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

By

Peter C. B. Phillips and Ying Wang

July 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2250

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

http://cowles.yale.edu/

When Bias Contributes to Variance: True Limit Theory in Functional Coefficient Cointegrating Regression*

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

[a]University of Auckland, [b]Yale University,
[c]Singapore Management University, [d]University of Southampton

July 16, 2020

Abstract

Limit distribution theory in the econometric literature for functional coefficient cointegrating (FCC) regression is shown to be incorrect in important ways, influencing rates of convergence, distributional properties, and practical work. In FCC regression the cointegrating coefficient vector $\beta(\cdot)$ is a function of a covariate z_t. The true limit distribution of the local level kernel estimator of $\beta(\cdot)$ is shown to have multiple forms, each form depending on the bandwidth rate in relation to the sample size and with an optimal convergence rate of $n^{3/4}$ which is achieved by letting the bandwidth have order $1/\sqrt{n}$ when z_t is scalar. Unlike stationary regression and contrary to the existing literature on FCC regression, the correct limit theory reveals that component elements from the bias and variance terms in the kernel regression can both contribute to variability in the asymptotics depending on the bandwidth behavior in relation to the sample size. The trade-off between bias and variance that is a common feature of kernel regression consequently takes a different and more complex form in FCC regression whereby balance is achieved via the dual-source of variation in the limit with an associated common convergence rate. The error in the literature arises because the random variability of the bias term has been neglected in earlier research. In stationary regression this random variability is of smaller order and can correctly be neglected in asymptotic analysis but with consequences for finite sample performance. In nonstationary regression, variability typically has larger order due to the nonstationary regressor and its omission leads to deficiencies and partial failure in the asymptotics reported in the literature. Existing results are shown to hold only in scalar covariate FCC regression and only when the bandwidth has order larger than $1/n$ and smaller than $1/\sqrt{n}$ for sample size n. The correct results in cases of a multivariate covariate z_t are substantially more complex and are not covered by any existing theory. Implications of the findings for inference, confidence interval construction, bandwidth selection, and stability testing for the functional coefficient are discussed. A novel self-normalized t-ratio statistic is developed which is robust with respect to bandwidth order and persistence in the regressor, enabling improved testing and confidence interval construction. Simulations show superior performance of this robust statistic that corroborate the finite sample relevance of the new limit theory in both stationary and nonstationary regressions.

JEL classification: C14; C22.

Keywords: Bandwidth selection; Bias variability; Functional coefficient cointegration; Kernel regression; Nonstationarity; Robust inference; Sandwich matrix.

*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.

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 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 FCC model of Xiao (2009) 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; Wang et al., 2019), 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. 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 $O(\frac{1}{\sqrt{n}})$ 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 $\hat{\beta}(z) = (\sum_{t=1}^n x_t y_t K_{tz}) / \sum_{t=1}^n x_t^2 K_{tz}$ 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.

(i) Stationary x_t

For stationary x_t and under commonly used regularity conditions for smooth FC regression, 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). It is therefore typically ignored in the limit theory. Indeed, under regularity conditions that enable use of laws of large numbers and triangular array martingale central limit theory, the three components of (1.3) and the signal function have the following standard asymptotic behavior (c.f., Li and Racine (2007), theorem 9.3) as $n \to \infty$ and $h \to 0$ with the effective sample size $nh \to \infty$

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

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

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

[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.

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

where $\mu_2(K) = \int s^2 K(s)ds$, $\nu_0(K) = \int K(s)^2 ds$, $\int K(s)ds = 1$, $\mathbb{E}x_t^2 = \sigma_x^2$, $\mathbb{E}u_t^2 = \sigma_u^2$, $C(z) = \frac{1}{2}\beta^{(2)}(z)f(z) + \beta^{(1)}(z)f^{(1)}(z)$ with $g^{(j)}$ signifying the j-th 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). 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 limit theory then 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 centred 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, as shown in Lemma B.1(c)(i) in the Appendix,

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

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. 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: it transpires that 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 the Appendix, 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 1 of the supplementary file 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.

We propose a new self-normalized t-ratio that incorporates both the traditional variance term and the random bias component. The statistic is constructed using a sandwich variance estimate that retains both asymptotically relevant and negligible components allowing them to figure automatically to whatever bandwidth order is used. In doing so, the self-normalized statistic accommodates all three categories of limit theory in Theorem 2.1 and is therefore robust to bandwidth order. This statistic turns out to be robust to the persistence properties of the regressor, so that the same statistic remains valid in the stationary model. Simulations show excellent performance compared to other approaches. Both the asymptotic theory and simulation findings recommend using this robust statistic for testing and confidence interval construction. When the regressor is nonstationary and covariate z_t is univariate the bandwidth order $h = O(n^{-1/2})$ gives rate optimality in estimation and performs well in inference.

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). 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. Section 3 discusses implications of these asymptotics for rate efficient estimation, inference and bandwidth selection. This section also develops the new robust T-ratio, a corresponding T^2 statistic, and reports simulation findings. Section 4 concludes. Proofs of the main results and key subsidiary lemmas are given in the Appendix. The online supplement accompanying the paper provides additional simulation evidence corroborating the new asymptotics. Testing the constancy of the functional coefficient is also studied in the online supplement where an easy-to-implement test statistic is considered and found to have finite sample size and power in accord with asymptotic theory.

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 $\xi_n = \eta_n\{1 + o_p(1)\}$. We use $\equiv_d$ to signify equivalence in distribution and, as in the usage 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 the subsequent analysis 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 $\mathcal{F}_t = \sigma\{\{x_s, z_s\}_{s=1}^{\infty}; u_t, u_{t-1}, \cdots\}$, $\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2$ 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 details the limit theory for the FC cointegrating 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 $nh^2 \to 0$,

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

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* $\beta^{(1)}(z) \neq 0$,

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

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* $\beta^{(1)}(z) \neq 0$,

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

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

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

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

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 (i) of all the items 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) Notably in the present case, the 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 small enough to be 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 the case of bandwidths for which $h \to 0$ slower than $1/\sqrt{n}$. Notably in this case the convergence rate of $\hat{\beta}(z)$ has the unusual form $\sqrt{\frac{n}{h}}$, which is $o(n^{\frac{3}{4}})$ and thus 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 $h = O(\frac{1}{\sqrt{n}})$. Notably, the bias term in (2.3) can be ignored in this case* *without any undersmoothing because $n^{\frac{3}{4}} \times h^2 = n^{-\frac{1}{4}} \to 0$. More importantly, the asymptotics in this case 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 of those two cases in which $h = O(\frac{1}{\sqrt{n}})$. Thus, this boundary case where $h = O(\frac{1}{\sqrt{n}})$ has the optimal convergence rate $O(n^{\frac{3}{4}})$ for $\hat{\beta}(z)$. Due to the mutual independence of $\{u_t\}$ and $\{z_t\}$, those two components are uncorrelated. This leads to the mixed normal distribution given in (2.3). Note that the constant c adjusts the relative contributions to the asymptotic variance that come from the random element in the bias function and the usual variance term.*

Remark 2.4. ***(Degeneracy)*** *From the definition of $\Omega_\beta(z)$, it is clear that if $\beta^{(1)}(z) = 0$ the limit distribution in (2.2) is degenerate, in which case there is a rise in the convergence rate. The simplest example occurs when $\beta(z) \equiv \beta$ is constant and the functional coefficient model is parametric. In this case, $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz} = 0$ and $\eta_t = 0$ for all t, so 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 $\mathcal{B}(z) = 0$. More discussion about this degenerate parametric case is given in Section 4 of the online supplement, where testing constancy is discussed. A general treatment of cases where the functional coefficient is locally flat to an arbitrary order in both stationary and cointegrating regressions is provided in Phillips and Wang (2020).*

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 forms of mixed normal limit theory, 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 distribution forms seem to suggest that (i) bandwidth specific formulations of the test statistic may be needed for inference, and (ii) major differences arise between stationary and nonstationary cases. However, construction of a general self-normalized test statistic for inference about the functional coefficient turns out to be possible and is applicable in all these cases. This construction of a robust test for inference about $\beta(z)$ is developed in the next section.*

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 the following theorem shows, it is possible to relax this requirement in the presence of a nonstationary regressor due to its stronger signal and the resulting enhancement of the regression signal strength that 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. Moreover, while nonstationarity may allow for very small bandwidths in the asymptotic development, practical issues in kernel smoothing inevitably 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.*

Some general discussion of this result and comparisons of convergence rates with Theorem 2.1 are in order. The condition $nh \to c \in [0, \infty)$ means that h tends to zero as fast or faster than $O(n^{-1})$. Moreover $nh \to c \in [0, \infty)$ implies $nh^2 \to 0$ and thereby matches the condition of Case (a) of Theorem 2.1, 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 naturally included subject to additional conditions that ensure computability of the estimate $\hat{\beta}(z)$, which in turn relies on positivity of the finite sample weighted regression signal $(\sum_t x_t x_t' K_{tz})$. 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, the bandwidth $h \to 0$ 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 is unaffected by the bandwidth.*

Remark 2.7. *($nh \to 0$ **and** $n^3h \to \infty$) 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.21) 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 $n^3h \to \infty$, in which case the kernel weighted signal matrix*

$$\sum_{t=1}^n x_t x_t' K_{tz} = \sum_{t=1}^n x_t x_t' \mathbb{E} K_{tz} + \sum_{t=1}^n x_t x_t' \zeta_{tK} = O_p(n^2h) + O_p(\sqrt{n^3h}) = O_p(\sqrt{n^3h}) \to \infty, \quad (2.6)$$

which means that persistent excitation still holds. The justification of (2.6) is as follows. Recall that $EK_{tz} = hf(z) + o(h)$ and $\sum_{t=1}^n x_t x_t' = O_p(n^2)$ as $n \to \infty$, so that $\sum_{t=1}^n x_t x_t' \mathbb{E} K_{tz} = O_p(n^2h)$. 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)

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

in the iid z_t case. Hence, $\sum_{t=1}^n x_t x_t' \zeta_{tK} = O_p(\sqrt{n^3h})$. Consequently, $\sum_{t=1}^n x_t x_t' K_{tz} = O_p(n^2h) + O_p\left(\sqrt{n^3h}\right) = O_p\left(\sqrt{n^3h}\right)$ when $nh \to 0$. In this case, we might 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*

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

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* $O_p(1/\sqrt{n})$ *term in the asymptotics and the* $o_p(1/\sqrt{n})$ *term in (2.9) comes from the term involving 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. *(*$n^3 h \to 0$*) Remark 2.7 establishes consistency when* $n^3 h \to \infty$*. We may well have expected inconsistency if* $n^3 h \to 0$ *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* $\sum_{t=1}^{n} x_t x_t' K_{tz} = O_p(\sqrt{n^3 h})$ *for* $nh \to 0$ *as before, yet now* $\sum_{t=1}^{n} x_t x_t' K_{tz} = o_p(1)$ *when* $n^3 h \to 0$ *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*

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

which holds even when $n^3 h \to 0$*. 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

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

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

Next consider the general case where z_t is multivariate of dimension q. Let $z = (z_1, \cdots, z_q)'$ and $z_t = (z_{1t}, \cdots, z_{qt})'$. For convenience in estimation we use the product kernel $K_{tzq} := K_q(z_t) := k_{tz_1} \times \cdots \times k_{tz_q}$ where $k_{tz_j} = k((z_{jt} - z_j)/h_j)$, $j = 1, \cdots, q$, $k(\cdot)$ is a symmetric second order kernel, and the $h_j, j = 1, \cdots, q$, 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

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

(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

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

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

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

(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 Part (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 still consistent even when $nh^q \to \infty$ fails. In this event, there is no invariance principle and the result corresponds to Theorem 2.2 when $q = 1$. Notably when $q > 2$ and $nh^q \to 0$, $nh^2 \to c \in (0, \infty)$ 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.

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 an accompanying central limit distribution theory.

3 Rate Efficient Estimation and Robust Inference

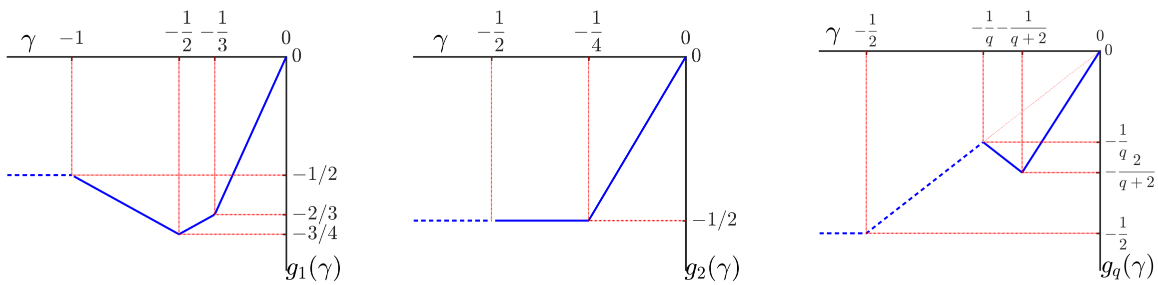
3.1 Optimal bandwidth order and rate efficient estimation

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 $-1/3 \leq \gamma < 0$. In view of Theorem 2.2, we have $g_1(\gamma) = -1/2$ for $\gamma \leq -1$. Collectively, we obtain

$$g_1(\gamma) = \begin{cases} -1/2 & \gamma \leq -1 & \text{No IP} \\ -(1+\gamma/2) & -1 < \gamma \leq -1/2 & \text{IP} \\ -\frac{1-\gamma}{2} & -1/2 < \gamma < -1/3 & \text{IP} \\ 2\gamma & -1/3 \leq \gamma < 0 & \text{IP} \end{cases} . \tag{3.1}$$

The function $g_1(\gamma)$ is plotted in Figure 1(a), in which the dashed part of the function depicts



(a) $q = 1$ (b) $q = 2$ (c) $q \geq 3$

Figure 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.

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

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

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

$$g_2(\gamma) = \begin{cases} -1/2 & \gamma \leq -1/2 & \text{No IP} \\ -1/2 & -1/2 < \gamma < -1/4 & \text{IP} \\ 2\gamma & -1/4 \leq \gamma < 0 & \text{IP} \end{cases}. \tag{3.2}$$

Figure 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

$$g_q(\gamma) = \begin{cases} -1/2 & \gamma \leq -1/2 & \text{No IP} \\ \gamma & -1/2 < \gamma \leq -1/q & \text{No IP} \\ -\frac{1+\gamma(q-2)}{2} & -1/q < \gamma \leq -1/(q+2) & \text{IP} \\ 2\gamma & -1/(q+2) < \gamma < 0 & \text{IP} \end{cases}, \tag{3.3}$$

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 Figure 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 Figure 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 $\sqrt{nh^{q-2}} \times h^2 = n^{\frac{1}{2}+\frac{\gamma_q^*(q-2)}{2}+2\gamma_q^*} = n^0 = O(1)$.

