frac{C}{\sqrt{k_N}}\right)^{\frac{k_N}{2}} \frac{\sqrt{k_N}\left(1+O\left(\frac{1}{k_N}\right)\right)}{2\sqrt{\pi}(C\sqrt{k_N}+2)} =$$

$$\exp\left(-\frac{C\sqrt{k_N}}{2} + \frac{k_N}{2}\log\left(1+\frac{C}{\sqrt{k_N}}\right)\right)\frac{\left(1+O\left(k_N^{-1/2}\right)\right)}{2\sqrt{\pi}C} =$$

$$\exp\left(-\frac{C^2}{4}+\frac{C^3}{6k_N^{1/2}} - \ldots\right)\frac{1}{2\sqrt{\pi}C}\left(1+O\left(k_N^{-1/2}\right)\right) \to \frac{e^{-\frac{C^2}{4}}}{2\sqrt{\pi}C} > 0,$$

but the firt probability by definition of $g(N)$ tends to zero as $N\to\infty$. It means that with necessity $g(N)\sqrt{k_N}\to\infty$. Now by the similar arguments but using asymptotic relation (8.15) instead of inequalities (8.13) we obtain

$$\frac{1}{2\varphi(N)} = \mathbb{P}\{\chi^2(k_N) \geqslant k_N(1+g(N))\} =$$

$$\exp\left\{-\frac{k_N}{2}\Big(g(N)-\log(1+g(N))\Big)\right\}\frac{C_1}{g(N)\sqrt{k_N}},$$

where $C_1 \approx \pi^{-1/2}$. Or taking logarithm from the both sides

$$\frac{2\log\varphi(N)}{k_N} = \Big(g(N) - \log(1+g(N))\Big) + \frac{2}{k_N}\log\left(\frac{g(N)\sqrt{k_N}}{2C_1}\right). \qquad (8.26)$$

Because $g(N)\sqrt{k_N} \to \infty$ we have

$$\frac{\log\left(\frac{g(N)\sqrt{k_N}}{2C_1}\right)}{k_N\big(g(N) - \log(1+g(N))\big)} \to 0 \quad \text{as} \quad N \to \infty$$

and therefore the right-hand side of (8.26) is asymptotically equivalent to $g(N) - \log(1+g(N))$ and does not exceed $2g(N_l)/A_0$ for some $A_0 > 0$. Thus we obtain the left inequality of (8.23). And the right inequality follows since for $x \geqslant 0$

$$x + \log(1+x) \geqslant \min\left\{x^2/6, x(1-\log(2))\right\} \geqslant \frac{\min\{x^2, x\}}{6}.$$

Suppose the condition (8.24) holds. It implies that k_N increases to infinity and it follows $g(N) \to 0$ because otherwise the right-hand side of (8.26) does not tend to zero. Therefore from (8.26) we obtain

$$\log\varphi(N) = \frac{g^2(N)k_N}{4}\left\{1 + o\big(g(N)\big) + \frac{2\log\big(g(N)\sqrt{k_N}/(2C_1)\big)}{g^2(N)k_N}\right\} \sim \frac{g^2(N)k_N}{4}$$

and (8.25) is stated. □

Now we state the similar lemma for the left tail.

Lemma 8.8 *Let a sequence of integers $\{k_N\}_{N\geqslant 1}$, $1 \leqslant k_N \leqslant N$ tends to infinity and let function $g(N)$ be defined from by the equation*

$$qu\left(\frac{1-\alpha_N}{2}, k_N\right) = k_N(1 - g(N)). \qquad (8.27)$$

There exist some positive constants A_0, A_1 such that for all sufficiently large N

$$1 - \big(\varphi(N)\big)^{-A_0/k_N} \leqslant g(N) \leqslant A_1\sqrt{\frac{\log\varphi(N)}{k_N}}. \qquad (8.28)$$

Moreover, if

$$\frac{\log\varphi(N)}{k_N} \to 0 \qquad (8.29)$$

then

$$g(N) \sim 2\sqrt{\frac{\log \varphi(N)}{k_N}} \quad \textit{as } N \to \infty. \tag{8.30}$$

Proof. The proof mostly repeats the proof of Lemma 8.7 but with small changes. Let $g(N)\sqrt{k_N} \leqslant C$. Then from the left-hand side inequality (8.19) and using Taylor expansion for logarithm we obtain

$$\mathbb{P}\Big\{\chi^2(k_N) \leqslant k_N(1-g(N))\Big\} \geqslant \mathbb{P}\Big\{\chi^2(k_N) \leqslant k_N - C\sqrt{k_N}\Big\} \geqslant$$

$$\exp\left(\frac{C\sqrt{k_N}}{2}\right)\left(1-\frac{C}{\sqrt{k_N}}\right)^{\frac{k_N}{2}} \frac{\sqrt{k_N}\left(1+O\left(\frac{1}{k_N}\right)\right)}{2\sqrt{\pi}(C\sqrt{k_N}-2)} =$$

$$\exp\left(\frac{C\sqrt{k_N}}{2}+\frac{k_N}{2}\log\left(1-\frac{C}{\sqrt{k_N}}\right)\right)\frac{\left(1+O\left(k_N^{-1/2}\right)\right)}{2\sqrt{\pi}C} =$$

$$\exp\left(-\frac{C^2}{4}-\frac{C^3}{6k_N^{1/2}}-\dots\right)\frac{1}{2\sqrt{\pi}C}\left(1+O\left(k_N^{-1/2}\right)\right) \to \frac{e^{-\frac{C^2}{4}}}{2\sqrt{\pi}C} > 0$$

but the first probability vanishing as N increases. Therefore the product $g(N)\sqrt{k_N} \to \infty$. Now by the similar arguments we have

$$\frac{1}{2\varphi(N)} = \mathbb{P}\{\chi^2(k) \leqslant k_N(1-g(N))\} =$$

$$\exp\left\{\frac{k_N}{2}\Big(g(N)+\log(1-g(N))\Big)\right\}\frac{C_1}{g(N)\sqrt{k_N}},$$

where C_1 lies between $\pi^{-1/2}/2$ and $\pi^{-1/2}$. And taking logarithm from the both sides

$$\frac{2\log\varphi(N)}{k_N} = -\Big(g(N)+\log(1-g(N))\Big)+\frac{2}{k_N}\log\left(\frac{g(N)\sqrt{k_N}}{2C_1}\right). \tag{8.31}$$

By definition $g(N) \leqslant 1$ while k_N and $g(N)\sqrt{k_N}$ increase to infinity. It implies that

$$\frac{\log\left(\frac{g(N)\sqrt{k_N}}{2C_1}\right)}{k_N\big(g(N)+\log(1-g(N))\big)} \to 0 \quad \text{as} \quad N \to \infty$$

and therefore the right-hand side of (8.31) is asymptotically equivalent to $-\big(g(N)+\log(1-g(N))\big)$ and does not exceed $-2\log(1-g(N))/A_0$ for some $A_0>0$. That is

$$\frac{A_0\log\varphi(N)}{k_N}\leqslant -\log(1-g(N))$$

or

$$\big(\varphi(N)\big)^{-A_0/k_N}=\exp\left(-\frac{A_0\log\varphi(N)}{k_N}\right)\geqslant 1-g(N)$$

and we obtain the left inequality in (8.28), while the right one follows from the fact that for $x\in[0,1)$

$$-\big(x+\log(1-x)\big)\geqslant\frac{x^2}{2}.$$

And if the condition (8.29) holds then $g(N)$ tends to zero too and therefore

$$\log\varphi(N)\approx\frac{g^2(N)k_N}{4}\left\{1+o\big(g(N)\big)+\frac{2\log\big(g(N)\sqrt{k_N}/(2C_1)\big)}{g^2(N)k_N}\right\}\sim\frac{g^2(N)k_N}{4}$$

□.

Remark. As follows from Lemmas 8.7 and 8.8 for any constant $C>0$

$$qu\left(\frac{1+\alpha_N}{2},k_N\right)\geqslant k_N+C\sqrt{k_N}\quad\text{and}\quad qu\left(\frac{1-\alpha_N}{2},k_N\right)\leqslant k_N-C\sqrt{k_N}\tag{8.32}$$

if $k_N\to\infty$ as $N\to\infty$. These inequalities are true for bounded k_N and $k_N-C\sqrt{k_N}>0$ by definition of quantiles.

