Change point detection in heteroscedastic time series

Tomasz Górecki\textsuperscript{a}, Lajos Horváth\textsuperscript{b}, Piotr Kokoszka\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a} Faculty of Mathematics and Computer Science, Adam Mickiewicz University, 61-614 Poznań, Poland
\textsuperscript{b} Department of Mathematics, University of Utah, Salt Lake City, UT 84112-0090, USA
\textsuperscript{c} Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA

\section*{ARTICLE INFO}

\begin{itemize}
\item Article history:
\item Received 27 November 2016
\item Revised 17 July 2017
\item Accepted 19 July 2017
\item Available online 27 July 2017
\end{itemize}

\section*{ABSTRACT}

Many time series exhibit changes both in level and in variability. Generally, it is more important to detect a change in the level, and changing or smoothly evolving variability can confound existing tests. A framework for testing for shifts in the level of a series which accommodates the possibility of changing variability is developed. The resulting tests are robust both to heteroskedasticity and serial dependence. They rely on a new functional central limit theorem for dependent random variables whose variance can change or trend in a substantial way. This new result is of independent interest as it can be applied in many inferential contexts applicable to time series. Its application to change point tests relies on a new approach which utilizes Karhunen–Loève expansions of the limit Gaussian processes. After presenting the theory in the most commonly encountered setting of the detection of a change point in the mean, it is shown how it can be extended to linear and nonlinear regression. Finite sample performance is examined by means of a simulation study and an application to yields on US treasury bonds.

© 2017 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.

\section*{1. Introduction}

In the most common change point paradigm, we consider the model

$$X_i = \mu_i + u_i, \quad 1 \leq i \leq N,$$

with mean zero errors, $Eu_i = 0$, and wish to test the no change in the mean null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_N.$$

The general alternative is that $H_0$ does not hold, but we target several change point alternatives discussed in Examples 2.1–2.3. Csörgö and Horváth (1997) provide an account of early results in change point detection based mainly on independent and identically distributed error terms and connect the likelihood method to maximally selected CUSUM. Aue and Horváth (2013) explain the extension of some of the classical results to time series setting. Jeng (2015) provides an overview of change point detection in finance. In change point research, usually the homoscedasticity of the errors is assumed. Inclán and Tiao (1994), Gombay and Horváth (1994), Davis et al. (1995), Lee and Park (2001), Deng and Perron (2008), Antoch et al. (1997), Berkes et al. (2009), Aue et al. (2009), Wied et al. (2012), Wied et al. (2013) and Zhou (2013) propose tests when the mean and/or the variance are changing under the alternative, i.e. heteroscedastic errors can occur only under the alternative.

\* Corresponding author.

\textsuperscript{E-mail address: Piotr.Kokoszka@colostate.edu (P. Kokoszka).}

http://dx.doi.org/10.1016/j.ecosta.2017.07.005

2452-3062/© 2017 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.
Dalla et al. (2015) and Xu (2015) point out that in some applications the errors are heteroscedastic, which should be taken into account when we test the validity of $H_0$. Our paper is related to their work. It rigorously derives a new class of tests which are valid under weak assumptions, which do not require any mixing conditions. Busetti and Taylor (2004), Cavaliere et al. (2011), Cavaliere and Taylor (2008), Hansen (1992) and Harvey et al. (2006) investigate change point tests when some type of nonstationarity is exhibited by the data.

The paper is organized as follows. Section 2 develops the asymptotic framework. The limit distributions of the tests statistics are nonstandard if error heteroskedasticity is allowed. These distributions can however be computed using suitable Karhunen–Loève expansions, which also lead to practical ways of computing the critical or $P$-values, as explained in Section 3. Section 4 explores the finite sample performance of the tests. Proofs of the asymptotic results are collected in Section 5.

2. Assumptions and limit results

We consider heteroscedastic errors:

**Assumption 2.1.** $u_i = u_{i,N} = a(i/N)e_i$, $1 \leq i \leq N,$

where the function $a$ satisfies.

**Assumption 2.2.** $a(t), 0 \leq t \leq 1,$ has bounded variation on $[0, 1]$.

(For the definition and properties of functions with bounded variation we refer e.g. to Hewitt and Stromberg, 1969.)

We allow a very general class of errors $e_i$, $-\infty < i < \infty$.

**Assumption 2.3.** $Ee_i = 0$ and $e_i$, $1 \leq i < \infty$, satisfy the functional central limit theorem, i.e. there is $\sigma > 0$ such that

$$N^{-1/2} \sum_{i=1}^{[N]} e_i \overset{\text{law}}{\rightarrow} \sigma W(t),$$

where $W(t)$, $0 \leq t < \infty$, denotes a Wiener process (standard Brownian motion).

We do not assume stationarity or any form of mixing for the error terms, they must merely satisfy a Central Limit Theorem, which is a minimal requirement for the existence of an asymptotic distribution of common test statistics. The extension of the functional central limit theorem to dependent sequences has been a major topic in probability theory, statistics as well as in econometrics. Billingsley (1968) provides a review of early results and it is still a basic reference to prove functional limit theorems. Hall and Heyde (1980) not only establish the FCLT for martingales and mixingales using the Skorokhod embedding scheme, but also provide results on the rate of convergence. The monographs of Bradley (2007) and Dedecker et al. (2007) provide introduction to mixing processes and comprehensive surveys. The proof of the functional central limit theorem under Assumption 3.4 is established, for example, in Aue et al. (2014). The essence of Assumption 2.3 is that we do not need to assume any specific conditions described in the monographs cited above, but rather a conclusion that any of them implies

The theory of testing in the various contexts studied below is based on the following result.

**Theorem 2.1.** If Assumptions 2.1–2.3 are satisfied, then

$$N^{-1/2} \sum_{i=1}^{[N]} u_i \overset{\text{law}}{\rightarrow} W(b(t)),$$

where $W(u)$, $0 \leq u < \infty$, is a Wiener process (standard Brownian motion) and

$$b(t) = \sigma^2 \int_0^t a^2(u)du.$$

**Theorem 2.1** is a major theoretical contribution of this paper. It establishes the asymptotic behavior of the partial sum process for dependent random variables with evolving variance without imposing any stationarity or mixing conditions on the errors $e_i$. The time transformed Wiener process has been known to appear as a limit since the 1950s. Limit theorems similar to our Theorem 2.1 under mixing assumptions are discussed in Hall and Heyde (1980) and Davidson (1994). We show that mixing conditions are actually not needed. Theorem 2.1 can be used in settings that extend beyond change-point detection, for example in various unit root and trend tests.

In Section 2.1, we show how Theorem 2.1 leads to a class of change point tests in the setting of a potential change in mean. Section 2.2 extends the scope of applicability to regression models.
2.1. Change point in the mean

Recall the definition of the CUSUM process:

$$Z_N(t) = N^{-1/2} \left( \sum_{t=1}^{\lfloor Nt \rfloor} X_t - \frac{\lfloor Nt \rfloor}{N} \sum_{t=1}^{N} X_t \right).$$

In the setting of iid normal errors, the maximally selected CUSUM statistic can be derived from the maximum likelihood principle. Test based on other functionals of the CUSUM process are often the simplest and most effective in more general settings, and are, in fact, the most often used change point detection procedures. Our testing procedures are based on functionals of the CUSUM process as well. However, in the setting specified by Assumptions 2.1–2.3, especially 2.1, the asymptotic behavior of this process is very different than in the usual case of homoskedastic errors. Understanding this behavior is necessary to derive the tests. The weak convergence of the CUSUM process is an immediate consequence of Theorem 2.1.

Corollary 2.1. If $H_0$ and Assumptions 2.1–2.3 are satisfied, then we have that

$$Z_N(t) \xrightarrow{[0,1]} \Gamma(t), \quad \text{where} \quad \Gamma(t) = W(b(t)) - tW(b(1)).$$

It is easy to see that $\Gamma(t)$, $0 \leq t \leq 1$, is a Gaussian process with $E\Gamma(t) = 0$ and

$$C(t, s) = E\Gamma(t)\Gamma(s) = b(t \wedge s) - tb(s) - sb(t) + tsb(1),$$

(2.1)

where $t \wedge s = \min(t, s)$.

In general, the computation of the distribution functions of functionals of the limit in Corollary 2.1 is nearly impossible due to the dependence on the unknown function $b(t)$. However, combining the Karhunen–Loéve expansion with principal component analysis we can approximate the distributions of $L^2$ functionals. This is done in Section 3.

The derivation of the $L^2$ functional of the standardized $Z_N(t)$ requires a bound on the correlation between the elements of the sequence $\{e_i, -\infty < i < \infty\}$.

Assumption 2.4. $|Ee_ie_j| = O((|k - l| + 1)^{-k})$, with some $\kappa > 1$.

Corollary 2.2. If $H_0$ holds and Assumptions 2.1–2.3 are satisfied, then we have that

$$\int_0^1 Z_N^2(t) \, dt \xrightarrow{D} \int_0^1 \Gamma^2(t) \, dt.$$  \hspace{1cm} (2.2)

If, in addition, Assumption 2.4 also holds, then we have that

$$\int_{1/N}^{1-1/N} \frac{Z_N^2(t)}{t(1-t)} \, dt \xrightarrow{D} \int_0^1 \frac{\Gamma^2(t)}{t(1-t)} \, dt.$$  \hspace{1cm} (2.3)

The statistic in (2.2) is a version of the Cramér–von Mises statistic while (2.3) is a modification of the Anderson–Darling statistic (cf. Shorack and Wellner, 1986, p. 148).

We would like to note that $Z_N(t)$ is not "symmetric" since by definition $Z_N(t) = 0$, if $0 < t < 1/N$ while $|Z_N(t)| > 0$ if $1 - 1/N < t < 1$. However,

$$\tilde{Z}_N(t) = N^{-1/2} \left( \sum_{t=1}^{\lfloor N(t+1)u \rfloor} X_t - \frac{N(t+1)u}{N} \sum_{t=1}^{N} X_t \right)$$

is "tied down" in a neighborhood of 0 as well as 1. Relation (2.3) can be replaced with

$$\int_0^1 \tilde{Z}_N^2(t) \, dt \xrightarrow{D} \int_0^1 \tilde{\Gamma}^2(t) \, dt.$$  \hspace{1cm} (2.4)

We conclude this section by establishing the asymptotic behavior of the tests statistics when $H_0$ does not hold. Let

$$d_k = \sum_{t=1}^{k} (\mu_t - \bar{\mu}), \quad 1 \leq k \leq N, \quad \text{with} \quad \bar{\mu} = \frac{1}{N} \sum_{t=1}^{N} \mu_t,$$

be the drift term of the CUSUM process. We impose the following minor restrictions on the expected values of the observations:

Assumption 2.5. $\max_{1 \leq k \leq N} |\mu_k| = O(1)$ and

$$d(u) = \lim_{N \to \infty} \frac{d_{\lfloor Nu \rfloor}}{N}, \quad 0 \leq u \leq 1,$$

exists. (We use $d_{\lfloor Nu \rfloor} = 0$, if $0 \leq u < 1/N$.)
Assumption 2.5 is automatically satisfied under $H_0$, since $d_k = 0$ for all $1 \leq k \leq N$, i.e. there is no drift. The following theorem will be used to establish the consistency of the tests.

**Theorem 2.2.** If Assumptions 2.1–2.3 and 2.5 are satisfied, then we have that

$$\frac{1}{N} \int_0^1 Z^2_N(t) dt \overset{p}{\to} \int_0^1 d^2(u) du. \quad (2.5)$$

If, in addition, Assumption 2.4 also holds, then we have that

$$\frac{1}{N} \int_{1/N}^{1-1/N} Z^2_N(t) \frac{dt}{t(1-t)} \overset{p}{\to} \int_0^1 d^2(t) \frac{dt}{t(1-t)}. \quad (2.6)$$

If

$$\int_0^1 d^2(u) du > 0 \quad (2.7)$$

holds, then

$$\int_0^1 Z^2_N(t) dt \overset{p}{\to} \infty \quad \text{and} \quad \int_{1/N}^{1-1/N} Z^2_N(t) \frac{dt}{t(1-t)} \overset{p}{\to} \infty.$$  

However, to establish the consistency, we have to carefully examine the asymptotic behavior of the estimated eigenvalues; this is done in Sections 3.1 and 3.2. In the following, condition (2.7) is adopted as the formal definition of $H_A$.

