Nonparametric Specification Testing for Nonlinear Time Series with Nonstationarity

Jiti Gao, Maxwell King, Zudi Lu and Dag Tjostheim
Nonparametric Specification Testing for Nonlinear Time Series with Nonstationarity

By Jiti Gao, Maxwell King, Zudi Lu, and Dag Tjøstheim

The University of Adelaide, Monash University, Curtin University of Technology
and The University of Bergen

Abstract

This paper considers a nonparametric time series regression model with a nonstationary regressor. We construct a nonparametric test for testing whether the regression is of a known parametric form indexed by a vector of unknown parameters. We establish the asymptotic distribution of the proposed test statistic. Both the setting and the results differ from earlier work on nonparametric time series regression with stationarity. In addition, we develop a bootstrap simulation scheme for the selection of suitable bandwidth parameters involved in the kernel test as well as the choice of simulated critical values. An example of implementation is given to show that the proposed test works in practice.

Key words: Integrated regressor, kernel test, nonparametric regression, nonstationary time series, random walk.

Abbreviated Title: Testing for Stationarity in Time Series.

---

1Jiti Gao is from the School of Economics, The University of Adelaide. Adelaide SA 5005, Australia. Email: jiti.gao@adelaide.edu.au.
1. **Introduction.** During the past two decades or so, there has been much interest in both theoretical and empirical analysis of long-run economic and financial time series data. Models and methods used have been based initially on parametric linear autoregressive moving average representations (Granger and Newbold 1977; Brockwell and Davis 1990; and many others) and then on parametric nonlinear time series models (see e.g. Tong 1990; Granger and Teräsvirta 1993; Fan and Yao 2003). Such parametric linear or nonlinear models, as already pointed out in existing studies, may be too restrictive in some cases. This leads to various nonparametric and semiparametric techniques being used to model nonlinear time series data with the focus of attention being on the case where the observed time series satisfies a type of stationarity. Both estimation and specification testing has been systematically examined in this situation (Robinson 1988, 1989; Masry and Tjøstheim 1995, 1997; Li and Wang 1998; Li 1999; Fan and Yao 2003; Gao 2007; Li and Racine 2007 and others).


In the field of model specification with nonstationarity, there are some existing studies (see, for example, Hong and Phillips 2005, Kasparis 2005, 2007 and Marmer 2008). All the cited papers consider specification testing in time series regression with unit-roots. The first two papers consider model specification testing in a cointegration setting, while the third paper discusses the applicability of the Bierens test in a class of nonlinear and nonstationary models and establishes some corresponding results. The last paper develops a functional form test in dealing with nonlinearity, nonstationarity and spurious forecasts. In the original version of this paper, the authors also propose using a nonparametric kernel test for nonstationarity in an autoregressive model.

In this paper, we are interested in considering a nonlinear time series of the form
\[
Y_t = m(X_t) + \sigma_0 e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \quad t = 1, 2, \ldots, T, \tag{1.1}
\]
where \(m(\cdot)\) is an unknown function defined over \(\mathbb{R}^1 = (-\infty, \infty)\), \(\sigma_0 > 0\) is an unknown parameter, \(\{u_t\}\) is a sequence of independent and identically distributed (i.i.d.) errors, and \(\{e_t\}\) is a sequence of martingale differences. We are then interested in testing the following null hypothesis:
\[
H_0 : P(m(X_t) = m_{\theta_0}(X_t)) = 1 \quad \text{for all} \quad t \geq 1, \tag{1.2}
\]
where \(m_{\theta_0}(x)\) is a known parametric function of \(x\) indexed by a vector of unknown parameters, \(\theta_0 \in \Theta\). Under \(H_0\), model (1.1) becomes a nonlinear parametric model of the form
\[
Y_t = m_{\theta_0}(X_t) + \sigma_0 e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \quad t = 1, 2, \ldots, T. \tag{1.3}
\]
Park and Phillips (2001) extensively discuss some estimation problems for this kind of parametric nonlinear time series model.

To the best of our knowledge, the problem of testing (1.2) for the case where \(\{X_t\}\) is nonstationary has not been discussed. This paper attempts to derive a simple kernel test for this kind of parametric specification of the conditional mean function when the regressors are integrated. In summary, the main contributions of this paper are:

(i) It proposes an asymptotically normal test procedure for model (1.2). Theoretical properties for the proposed test procedure are established.
In order to implement the proposed test in practice, we develop a new simulation procedure based on the assessment of both the size and power of the proposed test.

The rest of the paper is organised as follows. Section 2 establishes a nonparametric kernel test procedure as well as its asymptotic distribution. A bootstrap simulation scheme is proposed in Section 3. Section 4 shows how to implement the proposed test in practice. Section 5 concludes the paper with some remarks on extensions. Mathematical details are given in Appendix A. Additional details are available from Appendices B–D.

2. Nonparametric kernel test. Consider a test problem of the form:

$$H_0 : P(m(X_t) = m_{\theta_0}(X_t)) = 1 \text{ versus } H_1 : P(m(X_t) = m_{\theta_0}(X_t) + \Delta_T(X_t, \theta_1)) = 1$$ (2.1)

for all $t \geq 1$ and some $\theta_1 \in \Theta$ (a parameter space), where $\theta_0 \in \Theta$ denotes the true value of $\theta$ if $H_0$ is true, and $\Delta_T(\cdot, \theta_1)$ is a sequence of semiparametrically unknown functions to ensure that model (1.1) is a semiparametric time series model under $H_1$.

To construct a nonparametric kernel test, the main idea is to compare a parametric estimator of $m(\cdot)$ under $H_0$ with a nonparametric kernel estimator. In order to avoid introducing biases associated with nonparametric kernel estimation (Gao and King 2004), we use a smoothed version of the parametric estimator in the construction.

Similarly to existing studies for the stationary time series case (see, for example, Chapter 3 of Gao 2007), we propose using a kernel–based test of the form

$$M_T = M_T(h) = \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \hat{\epsilon}_s K_h(X_t - X_s) \hat{\epsilon}_t,$$ (2.2)

where $K_h(\cdot) = K(\cdot/h)$ with $K(\cdot)$ being a probability kernel function, $h$ is a bandwidth parameter and $\hat{\epsilon}_t = Y_t - m_{\hat{\theta}}(X_t)$, in which $\hat{\theta}$ is a consistent estimator of $\theta_0$ under $H_0$. In this paper, we consider $\hat{\theta}$ as the nonlinear least squares estimator of $\theta_0$ as defined in Park and Phillips (2001).

In order to establish an asymptotic distribution for $M_T$, we need to introduce the following assumption.

**Assumption 2.1.** (i) The sequence $\{u_t = X_t - X_{t-1}\}$ is a sequence of independent and identically distributed random errors with $E[u_t] = 0$, $E[u_t^2] = \sigma_u^2$ and $\mu_4 = E[u_t^4] < \infty$. The marginal density function of $\{u_t\}$ is symmetric. The characteristic function $\psi(\cdot)$ of $\{u_t\}$ satisfies $\int_{-\infty}^{\infty} |\psi(v)|dv < \infty$. 

4
(ii) The sequence \( \{e_t\} \) is a sequence of martingale differences satisfying \( E[e_t|\mathbb{B}_{t-1}] = 0 \), \( E[e_t^2|\mathbb{B}_{t-1}] = 1 \) a.s., \( E[e_t^3|\mathbb{B}_{t-1}] = 0 \) a.s. and \( 0 < \nu_4 = E[e_t^4|\mathbb{B}_{t-1}] < \infty \) a.s., where \( \mathbb{B}_{t-1} = \sigma\{e_s: 1 \leq s \leq t - 1\} \) is the \( \sigma \)-field generated by \( \{e_s: 1 \leq s \leq t - 1\} \).

(iii) The sequences \( \{u_s: s \geq 1\} \) and \( \{e_t: t \geq 1\} \) are mutually independent.

(iv) The function \( K(\cdot) \) is a symmetric and bounded probability density with compact support \( C(K) \). In addition, \( |K(x + y) - K(x)| \leq \Psi(x)|y| \) for all \( x \in C(K) \) and any given \( y \), where \( \Psi(x) \) is a non-negative bounded function for all \( x \in C(K) \).

Let \( Q_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (Y_t - m_\theta(X_t))^2 \). Define the nonlinear least squares estimator of \( \theta_0 \) as the minimizer of \( Q_T(\theta) \) over \( \theta \in \Theta: \hat{\theta} = \arg \min_{\theta \in \Theta} Q_T(\theta) \).

**Assumption 2.2.** (i) There are suitable unknown parameters \( \theta_0 \) and \( \sigma_0 > 0 \) such that model (1.3) under \( H_0 \) is the true identifiable model.

(ii) \( \lim_{T \to \infty} h = 0 \) and \( \limsup_{T \to \infty} T^{\frac{1}{2} - \delta_0} h = \infty \) for some \( 0 < \delta_0 < \frac{1}{5} \).

(iii) The function \( m_\theta(x) \) is differentiable with respect to \( \theta \) for each fixed \( x \). In addition, under \( H_i \) (\( i = 0, 1 \)) the following equations hold in probability: for \( 0 < \delta_0 < \frac{1}{5} \) (\( \tau \) denotes the transposed)

\[
\lim_{T \to \infty} \frac{R_{ij}(T)}{\sqrt{T^{\frac{1}{2} - 2\delta_0} h}} \left( (\hat{\theta} - \theta) \right)^\tau (\hat{\theta} - \theta_i) = 0, \tag{2.3}
\]

where for \( i = 0, 1; j = 1, 2 \), \( R_{ij}(T) = \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{T-s}} r_{ij}(s) \) with

\[
r_{ij}(s) = \int \left\{ \left( \frac{\partial m_\theta(x)}{\partial \theta} \right)^\tau \left( \frac{\partial m_\theta(x)}{\partial \theta} \right) \right\}^j \phi \left( \frac{x}{\sqrt{s}} \right) dx,
\]

in which \( \phi(\cdot) \) is the density function of the normal random variable \( N(0,1) \).

**Remark 2.1.** (i) Assumption 2.1(i) requires \( \{u_t\} \) to be independent and identically distributed in order to ensure that \( S_t = \sum_{i=1}^{t} X_i \) have independent increments for all \( t \geq 1 \). The last sentence of Assumption 2.1(i) imposes a mild condition on the characteristic function, and it holds in many cases. The condition \( \int_{-\infty}^{\infty} |\psi(v)|dv < \infty \) is to ensure certain convergence results. Let \( \phi_T(x) \) be the density function of \( \frac{1}{\sqrt{T\sigma}} \sum_{t=1}^{T} u_t \). Then Assumption 2.1(i) implies \( \sup_x |\phi_T(x) - \phi(x)| \to 0 \) as \( T \to \infty \), where \( \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \)

is the density function of the standard normal random variable \( N(0,1) \). The proof is standard (see, for example, Chapters 8 and 9 of Chow and Teicher 1988).

Assumption 2.1(ii) is quite standard in this kind of problem (see, for example, Assumption 2.1 of Park and Phillips 2001). Obviously, Assumption 2.1(ii) covers the case
where \( \{e_t\} \) is a sequence of independent and identically distributed errors. Assumption 2.1(iii) imposes the independence between \( \{e_s\} \) and \( \{u_t\} \) for all \( s, t \geq 1 \). Such an independence assumption is somewhat restrictive but may not be too unreasonable in this kind of nonstationary problem. Assumption 2.1(iv) is also quite standard in this kind of nonstationary situation.

(ii) Assumption 2.2(i) is to ensure that the true model (1.3) under \( H_0 \) is identifiable. Assumption 2.2(ii) imposes some minimum conditions on the bandwidth. Assumption 2.2(iii) imposes some technical conditions involving both the form of \( m_{\theta_0}(\cdot) \) and the rate of convergence of \( \hat{\theta} \) to \( \theta_0 \). For example, when \( m_{\theta_0}(x) = \alpha_0 + \beta_0 x \) and the rate of convergence of \( \hat{\theta} \) to \( \theta_0 \) is of \( o_p \left( T^{\frac{3}{8} + \frac{9}{2} h^{\frac{1}{2}}} \right) \), Assumption 2.2(iii) holds with \( i = 0 \). In the case where \( m_{\theta_1}(x) = \alpha_1 + \beta_1 x + \gamma_1 x (1 - \exp(-\lambda_1 x^2)) \) and the rate of convergence of \( \hat{\theta} \) to \( \theta_1 \) is of \( o_p \left( T^{\frac{3}{8} + \frac{9}{2} h^{\frac{1}{2}}} \right) \), Assumption 2.2(iii) holds with \( i = 1 \).

We state the main theorem of this section; its proof is given in Appendix A.

