WHEN BIAS CONTRIBUTES TO VARIANCE: TRUE LIMIT THEORY IN FUNCTIONAL COEFFICIENT COINTEGRATING REGRESSION

By

Peter C.B. Phillips

COWLES FOUNDATION PAPER NO. 1821

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2022

http://cowles.yale.edu/

Journal of Econometrics 232 (2023) 469–489

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

When bias contributes to variance: True limit theory in functional coefficient cointegrating regression[☆]

Peter C.B. Phillips [a,b,c,d,*], Ying Wang [e]

[a] University of Auckland, New Zealand
[b] Yale University, United States of America
[c] Singapore Management University, Singapore
[d] University of Southampton, United Kingdom
[e] Renmin University of China, China

ARTICLE INFO

Article history:
Received 5 November 2020
Received in revised form 19 September 2021
Accepted 19 September 2021
Available online 20 October 2021

JEL classification:
C14
C22

Keywords:
Bandwidth selection
Bias variability
Functional coefficient cointegration
Kernel regression
Nonstationarity

ABSTRACT

Limit distribution theory in the econometric literature for functional coefficient cointegrating regression is incorrect in important ways, influencing rates of convergence, distributional properties, and practical work. The correct limit theory reveals that components from both bias and variance terms contribute to variability in the asymptotics. The errors in the literature arise because random variability in the bias term has been neglected in earlier research. In stationary regression this random variability is of smaller order and can be ignored in asymptotic analysis but not without consequences for finite sample performance. Implications of the findings for rate efficient estimation are discussed. Simulations in the Online Supplement provide further evidence supporting the new limit theory in nonstationary functional coefficient regressions.

1. Introduction

Nonlinearities and parameter instabilities are commonly encountered phenomena in empirical research with both cross section and time series data. Modeling strategies in both cases have accordingly moved towards accommodating these features. A convenient mechanism for accomplishing such extensions is the use of functional coefficient (FC) regressions, which allow responses to explanatory variables to change in a systematic fashion according to movements in other relevant variables.

FC regression has provided a particularly useful tool for modeling comovement among nonstationary time series that may depart from strict parametric cointegration while retaining the essential property of stationary departures from long run linkages that characterize the data. Such functional coefficient cointegration (FCC) models were introduced in Xiao (2009). They embody notions of equilibrium that allow for responsive adjustment in the relationship to changes that occur

☆ Our thanks go to Serena Ng and the Associate Editor for helpful guidance on the revision and two referees for useful comments on directions for further research. The authors acknowledge Marsden Grant 16-UOA-239 support at the University of Auckland. Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 at Yale University and a Kelly Fellowship at the University of Auckland. Wang acknowledges support from the National Natural Science Foundation of China (Grant 72103197) and Beijing municipal fund for building world-class universities(disciplines) of Renmin University of China.

* Corresponding author at: Yale University, United States of America.
E-mail addresses: peter.phillips@yale.edu (P.C.B. Phillips), wangyingstat@gmail.com (Y. Wang).

https://doi.org/10.1016/j.jeconom.2021.09.007
0304-4076/

over time in relevant covariates. For instance, investment portfolios may realign in response to movements in interest rates or certain financial indices; or asset prices may relate to market fundamentals in a flexible manner that allows for the impact of relevant covariates, such as the profitability of alternative investments. In the last decade, models of this type have attracted much attention in the econometric literature, providing a flexible generalization of the cointegration concept and enabling econometric tests of strict fixed coefficient cointegration specifications in empirical work.

The prototypical FC model has the following form

$$y_t = x_t'\beta(z_t) + u_t \tag{1.1}$$

where the regressor x_t is a $d \times 1$ possibly nonstationary time series, the covariate z_t is a $q \times 1$ stationary time series and the error term u_t is a scalar stationary error process. This model has been extensively studied in the literature. An early paper by Cai et al. (2000) examined the stationary x_t case, Juhl (2005) examined the unit root autoregressive case, Xiao (2009) studied the model (1.1) with full rank $I(1)$ x_t, and Cai et al. (2009) allowed both $I(0)$ and $I(1)$ variables in x_t. Subsequent papers have developed specification tests for constant coefficients or strict cointegration (Sun et al., 2016), models with non-cointegrated structure (Sun et al., 2011), and applications where time varying volatility is relevant (Tu and Wang, 2019).

In all of this work, kernel weighted local least squares regression is employed to estimate the functional coefficient $\beta(\cdot)$. The derivation of the limit theory for these estimates follows standard lines for kernel regression asymptotics that were developed in the stationary case, while allowing for possible nonstationarity in the regressor x_t or certain components of x_t. In the prototypical case the limit theory is given as mixed normal and the results have been extensively used in the literature to develop test procedures for constant coefficients, confidence intervals for the functional coefficients, optimal bandwidth selection, and specification testing.

The present work shows that the limit theory given in this literature is incorrect in all cases where nonstationary regressors of integrated or near-integrated form are present in x_t. The errors originate from a missing term in the true limit theory that is associated with the random variability of the kernel regression bias. In stationary regression this term can be neglected as of smaller order than the usual variance expression. But in nonstationary regression, variability in the bias has larger order due to the nonstationary regressor. Its omission leads to failure in the reported asymptotic theory and the true limit distribution of the kernel regression estimator involves component elements from both the bias and the variance. The authors are not aware of other cases where such failures occur in nonparametric estimation, which helps to explain why the omission of potential bias effects on variance in nonparametric estimation has not been noticed in previous work.

In the present context of nonstationary regression, only in scalar FCC regression and only when the bandwidth is very small, viz., $o(1/\sqrt{n})$, are present results in the literature correct. That bandwidth restriction when z_t is scalar actually excludes optimal convergence, which occurs at the $n^{3/4}$ rate and requires the bandwidth setting $O(1/\sqrt{n})$ in estimation of $\beta(\cdot)$. Instead, optimal convergence leads to a limit distribution whose variance combines random elements from both the bias and variance terms in the regression. In short, we show that terms normally taken as 'bias' actually contribute to 'variance' and affect estimation and inference in material ways that have been neglected in earlier work.

The problem that arises in the existing limit theory can be explained simply in the model (1.1) when x_t is a scalar exogenous regressor, z_t is an independent univariate stationary process with smooth density $f(z)$, and u_t is a scalar stationary error process with zero mean and variance σ_u^2. The local level least squares estimate of $\beta(z)$ is $\hat{\beta}(z) = \left(\sum_{t=1}^n x_t y_t K_{tz}\right) / \sum_{t=1}^n x_t^2 K_{tz}$ for some suitable kernel function $K_{tz} = K((z_t - z)/h)$ with bandwidth h in the weighted regression. The estimate $\hat{\beta}(z)$ satisfies the usual decomposition into 'bias' and 'variance' terms, which in signal-normalized form is

$$\left(\sum_{t=1}^n x_t^2 K_{tz}\right)\left(\hat{\beta}(z) - \beta(z)\right) = \sum_{t=1}^n x_t^2[\beta(z_t) - \beta(z)]K_{tz} + \sum_{t=1}^n x_t u_t K_{tz}. \tag{1.2}$$

Limit theory is developed by analyzing each term on the right side of this equation in turn, as well as the behavior of the kernel weighted signal function $\sum_{t=1}^n x_t^2 K_{tz}$. To do so in a rigorous way requires the further decomposition of the right side as follows

$$\sum_{t=1}^n x_t^2 \mathbb{E}\xi_{\beta t} + \sum_{t=1}^n x_t^2 \eta_t + \sum_{t=1}^n x_t u_t K_{tz} \tag{1.3}$$

where $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz}$ and $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$. In (1.3), the first term in the decomposition leads in the conventional way to the 'deterministic' bias term[1] in the limit theory. The second term leads to a random element in the limit that is induced by the bias. It is neglect of this random element $\sum_{t=1}^n x_t^2 \eta_t$ that leads to the error in the literature. The relative magnitudes of the terms in (1.3) change for stationary and nonstationary regressors, as is now explained. But, as we will show in simulations, both terms are important in finite samples and affect inferential accuracy.

[1] The designation 'deterministic' is used because the bias term is actually deterministic in the limit only in the stationary case. In nonstationary regressor cases, the bias term has random elements that are induced by the asymptotic behavior of sample moments of the regressor x_t, which can influence the limit theory.

(i) Stationary x_t

In this case, the random element $\sum_{t=1}^{n} x_t^2\eta_t$ in the bias is of smaller order than the variance component in the third term $\sum_{t=1}^{n} x_t u_t K_{tz}$ of (1.3). In particular, under standard regularity conditions, the three components of (1.3) and the signal function have the following asymptotic behavior (c.f., Li and Racine (2007), theorem 9.3) as $n \to \infty$ and $h \to 0$ with the effective sample size $nh \to \infty$

$$\frac{1}{nh}\sum_{t=1}^{n} x_t^2 \mathbb{E}\xi_{\beta t} = h^2\sigma_x^2\mu_2(K)C(z) + o_p(h^2), \tag{1.4}$$

$$\sum_{t=1}^{n} x_t^2\eta_t = O_p(\sqrt{nh^3}), \tag{1.5}$$

$$\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_t u_t K_{tz} \rightsquigarrow \mathcal{N}\left(0,\, \nu_0(K)\sigma_u^2\sigma_x^2 f(z)\right), \tag{1.6}$$

$$\frac{1}{nh}\sum_{t=1}^{n} x_t^2 K_{tz} \to_p \sigma_x^2 f(z), \tag{1.7}$$

where $\mu_2(K) = \int s^2K(s)ds$, $\nu_0(K) = \int K(s)^2 ds$, $\int K(s)ds = 1$, $\mathbb{E}x_t^2 = \sigma_x^2$, $\mathbb{E}u_t^2 = \sigma_u^2$, $C(z) = \frac{1}{2}\beta^{(2)}(z)f(z) + \beta^{(1)}(z)f^{(1)}(z)$ with $g^{(j)}$ signifying the jth derivative of g, and where (1.5) is due to the fact that $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t \rightsquigarrow B_\eta(\cdot)$, as is shown in Lemma B.1(b) in Appendix B. Here and throughout the paper we use $\rightsquigarrow$ to signify weak convergence on the relevant probability space and $\to_p$ for convergence in probability. In view of (1.2) through (1.7), we can write

$$\begin{aligned}\hat{\beta}(z) - \beta(z) &= \frac{\sum_{t=1}^{n} x_t^2\mathbb{E}\xi_{\beta t} + \sum_{t=1}^{n} x_t^2\eta_t + \sum_{t=1}^{n} x_t u_t K_{tz}}{\sum_{t=1}^{n} x_t^2 K_{tz}} \\ &= O_p\left(\frac{nh^3 + \sqrt{nh^3} + \sqrt{nh}}{nh}\right) = O_p(h^2) + O_p(1/\sqrt{nh})\end{aligned} \tag{1.8}$$

because $\sqrt{nh^3} = o(\sqrt{nh})$. The standard FC regression limit theory follows, viz.,

$$\sqrt{nh}\left(\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z) + o_p(h^2)\right) = \frac{\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_t u_t K_{tz}}{\frac{1}{nh}\sum_{t=1}^{n} x_t^2 K_{tz}} + o_p(1) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu_0(K)\sigma_u^2}{\sigma_x^2 f(z)}\right), \tag{1.9}$$

giving the usual $\sqrt{nh}$ convergence rate for the suitably centered FC estimator $\hat{\beta}(z)$, the deterministic recentering bias function $h^2\mathcal{B}(z) = h^2\mu_2(K)C(z)/f(z)$, and a limiting normal distribution with variance $\nu_0(K)\frac{\sigma_u^2}{\sigma_x^2 f(z)}$. Notice that the second component of the bias, $\sum_{t=1}^{n} x_t^2\eta_t = O_p(\sqrt{nh^3})$, is $o_p(nh^3)$ provided $nh^3 \to \infty$, which holds for the usual optimal bandwidth choice $h = O(n^{-\frac{1}{5}})$ in stationary FC regression. Moreover, as evident in (1.5), the second component is $o_p(\sqrt{nh})$ whenever $h \to 0$, thereby ensuring that it is dominated by the variance term. In this stationary case, therefore, the random component of the bias function does not affect either the bias or the variance in the limit distribution of $\hat{\beta}(z)$.

(ii) Nonstationary x_t

In the nonstationary case with integrated or near-integrated regressor x_t the orders of magnitude of the components (1.4)–(1.7) change in critical ways that affect the balance in these components, thereby impacting the asymptotic behavior of $\hat{\beta}(z)$. First, nonstationarity in the regressor x_t changes signal strength. When x_t is a scalar unit root process and $nh \to \infty$ we have, from Lemma B.1(c)(i) in Appendix B,

$$\frac{1}{n^2h}\sum_{t=1}^{n} x_t^2 K_{tz} = \frac{1}{nh}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2 K_{tz} \rightsquigarrow \int B_x^2 \times f(z) \tag{1.10}$$

in place of (1.7), where $\frac{1}{\sqrt{n}}x_{\lfloor n\cdot\rfloor} \rightsquigarrow B_x(\cdot)$, Brownian motion with variance ω_x^2, and $\int$ denotes $\int_0^1$. In view of the standardization in (1.10), the FCC regression signal $\sum_{t=1}^{n} x_t^2 K_{tz}$ has stochastic order $O_p(n^2h)$ rather than $O_p(nh)$ and the requirement for consistency might therefore be thought to be $n\sqrt{h} \to \infty$ rather than $nh \to \infty$; or, upon appropriate standardization of x_t, the adjusted regression signal is $\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2 K_{tz} = O_p(nh)$, suggesting that the usual effective sample size condition $nh \to \infty$ is needed for consistency. However, the situation is considerably more subtle, as will be discussed in the paper: in fact, consistency continues to hold even when $nh \to \infty$ fails, as will be demonstrated in the paper in Theorem 2.2 and the following Remarks. Importantly, this is not the case when x_t is stationary as discussed in Remark 2.9.