We conclude with some examples of the alternative to which Theorem 2.2 applies.

**Example 2.1.** We say that there is exactly one change in the mean if $\mu_1 = \bar{\mu}_1, 1 \leq i \leq \lfloor N \theta \rfloor$, and $\mu_1 = \bar{\mu}_2, \lfloor N \theta \rfloor < i \leq N$, where $\bar{\mu}_1 \neq \bar{\mu}_2$ and $0 < \theta < 1$. In this case $\lfloor N \theta \rfloor$ is the time of change and

$$d(t) = \begin{cases} t \bar{\mu}_1 - t(\theta \bar{\mu}_1 + (1-\theta)\bar{\mu}_2), & \text{if } 0 \leq t \leq \theta, \\ \theta \bar{\mu}_1 + (t-\theta)\bar{\mu}_2 - t(\theta \bar{\mu}_1 + (1-\theta)\bar{\mu}_2), & \text{if } \theta \leq t \leq 1. \end{cases}$$

**Example 2.2.** If $\mu_1 = \bar{\mu}_1, \lfloor N \theta_{i-1} \rfloor < i \leq \lfloor N \theta_i \rfloor$, $1 \leq i \leq m+1$, $\theta_0 = 0, \theta_{m+1} = N$ and $\mu_i \neq \mu_{i'}$ for some $i \neq i'$, we have at most $m$ changes in the mean. With $\bar{\mu} = (\theta_1 - \theta_0)\bar{\mu}_1 + (\theta_2 - \theta_1)\bar{\mu}_2 + \cdots + (\theta_{m+1} - \theta_m)\bar{\mu}_{m+1}$, we get

$$d(t) = \begin{cases} t \bar{\mu}_1 - t\bar{\mu}, & \text{if } 0 \leq t \leq \theta_1, \\ \theta_1 \bar{\mu}_1 + (t-\theta_1)\bar{\mu}_2 - t\bar{\mu}, & \text{if } \theta_1 \leq t \leq \theta_2, \\ \vdots \\ \theta_m \bar{\mu}_1 + (\theta_2 - \theta_1)\bar{\mu}_2 + \cdots + (t - \theta_m)\bar{\mu}_{m+1} - t\bar{\mu}, & \text{if } \theta_m \leq t \leq \theta_{m+1}. \end{cases}$$

**Example 2.3.** Let $d(t)$ be a continuous function on $[0,1]$ and define $\mu_i = \bar{\mu}_1, 1 \leq i \leq \lfloor N \theta \rfloor$ and $\mu_i = \bar{d}(i/N)$, if $\lfloor N \theta \rfloor < i \leq N$. If $\bar{d}(t)$ is different from $\bar{\mu}_1$, the mean is constant before $\lfloor N \theta \rfloor$ but it is determined by $\bar{d}(t)$ after the time of change. Now

$$d(t) = \begin{cases} t \bar{\mu}_1 - t\left(\theta \bar{\mu}_1 + \int_0^1 \bar{d}(u) du\right), & \text{if } 0 \leq t \leq \theta, \\ \theta \bar{\mu}_1 + \int_0^t \bar{d}(u) du - t\left(\theta \bar{\mu}_1 + \int_0^1 \bar{d}(u) du\right), & \text{if } \theta \leq t \leq 1. \end{cases}$$

This example includes linearly or polynomially increasing/decreasing means after the change.

2.2. **Change point in the parameters of linear and nonlinear regression**

Section 2.1 focused on the theory of testing for a change point in mean in order to explain the essence of our approach and provide the details in that most common setting. In this section, we consider more general regression settings. The proofs use suitably defined model residuals which approximate the unobservable heteroskedastic errors $u_i$. The tests of Section 2.1 correspond to the residuals $\tilde{u}_i = X_i - \tilde{X}_N$, $1 \leq i \leq N$. Once the residuals are defined, the asymptotic arguments parallel those used to establish the results of Section 2.1, so we just outline the proofs.

Consider first the usual linear regression

$$X_i = x_i^T \beta_i + u_i, \quad 1 \leq i \leq N.$$  

We wish to test $H_0 : \beta_1 = \beta_2 = \cdots \beta_N$ against the change point alternative. The following standard assumption is made.

**Assumption 2.6.** The sequences $\{x_i\}$ and $\{u_i\}$ are independent. The sequence $\{x_i\}$ is stationary, ergodic, and $E\|x_0\|^2 < \infty$.  


We use the least squares estimator \( \hat{\beta}_N = A_N^{-1} X_N \), where
\[
A_N = \begin{bmatrix}
X_1^T \\
\vdots \\
X_N^T
\end{bmatrix}^T \begin{bmatrix}
X_1^T \\
\vdots \\
X_N^T
\end{bmatrix}, \quad X_i = [x_i(1), \ldots, x_i(p)]^T,
\]
where \( p \) is the dimension of the parameter vector. By the ergodicity of the regressors, \( A_N/N \xrightarrow{p} A_0 \) according to the ergodic theorem. The residuals are defined by
\[
\hat{u}_i = X_i - x_i^T \hat{\beta}_N, \quad 1 \leq i \leq N.
\]

**Theorem 2.3.** Under Assumptions 2.1–2.4, 2.6, and assuming that \( A_0 \) is nonsingular, Corollaries 2.1 and 2.2 remain valid for the residuals defined in (2.8).

We now turn to the nonlinear regression
\[
X_i = h(x_i, \theta_i) + u_i, \quad 1 \leq i \leq N,
\]
where \( \theta_i \) are \( p \)-dimensional parameter vectors, equal under \( H_0 \). The unknown common parameter vector is then \( \theta_0 \). It is estimated by minimizing
\[
L_N(\theta) = \sum_{i=1}^{N} (X_i - h(x_i, \theta))^2
\]
over the parameter space \( \Theta \). The following usual assumption is made.

**Assumption 2.7.** The parameter space \( \Theta \) is a compact subset of \( \mathbb{R}^p \) and \( \theta_0 \) is its interior point.

We impose the following assumption on the function \( h(\cdot, \cdot) \).

**Assumption 2.8.** The following conditions hold:
\[
sup_{\theta \in \Theta} Eh^2(x_0, \theta) < \infty, \\
\sup_{\theta \in \Theta} \left\| \frac{\partial^2}{\partial \theta^2} h(x_i, \theta) \right\| \leq M(x_i), \quad EM(x_0) < \infty
\]
and
\[
E \left\| \frac{\partial}{\partial \theta} h(x_0, \theta_0) \right\|^2 < \infty,
\]
\[
E[h(x_0, \theta_0) - h(x_0, \theta)]^2 > 0, \quad \text{if} \quad \theta \neq \theta_0.
\]

The conditions formulated in Assumption 2.8 ensure that under \( H_0 \) the differences between the functionals based on the unobservable errors \( u_i \) and the residuals
\[
\hat{u}_i = X_i - h(x_i, \hat{\theta})
\]
are asymptotically negligible in the sense that they do not affect the limits in Corollaries 2.1 and 2.2.

**Theorem 2.4.** Under Assumptions 2.1–2.4 and 2.7–2.8, Corollaries 2.1 and 2.2 remain valid for the residuals defined in (2.9).

The consistency of the tests in both linear and nonlinear regression settings can be established in a framework analogous to that considered in Section 2.1.

### 3. Computation of the limit distributions based on Karhunen–Loéve expansions

The Karhunen–Loéve expansion yields that
\[
\int_0^1 \Gamma^2(t) dt = \sum_{i=1}^{\infty} \lambda_i \xi_i^2, \quad (3.1)
\]
where \( \xi_1, \xi_2, \ldots \) are independent and identically distributed standard normal random variables, and \( \lambda_1 \geq \lambda_2 \geq \cdots \) are the eigenvalues of the operator associated with the kernel \( C(t, s) \) defined in (2.1). Specifically,
\[
\lambda_i \phi_i(t) = \int_0^1 C(t, s) \phi_i(s) ds, \quad 1 \leq i < \infty. \quad (3.2)
\]
The $\varphi_1, \varphi_2, \ldots$ are orthonormal functions, the eigenfunctions of $C(t, s)$. Similarly,

$$
\int_0^1 \frac{\Gamma^2(t)}{(1-t)} dt = \sum_{i=1}^{\infty} \tau_i \xi_i^2, 
$$

where $\tau_1 \geq \tau_2 \geq \cdots$ are the eigenvalues of

$$
D(t, s) = \frac{C(t, s)}{((1-t)(1-s))(1-t)^{1/2}}.
$$

The eigenvalues $\lambda_i$, $i \geq 1$, as well as $\tau_i$, $i \geq 1$, can be estimated from the sample. This is addressed in Sections 3.1 and 3.2. Using (3.1) or (3.3), we can obtain critical values for the Cramér–von Mises and Anderson–Darling statistics by proceeding as follows. If $\hat{C}_N(t, s)$ is an $L^2$ consistent estimate of $C(t, s)$, then the empirical eigenvalues $\hat{\lambda}_{N,1} \geq \hat{\lambda}_{N,2} \geq \hat{\lambda}_{N,3} \geq \cdots$ of $\hat{C}_N(t, s)$ can be used to approximate the sum on the right hand side of (3.1) by

$$
\hat{\lambda}_N = \sum_{i=1}^{N} \hat{\lambda}_{N,i} \xi_i^2.
$$

The empirical eigenvalues satisfy the integral equation

$$
\hat{\lambda}_{N,i} \hat{\varphi}_{N,i}(t) = \int_0^1 \hat{C}_N(t, s) \hat{\varphi}_{N,i}(s) ds,
$$

where $\hat{\varphi}_{N,i}(t), i \geq 1$, are orthonormal eigenfunctions. The same method works to approximate the distribution in (3.3).

We now turn to the details of the computation of the $\hat{\lambda}_{N,i}$ and the $\hat{\tau}_{N,i}$, first in the case of uncorrelated errors, then in the general case of correlated errors.

### 3.1. Estimation of the eigenvalues in case of uncorrelated errors

To illustrate our method, we first consider uncorrelated observations:

**Assumption 3.1.**

$$
Ee_i e_j = \begin{cases} 0, & \text{if } i \neq j, \\ \sigma^2, & \text{if } i = j. \end{cases}
$$

Assumption 3.1 holds for a large class of random variables, including GARCH-type sequences, Francq and Zakoian (2010). Let

$$
\tilde{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i
$$

and define

$$
H_N(u) = \frac{1}{N} \sum_{i=1}^{[Nu]} (X_i - \tilde{X}_N)^2, \quad 0 \leq u \leq 1.
$$

We estimate $C(t, s)$ with

$$
\hat{C}_N(t, s) = H_N(t \wedge s) - tH_N(s) - sH_N(t) + stH_N(1).
$$

Let

$$
g(u) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{[Nu]} (\mu_i - \mu)^2.
$$

It is clear that under $H_0$, $g(u) = 0$ for all $0 \leq u \leq 1$.

**Theorem 3.1.** We assume that Assumptions 2.1–2.3, 3.1 are satisfied and \( \{e_i, -\infty < i < \infty\} \) is a stationary and ergodic sequence with $Ee_0 = 0$ and $Ee_0^2 = \sigma^2 < \infty$.

