LIMIT THEORY FOR LOCALLY FLAT FUNCTIONAL COEFFICIENT REGRESSION

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

YING WANG
Renmin University of China

Functional coefficient (FC) regressions allow for systematic flexibility in the responsiveness of a dependent variable to movements in the regressors, making them attractive in applications where marginal effects may depend on covariates. Such models are commonly estimated by local kernel regression methods. This paper explores situations where responsiveness to covariates is locally flat or fixed. The paper develops new asymptotics that take account of shape characteristics of the function in the locality of the point of estimation. Both stationary and integrated regressor cases are examined. The limit theory of FC kernel regression is shown to depend intimately on functional shape in ways that affect rates of convergence, optimal bandwidth selection, estimation, and inference. In FC cointegrating regression, flat behavior materially changes the limit distribution by introducing the shape characteristics of the function into the limiting distribution through variance as well as centering. In the boundary case where the number of zero derivatives tends to infinity, near parametric rates of convergence apply in stationary and nonstationary cases. Implications for inference are discussed and a feasible pre-test inference procedure is proposed that takes unknown potential flatness into consideration and provides a practical approach to inference.

1. INTRODUCTION

Kernel approaches to nonparametric regression use localized versions of standard statistical methods to fit shape characteristics of nonlinear functions in statistical models. These methods have been extensively used in applied research across the social, business, and natural sciences. The methods are particularly useful in assessing the role of nonlinearities and parameter instabilities and are used in modeling cross section, time series, and panel data. An especially useful model

900

for which these methods have been developed is functional coefficient (FC) regression. Such regressions allow the responses of a dependent variable to depend locally in a systematic way on movements in other variables.

This paper demonstrates that the limit theory in FC regression depends on the functional shape[1] of the regression coefficient in ways that involve rates of convergence, asymptotic variance, bandwidth selection, and inference. Standard limit theory for kernel regression shows clearly how functional shape affects bias, which is well known to depend on the local first two derivatives of the regression function and the first derivative of the density of the covariate. The limit theory changes in material ways when these and possibly higher derivatives are zero at the point of estimation. Recent work on FC cointegrating regression (Phillips and Wang, 2021) pointed out dependence of the asymptotic variance on the first derivative of the functional coefficient in estimating cointegrating equations. But the effects of flat functional shape on the limit distribution, including the bias function and limiting variance, have not been explored in earlier nonparametric literature on FC regression in either stationary or nonstationary cases. There also appears to be no former research on the implications of flat functional shape on the limit theory for standard kernel density estimation or kernel regression.

The present paper provides results and methods that address these deficiencies. Most nonparametric research concerns functional shape and seeks to explore and comment upon shape characteristics that relate to underlying economic ideas. Examples include (i) functions of interest like elasticities that may be flat, increasing or decreasing in response to other variables; (ii) functions where extrema are of interest in which zero derivatives play a key role; or (iii) climate responses to policy changes on CO_2 emissions, where the magnitudes of departures from null effects are particularly relevant.

New asymptotics are developed in the paper that specifically involve shape characteristics of the function in the locality of the point of estimation. In particular, locally flat behavior in the coefficient function is shown to have a major effect on the form of the asymptotic distribution as well as the rate of convergence, with important differences between stationary and nonstationary regressions. Local flatness in the coefficient function at some point in the covariate space may be regarded as an intermediate case between the usual FC model and regression with a fixed coefficient, allowing for responses of the dependent variable to be unresponsive to movements in other variables at this point in their support. The primary focus in this paper is to develop asymptotics for FC regression under such flatness conditions. Related effects to those described here, as well as the methods provided, may be expected to apply in other nonparametric regression models where flatness occurs in nonlinear nonparametric regressions.

---

[1]The terminology "functional shape" here refers to the local shape of the function at a given point as captured by the magnitude of its derivatives at that point. The terminology "flat functional shape" or "locally flat behavior," which is used extensively in the paper, refers to the fact that the first $(L-1)$-th derivatives of the function at the local point are zeros while the L-th derivative is nonzero, for some integer $L \geq 1$. A precise statement is given in Theorem 2.1.

A further contribution of the paper is the development of a feasible pre-test inference method. Local flatness in a functional coefficient is typically unknown a priori, although it may be hypothesized on the basis of some underlying theory. Direct estimation of the degree of flatness is shown to be challenging. The feasible pre-test approach takes potential flatness into consideration and is shown to deliver efficiency gains in the flat region over naive estimation and inference where potential flatness is simply ignored. Pre-test inference is found to work well in simulations and is competitive in performance with oracle inference where local flatness information is assumed to be known.

The paper is organized as follows. The new limit theory is given in Section 2, which covers both stationary and nonstationary FC regression. Section 3 discusses the implications of the limit theory for inference and proposes a practically feasible pre-test inference method. Section 4 provides simulation evidence corroborating the asymptotics. Section 5 concludes. Proofs of the main results, several subsidiary lemmas, and computation details are given in the Appendix. Additional technical details are provided in the Supplementary Material to this paper. Throughout the paper we use the notation $\equiv_d$ to signify equivalence in distribution, $\sim_a$ to signify asymptotic equivalence, $\rightsquigarrow$ to denote weak convergence on the relevant probability space, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ to denote floor and ceiling functions, $[\cdot]$ to signify the rounded part of a real number, and $\mu_j(K) = \int_{\mathcal{K}} u^j K(u)du, v_j(K) = \int_{\mathcal{K}} u^j K^2(u)du$ for kernel moment functions, where $\mathcal{K}$ is the support of the kernel function K. According to the context, we use $:=$ and $=:$ to signify definitional equality. Unless otherwise indicated $\int$ denotes $\int_0^1$.

2. ASYMPTOTIC THEORY FOR LOCALLY FLAT FC ESTIMATION

The standard FC regression model is a simple extension of linear regression, taking the following form

$$y_t = x_t' \beta(z_t) + u_t, \tag{2.1}$$

in which the covariate z_t determines the strength or weakness of the response of y_t to the regressor x_t. The regressor x_t is a $p \times 1$ time series, which may be stationary or nonstationary. The covariate z_t is a $q \times 1$ time series and is commonly, although not always, assumed to be stationary. The error term u_t is a scalar stationary process. In view of its flexibility as a convenient extension of fixed parameter regression, the model has been extensively studied and applied in econometrics. A popular textbook reference is by Li and Racine (2007, Chap. 9.3). Many papers have studied estimation and inference in this model under various assumptions, including early work by Cai, Fan, and Yao (2000) and Tu and Wang (2020) on stationary regression and much subsequent work on nonstationary regressions covering both cointegrated and noncointegrated models (Juhl, 2005; Xiao, 2009; Cai, Li, and Park, 2009; Sun, Hsiao, and Li, 2011; Wang, Tu, and Chen, 2016; Tu and Wang, 2019; Wang, Phillips, and Tu, 2019; Tu and Wang, 2022).

Kernel weighted local least squares regression is a standard approach to estimate the functional coefficient $\beta(\cdot)$ in (2.1). The local level least squares estimate of $\beta(z)$ is $\hat{\beta}(z) = \left(\sum_{t=1}^n x_t x_t' K_{tz} \right)^{-1} \left(\sum_{t=1}^n x_t y_t K_{tz} \right)$ with kernel function $K_{tz} = K((z_t - z)/h)$ and bandwidth h. The estimate $\hat{\beta}(z)$ may be decomposed in the usual manner into "bias" and "variance" terms as

$$\left(\sum_{t=1}^n x_t x_t' K_{tz} \right) \left(\hat{\beta}(z) - \beta(z) \right) = \sum_{t=1}^n x_t x_t' [\beta(z_t) - \beta(z)] K_{tz} + \sum_{t=1}^n x_t u_t K_{tz}. \tag{2.2}$$

Under suitable regularity conditions the limit theory for $\hat{\beta}(z)$ is normal or mixed normal after standard corrections are employed for bias and suitable recentering or undersmoothing is employed (Phillips and Wang, 2021). These asymptotics lead to a theory of estimation and inference for both stationary, cointegrating, and mixed regressor cases. Our treatment extends the existing limit theory to address the impact of locally flat behavior in the regression coefficient function $\beta(\cdot)$. We start with the stationary case.

2.1. The FC Stationary Model

It is convenient for exposition to use a prototypical version of the model (2.1) in which the following conditions are assumed.

Assumption 1.

(i) The strictly stationary processes $\{x_t, z_t, u_t\}$ are α-mixing processes with mixing numbers $\alpha(j)$ that satisfy $\sum_{j \geq 1} j^c [\alpha(j)]^{1-2/\delta} < \infty$ for some $\delta > 2$, $c > \tau(1 - 2/\delta)$ and $\tau > 1$ with finite moments of order $p > 2\delta > 4$. Further, $\mathbb{E}(x_t x_t' | z_t = z) = \Sigma_{xx}(z) > 0$, $\mathbb{E}(u_t | x_t, z_t) = 0$, $\mathbb{E}(u_t^2 | x_t = x, z_t = z) = \sigma^2(x, z)$, and $\mathbb{E}(x_t x_t' \sigma^2(x_t, z_t) | z_t = z) = \Sigma_{xu}(z)$.
(ii) The density $f(z)$ of the scalar process z_t and the 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) The kernel function $K(\cdot)$ is a bounded probability density function symmetric about zero with $\mu_j(K) = \int_{\mathcal{K}} u^j K(u) du$, $v_j(K) = \int_{\mathcal{K}} u^j K^2(u) du$, and support $\mathcal{K}$ either $[-1, 1]$ or $\mathbb{R} = (-\infty, \infty)$.
(iv) $\beta(z)$ is a smooth function with uniformly bounded continuous derivatives to order $L + 1$ for some integer $L \geq 1$.
(v) $n \to \infty$ and $h \to 0$.

The stationarity conditions in Assumption (i) accord with earlier work on nonparametric and FC kernel regression for which the mixing requirements are commonly used to enable development of asymptotic theory in time series FC regression (e.g., Fan and Yao, 2008; Cai et al., 2000). A stronger mixing decay rate condition $c > \tau(1 - 2/\delta)$ for some $\delta > 2$ and $\tau > 1$ in (i) is used in place of the more usual condition $c > 1 - 2/\delta$ to assist in the nonparametric limit distribution theory under dependence. The exogeneity condition in (i) is convenient for the

limit theory. Relaxation of those conditions requires alternative methods such as FC instrumental variable methods and additional technical complications that are not within the goals of the present work to address. Heteroskedasticity is allowed as in Cai et al. (2000). The kernel assumptions in (iii) are commonly employed but when bandwidths are very small, as they are in some of the results herein, kernels with support $\mathcal{K}$ on the entire real line $\mathbb{R}$ are better suited, or other methods used to avoid finite sample failure in the kernel-weighted signal in the regression.

The smoothness conditions (iv) on $\beta(z_t)$ and its derivatives are needed for the theory developed here because the limiting bias expressions rely on higher order derivatives of $\beta(z_t)$. When the smoothness degree parameter L is unknown and estimated a stronger condition may be required to allow for potential overestimation of L in practice. Condition (v) is standard in nonparametric work and specific rate conditions involving (n, h) are given as needed in the results below. However, as shown in the analysis of limit behavior when $L \to \infty$, the optimal bandwidth may no longer satisfy the contraction condition $h \to 0$ in (v).

Our first result details the limit theory for the FC regression estimator $\hat{\beta}(z)$ in model (2.1) for the stationary case under locally flat conditions on the coefficient function.

THEOREM 2.1. *If Assumption 1 holds, if $\beta(z)$ has derivatives $\beta^{(\ell)}(z) = 0^2$ at z for all $\ell = 0, 1, 2, \ldots, L-1$ and some integer $L \geq 1$ for which $\beta^{(L)}(z) \neq 0$, then the following limit theory holds when $nh \to \infty$*

$$\sqrt{nh}\left\{\hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z) - o(h^{L^*})\right\} \rightsquigarrow \mathcal{N}(0, \Omega_S(z)), \tag{2.3}$$

where $L^* = (L+1)\mathbf{1}_{\{L=odd\}} + L\mathbf{1}_{\{L=even\}}$, $\Omega_S(z) = \frac{v_0(K)\sigma_u^2}{f(z)}\Sigma_{xx}^{-1}$,

$$\mathcal{B}_L(z) = \frac{\mu_{L^*}(K)}{f(z)}C_L(z) = \frac{\mathcal{G}_L(z)}{f(z)} \tag{2.4}$$

$\mu_{L^*}(K) = \mu_L(K) \times \mathbf{1}_{\{L=even\}} + \mu_{L+1}(K)\mathbf{1}_{\{L=odd\}}, \mathcal{G}_L(z) = \mu_{L^*}(K)C_L(z)$ *and*[3]

$$C_L(z) = \frac{f(z)\beta^{(L)}(z)}{L!}\mathbf{1}_{\{L=even\}} + \left[\frac{\beta^{(L)}(z)}{L!}f^{(1)}(z) + \frac{\beta^{(L+1)}(z)}{(L+1)!}f(z)\right]\mathbf{1}_{\{L=odd\}}. \tag{2.5}$$

Theorem 2.1 shows that flatness in the functional coefficient $\beta(\cdot)$ at z affects the limit theory of $\hat{\beta}(z)$ in the stationary x_t regressor case only through the bias function $h^{L^*}\mathcal{B}_L(z) = h^{L^*}\frac{\mathcal{G}_L(z)}{f(z)}$ in (2.3). The bias order $O(h^{L^*})$ and the functional form $\mathcal{G}_L(z)$ are affected. The bias function $\mathcal{G}_L(z)$ depends on the first two nonzero derivatives $\{\beta^{(\ell)}(z); \ell = L, L+1\}$ of $\beta(z)$, as well as the density $f(z)$ and its first derivative

---

[2] As a notational convention we define $\beta^{(0)}(z) \equiv 0$ without imposing any restriction on the value of the function $\beta(z)$, thereby including without restriction the standard case where the first derivative is nonzero, with $\beta^{(1)}(z) \neq 0$ and $L = 1$.

[3] In the specialized degenerate case where L is odd and $f^{(1)}(z) = 0$, $\beta^{(L+1)}(z) = 0$, it is immediately apparent that the bias is of smaller order than h^{L^*}, as mentioned by the co-editor.

$f^{(1)}(z)$, the latter appearing as is usual in nonparametric regression. When L is even the dependence is confined to the derivative $\beta^{(L)}(z)$ and the density $f(z)$. The limiting variance formula $\Omega_S(z) = \frac{v_0(K)\sigma_u^2}{f(z)} \Sigma_{xx}^{-1}$ is unchanged from the standard case without flatness and the convergence rate remains $\sqrt{nh}$. So, the effect of local flatness in $\beta(z)$ affects the limit theory of FC regression only via the bias function.

As L rises with an increasing degree of flatness in the regression coefficient at z, the bias function in (2.3), which is of order $O(h^{L^*})$, falls when $h \to 0$ as $n \to \infty$. When estimation bias falls it is natural to select a wider bandwidth to reduce variance. Correspondingly, the usual plug-in optimal bandwidth formula changes, with resulting adjustment to the convergence rate. This can be conveniently shown in the scalar coefficient function $\beta(z)$ case, for which the optimal bandwidth formula for minimizing asymptotic mean squared error can be deduced from (2.3) in the usual way, giving (using the scalar x_t case to illustrate)

$$h_{\text{opt}*} = \left(\frac{\Omega_S(z)}{2L^* \mathcal{B}_L(z)^2} \right)^{\frac{1}{2L^*+1}} \frac{1}{n^{1/(2L^*+1)}}. \tag{2.6}$$

In the conventional case where $L = 1$ and $L^* = 2$, we have the usual optimal bandwidth rate $h_{\text{opt}*} = O(n^{-\frac{1}{5}})$. More generally, and taking L to be even for convenience so that $L^* = L$ and $\mathcal{B}_L(z) = \left(\frac{\beta^{(L)}(z)}{L!} \right) f(z)\mu_L(K)$, we have

$$h_{\text{opt}*} = \left(\frac{L!\,(L-1)!\,\Omega_S(z)}{2[\mu_L(K)f(z)\beta^{(L)}(z)]^2} \right)^{1/(2L+1)} \frac{1}{n^{1/(2L+1)}} = \frac{c_L(z)}{n^{1/(2L+1)}} \tag{2.7}$$

where $c_L(z) = d_L(z)[L!\,(L-1)!]^{1/(2L+1)}$ with $d_L(z) = \left(\frac{\Omega_S(z)}{2[\mu_L(K)f(z)\beta^{(L)}(z)]^2} \right)^{1/(2L+1)}$. For instance, when the functional coefficient has the polynomial form $\beta(z) = \sum_{j=0}^{q} a_j z^{L+j}$ which is locally flat to order $L-1$ at $z = 0$ when $a_0 \neq 0$, we have $\beta^{(L)}(z) = L! \sum_{j=0}^{q} \frac{(L+j)!}{L!j!} a_j z^j$ and $\beta^{(L)}(0) = a_0 L! = O(L!)$. The same applies when the locally flat coefficient function $\beta(z)$ has the asymptotically regular form $\beta(z) \sim_a a_0 z^L$ as $L \to \infty$. In such cases, it is evident that $c_L(z) = O([L!/\beta^{(L)}(z)]^{1/L}) = O(1)$ as $L \to \infty$ and the optimal bandwidth $h_{\text{opt}*}$ in (2.7) approaches the nonshrinking rate $O(1/n^0) = O(1)$. Hence, for large L the associated optimal convergence rate is $\sqrt{nh_{\text{opt}*}}$ which approaches $\sqrt{n}$, giving a near-parametric convergence rate for extremely flat functions.

This behavior matches the heuristic that when a functional coefficient is nearly flat and bias is small from neighboring observation points, averaging over those observations by using a wider (or asymptotically nonshrinking) bandwidth is useful in reducing variance and thereby mean squared error. Note, however, that for this optimal choice of bandwidth as $L \to \infty$, in such cases we have $\mathcal{B}_L(z) = O(\beta^{(L)}(z)/L!) = O(1)$ as $L \to \infty$ so that $\sqrt{nh}h^{L^*}\mathcal{B}_L(z) = O\left(n^{1/2}h^{L+1/2}\right) = O\left(n^{1/2}n^{-\frac{L+1/2}{2L+1}}\right) = O(1)$ following (2.7) for the case that L is even. So the bias term is, as usual, not negligible for the optimal choice of bandwidth. What Theorem 2.1, formula (2.7), and this asymptotic bias analysis show is that when the coefficient

function is nearly flat in the neighborhood of the point of estimation, near parametric convergence rates are possible with the same limit normal distribution and variance as in other cases.

2.2. The FC Cointegrating Regression Model

For exposition we use a cointegrating regression equation with full rank $I(1)$ exogenous regressors and functional coefficients. The model is a prototype of more complex systems and provides results that show the impact of flat behavioral characteristics in the functional coefficients on rates of convergence, estimation, inference, and bandwidth selection in a nonstationary framework. These simplifying conditions enable the use of standard kernel-weighted least squares regression. Similar analyses to those given here will be needed in more complex modeling environments under endogeneity and cointegrated equations with possibly cointegrated or even functionally cointegrated regressors. Extensions to address such complexities would involve procedures such as "fully modified" FCC kernel regression. Some related FM methods have been designed for the time varying parameter framework of cointegration (Phillips, Li, and Gao, 2017; Li, Phillips, and Gao, 2016; Gao and Phillips, 2013) and may be developed for FC cointegrating models. But they are not the subject of the present work and are left for future research.

The following assumption modifies the conditions of Assumption 1 and provides for a simple cointegrating regression analogue of model (2.1).

Assumption 2.

(i) $\{x_t\}$ is a full rank unit root process 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 $\Sigma_{xx} > 0$. $\{u_t, z_t\}$ are strictly stationary 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 $c > \tau(1-2/\delta)$ and $\tau > 1$ with finite moments of order $p > 2\delta > 4$. $\{x_t\}$ and $\{z_t\}$ are independent processes. Further, $\mathbb{E}(u_t|x_s) = 0$ for all t,s, and $\mathbb{E}(u_t^2|x_t,z_t) = \sigma_u^2$.

(ii) Assumptions 1(ii)–(v) hold.