Second, as shown in Lemma B.1(d) in Appendix B, the random element in the bias component now converges to a stochastic integral at the rate $O_p(\sqrt{n^3h^3})$. Specifically, when x_t is a scalar unit root process and $nh \to \infty$, we have

$$\frac{1}{\sqrt{n^3h^3}} \sum_{t=1}^{n} x_t^2 \eta_t = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}}\right)^2 \frac{\eta_t}{\sqrt{nh^3}} \rightsquigarrow \int B_x^2 dB_\eta, \tag{1.11}$$

where $B_\eta(\cdot)$ is the Brownian motion limit of the partial sum process $\frac{1}{\sqrt{nh^3}} \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t$, as shown in Lemma B.1(b). The deterministic component of the bias is $O_p(n^2h^3)$ and satisfies

$$\frac{1}{n^2h^3} \sum_{t=1}^{n} x_t^2 \mathbb{E}\xi_{\beta t} = \frac{1}{nh^3} \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}}\right)^2 \mathbb{E}\xi_{\beta t} \rightsquigarrow \left(\int B_x^2\right) \mu_2(K) C(z),$$

analogous to the stationary case but with $\int B_x^2$ replacing σ_x^2. Vital for the correct limit theory, the variance component $\sum_{t=1}^{n} x_t u_t K_{tz} = O_p(n\sqrt{h})$ turns out to be dominated by the random component of the bias because $n\sqrt{h} = o(\sqrt{n^3h^3})$ whenever $nh^2 \to \infty$, i.e. whenever $h \to 0$ slower than $1/\sqrt{n}$. Importantly, this result and (1.10) hold whenever $nh \to \infty$. Similar results apply with suitable changes in the limit formulae when x_t is near integrated.

It follows that the random bias (second) component of the decomposition (1.3) dominates the variance (third) term and therefore determines the form of the limit distribution of $\hat{\beta}(z)$ whenever $h \to 0$ slower than $1/\sqrt{n}$, which is the usual case in kernel regression. When $h \to 0$ at precisely the $1/\sqrt{n}$ rate both the random bias and variance terms contribute to the asymptotics. This balance in the components of (1.3) is explored rigorously in what follows and the asymptotic consequences are given in Theorem 2.1 for the scalar z_t case. In Section 2 of the Online Supplement that accompanies the paper, we report simulations that show the relevance of these analytic findings on the relative magnitude of the components in (1.3) in finite samples. These computations highlight the differences between the stationary and nonstationary cases for practical work and the dominating role the random bias component plays when $h \to 0$ slower than $1/\sqrt{n}$. For multivariate FC regression with vector z_t, the limit theory is given in Theorem 2.3. This case involves further complications and is not a straightforward extension of the scalar covariate case, as might be inferred from the present literature. We therefore deal with the multivariate z_t case separately in the following development. The optimal bandwidth order and asymptotics of rate efficient estimation are also discussed.

To keep the exposition brief and focus on correcting limit theory in the literature, we confine analysis to local level estimation and work with the prototypical model (1.1). This model might be considered too simple to be empirically relevant; but it embodies the central characteristics of FC regression and enables us to focus on the key issue, which is to demonstrate the impact that kernel regression bias can have on variance even in the limit theory. Primary attention is given to the nonstationary case where x_t is a full rank integrated process independent of z_t and u_t but attention is also given to the stationary regressor case. More general cases with serially dependent errors, potentially cointegrated regressors, and endogeneity do not change the basic thrust of the present findings and full extensions to such cases are left for future work. The new limit theory is derived in Section 2, with attention given separately to univariate and multivariate z_t cases, and further attention to the implications of the asymptotics for rate efficient estimation. Section 3 concludes. Proofs of the asymptotic results for univariate z_t are given in Appendix A and a key technical lemma is given in Appendix B. Proofs of the asymptotic results for multivariate z_t and further simulations are provided in the Online Supplement that accompanies the paper – see Appendix C for details.

Throughout the paper we use the notation $\mu_j(K) = \int_{\mathcal{K}} u^j K(u)du$, and $\nu_j(K) = \int_{\mathcal{K}} u^j K^2(u)du$ for kernel moment functions, where $\mathcal{K}$ is the support of the kernel function K. The affix 'q' when it appears in μ_j^q and ν_j^q is used to indicate the dimension of z_t in the multivariate case. For any random variables ξ_n and η_n, $\xi_n \sim_a \eta_n$ means ξ_n and η_n are asymptotically equivalent, namely $\xi_n = \eta_n\{1+o_p(1)\}$ as $n \to \infty$. We use $\equiv_d$ to signify equivalence in distribution and, as above and unless otherwise indicated, $\int$ denotes $\int_0^1$. According to the context, we use $:=$ and $=:$ to signify definitional equality.

2. FC limit theory in cointegrated systems

We consider a cointegrating equation model with full rank $I(1)$ regressors and functional coefficients dependent on a stationary covariate. The model matches that of Xiao (2009) and is a prototype of more complex systems, including models with endogenous cointegrated regressors, models with both $I(0)$ and $I(1)$ or near integrated regressors, and models with functionally cointegrated regressors, as well as serially dependent errors. The analysis here is representative of the complexities that are involved in all these more complex triangular systems of cointegrated equations. The purpose of the present paper is to derive the correct limit theory for the prototype model as a foundation for subsequent analyses of more complex systems.

2.1. Univariate z_t

We first derive limit theory for the FC kernel estimator $\hat{\beta}(z) = \left(\sum_t x_t x_t' K_{tz}\right)^{-1} \left(\sum_t x_t y_t K_{tz}\right)$ in model (1.1) with univariate z_t. To avoid unnecessary complications in the asymptotics, it is convenient to use the following simplifying assumptions. Extensions to more general cases are discussed below but these are not needed for the purposes of the present contribution.

Assumption 1.

(i) $\{x_t\}$ is a full rank unit root process, with innovations $u_{xt} = \Delta x_t$ and initialization $x_0 = o_p(\sqrt{n})$, satisfying the functional law $\frac{1}{\sqrt{n}} x_{\lfloor n\cdot \rfloor} \rightsquigarrow B_x(\cdot)$, where B_x is vector Brownian motion with variance matrix $\Omega_{xx} > 0$; $\{u_t\}$ is a martingale difference sequence (mds) with respect to the filtration $\mathcal{F}_t = \sigma\{\{x_s, z_s\}_{s=1}^{\infty}; u_t, u_{t-1}, \ldots\}$, $\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2$ *a.s.* and $\mathbb{E}(u_t^6) < \infty$; and $\{z_t, u_{xt}\}$ are strictly stationary α-mixing processes with mixing numbers $\alpha(j)$ that satisfy $\sum_{j\geq 1} j^c[\alpha(j)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$ with finite moments of order $p > 2\delta > 4$.

(ii) The density $f(z)$ of z_t and joint density $f_{0,j}(s_0, s_j)$ of (z_t, z_{t+j}) are bounded above and away from zero over their supports with uniformly bounded and continuous derivatives to the second order.

(iii) $\{x_t\}$ and $\{z_t\}$ are mutually independent.

(iv) The kernel function $K(\cdot)$ is a bounded probability density function symmetric about zero with support $\mathcal{K}$ that is either $[-1, 1]$ or $\mathbb{R} = (-\infty, \infty)$.

(v) $\beta(z)$ is a smooth function with uniformly bounded continuous derivatives to the second order and $\mathbb{E}\|\beta(z_t)\|^2 + \mathbb{E}\|\beta^{(1)}(z_t)\|^2 + \mathbb{E}\|\beta^{(2)}(z_t)\|^2 < \infty$.

(vi) $n \to \infty$ and $h \to 0$.

The functional law in Assumption (i) is made for convenience and is assured by many primitive conditions (e.g., Phillips and Solo (1992)). The mds condition in (i) and the independence condition in (iii) are also convenient for the limit theory in the nonstationary case. They may be relaxed at the cost of technical complications but these would distract from the central purpose of the paper and are not pursued here. The α-mixing condition for $\{z_t, u_{xt}\}$ is a standard weak dependence condition that is useful in kernel regression and functional limit theory. Condition (iv) is standard, although relaxation of the symmetry condition leads to some changes in the results. In some cases where the bandwidths employed are very small it is convenient to use kernels whose support $\mathcal{K}$ is the entire real line $\mathbb{R}$, and this will be mentioned as required. The moment conditions (v) on $\beta(z_t)$ and the first two derivatives, $\{\beta^{(1)}(z_t), \beta^{(2)}(z_t)\}$ are needed for the limit theory developed below. Condition (vi) places minimal requirements on (n, h) and the following development uses various additional conditions. For instance, as discussed earlier in the context of the asymptotic behavior of the kernel weighted regression signal in (1.10) and that of the random component of the bias in (1.11), the effective sample size rate condition $nh \to \infty$ is needed for explicit limit results, just as it is in stationary nonparametric and functional coefficient regression. The effects on the various kernel weighted sample moments of relaxing this particular condition are explored in the technical derivations and are discussed in the paper. Other rate conditions are employed as needed.

Our first result gives limit theory for the FCC regression estimator $\hat{\beta}(z)$ in model (1.1) under specific conditions on the bandwidth in relation to the sample size.

Theorem 2.1. *Under Assumption 1, when $nh \to \infty$, the following hold:*

(a) if $nh^2 \to 0$,

$$n\sqrt{h}\left[\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z)\right] \rightsquigarrow \mathcal{MN}(0, \Omega_u(z)), \tag{2.1}$$

where $\mathcal{B}(z) = \mu_2(K)C(z)/f(z)$ and $C(z) = \frac{1}{2}\beta^{(2)}(z)f(z) + \beta^{(1)}(z)f^{(1)}(z)$;

(b) if $nh^2 \to \infty$ and $\beta^{(1)}(z) \neq 0$,

$$\sqrt{\frac{n}{h}}\left[\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z)\right] \rightsquigarrow \left(f(z)\int B_xB_x'\right)^{-1}\left(\int B_xB_x'dB_\eta\right) \equiv_d \mathcal{MN}(0, \Omega_\beta(z)), \tag{2.2}$$

where $B_\eta(\cdot)$ is Brownian motion with variance matrix $V_{\eta\eta} = \nu_2(K)f(z)\beta^{(1)}(z)\beta^{(1)}(z)'$;

(c) if $nh^2 \to c$ for some constant $c \in (0, \infty)$ and $\beta^{(1)}(z) \neq 0$,

$$n^{3/4}\left[\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z)\right] \rightsquigarrow \mathcal{MN}\left(0, c^{\frac{1}{2}}\Omega_\beta(z) + \frac{1}{c^{\frac{1}{2}}}\Omega_u(z)\right). \tag{2.3}$$

The (conditional) variance matrices in (2.1) and *(2.2) are as follows:*

$$\Omega_u(z) = \nu_0(K)\sigma_u^2 f^{-1}(z)\left(\int B_xB_x'\right)^{-1}, \tag{2.4}$$

$$\Omega_\beta(z) = \frac{\nu_2(K)}{f(z)}\left(\int B_xB_x'\right)^{-1}\left(\int B_xB_x'\left(B_x'\beta^{(1)}(z)\right)^2\right)\left(\int B_xB_x'\right)^{-1}. \tag{2.5}$$

Remark 2.1 (*Case (a)*)**.** (i) Case (a) is the result given in Xiao (2009) but without the condition $nh^2 \to 0$ that is made explicit here. As the proof of Theorem 2.1 makes clear, the limit theory (2.1) holds only when $nh^2 \to 0$, which requires a small bandwidth that goes to zero faster than $1/\sqrt{n}$. The proof of the theorem depends on the additional rate condition $nh \to \infty$, which is needed to establish central limit theory and functional laws that are given in all items labeled (i) of Lemma B.1 for kernel weighted partial sums of various time series. This condition is the usual effective sample size assumption made in kernel regression for stationary time series.

(ii) Nonstationarity of x_t raises the signal strength of the regression signal in (1.10), which leads to the $O(n\sqrt{h})$ convergence rate for $\hat{\beta}(z)$ given in (2.1). Consistency and some limit theory for $\hat{\beta}(z)$ may be expected to hold even when $h = o(1/n)$ and the usual effective sample size requirement $nh \to \infty$ fails. More extreme situations of such small bandwidths are considered below in Theorem 2.2.

(iii) When $nh^2 \to 0$ as in case (a), the bias term in the centering of $\hat{\beta}(z)$ in (2.1) is negligible and can be ignored in the limit theory since $n\sqrt{h} \times h^2 = o(nh^2) \to 0$. Further, when $h = o(1/\sqrt{n})$ the convergence rate of $\hat{\beta}(z)$ is $n\sqrt{h} = o(n^{\frac{3}{4}})$, and thereby always less than the optimal rate, which is shown to be $O(n^{\frac{3}{4}})$ in Case (c) under the additional condition $\beta^{(1)}(z) \neq 0$ on the derivative of the functional coefficient.

Remark 2.2 (*Case (b)*)**.** Cases (b)–(c) are new. Case (b) covers bandwidths for which $h \to 0$ slower than $1/\sqrt{n}$. The convergence rate of $\hat{\beta}(z)$ has the unusual form $\sqrt{\frac{n}{h}}$, which is $o(n^{\frac{3}{4}})$ and is again less than the optimal rate $O(n^{\frac{3}{4}})$. Inspection of (2.2) suggests that undersmoothing to eliminate the bias term $h^2\mathcal{B}(z)$ could be achieved in Case (b) by setting the bandwidth h so that $nh^3 \to 0$, as then $\sqrt{\frac{n}{h}} \times h^2 = \sqrt{nh^3} \to 0$. When $nh \to \infty$ Lemma B.1(b)(i) shows that $\frac{1}{\sqrt{nh^3}} \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t \rightsquigarrow B_\eta(\cdot)$ holds, where $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$, $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz}$. This functional law plays a key role in the weak convergence of the standardized sum $\frac{1}{\sqrt{n^3h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t$ to the stochastic integral $\int B_x B_x' dB_\eta$ that appears in (2.2). The proof of Theorem 2.1 shows that when $nh^2 \to \infty$, the limit theory is wholly determined by the random element in the bias function rather than the usual variance term, as mentioned in earlier remarks following (1.11). Because of its reliance on the bias function, the limit distribution in (2.2) depends on the functional coefficient derivative $\beta^{(1)}(z)$ and the result, including the rate of convergence $\sqrt{\frac{n}{h}}$, in turn relies on the non-zero derivative condition $\beta^{(1)}(z) \neq 0$.

