*Econometric Theory*, **0**, 2026, 1–53.
doi:10.1017/S0266466626100358

# NEW ASYMPTOTICS APPLIED TO FUNCTIONAL COEFFICIENT REGRESSION AND CLIMATE SENSITIVITY ANALYSIS

QIYING WANG
*The University of Sydney*

PETER C. B. PHILLIPS
*Yale University, University of Auckland, and Singapore Management University*

YING WANG
*Renmin University of China*

A general asymptotic theory is established for sample cross moments of nonstationary time series, allowing for long-range dependence and local unit roots. The theory provides a substantial extension of earlier results on nonparametric regression that include near-cointegrated nonparametric regression as well as spurious nonparametric regression. Many new models are covered by the limit theory, among which are functional coefficient regressions in which both regressors and the functional covariate are nonstationary. Simulations show finite sample performance matching well with the asymptotic theory and having broad relevance to applications, while revealing how dual nonstationarity in regressors and covariates raises sensitivity to bandwidth choice and the impact of dimensionality in nonparametric regression. An empirical example is provided involving climate data regression to assess Earth's climate sensitivity to $CO_2$, where nonstationarity is a prominent feature of both the regressors and covariates in the model. To our knowledge, this application is the first nonparametric empirical analysis to assess potential nonlinear impacts of $CO_2$ on Earth's climate while allowing for nonstationarity in both the regressors and covariates.

## 1. INTRODUCTION

In time-series econometrics, a key element in the asymptotics for nonstationary regression was the development of limit theory for the sample moments and cross moments of nonstationary processes. This development was largely completed by the mid 1990s enabling a full understanding of estimation and testing in linear

The authors thank the Editor Yixiao Sun and two referees for many excellent comments and suggestions that have significantly improved the article. Address correspondence to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, Australia, e-mail: qiying.wang@sydney.edu.au.

© The Author(s), 2026. Published by Cambridge University Press. 

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

regression with integrated and fractionally integrated time series, covering both cointegrated and spurious model specifications. A much wider class of econometric applications involves nonparametric and semiparametric regressions. But these models present more substantial difficulties in the development of asymptotic theory. Early foundational research addressing some of these complexities appeared in Park and Phillips (1999, 2000, 2001), Karlsen and Tjøstheim (2001), and Pötscher (2004); and later work on nonparametric cointegrated regression includes Wang and Phillips (2009a; hereafter, WP2009a), Wang and Phillips (2009b; hereafter, WP2009b), Wang and Phillips (2016), Phillips (2009), Phillips, Li, and Gao (2017), and Wang, Phillips, and Kasparis (2021).

To encompass a wide class of econometric applications, a long desired and necessary objective in the development of nonparametric regression asymptotics is a general local limit theory for statistics that take the following sample cross moment form:

$$S_n = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk}), \tag{1.1}$$

involving two nonstationary standardized arrays $X_{nk}$ and $Y_{nk}$, a continuous function $g(\cdot)$, and a function $f(\cdot)$ that includes kernels of the type commonly used in nonparametrics. Sample moments, such as (1.1), play a central role in a large proportion of econometric estimation and inference. An important aspect of the generality in (1.1) is the presence of two numerical sequences: a sample size $n \to \infty$ and a bandwidth-related sequence $c_n \to \infty$ for which $c_n/n \to 0$. These features are essential for full asymptotic analysis of the sample covariance functional in (1.1). The arrays $X_{nk}$ and $Y_{nk}$ that appear in this functional are suitably standardized time series for which the weak convergence

$$\left(Y_{n,[nt]}, X_{n,[nt]}\right) \Rightarrow (Y_t, X_t) \tag{1.2}$$

holds in the Skorokhod space $D_{\mathbb{R}^2}[0,1]$.

Asymptotic theory of $S_n$ in this general form was first considered in Phillips (2009), where the theory essentially required independence between the time series $X_{nk}$ and $Y_{nk}$, so there was no linkage between the variables. That work was primarily relevant to the study of spurious nonparametric regression and the analysis therein extended earlier work on linear spurious regression (Phillips, 1986), showing that all of the main asymptotic features of spurious regression coefficients, tests, and diagnostics carried over to the nonparametric case upon adjustment of convergence and divergence rates to take account of the nonparametric nature of the regression. In more recent work, Wang et al. (2021) investigated the asymptotics of $S_n$ allowing for an explicit linkage between the variables. Their work considered the situation where $X_{nk}$ is close to being linearly cointegrated with $Y_{nk}$ with an asymptotically constant coefficient and a stationary shift subject to an asymptotically negligible error.

The present article contributes to this past body of research by providing a general framework that encompasses all these models as well as various new

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

models and intermediate formulations, thereby providing a unified theory for use in econometric work involving nonparametric regressions of many different types. Our analysis here assumes that both $X_{nk}$ and $Y_{nk}$ are scalar. Use of scalar $X_{nk}$ is required for technical reasons that are discussed in **WP2009a**, but extensions to vector $Y_{nk}$ follow directly and are therefore not detailed. Further, in addition to (1.1), limit distribution theory in nonparametric regression typically also requires treatment of sample covariance statistics of the form

$$M_n = \left(\frac{c_n}{n}\right)^{1/2} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk})\, u_k, \tag{1.3}$$

where, with respect to an appropriate filtration $\{\mathcal{F}_k\}_{k\geq 1}$, $\mathbb{E}(u_k|\mathcal{F}_{k-1}) = 0$ so that $M_n$ is closely related to a martingale sequence, as discussed in detail in Section 2. The asymptotics for (1.3) are straightforward using Wang's (2014) extended martingale central limit theorem once the asymptotic theory and weak convergence of $S_n$ itself are established. The important element in that limit theorem is its use of weak convergence rather than convergence in probability of the sample moment (1.1). A limit distribution theory for more general versions of $M_n$ that are useful in practical work is given in Theorem 3.2 below. Some regression model implementations of that theory that reveal its generality are given in Section 4.

With results for the covariance functionals (1.1) and (1.3) in hand, the limit theory for nonparametric and semiparametric regression statistics follows in a straightforward manner, as shown later in the article. The limit theory also applies to spurious nonparametric regression, extending Phillips (2009), and various semiparametric functional coefficient models for time series (Gao and Phillips, 2013; Sun, Cai, and Li, 2013; Tu and Wang, 2022; Phillips and Wang, 2023) and panel regressions (Phillips and Wang, 2022). In particular, our methods apply to nonlinear cointegrating regressions in which the cointegrating coefficients are allowed to depend on covariates that may themselves be nonstationary. Such models are particularly useful in climate science applications of the type studied later in this article.

The remainder of the article proceeds as follows. Assumptions and preliminary results are given in Section 2. Section 3 gives the main results. Sections 4–6 provide regression applications, simulations, and empirics.[1] Our climate data application is, to the best of our knowledge, the first nonparametric empirical analysis to assess potential nonlinear impacts of $CO_2$ on Earth's climate sensitivity, allowing for nonstationarity in the regressors and the covariates that influence the coefficient responses to the regressors. Section 7 concludes. Proofs are given in Section 8 and Section 9 describes the bootstrap procedure employed in Section 5.

Throughout the article, we use conventional notation except where explicitly mentioned. For $x = (x_{ij})_{1\leq i\leq m,\, 1\leq j\leq k}$, $||x|| = \sum_{i=1}^{m}\sum_{j=1}^{k}|x_{ij}|$. $(n, K)_{\text{seq}} \to \infty$ denotes

[1] MATLAB code and data for the simulations and empirical application are available online under software at https://korora.econ.yale.edu/phillips/.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

that $n \to \infty$ followed by $K \to \infty$. $A^{\mathrm{T}}$ denotes the transpose of a vector or matrix $A$ and $\mathbb{I}$ is an identity matrix. We denote constants by $C, C_1, \ldots$, which may differ at each appearance.

## 2. ASSUMPTIONS AND PRELIMINARIES

Let $\epsilon_j, j \in \mathbb{Z}$, be a sequence of iid random variables with $\mathbb{E}\epsilon_0 = 0$ and $\mathbb{E}\epsilon_0^2 = 1$ and $\xi_j = \sum_{k=0}^{\infty} \phi_k \epsilon_{j-k}, j \geq 1$, be a linear process whose coefficients $\phi_k, k \geq 0$, satisfy one of the following (long memory or short memory) conditions:

**LM.** $\phi_k \sim k^{-\mu} a(k)$, where $1/2 < \mu < 1$ and $a(k)$ is a function slowly varying at $\infty$.

**SM.** $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

Suppose that $\limsup_{t\to\infty} t^{\delta} |\mathbb{E}e^{it\epsilon_0}| < \infty$ for some $\delta > 0$ throughout this article. This distributional smooth condition on $\epsilon_0$ is required to establish the convergence to local time for a partial sum process of $\xi_j$, as seen in the proof of the main result. To establish asymptotic theory for $S_n$, the following assumptions on $X_{nk}$ and $Y_{nk}$ are employed:

**A1** $X_{nk} = x_k/d_n$, where $x_k = \rho_n x_{k-1} + \xi_k$ with $\rho_n = 1 - \tau n^{-1}$ for some $\tau \geq 0$, $x_0 = o_P(\sqrt{n})$ and $d_n^2 = \mathbb{E}\big(\sum_{k=1}^{n} \xi_k\big)^2$;

**A2** $Y_{nk} = y_k/d_{yn}$, where $d_{yn}^2 = var(y_n) \to \infty$ and, on $D_{\mathbb{R}^3}[0,1]$,

$$\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_{-i},\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i,\ Y_{n,\lfloor nt \rfloor} \right) \Rightarrow \big(B_{-t},\ B_t,\ Y_t\big),$$

where $\{Y_t\}_{t\geq 0}$ is a path continuous Gaussian process, $\{B_t\}_{t\geq 0}$ is a standard Brownian motion, and $\{B_{-t}\}_{t\geq 0}$ is an independent copy of $\{B_t\}_{t\geq 0}$.

**A1** allows for the nearly integrated process $x_k$ to be generated either from short memory (under **SM**) or long memory (under **LM**) innovations, giving a general framework for practical work. It is well-known that

$$d_n^2 = \mathbb{E}|\sum_{k=1}^{n} \xi_k|^2 \ \sim\ \begin{cases} c_\mu\, n^{3-2\mu} a^2(n), & \text{under } \mathbf{LM}, \\ \phi^2\, n, & \text{under } \mathbf{SM}, \end{cases} \tag{2.1}$$

where $c_\mu$ is a constant (see, e.g., Wang, Lin, and Gulati, 2003). **A2** is a minor requirement on $Y_{nk}$ and the innovations $\epsilon_k$ that generate $X_{nk}$, which is close to being necessary for the derivation of asymptotics of $S_n$ when $X_{nk}$ has a linear structure of the type given in **A1**. No restrictions are imposed between the limit processes $Y_t$ and $B_t$ ($B_{-t}$), indicating that many types of potential linkages between $X_{nk}$ and $Y_{nk}$ are allowed so that $Y_{nk}$ has a similar structure to $X_{nk}$. For an explicit example of **A2** that is advantageous in applications, we write $\eta_j = (\eta_{j1}, \ldots, \eta_{jd})^{\mathrm{T}}$ and assume that

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$\lambda_j = (\epsilon_j, \eta_j^{\mathrm{T}}), j \in \mathbb{Z}$, is a sequence of iid random vectors on $\mathbb{R} \times \mathbb{R}^d$ for some integer $d \geq 1$ with $\mathbb{E}\lambda_0 = 0$ and $\mathbb{E}||\lambda_0||^2 < \infty$. Then, consider the following assumption for $Y_{nk}$:

**A2′** $Y_{nk} = y_k/d_{yn}$ with $d_{yn}^2 = var(y_n)$, where $y_k = \sum_{s=1}^{k} \alpha^{\mathrm{T}} v_s$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)^{\mathrm{T}} \in \mathbb{R}^p$, and $v_s = (v_{s1}, \ldots, v_{sp})^{\mathrm{T}}$ is a linear $p$-vector process with $p \leq d$ satisfying the following conditions:

(i) $v_s = \sum_{j=0}^{\infty} \Psi_j \eta_{s-j}$ with $p \times d$ coefficient matrices $\Psi_j$ satisfying $\sum_{j=0}^{\infty} j^{1/2} ||\Psi_j|| < \infty$ with matrix norm $||\cdot||$ and $\Psi := \sum_{j=0}^{\infty} \Psi_j$ of full rank $p$;

(ii) $\Omega := \mathbb{E}\left(\eta_1 \eta_1^{\mathrm{T}}\right)$ is a positive-definite matrix.

As shown in the following proposition, **A2** can be replaced by **A2′**. The structure of $y_k$ in **A2′** allows for a wide combination of multivariate linear processes, which are used in Section 4 for the regression application.

PROPOSITION 2.1. *Under* ***A2′****, on* $D_{\mathbb{R}^3}[0,1]$*, we have* $var(y_n)/n \to \Gamma := \alpha^{\mathrm{T}} \Psi \Omega \Psi^{\mathrm{T}} \alpha$ *and*

$$\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_{-i},\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i,\ Y_{n,\lfloor nt \rfloor} \right) \Rightarrow \left( B_{-t},\ B_t,\ Y_t \right) \tag{2.2}$$

*with* $Y_t = \Gamma^{-1/2} \alpha^{\mathrm{T}} \Psi\, \Lambda_t$*, where* $Z_t := (B_{-t}, B_t, \Lambda_t)$ *is a* $(2+d)$*-dimensional continuous Gaussian process with mean zero, independent increments, and covariance matrix:*

$$\Omega_t = t \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \gamma^{\mathrm{T}} \\ 0 & \gamma & \Omega \end{pmatrix} \quad \textit{with } \gamma = (\gamma_1, \ldots, \gamma_d)^{\mathrm{T}} \textit{ and } \gamma_k = E(\epsilon_0 \eta_{0k}).$$

We further mention that **A2**, together with standard functional limit theory (see Buchmann and Chan, 2007 or Theorem 2.21 of Wang, 2015 with a minor modification), implies the joint convergence

$$\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_{-i},\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i,\ X_{n,\lfloor nt \rfloor},\ Y_{n,\lfloor nt \rfloor} \right) \Rightarrow \left( B_{-t},\ B_t,\ X_t,\ Y_t \right), \tag{2.3}$$

on $D_{\mathbb{R}^4}[0,1]$, where $X_t$ is defined by

$$X_t = \widetilde{F}(t) + \tau \int_0^t e^{-\tau(t-s)} \widetilde{F}(s) ds, \tag{2.4}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

with $\widetilde{F}(t) = \begin{cases} F_{3/2-\mu}(t), & \text{under } \mathbf{LM}, \\ F_{1/2}(t), & \text{under } \mathbf{SM}, \end{cases}$ where $F_H(t)$ is a fractional Brownian motion with representation

$$F_H(t) = \kappa_H \int_{-\infty}^{t} \left[(t-u)_+^{H-1/2} - (-u)_+^{H-1/2}\right] dB_u,$$

where we use the notation $a_+ = \max\{a, 0\}$ and $\kappa_H$ is a standardizing constant so that $\mathbb{E}F_H^2(1) = 1$.

We next introduce the regularity conditions on $g(x), f(x)$, and $u_k$ that are required for the asymptotics of $S_n$ and $M_n$.

**A3** (i) $g(x)$ is a continuous function on $\mathbb{R}$.
(ii) $f(x)$ is a real function on $\mathbb{R}$ satisfying $\int_{-\infty}^{\infty} \big(|f(x)| + f^2(x)\big)dx < \infty$.

**A4** $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference with $\sigma_k^2 = \mathbb{E}\big(u_k^2 \mid \mathcal{F}_{k-1}\big) \to_{a.s.} \sigma^2 > 0$ as $k \to \infty$, and $\sup_{k\geq 1} \mathbb{E}\big[|u_k|^2 I(|u_k| \geq K) \mid \mathcal{F}_{k-1}\big] < \infty$, as $K \to \infty$, where

$$\mathcal{F}_k = \sigma\left(u_k, u_{k-1}, \ldots; (x_{k+1}, y_{k+1}), (x_k, y_k), \ldots\right)$$

is a $\sigma$-field so that $x_k$ and $y_k$ are adapted to $\mathcal{F}_{k-1}$.

**A3** (ii) is quite weak and close to being necessary for the asymptotic theory of $S_n$ and $M_n$ given in the next section, which was originally used in WP2009a in the case where $g(x) = 1$. **A4** ensures that $M_n = M_{nn}$, where $M_{nk} = \left(\frac{c_n}{n}\right)^{1/2} \sum_{j=1}^{k} g(Y_{nj}) f(c_n X_{nj})\, u_j, k \geq 1$, is a martingale with respect to the natural filtration $\mathcal{F} = \{\mathcal{F}_k\}_{k\geq 1}$, which is standard in the literature. When the $y_k$ given in **A2′** is used in $M_n$, we may take $\mathcal{F}_k = \sigma\left(u_k, u_{k-1}, \ldots; \lambda_{k+1}, \lambda_k, \ldots\right)$. It should be mentioned that **A4** essentially assumes that $\sigma_k^2 = \mathbb{E}\big(u_k^2 \mid \mathcal{F}_{k-1}\big)$ is asymptotically constant, which is convenient for the development of our main results. Note that $M_n = \left(\frac{c_n}{n}\right)^{1/2} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk})\, \sigma_k\, \hat{u}_k$, where $\hat{u}_k = \sigma_k^{-1} u_k$ satisfies that $\mathbb{E}\big(\hat{u}_k^2 \mid \mathcal{F}_{k-1}\big) = 1$. In regression applications, when $\sigma_k^2$ does not converge to a constant, we only need to replace the asymptotics of $S_n = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk})$ by that of $S_{1n} = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk}) \sigma_k$, as is similarly considered in Corollary 3.1 below. We remark that the conditional variances $\sigma_k$ are generally unknown and must be estimated (by $\widehat{\sigma}_k$, say), which then requires corresponding asymptotics of the statistic $\widehat{S}_{1n} = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk}) \widehat{\sigma}_k$. If $\widehat{\sigma}_k$ is suitably chosen so that $\widehat{\sigma}_k \in \mathcal{F}_k$, then Corollary 3.1 remains useful in establishing asymptotics for $\widehat{S}_{1n}$. The details are left for future work.

We finally mention that the diffusion $X_t$ defined by (2.4) is a fractional Ornstein–Uhlenbeck process having continuous local time $L_X(s,x)$ defined by

$$L_X(s,x) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^s I(|X_t - x| \leq \epsilon) dt, \quad a.s.$$

See Chapter 2.1.1 of Wang (2015), for instance. Since $L_X(s,x)$ is positive and increasing with respect to $s$, when $g(x)$ is a continuous function (i.e., **A3** (i) holds),

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$g(Y_s)$ is a continuous process and the following integral (Lebesgue–Stieltjes) exists:

$$S(x) := \int_0^1 g(Y_s) L_X(ds, x). \tag{2.5}$$

Moreover, from the continuity of $L_X(t,x)$, $S(x)$ is continuous and bounded almost surely. Hence, by dominated convergence for any $a(x)$ satisfying $\int_{-\infty}^{\infty} |a(x)|dx < \infty$, we have

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} a(x)\, S(\epsilon x) dx = S(0) \int_{-\infty}^{\infty} a(x) dx, \quad a.s. \tag{2.6}$$

The random integral $S(x)$ is involved in the limit distributions of $S_n$ and $M_n$, and (2.6) will be used in proofs of the main results without further indication.

## 3. MAIN RESULTS

We start with the following theorem and some extensions, which are subsequently used in the asymptotic analysis of sample covariance statistics that are useful in applications.

THEOREM 3.1. *Suppose **A1–A3** hold. Then, for any $c_n \to \infty$ and $c_n/n \to 0$, we have*

$$S_n = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) f(c_n X_{nk}) \to_D \mathbb{S} = S(0) \int_{-\infty}^{\infty} f(x) dx, \tag{3.1}$$

*and the following joint convergence also holds: on $D_{\mathbb{R}^5}[0,1]$,*

$$\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i,\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_{-i},\ X_{n,\lfloor nt \rfloor},\ Y_{n,\lfloor nt \rfloor},\ S_n \right) \Rightarrow \left( B_t,\ B_{-t},\ X_t,\ Y_t,\ \mathbb{S} \right). \tag{3.2}$$

**Remark 3.1.** When $g(x) = 1$, Theorem 3.1 reduces to Theorem 2.1 of WP2009a. See also Jeganathan (2004). In comparison with the latter papers, our condition on $X_{nk}$ is slightly stronger but more direct for practical applications. As is well known in the literature, the condition $c_n \to \infty$ is necessary for a limit result such as (3.1). Indeed, if $c_n = 1$, the limit distribution of $S_n$ is a standard stochastic integral, namely,

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n} g(Y_{nk}) f(X_{nk}) &\to_D \int_0^1 g(Y_t) f(X_t) dt \\ &= \int_{-\infty}^{\infty} f(s) \int_0^1 g(Y_t) L_X(dt, s) ds \neq S(0) \int_{-\infty}^{\infty} f(x) dx, \end{aligned} \tag{3.3}$$

where we have used the occupation time formula in (3.3). As shown in WP2009a, some simple sequential limit arguments help to reveal the nature of these

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

differences. Start by rewriting the average $S_n$ so that it is indexed by twin sequences $c_m$ and $n$ defining $S_{m,n} = \frac{c_m}{n}\sum_{k=1}^{n} g(Y_{nk})f(c_m X_{nk})$ and noting that $S_{m,n} = S_n$ when $m = n$. If we hold $c_m$ fixed as $n \to \infty$, then standard limit theory leads to the following limit behavior:

$$S_{m,n} \to_D c_m \int_{-\infty}^{\infty} f(c_m x) \int_0^1 g(Y_t) L_X(dt, x)dx$$
$$= \int_{-\infty}^{\infty} f(x) \int_0^1 g(Y_t) L_X(dt, x/c_m)dx := S_{m,\infty}. \tag{3.4}$$

It follows from (2.6) with $a(x) = f(x)$ that $S_{m,\infty} \to_{a.s.} S(0)\int_{-\infty}^{\infty} f(x)dx$ when $c_m \to \infty$ as $m \to \infty$, so that (3.1) may be regarded as a limiting version of (3.4). The goal and the technical difficulty are to turn this sequential argument as $n \to \infty$, followed by $m \to \infty$, into a joint limit argument so that $c_n$ may play an active role as a bandwidth parameter in density estimation and kernel regression.