The high level assumption (i) on the functional limit behavior of the regressor x_t is convenient, commonly used, and justified by standard primitive conditions (e.g., Phillips and Solo, 1992). The independence assumption between x_t and z_t is imposed to get useful mixed normal limit theory and a convenient inferential framework.[4] The exogeneity and homoskedasticity are also imposed as in Cai et al. (2009). There are ways of dealing with these restrictions, such as developing a fully modified version of the FC estimation procedure or some new version of IVX estimation (Phillips and Magdalinos, 2009) and limit theory tailored to nonstationary FC models. But these options deserve another paper. In further

---

[4]We conjecture that the independence and exogeneity conditions can be relaxed under additional summability conditions on the component time series. This relaxation will be explored in future work.

extensions of this type to FCC regression models, many of the findings of the present work on the effects of local flatness of the functional coefficient will be relevant and can be explored in future work. The remaining conditions are as in Assumption 1. The condition on $\beta(z_t)$ and its derivatives are needed in the nonstationary case because they figure in the development and appear in the asymptotic variance formula.

THEOREM 2.2. *If Assumption 2 holds, if $\beta(z)$ has derivatives $\beta^{(\ell)}(z) = 0$ at z for all $\ell = 0, 1, 2, \ldots, L-1$ and some integer $L \geq 1$ for which $\beta^{(L)}(z) \neq 0$, and $\mathbb{E}||\beta^{(L)}(z_t)||^2 < \infty$, then the following limit theory holds under the respective rate conditions indicated:*

(i) $n\sqrt{h} \left\{ \hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z) - o(h^{L^*}) \right\} \rightsquigarrow \mathcal{MN}\left(0, \Omega_{NS}(z)\right),$ **(2.8)**

   $\quad$ *if $nh^{2L} \to 0, nh \to \infty$,*

(ii) $\sqrt{\dfrac{n}{h^{2L-1}}} \left\{ \hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z) - o(h^{L^*}) \right\} \rightsquigarrow \mathcal{MN}\left(0, \Omega_L(z)\right),$

   $\quad$ *if $nh^{2L} \to \infty$,*   **(2.9)**

(iii) $n^{1-\frac{1}{4L}} \left\{ \hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z) - o(h^{L^*}) \right\}$

   $\rightsquigarrow c^{\frac{1}{2}-\frac{1}{4L}} \times \mathcal{MN}\left(0, \Omega_L(z)\right) + c^{-\frac{1}{4L}}\mathcal{MN}\left(0, \Omega_{NS}(z)\right)$

   $\equiv_d \mathcal{MN}\left(0, c^{1-\frac{1}{2L}}\Omega_L(z) + c^{-\frac{1}{2L}}\Omega_{NS}(z)\right),$ $\quad$ *if $nh^{2L} \to c \in (0, \infty)$,*   **(2.10)**

where $L^* = (L+1)\mathbf{1}_{\{L=odd\}} + L\mathbf{1}_{\{L=even\}}$,

$$\Omega_{NS}(z) = \frac{v_0(K)\sigma_u^2}{f(z)}\left(\int B_x B_x'\right)^{-1},$$ **(2.11)**

$$\Omega_L(z) = \frac{v_{2L}(K)}{(L!)^2 f(z)}\left(\int B_x B_x'\right)^{-1}\left(\int B_x B_x'\left(B_x'\beta^{(L)}(z)\right)^2\right)\left(\int B_x B_x'\right)^{-1},$$ **(2.12)**

and where the bias function $\mathcal{B}_L(z) = \frac{\mu_{L^*}(K)}{f(z)}C_L(z) = \frac{\mathcal{G}_L(z)}{f(z)}$, just as in Theorem 2.1.

The division of the limit theory of FC cointegrating regression into three categories was discovered in Phillips and Wang (2021) for the case where $L = 1$. Theorem 2.2 extends those results to the general case and reveals the effect on both the limit theory and the convergence rate of local flatness in the coefficient function at the point of estimation. As shown in Phillips and Wang (2021) and as is evident in the proof of Theorem 2.2, the presence of multiple categories to the limit theory arises because two different sources of variability occur in the

asymptotics – one from the random elements of the bias function and one from the sample covariance of the regressor and the equation error. Correspondingly, the form of the limit theory itself changes, according to the behavior of nh^{2L}.

Category (i) where $nh^{2L} \to 0$ is comparable to the stationary case, but with convergence rate $n\sqrt{h}$ that embodies the $O(\sqrt{n})$ order of the $I(1)$ regressor x_t and a limit variance matrix that replaces the stationary sample moment matrix limit Σ_{xx} with the corresponding quadratic functional $\int B_x B_x'$ for the nonstationary case in the limit matrix Ω_{NS} in (2.11). The bias function $h^{L^*} \mathcal{B}_L(z)$ in the centering of $\hat{\beta}(z)$ is identical to the stationary case and has the same order $O(h^{L^*})$. Mixed normal limit theory, but with different rates of convergence and different variance matrices, applies in cases (i), (ii) and the intermediate case (iii).

Remark 2.1 (Convergence-rate optimal bandwidth order). In case (iii) of Theorem 2.2 where $nh^{2L} \to c$ for some constant $c \in (0, \infty)$, the bandwidth $h \sim_a (c/n)^{\frac{1}{2L}}$ and then the convergence rate in case (ii) becomes $\sqrt{n/h^{2L-1}} = O\left(\sqrt{n^{1+\frac{2L-1}{2L}}}\right) = O\left(n^{1-\frac{1}{4L}}\right)$. Similarly, the convergence rate in case (i) becomes $n\sqrt{h} = O\left(n^{1-\frac{1}{4L}}\right)$ when $h \sim_a (c/n)^{\frac{1}{2L}}$. This duality between the two cases implies that the convergence rates in cases (i) and (ii) merge to the same $O\left(n^{1-\frac{1}{4L}}\right)$ rate for the intermediate situation where the bandwidth satisfies $nh^{2L} \to c$. In fact, the case $nh^{2L} \to c \in (0, \infty)$ yields the maximum convergence rate outcome for FCC regression because, for the boundary cases where $nh^{2L} \to 0$ or $nh^{2L} \to \infty$, we find that the respective convergence rates are $n\sqrt{h} = o\left(n^{1-\frac{1}{4L}}\right)$ and $\sqrt{n/h^{2L-1}} = o\left(n^{1-\frac{1}{4L}}\right)$. Thus, the FCC kernel regression convergence rate is optimal in the intermediate case where $nh^{2L} \to c \in (0, \infty)$. The associated convergence-rate optimal bandwidth, denoted h_{opt}, is $h_{\mathrm{opt}} \sim_a (c/n)^{\frac{1}{2L}} = O\left(n^{-\frac{1}{2L}}\right)$. The limit distribution is a mixture of the mixed normal component $\mathcal{MN}(0, \Omega_L(z))$ (which comes from the random element of the bias function) and the mixed normal component $\mathcal{MN}(0, \Omega_{NS}(z))$ (which comes from the usual equation error term). The coefficients in this mixture are $c^{\frac{1}{2}-\frac{1}{4L}}$ and $c^{-\frac{1}{4L}}$. For instance, if the locally flat function $\beta(z)$ has the asymptotically regular form $\beta(z) \sim_a a_0 z^L$ as $L \to \infty$, the variance matrix $\Omega_L(z) = O\left((\beta^{(L)}(z)/L!)^2\right) = O(1)$. Then the $c^{1-\frac{1}{2L}} \Omega_L(z)$ component in the asymptotic variance in (2.10) remains stable when L is large just like the $c^{-\frac{1}{2L}} \Omega_{NS}(z)$ component. Therefore the random element coming from the bias function cannot be ignored even when the functional coefficient is sufficiently flat at the point of estimation.

However, with $h = O\left(n^{-\frac{1}{2L}}\right)$, based on case (iii) of Theorem 2.2, the bias term cannot be neglected because it is of order $O\left(n^{1-\frac{1}{4L}} \times n^{-\frac{L^*}{2L}}\right) = O\left(n^{\frac{4L-2L^*-1}{4L}}\right) \to \infty$ when $L \geq 2$. Therefore when discussing the optimal bandwidth order, we need to take the bias effect into consideration, not only the convergence rate. This requires examination of the Mean Squared Error (MSE) optimal bandwidth order, as given next.

Remark 2.2 (Optimal bandwidth order). We explore the optimal bandwidth order with respect to Root Mean Squared Error (RMSE). Let $h = O(n^\gamma)$, $-1 < \gamma < 0$, and $\hat{\beta}(z) - \beta(z) = O(n^{g_L(\gamma)})$. The exponent function $g_L(\gamma)$ in the latter expression represents the order of the RMSE, which is determined by the maximum of the bias order and the standard deviation order. The subindex L in $g_L(\gamma)$ indicates that the RMSE order function varies with parameter L.

First consider the case where L is odd in which case $L^* = L + 1$. Based on result (i) of Theorem 2.2, when $nh^{2L} \to 0$ or equivalently $\gamma < -\frac{1}{2L}$, we have $\hat{\beta}(z) - \beta(z) = O_p\left(\frac{1}{n\sqrt{h}} + h^{L^*}\right) = O_p(n^{-1-\gamma/2} + n^{(L+1)\gamma})$. Then we have $g_L(\gamma) = \max\{-1 - \gamma/2, (L+1)\gamma\}$ when $\gamma < -\frac{1}{2L}$. Similarly, based on result (ii) of Theorem 2.2, we have $g_L(\gamma) = \max\{(L-1/2)\gamma - 1/2, (L+1)\gamma\}$ when $\gamma > -\frac{1}{2L}$. Following result (iii) of Theorem 2.2, we have $g_L(\gamma) = \max\left\{-1 + \frac{1}{4L}, (L+1)\gamma\right\}$ when $\gamma = -\frac{1}{2L}$. Straightforward analysis yields

$$g_1(\gamma) = \begin{cases} 2\gamma, & -1/3 \le \gamma < 0 \\ -\frac{1-\gamma}{2}, & -1/2 < \gamma < -1/3 \\ -(1+\gamma/2), & -1 < \gamma \le -1/2 \end{cases} \tag{2.13}$$

and

$$g_L(\gamma) = \begin{cases} (L+1)\gamma, & -\frac{2}{2L+3} \le \gamma < 0 \\ -1 - \frac{\gamma}{2}, & -1 < \gamma < -\frac{2}{2L+3} \end{cases} \tag{2.14}$$

for $L \ge 3$.

Similarly, suppose L is even, in which case $L^* = L$, and $g_L(\gamma)$ can be derived based on Theorem 2.2. Thus, if $nh^{2L} \to 0$ or equivalently $\gamma < -\frac{1}{2L}$, we have $\hat{\beta}(z) - \beta(z) = O_p(\frac{1}{n\sqrt{h}} + h^{L^*}) = O_p(n^{-1-\gamma/2} + n^{L\gamma})$. When $\gamma = -\frac{1}{2L}$, we have $g_L(\gamma) = \max\{-1 + \frac{1}{4L}, L\gamma\} = \max\{-1 + \frac{1}{4L}, -1/2\} = -1/2$; and when $\gamma > -\frac{1}{2L}$, we have $g_L(\gamma) = \max\{(L-1/2)\gamma - 1/2, L\gamma\}$. Standard calculations yield

$$g_L(\gamma) = \begin{cases} L\gamma, & -\frac{2}{2L+1} \le \gamma < 0 \\ -1 - \frac{\gamma}{2}, & -1 < \gamma < -\frac{2}{2L+1} \end{cases} \tag{2.15}$$

when L is even. Note that (2.14) and (2.15) can be combined as

$$g_L(\gamma) = \begin{cases} L^*\gamma, & -\frac{2}{2L^*+1} \le \gamma < 0 \\ -1 - \frac{\gamma}{2}, & -1 < \gamma < -\frac{2}{2L^*+1} \end{cases} \tag{2.16}$$

for $L \ge 2$.

The $g_L(\gamma)$ functions are plotted in Figure 1. Evidently, when $L = 1$, the RMSE optimal bandwidth order is $h_{opt*} = O(n^{-1/2})$, which equals the convergence-rate optimal bandwidth order h_{opt}. For $L \ge 2$, the optimal bandwidth is $h_{opt*} = O(n^{-\frac{2}{2L^*+1}})$, which is smaller than the convergence-rate optimal bandwidth order
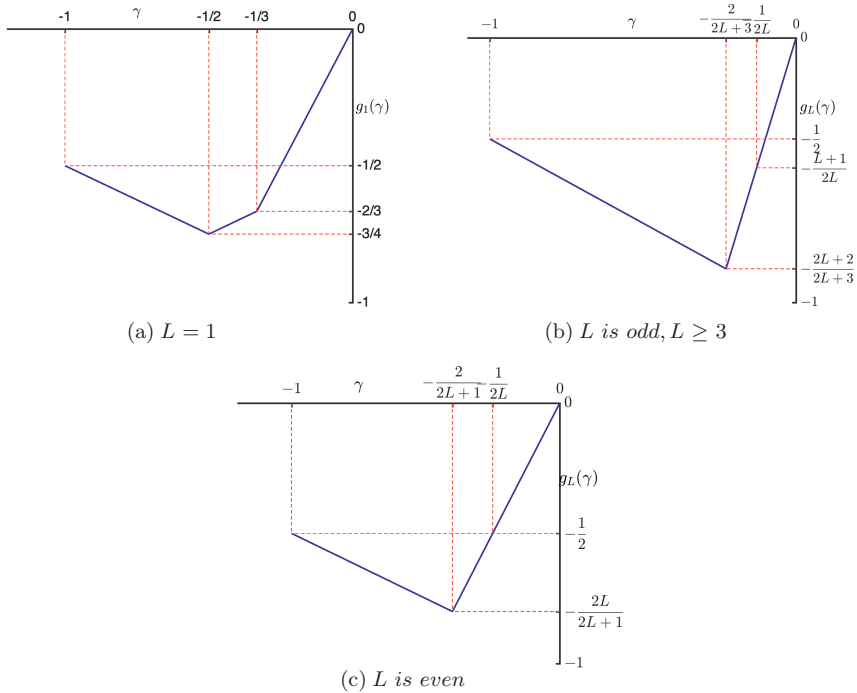
(a) $L = 1$

(b) L *is odd,* $L \geq 3$

(c) L *is even*

FIGURE 1. Plots of $g_L(\gamma)$ for $L \geq 1$. See online version for all color graphics.

$h_{opt} = O(n^{-\frac{1}{2L}})$ when $L \geq 2$. The discrepancy between these two optimal band-width rates is due to the fact that when $L \geq 2$ bias dominates variance in result (iii) of Theorem 2.2. To reduce bias, the RMSE optimal bandwidth prefers to select a smaller order. When L is large, we can see the order of h_{opt*}, viz., $-\frac{2}{2L^*+1}$, is close to zero and then h_{opt*} diminishes to zero at a very slow rate as $n \to \infty$. This outcome is consistent with heuristics as $\beta(z)$ is close to a constant function at the estimation point z when L is large; and estimation of an almost constant function requires only a very low degree of localization.

Remark 2.3 (MSE optimal bandwidth formula). The above analysis tells us that the RMSE optimal bandwidth order, or equivalently, the MSE optimal bandwidth order, is achieved within case (i) of Theorem 2.2. Taking the standard approach to optimal bandwidth selection that balances bias and variance (and using the scalar x_t case for convenience) leads to the following formula compared with the stationary regressor case given in (2.6)

$$h_{\text{opt*}} = \left(\frac{\Omega_{NS}(z)}{2L^* \mathcal{B}_L(z)^2} \right)^{\frac{1}{2L^*+1}} \frac{1}{n^{2/(2L^*+1)}} \equiv c_L(z) n^{-\frac{2}{2L^*+1}} \tag{2.17}$$

where $c_L(z) = \left(\frac{\Omega_{NS}(z)}{2L^* \mathcal{B}_L(z)^2}\right)^{\frac{1}{2L^*+1}}$. To illustrate, suppose L is even in which case

$$c_L(z) = \left(\frac{\Omega_{NS}(z)}{2L \mathcal{B}_L(z)^2}\right)^{\frac{1}{2L+1}} = \left(\frac{\nu_0(K)\sigma_u^2 L! (L-1)!}{2\mu_L(K)^2 f(z)(\int B_x^2)\beta^{(L)}(z)^2}\right)^{\frac{1}{2L+1}}. \qquad \textbf{(2.18)}$$

If, as before, the functional coefficient is flat at $z = 0$ with polynomial form $\beta(z) = \sum_{j=0}^q a_j z^{L+j}$ and $a_0 \neq 0$, then $\beta^{(L)}(z) = L! \sum_{j=0}^q \frac{(L+j)!}{L!j!} a_j z^j = O(L!)$ and $c_L(z) = O_p(1)$ as $L \to \infty$, the randomness of $c_L(z)$ arising from the presence of the quadratic functional $\int B_x^2$ in (2.18). Again, the optimal bandwidth $h_{opt*} = O_p(n^{-2/(2L+1)})$ and for large L the optimal bandwidth shrinks at a very slow rate and the associated optimal convergence rate $n\sqrt{h_{opt*}}$ approaches n, giving a near-parametric convergence rate for extremely flat functions. This suggests that a larger bandwidth is needed for large L. But in practice in both stationary and nonstationary cases, L is typically unknown, so $c_L(z)$ and the optimal bandwidth order are also unknown in the absence of information about $\beta(z)$ and its derivatives. While estimation of optimal bandwidths by cross validation or by the use of derivative function estimates is possible, these methods typically lead to very slow convergence rates in optimal bandwidth formulas even in the simplest cases (Hall and Marron, 1987; Hall et al., 1991). So the above findings are likely to be mainly of importance and use in theoretical work.

Remark 2.4 (Asymptotics with MSE optimal bandwidth). When $L = 1$, choice of the MSE optimal bandwidth order h_{opt*} leads to asymptotics that are determined according to case (iii) of Theorem 2.2 since $nh_{opt*}^{2L} = O(1)$ when $L = 1$. More specifically, the limit theory for $\hat{\beta}(z)$ is given by

$$n^{3/4}\left(\hat{\beta}(z) - \beta(z) - h_{opt*}^2 \mathcal{B}_1(z)\right) \rightsquigarrow \mathcal{MN}\left(0, c^{1/2}\Omega_L(z) + c^{-\frac{1}{2}}\Omega_{NS}(z)\right), \qquad \textbf{(2.19)}$$

which matches the result in Phillips and Wang (2021, Thm. 2.1(c)) for the standard case of no flatness in $\beta(z)$. In this case, the bias can be neglected because $n^{3/4} \times h_{opt*}^2 = O(n^{-1/4}) = o(1)$. When $L \geq 2$, we have $nh_{opt*}^{2L} \to 0$ and the limit theory is determined by case (i) of Theorem 2.2. Specifically with $h = h_{opt*} = O(n^{-\frac{2}{2L^*+1}})$, we have

$$n^{\frac{2L^*}{2L^*+1}}\left\{\hat{\beta}(z) - \beta(z) - h_{opt*}^{L^*}\mathcal{B}_L(z)\right\} \rightsquigarrow \mathcal{MN}\left(0, \Omega_{NS}(z)\right). \qquad \textbf{(2.20)}$$

In this case, the random bias component involving $\Omega_L(z)$ can be ignored asymptotically but the deterministic bias term cannot be neglected because $n^{\frac{2L^*}{2L^*+1}} \times h_{opt*}^{L^*} = O\left(n^{\frac{2L^*}{2L^*+1}} \times n^{-\frac{2L^*}{2L^*+1}}\right) = O(1)$. Using the MSE optimal bandwidth $h_{opt*} = O\left(n^{-\frac{2}{2L^*+1}}\right)$, the fastest convergence rate that $\hat{\beta}(z)$ can achieve is $O_p\left(n^{-\frac{2L^*}{2L^*+1}}\right)$ when $L \geq 2$. As $L \to \infty$, the fastest convergence rate approaches $O_p(n^{-1})$, leading to the parametric cointegrating regression convergence rate $\hat{\beta}(z) - \beta(z) = O_p(n^{-1})$

as $L \to \infty$. As in the stationary case, this matches heuristic arguments because $\beta(z)$ approaches a constant function at the estimation point z when $L \to \infty$.