3.2 Behavior of t-ratios constructed according to limit theory

To illustrate the differences in the asymptotics presented in Theorems 2.1, we conduct simulations of the distributions of the following standardized and bias corrected infeasible t-ratios

$$t_1(z) = \frac{n\sqrt{h}(\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z))}{\sqrt{\Omega_u(z)}}, \tag{3.4}$$

$$t_2(z) = \frac{\sqrt{n/h}(\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z))}{\sqrt{\Omega_\beta(z)}}, \tag{3.5}$$

$$t_3(z) = \frac{n^{3/4}(\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z))}{\sqrt{c^{-1/2}\Omega_u(z) + c^{1/2}\Omega_\beta(z)}}, \tag{3.6}$$

using the exact formulae for the asymptotic bias function $\mathcal{B}(z)$ and the true asymptotic variances $\Omega_u(z)$ and $\Omega_\beta(z)$. Note that c is a nonzero constant depending on the bandwidth formula. Since we adopt $h = \hat{\sigma}_z n^{-1/2}$ in this specific case where $nh^2 = O(1)$, we set $c = \sigma_z^2$, the variance of z_t. When $nh^2 \to 0$, we have from Theorem 2.1

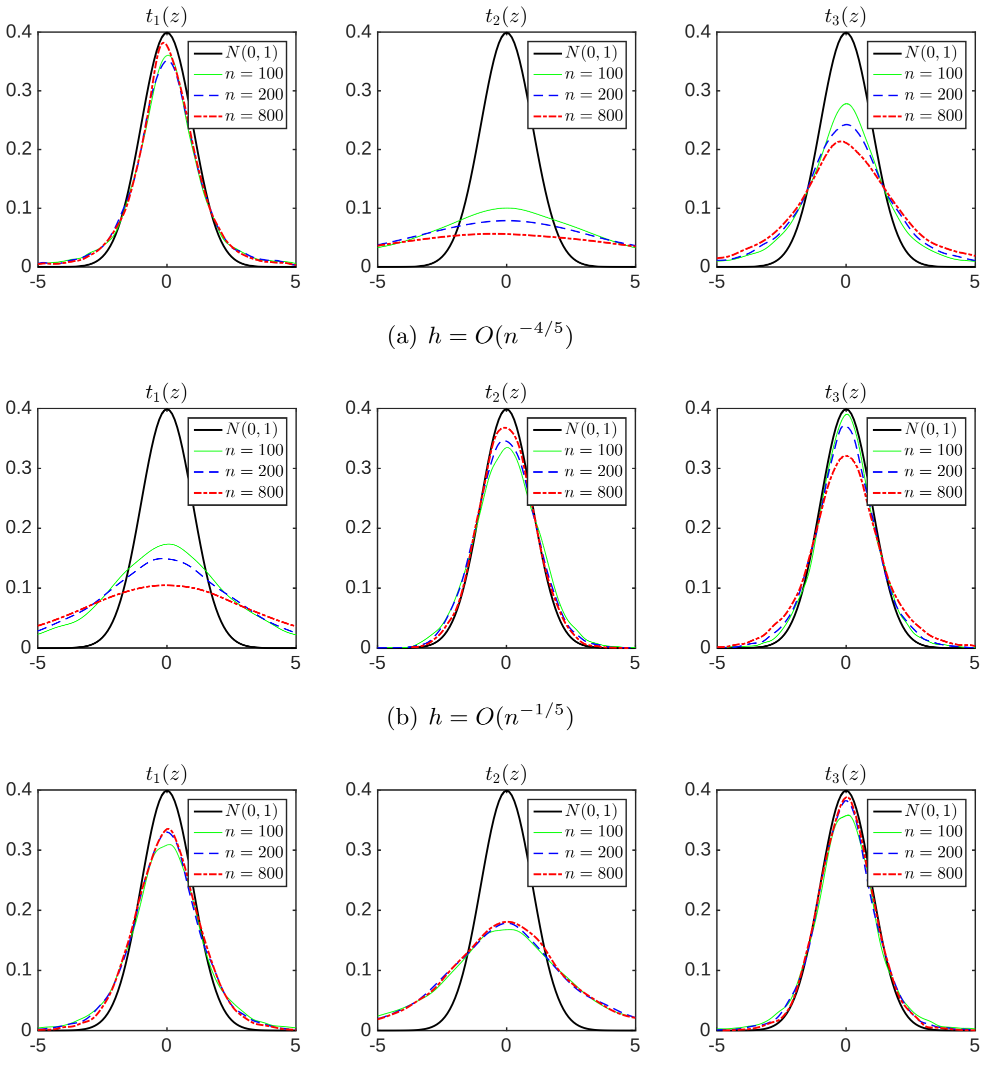
$$t_1(z) \rightsquigarrow N(0,1),\ t_2(z) = O_p\left(\frac{\sqrt{n/h}}{nh^{1/2}}\right) = O_p\left(\frac{1}{\sqrt{nh^2}}\right) \to \infty,\ t_3(z) = O_p\left(\frac{n^{\frac{3}{4}}}{nh^{\frac{1}{2}}}\right) = O_p\left(\frac{1}{(nh^2)^{\frac{1}{4}}}\right) \to \infty.$$

When $nh^2 = O(1)$, we have $t_1(z), t_2(z) = O_p(1)$ and $t_3(z) \rightsquigarrow N(0,1)$. When $nh^2 \to \infty$,

$$t_1(z) = O_p\left(\frac{n\sqrt{h}}{\sqrt{n/h}}\right) = O_p\left(\sqrt{nh^2}\right) \to \infty,\ t_2(z) \rightsquigarrow N(0,1),\ t_3(z) = O_p\left(\frac{n^{3/4}}{\sqrt{n/h}}\right) = O_p((nh^2)^{\frac{1}{4}}) \to \infty.$$

We plot the empirical densities of these t-ratios in Figure 2. The simulation design sets the innovations of x_t to be *iid* $\mathcal{N}(0,1)$, u_t to be *iid* $\mathcal{N}(0,1)$, and z_t to *iid* $\mathcal{U}(0,2)$, making these components mutually independent. The functional coefficient function is $\beta(z) = 1 + z^3$. We evaluate the t-ratios at $z = 1$. The bandwidth is given by $h = c_h \times \hat{\sigma}_z n^\gamma$ with $\gamma \in \{-4/5, -1/2, -1/5\}$ and the second order Epanechnikov kernel is used in estimation. We use the constant coefficient $c_h = 2$ when $\gamma = -4/5$, which mitigates the signal failure that can accompany extremely small bandwidths as discussed in Remark 2.8, and let $c_h = 1$ otherwise. Three sample sizes $n = 100, 200, 800$ are used. The number of replications is 10,000.

The main findings in Figure 2 are summarized as follows. First, when $h = O(n^{-4/5})$, which implies $nh^2 \to 0$, the empirical densities of $t_1(z)$ are very close to the standard normal density for all sample sizes, whereas the densities of $t_2(z)$ and $t_3(z)$ appear to diverge from the standard normal. The divergence of $t_3(z)$ is apparently much slower than that of $t_2(z)$ and this is due to its slower divergence rate in the asymptotic theory in this case. Second, when $h = O(n^{-1/5})$ which implies $nh^2 \to \infty$, panel (b) of Figure 2 shows that the distribution of $t_1(z)$ diverges significantly from the standard normal as n increases whereas the distribution of $t_2(z)$ appears to move slowly towards the standard normal. The density of $t_3(z)$ also diverges from standard normal but at a slower rate than $t_1(z)$. Third, when $h = O(n^{-1/2})$, which implies $nh^2 = O(1)$, the asymptotic theory suggests that all three t-ratios are $O_p(1)$, but only $t_3(z)$ has a standard normal limit distribution. This feature is corroborated by the appearance of the densities in panel (c) of Figure 2, where the empirical density of $t_3(z)$ is seen to be extremely close to standard normal for all sample sizes, whereas the densities of $t_1(z)$ and $t_2(z)$ are stable as n increases but are clearly differentiated from the standard normal for all sample sizes. Overall, these results support the analytic findings concerning the limit behavior of the three t-ratios with different bandwidth orders.



(a) $h = O(n^{-4/5})$

(b) $h = O(n^{-1/5})$

(c) $h = O(n^{-1/2})$

Figure 2: Empirical densities of the three t-ratios $t_i(z), i = 1, 2, 3$ defined in (3.4)-(3.6) shown against the standard $\mathcal{N}(0, 1)$ density.

One approach to practical implementation is to employ estimated bias and variance components in the formulation of $t_1(z), t_2(z), t_3(z)$ leading to estimated ratios of the form

$$\hat{t}_1(z) = \frac{n\sqrt{h}(\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z))}{\sqrt{\hat{\Omega}_u(z)}}, \tag{3.7}$$

$$\hat{t}_2(z) = \frac{\sqrt{n/h}(\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z))}{\sqrt{\hat{\Omega}_\beta(z)}}, \tag{3.8}$$

$$\hat{t}_3(z) = \frac{n^{3/4}(\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z))}{\sqrt{\hat{c}^{-1/2}\hat{\Omega}_u(z) + \hat{c}^{1/2}\hat{\Omega}_\beta(z)}}, \tag{3.9}$$

where $\hat{\mathcal{B}}(z)$, $\hat{\Omega}_u(z)$ and $\hat{\Omega}_\beta(z)$ are consistent estimates. Details regarding these estimates are provided in Section 3 of the online supplement, where the empirical densities of the estimated t-ratios are displayed in Figure S1. The patterns exhibited are very similar to those of Figure 2 across all panels, although some discrepancies are evident for sample size $n = 100$. Again these findings are broadly consistent with the asymptotic results of Theorem 2.1. In consequence, these estimated statistics suffer from the same sensitivities to bandwidth choice and divergences from standard $\mathcal{N}(0,1)$ limit theory as the infeasible statistics $\{t_1(z), t_2(z), t_3(z)\}$. This case-specific approach to inference is evidently unsatisfactory for practical work.

3.3 Robust self-normalized T-ratio and T^2 statistic

The statistics $\{\hat{t}_j(z) : j = 1,2,3\}$ in (3.7)-(3.9) all make explicit use of the rates of convergence and the explicit limit distribution theory of $\hat{\beta}(z)$ for the different bandwidth ranges. In doing so, this framework for inference differs substantially from the stationary case and is not well suited for empirical implementation because of the multiple asymptotic forms and the highly variable performance of each individual statistic over the different bandwidth ranges, documented in the online supplement. We therefore suggest a new approach to develop a robust statistic that covers all three nonstationary x_t cases as well as the stationary case.

The approach constructs a self-normalized statistic that includes the two potentially relevant ('bias' and 'variance') components that contribute to the variation of $\hat{\beta}(z)$, but without prior standardization on them. By combining these two sources of potential variation in the statistic the new construction succeeds in automatically standardizing each component according to the bandwidth employed, thereby covering all three versions of the limit theory in Theorem 2.1 as well as the stationary x_t case.

The new statistic uses a sandwich form for the variance matrix estimate in self-normalization. In the multiple regressor with single covariate z_t case, define the standardized and bias corrected estimation error

$$\hat{T}(z) = V_n(z)^{-\frac{1}{2}}[\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z)] \tag{3.10}$$

with self-normalizing sandwich form matrix $V_n(z) = A_{nz}^{-1}\Omega_n(z)A_{nz}^{-1}$ where

$$\Omega_n(z) = \nu_0(K)\hat{\sigma}_u^2 \sum_{t=1}^{n} x_t x_t' K_{tz} + \sum_{t=1}^{n} x_t x_t'(x_t'\hat{\beta}^{(1)}(z))^2(z_t - z)^2 K_{tz}^2, \tag{3.11}$$

and $A_{nz} = \sum_{t=1}^{n} x_t x_t' K_{tz}$. In (3.11) $\hat{\sigma}_u^2$, $\hat{\beta}^{(1)}(z)$, and $\hat{\mathcal{B}}(z)$ denote consistent estimates of the corresponding error variance, derivative, and bias components. These consistent estimates may be obtained in the usual manner, details of which are discussed later under practical implementation and in Section 3 of the online supplement.

Theorem 3.1 below shows that $\hat{T}(z)$ is robust and asymptotically pivotal with the same standard $\mathcal{N}(0, I_d)$ limit theory for each of the three bandwidth ranges (a), (b) and (c) of Theorem 2.1, as well as for stationary x_t. The statistic $\hat{T}(z)$ may therefore be used to form a robust test across bandwidths and persistence properties of the regressor x_t. Specifically, joint robust inference about $\beta(z)$ across bandwidths and for both stationary and nonstationary regressors is possible using a test based on a quadratic form in $\hat{T}(z)$, just as in Hotelling's T^2 statistic. This test statistic is

$$\hat{T}_2(z) = \hat{T}(z)'\hat{T}(z) = [\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z)]'V_n(z)^{-1}[\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z)], \tag{3.12}$$

which is asymptotically χ^2_d in both stationary and nonstationary regressor cases. To include the stationary case in the analysis we introduce the following Assumption.

Assumption 2. *(i) $\{u_t\}$ 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 $\mathbb{E}(u_t^4) < \infty$; and $\{x_t, z_t\}$ are strictly stationary and independent α-mixing scalar 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 $p > 2\delta > 4$ and $\mathbb{E}[x_t x_t'|z_t = z]$ is positive definite a.s.*

(ii) The density $f(z)$ of z_t, joint density $f_{0,j}(s_0, s_j)$ of (z_t, z_{t+j}), joint density $f(x, z)$ of (x_t, z_t), and conditional density $f(x|z)$ of x_t given $z_t = z$ 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.

Assumptions 2(i)-(iii) retain the martingale difference condition on the equation errors $\{u_t\}$ and the mutual independence condition of $\{x_t\}$ and $\{z_t\}$ that was made in Assumption 1(iii). The latter requirement was not critical in the nonstationary case but considerably complicates the limit theory by introducing dependences in the limit processes and additional terms in the limiting stochastic integral representations. In the stationary case, dependence between x_t and z_t can be handled by suitable conditioning arguments in the kernel limit theory and standard methods (Li and Racine, 2007, chapter 9.3). The proof of Theorem 3.1 applies when Assumption 2(iii) is relaxed and specialization to the mutual independence case provides concordance with the conditions for the nonstationary regressor case, thereby establishing robustness of the test statistic. Specifically, under Assumption 2 in the stationary case and the earlier conditions given in Assumptions 1 in the nonstationary case we have the following common result.

Theorem 3.1. *Under Assumptions 1(iv)-(vi) and either Assumptions 1(i)-(iii) or Assumptions 2(i)-(iii), when $nh \to \infty$, $\hat{T}(z) \rightsquigarrow \mathcal{N}(0, I_d)$ and $\hat{T}_2(z) \rightsquigarrow \chi^2_d$.*

Remark 3.1. *The simple idea leading to (3.10) and the statistic (3.12) comes from the following decomposition of the estimation error (see (A.8) in the proof of Theorem 2.1 and (1.8) in the scalar regressor case)*

$$\begin{aligned}
&\hat{\beta}(z) - \beta(z) - \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \\
&= \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t x_t' \eta_t + \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t u_t K_{tz} \\
&= \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \left\{\sum_{t=1}^{n} x_t x_t' \eta_t + \sum_{t=1}^{n} x_t u_t K_{tz}\right\}.
\end{aligned} \tag{3.13}$$

As evident from (3.13) and discussed in the Introduction, the two components within the right side braces both contribute to finite sample variation in $\hat{\beta}(z)$. Correspondingly, the central sandwich matrix $\Omega_n(z)$ in (3.11) within $V_n(z)$ is designed to capture finite sample variation from these two sources. This matrix accounts for all cases where the random element of the 'bias' term and/or the usual 'variance' term contribute to the limit distribution, thereby ensuring that the sandwich matrix $V_n(z)$ is a suitable normalizing matrix in all cases, covering stationary regression models as well as cointegration models. The data-based specification of $V_n(z)$ ensures automatic embodiment of these necessary components to cover the full asymptotic theory and no further normalization is needed.

Remark 3.2. *An alternative form of the sandwich matrix $\Omega_n(z)$ in (3.11) is*

$$\Omega_n^*(z) = \nu_0(K)\hat{\sigma}_u^2 \sum_{t=1}^n x_t x_t' K_{tz} + \sum_{t=1}^n x_t x_t' (x_t'[\hat{\beta}(z_t) - \hat{\beta}(z)]K_{tz})^2 \tag{3.14}$$

where the 'bias' term influence manifests immediately through the presence of the estimated approximation error $\hat{\beta}(z_t) - \hat{\beta}(z)$ in the second term of (3.14). The sandwich matrix $\Omega_n(z)$ uses instead the linear approximation $\hat{\beta}^{(1)}(z)(z_t - z)$ involving the estimated derivative $\hat{\beta}^{(1)}(z)$ which more directly focuses attention on behavior at the point of estimation z. Simulations (not reported) show that the use of the sandwich matrix $\Omega_n(z)$ rather than $\Omega_n^*(z)$ improves finite sample performance in the T ratio statistic, leading to the recommended form $\Omega_n(z)$ in (3.11). Importantly, neither of the two terms in the sandwich matrices $\Omega_n(z)$ and $\Omega_n^*(z)$ is normalized. As shown in the proof of Theorem 3.1, explicit normalization is unnecessary. The respective orders of magnitude of the two terms in (3.11) (and (3.14)) automatically determine which term plays a role in the asymptotic theory or indeed whether both terms are needed. Thus, when $nh^2 \to 0$ the first term in $\Omega_n(z)$ or $\Omega_n^*(z)$ dominates, mirroring the limit theory in Theorem 2.1(a); when $nh^2 \to \infty$ the second term in $\Omega_n(z)$ and $\Omega_n^*(z)$ dominates, mirroring the limit theory in Theorem 2.1(b); and when $nh^2 \to c \in (0, \infty)$ these terms have the same order and both influence the limit theory of $\hat{T}(z)$, thereby capturing the rate efficient behavior of the estimator $\hat{\beta}(z)$, mirroring Theorem 2.1(c). As shown in the proof of Theorem 3.1 in the stationary case only the first term of $\Omega_n(z)$ or $\Omega_n^*(z)$ figures in the limit theory. Nonetheless, finite sample behavior is in all cases influenced by the two terms. As the simulations reported below confirm, this analysis is corroborated in the good finite sample performance of the robust statistic $\hat{T}(z)$.