**Remark 3.2.** The key approach in the proof of Theorem 3.1 depends on the convergence of the following difference:

$$\Delta_n(\epsilon) := S_n - \frac{c_n}{n}\sum_{k=1}^{n} g(Y_{nk}) \int_{-\infty}^{\infty} f\big[c_n(X_{nk} + z\epsilon)\big]\phi(z)dz$$
$$= \frac{c_n}{n}\sum_{k=1}^{n} g(Y_{nk}) \int_{-\infty}^{\infty} \big\{f\big(c_n X_{nk}\big) - f\big[c_n(X_{nk} + z\epsilon)\big]\big\}\phi(z)dz,$$

as $n \to \infty$ followed by $\epsilon \to 0$, where $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. When $g(x) = 1$, $\Delta_n(\epsilon) \to_{\mathbb{P}} 0$ was established in Jeganathan (2004) and later by WP2009a under weak conditions. That result with $g(x) = 1$ cannot be directly used when $g(x) \neq 1$ due to the involvement of $Y_{nk}$. However, in view of the tightness of $\{Y_{n,[nt]}\}_{0\le t\le 1}$ and the continuity of $g(x)$, there exists $m = m_{n,\epsilon} \to \infty$ as $n \to \infty$ followed by $\epsilon \to 0$ so that

$$\max_{1\le j\le n/m} \max_{(j-1)m+1\le k\le jm} |g(Y_{nk}) - g(Y_{n,(j-1)m+1})| = o_P(1).$$

In consequence, letting $T_n = [n/m]$, the difference $\Delta_n(\epsilon)$ can be approximated by the following blocked sequence:

$$\Delta_{n1}(\epsilon) = \frac{c_n}{n}\sum_{j=2}^{T_n} g(Y_{n,(j-1)m+1}) \sum_{k=(j-1)m+1}^{jm} \int_{-\infty}^{\infty} \big\{f\big(c_n X_{nk}\big) - f\big[c_n(X_{nk} + z\epsilon)\big]\big\}\phi(z)dz.$$

When $g(x)$ is bounded, by similar arguments (differing only in details) as those in WP2009a, we have

$$|\Delta_{n1}(\epsilon)| \le \frac{C c_n}{n}\sum_{j=2}^{T_n} \Big| \sum_{k=(j-1)m+1}^{jm} \int_{-\infty}^{\infty} \big\{f\big(c_n X_{nk}\big) - f\big[c_n(X_{nk} + z\epsilon)\big]\big\}\phi(z)dz \Big| \to_P 0,$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

as $n \to \infty$ followed by $\epsilon \to 0$, Hence, $\Delta_n(\epsilon) \to_P 0$ is established by means of the convergence $\Delta_{n1}(\epsilon) \to_P 0$, since the boundedness of $g(x)$ can be removed by standard truncation methods as shown in the final part of the proof of Theorem 3.1.

Theorem 3.1 may be extended by allowing $S_n$ to include certain stationary variables. Indeed, if we define $\lambda_k = (\epsilon_k, \eta_{k1}, \ldots, \eta_{kd})$ as in Section 2, i.e., $\lambda_k, k \in \mathcal{Z}$, is a sequence of iid random vectors on $\mathbb{R} \times \mathbb{R}^d$ for some integer $d \geq 1$, and set

$$S_{1n} = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) f_1(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}),$$

where $m \geq 0$ is an integer and $f_1(\cdot, \cdot)$ is a real-valued function that is well defined in each of its arguments, we have the following corollary.

COROLLARY 3.1. *Suppose **A1**, **A2**, and **A3**(i) hold. Suppose that:*

***A5***. (i) *for any $x \in \mathbb{R}$ and $y \in \mathbb{R}^{(m+1)d+1}$ and for some $\beta > 0$, $|f_1(x,y)| \leq T(x)(1 + ||y||^{\beta})$, where $T(x)$ is a bounded and integrable function;*
(ii) $E||\lambda_0||^{2 \wedge (2\beta)} < \infty$.

*Then, for any $c_n \to \infty$ satisfying $c_n/n \to 0$, we have*

$$\left( \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \epsilon_k,\ \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \epsilon_{-k},\ X_{n,[nt]},\ Y_{n,[nt]},\ S_{1n} \right)$$
$$\Rightarrow \left( B_t,\ B_{-t},\ X_t,\ Y_t,\ \mathbb{S}_1 \right) \tag{3.5}$$

*on $D_{\mathbb{R}^5}[0,1]$, where $\mathbb{S}_1 = S(0) \int_{-\infty}^{\infty} \tilde{f}_1(x)dx$ with $\tilde{f}_1(x) = \mathbb{E} f_1(x, \lambda_m, \ldots, \lambda_0)$.*

**Remark 3.3.** The proof of Corollary 3.1 involves demonstrating negligibility of the residual difference

$$R_n = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) \left[ f_1\left(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}\right) - \tilde{f}_1(c_n X_{nk}) \right].$$

Using similar ideas, a further extension of this result is possible, along the lines of Theorem 4.1 in Wang et al. (2021), where $f_1(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m})$ in (3.5) is replaced by $f_1(c_n X_{nk}, \tilde{\lambda}_k)$ in which $\tilde{\lambda}_k$ is a linear process with innovations $\lambda_k$ and $f_1(x,y)$ satisfies certain smoothness conditions. This extension only involves some additional calculations rather than new ideas, so the details are omitted. Statistics, such as $S_{1n}$, appear in nonlinear cointegrating regression with additional stationary regressors. An example is given in Wang et al. (2021).

**Remark 3.4.** An earlier sequence of papers (Wang and Phillips, 2009a, 2009b, 2024) developed asymptotics of sample covariances, such as $(n/c_n)^{1/2} S_n$ or $(n/c_n)^{1/2} S_{1n}$ when $g(x) = 1$ in the "zero energy" case where $\int_{-\infty}^{\infty} f(x)dx = 0$ or

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$\int_{-\infty}^{\infty}\tilde{f}_1(x)dx = 0$. The proofs in those papers are technical and involve considerable care in the use of the zero energy nature of the functions. However, in the proof of Theorem 3.1 (and similarly for Corollary 3.1), the key idea is a block decomposition of the form as outlined in Remark 3.2 (also see (8.11) in the proof), where the estimation of a particular remainder term (e.g., $S_{1,n\epsilon}$ in the proof) depends on values of $|f(X_{nk})|$ rather than $f(X_{nk})$ due to the involvement of the nonstationary time series $Y_{nk}$, so that the advantages of a zero energy condition in obtaining sharp results are lost. Except for some special situations (e.g., where a martingale structure is imposed) as shown in Theorem 3.2 below, it is unclear at present how to bridge across these different techniques to achieve a completely general analysis of the zero energy case. So this task is left for future work.

### 3.1. Asymptotics for Sample Covariance Statistics

Regression applications often require limit theory for sample covariance statistics $M_n$ defined by (1.3) and $M_{1n}$ of the following form:

$$M_{1n} = \Big(\frac{c_n}{n}\Big)^{1/2}\sum_{k=1}^{n} g(Y_{nk})f_1\big[c_n(X_{nk}+c_n'x),\lambda_k,\ldots,\lambda_{k-m}\big]u_k,$$

where $x$ is fixed and the constant sequence $c_n' \to 0$ or $c_n' = 1$. Using Theorem 3.1 and Corollary 3.1, together with the extended martingale limit theorem given in Wang (2014), we obtain the second main result.

THEOREM 3.2. *If **A1**, **A2′**, **A3**(i), and **A4** hold, and if $f(x)$ is a bounded integrable real function, then*

$$\big(\widetilde{S}_n,\ M_n\big) \to_D \big(\widetilde{\mathbb{S}},\ \sigma\,\widetilde{\mathbb{S}}^{1/2}\,\mathbb{N}\big), \tag{3.6}$$

*for any $c_n \to \infty$ and $c_n/n \to 0$, where*

$$\widetilde{S}_n = \frac{c_n}{n}\sum_{k=1}^{n} g^2(Y_{nk})f^2\big(c_nX_{nk}\big), \quad \widetilde{\mathbb{S}} = \int_0^1 g^2(Y_s)L_X(ds,0)\int_{-\infty}^{\infty} f^2(x)dx$$

*and $\mathbb{N}$ is a standard normal variate independent of $X = \{X_t\}_{t\ge 0}$ and $Y = \{Y_t\}_{t\ge 0}$.*

*Similarly, if in addition to **A1**, **A2′**, **A3**(i), **A4**, and **A5**(i), $E||\lambda_0||^{4\beta+2} < \infty$, and $f_1(x)$ is a bounded integrable real function, then*

$$\big(\widetilde{S}_{1n},\ M_{1n}\big) \to_D \big(\widetilde{\mathbb{S}}_1,\ \sigma\,\widetilde{\mathbb{S}}_1^{1/2}\,\mathbb{N}\big), \tag{3.7}$$

*for any $c_n \to \infty$ and $c_n/n \to 0$, where*

$$\widetilde{S}_{1n} = \frac{c_n}{n}\sum_{k=1}^{n} g^2(Y_{nk})f_1^2\big[c_n(X_{nk}+c_n'x),\lambda_k,\ldots,\lambda_{k-m}\big],$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

*and*

$$\widetilde{\mathbb{S}}_1 = \int_0^1 g^2(Y_s) L_X(ds, x') \int_{-\infty}^{\infty} \mathbb{E} f_1^2(x, \lambda_m, \ldots, \lambda_0) dx, \text{ with } x' = \begin{cases} 0, & \text{if } c_n' \to 0 \\ -x, & \text{if } c_n' = 1. \end{cases}$$

**Remark 3.5.** Under the conditions of Theorem 3.2, if we let $M_n = M_{nn}$ and $M_{1n} = M_{1n,n}$ where, for $k \geq 1, M_{nk} = \left(\frac{c_n}{n}\right)^{1/2} \sum_{j=1}^{k} g(Y_{nj}) f(c_n X_{nj})\, u_j$ and

$$M_{1n,k} = \left(\frac{c_n}{n}\right)^{1/2} \sum_{j=1}^{k} g(Y_{nj}) f_1\left[c_n\,(X_{nj} + c_n' x), \lambda_j, \ldots, \lambda_{j-m}\right] u_j,$$

both $\{M_{nk}\}_{k\geq 1}$ and $\{M_{1n,k}\}_{k\geq 1}$ are martingales with respect to the filtration $\mathcal{F} = \{\mathcal{F}_k\}_{k\geq 1}$ with $\mathcal{F}_k = \sigma(u_k, u_{k-1}, \ldots; \lambda_{k+1}, \lambda_k, \ldots)$. Therefore, Theorem 3.2 may be readily employed in nonparametric or semi-parametric cointegrating regression contexts such as nonparametric predictive regressions, but it seems difficult to extend to the general sample covariance context, at least using presently available techniques. Indeed, even when $u_k = \epsilon_k$ so that $M_n$ and $M_{1n}$ both lead to endogeneity (as in the much simpler framework of WP2009b where $g(x) = 1$) it is unclear at present how to establish the limit theory of $M_n$ and $M_{1n}$ with general $g(x) \neq 1$, as explained in Remark 3.3. Further, the class of functions $f_1(\cdot, \cdot)$ in **A5(i)** or $f(\cdot)$ in **A3 (ii)** is restrictive for parametric regression contexts, where the model may require inclusion of regression functions that fail **A3(ii)** such as $f(x) = x^p$. These limitations indicate that the technical machinery employed in the present article is insufficient for tackling fully general asymptotics in nonlinear cointegration modeling.

**Remark 3.6.** Theorem 3.2 uses **A2′** to assist in proving asymptotic normality but this condition can be relaxed. For instance, Theorem 3.2 still holds only if $y_k$ satisfies **A2** with $y_k = l(\lambda_k, \lambda_{k-1}, \ldots)$, where $l(\ldots)$ is a well-defined measurable function of its components. The proof is similar to that of Theorem 3.2 by verifying the condition of Wang's (2014) extended martingale central limit theorem, and the details are therefore omitted.

## 4. REGRESSION APPLICATIONS

Nonlinear cointegrating regressions with functional coefficients were introduced in Xiao (2009) and Cai, Li, and Park (2009), where the authors suggested a model of the form

$$z_k = l_k^{\mathrm{T}} \beta(x_k) + u_k, \tag{4.1}$$

where $\{x_k, z_k, u_k\}$ are scalars, $l_k = (l_{k1}, \ldots, l_{kp})^{\mathrm{T}}$ is of dimension $p$ to be specified later, $\beta(\cdot) = \left(\beta_0(\cdot), \beta_1(\cdot), \ldots, \beta_{p-1}(\cdot)\right)^{\mathrm{T}}$ is a $p$-vector of unknown smooth functions $\beta_s(\cdot), s = 0, 1, \ldots, p-1$, defined on $\mathbb{R}$, and $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference so that $E\left(z_k | \mathcal{F}_{k-1}\right) = l_k^{\mathrm{T}} \beta(x_k)$. Xiao (2009) dealt with the case where the functional covariate $x_k$ in (4.1) is stationary; Cai et al. (2009) restricted the

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

model to cases where either the functional covariate $x_k$ in (4.1) or the regressor $l_k$ is stationary.

Extensions of the model (4.1) to cases where both $x_k$ and $l_k$ are nonstationary are considerably more complex but also have many relevant empirical applications, such as the climate sensitivity implementation of our methods in Section 6, where nonstationary regressors and covariates can normally be expected. The required limit theory for these extensions is given in the present section.

Some related versions of the model (4.1) can be found in Gao and Phillips (2013), Li et al. (2017), Hirukawa and Sakudo (2018), and Tu and Wang (2019, 2020). Further recent developments can be found in Tu and Wang (2022), Phillips and Wang (2023), Liang, Shen, and Wang (2023), and in a panel regression context by Phillips and Wang (2022). Phillips and Wang (2023) provided a correction to the limit theory in the previous nonstationary kernel regression literature, showing that conventional bias terms can influence both the asymptotic variance and the optimal convergence rate in functional coefficient nonstationary regression.

We develop a general limit theory of estimation and inference in model (4.1) for the local level kernel estimator $\widehat{\beta}_n(x)$ of $\beta(x)$ defined by

$$\begin{aligned}\widehat{\beta}_n(x) &= \arg\min_{\beta} \sum_{k=1}^{n} [z_k - l_k^{\mathrm{T}}\beta]^2 K\Big(\frac{x_k - x}{h}\Big) \\ &= \Big[\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\Big(\frac{x_k - x}{h}\Big)\Big]^{-1} \sum_{k=1}^{n} l_k z_k K\Big(\frac{x_k - x}{h}\Big), \end{aligned} \tag{4.2}$$

where $K(x)$ is a nonnegative real kernel function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$. The results in Section 3 are used here to accommodate dual nonstationarity where both regressors $l_k$ and covariate $x_k$ are nonstationary time series. For the same model and improved estimation of $\beta(x)$, this section also considers local linear estimation and provides asymptotics under additional restrictions between $x_k$ and $l_k$.

To avoid further complexity, except where explicitly mentioned, we employ in what follows the same notation and conditions as in previous sections on $x_k, u_k, \lambda_k$, and $v_k$. Let $x$ be a fixed constant in $\mathbb{R}$ and, in a neighborhood of $x$, consider the following local linear and quadratic conditions on the functional coefficient $\beta(\cdot) = \big(\beta_0(\cdot), \beta_1(\cdot), \ldots, \beta_{p-1}(\cdot)\big)^{\mathrm{T}}$:

**A6**. (i) $||\beta(y+x) - \beta(x) - \beta'(x)y|| \le C_x |y|^{1+\nu}$, or
(ii) $||\beta(y+x) - \beta(x) - \beta'(x)y - \frac{1}{2}\beta''(x)y^2|| \le C_x |y|^{2+\nu}$,
for some $0 < \nu \le 1$, where $C_x$ is a constant depending only on $x$.

We also make use of the following condition on the kernel function $K(\cdot)$ used in (4.2):

**A7**. (i) $K(x)$ is a positive real function with finite support satisfying $\int_{-\infty}^{\infty} K(s)ds = 1$, or

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

(ii) $K(x)$ is a positive real function with finite support satisfying $\int_{-\infty}^{\infty} K(x)dx = 1$ and $\int_{-\infty}^{\infty} xK(x)dx = 0$.

Under these conditions, the next result provides the asymptotic theory of $\widehat{\beta}_n(x)$.

THEOREM 4.1. *Suppose that **A1** and **A4** hold. Suppose $l_k = \sum_{s=1}^{k} v_s$, where $v_s = (v_{s1}, \ldots v_{sp})^{\mathrm{T}}, p \leq d$, is a linear vector process as defined as in **A2′**, so that $l_k$ is full rank $I(1)$. Then, under **A6**(i) and **A7(i)**, we have*

$$h^{-1}\left[\widehat{\beta}_n(x) - \beta(x)\right] \to_P \beta'(x) \int_{-\infty}^{\infty} sK(s)ds, \tag{4.3}$$

*for any $h \to 0$ satisfying $\min\{nh^2, nh/d_n\} \to \infty$. If **A6**(i) is replaced by **A6**(ii), we have*

$$\left(\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\left[(x_k - x)/h\right]\right)^{1/2} \left[\hat{\beta}_n(x) - \beta(x) - hL_{1n}\beta'(x) - \frac{1}{2}h^2 L_{2n}\beta''(x)\right]$$
$$\to_D \sigma_\beta \mathbb{N}_p, \tag{4.4}$$

*for any $h$ satisfying $n^2h^{5+2\nu}/d_n \to 0$ and $nh/d_n \to \infty$, where, for $j = 1$ and 2,*

$$L_{jn} = \left\{\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\left[(x_k - x)/h\right]\right\}^{-1} \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K_j\left[(x_k - x)/h\right], \quad K_j(s) = s^j K(s), \tag{4.5}$$

*$\sigma_\beta^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(x)dx$ and $\mathbb{N}_p \sim \mathcal{N}(0, \mathbb{I}_p)$ is a standard $p$-dimensional normal vector.*

**Remark 4.1.** Result (4.3) shows that, under linear local behavior of $\beta(\cdot)$ in conditions **A6**(i) and **A7**(i), $\widehat{\beta}_n(x)$ is a consistent estimator of $\beta(x)$ when $h \to 0$ and, as $n \to \infty$, $h$ satisfies $\min\{nh^2, nh/d_n\} \to \infty$. Observe that in this nonstationary functional regression context, two effective sample size conditions are used to achieve consistency: the first is $\sqrt{nh} \to \infty$ for the signal associated with the I(1) regressor $l_t$; and the second is $nh/d_n \to \infty$ for the signal associated with the local functional covariate $x_t$.[2],[3] Under local quadratic behavior, in place of the self-normalized limit

[2]See Theorem 3.1 and Remark 3.3 of WP2009a, where the condition $nh/d_n \to \infty$ was used and discussed in the context of nonstationary nonparametric regression with potential long memory. When the covariate is I(1) or near I(1), then $d_n = \sqrt{n}$ and the condition is $\sqrt{nh} \to \infty$, in contrast to the usual effective sample size condition $nh \to \infty$ that is employed in stationary kernel regression.

[3]Interestingly, formula (4.2) may be used for local estimation and inference in functional coefficient regression for a wide range of time series. The apparent simplicity in the procedure arises because of self-normalization explicit in the limit theory, which allows for a substantial range of possibilities in practice for estimation and inference. As well-established in existing practice with both stationary and nonstationary nonparametric and functional coefficient regression, care in bandwidth choice is important and applies here. Our results do not provide an optimal bandwidth result but the conditions for validity are clearly stated. As in normal practice, these conditions rely on some unknowns such as $d_n$; and higher-order expansions to assess convergence rates are unavailable and difficult to derive. In practice, it is typically wise to employ a bandwidth that is "too small," one that is "too large" and one or two "in between" in deciding a final choice.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

theory given in (4.4), we also have the (mixed) normal limit theory

$$\left(\frac{n^2h}{d_n}\right)^{1/2}\left[\hat{\beta}_n(x)-\beta(x)-hL_{1n}\beta'(x)-\frac{1}{2}h^2L_{2n}\beta''(x)\right]\to_D \sigma_\beta\,\mathcal{Y}^{-1/2}\,\mathbb{N}_p, \tag{4.6}$$

where $\mathcal{Y}$ is a random matrix independent of $\mathbb{N}_p$ defined by $\mathcal{Y}=\int_0^1 \Psi\,\Lambda_s\Lambda_s^{\mathrm{T}}\Psi L_X(ds,0)$. Here, $\Psi\,\Lambda_s$ is as in (2.2) with $\Psi$ of full rank and $\Lambda_s$ a Gaussian process with positive-definite variance matrix $s\Omega$, from **A2′(i)** and **A2(ii)**, so that the limiting signal matrix $\mathcal{Y}$ is positive definite almost surely.[4] Scaling the numerator and denominator of $L_{1n}$ and using Proposition 4.1 below gives

$$\frac{1}{n}\sum_{k=1}^n l_k l_k^{\mathrm{T}}K_1\big[(x_k-x)/h\big]=O_p[(nh/d_n)^{1/2}],$$

and from Theorem 3.1 with $c_n=d_n/h$, we have

$$\frac{d_n}{hn^2}\sum_{k=1}^n l_k l_k^{\mathrm{T}}K\left[d_n\left(\frac{x_k}{d_n}-\frac{x}{d_n}\right)/h\right]\to_D \mathcal{Y}.$$

It follows that the asymptotic impact of the $O(h)$ bias term on the distribution of $\hat{\beta}_n(x)$ is given by

$$\begin{aligned}
&\left(\sum_{k=1}^n l_k l_k^{\mathrm{T}}K\big[(x_k-x)/h\big]\right)^{1/2} hL_{1n}\\
&\quad=h\left(\sum_{k=1}^n l_k l_k^{\mathrm{T}}K\big[(x_k-x)/h\big]\right)^{-1/2}\sum_{k=1}^n l_k l_k^{\mathrm{T}}K_1\big[(x_k-x)/h\big]\\
&\quad=(hd_n)^{1/2}\left(\frac{d_n}{hn^2}\sum_{k=1}^n l_k l_k^{\mathrm{T}}K\big[(x_k-x)/h\big]\right)^{-1/2}\frac{1}{n}\sum_{k=1}^n l_k l_k^{\mathrm{T}}K_1\big[(x_k-x)/h\big]\\
&\quad=(hd_n)^{1/2}\times O_p\left(\frac{nh}{d_n}\right)^{1/2}=O_p(\sqrt{nh}),
\end{aligned} \tag{4.7}$$

which might diverge by virtue of the effective sample size condition $\sqrt{nh}\to\infty$. This indicates that the result (4.4) cannot generally be used in practice for inference.