Lemma 8.9 *Let $\delta(N)$ be a positive function and $\{k_N\}_{N\geqslant 1}$ is an integer sequence.*

If there exists such constant C that

$$\delta(N)\left(qu\left(\frac{1+\alpha_N}{2},k_N\right)-k_N+2\right)\leqslant C\tag{8.33}$$

then

$$\mathbb{P}\left\{\chi^2(k_N)\geqslant\big(1-\delta(N)\big)qu\left(\frac{1+\alpha_N}{2},k_N\right)\right\}\leqslant\frac{4e^{C/2}-7/2}{\varphi(N)}.\tag{8.34}$$

Proof. Denote $x = x(N) = qu\left(\frac{1+\alpha_N}{2}, k_N\right)$ and consider the integral

$$\int_{x(1-\delta(N))}^{x} f_{k_N}(y)dy,$$

where the density function f_k given by (8.16). Making the variable transformation $y = x(t+1)$ we have

$$\int_{x(1-\delta(N))}^{x} f_{k_N}(y)dy = xf_{k_N}(x)\int_{-\delta(N)}^{0} \exp\left(-\frac{xt}{2}\right)(1+t)^{\frac{k_N}{2}-1}\,dt. \tag{8.35}$$

Using the condition (8.33) we estimate the latter integral as

$$\int_{-\delta(N)}^{0} \exp\left(-\frac{xt}{2}+\frac{t(k_N-2)}{2}\right)dt = \frac{2x\left(1-\exp\left(-\delta(N)\frac{k_N-x-2}{2}\right)\right)}{k_N-x-2} =$$
$$\frac{2x\left(\exp\left(\delta(N)\frac{x-k_N+2}{2}\right)-1\right)}{x-k_N+2} \leqslant \frac{2x\left(\exp(C/2)-1\right)}{x-k_N+2}. \tag{8.36}$$

Because of the first relation in (8.32) we can use (8.13) from Lemma 8.5 that gives

$$\frac{x}{x-k_N+2} f_{k_N}(x) \leqslant 4\mathbb{P}\Big\{\chi^2(k_N) \geqslant x\Big\}. \tag{8.37}$$

If we notice that $\mathbb{P}\Big\{\chi^2(k_N) \geqslant x\Big\} = 1/(2\varphi(N))$ then combining (8.35), (8.36), and (8.37) we obtain

$$\mathbb{P}\left\{\chi^2(k_N) \geqslant \big(1-\delta(N)\big)qu\left(\frac{1+\alpha_N}{2}, k_N\right)\right\} =$$
$$\mathbb{P}\Big\{\chi^2(k_N) \geqslant x\Big\} + \int_{x(1-\delta(N))}^{x} f_{k_N}(y)dy \leqslant$$
$$\mathbb{P}\Big\{\chi^2(k_N) \geqslant x\Big\}\big(1+8\left(\exp(C/2)-1\right)\big) = \frac{1+8\left(\exp(C/2)-1\right)}{2\varphi(N)}. \qquad \square$$

8.3 Proofs for the volatility term

Below we require the relation between the lengths of intervals $\tilde{I}(x)$ and $I(x)$ defined in (5.5). We remark that by construction

$$|I(x,N)| \geqslant 1 \qquad \text{and} \qquad |\tilde{I}(x,N)| \geqslant \frac{1}{2N}$$

for all $x \in [0,1]$ and N and general relation is given by Theorem 8.3 and following

Lemma 8.10 *If the function σ is continuous on an interval $[x_1, x_2)$ then for any $\varepsilon > 0$ there exists some constant $A > 0$ such that*

$$\lim_{N\to\infty} \mathbb{P}\left\{ \inf_{x\in[x_1+\varepsilon, x_2-\varepsilon]} |I(x,N)| \geqslant A \min\left\{ \frac{N^{2/3}}{(\log\varphi(N))^{1/3}}, \varphi^{1/2}(N) \right\} \right\} = 1.$$

Proof. For every N we have the finite set $\mathcal{I}_1(N)$ of the intervals $I(\cdot, N)$ such that corresponding continuous intervals $\tilde{I}(\cdot, N)$ cover the interval $[x_1 + \varepsilon, x_2 - \varepsilon]$. If the assertion of the lemma does not hold then there exists some increasing sequence $\{N_k\}$, and a sequence of the intervals $I_k \in \mathcal{I}_1(N_k)$ such that

$$|I_k| \leqslant L_k - 1, \tag{8.38}$$

where

$$L_k = o\left(\min\left\{ \frac{N_k^{2/3}}{(\log\varphi(N_k))^{1/3}}, \varphi^{1/2}(N_k) \right\} \right). \tag{8.39}$$

Since $1/N \leqslant |\tilde{I}_k| \leqslant 1$ and by Theorem 8.3 with the function $g \equiv 1$ for some $A > 0$ we have

$$N_k|\tilde{I}_k| \min_{x\in[0,1]} \mu(x) \leqslant N_k \int_{\tilde{I}} \mu(dx) \leqslant$$
$$A\left(|I_k| + \sqrt{N_k|\tilde{I}_k| \log(1/|\tilde{I}_k|)} \right) \leqslant A\left(|I_k| + \sqrt{N_k \log N_k} \right)$$

and therefore for $C_0 = A/\min_{x\in[0,1]} \mu(x) > 0$ and some $C_1 > 0$

$$|\tilde{I}_k| \leqslant C_0 \left(\frac{1}{(N_k \log\varphi(N_k))^{1/3}} + \sqrt{\frac{\log N_k}{N_k}} \right) \leqslant \frac{C_1}{(N_k \log\varphi(N_k))^{1/3}}. \tag{8.40}$$

It implies that the length of the interval $\tilde{I}_k$ tends to zero and eventually its both ends lie in $[x_1, x_2]$. Let $j_k \equiv j(N_k)$ be the next point that follows the right-hand side end of the interval I_k and we consider the extended interval $I_k \cup \{j_k\}$. Notice that the corresponding values $X_{(j_k)}$ belongs to $[x_1, x_2)$ for k large enough. Because by construction

$$X_{(j_k)} - \sup\{x \in \tilde{I}_k\} \leqslant |\tilde{I}_k|$$

then

$$X_{(j_k)} - \inf\{x \in \tilde{I}_k\} \leqslant 2|\tilde{I}_k|$$

and therefore the length of corresponding extended continuous interval also tends to zero. Denoting

$$\sigma_-^2 = \min_{x \in \tilde{I}_k \cup X_{(j_k)}} \sigma^2(x) \quad \text{and} \quad \sigma_+^2 = \max_{x \in \tilde{I}_k \cup X_{(j_k)}} \sigma^2(x)$$

we can write

$$\frac{\sigma_-^2}{\sigma_+^2} = 1 - \frac{\sigma_+^2 - \sigma_-^2}{\sigma_+^2} \geqslant 1 - \frac{K_\sigma(\sigma_+ + \sigma_-)}{\sigma_+^2} 2|\tilde{I}_k| \geqslant 1 - C_2|\tilde{I}_k|,$$

where $C_2 = 4K_\sigma \max_{x\in[x_1,x_2)} \sigma(x)/\min_{x\in[x_1,x_2)} \sigma^2(x)$ and K_σ is the Lipshitz constant of σ. If $|I_k|$ is bounded then by Lemma 8.6

$$\frac{qu\left(\frac{1+\alpha_{N_k}}{2}, |I_k|\right)}{\log \varphi(N_k)} = O(1)$$

otherwise from Lemma 8.7 it follows

$$\frac{qu\left(\frac{1+\alpha_{N_k}}{2}, |I_k|\right) - |I_k|}{\log \varphi(N_k)} \lesssim \frac{\max\left\{\sqrt{|I_k| \log \varphi(N_k)}, \log \varphi(N_k)\right\}}{\log \varphi(N_k)} = O(1)$$

since $\log \varphi(N_k) \sim \log N_k$. Thus by (8.40)

$$|\tilde{I}_k| \left(qu\left(\frac{1+\alpha_{N_k}}{2}, |I_k|\right) - |I_k| + 2 \right) \lesssim \frac{1}{(N_k \log \varphi(N_k))^{1/3}} \to 0 \tag{8.41}$$

and we can use Lemma 8.9 with $\delta(N_k) = C_2|\tilde{I}_k|$.

Let $I_- = I_{-,k}$ and $I_+ = I_{+,k}$ be that subintervals of I_k where the lower σ_l^2 and upper σ_u^2 bounds are attained, that is

$$\frac{\sum_{i\in I_+} R_i^2}{qu\left(\frac{1-\alpha_{N_k}}{2}, |I_+|\right)} = \sigma_u^2\big(I_k \cup \{j_k\}\big) \quad \text{and} \quad \sigma_l^2\big(I_k \cup \{j_k\}\big) = \frac{\sum_{i\in I_-} R_i^2}{qu\left(\frac{1+\alpha_{N_k}}{2}, |I_-|\right)}.$$

Using Lemma 8.9 with some constant $C_3 > 0$ we have

$$\mathbb{P}\left\{\frac{\sum_{i\in I_-} R_i^2}{qu\left(\frac{1+\alpha_{N_k}}{2}, |I_-|\right)} \geqslant \sigma_-^2\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} \mathbb{P}\left\{\frac{\sum_{i\in I_-} \sigma_i^2 Z_i^2}{qu\left(\frac{1+\alpha_{N_k}}{2}, |I_-|\right)} \geqslant \sigma_-^2,\, |I_-| = l\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} \mathbb{P}\left\{\sum_{i\in I_-} Z_i^2 \geqslant \frac{\sigma_-^2}{\sigma_+^2} qu\left(\frac{1+\alpha_{N_k}}{2}, l\right),\, |I_-| = l\right\} =$$

$$\sum_{1\leqslant l\leqslant L_k} \mathbb{P}\left\{\max_{\substack{m,n\in I_k\cup\{j_k\}\\ n-m+1=l}} \sum_{m\leqslant i\leqslant n} Z_i^2 \geqslant \frac{\sigma_-^2}{\sigma_+^2} qu\left(\frac{1+\alpha_{N_k}}{2}, l\right),\, |I_-| = l\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} \mathbb{P}\left\{\bigcup_{\substack{m,n\in I_k\cup\{j_k\}\\ n-m+1=l}} \left(\sum_{m\leqslant i\leqslant n} Z_i^2 \geqslant \frac{\sigma_-^2}{\sigma_+^2} qu\left(\frac{1+\alpha_{N_k}}{2}, l\right)\right)\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} (L_k - l)\,\mathbb{P}\left\{\chi^2(l) \geqslant \frac{\sigma_-^2}{\sigma_+^2} qu\left(\frac{1+\alpha_{N_k}}{2}, l\right)\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} (L_k - l)\,\mathbb{P}\left\{\chi^2(l) \geqslant \left(1 - C_2|\tilde{I}_k|\right) qu\left(\frac{1+\alpha_{N_k}}{2}, l\right)\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} (L_k - l)\frac{C_3}{\varphi(N_k)} = \frac{C_3 L_k(L_k-1)}{2\varphi(N_k)}. \quad (8.42)$$