(i) If $H_0$ holds, then

$$
\int_0^1 \int_0^1 (\hat{C}_N(t, s) - C(t, s))^2 dt ds = o_P(1).
$$
(ii) If $H_A$ holds, and in addition Assumption 2.5 also holds, then
\[ \int_0^1 \int_0^1 (\tilde{C}_N(t, s) - C^*(t, s))^2 \, dt \, ds = o_P(1). \]

where
\[ C^*(t, s) = C(t, s) + g(t \wedge s) - tg(s) - sg(t) + tsg(1). \]

We obtain from Theorem 3.1 (see e.g. Lemma 2.2 in Horváth and Kokoszka, 2012 or Dunford and Schwartz, 1988) that under $H_0$
\[ \hat{\lambda}_{N,i} \xrightarrow{p} \lambda_i. \tag{3.6} \]

It is easy to see that $C^*(t, s)$ is a symmetric, non-negative definite function. Let $\lambda_1^* \geq \lambda_2^* \geq \lambda_3^* \geq \cdots$ be the eigenvalues of $C^*$. We conclude that under $H_A$
\[ \hat{\lambda}_{N,i} \xrightarrow{p} \lambda_i^*. \tag{3.7} \]

For any $0 < \alpha < 1$, we define $\hat{c}_N(\alpha)$ as
\[ \hat{c}_N(\alpha) = \min \{ x : P(\bar{H}_N \geq x) \leq \alpha \}, \]

where $\bar{H}_N$ is defined in (3.4). The empirical critical value $\hat{c}_N(\alpha)$ is asymptotically correct. It follows from Corollary 2.2, (3.1) and (3.6) that under $H_0$,
\[ \lim_{N \to \infty} P \left\{ \int_0^1 Z_N^2(t) \, dt \geq \hat{c}_N(\alpha) \right\} = \alpha. \]

By (2.5), (2.7) and (3.7), we conclude that under $H_A$,
\[ \lim_{N \to \infty} P \left\{ \int_0^1 Z_N^2(t) \, dt \geq \hat{c}_N(\alpha) \right\} = 1, \]

establishing the consistency of the Cramér–von Mises procedure. The same arguments apply to the Anderson–Darling procedure.

3.2. Estimation of the eigenvalues in case of correlated errors

In the general case of correlated errors, the kernel $C(t, s)$ is estimated by
\[ \tilde{C}_N(t, s) = \hat{g}_N(t \wedge s) - t\hat{g}_N(s) - s\hat{g}_N(t) + st\hat{g}_N(1), \quad 0 \leq s, t \leq 1, \tag{3.8} \]

where
\[ \hat{g}_N(u) = \hat{g}_{N[Nu]}, \quad 0 \leq u \leq 1. \]

and where $\hat{g}_{N[k]}$ is an estimator of the long-run variance based on the partial sample $X_1, X_2, \ldots, X_k, k \leq N$.

In the following, we establish the asymptotic validity of the tests, both under $H_0$ and $H_A$, for the commonly used kernel estimator $\hat{g}_{N,k}$.

For any $1 \leq k \leq N$ and $\ell, |\ell| < k$ we define
\[ \hat{y}_{N,k,\ell} = \sum_{i=1}^{k-\ell} \frac{1}{N} (X_i - \bar{X}_N)(X_{i+\ell} - \bar{X}_N), \quad \text{if } 0 \leq \ell \leq k - 1, \]
\[ \hat{y}_{N,k,\ell} = \frac{1}{N} \sum_{i=\ell+1}^k (X_i - \bar{X}_N)(X_{i+\ell} - \bar{X}_N), \quad \text{if } -(k-1) \leq \ell < 0. \]

Let
\[ \hat{g}_{N,k} = \sum_{\ell=-(k-1)}^{k-1} K(\ell/h) \hat{y}_{N,k,\ell}. \]

We assume standard conditions on the kernel $K$ and window (smoothing) parameter $h$:

Assumption 3.2. (i) $K(0) = 1$ (ii) $K(u) = K(-u) \geq 0$ (iii) $K(u) = 0$ if $|u| > c$ with some $c > 0$ (iv) $K$ is Lipschitz continuous on the real line

and

Assumption 3.3. $h = h(N) \to \infty$ and $h/N \to 0$. 

The study of the the estimator \( \hat{g}_N(t) \), \( 0 \leq t \leq 1 \), requires assumptions on the structure of the innovations \( e_i \), \(-\infty < i < \infty\). We assume that the \( e_i \)'s form a Bernoulli shift:

**Assumption 3.4.** (i) \( e_i = f(e_i, e_{i-1}, \ldots) \), where \( f \) is a measurable functional and \( e_i \), \(-\infty < i < \infty\), are independent and identically distributed random variables in some measure space.

(ii) \( E_0 = 0 \) and \( E|e_0|^4 < \infty \).

(iii) \[
(E|e_i, m - e_i|^4)^{1/4} = O(m^{-\alpha}) \quad \text{with some } \alpha > 2,
\]
where \( e_i, m = g(e_i, e_{i-1}, \ldots, e_{i-m}, e_{i+1}, e_{i+2}, \ldots) \), \(-\infty < i, m, \ell < \infty\), are independent and identically distributed copies of \( e_0 \).

We note that Assumption 3.4 implies that \( e_i \), \(-\infty < i < \infty\), is a stationary sequence and Assumptions 2.3 and 2.4 are also satisfied (cf. Berkes et al., 2013 and Jirak, 2013). To get the exact limit of \( C_N(t, s) \) under \( H_\lambda \) we need a further regularity condition:

**Assumption 3.5.**

\[
\max_{1 \leq t \leq c h} \frac{1}{N} \sum_{1 \leq i \leq k} |\mu_{i+k} - \mu_i| = o(1),
\]
where \( c \) is defined in Assumption 3.2.

It is easy to see that Examples 2.1–2.3 satisfy Assumption 3.5.

**Theorem 3.2.** We assume that Assumptions 2.1, 2.2 and 3.2–3.4 are satisfied.

(i) If \( H_0 \) holds, then

\[
\int_0^1 \int_0^1 \hat{C}_N(t, s) - C(t, s))^2 dt ds = o_P(1).
\]

(ii) If \( H_\lambda \) and, in addition, Assumption 2.5 hold, then

\[
\int_0^1 \int_0^1 \hat{C}_N(t, s) dt ds = O_P(h^2).
\]

(iii) If \( H_\lambda \) and, in addition, Assumptions 2.5 and 3.5 hold, then

\[
\int_0^1 \int_0^1 \left( \frac{1}{N} \hat{C}_N(t, s) - C^*(t, s) \right)^2 dt ds = o_P(1),
\]

where

\[
C^*(t, s) = (g(t \wedge s) - t g(s) - s g(t) + t s g(1)) \int_{-c}^c K(u) du.
\]

Let \( \tilde{\lambda}_{N,1} \geq \tilde{\lambda}_{N,2} \geq \ldots \) denote the eigenvalues of \( \hat{C}_N(t, s) \). It follows from Theorem 3.2 (i), analogously to (3.6), that

\[
\tilde{\lambda}_{N,i} \xrightarrow{p} \lambda_i,
\]

where \( \lambda_1 \geq \lambda_2 \geq \ldots \) are the eigenvalues of \( C(t, s) \) defined in (3.2). For any \( 0 < \alpha < 1 \), we now define \( \hat{c}_N(\alpha) \) as

\[
\hat{c}_N(\alpha) = \min \{ x : P(\hat{H}_N \geq x) \leq \alpha \}.
\]

where

\[
\hat{H}_N = \sum_{i=1}^N \tilde{\lambda}_{N,i} \hat{\xi}_i^2,
\]

and \( \hat{\xi}_1, \hat{\xi}_2, \ldots \) are independent standard normal random variables. It follows from Corollary 2.2, (3.1) and (3.12) that under \( H_0 \),

\[
\lim_{N \to \infty} P \left( \int_0^1 Z_N^2(t) dt \geq \hat{c}_N(\alpha) \right) = \alpha.
\]

However, the consistency of the procedure is more delicate, since the empirical eigenvalues \( \tilde{\lambda}_{N,i} \) might not have a finite limit as \( N \to \infty \). Indeed, under Assumption 2.5 we have that \( \tilde{\lambda}_{N,i}/h \) converges in probability to a finite limit. Since

\[
\tilde{\lambda}_{N,i} \hat{\phi}_{N,i}(t) = \int_0^1 \hat{C}_N(t, s) \hat{\phi}_{N,i}(s) ds,
\]
where the $\tilde{\varphi}_{N,i}(t)$’s are orthonormal functions, we get from (3.10) via the Cauchy–Schwartz inequality that

$$
\hat{\lambda}_{N,i}^2 = \int_0^1 \left( \int_0^1 \tilde{C}_N(t,s) \tilde{\varphi}_{N,i}(s) ds \right)^2 dt \\
\leq \int_0^1 \left( \int_0^1 \tilde{C}_{N,i}(s,t) ds \int_0^1 \tilde{\varphi}_{N,i}^2(s) ds \right) dt \\
= \int_0^1 \int_0^1 \tilde{C}_{N,i}(s,t) ds dt = O(h^2).
$$

Hence, under $H_i$, we have that $\hat{\lambda}_{N,i} = O_P(h)$, implying that for each $m$ and $0 < \alpha < 1$, $\tilde{H}_N = O_P(h)$ and therefore $\tilde{C}_N(\alpha) = O_P(h)$.

According to Theorem 2.2, if (2.7) holds, then $\int_0^1 Z^2_N(t) dt / N$ converges to a positive limit in probability, and since $N/h \to \infty$, we conclude

$$
\lim_{N \to \infty} P \left\{ \int_0^1 Z^2_N(t) dt \geq \tilde{C}_N(\alpha) \right\} = 1,
$$

establishing the consistency of the Cramér–von Mises procedure in case of correlated observations when Assumption 3.3 holds. Similar arguments apply to the Anderson–Darling procedure.

Our arguments leading to (3.13) show that if (2.7) is satisfied, then the power is a monotone function of $N/h$. If $N/h(N)$ is a monotone function of $N$, the power of our test will be monotone as well. For results on the monotonicity of change point tests we refer to Juhl and Xiao (2009).

4. Simulation study and application to US yields

The purpose of this section is to assess the finite sample performance of the proposed tests. After describing them in a systematic manner in Section 4.1, we explore in Section 4.2 their properties using simulated and real data.

4.1. Test procedures

For ease of reference, we begin by listing the test procedures introduced in this paper and in related work together with convenient abbreviations. We also provide their brief descriptions. The procedures are based on the following ingredients, which also form the building blocks of the abbreviations.

**Standard vs. Heteroskedastic.** In the standard approach we assume that $a(t) = 1$, i.e. we do not take the possible heteroskedasticity of the errors into account. In the heteroskedastic approach, the function $a(\cdot)$ is estimated as explained in the previous sections.

**Uncorrelated vs. Correlated.** In the uncorrelated case, we estimate the eigenvalues as described in Section 3.1, i.e. we assume that the observations are uncorrelated. In the correlated case, we estimate the eigenvalues as described in Section 3.2, i.e. we assume that the observations are correlated.

**C vs. A.** This refers to using either the Cramér–von Mises or the Anderson–Darling functional.

We also consider two methods studied by Dalla et al. (2015), which they denote VS+ and VS, but which we denote VSU and VSC to emphasize more clearly that VSU assumes uncorrelated errors, while VSC assumes correlated errors. Finally, we include one of the methods of Zhou (2013).

We now list the methods we study.

**SUC** (Standard Uncorrelated errors Cramér–von Mises.) Set

$$
\hat{T}_N = \int_0^1 Z^2_N(t) dt
$$

and denote by $\hat{\sigma}^2$ is the sample variance of the observations $X_i$. Then

$$
\frac{\hat{T}_N}{\hat{\sigma}^2} \overset{\mathcal{D}}{\to} \int_0^1 B^2(t) dt,
$$

where $B(\cdot)$ is the standard Brownian motion. The distribution of the right-hand side of (4.2) is known, Shorack and Wellner (1986).

**SCC** The only difference between this method and SUC is that in (4.2), $\hat{\sigma}^2$ is a consistent estimator of the long-run variance of the $X_i$. 