**Remark 4.2.** Liang et al. (2023) investigated the model (4.1) in the case where $l_k$ are stationary variables satisfying certain conditions; and WP2009a and Wang

[4]The nonsingularity of the limit matrix $\mathcal{Y}$ simplifies the asymptotic theory of estimation and inference in model (4.1) with multiple regressors in comparison to similar cointegrating regression models with time-varying parameter (TVP) coefficients of the form $z_k=l_k^{\mathrm{T}}\beta(k/n)+u_k$ in place of (4.1). In models of the latter type, the limiting signal matrix is singular because kernel regression asymptotics focus attention on a single point in time, thereby reducing full signal strength to a single direction determined by the limiting process $Y_t$, where $Y_{n,\lfloor nt\rfloor}\Rightarrow Y_t$ with $Y_{nk}=l_k/d_n$. This singularity leads to multiple rates of convergence in different (random) directions in the limit (see Phillips et al., 2017 for a full analysis of the asymptotic theory for this model of TVP cointegration).

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

and Phillips (2011) considered the pure nonparametric regression model in which $l_k = 1$. In comparison with (4.4) and (4.6), the results given in Liang et al. (2023), WP2009a, and Wang and Phillips (2011) have the form

$$\left(\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\big[(x_k - x)/h\big]\right)^{1/2} \left[\widehat{\beta}_n(x) - \beta(x) - \frac{h^2 \beta''(x)}{2} \int_{-\infty}^{\infty} s^2 K(s) ds\right] \to_D \sigma_\beta \mathbb{N}_p,$$

$$\left(\frac{nh}{d_n}\right)^{1/2} \left[\widehat{\beta}_n(x) - \beta(x) - \frac{h^2 \beta''(x)}{2} \int_{-\infty}^{\infty} s^2 K(s) ds\right] \to_D \left[\mathbb{E}(l_1 l_1^{\mathrm{T}})\right]^{-1} \sigma_\beta \mathcal{Y}^{-1/2} \mathbb{N}_p,$$

i.e., in (4.4) and (4.6), the linear term $hL_{1n}\beta'(x)$ disappears and $L_{2n}$ in the limit theory (4.4) can be replaced by $\int_{-\infty}^{\infty} s^2 K(s) ds$ when $l_k$ is stationary. The reason for the difference is that when $l_k$ is stationary $\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\big[(x_k - x)/h\big]$ has a reduced signal with divergence rate $\frac{nh}{d_n}$ rather than $\frac{n^2 h}{d_n}$, so that

$$\left(\frac{nh}{d_n}\right)^{1/2} hL_{1n}\beta'(x) = O_P(h) = o_P(1)$$

and $L_{2n}$ can be replaced by $\int_{-\infty}^{\infty} s^2 K(s) ds$ by virtue of a simple calculation. See Liang et al. (2023) for details. And when $l_k$ is nonstationary, Phillips and Wang (2023) found explicit limit formulas showing that the bias term can affect both the asymptotic variance and the convergence rates in nonstationary functional coefficient cointegrating regression even when $x_k$ is stationary. The present results provide further extension of these findings to the case of multiple nonstationary regressors and a nonstationary functional covariate, both of which are important in practical applications.

We may improve Theorem 4.1 by using local linear estimation (e.g., Fan and Gijbels, 1996) and imposing more restrictions between $x_k$ and $l_k$. The local linear estimator $\widehat{\beta}_L(x)$ of $\beta(x)$ in functional coefficient regression is defined by

$$\begin{pmatrix} \widehat{\beta}_L(x) \\ \widehat{\beta}_L'(x) \end{pmatrix} = \arg\min_{\beta, \beta_1} \sum_{k=1}^{n} \left\{z_k - l_k^{\mathrm{T}}[\beta + \beta_1 (l_k - x)]\right\}^2 K\left(\frac{x_k - x}{h}\right),$$

leading to

$$\widehat{\beta}_L(x) = \left(V_{n0} - V_{n1} V_{n2}^{-1} V_{n1}\right)^{-1} \sum_{k=1}^{n} \left[\mathbb{I} - V_{n1} V_{n2}^{-1} (x_k - x)\right] l_k z_k K\left(\frac{x_k - x}{h}\right), \tag{4.8}$$

where $V_{nj} = \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\left(\frac{x_k - z}{h}\right)(x_k - x)^j$ for $j = 0, 1,$ and 2.

THEOREM 4.2. *In addition to the conditions of Theorem 4.1, assume that $\epsilon_i, i \in \mathcal{Z}$ is independent of $(\eta_{i1}, \ldots, \eta_{id}), i \in \mathcal{Z}$. Then, under **A6**(ii) and **A7**(ii), we have*

$$\left(\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\big[(x_k - x)/h\big]\right)^{1/2} \left[\widehat{\beta}_L(x) - \beta(x) - \frac{h^2 \beta''(x)}{2} \int_{-\infty}^{\infty} s^2 K(s) ds\right]$$
$$\to_D \sigma_\beta \mathbb{N}_p, \tag{4.9}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

*for any $h$ satisfying* $\max\{n^2h^{5+2\nu}/d_n, nh^4\} \to 0$ *and* $nh/d_n \to \infty$, *where* $\sigma_\beta^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(x)dx$ *and* $\mathbb{N}_p \sim \mathcal{N}(0, \mathbb{I}_p)$ *is a standard p-dimensional normal vector.*

**Remark 4.3.** The independence condition between $x_k$ and $l_k$ is imposed to establish the following Proposition 4.1, which plays a key part in the proof of Theorem 4.2. Since $\int_{-\infty}^{\infty} m(x)dx = 0$ is assumed in this result, the sample moment $A_{n,m}$ given below in (4.10) is referred to as a "zero energy" statistic following the discussion in Remark 3.4. As explained there, we are currently unable to relax the independence condition between $x_k$ and $l_k$ even when only a rough estimate of the order of $A_{n,m}$ is needed.

PROPOSITION 4.1. *Under the conditions of Theorem 4.2, for any bounded real function $m(x)$ satisfying $\int_{-\infty}^{\infty} m(x)dx = 0$ and $\int_{-\infty}^{\infty} |xm(x)|dx < \infty$, we have*

$$A_{n,m} := \frac{1}{n}\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} m[(x_k - x)/h] = O_P\big[(nh/d_n)^{1/2}\big]. \tag{4.10}$$

**Remark 4.4.** In comparison with Theorem 4.1, the local linear estimator in model (4.1) with both $x_k$ and $l_k$ nonstationary has bias reducing properties under the additional assumption of independence between $x_k$ and $l_k$. We conjecture that such a bias reducing property still holds without the restriction of independence, but rigorous proof may be extremely challenging, as explained in Remark 4.3. This property matches that of stationary time-series regression, where both $x_k$ and $l_k$ are stationary but differs from the model in which $l_k$ is stationary and $x_k$ is nonstationary. In the latter case, Liang et al. (2023) proved that the bias reducing advantage in local linear estimation is lost, a limitation that was first noticed in the work of WP2009b and Wang and Phillips (2011) on nonstationary kernel regression.

**Remark 4.5.** For practical applications, model (4.1) can be readily extended to include an intercept in the regression with its own functional coefficient, as in a model of the following form:

$$z_k = \chi_k^{\mathrm{T}} \gamma(x_k) + u_k, \tag{4.11}$$

where $x_k$ and $u_k$ are defined as in **A1** and **A4**, respectively, $\chi_k = (\ell_{1k}^{\mathrm{T}}, l_k^{\mathrm{T}})^{\mathrm{T}}$ in which $\ell_{1k}^{\mathrm{T}} = (1, k, \ldots, k^q)$ is a vector of polynomial time trends, including an intercept and $l_k = \sum_{j=1}^{k} v_j$, where $v_j = (v_{j1}, \ldots v_{jp})^{\mathrm{T}}$, with $p \le d$, is a linear vector process defined as in **A2′**, and $\gamma(\cdot) = (\psi(\cdot)^{\mathrm{T}}, \beta(\cdot)^{\mathrm{T}})$ in which $\psi(\cdot) = (\psi_0(\cdot), \ldots, \psi_q(\cdot))$ satisfies the same condition as in **A6 (ii)**, i.e.,

$$||\psi(y+x) - \psi(x) - \psi'(x)y - \frac{1}{2}\psi''(x)y^2|| \le C_x |y|^{2+\nu},$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

for some $0 < \nu \leq 1$, where $C_x$ is a constant depending only on $x$. Let $\Psi \Lambda_t$ be given as in (2.2) and write

$$W_t = (1, t, \ldots, t^q, (\Psi \Lambda_t)^{\mathrm{T}}).$$

It is readily seen that, on $D_{\mathbb{R}^{1+q+p}}[0,1]$,

$$G_n^{-1} \chi_{\lfloor nt \rfloor} \Rightarrow W(t), \quad \text{where } G_n = \mathrm{diag}(1, n, \ldots, n^q, \sqrt{n}\mathbb{I}_p),$$

and $\int_0^1 g(W_t) L_X(ds, a)$ is well-defined for any continuous function $g(\ldots)$ on $\mathbb{R}^{1+q+p}$. Based on these facts, by the same arguments as those used in establishing Theorem 4.2, we have the following result for the local linear estimator $\widehat{\gamma}_L(x)$ of $\gamma(x)$ in model (4.11) defined by

$$\widehat{\gamma}_L(x) = \left(\widetilde{V}_{n0} - \widetilde{V}_{n1}\widetilde{V}_{n2}^{-1}\widetilde{V}_{n1}\right)^{-1} \sum_{k=1}^{n} \left[I - \widetilde{V}_{n1}\widetilde{V}_{n2}^{-1}(x_k - x)\right] \chi_k z_k K\left(\frac{x_k - x}{h}\right), \tag{4.12}$$

where $\widetilde{V}_{nj} = \sum_{k=1}^{n} \chi_k \chi_k^{\mathrm{T}} K\left(\frac{x_k - z}{h}\right)(x_k - x)^j$ for $j = 0, 1,$ and 2.

THEOREM 4.3. *In addition to the conditions of Theorem 4.1, assume that $\epsilon_i, i \in \mathcal{Z}$ is independent of $(\eta_{i1}, \ldots, \eta_{id}), i \in \mathcal{Z}$. Then, under **A6**(ii) and **A7**(ii), we have*

$$\left(\sum_{k=1}^{n} \chi_k \chi_k^{\mathrm{T}} K\left[(x_k - x)/h\right]\right)^{1/2} \left[\widehat{\gamma}_L(x) - \gamma(x) - \frac{h^2 \gamma''(x)}{2} \int_{-\infty}^{\infty} s^2 K(s) ds\right]$$
$$\to_D \sigma_\beta \mathbb{N}_{1+q+p}, \tag{4.13}$$

*for any $h$ satisfying $\max\{n^2 h^{5+2\nu}/d_n, nh^4\} \to 0$ and $nh/d_n \to \infty$, where $\sigma_\beta^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(x) dx$ and $\mathbb{N}_{1+q+p} \sim \mathcal{N}(0, \mathbb{I}_{1+q+p})$ is a standard $(1+q+p)$-dimensional normal vector.*

The bandwidth conditions in Theorem 4.3 seem reasonable for most applications. For instance, in the important case where $d_n = \sqrt{n}$, the conditions require that $h = o(1/n^{1/4})$ in conjunction with $\sqrt{n}h \to \infty$ which, as noted earlier, is the effective sample size condition in nonparametric regression with unit-root and local unit-root nonstationary time series. The estimator (4.12) and limit theory of Theorem 4.3 are particularly useful in applications where the regression model (4.11) includes an intercept and deterministic trend. For example, in the empirical study of Section 6, it is important to allow for a covariate $x_t$ representing atmospheric $CO_2$, which is a nonstationary time series that manifests both a stochastic trend and drift over the historical period of observation.

It should be mentioned that optimal bandwidth choice in nonlinear nonparametric regression faces complex challenges in the nonstationary time-series case. As noted in Wang and Phillips (2023), the classical mean squared error (MSE) bandwidth selection rule for stationary nonparametric regression is ill-suited to nonstationary data applications because the expectation is undefined. To assist in resolving this complication, Wang and Phillips (2023) introduced a conditional

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

MSE criterion in which the optimal bandwidth was found to be a random sequence depending on the sojourn time of the process. That result was obtained under exogenous regressor conditions. Due to the complexity of the present model, it is unclear whether or not a similar result can be established and exploration of optimal bandwidth choice in the present context is left for future research.

### 4.1. Specification Testing

This section constructs a simple specification test that is useful for a later empirical application. More general settings require certain technical extensions, which will be considered in subsequent work. Consider a nonlinear cointegrating regression model with functional coefficients

$$z_k = \beta_0(x_k) + \beta_1(x_k)\, h(y_k) + u_k, \tag{4.14}$$

in which $y_k$ and $u_k$ are defined as in **A2** and **A4**, respectively, $x_k$ satisfies **A1** with $\tau = 0$, $\beta_0(x)$ and $\beta_1(x)$ are unknown smooth functions defined on $\mathbb{R}$, and $h(x)$ is an asymptotically homogeneous function defined by

$$h(\lambda x) = v(\lambda)H(x) + R(\lambda, x) \tag{4.15}$$

so that $v(\lambda) > 0$ for all $\lambda > 0$, $H(x)$ is continuous on $\mathbb{R}$, and $\lim_{\lambda\to\infty} \sup_{|x|\le K} |R(\lambda, x) / v(\lambda)H(x)| = 0$ for any $K < \infty$. We aim to test the hypothesis:

$$H_0 : \beta_0(x) = \beta_{01} + \beta_{02}x, \quad \beta_1(x) = \beta_1 \quad \text{for } x \in R. \tag{4.16}$$

Write $G_k = (1, x_k, h(y_k))^{\mathrm{T}}$ and $\theta = (\beta_{01}, \beta_{02}, \beta_1)^{\mathrm{T}}$. As in Wang and Phillips (2016), we make use of the following test statistic:

$$T_n = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^{n} K[(x_k - x)/h] \left(z_k - \hat{\theta}_n G_k\right) \right\}^2 \pi(x)dx,$$

where $\pi(x)$ is a positive integrable function satisfying $\int_{-\infty}^{\infty}(1 + x^2)\pi(x)dx < \infty$, $K(x)$ is a kernel function having a compact support and satisfying that

$$|K(x) - K(y)| \le C|x - y|$$

whenever $|x - y|$ is sufficiently small, $h$ is a bandwidth satisfying $h \equiv h_n \to 0$ as the sample size $n \to \infty$, and $\hat{\theta}_n$ is an LS estimator of the unknown parameter $\theta$ under the null $H_0$, i.e.,

$$\hat{\theta}_n = \left( \sum_{k=1}^{n} G_k^{\mathrm{T}} G_k \right)^{-1} \sum_{k=1}^{n} G_k^{\mathrm{T}} z_k.$$

The statistic $T_n$ is a modification of the test statistic discussed by Härdle and Mammen (1993) for the random sample case, which was also used in Gao, Tjøstheim, and Yin (2012) for a nonlinear cointegrating model with a martingale error structure. We have the following result.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

THEOREM 4.4. *Under the null hypothesis $H_0$, we have*

$$T_{n,d} := \frac{d_n}{nh} T_n \to_D \tau_0 L_X(1,0), \tag{4.17}$$

*for any $h$ satisfying $nh^2 \log n/d_n \to 0$ and $n^{1-\delta_0}h/d_n \to \infty$, where*

$$\tau_0 = \sigma^2 \int_{-\infty}^{\infty} K^2(s)ds \int_{-\infty}^{\infty} \pi(x)dx \tag{4.18}$$

*and $\delta_0$ can be as small as required.*

**Remark 4.6.** As noted in Wang and Phillips (2016), in the short memory case (i.e., $x_k$ satisfies **A1** with $\tau = 0$ under **SM**), $d_n = \phi\sqrt{n}$ and result (4.17) reduces to

$$T_{n,0} = \frac{\phi}{\sqrt{n}h} T_n \to_D \tau_0 L_B(1,0). \tag{4.19}$$

When $u_k$ is a stationary martingale difference, $\sigma^2 = \mathbb{E}u_0^2$, which can be estimated by

$$\hat{\sigma}_n^2 = \frac{\sum_{k=1}^n \left(z_k - \hat{\theta}_n G_k\right)^2 K\left[(x_k - x)/h\right]}{\sum_{k=1}^n K\left[(x_k - x)/h\right]},$$

based on a localized version of the usual residual sum of squares. Further, $\phi$ can be estimated by a standard HAC method or other alternative methods, indicating that in the short memory case, the test $T_{n,0}$ is directly applicable. However, in the long memory case, the scaling in the statistic (4.17) relies on the expansion rate parameter $d_n^2 \sim c_\mu\, n^{3-2\mu}\, a^2(n)$, which in turn relies on the unknown value of $\mu$. Even in the simple case where $a(n)$ is constant and $d_n \sim cn^{\frac{1}{2}+d}$ for some constant $c$, the required scaling depends on the (typically unknown) value of the long memory parameter $d = 1 - \mu$. In consequence, the statistic $T_{n,d}$ is not well suited for practical implementation in this type of specification testing when the near integrated time series $x_k$ is driven by long memory innovations. We refer to Wang and Phillips (2016, Sect. 3) for further details on that complication.

**Remark 4.7.** To investigate asymptotic power, as in Wang and Phillips (2016), we may consider the following local alternative model:

$$H_1 : \beta_0(x) = \beta_{01} + \beta_{02}x + \rho_n m(x), \quad \beta_1(x) = \beta_1 \quad \text{for } x \in R, \tag{4.20}$$

where $\rho_n$ is a sequence of constants measuring local deviations from the null and $m(x)$ is a real function. Suppose that, in a neighborhood of $x$,

$$|m(x+y) - m(x)| \le C\,|y|^\gamma\, m_1(x),$$

for some $\gamma \in (0,1]$, where $m(x)$ and $m_1(x)$ are real functions satisfying that $\int_{-\infty}^{\infty}[1 + m^2(x) + m_1^2(x)]\,\pi(x)\,dx < \infty$ and $\int_{-\infty}^{\infty} m^2(x)\,\pi(x)\,dx > 0$. A similar argument to that used in Theorem 3.2 of Wang and Phillips (2016) yields the following

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

result under the alternative $H_1$:

$$\lim_{n\to\infty}\mathbb{P}\Big(\frac{d_n}{nh}T_n \geq t_0\Big) = 1,$$

for any $t_0 > 0$, any $h \to 0$ satisfying $n^{1-\delta_0}h/d_n \to \infty$, where $\delta_0$ can be as small as required, and any $\rho_n$ satisfying $nh\rho_n^2/d_n \to \infty$. This result indicates that the $T_n$ test has nontrivial power against local alternatives of the form (4.20) whenever $\rho_n \to 0$ at a rate that is slower than $[d_n/(nh)]^{1/2}$, as $nh/d_n \to \infty$.

**Remark 4.8.** By imposing more restrictions between $x_k$ and $y_k$, it is feasible to consider a different test statistic. In fact, as in Wang and Phillips (2012), we may introduce the kernel-smoothed test statistic

$$\widetilde{T}_n = \sum_{k,j=1,k\neq j}^{n} \hat{u}_k\hat{u}_j K\big[(x_k - x_j)/h\big],$$

where $\hat{u}_k = z_k - \hat{\theta}_n G_k^{\mathrm{T}}$. The following result provides a standard normal limit distribution free of any nuisance parameters, facilitating practical application.

THEOREM 4.5. *Suppose that $\{x_k\}_{k\geq 1}$ is independent of $\{y_k\}_{k\geq 1}$ and $x_k$ satisfies **A1** under **SM** with additional coefficient condition $\sum_{k=0}^{\infty} k^{1+\delta}|\phi_k| < \infty$ for some $\delta > 0$ and $\mathbb{E}\epsilon_0^6 < \infty$. Furthermore, in addition to **A4**, $\sup_{k\geq 1}\mathbb{E}\big(u_k^4|\mathcal{F}_{k-1}\big) < \infty$. Then, under the null hypothesis, we have*

$$\frac{\widetilde{T}_n}{\sqrt{2}\widetilde{\sigma}_n} \to_D \mathcal{N}(0,1), \tag{4.21}$$

*for any $h$ satisfying $nh^2 \to \infty$ and $nh^4\log^2 n \to 0$, where*

$$\widetilde{\sigma}_n^2 = \sum_{k,j=1,k\neq j}^{n} \hat{u}_k^2\hat{u}_j^2 K^2\big[(x_k - x_j)/h\big].$$

## 5. SIMULATIONS

This section explores the finite sample performance of the functional coefficient regression estimators proposed in Section 4. The following two models are considered:

- M1: $z_k = y_k\beta(x_k) + u_k$;
- M2: $z_k = \beta_0(x_k) + y_k^*\beta_1(x_k) + u_k$.