3. IMPLICATIONS FOR INFERENCE

3.1. Procedures for Inference

When L is known or is correctly hypothesized standard test statistics for inference about the functional coefficient $\beta(z)$ can be constructed in a standard way. Following Phillips and Wang (2021), but allowing now for local flatness in the coefficient function, we start with the matrix normalization[5]

$$\hat{T}(z;L) = \hat{V}_n(z;L)^{-1/2}\left[\hat{\beta}(z) - \beta(z) - h^{L^*}\hat{\mathcal{B}}_L(z)\right], \tag{3.1}$$

where $\hat{V}_n(z;L) = A_n(z)^{-1}\hat{\Omega}_n(z;L)A_n(z)^{-1}$ with $A_n(z) = \sum_{t=1}^n x_t x_t' K_{tz}$,

$$\hat{\Omega}_n(z;L) = v_0(K)\hat{\sigma}_u^2 \sum_{t=1}^n x_t x_t' K_{tz} + \sum_{t=1}^n x_t x_t' \left\{x_t' \frac{1}{L!}\hat{\beta}^{(L)}(z)(z_t - z)^L K_{tz}\right\}^2, \tag{3.2}$$

and

$$\hat{\mathcal{B}}_L(z) = \mu_{L^*}(K)\left\{\frac{\hat{\beta}^{(L)}(z)}{L!}1_{\{L=even\}} + \left[\frac{\hat{\beta}^{(L)}(z)}{L!}\frac{\hat{f}^{(1)}(z)}{\hat{f}(z)} + \frac{\hat{\beta}^{(L+1)}(z)}{(L+1)!}\right]1_{\{L=odd\}}\right\}. \tag{3.3}$$

The statistic $\hat{T}(z;L)$ follows the same design as the robust t-test statistic developed in Phillips and Wang (2021) for the nonflat case with $L = 1$.

The bias component $h^{L^*}\hat{\mathcal{B}}_L(z)$ in (3.1) and the second term of $\hat{\Omega}_n(z;L)$ in (3.2) are both infeasible in practical work unless L is known or is stated as part of a null hypothesis such as

$$\mathbb{H}_0 : \beta(z) = \beta_0, L = L_0. \tag{3.4}$$

Section 3.3.1 below discusses some of the difficulties involved in the direct estimation of the derivative order parameter L. An alternative way to determine L empirically is to test whether successive derivatives of $\beta(z)$ are zero at the point of estimation using consistent kernel estimates,[6] $\hat{\beta}^{(\ell)}(z)$, of the derivative functions $\beta^{(\ell)}(z)$ and conducting inference to detect zero derivatives at the point of interest. In most practical cases this procedure would involve examination of only the first derivative or first two derivatives ($\ell = 1, 2$). Section 3.3.3 provides an implementation of a two-step pre-test procedure. Fortunately, simulation evidence presented in Section 4.2 indicates that the two-step pre-test approach works well in

---

[5]To avoid dealing with higher order bias when performing inference in this section we assume that the higher order bias can be ignored after being scaled by the convergence rate. In the stationary case, condition $\sqrt{nh}h^{L^*+1} \to 0$ is adequate to ensure this. In the nonstationary case, condition $n\sqrt{h}h^{L^*+1} \to 0$ works for case (i). Conditions for cases (ii) and (iii) can be obtained in a similar fashion.

[6]Derivative estimates can be obtained in the usual way by employing higher order polynomial approximation.

terms of coverage compared with the infeasible test procedure that employs correct information about L.

Under $\mathbb{H}_0$ the statistic $\hat{T}(z;L_0)$ may be used to construct a robust Hotelling's T^2 type statistic based on the quadratic form $\hat{T}_2(z;L_0) = \hat{T}(z;L_0)'\hat{T}(z;L_0)$, so that

$$\hat{T}_2(z;L_0) = \left[\hat{\beta}(z) - \beta_0 - h^{L_0^*}\hat{\mathcal{B}}_{L_0}(z)\right]' \hat{V}_n(z;L_0)^{-1} \left[\hat{\beta}(z) - \beta_0 - h^{L_0^*}\hat{\mathcal{B}}_{L_0}(z)\right].$$

The following result shows that under the null $\mathbb{H}_0$ with use of the correct value of L the statistics $\hat{T}(z;L_0)$ and $\hat{T}_2(z;L_0)$ satisfy $\hat{T}(z;L_0) \rightsquigarrow \mathcal{N}(0,I_p)$ and $\hat{T}_2(z;L_0) \rightsquigarrow \chi_p^2$ as $n \to \infty$. This pivotal limit theory provides a basis for performing inference about $\beta(z)$ when the functional coefficient is locally flat and the flatness parameter L is correctly hypothesized. This approach covers both stationary and nonstationary regressor cases.

THEOREM 3.1. *Under either Assumption 1 or 2, when the null hypothesis $\mathbb{H}_0$ holds and $nh \to \infty$, $\hat{T}(z;L_0) \rightsquigarrow \mathcal{N}(0,I_p)$ and $\hat{T}_2(z;L_0) \rightsquigarrow \chi_p^2$.*

3.2. Test Power

When the null hypothesis is false and the true value of the functional coefficient $\beta(z) \neq \beta_0$ but the maintained hypothesis $L = L_0$ is correct, asymptotic power can be explored under local alternatives of the form

$$\mathbb{H}_{1,\beta} : \beta(z) = \beta_0 + \rho_n m(z),$$

where $m(z)$ is a p-vector function whose modulus is bounded away from the origin and ρ_n is a real sequence for which $\rho_n \to 0$. Let $\chi_p^2(\alpha)$ be the $1-\alpha$ right tail critical value of the χ_p^2 distribution. Then, under $\mathbb{H}_{1,\beta}$ we have

$$\lim_{n\to\infty} P\left(\hat{T}_2(z;L_0) > \chi_p^2(\alpha)\right) = 1, \tag{3.5}$$

for any ρ_n satisfying $\rho_n^2 nh \to \infty$ if x_t is stationary and Assumption 1 holds or for ρ_n satisfying $\rho_n^2 n^2 h \to \infty$ if x_t is nonstationary, $nh^{2L_0} \to 0$ and Assumption 2 holds. To prove (3.5) first consider the stationary case. In view of Theorem 2.1 we have, under $\mathbb{H}_{1,\beta}$,

$$\sqrt{nh}[\hat{\beta}(z) - \beta_0 - \rho_n m(z) - h^{L_0^*}\hat{\mathcal{B}}_{L_0}(z)] \rightsquigarrow \mathcal{N}(0,\Omega_S(z)).$$

Since $\frac{1}{nh}A_n(z) \to_p \Sigma_{xx}f(z)$ and

$$\frac{1}{nh}\hat{\Omega}_n(z;L_0) = v_0(K)\hat{\sigma}_u^2 \frac{1}{nh}\sum_{t=1}^n x_t x_t' K_{tz} + \frac{1}{nh}\sum_{t=1}^n x_t x_t' \left\{x_t' \frac{\hat{\beta}^{(L_0)}(z)}{L_0!}(z_t - z)^{L_0} K_{tz}\right\}^2$$

$$\to_p v_0(K)\sigma_u^2 f(z)\Sigma_{xx}, \tag{3.6}$$

just as in the proof of (A.25), we obtain

$$nh\hat{V}_n(z;L_0) = \left(\frac{1}{nh}A_n(z)\right)^{-1}\frac{1}{nh}\hat{\Omega}_n(z;L_0)\left(\frac{1}{nh}A_n(z)\right)^{-1} \to_p \Omega_S(z). \tag{3.7}$$

It follows that under $\mathbb{H}_{1,\beta}$ and Assumption 1

$$\hat{T}(z;L_0) = \hat{V}_n(z;L_0)^{-1/2}\left[\hat{\beta}(z) - \beta_0 - h^{L^*}\hat{B}_{L_0}(z)\right]$$

$$= \hat{V}_n(z;L_0)^{-1/2}\left[\hat{\beta}(z) - \beta_0 - \rho_n m(z) - h^{L^*}\hat{B}_{L_0}(z)\right] + \left[nh\hat{V}_n(z;L_0)\right]^{-1/2}\sqrt{nh}\rho_n m(z)$$

$$\sim_a \mathcal{N}\left(\Omega_S(z)^{-1/2}\sqrt{nh}\rho_n m(z), I_p\right), \tag{3.8}$$

so that (3.5) holds when $\rho_n^2 nh \to \infty$ and $m(z) \neq 0$ in the stationary case.

In the nonstationary case, the analysis can be carried out separately depending on the rate of nh^{2L_0}. We take $nh^{2L_0} \to 0$ as an example. When $nh^{2L_0} \to 0$ and $nh \to \infty$, from Theorem 2.2(i) under $\mathbb{H}_{1,\beta}$, we have

$$n\sqrt{h}\left\{\hat{\beta}(z) - \beta_0 - \rho_n m(z) - h^{L^*_0}\hat{B}_{L_0}(z)\right\} \rightsquigarrow \mathcal{MN}\left(0, \Omega_{NS}(z)\right), \tag{3.9}$$

where $\Omega_{NS}(z) = \frac{v_0(K)\sigma_u^2}{f(z)}\left(\int_0^1 B_x B_x'\right)^{-1}$. Now $\frac{1}{n^2h}A_n(z) \rightsquigarrow \int_0^1 B_x B_x' f(z)$ and $\frac{1}{n^2h}\hat{\Omega}_n(z;L_0) \rightsquigarrow v_0(K)\sigma_u^2 f(z)\int_0^1 B_x B_x'$, so that

$$n^2h\hat{V}_n(z;L_0) = \left(\frac{1}{n^2h}A_n(z)\right)^{-1}\frac{1}{n^2h}\hat{\Omega}_n(z;L_0)\left(\frac{1}{n^2h}A_n(z)\right)^{-1} \rightsquigarrow \Omega_{NS}(z). \tag{3.10}$$

Hence, under $\mathbb{H}_{1,\beta}$, Assumption 2 and with $nh^{2L_0} \to 0$ and $nh \to \infty$ we have

$$\hat{T}(z;L_0) = \hat{V}_n(z;L_0)^{-1/2}\left[\hat{\beta}(z) - \beta_0 - h^{L^*_0}\hat{B}_{L_0}(z)\right]$$

$$= \hat{V}_n(z;L_0)^{-1/2}\left[\hat{\beta}(z) - \beta_0 - \rho_n m(z) - h^{L^*_0}\hat{B}_{L_0}(z)\right] + \left[n^2h\hat{V}_n(z;L_0)\right]^{-1/2}\sqrt{n^2h}\rho_n m(z)$$

$$\sim_a \mathcal{MN}_m\left(\Omega_{NS}(z)^{-1/2}\sqrt{n^2h}\rho_n m(z), I_p\right), \tag{3.11}$$

where $\mathcal{MN}_m(\cdot,\cdot)$ signifies a mean mixture normal distribution.[7] The test statistic $\hat{T}_2(z;L_0)$ then diverges when $\rho_n^2 n^2 h \to \infty$ because it is asymptotically distributed as a mixture noncentral chi-square variate with the divergent noncentrality parameter $n^2h\rho_n^2 m(z)'\Omega_{NS}(z)^{-1}m(z) \to \infty$. It follows that (3.5) holds in the nonstationary case when $\rho_n^2 n^2 h \to \infty$ and $m(z) \neq 0$.

Results for $nh^{2L_0} \to \infty$ and $nh^{2L_0} \to c \in (0,\infty)$ can be obtained in similar ways and the details are omitted. In the case where $nh^{2L_0} \to \infty$, the condition

---

[7]A random p-vector ξ has a mean mixture normal density with covariance matrix I_p if the density of ξ is the mixture density $\frac{1}{(2\pi)^{p/2}}\int_{\vartheta\in\Theta}e^{-\frac{1}{2}(x-\vartheta)'(x-\vartheta)}dP(\vartheta)$ where $P(\vartheta)$ is the probability measure of the mean-mixing variate vector ϑ.

$n\rho_n^2/h^{2L_0-1} \to \infty$ is needed for the test to be consistent. When $nh^{2L_0} \to c \in (0, \infty)$ the test is consistent if $n^{1-\frac{1}{4L_0}}\rho_n \to \infty$ holds.

Before closing this section we point out that this test is not designed to detect alternatives specifically about L in either stationary or nonstationary cases. To illustrate the difficulties involved in such alternatives, we take the stationarity case with $nh^{2L_0} \to 0$ and consider the alternative $\mathbb{H}_{1,L} : \beta(z) = \beta_0$, $L_0 < L$ where the flatness degree L exceeds the hypothesized L_0. To examine test power observe that

$$
\sqrt{nh}\left\{\hat{\beta}(z) - \beta(z) - h^{L_0^*}\hat{\mathcal{B}}_{L_0}(z)\right\}
$$
$$
= \sqrt{nh}\left\{\hat{\beta}(z) - \beta(z) - h^{L^*}\hat{\mathcal{B}}_L(z)\right\} + \sqrt{nh}\left\{h^{L^*}\hat{\mathcal{B}}_L(z) - h^{L_0^*}\hat{\mathcal{B}}_{L_0}(z)\right\}.
$$

The order of the second term depends on both L and L_0, noting that the order of $\hat{\mathcal{B}}_{L_0}(z)$ depends on L_0 and L through the empirical estimates $\hat{\beta}^{(L_0)}(z)$ and $\hat{\beta}^{(L_0+1)}(z)$ that are used in the construction of the test statistic. Due to the fact that L is unknown under $\mathbb{H}_{1,L}$, the order of magnitude of the component $\sqrt{nh}\left\{h^{L^*}\hat{\mathcal{B}}_L(z) - h^{L_0^*}\hat{\mathcal{B}}_{L_0}(z)\right\}$ cannot be precisely determined and the power characteristics of the test are not known. Since these properties of the test in the case of departures L from the hypothesized L_0 are unknown, the statistic is not designed to test hypotheses concerning the flatness order L.

3.3. Challenges in Test Construction when *L* is Unknown

3.3.1. *Direct Estimation of L.*   The statistic $\hat{T}(z; L)$ cannot be used in practical work if L is unknown or is not part of the null or an explicit maintained hypothesis. A natural approach if this were not the case but if L were directly estimable (by $\hat{L}$, say) would be to employ plug-in estimates $\hat{\beta}^{(\hat{L})}(z)$ and $\hat{\beta}^{(\hat{L}+1)}(z)$ of the required derivatives of $\beta(z)$ in the bias and variance matrix components of $\hat{T}(z; L)$. However, as the analysis below reveals, in the general case of unknown L such a plug-in approach encounters difficulties because of the challenge of direct consistent estimation of L. Further, as earlier analysis reveals, the optimal bandwidth order in FC regression depends on the flatness degree parameter L. Since L is a higher order property of an unknown nonparametric function $\beta(z)$, this dependence poses a subtle question of how to determine the bandwidth h in estimation and inference.

In this respect, noting that $\beta(z_t) - \beta(z) = \frac{\beta^{(L)}(\tilde{z}_t)}{L!}(z_t - z)^L$ where $\tilde{z}_t$ lies on the line segment between z_t and z and $\beta^{(L)}(z) \neq 0$ by assumption, it follows that as $n \to \infty$ and $h \to 0$

$$
L_n^\dagger = \frac{1}{\log(h)}\log\left(\frac{\frac{1}{n}\sum_{t=1}^n |\beta(z_t) - \beta(z)|K_{tz}}{\frac{1}{n}\sum_{t=1}^n K_{tz}}\right) \to_p L,
$$

as shown in Part 1 in the Supplementary Material. Of course, $L_n^\dagger$ is an infeasible rate estimator reliant on the unknown function $\beta(\cdot)$ in a neighborhood of z. It has a slow logarithmic convergence rate with $L_n^\dagger - L = O_p(1/\log(h))$, so that

when $h = n^{-\delta}$ for some $\delta > 0$ we have $L_n^{\dagger} - L = O_p(1/\log(n))$. More specifically, $\log(h)(L_n^{\dagger} - L) = \log\left(\left|\frac{\beta^{(L)}(z)}{L!}\right|\right) + \log\left(\int |s|^L K(s) ds\right) + o_p(1)$, as shown in Part 1 of the Supplementary Material.

Setting $w_{tz} = K_{tz}/\sum_{t=1}^n K_{tz}$, this limit behavior suggests the following "plausible" practical estimate of L

$$\hat{L} = \frac{1}{\log(h)} \log\left(\sum_{t=1}^n \left|\hat{\beta}(z_t) - \hat{\beta}(z)\right| w_{tz}\right), \tag{3.12}$$

which can be computed using a preliminary bandwidth h satisfying $h \to 0$ and $nh \to \infty$. However, the estimator $\hat{L}$ is consistent only when bandwidth h is appropriately selected. Part 2 of the Supplementary Material provides a rigorous demonstration of this property of $\hat{L}$ in the cases $L = 1$ and $L = 2$. It is found that when $L = 1$, $\hat{L}$ is consistent only when condition $nh^3 \to c \in (0, \infty]$ is satisfied and when $L = 2$, the consistency condition is $nh^5 \to c \in (0, \infty]$. In general, we expect that when $nh^{2L+1} \to c \in (0, \infty]$ holds $\hat{L}$ would be consistent for $L \geq 1$. Hence, "appropriate selection" of the bandwidth for $\hat{L}$ to be consistent in general requires knowledge of L, making $\hat{L}$ infeasible in practice. Simulations that confirm these findings are also provided in Part 2 in the Supplementary Material.

A further complication that should be mentioned is that even if a consistent estimator of L were available, bias correction requires specification of the bandwidth factor $h(L) = h^L$ in $h^L \mathcal{B}_L(z)$, which presents additional difficulties. For example, whereas the infeasible estimator $L_n^{\dagger} \to_p L$, the consistency of $L_n^{\dagger}$ does not mean that $h^{L_n^{\dagger}} \sim_a h^L$. At the end of Part 2 in the Supplementary Material, we show that $h^{L_n^{\dagger}}$ is inconsistent for h^L. Thus the slow rate of convergence of $L_n^{\dagger}$ interferes with consistent estimation of the factor h^L needed for bias correction. In view of all these technical difficulties, the feasibility of direct consistent estimation of L requires further study and is left for future research.

3.3.2. *Adaptive Statistic Design.* This section comments briefly on the possibility of constructing an adaptive test statistic that does not require knowledge of L. The idea stems from Remarks 3.2 and 3.3 in Phillips and Wang (2021) where a statistic is developed that incorporates bias and variance matrix estimators that do not involve L but instead rely on local information about the function obtained by kernel estimation. In principle, it is straightforward to extend this idea to the case where $L > 1$. Take the stationary case as an example. The adaptive bias estimator is defined as

$$\hat{B}(z) = A_n(z)^{-1} \left(\sum_{t=1}^n x_t x_t'\right) \frac{1}{n} \sum_{t=1}^n \left[\hat{\beta}(z_t) - \hat{\beta}(z)\right] K\left(\frac{z_t - z}{h}\right),$$

where the sample average $\frac{1}{n} \sum_{t=1}^n \left[\hat{\beta}(z_t) - \hat{\beta}(z)\right] K\left(\frac{z_t - z}{h}\right)$ is introduced to approximate $\mathbb{E}[\beta(z_t) - \beta(z)]K_{tz}$. Unfortunately this adaptive bias estimator $\hat{B}(z)$ is not

consistent for the true bias when $L > 1$ because local kernel estimation in $\hat{\beta}(z_t) - \hat{\beta}(z)$ is insufficiently precise to capture the required derivative components. In consequence, the limit of $\hat{B}(z)$ has many additional terms when $L > 1$. Moreover, direct (bias correction) adjustment to achieve consistent bias estimation is not possible because the limit of $\hat{B}(z)$ depends on the unknown value of L. More details are provided in the Supplementary Material showing how the adaptive bias estimator fails in flat regions of the function where $L > 1$ in both stationary and nonstationary cases. There are further obstacles to inference in the adaptive bias estimator $\hat{B}(z)$ due to additional variation that affects the limit distribution of the bias centered term $\hat{\beta}(z) - \beta(z) - \hat{B}(z)$. In the nonstationary case, the variance of this term depends on L and $\beta^{(L)}(z)$, making it difficult to estimate the limit variance adaptively without introducing further bias effects. These complications combine to make it difficult to design an adaptive statistic in cases where the flatness degree is unknown, leaving this pursuit as a further challenge for future research.

3.3.3. *Pre-testing.*   In response to these challenges we propose a feasible pre-test method for practical work in cases where flatness may be suspected. The approach is to focus on the first two derivatives and determine L by pre-testing whether these derivatives are zero. Inference is implemented using the self-normalized statistic $\hat{T}(z; L)$ with L replaced by the pre-testing estimate. A two-step pre-test estimation procedure is suggested.

Step 1: Test $\mathbb{H}_{0,\beta^{(1)}} : \beta^{(1)}(z) = 0$. If $\mathbb{H}_{0,\beta^{(1)}}$ is rejected, conclude that $L = 1$ and the procedure stops. Otherwise, continue to Step 2;

Step 2: Test $\mathbb{H}_{0,\beta^{(2)}} : \beta^{(2)}(z) = 0$. If $\mathbb{H}_{0,\beta^{(2)}}$ is rejected, conclude that $L = 2$ and the procedure stops. Otherwise, we simply let $L = 3$.