Similarly for the upper bound σ_u^2

$$\mathbb{P}\left\{\frac{\sum_{i\in I_+} R_i^2}{qu\left(\frac{1-\alpha_{N_k}}{2}, |I_+|\right)} \leqslant \sigma_-^2\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} \mathbb{P}\left\{\frac{\sum_{i\in I_+} Z_i^2}{qu\left(\frac{1-\alpha_{N_k}}{2}, |I_+|\right)} \geqslant 1,\, |I_+| = l\right\} \leqslant \sum_{1\leqslant l\leqslant L_k} (L_k - l)\,\mathbb{P}\left\{\chi^2(l) \leqslant qu\left(\frac{1-\alpha_{N_k}}{2}, l\right)\right\} = \frac{L_k(L_k-1)}{4\varphi(N_k)} \quad (8.43)$$

By construction on $I_k \cup \{j_k\}$ holds $\sigma_u^2\big(I_k \cup \{j_k\}\big) \leqslant \sigma_l^2\big(I_k \cup \{j_k\}\big)$, that is

$$1 = \mathbb{P}\Big\{\sigma_u^2\big(I_k \cup \{j_k\}\big) \leqslant \sigma_l^2\big(I_k \cup \{j_k\}\big)\Big\}. \tag{8.44}$$

But from (8.42) and (8.43) it follows

$$\mathbb{P}\Big\{\sigma_u^2\big(I_k \cup \{j_k\}\big) \leqslant \sigma_l^2\big(I_k \cup \{j_k\}\big)\Big\} \leqslant \mathbb{P}\Big\{\sigma_u^2\big(I_k \cup \{j_k\}\big) \leqslant \sigma_-^2\Big\} + \\ \mathbb{P}\Big\{\sigma_-^2 \leqslant \sigma_l^2\big(I_k \cup \{j_k\}\big)\Big\} \leqslant \frac{(C_3 + 1/2)L_k(L_k - 1)}{2\varphi(N_k)}. \tag{8.45}$$

or combining with (8.44)

$$1 \leqslant \frac{(C_3 + 1/2)}{2} \cdot \frac{L_k^2}{\varphi(N_k)}$$

what contradicts (8.39). □

Lemma 8.11 *If the function σ is constant on an interval $[x_1, x_2)$ then for any $\varepsilon > 0$ there exists some constant $A > 0$ such that*

$$\lim_{N\to\infty} \mathbb{P}\left\{\inf_{x\in[x_1+\varepsilon, x_2-\varepsilon]} |I(x,N)| \geqslant A \min\left\{N, \varphi^{1/2}(N)\right\}\right\} = 1.$$

Proof. Because on $[x_1, x_2)$ the volatility σ is constant we do not need to use Lemma 8.9 which is ensured by (8.40) and (8.41). Therefore the factor

$$\frac{N^{2/3}}{(\log\varphi(N))^{1/3}}$$

is replaced by N. In other respects we repeat the proof of Lemma 8.10. □

Lemma 8.12 *If the volatility function σ is strictly monotone on some neighbourhood of point $x \in [0,1]$ and differentiable at x then for some constant $C > 0$ for almost every $\omega \in \Omega$ and all sufficiently large $N = N(\omega)$ holds*

$$|\tilde{I}(x,N)| \leqslant C \left(\frac{\log\varphi(N)}{N}\right)^{\frac{1}{2r+1}},$$

where $r \geqslant 1$ is such the smallest integer that $\sigma^{(r)}(x) \neq 0$.

Proof. It is only enough to consider the case $\sigma(x)$ is antitone. We can suppose that the volatility σ is monotone on the whole interval $\tilde{I}(x,N)$ because otherwise we consider instead of the interval $\tilde{I}$ its intersection with the corresponding neighbourhood on x.

If the statement of the lemma does not hold for $\omega \in \Omega$ then there exists some sequence $N_k = N_k(\omega)$ such that

$$\frac{|\tilde{I}(x,N_k)|}{\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{1}{2r+1}}} \to \infty \quad \text{as } N_k \to \infty, \tag{8.46}$$

and particularly

$$|\tilde{I}(x,N_k)| \geqslant 3\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{1}{2r+1}}.$$

For every N_k we define four points

$$x_0 - \text{ the left-hand end of the interval } \tilde{I}(x,N_k) \text{ and}$$
$$x_i = x_0 + i\cdot|\tilde{I}(x,N_k)|/3, \quad i = 1,2,3.$$

and choose intervals $\tilde{I}_1 \subseteq [x_0,x_1)$ and $\tilde{I}_3 \subseteq [x_2,x_3)$ such that

$$|\tilde{I}_1| = |\tilde{I}_3| = \left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{1}{2r+1}} \to 0. \tag{8.47}$$

Then for corresponding discrete intervals I_i, $i = 1,3$ by Corollary 8.3 for some $C_0 > 0$

$$\frac{\log\varphi(N_k)}{|I_i|} \leqslant \frac{\log\varphi(N_k)}{C_0^{-1}N_k|\tilde{I}_i| - \sqrt{N_k\log N_k}} =$$

$$\frac{\log\varphi(N_k)}{C_0^{-1}N_k\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{1}{2r+1}} - \sqrt{N_k\log N_k}} \leqslant \frac{\log\varphi(N_k)}{N_k\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{1}{2r+1}} C_0^{-1}/2} =$$

$$2C_0\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{2r}{2r+1}} \to 0 \qquad \text{as } N_k \to \infty.$$

Because by Corollary 8.3 also $|I_i| \leqslant C_0N_k|\tilde{I}_i| + \sqrt{N_k\log N_k}$ if fact

$$\frac{\log\varphi(N_k)}{|I_i|} \sim \left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{2r}{2r+1}}. \tag{8.48}$$

On the interval $\tilde{I}$ the inequality

$$\sigma_l^2(I) \leqslant \sigma_u^2(I)$$

is true by construction and by definition of the bounds it follows

$$\frac{\sum_{i\in I_1} R_i^2}{qu\left(\frac{1+\alpha_{N_k}}{2}, |I_1|\right)} \leqslant \sigma_l^2(I) \leqslant \sigma_u^2(I) \leqslant \frac{\sum_{i\in I_3} R_i^2}{qu\left(\frac{1-\alpha_{N_k}}{2}, |I_3|\right)}. \tag{8.49}$$

Using (8.8) for some $C_{\mu,\sigma} > 0$ and sufficiently large N_k

$$\sum_{i\in I_1} R_i^2 \geqslant \sum_{i\in I_1} \sigma_i^2 - C_{\mu,\sigma}\sqrt{-N_k|\tilde{I}_1|\log|\tilde{I}_1|} = \big(|I_1| - \beta_1\big)\sigma^2(x_1)$$

and

$$\sum_{i\in I_3} R_i^2 \leqslant \sum_{i\in I_3} \sigma_i^2 + C_{\mu,\sigma}\sqrt{-N_k|\tilde{I}_3|\log|\tilde{I}_3|} = \big(|I_3| + \beta_3\big)\sigma^2(x_2),$$

where

$$\beta_1 = \beta_1(N_k) = \frac{C_{\mu,\sigma}}{\sigma^2(x_1)}\sqrt{-N_k|\tilde{I}_1|\log|\tilde{I}_1|}$$

and

$$\beta_3 = \beta_3(N_k) = \frac{C_{\mu,\sigma}}{\sigma^2(x_2)}\sqrt{-N_k|\tilde{I}_3|\log|\tilde{I}_3|}$$

Notice that by (8.47), (8.48) and using $\log\varphi(N_k) \sim \log N_k$ we have

$$\frac{\beta_i}{\sqrt{|I_i|\log\varphi(N_k)}} \sim$$
$$\sqrt{\frac{\log\varphi(N_k)}{|I_i|}}\frac{1}{\log\varphi(N_k)} \cdot \sqrt{N_k\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{1}{2r+1}}\log\left(\frac{N_k}{\log\varphi(N_k)}\right)^{\frac{1}{2r+1}}} \sim$$
$$\left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{r}{2r+1}}\frac{N_k^{\frac{r}{2r+1}}(\log\varphi(N_k))^{\frac{r+1}{2r+1}}}{\log\varphi(N_k)} = 1.$$