In the simulations, the covariate $x_k$ is generated according to $x_k = x_{k-1} + \xi_k$. For the linear process $\xi_k = \sum_{j=0}^{\infty}\phi_j\epsilon_{k-j}$, we employ the model $(1-L)^d\xi_k = \epsilon_k$ with $0 \leq d < 0.5$ and innovations $\epsilon_k \sim_{iid} \mathcal{N}(0, 0.1^2)$. The regressor $y_k$ follows $y_k = y_{k-1} + v_k$, where $v_k = 0.5v_{k-1} + w_k$, and $w_k = (\epsilon_{yk} + \theta\epsilon_k)/\sqrt{1+\theta^2}$. The process $\epsilon_{yk}$ is iid standard normal and independent of $\epsilon_k$. The regressor $y_k^* = y_k + 0.1k$ so that $y_k^*$ is a unit-root process with additive drift that mimics the property

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

**TABLE 1.** Finite sample mean bias and standard error (in square brackets) of the local level estimator $\hat{\beta}_n(x)$ at point $x = 0$ for model M1 with $h = 0.5n^{\gamma}$

| | $\gamma = -3/10$ | | $\gamma = -1/5$ | |
|---|---|---|---|---|
| $n$ | $d = 0$ | $d = 0.4$ | $d = 0$ | $d = 0.4$ |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$ | | |
| 100 | 0.016 [0.076] | 0.013 [0.073] | 0.040 [0.142] | 0.033 [0.119] |
| 400 | 0.008 [0.041] | 0.007 [0.051] | 0.026 [0.089] | 0.022 [0.076] |
| 800 | 0.005 [0.031] | 0.004 [0.045] | 0.019 [0.065] | 0.016 [0.059] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 0, 0)$ | | |
| 100 | 0.017 [0.072] | 0.014 [0.071] | 0.043 [0.136] | 0.034 [0.115] |
| 400 | 0.008 [0.040] | 0.006 [0.051] | 0.026 [0.084] | 0.022 [0.073] |
| 800 | 0.005 [0.031] | 0.004 [0.045] | 0.020 [0.061] | 0.017 [0.058] |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 1, 1)$ | | |
| 100 | 0.017 [0.074] | 0.010 [0.070] | 0.040 [0.143] | 0.031 [0.119] |
| 400 | 0.007 [0.039] | 0.005 [0.048] | 0.027 [0.090] | 0.020 [0.074] |
| 800 | 0.005 [0.029] | 0.004 [0.040] | 0.019 [0.064] | 0.016 [0.057] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 1, 1)$ | | |
| 100 | 0.016 [0.071] | 0.013 [0.068] | 0.042 [0.135] | 0.035 [0.116] |
| 400 | 0.007 [0.037] | 0.005 [0.046] | 0.026 [0.083] | 0.020 [0.072] |
| 800 | 0.005 [0.028] | 0.003 [0.039] | 0.020 [0.061] | 0.016 [0.056] |

of the regressor $\ln CO_{2,k}$ in the empirical study of climate sensitivity reported below in Section 6. The regression error has the composite form $u_k = (e_k + \lambda_1\epsilon_{yk} + \lambda_2\epsilon_k)/\sqrt{1+\lambda_1^2+\lambda_2^2}$, where the $e_k$ are iid standard normal and independent of the innovations $(\epsilon_{yk}, \epsilon_k)$, a formulation that accommodates the possibility of endogeneity in the regressor $y_k$. The setting $x_1 = 0$ is used to examine pointwise estimation accuracy at $x = 0$. To avoid problems of a weak signal in the regressor in nonparametric estimation of model M1, the initialization $y_0 = 10$ is employed, which avoids occasional clustering of near zero observations in small samples. For M1, we consider the functional coefficient $\beta(x) = 1 + x + x^2$, and for M2, we consider $\beta_0(x) = 1 + x + x^2$ and $\beta_1(x) = 1 - x - x^2$. A Gaussian kernel is used throughout and the number of replications is 50,000.

The pointwise estimation accuracy of the local level estimator $\hat{\beta}_n(x)$ in model M1 is reported in Table 1. The two figures in each entry are the finite sample bias and standard error (shown in square brackets). A fixed bandwidth $h = C_h n^{\gamma}$ is used, where the constant scale coefficient $C_h$ and power parameter $\gamma$ are indicated in the table. Here, and later for model M2, a fixed bandwidth $h = C_h n^{\gamma}$ is used in demonstrating the performance of the local level and local linear estimators. The usual bandwidth formula $h = C_h \hat{\sigma}_x n^{\gamma}$ is inappropriate in the present context because the standard deviation of a nonstationary regressor or covariate diverges and this typically leads to bandwidth divergence.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

The left panel with $\gamma = -3/10$ meets the bandwidth condition given in Theorem 4.1, whereas the right panel with $\gamma = -1/5$ does not. When $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$, $x_k$ and $y_k$ are independent and both are independent of the equation error $u_k$. Other scenarios demonstrate finite sample performance when that independence is removed. The local level estimator seems to work well with diminishing bias and standard error as $n$ increases in all the scenarios considered. The performance is not sensitive to the correlation parameters $(\theta, \lambda_1, \lambda_2)$, thereby indicative of robustness, including robustness to endogeneity in the regression, a result that mirrors earlier findings on nonparametric cointegrating regression in WP2009b. A larger bandwidth choice with higher $\gamma$ seems to make both bias and standard error worse, as may be expected. On the other hand, a larger memory parameter $d$ raises signal strength and helps to reduce the bias. These results are obtained using a fixed given bandwidth. It might be expected that improved results could be achieved with variable bandwidth choices better suited to the specific generating mechanism. The present limit theory only provides general guidance and optimal bandwidth results are well known to be scarce in nonstationary nonparametric regression, although Wang and Phillips (2023) provide some findings for nonlinear cointegrated regression. The difficulties of developing a general theory of optimal bandwidth choice are substantially greater here with nonstationarity in the covariate as well as the regressor.

Table 2 collects parallel results for the local linear estimator $\hat{\beta}_L(x)$. The local linear estimator is much more sensitive to bandwidth than the local level estimator. Poor bandwidth choice may cause the bias and/or standard error to increase and make the estimator seem inconsistent. When $\gamma = -1/5$, which does not satisfy the bandwidth condition in Theorem 4.2, the bias seems to increase from $n = 100$ to 400, and then decrease from $n = 400$ to 800. We have tried increasing the sample size and confirmed that the bias is decreasing, but at a very slow rate. Results with $\gamma = -3/10$ seem satisfactory. The results again seem insensitive to the correlation parameter $(\theta, \lambda_1, \lambda_2)$, showing that the local linear estimator retains robustness to endogeneity. The bias is smaller and the standard error is larger when there is stronger memory in $\xi_k$ (larger $d$). Compared to the results in Table 1, the findings for the local linear estimator indicate that it outperforms the local level estimator in terms of both bias and standard deviation under the same bandwidth conditions.

We next consider model M2. The finite sample bias and standard error of the local level estimator for the two functional coefficients are collected in Table 3. We use $h = 1.5n^{-3/10}$ with an increased scale coefficient $C_h = 1.5$, which helps to avoid near singularity in the weighted signal matrix. The results show that the slope coefficient $\beta_1(x)$ is more accurately estimated than the intercept coefficient $\beta_0(x)$, reflecting the different signal strengths in the corresponding regressors. Also noted is that estimation accuracy deteriorates as the memory parameter $d$ increases. The explanation is simply that with stronger memory in $\xi_k = \Delta x_k$, the covariate $x_k$ wanders more extensively, leading to fewer observations being available in local estimation. Estimation is again seen to be insensitive to the correlation parameters $(\theta, \lambda_1, \lambda_2)$, showing that robustness to endogeneity applies in the estimation of

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

**TABLE 2.** Finite sample mean bias and standard error (in square brackets) of the local linear estimator $\hat{\beta}_L(x)$ at point $x = 0$ for model M1 with $h = 0.5n^{\gamma}$

| | $\gamma = -3/10$ | | $\gamma = -1/5$ | |
|---|---|---|---|---|
| $n$ | $d = 0$ | $d = 0.4$ | $d = 0$ | $d = 0.4$ |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$ | | |
| 100 | 0.006 [0.037] | 0.005 [0.055] | 0.006 [0.040] | 0.010 [0.053] |
| 400 | 0.005 [0.029] | 0.004 [0.048] | 0.012 [0.027] | 0.013 [0.043] |
| 800 | 0.004 [0.025] | 0.003 [0.044] | 0.012 [0.023] | 0.011 [0.038] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 0, 0)$ | | |
| 100 | 0.006 [0.036] | 0.005 [0.054] | 0.007 [0.039] | 0.010 [0.053] |
| 400 | 0.005 [0.030] | 0.004 [0.049] | 0.013 [0.027] | 0.012 [0.043] |
| 800 | 0.004 [0.027] | 0.003 [0.045] | 0.012 [0.023] | 0.011 [0.038] |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 1, 1)$ | | |
| 100 | 0.006 [0.030] | 0.003 [0.046] | 0.006 [0.036] | 0.008 [0.047] |
| 400 | 0.005 [0.024] | 0.003 [0.040] | 0.013 [0.024] | 0.012 [0.036] |
| 800 | 0.004 [0.021] | 0.002 [0.036] | 0.012 [0.020] | 0.011 [0.032] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 1, 1)$ | | |
| 100 | 0.005 [0.031] | 0.003 [0.045] | 0.006 [0.035] | 0.007 [0.046] |
| 400 | 0.005 [0.025] | 0.003 [0.040] | 0.013 [0.024] | 0.011 [0.037] |
| 800 | 0.004 [0.022] | 0.002 [0.038] | 0.012 [0.019] | 0.011 [0.032] |

model M2 also. Table 4 reports similar results for the local linear estimator $\widetilde{\widetilde{\beta}}_L(x)$. The estimator performance is extremely sensitive to bandwidth, especially when $d = 0.4$. A larger scale coefficient $C_h = 2$ is used to avoid singularity in the weighted signal matrix. Results in the SM case with $d = 0$ show monotonically decreasing bias and standard error. In the LM case with $d = 0.4$, the bias sometimes does not decay monotonically. These simulation findings suggest that estimator performance in this scenario is extremely sensitive to bandwidth. Consequently, it appears that finding an appropriate bandwidth selection rule that applies in all situations with a diverse range of nonstationarity in the variables is a challenging problem for future investigation.

In terms of the overall estimation accuracy of the functional coefficient curves, Figure 1 provides an illustrative example based on a typical replication and use of a fixed bandwidth $h = C_h n^{-3/10}$ with $C_h \in \{0.5, 1, 2\}$. Bandwidth choice clearly affects performance and local linear estimation generally outperforms local level estimation. For model M2, the slope coefficient $\beta_1(x)$ is estimated more precisely than the intercept coefficient $\beta_0(x)$, as expected.

Further, to illustrate how well asymptotic results approximate the finite sample behavior of the estimators, we plot the empirical densities of the self-normalized quantities studied in Theorems 4.1–4.3. Specifically, the empirical densities of the left-hand side quantities in equations (4.4), (4.9), and (4.13) are plotted and

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

**TABLE 3.** Finite sample mean bias and standard error (in square brackets) of the local level estimator at point $x = 0$ for model M2 with bandwidth $h = 1.5n^{-3/10}$

| | $d = 0$ | | $d = 0.4$ | |
|---|---|---|---|---|
| $n$ | $\beta_0(x)$ | $\beta_1(x)$ | $\beta_0(x)$ | $\beta_1(x)$ |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$ | | |
| 100 | 1.827 [6.661] | −0.235 [0.566] | 2.484 [10.405] | −0.299 [0.958] |
| 400 | 0.821 [4.910] | −0.112 [0.326] | 0.820 [6.795] | −0.115 [0.623] |
| 800 | 0.472 [4.313] | −0.070 [0.236] | 0.460 [6.001] | −0.070 [0.554] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 0, 0)$ | | |
| 100 | 2.711 [7.778] | −0.306 [0.689] | 3.905 [13.517] | −0.433 [1.277] |
| 400 | 1.194 [5.227] | −0.140 [0.380] | 1.287 [8.893] | −0.158 [0.843] |
| 800 | 0.711 [4.473] | −0.084 [0.277] | 0.762 [7.939] | −0.098 [0.759] |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 1, 1)$ | | |
| 100 | 1.867 [6.724] | −0.234 [0.566] | 2.379 [10.200] | −0.290 [0.931] |
| 400 | 0.747 [4.874] | −0.108 [0.324] | 0.649 [7.697] | −0.097 [0.714] |
| 800 | 0.449 [4.271] | −0.068 [0.235] | 0.309 [6.434] | −0.053 [0.586] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 1, 1)$ | | |
| 100 | 2.702 [7.849] | −0.302 [0.685] | 3.584 [13.390] | −0.400 [1.259] |
| 400 | 1.193 [5.180] | −0.139 [0.379] | 1.096 [8.888] | −0.136 [0.841] |
| 800 | 0.727 [4.396] | −0.083 [0.272] | 0.534 [7.742] | −0.075 [0.729] |

shown against the respective limiting normal density. The unknown $\beta(x)$, $\beta'(x)$, $\beta''(x)$, and $\sigma^2$ are replaced with true values. The results are displayed in Figure 2. The findings here are not sensitive to $d$ and $(\theta, \lambda_1, \lambda_2)$ but they are sensitive in estimation accuracy to bandwidth, as we have observed. As is well-known, the usual advantage of local linear estimation is to reduce bias and improve performance near the boundary. It is interesting to notice that here, while the bias terms are included to ensure the asymptotic normality as given in Theorems 4.1–4.3, the density of the local level estimator appears to be better approximated by limit behavior than that of the local linear estimator, as indicated in Figure 2(a). In sum, the finite sample densities are generally close to and appear to be converging to the asymptotic distribution, corroborating the limit theory of Theorems 4.1–4.3.

These simulation findings in the present functional coefficient model setting, where both the regressor $y_k$ and the smoothing covariate $x_k$ are nonstationary, show that nonparametric estimation can be very sensitive to bandwidth. This sensitivity is found to become severe when there are multiple regressors and local linear methods are used. Poor choice of bandwidth can worsen finite sample performance, sometimes with increasing bias and/or variance even as the sample size rises from $n = 100$ to $n = 400$, and larger sample sizes are needed to obtain improvements in performance. This high sensitivity is primarily due to the random wandering nature of the covariate $x_k$ and its functional interaction with the nonstationary regressor $y_k$ by virtue of the nonparametric treatment of the functional coefficient.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

**TABLE 4.** Finite sample mean bias and standard error (in square brackets) of the local linear estimator at point $x = 0$ for model M2 with bandwidth $h = 2n^{-3/10}$

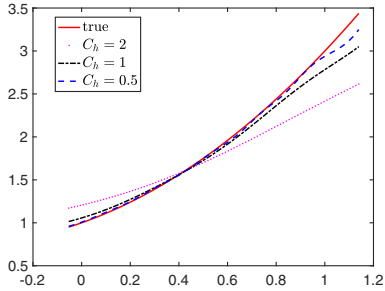
| | $d = 0$ | | $d = 0.4$ | |
|---|---|---|---|---|
| $n$ | $\beta_0(x)$ | $\beta_1(x)$ | $\beta_0(x)$ | $\beta_1(x)$ |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$ | | |
| 100 | −1.666 [3.748] | 0.164 [0.333] | −1.962 [7.781] | 0.192 [0.724] |
| 400 | −0.561 [2.824] | 0.035 [0.202] | −0.069 [5.561] | −0.008 [0.526] |
| 800 | −0.157 [2.237] | −0.004 [0.138] | 0.107 [5.946] | −0.028 [0.568] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 0, 0)$ | | |
| 100 | −1.912 [4.286] | 0.187 [0.395] | −2.192 [9.820] | 0.217 [0.937] |
| 400 | −0.634 [3.019] | 0.044 [0.233] | −0.073 [7.208] | −0.005 [0.695] |
| 800 | −0.188 [2.354] | −0.000 [0.162] | 0.131 [8.535] | −0.029 [0.821] |
| | | $(\theta, \lambda_1, \lambda_2) = (0, 1, 1)$ | | |
| 100 | −1.711 [3.720] | 0.169 [0.336] | −2.449 [7.807] | 0.241 [0.738] |
| 400 | −0.585 [2.806] | 0.038 [0.204] | −0.512 [5.250] | 0.037 [0.508] |
| 800 | −0.182 [2.219] | −0.001 [0.138] | −0.279 [5.119] | 0.012 [0.488] |
| | | $(\theta, \lambda_1, \lambda_2) = (1, 1, 1)$ | | |
| 100 | −1.967 [4.261] | 0.194 [0.398] | −2.875 [9.780] | 0.285 [0.944] |
| 400 | −0.664 [2.998] | 0.048 [0.234] | −0.700 [6.590] | 0.058 [0.645] |
| 800 | −0.221 [2.325] | 0.004 [0.161] | −0.407 [6.910] | 0.025 [0.667] |

Standard cross validation, which is not validated in the present setting of dual-source nonstationarity, does not materially improve matters even if implemented within each replication. Use of small bandwidths typically increases the risk of the locally weighted signal matrix being close to singular, leading to a form of ill-posed inverse problem analogous to, but originating differently from that which arises in certain microeconometric instrumental variable regressions (e.g., Hall and Horowitz, 2005). Problems of this type also arise in nonparametric regression whenever signal information is limited to very few observations, as when strong effects of long memory are present in the data (e.g., Wang and Phillips, 2024). In such circumstances, estimation can become volatile and unreliable, which in turn interferes with the performance of methods such as cross validation. It therefore appears that such problems as ill-posed inversions and the curse of dimensionality in nonparametric estimation are exacerbated in the functional coefficient dual-source nonstationary setting. Tackling these issues systematically using methods, such as regularization, is a challenging task to be addressed in future research.

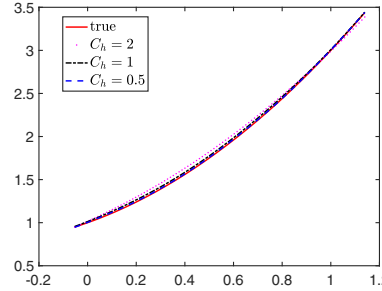
### 5.1. Test Performance

This section examines the finite sample performance of the two test statistics proposed in Section 4.1. First, convergence in both (3.3) and (4.21) is found to

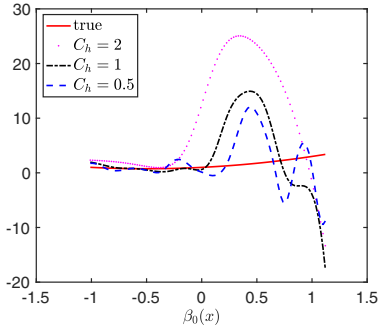
https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

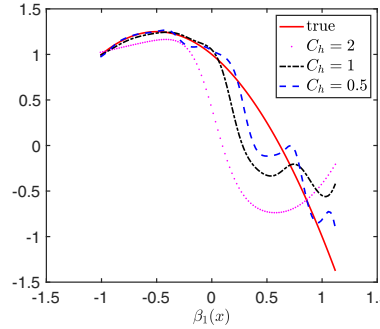



(a) model M1, local level estimate

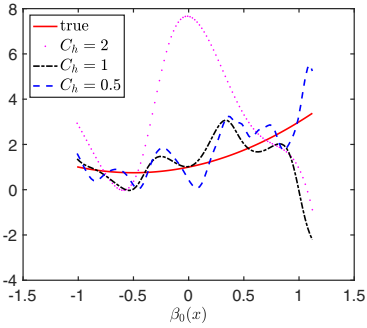


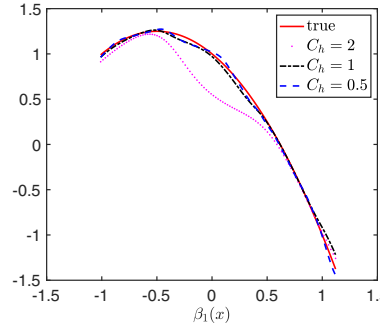

(b) model M1, local linear estimate




(c) model M2, local level estimate



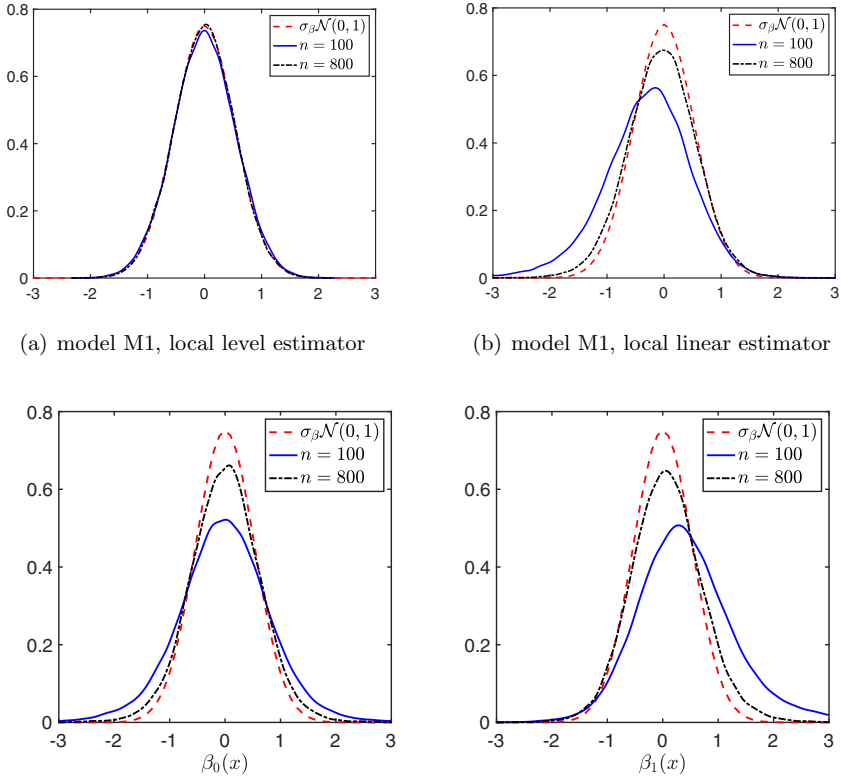




(d) model M2, local linear estimate

**FIGURE 1.** A representative demonstration of the curve estimates with $n = 400$, $h = C_h n^{-3/10}$, $d = 0$, and $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$.

be slow and the finite sample density of $T_{n,0}$ is more concentrated at the origin than that of the limit $\tau_0 L_X(1, 0)$, which results in severe under-sizing. Second, the density of $\widetilde{T}_n/\sqrt{2\tilde{\sigma}_n^2}$ is biased to the left and has smaller variance than the standard

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press




(a) model M1, local level estimator

(b) model M1, local linear estimator

(c) model M2, local linear estimator

**FIGURE 2.** A representative demonstration of the empirical densities of the self-normalized quantities given in Theorems 4.1–4.3 (equations (4.4), (4.9), and (4.13)) with $h = 0.5n^{-3/10}$, $d = 0$, and $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$.

normal, leading to bias distortion. The wild bootstrap procedure is adopted to improve the finite sample performance of these statistics. Our experiments focus on model M2 to provide more useful insight on the performance of the tests that are employed in the empirical work.