The procedure stops at $L = 3$ for simplicity and because this measure of flatness should suffice in most practical situations. Significance tests involving the derivatives are constructed as follows. To test $\mathbb{H}_{0,\beta^{(1)}}$ we use limit theory for the derivative estimator $\hat{\beta}^{(1)}(z)$. In our ongoing work (Wang and Phillips, 2022) using local p-th order polynomial estimation, the asymptotic theory for derivative estimators $\hat{\beta}^{(k)}(z)$ $(k = 0, 1, 2, \ldots, p)$ in FC regressions are obtained. For example, with nonstationary x_t and local linear $(p = 1)$ estimation, the asymptotic theory for $\hat{\beta}^{(1)}(z)$ $(k = 1)$ is given by

$$\hat{\beta}^{(1)}(z) - \beta^{(1)}(z) - h^2 f^{-1}(z)\mu_2(K)^{-1}\mu_4(K)B_{2p}(z) \sim_a \sqrt{\frac{h}{n}}\mathcal{MN}\left(0, \Omega_{\beta,1,1}(z)\right)$$

$$+ \frac{1}{n\sqrt{h^3}}\mathcal{MN}\left(0, \Omega_{u,1,1}(z)\right), \tag{3.13}$$

where $\sim_a$ signifies asymptotic equivalence, $B_{2p}(z) = \frac{\beta^{(p+1)}(z)}{(p+1)!}f^{(1)}(z) + \frac{\beta^{(p+2)}(z)}{(p+2)!}f(z)$, $\Omega_{\beta,1,1}(z) = \nu_6(K)\mu_2(K)^{-2}f^{-1}(z)B_{(x,2)}^{-1}\int\left[B_x(B_x'\frac{\beta^{(2)}(z)}{2})^2B_x'\right]B_{(x,2)}^{-1}$, $\Omega_{u,1,1}(z) = \nu_2(K)\mu_2(K)^{-2}\sigma_u^2 f^{-1}(z)B_{(x,2)}^{-1}$, and $B_{(x,2)} = \int B_x B_x'$. From (3.13), it is not hard to see that the optimal bandwidth order that balances bias and asymptotic variance is $n^{-2/7}$. Moreover, with this optimal bandwidth order the first term on the right

side of (3.13) is negligible relative to the second term as $\sqrt{h/n} = o(\frac{1}{n\sqrt{h^3}})$ when $h = O(n^{-2/7})$. Further, result (3.13) remains asymptotically valid when local flatness exists as in that situation the first term is of smaller order than $\sqrt{h/n}$ and is again negligible when the optimal bandwidth order $h = O(n^{-2/7})$ is used. Hence the t-ratio for testing $\mathbb{H}_{0,\beta^{(1)}} : \beta^{(1)}(z) = 0$ can be constructed based on (3.13). The construction follows that of the t-ratio given in (3.1). Estimation of the asymptotic variance in (3.13) can be constructed in the same way as (3.2) since the asymptotic variances $\Omega_{\beta,1,1}(z)$ and $\Omega_{u,1,1}(z)$ in (3.13) take similar forms to those in (2.11) and (2.12) and are not repeated. For testing $\mathbb{H}_{0,\beta^{(2)}} : \beta^{(2)}(z) = 0$, the asymptotic limit theory takes a more complicated form and is given in Appendix D.

Denoting the pre-test estimator of L obtained in this way by $\hat{L}^{pre}$, tests can be constructed using the plug-in statistic $\hat{T}(z; \hat{L}^{pre})$. Further, with $\hat{L}^{pre}$ available, the relevant optimal bandwidth order can be employed, based on (2.6) for stationary x_t and Remark 2.2 for nonstationary x_t. In this way bandwidth is adaptively determined to account for any local flatness of the function. Simulation performance of the feasible statistic $\hat{T}(z; \hat{L}^{pre})$ is reported in Section 4.2.

4. SIMULATIONS

The simulation experiments that follow employ a simple prototypical framework for evaluating the adequacy of the asymptotic theory. We explore the behavior of the FC estimators and the adequacy of the limit theory in locally flat and nonflat regions of the function. The following sections consider estimation and inference in stationary and nonstationary cases, separately.

4.1. Estimation

Nonstationary x_t

In the first experiment the model (2.1) is used with a single $I(1)$ exogenous regressor x_t generated as a random walk with $iid\ \mathcal{N}(0, \sigma_x^2)$ innovations ϵ_{xt} and zero initialization x_0, $iid\ \mathcal{N}(0, \sigma_u^2)$ equation errors u_t, and $iid\ U[-1, 2]$ covariates z_t. We set $\sigma_x^2 = 1$ and $\sigma_u^2 = 1$. Throughout the simulations, the number of replications used is 10,000 and the coefficient function is the quartic $\beta(z) = z^4$, for which the first three derivatives at $z_1 = 0$ are zero, $\beta^{(4)}(z_1) = 4!$ and $\beta^{(1)}(z_2) = 4z_2^3 = 4$ at $z_2 = 1$.

Figure 2 shows the mean bias (plotted in the left panel), standard deviation (plotted in the middle panel) and RMSE (plotted in the right panel) for $\hat{\beta}(z)$ calculated at the points $\{z = 0, 1\}$ using samples of size $n = 100, 400$ and 800, based on 10,000 replications. In estimation we employ a Gaussian kernel and the bandwidth formula $h = \hat{\sigma}_z \times n^\gamma$. The range $-0.90 \leq \gamma \leq -0.05$ is used to meet the condition $nh \to \infty$ and to avoid extremely small bandwidths for which there is considerable imprecision in the simulation estimates, as is evident in the plotted curves for the standard deviation and RMSE near the left limit of the domain of
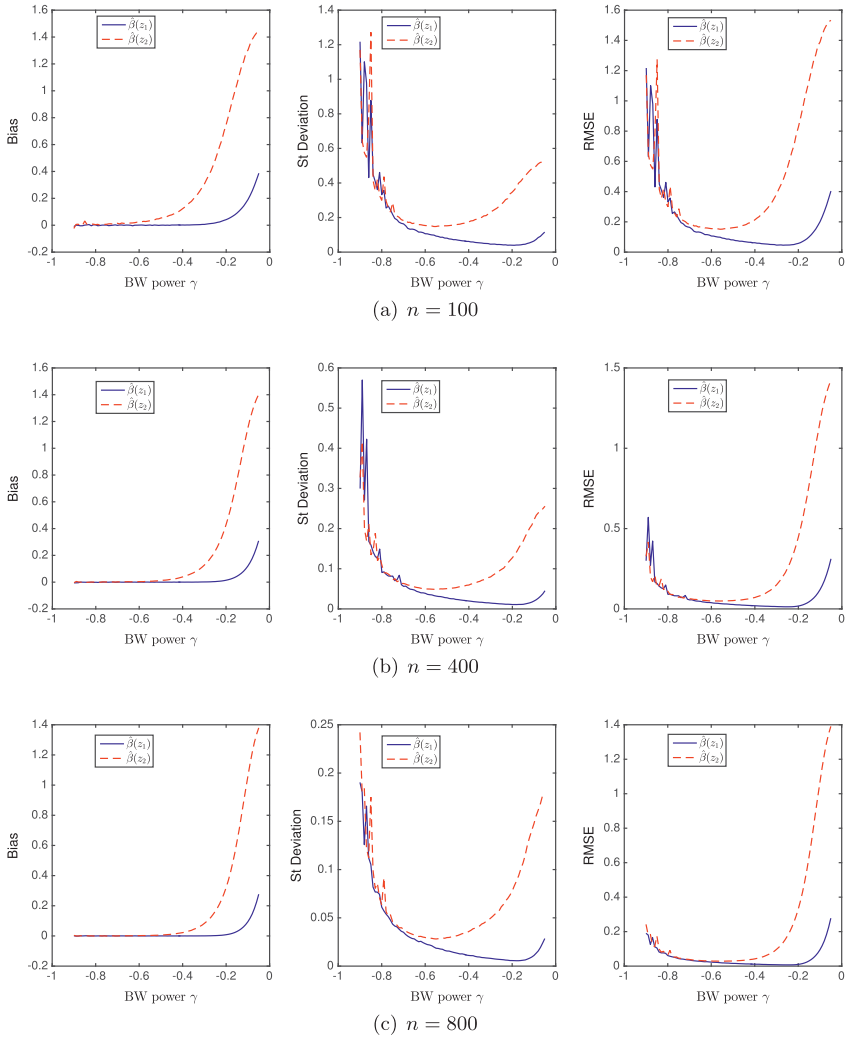
FIGURE 2. Nonstationary case: bias, standard deviation and RMSE plots for FCC estimator $\hat{\beta}(z)$ at points $z_1 = 0$ and $z_2 = 1$ for the quartic coefficient function $\beta(z) = z^4$. The figures show bias, standard deviation, and RMSE in the left, middle, and right panels as functions of bandwidth power γ ($-0.9 \leq \gamma \leq -0.05$) in $h = \hat{\sigma}_z \times n^{\gamma}$ for Model (2.1) with $I(1)$ regressor x_t and sample size $n = 100, 400,$ and 800.

definition.[8] The plots show significant differences in estimator behavior between the two points of estimation $\{z = 0, 1\}$, which we summarize as follows.

---

[8]This imprecision is related to the fact that when $nh \rightarrow c < \infty$ the asymptotic theory changes and no invariance principle applies. Readers are referred to Phillips and Wang (2021) for further analysis and discussion of this phenonomenon.

(i) Bias increases as the bandwidth widens and the bandwidth power $\gamma \to 0$. For very wide bandwidths, estimates at both $z_1 = 0$ and $z_2 = 1$ suffer large bias. However, bias is smaller and usually much smaller at the point $z_1 = 0$ of locally flat functional form than at point $z_2 = 1$. These findings all match the asymptotic theory in Theorem 2.2, which shows that bias has order h^{L^*}, which is h^4 when $z_1 = 0$ where $L^* = L = 4$, compared with h^2 when $z_2 = 1$ where $L^* = L + 1 = 2$ with $L = 1$.

(ii) Standard deviation rises in estimation at both points of estimation as the bandwidth becomes very small when $\gamma \to -1$ or as bandwidth becomes very large when $\gamma \to 0$. This outcome corresponds to asymptotic theory where there are three convergence rates for the cases given in Theorem 2.2, where it is shown that the highest convergence rate (or minimum standard deviation) occurs in the intermediate bandwidth contraction case with $h = O(n^{-\frac{1}{2L}})$. When the bandwidth is very small (γ close to -1), considerable volatility in the standard deviation estimates was found even with a large number of replications, particularly for smaller sample sizes. We therefore only report results for $\gamma \geq -0.90$ and some volatility in the estimates is evident in the graphics close to this lower limit. The standard deviation of $\hat{\beta}(z)$ at $z_2 = 1$ is seen to be substantially greater than that at $z_1 = 0$ except for small bandwidths, again matching the limit theory.

(iii) The RMSE curves demonstrate similar U-shaped patterns to those of the standard deviation curves. This simulation evidence corroborates the analysis in Remark 2.2, where it is shown that the RMSE order $g_L(\gamma)$ has a check function shape with $L = 4$. Further, the RMSE is considerably lower when $\beta(z)$ is flat at $z_1 = 0$ than when the coefficient function is rising at $z_2 = 1$. These gains hold throughout a wide range of bandwidth powers except for smaller bandwidths.

(iv) Across panels (a)–(c) in Figure 2, the main impact of larger sample sizes is the anticipated reduction in the bias, standard deviation, and RMSE, which applies to both $z_1 = 0$ and $z_2 = 1$ cases and across all bandwidth powers.[9]

To better illustrate the optimal bandwidth order discussed in Remarks 2.1 and 2.2, we report the bandwidth power values corresponding to the minimum points of the standard deviation and RMSE curves from the simulations in Figure 2. Results are collected in Table 1 under the panel headed "x_t is nonstationary". According to Remark 2.1, the convergence-rate, or equivalently, the standard-deviation optimal bandwidth order is achieved at $-\frac{1}{2L}$, which is $-\frac{1}{8} \approx -0.13$ for $z_1 = 0$ ($L = 4$) and $-\frac{1}{2}$ for $z_2 = 1$ ($L = 1$). Following Remark 2.2, the RMSE optimal bandwidth order is $-\frac{2}{2L+1} = -\frac{2}{9} \approx -0.22$ for $z_1 = 0$ ($L = 4$) and $-\frac{1}{2}$ for $z_2 = 1$ ($L = 1$). These are the figures reported in the last row of Table 1 for $n = \infty$.

---

[9]The numbers in this row are the optimal bandwidth orders based on the asymptotic theory as $n \to \infty$. For the case that x_t is nonstationary, the standard-deviation optimal bandwidth order is the convergence-rate optimal bandwidth order analyzed in Remarks 2.1. It is given as $-\frac{1}{2L}$, which is $-\frac{1}{8} \approx -0.13$ for $z_1 = 0$ ($L = 4$) and $-\frac{1}{2}$ for $z_2 = 1$ ($L = 1$). The RMSE optimal bandwidth follows Remark 2.2, which is $-\frac{2}{2L+1} = -\frac{2}{9} \approx -0.22$ for $z_1 = 0$ ($L = L^* = 4$) and $-\frac{1}{2}$ for $z_2 = 1$ ($L = 1$). For the case that x_t is stationary, the RMSE optimal bandwidth order is $-\frac{1}{2L^*+1}$ as given in (2.6). Then the true value for $z_1 = 0$ ($L = L^* = 4$) is $-\frac{1}{9} \approx -0.11$ and that for $z_2 = 1$ ($L = 1, L^* = 2$) is $-\frac{1}{5}$.

TABLE 1. Finite sample optimal bandwidth order estimates

| | x_t is nonstationary | | | | x_t is stationary | |
| | StDev optimal | | RMSE optimal | | RMSE optimal | |
	$z_1 = 0$	$z_2 = 1$	$z_1 = 0$	$z_2 = 1$	$z_1 = 0$	$z_2 = 1$
$n = 100$	-0.19	-0.56	-0.27	-0.56	-0.20	-0.40
$n = 400$	-0.18	-0.56	-0.26	-0.56	-0.17	-0.38
$n = 800$	-0.18	-0.55	-0.25	-0.55	-0.17	-0.36
$n = \infty$	-0.13	-0.50	-0.22	-0.50	-0.11	-0.20

Only when $L = 1$ are these two optimal bandwidth orders the same both here and for $z_2 = 1$ in Table 1. When $L = 4$, the convergence-rate optimal bandwidth power is larger than the RMSE optimal bandwidth power. In Table 1 it is evident that for $z_1 = 0$, the standard-deviation optimal bandwidth order estimates are larger than the RMSE optimal bandwidth order estimates. Moreover, as the sample size n increases, the optimal bandwidth order estimates approach the corresponding limit values reported in the last row for $n = \infty$. These results again corroborate the analysis in Remarks 2.1 and 2.2 showing that the RMSE optimal bandwidth rate equals the convergence-rate optimal bandwidth order when $L = 1$ or is less than the convergence-rate optimal bandwidth order when $L \geq 2$.

Stationary x_t

In the second experiment the same model (2.1) is used but with a stationary exogenous regressor x_t generated by the autoregression $x_t = \theta x_{t-1} + \epsilon_{xt}$ with *iid* $\mathcal{N}(0, \sigma_x^2)$ innovations ϵ_{xt} and zero initialization x_0, *iid* $\mathcal{N}(0, \sigma_u^2)$ equation errors u_t, and *iid* $U[-1, 2]$ covariates z_t. We set $\sigma_x^2 = 1$, $\sigma_u^2 = 1$, and $\theta = 0.5$. Again 10,000 replications are employed. The results for bias, standard deviation and RMSE are shown in Figure 3. The plots for the stationary case mirror those in Figure 2 for the FCC case. The imprecision in the simulation estimates at small bandwidths is more severe than in the nonstationary case and results are accordingly reported here for the reduced bandwidth power region $-0.8 \leq \gamma \leq -0.05$. The findings for the stationary case are summarized below.

(v) The main difference with the nonstationary model occurs in the standard deviation curves. Different from the nonstationary case, Theorem 2.1 shows that the convergence rate on the left hand side is unaffected by the local flatness parameter L or the bandwidth rate condition nh^{2L}. We therefore expect to see monotonously decreasing standard deviation curves for both points of estimation $\{z = 0, 1\}$ as the bandwidth power γ increases. From the middle panel of Figure 3, we observe that the standard deviation curve for $z_1 = 0$ indeed shows a decreasing pattern as γ increases to 0, but that for $z_2 = 1$ the curve starts to rise slightly when γ is close to 0. This is explained by the randomness that is present in the bias function in finite samples. Although the randomness in the bias function is of smaller order
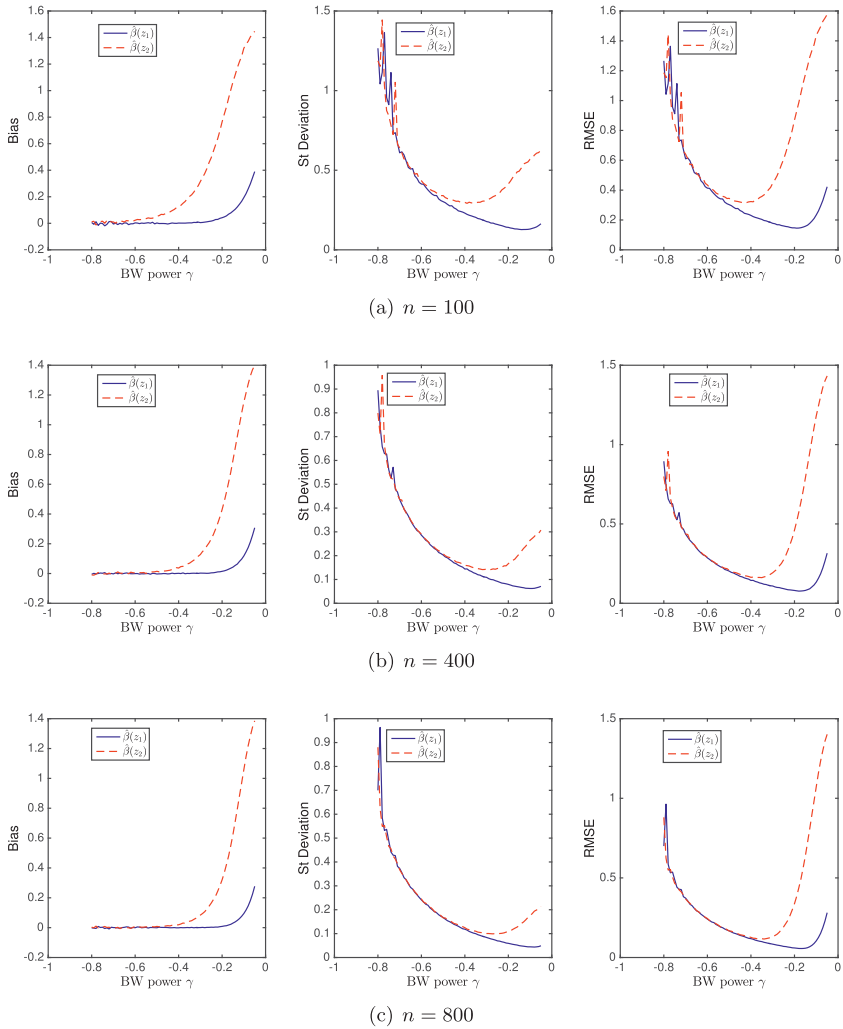
(a) $n = 100$

(b) $n = 400$

(c) $n = 800$

FIGURE 3. Stationary case: bias, standard deviation and RMSE plots for the FC estimator $\hat{\beta}(z)$ at points $z_1 = 0$ and $z_2 = 1$ for the quartic coefficient function $\beta(z) = z^4$. The figures show bias, standard deviation, and RMSE in the left, middle and right panels as functions of bandwidth power γ $(-0.80 \leq \gamma \leq -0.05)$ in $h = \hat{\sigma}_z \times n^{\gamma}$ for Model (2.1) with a stationary autoregressive regressor x_t with autoregressive coefficient $\theta = 0.5$ and sample size $n = 100, 400$ and 800.

than that of the usual error term asymptotically and therefore does not figure in the limit theory, it can still affect finite sample performance. Moreover, the finite sample effects are less severe when the functional coefficient is locally flatter (with larger L) because the bias is smaller when L is larger. This explains why a marked

rise in the curve is only observed towards the right limit near $\gamma = 0$ of the domain of definition of the standard deviation curves for $z_2 = 1$ but not for the curves for $z_1 = 0$.