HUC (Heteroskedastic Uncorrelated errors Cramér–von Mises.) The test statistic is $T_N$ defined by (4.1). Its limit distribution is approximated by the empirical distribution of the random variable

$$T(m) = \sum_{i=1}^{m} \hat{\lambda}_i \xi_i^2.$$  

The $\xi_j$ are independent standard normal. The $\hat{\lambda}_i$ satisfy

$$\hat{\lambda}_i \hat{\phi}_i(t) = \int_0^1 \hat{C}(t, s) \hat{\phi}_i(s) ds,$$

where $\hat{C}$ is given by (3.5). The $P$-value is computed as

$$P = \frac{1}{R} \# \{ r : T(m)_r > \hat{T}_N \},$$

where $T(m)_r$, $r = 1, 2, \ldots, R$, are independent replications of $T(m)$.

HCC (Heteroskedastic Correlated errors Cramér–von Mises.) Conceptually, the only difference between this method and SUC is that $\hat{C}$ is replaced by $\hat{C}$ given by (3.8). We note that $\hat{\rho}_{N,k}$ is the estimator of the long-run variance of $X_1, X_2, \ldots, X_k$, and any suitable estimator can be used. To enhance comparison, we used the spectral estimator employed by Dalla et al. (2015). We thank Dr. V. Dalla for making the code available.

SUA (Standard uncorrelated errors Anderson–Darling.) Set

$$\hat{U}_N = \int_{1/N}^{1-1/N} \frac{Z_N^2(t)}{t(1-t)} dt$$

(4.3)

and denote by $\hat{\sigma}_N^2$ the sample variance of the $X_i$. Then

$$\frac{\hat{U}_N}{\hat{\sigma}_N^2} \xrightarrow{D} \int_0^1 \frac{B^2(t)}{t(1-t)} dt,$$  

(4.4)

where $B(\cdot)$ is the standard Brownian motion. The distribution of the right-hand side of (4.2) is known, Shorack and Wellner (1986).

SCA The only difference between this method and SUA is that in (4.4), $\hat{\sigma}^2$ is a consistent estimator of the long-run variance of the $X_i$.

HUA (Heteroskedastic Uncorrelated errors Anderson–Darling.) The test statistic is $\hat{U}_N$ defined by (4.3). Its limit distribution is approximated by the empirical distribution of the random variable

$$U(m) = \sum_{i=1}^{m} \tilde{\xi}_i \tilde{\xi}_i^2.$$  

The $\tilde{\xi}_j$ are independent standard normal. The $\tilde{\xi}_i$ satisfy

$$\tilde{\xi}_i \tilde{\phi}_i(t) = \int_{1/N}^{1-1/N} \hat{D}(t, s) \hat{\phi}_i(s) ds,$$

where

$$\hat{D}(t, s) = \frac{\hat{C}(t, s)}{\sqrt{t(1-t)(1-s))(1-s))}.$$  

with $\hat{C}$ given by (3.5). The $P$-value is computed as

$$P = \frac{1}{R} \# \{ r : U(m)_r > \hat{U}_N \},$$

where $U(m)_r$, $r = 1, 2, \ldots, R$, are independent replications of $U(m)$.

HCA (Heteroskedastic correlated errors Anderson–Darling.) The only difference between this method and HUA is that $\hat{C}$ is replaced by $\tilde{C}$ given by (3.8). The specific implementation is the same as for the HCC method.

VSU (VS statistic uncorrelated errors) The test statistic is

$$\hat{V}_N = \frac{1}{\hat{\sigma}^2 N^2} \sum_{k=1}^{N} \left( S_k' - \bar{S} \right)^2, \quad S_k' = \sum_{i=1}^{k} (X_i - \bar{X}),$$

where $\hat{\sigma}^2$ is the sample variance of the observations $X_i$, $1 \leq i \leq N$. Its null distribution is approximated by the distribution of the random variable

$$V(m) = \sum_{k=1}^{m} \frac{\chi^2(2)}{4\pi^2 k^2},$$

where $\chi^2(2)$ is the chi-squared distribution with 2 degrees of freedom.
Table 4.1
List of models used to evaluate empirical size. The modulating functions are as follows: \( a_1 = i/2N; a_2 = 0.25(0 \leq i \leq 0.5N) + 0.5[0.5N < i \leq N] \); the \( R_i \) are GARCH(1,1) processes with \( \omega = 10^{-6}, \alpha = 0.2, \beta = 0.5 \), alternatively \( \beta = 0.7 \).

<table>
<thead>
<tr>
<th>Model</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-1</td>
<td>( X_i = a_1 R_i )</td>
</tr>
<tr>
<td>G-2</td>
<td>( X_i = a_2 R_i )</td>
</tr>
<tr>
<td>G-1A</td>
<td>( X_i = a_1 (</td>
</tr>
<tr>
<td>G-2A</td>
<td>( X_i = a_2 (</td>
</tr>
</tbody>
</table>

where the \( \chi^2_2 \) are iid chi-square with 2 degrees of freedom. If \( \hat{V}_N^2 \) is the observed value of the statistic, then the \( P \)-value is computed as

\[
P = \frac{1}{R} \# \{ r : V(m_r) \geq \hat{V}_N^2 \},
\]

where \( V(m_r), r = 1, 2, \ldots, R \), are independent replications of \( V(m) \).

VSC (VS statistic correlated errors) The only difference between this method and VSU is that in the definition of the test statistics, say \( \hat{V}_N \), \( \hat{\sigma}^2 \) is replaced by a consistent estimator of the long-run variance of the observations \( X_i, 1 \leq i \leq N \), i.e. by \( \hat{g}_{N,N} \) in the notation of Section 3.2.

RB2 One of the bootstrap methods of Zhou (2013), which is described in his paper.

We emphasize that, in contrast to the H-methods introduced in this paper, the common asymptotic distribution of the statistics \( \hat{V}_N \) and \( \hat{V}_S \) does not depend on the data. These statistics do not directly take possible heteroskedasticity into account; their applicability is based on empirically and theoretically established relative insensitivity to heteroskedastic errors. The methods of Zhou (2013) do not aim at utilizing a pivotal limit, but rather approximate it be bootstrap procedures.

4.2. Empirical rejection rates and application to US yields

All rejection rates reported in this section are based on one thousand replications. We analyzed the size and power of the tests for all models considered by Dalla et al. (2015). Regarding the empirical size, our HU tests have similar size as the VSU test; the HC tests have size similar to the VSC test. Generally, the differences in empirical size within these two categories of tests are within the standard error of the rejection rates. For illustration, Tables 4.2 and 4.3 provide selected results in case of correlated and heteroskedastic errors, the most general case. With prior knowledge that the errors are uncorrelated, the U-methods can be expected to have correct size only if the observations have the form \( X_i = a_1 R_i \). The \( R_i \) are realizations of a GARCH process. Without any prior knowledge about correlation and heteroskedasticity of the errors, only methods HCC, HCA and VSC should be applied. In most cases, there is no clear advantage of any of these methods over the other. If the observations have heavy tails, the case of \( X_i = a_1 (R_i^2 - E R_i^2) \) with \( \beta = 0.7 \) the VSC method overrejects, the empirical size is over 8% at the nominal size of 5%. Generally, our HC methods tend to have size slightly smaller than the nominal size, the VSC method a somewhat larger size. The method RB2 overrejects in case of models and sample sizes we considered.

Despite the oversized rejection rates under \( H_0 \), the VSC method often has smaller power the our HC methods. This illustrated in Table 4.5. Only the C methods are included, the rejection rates for the A methods are very similar. It is however possible to find cases in which the VSC method dominates our HC methods. In our simulations, this happens if the variance or the errors drops a lot. In the cases of \( a_{44} \) the variance of the errors drops from 1 to 1/16 in the second half of the sample.
The HC methods apparently “keep” the larger estimates of the variances based on initial realizations $X_1, X_2, \ldots, X_k$. These much larger initial variances suppress the values of the HC test statistics, resulting in smaller power.

To shed more light on the power behavior of these tests, we apply them to time series of yields on US treasury bills, which are shown in Fig. 4.1. There is an obvious drop in the yields, $H_A$ is true, which is accompanied by a drop in variance. Such a scenario might appear to favor the VSC method. However, as Table 4.6 shows, it does not detect a fairly obvious change. Our methods detect the change in 3 month yields with statistical significance, and a change in 12 month yields with borderline significance ($P$-values about 5%). Method RB2 also detects a change. This is expected because it severely overrejects even if there is no change in mean.

5. Proofs of the asymptotic results

5.1. Proofs of the results of Section 2

Proof of Theorem 2.1. Let $S(k) = \sum_{t=1}^k e_t$ and $S(0) = 0$. By Abel’s summation formula, we have

$$\sum_{t=1}^k u_t = \sum_{t=1}^k a(t/N)e_t$$

$$= a(k/N)S(k) - \sum_{t=1}^{k-1} S(\ell)(a((\ell + 1)/N) - a(\ell/N)), \quad 1 \leq k \leq N.$$

Since under Assumption 2.3,

$$N^{-1/2}S([Nt]) \overset{\mathcal{D}}{\rightarrow} \sigma W(t),$$

we have

$$a(k/N)S(k) - \sum_{t=1}^{k-1} S(\ell)(a((\ell + 1)/N) - a(\ell/N)) \overset{\mathcal{D}}{\rightarrow} \sigma W(k).$$

Therefore,

$$\sum_{t=1}^k u_t \overset{\mathcal{D}}{\rightarrow} \sigma W(k).$$

Table 4.3
Empirical sizes under nonlinear dependence (at 5% nominal level), $N = 512$. Models are specified in Table 4.1, top panel $\beta = 0.5$, bottom panel $\beta = 0.7$.

<table>
<thead>
<tr>
<th>Method</th>
<th>SUC</th>
<th>SCC</th>
<th>HUC</th>
<th>HCC</th>
<th>SLA</th>
<th>SCA</th>
<th>HIJA</th>
<th>HCA</th>
<th>VSU</th>
<th>VSC</th>
<th>RB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-1</td>
<td>6.8</td>
<td>7.6</td>
<td>3.9</td>
<td>5.1</td>
<td>8.0</td>
<td>7.5</td>
<td>4.1</td>
<td>5.4</td>
<td>5.9</td>
<td>5.0</td>
<td>6.9</td>
</tr>
<tr>
<td>GA-1</td>
<td>26.7</td>
<td>9.5</td>
<td>23.7</td>
<td>6.3</td>
<td>35.3</td>
<td>9.4</td>
<td>25.6</td>
<td>6.4</td>
<td>35.4</td>
<td>6.9</td>
<td>17.5</td>
</tr>
<tr>
<td>G-2</td>
<td>32.4</td>
<td>10.4</td>
<td>25.3</td>
<td>4.1</td>
<td>37.5</td>
<td>12.3</td>
<td>26.6</td>
<td>4.3</td>
<td>39.4</td>
<td>6.0</td>
<td>17.4</td>
</tr>
<tr>
<td>GA-2</td>
<td>5.9</td>
<td>6.2</td>
<td>4.9</td>
<td>4.6</td>
<td>4.9</td>
<td>5.9</td>
<td>5.4</td>
<td>5.0</td>
<td>4.7</td>
<td>5.3</td>
<td>6.3</td>
</tr>
<tr>
<td>G-3</td>
<td>26.3</td>
<td>6.0</td>
<td>23.4</td>
<td>4.6</td>
<td>29.2</td>
<td>7.4</td>
<td>26.1</td>
<td>5.4</td>
<td>37.1</td>
<td>7.1</td>
<td>14.3</td>
</tr>
<tr>
<td>GA-3</td>
<td>28.7</td>
<td>6.7</td>
<td>29.3</td>
<td>5.3</td>
<td>32.4</td>
<td>7.2</td>
<td>31.1</td>
<td>5.6</td>
<td>43.5</td>
<td>6.7</td>
<td>13.5</td>
</tr>
<tr>
<td>G-4</td>
<td>6.1</td>
<td>5.5</td>
<td>5.2</td>
<td>4.2</td>
<td>8.2</td>
<td>7.9</td>
<td>5.1</td>
<td>5.4</td>
<td>6.5</td>
<td>5.4</td>
<td>8.2</td>
</tr>
<tr>
<td>GA-4</td>
<td>61.6</td>
<td>9.9</td>
<td>52.8</td>
<td>6.1</td>
<td>67.7</td>
<td>9.9</td>
<td>56.2</td>
<td>4.5</td>
<td>75.9</td>
<td>5.7</td>
<td>55.7</td>
</tr>
<tr>
<td>G-5</td>
<td>65.4</td>
<td>14.6</td>
<td>60.3</td>
<td>6.9</td>
<td>71.1</td>
<td>15.6</td>
<td>63.4</td>
<td>5.9</td>
<td>78.6</td>
<td>8.3</td>
<td>54.5</td>
</tr>
<tr>
<td>GA-5</td>
<td>6.5</td>
<td>5.9</td>
<td>4.9</td>
<td>5.2</td>
<td>6.7</td>
<td>5.9</td>
<td>5.2</td>
<td>5.4</td>
<td>5.6</td>
<td>5.5</td>
<td>7.5</td>
</tr>
<tr>
<td>G-6</td>
<td>59.2</td>
<td>8.5</td>
<td>55.9</td>
<td>4.4</td>
<td>63.7</td>
<td>7.6</td>
<td>62.1</td>
<td>4.5</td>
<td>75.0</td>
<td>5.7</td>
<td>55.9</td>
</tr>
<tr>
<td>GA-6</td>
<td>64.3</td>
<td>9.4</td>
<td>60.7</td>
<td>5.2</td>
<td>88.8</td>
<td>8.9</td>
<td>65.1</td>
<td>5.7</td>
<td>80.0</td>
<td>8.6</td>
<td>54.8</td>
</tr>
</tbody>
</table>