Remark 3.3. *To fix ideas in the nonstationary case, it is convenient to write (3.13) in standardized and centered form as*

$$\begin{aligned}
&\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h} \sum_{t=1}^n x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^3} \sum_{t=1}^n x_t x_t' \mathbb{E}\xi_{\beta t} \\
=&\frac{1}{n\sqrt{h}} \left(\frac{1}{n^2 h} \sum_{t=1}^n x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^n x_t u_t K_{tz} + \sqrt{\frac{h}{n}} \left(\frac{1}{n^2 h} \sum_{t=1}^n x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^n x_t x_t' \eta_t.
\end{aligned} \tag{3.15}$$

Then, if $nh^2 \to 0$ the first term of (3.15) dominates because $\frac{h}{n} = o(\frac{1}{n^2h})$; similarly, if $nh^2 \to \infty$, the second term dominates because $\frac{1}{n^2h} = o(\frac{h}{n})$; and when $nh^2 \to c \in (0, \infty)$ both terms are $O(n^{-3/4})$ because $\frac{1}{n\sqrt{h}} \sim_a \frac{1}{c^{1/4}n^{3/4}}$ and $\sqrt{\frac{h}{n}} \sim_a \frac{c^{1/4}}{n^{3/4}}$. The sandwich matrix $\Omega_n(z)$ in (3.11) accommodates all these possibilities by the inclusion of an estimated sample variance term for each component that contributes to the variation of $\hat{\beta}(z)$. Thus, $\nu_0(K)\hat{\sigma}_u^2 \sum_{t=1}^n x_t x_t' K_{tz}$ estimates the conditional variance of the martingale $\sum_{t=1}^n x_t u_t K_{tz}$ and $\sum_{t=1}^n x_t x_t'(x_t'\hat{\beta}^{(1)}(z))^2(z_t - z)^2 K_{tz}^2$ estimates the conditional variance of the martingale $\sum_{t=1}^n x_t x_t' \eta_t$. The composition of these components in $\Omega_n(z)$ may then be considered an estimate of the overall martingale conditional variance, making it a suitable variance matrix in the sandwich form $V_n(z)$ for self-normalization of the estimation error of $\hat{\beta}(z)$. A similar rationale for the use of $\Omega_n(z)$ applies in the stationary case, thereby justifying the robustness of the statistic to persistence properties of the regressor.

Remark 3.4. *These heuristic arguments suggest that the approach employed here to produce a test statistic that is robust to bandwidth choice and persistence in the regressor may be used in more general cases such as those where there may be conditional or unconditional heterogeneity in the equation error u_t. Extensions of the present framework to such cases may be considered in future work.*

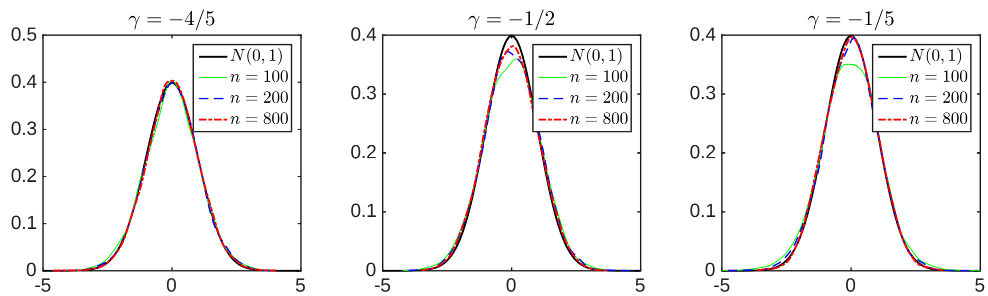
Simulations were conducted to assess the finite sample performance of the self-normalized t-ratio $\hat{T}(z)$. We focus here on the scalar regressor case where $\hat{T}(z)$ is a scalar test statistic. In the multiple regressor case, the elements of $\hat{T}(z)$ may be used for individual tests about the components $\beta_i(z)$ and the $\hat{T}_2(z)$ statistic may be used for testing joint hypotheses about $\beta(z)$.

Figure 3 shows the finite sample densities of $\hat{T}(z)$ under different bandwidths with panel (a) for nonstationary x_t and panel (b) for stationary x_t. The DGP for the nonstationary case is the same as that for Figure 2 and Figure S1 in the online supplement. So is the bandwidth used in the computation. In the stationary model the regressor x_t follows the autoregression $x_t = 0.5x_{t-1} + \epsilon_t$ where ϵ_t is $iid\mathcal{N}(0, 1)$. From Figure 3 the densities of the self-normalized t-ratio $\hat{T}(z)$ are all very close to that of standard normal except for some discrepancy in the stationary case when $\gamma = -1/5$. Further, comparing panel (a) of Figure 3 with the diagonal plots in Figure S1 of the online supplement, which demonstrate the performance of the correctly standardized t-ratios that apply under specific bandwidth orders, we find that $\hat{T}(z)$ outperforms the correctly standardized individual t-ratios under all three bandwidth orders and the improvement is more prominent for small γ. Similarly, comparison between panel (b) of Figure 3 and Figure 4, which shows the finite sample densities of the conventional t-ratio in the stationary case[2], suggests that $\hat{T}(z)$ continues to dominate the conventional t-ratio in performance accuracy. Only in the case of bandwidth $\gamma = -1/5$ are the results comparable in the stationary case. Overall, the self-normalized t-ratio $\hat{T}(z)$ outperforms the other approaches almost uniformly and shows strong robustness to bandwidth and persistence properties of x_t. These simulation findings support the analytic results in recommending the use of $\hat{T}(z)$ and $\hat{T}_2(z)$ in applications.

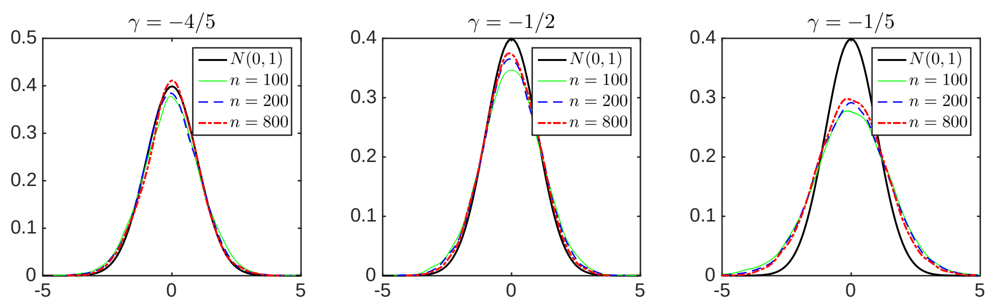
Practical implementation of $\hat{T}(z)$ relies on computing some estimated components, which we now clarify. The variance σ_u^2 is estimated by $n^{-1}\sum_t \hat{u}_t^2$ where $\hat{u}_t = y_t - x_t\hat{\beta}(z_t)$ and $\hat{\beta}(z_t)$ is estimated via local level estimation with bandwidth $\hat{\sigma}_z n^{-1/2}$. The choice $\gamma = -1/2$ is rate

[2]The conventional t-ratio is computed as $\sqrt{nh}[\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z)]/\left\{\hat{\sigma}_u^2\nu_0(K)\hat{f}^{-1}(z)(\mathbb{E}x_t^2)^{-1}\right\}^{1/2}$, which matches the construction of $\hat{t}_1(z)$ in the nonstationary case.

efficient in the nonstationary case and achieves undersmoothing in the stationary case. Bias estimation of $\mathcal{B}(z)$ requires estimation of $f(z)$, $f^{(1)}(z)$, $\beta^{(1)}(z)$ and $\beta^{(2)}(z)$. The traditional kernel density estimator is used to estimate $f(z)$ employing the conventional optimal bandwidth order $n^{-1/5}$. The density derivative $f^{(1)}(z)$ is estimated by the derivative of the kernel density estimator, which is $-(nh^2)^{-1}\sum_{t=1}^{n} K^{(1)}((z_t - z)/h)$, with the conventional optimal bandwidth order $n^{-1/7}$ used for that purpose. Local linear estimation was used to estimate the first order derivative $\beta^{(1)}(z)$ and the second derivative $\beta^{(2)}(z)$ was estimated by local quadratic estimation.[3]



(a) x_t is nonstationary



(b) x_t is stationary

Figure 3: Empirical densities of the self-normalized t-ratio $\hat{T}(z)$ in (3.10) with the asymptotic standard $\mathcal{N}(0,1)$ density for reference.

To appreciate the benefit of using the robust statistics $\hat{T}(z)$ and $\hat{T}_2(z)$ for inference across the support of z_t, we computed average coverage rates of confidence intervals constructed by inverting the respective t-ratio statistics. Computations used the same simulation design and involved 10,000 replications at grid points with a step length of 0.1 over the interval $[0.1, 1.9]$

[3] For nonstationary models no results in the present literature guide bandwidth choices for estimating derivatives of functional coefficients. The bandwidths employed in the simulations were based on the optimizing order obtained from calculations of the bias and variance orders by the authors. For instance, in local linear estimation of the first order derivative $\beta^{(1)}(z)$, balancing bias and variance in the stationary regressor case leads to the bandwidth order $n^{-1/7}$, just as in the standard stationary nonparametric model setting (Li and Racine, 2007, Theorem 2.9). For the nonstationary regressor case, similar calculations yield an optimal rate of $n^{-2/7}$. We therefore employed the intermediate rate $n^{-1.5/7}$ in the simulations to cover both stationary and nonstationary regressor cases. For present purposes consistent estimation of the bias function $\mathcal{B}(z)$ only requires consistent estimation of $f(z)$, $f^{(1)}(z)$, $\beta^{(1)}(z)$ and $\beta^{(2)}(z)$. So rate-efficient bandwidth order selection is likely of secondary importance and this was confirmed in simulation performance. Note that no derivative estimation is needed for the sandwich form $\Omega_n^*(z)$ in the construction of $\hat{T}(z)$.

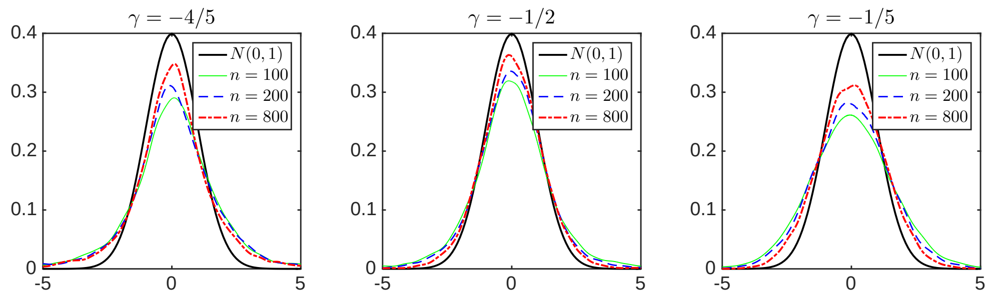


Figure 4: Empirical densities of the conventional t-ratio in the stationary case with the standard $\mathcal{N}(0,1)$ density for reference.

in the support of z_t. The results for a nominal asymptotic 95% confidence band are collected in Table 1.

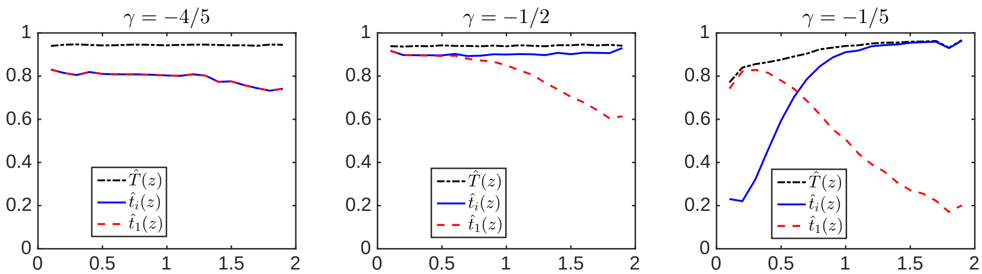
In the nonstationary case, the entries in the left panel of Table 1 use the incorrect limit theory in the literature for which the limit distribution given in (2.1) is used to construct confidence intervals for all bandwidth orders. So the coverage rates shown for the bandwidth parameters $\gamma = -1/2$ and $\gamma = -1/5$ are based on the wrong limit distribution in this panel. The results show severe undercoverage for both bandwidths. The problem of undercoverage is exacerbated as γ increases because the omitted random bias contribution to variation becomes more important for larger bandwidths. Even more severe distortion is evident when $\gamma = -1/5$ because in this case coverage rates decrease rather than increase as the sample size rises. This additional distortion occurs because the neglected random bias contribution turns out to be of greater importance than the traditional variance term. The confidence intervals constructed from $\hat{t}_1(z)$ therefore shrink at a faster rate as the sample size rises, leading to decreased coverage.

Table 1: Average Coverage Rates of the 95% Confidence Intervals

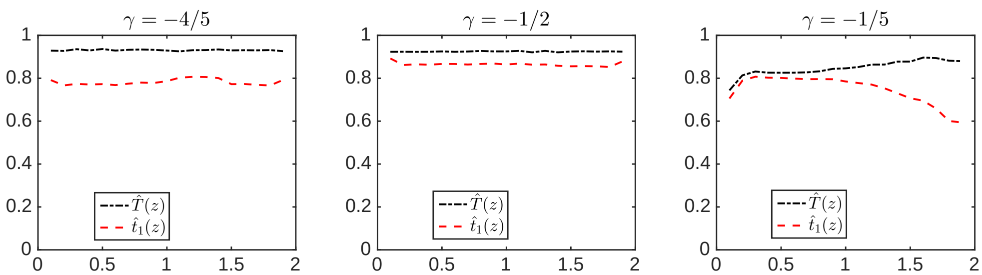
	Based on $\hat{t}_1(z)$			Based on correct $\hat{t}_i(z)$			Based on $\hat{T}(z)$		
γ	-4/5	-1/2	-1/5	-4/5	-1/2	-1/5	-4/5	-1/2	-1/5
				x_t is nonstationary					
$n = 100$	0.76	0.77	0.52	0.76	0.90	0.74	0.94	0.93	0.90
$n = 200$	0.79	0.80	0.51	0.79	0.91	0.76	0.94	0.94	0.91
$n = 400$	0.82	0.82	0.49	0.82	0.92	0.78	0.95	0.94	0.92
$n = 800$	0.84	0.83	0.45	0.84	0.93	0.80	0.95	0.95	0.93
				x_t is stationary					
$n = 100$	0.75	0.82	0.70	-	-	-	0.92	0.91	0.83
$n = 200$	0.78	0.87	0.75	-	-	-	0.93	0.92	0.85
$n = 400$	0.81	0.89	0.78	-	-	-	0.94	0.93	0.86
$n = 800$	0.83	0.91	0.81	-	-	-	0.94	0.94	0.87

Note: Correct $\hat{t}_i(z)$ involves normalization for that bandwidth region.

The middle panel of Table 1 shows results based on the limit theory presented in Theorem 2.1 for the nonstationary regressor case, namely the correctly standardized t-ratios $\hat{t}_i(z), i = 1, 2, 3$ where each statistic is normalized according to the limit theory that applies for the particular



(a) x_t is nonstationary



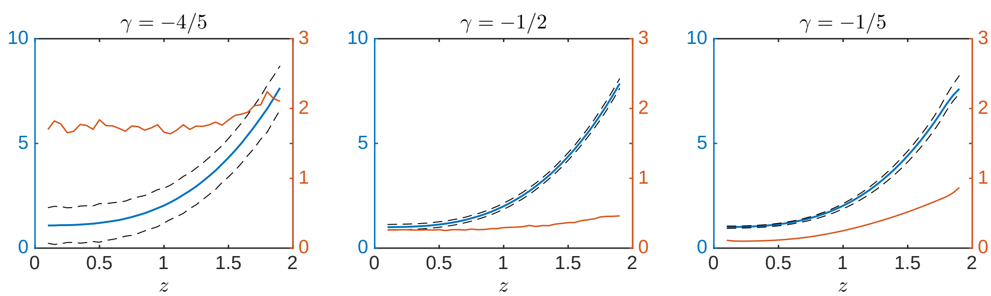
(b) x_t is stationary

Figure 5: Coverage rate curves of the 95% confidence bands over the support of z_t with different bandwidth orders and $n = 200$ based on 10,000 replications: the black dash-dotted line is based on the robust statistic $\hat{T}(z)$, the blue solid line is based on the correct limit theory in Theorem 2.1 for the separate bandwidths, and the red dashed line is based on the traditional result in (2.1).