From (8.49) by Lemmas 8.8 and 8.7 it follows

$$\frac{(|I_1| - \beta_1)\sigma^2(x_1)}{|I_1| + 2\sqrt{|I_1|\log\varphi(N_k)}} \leqslant \frac{(|I_3| + \beta_3)\sigma^2(x_2)}{|I_3| - 2\sqrt{|I_3|\log\varphi(N_k)}}$$

or

$$\frac{|I_1| - \beta_1}{|I_1| + 2\sqrt{|I_1| \log \varphi(N_k)}} \cdot \frac{|I_3| - 2\sqrt{|I_3| \log \varphi(N_k)}}{|I_3| + \beta_3} \leqslant \frac{\sigma^2(x_2)}{\sigma^2(x_1)} \tag{8.50}$$

for all N_k large enough. For the fractions on the left-hand side we can write

$$\frac{|I_1| - \beta_1}{|I_1| + 2\sqrt{|I_1| \log \varphi(N_k)}} = 1 - \frac{2\sqrt{|I_1| \log \varphi(N_k)} + \beta_1}{|I_1| + 2\sqrt{|I_1| \log \varphi(N_k)}} \gtrsim 1 - C_1' \sqrt{\frac{\log \varphi(N_k)}{|I_1|}}$$

and

$$\frac{|I_3| - 2\sqrt{|I_3| \log \varphi(N_k)}}{|I_3| + \beta_3} = 1 - \frac{\beta_3 + 2\sqrt{|I_3| \log \varphi(N_k)}}{|I_3| + \beta_3} \gtrsim 1 - C_1'' \sqrt{\frac{\log \varphi(N_k)}{|I_1|}}$$

and its product asymtotically also not less then $1 - C_1 \sqrt{\frac{\log \varphi(N_k)}{|I_1|}}$ for some $C_1 > 0$. Thus from (8.50)

$$1 - C_1 \sqrt{\frac{\log \varphi(N_k)}{|I_1|}} \leqslant \frac{\sigma^2(x_2)}{\sigma^2(x_1)} \tag{8.51}$$

This inequality is impossible if $x_2 - x_1 \nrightarrow 0$ because by decreasing of σ the right-hand fraction remains strictly less than 1 while the left-hand expression tends to 1 as N_k increases. Thus $x_2 - x_1$ and $|\tilde{I}(x, N)|$ tends to zero. A Taylor expansion gives

$$\frac{\sigma^2(x_2)}{\sigma^2(x_1)} = \frac{\sigma^2(x) + \frac{2\sigma(x)\sigma^{(r)}(x)}{r!}(x_2 - x)^r + o(|x_2 - x|^r)}{\sigma^2(x) + \frac{2\sigma(x)\sigma^{(r)}(x)}{r!}(x_1 - x)^r + o(|x_1 - x|^r)}$$

Because σ antitone on $\tilde{I}(x, N_k)$ the expression $\frac{2\sigma^{(r)}(x)}{\sigma(x)r!}((x_1 - x)^r - (x_2 - x)^r)$ is positive, the fraction $\frac{2\sigma^{(r)}(x)}{\sigma(x)r!}$ and the difference $((x_1 - x)^r - (x_2 - x)^r)$ are negative, and r is odd. It implies

$$|(x_1 - x)^r - (x_2 - x)^r| \geqslant \frac{\max\{|x_1 - x|^r, |x_2 - x|^r\}}{2} \geqslant \frac{1}{2}\left|\frac{x_2 - x_1}{2}\right|^r = \frac{|\tilde{I}(x, N_k)|^r}{2 \cdot 6^r}$$

and consequently

$$\frac{\sigma^2(x) + \frac{2\sigma(x)\sigma^{(r)}(x)}{r!}(x_2 - x)^r + o(|x_2 - x|^r)}{\sigma^2(x) + \frac{2\sigma(x)\sigma^{(r)}(x)}{r!}(x_1 - x)^r + o(|x_1 - x|^r)} =$$

$$1 - \frac{\frac{2\sigma^{(r)}(x)}{\sigma(x)r!}((x_1 - x)^r - (x_2 - x)^r) + o(|x_1 - x|^r) + o(|x_2 - x|^r)}{1 + \frac{2\sigma^{(r)}(x)}{\sigma(x)r!}(x_1 - x)^r + o(|x_1 - x|^r)} \leqslant$$

$$1 - \frac{\frac{2\sigma^{(r)}(x)}{\sigma(x)r!}((x_1 - x)^r - (x_2 - x)^r)}{2} \leqslant 1 - \frac{\sigma^{(r)}(x)}{\sigma(x)r!}\frac{|\tilde{I}(x, N_k)|^r}{2 \cdot 6^r}.$$

Therefore for some $C_2 > 0$

$$\frac{\sigma^2(x_2)}{\sigma^2(x_1)} \leqslant 1 - C_2|\tilde{I}(x, N_k)|^r \tag{8.52}$$

and by (8.51) and (8.48)

$$C_2|\tilde{I}(x, N_k)|^r \leqslant C_1\sqrt{\frac{\log\varphi(N_k)}{|I_1|}} \sim \left(\frac{\log\varphi(N_k)}{N_k}\right)^{\frac{r}{2r+1}}$$

and (8.46) can not be true. □

Lemma 8.13 *If the volatility σ is differentiable on an interval $[x_1, x_2)$ and*

$$\inf_{x\in[x_1,x_2)} |\sigma'(x)| > 0 \tag{8.53}$$

then for any $\varepsilon > 0$ and some constant $C > 0$ for almost every $\omega \in \Omega$

$$\sup_{x\in[x_1+\varepsilon,x_2-\varepsilon)} |\tilde{I}(x, N)| \leqslant C\left(\frac{\log\varphi(N)}{N}\right)^{1/3}$$

for all sufficiently large $N = N(\omega)$.

Proof. If we suppose that the statement of the lemma does not hold then as in Lemma 8.10 we can find an increasing sequence N_k and a sequence of the intervals $\tilde{I}_k$ from the covering family $\mathcal{I}_1(N_k)$ such that

$$\frac{|\tilde{I}_k|}{\left(\log\varphi(N_k)/N_k\right)^{1/3}} \to \infty \quad \text{as } N_k \to \infty.$$

If $|\tilde{I}_k|$ does not tend to zero then there exist $x \in [x_1, x_2)$ such that $x \in \tilde{I}_k$, that is $\tilde{I}(x, N_k) = \tilde{I}_k$ infinitely often. But for x Lemma 8.12 holds with $r = 1$ and therefore

$$\lim_{N_k \to \infty} |\tilde{I}(x, N_k)| = 0$$

and from some N_k the both ends of $|\tilde{I}(x, N_k)|$ lie in $[x_1, x_2)$. Futher on we exactly repeat the reasoning from the proof of Lemma 8.12 for sequence $\tilde{I}_k$ with any $y_k \in \tilde{I}_k$. The proof of the Lemma 8.12 depends on y_k only in the part of obtaining the relation (8.52). In this lemma such inequality with $r = 1$ is ensured uniformly in $y \in [x_1, x_2)$ by the condition (8.53). $\square$

Remark. The assertion of the lemma remains valid if the volatility σ is not differentiable on $[x_1, x_2)$ but only continuous and strictly monotone and the following holds:

$$\inf_{x,y \in [x_1, x_2)} \frac{|\sigma(y) - \sigma(x)|}{|x - y|} > 0.$$

Lemma 8.14 *Let x be a point of discontinuity of σ and let $\tilde{I}_l(x, N) = \{y \in \tilde{I}(x, N) \,|\, y < x\}$ and $\tilde{I}_r(x, N) = \{y \in \tilde{I}(x, N) \,|\, y \geqslant x\}$. Then for some constant A with probability one*

$$\lim_{N \to \infty} \frac{\min\left\{|I_l(x, N)|, |I_r(x, N)|\right\}}{\log \varphi(N)} \leqslant A.$$

Proof. Suppose that $\sigma(x-) > \sigma(x)$. Let firstly $\tilde{I}_l$ and $\tilde{I}_r$ be the following: $\tilde{I}_l(x, N) = [x_l, x)$ and $\tilde{I}_r(x, N) = [x, x_r)$, where

$$x_l = \inf\{y \in \tilde{I}(x, N) \,|\, y < x \text{ and } \sigma^2(y) \geqslant (\sigma^2(x) + 2\sigma^2(x-))/3\} \qquad \text{and}$$
$$x_r = \sup\{y \in \tilde{I}(x, N) \,|\, y \geqslant x \text{ and } \sigma^2(y) \leqslant (2\sigma^2(x) + \sigma^2(x-))/3\}.$$

Now we only consider that N for which $\tilde{I}_l(x, N) \neq \emptyset$, that is interval $\tilde{I}(x, N)$ begins not at x but on the left-hand side of it. Otherwise the statement of the lemma is obviously true. If the assertion of the lemma is false then for some sequence N_k

$$\frac{\min\left\{|I_l(x, N)|, |I_r(x, N)|\right\}}{\log \varphi(N_k)} \to \infty$$

and we can apply asymptotic quantilies values (8.25) and (8.30) from Lemmas 8.7 and 8.8 respectively. Using the similar arguments as in Lemma 8.12

we obtain (8.49) with $I_1 = I_l$ and $I_3 = I_r$ and consequently (8.50) while $x_1 = x_l$, $x_2 = x_r$ and

$$\frac{\beta_i}{\sqrt{|I_i|\log\varphi(N_k)}} = O(1)$$

as $N_k \to \infty$. For the fractions on the left-hand side we can write

$$\frac{|I_1| - \beta_1}{|I_1| + 2\sqrt{|I_1|\log\varphi(N_k)}} = 1 - \frac{2\sqrt{|I_1|\log\varphi(N_k)} + \beta_1}{|I_1| + 2\sqrt{|I_1|\log\varphi(N_k)}} \sim$$
$$1 - \frac{C_1'\sqrt{\log(\varphi(N_k))/|I_1|}}{1 + 2\sqrt{\log(\varphi(N_k))/|I_1|}} = 1 - o(1)$$

and similarly

$$\frac{|I_3| - 2\sqrt{|I_3|\log\varphi(N_k)}}{|I_3| + \beta_3} = 1 - \frac{\beta_3 + 2\sqrt{|I_3|\log\varphi(N_k)}}{|I_3| + \beta_3}$$
$$1 - \frac{C_2''\sqrt{\log(\varphi(N_k))/|I_3|}}{1 + \beta_3/|I_3|} = 1 - o(1).$$