Table 4.4
Models used in Table 4.B to assess empirical power: $\mu_1 = 0.51(i > 0.5N)$, $a_{13} = i/(2N)$: $a_{24} = 0.25I[0 < i \leq 0.5N] + 0.5I[0.5N < i \leq N]$. $a_{35} = 1 + 3I[i > 0.5N]$. $a_{46} = 1 - 0.75I[i > 0.5N]$.

<table>
<thead>
<tr>
<th>Model</th>
<th>$X_i$</th>
<th>$Y_i$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_i = \mu_i + \sin(\pi i/N)$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>2</td>
<td>$X_i = \mu_i + \sin(\pi i/N)$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>3</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>4</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>5</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>6</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>7</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>8</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>9</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>10</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>11</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>12</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>13</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>14</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
<tr>
<td>15</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>16</td>
<td>$X_i = \mu_i + a_{12}Y_i$</td>
<td>$Y_i = \text{AR}(1)$</td>
<td>$\mu_1 = 0.51(i &gt; 0.5N)$</td>
<td>$\alpha = 0.2$</td>
<td>$\beta = 0.7$</td>
</tr>
</tbody>
</table>
Fig. 4.1. Yields over a one year period on US treasury bills with maturities of 3 and 12 months.
Table 4.5
Empirical power under nonlinear dependence (at 5% nominal level). Models are specified in Table 4.4.

<table>
<thead>
<tr>
<th>Model</th>
<th>SUC</th>
<th>SCC</th>
<th>HUC</th>
<th>HCC</th>
<th>VSU</th>
<th>VSC</th>
<th>RB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>88.1</td>
<td>28.9</td>
<td>95.0</td>
<td>38.8</td>
<td>84.2</td>
<td>17.4</td>
<td>64.3</td>
</tr>
<tr>
<td>Model 2</td>
<td>98.2</td>
<td>31.6</td>
<td>100.0</td>
<td>34.6</td>
<td>89.3</td>
<td>17.2</td>
<td>97.9</td>
</tr>
<tr>
<td>Model 3</td>
<td>100.0</td>
<td>5.0</td>
<td>100.0</td>
<td>88.8</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 4</td>
<td>100.0</td>
<td>2.3</td>
<td>100.0</td>
<td>71.3</td>
<td>100.0</td>
<td>0.1</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 5</td>
<td>100.0</td>
<td>60.7</td>
<td>100.0</td>
<td>53.0</td>
<td>100.0</td>
<td>22.1</td>
<td>98.4</td>
</tr>
<tr>
<td>Model 6</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>99.9</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 7</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 8</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 9</td>
<td>98.9</td>
<td>65.7</td>
<td>98.8</td>
<td>48.4</td>
<td>99.1</td>
<td>27.7</td>
<td>92.3</td>
</tr>
<tr>
<td>Model 10</td>
<td>100.0</td>
<td>16.6</td>
<td>100.0</td>
<td>37.6</td>
<td>100.0</td>
<td>1.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 11</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>99.9</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 12</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 13</td>
<td>41.1</td>
<td>8.5</td>
<td>40.0</td>
<td>7.4</td>
<td>54.2</td>
<td>6.3</td>
<td>21.6</td>
</tr>
<tr>
<td>Model 14</td>
<td>14.8</td>
<td>13.9</td>
<td>13.3</td>
<td>8.3</td>
<td>12.1</td>
<td>9.7</td>
<td>15.8</td>
</tr>
<tr>
<td>Model 15</td>
<td>80.8</td>
<td>36.3</td>
<td>80.6</td>
<td>7.0</td>
<td>84.2</td>
<td>19.0</td>
<td>49.8</td>
</tr>
<tr>
<td>Model 16</td>
<td>94.3</td>
<td>61.1</td>
<td>94.0</td>
<td>6.1</td>
<td>88.2</td>
<td>41.8</td>
<td>87.7</td>
</tr>
</tbody>
</table>

N = 128

<table>
<thead>
<tr>
<th>Model</th>
<th>SUC</th>
<th>SCC</th>
<th>HUC</th>
<th>HCC</th>
<th>VSU</th>
<th>VSC</th>
<th>RB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>100.0</td>
<td>98.7</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td>90.2</td>
<td>99.2</td>
</tr>
<tr>
<td>Model 2</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>98.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 3</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 4</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 5</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 6</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 7</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 8</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 9</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>99.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 10</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 11</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 12</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Model 13</td>
<td>33.5</td>
<td>16.4</td>
<td>51.7</td>
<td>65.6</td>
<td>64.7</td>
<td>13.2</td>
<td>17.8</td>
</tr>
<tr>
<td>Model 14</td>
<td>38.9</td>
<td>38.2</td>
<td>37.3</td>
<td>80.2</td>
<td>32.9</td>
<td>31.1</td>
<td>36.2</td>
</tr>
<tr>
<td>Model 15</td>
<td>99.9</td>
<td>92.6</td>
<td>99.8</td>
<td>39.6</td>
<td>96.9</td>
<td>86.1</td>
<td>92.3</td>
</tr>
<tr>
<td>Model 16</td>
<td>100.0</td>
<td>99.9</td>
<td>100.0</td>
<td>93.1</td>
<td>100.0</td>
<td>98.6</td>
<td>99.9</td>
</tr>
</tbody>
</table>

N = 512

Table 4.6
P-values for time series of yields in Fig. 4.1.

<table>
<thead>
<tr>
<th></th>
<th>HCC</th>
<th>HCA</th>
<th>VSC</th>
<th>RB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>0.0243</td>
<td>0.0347</td>
<td>0.1072</td>
<td>0.0000</td>
</tr>
<tr>
<td>12 month</td>
<td>0.0545</td>
<td>0.0503</td>
<td>0.1073</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

by the Skorohod–Dudley–Wichura representation (cf. Shorack and Wellner, 1986, p. 47) we can define Wiener processes $W_N(t)$ such that

$$
\max_{1 \leq k \leq N} |S(k) - \sigma W_N(k)| = o_P(N^{1/2}).
$$

Hence, by Assumption 2.2,

$$
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} u_{\ell} - \sigma \left( a(k/N)W_N(k) - \sum_{\ell=1}^{k} W_N(\ell) (a((\ell + 1)/N) - a(\ell/N)) \right) \right|
\leq \max_{1 \leq k \leq N} |S(k) - \sigma W_N(k)|
+ \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} (S(\ell) - \sigma W_N(\ell))(a((\ell + 1)/N) - a(\ell/N)) \right|
= o_P(N^{1/2}) \sup_{0 \leq t \leq 1} \left| a(t) + o_P(N^{1/2}) \sum_{\ell=1}^{N-1} |a((\ell + 1)/N) - a(\ell/N)| \right|
= o_P(N^{1/2}).
$$
By the Jordan decomposition theorem (cf. Hewitt and Stromberg, 1969, p. 266), there are two nondecreasing functions such that \( a(x) = a_1(x) - a_2(x) \). Focusing on the function \( a_1 \), we can write

\[
\sum_{\ell=1}^{k-1} W_N(\ell) (a_1((\ell + 1)/N) - a_1(\ell/N)) = \sum_{\ell=1}^{k-1} W_N(\ell) \int_{\ell}^{\ell+1} a_1(x/N) \, dx = \int_0^k W_N(x)da_1(x/N) + \sum_{\ell=1}^{k-1} \int_{\ell}^{\ell+1} (W_N(\ell) - W_N(x))da_1(x/N)
\]

By the modulus of continuity of the Wiener process (cf. Csörgö and Révész, 1981, p. 29, Lemma 1.2.1), we have that

\[
\sup_{0 \leq u \leq N} \sup_{0 \leq x \leq 1} |W_N(u) - W_N(u + x)| = O_p((\log N)^{1/2}).
\]

Integration by parts gives

\[
W_N(k) a_1(k/N) - \int_0^k W_N(x)da_1(x/N) = \int_0^k a_1(x/N)dW_N(x)
\]

and therefore

\[
\left| a_1 (k/N)W_N(k) - \sum_{\ell=1}^{k} W_N(\ell) (a_1((\ell + 1)/N) - a_1(\ell/N)) - \int_0^k a_1(x/N)dW_N(x) \right| = O_p((\log N)^{1/2}).
\]

Similarly,

\[
\left| a_2 (k/N)W_N(k) - \sum_{\ell=1}^{k} W_N(\ell) (a_2((\ell + 1)/N) - a_2(\ell/N)) - \int_0^k a_2(x/N)dW_N(x) \right| = O_p((\log N)^{1/2}),
\]

resulting in

\[
\max_{1 \leq k \leq N} \left| \sum_{i=1}^{k} u_i - \sigma \int_0^k a(x/N)dW_N(x) \right| = o_p(N^{1/2}).
\]

Let

\[
U_N(t) = \int_0^t a(x/N)dW_N(x), \ 0 \leq t \leq N.
\]

It is easy to see that

\[
\{U_N(t), \ 0 \leq t \leq N\} \overset{D}{=} \left\{ W\left( \int_0^t a^2(x/N)dx \right), \ 0 \leq t \leq N \right\},
\]

where \( W(\cdot) \) is a Wiener process. Next, we note that

\[
\max_{0 \leq k \leq N-1} \sup_{0 \leq \ell \leq 1} \left| W\left( \int_0^{k+\ell} a^2(x/N)dx \right) - W\left( \int_0^{k} a^2(x/N)dx \right) \right| = O_p((\log N)^{1/2}),
\]

since

\[
\max_{0 \leq k \leq N-1} \sup_{0 \leq \ell \leq 1} \int_{k}^{k+\ell} a^2(x/N)dx \leq 4(a_1^2(1) + a_2^2(1)).
\]

Thus, we conclude that

\[
\sup_{0 \leq t \leq N} \left| N^{-1/2} \sum_{i=1}^{[Nt]} u_i - \frac{\sigma}{N^{1/2}} \int_0^{Nt} a(x/N)dW_N(x) \right| = o_p(1).
\]
Since
\[ \int_0^{tN} a(x/N) dW_N(x) = \int_0^t a(z) dW_N(Nz), \]
by the scale transformation of the Wiener process, we get that
\[ \left\{ N^{-1/2} \int_0^{tN} a(x/N) dW_N(x), 0 \leq t \leq 1 \right\} \overset{\text{D}}{=} \left\{ \int_0^t a(x) dW(x), 0 \leq t \leq 1 \right\}. \]
Computing the covariance function, one can easily verify that
\[ \left\{ \sigma \int_0^t a(x) dW(x), 0 \leq t \leq 1 \right\} \overset{\text{D}}{=} \left\{ W(b(t)), 0 \leq t \leq 1 \right\}, \]
completing the proof of Lemma 2.1. \( \square \)