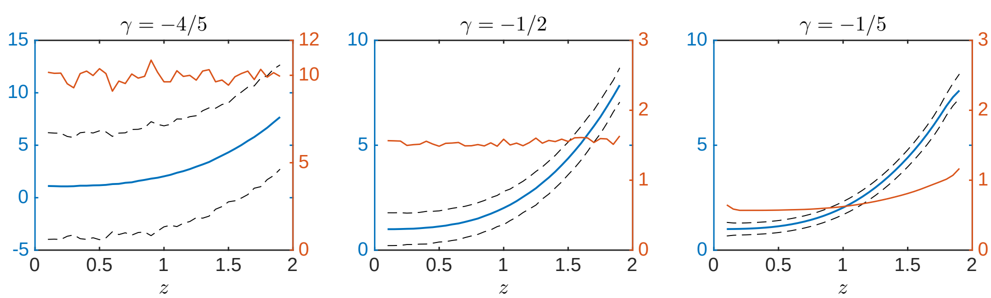
bandwidth region. When $\gamma = -1/2$, the coverage rates are very close to the nominal rate and converge to the nominal rate as the sample size increases, thereby corroborating the new asymptotic theory. When $\gamma = -4/5$, the coverage rate is smaller than the nominal rate but increases with the sample size. This discrepancy from the nominal rate reflects the finite sample effect of ignoring the variation, measured asymptotically by $\Omega_\beta(z)$, associated with the random bias component even though this variation is negligible asymptotically for this bandwidth. Similarly, the undercoverage when $\gamma = -1/5$ is caused by ignoring the variance component $\Omega_u(z)$ in finite samples. The right panel of Table 1 is based on the self-normalized t-ratio $\hat{T}(z)$. The coverage rates are all very close to the nominal rate even when sample size is small, demonstrating strong uniform improvement over that in the middle panel for all bandwidth orders.

In the stationary case, the left panel shows the result based on the usual t-ratio, i.e., the t-ratio plotted in Figure 4. Although this usual t-ratio is asymptotically standard $\mathcal{N}(0,1)$ for all three bandwidth orders, confidence intervals based on it evidently still suffer from severe undercoverage problems for each bandwidth. But as the right panel shows, the self-normalized t-ratio $\hat{T}(z)$ raises the coverage rate uniformly for all bandwidth choices. The improvement is more profound when $\gamma = -4/5$ and $\gamma = -1/2$, for which the coverage rates of the confidence intervals based on $\hat{T}(z)$ are all close to the nominal rate. Even in the case where $\gamma = -1/5$, use of $\hat{T}(z)$ delivers improvements over the usual t-ratio.

Figure 5 plots coverage rate curves over the support of z_t for sample size $n = 200$, which is representative of results for other sample sizes. Evidently, when x_t is nonstationary, the coverage rate curves based on the robust statistic $\hat{T}(z)$ (given by the black dash-dotted line) are the highest across all three bandwidths and are very close to the nominal level when $\gamma = -4/5$ and $-1/2$, consistent with the best performance in Table 1. When $\gamma = -4/5$ the red dashed line (based on existing limit theory and $\hat{t}_1(z)$) overlaps with the blue solid line (based on the correct limit theory in this paper) because for this bandwidth the existing limit theory is correct and given by (2.1). But when $\gamma = -1/2$ and $-1/5$ the red dashed lines show a decreasing pattern of coverage probability as z increases. This pattern is explained by the fact that the derivative $\beta^{(1)}(z)$ is an increasing function of z and so too is $\Omega_\beta(z)$, which depends on this derivative. Hence, the consequence of neglecting the additional variance term $\Omega_\beta(z)$ when computing the confidence bands becomes far worse when z is large, leading to more severe undercoverage. Importantly, when $\gamma = -4/5$ and $-1/2$, the blue solid lines and the black dash-dotted lines are very flat over the entire support of z_t. But when $\gamma = -1/5$ these curves reveal undercoverage in the region of small z. This undercoverage is much less severe for inference based on the robust statistic $\hat{T}(z)$, shown by the black dash-dotted line. Low coverage manifested by the blue solid line for small z occurs because the specific design function $\beta(z) = 1 + z^3$ used in this simulation has derivative function $\beta^{(1)}(z) = 3z^2$ which is small and tends to zero as $z \to 0$, so that $\Omega_\beta(z)$ is correspondingly small and tends to zero with z. Hence, neglecting $\Omega_u(z)$ as the statistic $t_i(z)$ does when $\gamma = -1/5$ leads to particularly low coverage when z is small as well as the low average coverage rate shown in Table 1 under the panel 'Based on correct $\hat{t}_i(z)$'. In the extreme case where $\Omega_\beta(z) = 0$, we have the degeneracy problem discussed in Remark 2.4. Degeneracy leads to a different limit theory, which will be considered in a separate study.



(a) x_t is nonstationary



(b) x_t is stationary

Figure 6: Left scale: Plots of $\hat{\beta}(z)$ (solid blue curve) with 95% confidence bands based on the self-normalized t-ratio $\hat{T}(z)$ (dashed black lines) for different bandwidth orders and $n = 200$ from 10,000 replications. Right scale and curve (solid orange line): confidence interval width.

Panel (b) of Figure 5 shows the coverage rate curves in the stationary case. The red dashed line is again based on the usual t-ratio. Evidently, across all three bandwidths the coverage rate curves based on the robust statistic $\hat{T}(z)$ are uniformly higher over the entire support of z_t than those constructed from the usual t-ratio, giving coverage rates that are very close to the nominal rate when $\gamma = -4/5$ and $-1/2$. For $\gamma = -1/5$ the curves based on $\hat{T}(z)$ show some undercoverage, as in the nonstationary case and for the same reason given above for small z, while still offering better coverage than intervals based on the usual statistic. These results reveal the gains from using the robust statistic $\hat{T}(z)$ for inference in the stationary case rather than the usual t-ratio.

Figure 6 displays the average fitted functional coefficient curves and the corresponding 95% confidence bands based on the self-normalized statistic $\hat{T}(z)$ with panel (a) showing results for nonstationary x_t and panel (b) for stationary x_t. The widths of the corresponding confidence bands are plotted using the right scale. The curves are computed at grid points from 0.1 to 1.9 by step length 0.05 for sample size $n = 200$ with an Epanechnikov kernel and 10,000 replications. Evidently, for both stationary and nonstationary x_t, the confidence bands are the widest when $\gamma = -4/5$, suggesting the coefficient estimates have the largest amount of variation in this case, or equivalently, that estimation accuracy is worse when $\gamma = -4/5$. In the nonstationary case, estimation is most efficient when $\gamma = -1/2$ with the sharpest confidence bands shown in the middle plot of panel (a) of Figure 6, and the close-to-nominal coverage rates over the whole support of z_t evident in the middle plot of panel (a) of Figure 5. These findings corroborate the limit theory which establishes the rate-efficient order $\gamma = -1/2$ in the nonstationary case in Section 3.1 and results given in Table S2 of the online supplement where the IMSE is found to be smallest when $\gamma = -1/2$. In the stationary regressor case, the confidence bands are narrowest when $\gamma = -1/5$ suggesting estimation is most efficient when $\gamma = -1/5$. This outcome is consistent with the conventional result that the rate-efficient bandwidth order in stationary FC estimation is $\gamma = -1/5$. But while the confidence band is a bit wider when $\gamma = -1/2$ than when $\gamma = -1/5$ in the stationary case, it has coverage rate closer to the nominal rate as is evident in panel (b) of Figure 5. Based on coverage accuracy $\gamma = -1/2$ is therefore a good bandwidth choice for both stationary and nonstationary cases.

In sum, these simulations reinforce the analytic findings that $\hat{T}(z)$ has good performance characteristics compared with other approaches, demonstrates robustness to the persistence properties of the regressor x_t, and provides good coverage probability in confidence band construction with the bandwidth rate $\gamma = -1/2$ that is optimal for estimating the functional coefficient in the nonstationary case.

4 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 trade-offs, the present paper shows how useful the (normally ignored) random elements of the bias component can be in sharpening accuracy in estimation and reliability in inference. The analysis of nonstationary functional coefficient models reveals that these elements both figure in the limit theory variance. Taking them into account can deliver rate efficient estimation and provide substantial gains in confidence interval accuracy in finite samples. A key development is the construction of robust statistics that embody the effects of both bias and variance in a normalization that takes a new sandwich matrix form. The finite sample gains from this construction that are evident in nonstationary regressions are shown to carry over to stationary models. It may therefore be expected that this approach to inference will 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 as those shown 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. In addition, similar considerations to those raised here apply to other nonparametric estimators such as local polynomial estimators. Extensions of the results given in the present paper 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

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

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 $nh \to \infty$

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

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), 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)

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

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 $\sum_{t=1}^{n} x_t x_t' K_{tz} = O_p(\sqrt{n^3 h})$ 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),

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

where B_η is vector Brownian motion with variance matrix $Var(B_\eta) = \nu_2(K) f(z) [\beta^{(1)}(z) \beta^{(1)}(z)']$. 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

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

which we write in standardized form as

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

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

$$n\sqrt{h} \left(\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^3} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \right)$$

$$= \sqrt{nh^2}\left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n}x_tx_t'\eta_t + \left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_tu_tK_{tz},$$

$$= o_p(1) + \left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_tu_tK_{tz} \tag{A.10}$$

$$\rightsquigarrow \left(f(z)\int B_xB_x'\right)^{-1}\left(\int B_xdB_{uK}\right) \equiv_d \mathcal{MN}\left(0, \frac{\nu_0(K)\sigma_u^2}{f(z)}\left(\int B_xB_x'\right)^{-1}\right), \tag{A.11}$$

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)

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

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

$$\sqrt{\frac{n}{h}}\left(\hat{\beta}(z) - \beta(z) - h^2\left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{n^2h^3}\sum_{t=1}^{n}x_tx_t'\mathbb{E}\xi_{\beta t}\right)$$

$$= \left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n}x_tx_t'\eta_t + \frac{1}{\sqrt{nh^2}}\left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_tu_tK_{tz} \tag{A.13}$$

$$= \left(\frac{1}{n^2h}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h^3}}\sum_{t=1}^{n}x_tx_t'\eta_t + o_p(1)$$

$$\rightsquigarrow \left(f(z)\int B_xB_x'\right)^{-1}\left(\int B_xB_x'dB_\eta\right) \tag{A.14}$$

$$\equiv_d \mathcal{MN}\left(0, \frac{\nu_2(K)}{f(z)}\left(\int B_xB_x'\right)^{-1}\int B_xB_x'\left(B_x'\beta^{(1)}(z)\right)^2\left(\int B_xB_x'\right)^{-1}\right), \tag{A.15}$$

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

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

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

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

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

$$\begin{aligned} & n^{\frac{3}{4}}\left(\hat{\beta}(z) - \beta(z) - h^2\left(\frac{1}{n^2 h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\frac{1}{n^2 h^3}\sum_{t=1}^n x_t x_t' \mathbb{E}\xi_{\beta t}\right) \\ &= \left(\frac{1}{n^2 h}\sum_{t=1}^n x_t x_t' K_{tz}\right)^{-1}\left[(nh^2)^{\frac{1}{4}}\left(\frac{1}{\sqrt{n^3 h^3}}\sum_{t=1}^n x_t x_t' \eta_t\right) + \frac{1}{(nh^2)^{\frac{1}{4}}}\left(\frac{1}{n\sqrt{h}}\sum_{t=1}^n x_t u_t K_{tz}\right)\right] \end{aligned} \tag{A.19}$$

$$\begin{aligned} & \rightsquigarrow \left(f(z)\int B_x B_x'\right)^{-1}\left(c^{\frac{1}{4}}\int B_x B_x' dB_\eta + \frac{1}{c^{\frac{1}{4}}}\int B_x dB_{uK}\right) \\ & \equiv_d \mathcal{MN}\left(0, c^{\frac{1}{2}}\Omega_\beta(z)\right) + \mathcal{MN}\left(0, \frac{1}{c^{\frac{1}{2}}}\Omega_u(z)\right) = \mathcal{MN}\left(0, c^{\frac{1}{2}}\Omega_\beta(z) + \frac{1}{c^{\frac{1}{2}}}\Omega_u(z)\right), \end{aligned} \tag{A.20}$$

where $\Omega_\beta(z) = \frac{\nu_2(K)}{f(z)}\left(\int B_x B_x'\right)^{-1}\left(\int B_x B_x'\left(B_x'\beta^{(1)}(z)\right)^2\right)\left(\int B_x B_x'\right)^{-1}$, $\Omega_u(z) = \nu_0(K)\sigma_u^2 f^{-1}(z)\left(\int B_x B_x'\right)^{-1}$. 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

$$\left(\hat{\beta}(z)-\beta(z)\right)=\left(\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\sum_{t=1}^{n}x_tx_t'\mathbb{E}\xi_{\beta t} + \left(\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\sum_{t=1}^{n}x_tx_t'\eta_t+\left(\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\sum_{t=1}^{n}x_tu_tK_{tz}. \tag{A.21}$$

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

$$\begin{aligned}\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}&=\frac{\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\mathbb{E}K_{tz}}{h}+\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}}\\&=\frac{\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z)+O(h^2)\right\}+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{h}}.\end{aligned} \tag{A.22}$$

When $nh\to c\in[0,\infty)$ the 'usual' effective sample size nh is asymptotically deficient. In this case, the first term of (A.22) 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

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

By stationarity of z_t and Markov's inequality $\mathbb{P}\left(|\zeta_{tKh}|>M\right)\le\mathbb{E}\zeta_{tKh}^2/M^2=M^{-2}\left\{f(z)\nu_0(K)+O(h)\right\}$, so that for every $\epsilon>0$ there exists a constant M_ϵ such that $\sup_{h\to0}\mathbb{P}\left(|\zeta_{tKh}|>M_\epsilon\right)<\epsilon$ and $\zeta_{tKh}=O_p(1)$ uniformly in t as $h\to0$. 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

$$\frac{2}{n}\sum_{s>t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_s}{\sqrt{n}}\right)^2\right\}\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\frac{\zeta_{sK}}{\sqrt{h}}\right)=\frac{2}{n}\sum_{t=1}^{n}\sum_{j=1}^{n-t}\mathbb{E}\left\{\left(\frac{x_t}{\sqrt{n}}\right)^2\left(\frac{x_{t+j}}{\sqrt{n}}\right)^2\right\}\mathbb{E}\left(\frac{\zeta_{tK}}{\sqrt{h}}\frac{\zeta_{t+jK}}{\sqrt{h}}\right)$$

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

where we use the fact that $\mathbb{E}K_{tz}K_{t+jz} = \int\int K\left(\frac{s_0-z}{h}\right)K\left(\frac{s_j-z}{h}\right)f_{0,j}(s_0,s_j)\,ds_0ds_j = \int\int K(p_0)K(p_j)$ $\times f_{0,j}(z+p_0h, z+p_jh)\,dp_0dp_jh^2 = h^2 f_{0,j}(z,z)+o\left(h^2\right)$ and so $\gamma_\zeta(j) = \mathbb{E}K_{tz}K_{t+jz}-\mathbb{E}K_{tz}\mathbb{E}K_{t+jz} = h^2 f_{0,j}(z,z)+o\left(h^2\right)$. From these calculations of the mean and variance, it follows that the second term of (A.22)

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\zeta_{tK}}{\sqrt{h}} = O_p(1) \tag{A.23}
$$

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

To prove consistency of $\hat{\beta}(z)$ we consider each term on the right side of (A.21) 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

$$
\begin{aligned}
&\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'\mathbb{E}\xi_{\beta t} \\
&=\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{h^3\mu_2(K)C(z)+o(h^3)}{\sqrt{h}} \\
&=\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{h^{5/2}}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{\mu_2(K)C(z)+o(1)\right\} \\
&=\left(\frac{1}{\sqrt{n^3h}}\sum_{t=1}^{n}x_tx_t'K_{tz}\right)^{-1}\frac{h^2\sqrt{nh}}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{\mu_2(K)C(z)+o(1)\right\} \\
&=O_p\left(h^2\sqrt{nh}\right).
\end{aligned}
\tag{A.24}
$$

(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.23) and Lemma B.1(b)(ii) we find that

$$
\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}}\right)^2\frac{\eta_t}{\sqrt{nh^3}} = O_p(1), \tag{A.25}
$$

so that $\sum_{t=1}^{n} x_t x_t' \eta_t = O_p\left(\sqrt{n^3 h^3}\right)$ and then

$$\left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' \eta_t = \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \times O_p\left(\frac{\sqrt{n^3 h^3}}{\sqrt{n^3 h}}\right) = O_p(h). \tag{A.26}$$

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

$$\mathbb{V}\text{ar}\left(\sum_{t=1}^{n} x_t u_t K_{tz}\right) = \sigma_u^2 \sum_{t=1}^{n} \mathbb{E}\left(x_t^2\right) \mathbb{E}\left(K_{tz}^2\right) = O\left(n^2 h\right),$$

so that $\sum_{t=1}^{n} x_t u_t K_{tz} = O_p\left(n\sqrt{h}\right)$ and

$$\begin{aligned} & \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t u_t K_{tz} \\ = & \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \times O_p\left(\frac{n\sqrt{h}}{\sqrt{n^3 h}}\right) = O_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{A.27}$$