Therefore instead of (8.51) we obtain

$$1 - o(1) \leqslant \frac{\sigma^2(x_2)}{\sigma^2(x_1)} = \frac{2\sigma^2(x) + \sigma^2(x-)}{\sigma^2(x) + 2\sigma^2(x-)} = 1 - \frac{\sigma^2(x-) - \sigma^2(x)}{\sigma^2(x) + 2\sigma^2(x-)},$$

which is asymptotically impossible because the value on the right-hand side remains strictly less than 1 while the left-hand expression tends to 1. Also it implies that length of the minor of the intervals $\tilde{I}_l$ and $\tilde{I}_r$ tend to zero and that interval eventually contains all y from $\tilde{I}(x, N)$ that smaller or larger than x respectively. □

Proof of Theorem 5.1. Case 1. Because of the condition (5.7) the volatility σ is monotone and we only consider the case of σ is increasing on $[x_1, x_2)$. For every N we have the finite family $\tilde{\mathcal{I}}_1(N)$ of the intervals $\tilde{I}(\cdot, N)$ that cover the interval $[x_1 + \varepsilon, x_2 - \varepsilon)$. Let $\tilde{I} \in \tilde{\mathcal{I}}_1(N)$ be such interval that

$$\sup_{x\in[x_1+\varepsilon, x_2-\varepsilon)} \left|\sigma^2(x) - \hat{\sigma}^2(x)\right| = \sup_{x\in\tilde{I}} \left|\sigma^2(x) - \hat{\sigma}^2(\tilde{I})\right|.$$

We denote this interval $\tilde{I}_N$ and notice that probabiliy of the event

$$\Omega_{N,1} = \{\text{the both ends of } \tilde{I}_N \text{ lie in } [x_1, x_2)\}$$

tends to one. It follows from Lemma 8.13 because the lengths, say, of $\tilde{I}(x_1 + \varepsilon/2, N)$ and $\tilde{I}(x_2 - \varepsilon/2, N)$ tend to zero. Also it is clear, that the supremum $\sup_{x\in\tilde{I}}\left|\sigma^2(x) - \hat{\sigma}^2(\tilde{I})\right|$ is attained at either the left or the right end of $\tilde{I}_N$. Futher we denote that end as x_+. Then we have

$$\sup_{x\in[x_1+\varepsilon,x_2-\varepsilon)}\left|\sigma^2(x)-\hat{\sigma}^2(x)\right| = \left|\sigma^2(x_+)-\hat{\sigma}^2(\tilde{I}_N)\right| = \left|\sigma^2(x_+) - \frac{\sum_{i\in I_N} R_i^2}{|I_N|}\right| \leqslant \frac{\left|\sum_{i\in I_N}\sigma_i^2(Z_i^2-1)\right|}{|I_N|} + \frac{\left|\sum_{i\in I_N}(\sigma^2(x_+)-\sigma_i^2)\right|}{|I_N|}. \quad (8.54)$$

On $\Omega_{N,1}$ for the second fraction the remark after Lemma 8.13 yields

$$\frac{\left|\sum_{i\in I_N}(\sigma^2(x_+)-\sigma_i^2)\right|}{|I_N|} \leqslant \frac{K_\sigma|\tilde{I}_N||I_N|}{|I_N|} = K_\sigma|\tilde{I}_N| \leqslant K_\sigma C_1\left(\frac{\log\varphi(N)}{N}\right)^{1/3} \quad (8.55)$$

for the Lipshitz constant K_σ and some $C_1 > 0$. Also by Lemma 8.10 for some positive C_2 probability of the event

$$\Omega_{N,2} = \left\{|I_N| \geqslant C_2\min\left\{\frac{N^{2/3}}{(\log\varphi(N))^{1/3}}, \varphi^{1/2}(N)\right\}\right\}$$

tends to 1. On $\Omega_{N,2}$ holds $\log N/|I_N| \to 0$ and applying Lemma 8.4 we can find constant $C_3 > 0$ such that

$$\frac{\left|\sum_{i\in I_N}\sigma_i^2(Z_i^2-1)\right|}{|I_N|} \leqslant C_3\max_{x\in[x_1,x_2)}\sigma^2(x)\sqrt{\frac{\log\varphi(N)}{|I_N|}}$$

for sufficiently large N. Here we use that $\log\varphi(N) \sim \log N$. And because on $\Omega_{N,2}$

$$\sqrt{\frac{\log\varphi(N)}{|I_N|}} \leqslant \sqrt{C_2}\max\left\{\frac{(\log\varphi(N))^{2/3}}{N^{1/3}}, \frac{(\log\varphi(N))^{1/2}}{\varphi^{1/4}(N)}\right\}$$

Combining the latter with (8.55) and (8.54) we obtain (5.8).

Case 2. In the same way as in the Case 1 for every N we define the interval $\tilde{I}_N$ but now it is not necessarily that the ends of $\tilde{I}_N$ belong to $[x_1, x_2)$. From Lemma 8.11 it follows that for some constant $C_1 > 0$ probability of the event

$$\Omega_{N,1} = \left\{|I_N|^{-1} \leqslant C_1^2\max\left\{N^{-1}, \varphi^{-1/2}(N)\right\}\right\} \quad (8.56)$$

tends to one as N increases. We only consider the worth situation when the points x_1 and x_2 are points of discontinuity of the function σ and the interval $\tilde{I}_N$ covers for every N the whole interval $[x_1, x_2)$. Split the interval $\tilde{I}_N$ into three parts: $\tilde{I}_{N,l} = \{x \in \tilde{I}_N \mid x < x_1\}$, $\tilde{I}_{N,r} = \{x \in \tilde{I}_N \mid x \geqslant x_2\}$, and $\tilde{I}_{N,c} = \tilde{I}_N \Big\backslash (\tilde{I}_{N,l} \cup \tilde{I}_{N,r}) \equiv [x_1, x_2)$. Then because both σ and $\hat{\sigma}$ are constant on $[x_1, x_2)$ we have

$$\sup_{x\in[x_1+\varepsilon, x_2-\varepsilon)} \Big|\sigma^2(x) - \hat{\sigma}^2(x)\Big| = \Big|\sigma^2(x_1) - \hat{\sigma}^2(x_1)\Big| = \left|\sigma^2(x_1) - \frac{\sum_{i\in I_N} R_i^2}{|I_N|}\right| \leqslant$$
$$\frac{\left|\sum_{i\in I_{N,l}\cup I_{N,r}}(R_i^2 - \sigma_i^2 + \sigma_i^2 - \sigma^2(x_1)) + \sum_{i\in I_{N,c}}(R_i^2 - \sigma^2(x_1))\right|}{|I_N|} \leqslant$$
$$\frac{2\sigma_+^2 k_N + \left|\sum_{i\in I_N}\sigma_i^2(Z_i^2-1)\right|}{|I_N|} \leqslant \sigma_+^2 \frac{2k_N + \left|\sum_{i\in I_N}(Z_i^2-1)\right|}{|I_N|}, \quad (8.57)$$

where we put $k_N = |I_{N,l} \cup I_{N,r}|$ and $\sigma_+^2 = \max_{x\in[0,1]} \sigma^2(x)$. Because $|\tilde{I}_{N,c}| \equiv x_2 - x_1 > 0$ by Lemma 8.14 there exists constants $C_2 > 0$ such that for the event

$$\Omega_{N,2} = \Big\{k_N = |I_{N,l}| + |I_{N,r}| \leqslant C_2 \log\varphi(N)\Big\}$$

holds $\mathbb{P}\big\{\Omega_{N,2}\big\} \to 1$. On $\Omega_{N,1}$ we have $\log N/|I_N| \to 0$ and therefore from Lemma 8.4 for some $C_3 > 0$ and from (8.56) it follows

$$\frac{\left|\sum_{i\in I_N}(Z_i^2-1)\right|}{|I_N|} \leqslant C_3\sqrt{\frac{\log\varphi(N)}{|I_N|}} \leqslant C_3 C_1 \frac{\sqrt{\log\varphi(N)}}{\min\{N^{1/2}, \varphi^{-1/4}(N)\}}$$

for sufficiently large N. And because on $\Omega_{N,1}$

$$\frac{\log\varphi(N)}{|I_N|} = o\left(\sqrt{\frac{\log\varphi(N)}{|I_N|}}\right)$$

the latter together with (8.57) implies

$$\mathbb{P}\left\{\sup_{x\in[x_1+\varepsilon, x_2-\varepsilon)} \Big|\sigma^2(x) - \hat{\sigma}^2(x)\Big| \leqslant \frac{2\sigma_+^2 C_3 C_1\sqrt{\log\varphi(N)}}{\min\{N^{1/2}, \varphi^{-1/4}(N)\}},\ \Omega_{N,1},\ \Omega_{N,2}\right\} \to 1$$

as $N \to \infty$. □

8.4 Proofs for the drift term

Proof of Theorem 6.1 repeats that of Theorem 3.2 from Davies and Kovac (2001) □.