**Proof of Corollary 2.1.** It follows immediately from Theorem 2.1. \( \square \)

**Proof of Corollary 2.2.** The convergence in distribution in (2.2) is an immediate consequence of the continuous mapping theorem and Corollary 2.1 (cf. Billingsley, 1968)

The result in Corollary 2.1 can be restated by an application of the Skorohod–Dudley–Wichura representation (cf. Shorack and Wellner, 1986, p. 47) that there are Gaussian processes \( \Gamma_N(t) \) distributed as \( \Gamma(t) \) for each \( N \) such that
\[
\sup_{0 \leq t \leq 1} |Z_N(t) - \Gamma_N(t)| = o_p(1). \tag{5.1}
\]
Let \( 0 < \delta < 1/2 \). We write by the Cauchy–Schwartz inequality that
\[
\left| \int_{1/N}^{1-1/N} Z^2_N(t) \frac{dt}{t(1-t)} - \int_{1/N}^{1-1/N} \frac{\Gamma^2_N(t)}{t(1-t)} dt \right|
\leq \int_{1/N}^{1-1/N} |Z_N(t) - \Gamma_N(t)| |Z_N(t)| + |\Gamma_N(t)| \frac{dt}{(1-t)^{1+2\delta}}
\leq 2 \left( \int_{1/N}^{1-1/N} \frac{Z_N^2(t)}{(1-t)^{1+2\delta}} dt \right)^{1/2}
\times \left( \int_{1/N}^{1-1/N} \frac{\Gamma_N^2(t)}{(1-t)^{1+2\delta}} dt + \int_{1/N}^{1-1/N} \frac{\Gamma_N^2(t)}{(1-t)^{1+2\delta}} dt \right)^{1/2}.
\]
It follows from (5.1) that
\[
\int_{1/N}^{1-1/N} \frac{(Z_N(t) - \Gamma_N(t))^2}{(1-t)^{1+2\delta}} dt \leq \sup_{0 \leq t \leq 1} |Z_N(t) - \Gamma_N(t)| \int_0^1 (t(1-t))^{-1+2\delta} dt = o_p(1).
\]
Next we show that
\[
\int_{1/N}^{1-1/N} \frac{Z^2_N(t)}{(1-t)^{1+2\delta}} dt = O_p(1) \quad \text{and} \quad \int_{1/N}^{1-1/N} \frac{\Gamma^2_N(t)}{(1-t)^{1+2\delta}} dt = O_p(1). \tag{5.2}
\]
We note that \( E\Gamma_N^2(t) = b(t) - 2tb(t) + t^2b(1) = (b(t) - b(1))(1 - 2t) + (1 - t)^2b(1) \) and therefore
\[
E\Gamma_N^2(t) \leq c_1 t(1-t) \quad \text{with some constant} \quad c_1 \tag{5.3}
\]
resulting in
\[
E \int_{1/N}^{1-1/N} \frac{\Gamma^2_N(t)}{(1-t)^{1+2\delta}} dt \leq c_1 \int_0^1 (t(1-t))^{-\delta} dt,
\]
which proves the second half of (5.2). Using Assumption 2.2 and \( |Ee_t| = O((|k - i| + 1)^{-k}) \), we get
\[
E \left( \sum_{i=1}^m u_i \right)^2 = \sum_{i=1}^m a(i/N) a(i/N) Ee_i e_i \leq c_2 \sum_{i=1}^m \sum_{i=1}^m |Ee_i e_i| \leq c_3 (m - k)
\]
with some constants \( c_2 \) and \( c_3 \). Hence \( E\Gamma_N^2(t) \leq c_4 t(1-t) \) for all \( 1/N \leq t \leq 1 - 1/N \) which implies immediately that
\[
E \int_{1/N}^{1-1/N} \frac{Z^2_N(t)}{(1-t)^{1+2\delta}} dt = O(1).
\]
completing the proof of the first part of (5.2) via Markov’s inequality. We obtain from (5.3) that
\[ \int_0^{1/N} \frac{\Gamma^2_N(t)}{t(1-t)} dt = o_p(1) \text{ and } \int_{1-1/N}^1 \frac{\Gamma^2_N(t)}{t(1-t)} dt = o_p(1), \]
which yields
\[ \int_{1/N}^{1-1/N} \frac{\Gamma^2_N(t)}{t(1-t)} dt \overset{p}{\to} \int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt. \]
since the distribution of \( \Gamma_N(t) \) does not depend in \( N \). □

**Proof of Theorem 2.2.** It follows from the definition of \( X_t \) that
\[ Z_N(t) = N^{-1/2} \left( \sum_{\ell=1}^{\lfloor N \rfloor} u_{\ell} - \frac{\lfloor N \rfloor}{N} \sum_{\ell=1}^N u_{\ell} \right) + N^{-1/2} d_{\lfloor N \rfloor}. \]
By *Theorem 2.1*, we have that
\[ \sup_{0 \leq t \leq 1} N^{-1/2} \left| \sum_{\ell=1}^{\lfloor N \rfloor} u_{\ell} - \frac{\lfloor N \rfloor}{N} \sum_{\ell=1}^N u_{\ell} \right| = O_p(1) \]
and by *Assumption 2.5*
\[ \sup_{0 \leq t \leq 1} N^{-1/2} \left| \sum_{\ell=1}^{\lfloor N \rfloor} u_{\ell} - \frac{\lfloor N \rfloor}{N} \sum_{\ell=1}^N u_{\ell} \right|^2 N^{-1/2} d_{\lfloor N \rfloor} = O_p(N^{1/2}). \]
Hence, (2.5) follows from the definition of \( d(u) \) and *Assumption 2.5* via the Lebesgue dominated convergence theorem (cf. *Hewitt and Stromberg*, 1969, p. 172).

Similar arguments can be used to prove (2.5) and therefore the details are omitted. □

**Proof of Theorem 2.3.** We have the usual representation
\[ \hat{\beta}_N - \beta_0 = A_N^{-1} Z_N, \]
where \( \beta_0 \) is the true parameter vector under \( H_0 \) and \( Z_N = [Z_N(1), \ldots, Z_N(p)]^\top \) with \( Z_N(j) = \sum_{\ell=1}^N x_{\ell}(j) u_{\ell}, \quad 1 \leq j \leq p. \)
*Assumption 2.2* yields that \( a(\cdot) \) is bounded and therefore by *Assumptions 2.4* and 2.6 we conclude that
\[ EZ_N(j) = \sum_{\ell=1}^N \sum_{k=1}^N E[x_{\ell}(j)x_k(j)u_{\ell}u_k] \leq \sum_{\ell=1}^N \sum_{k=1}^N |E[x_{\ell}(j)x_k(j)]||E[u_{\ell}u_k]| \]
\[ \leq O(1) \sum_{\ell=1}^N \sum_{k=1}^N (Ex_{\ell}^2(j)Ex_k^2(j))^{1/2}(|k-\ell|+1)^{-\kappa} = O(N). \]
Since \( A_N^{-1} = O_p(1/N) \), we obtain that
\[ N^{1/2} \| \hat{\beta}_N - \beta_0 \| = O_p(1). \quad (5.4) \]
Using *Assumption 2.6*, we get via the ergodic theorem that
\[ \frac{1}{N} \max_{1 \leq i \leq N} \left\| \sum_{j=1}^i x_j - \frac{i}{N} \sum_{j=1}^N x_j \right\| = o_p(1). \quad (5.5) \]
It follows from (5.4) and (5.5) that
\[ N^{-1/2} \max_{1 \leq i \leq N} \left| \sum_{j=1}^i \hat{u}_j - \frac{i}{N} \sum_{j=1}^N \hat{u}_j \right| - \left( \sum_{j=1}^i u_j - \frac{i}{N} \sum_{j=1}^N u_j \right) = o_p(1), \]
completing the proof of *Theorem 2.3*. □

**Proof of Theorem 2.4.** First, we write
\[ L_N(\theta) = \sum_{i=1}^N u_i^2 + V_N(\theta), \]
\[ V_N(\theta) = 2 \sum_{i=1}^N u_i(h(x_i, \theta) - h(x_i, \theta_0)) + \sum_{i=1}^N (h(x_i, \theta_0) - h(x_i, \theta))^2. \]
and the location of the minimum of $L_N$ and $V_N$ is the same. Using Assumption 2.8 and the ergodic, we get that

$$\sup_{\theta \in \Theta} \left| \frac{1}{N} L_N(\theta) - V(\theta) \right| = o(1) \quad \text{a.s.} \quad \text{where} \quad V(\theta) = E(h(x_0, \theta_0) - h(x_0, \theta))^2.$$  

Since $V(\theta)$ has a unique maximum at $\theta_0$, standard arguments yield (c.f. Pfanzagl, 1994) that

$$\hat{\theta}_N \rightarrow \theta_0 \quad \text{a.s.} \quad \text{where} \quad \text{Supposing \ the \ mean \ value \ theorem \ coordinatewise, we \ obtain \ that} \quad \frac{\partial}{\partial \theta} L_N(\theta_0) = \frac{\partial}{\partial \theta} L_N(\hat{\theta}_N) = G_N(\theta_0 - \hat{\theta}_N)^T$$

and

$$\frac{1}{N} G_N \rightarrow \frac{\partial^2}{\partial \theta^2} V(\theta_0) \quad \text{a.s.}$$

Following the proof of Theorem 2.3 one can verify that

$$E \left\| \frac{\partial}{\partial \theta} L_N(\theta_0) \right\|^2 = O(N)$$

and therefore

$$N^{1/2} \left\| \hat{\theta}_N - \theta_0 \right\| = O_P(1). \quad (5.7)$$

Using a two term Taylor expansion with the ergodic theorem and (5.7) we get that

$$\sum_{j=1}^{i} \hat{u}_j - \frac{i}{N} \sum_{j=1}^{N} \hat{u}_j = \sum_{j=1}^{i} u_j - \frac{i}{N} \sum_{j=1}^{N} u_j + R_{i,1} + R_{i,2}$$

with

$$R_{i,1} = \left( \sum_{j=1}^{i} \frac{\partial}{\partial \theta} h(x_j, \theta_0) - \frac{i}{N} \sum_{j=1}^{N} \frac{\partial}{\partial \theta} h(x_j, \theta_0) \right)^T (\theta_0 - \hat{\theta}_N)$$

and

$$N^{-1/2} \max_{1 \leq i \leq N} |R_{i,2}| = o_p(1).$$

Hence, the proof is complete. □

5.2. Proofs of the results in Section 3

**Lemma 5.1.** If Assumptions 2.1–2.3, 2.5, 3.1 are satisfied and \( \{e_i, -\infty < i < \infty \} \) is a stationary and ergodic sequence with $Ee_0 = 0$ and $Ee_0^2 = \sigma^2$, then we have that

$$\max_{1 \leq k \leq N} \left| H_N(k/N) - \frac{1}{N} \left( \sum_{i=1}^{k} u_i^2 + \sum_{i=1}^{k} (\mu - \hat{\mu})^2 \right) \right| = O_P(N^{-1/2} \log N) \quad (5.8)$$

and

$$\int_0^1 (H_N(u) - (b(u) + g(u)))^2 dt = o_P(1). \quad (5.9)$$
Proof. It is easy to see that

$$NH_N(k/N) = \sum_{\ell=1}^{k} u_{\ell}^2 + \sum_{\ell=1}^{k} (\mu_{\ell} - \bar{\mu})^2 + 2 \sum_{\ell=1}^{k} u_{\ell} (\mu_{\ell} - \bar{\mu}) + 2(\bar{\mu} - \bar{X}_N) \sum_{\ell=1}^{k} u_{\ell}$$

$$+ 2(\bar{\mu} - \bar{X}_N) \sum_{\ell=1}^{k} (\mu_{\ell} - \bar{\mu}) + k(\bar{\mu} - \bar{X}_N)^2.$$ 