In the short memory case with $d = 0$, the bootstrap size and local power of the two statistics are reported in Table 5. The bandwidth is $h = n^{-1/3}$, which meets the requirements in Theorems 4.4 and 4.5. We consider $H_0 : \beta_0(x) = 1 + z, \beta_1(x) = 1$ and $m(x) = x^2$, $\rho_n = n^{-1/15}$, which satisfy the conditions in Remark 4.7. From Table 5, it is evident that both tests have good size and local power performance. The $\widetilde{T}_n$ statistic is slightly undersized when $n = 100$ but has good size when $n = 400$. The $T_{n,d}$ statistic seems somewhat more powerful than the $\widetilde{T}_n$ statistic. Importantly, dependence between $x_k$ and $y_k$ and endogeneity in $(x_k, y_k)$ have

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

**TABLE 5.** Finite sample size and local power of bootstrap procedure in model M2 with $d = 0$

| | $T_{n,d}$ | | | | | | $\widetilde{T}_n$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *size* | | | *Local power* | | | *Size* | | | *Local power* | | |
| $n$ | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| | | | | | | $(\theta, \lambda_1, \lambda_2) = (0,0,0)$ | | | | | | |
| 100 | 1.6 | 5.9 | 11.7 | 6.7 | 17.9 | 26.2 | 0.6 | 2.4 | 6.7 | 4.9 | 10.5 | 15.0 |
| 400 | 0.7 | 5.3 | 10.9 | 78.9 | 87.1 | 90.9 | 0.9 | 4.9 | 9.6 | 72.0 | 81.1 | 85.3 |
| | | | | | | $(\theta, \lambda_1, \lambda_2) = (1,0,0)$ | | | | | | |
| 100 | 0.7 | 5.5 | 10.0 | 7.3 | 16.7 | 24.3 | 0.5 | 3.0 | 6.1 | 5.2 | 11.6 | 17.3 |
| 400 | 1.1 | 6.0 | 10.5 | 78.7 | 87.0 | 90.0 | 0.6 | 4.9 | 9.5 | 73.1 | 82.3 | 85.9 |
| | | | | | | $(\theta, \lambda_1, \lambda_2) = (0,1,1)$ | | | | | | |
| 100 | 0.7 | 5.0 | 10.7 | 10.7 | 23.9 | 32.0 | 0.3 | 2.6 | 5.7 | 6.8 | 14.2 | 20.4 |
| 400 | 0.7 | 3.9 | 8.0 | 85.8 | 93.1 | 96.1 | 1.0 | 5.2 | 9.9 | 82.5 | 89.0 | 91.5 |
| | | | | | | $(\theta, \lambda_1, \lambda_2) = (1,1,1)$ | | | | | | |
| 100 | 0.6 | 5.3 | 11.4 | 9.8 | 21.3 | 29.2 | 0.4 | 2.9 | 5.9 | 7.0 | 13.9 | 20.2 |
| 400 | 1.1 | 5.1 | 10.4 | 85.9 | 92.9 | 94.4 | 0.8 | 4.7 | 10.1 | 80.5 | 87.7 | 90.2 |

negligible effects on test performance. These finite sample properties of the tests support the use of the bootstrap procedure in empirical work such as the climate data sensitivity application in Section 6.

We also examined the performance of the bootstrap procedure in the LM case with $d = 0.4$. This implementation works under the assumption of SM innovations (with $d = 0$), whereas the true generating mechanism has $d = 0.4$. The results are collected in Table 6 with $(\theta, \lambda_1, \lambda_2) = (0,0,0)$. Results for other combinations of $(\theta, \lambda_1, \lambda_2)$ are similar and are omitted. Evidently, both statistics are seen to be slightly undersized when $n$ is small. When $n = 400$, the statistic $T_{n,d}$ still suffers from undersizing but $\widetilde{T}_n$ has good size performance. This outcome matches the conjecture in Wang and Phillips (2016) that $T_{n,d}$ would become more conservative as $d$ increases. What may be unexpected in the findings is the high power performance. Since the statistics themselves are not designed to accommodate long memory, and the limit theory of the bootstrap is not investigated here, these properties are issues left for future research.

## 6. APPLICATION TO GLOBAL CLIMATE SENSITIVITY

As an application of our methods, we consider the problem of estimating Earth's climate sensitivity to a given increase in atmospheric $CO_2$ concentration. As described in Storelvmo et al. (2018), this is an issue on which there is much ongoing research, primarily with the use of large-scale global climate models. In these global climate modeling exercises, analysis relies on computer simulation data

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

**TABLE 6.** Finite sample size and local power performance in model M2 with $d = 0.4$ and $(\theta, \lambda_1, \lambda_2) = (0, 0, 0)$

| | $T_{n,d}$ | | | $\widetilde{T}_n$ | | |
|---|---|---|---|---|---|---|
| $n$ | 1% | 5% | 10% | 1% | 5% | 10% |
| | | | *Size* | | | |
| 100 | 0.7 | 3.6 | 8.5 | 0.3 | 3.2 | 7.9 |
| 200 | 0.6 | 3.3 | 8.9 | 0.9 | 4.9 | 10.3 |
| 400 | 0.3 | 4.4 | 8.2 | 1.4 | 5.4 | 11.1 |
| | | | *Local power* | | | |
| 100 | 75.8 | 87.0 | 91.2 | 79.9 | 85.2 | 87.9 |
| 200 | 97.9 | 99.9 | 99.9 | 99.3 | 99.4 | 99.6 |
| 400 | 99.0 | 99.7 | 99.7 | 100.0 | 100.0 | 100.0 |

generated from immensely detailed models of Earth's climate using an ensemble of initializations of the variables that help to assess model sensitivity. An alternative approach that instead relies on observational data is to use econometric methods to fit much simpler dynamic panel models from which parameter estimates may be obtained to assess the impact on climate of rising atmospheric $CO_2$ concentrations. The methodology has been developed in recent work by Magnus, Melenberg, and Muris (2011), Storelvmo et al. (2016), and Phillips, Leirvik, and Storelvmo (2020). An advantage of the approach is that observationally based confidence intervals may be constructed for the key parameters that are involved in measuring climate change dynamics and the long-term impact of rising $CO_2$ concentrations. Recent research (Storelvmo et al., 2025) indicates that the use of observational data, rather than climate model simulations from multiple ensembles, helps to narrow the confidence interval in the estimation of the Earth's climate sensitivity.

Our application here uses three observational data sets: temperature ($T_{it}$), surface level solar radiation ($R_{it}$), and $CO_2$ equivalent greenhouse gas concentrations ($CO_{2,t}$). The temperature and surface radiation data record time series at multiple surface stations and thereby conform to usual panel data with individual station and time series observations, whereas the $CO_2$ data varies only over the temporal dimension. Since a primary goal of empirical research with this data is to measure the recent historical impact of aggregate $CO_2$ on Earth's climatic temperature, the panel model framework necessarily involves the use of time and station level series in conjunction with possible communal variables that affect climate in aggregate. Such communal variables in the present case are aggregate $CO_2$ levels and aggregate solar radiation levels measured at the Earth's surface.

It is convenient in this application to use the same base dataset as in Phillips et al. (2020), which is recorded over the 42-year period from 1964 to 2005 over 1,484 land-based observation stations. Phillips et al. (2020) use aggregate data in their empirical work, basing their formulation on an aggregated version of the panel

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

model in Storelvmo et al. (2016) to focus directly on Earth's climate sensitivity to $CO_2$. As explained in Phillips et al. (2020), there are technical challenges in using a panel model (particularly nonstationarity in the data, co-relationships among the time series, and weak dependence in the innovations) to achieve consistent estimation and valid asymptotic inference of this global quantity. These considerations led to an aggregate model that was developed under specific conditions concerning aggregation (Phillips et al., 2020, Assumption A(i)–(v)). That model relates closely to the models given below in (6.1) and (6.2). Using these models and aggregated data in the present work enables comparisons of various nonlinear functional coefficient specifications with the linear specifications employed in Phillips et al. (2020). For example, we are able to assess whether the impact of $CO_2$ works nonlinearly through the functional coefficients or simply linearly as a common time effect regressor. For more information about the data, see Phillips et al. (2020) and Storelvmo et al. (2016) and the references therein.[5]

Figure 3 plots the aggregated time-series data.[6] An obvious feature of these aggregate data series is their nonstationarity. Trend characteristics of varying types are evident in the global temperature, radiation, and $CO_2$ series, with greater year-to-year volatility in temperature and radiation than in $CO_2$. These series were modeled in Phillips et al. (2020) by allowing for stochastic and deterministic linear trends as well as linear cointegrating linkages among the three aggregate series. As might be expected, much greater variation occurs in the disaggregated station-level data. Linear cointegrating regression analysis was used in Phillips et al. (2020) in studying the aggregate series and an asymptotic theory for the estimated coefficients was obtained, including asymptotics for a composite parameter that measures climate sensitivity to $CO_2$.

In Magnus et al. (2011) and Storelvmo et al. (2016), disaggregated station-level data were employed, standard linear dynamic panel within group and GMM methods were used in estimation, and no attention was paid to possible nonstationarity in the data or nonlinearity in the relationships. The present application utilizes the nonparametric regression modeling methodology developed here, allowing for the presence of functional regression coefficients that depend on an aggregate variable with nonstationary trending characteristics. The goal of the application is to assess the impact of the global influences on Earth's temperature of $CO_2$ and surface (downwelling) solar radiation and to determine whether linear specifications of the type used in earlier research are justified. The implementation of this article's methodology that follows specifically addresses nonstationarity in the aggregate data. So estimates and confidence intervals obtained in fitting the functional

[5]For a general discussion of Earth's climate sensitivity to greenhouse gases and for references to recent work in this field of climate science, readers are also referred to Storelvmo et al. (2018).

[6]Minor differences between the $CO_2$ graphic in Figure 3 and that shown in Phillips et al. (2020) arise because more decimal places are used here than in Phillips et al. (2020).

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

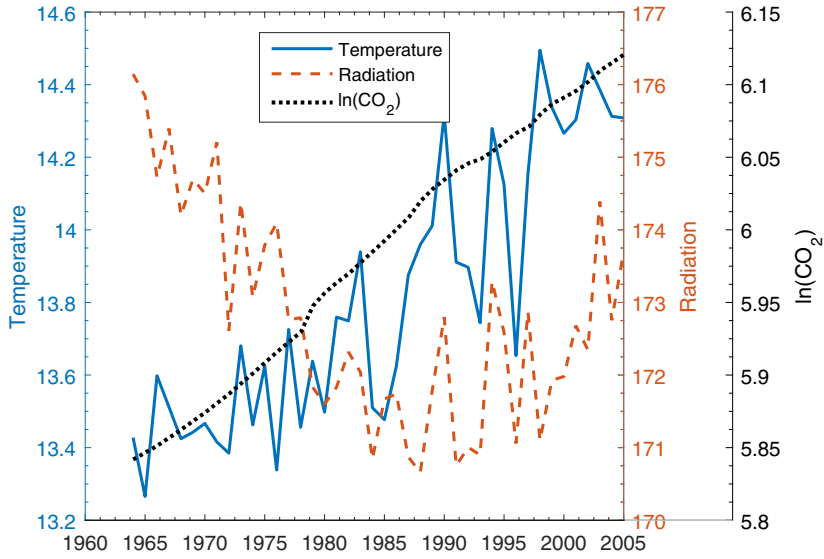



**FIGURE 3.** Time-series plot of station-averaged temperature ($\overline{T}_t$, °Celsius; blue solid), downward surface radiation ($\bar{R}_t$, Watts per $m^2$; red dashed), and logarithms of $CO_2$ (Pg: metric gigatons; black dotted with circle markers) ranging from 1964 to 2005.

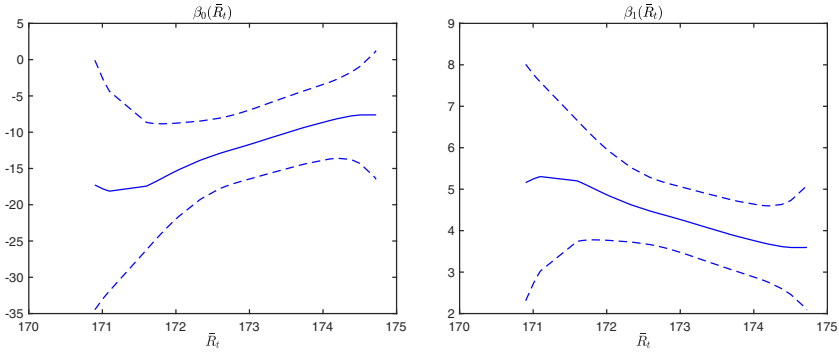
coefficient regression (6.1) in our empirical work below are supported by the theory of the present article.[7]

We first investigate the relation among these three variables using the aggregated time-series data plotted in Figure 3. The nonstationarity in each of these variables is plainly evident in the figure and confirmed in the analysis reported in Phillips et al. (2020). Both $\overline{T}_t$ and $\bar{R}_t$ were found to be well characterized as unit-root processes and $\ln CO_2$ as a unit-root process with drift. Upon suitable normalization, these variables therefore satisfy the requirements of Assumptions **A1**–**A3**.[8]

[7] But, as noted below, tests for linearity in the specification are not rigorously covered by the present theory.

[8] In the case of $CO_2$, the corresponding limit process under certain conditions is a Brownian motion with drift and in that case the limit theory of Theorem 4.1 holds with appropriate adjustment to the limit process. In particular, the Brownian motion limit process with drift is produced when the generating mechanism has a weak (local to zero) trend as in a specification of the form $\ln CO_{2,t} = \delta_n + \ln CO_{2,t-1} + e_t$ with $\delta_n = \frac{\delta}{\sqrt{n}}$ for some constant drift coefficient $\delta$ and stationary input process $e_t$. For the data used in the present study, $n = 42$ and $\frac{1}{n}\sum_{t=1}^{n} \Delta \ln(CO_{2,t}) = 0.0068$, giving an estimate $\hat{\delta} = 0.044$, which is in line with a local to zero trend specification. In such cases, the limit theory in Theorem 4.1 continues to hold in its present form. The reason is that the quadratic variation and hence the local time of a Brownian motion with drift, such as $B(r) + \delta r$, are the same as the quadratic variation and local time of the Brownian motion $B(r)$. If the discrete time drift is not local, then a linear or suitable time polynomial may be added to the functional coefficient regression given in (6.1) below, as discussed in Remark 4.5 with associated changes in the limit theory that involve $L_2$ projection residuals so that Theorem 4.3 applies. These adjustments are analogous to those in the analysis of Phillips et al. (2020) and simpler findings in the early unit-root linear cointegrating regression literature (Park and Phillips, 1988, 1989).

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press




(a) $\beta_0(\bar{R}_t)$ (b) $\beta_1(\bar{R}_t)$

**FIGURE 4.** Estimated functional coefficients of model (6.1).

To explore the sensitivity of temperature to $CO_2$, earlier work in the literature has used linear regression of temperature on $CO_2$ via linear modeling or linear cointegration modeling, as in Phillips et al. (2020). To allow for a possibly nonlinear coefficient impact of $CO_2$ on temperature, where aggregate levels of downwelling radiation may potentially affect the impact of $CO_2$, we use the functional coefficient regression formulation

$$\overline{T}_t = \beta_0(\bar{R}_t) + \beta_1(\bar{R}_t)\ln(CO_{2,t}) + e_t, \tag{6.1}$$

where $\overline{T}_t$ and $\bar{R}_t$ denote the cross-station average of $T_{it}$ and $R_{it}$, respectively. The estimated functional coefficients from local linear regression and the corresponding asymptotic 95% confidence bands based on the limit theory in Theorem 4.3 are plotted in Figure 4. The bandwidth used is $h = 2n^{-3/10}$, following the guidance obtained from Table 4 in the simulations. The fitted functional coefficient estimates show that the intercept function $\beta_0(\bar{R}_t)$ exhibits an upward trend with radiation, suggesting that the level impact on temperature rises with solar radiation, as expected. On the other hand, the slope coefficient function $\beta_1(\bar{R}_t)$ exhibits a downward trend, suggesting lower correlation between temperature and $CO_2$ when radiation is higher. This outcome may be partly explained by the fact that downwelling solar radiation rises when atmospheric conditions are clearer with less pollutants (like sulfur dioxide), and in such cases, the greenhouse gas effects of rising $CO_2$ on temperature may be attenuated because of greater infrared radiation into space and aerosol/cloud interactions (Wild, 2012).

To see whether the linear relation assumption between temperature and $CO_2$ is supported by the data, we consider the following partial linear model:

$$\overline{T}_t = \beta_0(\bar{R}_t) + \beta_1\ln(CO_{2,t}) + v_t. \tag{6.2}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

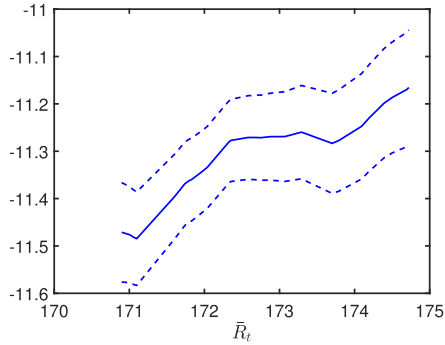



**FIGURE 5.** Estimated functional coefficient $\beta_0(\bar{R}_t)$ of model (6.2).

The estimated curve of $\beta_0(\bar{R}_t)$ is plotted in Figure 5 with 95% confidence bands, and the estimate of $\beta_1$ is 4.20 with the 95% confidence interval [3.47, 4.92].[9] The estimated curve $\beta_0(\bar{R}_t)$ in Figure 5 shows a clear upward trend, revealing strong positive effects from radiation to temperature. To test the constancy of $\beta_1(\bar{R}_t)$ in model (6.1), or equivalently, the linear relation assumption between temperature and $CO_2$, we employ the generalized likelihood ratio (GLR) test of Fan, Zhang, and Zhang (2001) to test the null model of (6.2) against the alternative model of (6.1). The test gives a $p$-value of 0.52,[10] suggesting a failure to reject the linear relationship between temperature and $CO_2$ that embodies the positive impact of radiation on temperature through the radiation-dependent intercept.

To further investigate the nature of the impact of radiation on temperature, we consider testing the constancy of $\beta_0(\bar{R}_t)$ in model (6.2). The GLR test gives a $p$-value of 0.019, suggesting rejection of the commonly used simple linear model $\overline{T}_t = \beta_0 + \beta_1 \ln(CO_{2,t}) + \epsilon_t$. This result strongly demonstrates the significant role of radiation on temperature. Moreover, from the estimated curve shown in Figure 5 for the intercept function, it seems that radiation has an approximate linear impact on temperature. We therefore proceed to test whether $\beta_0(\bar{R}_t)$ in model (6.2) can be accepted as a linear function of $\bar{R}_t$. That is, we test a linear model of the following form:

$$\overline{T}_t = \alpha_0 + \alpha_1 \bar{R}_t + \beta_1 \ln(CO_{2,t}) + \varepsilon_t, \tag{6.3}$$

against the partial linear model in (6.2). Using the GLR test gives a $p$-value of 0.75, suggesting that the linear specification in (6.3) adequately captures the dependence

[9]The confidence bands in Figure 5 and the confidence interval for $\beta_1$ are only approximate because the asymptotic theory for partial linear regression with nonstationary data as in model (6.2) is not strictly covered by the present theory. Indeed, according to the tests in Phillips et al. (2020), $\bar{R}_t$ has a stochastic trend and $\ln(CO_{2,t})$ has a stochastic trend with drift (cf. footnote 8).

[10]This $p$-value may not be reliable since the validity of the GLR test in the current nonstationary scenario has not been justified. This qualification applies to other GLR test results reported in the article.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

of temperature on radiation and $CO_2$. In addition, the two test statistics proposed in Section 4.1 were used to test the null model of (6.3) against the functional coefficient model (6.1). Both statistics returned $p$-values greater than 0.4, which again support the linear model specification in (6.3). The linear model in (6.3) has precisely the form of the global linear (cointegrating) relation among the variables that was used in Phillips et al. (2020).

These empirical findings show that functional coefficient regression models can be helpful in guiding and justifying parametric specifications when both nonstationary covariates and nonstationary regressors are present. The nonparametric regression results and functional specification tests obtained provide empirical support for the aggregate linear models that have recently been used in climate econometric research to study the impact on temperature of $CO_2$ greenhouse gases and downwelling solar radiation.

## 7. CONCLUSION

While this article adds to existing limit theory and widens the field of potential applications of nonlinear regressions with nonstationary time series, there is scope for further extension. Some additions that would extend the present theory involve relaxing the mds condition on the equation error $u_t$, developing bias reduction methods, and coping with further induced endogeneity. In the context of TVP cointegration, Phillips et al. (2017) developed a version of fully modified regression that worked successfully in that context. General forms of the regression model (4.1) with nonlinear functional coefficient vector $\beta(x_k)$ and stationary errors have considerably more complex elements and it seems that new methods may be needed to achieve these extensions. This work is left for future research.