For the proof of Theorems 6.2 and 6.4 we notice that

$$y_N(x+\delta) - y_N(x) = \big(f_N(x+\delta) - f_N(x)\big) + \big(s_N(x+\delta) - s_N(x)\big). \qquad (8.58)$$

Proof of 1 of Theorem 6.2. Let $s_N(\cdot)$ be convex until it reaches p_1^l. Then for $t_1 = p_1^r - r_1^l$ and arbitrary small $\delta > 0$ as a property of the taut string it follows

$$t_1 = \operatorname*{argmax}_{0 \leqslant t \leqslant \delta} \frac{\Big(y_N(p_1^l + t) - C_{\mathcal{T}}/\sqrt{N}\Big) - \Big(y_N(p_1^l) + C_{\mathcal{T}}/\sqrt{N}\Big)}{t}$$

as N tends to infinity. We put $p = p_1^e - p_1^l$ and can write

$$t_1 = \operatorname*{argmax}_{0 \leqslant t \leqslant \delta} \frac{y_N(p_1^e - p + t) - y_N(p_1^e - p) - 2C_{\mathcal{T}}/\sqrt{N}}{t}$$

By (8.58)

$$\begin{aligned} y_N(p_1^e - p + t) - y_N(p_1^e - p) = {} & \frac{\sqrt{N}(s_N(p_1^e - p + t) - s_N(p_1^e - p))}{\sqrt{N}} + \\ & \big(f_N(p_1^e - p + t) - f_N(p_1^e)\big) - \big(f_N(p_1^e - p) - f_N(p_1^e)\big) \end{aligned}$$

New we apply for the first summand (8.7) and a Taylor expansion of order two to the second and third summands. It gives

$$\begin{aligned} & \frac{y_N(p_1^e - p + t) - y_N(p_1^e - p)}{t} = \\ & \qquad \frac{o(1)}{\sqrt{N}} + f(p_1^e) + f^{(2)}(p_1^e)\frac{t^2 - 3tp + 3p^2}{6}\big(1 + o(1)\big) \end{aligned}$$

where all $o(1) \to 0$ as $\max\{p, t\} \to 0$. In other words

$$t_1 = \operatorname*{argmax}_{0 \leqslant t \leqslant \delta} \left(f^{(2)}(p_1^e)\frac{t^2 - 3tp}{6}\big(1 + o(1)\big) - \frac{2C_{\mathcal{T}}}{t\sqrt{N}}\big(1 + o(1)\big) \right)$$

that implies

$$\frac{(2t_1 - 3p)}{6} f^{(2)}(p_1^e)\big(1 + o(1)\big) = -\frac{2C_{\mathcal{T}}}{t_1^2\sqrt{N}}\big(1 + o(1)\big). \tag{8.59}$$

and because of $f^{(2)}(p_1^e) < 0$ we get $3p \leqslant 2t_1\big(1 + o(1)\big)$ and consequently $p_1^e = p_1^l + p < p_1^l + t_1 = p_1^r$. Using similar reasoning for the other direction we have

$$t_1 = \operatorname*{argmax}_{0\leqslant t\leqslant\delta} \frac{\Big(y_N(p_1^r) - C_{\mathcal{T}}/\sqrt{N}\Big) - \Big(y_N(p_1^r - t) + C_{\mathcal{T}}/\sqrt{N}\Big)}{t}$$

and denoting $p_1^r = p_1^e + p*$ we obtain

$$\frac{(2t_1 - 3p*)}{6} f^{(2)}(p_1^e)\big(1 + o(1)\big) = -\frac{2C_{\mathcal{T}}}{t_1^2\sqrt{N}}\big(1 + o(1)\big). \tag{8.60}$$

that implies $p_1^e > p_1^l$ and hence $p_1^e \in [p_1^l, p_1^r]$.

Proof of 2. Adding (8.59) and (8.60) with remark $t_1 = p + p*$ yields

$$\frac{t_1}{6} f^{(2)}(p_1^e)\big(1 + o(1)\big) = -\frac{2C_{\mathcal{T}}}{t_1^2\sqrt{N}}\big(1 + o(1)\big)$$

or

$$t_1^3 f^{(2)}(p_1^e)\big(1 + o(1)\big) = -\frac{24C_{\mathcal{T}}}{\sqrt{N}}\big(1 + o(1)\big)$$

that implies

$$t_1 \sim (24C_{\mathcal{T}})^{1/3}|f^{(2)}(p_1^e)|^{-1/3}N^{-1/6}.$$

Proof of 3. We only consider the first two knots x_1, x_2 and suppose $\varepsilon_N(\cdot)$ and $f_N(\cdot)$ are convex on interval (x_1, x_2). Another cases can be checked in a similar way. Let $t_0 = x_2 - x_1$, then

$$t_0 = \operatorname*{argmin}_{0\leqslant t\leqslant\delta} \frac{y_N(x_1 + t) - y_N(x_1)}{t} \tag{8.61}$$

From (8.58) using a Taylor expansion we can write

$$\frac{y_N(x_1 + t) - y_N(x_1)}{t} = f(x_1) + f^{(1)}(x_1)\frac{t}{2} + \frac{\varepsilon_N(x_1 + t) - \varepsilon_N(x_1)}{t} + O\left(t^2\right).$$

Uniformly in $x \in [0,1]$ and small t, say $0 \leqslant t \leqslant N^{-1/10}$, and for some positive constant A the modulus of continuity (8.7)gives

$$-A\sqrt{-t\log t} \leqslant \sqrt{N}\big(s_N(x+t) - s_N(x)\big) \leqslant A\sqrt{-t\log t}.$$

On taking

$$t = a|f^{(1)}(x_1)|^{-2/3}\left(\frac{\log N}{N}\right)^{1/3} \tag{8.62}$$

we obtain

$$\frac{y_N(x_1+t) - y_N(x_1)}{t} \geqslant f(x_1) + \frac{a}{2}|f^{(1)}(x_1)|^{1/3}\left(\frac{\log N}{N}\right)^{1/3}$$
$$- \frac{A}{\sqrt{a}}|f^{(1)}(x_1)|^{1/3}\left(\frac{\log N}{N}\right)^{1/3} + O\left(|f^{(1)}(x_1)|^{4/3}\left(\frac{\log N}{N}\right)^{2/3}\right).$$

From the statement 2 of the theorem it follows $\left|f^{(1)}(x_1)\right| \geqslant AN^{-1/6}$ for all $C_{\mathcal{T}}$ larger than some C_0. Therefore for the last term we have

$$O\left(|f^{(1)}(x_1)|^{4/3}\left(\frac{\log N}{N}\right)^{2/3}\right) \leqslant O\left(\frac{(\log N)^{2/3}}{N^{4/9}}\right) = o\left(\left(\frac{\log N}{N}\right)^{1/3}\right)$$

and it may be ignored. Thus for sufficiently large a

$$\frac{y_N(x_1+t) - y_N(x_1)}{t} \geqslant f(x_1) + \frac{a}{4}|f^{(1)}(x_1)|^{1/3}\left(\frac{\log N}{N}\right)^{1/3}. \tag{8.63}$$

Analogously the upper bound is given as follows

$$\frac{y_N(x_1+t) - y_N(x_1)}{t} \leqslant f(x_1) + a|f^{(1)}(x_1)|^{1/3}\left(\frac{\log N}{N}\right)^{1/3} \tag{8.64}$$

with the same sufficiently large a as in (8.63). It implies that the local minimum in (8.61) is attained at the point $x_1 + t$ where t is set in (8.62), i.e.

$$t = O\left(|f^{(1)}(x_1)|^{-2/3}\left(\frac{\log N}{N}\right)^{1/3}\right).$$

This asserts 3 of the theorem.

Proof of 4. Let at first x be a knot x_i of $s_N(\cdot)$ such that the interval (x_i, x_{i+1}) does not contain a local extreme value of $s_N(\cdot)$ and let s_N be convex at x_i. By the definition of the taut string we have

$$s_N(x_i) \leqslant \frac{y_N(x_i + t) - y_N(x_i)}{t}$$

for any $t \leqslant x_{i+1} - x_i$. In the similar way as in the proof of 3 of the theorem by a Taylor expansion and (8.58) for

$$t = a|f^{(1)}(x_1)|^{-2/3} \left(\frac{\log N}{N}\right)^{1/3}$$

we can write

$$s_N(x_i) \leqslant f(x_i) + A|f^{(1)}(x_i)|^{1/3} \left(\frac{\log N}{N}\right)^{1/3}. \tag{8.65}$$

Analogously using

$$s_N(x_i) \leqslant \frac{y_N(x_i + t) - y_N(x_i)}{t}$$

we get

$$s_N(x_i) \geqslant f(x_i) - A|f^{(1)}(x_i)|^{1/3} \left(\frac{\log N}{N}\right)^{1/3}. \tag{8.66}$$

that together with (8.65) yields

$$|f(x_i) - s_N(x_i)| = O\left(|f^{(1)}(x_i)|^{1/3} \left(\frac{\log N}{N}\right)^{1/3}\right).$$

Now for a point x from interval $[x_i, x_{i+1}]$ from 3 of the theorem we have

$$|f(x) - f(x_i)| \leqslant |f^{(1)}(x_i)| \cdot A|f^{(1)}(x_i)|^{-2/3} \left(\frac{\log N}{N}\right)^{1/3} = A\left(|f^{(1)}(x_i)|\frac{\log N}{N}\right)^{1/3}$$

and further because $s_N(\cdot)$ is constant on $[x_i, x_{i+1}]$

$$|f(x) - s_N(x)| = |f(x) - s_N(x_i)| \leqslant |f(x) - f(x_i)| + |f(x_i) - s_N(x_i)| \leqslant A_1\left(|f^{(1)}(x_i)|\frac{\log N}{N}\right)^{1/3} \square.$$