Remark 2.3 (*Case (c)*)**.** Case (c) yields the optimal convergence rate $O(n^{\frac{3}{4}})$ which holds when $nh^2 \to c$ for some constant $c \in (0, \infty)$ and $\beta^{(1)}(z) \neq 0$. The bandwidth that achieves this optimal rate is $h = O(\frac{1}{\sqrt{n}})$ and the bias term in (2.3) can be ignored without any undersmoothing because $n^{\frac{3}{4}} \times h^2 = n^{-\frac{1}{4}} \to 0$. More importantly, the asymptotics involve a composite form of two components, which are made explicit in the proof — see (A.19). Those two terms correspond to Cases (b) and (a), respectively, and are, in fact, boundary versions in which $h = O(\frac{1}{\sqrt{n}})$. This boundary at $h = O(\frac{1}{\sqrt{n}})$ delivers the optimal convergence rate $O(n^{\frac{3}{4}})$ for $\hat{\beta}(z)$. By mutual independence of $\{u_t\}$ and $\{z_t\}$, the two contributing components are uncorrelated, giving the mixed normal distribution of (2.3). The constant c adjusts the relative contributions to the asymptotic variance that come from the bias and the usual variance term.

Remark 2.4 (*Degeneracy*)**.** When the derivative $\beta^{(1)}(z) = 0$ it is clear from $\Omega_\beta(z)$ that the limit distribution in (2.2) is degenerate and the convergence rate rises. The simplest example occurs when $\beta(z) \equiv \beta$ is constant and the functional coefficient model is parametric. So $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz} = 0$ and $\eta_t = 0$ for all t, and there is no approximation error bias in the limit theory. The limit distribution of $\hat{\beta}(z)$ is then determined completely by the variance component and the result in (2.1) holds with $\mathcal{B}(z) = 0$. This degenerate case is discussed in Phillips and Wang (2020) where a test statistic is proposed to check the constancy of the functional coefficient. A general asymptotic treatment of locally flat functional coefficient regression is provided in Phillips and Wang (2021).

Remark 2.5 (*Implications for Robustness*)**.** In case (a) where $nh^2 \to 0$ the limit result may be interpreted as the nonstationary analogue in terms of both bias and variance of the stationary case, albeit up to rates of convergence and the limiting form of the regression sample moment matrix. But this match between the stationary and nonstationary cases holds only when $nh^2 \to 0$. Depending on the bandwidth employed in estimation, the true limit theory has three clearly different mixed normal limits, only one of which delivers rate efficient estimation and this occurs at the precise bandwidth rate $h = O(n^{-1/2})$ which is excluded in case (a). The three limit distributional forms reveal major differences between stationary and nonstationary FCC limit theory and seem to suggest that bandwidth specific formulations may be needed for inference. Notwithstanding these complications, construction of a robust self-normalized test statistic for inference about $\beta(z)$ is possible and applies to stationary and nonstationary cases, as shown in the original version of this paper (Phillips and Wang, 2020). This robust approach to inference will be analyzed in subsequent work.

Theorem 2.1 allows for bandwidths that satisfy $h \to 0$ slower than $1/n$, thereby ensuring that $nh \to \infty$. As mentioned in Remark 2.1, this is a stationary time series effective sample size requirement that enables the use of kernel limit theory for kernel weighted stationary time series. As in the following theorem, it is possible to relax this requirement due to the stronger signal of a nonstationary regressor and the resulting amplification of the regression signal weakens restrictions on the bandwidth. But when $nh \not\to \infty$ the conditions that assure central limit theory break down and no invariance principle (IP) applies even though FCC regression may still be consistent. While nonstationarity may allow for small bandwidths in the asymptotic development, practical issues in kernel smoothing do affect computability and finite sample behavior, almost always requiring use of a kernel $K(\cdot)$ with support $\mathcal{K} = \mathbb{R}$, as discussed earlier in connection with Assumption 1(iv).

Theorem 2.2. *Under Assumption 1, if $nh \to c$ for some $c \in [0, \infty)$, then $\hat{\beta}(z) \to_p \beta(z)$ and $\sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) = O_p(1)$ but no invariance principle applies.*

The condition $nh \to c \in [0, \infty)$ means that h tends to zero as fast or faster than $O(n^{-1})$, so that $nh^2 \to 0$ thereby matching the condition of case (a) of Theorem 2.1 but removing the effective sample size condition $nh \to \infty$ and allowing even smaller bandwidths. Theorem 2.1(a) allows for bandwidths in the region $O(n^{-1}) < h < O(n^{-1/2})$ whereas Theorem 2.2 allows for bandwidths $h \leq O(n^{-1})$. Such smaller bandwidth rates are only included subject to computability of $\hat{\beta}(z)$, which in turn relies on positivity of the finite sample weighted regression signal $\left(\sum_t x_t x_t' K_{tz}\right)$. More detailed comments on this matter and other aspects of Theorem 2.2 follow.

Remark 2.6 (*The Intermediate Case $nh \to c \in (0, \infty)$*)**.** From Theorem 2.2 when $nh \to c \in (0, \infty)$, the convergence rate of $\hat{\beta}(z)$ is $\sqrt{n}$. The rate of convergence of $\hat{\beta}(z)$ from Theorem 2.1(a) is $n\sqrt{h} = \sqrt{n}\sqrt{nh}$, which exceeds $\sqrt{n}$ since $nh \to \infty$ in Theorem 2.1(a). Thus, when the stationary process effective sample size nh diverges, $h \to 0$ slower than $\frac{1}{n}$ and the convergence rate of $\hat{\beta}(z)$ rises from $\sqrt{n}$ to $n\sqrt{h}$. The bandwidth then plays a role in determining the convergence rate. But when $h \to 0$ as fast or faster than $\frac{1}{n}$ the convergence rate of $\hat{\beta}(z)$ is $\sqrt{n}$ and unaffected by bandwidth.

Remark 2.7 ($nh \to 0$ *and* $n^3 h \to \infty$)**.** We may well wonder why there is no reduction in the convergence rate below $\sqrt{n}$ or even a failure of consistency if $h \to 0$ faster than $1/n$. In this case, it turns out that in the decomposition of $\hat{\beta}(z) - \beta(z)$ (see (A.20) in the Appendix or (1.8) in the scalar x_t case) the terms involving the approximation error $\beta(z_t) - \beta(z)$ are small enough to be neglected and dominated by $\left(\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^n x_t u_t K_{tz}$. Suppose, for instance, that $n^3 h \to \infty$, in which case

$$\sum_{t=1}^n x_t x_t' K_{tz} = \sum_{t=1}^n x_t x_t' \mathbb{E} K_{tz} + \sum_{t=1}^n x_t x_t' \zeta_{tK} = O_p(\sqrt{n^3 h}) \to \infty, \tag{2.6}$$

which means that persistent excitation still holds. The justification of (2.6) is as follows. Recall that $\mathbb{E} K_{tz} = hf(z) + o(h)$ and $\sum_{t=1}^n x_t x_t' = O_p(n^2)$ as $n \to \infty$, so that $\sum_{t=1}^n x_t x_t' \mathbb{E} K_{tz} = O_p(n^2 h)$. The term $\sum_{t=1}^n x_t x_t' \zeta_{tK}$ has zero mean and variance (using the scalar regressor case for convenience of exposition)

$$\begin{aligned} \mathbb{E}\left(\sum_{t=1}^n x_t^2 \zeta_{tK}\right)^2 &= \sum_{t=1}^n \mathbb{E} x_t^4 \mathbb{E} \zeta_{tK}^2 = 3 \sum_{t=1}^n t^2 \omega_x^4 \times \{h \nu_0(K) f(z) + o(h)\} \\ &= 3n^3 h \times \frac{1}{n} \sum_{t=1}^n \left(\frac{t}{n}\right)^2 \times \omega_x^4 \nu_0(K) f(z) = O(n^3 h) \end{aligned} \tag{2.7}$$

in the *iid* z_t case. Hence, $\sum_{t=1}^n x_t x_t' \zeta_{tK} = O_p(\sqrt{n^3 h})$. Consequently, $\sum_{t=1}^n x_t x_t' K_{tz} = O_p(n^2 h) + O_p\left(\sqrt{n^3 h}\right) = O_p\left(\sqrt{n^3 h}\right)$ when $nh \to 0$. We might then expect the $\sqrt{n}$ convergence rate (corresponding to the intermediate case $nh \to c \in (0, \infty)$) to be reduced in line with the diminished signal. However, calculation shows the variance matrix of the critical covariance term $\sum_{t=1}^n x_t u_t K_{tz}$ to be

$$\begin{aligned} \mathbb{E}\left(\sum_{t=1}^n x_t u_t K_{tz}\right)\left(\sum_{t=1}^n x_t u_t K_{tz}\right)' &= \sum_{t=1}^n \mathbb{E}\left(x_t x_t'\right) \mathbb{E}\left(u_t^2\right) \mathbb{E}\left(K_{tz}^2\right) \\ &= h \sum_{t=1}^n t \times \Omega_{xx} \sigma_u^2 \{f(z) \nu_0(K) + o(1)\} \\ &= n^2 h \left(\frac{1}{n^2} \sum_{t=1}^n t\right) \times \Omega_{xx} \sigma_u^2 \{f(z) \nu_0(K) + o(1)\}, \end{aligned} \tag{2.8}$$

where Ω_{xx} is the long run variance matrix of Δx_t. Since $\mathbb{E}\left(\sum_{t=1}^n x_t u_t K_{tz}\right) = 0$ and $\mathbb{V}ar\left(\sum_{t=1}^n x_t u_t K_{tz}\right) = O_p\left(n^2 h\right)$, it follows that $\sum_{t=1}^n x_t u_t K_{tz} = O_p\left(\sqrt{n^2 h}\right)$. In this case, we deduce that

$$\hat{\beta}(z) - \beta(z) = \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^3 h}} \sum_{t=1}^n x_t u_t K_{tz} + o_p\left(1/\sqrt{n}\right) = O_p\left(1/\sqrt{n}\right). \tag{2.9}$$

The estimator $\hat{\beta}(z)$ is then $\sqrt{n}$ consistent because the first member on the right side of (2.9) is the dominant $O_p\left(1/\sqrt{n}\right)$ term in the asymptotics and the $o_p\left(1/\sqrt{n}\right)$ term in (2.9) comes from the term involving the approximation error $\beta(z_t) - \beta(z)$. More detailed justification regarding the $o_p(1/\sqrt{n})$ term can be found in the proof of Theorem 2.2.

Remark 2.8 ($n^3h \to 0$)**.** Remark 2.7 establishes consistency when $n^3h \to \infty$. We may well have expected inconsistency if $n^3h \to 0$ or $h = o\left(1/n^3\right)$ because in that event the kernel weighted signal does not deliver persistent excitation. Indeed, in this event (2.7) continues to hold and $\sum_{t=1}^n x_t x_t' K_{tz} = O_p(\sqrt{n^3h})$ for $nh \to 0$ as before, yet now $\sum_{t=1}^n x_t x_t' K_{tz} = o_p(1)$ when $n^3h \to 0$ and the signal matrix fails the persistent excitation condition. Nonetheless, conditioning on $\mathcal{F}_{x,z} = \sigma\left\{x_t, z_t\right\}_1^\infty$ and using the scalar regressor case for convenience of exposition, we see that

$$\mathbb{V}ar\left(\left.\frac{\sum_{t=1}^n x_t u_t K_{tz}}{\sum_{t=1}^n x_t^2 K_{tz}}\right|_{\mathcal{F}_{x,z}}\right) = \frac{\sum_{t=1}^n x_t^2 K_{tz}^2 \sigma_u^2}{\left(\sum_{t=1}^n x_t^2 K_{tz}\right)^2} = \frac{n^2h\left(\frac{1}{n}\sum_{t=1}^n\left(\frac{x_t}{\sqrt{n}}\right)^2 \frac{\mathbb{E}K_{tz}^2}{h}\sigma_u^2\right)}{n^3h\left(\frac{1}{n}\sum_{t=1}^n\left(\frac{x_t}{\sqrt{n}}\right)^4 \frac{\mathbb{E}K_{tz}^2}{h}\right)} = O_p\left(\frac{1}{n}\right) \to 0, \tag{2.10}$$

which holds even when $n^3h \to 0$. In view of (2.10) consistency appears to hold irrespectively of whether the rate $h \to 0$ so fast that the persistent excitation condition fails. Of course, if $h \to 0$ too fast and the kernel support is compact then for finite n the signal is zero with positive probability, viz., $\mathbb{P}\left(\sum_{t=1}^n x_t^2 K_{tz} = 0\right) > 0$ and kernel estimation of the functional coefficient will fail. Even for Gaussian and other kernels with infinite support the signal $\sum_{t=1}^n x_t^2 K_{tz}$ may be so small as to prevent or inhibit calculation in such cases. Nonetheless, the result indicates that nonstationarity in the regressor continues to have a powerful influence on the asymptotic properties of functional coefficient regression estimator $\hat{\beta}(z)$ even when kernel weighted signal strength is no longer asymptotically infinite.