**Theorem 2.1.** Consider model (1.1). Suppose that Assumptions 2.1–2.2 hold with \( i = 0 \) in Assumption 2.2(iii). Then under \( H_0 \)

\[
\hat{L}_T = \hat{L}_T(h) = \frac{M_T(h)}{\hat{\sigma}_T} \rightarrow_D N(0, 1) \quad \text{as} \quad T \rightarrow \infty,
\]

where \( \hat{\sigma}_T^2 = 2 \sum_{t=1}^{T} \sum_{s=1,s\neq t}^{T} \hat{\epsilon}_s^2 K_h^2(X_s - X_t) \hat{\epsilon}_t^2 \).

As shown in Theorem 2.1, \( \hat{L}_T(h) \) converges in distribution to standard normality as \( T \rightarrow \infty \). Existing studies for the stationary time series case show that the finite sample performance of the size function of this type of nonparametric kernel based test is not very good when using a normal distribution to approximate the exact finite–sample distribution of the test under consideration. In order to improve the finite sample performance of \( \hat{L}_T(h) \), we propose using a bootstrap simulation method. Such a method is known to work quite well in the stationary case. For each given bandwidth satisfying certain theoretical conditions, instead of using an asymptotic value of \( l_{0.05} = 1.645 \) at the 5\% level for example, we use a simulated critical value for computing the size function and then the power function. An optimal bandwidth is chosen such that the power function is maximized at the optimal bandwidth. Our finite–sample studies show that there is little size distortion when using such a simulated critical value. Such issues are discussed in detail in Section 3 below.
3. Bootstrap simulation scheme. This section discusses how to simulate a critical value for the implementation of \( \hat{L}_T(h) \) in each case. We then examine its finite sample performance through using one example in Section 4 below.

Before we look at how to implement \( \hat{L}_T(h) \) in practice, we propose the following simulation scheme.

**Simulation Scheme 3.1:** The exact \( \alpha \)-level critical value, \( l_\alpha(h) \) (\( 0 < \alpha < 1 \)) is the \( 1 - \alpha \) quantile of the exact finite-sample distribution of \( \hat{L}_T(h) \). Because there are unknown quantities, such as parameters and functions, we cannot evaluate \( l_\alpha(h) \) in practice. We therefore suggest choosing an approximate \( \alpha \)-level critical value, \( \hat{l}_\alpha(h) \), by using the following simulation procedure:

(i) For each \( t = 1, 2, \ldots, T \), generate \( Y_t^* = m_\theta(X_t) + \hat{\sigma}_0 \epsilon_t^* \), where the original sample \( (X_1, \ldots, X_T) \) acts in the resampling as a fixed design, \( \{\epsilon_t^*\} \) is sampled independently either from a pre-specified distribution or using a nonparametric bootstrap method, and \( \hat{\sigma}_0 \) is an initial consistent estimator of \( \sigma_0 \), and \( \hat{\theta} \) is the nonlinear least squares estimator of \( \theta_0 \) based on the original sample.

(ii) Use the data set \( \{(Y_t^*, X_t) : t = 1, 2, \ldots, T\} \) to re-estimate \( (\theta_0, \sigma_0) \). Denote the resulting estimate by \( (\hat{\theta}^*, \hat{\sigma}^*) \). Compute the statistic \( \hat{L}_T^*(h) \) that is the corresponding version of \( \hat{L}_T(h) \) by replacing \( (\hat{\theta}, \hat{\sigma}) \) and \( \{(Y_t, X_t) : 1 \leq t \leq T\} \) with \( (\hat{\theta}^*, \hat{\sigma}^*) \) and \( \{(Y_t^*, X_t) : 1 \leq t \leq T\} \) on the right-hand side of \( \hat{L}_T(h) \).

(iii) Repeat the above steps \( M \) times and produce \( M \) versions of \( \hat{L}_T^*(h) \) denoted by \( \hat{L}_{Tm}^*(h) \) for \( m = 1, 2, \ldots, M \). Use the \( M \) values of \( \hat{L}_{Tm}^*(h) \) to construct their empirical bootstrap distribution function. The bootstrap distribution of \( \hat{L}_T^*(h) \) given \( W_T = \{(X_t, Y_t) : 1 \leq t \leq T\} \) is defined by \( P^* \left( \hat{L}_T^*(h) \leq x \right) = P \left( \hat{L}_T^*(h) \leq x | W_T \right) \). Let \( l_{\alpha}^*(h) \) satisfy \( P^* \left( \hat{L}_T^*(h) \geq l_{\alpha}^*(h) \right) = \alpha \) and then estimate \( l_\alpha(h) \) by \( \hat{l}_\alpha(h) \).

(iv) Define the size and power functions by

\[
\alpha(h) = P \left( \hat{L}_T(h) \geq l_{\alpha}(h) | H_0 \right) \quad \text{and} \quad \beta(h) = P \left( \hat{L}_T(h) \geq l_{\alpha}(h) | H_1 \right).
\]

Let \( \mathcal{H} = \{h : \alpha - \varepsilon \leq \alpha(h) \leq \alpha + \varepsilon\} \) for \( 0 < \varepsilon \leq \frac{\alpha}{10} \). Choose an optimal bandwidth \( \hat{h}_{\text{test}} \) such that

\[
\hat{h}_{\text{test}} = \arg\max_{h \in \mathcal{H}} \beta(h). \quad (3.1)
\]

We use \( l_{\alpha}^*(\hat{h}_{\text{test}}) \) when computing the size and power values of \( \hat{L}_T(\hat{h}_{\text{test}}) \) in each case.
To study the power function of \( \hat{L}_T(h) \), we specify the form of alternatives as follows:

\[
H_1 : \quad P(m(X_t) = m_{\theta_1}(X_t) + \Delta_T(X_t, \theta_1)) = 1, \quad (3.2)
\]

where \( \Delta_T(x, \theta_1) \) is a sequence of semiparametrically unknown functions satisfying certain conditions in Assumption 3.2 below. Under \( H_1 \), model (1.1) becomes

\[
Y_t = m(X_t) + \epsilon_t = m_{\theta_1}(X_t) + \Delta_T(X_t, \theta_1) + \epsilon_t, \quad (3.3)
\]

where \( \Delta_T(x, \theta_1) \) can be consistently estimated by \( \hat{\Delta}_T(x, \hat{\theta}_1) \), in which \( \hat{\theta}_1 \) minimizes

\[
\sum_{t=1}^{T} (Y_t - m_{\theta_1}(X_t) - \hat{\Delta}_T(X_t, \theta_1))^2, \quad (3.4)
\]

and \( \hat{\Delta}_T(x, \theta_1) = \frac{\sum_{t=1}^{T} K_{bcv}(x_t-x)(Y_t-m_{\theta_1}(X_t))}{\sum_{t=1}^{T} K_{bcv}(x_t-x)} \) with \( \hat{b}_{cv} \) being chosen by a conventional cross-validation estimation method.

In addition to Assumption 2.1, we need the following conditions.

**Assumption 3.1.** (i) There are consistent estimators \( \hat{\sigma}^* \) and \( \hat{\sigma} \) such that as \( T \to \infty \)

\[
\hat{\sigma} - \sigma_0 \to_P 0 \quad \text{and} \quad \hat{\sigma}^* - \tilde{\sigma} \to_P 0.
\]

(ii) Let \( H_0 \) be true. Then the following equation holds in probability: for \( 0 < \delta_0 < \frac{1}{5} \)

\[
\lim_{T \to \infty} \frac{\hat{R}_j(T)h}{\sqrt{(T^{3/2-2\delta_0}h)}} (\hat{\theta}^* - \hat{\theta})^T (\hat{\theta}^* - \hat{\theta})^j = 0, \quad (3.5)
\]

where for \( j = 1, 2 \), \( \hat{R}_j(T) = \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \hat{r}_j(s) \) with

\[
\hat{r}_j(s) = \int \left\{ \left( \frac{\partial m_{\hat{\theta}}(x)}{\partial \theta} \right)^T \left( \frac{\partial m_{\hat{\theta}}(x)}{\partial \theta} \right) \right\}^j \phi \left( \frac{x}{\sqrt{s}} \right) dx.
\]

**Assumption 3.2.** Let \( H_1 \) be true. Suppose that Assumption 2.2(iii) holds with \( i = 1 \). In addition, the following equation holds in probability: for \( 0 < \delta_0 < \frac{1}{5} \)

\[
\lim_{T \to \infty} \frac{D(T)\sqrt{h}}{T^{3/2-\delta_0}} = \infty, \quad (3.6)
\]
where \( D(T) = \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} C_T(s) \) with \( C_T(s) = \int \Delta^2_T(x, \theta_1) \phi \left( \frac{x}{\sqrt{s}} \right) dx \).

Assumption 3.1(i) imposes only mild consistency conditions on \( \hat{\sigma}^* \) and \( \hat{\sigma} \) to ensure that the bootstrap critical value \( t^*_\alpha(h) \) is asymptotically correct \( \alpha \)-level critical value under any model in \( H_0 \). Similarly to Corollary 4.4 of Park and Phillips (2001), one may impose conditions on the local integrability or the integrability of \( m_\theta(\cdot) \) to ensure that Assumption 3.1(i) holds. Assumption 3.1(ii) corresponds to Assumption 2.2(iii) with \( i = 0 \). Similarly to Remark 2.1(iii), it can be verified that Assumption 3.1(ii) holds when \( m_\theta(x) \) belongs to a class of parametric functions.

Assumption 3.2 requires that the 'distance' between \( H_0 \) and \( H_1 \) is large enough to ensure that the test is consistent under \( H_1 \). Similarly to Assumption 2.2(iii), Assumption 3.2 involves both the form of \( m_{\theta_1}(x) \) under \( H_1 \) and the rate of convergence of \( \hat{\theta} \) to \( \theta_1 \) when the form of \( m(x) \) is chosen as \( m(x) = m_{\theta_1}(x) + \Delta_T(x, \theta_1) \). In both theory and practice, various forms may be considered for \( m_{\theta_1}(\cdot) \) and \( m(\cdot) \). For example, we consider the following forms:

\[
H_0: \quad m_{\theta_0}(x) = \alpha_0 + \beta_0 x \quad \text{versus} \quad H_1: \quad m(x) = m_{\theta_1}(x) + \Delta_T(x, \theta_1) = \alpha_1 + \beta_1 x + \gamma_1 x^2, \tag{3.7}
\]

where \( \theta_0 = (\alpha_0, \beta_0) \) is estimated by \( \hat{\theta} \), and \(-\infty < \alpha_1, \beta_1, \gamma_1 < \infty \) are unknown parameters. In this case, in order to verify Assumption 3.2, it suffices to show that as \( T \to \infty \)

\[
\frac{E \left[ \sum_{t=2}^{T} \sum_{s=1}^{t-1} X^2_s K_h(X_t - X_s) X^2_t \right]}{T^{2-\delta_0}} \to \infty, \tag{3.8}
\]

which follows from (letting \( X_{st} = X_t - X_s \))

\[
\sum_{t=2}^{T} \sum_{s=1}^{t-1} E \left[ X^2_s K \left( \frac{X_t - X_s}{h} \right) X^2_t \right] = \sum_{t=2}^{T} \sum_{s=1}^{t-1} E \left[ X^2_s K \left( \frac{X_t - X_s}{h} \right) (X_s + X_t - X_s)^2 \right] 
\]

\[
= \sum_{t=2}^{T} \sum_{s=1}^{t-1} \int \int x^2_s K \left( \frac{x_{st}}{h} \right) (x_s + x_{st})^2 f_s(x_s) f_{st}(x_{st}) dx_s dx_{st} 
\]

letting \( y_s = x_s \) and \( y_{st} = \frac{x_{st}}{h} \)

\[
= h \sum_{t=2}^{T} \sum_{s=1}^{t-1} \int \int y^2_s K(y_s) (y_s + y_{st}h)^2 f_s(y_s) f_{st}(y_{st}h) dy_s dy_{st} 
\]

\[
= h(1 + o(1)) \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \int \int x^4 K(y) g_s \left( \frac{x}{\sqrt{s}} \right) g_{st} \left( \frac{y h}{\sqrt{t-s}} \right) dx dy 
\]

\[
= \phi(0) h(1 + o(1)) \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \left( \int x^4 \phi \left( \frac{x}{\sqrt{s}} \right) dx \right) \int K(y) dy 
\]

9
\[ = \phi(0) \ h(1 + o(1)) \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{s \sqrt{t-s}} \int x^4 \phi \left( \frac{x}{\sqrt{s}} \right) dx = C \ T^{\frac{7}{2}} \ h(1 + o(1)), \quad (3.9) \]

using the normal distribution approximation method as outlined in the proof of Lemma A.1 in Appendix A below, where \( f_s(\cdot) \) denotes the density function of \( X_s \) and \( f_{st}(\cdot) \) denotes the density function of \( X_t - X_s \), and \( g_s(\cdot) \) denotes the density function of \( \frac{X_s}{\sqrt{s}} \) and \( g_{st}(\cdot) \) denotes the density function of \( \frac{X_t - X_s}{\sqrt{t-s}} \). This shows that Assumption 3.2 holds.