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.21) and scaling the estimation error by $\sqrt{n}$ yields the following when $nh \to c \in [0, \infty)$

$$\begin{aligned} \sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) &= \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} \\ &+ \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' \eta_t + \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t u_t K_{tz} \\ &= O_p\left(\sqrt{n} \times h^2 \sqrt{nh}\right) + O_p\left(\sqrt{n} \times h\right) + \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{\sqrt{n}}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t u_t K_{tz} \\ &= o_p(1) + \left(\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{n^2 h}} \sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1), \end{aligned} \tag{A.28}$$

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

Proof of Theorem 2.3

Part (a) The analysis follows the proof of Theorem 2.1, with changes to accommodate multivariate z_t. We outline the key elements of the derivation here. By standard theory, e.g., Li and Racine (2007), we have $\mathbb{E}K_{tzq} = h^q f(z) + o(h^q)$ and $\mathbb{E}K_{tzq}^2 = h^q f(z) \nu_0^q(k) + o(h^q)$. Then, setting $\zeta_{tKq} = K_{tzq} - \mathbb{E}K_{tzq}$, we have $\mathbb{V}ar(\zeta_{tKq}) = O(h^q)$ and $\zeta_{tKq} = O_p(\sqrt{h^q})$. Next, similar to the calculations in Lemma B.1(b) we find for $q \geq 1$

$$\mathbb{E}\xi_{\beta tq} = \mathbb{E}[\beta(z_t) - \beta(z)]K_{tzq} = \int [\beta(z_t) - \beta(z)] K_q(z_t) f(z_t) dz_t$$

$$= h^q \int [\beta(z+hs) - \beta(z)] K_q(s) f(z+sh) ds$$

$$= h^{q+2} \int \sum_{j=1}^{q} \left(\left[\beta_j^{(1)}(z) f_j^{(1)}(z) + \frac{1}{2} \beta_{jj}^{(2)}(z) f(z) \right] s_j^2 \right) K_q(s) ds + o(h^{q+2})$$

$$= h^{q+2} \sum_{j=1}^{q} \left[\beta_j^{(1)}(z) f_j^{(1)}(z) + \frac{1}{2} \beta_{jj}^{(2)}(z) f(z) \right] \mu_2^q(k) + o(h^{q+2})$$

$$= h^{q+2} D(z) \mu_2^q(k) + o(h^{q+2}), \tag{A.29}$$

where $D(z) = \sum_{j=1}^{q} \left[\beta_j^{(1)}(z) f_j^{(1)}(z) + \frac{1}{2} \beta_{jj}^{(2)}(z) f(z) \right]$ and $\mu_2^q(k) = \int \Pi_{j=1}^q s_j^2 k(s_j) ds_j = (\mu_2(k))^q$. In a similar fashion, $\mathbb{E}\xi_{\beta tq}\xi_{\beta tq}' = h^{q+2} \sum_{j=1}^{q} [\beta_j^{(1)}(z) \beta_j^{(1)}(z)'] f(z) \nu_2^q(k) + o(h^{q+2})$, where $\nu_2^q(k) = \int \Pi_{j=1}^q k(s_j)^2 s_j^2 ds_j = (\nu_2(k))^q$.

The multivariate z_t version of (A.1) is

$$\left(\sum_{t=1}^{n} x_t x_t' K_{tzq} \right) \left(\hat{\beta}(z) - \beta(z) \right) = \sum_{t=1}^{n} x_t x_t' [\beta(z_t) - \beta(z)] K_{tzq} + \sum_{t=1}^{n} x_t u_t K_{tzq}.$$

$$= \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta tq} + \sum_{t=1}^{n} x_t x_t' \eta_{tq} + \sum_{t=1}^{n} x_t u_t K_{tzq}, \tag{A.30}$$

where $\eta_{tq} = \xi_{\beta tq} - \mathbb{E}\xi_{\beta tq}$ for $q \geq 1$. As in the scalar z_t case, we analyze each of the components of (A.30) in turn. Starting with the signal matrix we have

$$\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} = \frac{1}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \mathbb{E}\left(\frac{K_{tzq}}{h^q} \right) + \frac{1}{\sqrt{nh^q}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}},$$

where $\zeta_{tKq} = O_p(\sqrt{h^q})$, from above. We deduce that when $nh^q \to \infty$

$$\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} = \frac{1}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \{f(z) + o(1)\} + \frac{1}{\sqrt{nh^q}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}} \rightsquigarrow \left(\int B_x B_x' \right) f(z), \tag{A.31}$$

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

$$\sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}} \rightsquigarrow \int B_x B_x' dB_{\zeta Kq}, \tag{A.32}$$

holds, just as shown in the scalar z_t case in Lemma B.1(d) with $q = 1$.

By contrast, when $nh^q \to c$ for some $c \in [0, \infty)$, we have in place of (A.31) and (A.32)

$$\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} = \frac{\sqrt{nh^q}}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{K_{tzq}}{h^q}$$

$$= \frac{\sqrt{nh^q}}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \{f(z) + o(1)\} + \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}}$$

$$\sim_a \sqrt{c} f(z) \int B_x B_x' + \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}} = O_p(1). \tag{A.33}$$

No invariance principle holds here because $\sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{h^q}}$ has zero mean and finite variance matrix asymptotically, just as in Lemma B.1(c)(ii) when $q = 1$, and $\frac{1}{\sqrt{nh^q}} \sum_{t=1}^{n} \zeta_{tKq} = O_p(1)$ but does not satisfy a central limit theorem, just as in Lemma B.1(a)(ii) when $q = 1$. Results analogous to those given in the scalar z_t case now apply. As witnessed in the scalar z_t case, no invariance principle applies, which complicates inference, even though $\hat{\beta}(z)$ may be consistent. This case is pursued in detail in the proof of part (b) below. We now proceed under the condition that $nh^q \to \infty$, as in the statement of Theorem 2.3 (a).

From (A.29) we have $\mathbb{E}\xi_{\beta tq} = h^{q+2}\mu_2^q(k)D(z) + o(h^{q+2})$ and so the first term in (A.30) is, upon normalization and as in Lemma B.1(b)(i),

$$\frac{1}{nh^{q+2}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \mathbb{E}\left(\xi_{\beta tq}\right) \rightsquigarrow \mu_2^q(k) \left(\int B_x B_x' \right) D(z), \tag{A.34}$$

with $D(z) = \sum_{j=1}^{q} \left[\beta_j^{(1)}(z) f_j^{(1)}(z) + \frac{1}{2}\beta_{jj}^{(2)}(z) f(z) \right]$. The second term of (A.30) is, upon normalization and with derivations mirroring those in the proof of Lemma B.1(d)(i),

$$\frac{1}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} = \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\eta_{tq}}{\sqrt{nh^{q+2}}} \rightsquigarrow \int B_x B_x' dB_{\eta q}, \tag{A.35}$$

where $B_{\eta q}$ is vector Brownian motion with variance matrix $V_{\eta \eta q} = \nu_2^q(k) f(z) \sum_{j=1}^{q} [\beta_j^{(1)}(z) \beta_j^{(1)}(z)']$.

The final term of (A.30) is, upon normalization and using the analogue for $q \geq 1$ of Lemma B.1(a)(i),

$$\frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq} = \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{u_t K_{tzq}}{\sqrt{nh^q}} \rightsquigarrow \int B_x dB_{uKq}, \tag{A.36}$$

where B_{uKq} is the limit Brownian motion of $\frac{1}{\sqrt{nh^q}} \sum_{t=1}^{\lfloor n \cdot \rfloor} u_t K_{tzq}$ with variance $\sigma_u^2 f(z) \nu_0^q(k)$. Standardizing by the weighted signal matrix and recentering (A.30) we have

$$\begin{aligned} &\hat{\beta}(z) - \beta(z) - \left(\sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta tq} \\ &= \left(\sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \sum_{t=1}^{n} x_t x_t' \eta_{tq} + \left(\sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \sum_{t=1}^{n} x_t u_t K_{tzq}, \end{aligned} \tag{A.37}$$

which we rewrite as

$$\begin{aligned} &\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n^2 h^{q+2}} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta tq} \\ =& \sqrt{\frac{1}{nh^{q-2}}} \left(\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} + \frac{1}{n\sqrt{h^q}} \left(\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq}. \end{aligned} \tag{A.38}$$

Rescaling (A.38) by $\sqrt{nh^{q-2}}$ gives

$$\sqrt{nh^{q-2}}\left(\hat{\beta}(z)-\beta(z)-h^2\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{n^2h^{q+2}}\sum_{t=1}^n x_tx_t'\mathbb{E}\xi_{\beta tq}\right)$$
$$=\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^n x_tx_t'\eta_{tq}+\frac{\sqrt{nh^{q-2}}}{n\sqrt{h^q}}\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{n\sqrt{h^q}}\sum_{t=1}^n x_tu_tK_{tzq}$$
$$=\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^n x_tx_t'\eta_{tq}+\frac{1}{\sqrt{nh^2}}\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{n\sqrt{h^q}}\sum_{t=1}^n x_tu_tK_{tzq}$$
$$=\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^n x_tx_t'\eta_{tq}+o_p(1), \tag{A.39}$$

because $nh^2\to\infty$ which is implied by $nh^q\to\infty$ with $q\geq 2$. Combining (A.39) with (A.35) and (A.31), we obtain

$$\sqrt{nh^{q-2}}\left(\hat{\beta}(z)-\beta(z)-h^2\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{n^2h^{q+2}}\sum_{t=1}^n x_tx_t'\mathbb{E}\xi_{\beta tq}\right)$$
$$=\left(\frac{1}{n^2h^q}\sum_{t=1}^n x_tx_t'K_{tzq}\right)^{-1}\frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^n x_tx_t'\eta_{tq}+o_p(1)$$
$$\rightsquigarrow\left(f(z)\int B_xB_x'\right)^{-1}\left(\int B_xB_x'dB_{\eta q}\right)$$
$$\equiv_d \mathcal{MN}\left(0,\ \frac{\nu_2^q(k)}{f(z)}\left(\int B_xB_x'\right)^{-1}\int B_xB_x'\left(B_x'\beta^{(1)}(z)\right)^2\left(\int B_xB_x'\right)^{-1}\right), \tag{A.40}$$

where $B_{\eta q}$ is Brownian motion with variance matrix $V_{\eta\eta q}$ and $\beta_j^{(1)}(z)=\partial\beta(z)/\partial z_j$, as above. To verify the representation in the final line of (A.40) we note that conditional on B_x the process $B_x'B_{\eta q}$ is scalar Brownian motion with variance $B_x'V_{\eta\eta q}B_x=\nu_2^q(k)f(z)\sum_{j=1}^q[B_x'\beta_j^{(1)}(z)]^2$. Thus, in view of the independence of B_x and $B_{\eta q}$, we have the following mixed normal representation

$$\int B_xB_x'dB_{\eta q}\equiv_d \mathcal{MN}\left(0,\ \nu_2^q(k)f(z)\int B_xB_x'\sum_{j=1}^q\left(B_x'\beta_j^{(1)}(z)\right)^2\right). \tag{A.41}$$

Combining (A.40) with (A.31) and (A.34) for explicit representation of the bias component, we have

$$\sqrt{nh^{q-2}}\left(\hat{\beta}(z)-\beta(z)-h^2\frac{\nu_2^q(k)}{f(z)}D(z)\right)\rightsquigarrow\left(f(z)\int B_xB_x'\right)^{-1}\left(\int B_xB_x'dB_{\eta q}\right)$$
$$\equiv_d \mathcal{MN}\left(0,\frac{\nu_2^q(k)}{f(z)}\left(\int B_xB_x'\right)^{-1}\int B_xB_x'\sum_{j=1}^q\left(B_x'\beta_j^{(1)}(z)\right)^2\left(\int B_xB_x'\right)^{-1}\right), \tag{A.42}$$

which leads to the limit theory (2.17) and the mixed normal representation (2.18), proving the stated result. The limit theory (A.42) requires $nh^q\to\infty$ which, as indicated earlier, is needed to establish weak convergence to the stochastic integral (A.35) that plays a key role in determining the limit (2.17). The effect of relaxing this condition is considered next. ■

We take the case where $nh^q \to c \in [0, \infty)$ and start the analysis with some preliminaries. As noted above in the discussion of (A.33)

$$\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \sim_a \sqrt{c} f(z) \int B_x B_x' + \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{nh^q}} = O_p(1), \tag{A.43}$$

with no invariance principle holding because $nh^q \not\to \infty$. Next, in place of (A.35), (A.32), and (A.36) we have

$$\frac{1}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\eta_{tq}}{\sqrt{h^{q+2}}} = O_p(1), \tag{A.44}$$

$$\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' \zeta_{tKq} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\zeta_{tKq}}{\sqrt{h^q}} = O_p(1), \tag{A.45}$$

$$\frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{u_t K_{tzq}}{\sqrt{h^q}} = O_p(1), \tag{A.46}$$

and again no invariance principles hold, as established in Lemma B.1(a)(ii),(b)(ii), and (d)(ii) for the case $q = 1$. It follows that in place of the decomposition leading to (A.38) we now have

$$\hat{\beta}(z) - \beta(z) - h^2 \left(\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n^2 h^{q+2}} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta tq} \tag{A.47}$$

$$= \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} \times \frac{\sqrt{n^3 h^{q+2}}}{\sqrt{n^3 h^q}}$$

$$+ \frac{n\sqrt{h^q}}{\sqrt{n^3 h^q}} \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq}$$

$$= h \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{\sqrt{n^3 h^{q+2}}} \sum_{t=1}^{n} x_t x_t' \eta_{tq} + \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n\sqrt{h^q}} \sum_{t=1}^{n} x_t u_t K_{tzq} \tag{A.48}$$

$$= O_p(h) + O_p(1/\sqrt{n}).$$

Accordingly, the asymptotics rely on one or other or both of the terms in (A.48). For the bias term in (A.47) we observe that

$$h^2 \left(\frac{1}{n^2 h^q} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n^2 h^{q+2}} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta tq}$$

$$= h^2 \frac{n^2 h^q}{\sqrt{n^3 h^q}} \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\mathbb{E}\xi_{\beta tq}}{h^{q+2}}$$

$$= h^2 \sqrt{nh^q} \left(\frac{1}{\sqrt{n^3 h^q}} \sum_{t=1}^{n} x_t x_t' K_{tzq} \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\mathbb{E}\xi_{\beta tq}}{h^{q+2}}$$

$$= O_p(h^2\sqrt{nh^q}).$$

Part (b) In this case we have $nh^q \to 0$.

(1) The $q = 1$ case is covered in Theorem 2.2.

(2) When $q = 2$, evidently $nh^2 \to 0$ and $h = o(1/\sqrt{n})$ so that the second term in (A.48) dominates. We then have

$$\sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) \sim_a \left(\frac{1}{\sqrt{n^3h^q}}\sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \frac{1}{n\sqrt{h^q}}\sum_{t=1}^{n} x_t u_t K_{tzq} = O_p(1). \tag{A.49}$$

In this case, the bias is neglected since $\sqrt{n}h^2\sqrt{nh^q} \to 0$.

(3) When $q > 2$, we deduce three cases for nh^2:

(i) $nh^2 \to 0$. In this situation (A.49) continues to hold.

(ii) $nh^2 \to c \in (0, \infty)$. In this situation both terms of (A.48) contribute to the asymptotics. We have

$$\sqrt{n}\left(\hat{\beta}(z) - \beta(z)\right) \sim_a \left(\frac{1}{\sqrt{n^3h^q}}\sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \left(c^{1/2}\frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^{n} x_t x_t' \eta_{tq} + \frac{1}{n\sqrt{h^q}}\sum_{t=1}^{n} x_t u_t K_{tzq}\right) = O_p(1). \tag{A.50}$$

The bias term is neglected because $\sqrt{n}h^2\sqrt{nh^q} \to 0$ and the $O_p(1)$ order of $\frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^{n} x_t x_t' \eta_{tq} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\eta_{tq}}{\sqrt{h^{q+2}}}$ follows as in Lemma B.1 (c)(ii) and (d)(ii).

(iii) $nh^2 \to \infty$. In this situation, the first term in (A.48) dominates. We have

$$\frac{1}{h}\left(\hat{\beta}(z) - \beta(z)\right) \sim_a \left(\frac{1}{\sqrt{n^3h^q}}\sum_{t=1}^{n} x_t x_t' K_{tzq}\right)^{-1} \frac{1}{\sqrt{n^3h^{q+2}}}\sum_{t=1}^{n} x_t x_t' \eta_{tq} = O_p(1). \tag{A.51}$$

The bias term is neglected because $\frac{1}{h}h^2\sqrt{nh^q} \to 0$.