Remark 2.9 (*Stationary Case*)**.** For comparison, consider the stationary scalar x_t and *iid* z_t case where, when $nh \to c \in [0, \infty)$, we have

$$\frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t^2 K_{tz} = \sqrt{nh}\frac{1}{n}\sum_{t=1}^n x_t^2 \frac{\mathbb{E}K_{tz}}{h} + \frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t^2 \zeta_{tK} = O_p(1), \tag{2.11}$$

$$\frac{1}{nh^3}\sum_{t=1}^n x_t^2 \mathbb{E}\xi_{\beta t} = \frac{1}{n}\sum_{t=1}^n x_t^2 \frac{\mathbb{E}\xi_{\beta t}}{h^3} = O_p(1), \tag{2.12}$$

$$\frac{1}{\sqrt{nh^3}}\sum_{t=1}^n x_t^2 \eta_t = \frac{1}{\sqrt{n}}\sum_{t=1}^n x_t^2 \frac{\eta_t}{\sqrt{h^3}} = O_p(1), \tag{2.13}$$

$$\frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t u_t K_{tz} = \frac{1}{\sqrt{n}}\sum_{t=1}^n x_t u_t \frac{K_{tz}}{\sqrt{h}} = O_p(1). \tag{2.14}$$

Then

$$\begin{aligned}
\hat{\beta}(z) - \beta(z) &= \left(\sum_{t=1}^n x_t^2 K_{tz}\right)^{-1}\left(\sum_{t=1}^n x_t^2 \mathbb{E}\xi_{\beta t} + \sum_{t=1}^n x_t^2 \eta_t + \sum_{t=1}^n x_t u_t K_{tz}\right)\\
&= \left(\frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t^2 K_{tz}\right)^{-1}\left(\sqrt{nh^5}\frac{1}{nh^3}\sum_{t=1}^n x_t^2 \mathbb{E}\xi_{\beta t} + h\frac{1}{\sqrt{nh^3}}\sum_{t=1}^n x_t^2 \eta_t + \frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t u_t K_{tz}\right)\\
&= \left(\frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t^2 K_{tz}\right)^{-1}\left(o_p(1) + \frac{1}{\sqrt{nh}}\sum_{t=1}^n x_t u_t K_{tz}\right) = O_p(1).
\end{aligned} \tag{2.15}$$

The bias terms are evidently negligible in the above calculations because $nh^5 \to 0$ and $h \to 0$. In addition, conditional on $\mathcal{F}_{x,z} = \sigma\{x_t, z_t\}_1^\infty$ the conditional error variance is

$$\mathbb{V}ar\left(\hat{\beta}(z) - \beta(z)\Big|_{\mathcal{F}_{x,z}}\right) = \mathbb{V}ar\left(\left.\frac{\sum_{t=1}^n x_t u_t K_{tz}}{\sum_{t=1}^n x_t^2 K_{tz}}\right|_{\mathcal{F}_{x,z}}\right) = \frac{\sum_{t=1}^n x_t^2 K_{tz}^2 \sigma_u^2}{\left(\sum_{t=1}^n x_t^2 K_{tz}\right)^2} \neq 0, \tag{2.16}$$

and $\hat{\beta}(z)$ is evidently inconsistent in the stationary case. Unlike the nonstationary case, there is no asymptotic divergence between the stochastic order of the regressor x_t appearing in the sample covariance $\sum_{t=1}^n x_t u_t K_{tz}$ and that of the squared regressor x_t^2 that appears in the signal $\sum_{t=1}^n x_t^2 K_{tz}$. It is these differences in the stochastic order implications of the regressor that lead to major differences regarding consistency between the stationary and nonstationary cases under rapid bandwidth shrinkage when $nh \to c \in [0, \infty)$.

2.2. Multivariate z_t

In the general case where z_t is multivariate of dimension q, let $z = (z_1, \ldots, z_q)'$ and $z_t = (z_{1t}, \ldots, z_{qt})'$. We use the product kernel $K_{tzq} := K_q(z_t) := k_{tz_1} \times \cdots \times k_{tz_q}$ where $k_{tz_j} = k((z_{jt} - z_j)/h_j)$, $j = 1, \ldots, q$, $k(\cdot)$ is a symmetric second order kernel, and the $h_j, j = 1, \dots, q$, are individual bandwidths that are assumed to be the same up to a constant. The functional coefficient estimator now has the form $\hat{\beta}(z) = \left(\sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \sum_{t=1}^{n} x_t y_t K_{tzq}$. For notational simplicity, we use h to denote the common bandwidth. Let $\mu_j(k) = \int u^j k(u)du$ and $\nu_j(k) = \int u^j k^2(u)du$.

Theorem 2.3. *Under Assumption 1 and $\beta^{(1)}(z) \neq 0$, the following hold:*

(a) If $nh^q \to \infty$, we have

(1) when $q = 1$, see Theorem 2.1;
(2) when $q \geq 2$,

$$\sqrt{nh^{q-2}}\left(\hat{\beta}(z) - \beta(z) - h^2 \mu_2^q(k) D(z) f^{-1}(z)\right) \rightsquigarrow \left(f(z)\int B_x B_x'\right)^{-1}\left(\int B_x B_x' dB_{\eta q}\right) \tag{2.17}$$

$$\equiv_d \mathcal{MN}\left(0, \frac{\nu_2^q(k)}{f(z)}\left(\int B_x B_x'\right)^{-1}\int B_x B_x' \sum_{j=1}^{q}\left(B_x' \beta_j^{(1)}(z)\right)^2 \left(\int B_x B_x'\right)^{-1}\right), \tag{2.18}$$

where $D(z) = \sum_{j=1}^{q}\left[\beta_j^{(1)}(z) f_j^{(1)}(z) + \frac{1}{2}\beta_{jj}^{(2)}(z) f(z)\right]$, $\beta_j^{(1)}(z) = \partial\beta(z)/\partial z_j$ $\beta_{jj}^{(2)}(z) = \partial^2\beta(z)/\partial z_j^2$, and $B_{\eta q}(\cdot)$ is d-vector Brownian motion with variance matrix
$V_{\eta\eta q} = \nu_2^q(k) f(z) \sum_{j=1}^{q} \beta_j^{(1)}(z) \beta_j^{(1)}(z)'$.

(b) If $nh^q \to 0$, then

(1) when $q = 1$, see Theorem 2.2;
(2) when $q = 2$, we have

$$\sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) \sim_a \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq} = O_p(1); \tag{2.19}$$

(3) when $q > 2$, we have
(i) if $nh^2 \to 0$, then (2.19) continues to hold;
(ii) if $nh^2 \to c \in (0, \infty)$, then we have

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) &\sim_a \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \left(\frac{c^{1/2}}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} + \frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq}\right) \\ &= O_p(1);\end{aligned} \tag{2.20}$$

(iii) if $nh^2 \to \infty$, then

$$\frac{1}{h}\left(\hat{\beta}(z) - \beta(z)\right) \sim_a \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \frac{1}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} = O_p(1). \tag{2.21}$$

(c) If $nh^q \to c \in (0, \infty)$, then

(1) when $q = 1$, see Theorem 2.2;
(2) when $q = 2$, then (2.20) continues to hold;
(3) when $q > 2$, (2.21) continues to hold.

Theorem 2.3 is the multivariate extension of Theorems 2.1 and 2.2. From case (a), observe that when the condition $nh^q \to \infty$ holds and $q = 2$, the convergence rate is $\sqrt{n}$, irrespective of h. The bias can be ignored in this case when the undersmoothing condition $nh^{q+2} = nh^4 \to 0$ holds. When $q > 2$, the convergence rate is $\sqrt{nh^{q-2}}$ and declines as q increases, just as it does in the multidimensional z_t case for stationary time series (Li and Racine, 2007). Further, when $q \geq 2$, $\hat{\beta}(z)$ has the limit distribution in (2.18) with a sandwich form variance matrix that relies on the first derivatives $\{\beta_j^{(1)}(z) = \partial\beta(z)/\partial z_j\}_{j=1}^{q}$, analogous to case (b) of Theorem 2.1 where $q = 1$ and the convergence rate is $\sqrt{n/h}$. If these derivatives are zero at the point of estimation z, then $\hat{\beta}(z)$ has faster convergence rate than $\sqrt{nh^{q-2}}$ and its limit distribution depends on higher derivatives of the functional coefficient $\beta(z)$. This flat derivative case involves further complexities and is studied elsewhere.

Cases (b) and (c) of Theorem 2.3 show that $\hat{\beta}(z)$ is consistent even when $nh^q \to \infty$ fails. In this event, there is no invariance principle and the result matches Theorem 2.2 when $q = 1$. Notably, when $q > 2$ and $nh^q \to 0$ with $nh^2 \to c \in (0, \infty)$ in case (b)(3)(ii) or when $q = 2$ and $nh^q \to c \in (0, \infty)$ in case (c)(2), the limit behavior is described by (2.20), for which no invariance principle applies but where, like Theorem 2.1(c), both bias and variance terms contribute to large sample behavior.

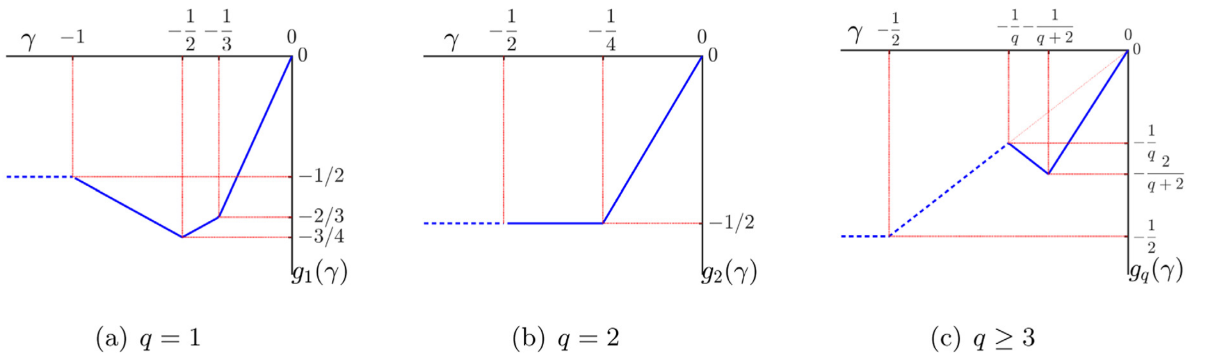

Fig. 1. Plots of the function $g_q(\gamma)$. The solid (blue) line depicts the region where an invariance principle holds in the limit theory and the dashed (blue) line depicts the region where no invariance principle applies.

Analogous to the condition $nh \to \infty$ in the case of $q = 1$, the condition $nh^q \to \infty$ is needed to establish functional laws for normalized partial sums of stationary elements involving the kernel weights that enter the asymptotics, such as $\zeta_{tK} = K_{tz} - \mathbb{E}K_{tz}$ and $\zeta_{tKq} = K_{tzq} - \mathbb{E}K_{tzq}$. Weak convergence of such quantities fails when $nh^q \not\to \infty$. The result is consistent estimation but without accompanying limit distribution theory.

2.3. Optimal bandwidth order and rate efficient estimation

This section explores the implications of the new limit theory on bandwidth selection and the convergence rate of the local level estimator $\hat{\beta}(z)$. Suppose $h = O(n^{\gamma})$ with $\gamma < 0$ and the estimation error $\hat{\beta}(z) - \beta(z) = O_p(n^{g_q(\gamma)})$, where $g_q(\gamma)$ is a function of γ and the subindex q indicates dependence on the dimension of z_t. The optimal bandwidth order, denoted γ_q^*, is the order for which $g_q(\gamma)$ achieves its minimum value and delivers the optimal convergence rate $\hat{\beta}(z) - \beta(z) = O_p(n^{g_q(\gamma_q^*)})$. To facilitate comparisons that are meaningful for inference it is convenient to require that the rate $n^{g_q(\gamma)}$ is such that an invariance principle (IP) holds when $\gamma = \gamma_q^*$.

We first look at the case where $q = 1$. According to Theorem 2.1(a), we have $-1 < \gamma < -1/2$ and $g_1(\gamma) = -(1+\gamma/2)$. When $\gamma = -1/2$, we have $nh^2 = O(1)$ and $g_1(\gamma) = -3/4$ based on Theorem 2.1(c). Theorem 2.1(b) deals with the case where $-1/2 < \gamma < 0$ and then $g_1(\gamma) = \max\{2\gamma, -\frac{1-\gamma}{2}\}$. It follows that $g_1(\gamma) = -\frac{1-\gamma}{2}$ when $-1/2 < \gamma < -1/3$ and $g_1(\gamma) = 2\gamma$ when $-1/3 \le \gamma < 0$. In view of Theorem 2.2, we have $g_1(\gamma) = -1/2$ for $\gamma \le -1$. Collectively, we obtain

$$g_1(\gamma) = \begin{cases} -1/2 & \gamma \le -1 & \text{No IP} \\ -(1+\gamma/2) & -1 < \gamma \le -1/2 & \text{IP} \\ -\frac{1-\gamma}{2} & -1/2 < \gamma < -1/3 & \text{IP} \\ 2\gamma & -1/3 \le \gamma < 0 & \text{IP} \end{cases}. \tag{2.22}$$

The function $g_1(\gamma)$ is plotted in Fig. 1(a), in which the dashed part of the function depicts regions where no IP holds, including the boundary point where the solid line commences. Evidently when $\gamma = -1/2$, the function $g_1(\gamma)$ achieves its minimum $-3/4$, the optimal bandwidth order is $O(n^{-1/2})$, $\hat{\beta}(z)$ achieves its fastest rate of convergence $n^{-3/4}$, and the mixed normal limit theory of Theorem 2.1(c) applies. In this case, the bias in (2.3) can be neglected because $n^{3/4} \times h^2 = n^{-1/4} \to 0$, and the optimal limit theory when $q = 1$ is given by

$$n^{3/4}\left[\hat{\beta}(z) - \beta(z)\right] \rightsquigarrow \mathcal{MN}\left(0,\ c^{\frac{1}{2}}\Omega_{\beta}(z) + \frac{1}{c^{\frac{1}{2}}}\Omega_u(z)\right),$$

which is attained with $h = O(n^{-1/2})$ and where the constant $c > 0$ is given by the limit $nh^2 \to c$.