Proof of 5. Consider one of K intervals $[p_i^l, p_i^r]$ and suppose that $s_N(p_i^l)$ lies on the upper border ub_N while $s_N(p_i^r)$ on the lower lb_N and $s_N(p_i^l) \leqslant s_N(p_i^r)$. Using the fact $f^{(1)}(p_i^e) = 0$ for interval $[p_i^l, p_i^r] \ni p_i^e$ we can apply similar arguments as in the proof of 4 of the theorem and a Taylor expansion of order three, which gives

$$s_N(p_i^e) \leqslant f(p_i^e) + \frac{t^2}{6} f^{(2)}(p_i^e) + A\sqrt{-\frac{\log t}{Nt}} + O(t^3)$$

and

$$s_N(p_i^e) \geqslant f(p_i^e) + \frac{t^2}{6} f^{(2)}(p_i^e) - A\sqrt{-\frac{\log t}{Nt}} - \frac{2C_{\mathcal{T}}}{Nt} - O(t^3)$$

for any $t > 0$ such that $p_i^e + t \leqslant p_i^r$. On setting

$$t = A_1 C_{\mathcal{T}}^{1/3} |f^{(2)}(p_i^e)|^{-1/3} N^{-1/6}$$

we obtain

$$|f(p_i^e) - s_N(p_i^e)| = \frac{A_1}{6} C_{\mathcal{T}}^{2/3} |f^{(2)}(p_i^e)|^{1/3} N^{-1/3} \big(1 + o(1)\big) \tag{8.67}$$

where $o(1) \to 0$ as $N \to \infty$. Because for any $x \in [p_i^l, p_i^r]$ from 2 of the theorem we have

$$|f(x) - f(p_i^e)| = |f^{(2)}(p_i^e)| \big(1 + o(1)\big) \left(\frac{A_3 C_{\mathcal{T}}^{1/3}}{|f^{(2)}(p_i^e)|^{1/3} N^{1/6}} \right)^2 \leqslant \\ A_4 |f^{(2)}(p_i^e)|^{1/3} N^{-1/3} C_{\mathcal{T}}^{2/3} \tag{8.68}$$

then for any $t \in [p_i^l, p_i^r]$

$$|f(x) - s_N(x)| = |f(p_i^e) - s_N(p_i^e)| \leqslant \\ |f(x) - f(p_i^e)| + |f(p_i^e) - s_N(p_i^e)| \leqslant A_5 |f^{(2)}(p_i^e)|^{1/3} N^{-1/3} C_{\mathcal{T}}^{2/3} \qquad \square.$$

Proof of 1 of Theorem 6.4. If there is no observations in the interval $[t_{1,N}, t_{2,N})$ then multiresolution coefficient $w(t_{1,N}, t_{2,N})$ equals to zero and (6.13) holds. In the case when the number of X_j's in $[t_{1,N}, t_{2,N})$ is bounded for all N the statement of the theorem follows from (6.14). Let now the number of observations in the interval $\sharp(t_{1,N}, t_{2,N})$ tends to infinity as N increases. Then using the definition of the taut string S_N and (6.3) we can write

$$w(t_{1,N}, t_{2,N}) = \frac{\sqrt{Nh}\Big(\big(Y_N(t_{2,N}) - S_N(t_{2,N})\big) - \big(Y_N(t_{1,N}) - S_N(t_{1,N})\big)\Big)}{\sqrt{\int_{t_{1,N}}^{t_{2,N}} \sigma^2(v)\mu(v)dv}}. \quad (8.69)$$

For the nominator we have

$$\begin{aligned}\Big|\big(Y_N(t_{2,N}) - S_N(t_{2,N})\big) - \big(Y_N(t_{1,N}) - S_N(t_{1,N})\big)\Big| \leqslant \\ |\widehat{S}_N(t_{2,N}) - \widehat{S}_N(t_{1,N})| + |(F_N(t_{2,N}) - S_N(t_{2,N})) - (F_N(t_{1,N}) - S_N(t_{1,N}))|. \quad (8.70)\end{aligned}$$

By (8.7) for some $C_{\mu,\sigma} > 0$

$$\begin{aligned}|\widehat{S}_N(t_{2,N}) - \widehat{S}_N(t_{1,N})| \leqslant C_{\mu,\sigma}\sqrt{\frac{-(t_{2,N} - t_{1,N})\log(t_{2,N} - t_{1,N})}{N}} \leqslant \\ C_{\mu,\sigma}\sqrt{(t_{2,N} - t_{1,N})}\sqrt{\frac{\log N}{N}} \quad (8.71)\end{aligned}$$

because by construction $(t_{2,N} - t_{1,N}) \geqslant 1/N$. As above we set

$$M_+ = \max_{t\in[0,1]} \sigma^2(t)\mu(t) \qquad \text{and} \qquad M_- = \min_{t\in[0,1]} \sigma^2(t)\mu(t)$$

and notice $M_- > 0$. Further by Theorem 8.3 for some constants A_1, A_2, A_3 and large N

$$\begin{aligned}&|(F_N(t_{2,N}) - S_N(t_{2,N})) - (F_N(t_{1,N}) - S_N(t_{1,N}))| \leqslant \\ &\left|\int_{t_{1,N}}^{t_{2,N}} f(v) - s_N(v)dv\right| + \left|\int_{t_{1,N}}^{t_{2,N}} f(v)dv - (F_N(t_{2,N}) - F_N(t_{1,N})\right| \leqslant \\ &(t_{2,N} - t_{1,N})\sup_{t\in\mathcal{Y}_N}|f(t) - s_N(t)| + A_1\sqrt{\frac{-(t_{2,N} - t_{1,N})\log(t_{2,N} - t_{1,N})}{N}} \leqslant \\ &(t_{2,N} - t_{1,N})A_2\sup_{t\in\mathcal{Y}_N}|f^{(1)}(t)|\left(\frac{\log N}{N}\right)^{1/3} + A_1\sqrt{\frac{(t_{2,N} - t_{1,N})\log N}{N}} \leqslant \\ &A_3\sqrt{(t_{2,N} - t_{1,N})}\sqrt{\frac{\log N}{N}}\end{aligned}$$

where we used 3 of Theorem 6.2 and that fact that $(t_{2,N} - t_{1,N}) \leqslant \left(\frac{\log N}{N}\right)^{1/3}$. Thus the latter combined with (8.71) and (8.69) implies for some $A_4 > 0$

$$|w(t,\delta)| \leqslant \frac{\sqrt{Nh}(C_{\mu,\sigma} + A_3)\sqrt{t_{2,N} - t_{1,N}}\sqrt{\log N/N}}{\sqrt{\int_{t_{1,N}}^{t_{2,N}} \sigma^2(v)\mu(v)dv}} \leqslant$$
$$\frac{\sqrt{h}(C_{\mu,\sigma} + A_3)\sqrt{t_{2,N} - t_{1,N}}\sqrt{\log N}}{\sqrt{M_-}\sqrt{t_{2,N} - t_{1,N}}} = A_4\sqrt{\log N}$$

and (6.13) holds.

Proof of 2 of Theorem 6.4 follows from 1 of this theorem and 2 of Theorem 6.2.

Proof of 3 of Theorem 6.4. By the definition of the taut string for any $t \in [0, 1]$

$$|y_N(t) - s_N(t)| \leqslant \frac{C_{\mathcal{T}}}{\sqrt{N}}$$

and using (8.69) it implies

$$|w(t_{1,N}, t_{2,N})| \leqslant \frac{\sqrt{N} 2 \frac{C_{\mathcal{T}}}{\sqrt{N}}}{\sqrt{\int_{t_{1,N}}^{t_{2,N}} \sigma^2(v)\mu(v)dv}} \leqslant \frac{2C_{\mathcal{T}}}{\sqrt{M_-(t_{2,N} - t_{1,N})}}.$$

Therefore (6.13) holds as

$$t_{2,N} - t_{1,N} \geqslant \frac{4C_{\mathcal{T}}^2}{\tau M_- \log N}.$$

Proof of 4 of Theorem 6.4. From (8.70) for the multiresolution coefficient we have

$$|w(t_1, t_2)| \geqslant \frac{\sqrt{Nh}\left((t_{2,N} - t_{1,N}) \inf_{t \in [t_1, t_2]} |f(t') - s_N(t')| - |s_N(t_2) - s_N(t_1)|\right)}{\sqrt{\int_{t_1}^{t_2} \sigma^2(v)\mu(v)dv}}. \tag{8.72}$$

From statement 2 of Theorem 6.2 for any extreme interval $[p_i^l,\, p_i^r]$ there exists $\delta_N > A_1 N^{-1/6}$ such that

$$[p_i^e - \delta_N,\, p_i^e + \delta_N] \subset [p_i^l,\, p_i^r]$$

and by (8.67) we can suppose

$$\inf_{t\in[p_i^e-\delta_N,\, p_i^e+\delta_N]} |f(t)-s_N(t)| \geqslant A_2 N^{-1/3} \tag{8.73}$$

for some constants A_1, A_2 and all sufficiently large N. Set $t_{1,N} = p_i^e - \delta_N$ and $t_{2,N} = p_i^e + 2\delta_N$. Then (8.72) and (8.73) using (8.7) yield

$$|w(t_{1,N}, t_{2,N})| \geqslant \frac{\sqrt{Nh}\left(2\delta_N A_2 N^{-1/3} - |\hat{s}_N(t_{2,N}) - \hat{s}_N(t_{1,N})|\right)}{\sqrt{\int_{t_{1,N}}^{t_{2,N}} \sigma^2(v)\mu(v)dv}} \geqslant$$

$$\frac{\sqrt{h}\left(2\delta_N A_2 N^{1/6} - C'_{\mu,\sigma}\sqrt{-\delta_N \log\delta_N}\right)}{\sqrt{2\delta_N M_+}} \geqslant \sqrt{\frac{h}{2M_+}}\Bigg(2A_2N^{1/6}\sqrt{\delta_N} -$$

$$C'_{\mu,\sigma}\sqrt{-\log\delta'_N}\Bigg) \geqslant A_3\left(A_2 N^{1/6}\sqrt{A_1 N^{-1/6}} - C'_{\mu,\sigma}\sqrt{\log(N^{1/6}/A_1)}\right) \geqslant$$

$$A_4 N^{1/12} - A_5\sqrt{\log N} \geqslant A_4 N^{1/12} \geqslant \sqrt{\tau \log N}$$

for any τ and all sufficiently large N □.