In general, we may consider testing various classes of parametric functions under \( H_0 \) against nonparametric and/or semiparametric alternatives under \( H_1 \). This is both theoretically justifiable and practically implementable, because, as demonstrated by Park and Phillips (2001, Theorems 5.1 and 5.2), the rate of convergence for one class can be different from that for another class.

We state the following results of this section.

**Theorem 3.1.** (i) Assume that the conditions of Theorem 2.1 hold. In addition, if Assumption 3.1 holds, then under \( H_0 \), we have
\[ \lim_{T \to \infty} P(\hat{L}_T(h) \geq l^*_\alpha(h)) = \alpha. \]

(ii) Assume that the conditions of Theorem 2.1 hold. In addition, if Assumptions 3.1 and 3.2 hold, then under \( H_1 \), we have
\[ \lim_{T \to \infty} P(\hat{L}_T(h) \geq l^*_\alpha(h)) = 1. \]

The proof of Theorem 3.1 is given in Appendix A. Theorem 3.1(i) implies that each \( l^*_\alpha(h) \) is an asymptotically correct \( \alpha \)–level critical value under any model in \( H_0 \), while Theorem 3.1(ii) shows that \( \hat{L}_T \) is asymptotically consistent. In Section 4 below, we illustrate Theorem 3.1 using a simulated example.

4. **An example of implementation.** This section studies the finite–sample properties of the size and power functions of the proposed test.

**Example 4.1.** Consider a nonlinear time series model of the form
\[ Y_t = m(X_t, \theta) + e_t \quad \text{and} \quad X_t = X_{t-1} + u_t, \quad t = 1, 2, \cdots, \quad (4.1) \]
where \( \{e_t\} \) is a sequence of i.i.d. \( N(0,1) \), \( \{u_t\} \) is also a sequence of i.i.d. \( N(0,1) \), \( X_0 = 0 \), and the forms of \( m(x, \theta) \) are given as follows:
\[
\begin{align*}
H_0 : \ m(x, \theta_0) &= \theta_0 \ x \quad \text{versus} \quad H_1 : \ m(x, \theta_1) = \theta_{11} x + \theta_{12} x^2 \quad \text{and} \\
H_0 : \ m(x, \theta_0) &= \theta_0 \ x \quad \text{versus} \quad H_1 : \ m(x, \theta_1) = \theta_{21} x + \theta_{22} x \left(1 - e^{-\theta_{23} x^2}\right) \quad (4.2) \quad (4.3)
\end{align*}
\]
where the \( \theta \)'s are chosen as follows: Case 1: \( \theta_0 = \theta_{11} = \theta_{21} = 1 \) and \( \theta_{12} = \theta_{22} = \theta_{23} = 0.08; \) Case 2: \( \theta_0 = \theta_{11} = \theta_{21} = 1 \) and \( \theta_{12} = \theta_{22} = 0.05. \) Note that Assumptions 2.2 and 3.2 both hold in this case. The form of \( m(x, \theta_1) \) in (4.3) has been used in Kapetanios, Shin and Snell (2003).

In this section, we use an ordinary least squares (OLS) method to estimate the unknown parameters for models under \( H_0 \) and the proposed semiparametric estimation method in (3.4) for the unknown parameters and functions under \( H_1. \) In order to compare the performance of the proposed test based on different bandwidths, we evaluate the finite–sample performance of the proposed test associated with both the power–based optimal bandwidth \( \hat{h}_{\text{test}} \) in (3.1) and an estimation–based optimal bandwidth of the form

\[
\hat{h}_{\text{cv}} = \arg \min_{h \in \mathcal{H}_T} \frac{1}{T} \sum_{i=1}^{T} (Y_i - \hat{m}_{\cdot i}(X_i; h))^2,
\]

in which \( \hat{m}_{\cdot i}(X_i; h) = \frac{\sum_{j=1, \neq i}^{T} K \left( \frac{x_j - x_i}{h} \right) y_j}{\sum_{l=1, \neq i}^{T} K \left( \frac{x_l - x_i}{h} \right)} \).

with \( K(x) = |x| I_{[-1,1]}(x) \) and \( \mathcal{H}_T = [T^{-1}, C_H] \) is chosen such that both small and relatively large bandwidth values may be selected, where \( C_H \) is some positive constant.

Note that each of \( \hat{h}_{\text{test}} \) and \( \hat{h}_{\text{cv}} \) has one version under \( H_0, \) but both have two different versions for Cases 1 and 2 under \( H_1. \) To use some simple notation, we introduce \( h_{i\text{test}} = \hat{h}_{\text{test}} \) and \( h_{i\text{cv}} = \hat{h}_{\text{cv}} \) for \( i = 0, 1, 2 \) to represent \( h_{0\text{test}} \) and \( h_{0\text{cv}} \) under \( H_0, \) and \( h_{i\text{test}} \) and \( h_{i\text{cv}} \) under \( H_1 \) for Cases \( i \) with \( i = 1, 2. \) We then define \( L_{i\text{test}} = \hat{L}_T(h_{i\text{test}}) \) and \( L_{i\text{cv}} = \hat{L}_T(h_{i\text{cv}}) \) for \( i = 0, 1, 2. \) For \( i = 0, 1, 2, \) let \( f_{i\text{test}} \) denote the frequency of \( L_{i\text{test}} > l^*_{\alpha}(h_{i\text{test}}) \) and \( f_{i\text{cv}} \) denote the frequency of \( L_{i\text{cv}} > l^*_{\alpha}(h_{i\text{cv}}). \) In Tables 4.1–4.3 below, we consider cases where the number of replications of each of the sample versions of the size and power functions was \( M = 1000, \) each with \( B = 250 \) number of bootstrapping resamples \( \{e_t^*\} \) (involved in the Simulation Scheme 3.1 in Section 3 above) from the standard normal distribution \( N(0,1), \) and the simulations were done for the cases of \( T = 80, 200, 500 \) and 800.

11
Table 4.1. Simulated sizes and power values at the 1% level

<table>
<thead>
<tr>
<th>Null Hypothesis Is True</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T</strong></td>
<td>$f_{0cv}$</td>
<td>$f_{0test}$</td>
</tr>
<tr>
<td>80</td>
<td>0.0090</td>
<td>0.0080</td>
</tr>
<tr>
<td>200</td>
<td>0.0080</td>
<td>0.0060</td>
</tr>
<tr>
<td>500</td>
<td>0.0110</td>
<td>0.0160</td>
</tr>
<tr>
<td>800</td>
<td>0.0130</td>
<td>0.0090</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Null Hypothesis Is False</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test</strong></td>
<td>Model (4.2)</td>
<td>Model (4.3)</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>$f_{1cv}$</td>
<td>$f_{2cv}$</td>
</tr>
<tr>
<td>80</td>
<td>0.0120</td>
<td>0.0100</td>
</tr>
<tr>
<td>200</td>
<td>0.0460</td>
<td>0.0380</td>
</tr>
<tr>
<td>500</td>
<td>0.3180</td>
<td>0.2300</td>
</tr>
<tr>
<td>800</td>
<td>0.6160</td>
<td>0.5230</td>
</tr>
</tbody>
</table>

Table 4.2. Simulated sizes and power values at the 5% level

<table>
<thead>
<tr>
<th>Null Hypothesis Is True</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T</strong></td>
<td>$f_{0cv}$</td>
<td>$f_{0test}$</td>
</tr>
<tr>
<td>80</td>
<td>0.0400</td>
<td>0.0420</td>
</tr>
<tr>
<td>200</td>
<td>0.0560</td>
<td>0.0530</td>
</tr>
<tr>
<td>500</td>
<td>0.0480</td>
<td>0.0540</td>
</tr>
<tr>
<td>800</td>
<td>0.0520</td>
<td>0.0460</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Null Hypothesis Is False</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test</strong></td>
<td>Model (4.2)</td>
<td>Model (4.3)</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>$f_{1cv}$</td>
<td>$f_{2cv}$</td>
</tr>
<tr>
<td>80</td>
<td>0.0580</td>
<td>0.0550</td>
</tr>
<tr>
<td>200</td>
<td>0.1230</td>
<td>0.0990</td>
</tr>
<tr>
<td>500</td>
<td>0.4990</td>
<td>0.4070</td>
</tr>
<tr>
<td>800</td>
<td>0.7520</td>
<td>0.6740</td>
</tr>
</tbody>
</table>
Table 4.3. Simulated sizes and power values at the 10% level

<table>
<thead>
<tr>
<th>T</th>
<th>( f_{0cv} )</th>
<th>( f_{0test} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.0930</td>
<td>0.0920</td>
</tr>
<tr>
<td>200</td>
<td>0.1050</td>
<td>0.1060</td>
</tr>
<tr>
<td>500</td>
<td>0.0930</td>
<td>0.0970</td>
</tr>
<tr>
<td>800</td>
<td>0.1090</td>
<td>0.0990</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Model (4.2)</th>
<th>Model (4.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>( f_{1cv} )</td>
<td>( f_{1test} )</td>
</tr>
<tr>
<td>80</td>
<td>0.1110</td>
<td>0.1030</td>
</tr>
<tr>
<td>200</td>
<td>0.2130</td>
<td>0.1730</td>
</tr>
<tr>
<td>500</td>
<td>0.6150</td>
<td>0.5300</td>
</tr>
<tr>
<td>800</td>
<td>0.8220</td>
<td>0.7610</td>
</tr>
</tbody>
</table>

Tables 4.1–4.3 all show that both the proposed test and the proposed Simulation Scheme are implementable and work well numerically for the co-integration case. First of all, the augmented test based on \( \hat{h}_{test} \) is more powerful than that associated with \( \hat{h}_{cv} \) in each individual case. Second, Tables 4.1–4.3 show that the proposed test is applicable to test both linear and nonlinear alternatives. Third, Tables 4.1–4.3 show that the proposed test still has power even when the 'distance' between the null and an alternative is made deliberately close. For example, when \( \theta_{12} \) and \( \theta_{22} \) are made as small as 5% and the sample size is as medium as \( T = 80 \), the proposed test still has a power value greater than the nominal level in each case. Finally, Tables 4.1–4.3 also show that the power increases when the 'distance' between the null hypothesis and an alternative increases.

5. Conclusion and extensions. We have proposed a new nonparametric test for the conditional mean function when the regressors are integrated. The asymptotic normal distribution of the proposed test statistic has been established. In addition, we have also proposed a Simulation Scheme to implement the proposed test in practice. The finite–sample results show that both the proposed test and the Simulation Scheme are practically applicable and implementable.
As briefly mentioned in the introductory section, we may also consider testing the conditional variance nonparametrically. Furthermore, both the conditional mean and the conditional variance functions may be specified simultaneously. The main idea is as follows. To test
\[ H_{01}: \ P (m(X_t) = m_{\theta_0}(X_t) \text{ and } \sigma(X_t) = \sigma_{\theta_0}(X_t)) = 1, \] (5.1)

we may use a kernel–based test of the form
\[ L_T(h) = \sum_{t=1}^{T} \sum_{s=1,s\neq t}^{T} [U_s K_{h_1}(X_s - X_t) U_t + V_s G_{h_2}(X_s - X_t) V_t], \] (5.2)

where \( h = (h_1, h_2) \) is a pair of bandwidth parameters, \( K(\cdot) \) and \( G(\cdot) \) are both probability kernel functions, \( U_t = Y_t - m_{\hat{\theta}}(X_t) \), \( V_t = U_t^2 - \sigma_{\hat{\theta}}^2(X_t) \) and \( \hat{\theta} \) is an estimator of \( \theta_0 \) under \( H_{01} \). Analogously to Theorem 2.1, we may establish a corresponding theorem for \( L_T(h) \).

As the detail for this case is extremely lengthy and technical, we leave this issue for future study.

Another important extension would be to the case where \( X_t = (X_{t1}, \ldots, X_{td}) \) in (1.1) is a vector of \( d \)-dimensional nonstationary sequences. In this case, we are interested in testing
\[ H_{02}: \ P \left( m(X_t) = \sum_{i=1}^{d} m_{i\theta_0}(X_{ti}) \right) = 1 \text{ for all } t \geq 1, \] (5.3)

where each \( m_{i\theta_0}(\cdot) \) is a known function indexed by \( \theta_0 \). Detailed construction of such a test would involve some estimation procedures for additive models as used in Gao, Lu and Tjøstheim (2006) in the stationary spatial case. Since such an extension is not straightforward, we leave it as a future topic.