Part (c) In this case we have $nh^q \to c \in (0, \infty)$.

(1) When $q = 1$ the case is covered in Theorem 2.2.

(2) When $q = 2$ we have $nh^2 \to c \in (0, \infty)$. In this situation, (A.50) continues to hold.

(3) When $q > 2$ then $nh^2 \to \infty$ and so (A.51) holds. ■

Proof of Theorem 3.1

We consider the nonstationary and stationary regressor cases in turn. The statistic studied in all cases is $\hat{T}(z) = V_n(z)^{-1/2}\left(\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z)\right)$ where

$$V_n(z) = A_{nz}^{-1}\left\{\hat{\sigma}_u^2\nu_0(K)\sum_{t=1}^{n} x_t x_t' K_{tz} + \sum_{t=1}^{n} x_t x_t'\left[x_t'\hat{\beta}^{(1)}(z)\,(z_t - z)\,K_{tz}\right]^2\right\} A_{nz}^{-1}, \tag{A.52}$$

with $A_{nz} = \sum_{t=1}^{n} x_t x_t' K_{tz}$. The estimates $\hat{\sigma}_u^2$, $\hat{\beta}^{(1)}(z)$ and $\hat{\mathcal{B}}(z)$ are obtained in the usual manner, as discussed in Section 3 of the online supplement.

Nonstationary $\mathbf{x_t}$ We establish the result for the three bandwidth ranges given in Theorem 2.1 in turn. The derivations make extensive use of the results in Lemma B.1.

Case (a) Here $nh^2 \to 0$. From Lemma B.1(c) we have

$$\frac{1}{n^2 h}\sum_{t=1}^{n} x_t x_t' K_{tz} \rightsquigarrow f(z)\int B_x B_x', \tag{A.53}$$

and by consistency of the derivative kernel estimate $\hat{\beta}^{(1)}(z) \to_p \beta^{(1)}(z)$ as $n \to \infty$ it follows that

$$\begin{aligned}
&\frac{1}{n^3h^3}\sum_{t=1}^{n} x_t x_t' \left[x_t'\hat{\beta}^{(1)}(z)(z_t - z)K_{tz}\right]^2 \\
= \; & \frac{1}{nh}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left[\frac{x_t'}{\sqrt{n}}\hat{\beta}^{(1)}(z)\right]^2\left(\frac{z_t-z}{h}\right)^2 K\left(\frac{z_t-z}{h}\right)^2 \\
= \; & \frac{1}{nh}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left[\frac{x_t'}{\sqrt{n}}\beta^{(1)}(z)\right]^2 \mathbb{E}\left\{\left(\frac{z_t-z}{h}\right)^2 K\left(\frac{z_t-z}{h}\right)^2\right\}\{1+o_p(1)\} \\
= \; & f(z)\nu_2(K)\frac{1}{n}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left[\frac{x_t'}{\sqrt{n}}\beta^{(1)}(z)\right]^2\{1+o_p(1)\} \\
\rightsquigarrow \; & f(z)\nu_2(K)\int B_x B_x'\left[B_x'\beta^{(1)}(z)\right]^2,
\end{aligned} \tag{A.54}$$

using standard weak convergence methods. Since $\hat{\sigma}_u^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2 \to_p \sigma_u^2$ and in view of (A.53) and (A.54) we deduce that when $nh^2 \to 0$

$$\begin{aligned}
n^2hV_n(z) \; &= \; \left(\frac{A_{nz}}{n^2h}\right)^{-1}\left\{\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^{n}x_tx_t'K_{tz}}{n^2h} + \frac{nh^2}{n^3h^3}\sum_{t=1}^{n}x_tx_t'\left[x_t'\hat{\beta}^{(1)}(z)(z_t-z)K_{tz}\right]^2\right\}\left(\frac{A_{nz}}{n^2h}\right)^{-1} \\
&= \; \left(\frac{A_{nz}}{n^2h}\right)^{-1}\left\{\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^{n}x_tx_t'K_{tz}}{n^2h} + O_p\left(nh^2\right)\right\}\left(\frac{A_{nz}}{n^2h}\right)^{-1} \\
&\rightsquigarrow \; \left(f(z)\int B_xB_x'\right)^{-1}\left\{\sigma_u^2\nu_0(K)f(z)\int B_xB_x'\right\}\left(f(z)\int B_xB_x'\right)^{-1} \\
&= \; \frac{\sigma_u^2\nu_0(K)}{f(z)}\left(\int B_xB_x'\right)^{-1} = \Omega_u(z).
\end{aligned} \tag{A.55}$$

The required result in this case now follows directly from Theorem 2.1(a) because

$$\begin{aligned}
\hat{T}(z) &= \left(n^2hV_n(z)\right)^{-1/2}\left\{n\sqrt{h}\left[\hat{\beta}(z)-\beta(z)-h^2\hat{\mathcal{B}}(z)\right]\right\} \\
&\sim_a \Omega_u(z)^{-1/2}\left\{n\sqrt{h}\left[\hat{\beta}(z)-\beta(z)-h^2\mathcal{B}(z)\right]\right\} \rightsquigarrow \mathcal{N}(0, I_d).
\end{aligned} \tag{A.56}$$

Case (b) Here $nh^2 \to \infty$ and in place of (A.55) we have, using (A.53) and (A.54),

$$\begin{aligned}
\frac{n}{h}V_n(z) \; &= \; \left(\frac{A_{nz}}{n^2h}\right)^{-1}\left\{\frac{n}{h}\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^{n}x_tx_t'K_{tz}}{n^4h^2} + \frac{n}{h}\frac{1}{n^4h^2}\sum_{t=1}^{n}x_tx_t'\left[x_t'\hat{\beta}^{(1)}(z)(z_t-z)K_{tz}\right]^2\right\}\left(\frac{A_{nz}}{n^2h}\right)^{-1} \\
&= \; \left(\frac{A_{nz}}{n^2h}\right)^{-1}\left\{\frac{1}{nh^2}\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^{n}x_tx_t'K_{tz}}{n^2h} + \frac{1}{n^3h^3}\sum_{t=1}^{n}x_tx_t'\left[x_t'\hat{\beta}^{(1)}(z)(z_t-z)K_{tz}\right]^2\right\}\left(\frac{A_{nz}}{n^2h}\right)^{-1}
\end{aligned}$$

$$\begin{aligned}
&\rightsquigarrow \left(f(z)\int B_x B_x'\right)^{-1}\left\{f(z)\nu_2(K)\int B_x B_x'\left[B_x'\beta^{(1)}(z)\right]^2\right\}\left(f(z)\int B_x B_x'\right)^{-1} \\
&= \frac{\nu_2(K)}{f(z)}\left(\int B_x B_x'\right)^{-1}\left\{\int B_x B_x'\left[B_x'\beta^{(1)}(z)\right]^2\right\}\left(\int B_x B_x'\right)^{-1} = \Omega_\beta(z). \qquad \text{(A.57)}
\end{aligned}$$

The required result in this case now follows from Theorem 2.1(b) because

$$\begin{aligned}
\hat{T}(z) &= \left(\frac{n}{h}V_n(z)\right)^{-1/2}\left\{\sqrt{\frac{n}{h}}\left[\hat{\beta}(z)-\beta(z)-h^2\hat{\mathcal{B}}(z)\right]\right\} \\
&\sim_a \Omega_\beta(z)^{-1/2}\left\{\sqrt{\frac{n}{h}}\left[\hat{\beta}(z)-\beta(z)-h^2\mathcal{B}(z)\right]\right\} \rightsquigarrow \mathcal{N}(0, I_d), \qquad \text{(A.58)}
\end{aligned}$$

when $nh^2 \to \infty$.

Case (c) Here $nh^2 \to c \in (0,\infty)$ and we have

$$\begin{aligned}
n^{3/2}V_n(z) &= \left(\frac{A_{nz}}{n^2h}\right)^{-1}\left\{n^{3/2}\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^n x_t x_t' K_{tz}}{n^4h^2} + n^{3/2}\frac{1}{n^4h^2}\sum_{t=1}^n x_t x_t'\left[x_t'\hat{\beta}^{(1)}(z)(z_t - z)K_{tz}\right]^2\right\}\left(\frac{A_{nz}}{n^2h}\right)^{-1} \\
&= \left(\frac{A_{nz}}{n^2h}\right)^{-1}\left\{\frac{1}{\sqrt{nh^2}}\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^n x_t x_t' K_{tz}}{n^2h} + \frac{\sqrt{nh^2}}{n^3h^3}\sum_{t=1}^n x_t x_t'\left[x_t'\hat{\beta}^{(1)}(z)(z_t - z)K_{tz}\right]^2\right\}\left(\frac{A_{nz}}{n^2h}\right)^{-1} \\
&\rightsquigarrow \left(f(z)\int B_x B_x'\right)^{-1}\left\{\frac{\sigma_u^2\nu_0(K)f(z)}{\sqrt{c}}\int B_x B_x' + \sqrt{c}\nu_2(K)f(z)\int B_x B_x'\left[B_x'\beta^{(1)}(z)\right]^2\right\}\left(f(z)\int B_x B_x'\right)^{-1} \\
&= \frac{1}{f(z)}\left(\int B_x B_x'\right)^{-1}\left\{\frac{\sigma_u^2\nu_0(K)}{c^{1/2}}\int B_x B_x' + c^{1/2}\nu_2(K)\int B_x B_x'\left[B_x'\beta^{(1)}(z)\right]^2\right\}\left(\int B_x B_x'\right)^{-1} \\
&= \frac{1}{c^{1/2}}\Omega_u(z) + c^{1/2}\Omega_\beta(z). \qquad \text{(A.59)}
\end{aligned}$$

The required result in this case now follows from Theorem 2.1(c) because

$$\begin{aligned}
\hat{T}(z) &= \left(n^{3/2}V_n(z)\right)^{-1/2}\left\{n^{3/4}\left[\hat{\beta}(z)-\beta(z)-h^2\hat{\mathcal{B}}(z)\right]\right\} \\
&\sim_a \left\{\frac{1}{c^{1/2}}\Omega_u(z) + c^{1/2}\Omega_\beta(z)^{-1/2}\right\}^{-1/2}\left\{n^{3/4}\left[\hat{\beta}(z)-\beta(z)-h^2\mathcal{B}(z)\right]\right\} \rightsquigarrow \mathcal{N}(0, I_d), \\
&\qquad \text{(A.60)}
\end{aligned}$$

when $nh^2 \to c \in (0,\infty)$. Thus, the statistic $\hat{T}(z)$ has robust standard $\mathcal{N}(0, I_d)$ asymptotics for all bandwidth choices covering cases (a), (b) and (c), thereby including the rate efficient case where the bandwidth $h = O\left(n^{-1/2}\right)$ and convergence rate is $O(n^{3/4})$.

Stationary $\mathbf{x_t}$ We start with the decomposition

$$A_{nz}[\hat{\beta}(z)-\beta(z)] = \sum_t x_t x_t'[\beta(z_t)-\beta(z)]K_{tz} + \sum_t x_t u_t K_{tz}, \qquad \text{(A.61)}$$

or, equivalently,

$$A_{nz}\left\{\hat{\beta}(z)-\beta(z)-A_{nz}^{-1}\sum_t \mathbb{E}[x_t x_t'[\beta(z_t)-\beta(z)]K_{tz}|z_t = z]\right\}$$

$$= \sum_t \left\{ x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} - \mathbb{E}[x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} | z_t = z] \right\} + \sum_t x_t u_t K_{tz}. \tag{A.62}$$

From Lemma B.2 we have

$$A_{nz} = \sum_t x_t x_t' K_{tz} \sim_a nh \mathbb{E}[x_t x_t' | z_t = z] f(z) =: nh \mathcal{M}_z, \tag{A.63}$$

$$A_{nz}^{-1} \sum_t \mathbb{E}[x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} | z_t = z] \sim_a h^2 \mu_2(K) \mathcal{M}_z^{-1} C_S(z) =: h^2 \mathcal{B}_S(z), \tag{A.64}$$

$$\frac{1}{\sqrt{nh^3}} \sum_{t=1}^n \{x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} - \mathbb{E}\{x_t x_t' [\beta(z_t) - \beta(z)] K_{tz}\}\} \rightsquigarrow \mathcal{N}\left(0, \nu_2(K) f(z) \mathbb{E}\{x_t x_t' [x_t' \beta^{(1)}(z)]^2 | z_t = z\}\right), \tag{A.65}$$

$$\frac{1}{\sqrt{nh}} \sum_t x_t u_t K_{tz} \rightsquigarrow \mathcal{N}(0, \nu_0(K) \sigma_u^2 f(z) \mathbb{E}[x_t x_t' | z_t = z]), \tag{A.66}$$

with $\mathcal{M}_z = f(z)\mathbb{E}[x_t x_t' | z_t = z]$, $\mathcal{B}_{\mathcal{S}}(z) = \mu_2(K)\mathcal{M}_z^{-1} C_S(z)$ and

$$C_S(z) = \mathbb{E}\left\{ x_t x_t' \left[\beta^{(1)}(z) \frac{f_z(x_t, z)}{f(x_t | z_t = z)} + \frac{1}{2} \beta^{(2)}(z) f(z) \right] \middle| z_t = z \right\}, \tag{A.67}$$

where $f_z(s, z) = \partial f(s, z) / \partial z$, $f(x_t, z_t)$ is the joint density of (x_t, z_t) at (s, z) and $f(x_t | z_t = z)$ is the conditional density of x_t given $z_t = z$. Importantly, as shown in Lemma B.2, when x_t and the covariate z_t are independent we have

$$C_S(z) = \mathbb{E}\{x_t x_t'\} \left[\beta^{(1)}(z) f^{(1)}(z) + \frac{1}{2} \beta^{(2)}(z) f(z) \right], \ \mathcal{M}_z = f(z) \mathbb{E}[x_t x_t'], \tag{A.68}$$

so that

$$h^2 \mathcal{B}_S(z) = h^2 \mu_2(K) \mathcal{M}_z^{-1} C_S(z) = h^2 \frac{\mu_2(K)}{f(z)} \left[\beta^{(1)}(z) f^{(1)}(z) + \frac{1}{2} \beta^{(2)}(z) f(z) \right] = h^2 \mathcal{B}(z), \tag{A.69}$$

which reproduces the standard deterministic bias function in the stationary case (1.9).