(vi) For both RMSE curves, there is also a clear minimum RMSE bandwidth choice as in the nonstationary case. Furthermore, the curves indicate that the minimum RMSE bandwidth power γ is larger for estimation at $z_1 = 0$ than at $z_2 = 1$. Direct evidence of this difference is given by the estimates of the RMSE optimal bandwidth power reported in Table 1 under the panel "x_t is stationary". These findings corroborate the analysis concerning the optimal RMSE bandwidth order following Theorem 2.1.

(vii) The plots in Figure 3 show the finite sample gains in estimation that occur from local flatness of the functional coefficient. These gains occur for bandwidths large enough to be well beyond the region where there is imprecision in the simulation estimates of the standard deviation and RMSE.

4.2. Inference

This Section reports findings on the finite sample performance of the t-ratios discussed in Section 3. Four statistics are considered: (i) the oracle t-ratio $T(z; true\ L)$, which assumes L, the derivatives $(\beta^{(L)}(z), \beta^{(L+1)}(z))$, and other components σ_u^2, $f(z)$ and $f^{(1)}(z)$ are known; (ii) the infeasible statistic $\hat{T}(z; true\ L)$ in which the true value of L is used and other components are estimated; (iii) the "naive" statistic $\hat{T}(z; L = 1)$ where $L = 1$ is used as the simplest case without any attention to potential flatness; and (iv) the pre-test statistic $\hat{T}(z; \hat{L}^{pre})$. The oracle and infeasible statistics provide two baselines to assess the relative performance of the naive and pre-testing statistic for comparative purposes. The same generating mechanism is used as in the previous section and we again consider the two evaluation points $z_1 = 0$ (with $L = 4$) and $z_2 = 1$ (with $L = 1$). The bandwidth formula $h = \hat{\sigma}_z n^\gamma$ and a second order Epanechnikov kernel are employed throughout the computations. The naive t-ratio $\hat{T}(z; L = 1)$ uses $\gamma = -1/2$ for nonstationary x_t and $\gamma = -1/5$ for stationary x_t in the computation of $\hat{\beta}(z)$ and K_{tz}. The infeasible statistic $\hat{T}(z; true\ L)$ is identical to the naive choice when $L = 1$. When $L = 4$, it uses the optimal order $\gamma = -1/(2L^* + 1)$ for stationary x_t and $\gamma = -2/(2L^* + 1)$ for nonstationary x_t for the computation of $\hat{\beta}(z)$ and K_{tz}. Given a known L or the estimate $\hat{L}^{pre}$, the oracle t-ratio $T(z; true\ L)$ and the pre-test statistic $\hat{T}(z; \hat{L}^{pre})$ also use the optimal bandwidth order in the computation of $\hat{\beta}(z)$ and K_{tz}. Other details concerning computation are given in Appendix C.

The empirical densities of these t-ratio statistics are shown in Figure 4 for stationary x_t and in Figure 5 for nonstationary x_t. From Figure 4, at the flat point $z_1 = 0$ the densities of the oracle statistic are evidently extremely close to the standard normal. Densities of the naive, infeasible, and pre-test statistics show some discrepancy from the standard normal, but the distribution of the infeasible statistic is closer to standard normal when the sample size is large. The improvement of the infeasible over the naive and pre-test statistics reveals
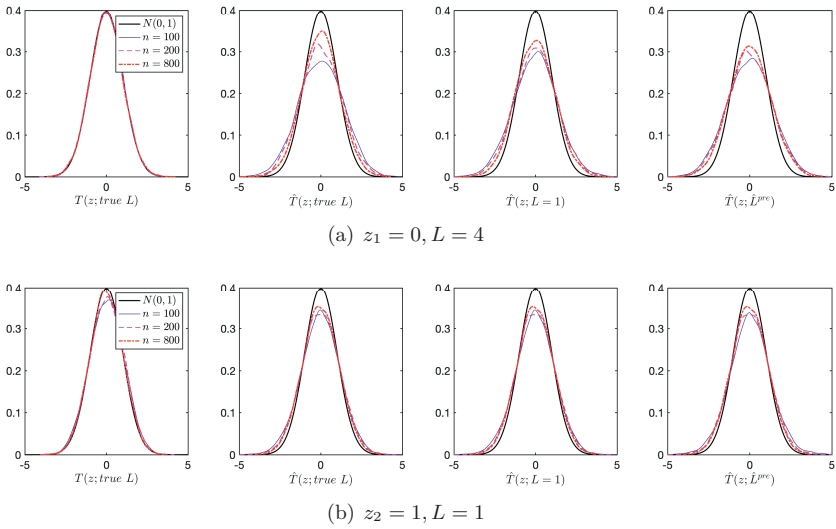
(a) $z_1 = 0, L = 4$

(b) $z_2 = 1, L = 1$

FIGURE 4. Empirical densities of the self-normalized t-ratio $T(z; true\ L)$, $\hat{T}(z; true\ L)$, $\hat{T}(z; L = 1)$, and $\hat{T}(z; \hat{L}^{pre})$ when x_t is stationary.

(a) $z_1 = 0, L = 4$
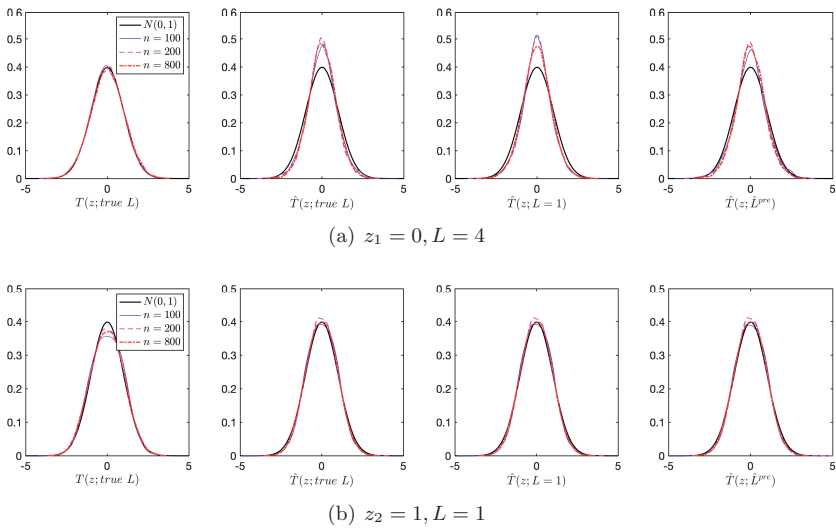

(b) $z_2 = 1, L = 1$

FIGURE 5. Empirical densities of the self-normalized t-ratio $T(z; true\ L)$, $\hat{T}(z; true\ L)$, $\hat{T}(z; L = 1)$, and $\hat{T}(z; \hat{L}^{pre})$ when x_t is nonstationary.

the gains from knowledge of L, or equivalently, the consequences of ignoring or incorrectly estimating local flatness in the coefficient function at the point of flatness. The naive and pre-test methods share similar performance. At the nonflat point, the oracle statistic is the one closest to standard normal. The naive and infeasible statistics are identical in this case because the true value of L is 1. The pre-test statistic is competitive to these. Compared to the performance at the flat point $z_1 = 0$, the densities are closer to standard normal at the nonflat point. In the nonstationary case in Figure 5, the oracle statistic is again extremely close to standard normal at the flat point. The naive, infeasible and pre-test statistics are competitive in performance, although all are too densely distributed at the origin. The pre-test statistic is slightly closer to standard normal. At the nonflat point, the naive and the infeasible distributions are again identical and their performance is competitive to that of the oracle statistic. The pre-test statistic shares the same competitive performance as the naive and infeasible statistics. In summary, ignoring local flatness in the coefficient function seems to cause some efficiency loss at the flat point and the pre-test method appears to recover slightly this lost efficiency in the nonstationary case. To see these features more clearly and consider their implications for inference we explore the coverage rates and confidence interval lengths of the associated tests.

Table 2 reports coverage rates and lengths of the confidence intervals constructed at the two points $z_1 = 0$ and $z_2 = 1$ using the four statistics $T(z; true\ L)$, $\hat{T}(z; true\ L)$, $\hat{T}(z; L = 1)$, and $\hat{T}(z; \hat{L}^{pre})$. Results are based on 20,000 replications. In all situations considered, the oracle statistic has the best performance with coverage rates close to the nominal level and confidence interval length the narrowest in most cases. The infeasible statistic has the second best performance. In the nonstationary case, it has coverage rates close to nominal levels but with confidence intervals slightly wider than those of the oracle. In the stationary case, the efficiency loss of the infeasible statistic versus the oracle is manifest in the lower coverage rate, while the confidence interval can be slightly narrower than the oracle. The naive method is identical to the infeasible method when $L = 1$. When $L = 4$, the naive method shares similar coverage rates with the infeasible method, but with much wider confidence interval lengths, especially when x_t is nonstationary. This finding reflects the efficiency loss of the naive method caused by ignoring local flatness when it is present. The pre-test method has very similar performance to the infeasible method, especially when $L = 1$. When $L = 4$, pre-testing achieves similar coverage rates with the infeasible statistic at the cost of slightly wider confidence intervals. It is worth noting that in the case of $L = 4$, the pre-test method significantly outperforms the naive method with much narrower confidence intervals. This finding reflects the efficiency gain of the estimator $\hat{L}^{pre}$ compared to simply treating L as 1. In sum, the pre-testing and infeasible statistics share similar coverage rates but with slightly wider confidence bands around the flat region for the pre-test statistic. Pre-testing appears to deliver significant efficiency gains over the naive method with much narrower confidence bands in

TABLE 2. Coverage rates and confidence interval length (in brackets) at points (z_1, z_2) based on the oracle t-ratio $T(z; true\ L)$, the infeasible t-ratio $\hat{T}(z; true\ L)$, the naive t-ratio $\hat{T}(z; L=1)$, and the pre-test t-ratio $\hat{T}(z; \hat{L}^{pre})$

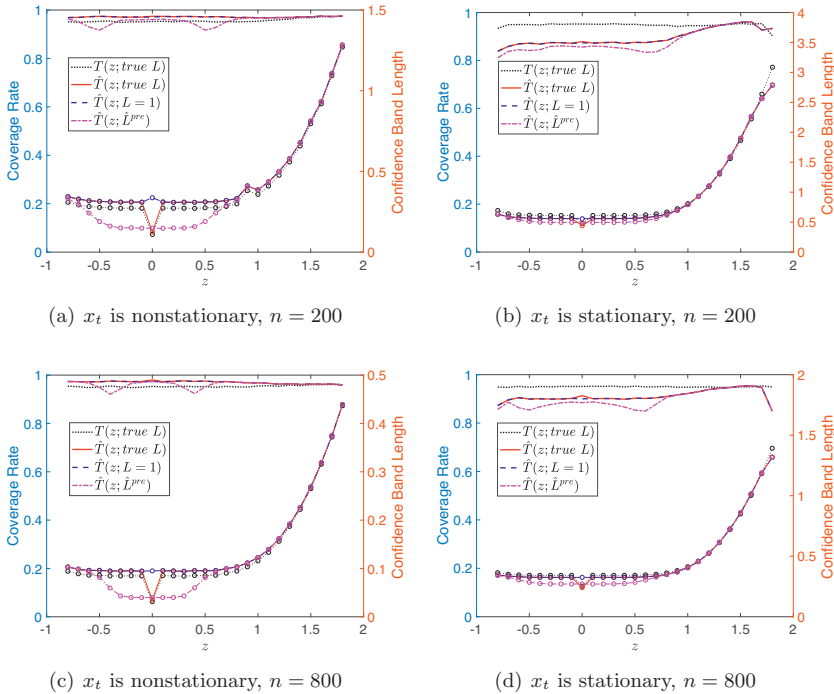|	x_t is nonstationary				x_t is stationary			
n	$T(z; true\ L)$	$\hat{T}(z; true\ L)$	$\hat{T}(z; L=1)$	$\hat{T}(z; \hat{L}^{pre})$	$T(z; true\ L)$	$\hat{T}(z; true\ L)$	$\hat{T}(z; L=1)$	$\hat{T}(z; \hat{L}^{pre})$
	$z_1 = 0\ (L=4)$				$z_1 = 0\ (L=4)$			
100	0.953	0.960	0.965	0.952	0.952	0.847	0.846	0.840
	(0.205)	(0.235)	(0.628)	(0.298)	(0.660)	(0.618)	(0.742)	(0.680)
200	0.950	0.972	0.974	0.962	0.951	0.879	0.872	0.858
	(0.109)	(0.125)	(0.386)	(0.149)	(0.476)	(0.441)	(0.559)	(0.494)
800	0.949	0.977	0.974	0.970	0.951	0.912	0.899	0.882
	(0.031)	(0.035)	(0.094)	(0.040)	(0.252)	(0.240)	(0.325)	(0.270)
	$z_2 = 1\ (L=1)$				$z_2 = 1\ (L=1)$			
100	0.953	0.961	0.961	0.955	0.943	0.905	0.905	0.890
	(0.768)	(0.839)	(0.839)	(0.826)	(1.154)	(1.140)	(1.140)	(1.123)
200	0.954	0.969	0.969	0.969	0.950	0.918	0.918	0.916
	(0.356)	(0.384)	(0.384)	(0.383)	(0.808)	(0.794)	(0.794)	(0.793)
800	0.955	0.969	0.969	0.969	0.947	0.920	0.920	0.920
	(0.116)	(0.123)	(0.123)	(0.123)	(0.413)	(0.406)	(0.406)	(0.406)

FIGURE 6. Coverage rate (left scale) and length (right scale, with lines marked by circles) of the 95% confidence bands over the support of z_t for $n = 200$ and $n = 800$, from 20,000 replications.

the flat case. These results indicate a preference for the pre-testing approach among the feasible statistics.

Figure 6 further demonstrates the overall performance of the four statistics in the support of z_t. The coverage rate curves and lengths of the confidence intervals are plotted over the support $[-1, 2]$ of z_t for sample sizes $n = 200$ and $n = 800$. It is evident that the pre-test statistic has narrower confidence bands in the flat region (the vicinity of the flat point $z = 0$) and the phenomenon is stronger in the nonstationary case. These findings are explained by the results in Figure 7, which shows the empirical frequencies of the pre-test estimator $\hat{L}^{pre}$, where it is clear that $\hat{L}^{pre}$ tends to over-estimate L in the flat region. The over-estimation is more severe in the nonstationary case, although this over-estimation is mitigated by increasing the sample size. At the same time, Figure 6 shows that the coverage rate of the pre-test statistic is competitive with the naive and infeasible statistics in the flat region. As we move away from the flat point $z = 0$, we find: (i) that the confidence bands of the pre-test statistic grow wider and finally merge with the other three methods; and (ii) that the coverage rate suffers a very small drop and then merges with the naive and infeasible methods. The small drop in coverage rate may be caused by the inferior performance of the derivative significance test in the shape
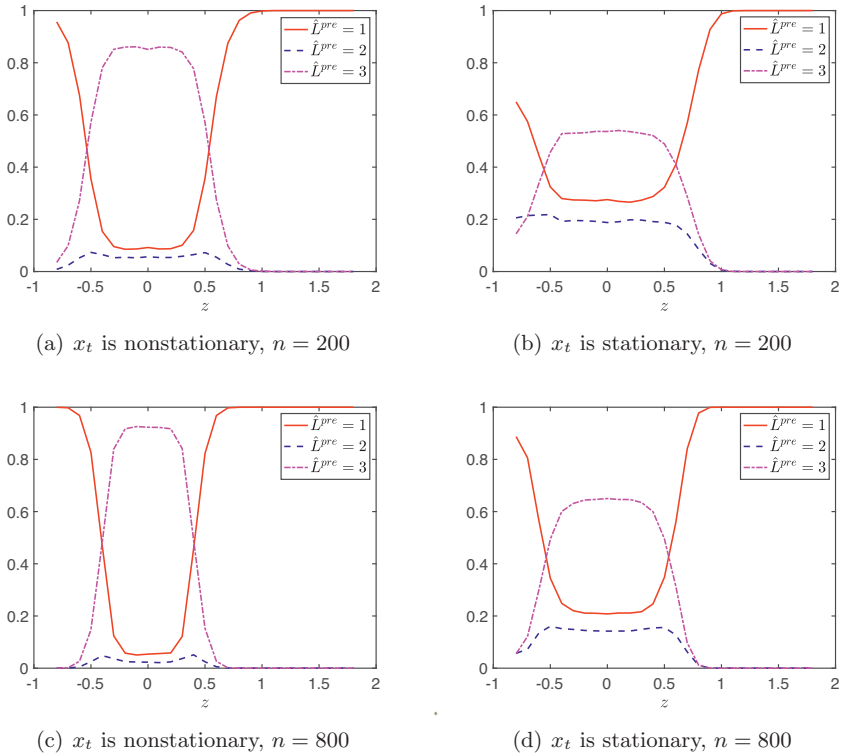
(a) x_t is nonstationary, $n = 200$

(b) x_t is stationary, $n = 200$

(c) x_t is nonstationary, $n = 800$

(d) x_t is stationary, $n = 800$

FIGURE 7. Empirical frequency of the pre-test estimator $\hat{L}^{pre}$ over the support of z_t for $n = 200$ and $n = 800$, from 20,000 replications.

transformation region of the coefficient function. In sum, the pre-test method has much narrower confidence bands around the flat region of the function at the cost of slightly lower coverage rates in the transformation region. In terms of overall performance, these results provide some promising support for practical use of the pre-test method.

5. CONCLUSION

This paper extends existing limit theory in functional coefficient regression to accommodate locally constant coefficients in the regression model (2.1), allowing for both stationary and nonstationary regressors x_t. The findings show that, in the stationary case, the primary effects on the limit theory involve estimation bias, which in turn affects optimal bandwidth choice and optimal convergence rates. In the nonstationary case, both bias and dispersion are affected in the limit theory. As a result, the conditions that separate the limit theory into three different categories are affected by the flatness degree parameter. In particular, both bias and variance depend on the number $(L-1)$ of zero derivatives in the

coefficient function, with consequential effects on optimal bandwidth choice and rates of convergence. In the boundary case where $L \to \infty$ near parametric rates of convergence apply for both stationary and nonstationary cases. In both cases, locally flat functional coefficients make wider bandwidth choices beneficial compared with those implied by standard limit theory. But optimal bandwidth choice is complicated by the fact that bias-variance trade-offs may not correspond to optimal convergence rates and bias correction is more complex due to the locally flat behavior of the coefficient function.

In closing it is worth mentioning that extensions of the type given here are relevant to existing asymptotic theory for nonparametric estimation whenever locally flat functional behavior is present in other models such as probability densities, models with nonstationary regressors that are more complex than $I(1)$ processes and models with time varying parameters. Common practice in the latter models, for instance, is to use weak trend formulations of the parameters, leading to time dependent coefficients of the form $\beta(\frac{t}{n})$. Trend formulations of this type in both stationary and nonstationary systems will lead to asymptotics that involve extensions similar to those developed here, particularly in the stationary regressor case where bias expressions, bias order, and optimal bandwidth choice will all be influenced by flatness in the function. Similarly, in time varying parameter cointegrated systems of the type studied in Phillips et al. (2017), the limit theory will be affected by locally flat regions of the coefficient function. An important simplification in both these cases is that the coefficient function $\beta(\cdot)$ is deterministic, which means that the bias component affects centering but will not contribute directly to variability and the form of the limit distribution, as it can do in models with nonstationary regressors. These are some extensions of the present theory that seem worthy of full investigation in future research.

In all of the above models, any regions of flatness in the function being estimated are typically unknown a priori, including the degree of local flatness, just as the function itself is unknown. Our analysis shows that in such cases the formulas based on standard asymptotics that are used to measure bias and variance in nonparametric estimation are only approximate and rates of convergence may be wrong, especially in cases of nonstationary regressors. After extensive attempts we have found it extremely challenging to devise a direct inference procedure that accurately accommodates information about unknown locally flat characteristics of a functional coefficient. Fortunately, the two-step pre-test procedure is found to work well and can achieve evident efficiency gains at the flat region over the naive approach that simply ignores the local flatness problem. Empirical estimation of the precise degree of local flatness and improved inferential procedures that take account of potential flatness both merit further research.

APPENDIX

This Appendix has four sections. Section A provides proofs of the main results in the paper, Section B contains proofs of useful supporting lemmas, Section C provides additional

computational details, Section D outlines some limit theory associated with testing that is based on local quadratic estimation.