6. Acknowledgments. The authors also acknowledge useful comments from seminar participants, Bruce Hansen, Yongmiao Hong, Peter Phillips and Robert Taylor in particular. Thanks also go to Gowry Sriananthakumar and Jiying Yin for their excellent computing assistance and the Australian Research Council for its continuing support of the Discovery Grants under Grant Numbers: DP0558602 and DP0879088.
1 Appendix A

This appendix provides mathematical details for the proofs of the main theorems and their associated lemmas.

To avoid notational complication, we introduce the following notation. Let \( a_{st} = K_h(X_t - X_s) \), \( \epsilon_t = \sigma_0 \epsilon_t \) and \( \eta_t = 2 \sum_{s=1}^{t-1} a_{st} \epsilon_s \). Recall \( \lambda_t(\theta_0) = m_{\theta_0}(X_t) - m_\theta(X_t) \).

Observe that under \( H_0 \)

\[
M_T(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t} \tilde{\epsilon}_s K_h(X_t - X_s) \tilde{\epsilon}_t = \sum_{t=1}^{T} \sum_{s=1,\neq t} \epsilon_s K_h(X_s - X_t) \epsilon_t \\
+ \sum_{t=1}^{T} \sum_{s=1,\neq t} \lambda_s(\theta_0) K_h(X_t - X_s) \lambda_t(\theta_0) + 2 \sum_{t=1}^{T} \sum_{s=1,\neq t} \epsilon_s K_h(X_t - X_s) \lambda_t(\theta_0) \\
\equiv M_{T1} + M_{T2} + M_{T3}, \tag{A.1}
\]

where \( \tilde{\epsilon}_t \) is defined as in the main text. Now, define

\[
\sigma_T^2 = \sum_{t=1}^{T} \sum_{s=1,\neq t} \tilde{\epsilon}_s^2 K_h^2(X_t - X_s) \tilde{\epsilon}_t^2 = \sum_{t=1}^{T} \sum_{s=1,\neq t} \epsilon_s^2 K_h^2(X_t - X_s) \epsilon_t^2 \\
+ 2 \sum_{t=1}^{T} \sum_{s=1,\neq t} \lambda_s^2(\theta_0) K_h^2(X_t - X_s) \lambda_t^2(\theta_0) + \hat{R}_T, \tag{A.2}
\]

where \( \hat{R}_T \) is the remainder term given by

\[
\hat{R}_T = \sigma_T^2 - 2 \sum_{t=1}^{T} \sum_{s=1,\neq t} \epsilon_s^2 K_h^2(X_t - X_s) \epsilon_t^2 - 2 \sum_{t=1}^{T} \sum_{s=1,\neq t} \lambda_s^2(\theta_0) K_h^2(X_t - X_s) \lambda_t^2(\theta_0).
\]

In view of (A.1) and (A.2), to prove Theorem 2.1, it suffices to show that as \( T \to \infty \)

\[
\frac{M_{T1}}{\sigma_T} \to_D N(0, 1), \tag{A.3}
\]

\[
\frac{M_{Ti}}{\sigma_T} \to_P 0 \quad \text{for } i = 2, 3, \tag{A.4}
\]

\[
\frac{\sigma_T^2 - \hat{\sigma}_T^2}{\sigma_T^2} \to_P 0, \tag{A.5}
\]

where \( \hat{\sigma}_T^2 = 2 \sum_{t=1}^{T} \sum_{s=1,\neq t} \epsilon_s^2 a_{st} \epsilon_t^2 \).

We will return to the proof of (A.4) and (A.5) in the second half of this appendix after having proved Lemmas A.1–A.3. In order to prove (A.3), we need to introduce a stochastic normalization procedure before we may apply Corollary 3.1 of Hall and Heyde (1980, p.58) to our case.

Let \( C_{10} = 2 \sigma_0^4 \int K^2(u)du \) and define a random variable of the form

\[
\sigma_{10}^2 = C_{10} \ N(T) \ Th, \tag{A.6}
\]
in which $N(T)$ has the same definition as $T(n)$ defined in Karlsen and Tjøstheim (2001). It is the number of regenerations for the Markov chain \{\(X_t\)\}. Note that we use $\sigma^2_{10}$ to express the explicit function of the random variable $N(T)$ for notational simplicity. More details about the definition of $N(T)$ are available from Appendix B below. In addition, it also follows from the Appendix B that the following inequality

\[ T^{\frac{1}{2}-\delta_0} \leq N(T) \leq T^{\frac{1}{2}+\delta_0} \] (A.7)

holds almost surely for $T$ large enough and all $0 < \delta_0 < \frac{1}{5}$.

As shown in Lemma A.3 below, we have as $T \to \infty$

\[
\frac{\tilde{\sigma}^2_T}{\sigma_{10}^2} \to P 1. \tag{A.8}
\]

In view of (A.8), to prove (A.3), it suffices to show that as $T \to \infty$

\[
\frac{M_{T1}}{\sigma_{10}} \to D N(0,1). \tag{A.9}
\]

We now start to prove (A.9). Before verifying the conditions of Corollary 3.1 of Hall and Heyde (1980), we introduce the following notation.

Let $U_{Tt} = \frac{e_t}{\sigma_{10}}$ and $\Omega_{T,t} = \sigma\{U_{Tt} : 1 \leq t \leq s\}$ be the $\sigma$-field generated by $\{U_{Tt} : 1 \leq t \leq s\}$. Since $N(T)$ is independent of $\{e_t : 1 \leq t \leq T\}$ by construction, $E[U_{Tt}|\Omega_{T,t-1}] = 0$. By Corollary 3.1 of Hall and Heyde (1980), in order to prove (A.9), it suffices to show that for all $\delta > 0$,

\[
\sum_{t=2}^{T} E\left[U_{Tt}^2 I_{\{|U_{Tt}|>\delta\}}\right] |\Omega_{T,t-1} \to P 0, \tag{A.10}
\]

\[
\sum_{t=2}^{T} E\left[U_{Tt}^2 \right] |\Omega_{T,t-1} \to P 1. \tag{A.11}
\]

In view of the definition of $\{U_{Tt}\}$, in order to verify (A.10) and (A.11), it suffices to show that as $T \to \infty$

\[
\frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \eta_t^1 \to_P 0, \tag{A.12}
\]

\[
\frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \eta_t^2 \to_P 1. \tag{A.13}
\]

The proofs of (A.12) and (A.13) are given in Lemmas A.2 and A.3 respectively.

**A.1. Lemmas.** Assumption 2.1(i) already assumes that $\{u_i\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and has a symmetric probability density function. Now we let $f(x)$ and $f_{st}(x)$ be the density functions of $u_i$ and $X_{st} = \frac{X_s + X_t}{2}$ for $s, t \in \mathbb{Z}$.
for where $\nu$ to evaluate $s,t$ mutually independent for all $g$ following notation: this appendix. For $i$ it sup $st$ that $\sigma^2_1 = \text{var} \{X_{s,t}\}$ be the density function of $V_{s,t} = \frac{X_{s,t}}{\sqrt{t-s}}$. Clearly $f_{s,t}(x) = g_{s,t} \left( \frac{x}{\sqrt{t-s}} \right) \frac{1}{\sqrt{t-s}}$, and by utilising the usual normal approximation of $V_{s,t} \to_D N(0,1)$ as $t-s \to \infty$ under the conventional central limit theorem conditions, it follows from Assumption 2.1(i) that $\sup_{x \in R^1} |g_{s,t}(x) - \phi(x)| \to 0$ as $t-s \to \infty$. Thus, $\sup_{x \in R^1} |g_{s,t} \left( \frac{x}{\sqrt{t-s}} \right) - \phi \left( \frac{x}{\sqrt{t-s}} \right)\left| \to 0 \right.$ as $t-s \to \infty$, where $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$.

Another key condition used in the following proofs is that $\{e_s\}$ and $\{u_t\}$ are assumed to be mutually independent for all $s, t \geq 1$. In order to complete the proof of Theorem 2.1, we need to evaluate $\sigma^2_{T_1} = \text{var}(M_{T_1}(h))$. Recall $e_s = \sigma_0 e_s$, $X_{s,t} = X_t - X_s = \sum_{i=s+1}^{t} u_i$ and define,

$$\xi_{s,t} = K_h(X_{s,t}) e_s e_t \text{ with } \lambda_s(\theta_0) = m_{\theta_0}(X_s) - m_{\theta_0}(X_s).$$

We assume without loss of generality that $\sigma^2_2 = E[u^2_t] \equiv 1$ and $\sigma^2_0 = E[e^2_s] \equiv 1$ throughout this appendix. For $i = 1, \cdots, 4$, $1 \leq s < t \leq T$ and $1 \leq s_2 < s_1 < t \leq T$, we introduce the following notation:

$$B_i(s,t) = E \left[ K_h(X_{s,t}) \right] = h \int K_i(x) f_{s,t}(xh) \, dx$$

$$B_i(s_1, s_2, t) = E \left[ K_h(X_{s_1,t})K_h(X_{s_1,t} + X_{s_2,s_1}) \right]$$

$$C_{cdpq}(s,t) = E \left[ \epsilon_s^c \epsilon_t^d (\epsilon^2_s - 1)^p (\epsilon^2_t - 1)^q \right]$$

where $\nu_4 = E[\epsilon^4_t]$. Since $\{e_t\}$ and $\{u_s\}$ are assumed to be mutually independent for all $s, t$, we can obtain that for $T$ large enough,

$$\sigma^2_{T_1} = \text{var} [M_{T_1}(h)] = 4(1 + o(1)) \sum_{t=2}^{T} \sum_{s=1}^{t-1} E \left[ \xi^2_{s,t} \right] = 4\sigma^2_0(1 + o(1)) \sum_{t=2}^{T} \sum_{s=1}^{t-1} B_2(s,t) C_{2200}(s,t).$$
Lemma A.1 below derives the order of $\sigma^2 T_1$ and shows that the rate of $\sigma^2 T_1$ diverging to $\infty$ is slower than $T^2 h$, which is the corresponding rate for the stationary case.

**Lemma A.1.** Assume that the conditions of Theorem 2.1 hold. Then as $T \to \infty$

$$\sigma^2 T_1 = \text{var} [M_{T_1}(h)] = C_0 T^{\frac{3}{2}} h (1 + o(1)), \hspace{1cm} (A.18)$$

where $C_0 = \frac{16\sigma_0^4}{3\sqrt{2\pi}} \int K^2(u)du$.

**Proof:** Choose some positive integer $\Pi_T \geq 1$ such that $\Pi_T \to \infty$ and $\frac{\Pi_T}{\sqrt{T} h} \to 0$ as $T \to \infty$.

Observe that

$$\sum_{t=2}^{T} \sum_{s=1}^{T-1} E[a_{st}^2] = \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} E[a_{st}^2] = A_{1T} + A_{2T}, \hspace{1cm} (A.19)$$

where $A_{1T} = \sum_{s=1}^{T-1} \sum_{1 \leq (t-s) \leq \Pi_T} E[a_{st}^2] = O(T \Pi_T) = o(T^{3/2} h)$ using the fact that $E[a_{st}^2] \leq k_0^2$ due to the boundedness of the kernel $K(\cdot)$ by a constant $k_0 > 0$.

Using (A.15), we have

$$A_{2T} = \sum_{s=1}^{T-1} \sum_{\Pi_T+1 \leq (t-s) \leq T-1} E[a_{st}^2]$$

$$= \sum_{s=1}^{T-1} \sum_{\Pi_T+1 \leq (t-s) \leq T-1} \frac{h}{\sqrt{t-s}} \int K^2(x) g_{st} \left( \frac{xh}{\sqrt{t-s}} \right) dx$$

$$= d_0 h(1 + o(1)) \int K^2(x) dx \sum_{s=1}^{T-1} \sum_{\Pi_T+1 \leq (t-s) \leq T-1} \frac{1}{\sqrt{t-s}}$$

$$= \frac{4 \int K^2(y)dy}{3} d_0 T^{\frac{3}{2}} h(1 + o(1)), \hspace{1cm} (A.20)$$

where $d_0 = \frac{1}{\sqrt{2\pi}}$.