Simulations are generally supportive of the functional coefficient limit theory in the presence of both a nonstationary regressor and smoothing covariate. But our findings reveal that finite sample performance is affected by reduced local signal strength and increased sensitivity to bandwidth choice. These issues are exacerbated in multiple regressor cases and in the use of local linear estimation. They arise from the random wandering nature of the covariate $x_k$ and its functional interaction with the nonstationary regressor $y_k$. The combined impact of this dual-source nonstationarity in regression is to reduce kernel-weighted signal strength in the relevant estimation locality by virtue of the nonparametric treatment of the functional coefficient. Small bandwidths typically reduce the number of relevant observations and raise the risk of singularity in the locally weighted signal matrix. Increasing the bandwidth helps to resolve the ill-posed inversion but also raises bias in the regression. Thus, while the originating source differs from existing ill-posed inverse problems in econometrics, some of the effects turn out to be similar. Addressing this issue and the curse of dimensionality in nonstationary nonparametric regression is a further challenge for future research.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

## 8. PROOFS

**Proof of Proposition 2.1.** It is routine to see that $var(y_n)/n \to \Gamma$, where $\Gamma = \alpha^{\mathrm{T}}\Psi\,\Omega\Psi^{\mathrm{T}}\alpha$. As for (2.2), by **A2′**(ii), we have

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_{-i},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_j\right)\Rightarrow Z_t := (B_{-t}, B_t, \Lambda_t), \tag{8.1}$$

where $Z = \{Z_t\}_{t\geq 0}$ is a $(2+d)$-dimensional Gaussian process with mean zero, independent increments, and covariance matrix

$$\Omega_t = t\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \gamma^{\mathrm{T}} \\ 0 & \gamma & \Omega \end{pmatrix}, \quad \text{where } \gamma = (\gamma_1, \ldots, \gamma_d)^{\mathrm{T}} \text{with } \gamma_k = E(\epsilon_0\eta_{0k}).$$

Therefore, it follows from **A2′**(i) and Phillips and Solo (1992) that

$$\frac{y_k}{\sqrt{n}} = \alpha^{\mathrm{T}}\Psi\,\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\eta_j + o_P(1)$$

and, on $D_{\mathbb{R}^3}[0,1]$,

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_{-i},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i,\ \frac{y_{[nt]}}{\sqrt{n}}\right)\Rightarrow(B_{-t},\ B_t,\ \alpha^{\mathrm{T}}\Psi\,\Lambda_t), \tag{8.2}$$

implying (2.2). □

**Proof of Theorem 3.1.** We only prove (3.1) since the joint convergence (3.2) is immediate after some routine notation changes in view of (2.3). We first assume that $g(x)$ is bounded and continuous on $\mathbb{R}$. This restriction will be removed later. Write

$$L_{1n,\epsilon} = \frac{1}{n}\sum_{k=1}^{n} g(Y_{nk})\,\phi_\epsilon(X_{nk})\int_{-\infty}^{\infty} f(x)dx,$$

$$L_{2n,\epsilon} = \frac{c_n}{n}\sum_{k=1}^{n} g(Y_{nk})\int_{-\infty}^{\infty} f\big[c_n(X_{nk}+z\epsilon)\big]\phi(z)dz,$$

where $\phi_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}}\exp\big\{-\frac{x^2}{2\epsilon^2}\big\}$ and $\phi(x) = \phi_1(x)$. Recalling (2.3), it follows from the continuous mapping theorem, the occupation time formula, and (2.6) with $a(x) = \phi(x)$ that

$$\begin{aligned} L_{1n,\epsilon} &\Rightarrow \int_0^1 g(Y_t)\,\phi_\epsilon(X_t)dt\int_{-\infty}^{\infty} f(x)dx \quad (\text{as } n\to\infty) \\ &= \int_{-\infty}^{\infty}\int_0^1 g(Y_t)\,\phi_\epsilon(x)\,L_X(dt,x)\,dx\int_{-\infty}^{\infty} f(x)dx \\ &\qquad (\text{by the occupation time formula}) \end{aligned}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$$= \int_{-\infty}^{\infty} \phi(x) \int_0^1 g(Y_t) L_X(dt, \epsilon x) dx \int_{-\infty}^{\infty} f(x) dx$$
$$\to_{a.s.} \int_0^1 g(Y_t) L_X(dt, 0) \int_{-\infty}^{\infty} \phi(x) dx \int_{-\infty}^{\infty} f(x) dx = \mathbb{S}, \quad \text{as } \epsilon \to 0.$$

Result (3.1) will follow if we prove: for any $\epsilon > 0$,

$$L_{1n,\epsilon} - L_{2n,\epsilon} = o_P(1), \tag{8.3}$$

and, as $n \to \infty$ followed by $\epsilon \to 0$,

$$S_n - L_{2n,\epsilon} = o_P(1). \tag{8.4}$$

We start with some basic preliminaries before proving (8.3) and (8.4). Recall **A1** and $\limsup_{t\to\infty} t^{\delta} |\mathbb{E}e^{it\epsilon_0}| < \infty$ for some $\delta > 0$. It follows from Theorem 2.18 of Wang (2015) that, for all $k \geq 1$, $x_k/d_k$ has a density $p_k(x)$ which is uniformly bounded by a constant $C$ and, conditional on $\mathcal{F}_{\epsilon j} := \sigma(\epsilon_j, \epsilon_{j-1}, \ldots)$, $(x_k - x_j)/d_{k-j}$ has a density $p_{kj}(x)$ satisfying

$$\sup_x |p_{kj}(x+u) - p_{kj}(x)| \leq C \min\{|u|, 1\}, \tag{8.5}$$

for all $j \geq 1$ and $k - j \geq n_0 \geq 1$, where the fixed constant $n_0$ is chosen so that $x_k - x_j \neq 0$. Since the finite items have no impact on the asymptotics, without loss of generality, we assume $n_0 = 1$ in the following proof. Using these facts, for any integrable function $l(x)$ and constant sequence $a_n > 0$, we have

$$\mathbb{E}|l(a_n X_{nk})| = \int_{-\infty}^{\infty} |l(a_n d_k d_n^{-1} x)| p_k(x) dx \ \leq \ C a_n^{-1} (d_n/d_k) \int_{-\infty}^{\infty} |l(x)| dx, \tag{8.6}$$

for all $k \geq 1$, and

$$\begin{aligned}
&\left|\mathbb{E}\left(\left\{l(a_n X_{nk}) - l[a_n(X_{nk} + z\epsilon)]\right\} \mid \mathcal{F}_{\epsilon j}\right)\right| \\
&= \left| \int_{-\infty}^{\infty} \left\{ l(a_n X_{nj} + \frac{a_n d_{k-j}}{d_n} y) - l\Big[a_n(X_{nj} + z\epsilon) + \frac{a_n d_{k-j}}{d_n} y\Big] \right\} p_{kj}(y) dy \right| \\
&= a_n^{-1} (d_n/d_{k-j}) \left| \int_{-\infty}^{\infty} \left\{ l(a_n X_{nj} + y) - l\big[a_n(X_{nj} + z\epsilon) + y\big] \right\} p_{kj}\Big(\frac{y d_n}{a_n d_{k-j}}\Big) dy \right| \\
&\leq a_n^{-1} (d_n/d_{k-j}) \int_{-\infty}^{\infty} |l(y)| \left| p_{kj}\Big(\frac{y - a_n X_{nj}}{a_n d_{k-j}/d_n}\Big) - p_{kj}\Big(\frac{y - a_n(X_{nj} + z\epsilon)}{a_n d_{k-j}/d_n}\Big) \right| dy \\
&\leq C a_n^{-1} (d_n/d_{k-j}) \min\{1, d_n |z| \epsilon / d_{k-j}\},
\end{aligned} \tag{8.7}$$

for all $j \geq 1$ and $k - j \geq 1$. Furthermore, by recalling that $a(k)$ is slowly varying at infinity, it follows from (2.1) that, for all $1 \leq l + 1 \leq j \leq n$,

$$\sum_{k=l+1}^{j} d_k^{-1} \leq C(j-l) d_j^{-1}, \qquad \sum_{k=1}^{j} d_k^{-2} \ \leq \ C \begin{cases} 1, & \text{under } \mathbf{LM}, \\ \log j, & \text{under } \mathbf{SM}, \end{cases} \tag{8.8}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

and, for all $l \le j$ and $j$ sufficiently large,

$$C_1 (l/j)^{1-\delta} \le d_l/d_j \le C_2 (l/j)^{1/2}, \tag{8.9}$$

for some $\mu - 1/2 > \delta > 0$, where $C_1 > 0$ and $C_2 > 0$ are two absolute constants.

Return to the proofs of (8.3) and (8.4). Equation (8.3) is simple. In fact, by the boundedness of $g(x)$ and the fact that

$$L_{2n,\epsilon} = \frac{1}{n} \sum_{k=1}^{n} g(Y_{nk}) \int_{-\infty}^{\infty} f(y) \phi_\epsilon (y/c_n - X_{nk}) dy,$$

it follows from (8.6) with $a_n = 1$ and (8.8) with $l = 0$ and $j = n$ that, for any $\epsilon > 0$,

$$\begin{aligned}
\mathbb{E}\,|L_{1n,\epsilon} - L_{2n,\epsilon}| &\le \frac{C}{n} \sum_{k=1}^{n} \int_{-\infty}^{\infty} |f(y)|\, \mathbb{E}\left|\phi_\epsilon (y/c_n - X_{nk}) - \phi_\epsilon (X_{nk})\right| dy \\
&\le \frac{C d_n}{n} \sum_{k=1}^{n} d_k^{-1} \int_{-\infty}^{\infty} |f(y)| \int_{-\infty}^{\infty} |\phi_\epsilon (y/c_n - x) - \phi_\epsilon (x)| dx\, dy \\
&\le C \Big[ \int_{(|y| \ge \sqrt{c_n})} |f(y)| dy + \int_{|y| < \sqrt{c_n}} |f(y)| \int_{-\infty}^{\infty} |\phi_\epsilon (y/c_n - x) - \phi_\epsilon (x)| dx\, dy \Big] \\
&\le C \Big\{ \int_{|y| \ge \sqrt{c_n}} |f(y)| dy + \big[ (\epsilon \sqrt{c_n})^{-1} + (\epsilon^2 c_n)^{-1} \big] \int_{-\infty}^{\infty} |f(y)| dy \Big\} \\
&\to 0, \quad \text{as } n \to \infty,
\end{aligned}$$

due to $c_n \to \infty$ and $\int_{-\infty}^{\infty} |f(y)| dy < \infty$, where we have used the fact: for any $\delta > 0$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} |\phi_\epsilon (\delta - x) - \phi_\epsilon (x)| dx \\
&\quad \le \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |e^{-(\delta - \epsilon x)^2/2\epsilon^2} - e^{-x^2/2}| dx \le C \big[ \delta/\epsilon + (\delta/\epsilon)^2 \big].
\end{aligned}$$

We next consider (8.4). First note that, by the tightness of $\{Y_{n,[nt]}\}_{0 \le t \le 1}$ and continuity of $g(x)$,

$$\max_{1 \le j \le n/m} \max_{(j-1)m+1 \le k \le jm} |g(Y_{nk}) - g(Y_{n,(j-1)m+1})| = o_P(1), \tag{8.10}$$

as $n \to \infty$ followed by $\epsilon \to 0$, where $m = \lfloor n/K_\epsilon \rfloor$ and $K_\epsilon \to \infty$ as $\epsilon \to 0$. We may take $K_\epsilon \to \infty$ sufficiently slow so that $\epsilon^\eta K_\epsilon \to 0$ for any $\eta > 0$. Let $T_n = \lfloor n/m \rfloor$. We may write

$$\begin{aligned}
S_n &= \frac{c_n}{n} \sum_{j=2}^{T_n} \sum_{k=(j-1)m+1}^{jm} g(Y_{nk}) f(c_n X_{nk}) + S_{2n,\epsilon} \\
&= \frac{c_n}{n} \sum_{j=2}^{T_n} g(Y_{n,(j-1)m+1}) \sum_{k=(j-1)m+1}^{jm} f(c_n X_{nk})
\end{aligned}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$$+\frac{c_n}{n}\sum_{j=2}^{T_n}\sum_{k=(j-1)m+1}^{jm}\big[g(Y_{nk})-g(Y_{n,(j-1)m+1})\big]f(c_nX_{nk})+S_{2n,\epsilon}$$

$$:=\frac{c_n}{n}\sum_{j=2}^{T_n}g(Y_{n,(j-1)m+1})\sum_{k=(j-1)m+1}^{jm}f(c_nX_{nk})+S_{1n,\epsilon}+S_{2n,\epsilon}, \tag{8.11}$$

where

$$|S_{2n,\epsilon}|\le\frac{Cc_n}{n}\Big(\sum_{k=1}^{m}+\sum_{k=T_nm+1}^{n}\Big)|f(c_nX_{nk})|.$$

It follows from (8.6) with $a_n=c_n$ and (8.8) and (8.9) that

$$\mathbb{E}|S_{2n,\epsilon}|\le Cn^{-1}d_n\Big(\sum_{k=1}^{m}d_k^{-1}+\sum_{k=n-m}^{n}d_k^{-1}\Big)$$
$$\le C_1m/n(1+d_n/d_m)\le C_1(m/n)^{1/2}\le C_1K_\epsilon^{-1/2},$$

i.e., $S_{2n,\epsilon}\to_P 0$ as $n\to\infty$ followed by $\epsilon\to 0$. Similarly, we have $\frac{c_n}{n}\sum_{k=1}^{n}\big|f(c_nX_{nk})\big|=O_P(1)$, from which it follows from (8.10) that

$$|S_{1n,\epsilon}|\le\max_{1\le j\le n/m}\max_{(j-1)m+1\le k\le jm}|Y_{nk}-Y_{n,(j-1)m+1}|\,\frac{c_n}{n}\sum_{k=1}^{n}\big|f(c_nX_{nk})\big|\to_P 0,$$

Taking these estimates into (8.11), we obtain

$$S_n=\frac{c_n}{n}\sum_{j=2}^{T_n}g(Y_{n,(j-1)m+1})\sum_{k=(j-1)m+1}^{jm}f(c_nX_{nk})+S_{n,\epsilon}, \tag{8.12}$$

where $S_{n,\epsilon}=o_P(1)$ as $n\to\infty$ followed by $\epsilon\to 0$. Similarly, we have

$$L_{2n,\epsilon}=\frac{c_n}{n}\sum_{j=2}^{T_n}g(Y_{n,(j-1)m+1})\sum_{k=(j-1)m+1}^{jm}\int_{-\infty}^{\infty}f\big[c_n(X_{nk}+z\epsilon)\big]\phi(z)dz+L_{n,\epsilon}, \tag{8.13}$$

where $L_{n,\epsilon}=o_P(1)$ as $n\to\infty$ followed by $\epsilon\to 0$.

In terms of (8.12) and (8.13), the boundedness of $g(x)$ and $\int_{-\infty}^{\infty}\phi(z)dz=1$, result (8.4) will follow if we prove:

$$\frac{c_n}{n}\sum_{j=2}^{T_n}|I_{jm}(\epsilon)|\to 0, \tag{8.14}$$

as $n\to\infty$ followed by $\epsilon\to 0$, where

$$I_{jm}(\epsilon)=\int_{-\infty}^{\infty}\sum_{k=(j-1)m+1}^{jm}\big\{f(c_nX_{nk})-f\big[c_n(X_{nk}+z\epsilon)\big]\big\}\phi(z)dz.$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

To this end, write $Z_{nk}(z,\epsilon) = f(c_n X_{nk}) - f[c_n(X_{nk} + z\epsilon)]$ and $\Lambda_{n,j}(z,\epsilon) = \sum_{k=(j-1)m+1}^{jm} Z_{nk}(z,\epsilon)$. Note that

$$\frac{c_n}{n}\sum_{j=2}^{T_n}\mathbb{E}\,|I_{jm}(\epsilon)| \le \sum_{j=2}^{T_n}\int_{-\infty}^{\infty}\frac{c_n}{n}\,\mathbb{E}\left|\Lambda_{n,j}(z,\epsilon)\right|\phi(z)\,dz. \tag{8.15}$$

The proof of (8.14) involves the calculations of $\mathbb{E}\left|\Lambda_{n,j}(z,\epsilon)\right|$ in cases that $|z| \ge \log\epsilon^{-1}$ or $|z| \le \log\epsilon^{-1}$ so that $\mathbb{E}\,|I_{jm}(\epsilon)|$ has a required order as $n \to \infty$ followed by $\epsilon \to 0$. Such a calculation is similar to (2.69) of Wang (2015) but with different details under the recent formation of $m = \lfloor n/K_\epsilon \rfloor$. In fact, it follows from (8.6) with $a_n = c_n$ and (8.8) and (8.9) that, uniformly for all $z \in R$ and $2 \le j \le T_n$,

$$\begin{aligned}\sum_{k=(j-1)m+1}^{jm}\left[\mathbb{E}\,|Z_{nk}(z,\epsilon)| + \mathbb{E}\,Z_{nk}^2(z,\epsilon)\right] &\le C c_n^{-1} d_n \sum_{k=(j-1)m+1}^{jm} d_k^{-1}\\ &\le C c_n^{-1} m d_n/d_{jm} \le C c_n^{-1} m K_\epsilon^{1-\delta}.\end{aligned} \tag{8.16}$$

Similarly, it follows from (8.6) with $a_n = c_n$ and (8.7)–(8.9) that, for $(j-1)m \le i \le jm$ and for any $|z| \le \log^{-1}\epsilon$ (letting $\sum_{k=i}^{l} = 0$ if $i > l$),

$$\begin{aligned}&\sum_{k=i+1}^{jm}\left|E\left\{Z_{nk}(z,\epsilon)\,Z_{ni}(z,\epsilon)\right\}\right|\\ &\le \mathbb{E}\left\{|Z_{ni}(z,\epsilon)|\sum_{k=i+1}^{jm}\left|\mathbb{E}\left(Z_{nk}(z,\epsilon)\mid\mathcal{F}_i\right)\right|\right\}\\ &\le C c_n^{-2}(d_n/d_i)\left[\sum_{k=i+1}^{i+\epsilon m} d_n/d_{k-i} + \epsilon\log^{-1}\epsilon\sum_{k=i+\epsilon m}^{jm}(d_n/d_{k-i})^2\right]\\ &\le C c_n^{-2}(d_n/d_i)\left[\epsilon\, m\, d_n/d_{\epsilon m} + \epsilon\, m\log^{-1}\epsilon\,(d_n/d_m)\,(d_n/d_{\epsilon m})\right]\\ &\le C c_n^{-2}(d_n/d_i)\, m\,\epsilon^{\delta}\log^{-1}\epsilon\, K_\epsilon^{2(1-\delta)},\end{aligned} \tag{8.17}$$

where we have used the facts: by (8.9), $d_n/d_m \le C(n/m)^{1-\delta} \le C K_\epsilon^{1-\delta}$ and

$$d_n/d_{\epsilon m} \le C\epsilon^{\delta-1}(n/m)^{1-\delta} \le C\epsilon^{\delta-1}K_\epsilon^{1-\delta}.$$

In terms of (8.16), uniformly for all $z \in R$ and $2 \le j \le T_n$, we have

$$\mathbb{E}\left|\Lambda_{n,j}(z,\epsilon)\right| \le \sum_{k=(j-1)m+1}^{jm}\mathbb{E}\,|Z_{nk}(z,\epsilon)| \le C c_n^{-1} m K_\epsilon^{1-\delta}.$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

It follows from (8.16) and (8.17) that, for any $|z| \le \log^{-1}\epsilon$ and $2 \le j \le T_n$,

$$\begin{aligned}
\big(\mathbb{E}\,|\Lambda_{n,j}(z,\epsilon)|\big)^2 \le \mathbb{E}\,|\Lambda_{n,j}(z,\epsilon)|^2 &= \mathbb{E}\Big[\sum_{k=(j-1)m+1}^{jm} Z_{nk}(z,\epsilon)\Big]^2 \\
&\le \sum_{k=(j-1)m+1}^{jm} \mathbb{E}Z_{nk}^2(z,\epsilon) + 2\sum_{i=(j-1)m+1}^{jm}\sum_{k=i+1}^{jm} \big|\mathbb{E}\big\{Z_{nk}(z,\epsilon)\,Z_{ni}(z,\epsilon)\big\}\big| \\
&\le C\,c_n^{-1} m K_\epsilon^{1-\delta} + C m c_n^{-2}\epsilon^{\delta}\log^{-1}\epsilon\, K_\epsilon^{2(1-\delta)} \sum_{k=(j-1)m+1}^{jm} d_n/d_i \\
&\le C\,c_n^{-1} m K_\epsilon^{1-\delta} + C m^2 c_n^{-2}\epsilon^{\delta}\log^{-1}\epsilon\, K_\epsilon^{3(1-\delta)}.
\end{aligned}$$

Taking these estimates into (8.15) and recalling that $K_\epsilon \to \infty$ is so slow so that $\epsilon^\eta K_\epsilon \to 0$ for any $\eta > 0$, we have

$$\begin{aligned}
\frac{c_n}{n}\sum_{j=1}^{T_n}\mathbb{E}\,|I_{jm}(\epsilon)| &\le \sum_{j=2}^{T_n}\int_{-\infty}^{\infty}\frac{c_n}{n}\mathbb{E}\,|\Lambda_{n,j}(z,\epsilon)|\,\phi(z)\,dz \\
&\le \frac{c_n}{n}\sum_{j=2}^{T_n}\int_{|z|\ge\log^{-1}\epsilon}\mathbb{E}\,|\Lambda_{n,j}(z,\epsilon)|\,\phi(z)\,dz \\
&\qquad + \frac{c_n}{n}\sum_{j=2}^{T_n}\int_{|z|<\log^{-1}\epsilon}\big(\mathbb{E}\,|\Lambda_{n,j}(z,\epsilon)|^2\big)^{1/2}\phi(z)\,dz \\
&\le C K_\epsilon^{1-\delta}\int_{|z|\ge\log^{-1}\epsilon}\phi(z)\,dz + \epsilon^{\delta/2}\log^{-1/2}\epsilon\, K_\epsilon^{3(1-\delta)/2} \to 0,
\end{aligned}$$

as $n \to \infty$ followed by $\epsilon \to 0$. This proves (8.14) and completes the proof of (3.1) under the additional assumption that $g(x)$ is bounded on $\mathbb{R}$.