It follows from Theorem 2.1 that

$$\bar{X}_N - \bar{\mu} = \frac{1}{N} \sum_{\ell=1}^{N} u_{\ell} = O_P(N^{-1/2})$$

and

$$\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} u_{\ell} \right| = O_P(N^{1/2})$$

and therefore

$$\max_{1 \leq k \leq N} \left| \bar{\mu} - \bar{X}_N \right| = O_P(1),$$

$$\max_{1 \leq k \leq N} \left| (\bar{\mu} - \bar{X}_N) \sum_{\ell=1}^{k} (\mu_{\ell} - \bar{\mu}) \right| = O_P(N^{1/2})$$

and

$$\max_{1 \leq k \leq N} k(\bar{\mu} - \bar{X}_N)^2 = O_P(1).$$

Using Assumptions 2.2, 2.5 and 3.1, we get that

$$E \left( \sum_{\ell=m}^{k} u_{\ell} (\mu_{\ell} - \bar{\mu}) \right)^2 \leq c_1 (k - m) \quad \text{for all } 1 \leq k \leq m \leq N$$

with some $c_1$ and therefore by Menshov’s inequality (cf. Billingsley, 1968, p. 102) that

$$\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} u_{\ell} (\mu_{\ell} - \bar{\mu}) \right| = O_P(N^{1/2} \log N).$$

Next we show that

$$\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} (u_{\ell}^2 - \sigma^2 a^2(\ell/N)) \right| = o_P(N). \quad (5.10)$$

Set

$$S(0) = 0 \quad \text{and} \quad S(k) = \sum_{\ell=1}^{k} (e_{\ell}^2 - \sigma^2).$$

By Abel’s summation formula, we have

$$\sum_{\ell=1}^{k} (u_{\ell}^2 - \sigma^2 a^2(\ell/N)) = \sum_{\ell=1}^{k} a^2(\ell/N)(S(\ell) - S(\ell - 1))$$

$$= S(N) a^2(1) - \sum_{\ell=1}^{k-1} S(\ell) (a^2(\ell/N) - a^2(\ell/N)).$$

It follows from the ergodic theorem (cf. Breiman, 1968, p. 118) that

$$\lim_{k \to \infty} \frac{1}{k} S(k) = 0 \quad \text{a.s.}$$
For any $\delta > 0$, there is a random variable $k^* = k^*(\omega)$ such that $|S(k)| \leq \delta k$ if $k \leq k^*$ and therefore for $k > k^*$

$$\frac{1}{N} \sum_{\ell=1}^{k-1} S_\ell (a^2 ((\ell + 1)/N) - a^2 (\ell/N))$$

is bounded.

$$\leq \frac{1}{N} \sum_{\ell=1}^{k} |S_\ell (a^2 ((\ell + 1)/N) - a^2 (\ell/N))| + \frac{1}{N} \sum_{\ell=k+1}^{N-1} |S_\ell| |a^2 ((\ell + 1)/N) - a^2 (\ell/N)|$$

$$= o_p(1) + \delta \sum_{\ell=1}^{N-1} |a^2 ((\ell + 1)/N) - a^2 (\ell/N)|.$$ 

It follows from Assumption 2.2 that $a^2(t)$ also has bounded variation on $[0, 1]$. Since $\delta$ can be as small as we want, the proof of (5.10) is complete. Observing that

$$\int_0^1 \left( \frac{1}{N} \sum_{\ell=1}^{N} a^2(\ell/N) - \int_0^t a^2(s) ds \right)^2 dt = o(1),$$

the proof of (5.9) is complete. \qed

**Proof of Theorem 3.1.** It is an immediate consequence of Lemma 5.1 and the definition of $\hat{C}_N(t, s)$. \qed

It follows from the definition of $\hat{y}_{N,k,\ell}$ that for $\ell \geq 0$

$$N\hat{y}_{N,k,\ell} = r_{k,\ell,1} + \cdots + r_{k,\ell,9},$$

where

$$r_{k,\ell,1} = \sum_{i=1}^{k-\ell} (X_i - \mu)(X_{i+\ell} - \mu_{i+\ell}), \quad r_{k,\ell,2} = (\bar{\mu} - \bar{X}_N) \sum_{i=1}^{k-\ell} (X_i - \mu_i)$$

$$r_{k,\ell,3} = (\bar{\mu} - \bar{X}_N) \sum_{i=1}^{k-\ell} (X_{i+\ell} - \mu_{i+\ell}), \quad r_{k,\ell,4} = (k - \ell)(\bar{\mu} - \bar{X}_N)^2$$

$$r_{k,\ell,5} = \sum_{i=1}^{k-\ell} (X_i - \mu_i)(\mu_{i+\ell} - \bar{\mu}), \quad r_{k,\ell,6} = \sum_{i=1}^{k-\ell} (X_{i+\ell} - \mu_{i+\ell})(\mu_1 - \bar{\mu})$$

$$r_{k,\ell,7} = (\bar{\mu} - \bar{X}_N) \sum_{i=1}^{k-\ell} (\mu_i - \bar{\mu}), \quad r_{k,\ell,8} = (\bar{\mu} - \bar{X}_N) \sum_{i=1}^{k-\ell} (\mu_{i+\ell} - \bar{\mu})$$

and

$$r_{k,\ell,9} = \sum_{i=1}^{k-\ell} (\mu_i - \bar{\mu})(\mu_{i+\ell} - \bar{\mu}).$$

**Lemma 5.2.** If $H_0$, Assumptions 2.1, 2.2 and 3.2-3.4 are satisfied, then we have that

$$\int_0^1 (\hat{g}_N(t) - b(t))^2 = o_p(1).$$

**Proof.** It is easy to see that

$$Er_{k,\ell,1} = \sum_{i=1}^{k-\ell} Eu_{i+\ell} = \sum_{i=1}^{k-\ell} a(i/N)a((i+\ell)/N)r(\ell), \quad \text{where } r(\ell) = Ee_\ell e_\ell,$$

resulting in

$$\frac{1}{N} \sum_{\ell=1}^{k} K(\ell/h) \frac{1}{N} r_{k,\ell,1} = \frac{1}{N} \sum_{\ell=1}^{k} K(\ell/h)r(\ell) \sum_{i=1}^{k-\ell} a(i/N)a((i+\ell)/N).$$

Let $M$ be a positive integer. We have for $k \geq M$ that

$$F_0(k) = \sum_{\ell=1}^{k} K(\ell/h)r(\ell) \frac{1}{N} \sum_{i=1}^{k-\ell} a(i/N)a((i+\ell)/N)$$
\[-\sum_{\ell=1}^{M} K(\ell/h) r(\ell) \frac{1}{N} \sum_{i=1}^{k-\ell} a(i/N) a((i+\ell)/N) \leq c_1 \sum_{\ell=M+1}^{\infty} |r(\ell)| \]

with some constant $c_1$ since by Assumptions 2.2 and 3.2 $K$ and $a$ are bounded functions. If $1 \leq k \leq M$, then

\[ \left| \sum_{\ell=1}^{k} K(\ell/h) r(\ell) \frac{1}{N} \sum_{i=1}^{k-\ell} a(i/N) a((i+\ell)/N) \right| \leq c_2 \frac{M^2}{N}. \]

Hence, we have

\[ \max_{1 \leq k \leq N} |F_N(k)| \leq c_1 \sum_{\ell=M+1}^{\infty} |r(\ell)| + c_2 \frac{M^2}{N}. \] (5.11)

Using Assumptions 2.2 and 3.2, we have

\[ \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{M} (K(\ell/h) - 1) r(\ell) \frac{1}{N} \sum_{i=1}^{k-\ell} a(i/N) a((i+\ell)/N) \right| \leq c_3 \frac{M^2}{N}. \] (5.12)

with some constant $c_3$ and

\[ \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{M} r(\ell) \frac{1}{N} \left( \sum_{i=1}^{k-\ell} a(i/N) a((i+\ell)/N) - \sum_{i=1}^{k} a(i/N) a((i+\ell)/N) \right) \right| \]

\[ \leq c_4 \left( \sum_{\ell=1}^{\infty} |r(\ell)| \right) \frac{M}{N}. \]

Next, we note that

\[ \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{M} r(\ell) \frac{1}{N} \sum_{i=1}^{k} a(i/N) (a((i+\ell)/N) - a(i/N)) \right| \]

\[ \leq c_5 \frac{1}{N} \max_{1 \leq \ell \leq N} \sum_{i=1}^{N} |a((i+\ell)/N) - a(i/N)| \]

\[ \leq c_6 \frac{M}{N}. \]

by Assumption 2.2, where $c_5$ and $c_6$ are constants and $a(u) = 0$ for $u > 1$. Finally,

\[ \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{M} r(\ell) \frac{1}{N} \sum_{i=1}^{k} a^2(i/N) - \sum_{\ell=1}^{\infty} r(\ell) \frac{1}{N} \sum_{i=1}^{k} a^2(i/N) \right| \leq c_7 \sum_{\ell=M+1}^{\infty} |r(\ell)|. \] (5.15)

with some constant $c_7$. Putting together (5.11)–(5.15), we conclude that

\[ \limsup_{N \to \infty} \left| E \sum_{\ell=1}^{k} K(\ell/h) \frac{1}{N} r_{k,\ell} - \left( \sum_{\ell=1}^{\infty} r(\ell) \right) \frac{1}{N} \sum_{i=1}^{k} a^2(i/N) \right| \leq (c_1 + c_7) \sum_{\ell=M+1}^{\infty} |r(\ell)|, \]

and since we can take $M$ as large as we want to we obtain

\[ \lim_{N \to \infty} \left| E \sum_{\ell=1}^{k} K(\ell/h) \frac{1}{N} r_{k,\ell} - \left( \sum_{\ell=1}^{\infty} r(\ell) \right) \frac{1}{N} \sum_{i=1}^{k} a^2(i/N) \right| = 0. \]

Next we show that

\[ \frac{1}{N} \sum_{k=1}^{N} \left( \sum_{\ell=1}^{k} K(h/\ell) \frac{1}{N} (r_{k,\ell} - E_{k,\ell,1}) \right)^2 = o_P(1). \] (5.16)

We observe that by the stationarity of the $e_i$'s and Assumptions 2.2 and 3.2.

\[ E \frac{1}{N} \sum_{k=1}^{N} \left( \sum_{\ell=1}^{k} K(h/\ell) \frac{1}{N} (r_{k,\ell} - E_{k,\ell,1}) \right)^2 \]
Let
\[ G_{1,k} = \{(j, \ell, \ell') : \text{ch} + 1 \leq j \leq k, 1 \leq \ell, \ell' \leq \text{ch}\} \]
\[ G_{2,k} = \{(j, \ell, \ell') : 1 \leq j \leq \text{ch}, 1 \leq \ell, \ell' \leq \text{ch}\} \]
and define
\[ Q_{1,k} = \sum_{(j,\ell,\ell') \in G_{1,k}} |E[e_0 e_\ell e_{j+\ell'} - r(\ell) r(\ell')]|, \]
\[ Q_{2,k} = \sum_{(j,\ell,\ell') \in G_{2,k}} |E[e_0 e_\ell e_{j+\ell'} - r(\ell) r(\ell')]|. \]

Next we define
\[ \tilde{e}_{j,j-\ell-1} = f(e_j, e_{j-1}, \ldots, e_{\ell+1}, e_\ell, e_{\ell+1}, \ldots) \]
and
\[ \tilde{e}_{j+\ell', j+\ell'+1} = f(e_{j+\ell'}, e_{j+\ell'+1}, \ldots, e_{\ell+1}, e_\ell, e_{\ell+1}, \ldots), \]
where \( e_{j'} \) are independent copies of \( e_0 \), independent of \( e_j \) and \( e_{j'} \). It follows from Assumption 3.4 that \( (e_0, e_\ell) \) is independent of \( (\tilde{e}_{j,j-\ell-1}, \tilde{e}_{j+\ell', j+\ell'+1-1}) \). Also, according to the construction the vectors \( (e_j, e_{j+\ell'}) \) and \( (\tilde{e}_{j,j-\ell-1}, \tilde{e}_{j+\ell', j+\ell'+1-1}) \) have the same distribution. Note that
\[ E[e_0 e_\ell \tilde{e}_{j,j-\ell-1} \tilde{e}_{j+\ell', j+\ell'+1-1}] = r(\ell) r(\ell'). \]