When z_t is of dimension q, similar analyses can be conducted based on Theorem 2.3. When $q = 2$, we have

$$g_2(\gamma) = \begin{cases} -1/2 & \gamma \le -1/2 & \text{No IP} \\ -1/2 & -1/2 < \gamma < -1/4 & \text{IP} \\ 2\gamma & -1/4 \le \gamma < 0 & \text{IP} \end{cases}. \tag{2.23}$$

Fig. 1(b) plots the function $g_2(\gamma)$ for $q = 2$. The optimal choice of γ in this case is evidently $\gamma_2^* \in (-1/2, -1/4]$. Within this range for γ we have $\sqrt{n}$ consistency and asymptotic mixed normality, as given in (2.18). The bias term can again be ignored when $\gamma_2^* \in (-1/2, -1/4)$ because $\sqrt{nh^{q-2}} \times h^2 = n^{1/2+2\gamma} \to 0$ when $\gamma < -1/4$.

For higher dimensions with $q \geq 3$, following Theorem 2.3 we deduce that

$$g_q(\gamma) = \begin{cases} -1/2 & \gamma \leq -1/2 & \text{No IP} \\ \gamma & -1/2 < \gamma \leq -1/q & \text{No IP} \\ -\frac{1+\gamma(q-2)}{2} & -1/q < \gamma \leq -1/(q+2) & \text{IP} \\ 2\gamma & -1/(q+2) < \gamma < 0 & \text{IP} \end{cases}, \tag{2.24}$$

where the final two (IP) convergence rates come from Theorem 2.3(a)(2), the last involving the order of the bias term. The plot of $g_q(\gamma)$ for $q \geq 3$ is shown in Fig. 1(c). Under the premise that an invariance principle holds in the limit, the optimal bandwidth order that balances bias and variance is obtained with parameter setting $\gamma_q^* = -1/(q+2)$, for which the convergence rate is $n^{2/(q+2)}$. As is evident in Fig. 1(c), some smaller bandwidths with $\gamma \leq -1/q$ may lead to a faster rate of convergence in estimation than is achieved at $\gamma = -1/(q+2)$, but such rates sacrifice invariance principle asymptotics in the limit. For convenience in practical work, the optimal bandwidth order parameter setting $\gamma_q^* = -1/(q+2)$ is therefore suggested in this case. The corresponding optimal limit distribution theory is given by (2.18) and here the bias cannot be neglected because $\sqrt{nh^{q-2}} \times h^2 = n^{\frac{1}{2}+\frac{\gamma_q^*(q-2)}{2}+2\gamma_q^*} = n^0 = O(1)$.

3. Conclusion

Since the earliest work on spectral density estimation for stationary time series it has been traditional in nonparametric work to separate bias and variation in the analysis of nonparametric estimation and inference, emphasizing trade-offs between them that need to be balanced in applications. In contrast to such separation, the present paper shows how useful the random elements of the bias component that are normally ignored can be in sharpening accuracy in estimation. The analysis of nonstationary functional coefficient models reveals that these elements figure even in the limit theory variance and they are essential to rate efficient estimation. The next step in this research is to enhance inference via robust standard error estimation and test statistic construction in a way that utilizes the new limit theory, embodying all the random contributions to variance in a suitable normalization. The original version of this paper (Phillips and Wang, 2020) outlined a new approach to inference using a self-normalized test statistic that is robust with respect to bandwidth order and persistence in the regressor. The limit theory in the present paper should prove useful in developing this adaptive approach to inference and may prove useful in other areas of nonparametric estimation and inference.

The analysis given here confines attention to local level estimation and the functional coefficient cointegrating regression model (1.1) where x_t is a full rank integrated process. Corrections to the existing literature that are shown to apply in this prototypical model are also relevant in other functional coefficient models. Many extensions of the present development are possible. These include models with stationary and nonstationary regressors, near integrated or cointegrated regressors, endogeneity, and error processes more general than martingale differences. In all these cases similar influences to those demonstrated here arise from the presence of random variability in the bias term. In particular, models such as (1.1) where the regressors x_t have both $I(1)$ and $I(0)$ components (Cai et al., 2009) suffer the same difficulties as those presented here for the full rank $I(1)$ case; and models with multiple covariates z_t encounter similar complexities in the development of the correct limit theory to those analyzed in Theorem 2.3.

Primary among the effects that govern the correct limit theory are: (i) more complex trade-offs involving the bias and variance components in the limit theory; (ii) new optimal rates of convergence; (iii) multiple limit theory results that depend intimately on bandwidth choice; (iv) much greater complexity in models with functional coefficients involving high dimensional covariates; and (v) cases of consistent estimation where the usual effective sample size condition fails but no invariance principle limit theory holds. Similar considerations to those raised here apply to other nonparametric estimators such as local polynomial estimators. Extensions of the results to encompass these various complexities are left for future work.

Appendix A. Proofs of theorems

Proof of Theorem 2.1. We analyze the components in the following normalized decomposition of the estimation error

$$\begin{aligned}\left(\sum_{t=1}^n x_t x_t' K_{tz}\right)\left(\hat\beta(z) - \beta(z)\right) &= \sum_{t=1}^n x_t x_t'[\beta(z_t) - \beta(z)]K_{tz} + \sum_{t=1}^n x_t u_t K_{tz} \\ &= \sum_{t=1}^n x_t x_t' \mathbb{E}\xi_{\beta t} + \sum_{t=1}^n x_t x_t' \eta_t + \sum_{t=1}^n x_t u_t K_{tz},\end{aligned} \tag{A.1}$$

as in the scalar regressor case (1.3), with $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz}$ and $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$. Starting with the kernel weighted signal matrix, we have

$$\frac{1}{n^2 h}\sum_{t=1}^n x_t x_t' K_{tz} = \frac{1}{nh}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\mathbb{E}(K_{tz}) + \frac{1}{nh}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\zeta_{tK}, \tag{A.2}$$

where $\zeta_{tK} = K_{tz} - \mathbb{E}(K_{tz})$ and $\mathbb{E}K_{tz} = h\int K(r)f(z+rh)\,dr = hf(z) + O(h^3)$. Since $\mathbb{E}K_{tz}^2 = h\int K^2(r)f(z+rh)\,dr = hf(z)\int K^2(r)dr + o(h) = hf(z)\nu_0(K) + o(h)$, where $\nu_j(K) = \int u^j K^2(u)du$, it follows that $\mathbb{V}ar(\zeta_{tK}) = \mathbb{E}K_{tz}^2 - (\mathbb{E}K_{tz})^2 = O(h)$ and so $\zeta_{tK} = O_p(\sqrt{h})$. We deduce that when $nh \to \infty$

$$\frac{1}{n^2h}\sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{1}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z)+O(h^2)\right\} + \frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}}$$
$$= \frac{1}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z)+O(h^2)\right\} + O_p\left(\frac{1}{\sqrt{nh}}\right) \rightsquigarrow \left(\int B_x B_x'\right) f(z), \tag{A.3}$$

which follows because (i) $n^{-1/2}x_{\lfloor n\cdot\rfloor} \rightsquigarrow B_x(\cdot)$ by assumption, (ii) $(nh)^{-1/2}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK} \rightsquigarrow B_{\zeta K}(\cdot)$ from Lemma B.1(a) in Appendix B, and (iii) weak convergence to the matrix stochastic integral

$$\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}} \rightsquigarrow \int B_x B_x' dB_{\zeta K}, \tag{A.4}$$

holds, as shown in Lemma B.1(d).

When $nh \to c$ for some $c \in [0,\infty)$ we have in place of (A.3)

$$\frac{\sqrt{nh}}{n^2h}\sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z)+O(h^2)\right\} + \sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}}$$
$$= O_p(\sqrt{nh}) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{h}} = O_p(1), \tag{A.5}$$

and no invariance principle applies. The failure occurs because although $\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\zeta_{tK} = O_p(1)$ it does not satisfy a central limit theorem and, correspondingly, the functional law given in Lemma B.1(a)(i) fails, as explained in the proof of Lemma B.1(a)(ii). As a result of (A.5), the kernel weighted signal matrix $\sum_{t=1}^{n} x_t x_t' K_{tz} = O_p(\sqrt{n^3h})$ when $nh \to c \in [0,\infty)$. As discussed later in the proof of Theorem 2.2, it turns out that in this case where $nh \not\to \infty$ the estimator $\hat{\beta}(z)$ is still consistent but does not satisfy an invariance principle as $n \to \infty$. In what follows in the present proof, we proceed under the condition that $nh \to \infty$.

Next, from the proof of Lemma B.1(b), we have $\mathbb{E}\xi_{\beta t} = h^3\mu_2(K)C(z) + o(h^3)$ and so the first term in (A.1) is, upon normalization and use of standard weak convergence methods,

$$\frac{1}{nh^3}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\mathbb{E}\left(\xi_{\beta t}\right) \rightsquigarrow \mu_2(K)\left(\int B_x B_x'\right) C(z), \tag{A.6}$$

with $C(z) = \frac{1}{2}\beta^{(2)}(z)f(z) + \beta^{(1)}(z)f^{(1)}(z)$. The second term of (A.1) is, upon normalization and using Lemma B.1(d),

$$\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n} x_t x_t' \eta_t = \sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\eta_t}{\sqrt{nh^3}} \rightsquigarrow \int B_x B_x' dB_\eta,$$

where B_η is vector Brownian motion with variance matrix $\mathbb{V}ar(B_\eta) = \nu_2(K)f(z)[\beta^{(1)}(z)\beta^{(1)}(z)']$. The final term of (A.1) is, upon normalization and using Lemma B.1(a),

$$\frac{1}{n\sqrt{h}}\sum_{t=1}^{n} x_t u_t K_{tz} = \sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{u_t K_{tz}}{\sqrt{nh}} \rightsquigarrow \int B_x dB_{uK}, \tag{A.7}$$

where $B_{uK}(r)$ is the limit Brownian motion of $\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor} u_t K_{tz}$ with variance $\sigma_u^2 f(z)\nu_0(K)$. Standardizing by the weighted signal matrix and recentering (A.1) we have the estimation error decomposition

$$\hat{\beta}(z) - \beta(z) - \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1}\sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t}$$
$$= \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1}\sum_{t=1}^{n} x_t x_t' \eta_t + \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1}\sum_{t=1}^{n} x_t u_t K_{tz}, \tag{A.8}$$

which we write in standardized form as

$$\begin{aligned}
&\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^3} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \\
=&\sqrt{\frac{h}{n}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t + \frac{1}{n\sqrt{h}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz}. \qquad (A.9)
\end{aligned}$$

We now consider various cases depending on the bandwidth contraction rates in relation to the sample size.

Part (a)

In this case where $nh^2 \to 0$ the bandwidth $h = o(1/\sqrt{n})$. Upon rescaling (A.9) by $n\sqrt{h}$ and using results (A.3)–(A.7) we then have

$$\begin{aligned}
&n\sqrt{h} \left(\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^3} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \right) \\
&= \sqrt{nh^2} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t + \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz}, \\
&= o_p(1) + \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz} \qquad (A.10) \\
&\rightsquigarrow \left(f(z) \int B_x B_x' \right)^{-1} \left(\int B_x dB_{uK} \right) \equiv_d \mathcal{MN} \left(0, \frac{\nu_0(K)\sigma_u^2}{f(z)} \left(\int B_x B_x' \right)^{-1} \right), \qquad (A.11)
\end{aligned}$$

the mixed normality following from the independence of B_x and B_{uK}. Joint weak convergence of the numerator and denominator components of the matrix quotient in the second term of (A.10) follows from Lemma B.1(f). In view of (A.3) and (A.6)

$$\left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^3} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \rightsquigarrow \frac{\mu_2(K) C(z)}{f(z)}, \qquad (A.12)$$

giving the bias function and leading to the stated result (2.1) for case (a).

Part (b)

When $nh^2 \to \infty$ the bandwidth goes to zero slower than $O(1/\sqrt{n})$. We now rescale (A.9) by $\sqrt{n/h}$, giving

$$\begin{aligned}
&\sqrt{\frac{n}{h}} \left(\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^3} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \right) \\
&= \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t + \frac{1}{\sqrt{nh^2}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz} \qquad (A.13) \\
&= \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t + o_p(1) \\
&\rightsquigarrow \left(f(z) \int B_x B_x' \right)^{-1} \left(\int B_x B_x' dB_\eta \right) \qquad (A.14) \\
&\equiv_d \mathcal{MN} \left(0, \frac{\nu_2(K)}{f(z)} \left(\int B_x B_x' \right)^{-1} \int B_x B_x' \left(B_x' \beta^{(1)}(z) \right)^2 \left(\int B_x B_x' \right)^{-1} \right), \qquad (A.15)
\end{aligned}$$

using Lemma B.1(c) and (A.3) and where $B_\eta(\cdot)$ is Brownian motion with variance matrix $V_{\eta\eta} = \nu_2(K) f(z) \beta^{(1)}(z) \beta^{(1)}(z)'$. The weak convergence $\frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t \rightsquigarrow \left(\int B_x B_x' dB_\eta \right)$ in (A.14) depends on the functional law $\frac{1}{\sqrt{nh^3}} \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t \rightsquigarrow B_\eta(\cdot)$, as shown in the proof of Lemma B.1(b). Joint weak convergence of the respective components in (A.14) follows from Lemma B.1(f).