Finally, we remove the additional restriction on the boundedness of $g(x)$. For each $N > 0$, $g_N(x) = g(x)\xi_N(x)$ with

$$\xi_N(x) = \begin{cases} 1 & |x| \le N \\ 2 - |x|/N & N < |x| < 2N \\ 0 & |x| \ge 2N. \end{cases}$$

Let $S_{n,N} = \frac{c_n}{n}\sum_{k=1}^n g_N(Y_{nk}) f(c_n X_{nk})$. Since $g_N(x)$ is bounded and continuous, the first part of the proof yields that, for each $N > 0$,

$$S_{n,N} \to_D \mathbb{S}^N := \int_0^1 g_N(Y_s)L_X(ds,0)\int_{-\infty}^{\infty} f(t)dt.$$

This implies (3.1) without assuming the boundedness of $g(x)$ since

$$\begin{aligned}
&\mathbb{P}\big(S_n \ne S_{n,N}\big) + \mathbb{P}\big(\mathbb{S} \ne \mathbb{S}^N\big) \\
&\quad \le \mathbb{P}\big(\max_{1\le k\le n}|Y_{nk}| \ge N\big) + \mathbb{P}\big(\sup_{0\le t\le 1}|Y_t| \ge N\big) \to 0,
\end{aligned}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

as $n \to \infty$ followed by $N \to \infty$, due to the facts that $Y_{n,[nt]} \Rightarrow Y_t$ on $D_{\mathbb{R}}[0,1]$ and that the limit Gaussian process $Y_t$ is path continuous. The proof of Theorem 3.1 is now complete. □

**Proof of Corollary 3.1.** As in Theorem 3.1, without loss of generality, we assume that $g(x)$ is bounded and continuous. Recall $\tilde{f}_1(x) = \mathbb{E}f_1(x, \lambda_m, \ldots, \lambda_0)dx$ and write

$$R_n = \frac{c_n}{n} \sum_{k=1}^{n} g(Y_{nk}) \big[ f_1\big(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}\big) - \tilde{f}_1(c_n X_{nk}) \big].$$

Since $\tilde{f}_1(x)$ is bounded and integrable, by Theorem 3.1, (3.5) will follow if we prove:

$$R_n = o_P(1). \tag{8.18}$$

To this end, we first note that (8.6) can be extended to include certain stationary variables, i.e., the following result still holds:

$$\mathbb{E}|l(a_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m})| \le C a_n^{-1}(d_n/d_k) \int_{-\infty}^{\infty} \mathbb{E}|l(x, \lambda_m, \lambda_{m-1}, \ldots, \lambda_0)|dx,$$

for any $k \ge m + m_0$ and some fixed $m_0 > 0$. See Lemma 7.1 of Wang et al. (2021), for instance. Now, the same argument as in the proof of (8.11) yields that

$$\begin{aligned} R_n &= \frac{c_n}{n} \sum_{j=2}^{T_n} g(Y_{n,(j-1)m+1}) \sum_{k=(j-1)m+1}^{jm} \big[ f_1\big(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}\big) - \tilde{f}_1(c_n X_{nk}) \big] \\ &\quad + R_{n,K}, \\ &:= \frac{m}{n} \sum_{j=2}^{T_n} g(Y_{n,(j-1)m+1}) \frac{c_n}{m} L_{mj} + R_{n,K}, \end{aligned} \tag{8.19}$$

where $m = \lfloor n/K \rfloor$, $T_n = \lfloor n/m \rfloor$, and $R_{n,K} \to_P 0$, as $(n,K)_{\text{seq}} \to \infty$. As in the proof of (7.37) with $i = 2$ in Wang et al. (2021), we have

$$\frac{c_n}{m} \mathbb{E}|L_{mj}| \le \big(\mathbb{E}\big|\frac{c_n}{m} L_{mj}\big|^2\big)^{1/2} \to 0,$$

uniformly for $2 \le j \le T_n \le K+1$ as $(n,K)_{\text{seq}} \to \infty$. It follows from the boundedness of $g(x)$ that

$$\frac{m}{n} \sum_{j=2}^{T_n} |g(Y_{n,(j-1)m+1})| \, \mathbb{E}\big|\frac{c_n}{m} L_{mj}\big| \le C \max_{2 \le j \le K+1} \frac{c_n}{m} \mathbb{E}|L_{mj}| \to 0,$$

i.e., $\frac{m}{n} \sum_{j=2}^{T_n} g(Y_{n,(j-1)m+1}) \frac{c_n}{m} L_{mj} \to_P 0$, as $(n,K)_{\text{seq}} \to \infty$. Taking this estimate into (8.19), we obtain (8.18), which completes the proof of Corollary 3.1. □

**Proof of Theorem 3.2.** We only prove (3.7) and assume $c_n' = 0$. The extension to general $c_n'$ is standard (e.g., WP2009a) and (3.6) is similar except simpler.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

Note that $M_{1n} = \sum_{k=1}^{n} m_{nk} u_k$, where $m_{nk} = \left(\frac{c_n}{n}\right)^{1/2} g(Y_{nk}) f_1(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m})$ depending only on $\lambda_k, \lambda_{k-1}, \ldots$. We may establish (3.7) by using the extended martingale limit theorem given in Wang (2014). To this end, let $\tilde{g}_\eta(x) = |g(x)|^{2+\eta}$ and $\tilde{f}_\eta(x,y) = |f_1(x,y)|^{2+\eta}$. Due to **A5**(i), there exists a bounded and integrable function $\tilde{T}(x)$ so that, for any $-1 \le \eta \le 1/\beta$,

$$|\tilde{f}_\eta(x,y)| \le \tilde{T}(x)(1+|y|^{(2+\eta)\beta}),$$

i.e., **A5**(i) still holds if we replace $f_1$ by $\tilde{f}_\eta$ and $\beta$ by $(2+\eta)\beta$. Since $\tilde{g}_\eta(x)$ is still continuous, it follows from Corollary 3.1 (with $g$ and $f_1$ replaced by $\tilde{g}_\eta$ and $\tilde{f}_\eta$, respectively) that

$$\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_{-k},\ X_{n,[nt]},\ Y_{n,[nt]},\ \frac{c_n}{n}\sum_{k=1}^{n}\tilde{g}_\eta(Y_{nk})\tilde{f}_\eta\big(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}\big)\Big)$$
$$\Rightarrow \Big(B_t,\ B_{-t},\ X_t,\ Y_t,\ \int_0^1 \tilde{g}_\eta(Y_s) dL_X(s,0)\, \mathbb{E}\tilde{f}_\eta(x, \lambda_m, \lambda_{m-1}, \ldots, \lambda_0) dx\Big), \qquad \textbf{(8.20)}$$

for any $-1 \le \eta \le 1/\beta$. Consequently, it is readily seen that

$$\max_{1\le k\le 1} |m_{nk}| \le \Big[\big(\frac{c_n}{n}\big)^{1+1/(2\beta)} \sum_{k=1}^{n} \tilde{g}_{1/\beta}(Y_{nk}) \tilde{f}_{1/\beta}\big(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}\big)\Big]^{2\beta/(1+2\beta)} = o_P(1)$$

and

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}|m_{nk}| = c_n^{-1/2}\frac{c_n}{n}\sum_{k=1}^{n}\tilde{g}_{-1}(Y_{nk})\tilde{f}_{-1}\big(c_n X_{nk}, \lambda_k, \ldots, \lambda_{k-m}\big) = o_P(1),$$

due to $c_n \to \infty$ and $c_n/n \to 0$. Result (3.7) now follows from Theorem 2.1 of Wang (2014). □

**Proof of Theorem 4.1.** We first prove (4.3). Let $D_n = \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K[(x_k - x)/h]$. We have

$$\hat{\beta}_n(x) - \beta(x) - hL_{1n}\beta'(x) = D_n^{-1}\sum_{k=1}^{n} l_k K\left(\frac{x_k - x}{h}\right) u_k + D_n^{-1} R_n, \qquad \textbf{(8.21)}$$

where, under **A6**(i),

$$||R_n|| \le \sum_{k=1}^{n} ||l_k l_k^{\mathrm{T}}|| \,\big|\big|\beta(x_k) - \beta(x) - \beta'(x)(x_k - x)\big|\big| K[(x_k - x)/h]$$
$$\le C h^{1+\nu} \sum_{k=1}^{n} ||l_k\, l_k^{\mathrm{T}}|| K\left(\frac{x_k - x}{h}\right).$$

Let $X_{nk} = x_k/d_n, c_n = d_n/h$, $\widetilde{Y}_{nk} = \alpha^{\mathrm{T}} l_k/\sqrt{n} = \sum_{s=1}^{k} \alpha^{\mathrm{T}} v_s/\sqrt{n}$, and $\widetilde{Y}_t = \alpha^{\mathrm{T}} \Psi\, \Lambda_t$, where $\alpha \in \mathbb{R}^p$ and $\Lambda_t$ is defined as in Proposition 2.1. It follows from Proposition 2.1, Theorem 3.2, and the continuous mapping theorem that

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$$\frac{c_n}{n^2}\alpha^{\mathrm{T}} D_n \alpha = \frac{c_n}{n}\sum_{k=1}^{n} \widetilde{Y}_{nk}^2 K[c_n(X_{nk}-x/d_n)] \to_D \int_0^1 \widetilde{Y}_t^2 L_X(dt,0), \tag{8.22}$$

indicating that $||D_n^{-1}|| = O_P(c_n/n^2)$ due to the fact that

$$\mathbb{P}\Big(\int_0^1 \widetilde{Y}_t^2 L_X(dt,0) \neq 0\Big) = 1, \text{ for any } \alpha \in \mathbb{R}^p \text{ and } \alpha \neq 0, \tag{8.23}$$

by using (8.1) and (8.2). Similarly, when $c_n \to \infty$ and $c_n/n \to 0$, it follows from Theorem 3.2 that

$$\frac{c_n}{n^2}\sum_{k=1}^{n} ||l_k l_k^{\mathrm{T}}|| K\left(\frac{x_k - x}{h}\right) = O_P(1),$$

and, jointly with (8.22),

$$\Big(\frac{c_n}{n^2}\Big)^{1/2}\sum_{k=1}^{n} \alpha^{\mathrm{T}} l_k K\left(\frac{x_k - x}{h}\right) u_k = \Big(\frac{c_n}{n}\Big)^{1/2}\sum_{k=1}^{n} \widetilde{Y}_{nk} K[c_n(X_{nk}-x/d_n)]\, u_k$$
$$\to_D \Big[\int_0^1 \widetilde{Y}_t^2 L_X(dt,0)\Big]^{1/2} \mathcal{N}(0,\sigma_\beta^2), \tag{8.24}$$

and, jointly with (8.22),

$$\frac{c_n}{n^2}\sum_{k=1}^{n} (\alpha^{\mathrm{T}} l_k)^2 K_1\big[(x_k - x)/h\big] = \frac{c_n}{n}\sum_{k=1}^{n} \widetilde{Y}_{nk}^2 K_1[c_n(X_{nk}-x/d_n)]$$
$$\to_D \int_0^1 \widetilde{Y}_t^2 L_X(dt,0) \int K_1(s)ds. \tag{8.25}$$

As a consequence, we have $D_n^{-1} R_n = O_P(h^{1+\nu})$,

$$D_n^{-1}\sum_{k=1}^{n} l_k u_k K\left(\frac{x_k - x}{h}\right) = O_P\big[(c_n/n^2)^{1/2}\big] = O_P\big[(d_n/n^2 h)^{1/2}\big] = o_P(h)$$

whenever $\min\{nh^2, nh/d_n\} \to \infty$, and $L_{1n} \to_P \mathbb{I} \int sK(s)ds$. Taking these estimates into (8.21), we obtain (4.3).

We next prove (4.4). Similar to (8.21), we have

$$D_n^{1/2}\Big[\hat{\beta}_n(x) - \beta(x) - hL_{1n}\beta'(x) - \frac{1}{2}h^2 L_{2n}\beta''(x)\Big]$$
$$= D_n^{-1/2}\sum_{k=1}^{n} l_k u_k K\left(\frac{x_k - x}{h}\right) + D_n^{-1/2} R_{1n}, \tag{8.26}$$

where, under **A6**(ii),

$$||R_n|| \le \sum_{k=1}^{n} ||l_k l_k^{\mathrm{T}}|| \Big|\beta(x_k) - \beta(x) - \beta'(x)(x_k - x) - \frac{1}{2}\beta''(x)(x_k - x)^2\Big| K[(x_k - x)/h]$$
$$\le C h^{2+\nu}\sum_{k=1}^{n} ||l_k l_k^{\mathrm{T}}|| K\left(\frac{x_k - x}{h}\right),$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

i.e., whenever $n^2h^{5+2\nu}/d_n \to 0$,

$$||D_n^{-1/2}R_{1n}|| \le O_P(1)\,h^{2+\nu}(n^2h/d_n)^{1/2} = o_P(1).$$

Since (8.24) holds jointly with (8.22), result (4.4) follows from (8.26) and the continuous mapping theorem. The proof of Theorem 4.1 is now complete. □

**Proof of Theorem 4.2.** Since

$$\widehat{\beta}_L(x) = \left(V_{n0} - V_{n1}V_{n2}^{-1}V_{n1}\right)^{-1}\sum_{k=1}^{n}\left[\mathbb{I} - V_{n1}V_{n2}^{-1}(x_k - x)\right]l_k z_k K\left(\frac{x_k - x}{h}\right),$$

we have

$$\sum_{k=1}^{n}\left[\mathbb{I} - V_{n1}V_{n2}^{-1}(x_k - x)\right]l_k l_k^{\mathrm{T}} K[(x_k - x)/h](x_k - x)$$
$$= \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K[(x_k - x)/h](x_k - x) - V_{n1}V_{n2}^{-1}\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K[(x_k - x)/h](x_k - x)^2 = 0,$$

where $V_{nj} = \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K\left(\frac{x_k - z}{h}\right)(x_k - x)^j$ for $j = 0, 1,$ and 2. Then, similar to $\hat{\beta}_n(x)$, we may write

$$\widehat{\beta}_L(x) - \beta(x) - \frac{1}{2}h^2\beta''(x)\int s^2K(s)ds = \Delta_n^{-1}\left[P_n + \frac{1}{2}h^2T_n\beta''(x) + R_{2n}\right], \qquad \textbf{(8.27)}$$

where $\Delta_n = V_{n0} - V_{n1}V_{n2}^{-1}V_{n1}$,

$$P_n = \sum_{k=1}^{n}\left[\mathbb{I} - V_{n1}V_{n2}^{-1}(x_k - x)\right]l_k K[(x_k - z)/h]u_k,$$

$$T_n = \sum_{k=1}^{n}\left[\mathbb{I} - V_{n1}V_{n2}^{-1}(x_k - x)\right]l_k l_k^{\mathrm{T}} K_2[(x_k - x)/h] - \Delta_n\int s^2K(s)ds$$
$$= \sum_{k=1}^{n}\left[\mathbb{I} - V_{n1}V_{n2}^{-1}(x_k - x)\right]l_k l_k^{\mathrm{T}}\widetilde{K}_2[(x_k - x)/h],$$

with $K_j(s) = s^jK(s), j = 1, 2,$ and $\widetilde{K}_2(s) = K_2(s) - K(s)\int K_2(t)dt$, and, under **A6**(ii),

$$||R_{2n}|| \le \sum_{k=1}^{n}||l_k l_k^{\mathrm{T}}||\,||\left[\mathbb{I} - V_{n1}V_{n2}^{-1}(x_k - x)\right]||$$
$$\times\left|\left|\beta(x_k) - \beta(x) - \beta'(x)(x_k - x) - \frac{1}{2}\beta''(x)(x_k - x)^2\right|\right|K[(x_k - x)/h]$$
$$\le Ch^{2+\nu}\sum_{k=1}^{n}||l_k\,l_k^{\mathrm{T}}||\,||\left[I - V_{n1}V_{n2}^{-1}(x_k - x)\right]||\,K[(x_k - x)/h].$$

We can now use Proposition 4.1, together with some similar arguments to those given in the proof of Theorem 4.1, to establish (4.9). Write $D_n = \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K[(x_k -$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$x)/h]$ and

$$D_{jn} = \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} K_j[(x_k - x)/h], \quad j = 1, 2.$$

As in the proof of Theorem 4.1, we have $D_n^{-1} + D_{2n}^{-1} = O_P(d_n/n^2h)$ and, for any function $m(x)$ having finite support,

$$\sum_{k=1}^{n} l_k l_k^{\mathrm{T}} m[(x_k - x)/h] = O_P(n^2h/d_n),$$
$$\sum_{k=1}^{n} l_k m[(x_k - x)/h]\, u_k = O_P\big[(n^2h/d_n)^{1/2}\big].$$

These facts, together with Proposition 4.1 (e.g., $D_{1n} = O_P\big[\sqrt{n}\,(n^2h/d_n)^{1/2}\big]$ as $\int K_1(t)dt = 0$ and $nh^4 \to 0$), yield that

$$\Delta_n = h^2 D_{2n} D_n - h^2 D_{1n}^2 = h^2 D_{2n} D_n \Big[1 + O_P\big(d_n/nh\big)\Big], \tag{8.28}$$

$$\begin{aligned} P_n &= h^2 D_{2n} \sum_{k=1}^{n} l_k K[(x_k - x)/h] u_k - h^2 D_{1n} \sum_{k=1}^{n} l_k K_1[(x_k - x)/h] u_k \\ &= h^2 D_{2n} \Big[ \sum_{k=1}^{n} l_k K[(x_k - x)/h] u_k - O_P(n^{1/2}) \Big] \end{aligned} \tag{8.29}$$

$$\begin{aligned} T_n &= \sum_{k=1}^{n} \big[\mathbb{I} - V_{n1} V_{n2}^{-1}(x_k - x)\big] l_k l_k^{\mathrm{T}} \widetilde{K}_2[(x_k - x)/h] \\ &= h^2 D_{2n} \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} \widetilde{K}_2[(x_k - x)/h] - h^2 D_{1n} \sum_{k=1}^{n} l_k l_k^{\mathrm{T}} \widehat{K}_2[(x_k - x)/h] \\ &\qquad \big[\widehat{K}_2(x) = x\widetilde{K}_2(x) \text{ and recall } \int \widetilde{K}_2(t)dt = 0\big] \\ &= h^2 D_{2n} O_P\big[\sqrt{n}\,(n^2h/d_n)^{1/2}\big], \end{aligned} \tag{8.30}$$

and

$$\begin{aligned} ||R_{2n}|| &\le C h^{2+\nu} \sum_{k=1}^{n} ||l_k l_k^{\mathrm{T}}|| \, || \big[\mathbb{I} - V_{n1} V_{n2}^{-1}(x_k - x)\big] || \, K\left(\frac{x_k - x}{h}\right) \\ &\le C h^{2+\nu} \Big[h^2 ||D_{2n}|| + h^2 ||D_{1n}||\Big] \sum_{k=1}^{n} ||l_k l_k^{\mathrm{T}}|| \, K\left(\frac{x_k - x}{h}\right) \\ &\le h^2 ||D_{2n}|| \, O_P\big[h^{2+\nu}\,(n^2h/d_n)\big]. \end{aligned} \tag{8.31}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

Taking these estimates into (8.27), we obtain that

$$
\begin{aligned}
&D_n^{1/2}\big[\widehat{\beta}_L(x)-\beta(x)-\frac{1}{2}h^2\beta''(x)\int s^2K(s)ds\big]\\
&\quad=\frac{D_n^{-1/2}}{1+O_P\big(d_n/nh\big)}\Big\{\sum_{k=1}^{n} l_kK[(x_k-x)/h]u_k+O_P\big[h^{2+\nu}\,(n^2h/d_n)\big]\\
&\qquad\quad+O_P(n^{1/2})+O_P\big[\sqrt{n}h^2\,(n^2h/d_n)^{1/2}\big]\Big\}\\
&\quad=\frac{D_n^{-1/2}}{1+O_P\big(d_n/nh\big)}\sum_{k=1}^{n} l_kK[(x_k-x)/h]u_k\\
&\qquad+O_P\big[(n^2h^{5+2\nu}/d_n)^{1/2}\big]+O_P(\sqrt{n}h^2)+O_P\big[(d_n/nh)^{1/2}\big]\\
&\qquad\to_D\sigma_\beta\,\mathbb{N}_p, \qquad\qquad (8.32)
\end{aligned}
$$

for any $h$ satisfying $n^2h^{5+2\nu}/d_n\to 0, nh^4\to 0$ and $nh/d_n\to\infty$, i.e., we have (4.9). The proof of Theorem 4.2 is now complete. □

**Proof of Proposition 4.1.** It suffices to show that, for any $\alpha\in R^p$,

$$
A_n:=\alpha^{\mathrm{T}}A_{n,m}\alpha=\sum_{k=1}^{n}\widetilde{Y}_{nk}^2\,m[(x_k-x)/h]=O_P\big[(nh/d_n)^{1/2}\big], \qquad (8.33)
$$

where $\widetilde{Y}_{nk}=\alpha^{\mathrm{T}}l_k/\sqrt{n}$. We start with some preliminaries. First recall that, conditional on $\mathcal{F}_{\epsilon j}=\sigma(\epsilon_j,\epsilon_{j-1},\ldots),(x_k-x_j)/d_{k-j}$ has a bounded density function $p_{kj}(x)$ satisfying (8.5) for all $j\ge 1$ and $k-j\ge n_0$ for some $n_0\ge 1$ as seen in the proof of Theorem 3.1. This yields the following, for all $j\ge 1$ and $k-j\ge n_0$,

$$
\begin{aligned}
\mathbb{E}\big(m[(x_k-x)/h]|\mathcal{F}_{\epsilon j}\big)&=\int_{-\infty}^{\infty}m\Big(\frac{yd_{k-j}}{h}+\frac{x_j-x}{h}\Big)p_{kj}(y)dy\\
&=\frac{h}{d_{k-j}}\int_{-\infty}^{\infty}m(y)\,p_{kj}\Big(\frac{yh}{d_{k-j}}-\frac{x_j-x}{h}\Big)dy.
\end{aligned}
$$

Then, by using (8.5) and $\int_{-\infty}^{\infty}m(y)dy=0$, we have

$$
\begin{aligned}
&\big|\mathbb{E}\big(m[(x_k-x)/h]|\mathcal{F}_{\epsilon j}\big)\big|\\
&\quad\le\frac{h}{d_{k-j}}\int_{-\infty}^{\infty}m(y)\,\Big|p_{kj}\Big(\frac{yh}{d_{k-j}}-\frac{x_j-x}{h}\Big)-p_{kj}\Big(-\frac{x_j-x}{h}\Big)\Big|\,dy\\
&\quad\le C\,\frac{h^2}{d_{k-j}^2}\int_{-\infty}^{\infty}|y\,m(y)|dy\le C_1\frac{h^2}{d_{k-j}^2}, \qquad (8.34)
\end{aligned}
$$

for all $j\ge 1$ and $k-j\ge n_0$.