A. PROOFS OF THE THEOREMS

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

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

$$= \sum_{t=1}^{n} \mathbb{E}(x_t x_t' \xi_{\beta t}) + \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E}(x_t x_t' \xi_{\beta t})] + \sum_{t=1}^{n} x_t u_t K_{tz}, \qquad \textbf{(A.1)}$$

with $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz}$. Starting with the kernel-weighted signal matrix, we have

$$\frac{1}{nh}\sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{1}{nh}\sum_{t=1}^{n} \mathbb{E}\left(x_t x_t' K_{tz}\right) + \frac{1}{nh}\sum_{t=1}^{n}[x_t x_t' K_{tz} - \mathbb{E}(x_t x_t' K_{tz})]. \qquad \textbf{(A.2)}$$

Simple calculations give $\mathbb{E}K_{tz} = h \int K(r)f(z+rh)\,dr = hf(z) + O(h^3)$, and $\mathbb{E}K_{tz}^2 = h \int K^2(r) \times f(z+rh)\,dr = hf(z) \int K^2(r)dr + o(h) = hf(z)v_0(K) + o(h)$, where $v_j(K) = \int u^j K^2(u)du$. It follows that $\mathbb{V}ar(K_{tz} - \mathbb{E}K_{tz}) = \mathbb{E}K_{tz}^2 - (\mathbb{E}K_{tz})^2 = O(h)$ and so $K_{tz} - \mathbb{E}K_{tz} = O_p(\sqrt{h})$. We deduce that $\mathbb{V}ar(x_t x_t' K_{tz} - \mathbb{E}(x_t x_t' K_{tz})) = O(h)$ and $x_t x_t' K_{tz} - \mathbb{E}(x_t x_t' K_{tz}) = O_p(\sqrt{h})$. As a result, when $nh \to \infty$, we have $\frac{1}{nh}\sum_{t=1}^{n}[x_t x_t' K_{tz} - \mathbb{E}(x_t x_t' K_{tz})] = O_p(1/\sqrt{nh}) = o_p(1)$ since $\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}[x_t \otimes x_t K_{tz} - \mathbb{E}x_t \otimes x_t K_{tz}] \rightsquigarrow \mathcal{N}\left(0, v_0(K)f(z)\mathbb{E}\left[x_t x_t' \otimes x_t x_t'|z_t = z\right]\right)$ from Lemma B.1(d)(i). Then we can get

$$\frac{1}{nh}\sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{1}{nh}\sum_{t=1}^{n} \mathbb{E}[K_{tz}\mathbb{E}\left(x_t x_t'|z_t\right)] + o_p(1) \to_p \Sigma_{xx}(z)f(z). \qquad \textbf{(A.3)}$$

Next, from the proof of Lemmas B.2(c) and (B.5), we have

$$\mathbb{E}\xi_{\beta t} = h^{L^*+1}C_L(z) + o(h^{L^*+1}),$$

where $L^* = L \times \mathbf{1}_{\{L=\text{even}\}} + (L+1)\mathbf{1}_{\{L=\text{odd}\}}$, $\mu_{L^*}(K) = \mu_L(K) \times \mathbf{1}_{\{L=\text{even}\}} + \mu_{L+1}(K)\mathbf{1}_{\{L=\text{odd}\}}$ and $C_L(z)$ defined in (2.5). Upon normalization and using Lemma B.2(c), the first term in (A.1) is then

$$\frac{1}{nh^{L^*+1}}\sum_{t=1}^{n} \mathbb{E}\left(x_t x_t' \xi_{\beta t}\right) \to_p \Sigma_{xx}(z)C_L(z). \qquad \textbf{(A.4)}$$

The second term of (A.1) is, upon normalization and using Lemma B.2(b)(i),

$$\frac{1}{\sqrt{nh^{2L+1}}}\sum_{t=1}^{n}[x_t x_t' \xi_{\beta t} - \mathbb{E}(x_t x_t' \xi_{\beta t})] \rightsquigarrow \mathcal{N}\left(0, \frac{v_{2L}(K)f(z)}{(L!)^2}\mathbb{E}\left[(x_t' \beta^{(L)}(z))^2 x_t x_t'|z_t = z\right]\right),$$

$$\textbf{(A.5)}$$

provided $nh \to \infty$. Otherwise $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E}(x_t x_t' \xi_{\beta t})] = O_p(1)$ but no central limit theorem holds, as shown in Lemma B.2(b)(ii). The final term of (A.1) is, after suitable normalization and using Lemma B.1(c)(i),

$$\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} \rightsquigarrow \mathcal{N}(0, v_0(K) f(z) \Sigma_{xu}(z)), \tag{A.6}$$

provided $nh \to \infty$. Otherwise from Lemma B.1(c)(ii), $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1)$ but no invariance principle applies.

Standardizing by the weighted signal matrix and recentering (A.1) we have the estimation error decomposition

$$\hat{\beta}(z) - \beta(z) - \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} \mathbb{E} x_t x_t' \xi_{\beta t} = \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E} x_t x_t' \xi_{\beta t}]$$
$$+ \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t u_t K_{tz}, \tag{A.7}$$

or, with each component appropriately standardized, as

$$\hat{\beta}(z) - \beta(z) - h^{L^*} \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{nh^{L^*+1}} \sum_{t=1}^{n} \mathbb{E} x_t x_t' \xi_{\beta t}$$
$$= \sqrt{\frac{h^{2L-1}}{n}} \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E} x_t x_t' \xi_{\beta t}]$$
$$+ \frac{1}{\sqrt{nh}} \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz}. \tag{A.8}$$

Using (A.4), (A.5) and (A.6) in (A.8), we have

$$\sqrt{nh} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{nh^{L^*+1}} \sum_{t=1}^{n} \mathbb{E} x_t x_t' \xi_{\beta t} \right)$$
$$= h^{L} \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E} x_t x_t' \xi_{\beta t}]$$
$$+ \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz}$$
$$= O_p(h^{L}) + \left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} \tag{A.9}$$

$$\leadsto \mathcal{N}\left(0, \frac{v_0(K)}{f(z)} \Sigma_{xx}^{-1}(z) \Sigma_{xu}(z) \Sigma_{xx}^{-1}(z)\right). \tag{A.10}$$

Using (A.3) and (A.4) we have

$$\left(\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{nh^{L^*+1}} \sum_{t=1}^{n} \mathbb{E} x_t x_t' \xi_{\beta t} \to_p \frac{\mathcal{G}_L(z)}{f(z)}, \tag{A.11}$$

which leads to

$$\sqrt{nh}\left(\hat{\beta}(z) - \beta(z) - h^{L^*} \mathcal{B}_L(z) - o(h^{L^*})\right) \leadsto \mathcal{N}\left(0, \frac{v_0(K)}{f(z)} \Sigma_{xx}^{-1}(z) \Sigma_{xu}(z) \Sigma_{xx}^{-1}(z)\right) \tag{A.12}$$

where $\mathcal{B}_L(z) = \frac{\mathcal{G}_L(z)}{f(z)}$, giving the stated result for the first part, which holds whenever $nh \to \infty$ ensuring the central limit theorem (A.12).

In cases where $nh \to c \in [0, \infty)$, in view of Lemmas B.1(c)(ii) and B.2(b)(ii), we still have $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1)$ and $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E} x_t x_t' \xi_{\beta t}] = O_p(1)$ although no invariance principle holds. Further, in view of Lemma B.2(c) we have $\sum_{t=1}^{n} \mathbb{E} x_t x_t' \xi_{\beta t} = O_p(nh^{L^*+1})$. But the signal matrix $\sum_{t=1}^{n} x_t x_t' K_{tz} = O_p(\sqrt{nh})$ and so the last term in (A.7) is $O_p(1)$. Therefore $\hat{\beta}(z)$ is inconsistent when $nh \to c \in [0, \infty)$. $\qquad \square$

Proof of Theorem 2.2.
Case (i) We start with the decomposition

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

$$= \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t x_t' \eta_t + \left(\sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \sum_{t=1}^{n} x_t u_t K_{tz}, \tag{A.13}$$

which can be obtained in the same way as (A.7). Rescale the components according to their asymptotic behavior, as determined in Lemma B.3, so that

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

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

$$+ \frac{1}{n\sqrt{h}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz}\right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz}. \tag{A.14}$$

Then, since $nh^{2L} \to 0$ in this case, we rescale the equation by $n\sqrt{h}$, giving

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

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

$$
+ \left(\frac{1}{n^2 h}\sum_{t=1}^{n}x_t x_t' K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_t u_t K_{tz}
$$

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

$$
\rightsquigarrow \left(f(z)\int B_x B_x'\right)^{-1}\left(\int B_x dB_{uK}\right) \equiv_d \mathcal{MN}\left(0, \frac{v_0(K)\sigma_u^2}{f(z)}\left(\int B_x B_x'\right)^{-1}\right),
$$
(A.15)

the mixed normality following from Lemma B.3(d)(i). In view of Lemmas B.3(b) and B.3(c)(i) we have

$$
\left(\frac{1}{n^2 h}\sum_{t=1}^{n}x_t x_t' K_{tz}\right)^{-1}\frac{1}{n^2 h^{L^*+1}}\sum_{t=1}^{n}x_t x_t' \mathbb{E}\xi_{\beta t} \to_p \frac{\mathcal{G}_L(z)}{f(z)},
$$
(A.16)

for the bias function. Hence,

$$
n\sqrt{h}\left(\hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z) - o(h^{L^*})\right) \rightsquigarrow \mathcal{MN}(0, \Omega_{NS}(z)),
$$
(A.17)

with $\Omega_{NS}(z) = \frac{v_0(K)\sigma_u^2}{f(z)}\left(\int B_x B_x'\right)^{-1}$, as given in the stated result (2.8) for case (i).

Case (ii)

When $nh^{2L} \to \infty$ the bandwidth goes to zero slower than $O(\frac{1}{\sqrt{n^{1/2L}}})$. To derive the limit theory in this case, rescale (A.14) by $\sqrt{n/h^{2L-1}}$, giving

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

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

$$
+ \frac{1}{\sqrt{nh^{2L}}}\left(\frac{1}{n^2 h}\sum_{t=1}^{n}x_t x_t' K_{tz}\right)^{-1}\frac{1}{n\sqrt{h}}\sum_{t=1}^{n}x_t u_t K_{tz}
$$

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

$$
\rightsquigarrow \left(f(z) \int B_x B_x' \right)^{-1} \left(\int B_x B_x' dB_{\eta,L} \right) \tag{A.18}
$$

$$
\equiv_d \ \mathcal{MN} \left(0, \frac{v_{2L}(K)}{f(z)(L!)^2} \left(\int B_x B_x' \right)^{-1} \int B_x B_x' \left(B_x' \beta^{(L)}(z) \right)^2 \left(\int B_x B_x' \right)^{-1} \right), \tag{A.19}
$$

using Lemma B.3(a) and (c), where $B_{\eta,L}$ is Brownian motion with variance matrix $V_{\eta\eta,L} = \frac{v_{2L}(K)f(z)}{(L!)^2} \beta^{(L)}(z)\beta^{(L)}(z)'$.

Since $B_{\eta,L}$ is singular Brownian motion whenever $p > 1$ we may write the inner product $B_x(r)'B_{\eta,L}(r)$ in the equivalent form $B_x(r)'B_{\eta,L}(r) = \left(B_x(r)'\beta^{(L)}(z) \right) B_{f,L}(r)$, where $B_{f,L}$ is scalar Brownian motion with variance $\frac{v_{2L}(K)f(z)}{(L!)^2}$. Then in view of the independence of B_x and $B_{\eta,L}$ we have

$$
\int B_x B_x' dB_{\eta,L} \equiv_d \mathcal{MN} \left(0, \frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x' \left(B_x' \beta^{(L)}(z) \right)^2 \right), \tag{A.20}
$$

which leads to the mixed normal limit distribution given in (A.19). Combining this result with the bias function evaluation obtained earlier in (A.16) yields

$$
\sqrt{\frac{n}{h^{2L-1}}} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \mathcal{B}_L(z) - o(h^{L^*}) \right) \rightsquigarrow \left(f(z) \int B_x B_x' \right)^{-1} \left(\int B_x B_x' dB_{\eta,L} \right)
$$

$$
\equiv_d \mathcal{MN} \left(0, \Omega_L(z) \right), \tag{A.21}
$$

where $\Omega_L(z) = \frac{v_{2L}(K)f(z)}{(L!)^2} \left(\int B_x B_x' \right)^{-1} \int B_x B_x' \left(B_x' \beta^{(L)}(z) \right)^2 \left(\int B_x B_x' \right)^{-1}$, giving the stated result (ii) of Theorem 2.2.

Case (iii)

Since $nh^{2L} \to c$ for some constant $c \in (0, \infty)$, $h \sim_a (c/n)^{\frac{1}{2L}}$ and then $\sqrt{n/h^{2L-1}} = O(\sqrt{n^{1+\frac{2L-1}{2L}}}) = O(n^{1-\frac{1}{4L}}) = n\sqrt{h}$. It follows that the first and second terms on the right side of (A.14) have the same order and both therefore appear to contribute to the asymptotics. So, upon rescaling (A.14) by $n^{1-\frac{1}{4L}}$ we find that

$$
n^{1-\frac{1}{4L}} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^2 h^{L^*+1}} \sum_{t=1}^{n} x_t x_t' \mathbb{E} \xi_{\beta t} \right)
$$

$$
= (nh^{2L})^{\frac{1}{2}-\frac{1}{4L}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^{2L+1}}} \sum_{t=1}^{n} x_t x_t' \eta_t
$$

$$+ \frac{1}{(nh^{2L})^{\frac{1}{4L}}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz},$$

$$= c^{\frac{1}{2} - \frac{1}{4L}} \times \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{\sqrt{n^3 h^{2L+1}}} \sum_{t=1}^{n} x_t x_t' \eta_t$$

$$+ c^{-\frac{1}{4L}} \left(\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} \right)^{-1} \frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz}. \tag{A.22}$$

The asympototics are then jointly determined by the two terms of (A.22). Conditional on $\mathcal{F}_x$, these terms are uncorrelated as the conditional covariance involves the matrix

$$\mathbb{E} \left(\frac{1}{\sqrt{n^3 h^{2L+1}}} \sum_{t=1}^{n} x_t x_t' \eta_t \right) \left(\frac{1}{n\sqrt{h}} \sum_{t=1}^{n} x_t u_t K_{tz} \right)' = \frac{1}{\sqrt{n^5 h^{L+1}}} \sum_{t,s=1}^{n} \mathbb{E} \left(x_t x_s' (x_t' \eta_t u_s K_{sz}) \right) = 0. \tag{A.23}$$

From (A.22) and the bias function calculation (A.16) which continues to hold, it follows that when $nh^{2L} \to c > 0$

$$n^{1 - \frac{1}{4L}} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \mathcal{B}_L(z) - o(h^{L^*}) \right)$$

$$\sim_a c^{\frac{1}{2} - \frac{1}{4L}} \times \left(f(z) \int B_x B_x' \right)^{-1} \left(\int B_x B_x' dB_{\eta, L} \right) + c^{-\frac{1}{4L}} \left(f(z) \int B_x B_x' \right)^{-1} \left(\int B_x dB_{uK} \right)$$

$$\equiv_d c^{\frac{1}{2} - \frac{1}{4L}} \times \mathcal{MN}(0, \Omega_L(z)) + c^{-\frac{1}{4L}} \mathcal{MN}(0, \Omega_{NS}(z))$$

$$= \mathcal{MN} \left(0, c^{1 - \frac{1}{2L}} \Omega_L(z) + c^{-\frac{1}{2L}} \Omega_{NS}(z) \right). \tag{A.24}$$

This proves result (iii) of Theorem 2.2.   □

Proof of Theorem 3.1.
(i) Stationary x_t We assume L is known. Using Lemma B.1, $\hat{\sigma}_u^2 \to_p \sigma_u^2$, and any consistent derivative estimator $\hat{\beta}^{(L)}(z)$ of $\beta^{(L)}(z)$, we have

$$\hat{\Omega}_n(z; L) = v_0(K) \hat{\sigma}_u^2 \sum_{t=1}^{n} x_t x_t' K_{tz} + \sum_{t=1}^{n} x_t x_t' \left\{ x_t' \frac{1}{L!} \hat{\beta}^{(L)}(z)(z_t - z)^L K_{tz} \right\}^2$$

$$\sim_a nh \left(\frac{v_0(K) \sigma_u^2}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz} + \frac{1}{nh} \sum_{t=1}^{n} x_t x_t' \left\{ x_t' \frac{\beta^{(L)}(z)}{L!} (z_t - z)^L K_{tz} \right\}^2 \right)$$

$$\sim_a nh \left(v_0(K) \sigma_u^2 f(z) \Sigma_{xx} + \frac{1}{h} \mathbb{E} \left[x_t x_t' \left\{ x_t' \frac{\beta^{(L)}(z)}{L!} (z_t - z)^L K_{tz} \right\}^2 \right] \right)$$

$$\sim_a nh \left(v_0(K) \sigma_u^2 f(z) \Sigma_{xx} + \frac{1}{h} \mathbb{E} \left[x_t x_t' \left(x_t' \frac{\beta^{(L)}(z)}{L!} \right)^2 \right] \int \left\{ (z_t - z)^L K \left(\frac{z_t - z}{h} \right) \right\}^2 f(z_t) \, dz_t \right)$$

$$\sim_a nh \left(v_0(K)\sigma_u^2 f(z) \Sigma_{xx} + \mathbb{E}\left[x_t x_t' \left(x_t' \frac{\beta^{(L)}(z)}{L!} \right)^2 \right] \int \left\{ (sh)^L K(s) \right\}^2 f(z+sh)\, ds \right)$$

$$\sim_a nh \left(v_0(K)\sigma_u^2 f(z) \Sigma_{xx} + h^{2L} \frac{v_{2L}(K)f(z)}{(L!)^2} \mathbb{E}\left[x_t x_t' \left(x_t' \beta^{(L)}(z) \right)^2 \right] \right)$$

$$\sim_a nh v_0(K)\sigma_u^2 f(z) \Sigma_{xx}. \tag{A.25}$$

Then

$$nh\hat{V}_n(z;L) = \left[\frac{1}{nh} A_n(z) \right]^{-1} \left[\frac{1}{nh} \hat{\Omega}_n(z;L) \right] \left[\frac{1}{nh} A_n(z) \right]^{-1}$$

$$\to_p [f(z)\Sigma_{xx}]^{-1} \left[v_0(K)\sigma_u^2 f(z)\Sigma_{xx} \right] [f(z)\Sigma_{xx}]^{-1} = \frac{v_0(K)\sigma_u^2}{f(z)} \Sigma_{xx}^{-1} = \Omega_S(z). \tag{A.26}$$

Combining (A.26) and Theorem 2.1 gives

$$\hat{T}(z;L) = \hat{V}_n(z;L)^{-1/2} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \hat{\mathcal{B}}_L(z) \right)$$

$$\sim_a [\Omega_S(z)]^{-1/2} \sqrt{nh} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \mathcal{B}_L(z) \right) \rightsquigarrow \mathcal{N}\left(0, I_p \right),$$

and $\hat{T}_2(z;L) \rightsquigarrow \chi_p^2$ follows. Further, in view of (A.25) in the stationary case, the simpler estimate $\tilde{\Omega}_n(z;L) = v_0(K)\hat{\sigma}_u^2 \sum_{t=1}^n x_t x_t' K_{tz}$, which is based solely on the variance term, can be employed and the same limit theory applies.

(ii) Nonstationary x_t We again assume that L is known. We analyze each case of the Theorem in turn.