Further, since $B_\eta(r)$ is singular Brownian motion whenever $d > 1$ we may write the inner product $B_x(r)' B_\eta(r)$ in the equivalent form $B_x(r)' B_\eta(r) = \left(B_x(r)' \beta^{(1)}(z) \right) B_f(r)$, where B_f is scalar Brownian motion with variance $\nu_2(K) f(z)$. Then, in view of the independence of B_x and B_η, we have

$$\int B_x B_x' dB_\eta \equiv_d \mathcal{MN}\left(0,\, \nu_2(K) f(z) \int B_x B_x' \left(B_x' \beta^{(1)}(z)\right)^2\right), \tag{A.16}$$

which leads to the mixed normal limit distribution given in (A.15) and the stated result (b).

Part (c)

If $nh^2 \to c$ for some constant $c \in (0, \infty)$, then $h \sim_a \sqrt{c/n}$ and $\sqrt{n/h} = O(n^{\frac{3}{4}}) = n\sqrt{h}$. So the convergence rates in cases (a) and (b) of the Theorem are then the same $O(n^{\frac{3}{4}})$ rate. Correspondingly, the first and second terms on the right side of (A.9) appear to have the same order and both appear to contribute to the asymptotics. In this event, upon rescaling (A.9) by $n^{\frac{3}{4}}$ we find that

$$\begin{aligned}
&n^{\frac{3}{4}}\left(\hat\beta(z) - \beta(z) - h^2\left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{n^2h^3}\sum_{t=1}^n x_t x_t' \mathbb{E}\xi_{\beta t}\right)\\
&= (nh^2)^{\frac{1}{4}}\left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_t x_t' \eta_t + \frac{1}{(nh^2)^{\frac{1}{4}}}\left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_t u_t K_{tz},\\
&= c^{1/4}\left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_t x_t' \eta_t + \frac{1}{c^{\frac{1}{4}}}\left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_t u_t K_{tz}.
\end{aligned} \tag{A.17}$$

The asymptotics are jointly determined by the two terms of (A.17). Conditional on $\mathcal{F}_x$, these terms are uncorrelated as their conditional covariance matrix is

$$\mathbb{E}\left(\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_t x_t' \eta_t\right)\left(\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_t u_t K_{tz}\right)' = \frac{1}{n^2\sqrt{nh^2}}\sum_{t,s=1}^n \mathbb{E}\left(x_t x_s' (x_t' \eta_t u_s K_{sz})\right) = 0. \tag{A.18}$$

Using Lemma B.1(d)(ii) and (e), we find that since $nh \to \infty$ and $nh^2 \to c > 0$

$$\begin{aligned}
&n^{\frac{3}{4}}\left(\hat\beta(z) - \beta(z) - h^2\left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{n^2h^3}\sum_{t=1}^n x_t x_t' \mathbb{E}\xi_{\beta t}\right)\\
&= \left(\frac{1}{n^2h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\left[(nh^2)^{\frac{1}{4}}\left(\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_t x_t' \eta_t\right) + \frac{1}{(nh^2)^{\frac{1}{4}}}\left(\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_t u_t K_{tz}\right)\right]\\
&\rightsquigarrow \left(f(z)\int B_x B_x'\right)^{-1}\left(c^{\frac{1}{4}}\int B_x B_x' dB_\eta + \frac{1}{c^{\frac{1}{4}}}\int B_x dB_{uK}\right)\\
&\equiv_d \mathcal{MN}\left(0, c^{\frac{1}{2}}\Omega_\beta(z)\right) + \mathcal{MN}\left(0, \frac{1}{c^{\frac{1}{2}}}\Omega_u(z)\right) = \mathcal{MN}\left(0, c^{\frac{1}{2}}\Omega_\beta(z) + \frac{1}{c^{\frac{1}{2}}}\Omega_u(z)\right),
\end{aligned} \tag{A.19}$$

where $\Omega_\beta(z) = \frac{\nu_2(K)}{f(z)}\left(\int B_x B_x'\right)^{-1}\left(\int B_x B_x'\left(B_x'\beta^{(1)}(z)\right)^2\right)\left(\int B_x B_x'\right)^{-1}$, and $\Omega_u(z) = \nu_0(K)\sigma_u^2 f^{-1}(z)\left(\int B_x B_x'\right)^{-1}$. Joint weak convergence of the three matrix components in (A.19) holds in view of Lemma B.1(f). It follows that $\hat\beta(z)$ is $O(n^{\frac{3}{4}})$ convergent. ■

Proof of Theorem 2.2. Using the same notation as earlier, we analyze the decomposed estimation error

$$\begin{aligned}
\left(\hat\beta(z) - \beta(z)\right) &= \left(\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\sum_{t=1}^n x_t x_t' \mathbb{E}\xi_{\beta t} + \left(\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\sum_{t=1}^n x_t x_t' \eta_t\\
&\quad + \left(\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\sum_{t=1}^n x_t u_t K_{tz}.
\end{aligned} \tag{A.20}$$

The kernel weighted signal matrix under $\sqrt{n^3h}$ normalization has the following form

$$\begin{aligned}
\frac{1}{\sqrt{n^3h}}\sum_{t=1}^n x_t x_t' K_{tz} &= \frac{\sqrt{nh}}{n}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\mathbb{E}K_{tz}}{h} + \sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}}\\
&= \frac{\sqrt{nh}}{n}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z) + O(h^2)\right\} + \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{h}}.
\end{aligned} \tag{A.21}$$

When $nh \to c \in [0,\infty)$ the 'usual' effective sample size nh is asymptotically deficient. In this case, the first term of (A.21) satisfies $\frac{\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z)+O(h^2)\right\} \rightsquigarrow \sqrt{c}f(z)\int B_xB_x'$ and is therefore $O_p(1)$ if $c>0$ and $o_p(1)$ if $c=0$. To analyze the second term we proceed as follows. Since x_t is full rank $I(1)$ it is sufficient to consider the scalar case, which we write as $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}$. Since $\mathbb{E}K_{tz}=h\int K(r)f(z+rh)\,dr=hf(z)+O(h^3)$ and $\mathbb{E}K_{tz}^2=h\int K^2(r)f(z+rh)\,dr=hf(z)\int K^2(r)dr+O(h^3)=hf(z)\nu_0(K)+o(h)$ it follows that $\zeta_{tKh}:=\frac{\zeta_{tK}}{\sqrt{h}}=\frac{K_{tz}-\mathbb{E}(K_{tz})}{\sqrt{h}}$ is a zero mean triangular array with variance

$$\begin{aligned}\sigma_{\zeta h}^2 &= \mathbb{V}ar(\zeta_{tK}/\sqrt{h})=\left\{\mathbb{E}K_{tz}^2-(\mathbb{E}K_{tz})^2\right\}/h\\ &= \int K^2(r)f(z+rh)\,dr-h\left(\int K(r)f(z+rh)\,dr\right)^2=f(z)\nu_0(K)+O(h).\end{aligned}$$

By stationarity of z_t and Markov's inequality $\mathbb{P}(|\zeta_{tKh}|>M)\le\mathbb{E}\zeta_{tKh}^2/M^2=M^{-2}\{f(z)\nu_0(K)+O(h)\}$, so that for every $\epsilon>0$ there exists a constant M_ϵ such that $\sup_{h\to 0}\mathbb{P}(|\zeta_{tKh}|>M_\epsilon)<\epsilon$ and $\zeta_{tKh}=O_p(1)$ uniformly in t as $h\to 0$. It is shown in Lemma B.1(a)(ii) that while the normalized sum $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\zeta_{tK}}{\sqrt{h}}=O_p(1)$ it does not satisfy a central limit theorem because $nh\not\to\infty$ and the Lindeberg condition fails.

Next, by independence of x_t and z_t, we have $\mathbb{E}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\left(\frac{x_t}{\sqrt{n}}\right)^2\mathbb{E}\frac{\zeta_{tK}}{\sqrt{h}}=0$ and, when z_t is serially independent, so is ζ_{tK}. Thus,

$$\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}\right)^2=\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\frac{x_t}{\sqrt{n}}\right)^4\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\right)^2\to\nu_0(K)f(z)\int\mathbb{E}B_x^4>0\ \ a.s.,$$

as cross product terms are all zero for independent $\{z_t\}$. When z_t is serially dependent we have the additional cross product terms

$$\begin{aligned}&\frac{2}{n}\sum_{s>t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_s}{\sqrt{n}}\right)^2\right\}\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\frac{\zeta_{sK}}{\sqrt{h}}\right)=\frac{2}{n}\sum_{t=1}^{n}\sum_{j=1}^{n-t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_{t+j}}{\sqrt{n}}\right)^2\right\}\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\frac{\zeta_{t+jK}}{\sqrt{h}}\right)\\
&=\frac{2}{n}\sum_{t=1}^{n}\sum_{j=1}^{n-t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_{t+j}}{\sqrt{n}}\right)^2\right\}\frac{\gamma_\zeta(j)}{h}=\frac{2}{n}\sum_{t=1}^{n}\sum_{j=1}^{n-t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_{t+j}}{\sqrt{n}}\right)^2\right\}\frac{h^2f_{0,j}(z,z)+o\left(h^2\right)}{h}\\
&\sim_a\frac{2nh}{n^2}\sum_{t=1}^{n}\sum_{j=1}^{n-t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_{t+j}}{\sqrt{n}}\right)^2\right\}f_{0,j}(z,z)\\
&\le 2nh\times\sup_{j\ge 1}f_{0,j}(z,z)\times\mathbb{E}\left\{\int_0^1B_x(r)^2\int_r^1B_x^2(s)\,dsdr\right\}\\
&\to 2c\sup_{j\ge 1}f_{0,j}(z,z)\times\mathbb{E}\left\{\int_0^1B_x(r)^2\int_r^1B_x^2(s)\,dsdr\right\},\end{aligned}$$

where we use the fact that $\mathbb{E}K_{tz}K_{t+jz}=\iint K\left(\frac{s_0-z}{h}\right)K\left(\frac{s_j-z}{h}\right)f_{0,j}\left(s_0,s_j\right)ds_0ds_j=\iint K\left(p_0\right)K\left(p_j\right)\times f_{0,j}\left(z+p_0h,z+p_jh\right)dp_0dp_jh^2=h^2f_{0,j}(z,z)+o\left(h^2\right)$ and so $\gamma_\zeta(j)=\mathbb{E}K_{tz}K_{t+jz}-\mathbb{E}K_{tz}\mathbb{E}K_{t+jz}=h^2f_{0,j}(z,z)+o\left(h^2\right)$. From these calculations of the mean and variance, it follows that the second term of (A.21)

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}=O_p(1) \tag{A.22}$$

and then the kernel weighted signal $\sum_{t=1}^{n}x_t^2K_{tz}=O_p\left(\sqrt{n^3h}\right)$.

To prove consistency of $\hat{\beta}(z)$ we consider each term on the right side of (A.20) in turn.

(i) Using $\xi_{\beta t}=[\beta(z_t)-\beta(z)]K_{tz}$ we have, as shown in Lemma B.1(b)(i), $\mathbb{E}\xi_{\beta t}=h^3\mu_2(K)C(z)+o(h^3)$ and then

$$\begin{aligned}&\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'\mathbb{E}\xi_{\beta t}\\
&=\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{h^3\mu_2(K)C(z)+o(h^3)}{\sqrt{h}}\\
&=\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{h^{5/2}}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\{\mu_2(K)C(z)+o(1)\}\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{h^2\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{\mu_2(K)C(z)+o(1)\right\} \\
&= O_p\left(h^2\sqrt{nh}\right).
\end{aligned} \tag{A.23}
$$

(ii) Next, using $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$ we may show that $\frac{\eta_t}{\sqrt{h^3}} = O_p(1)$ uniformly in t as $h \to 0$ using the results of Lemma B.1(b) and by arguments similar to those used above in proving that $\frac{\zeta_{tK}}{\sqrt{h}} = O_p(1)$ uniformly in t as $h \to 0$. As in the proof of (A.22) and Lemma B.1(b)(ii) we find that $\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2 \frac{\eta_t}{\sqrt{nh^3}} = O_p(1)$, so that $\sum_{t=1}^{n} x_t x_t' \eta_t = O_p\left(\sqrt{n^3h^3}\right)$ and then

$$
\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' \eta_t = \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \times O_p\left(\frac{\sqrt{n^3h^3}}{\sqrt{n^3h}}\right) = O_p(h).
$$

It follows that $\left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t x_t' \eta_t = O_p(h)$.

(iii) We have $\mathbb{E}\sum_{t=1}^{n} x_t u_t K_{tz} = 0$ and $\mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n} x_t u_t K_{tz}\right) = \sigma_u^2 \sum_{t=1}^{n}\mathbb{E}\left(x_t^2\right)\mathbb{E}\left(K_{tz}^2\right) = O\left(n^2h\right)$, so that $\sum_{t=1}^{n} x_t u_t K_{tz} = O_p\left(n\sqrt{h}\right)$ and

$$
\begin{aligned}
&\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t u_t K_{tz} \\
&= \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \times O_p\left(\frac{n\sqrt{h}}{\sqrt{n^3h}}\right) = O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned} \tag{A.24}
$$

Note that in the present case where $nh \to c \in [0,\infty)$, the normalized sum $\sum_{t=1}^{n}\frac{u_t K_{tz}}{\sqrt{nh}}$ does not satisfy a central limit theorem because $nh \not\to \infty$, as explained in Lemma B.1(a)(ii), and correspondingly $\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{u_t K_{tz}}{\sqrt{nh}} = O_p(1)$, but does not converge weakly to a stochastic integral.