We return to the proof of (8.34). For fixed $N\ge 1$, let $\widehat{Y}_{kN}=\widetilde{Y}_{nk}I(|\widetilde{Y}_{nk}|\le N)$ and $B_{nN}=\sum_{k=1}^{n}\widehat{Y}_{kN}^2\,m[(x_k-x)/h]$. Using the independence between $\widetilde{Y}_{nk}$ and $x_k$, (8.6)

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

and (8.34), we have

$$\begin{aligned}
\mathbb{E}B_{nN}^2 &= \sum_{k=1}^n \sum_{j=1}^n \mathbb{E}\Big\{\widehat{Y}_{kN}^2 \widehat{Y}_{jN}^2 m[(x_k-x)/h]m[(x_j-x)/h]\Big\} \\
&= \sum_{k=1}^n \sum_{j=1}^n \mathbb{E}\Big(\widehat{Y}_{kN}^2 \widehat{Y}_{jN}^2\Big)\mathbb{E}\Big\{m[(x_k-x)/h]m[(x_j-x)/h]\Big\} \\
&\le N^4\Big(\sum_{|k-j|\le n_0} \big|\mathbb{E}\, m[(x_k-x)/h]m[(x_j-x)/h]\big| \\
&\quad + 2\sum_{j=1}^n \sum_{k=j+n_0}^n \mathbb{E}\Big\{\big|m[(x_j-x)/h]\big|\,\big|\mathbb{E}\big\{m[(x_k-x)/h]|\mathcal{F}_{\epsilon j}\big\}\big|\Big\}\Big) \\
&\le CN^4\Big(h\sum_{|k-j|\le n_0} d_k^{-1} + h^2\sum_{j=1}^n \sum_{k=j+n_0}^n d_j^{-1}d_{k-j}^{-2}\Big) \\
&\le CN^4\Big(n_0 h\sum_{k=1}^n d_k^{-1} + h^2\sum_{j=1}^n d_j^{-1}\sum_{k=1}^n d_k^{-2}\Big) \\
&\le C_1N^4 nh/d_n\Big(1+h\sum_{k=1}^n d_k^{-2}\Big) \le C_2N^4 nh/d_n, \qquad \textbf{(8.35)}
\end{aligned}$$

where we have used (8.8), indicating that $h\sum_{k=1}^n d_k^{-2} \le Ch\log n = o(1)$ by $nh^4 \to 0$. Note that, as $n \to \infty$ first and then $N \to \infty$,

$$P\big(\max_{1\le k\le n}|\widetilde{Y}_{nk}| \ge N\big) \to P\big(\sup_{0\le t\le 1}|\widetilde{Y}_t| \ge N\big) \to 0,$$

since $\widetilde{Y}_{n,[nt]} \Rightarrow \widetilde{Y}_t$ on $D_{\mathbb{R}}[0,1]$ as given in the proof of Theorem 4.1. This, together with (8.35), implies that, for any $\epsilon > 0$, there exist $N_\epsilon$ and $K_\epsilon > 0$ so that

$$\begin{aligned}
P(A_n \ge K_\epsilon (nh/d_n)^{1/2}) &\le P(A_n \ne B_{nN_\epsilon}) + P(B_{nN_\epsilon} \ge K_\epsilon (nh/d_n)^{1/2}) \\
&\le P\big(\max_{1\le k\le n}|\widetilde{Y}_{nk}| \ge N_\epsilon\big) + C_2K_\epsilon^{-2}N_\epsilon^4 \le \epsilon
\end{aligned}$$

for all $n$ sufficiently large. This yields (8.33) and therefore completes the proof of Proposition 4.1. □

**Proof of Theorem 4.4.** Without loss of generality, assume that $R(\lambda,x) \equiv 0$ in (4.15). The extension to the general situation is standard and involves only routine calculations. Write $\Gamma_n = diag(1, d_n, v(d_{yn}))$, where $d_n$ and $d_{yn}$ are defined as in **A1** and **A2**. Note that

$$\hat{\theta}_n = \Big(\sum_{k=1}^n G_k^{\mathrm{T}}G_k\Big)^{-1}\sum_{k=1}^n G_k^{\mathrm{T}}z_k = \theta + \Big(\sum_{k=1}^n G_k^{\mathrm{T}}G_k\Big)^{-1}\sum_{k=1}^n G_k^{\mathrm{T}}u_k\,. \qquad \textbf{(8.36)}$$

It is readily seen by standard arguments that

$$\Gamma_n(\hat{\theta}_n - \theta) = O_P(n^{-1/2}). \qquad \textbf{(8.37)}$$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

Using (8.37) and recalling (4.14), we have

$$T_n = \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] \big[ u_k + (\theta - \hat{\theta}_n) G_k \big] \Big\}^2 \pi(x) dx$$
$$= T_{1n} + T_{2n} + T_{3n}, \tag{8.38}$$

where $T_{1n} = \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] u_k \Big\}^2 \pi(x) dx$, $|T_{3n}|^2 \le 4 T_{1n} T_{2n}$ by Hölder's inequality and

$$T_{2n} = \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] \big[ (\theta - \hat{\theta}_n) G_k \big] \Big\}^2 \pi(x) dx$$
$$= O_P(n^{-1}) \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] \, \big\| \Gamma_n^{-1} G_k \big\| \Big\}^2 \pi(x) dx.$$

In terms of (8.38), the result (3.3) will follow if we prove

$$\frac{d_n}{nh} T_{1n} \to_D \tau_0 L_X(1,0) \tag{8.39}$$

and

$$\Delta_n := \Big(\frac{d_n}{nh}\Big)^2 \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] \, \big\| \Gamma_n^{-1} G_k \big\| \Big\}^2 \pi(x) dx = O_P(1). \tag{8.40}$$

The approach to proving (8.39) is the same as that in the proof of Proposition 7.3 of Wang and Phillips (2016) with minor modifications, so the details are omitted. To show (8.40), let $\widehat{G}_k = G_k I(|x_k|/d_n \le N, |y_k|/d_{yn} \le N)$ and

$$\widehat{\Delta}_n = \Big(\frac{d_n}{nh}\Big)^2 \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] \, \| \| \Gamma_n^{-1} \widehat{G}_k \| \| \Big\}^2 \pi(x) dx.$$

Recall that $h(y_k) = v(d_{yn}) H(y_k/d_{yn})$ and $H(x)$ is a continuous function. It is readily seen that, for each $N \ge 1$,

$$\widehat{\Delta}_n \le C_N \Big(\frac{d_n}{nh}\Big)^2 \int_{-\infty}^{\infty} \Big\{ \sum_{k=1}^{n} K[(x_k - x)/h] \Big\}^2 \pi(x) dx = O_P(1),$$

where $C_N$ is a constant depending only on $N$ and we have used (7.12) with $m(x) = 1$ in Proposition 7.3 of Wang and Phillips (2016). This implies (8.40) since, by **A1** and **A2**,

$$P(\Delta_n \ne \widehat{\Delta}_n) \le P(\max_{1 \le k \le n} |x_k| \ge d_n N) + P(\max_{1 \le k \le n} |y_k| \ge d_{yn} N) \to 0,$$

as $N \to \infty$.The proof of Theorem 4.4 is now complete. □

**Proof of Theorem 4.5.** Without loss of generality, assume that $\sigma^2 = 1$ in **A4** and $R(\lambda, x) \equiv 0$ in (4.15). As in the proof of Theorem 4.4, by letting $\Gamma_n =$

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$diag(1, d_n, v(d_{yn}))$, we have

$$\begin{aligned}\widetilde{T}_n =& 2\sum_{k=2}^{n} u_k A_{nk} + 2(\theta-\hat{\theta})\sum_{\substack{j,k=1\\ j\neq k}}^{n} u_j G_k K\big[(x_k-x_j)/h\big] \\ &+ \sum_{\substack{j,k=1\\ j\neq k}}^{n} (\theta-\hat{\theta}_n) G_k (\theta-\hat{\theta}_n) G_j K\big[(x_k-x_j)/h\big] \\ =& 2\sum_{k=2}^{n} u_k A_{nk} + O_P(n^{-1/2})||\widetilde{T}_{1n}|| + O_P(n^{-1})\widetilde{T}_{2n}, \end{aligned} \tag{8.41}$$

where $A_{nk} = \sum_{j=1}^{k-1} u_j K\big[(x_k-x_j)/h\big]$, $\widetilde{T}_{1n} = \sum_{\substack{j,k=1\\ j\neq k}}^{n} u_j \Gamma_n^{-1} G_k K\big[(x_k-x_j)/h\big]$, and

$$|\widetilde{T}_{2n}| \leq \sum_{\substack{j,k=1\\ j\neq k}}^{n} ||\Gamma_n^{-1} G_k||\,||\Gamma_n^{-1} G_j||\, K\big[(x_k-x_j)/h\big].$$

Similarly, we have

$$\begin{aligned}\widetilde{\sigma}_n^2 =& \sum_{\substack{j,k=1\\ j\neq k}}^{n} u_k^2 u_j^2 K^2\big[(x_k-x_j)/h\big] + 2\sum_{\substack{j,k=1\\ j\neq k}}^{n} (\hat{u}_k^2 - u_k^2)\hat{u}_j^2 K^2\big[(x_k-x_j)/h\big] \\ :=& \widetilde{\sigma}_{1n}^2 + O(n^{-1/2})\widetilde{\sigma}_{2n}^2, \end{aligned} \tag{8.42}$$

where

$$\begin{aligned}\widetilde{\sigma}_{2n}^2 \leq& \sum_{\substack{j,k=1\\ j\neq k}}^{n} ||\Gamma_n^{-1} G_k||\,|\hat{u}_k + u_k|\,\hat{u}_j^2 K^2\big[(x_k-x_j)/h\big] \\ \leq& 4\sum_{\substack{j,k=1\\ j\neq k}}^{n} ||\Gamma_n^{-1} G_k||\big[|u_k| + O_P(n^{-1/2})||\Gamma_n^{-1} G_k||\big] \\ &\big[|u_j|^2 + O_P(n^{-1})||\Gamma_n^{-1} G_j||^2\big] K^2\big[(x_k-x_j)/h\big]. \end{aligned}$$

Recalling $\Gamma_n^{-1} G_k = (1, x_k/d_n, H(y_k/d_{yn})^{\mathrm{T}}$ and $d_n^2 \asymp n$ under **SM**, we may prove

$$||\widetilde{T}_{1n}|| = O_P\big(n^{5/4}h^{3/4}\big), \tag{8.43}$$

$$\widetilde{T}_{2n} + \widetilde{\sigma}_{2n}^2 = O_P(n^{3/2}h). \tag{8.44}$$

To this end, let $\widehat{H}_N(x) = H(x)I(|x| \leq N)$ and $\widehat{G}_k = (1, x_k/d_n, \widehat{H}_N(y_k/d_{yn}))^{\mathrm{T}}$. Since $\{x_k\}_{k\geq 1}$ is independent of $\{y_k\}_{k\geq 1}$, the same argument as in the proof of Proposition 6.2 of Wang and Phillips (2012) yields that, for each $N \geq 1$,

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

$$
\begin{aligned}
||\widehat{T}_{1n}|| &:= || \sum_{\substack{j,k=1\\ j\neq k}}^{n} u_j \Gamma_n^{-1} \widehat{G}_k K\big[(x_k - x_j)/h\big]|| \\
&\leq C n^{-1/2} || \sum_{\substack{j,k=1\\ j\neq k}}^{n} u_j x_k K\big[(x_k - x_j)/h\big]|| \\
&\quad + || \sum_{\substack{j,k=1\\ j\neq k}}^{n} u_j \big(1, \widehat{H}_N(y_k/d_{yn})\big)^{\mathrm{T}} K\big[(x_k - x_j)/h\big]|| \\
&= O_P\big(n^{5/4} h^{3/4}\big).
\end{aligned}
$$

This implies (8.43), i.e., $||\widetilde{T}_{1n}|| = O_P\big(n^{5/4}h^{3/4}\big)$, since, by **A2**,

$$
\mathbb{P}\big(\widetilde{T}_{1n} \neq \widehat{T}_{1n}\big) \leq \mathbb{P}\big(\max_{1\leq k\leq n} |y_k/d_n| \geq N\big) \to 0,
$$

as $(n,N)_{\mathrm{seq}} \to \infty$. The proof of (8.44) is similar. In fact, it follows from Proposition 6.1 of Wang and Phillips (2012) that, for each $N \geq 1$,

$$
\begin{aligned}
||\widehat{T}_{1n}|| :=& || \sum_{\substack{j,k=1\\ j\neq k}}^{n} ||\Gamma_n^{-1} \widehat{G}_k||\, ||\Gamma_n^{-1} \widehat{G}_j||\, K\big[(x_k - x_j)/h\big]|| \\
&\leq \sum_{\substack{j,k=1\\ j\neq k}}^{n} \big(1 + N + |x_k|/\sqrt{n}\big) K\big[(x_k - x_j)/h\big] = O_P(n^{3/2}h),
\end{aligned}
$$

indicating $\widetilde{T}_{2n} = o_P(n^{3/2}h)$. Similarly, we also have $\widetilde{\sigma}_{2n}^2 = O_P(n^{3/2}h)$.

Taking (8.43) and (8.44) into (8.41) and (8.42), we obtain

$$
\widetilde{T}_n = 2 \sum_{k=2}^{n} u_k A_{nk} + O_P\big[(nh)^{3/4}\big] \tag{8.45}
$$

and

$$
\widetilde{\sigma}_n^2 = \widetilde{\sigma}_{1n}^2 + O_P(nh) = 2 \sum_{t=2}^{n} A_{nt}^2 + o_P(n^{3/2}h), \tag{8.46}
$$

where we have used (3.6) of Wang and Phillips (2012) in the second step of (8.46). Result (4.21) now follows from (8.45)- and (8.46) and Theorem 3.3 of Wang and Phillips (2012). □

## 9. BOOTSTRAP PROCEDURE

This section gives the bootstrap procedure used in Section 5.1. The steps in the procedure for statistic $T_{n,0}$ are as follows:

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

- Step (i): Based on the observed sample $\{z_k, x_k, y_k\}_{k=1}^n$, estimate the null restricted model and obtain the residuals $\hat{u}_k = z_k - G_k^{\mathrm{T}}\hat{\theta}_n$, where $\hat{\theta}_n$ is an LS estimator of $\theta$ under the null. Compute the test statistic $T_{n,0}$ based on equation (4.19).
- Step (ii): Construct the bootstrap residuals $\{u_k^\star\}$ by $u_k^\star = \hat{u}_k e_k$, where $e_k$ follows iid standard normal. Compute $z_k^\star = G_k^{\mathrm{T}}\hat{\theta}_n + u_k^\star$ for $k = 1, \ldots, n$.
- Step (iii): Based on the bootstrap sample $\{z_k^\star, x_k, y_k\}_{k=1}^n$, compute the bootstrap statistic $T_{n,0}^\star$ as in Step (i).
- Step (iv): Repeat Steps (ii) and (iii) a large number of times, say, $B$ times, and obtain $\{T_{n,0}^{\star(b)}\}_{b=1}^B$. Obtain the empirical upper $\alpha$-percentile of the bootstrap statistic, $c_\alpha$. Reject the null hypothesis if $T_{n,0} > c_\alpha$.

The bootstrap procedure for $\widetilde{T}_n$ is similar to the above procedure, except for the last step where the null hypothesis is rejected if $|\widetilde{T}_n| > c_{\alpha/2}$ since $\widetilde{T}_n$ is a two-sided test.

## COMPETING INTEREST STATEMENT

The authors declare that no competing interests exist.

## FUNDING STATEMENT

Peter C.B. Phillips acknowledges research support from the NSF (Grant No. SES 18-50860), a Kelly Fellowship at the University of Auckland, and an LKC Fellowship at Singapore Management University. Ying Wang acknowledges research support from the National Natural Science Foundation of China (Grant Nos. 72103197, 72573174, and 72373150).

### *REFERENCES*

Buchmann, B., & Chan, N. H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Annals of Statistics*, 35, 2001–2017.

Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148, 101–113.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66* (Vol. 66). CRC Press.

Fan, J., Zhang, C., & Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics*, 29, 153–193.

Gao, J., & Phillips, P. C. B. (2013). Functional coefficient nonstationary regression. Cowles Foundation Discussion Paper No. 1911.

Gao, J., Tjøstheim, D., & Yin, J. (2012). Model specification between parametric and nonparametric cointegration. Working Paper.

Hall, P., & Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics*, 33, 2904–2929.

Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21, 1926–1947.

Hirukawa, M., & Sakudo, M. (2018). Functional-coefficient cointegration models in the presence of deterministic trends. *Econometric Reviews*, 37, 507–533.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

Jeganathan, P. (2004). Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *Annals of Probability*, 32, 1771–1795.

Karlsen, H. A., & Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics*, 29, 372–416.

Li, K., Li, D., Liang, Z., & Hsiao, C. (2017). Estimation of semi-varying coefficient models with nonstationary regressors. *Econometric Reviews*, 36, 354–369.

Liang, H., Shen, Y., &Wang, Q. (2023). Functional-coefficient cointegrating regression with endogeneity. In Y. Chang, J. Sokbae Lee, & I. Miller (Eds.), *Essays in honor of Joon Y. Park: Econometric theory* (pp. 372–416). Emerald Publishing Limited.

Magnus, J. R., Melenberg, B., & Muris, C. (2011). Global warming and local dimming: The statistical evidence. *Journal of the American Statistical Association*, 106, 452–464.

Park, J., & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4, 468–498.

Park, J. Y., & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, 5, 95–131.

Park, J. Y., & Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of nonlinear time series. *Econometric Theory*, 15, 269–298.

Park, J. Y., & Phillips, P. C. B. (2000). Nonstationary binary choice. *Econometrica*, 68, 1249–1280.

Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, 69, 117–161.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33, 311–340.

Phillips, P. C. B. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory*, 25, 1466–1497.

Phillips, P. C. B., Leirvik, T., & Storelvmo, T. (2020). Econometric estimates of Earth's transient climate sensitivity. *Journal of Econometrics*, 214, 6–32. Annals Issue: Econometric Models of Climate Change.

Phillips, P. C. B., Li, D., & Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196, 180–195.

Phillips, P. C. B., & Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, 20, 971–1001.

Phillips, P. C. B., & Wang, Y. (2022). Functional coefficient panel modeling with communal smoothing covariates. *Journal of Econometrics*, 227, 371–407.

Phillips, P. C. B., & Wang, Y. (2023). When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Journal of Econometrics*, 232, 469–489.

Pötscher, B. M. (2004). Nonlinear functions and convergence to Brownian motion: Beyond the continuous mapping theorem. *Econometric Theory*, 20, 1–22.

Storelvmo, T., Heede, U. K., Leirvik, T., Phillips, P. C. B., Arndt, P., & Wild, M. (2018). Lethargic response to aerosol emissions in current climate models. *Geophysical Research Letters*, 45, 9814–9823.

Storelvmo, T., Leirvik, T., Lohmann, U., Phillips, P. C. B., & Wild, M. (2016). Disentangling greenhouse warming and aerosol cooling to reveal earth's climate sensitivity. *Nature Geoscience*, 9, 286–289.

Storelvmo, T., Yuan, M., Leirvik, T., Alterskjær, K., Phillips, P. C. B., & Smith, C. (2025). Assessing the robustness and implications of econometric estimates of climate sensitivity. *Environmental Research Letters*, 20, 24055.

Sun, Y., Cai, Z., & Li, Q. (2013). Semiparametric functional coefficient models with integrated covariates. *Econometric Theory*, 29, 659–672.

Tu, Y., & Wang, Y. (2019). Functional coefficient cointegration models subject to time–varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, 8, 1401–1423.

Tu, Y., & Wang, Y. (2020). Adaptive estimation of heteroskedastic functional-coefficient regressions with an application to fiscal policy evaluation on asset markets. *Econometric Reviews*, 39, 299–318.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press

Tu, Y., & Wang, Y. (2022). Spurious functional-coefficient regression models and robust inference with marginal integration. *Journal of Econometrics*, 229, 396–421.
Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory*, 30, 509–535.
Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression*. World Scientific.
Wang, Q., Lin, Y.-X., & Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, 19, 143–164.
Wang, Q., & Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25, 710–738.
Wang, Q., & Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, 77, 1901–1948.
Wang, Q., & Phillips, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory*, 27, 235–259.
Wang, Q., & Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, 40, 727–758.
Wang, Q., & Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory*, 32, 359–401.
Wang, Q., & Phillips, P. C. B. (2023). Optimal bandwidth selection in nonlinear cointegrating regression. *Econometric Theory*, 39, 1325–1337.
Wang, Q., & Phillips, P. C. B. (2024). A general limit theory for nonlinear functionals of nonstationary time series. *Econometric Theory*, 1–58, DOI: https://doi.org/10.1017/S0266466624000276.
Wang, Q., Phillips, P. C. B., & Kasparis, I. (2021). Latent variable nonparametric cointegrating regression. *Econometric Theory*, 37, 138–168.
Wild, M. (2012). Enlightening global dimming and brightening. *Bulletin of the American Meteorological Society*, 93, 27–37.
Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152, 81–92.

https://doi.org/10.1017/S0266466626100358 Published online by Cambridge University Press