Equations (A.19) and (A.20) imply that for for $T$ large enough

$$\sum_{t=2}^{T} \sum_{s=1}^{T-1} E[a_{st}^2] = \frac{4 \int K^2(y)dy}{3\sqrt{2\pi}} T^{\frac{3}{2}} h(1 + o(1)). \hspace{1cm} (A.21)$$

Therefore, it follows that for $T \to \infty$

$$4 \sum_{t=2}^{T} \sum_{s=1}^{T-1} E \left[ \xi_{st}^2 \right] = 4\sigma_0^4 \sum_{t=2}^{T} \sum_{s=1}^{T-1} B_2(s, t) C_{2200}(s, t) = C_0 T^{\frac{7}{2}} h (1 + o(1)), \hspace{1cm} (A.22)$$

where $C_0 = \frac{16\sigma_0^4}{3\sqrt{2\pi}} \int K^2(u)du$. Thus, the proof of Lemma A.1 is completed.
For $0 < \delta_0 < \frac{1}{5}$, recall $C_{10}$ as defined in (A.6) and let $\sigma_{20}^2 = C_{10} T^{\frac{1}{2} - \delta_0} Th$. We now have the following lemma.

**Lemma A.2.** Under the conditions of Theorem 2.1, we have as $T \to \infty$

$$\frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \eta_t^4 \to P 0.$$ (A.23)

**Proof.** In view of (A.7), we have for $T$ large enough and any given $\delta > 0$

$$P \left( \frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \eta_t^4 > \delta \right) \leq \frac{1}{\sigma_{20}^2} \sum_{t=2}^{T} E \left[ \eta_t^4 \right] + o(1).$$ (A.24)

Thus, in order to prove (A.23), we need only to show that

$$\frac{1}{\sigma_{20}^2} \sum_{t=2}^{T} E \left[ \eta_t^4 \right] \to 0.$$ (A.25)

Observe that

$$E \left[ \eta_t^4 \right] = 16 \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} \sum_{s_3=1}^{t-1} \sum_{s_4=1}^{t-1} E \left[ a_{s_1 t} a_{s_2 t} a_{s_3 t} a_{s_4 t} \epsilon_{s_1} \epsilon_{s_2} \epsilon_{s_3} \epsilon_{s_4} \right].$$ (A.26)

Since Assumption 2.1 imposes the mutual independence on $\{u_s\}$ and $\{e_t\}$ for all $s, t \geq 1$, in order to prove (A.23), it suffices to show that as $T \to \infty$

$$\frac{1}{\sigma_{20}^2} \sum_{t=2}^{T} \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} \sum_{s_1 \neq s_2}^{t-1} E \left[ a_{s_1 t}^2 a_{s_2 t}^2 \right] \to 0,$$ (A.27)

$$\frac{1}{\sigma_{20}^2} \sum_{t=2}^{T} \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} E \left[ a_{s_1 t}^4 \right] \to 0.$$ (A.28)

To prove (A.27), using (A.16) we have

$$\sum_{t=2}^{T} \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} E \left[ a_{s_1 t}^2 a_{s_2 t}^2 \right] = 4 h^2 \sum_{t=3}^{T} \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{t-1} \frac{1}{\sqrt{t-s_1}} \frac{1}{\sqrt{s_1-s_2}} \int K^2(x)K^2(x+y)$$

$$\times g_{s_1 t} \left( \frac{xh}{\sqrt{t-s_1}} \right) g_{s_2 t} \left( \frac{yh}{\sqrt{s_1-s_2}} \right) dx dy.$$
\[= 4h^2(1 + o(1))J_{02}^2 d_0^2 \sum_{t=3}^T \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{t-1-1} \frac{1}{\sqrt{t - s_1}} \frac{1}{\sqrt{s_1 - s_2}} \]

\[= C \ T^2 \ h^2 = o \left( T^{3-2\delta_0} h^2 \right) = o \left( \sigma_{20}^4 \right) \] (A.29)

using the assumption that \( \lim_{T \to \infty} T^{1-\delta_0} h = \infty \) for \( 0 < \delta_0 < \frac{1}{5} \), where \( C > 0 \) is some constant, \( J_{02} = \int K^2(u)du \) and \( d_0 = \frac{1}{\sqrt{2\pi}} \).

Similarly to (A.20), using (A.15) we have

\[\sum_{t=2}^T \sum_{s=1}^{t-1} E[a^4_{st}] = \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{h}{\sqrt{t - s}} \int K^4(x) g_{st} \left( \frac{xh}{\sqrt{t - s}} \right) dx \]

\[= C_0 \ h(1 + o(1)) \int K^4(x) dx \cdot \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{1}{\sqrt{t - s}} \]

\[= C \ T^2 h \ (1 + o(1)) = o \left( T^{3-2\delta_0} h^2 \right) = o \left( \sigma_{20}^4 \right) \] (A.30)

using the assumption that \( \lim_{T \to \infty} T^{1-\delta_0} h = \infty \), where \( C > 0 \) is some constant.

Equations (A.29) and (A.30) complete the proofs of (A.27) and (A.28). This completes the proof of Lemma A.2.

**Lemma A.3.** Let the conditions of Theorem 2.1 hold. Then as \( T \to \infty \)

\[\frac{1}{\sigma_{10}^2} \sum_{t=2}^T \eta_t^2 \to P 1. \] (A.31)

**Proof.** Observe that

\[\sum_{t=2}^T \sum_{s=1}^{t-1} a^4_{st} = \sum_{t=2}^T \left( 2 \sum_{s=1}^{t-1} a^2_{st} \epsilon_s \right)^2 \]

\[= 4 \sum_{t=2}^T \sum_{s=1}^{t-1} a^2_{st} \epsilon_s^2 + 4 \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{s=1}^{t-1} \epsilon_s a_{st} a_{s't'} \epsilon_{s't'} \] (A.32)

We first show that as \( T \to \infty \)

\[\frac{4}{\sigma_{10}^2} \sum_{t=2}^T \sum_{s=1}^{t-1} a^2_{st} \epsilon_s^2 \to P 1. \] (A.33)

Similarly to the proofs of Lemmas A.1 and A.2, it can be shown that

\[\sum_{t=2}^T \left( \sum_{s=1}^{t-1} a^2_{st} \epsilon_s^2 - 1 \right) = o_{P} \left( \sigma_{10}^2 \right) \] (A.34)

using the assumption that \( \{ \epsilon_t \} \) is assumed to be independent of \( \{ u_s \} \) for all \( s, t \) and \( E[\epsilon_t^2] = 1 \).

In view of (A.34), in order to prove (A.33), it suffices to show that as \( T \to \infty \)

\[\frac{4}{\sigma_{10}^2} \sum_{t=2}^T \sum_{s=1}^{t-1} a^2_{st} = \frac{2}{\sigma_{10}^2} \sum_{t=1}^T \sum_{s=1}^{t-1} a^2_{st} \to P 1. \] (A.35)
Let \( Q(u) = \frac{K^2(u)}{\int K^2(u) \, du} \). Then \( Q(\cdot) \) is a probability kernel. According to Lemma C.1 in Appendix C of Gao et al (2008), we have that as \( T \to \infty \)

\[
\frac{1}{N(T)h} \sum_{s=1}^{T} Q \left( \frac{X_s - x}{h} \right) \to P 1
\]  

(A.36)

uniformly in all \( x \in \mathbb{R}^1 \), where we have used the result that the invariant measure of the random walk \( \{X_t\} \) can be taken to be Lebesgue measure with corresponding density \( p(x) \equiv 1 \). The uniform convergence in (A.36) strengthens the point–wise convergence of Theorem 5.1 of Karlsen and Tjøstheim (2001) in the random walk case. We refer to Appendices B–D for more details.

Thus, the proof of (A.35) follows from (A.36) and

\[
\frac{2}{\sigma_{10}^2} \sum_{t=1}^{T} \sum_{s=1}^{T} a_{st}^2 = \frac{2}{T} \sum_{t=1}^{T} \left( \frac{1}{N(T)h} \sum_{s=1}^{T} K^2 \left( \frac{X_s - X_t}{h} \right) \right) \to P 1
\]  

(A.37)

as \( T \to \infty \).

In view of (A.31) and (A.32), in order to complete the proof of (A.31), we need to show that

\[
\frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \sum_{s_1=1}^{t-1} \sum_{s_2=1, s_2 \neq s_1}^{t-1} \epsilon_{s_1} a_{s_1 t} a_{s_2 t} \epsilon_{s_2} \to P 0 \quad \text{as} \quad T \to \infty.
\]  

(A.38)

Similarly to (A.24), the proof of (A.38) follows from

\[
\frac{1}{\sigma_{20}^2} \mathbb{E} \left[ \sum_{t=2}^{T} \sum_{s_1=1}^{t-1} \sum_{s_2=1, s_2 \neq s_1}^{t-1} \epsilon_{s_1} a_{s_1 t} a_{s_2 t} \epsilon_{s_2} \right]^2 \to 0,
\]  

(A.39)

which, using the same arguments as in (A.25)–(A.30) and the fact that \( \{\epsilon_s\} \) is a sequence of martingale differences and also independent of \( \{u_t\} \), follows from

\[
\sum_{t_1=2}^{T} \sum_{t_2=1}^{T} \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} E[a_{s_1 t_1} a_{s_2 t_2} a_{s_1 t_2} a_{s_2 t_1}] = O \left( T^5 h^3 \right) = o \left( \sigma_{20}^4 \right)
\]  

(A.40)

This therefore completes the proof of Lemma A.3.

**A.2. Proof of Theorem 2.1.** In view of (A.3), to complete the proof of Theorem 2.1, it suffices to prove (A.4) and (A.5). We only give the proof of (A.4) since the proof of (A.5) is very similar.
Taylor expansions of \( m_\theta(x) \) with respect to \( \theta \) at \( \theta_0 \) imply

\[
m_\theta(x) - m_{\theta_0}(x) = \left( \frac{\partial m_{\theta_0}(x)}{\partial \theta} \right)^\tau (\theta - \theta_0) + o_P(||\theta - \theta_0||) \tag{A.41}
\]

for each given \( x \). Thus, in order to prove (A.4), using the same arguments as in (A.24), it suffices to show that

\[
(\hat{\theta} - \theta_0)^\tau \sum_{t=1}^T \sum_{s=1}^T \left( \frac{\partial m_{\theta_0}(X_s)}{\partial \theta} \right)^\tau \left( \frac{X_t - X_s}{h} \right) \left( \frac{\partial m_{\theta_0}(X_t)}{\partial \theta} \right)^\tau (\hat{\theta} - \theta_0) = o_P(\sigma_{10}). \tag{A.42}
\]

Note that using the same arguments as in (A.24), the proof of (A.42) follows when (A.42) holds with \( \sigma_{10} \) replaced by \( \sigma_{20} \).

In order to do so, we first evaluate the following quantity. Straightforward calculations imply that for \( T \) large enough (letting \( Y_1 = X_s \) and \( Y_{12} = X_t - X_s \) and then \( X_1 = y_1 \) and \( X_2 = \frac{y_2}{h} \))

\[
\sqrt{T} \left( \frac{X_t - X_s}{h} \right) \left( \frac{\partial m_{\theta_0}(X_t)}{\partial \theta} \right)^\tau \left( \frac{X_t - X_s}{h} \right) \left( \frac{\partial m_{\theta_0}(X_t)}{\partial \theta} \right)^\tau (\hat{\theta} - \theta_0) = o_P(\sigma_{10}). \tag{A.42}
\]

This, along with Assumption 2.2(iii) with \( j = 1 \) and the Markov inequality, implies that (A.42) holds with \( \sigma_{10} \) replaced by \( \sigma_{20} \). This therefore proves (A.4) for \( i = 2 \).

Meanwhile, it follows from (A.3) that

\[
\frac{1}{\sigma_{10}} \sum_{t=1}^T \sum_{s=1}^T \epsilon_s K \left( \frac{X_t - X_s}{h} \right) \epsilon_t = O_P(1). \tag{A.44}
\]

Thus, the proof of (A.4) for \( i = 3 \) follows from (A.42)–(A.44) and

\[
\left| \sum_{t=1}^T \sum_{s=1}^T \epsilon_s K \left( \frac{X_t - X_s}{h} \right) K \left( \frac{X_t - X_s}{h} \right) \lambda_t(\theta_0) \right|^2 \leq \sum_{t=1}^T \sum_{s=1}^T \epsilon_s K \left( \frac{X_t - X_s}{h} \right) \epsilon_t \times \sum_{t=1}^T \sum_{s=1}^T \lambda_s(\theta_0) K \left( \frac{X_t - X_s}{h} \right) \lambda_t(\theta_0) = O_P(\sigma_{10}) \cdot o_P(\sigma_{10}) = o_P(\sigma_{10}^2). \tag{A.45}
\]

22
Similarly to (A.41)–(A.43), using Assumption 2.2(iii) with \( j = 2 \), one may verify (A.5).

### A.3. Proof of Theorem 3.1

Using

\[
\tilde{c}_t^* \equiv Y_t^* - m_{\tilde{\theta}}(X_t) = m_{\tilde{\theta}}(X_t) - m_{\tilde{\theta}^*}(X_t) + \tilde{\sigma}_0 e_t^*,
\]

we have

\[
M^*_T(h) \equiv \sum_{t=1}^T \sum_{s=1, s \neq t}^{T} \tilde{c}_s^* K_h(X_s - X_t) \tilde{c}_t^* = \sum_{t=1}^T \sum_{s=1, s \neq t}^{T} \tilde{\sigma}_0 e_s^* K_h(X_s - X_t) \tilde{\sigma}_0 e_t^* + \sum_{t=1}^T \sum_{s=1, s \neq t}^{T} \lambda_s^* K_h(X_s - X_t) \lambda_t^* + 2 \sum_{t=1}^T \sum_{s=1, s \neq t}^{T} \tilde{\sigma}_0 e_s^* K_h(X_s - X_t) \lambda_t^*, \tag{A.46}
\]

where \( \lambda_t^* = m_{\tilde{\theta}}(X_t) - m_{\tilde{\theta}^*}(X_t) \).