Combining (i), (ii) and (iii) with (A.20) and scaling the estimation error by $\sqrt{n}$ yields the following when $nh \to c \in [0,\infty)$

$$
\begin{aligned}
\sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) &= \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \\
&+ \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' \eta_t + \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t u_t K_{tz} \\
&= O_p\left(\sqrt{n}\times h^2\sqrt{nh}\right) + O_p\left(\sqrt{n}\times h\right) + \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t u_t K_{tz} \\
&= o_p(1) + \left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^2h}}\sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1),
\end{aligned} \tag{A.25}
$$

so that $\hat{\beta}(z)$ is $\sqrt{n}$ convergent but without an invariance principle. ■

Appendix B. A key lemma

Lemma B.1. *Under Assumption 1, the following hold as $n \to \infty$:*

(a) (i) If $nh \to \infty$, $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_t K_{tz}\} \rightsquigarrow \{B_{\zeta K}(\cdot), B_{uK}(\cdot)\}$, where $\{B_{\zeta K}, B_{uK}\}$ are independent Brownian motions with respective variances $\nu_0(K)f(z)$, and $\nu_0(K)\sigma_u^2 f(z)$, with $\zeta_{tK} = K_{tz} - \mathbb{E}K_{tz}$ and $K_{tz} = K(\frac{z_t - z}{h})$;
(ii) If $nh \to c \in [0,\infty)$, then $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_t K_{tz}\} = O_p(1)$ but no invariance principle holds.

(b) (i) If $nh \to \infty$ and $\beta^{(1)}(z) \neq 0$, $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t \rightsquigarrow B_\eta(\cdot)$, Brownian motion with variance matrix $V_{\eta\eta} = \nu_2(K)f(z)\beta^{(1)}(z)\beta^{(1)}(z)'$, where $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$, $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz}$;
(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t = O_p(1)$, but no invariance principle holds.

(c) (i) If $nh \to \infty$, $\frac{1}{n^2h}\sum_{t=1}^{n} x_t x_t' K_{tz} \rightsquigarrow \left(\int B_x B_x'\right) f(z)$;
(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n} x_t x_t' K_{tz} = O_p(1)$ but no invariance principle holds.

(d) (i) If $nh \to \infty$, $\frac{1}{\sqrt{n^3h}}\sum_{t=1}^n x_tx_t'\zeta_{tK} \rightsquigarrow \int B_xB_x'dB_{\zeta K}$, *and* $\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_tx_t'\eta_t \rightsquigarrow \int B_xB_x'dB_\eta \equiv_d \mathcal{MN}\left(0, \int B_x(r)B_x(r)'\left(B_x(r)'\beta^{(1)}(z)\right)^2\right)$;

(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{\sqrt{n^3h}}\sum_{t=1}^n x_tx_t'\zeta_{tK} = O_p(1)$ *but no invariance principle holds, and* $\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_tx_t'\eta_t = O_p(1)$ *but no invariance principle holds;*

(e) (i) If $nh \to \infty$, $\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_tu_tK_{tz} \rightsquigarrow \int B_xdB_{uK}$;

(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_tu_tK_{tz} = O_p(1)$ *but no invariance principle holds;*

(f) If $nh \to \infty$, $X_{u,n} = \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor} u_tK_{tz}$, $X_{\zeta,n} = \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}$, *and* $X_{\eta,n} = \frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t$, *then the following joint convergence holds*

$$\left\{X_{u,n}, X_{\zeta,n}, X_{\eta,n}, \frac{1}{n^2h}\sum_{t=1}^n x_tx_t'K_{tz}, \frac{1}{n\sqrt{h}}\sum_{t=1}^n x_tu_tK_{tz}, \frac{1}{\sqrt{n^3h}}\sum_{t=1}^n x_tx_t'\zeta_{tK}, \frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^n x_tx_t'\eta_t\right\}$$
$$\rightsquigarrow \{B_{uK}(\cdot), B_{\zeta K}(\cdot), B_\eta(\cdot), \left(\int B_xB_x'\right)f(z), \int B_xdB_{uK}, \int B_xB_x'dB_{\zeta K}, \int B_xB_x'dB_\eta\}.$$

Proof of Lemma B.1.

Part (a) (i) The joint limit result stated for $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{tz}\}$ is standard for partial sums involving kernel functions of strictly stationary weakly dependent time series (Xiao, 2009; Sun et al., 2011). Straightforward calculations in the present case show that $\mathbb{E}K_{tz} = hf(z) + o(h)$, and $\mathbb{E}K_{tz}^2 = hf(z)\nu_0(K) + o(h)$, so that $\mathbb{V}ar(\zeta_{tK}) = hf(z)\nu_0(K) + o(h)$ and $\zeta_{tK} = K_{tz} - \mathbb{E}(K_{tz}) = O_p(\sqrt{h})$. Further, $\mathbb{V}ar(u_tK_{tz}) = h\nu_0(K)\sigma_u^2f(z) + o(h)$ and $\mathbb{E}(u_tK_{tz}K_{sz}) = 0$ for all t and s. So the standardized partial sums processes $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{tz}\}$ are uncorrelated, uniformly tight, and the stated joint functional law follows by standard weak convergence methods for triangular arrays (e.g., Davidson (1994, Theorem 27.17 for martingale difference arrays, and chapter 29.3 for dependent arrays)). The resulting limit processes $\left(B_{\zeta K}(r), B_{uK}(r)\right)$ are independent with respective variances $\nu_0(K)f(z)$ and $\nu_0(K)\sigma_u^2f(z)$. The effective sample size condition $nh \to \infty$ is required for this result.

Part (a) (ii) If $nh \to c \in [0,\infty)$ then the effective sample size condition $nh \to \infty$ fails. In this case, $\left(\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{tz}\right) = O_p(1)$ but no invariance principle applies because of failure in the Lindeberg condition. To demonstrate, it is sufficient to consider the case of $\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}$ and *iid* $\{z_t\}$. In this case the stability condition $\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\zeta_{tK}}{\sqrt{h}}\right)^2 = \frac{1}{n}\sum_{t=1}^n\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\right)^2 = f(z)\nu_0(K) + O(h)$ is satisfied but the Lindeberg condition fails. To see this, note that $\zeta_{tK} = K(\frac{z_t-z}{h}) - \mathbb{E}K(\frac{z_t-z}{h}) = K(\frac{z_t-z}{h}) + O(h)$. Given $\epsilon > 0$, we have

$$\frac{1}{n}\sum_{t=1}^n\mathbb{E}\left\{\left(\frac{\zeta_{tK}}{\sqrt{h}}\right)^2 1_{[|\zeta_{tK}|>\epsilon\sqrt{nh}]}\right\} = \int\frac{[K(\frac{z_t-z}{h}) + O(h)]^2}{h}1_{\left[\left|K(\frac{z_t-z}{h})+O(h)\right|>\epsilon\sqrt{nh}\right]}f(z_t)dz_t$$
$$= \int (K(p) + O(h))^2\, 1_{[|K(p)+O(h)|>\epsilon\sqrt{nh}]}f(z+ph)dp$$
$$\to \begin{cases} f(z)\nu_0(K) > 0 & \text{if } nh \to 0 \\ \int K^2(p)1_{[|K(p)|>\epsilon\sqrt{c}]}dpf(z) > 0 & \text{if } nh \to c \in (0,\infty) \end{cases}.$$

A similar proof applies in the case of $\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{tz}$.

Part (b) (i) We compute the first and second moments of $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$ and show that $\eta_t = O_p(h^{3/2})$. First

$$\mathbb{E}\xi_{\beta t} = \mathbb{E}[\beta(z_t) - \beta(z)]K_{tz} = \int_{-1}^1[\beta(s) - \beta(z)]K((s-z)/h)f(s)ds$$
$$= h\int_{-1}^1[\beta(z+hp) - \beta(z)]K(p)f(z+hp)dp$$
$$= h^3[\frac{1}{2}\beta^{(2)}(z)f(z) + \beta^{(1)}(z)f^{(1)}(z)]\int_{-1}^1 p^2K(p)dp + o(h^3)$$
$$= h^3C(z)\mu_2(K) + o(h^3), \tag{B.1}$$

with $C(z) = \frac{1}{2}\beta^{(2)}(z)f(z) + \beta^{(1)}(z)f^{(1)}(z)$. Next

$$\mathbb{E}\xi_{\beta t}\xi_{\beta t}' = \mathbb{E}[(\beta(z_t) - \beta(z))(\beta(z_t) - \beta(z))'K^2(\frac{z_t-z}{h})]$$
$$= h\int_{-1}^1(\beta(z+hs) - \beta(z))(\beta(z+hs) - \beta(z))'K^2(s)f(z+hs)ds$$

$$= h^3[\beta^{(1)}(z)\beta^{(1)}(z)']f(z)\int_{-1}^{1} s^2K^2(s)ds + o(h^3)$$
$$= h^3[\beta^{(1)}(z)\beta^{(1)}(z)']f(z)\nu_2(K) + o(h^3). \tag{B.2}$$

It follows that

$$\mathbb{V}ar(\eta_t) = \mathbb{E}\xi_{\beta t}\xi_{\beta t}' - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})' = h^3\nu_2(K)f(z)[\beta^{(1)}(z)\beta^{(1)}(z)'] + o(h^3), \tag{B.3}$$

and $\eta_t = O_p(h^{3/2})$. Next, in view of (B.1) the serial covariances satisfy

$$\mathbb{C}ov(\xi_{\beta t}, \xi_{\beta t+j}) = \mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - \left(\mathbb{E}\xi_{\beta t}\right)\left(\mathbb{E}\xi_{\beta t+j}\right)' = \mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' + O(h^6),$$

and by virtue of the strong mixing of z_t, measurability of $\beta(\cdot)$, and Davydov's lemma the covariances satisfy the bound

$$|\mathbb{C}ov(\xi_{\beta t}, \xi_{\beta t+j})| \le 12\left(\mathbb{E}\left|\xi_{\beta t}\right|^{\delta}\right)^{2/\delta}|\alpha(j)|^{1-2/\delta} = A_\beta h^{2+2/\delta}|\alpha(j)|^{1-2/\delta} + o(h^{2+2/\delta}), \tag{B.4}$$

where $A_\beta = 12(\int|\beta^{(1)}(\tilde{z}_p)|^{\delta}|p|^{\delta}K(p)^{\delta}dpf(z))^{2/\delta}$, since $\mathbb{E}\left|\xi_{\beta t}\right|^{\delta} = h^{1+\delta}\int|\beta^{(1)}(\tilde{z}_p)|^{\delta}|p|^{\delta}K(p)^{\delta}dpf(z) + o(h^{1+\delta})$ in a similar way to (B.1), and where $\tilde{z}_p$ is on the line segment connecting z and $z+hp$. Further, for $j \neq 0$ and using the joint density $f_{0,j}(s_0, s_j)$ of $\left(z_t, z_{t+j}\right)$ we have

$$\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' = \mathbb{E}[(\beta(z_t)-\beta(z))\left(\beta(z_{t+j})-\beta(z)\right)'K_{tz}K_{t+j,z}]$$
$$= \iint(\beta(s_0)-\beta(z))\left(\beta(s_j)-\beta(z)\right)'K\left(\frac{s_0-z}{h}\right)K\left(\frac{s_j-z}{h}\right)f_{0,j}(s_0,s_j)ds_0ds_j$$
$$= h^2\iint(\beta(z+hp_0)-\beta(z))(\beta(z+hp_j)-\beta(z))'K(p_0)K(p_j)f_{0,j}(z+hp_0,z+hp_j)dp_0dp_j$$
$$= h^6\left\{\frac{1}{4}[\beta^{(2)}(z)][\beta^{(2)}(z)]'f_{0,j}(z,z) + [\beta^{(1)}(z)][\beta^{(1)}(z)]'\frac{\partial^2 f_{0,j}}{\partial s_0\partial s_j}(z,z) + \frac{1}{2}[\beta^{(1)}(z)][\beta^{(2)}(z)]'\frac{\partial f_{0,j}}{\partial s_0}(z,z)\right.$$
$$\left.+\frac{1}{2}[\beta^{(2)}(z)][\beta^{(1)}(z)]'\frac{\partial f_{0,j}}{\partial s_j}(z,z)\right\}[\mu_2(K)]^2 + o(h^6). \tag{B.5}$$

We now deduce that the long run variance matrix of η_t is

$$\mathbb{V}^{LR}(\eta_t) = \mathbb{E}\left[\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{n}\eta_t\right]\left[\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{n}\eta_t\right]' = \frac{1}{nh^3}\sum_{t=1}^{n}\mathbb{E}\eta_t\eta_t' + \frac{1}{nh^3}\sum_{t\neq s}\mathbb{E}\eta_t\eta_s'$$
$$= \frac{1}{h^3}\mathbb{E}\eta_t\eta_t' + o(1) \to \nu_2(K)f(z)[\beta^{(1)}(z)\beta^{(1)}(z)'] =: V_{\eta\eta}, \tag{B.6}$$

which follows from (B.3) and standard arguments concerning the $o(1)$ magnitude of the sum of the autocovariances of kernel weighted stationary processes. In particular, from the α mixing property of z_t and using a sum splitting argument and results (B.1), (B.4) and (B.5) above, we have

$$\frac{1}{nh^3}\sum_{t\neq s}\mathbb{E}\eta_t\eta_s' = \frac{1}{h^3}\sum_{j=-n+1,j\neq 0}^{n-1}\left[1-\frac{|j|}{n}\right][\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - \left(\mathbb{E}\xi_{\beta t}\right)\left(\mathbb{E}\xi_{\beta t+j}\right)']$$
$$= \frac{1}{h^3}\sum_{j=-M,j\neq 0}^{M}\left[1-\frac{|j|}{n}\right][\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - \left(\mathbb{E}\xi_{\beta t}\right)\left(\mathbb{E}\xi_{\beta t}\right)'] + \frac{1}{h^3}\sum_{M<|j|<n}\left(1-\frac{|j|}{n}\right)[\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - \left(\mathbb{E}\xi_{\beta t}\right)\left(\mathbb{E}\xi_{\beta t}\right)']$$
$$= O\left(\frac{Mh^6}{h^3}\right) + O\left(\frac{1}{h^3}\left(\mathbb{E}\left|\xi_{\beta t}\right|^{\delta}\right)^{2/\delta}\sum_{M<|j|<n}\alpha_j^{1-2/\delta}\right) = O\left(Mh^3\right) + O\left(\frac{h^{2\frac{1+\delta}{\delta}}}{h^3M^a}\sum_{M<|j|<\infty}j^a\alpha_j^{1-2/\delta}\right)$$
$$= O\left(Mh^3\right) + O\left(\frac{h^{\frac{2}{\delta}}}{hM^a}\sum_{M<|j|<\infty}j^a\alpha_j^{1-2/\delta}\right) = O\left(Mh^3\right) + O\left(\frac{1}{h^{1-2/\delta}M^a}\sum_{M<|j|<\infty}j^a\alpha_j^{1-2/\delta}\right)$$
$$= O\left(Mh^3\right) + o\left(\frac{1}{(Mh)^{1-2/\delta}}\right) = o\left(1\right),$$

for a suitable choice of $M \to \infty$ such that $Mh \to \infty$ $Mh^3 \to 0$ and $\frac{M}{n} \to 0$ and with $a > 1-2/\delta$ and $\delta > 2$. It then follows by arguments similar to the central limit theory for weakly dependent kernel regression in Robinson (1983), Masry and Fan (1997), and Fan and Yao (2003, theorem 6.5) that the standardized partial sum process of η_t

satisfies a triangular array functional law giving $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t \rightsquigarrow B_\eta(\cdot)$, where B_η is vector Brownian motion with variance matrix $V_{\eta\eta}=\nu_2(K)f(z)\beta^{(1)}(z)\beta^{(1)}(z)'$. The effective sample size condition $nh\to\infty$ is required for this result.