Chapter 9

Final remarks. Future research

The model that we have considered in this thesis is a homogeneous one. One of the possible directions for the future researches may be a construction of a model with the drift and/or volatility coefficients also directly depending on the time parameter.

The taut string method used in this thesis for the estimation of the drift term and the invariant density function minimizes the number of local extrema of the estimated function. At the same time the method, which have been used for the estimation of the volatility coefficient, provides the minimal number of constancy intervals. It looks very interesting to adopt the ideas and technique of the latter method for an estimation of the drift function.

And one more reasonable idea is to use the functional relation between the invariant density, the drift, and the volatility functions deducible from 2.1 and 2.4: provided the invariant density and the volatility we can then compute the drift function as

$$b(x) = \frac{1}{2\mu(x)} \frac{d\big(\mu(x)\sigma^2(x)\big)}{dx}.$$

But because the derivatives of the invariant density and the volatility are involved it is required to consider a smoothing of $\mu(\cdot)$ and $\sigma(\cdot)$ or the product $\mu\sigma^2$. We plan to use for this the technique like that proposed in Meise (2007).

Bibliography

[1] Abramowitz, M., and Stegun, I.A. (1992) *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Reprint of the 1972 edition. New York: Dover Publications.

[2] Aït-Sahalia, Y. (1996) Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica*, **64**, No.3, 527-560.

[3] Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9**, 385-426.

[4] Arnold L. (1974) *Stochastic Differential Equations: Theory And Application.* New York: John Wiley and Sons.

[5] Barlow, R., Bartholomew, D., Bremmer, J., and Brunk, H. (1972) *Statistical Inference under Other Restrictions.* New York: John Wiley and Sons.

[6] Billingsley, P. (1968) *Convergence of probability measures.* New York: John Wiley and Sons.

[7] Black, F., and Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637-654.

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

[9] Brown, S.J., and Dybvig, P.H. (1986) The empirical implications of the Cox, Ingersoll, Ross theory of the term strusture of interest rates. *Journal of Finance*, **41**, 617-630.

[10] Cai, Z., Hong, Y. (2003) Nonparametric methods in continuous-time finance: A selective review. In *Recent Advances and Trends in Nonparametric Statistics*, (M.G. Akritas and D.N. Politis, eds.), 283-302. North-Holland, Amsterdam.

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

[12] Chapman, D., and Pearson, N. (2000) Is the short rate drfit actually nonlinear? *Journal of Finance*, **55**, 355-388.

[13] Chen, M.F. (2001) Explicit criteria for several types of ergodicity. *Chinese Journal of Applied Probability and Statistics*, **17**, 1-8.

[14] Chen, M.F. (2002) Ergodic convergence rates of Markov processes – Eigenvalues, inequalitites and ergodic theory. *ICM 2002*, **III**, 41-52.

[15] Chen, X. (1999) Limit theorems for functionals of ergodic Markov chains with general state space. *Memoirs of the American Mathematical Society*, **138**, No. 664.

[16] Courtadon, G. (1982) The pricing of options on default-free bonds. *Journal of Financial and Quantitative Analysis*, **17**, 75-100.

[17] Constantinides, G.M. (1992) A theory of the nominal term structure of interest rates. *Review of financial Studies*, **5**, 531-552.

[18] Cox, J.C. (1975) Notes on option pricing I: Constant elasticity of variance diffusions. *Working paper*, Stanford University.

[19] Cox, J.C., Ingersoll, J.E., and Ross, S.A. (1980) An analysis of variable rate loan contracts. *Journal of Finance*, **35**, 389-403.

[20] Cox, J.C., Ingersoll, J.E., and Ross, S.A. (1985) A theory of the term structure of interest rate. *Econometrica*, **53**, 385-407.

[21] Davies, P.L. (2006) Long range financial data and model choice. *Universität Dortmund, Technical Report*, **21**/2006.

[22] Davies, P.L., and Kovac, A. (2001) Local extremes, runs, strings and multiresolutions (with discussion). *The Annals of Statistics*, **29**, No.1, 1-65.

[23] Davies, P.L., and Kovac, A. (2004) Densitites, spectral densities and modality. *The Annals of Statistics*, **32**, No.3, 1093-1136.

[24] Dothan, L.U.(1978) On the term structure of interest rate. *Journal of Financial Economics*, **6**, 59-69.

[25] Dudley, R. M. (1987) Universal Dunsker classes and metric entropy. *The Annals of Probability*, **15**, 1306-1326.

[26] Duffie, D., and Kan, R. (1993) A yield factor model of interest rates. *Working paper*, Stanford University.

[27] Fan, J. (2005a) A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**, No.4, 317-337.

[28] Fan, J. (2005b) Rejoinder: A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**, No.4, 351-357.

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

[30] Genest, C., Ghoudi, K. and Remillard, B. (1996) A note on tightness. *Statistics & Probability Letters*, **27**, 331-339.

[31] Genon-Catalot, V., Laredo, C., and Picard, D. (1992) Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics*, **19**, 317-335.

[32] Gibbons, M.R., and Ramaswamy, K. (1993) A test of the Cox, Ingersoll and Ross model of the term structure. *Review of Financial Studies*, **6**, 619-658.

[33] Gihman, I. and Skorohod, A. (1972) *Stochastic Differential Equations.* New York: Springer Verlag.

[34] Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, 135-163.

[35] Honoré, P. (1997) Maximum likelihood estimation of non-linear continuous-time term-structure models. *Working parper*, 1997-7, Department of Finance, Aarhus School of Business.

[36] Höhenrieder, C. (2007) Nichtparametrische Volatilitäts- und Trendapproximation von Finanzdaten *Working paper.*

[37] Jacod, J. (2000) Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics*, **27**, 83-96.

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

[39] Karlin, S., and Taylor, H.M. (1981) *A second course in stochastic processes.*, New York: Academic Press.

[40] Kelly, L., Platen, E., and Søorensen, M. (2004) Estimation for discretely observed diffusions using transform functions. *Journal of Applied Probability*, **41A**, 99-118.

[41] Kloeden, P.E., Platen, E., Schurz, H., and Søorensen, M. (1996) On effects of discretization on estimators of drift parameters for diffusion processes. *Journal of Applied Probability*, **33**, 1061-1076.

[42] Kutoyants, Y. (1984) *Parameter estimation for stochastic processes.*, Berlin: Helderman Verlag.

[43] Liptser, R., and Shiryaev, A. (2001) *Statistics of random processes: II. Applications.* (2nd, revised and expanded edition), New York: Springer Verlag.

[44] Mao, Y.H. (2006) Convergence rates in strong ergodicity for Markov processes. *Stochastic Processes and their Applications*, **116**, 1964-1976.

[45] Markus, M.B., and Shepp, L.A. (1971) Sample behavior of Gaussian processes. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **2**, 423-442.

[46] Marsh, T.A., and Rosenfeld, E.R. (1983) Stochastic processes for interest rates and equilibrium bond prices. *Journal of Finance*, **38**, 635-646.

[47] Meise, M. (2007) Density smoothing. *Working paper.*

[48] Mykland, P.A., and Zhang, L. (2005) Comment: A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**, No.4, 347-350.

[49] Nummelin, E. (1984) *General irreducible Markov chains and non-negetive operators.* Cambridge: Cambridge University Press.

[50] Petrov, V.V. (1975) *Sums of Independent Random Variables.*, Berlin: Springer Verlag.

[51] Philliphs, P.C.B., and Yu, J. (2005) Comment: A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**, No.4, 338-343.

[52] Pollard, D. (1984) *Convergence of Stochastic Processes.* New York: Springer Verlag.

[53] Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies*, **11**, 449-487.

[54] Reiß, M. (2006) Nonparametric volatility estimation on the real line from low-frequency data. In *The Art of Semiparametrics*, (S. Sperlich, W. Hrdle, G. Aydinli Ed.), 32-48, Physica, Heidelberg.

[55] Sørensen, H. (2004) Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, **72**, No.3, 337-354.

[56] Shoji, I. (2002) Nonparametric state estimation of diffusion processes. *Biometrika*, **89**, No.2, 451-456.

[57] Shoji, I. (2004) Nonparametric estimation of diffusion coefficients from discretely observed data. *2004 SRCCS Probability Workshop*, Seoul National University, No.11.

[58] Sørensen, M. (2005) Comment: A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**, No.4, 344-346.

[59] Spokoiny, V. (2003) Adaptive drift estimation for nonparametric diffusion model. *The Annals of Statistics*, **28**, 815-836.

[60] Tricomi, F.C. (1950) Asymtotische Eigenschaften der unvollstndigen Gammafunktion. *Mathematische Zeitschrift*, **53**, 136-148.

[61] Tsai, T.H. (1998) The uniform CLT and LIL for Markov chains. *PhD thesis*, University of Wisconsin, Madison.

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