Using Assumptions 2.1–2.2 and 3.1, in view of the notation of \( \tilde{L}_T^*(h) \) introduced in the Simulation Scheme 3.1 above Assumption 3.1 as well as the proof of Theorem 2.1, we may show that as \( T \to \infty \)

\[
P^* \left( \tilde{L}_T^*(h) \leq x \right) \to \Phi(x) \quad \text{for all } x \in (-\infty, \infty) \tag{A.47}
\]

holds in probability with respect to the distribution of the original sample \( \{(X_t, Y_t) : 1 \leq t \leq T\} \).

In detail, in order to prove (A.47), using the fact that \( \{e_s^*\} \) and \( \{(X_t, Y_t)\} \) are independent for all \( s, t \geq 1 \), we may show that the proofs of Lemmas A.2 and A.3 all remain true by successive conditioning arguments.

Let \( z_\alpha \) be the \( 1 - \alpha \) quantile of \( \Phi(\cdot) \) such that \( \Phi(z_\alpha) = 1 - \alpha \). Then it follows from (A.47) that as \( T \to \infty \)

\[
P^* \left( \tilde{L}_T^*(h) \geq z_\alpha \right) \to 1 - \Phi(z_\alpha) = \alpha. \tag{A.48}
\]

This, together with the construction that \( P^* \left( \tilde{L}_T^*(h) \geq l_\alpha^* \right) = \alpha \), implies that as \( T \to \infty \)

\[
l_\alpha^* - z_\alpha \to p 0. \tag{A.49}
\]

Using the conclusion of Theorem 2.1 and (A.47) again, we have as \( T \to \infty \)

\[
P^* \left( \tilde{L}_T^*(h) \leq x \right) - P \left( \tilde{L}_T(h) \leq x \right) \to p 0 \quad \text{for all } x \in (-\infty, \infty). \tag{A.50}
\]

This, along with the construction that \( P^* \left( \tilde{L}_T^*(h) \geq l_\alpha^* \right) = \alpha \) again, shows that as \( T \to \infty \)

\[
\lim_{T \to \infty} P \left( \tilde{L}_T(h) \geq l_\alpha^* \right) = \alpha \tag{A.51}
\]

holds. Therefore the conclusion of Theorem 3.1(i) is proved.
To prove Theorem 3.1(ii), we need to decompose $M_T(h)$ as follows:

$$M_T(h) = \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \hat{e}_s K(X_{st}) \hat{e}_t = \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \epsilon_s(\theta_1) K_h(X_{st}) \epsilon_t(\theta_1)$$

$$+ \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \lambda_s(\theta_1) K_h(X_{st}) \lambda_t(\theta_1) + 2 \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \lambda_s(\theta_1) K_h(X_{st}) \epsilon_t(\theta_1),$$

where $\epsilon_t(\theta_1) = Y_t - m(X_t)$ and $\lambda_t(\theta_1) = m(X_t) - m_{\hat{\theta}}(X_t)$ under $H_1$.

By the proof of Theorem 2.1, in order to prove Theorem 3.2(ii), it suffices to show that under $H_1$

$$\sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \lambda_s(\theta_1) K_h(X_{st}) \lambda_t(\theta_1) \sigma_{10} \rightarrow \sigma_{20},$$

(A.52)

Using Taylor expansions to $m_\theta(\cdot)$ with respect to $\theta$, we have

$$m(X_t) - m_{\hat{\theta}}(X_t) = \Delta_T(X_t, \theta_1) + m_\theta(X_t) - m_{\hat{\theta}}(X_t)$$

$$= \Delta_T(X_t, \theta_1) + (\theta_1 - \hat{\theta}) \tau \frac{\partial m_\theta(X_t)}{\partial \theta} |_{\theta = \theta_1}. $$

(A.53)

In view of (A.52), using Assumption 2.2(iii) with $i = 1$, in order to prove (A.52), it suffices to show that

$$\sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} E \left[ \Delta_T(X_s, \theta_1) K_h(X_t - X_s) \Delta_T(X_t, \theta_1) \right] \sigma_{20} \rightarrow \sigma_{20},$$

(A.54)

which follows from Assumption 3.2. This completes the proof of Theorem 3.1(ii).

2 Appendix B

To make this paper more self-contained, we summarize the definitions of some terms as well as some facts in Markov theory in this section. We still adopt the notations used in Nummelin (1984) and Karlsen and Tjøstheim (2001).

Let $\{X_t, t \geq 0\}$ be a class of Markov chains with transition probability $P$ and state space $(E, \mathcal{E})$, and $\phi$ be a measure on $(E, \mathcal{E})$. $\{X_t, t \geq 0\}$ is said to be $\phi$–irreducible if each $\phi$–positive set $A$ is communicating with the whole state space $E$, i.e.

$$\sum_{n=1}^{\infty} P^n(x, A) > 0, \text{ for all } x \in E \text{ whenever } \phi(A) > 0.$$

Denote the class of nonnegative measurable functions with $\phi$–positive support by $\mathcal{E}^+$. For a set $A \in \mathcal{E}$, we write $A \in \mathcal{E}^+$ if $1_A \in \mathcal{E}^+$, where $1_A$ stands for the indicator function of the set $A$. 

24
The chain \( \{X_t\} \) is Harris recurrent if for all \( A \in \mathcal{E}^+ \), \( x \in \mathbf{E} \),

\[
P(S_A < \infty | X_0 = x) \equiv 1, \quad S_A = \min\{t \geq 1, \ X_t \in A\},
\]
or equivalently, if given a neighborhood \( N_x \) of \( x \), \( x \in \mathbf{E} \), with \( \phi(N_x) > 0 \), \( \{X_t\} \) will return to \( N_x \) with probability one. This is what makes asymptotics for our semi-parametric estimation possible. So in the following, we will always assume that \( \{X_t\} \) is \( \psi \)–irreducible Harris recurrent.

Let \( \eta \) be a nonnegative measurable function and \( \lambda \) be a measure. We define the kernel \( \eta \otimes \lambda \) by

\[
\eta \otimes \lambda(x, A) = \eta(x)\lambda(A), \quad (x, A) \in (\mathbf{E}, \mathcal{E}).
\]

If \( K \) is a kernel, we define the function \( K\eta \), the measure \( \lambda K \) and the number \( \lambda\eta \) by

\[
K\eta(x) = \int K(x, dy)\eta(y), \quad \lambda K(A) = \int \lambda(dx)K(x, A), \quad \lambda\eta = \int \lambda(dx)\eta(x).
\]

The convolution of two kernels \( K_1 \) and \( K_2 \) is defined by

\[
K_1K_2(x, A) = \int K_1(x, dy)K_2(y, A).
\]

A function \( \eta \in \mathcal{E}^+ \) is said to be a small function if there exist a measure \( \lambda \), a positive constant \( b \) and an integer \( m \geq 1 \), so that \( P^m \geq b\eta \otimes \lambda \).

And if \( \lambda \) satisfies the above inequality for some \( \eta \in \mathcal{E}^+ \), \( b > 0 \) and \( m \geq 1 \), then \( \lambda \) is called a small measure. A set \( A \) is small if \( 1_A \) is a small function. By Theorem 2.1 and Proposition 2.6 in Nummelin (1984), we know that for a \( \phi \)–irreducible Markov chain, there exists a minorization inequality: there are a small function \( s \), a probability measure \( \nu \) and an integer \( m_0 \geq 1 \) such that \( P^{m_0} \geq s \otimes \nu \).

As pointed out by Karlsen and Tjøstheim (2001), it causes some technical difficulties to have \( m_0 > 1 \) and it is not a severe restriction to assume \( m_0 = 1 \). So in the paper, we always assume that the minorization inequality

\[
P \geq s \otimes \nu \tag{B.1}
\]

holds with \( \nu(\mathbf{E}) = 1, \ 0 \leq s(x) \leq 1, \ x \in \mathbf{E} \).

We apply the method of Karlsen and Tjøstheim (2001) to our case in this paper. In this method, an important role is played by the split chain, which can be constructed when the minorization inequality (B.1) holds. This allows for the decomposition of the chain into identically distributed main parts and remaining parts that are asymptotically negligible. Denote

\[
Q(x, A) = (1 - s(x))^{-1}(P(x, A) - s(x)\nu(A))1(s(x) < 1) + 1_A(x)1(s(x) = 1).
\]
Then the transition probability $P(x, A)$ can be decomposed as

$$P(x, A) = (1 - s(x))Q(x, A) + s(x)\nu(A).$$

When (B.1) holds, it can be verified that $Q$ is a transition probability. As $0 \leq s(x) \leq 1$ and $\nu(E) = 1$, $P$ can be seen as a mixture of the transition probability $Q$ and the small measure $\nu$. Since $\nu$ is independent of $x$, the chain regenerates each time when $\nu$ is chosen with probability $s(\nu)$. For more details, we refer to Nummelin (1984).

We introduce the split chain $\{(X_t, Z_t), t \geq 0\}$, where the auxiliary chain $\{Z_t\}$ only takes the values 0 and 1. Given $X_t = x$, $Z_{t-1} = z_{t-1}$, $Z_t$ takes the value 1 with probability $s(x)$. The distribution of $\{(X_t, Z_t), t \geq 0\}$ is determined by its initial distribution $\lambda$, the transition probability $P$ and $(s, \nu)$. We use $P_\lambda$ and $E_\lambda$ for the distribution and expectation of the Markov chain with initial distribution $\lambda$. When $\lambda = \delta_x$ we write $P_x$ in stead of $P_{\delta_x}$, which is the conditional distribution of $(Z_0, \{(X_t, Z_t), t \geq 1\})$ given $X_0 = x$. When $\lambda = \delta_\alpha(x, y)$, i.e., $X_0 = x$ for arbitrary $x \in E$ and $Z_0 = 1$, then we write $P_\alpha$ and $E_\alpha$.

As shown in Karlsen and Tjøstheim (2001), if we let

$$\pi_s = \nu G_{s,\nu}, \text{ where } G_{s,\nu} = \sum_{n=0}^{\infty} (P - s \otimes \nu)^n,$$

then $\pi_s = \pi_s P$, which implies that $\pi_s$ is an invariant measure.

We then give some definitions of the stopping times of the Markov chain. Let

$$\tau = \tau_\alpha = \min\{t \geq 0 : Z_t = 1\}$$

and

$$S_\alpha = \min\{t \geq 1 : Z_t = 1\}. \quad \text{(B.4)}$$

As $\{(X_t, Z_t), t \geq 0\}$ is Harris recurrent, $P_\alpha(S_\alpha < \infty) = 1$. Moreover, define

$$\tau_k = \begin{cases} 
\inf\{t \geq 0 : Z_t = 1\}, & k = 0, \\
\inf\{t > \tau_{k-1} : Z_t = 1\}, & k \geq 1,
\end{cases} \quad \text{(B.5)}$$

and denote the total number of regenerations in the time interval $[0, T]$ by $N(T)$, that is,

$$N(T) = \begin{cases} 
\max\{t : \tau_t \leq T\}, & \text{if } \tau_0 \leq T, \\
0, & \text{otherwise},
\end{cases} \quad \text{(B.6)}$$

where $\tau_0$ is defined as follows:

$$P_x(\tau_0 = t) = (P - s \otimes \nu)^t s(x), \quad t \geq 0,$$

$$P_\alpha(S_\alpha = t) = \nu(P - s \otimes \nu)^{t-1} s, \quad t \geq 1.$$
It is noted that $N(T)$ has the same properties as $T(n)$ defined in Karlsen and Tjøstheim (2001). As in Lemma 3.3 of Karlsen and Tjøstheim (2001), we define

$$\tilde{N}(T) = \sum_{t=0}^{T} Z_t \quad \text{and} \quad N(T) = \left(\tilde{N}(T) - 1\right) I\left(\tilde{N}(T) > 0\right).$$

(B.7)

It also follows from Lemma 3.4 of Karlsen and Tjøstheim (2001) that the following inequality

$$T^{\frac{1}{2} - \frac{\delta_0}{8}} \leq N(T) \leq T^{\frac{1}{2} + \frac{\delta_0}{8}}$$

(B.8)

holds almost surely for all $0 < \delta_0 < \frac{1}{5}$.