Using these results in (A.62) gives

$$A_{nz}[\hat{\beta}(z) - \beta(z) - h^2 \mathcal{B}_S(z)] \sim_a \sqrt{nh^3} \mathcal{N}\left(0, f(z) \nu_2(K) \mathbb{E}\{x_t x_t' [x_t' \beta^{(1)}(z)]^2 | z_t = z\}\right) + \sqrt{nh} \mathcal{N}\left(0, \nu_0(K) \sigma_u^2 f(z) \mathbb{E}[x_t x_t' | z_t = z]\right). \tag{A.70}$$

Since u_t is a martingale difference with respect to the filtration $\mathcal{F}_t = \sigma\{\{x_s\}_{s=1}^\infty; \{z_s\}_{s=1}^\infty; u_t, u_{t-1}, \cdots\}$, the covariance between $\frac{1}{\sqrt{nh^3}} \sum_t \{x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} - \mathbb{E} x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} | z_t = z]\}$ and $\frac{1}{\sqrt{nh}} \sum_t x_t u_t K_{tz}$ is zero. So the two normal distributions on the right side of (A.70) are independent. Hence, as $nh \to \infty$,

$$\begin{aligned} & A_{nz}[\hat{\beta}(z) - \beta(z) - h^2 \mathcal{B}_S(z)] \\ & \sim_a \mathcal{N}\left(0, nh^3 f(z) \nu_2(K) \mathbb{E}[x_t x_t' (x_t' \beta^{(1)}(z))^2 | z_t = z] + nh \nu_0(K) \sigma_u^2 f(z) \mathbb{E}[x_t x_t' | z_t = z]\right) \\ & = \sqrt{nh} \mathcal{N}\left(0, h^2 f(z) \nu_2(K) \mathbb{E}[x_t x_t' (x_t' \beta^{(1)}(z))^2 | z_t = z] + \nu_0(K) \sigma_u^2 f(z) \mathbb{E}[x_t x_t' | z_t = z]\right). \end{aligned} \tag{A.71}$$

With $h \to 0$ and in view of (A.63), we have

$$\begin{aligned}\sqrt{nh}[\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}_S(z)] &\sim_a \left(\frac{A_{nz}}{nh}\right)^{-1} \mathcal{N}\left(0, \nu_0(K)\sigma_u^2 f(z)\mathbb{E}[x_t x_t' | z_t = z]\right), \\ &\rightsquigarrow \left(f(z)\mathbb{E}[x_t x_t'|z_t = z]\right)^{-1} \mathcal{N}\left(0, \nu_0(K)\sigma_u^2 f(z)\mathbb{E}[x_t x_t'|z_t = z]\right) \\ &= \mathcal{N}\left(0, \frac{\nu_0(K)\sigma_u^2}{f(z)}\left(\mathbb{E}[x_t x_t'|z_t = z]\right)^{-1}\right) =: \mathcal{N}\left(0, \Omega_S(z)\right).\end{aligned} \tag{A.72}$$

The components $\sigma_u^2 f(z)\mathbb{E}[x_t x_t'|z_t = z]$ and $f(z)\nu_2(K)\mathbb{E}[x_t x_t'(x_t'\beta^{(1)}(z))^2|z_t = z]$ in (A.71) can be consistently estimated by $\hat{\sigma}_u^2 \frac{1}{nh}\sum_t x_t x_t' K_{tz}$ and $\frac{1}{nh^3}\sum_t x_t x_t'(x_t'\beta^{(1)}(z))^2(z_t - z)^2 K_{tz}^2$, respectively. So $\frac{h^2}{nh^3}\sum_t x_t (x_t'\beta^{(1)}(z))^2 x_t'(z_t - z)^2 K_{tz}^2 + \nu_0(K)\hat{\sigma}_u^2 \frac{1}{nh}\sum_t x_t x_t' K_{tz}$ is a suitable estimate of the variance matrix in (A.71) for self normalization. In particular, we have

$$\begin{aligned}nhV_n(z) &= \left(\frac{A_{nz}}{nh}\right)^{-1}\left\{\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^n x_t x_t' K_{tz}}{nh} + \frac{h^2}{nh^3}\sum_{t=1}^n x_t x_t' \left[x_t'\hat{\beta}^{(1)}(z)\left(z_t - z\right)K_{tz}\right]^2\right\}\left(\frac{A_{nz}}{nh}\right)^{-1} \\ &= \left(\frac{A_{nz}}{nh}\right)^{-1}\left\{\hat{\sigma}_u^2\nu_0(K)\frac{\sum_{t=1}^n x_t x_t' K_{tz}}{nh} + O_p\left(h^2\right)\right\}\left(\frac{A_{nz}}{nh}\right)^{-1} \\ &\rightsquigarrow \left(f\left(z\right)\mathbb{E}[x_t x_t'|z_t = z]\right)^{-1}\left\{\nu_0(K)\sigma_u^2 f\left(z\right)\mathbb{E}[x_t x_t'|z_t = z]\right\}\left(f\left(z\right)\mathbb{E}[x_t x_t'|z_t = z]\right)^{-1} \\ &= \frac{\nu_0(K)\sigma_u^2}{f\left(z\right)}\left(\mathbb{E}[x_t x_t'|z_t = z]\right)^{-1} = \Omega_S(z).\end{aligned} \tag{A.73}$$

Using a consistent estimate $\hat{\mathcal{B}}_S(z)$ of $\mathcal{B}_S(z)$, it then follows that the t-ratio defined in (3.10) has a standard normal $\mathcal{N}(0, I_d)$ distribution in the limit. Thus,

$$\begin{aligned}\hat{T}(z) &= \left(nhV_n(z)\right)^{-1/2}\left\{\sqrt{nh}\left[\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}_S(z)\right]\right\} \\ &\sim_a \Omega_S(z)^{-1/2}\left\{\sqrt{nh}\left[\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}_S(z)\right]\right\} \rightsquigarrow \mathcal{N}\left(0, I_d\right).\end{aligned} \tag{A.74}$$

When x_t and z_t are mutually independent as in Assumption 2 (iii), in place of (A.63) we have $\frac{1}{nh}A_{nz} \to_p \mathbb{E}(x_t x_t' K_{tz}) = \Sigma_{xx} f(z)$ with $\Sigma_{xx} > 0$ and $\Omega_S(z) = \frac{\nu_0(K)\sigma_u^2}{f(z)}\Sigma_{xx}^{-1}$. Then $\mathcal{B}_S(z)$ may be replaced by $\mathcal{B}(z)$ as in (A.69) and we have the specialization

$$\begin{aligned}\hat{T}(z) &= \left(nhV_n(z)\right)^{-1/2}\left\{\sqrt{nh}\left[\hat{\beta}(z) - \beta(z) - h^2\hat{\mathcal{B}}(z)\right]\right\} \\ &\sim_a \Omega_S(z)^{-1/2}\left\{\sqrt{nh}\left[\hat{\beta}(z) - \beta(z) - h^2\mathcal{B}(z)\right]\right\} \rightsquigarrow \mathcal{N}\left(0, I_d\right),\end{aligned} \tag{A.75}$$

with the same limit theory as in the nonstationary regressor case. This correspondence ensures that the robust test statistics $\hat{T}(z)$ and $\hat{T}_2(z)$ have the same limiting form in both stationary and nonstationary cases under comparable conditions.

$\hat{T}_2(z)$ asymptotics

The limit distribution $\hat{T}_2(z) \rightsquigarrow \chi_d^2$ now follows directly from the limit theory $\hat{T}(z) \rightsquigarrow \mathcal{N}(0, I_d)$ given above in both stationary and nonstationary cases. ■

B Useful Lemmas

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 $nh \to c \in [0,\infty)$, then $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot \rfloor}\eta_t = O_p(1)$, but no invariance principle holds.

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

(d) (i) If $nh \to \infty$, $\frac{1}{\sqrt{n^3 h}}\sum_{t=1}^{n} x_t x_t' \zeta_{tK} \rightsquigarrow \int B_x B_x' dB_{\zeta K}$, and $\frac{1}{\sqrt{n^3 h^3}}\sum_{t=1}^{n} x_t x_t' \eta_t \rightsquigarrow \int B_x B_x' dB_\eta \equiv_d \mathcal{MN}\left(0, \int B_x(r)B_x(r)'\left(B_x(r)'\beta^{(1)}(z)\right)^2\right)$;
(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{\sqrt{n^3 h}}\sum_{t=1}^{n} x_t x_t' \zeta_{tK} = O_p(1)$ but no invariance principle holds, and $\frac{1}{\sqrt{n^3 h^3}}\sum_{t=1}^{n} x_t x_t' \eta_t = O_p(1)$ but no invariance principle holds;

(e) (i) If $nh \to \infty$, $\frac{1}{n\sqrt{h}}\sum_{t=1}^{n} x_t u_t K_{tz} \rightsquigarrow \int B_x dB_{uK}$;
(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{n\sqrt{h}}\sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1)$ but no invariance principle holds.

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

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

Proof of Lemma B.1

Part (a) (i) The joint limit result stated for $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot \rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot \rfloor}u_t K_{tz}\}$ 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, $\mathbb{V}ar(u_t K_{tz}) = h\nu_0(K)\sigma_u^2 f(z) + o(h)$ and $\mathbb{E}(u_t K_{tz} K_{sz}) = 0$ for all t and s. So the standardized partial sums processes $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot \rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot \rfloor}u_t K_{tz}\}$ 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

$(B_{\zeta K}(r), B_{uK}(r))$ are independent with respective variances $\nu_0(K)f(z)$ and $\nu_0(K)\sigma_u^2 f(z)$. 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, $\left(\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_t K_{tz}\right) = O_p(1)$ 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

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\{\left(\frac{\zeta_{tK}}{\sqrt{h}}\right)^2 1_{\left[|\zeta_{tK}| > \epsilon\sqrt{nh}\right]}\right\} = \int \frac{[K(\frac{z_t - z}{h}) + O(h)]^2}{h} 1_{\left[\left|K(\frac{z_t - z}{h}) + O(h)\right| > \epsilon\sqrt{nh}\right]} f(z_t)dz_t$$

$$= \int (K(p) + O(h))^2\, 1_{\left[|K(p)+O(h)| > \epsilon\sqrt{nh}\right]} f(z + ph)dp$$

$$\to \begin{cases} f(z)\nu_0(K) > 0 & \text{if } nh \to 0 \\ \int K^2(p) 1_{\left[|K(p)| > \epsilon\sqrt{c}\right]} dp f(z) > 0 & \text{if } nh \to c \in (0, \infty) \end{cases}.$$

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

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

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

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

$$\begin{aligned}
\mathbb{E}\xi_{\beta t}\xi_{\beta t}' &= \mathbb{E}[(\beta(z_t) - \beta(z))(\beta(z_t) - \beta(z))'K^2(\frac{z_t - z}{h})] \\
&= h\int_{-1}^{1}(\beta(z + hs) - \beta(z))(\beta(z + hs) - \beta(z))'K^2(s)f(z + hs)ds \\
&= h^3[\beta^{(1)}(z)\beta^{(1)}(z)']f(z)\int_{-1}^{1}s^2K^2(s)ds + o(h^3) \\
&= h^3[\beta^{(1)}(z)\beta^{(1)}(z)']f(z)\nu_2(K) + o(h^3).
\end{aligned} \tag{B.2}$$

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

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

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

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

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 (z_t, z_{t+j}) we have

$$\begin{aligned}
\mathbb{E}\xi_{\beta t}\xi'_{\beta t+j} &= \mathbb{E}[(\beta(z_t) - \beta(z))(\beta(z_{t+j}) - \beta(z))' K_{tz}K_{t+j,z}] \\
&= \int\int(\beta(s_0) - \beta(z))(\beta(s_j) - \beta(z))' K\left(\frac{s_0 - z}{h}\right)K\left(\frac{s_j - z}{h}\right)f_{0,j}(s_0, s_j)ds_0ds_j \\
&= h^2\int\int(\beta(z + hp_0) - \beta(z))(\beta(z + hp_j) - \beta(z))'K(p_0)K(p_j)f_{0,j}(z + hp_0, z + hp_j)dp_0dp_j \\
&= h^4[\beta^{(1)}(z)][\beta^{(1)}(z)]'f_{0,j}(z, z)\int p_1K(p_1)dp_1\int p_2K(p_2)dp_2 \\
&\quad + h^5\left\{\frac{1}{2}\left([\beta^{(1)}(z)][\beta^{(2)}(z)]' + [\beta^{(2)}(z)][\beta^{(1)}(z)]'\right)f_{0,j}(z, z)\right. \\
&\quad \left. +[\beta^{(1)}(z)][\beta^{(1)}(z)]'\left[\frac{\partial f_{0,j}}{\partial s_0}(z, z) + \frac{\partial f_{0,j}}{\partial s_j}(z, z)\right]\right\}\int p_1K(p_1)dp_1\int p_2^2K(p_2)dp_2 \\
&\quad + h^6\left\{\frac{1}{4}[\beta^{(2)}(z)][\beta^{(2)}(z)]'f_{0,j}(z, z) + [\beta^{(1)}(z)][\beta^{(1)}(z)]'\frac{\partial^2 f_{0,j}}{\partial s_0\partial s_j}(z, z) + \frac{1}{2}[\beta^{(1)}(z)][\beta^{(2)}(z)]'\frac{\partial f_{0,j}}{\partial s_0}(z, z)\right. \\
&\quad \left. +\frac{1}{2}[\beta^{(2)}(z)][\beta^{(1)}(z)]'\frac{\partial f_{0,j}}{\partial s_j}(z, z)\right\}\int p_1^2K(p_1)dp_1\int p_2^2K(p_2)dp_2 + o\left(h^6\right) \\
&= h^6\left\{\frac{1}{4}[\beta^{(2)}(z)][\beta^{(2)}(z)]'f_{0,j}(z, z) + [\beta^{(1)}(z)][\beta^{(1)}(z)]'\frac{\partial^2 f_{0,j}}{\partial s_0\partial s_j}(z, z) + \frac{1}{2}[\beta^{(1)}(z)][\beta^{(2)}(z)]'\frac{\partial f_{0,j}}{\partial s_0}(z, z)\right. \\
&\quad \left. +\frac{1}{2}[\beta^{(2)}(z)][\beta^{(1)}(z)]'\frac{\partial f_{0,j}}{\partial s_j}(z, z)\right\}[\mu_2(K)]^2 + o(h^6).
\end{aligned} \tag{B.5}$$

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

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

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

$$\frac{1}{nh^3}\sum_{t\neq s}\mathbb{E}\eta_t\eta_s' = \frac{1}{h^3}\sum_{j=-n+1, j\neq 0}^{n-1}\left[1 - \frac{|j|}{n}\right][\mathbb{E}\xi_{\beta t}\xi'_{\beta t+j} - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t+j})']$$

$$= \frac{1}{h^3}\sum_{j=-M, j\neq 0}^{M}\left[1-\frac{|j|}{n}\right]\left[\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})'\right] + \frac{1}{h^3}\sum_{M<|j|<n}\left(1-\frac{|j|}{n}\right)\left[\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})'\right]$$

$$= O\left(\frac{Mh^6}{h^3}\right) + O\left(\frac{1}{h^3}\left(\mathbb{E}\left|\xi_{\beta t}\right|^{\delta}\right)^{2/\delta}\sum_{M<|j|<n}\alpha_j^{1-2/\delta}\right) = O\left(Mh^3\right) + O\left(\frac{h^{2\frac{1+\delta}{\delta}}}{h^3 M^a}\sum_{M<|j|<\infty} j^a \alpha_j^{1-2/\delta}\right)$$

$$= O\left(Mh^3\right) + O\left(\frac{h^{\frac{2}{\delta}}}{hM^a}\sum_{M<|j|<\infty} j^a \alpha_j^{1-2/\delta}\right) = O\left(Mh^3\right) + O\left(\frac{1}{h^{1-2/\delta}M^a}\sum_{M<|j|<\infty} j^a \alpha_j^{1-2/\delta}\right)$$

$$= O\left(Mh^3\right) + o\left(\frac{1}{(Mh)^{1-2/\delta}}\right) = o\left(1\right),$$

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

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

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

$$\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t = O_p(1), \tag{B.8}$$

but with no invariance principle applying. This result mirrors the finding in Part (a)(ii) for $\{\frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}u_t K_{tz}\}$. 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 $\mathbb{E}\xi_{\beta t} = h^3 C(z)\mu_2(K) + o(h^3)$ 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.8). 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

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\{\left(\frac{\eta_t}{\sqrt{h^3}}\right)^2 1_{\left[|\eta_t|>\epsilon\sqrt{nh^3}\right]}\right\} = \int \frac{[[\beta(z_t)-\beta(z)]K_{tz} + O(h^3)]^2}{h^3} 1_{\left[|[\beta(z_t)-\beta(z)]K_{tz}+O(h^3)|>\epsilon\sqrt{nh^3}\right]} f(z_t)dz_t$$

$$= \int \frac{[\beta^{(1)}(z)hpK(p) + O(h^2)]^2}{h^2} 1_{\left[|\beta^{(1)}(z)hpK(p)+O(h^2)|>\epsilon\sqrt{nh^3}\right]} f(z+ph)dp$$

$$= \left(\beta^{(1)}(z)\right)^2 f(z) \int p^2 K^2(p) 1_{\left[|\beta^{(1)}(z) pK(p)| > \epsilon\sqrt{nh}\right]} dp + O(h)$$

$$\to \begin{cases} f(z)\left(\beta^{(1)}(z)\right)^2 \nu_2(K) > 0 & \text{if } nh \to 0 \\ f(z)\left(\beta^{(1)}(z)\right)^2 \int p^2 K^2(p) 1_{\left[|pK(p)\beta^{(1)}(z)| > \epsilon\sqrt{c}\right]} dp > 0 & \text{if } nh \to c \in (0,\infty) \end{cases},$$

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

$$\begin{aligned} \frac{\sqrt{nh}}{n^2 h}\sum_{t=1}^n x_t x_t' K_{tz} &= \frac{\sqrt{nh}}{n}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\left\{f(z) + O(h^2)\right\} + \sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{nh}} \\ &\sim_a cf(z)\int B_x B_x' + \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{x_t}{\sqrt{n}}\frac{x_t'}{\sqrt{n}}\frac{\zeta_{tK}}{\sqrt{h}} + o_p(1) = O_p(1). \end{aligned} \tag{B.9}$$

The second term of (B.9) 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

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

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

$$\left(\frac{1}{\sqrt{n}}x_{\lfloor n\cdot\rfloor}, \quad \frac{1}{\sqrt{nh}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK}, \quad \frac{1}{\sqrt{nh^3}}\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t\right) \rightsquigarrow \left(B_x(\cdot), \quad B_{\zeta K}(\cdot), \quad B_\eta(\cdot)\right), \tag{B.10}$$

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,

$$\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' \zeta_{tK} = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\zeta_{tK}}{\sqrt{nh}} \rightsquigarrow \int B_x B_x' dB_{\zeta K}, \tag{B.11}$$

$$\frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\eta_t}{\sqrt{nh^3}} \rightsquigarrow \int B_x B_x' dB_\eta \tag{B.12}$$

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.,

$$\int B_x \otimes B_x dB_{\zeta K} \equiv_d \mathcal{MN} \left(0, \nu_0(K) f(z) \int B_x B_x' \otimes B_x B_x' \right), \tag{B.13}$$

$$\int B_x B_x' dB_\eta \equiv_d \mathcal{MN} \left(0, \nu_2(K) f(z) \int B_x B_x' \left(B_x(r)' \beta^{(1)}(z) \right)^2 \right), \tag{B.14}$$

and the stated result (i) holds.