Case (a) Using Lemma B.3(c) we have $\frac{1}{n^2 h} \sum_{t=1}^n x_t x_t' K_{tz} \rightsquigarrow f(z) \int B_x B_x'$. In place of (A.25) and again using a consistent derivative estimator $\hat{\beta}^{(L)}(z) \to_p \beta^{(L)}(z)$ and $\hat{\sigma}_u^2 \to_p \sigma_u^2$, we now have

$$\hat{\Omega}_n(z;L) = v_0(K)\hat{\sigma}_u^2 \sum_{t=1}^n x_t x_t' K_{tz} + \sum_{t=1}^n x_t x_t' \left\{ x_t' \frac{1}{L!} \hat{\beta}^{(L)}(z)(z_t - z)^L K_{tz} \right\}^2$$

$$\sim_a n^2 h \left(\frac{v_0(K)\sigma_u^2}{n^2 h} \sum_{t=1}^n x_t x_t' K\left(\frac{z_t - z}{h} \right) + \frac{h^{2L}}{n^2 h} \sum_{t=1}^n x_t x_t' \left\{ x_t' \frac{\beta^{(L)}(z)}{L!} \left(\frac{z_t - z}{h} \right)^L K\left(\frac{z_t - z}{h} \right) \right\}^2 \right)$$

$$\sim_a n^2 h \left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + nh^{2L-1} \left[\frac{1}{n} \sum_{t=1}^n \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \left\{ \frac{x_t'}{\sqrt{n}} \frac{\beta^{(L)}(z)}{L!} \left(\frac{z_t - z}{h} \right)^L K\left(\frac{z_t - z}{h} \right) \right\}^2 \right] \right)$$

$$\sim_a n^2 h \left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + nh^{2L-1} \left[\int B_x B_x \left\{ B_x' \frac{\beta^{(L)}(z)}{L!} \right\}^2 \mathbb{E}\left\{ \left(\frac{z_t - z}{h} \right)^{2L} K\left(\frac{z_t - z}{h} \right)^2 \right\} \right] \right)$$

$$\sim_a n^2 h \left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + nh^{2L-1} \left[\int B_x B_x \left\{ B_x' \frac{\beta^{(L)}(z)}{L!} \right\}^2 \int p^{2L} K(p)^2 f(z+ph)\, dp h \right] \right)$$

$$\sim_a n^2 h \left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + nh^{2L} \frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{ B_x' \beta^{(L)}(z) \right\}^2 \right) \tag{A.27}$$

$$\sim_a n^2 h v_0(K)\sigma_u^2 f(z) \int B_x B_x' \quad \text{when } nh^{2L} \to 0. \tag{A.28}$$

It follows that in this case

$$
\begin{aligned}
n^2 h \hat{V}_n(z;L) &= \left[\frac{1}{n^2 h} A_n(z)\right]^{-1} \left[\frac{1}{n^2 h} \hat{\Omega}_n(z;L)\right] \left[\frac{1}{n^2 h} A_n(z)\right]^{-1} \\
&\leadsto \left[f(z) \int B_x B_x'\right]^{-1} \left[v_0(K)\sigma_u^2 f(z) \int B_x B_x'\right] \left[f(z) \int B_x B_x'\right]^{-1} \\
&= \frac{v_0(K)\sigma_u^2}{f(z)} \left(\int B_x B_x'\right)^{-1} = \Omega_{NS}(z),
\end{aligned}
\tag{A.29}
$$

from which we deduce from Theorem 2.2(i) that

$$
\begin{aligned}
\hat{T}(z;L) &= \hat{V}_n(z;L)^{-1/2} \left(\hat{\beta}(z) - \beta(z) - h^{L^*}\hat{\mathcal{B}}_L(z)\right) \\
&\sim_a [\Omega_{NS}(z)]^{-1/2} n\sqrt{h}\left(\hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z)\right) \leadsto \mathcal{N}\left(0, I_p\right).
\end{aligned}
$$

Then $\hat{T}_2(z;L) \leadsto \chi_p^2$, as required.

Case (b) When $nh^{2L} \to \infty$ we have by calculations similar to those leading to (A.27)

$$
\begin{aligned}
\hat{\Omega}_n(z;L) &= v_0(K)\hat{\sigma}_u^2 \sum_{t=1}^{n} x_t x_t' K_{tz} + \sum_{t=1}^{n} x_t x_t' \left\{x_t' \frac{1}{L!}\hat{\beta}^{(L)}(z)(z_t - z)^L K_{tz}\right\}^2 \\
&\sim_a n^2 h\left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + nh^{2L}\frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{B_x'\beta^{(L)}(z)\right\}^2\right) \\
&= n^3 h^{2L+1}\left(\frac{1}{nh^{2L}} v_0(K)\sigma_u^2 f(z) \int B_x B_x' + \frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{B_x'\beta^{(L)}(z)\right\}^2\right) \\
&\sim_a n^3 h^{2L+1}\frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{B_x'\beta^{(L)}(z)\right\}^2.
\end{aligned}
$$

Hence, in this case

$$
\begin{aligned}
\frac{n}{h^{2L-1}}\hat{V}_n(z;L) &= \left[\frac{1}{n^2 h}A_n(z)\right]^{-1} \left[\frac{1}{n^3 h^{2L+1}}\hat{\Omega}_n(z;L)\right] \left[\frac{1}{n^2 h}A_n(z)\right]^{-1} \\
&\sim_a \left[f(z)\int B_x B_x'\right]^{-1} \left[\frac{v_{2L}(K)f(z)}{(L!)^2}\int B_x B_x\left\{B_x'\beta^{(L)}(z)\right\}^2\right]\left[f(z)\int B_x B_x'\right]^{-1} \\
&= \frac{v_{2L}(K)}{(L!)^2 f(z)}\left(\int B_x B_x'\right)^{-1}\left(\int B_x B_x\left\{B_x'\beta^{(L)}(z)\right\}^2\right)\left(\int B_x B_x'\right)^{-1} = \Omega_L(z),
\end{aligned}
\tag{A.30}
$$

and from Theorem 2.2(ii) we deduce that

$$
\begin{aligned}
\hat{T}(z;L) &= \hat{V}_n(z;L)^{-1/2}\left(\hat{\beta}(z) - \beta(z) - h^{L^*}\hat{\mathcal{B}}_L(z)\right) \\
&= \left[\frac{n}{h^{2L-1}}\hat{V}_n(z;L)\right]^{-1/2}\sqrt{\frac{n}{h^{2L-1}}}\left(\hat{\beta}(z) - \beta(z) - h^{L^*}\hat{\mathcal{B}}_L(z;L)\right) \\
&\sim_a [\Omega_L(z)]^{-1/2}\sqrt{\frac{n}{h^{2L-1}}}\left(\hat{\beta}(z) - \beta(z) - h^{L^*}\mathcal{B}_L(z)\right) \leadsto \mathcal{N}\left(0, I_p\right),
\end{aligned}
$$

and $\hat{T}_2(z;L) \leadsto \chi_p^2$, as required.

Case (c) When $nh^{2L} \to c \in (0,\infty)$ we have by calculations similar to those leading to (A.27)

$$\hat{\Omega}_n(z;L) = v_0(K)\hat{\sigma}_u^2 \sum_{t=1}^n x_t x_t' K_{tz} + \sum_{t=1}^n x_t x_t' \left\{ x_t' \frac{1}{L!} \hat{\beta}^{(L)}(z)(z_t - z)^L K_{tz} \right\}^2$$

$$\sim_a n^2 h \left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + nh^{2L} \frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{ B_x' \beta^{(L)}(z) \right\}^2 \right)$$

$$\sim_a n^2 h \left(v_0(K)\sigma_u^2 f(z) \int B_x B_x' + c \frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{ B_x' \beta^{(L)}(z) \right\}^2 \right).$$

Then

$$n^2 h \hat{V}_n(z;L) = \left[\frac{1}{n^2 h} A_n(z) \right]^{-1} \left[\frac{1}{n^2 h} \hat{\Omega}_n(z;L) \right] \left[\frac{1}{n^2 h} A_n(z) \right]^{-1}$$

$$\sim_a \left[f(z) \int B_x B_x' \right]^{-1} \left[v_0(K)\sigma_u^2 f(z) \int B_x B_x' \right.$$

$$\left. + c \frac{v_{2L}(K)f(z)}{(L!)^2} \int B_x B_x \left\{ B_x' \beta^{(L)}(z) \right\}^2 \right] \left[f(z) \int B_x B_x' \right]^{-1}$$

$$= \Omega_{NS}(z) + c\Omega_L(z). \tag{A.31}$$

Hence, when $nh^{2L} \to c$ or $h \sim_a (c/n)^{\frac{1}{2L}}$ and $n^2 h \sim_a c^{\frac{1}{2L}} n^{2-\frac{1}{2L}}$ as $n \to \infty$ we have

$$n^{2-\frac{1}{2L}} \hat{V}_n(z;L) \rightsquigarrow c^{-\frac{1}{2L}} \Omega_{NS}(z) + c^{1-\frac{1}{2L}} \Omega_L(z).$$

From Theorem 2.2(iii) it now follows that when $nh^{2L} \to c$ we have

$$\hat{T}(z;L) = \hat{V}_n(z;L)^{-1/2} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \hat{\mathcal{B}}_L(z) \right)$$

$$= \left[n^{2-\frac{1}{2L}} \hat{V}_n(z;L) \right]^{-1/2} n^{1-\frac{1}{4L}} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \hat{\mathcal{B}}_L(z) \right)$$

$$\sim_a [c^{-\frac{1}{2L}} \Omega_{NS}(z) + c^{1-\frac{1}{2L}} \Omega_L(z)]^{-1/2} n^{1-\frac{1}{4L}} \left(\hat{\beta}(z) - \beta(z) - h^{L^*} \mathcal{B}_L(z) \right) \rightsquigarrow \mathcal{N}\left(0, I_p \right),$$

and again $\hat{T}_2(z;L) \rightsquigarrow \chi_p^2$, as required. $\qquad\square$

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 $v_0(K)f(z)$, and $v_0(K)\sigma_u^2(z)f(z)$, with $\zeta_{tK} = K_{tz} - \mathbb{E}K_{tz}$, $K_{tz} = K(\frac{z_t - z}{h})$, and $\sigma_u^2(z) = \mathbb{E}(u_t^2 | z_t = z)$;*

(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 $nh \to \infty$, $\frac{1}{nh} \sum_{t=1}^n x_t x_t' K_{tz} \to_p \Sigma_{xx}(z) f(z)$;*

(ii) If $nh \to c \in [0,\infty)$, $\frac{1}{\sqrt{nh}} \sum_{t=1}^n x_t x_t' K_{tz} = O_p(1)$ but no invariance principle holds.

(c) (i) If $nh \to \infty$, $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} \leadsto \mathcal{N}(0, v_0(K)f(z)\Sigma_{xu}(z))$;

    (ii) If $nh \to c \in [0, \infty)$, $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1)$ but no invariance principle holds.

(d) (i) If $nh \to \infty$, $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n}[x_t \otimes x_t K_{tz} - \mathbb{E}(x_t \otimes x_t K_{tz})] \leadsto \mathcal{N}(0, v_0(K)f(z)$
$\mathbb{E}[x_t x_t' \otimes x_t x_t' | z_t = z])$;

    (ii) If $nh \to c \in [0, \infty)$, $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n}[x_t \otimes x_t K_{tz} - \mathbb{E}(x_t \otimes x_t K_{tz})] = O_p(1)$ but no invariance principle holds.

Proof of Lemma B.1.

Part (a) (i) and (ii) See Lemma B.1(a) of Phillips and Wang (2021). For later use, note that $\mathbb{E}K_{tz} = hf(z) + o(h)$, and $\mathbb{E}K_{tz}^2 = hf(z)v_0(K) + o(h)$, so that $\mathbb{V}ar(\zeta_{tK}) = hf(z)v_0(K) + o(h)$ and $\zeta_{tK} = K_{tz} - \mathbb{E}(K_{tz}) = O_p(\sqrt{h})$. Further, $\mathbb{V}ar(u_t K_{tz}) = hv_0(K)\sigma_u^2(z)f(z) + o(h)$ and $\mathbb{E}u_t K_{tz}^2 = 0$, so that the limit processes $(B_{\zeta K}(r), B_{uK}(r))$ are independent. The functional laws follow by standard weak convergence methods when $nh \to \infty$ and (ii) follows by showing the $O_p(1)$ property directly whereas the CLT does not hold because of failure of the Lindeberg condition, just as in the proof of Phillips and Wang (2021, Lem. B.1(a) (ii)).

Part (b) (i) We have

$$\frac{1}{nh} \sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{1}{n} \sum_{t=1}^{n} \frac{\mathbb{E}(x_t x_t' K_{tz})}{h} + \frac{1}{nh} \sum_{t=1}^{n}[x_t x_t' K_{tz} - \mathbb{E}(x_t x_t' K_{tz})]$$

$$= \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[h^{-1} K_{tz}\mathbb{E}(x_t x_t' | z_t)] + O_p\left(\frac{1}{\sqrt{nh}}\right) \to_p \Sigma_{xx}(z)f(z), \qquad \text{(B.1)}$$

as in (A.3) by virtue of Lemma B.1(d)(i) and the law of large numbers. For (ii), when $nh \to c \in [0, \infty)$ we have $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n}[x_t x_t' K_{tz} - \mathbb{E}(x_t x_t' K_{tz})] = O_p(1)$ as in Lemma B.1(d)(ii) with no invariance principle holding. Then, as in (B.1), $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t x_t' K_{tz} = \sqrt{nh} \times O_p(1) + O_p(1) = O_p(1)$, as stated.

Part (c) Result (i) follows by the central limit theorem for α-mixing processes. Stability holds because the conditional variance matrix is

$$\mathbb{V}ar\left\langle \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} x_t u_t K_{tz} \right\rangle \to_p v_0(K)f(z)\mathbb{E}[x_t x_t' \sigma^2(x_t, z_t) | z_t = z] \equiv v_0(K)f(z)\Sigma_{xu}(z). \qquad \text{(B.2)}$$

Setting $w_{tK} = x_t u_t K_{tz}$ and noting that $\mathbb{E}\left\| \frac{w_{tK}}{\sqrt{h}} \right\|^2 < \infty$ we have, given $\epsilon > 0$,

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left\{ \left\| \frac{w_{tK}}{\sqrt{h}} \right\|^2 \mathbf{1}_{\left[\|w_{tK}\| > \epsilon\sqrt{nh} \right]} \right\} \to 0 \qquad \text{(B.3)}$$

and the Lindeberg condition holds when $nh \to \infty$, giving the stated result. Part (ii) follows because, although the stability condition continues to hold as in (B.2), the Lindeberg condition fails when $nh \to c \in [0, \infty)$ as (B.3) no longer tends to zero. To see this, it suffices to look at the simple case with scalar x_t, $iid\{(u_t, z_t)\}$ and independent, strictly stationary components with respective densities $\{f_x(x), f_u(u), f(s)\}$ we have, given $\epsilon > 0$

and $nh \to c \in [0, \infty)$

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \left\{ \left(\frac{w_{tK}}{\sqrt{h}} \right)^2 1_{\left[|w_{tK}| > \epsilon \sqrt{nh} \right]} \right\}$$

$$= \frac{1}{h} \int x^2 u^2 K \left(\frac{s-z}{h} \right)^2 f_x(x) f_u(u) f(s) 1_{\left[\left| xuK\left(\frac{s-z}{h}\right) \right| > \epsilon \sqrt{nh} \right]} dx\,du\,ds$$

$$= \int x^2 u^2 K(p)^2 f_x(x) f_u(u) f(z+ph) 1_{\left[|xuK(p)| > \epsilon \sqrt{nh} \right]} dx\,du\,dp$$

$$\to \begin{cases} v_0(K) f(z) \mathbb{E}(x_t^2 u_t^2) > 0 & \text{if } nh \to 0 \\ \int x^2 u^2 K(p)^2 f_x(x) f_u(u) f(z) 1_{\left[|xuK(p)| > \epsilon \sqrt{c} \right]} dx\,du\,dp > 0 & \text{if } nh \to c \in (0, \infty) \end{cases},$$

leading to failure in the Lindeberg condition.

Part (d) (i) and (ii) of Part (d) follow in the same way as (i) and (ii) of Part (c). □

LEMMA B.2. *Under Assumption 1, if $\beta^{(\ell)}(z) = 0$ for $\ell = 0, 1, \cdots, L-1$ and $\beta^{(L)}(z) \neq 0$, then the following hold as $n \to \infty$ and $h \to 0$:*

(a) *(i) If $nh \to \infty$, $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{\lfloor n \cdot \rfloor} \eta_t \rightsquigarrow B_{\eta, L}(\cdot)$, where $B_{\eta, L}(\cdot)$ is Brownian motion with variance matrix $V_{\eta\eta, L} = \frac{v_{2L}(K)f(z)}{(L!)^2} \beta^{(L)}(z) \beta^{(L)}(z)'$, with $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$ and $\xi_{\beta t} = [\beta(z_t) - \beta(z)] K_{tz}$;*
(ii) If $nh \to c \in [0, \infty)$, then $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{\lfloor n \cdot \rfloor} \eta_t = O_p(1)$, but no invariance principle holds.

(b) *(i) If $nh \to \infty$, $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E} x_t x_t' \xi_{\beta t}] \rightsquigarrow \mathcal{N} \left(0, \frac{v_{2L}(K)f(z)}{(L!)^2} \right.$
$\left. \mathbb{E} \left[(x_t' \beta^{(L)}(z))^2 x_t x_t' | z_t = z \right] \right)$;*
(ii) If $nh \to c \in [0, \infty)$, $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} [x_t x_t' \xi_{\beta t} - \mathbb{E} x_t x_t' \xi_{\beta t}] = O_p(1)$, but no invariance principle holds.

(c) *$\frac{1}{nh^{L^*+1}} \sum_{t=1}^{n} \mathbb{E} x_t x_t' \xi_{\beta t} \to_p \mathcal{G}_L(z) \Sigma_{xx}(z)$, where $L^* = L \times 1_{\{L=even\}} + (L+1) 1_{\{L=odd\}}$, $\mathcal{G}_L(z) = \mu_{L^*}(K) C_L(z)$ with $\mu_{L^*}(K) = \mu_L(K) 1_{\{L=even\}} + \mu_{L+1}(K) 1_{\{L=odd\}}$, and*

$$C_L(z) = \frac{f(z)\beta^{(L)}(z)}{L!} 1_{\{L=even\}} + \left[\frac{\beta^{(L)}(z)}{L!} f^{(1)}(z) + \frac{\beta^{(L+1)}(z)}{(L+1)!} f(z) \right] 1_{\{L=odd\}}. \quad \textbf{(B.4)}$$

Proof of Lemma B.2. Part (a) (i) Phillips and Wang (2021) proved a related result when $L = 1$ in their Lemma B.1(b). A similar argument is employed here. But for general L we need to compute the first and second moments of $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$ and deal with the precise form of the local behavior of the coefficient function $\beta(\cdot)$ in the neighborhood of the point of estimation z. To this end, the proof uses the following Taylor representations

$$\beta(z+ph) - \beta(z) = \beta^{(L)}(z) \frac{p^L h^L}{L!} + \beta^{(L+1)}(\tilde{z}_p) \frac{p^{L+1} h^{L+1}}{(L+1)!}$$

$$f(z+ph) = f(z) + f^{(1)}(\breve{z}_p) ph,$$

where $\left\{ \beta^{(j)}(z) = 0; j = 1, \ldots, L-1 \right\}$ and with $\tilde{z}_p$ and $\breve{z}_p$ on the line segment connecting z_t and z. The first and second moments of η_t may now be deduced. Specifically, using the

symmetry of K, we have

$$
\mathbb{E}\xi_{\beta t} = \mathbb{E}[\beta(z_t) - \beta(z)]K_{tz} = \int_{-1}^{1}[\beta(s) - \beta(z)]K((s-z)/h)f(s)ds
$$

$$
= h\int_{-1}^{1}[\beta(z+ph) - \beta(z)]K(p)f(z+ph)dp
$$

$$
= h\int_{-1}^{1}\left[\frac{h^L p^L}{L!}\beta^{(L)}(z) + \frac{h^{L+1}p^{L+1}}{(L+1)!}\beta^{(L+1)}(\check{z}_p)\right]K(p)\left[f(z) + f^{(1)}(\check{z}_p)ph\right]dp
$$

$$
= \left\{h\int_{-1}^{1}\frac{h^L}{L!}\beta^{(L)}(z)f(z)p^L K(p)dp + o(h^{L+1})\right\}\times\mathbf{1}_{\{L=\text{even}\}}
$$

$$
+ \left\{h\int_{-1}^{1}\left[\frac{h^{L+1}}{L!}\beta^{(L)}(z)f^{(1)}(\check{z}_p) + \frac{h^{L+1}}{(L+1)!}\beta^{(L+1)}(\check{z}_p)f(z)\right]p^{L+1}K(p)dp + o(h^{L+2})\right\}
$$

$$
\times\mathbf{1}_{\{L=\text{odd}\}}
$$

$$
= \left\{h^{L+1}\mu_L(K)\frac{\beta^{(L)}(z)}{L!}f(z) + o(h^{L+1})\right\}\times\mathbf{1}_{\{L=\text{even}\}}
$$

$$
+ \left\{h^{L+2}\mu_{L+1}(K)\left[\frac{\beta^{(L)}(z)}{L!}f^{(1)}(z) + \frac{\beta^{(L+1)}(z)}{(L+1)!}f(z) + o(1)\right] + o(h^{L+2})\right\}\times\mathbf{1}_{\{L=\text{odd}\}}
$$

$$
= h^{L^*+1}\mu_{L^*}(K)C_L(z) + o(h^{L^*+1}) =: h^{L^*+1}\mathcal{G}_L(z) + o(h^{L^*+1}), \tag{B.5}
$$

where L^*, $\mu_{L^*}(K)$, $\mathcal{G}_L(z) = \mu_{L^*}(K)C_L(z)$, and $C_L(z)$ are defined in the statement of the Lemma. Next

$$
\mathbb{E}\xi_{\beta t}\xi'_{\beta t} = \mathbb{E}[(\beta(z_t) - \beta(z))(\beta(z_t) - \beta(z))'K((z_t - z)/h)^2]
$$

$$
= h\int_{-1}^{1}(\beta(z+hs) - \beta(z))(\beta(z+hs) - \beta(z))'K(s)^2 f(z+hs)ds
$$

$$
= h\int_{-1}^{1}\left[\frac{h^L s^L}{L!}\beta^{(L)}(z) + o\left(h^L\right)\right]\left[\frac{h^L s^L}{L!}\beta^{(L)}(z) + o\left(h^L\right)\right]' K(s)^2[f(z) + o(1)]ds
$$

$$
= \frac{h^{2L+1}}{(L!)^2}f(z)\beta^{(L)}(z)\beta^{(L)}(z)'\int_{-1}^{1}s^{2L}K^2(s)ds + o(h^{2L+1})
$$

$$
= \frac{h^{2L+1}}{(L!)^2}v_{2L}(K)f(z)\beta^{(L)}(z)\beta^{(L)}(z)' + o(h^{2L+1}).
$$