Note that $N(T)$ is independent of $\{e_t\}$ by construction when $\{X_s\}$ and $\{e_t\}$ are assumed to be independent in the model $Y_t = m(X_t) + e_t$.

### 3 Appendix C

This appendix provides a useful lemma. The lemma is concerned with uniform strong convergence of nonparametric kernel density estimate of a sequence of nonstationary time series of the form $X_t = X_{t-1} + u_t$.

In the stationary time series case, Hansen (2008) establishes uniform strong convergence with rates for both nonparametric density and regression estimates.

Recall the definition of $N(T)$ from Appendix B and define

$$\hat{f}(x) = \hat{f}_s(x) = \frac{1}{N(T)h} \sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right).$$

(C.1)

In Theorem 5.1 of Karlsen and Tjøstheim (2001), the pointwise convergence of $\hat{f}(x)$ to $f(x)$ was established. In the following lemma, uniform strong convergence on $R^1$ is obtained.

**Lemma C.1.** Let Assumption 2.1 hold. Then

$$\sup_{x \in R^1} \left| \hat{f}(x) - 1 \right| = o(1), \quad \text{almost surely (a.s.).}$$

(C.2)

**Proof:** For $0 < \delta_0 < \frac{1}{5}$ as chosen in Assumption 2.2(ii), let $A_T = C \cdot T^{\frac{1}{4\delta_0}}$ for some $C > 0$. In order to prove (C.2), we first show that for $T$ large enough

$$\sup_{|x| \leq A_T} \left| \hat{f}(x) - 1 \right| = o(1), \quad \text{a.s..}$$

(C.3)
For 0 < \delta_0 < \frac{1}{8}, observe the following relationship between events

\[
\begin{align*}
\sup_{|x| \leq A_T} \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) - 1 \to 0 \\
\sup_{|x| \leq A_T} \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) - 1 \to 0,
\end{align*}
\]

by a finite number of subsets \( T_k \delta \), where \( T_k \) is chosen such that \( \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) - 1 \to 0\), \( T_{\frac{1}{2} + \frac{\delta_0}{4}} \leq N(T) \leq T_{\frac{1}{2} + \frac{\delta_0}{4}} \) \( \cup \left\{ \sup_{|x| \leq A_T} \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) - 1 \to 0, \right. \)

\[
\left. N(T) < T_{\frac{1}{2} - \frac{\delta_0}{4}} \ \text{or} \ N(T) > T_{\frac{1}{2} + \frac{\delta_0}{4}} \right\}.
\]  

(C.4)

In view of equation (A.15) and (C.4) above, in order to prove (C.3), it suffices to show that for any \( \eta > 0 \),

\[
P \left\{ \sup_{|x| \leq A_T} \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) - 1 > \eta, \ T_{\frac{1}{2} - \frac{\delta_0}{4}} \leq N(T) \leq T_{\frac{1}{2} + \frac{\delta_0}{4}}, \ i.o. \right\} = 0.
\]  

(C.5)

Let \( S(T) = [-A_T, A_T] \) and \( S_0 = [-1, 1] \). Since \( S_0 \) is a compact set, it can be covered by a finite number of subsets \( \{S_i\} \) centered at \( x_i \) with radius \( T_{\frac{1}{2} - \frac{\delta_0}{8}} \), where \( k \) is chosen such that \( k\delta_0 > 2 + \frac{5}{4}\delta_0 + \frac{1}{2\delta_0} \). Denote \( \Pi_T \) the number of such sets covering \( S(T) \), then

\[
\Pi_T = C \cdot \left( T_{\frac{1}{2} + \frac{\delta_0}{4}} \left( T_{\frac{1}{2} - \frac{\delta_0}{4}} \right)^{2k-1} \right)
\] 

for some \( C > 0 \). Thus, we have

\[
\sup_{|x| \leq A_T} \left| \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) - 1 \right| \leq \max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x_j}{h} \right) - 1 \right| \]

\[
+ \max_{1 \leq j \leq \Pi_T} \sup_{x \in S_j} \frac{1}{N(T)h} \sum_{t=1}^{T} \left| K \left( \frac{X_t - x_j}{h} \right) - K \left( \frac{X_t - x}{h} \right) \right|
\]

\[
\leq \max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x_j}{h} \right) - 1 \right| \]

\[
+ \max_{1 \leq j \leq \Pi_T} \frac{1}{N(T)h} \sum_{t=1}^{T} \Psi \left( \frac{X_t - x_j}{h} \right) \max_{1 \leq j \leq \Pi_T} \sup_{x \in S_j} \left| x_j - x \right|
\].

Since \( k\delta_0 > 2 + \frac{5}{4}\delta_0 + \frac{1}{2\delta_0} \), equation (A.15) and Assumption 2.1(iv) imply

\[
\max_{1 \leq j \leq \Pi_T} \frac{1}{N(T)h} \sum_{t=1}^{T} \Psi \left( \frac{X_t - x_j}{h} \right) \max_{1 \leq j \leq \Pi_T} \sup_{x \in S_j} \left| x_j - x \right|
\]

28
\[
\leq O \left( \Pi_T^{-1} T \left( T^{\frac{1}{2}} - \frac{8}{3} h \right)^{-1} h^{-2} \right)
\]
\[
\leq O \left( T^{-\frac{1}{4} \delta_0} \left( T^{\frac{1}{2}} - \frac{8}{3} h \right)^{-2} T^{\frac{1}{2} + \frac{8}{3} h} h^{-2} \right)
\]
\[
= O \left( T^{-\frac{1}{4} \delta_0} \left( T^{\frac{1}{2}} - \frac{8}{3} h \right)^{-2} T^{\frac{1}{2} - \frac{8}{3} \delta_0} \right)
\]
\[
= O \left( T^{-\frac{1}{2} + \frac{2}{3} \delta_0} + \frac{8}{3} h \right)
\]
\[
= o \left( T^{-\frac{1}{2} + \frac{2}{3} \delta_0} + \frac{8}{3} h \right) = o(1), \ a.s..
\]

Thus, in order to prove (C.5), it suffices to show that for \( T \) large enough
\[
P \left\{ \max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T) h} \sum_{t=1}^{T} K \left( \frac{X_t - x_j}{h} \right) - 1 \right| > \frac{\eta}{2} \right\} = o(1), \ a.s.
\]
\[
\leq O \left( T^{-\frac{1}{4} \delta_0} \left( T^{\frac{1}{2}} - \frac{8}{3} h \right)^{-2} T^{\frac{1}{2} + \frac{8}{3} h} h^{-2} \right)
\]
\[
= O \left( T^{-\frac{1}{4} \delta_0} \left( T^{\frac{1}{2}} - \frac{8}{3} h \right)^{-2} T^{\frac{1}{2} - \frac{8}{3} \delta_0} \right)
\]
\[
= O \left( T^{-\frac{1}{2} + \frac{2}{3} \delta_0} + \frac{8}{3} h \right)
\]
\[
= o \left( T^{-\frac{1}{2} + \frac{2}{3} \delta_0} + \frac{8}{3} h \right) = o(1), \ a.s.
\]

Similarly to equation (3.3) of Karlsen, Myklebust and Tjøstheim (2007), we define
\[
Z_{k,j} = \begin{cases} 
\sum_{i=0}^{n_0} \frac{1}{h} K \left( \frac{X_i - x_j}{h} \right), & k = 0, \\
\sum_{i=k-1+1}^{n_0} \frac{1}{h} K \left( \frac{X_i - x_j}{h} \right), & k \geq 1, \\
\sum_{i=n(T)+1}^{T-1} \frac{1}{h} K \left( \frac{X_i - x_j}{h} \right), & k = (T).
\end{cases}
\]

Then for all \( j \geq 1 \),
\[
\sum_{s=1}^{T} \frac{1}{h} K \left( \frac{X_s - x_j}{h} \right) = Z_{0,j} + \sum_{k=1}^{N(T)} Z_{k,j} + Z_{(T),j}.
\]

It follows from Nummelin (1984) that \( \{Z_{k,j}, \ k \geq 1\} \) is a sequence of i.i.d. random variables for fixed \( j \). Since \( \{X_t\} \) is a random walk process, moreover, we have
\[
\mu(K_{j,h}) \equiv E [Z_{k,j}] = \int K(u) du + o(1) = 1 + o(1)
\]
using (5.6) of Lemma 5.1 of Karlsen and Tjøstheim (2001).

In view of (C.8)–(C.10), in order to prove (C.7), it suffices to show that
\[
\max_{1 \leq j \leq N(T)} \frac{1}{N(T)} \sum_{k=1}^{N(T)} (Z_{k,j} - E [Z_{k,j}]) \rightarrow 0, \ a.s.
\]
\[
\max_{1 \leq j \leq N(T)} \frac{1}{N(T)} Z_{0,j} \rightarrow 0, \ a.s.
\]
\[
\max_{1 \leq j \leq N(T)} \frac{1}{N(T)} Z_{(T),j} \rightarrow 0, \ a.s..
\]
We first prove (C.11) using Bernstein’s inequality and the truncation method. By Lemma D.1 in Appendix D below, we have

$$\max_j E \left[ Z_{k,j}^{2p} \right] \leq C \ h^{-2p+1}$$  \hfill (C.14)

for some constant $0 < C < \infty$ (independent of $T$) and any $p \geq 1$.

For the chosen $0 < \delta_0 < \frac{1}{4}$, define

$$Z_{k,j} = Z_{k,j} I \left( |Z_{k,j}| \leq T^{1/2 - \delta_0} \right) \text{ and } \tilde{Z}_{k,j} = Z_{k,j} - Z_{k,j}.$$

Taking $p = k$ in (C.14) and choosing $k$ such that $k\delta_0 > 2 + \frac{5}{4} \delta_0 + \frac{1}{2\delta_0}$, we have for $T_0$ large enough

$$\sum_{T=T_0}^{\infty} \Pi_T \sum_{j=1}^{\Pi_T} P \left\{ \max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T)} \sum_{k=1}^{N(T)} \left( \tilde{Z}_{k,j} - E \left[ \tilde{Z}_{k,j} \right] \right) \right| > \frac{\eta}{2}, \ T^{1/2 - \delta_0} \leq N(T) \leq T^{1/2 + \delta_0} \right\}$$

$$\leq \sum_{T=T_0}^{\infty} \Pi_T \sum_{j=1}^{\Pi_T} P \left\{ \frac{1}{N(T)} \sum_{k=1}^{N(T)} \left( |\tilde{Z}_{k,j}| + E \left| \tilde{Z}_{k,j} \right| \right) > \frac{\eta}{2}, \ T^{1/2 - \delta_0} \leq N(T) \leq T^{1/2 + \delta_0} \right\}$$

$$\leq \sum_{T=T_0}^{\infty} \Pi_T \sum_{j=1}^{\Pi_T} P \left\{ \max_{1 \leq k \leq N(T)} \left( |\tilde{Z}_{k,j}| + E \left| \tilde{Z}_{k,j} \right| \right) > \frac{\eta}{2}, \ T^{1/2 - \delta_0} \leq N(T) \leq T^{1/2 + \delta_0} \right\}$$

$$\leq \sum_{T=T_0}^{\infty} \Pi_T \sum_{j=1}^{\Pi_T} P \left\{ \max_{1 \leq k \leq T^{1/2 + \delta_0 / 8}} \left( |\tilde{Z}_{k,j}| + E \left| \tilde{Z}_{k,j} \right| \right) > \frac{\eta}{2} \right\}$$  \hfill (C.15)

$$\leq c_1 \sum_{T=T_0}^{\infty} T^{1/2 + \delta_0 / 8} \sum_{j=1}^{\Pi_T} P \left\{ |Z_{1,j}| > c_2 \ T^{1/2 - \delta_0 / 4} \right\}$$

$$\leq c_3 \sum_{T=T_0}^{\infty} T^{1/2 + \delta_0 / 8} \sum_{j=1}^{\Pi_T} E \left| Z_{1,j} \right|^{2k} \ T^{-2k(1/2 - \delta_0 / 4)}$$

$$\leq c_4 \sum_{T=T_0}^{\infty} \Pi_T \ T^{1/2 + \delta_0 / 8} \ h^{1-2k} \ T^{-2k(1/2 - \delta_0 / 4)}$$