Part (b) (ii) When $nh\to c\in[0,\infty)$ we prove that

$$\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t = O_p(1), \tag{B.7}$$

but with no invariance principle applying. This result mirrors the finding in Part (a)(ii) for $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK},\ \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{tz}\}$. In the present case and without loss of generality, let x_t be scalar and $\{z_t\}$ be *iid*, so that $\eta_t=\xi_{\beta t}-\mathbb{E}\xi_{\beta t}=\xi_{\beta t}+O(h^3)$, since $\mathbb{E}\xi_{\beta t}=h^3C(z)\mu_2(K)+o(h^3)$ from (B.1). The martingale stability condition

$$\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\eta_t}{\sqrt{h^3}}\right)^2=\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\frac{\eta_t}{\sqrt{h^3}}\right)^2=\nu_2(K)f(z)\left(\beta^{(1)}(z)\right)^2+O(h),$$

is satisfied so that $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{n}\eta_t=O_p(1)$, giving (B.7). But the Lindeberg condition fails and no invariance principle holds. The proof is similar to that of Part (a)(ii) but has additional complications due to the form of the sequence η_t. First note that $\eta_t=[\beta(z_t)-\beta(z)]K_{tz}+O(h^3)$. Then, given $\epsilon>0$, $nh\not\to\infty$ and $\beta^{(1)}(z)\neq 0$, we find that

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\{\left(\frac{\eta_t}{\sqrt{h^3}}\right)^2 1_{\left[|\eta_t|>\epsilon\sqrt{nh^3}\right]}\right\}\\
&=\int\frac{[[\beta(z_t)-\beta(z)]K_{tz}+O(h^3)]^2}{h^3}1_{\left[|[\beta(z_t)-\beta(z)]K_{tz}+O(h^3)|>\epsilon\sqrt{nh^3}\right]}f(z_t)dz_t\\
&=\int\frac{[\beta^{(1)}(z)hpK(p)+O(h^2)]^2}{h^2}1_{\left[|\beta^{(1)}(z)hpK(p)+O(h^2)|>\epsilon\sqrt{nh^3}\right]}f(z+ph)dp\\
&=\left(\beta^{(1)}(z)\right)^2f(z)\int p^2K^2(p)1_{\left[|\beta^{(1)}(z)pK(p)|>\epsilon\sqrt{nh}\right]}dp+O(h)\\
&\to\begin{cases} f(z)\left(\beta^{(1)}(z)\right)^2\nu_2(K)>0 & \text{if } nh\to 0\\ f(z)\left(\beta^{(1)}(z)\right)^2\int p^2K^2(p)1_{\left[|pK(p)\beta^{(1)}(z)|>\epsilon\sqrt{c}\right]}dp>0 & \text{if } nh\to c\in(0,\infty)\end{cases},
\end{aligned}$$

and the Lindeberg condition fails in both cases since $\beta^{(1)}(z)\neq 0$.

Part (c) (i) This result (i) is established using standard methods in (A.3) in the proof of Theorem 2.1.

Part (c) (ii) As in (A.5) in the proof of Theorem 2.1, when $nh\to c\in[0,\infty)$ we have the following decomposition

$$\begin{aligned}
\frac{\sqrt{nh}}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}&=\frac{\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z)+O(h^2)\right\}+\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}}\\
&\sim_a cf(z)\int B_xB_x'+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{h}}+o_p(1)=O_p(1).
\end{aligned}\tag{B.8}$$

The second term of (B.8) is $O_p(1)$ but with no invariance principle. To see this, we proceed in a similar fashion to Part (b) (ii). For convenience and without loss of generality, let x_t be scalar and z_t be *iid*. We then have $\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}\right)=0$ and

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}\right)^2&=\mathbb{E}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^4\right)\times\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\right)^2\\
&=\mathbb{E}\left(\int B_x^4\right)\times\{f(z)\nu_0(K)+O(h)\}=O(1),
\end{aligned}$$

so that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}}=O_p(1)$, as required. No invariance principle holds in this case because $\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}=O_p(1)$ without an invariance principle when $nh\to c\in[0,\infty)$ by virtue of Part (a)(ii).

Part (d) (i) By Assumption 1, Lemma B.1(a) and (b) and when $nh\to\infty$ we have the joint convergence

$$\left(\frac{1}{\sqrt{n}}x_{\lfloor n\cdot\rfloor},\quad \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK},\quad \frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t\right)\rightsquigarrow\left(B_x(\cdot),\quad B_{\zeta K}(\cdot),\quad B_\eta(\cdot)\right), \tag{B.9}$$

where the Brownian motions $\{B_x,B_{\zeta K},B_\eta\}$ are independent by virtue of (i) the exogeneity of x_t and (ii) the independence of $\{B_{\zeta K},B_\eta\}$. The latter follows from the fact that the contemporaneous covariance $\mathbb{E}\zeta_{tK}\eta_t=h^3\nu_2(K)[\frac{1}{2}\beta^{(2)}(z)f(z)+\beta^{(1)}(z)f^{(1)}(z)]+O(h^4)=O(h^3)$ and the cross serial covariance $\mathbb{E}\zeta_{tK}\eta_{t+j}=O(h^4)$ for $j\neq 0$, so that combined with the weak dependence of z_t and an argument along the same lines as that leading to (B.6) we have $\mathbb{E}\left(\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}\times\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t\right)$ $=\frac{1}{h^2}\mathbb{E}\left(\zeta_{tK}\eta_t\right)+o(1)=o(1)$. Convergence to the stochastic integral limits,

$$\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'\zeta_{tK}=\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\right)\frac{\zeta_{tK}}{\sqrt{nh}}\rightsquigarrow\int B_xB_x'dB_{\zeta K},\tag{B.10}$$

$$\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n}x_tx_t'\eta_t=\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\right)\frac{\eta_t}{\sqrt{nh^3}}\rightsquigarrow\int B_xB_x'dB_{\eta}\tag{B.11}$$

then follows by a triangular array extension of Ibragimov and Phillips (2008, theorem 4.3) when $nh\to\infty$. Both stochastic integrals have mixed normal distributions, viz.,

$$\int B_x\otimes B_xdB_{\zeta K}\equiv_d\mathcal{MN}\left(0,\,\nu_0(K)f(z)\int B_xB_x'\otimes B_xB_x'\right),\tag{B.12}$$

$$\int B_xB_x'dB_{\eta}\equiv_d\mathcal{MN}\left(0,\,\nu_2(K)f(z)\int B_xB_x'\left(B_x(r)'\beta^{(1)}(z)\right)^2\right),\tag{B.13}$$

and the stated result (i) of Part (d) holds.

Part (d) (ii) When the rate condition $nh\to\infty$ fails and, instead $nh\to c\in[0,\infty)$ applies, it follows from Part (a)(ii) and Part (b)(ii) that $\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\zeta_{tK}=O_p(1)$ and $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{n}\eta_t=O_p(1)$, respectively, but with no invariance principles holding. Correspondingly, in place of (B.10) and (B.11), we have in the same manner as before in the proof of Part (c)(ii)

$$\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'\zeta_{tK}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\right)\frac{\zeta_{tK}}{\sqrt{h}}=O_p(1),\tag{B.14}$$

$$\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n}x_tx_t'\eta_t=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\right)\frac{\eta_t}{\sqrt{h^3}}=O_p(1),\tag{B.15}$$

again without invariance principles.

Part (e) (i) Write $\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_tu_tK_{tz}=\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)\left(\frac{u_tK_{tz}}{\sqrt{nh}}\right)\rightsquigarrow\int B_xdB_{uK}$, and the result follows by standard limit theory directly from Part (a), the mutual independence of x_t, u_t and z_t, and an array extension of Ibragimov and Phillips (2008, theorem 4.3).

Part (e) (ii) If $nh\to c\in[0,\infty)$, it follows from Part (a) (ii) that $\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}u_tK_{tz}=O_p(1)$ but no invariance principle holds. In a similar fashion and as in Parts (c)(ii) and (d)(ii), we deduce that $\frac{1}{\sqrt{n^2h}}\sum_{t=1}^{n}x_tu_tK_{tz}=\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\frac{u_tK_{tz}}{\sqrt{nh}}\right)=O_p(1)$ with no invariance principle holding.

Part (f) By Assumption 1 and Lemma B.1(a), (b), (d) when $n\to\infty$ and $nh\to\infty$ we have the joint weak convergence

$$\left(\frac{1}{\sqrt{n}}x_{\lfloor n\cdot\rfloor},\quad\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{zt},\quad\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK},\quad\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t\right)\rightsquigarrow\left(B_x(\cdot),\quad B_{uK}(\cdot),\quad B_{\zeta K}(\cdot),\quad B_{\eta}(\cdot)\right),$$

where the Brownian motions $\{B_x,B_{uK},B_{\zeta K},B_{\eta}\}$ are independent by virtue of the exogeneity of x_t and z_t and the independence of $\{B_x,B_{\zeta K},B_{\eta}\}$. It then follows by a triangular array extension of joint weak convergence to stochastic integrals for α-mixing time series (Liang et al., 2016, theorem 3.1) that

$$\left\{\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz},\,\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_tu_tK_{tz},\,\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'\zeta_{tK},\,\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n}x_tx_t'\eta_t\right\}$$
$$\rightsquigarrow\left\{\int B_xB_x'f(z),\,\int B_xdB_{uK},\,\int B_xB_x'dB_{\zeta K},\,\int B_xB_x'dB_{\eta}\right\}.$$

The conditions of Liang et al. (2016, theorem 3.1) require sixth moments of the component innovations and α mixing numbers that decay according to a power law $\alpha(j)=\frac{1}{j^{\gamma}}$ with $\gamma>6$. This condition is satisfied by the mixing conditions of Assumption 1 when $\delta=3>2$ and $c=\frac{1}{2}>1-\frac{2}{\delta}=\frac{1}{3}$ and $\alpha(j)=\frac{1}{j^{\gamma}}$ with $\gamma=6(1+\epsilon)>6$ for some $\epsilon>0$. For in that case, the summability condition $\sum_{j\geq1}j^{c}[\alpha(j)]^{1-2/\delta}=\sum_{j\geq1}\frac{1}{j^{\frac{\gamma}{3}-\frac{1}{2}}}=\sum_{j\geq1}\frac{1}{j^{\frac{3}{2}+2\epsilon}}<\infty$ holds and the innovations have finite moments of order $p>2\delta=6$. ■

Appendix C. Online supplement

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2021.09.007. This material includes additional proofs and simulations.

References

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. J. Amer. Statist. Assoc. 95 (451), 941–956.
Cai, Z., Li, Q., Park, J.Y., 2009. Functional-coefficient models for nonstationary time series data. J. Econometrics 148 (2), 101–113.
Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. OUP Oxford.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer Science & Business Media.
Ibragimov, R., Phillips, P.C.B., 2008. Regression asymptotics using martingale convergence methods. Econometric Theory 24 (4), 888–947.
Juhl, T., 2005. Functional-coefficient models under unit root behaviour. Econom. J. 8 (2), 197–213.
Li, Q., Racine, J.S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press.
Liang, H., Phillips, P.C.B., Wang, H., Wang, Q., 2016. Weak convergence to stochastic integrals for econometric applications. Econometric Theory 32 (6), 1349–1375.
Masry, E., Fan, J., 1997. Local polynomial estimation of regression functions for mixing processes. Scand. J. Stat. 24 (2), 165–179.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Ann. Statist. 971–1001.
Phillips, P.C.B., Wang, Y., 2020. When bias contributes to variance: true limit theory in functional coefficient cointegrating regression. In: Cowles Foundation Discussion Paper No. 2250, Yale University.
Phillips, P.C.B., Wang, Y., 2021. Limit theory in locally flat functional coefficient regression. Working Paper, Yale University.
Robinson, P.M., 1983. Nonparametric estimators for time series. J. Time Series Anal. 4 (3), 185–207.
Sun, Y., Cai, Z., Li, Q., 2016. A consistent nonparametric test on semiparametric smooth coefficient models with integrated time series. Econometric Theory 32 (4), 988–1022.
Sun, Y., Hsiao, C., Li, Q., 2011. Measuring correlations of integrated but not cointegrated variables: A semiparametric approach. J. Econometrics 164 (2), 252–267.
Tu, Y., Wang, Y., 2019. Functional coefficient cointegration models subject to time–varying volatility with an application to the purchasing power parity. Oxf. Bull. Econ. Stat. 81 (6), 1401–1423.
Xiao, Z., 2009. Functional-coefficient cointegration models. J. Econometrics 152 (2), 81–92.