Part (d) (ii) When the rate conditions $nh \to \infty$ 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.11) and (B.12), we have in the same manner as before in the proof of Part (c)(ii)

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

$$\frac{1}{\sqrt{n^3 h^3}} \sum_{t=1}^{n} x_t x_t' \eta_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\eta_t}{\sqrt{h^3}} = O_p(1), \tag{B.16}$$

again without invariance principles. ■

Part (e) (i) Write

$$\frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz} = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \right) \left(\frac{u_t K_{tz}}{\sqrt{nh}} \right) \rightsquigarrow \int B_x dB_{uK},$$

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 $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} u_t K_{tz} = O_p(1)$ but no invariance principle holds. In a similar fashion and as in Parts (c)(ii) and (d)(ii), we deduce that

$$\frac{1}{\sqrt{n^2 h}} \sum_{t=1}^{n} x_t u_t K_{tz} = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{u_t K_{tz}}{\sqrt{nh}} \right) = O_p(1), \tag{B.17}$$

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

$$\left(\tfrac{1}{\sqrt{n}}x_{\lfloor n\cdot\rfloor},\quad \tfrac{1}{\sqrt{nh}}\textstyle\sum_{t=1}^{\lfloor n\cdot\rfloor}u_tK_{zt},\quad \tfrac{1}{\sqrt{nh}}\textstyle\sum_{t=1}^{\lfloor n\cdot\rfloor}\zeta_{tK},\quad \tfrac{1}{\sqrt{nh^3}}\textstyle\sum_{t=1}^{\lfloor n\cdot\rfloor}\eta_t\right) \rightsquigarrow \left(B_x(\cdot),\quad B_{uK}(\cdot),\quad B_{\zeta K}(\cdot),\quad B_\eta(\cdot)\right), \tag{B.18}$$

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

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

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

$$\sum_{j\geq 1} j^c[\alpha(j)]^{1-2/\delta} = \sum_{j\geq 1}\frac{1}{j^{\frac{\gamma}{3}-\frac{1}{2}}} = \sum_{j\geq 1}\frac{1}{j^{\frac{3}{2}+2\epsilon}} < \infty \tag{B.19}$$

holds and the innovations have finite moments of order $p > 2\delta = 6$. ■

Lemma B.2. *Under Assumption 1(iv)-(vi) and Assumption 2(i)(ii), when $nh \to \infty$ and $h \to 0$ the following hold:*

(a) $\frac{1}{nh}\sum_{t=1}^{n}x_tx_t'K_{tz} \xrightarrow{p} f(z)\mathbb{E}[x_tx_t'|z_t = z]$,

(b) $\mathbb{E}\left(x_tx_t'[\beta(z_t) - \beta(z)]K_{tz}\right) = h^3\mu_2(K)C_S(z) + o\left(h^3\right)$,

(c) $A_{nz}^{-1}\sum_{t=1}^{n}\mathbb{E}[x_tx_t'[\beta(z_t) - \beta(z)]K_{tz}|z_t = z] \sim_a h^2\mathcal{B}_S(z)$,

(d) $\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{n}\{x_tx_t'[\beta(z_t)-\beta(z)]K_{tz}-\mathbb{E}\left(x_tx_t'[\beta(z_t) - \beta(z)]K_{tz}\right)\} \rightsquigarrow \mathcal{N}\left(0, \nu_2(K)f(z)\mathbb{E}\{x_tx_t'[x_t'\beta^{(1)}(z)]^2|z_t = z\}\right)$,

(e) $\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}x_tu_tK_{tz} \rightsquigarrow \mathcal{N}(0, \nu_0(K)\sigma_u^2f(z)\mathbb{E}[x_tx_t'|z_t = z])$,

with $A_{nz} = \sum_{t=1}^{n}x_tx_t'K_{tz}$, $\mathcal{M}_z = f(z)\mathbb{E}[x_tx_t'|z_t = z]$, $\mathcal{B}_S(z) = \mu_2(K)\mathcal{M}_z^{-1}C_S(z)$,

$$C_S(z) = \mathbb{E}\left\{x_tx_t'\left[\beta^{(1)}(z)\frac{f_z\left(x_t, z\right)}{f(x_t|z_t = z)} + \frac{1}{2}\beta^{(2)}(z)f\left(z\right)\right]\Bigg|z_t = z\right\} \tag{B.20}$$

and where $f_z\left(s, z\right) = \partial f\left(s, z\right)/\partial z$, $f(x_t, z_t)$ *is the joint density of* (x_t, z_t) *at* (s, z) *and* $f(x_t|z_t = z)$ *is the conditional density of* x_t *given* $z_t = z$. *When Assumption 2(iii) holds and* x_t *and* z_t *are mutually independent* $\mathcal{B}_S(z) = \mathcal{B}(z) = \frac{\mu_2(K)}{f(z)}\left[\beta^{(1)}(z)f^{(1)}(z) + \frac{1}{2}\beta^{(2)}(z)f(z)\right]$.

Proof. (a)

$$\frac{1}{nh}\sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{1}{nh}\sum_{t=1}^{n}\mathbb{E}[x_t x_t' K_{tz}] + \frac{1}{nh}\sum_{t=1}^{n}\{x_t x_t' K_{tz} - \mathbb{E}[x_t x_t' K_{tz}]\}$$
$$= \frac{1}{nh}\sum_{t=1}^{n}\mathbb{E}\Big\{[x_t x_t'|z_t = z]\mathbb{E}K_{tz}\Big\} + O_p\left(\frac{1}{\sqrt{nh}}\right) \tag{B.21}$$
$$= \frac{1}{nh}\sum_{t=1}^{n}\mathbb{E}[x_t x_t'|z_t = z] h f(z) + O_p\left(\frac{1}{\sqrt{nh}}\right)$$
$$= \mathbb{E}[x_t x_t'|z_t = z] f(z) + o_p(1), \tag{B.22}$$

proving (a). The second equality (B.21) is justified as follows. Taking the scalar x_t case with no loss of generality, we have $\mathbb{E}[x_t^2 K_{tz}]^2 = \mathbb{E}[x_t^4|z_t = z] h f(z)\nu_0(K) = O(h)$ and $\mathbb{V}ar(x_t^2 K_{tz}) = O(h)$. Then, by central limit theory for conventional kernel-weighted centred sums of stationary processes, we have $\sum_{t=1}^{n}\{x_t x_t' K_{tz} - \mathbb{E}[x_t x_t' K_{tz}]\} = O_p(\sqrt{nh})$, giving (B.21).

(b) The following decomposition is useful for analyzing the bias component

$$\sum_{t=1}^{n} x_t x_t'[\beta(z_t) - \beta(z)]K_{tz} = \sum_{t=1}^{n}\mathbb{E}\{x_t x_t'[\beta(z_t) - \beta(z)]K_{tz}\}$$
$$+ \sum_{t=1}^{n}\{x_t x_t'[\beta(z_t) - \beta(z)]K_{tz} - \mathbb{E}\{x_t x_t'[\beta(z_t) - \beta(z)]K_{tz}\}\}. \tag{B.23}$$

We first evaluate $\mathbb{E}\{x_t x_t'[\beta(z_t) - \beta(z)]K_{tz}\}$. For some $\tilde{z}_t$ on the line segment between z_t and z, and some $\widetilde{rh}$ on the line segment between rh and 0, we have

$$\mathbb{E}\{x_t x_t'[\beta(z_t) - \beta(z)]K_{tz}\} = \mathbb{E}\left(x_t x_t'[\beta^{(1)}(z)(z_t - z) + \frac{1}{2}\beta^{(2)}(\tilde{z}_t)(z_t - z)^2]K_{tz}\right)$$
$$= \int_s\int_p ss'[\beta^{(1)}(z)(p - z) + \frac{1}{2}\beta^{(2)}(\tilde{z}_t)(p - z)^2]K\left(\frac{p - z}{h}\right) f(s, p)\, ds dp$$
$$= \int_s\int_r ss'[\beta^{(1)}(z) rh + \frac{1}{2}\beta^{(2)}(z + \widetilde{rh}) r^2 h^2]K(r) f(s, z + rh)\, ds dr h$$
$$= \int_s\int_r ss'[\beta^{(1)}(z) rh + \frac{1}{2}\beta^{(2)}(z + \widetilde{rh}) r^2 h^2]\left[f(s, z) + f_z\left(s, z + \widetilde{rh}\right) rh\right] K(r)\, ds dr h$$
$$= h^3\int_s\int_r ss'[\beta^{(1)}(z) f_z(s, z) + \frac{1}{2}\beta^{(2)}(z) f(s, z)]K(r) r^2 ds dr + o\left(h^3\right)$$
$$= h^3\mu_2(K)\int_s ss'[\beta^{(1)}(z) f_z(s, z) + \frac{1}{2}\beta^{(2)}(z) f(s, z)] ds + o\left(h^3\right)$$
$$= h^3\mu_2(K)\int_s ss'[\beta^{(1)}(z)\frac{f_z(s, z)}{f(s|z)} + \frac{1}{2}\beta^{(2)}(z) f(z)] f(s|z)\, ds + o\left(h^3\right)$$
$$= h^3\mu_2(K)\,\mathbb{E}\left\{x_t x_t'\left[\beta^{(1)}(z)\frac{f_z(x_t, z)}{f(x_t|z_t = z)} + \frac{1}{2}\beta^{(2)}(z) f(z)\right]\Bigg| z_t = z\right\} + o\left(h^3\right)$$
$$=: h^3\mu_2(K)\, C_S(z) + o\left(h^3\right), \tag{B.24}$$

where $C_S(z)$ is given in (B.20), thereby establishing (b).

(c) The bias function now follows from (B.22) and (B.24)

$$A_{nz}^{-1}\sum_{t=1}^{n}\mathbb{E}[x_t x_t'[\beta(z_t)-\beta(z)]K_{tz}|z_t=z] \sim_a [nh\mathbb{E}[x_t x_t'|z_t=z]f(z)]^{-1}nh^3\mu_2(K)C_S(z)$$
$$= h^2\mu_2(K)\mathcal{M}_z^{-1}C_S(z) = h^2\mathcal{B}_S(z), \tag{B.25}$$

Note that when x_t and the covariate z_t are independent we have

$$C_S(z) = \mathbb{E}\left\{x_t x_t'\right\}\left[\beta^{(1)}(z)f^{(1)}(z)+\frac{1}{2}\beta^{(2)}(z)f(z)\right],\ \mathcal{M}_z = f(z)\mathbb{E}[x_t x_t'], \tag{B.26}$$

giving

$$h^2\mathcal{B}_S(z) = h^2\mu_2(K)\mathcal{M}_z^{-1}C_S(z) = h^2\frac{\mu_2(K)}{f(z)}\left[\beta^{(1)}(z)f^{(1)}(z)+\frac{1}{2}\beta^{(2)}(z)f(z)\right] = h^2\mathcal{B}(z), \tag{B.27}$$

reproducing the standard deterministic bias function of the stationary case in (1.9) and that of the nonstationary case in Theorem 2.1. Hence, under the same independence condition between x_t and z_t, the bias function $\mathcal{B}_S(z)$ in the stationary case aligns precisely with the nonstationary version $\mathcal{B}(z)$. This correspondence ensures that the robust test statistics $\hat{T}(z)$ and $\hat{T}_2(z)$ have the same limiting form in both stationary and nonstationary cases under those comparable conditions.

(d) The elements of the sum in the second term on the right side of (B.23) are stationary with zero mean and variance matrix of order $O(h^3)$. In particular,

$$\begin{aligned}
&\mathbb{E}\left\{x_t x_t'\left[\beta(z_t)-\beta(z)\right]K_{tz}\right\}\left\{K_{tz}[\beta(z_t)-\beta(z)]'x_t x_t'\right\} = \mathbb{E}\left\{x_t\left[x_t'\beta^{(1)}(\tilde{z})\left(z_t-z\right)\right]^2 x_t' K_{tz}^2\right\}\\
&= \mathbb{E}\left\{x_t\left[hx_t'\beta^{(1)}(\tilde{z})\left(\frac{z_t-z}{h}\right)\right]^2 x_t' K^2\left(\frac{z_t-z}{h}\right)\right\}\\
&= h^3\int_s\int_r s[s'\beta^{(1)}(z+\widetilde{rh})]^2 s' f\left(s,z+rh\right)K^2\left(r\right)r^2 ds dr + o\left(h^3\right)\\
&= h^3\nu_2\left(K\right)\int_s s[s'\beta^{(1)}(z)]^2 s' f\left(s,z\right)ds + o\left(h^3\right)\\
&= h^3\nu_2\left(K\right)f\left(z\right)\int_s s[s'\beta^{(1)}(z)]^2 s' f\left(s|z\right)ds + o\left(h^3\right)\\
&= h^3\nu_2\left(K\right)f\left(z\right)\mathbb{E}\left\{x_t[x_t'\beta^{(1)}(z)]^2 x_t'|z_t=z\right\} + o\left(h^3\right).
\end{aligned} \tag{B.28}$$

It follows that the variance matrix of the time series $\{x_t x_t'[\beta(z_t)-\beta(z)]K_{tz} - \mathbb{E}\{x_t x_t'[\beta(z_t)-\beta(z)]K_{tz}\}$ is $h^3\nu_2\left(K\right)f\left(z\right)\mathbb{E}\left\{x_t x_t'[x_t'\beta^{(1)}(z)]^2|z_t=z\right\}+o\left(h^3\right)$. Then, by standard central limit theory for kernel-weighted centred stationary processes, we have

$$\frac{1}{\sqrt{nh^3}}\sum_{t=1}^{n}\{x_t x_t'[\beta(z_t)-\beta(z)]K_{tz}-\mathbb{E}\{x_t x_t'[\beta(z_t)-\beta(z)]K_{tz}\}\} \rightsquigarrow \mathcal{N}\left(0,\nu_2(K)f(z)\mathbb{E}\{x_t x_t'[x_t'\beta^{(1)}(z)]^2|z_t=z\}\right), \tag{B.29}$$

giving result (d).

(e) Since u_t is a martingale difference with $\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2$ $a.s.$ and $\mathcal{F}_t = \sigma\{\{x_s, z_s\}_1^\infty; u_t, u_{t-1}, \ldots\}$, we have $\mathbb{E}(x_t u_t K_{tz}) = 0$ and

$$\mathbb{E}(x_t x_t' u_t^2 K_{tz}^2) = \sigma_u^2 \mathbb{E}[x_t x_t' | z_t = z] \mathbb{E}(K_{tz}^2) = h\nu_0(K)\sigma_u^2 f(z)\mathbb{E}[x_t x_t' | z_t = z] = O(h).$$

Hence $x_t u_t K_{tz} = O_p(\sqrt{h})$. Then, by standard central limit theory for kernel-weighted centred stationary processes we have $\frac{1}{\sqrt{nh}} \sum_{t=1}^n x_t u_t K_{tz} \rightsquigarrow \mathcal{N}(0, f(z)\nu_0(K)\sigma_u^2 \mathbb{E}[x_t x_t' | z_t = z])$, as stated.
■

References

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Ibragimov, R. and 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. *The Econometrics Journal*, 8(2):197–213.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Liang, H., Phillips, P. C. B., Wang, H., and Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory*, 32(6):1349–1375.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, 24(2):165–179.

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

Phillips, P. C. B. and Wang, Y. (2020). Limit theory in locally flat functional coefficient regression. *Working Paper, Yale University.*

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3):185–207.

Sun, Y., Cai, Z., and 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., and Li, Q. (2011). Measuring correlations of integrated but not cointegrated variables: A semiparametric approach. *Journal of Econometrics*, 164(2):252–267.

Tu, Y. and Wang, Y. (2019). Functional coefficient cointegration models subject to time–varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, 81(6):1401–1423.

Wang, Y., Phillips, P. C. B., and Tu, Y. (2019). Limit theory and inference in non-cointegrated functional coefficient regression. *Working Paper, University of Auckland*.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.