$$\leq c_5 \sum_{T=T_0}^{\infty} \left( T^{1/2 \ 2k - \delta_0 / 2} \right) \ h^{2k-1} \ T^{1/2 + \delta_0 / 8} \ h^{1-2k} \ T^{-2k(1/2 - \delta_0 / 4)}$$

$$\leq c_6 \sum_{T=T_0}^{\infty} \ T^{-\left( \frac{1}{2} \delta_0 - \frac{5}{8} \delta_0 - \frac{1}{4 \delta_0} \right)} < \infty$$

using the fact that $\{Z_{k,j}\}$ is a sequence of i.i.d. random variables for each fixed $j$, where $c_i$ for $1 \leq i \leq 6$ are all positive constants.
To deal with $Z_{k,j}$, for $T_0$ large enough we apply Bernstein’s inequality as follows:

$$\sum_{T=T_0}^{\infty} P \left\{ \max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T)} \sum_{k=1}^{N(T)} (Z_{k,j} - E[Z_{k,j}]) \right| > \frac{\eta}{2}, T^{\frac{1}{2} - \frac{40}{8}} \leq N(T) \leq T^{\frac{1}{2} + \frac{40}{8}} \right\}$$

$$\leq C_0 \sum_{T=T_0}^{\infty} \sum_{j=1}^{\Pi_T} \sum_{l=C_1 T^{\frac{1}{2} - \frac{40}{8}}} \left| \frac{1}{l} \sum_{k=1}^{l} \left| \frac{Z_{k,j} - E[Z_{k,j}]}{N(T)} \right| > \frac{\eta}{2} \right\}$$

$$\leq C_3 \sum_{T=T_0}^{\infty} \sum_{j=1}^{\Pi_T} \sum_{l=C_1 T^{\frac{1}{2} - \frac{40}{8}}} \exp \left\{ -l C_4 T^{-\frac{1}{2} + \frac{40}{8}} \right\}$$

$$\leq C_5 \sum_{T=T_0}^{\infty} \sum_{j=1}^{\Pi_T} T^{\frac{1}{2} + \frac{40}{8}} \exp \left\{ -C_6 T^{\frac{80}{8}} \right\} < \infty$$

using $\Pi_T = C_7 \left( T^{\frac{1}{80}} \left( T^{\frac{1}{2} - \frac{40}{8}} - h \right)^{2k-1} \right)$, where $C_i$ for $0 \leq i \leq 7$ are all positive constants.

The above equations (C.15) and (C.16) as well as Borel–Cantelli Lemma imply that (C.11) is proved.

Similarly, using (C.14) and (A.15), we can show that as $T \to \infty$

$$\max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T)} |Z_{0,j}| \right| = o(1) \text{ a.s.} \quad \text{(C.17)}$$

$$\max_{1 \leq j \leq \Pi_T} \left| \frac{1}{N(T)} |Z_{T,j}| \right| = o(1) \text{ a.s.} \quad \text{(C.18)}$$

By (C.11)–(C.18), the proof of (C.3) is completed. We now come back to prove (C.2).

Let

$$\tilde{f}(x) = \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) I[|X_t| \leq A_T]$$

$$\bar{f}(x) = \frac{1}{N(T)h} \sum_{t=1}^{T} K \left( \frac{X_t - x}{h} \right) I[|X_t| > A_T].$$

In view of (C.3), in order to prove (C.2), it suffices to show that

$$\sup_{|x| > 2A_T} \left| \tilde{f}(x) - E[\tilde{f}(x)] \right| = o(1), \text{ a.s.,} \quad \text{(C.20)}$$

$$\sup_{|x| \leq A_T} \left| \bar{f}(x) - E[\bar{f}(x)] \right| = o(1), \text{ a.s.,} \quad \text{(C.21)}$$

$$\sup_{A_T < |x| \leq 2A_T} \left| \tilde{f}(x) - E[\tilde{f}(x)] \right| = o(1), \text{ a.s.,} \quad \text{(C.22)}$$

$$\sup_{x \in R^k} \left| \bar{f}(x) - E[\bar{f}(x)] \right| = o(1), \text{ a.s..} \quad \text{(C.23)}$$
Since the assumptions of $|x| > 2A_T$ and $|X_t| \leq A_T$ imply $|X_t - x| \geq A_T$, the fact that $K(\cdot)$ has compact support implies that the following equations hold almost surely:

\[
\sup_{|x| > 2A_T} \left| \tilde{f}(x) \right| \leq \frac{1}{N(T)h} \sum_{t=1}^{T} \sup_{|X_t-x| \geq A_T} K \left( \frac{X_t-x}{h} \right) = o(1), \quad \text{(C.24)}
\]

\[
\sup_{|x| > 2A_T} \left| E \left( \tilde{f}(x) \right) \right| \leq E \left( \sup_{|x| > 2A_T} \left| \tilde{f}(x) \right| \right) = o(1),
\]

\[
\sup_{|x| > 2A_T} \left| \bar{f}(x) - E \left[ \tilde{f}(x) \right] \right| \leq \sup_{|x| > 2A_T} \left| \bar{f}(x) \right| + \sup_{|x| > 2A_T} \left| E \left( \tilde{f}(x) \right) \right| = o(1).
\]

This completes the proof of (C.20). The proof of (C.21) follows from that of (C.3). The proof of (C.22) follows from

\[
\sup_{A_T < |x| \leq 2A_T} \left| \tilde{f}(x) - E \left[ \tilde{f}(x) \right] \right| \leq \sup_{|x| \leq 2A_T} \left| \tilde{f}(x) - E \left[ \tilde{f}(x) \right] \right| = o(1), \quad \text{a.s.,} \quad \text{(C.25)}
\]

which follows from the proof of (C.3).

To prove (C.23), in view of (C.7) and using the boundedness of $K(\cdot)$, we have that for any given $\delta > 0$,

\[
P \left( \sup_{x \in \mathbb{R}^2} |\tilde{f}(x)| > \frac{\delta}{2} \right) \leq P \left( \sup_{x \in \mathbb{R}^2} |\tilde{f}(x)| > \frac{1}{2}\delta, \quad T^{\frac{1}{2} - \frac{\delta}{8}} \leq N(T) \leq T^{\frac{1}{2} + \frac{\delta}{8}} \right)
\]

\[
+ \quad P \left( N(T) < T^{\frac{1}{2} - \frac{\delta}{8}} \text{ or } N(T) > T^{\frac{1}{2} + \frac{\delta}{8}} \right)
\]

\[
\leq \frac{2}{\delta} E \left[ \frac{1}{T^{\frac{1}{2} - \frac{\delta}{8}} h} \sum_{t=1}^{T} K \left( \frac{X_t-x}{h} \right) 1[|X_t| > A_T] \right] + o(1)
\]

\[
= \frac{2}{\delta} \frac{1}{T^{\frac{1}{2} - \frac{\delta}{8}} h} E \left[ \sum_{x \in \mathbb{R}^2} K \left( \frac{X_t-x}{h} \right) 1[|X_t| > A_T] \right] + o(1)
\]

\[
\leq \frac{2}{\delta} \frac{1}{T^{\frac{1}{2} - \frac{\delta}{8}} h} \sum_{t=1}^{T} E \left[ K \left( \frac{X_t-x}{h} \right) 1[|X_t| > A_T] \right] + o(1)
\]

\[
\leq \frac{C}{T^{\frac{1}{2} - \frac{\delta}{8}} h} \sum_{t=1}^{T} E[|X_t| > A_T] + o(1)
\]

\[
= \frac{C}{T^{\frac{1}{2} - \frac{\delta}{8}} h} \sum_{t=1}^{T} P(|X_t| > A_T) + o(1)
\]

\[
\leq \frac{C}{A_T^2 T^{\frac{1}{2} - \frac{\delta}{8}} h} \sum_{t=1}^{T} E[X_t^2] + o(1) \leq C \frac{T^2}{A_T^2 T^{\frac{1}{2} - \frac{\delta}{8}} h} + o(1)
\]

\[
\leq O \left( \frac{1}{T^{\frac{1}{2} - \frac{\delta}{8}} h} \right) + o(1) \leq O \left( \frac{1}{T^{\frac{1}{2} - \frac{\delta}{8}} h} \right) + o(1) = o(1)
\]

32
using Assumption 2.2(ii) and the fact that $\sum_{t=1}^{T} E [X_t^2] = O (T^2)$ under $X_t = X_{t-1} + u_t$, where $C > 0$ is some constant and $0 < \delta_0 < \frac{1}{2}$ is as chosen in Assumption 2.2(ii).

Finally, using the same arguments as in (C.26) we have
\[
\sup_{x \in \mathbb{R}^1} E \left[ f(x) \right] \leq E \left( \sup_{x \in \mathbb{R}^1} f(x) \cdot I \left[ T^{\frac{1}{2}} - \frac{\delta_0}{\delta} \leq N(T) \leq T^{\frac{1}{2}} + \frac{\delta_0}{\delta} \right] \right) + E \left( \sup_{x \in \mathbb{R}^1} f(x) \cdot I \left[ N(T) < T^{\frac{1}{2}} - \frac{\delta_0}{\delta} \text{ or } N(T) > T^{\frac{1}{2}} + \frac{\delta_0}{\delta} \right] \right) = o(1).
\]

This completes the proof of (C.23) and therefore that of Lemma C.1.

4 Appendix D

This appendix D shows that equation (C.14) above can be proved using the same arguments as in the proof of Lemma 5.2 of KT (2001).

**Lemma D.1.** Let the conditions of their Lemma 5.2 of KT (2001) hold. Then
\[
\sup_{x} E \left[ \left| U^{2m}(g_h) \right| \right] \leq d_m h^{-2m+1}
\]
for some sequence $d_m > 0$.

**Proof:** The main issue is to deal with the inequalities in the middle of page 404. Note that in our case the function $\xi_0 \equiv 1$, so that $c_2$ is independent of $x$. Similarly,
\[
c_1 K_{x,h}^{l_i}(y) = \frac{c_1}{h^{l_i}} K_{x,h}^{l_i} \left( \frac{x-y}{h} \right) \leq c_2 \frac{1}{h^{l_i}} I_{N_x}(y),
\]
where $c_2 = c_1 (\sup |K|)^{l_i}$ and $I_{N_x}(y) = I(|y - x| \leq h)$.

By definition of $G_{s,\nu}$ of (3.6) and (3.8) of the KT paper,
\[
G_{s,\nu} I_{N_x}(y) = E_y \left( \sum_{n=0}^{\tau} I_{N_x}(X_n) \right).
\]

From Remark 5.1, $\sup_y E_y (\sum_{n=0}^{\tau} I_{N_x}(X_n)) \leq M_x$. We thus need to show that there is an absolute constant $M$ such that $\sup_x |M_x| \leq M$. To do so, we consider only the random walk...
case, since this is the case we have considered in the main parts of the current paper. By symmetry
\[ E_y \left( \sum_{n=0}^{\tau} \mathcal{I}_{\mathcal{N}_x}(X_n) \right) \leq E_x \left( \sum_{n=0}^{\tau} \mathcal{I}_{\mathcal{N}_x}(X_n) \right) = E_0 \left( \sum_{n=0}^{\tau} \mathcal{I}_{\mathcal{N}_x}(X_n) \right), \]
which is independent of \( x \), where \( E_y \) means that the expectation is taken with the initial condition \( X_0 = y \).

It then follows trivially that there is some \( 0 < M < \infty \) such that \( \sup_x c_3(x) \leq M \) involved in the inequality \( G_{s,v} I_{|\gamma|} \), on page 404 of the KT paper. Repeating this argument as in the KT (2001) paper,
\[ E_{\nu} J_{k,l} \leq c_4 h^{- \sum_{i=2}^{k} l_i} \pi_s K_{x,h}^{l_i}, \]
where \( c_4 \) is independent of \( x \) by above reasoning, and where
\[ \pi_s K_{x,h}^{l_i} = \int \frac{1}{h} K \left( \frac{x-y}{h} \right) p_s(y)dy \leq \frac{1}{h} \sup_y |p_s(y)| \cdot \int I_{\mathcal{N}_x}(y)dy \equiv \frac{c_5 h}{h} \]
\[ = c_5 h^{l_i+1} \]
independent of \( x \). Along with the same lines as in the proof of Lemma 5.2 of KT (2001), this means that we have strengthened Lemma 5.2 of KT (2001) to
\[ \sup_x E \left[ \left| U^{2m}(|\gamma|) \right| \right] \leq d_m h^{-2m+1}. \quad \text{(D.1)} \]

This finally completes an outline of the proof of Lemma D.1.

References