It follows from Assumption 3.4
\[ (E(e_j - \tilde{e}_{j,j-\ell-1})^4)^{1/4} \leq c_8 (j - \ell)^{-\alpha} \]
and
\[ (E(e_{j+\ell'} - \tilde{e}_{j+\ell', j+\ell'+1-1})^4)^{1/4} \leq c_8 (j + \ell' - \ell)^{-\alpha} \]
with some constant \( c_8 \) for all \( (j, \ell, \ell') \in G_{1,k} \). Hence the Cauchy–Schwartz inequality yields
\[ |E[e_0 e_\ell e_{j+\ell'} - \tilde{e}_{j,j-\ell-1} \tilde{e}_{j+\ell', j+\ell'+1-1}]| \]
\[ \leq (E[e_0^4 e_\ell^4] E[e_{j+\ell'}^4] (j - \ell)^{-\alpha} + (j + \ell' - \ell)^{-\alpha}) \]
\[ \leq 2c_8 (E[\tilde{e}_{j,j-\ell-1}^4])^{1/4} (j - \ell)^{-\alpha} \]
for all \( (j, \ell, \ell') \in G_{1,k} \). Thus, we get with \( c_9 = 2c_8 (E[\tilde{e}_{j,j-\ell-1}^4])^{1/4} \) that
\[ Q_{1,k} \leq c_9 \sum_{(j,\ell,\ell') \in G_{1,k}} \int_{ch+1}^{\infty} (x-y)^{-\alpha} dy dx \leq c_{11} h \]
with some constants \( c_{10} \) and \( c_{11} \).

We note that
\[ Q_{2,k} \leq \sum_{(j,\ell,\ell') \in G_{2,k}} |E[e_0 e_\ell e_{j+\ell'}]| + \sum_{(j,\ell,\ell') \in G_{2,k}} |E[e_0 e_\ell] E[e_{j+\ell'}]| \]
and
\[ \sum_{(j,\ell,\ell') \in G_{2,k}} |E[e_0 e_\ell] E[e_{j+\ell'}]| \leq ch \left( \sum_{\ell=1}^{\infty} |E[e_0 e_\ell]| \right)^2. \]
Let $e_{j,m}$ be the random variables defined in Assumption 3.4. We get for all $0 \leq s \leq t \leq v \leq 2ch$ that
\[
e_0 e_s e_v = e_0 e_{s,s}(e_t - e_{t,t-s})(e_v - e_{v,v-t}) + e_0 e_{s,s}(e_t - e_{t,t-s})e_{v,v-t}
+ e_0 e_{s,s,e_t,t-s}(e_v - e_{v,v}) + e_0 e_{s,s}(e_t - e_{t,t-s})e_v
+ e_0 (e_s - e_{s,s})e_{t,t-s}e_{v,v-t} + e_0 (e_s - e_{s,s})e_{t,t-s}(e_v - e_{v,v-t}).
\]
The definition of $e_{j,m}$ yields that $e_{v,v-t}$ is independent of $e_0 e_{s,s}(e_t - e_{t,t-s})$, $e_0$ is independent of $e_{s,s} e_{t,t-s} e_{v,v}$ and $e_{v,v-t}$ is independent of $e_0 (e_s - e_{s,s}) e_{t,t-s}$ and therefore
\[
E[e_0 e_{s,s}(e_t - e_{t,t-s})e_{v,v-t}] = 0, \quad E[e_0 e_{s,s} e_{t,t-s} e_{v,v}] = 0
\]
and $E[e_0 (e_s - e_{s,s}) e_{t,t-s} e_{v,v-t}] = 0$.

Using Assumption 3.4, we obtain that
\[
\sum_{1 \leq s \leq t \leq v \leq 2ch} E|e_0 e_{s,s}(e_t - e_{t,t-s})(e_v - e_{v,v-t})| \leq (Ee_0^4)^{1/2} \sum_{1 \leq s \leq t \leq v \leq 2ch} (E(e_t - e_{t,t-s})^4)^{1/4}(E(e_v - e_{v,v-t})^4)^{1/4} = O(h).
\]

Similarly,
\[
\sum_{1 \leq s \leq t \leq v \leq 2ch} E|e_0 (e_s - e_{s,s})(e_t - e_{t,t-s})e_v| = O(h)
\]
and
\[
\sum_{1 \leq s \leq t \leq v \leq 2ch} E|e_0 (e_s - e_{s,s})e_{t,t-s}(e_v - e_{v,v-t})| = O(h)
\]
and
\[
\sum_{1 \leq s \leq t \leq v \leq 2ch} E|e_0 e_{s,s} e_{t,t-s}(e_v - e_{v,v})| = O(1) \int_0^{2ch} \int_s^{2ch} \int_t^{2ch} v^{-\alpha} dv dt ds = O(h).
\]

Thus we conclude
\[
Q_{2,k} \leq C_2 h,
\]
which completes the proof of
\[
E \frac{1}{N} \sum_{k=1}^N \left( \sum_{i=1}^k K(h/i) \frac{1}{N}(r_{k,i,1} - Er_{k,i,1}) \right)^2 = O(h/N)
\]
and therefore (5.16) follows via Markov’s inequality.

Theorem 2.1 yields that
\[
\max_{1 \leq k \leq N} \left| \sum_{i=1}^{k-1} K(h/i) \frac{1}{N} r_{k,i,2} \right| \leq \frac{ch}{N^2} \sum_{i=1}^N u_i \max_{|u| \leq 1} |K(u)| \max_{1 \leq k \leq N} \sum_{i=1}^k u_i = O_p(h/N).
\]

Similar argument gives
\[
\max_{1 \leq k \leq N} \left| \sum_{i=1}^{k-1} K(h/i) \frac{1}{N} r_{k,i,3} \right| = O_p(h/N).
\]

Theorem 2.1 and Assumption 3.2 yield
\[
\max_{1 \leq k \leq N} \left| \sum_{i=1}^k K(h/N) \frac{1}{N} r_{k,i,4} \right| = O_p(h/N).
\]
\[
\leq \frac{1}{N^2} \left( \sum_{i=1}^{N} u_i \right)^2 \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} K(\ell/k) (k-\ell) \right| = O_P(h/N).
\]

Since \( r_{k,\ell} = r_{k,\ell} = r_{k,\ell} = r_{k,\ell} = 0 \) under \( H_0 \) we proved that
\[
\int_0^1 \left( \sum_{\ell=1}^{[Nu]-1} K(\ell/h) \hat{\gamma}_{Nu,\ell} - \sum_{\ell=-\infty}^{0} r(\ell) \int_0^u a^2(v) dv \right) du = O_p(1). \tag{5.17}
\]

Similar arguments show that
\[
\int_0^1 \left( \sum_{\ell=-([Nu]-1)}^{0} K(\ell/h) \hat{\gamma}_{Nu,\ell} - \sum_{\ell=-\infty}^{0} r(\ell) \int_0^u a^2(v) dv \right) du = O_p(1). \tag{5.18}
\]

Now Lemma 5.2 follows from (5.17) and (5.18). \( \square \)

**Lemma 5.3.** If \( H_0 \), Assumptions 2.1, 2.2 and 3.2–3.4 are satisfied, then we have
\[
\int_0^1 \hat{g}^2_N(u) du = O_P(h^2).
\]

If in addition, Assumption
\[
\int_0^1 \left( \frac{1}{\hat{r}} \hat{g}_N(u) - g(u) \int_{-c}^c K(v) dv \right)^2 du = O_p(1).
\]

**Proof.** Following the proof of Theorem 2.1 one can show that
\[
\max_{1 \leq k \leq N} \max_{1 \leq \ell \leq k} \left| \sum_{i=1}^{k-\ell} (X_i - \mu_i)(\mu_{i+\ell} - \tilde{\mu}) \right| = \max_{1 \leq k \leq N} \max_{1 \leq \ell \leq k} \left| \sum_{i=1}^{k-\ell} u_i (\mu_{i+\ell} - \tilde{\mu}) \right| = O_p(N^{1/2}),
\]
and therefore
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} K(\ell/k) \frac{1}{N} r_{k,\ell,5} \right| = O_p(h/N^{1/2}).
\]

Similarly,
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} K(\ell/k) \frac{1}{N} r_{k,\ell,6} \right| = O_p(h/N^{1/2}).
\]

Combining Theorem 2.1 and Assumption 2.5, we conclude
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} K(\ell/h) \frac{1}{N} r_{k,\ell,5} \right| \leq \frac{1}{N^3} \left( \sum_{i=1}^{N} u_i \right)^2 \max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} K(\ell/h) \sum_{j=1}^{k-\ell} (\mu_i - \tilde{\mu}) \right| = O_p(h/N)
\]
and
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k} K(\ell/h) \frac{1}{N} r_{k,\ell,8} \right| = O_p(h/N)
\]

We note that by Assumptions 2.5 and 3.2
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k-1} K(\ell/h) \frac{1}{N} \sum_{i=1}^{k-\ell} (\mu_i - \tilde{\mu})(\mu_{i+\ell} - \tilde{\mu}) \right| = O(h),
\]
and therefore by the proof of Lemma 5.2 we get
\[
\int_0^1 \left( \sum_{\ell=1}^{[Nu]-1} K(\ell/h) \hat{\gamma}_{Nu,\ell} \right)^2 du = O_p(h^2). \tag{5.19}
\]
Similarly, 
\[
\int_0^1 \left( \sum_{\ell=1}^{[\nu u]} K(\ell/h) \hat{\gamma}_{N,\ell} \right)^2 = O_p(h^2),
\]
(5.20)
and now the first part of Lemma 5.3 follows from (5.19) and (5.20).

Using Assumption 3.5, we conclude
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k-1} K(\ell/h) \frac{1}{N} \sum_{i=1}^{k-\ell} (\mu_i - \bar{\mu})(\mu_{i+\ell} - \bar{\mu}) \right| = o(h),
\]
and by Assumption 2.5
\[
\max_{1 \leq k \leq N} \left| \sum_{\ell=1}^{k-1} K(\ell/h) \frac{1}{N} \sum_{i=1}^{k-\ell+1} (\mu_i - \bar{\mu}) \right| = o(1).
\]

It follows from Assumption 3.2 that for all 0 < u < 1,
\[
\int_0^1 K(\ell/h) \to \int_0^c K(u)du.
\]

Thus we conclude
\[
\int_0^1 \left( \sum_{\ell=1}^{[\nu u]^{-1}} K(\ell/h) \frac{1}{N} \sum_{i=1}^{[\nu u]^{-1} - \ell} (\mu_i - \bar{\mu})(\mu_{i+\ell} - \bar{\mu}) - g(u) \int_0^c K(v)dv \right)^2 du = o(1),
\]
and therefore we can replace (5.19) and (5.20) with the more precise
\[
\int_0^1 \left( \sum_{\ell=1}^{[\nu u]^{-1}} K(\ell/h) \hat{\gamma}_{N,\ell} - g(u) \int_0^c K(v)dv \right)^2 du = o_P(1)
\]
and
\[
\int_0^1 \left( \sum_{\ell=1}^{0} K(\ell/h) \hat{\gamma}_{N,\ell} - g(u) \int_{-c}^0 K(v)dv \right)^2 du = o_P(1).
\]

This completes the proof of Lemma 5.3. □

Proof of Theorem 3.2. The result in (3.9) follows from Lemma 5.2 while 3.10 and 3.11 are immediate consequences of Lemma 5.3. □

References