It follows that

$$
\mathbb{V}ar(\eta_t) = \mathbb{E}\xi_{\beta t}\xi'_{\beta t} - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})' = \frac{h^{2L+1}}{(L!)^2}v_{2L}(K)f(z)\beta^{(L)}(z)\beta^{(L)}(z)' + o(h^{2L+1}), \tag{B.6}
$$

and $\eta_t = O_p\left(h^{L+\frac{1}{2}}\right)$. Next, in view of (B.5) the serial covariances satisfy

$$
\mathbb{C}ov(\xi_{\beta t}, \xi_{\beta t+j}) = \mathbb{E}\xi_{\beta t}\xi'_{\beta t+j} - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})' = \mathbb{E}\xi_{\beta t}\xi'_{\beta t+j} + O(h^{2L^*+2})
$$

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

$$|\mathbb{C}ov(\xi_{\beta t}, \xi_{\beta t+j})| \leq 8 \left(\mathbb{E} \left| \xi_{\beta t} \right|^\delta \right)^{2/\delta} |\alpha(j)|^{1-2/\delta} = A_\beta h^{2L+2/\delta} |\alpha(j)|^{1-2/\delta} + o(h^{2L+2/\delta}),$$

(B.7)

where $A_\beta = 8(\int \left| \frac{\beta^{(L)}(\tilde{z}_p)}{L!} \right|^\delta |p|^{\delta L} K(p)^\delta dp f(z))^{2/\delta}$, since $\mathbb{E} \left| \xi_{\beta t} \right|^\delta = h^{1+L\delta} \int \left| \frac{\beta^{(L)}(\tilde{z}_p)}{L!} \right|^\delta$ $|p|^{\delta L} K(p)^\delta dp f(z) + o(h^{1+L\delta})$ in a similar way to (B.5), 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

$$\mathbb{E}\xi_{\beta t}\xi'_{\beta t+j} = \mathbb{E}[(\beta(z_t) - \beta(z))(\beta(z_{t+j}) - \beta(z))' K_{tz}K_{t+j,z}]$$

$$= \int \int (\beta(s_0) - \beta(z))(\beta(s_j) - \beta(z))' K\left(\frac{s_0 - z}{h}\right) K\left(\frac{s_j - z}{h}\right) f_{0,j}(s_1, s_j) ds_0 ds_j$$

$$= h^2 \int \int (\beta(z + hp_0) - \beta(z))(\beta(z + hp_j) - \beta(z))' K(p_0) K(p_j) f_{0,j}(z + hp_0, z + hp_j) dp_0 dp_j$$

$$= \left\{ h^{2L+2} \frac{\beta^{(L)}(z)}{L!} \frac{\beta^{(L)}(z)'}{L!} f_{0,j}(z, z) \int p_1^L K(p_1) dp_1 \int p_2^L K(p_2) dp_2 + o(h^{2L+2}) \right\}$$

$$\times \mathbf{1}_{\{L=\text{even}\}}$$

$$+ \left\{ h^{2L+4} \left[\frac{\beta^{(L+1)}(z)}{(L+1)!} \frac{\beta^{(L+1)}(z)'}{(L+1)!} f_{0,j}(z, z) + \frac{\beta^{(L+1)}(z)}{(L+1)!} \frac{\beta^{(L)}(z)'}{L!} \frac{\partial f_{0,j}}{\partial s_j}(z, z) \right. \right.$$

$$\left. + \frac{\beta^{(L)}(z)}{L!} \frac{\beta^{(L+1)}(z)'}{(L+1)!} \frac{\partial f_{0,j}}{\partial s_0}(z, z) + \frac{\beta^{(L)}(z)}{L!} \frac{\beta^{(L)}(z)'}{L!} \frac{\partial^2 f_{0,j}}{\partial s_0 \partial s_j}(z, z) \right]$$

$$\int p_1^{L+1} K(p_1) dp_1 \int p_2^{L+1} K(p_2) dp_2$$

$$\left. + o(h^{2L+4}) \right\} \times \mathbf{1}_{\{L=\text{odd}\}}$$

$$= O(h^{2L^*+2}) \leq O(h^6) \text{ for } L \geq 1.$$

(B.8)

We now deduce that the long run variance matrix of η_t, or variance matrix of the standardized partial sum $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^n \eta_t$, is

$$\mathbb{V}^{LR}(\eta_t) = \mathbb{E} \left[\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^n \eta_t \right] \left[\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^n \eta_t \right]'$$

$$= \frac{1}{nh^{2L+1}} \sum_{t=1}^n \mathbb{E}\eta_t\eta'_t + \frac{1}{nh^{2L+1}} \sum_{t \neq s} \mathbb{E}\eta_t\eta'_s$$

$$= \frac{1}{h^{2L+1}} \mathbb{E}\eta_t\eta'_t + o(1) \rightarrow \frac{v_{2L}(K)f(z)}{(L!)^2} \beta^{(L)}(z)\beta^{(L)}(z)' =: V_{\eta\eta, L},$$

(B.9)

which follows from (B.6) 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.5), (B.7) and

(B.8) above, we have

$$\frac{1}{nh^{2L+1}} \sum_{t \neq s} \mathbb{E}\eta_t \eta_s' = \frac{1}{h^{2L+1}} \sum_{j=-n+1, j\neq 0}^{n-1} \left[1 - \frac{|j|}{n}\right] [\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})']$$

$$= \frac{1}{h^{2L+1}} \sum_{j=-M, j\neq 0}^{M} \left[1 - \frac{|j|}{n}\right] [\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})']$$

$$+ \frac{1}{h^{2L+1}} \sum_{M < |j| < n} \left(1 - \frac{|j|}{n}\right) [\mathbb{E}\xi_{\beta t}\xi_{\beta t+j}' - (\mathbb{E}\xi_{\beta t})(\mathbb{E}\xi_{\beta t})']$$

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

$$= O\left(Mh \times \mathbf{1}_{\{L=\text{even}\}} + Mh^3 \times \mathbf{1}_{\{L=\text{odd}\}}\right) + O\left(\frac{h^{2\frac{1+L\delta}{\delta}}}{h^{2L+1}M^c} \sum_{M < |j| < \infty} j^c \alpha_j^{1-2/\delta}\right)$$

$$= O\left(Mh \times \mathbf{1}_{\{L=\text{even}\}} + Mh^3 \times \mathbf{1}_{\{L=\text{odd}\}}\right) + O\left(\frac{1}{h^{1-2/\delta}M^c} \sum_{M < |j| < \infty} j^c \alpha_j^{1-2/\delta}\right)$$

$$= O\left(Mh \times \mathbf{1}_{\{L=\text{even}\}} + Mh^3 \times \mathbf{1}_{\{L=\text{odd}\}}\right) + o\left(\frac{1}{(M^{\frac{c}{1-2/\delta}}h)^{1-2/\delta}}\right) = o(1), \quad \textbf{(B.10)}$$

for a suitable choice of $M \to \infty$ such that $Mh \to 0$, $M^\tau h \to \infty$, and $\frac{M}{n} \to 0$, with $\tau > 1$, $c > \tau(1 - 2/\delta)$ 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 (2008, Thm. 6.5) that the standardized partial sum process of η_t satisfies a triangular array functional law giving

$$\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t \rightsquigarrow B_{\eta,L}(\cdot), \quad \textbf{(B.11)}$$

where $B_{\eta,L}$ is vector Brownian motion with variance matrix $V_{\eta\eta,L} = \frac{v_{2L}(K)f(z)}{(L!)^2}\beta^{(L)}(z)$ $\beta^{(L)}(z)'$. The effective sample size condition $nh \to \infty$ is required for this result.

   Part (a) (ii) Otherwise, when $nh \to c \in [0, \infty)$ we have

$$\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t = O_p(1), \quad \textbf{(B.12)}$$

but no invariance principle applies. Taking scalar x_t and $iid\{z_t\}$ for ease of notation, the stability condition

$$\mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\eta_t}{\sqrt{h^{2L+1}}}\right)^2 = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left(\frac{\eta_t}{\sqrt{h^{2L+1}}}\right)^2 = v_2(K)f(z)\left(\frac{\beta^{(L)}(z)}{L!}\right)^2 + O(h),$$

is satisfied so that $\frac{1}{\sqrt{nh^{2L+1}}}\sum_{t=1}^{n}\eta_t = O_p(1)$ but the Lindeberg condition fails. To show this, note that $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t} = \xi_{\beta t} + O(h^{L^*+1})$. Given $\epsilon > 0$, $nh \nrightarrow \infty$ and $\beta^{(L)}(z) \neq 0$ imply

$$
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\{\left(\frac{\eta_t}{\sqrt{h^{2L+1}}}\right)^2 \mathbf{1}_{\left[|\eta_t|>\epsilon\sqrt{nh^{2L+1}}\right]}\right\}
$$

$$
= \int \frac{[[\beta(z_t)-\beta(z)]K_{tz}+O(h^{L^*+1})]^2}{h^{2L+1}}\mathbf{1}_{\left[|[\beta(z_t)-\beta(z)]K_{tz}+O(h^{L^*+1})|>\epsilon\sqrt{nh^{2L+1}}\right]}f(z_t)dz_t
$$

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

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

$$
\rightarrow \begin{cases} \left(\frac{\beta^{(L)}(z)}{L!}\right)^2 f(z)v_{2L}(K) > 0 & \text{if } nh \rightarrow 0, \\[2ex] \left(\frac{\beta^{(L)}(z)}{L!}\right)^2 f(z)\int p^{2L}K(p)^2\mathbf{1}_{\left[|\frac{\beta^{(L)}(z)}{L!}p^L K(p)|>\epsilon\sqrt{c}\right]}dp > 0 & \text{if } nh \rightarrow c. \end{cases}
$$

Part (b) This part follows in essentially the same way as Part (a) and the proof is omitted.

Part (c) From (B.5) we have

$$
\mathbb{E}\xi_{\beta t} = h^{L^*+1}\mu_{L^*}(K)C_L(z) + o(h^{L^*+1}) = h^{L^*+1}\mathcal{G}_L(z) + o(h^{L^*+1}) \tag{B.13}
$$

from which it follows directly that $\frac{1}{nh^{L^*+1}}\sum_{t=1}^{n}\mathbb{E}x_t x_t'\xi_{\beta t} = \frac{1}{nh^{L^*+1}}\sum_{t=1}^{n}\mathbb{E}[\xi_{\beta t}\mathbb{E}(x_t x_t'|z_t)]$ $\rightarrow_p \mathcal{G}_L(z)\Sigma_{xx}(z)$, as required.   □

LEMMA B.3. *Under Assumption 2 and if $\beta^{(\ell)}(z) = 0$ for $\ell = 0, 1, \cdots, L-1$ and $\beta^{(L)}(z) \neq 0$, the following hold as $n \rightarrow \infty$ and $h \rightarrow 0$:*

(a) *(i) If $nh \rightarrow \infty$, $\frac{1}{\sqrt{n^3 h}}\sum_{t=1}^{n}x_t x_t'\zeta_{tK} \rightsquigarrow \int B_x B_x' dB_{\zeta K} \equiv_d \mathcal{MN}\left(0, v_0(K)f(z)\int B_x B_x' \otimes B_x B_x'\right)$, and $\frac{1}{\sqrt{n^3 h^{2L+1}}}\sum_{t=1}^{n}x_t x_t'\eta_t \rightsquigarrow \int B_x B_x' dB_{\eta,L} \equiv_d \mathcal{MN}\left(0, \frac{v_{2L}(K)f(z)}{(L!)^2}\int B_x B_x'\right.$ $\left[B_x'\beta^{(L)}(z)\right]^2$), where $\zeta_{tK} = K_{tz} - \mathbb{E}K_{tz}$, and $\eta_t = \xi_{\beta t} - \mathbb{E}\xi_{\beta t}$ with $\xi_{\beta t} = [\beta(z_t) - \beta(z)]K_{tz}$;*
   (ii) If $nh \rightarrow c \in [0, \infty]$, $\frac{1}{\sqrt{n^3 h^{2L+1}}}\sum_{t=1}^{n}x_t x_t'\eta_t = O_p(1)$, and $\frac{1}{\sqrt{n^3 h}}\sum_{t=1}^{n}x_t x_t'\zeta_{tK} = O_p(1)$, both without invariance principles holding.

(b) *With $C_L(z)$ defined as in (B.4) and $\mathcal{G}_L(z) = \mu_{L^*}(K)C_L(z)$,*

$$
\frac{1}{n^2 h^{L^*+1}}\sum_{t=1}^{n}x_t x_t'\mathbb{E}\xi_{\beta t} \rightsquigarrow \mu_{L^*}(K)C_L(z)\int B_x B_x' = \mathcal{G}_L(z)\int B_x B_x', \tag{B.14}
$$

   where $L^* = L \times \mathbf{1}_{\{L=even\}} + (L+1)\mathbf{1}_{\{L=odd\}}$.

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

(d) *(i) If* $nh \to \infty$, $\frac{1}{\sqrt{n^2 h}} \sum_{t=1}^{n} x_t u_t K_{tz} \rightsquigarrow \int B_x dB_{uK} \equiv_d \mathcal{MN} \left(0, v_0(K) \sigma_u^2 f(z) \int B_x B_x' \right)$;
   (ii) If $nh \to c \in [0, \infty)$, $\frac{1}{\sqrt{n^2 h}} \sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1)$ *but no invariance principle holds.*

Proof of Lemma B.3. Part (a) (i) First observe that when $nh \to \infty$ we have the joint convergence

$$\left(\frac{1}{\sqrt{n}} x_{\lfloor n \cdot \rfloor}, \quad \frac{1}{\sqrt{nh}} \sum_{t=1}^{\lfloor n \cdot \rfloor} \zeta_{tK}, \quad \frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{\lfloor n \cdot \rfloor} \eta_t \right) \rightsquigarrow \left(B_x(\cdot), \quad B_{\zeta K}(\cdot), \quad B_{\eta, L}(\cdot) \right),$$
(B.15)

by virtue of Assumption 1, Lemmas B.1(a)(i) and B.2(a)(i), where the Brownian motion $\{B_x\}$ is independent of $\{B_{\zeta K}, B_{\eta, L}\}$ by virtue of the exogeneity of x_t. The covariance between $B_{\zeta K}$ and $B_{\eta, L}$ is $\frac{\beta^{(L)}(z)}{L!} v_L(K) f(z)$. This follows from the fact that the contemporaneous covariance $\mathbb{E} \zeta_{tK} \eta_t = \mathbb{E} K_{tz}^2 (\beta(z_t) - \beta(z)) - \mathbb{E} K_{tz} \mathbb{E} K_{tz} (\beta(z_t) - \beta(z)) = h^{L+1} \frac{\beta^{(L)}(z)}{L!} v_L(K) f(z) + O(h^{L^*+2}) = O(h^{L+1})$ and the cross serial covariance $\mathbb{E} \zeta_{tK} \eta_{t+j} = O(h^{L^*+2})$ 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.9) we have

$$\mathbb{E} \left(\frac{1}{\sqrt{nh}} \sum_{t=1}^{\lfloor n \cdot \rfloor} \zeta_t \times \frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{\lfloor n \cdot \rfloor} \eta_t \right)$$
$$= \frac{1}{h^{L+1}} \mathbb{E} \left(\zeta_{tK} \eta_t \right) \{1 + o(1)\} = \frac{\beta^{(L)}(z)}{L!} v_L(K) f(z) \{1 + o(1)\}.$$

Weak convergence to the stochastic integral limits

$$\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' \zeta_{tK} = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\zeta_{tK}}{\sqrt{nh}} \rightsquigarrow \int B_x B_x' dB_{\zeta K},$$
(B.16)

$$\frac{1}{\sqrt{n^3 h^{2L+1}}} \sum_{t=1}^{n} x_t x_t' \eta_t = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\eta_t}{\sqrt{nh^{2L+1}}} \rightsquigarrow \int B_x B_x' dB_{\eta, L}$$
(B.17)

then follows directly from Ibragimov and Phillips (2008, Thm. 4.3) when $nh \to \infty$, respectively. Both stochastic integrals have mixed normal distributions, viz.,

$$\int B_x \otimes B_x dB_{\zeta K} \equiv_d \mathcal{MN} \left(0, v_0(K) f(z) \int B_x B_x' \otimes B_x B_x' \right),$$
(B.18)

$$\int B_x B_x' dB_{\eta, L} \equiv_d \mathcal{MN} \left(0, \frac{v_{2L}(K) f(z)}{(L!)^2} \int B_x B_x' \left(B_x(r)' \beta^{(L)}(z) \right)^2 \right),$$
(B.19)

and the stated result holds.
   Part (a) (ii) When the rate conditions $nh \to \infty$ fails and, instead $nh \to c \in [0, \infty)$ applies, it follows from Lemmas B.1(a)(ii) and B.2(a)(ii) that $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \zeta_{tK} = O_p(1)$ and $\frac{1}{\sqrt{nh^{2L+1}}} \sum_{t=1}^{n} \eta_t = O_p(1)$, respectively, but with no invariance principles holding. Correspondingly, in place of (B.16) and (B.17), we have

$$\frac{1}{\sqrt{n^3 h}} \sum_{t=1}^{n} x_t x_t' \zeta_{tK} = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\zeta_{tK}}{\sqrt{nh}} = O_p(1), \tag{B.20}$$

$$\frac{1}{\sqrt{n^3 h^{2L+1}}} \sum_{t=1}^{n} x_t x_t' \eta_t = \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \right) \frac{\eta_t}{\sqrt{nh^{2L+1}}} = O_p(1), \tag{B.21}$$

again without invariance principles.

Part (b). Using (B.13) and standard weak convergence methods we have

$$\frac{1}{n^3 h^{L^*+1}} \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} = \frac{1}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\mathbb{E}\xi_{\beta t}}{h^{L^*+1}} \rightsquigarrow \mathcal{G}_L(z) \int B_x B_x',$$

giving the stated result.

Part (c). Using (B.20), Part (i) follows as in Phillips and Wang (2021, Lem. B.1(c) (i)), giving

$$\frac{1}{n^2 h} \sum_{t=1}^{n} x_t x_t' K_{tz} = \frac{1}{n} \sum_{t=1}^{n} \frac{x_t}{\sqrt{n}} \frac{x_t'}{\sqrt{n}} \frac{\mathbb{E}K_{tz}}{h} + o_p(1) \rightsquigarrow \int B_x B_x' f(z). \tag{B.22}$$

The proof of (ii) follows as in Phillips and Wang (2021, Lem. B.1(c) (ii)).

Part (d). (i) As in Part (a)(i), when $nh \to \infty$ we have $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n\lfloor \cdot \rfloor} u_t K_{tz} \rightsquigarrow B_{uK}(\cdot)$ as in Lemma B.1(a) and then weak convergence to the stochastic integral $\frac{1}{\sqrt{n^2 h}} \sum_{t=1}^{n} x_t u_t K_{tz} \rightsquigarrow \int B_x dB_{uK}$ follows directly from Ibragimov and Phillips (2008, Thm. 4.3). In view of the independence between B_x and B_{uK}[10], we have $\int B_x dB_{uK} \equiv_d \mathcal{MN}\left(0, v_0(K)\sigma_u^2 f(z) \int B_x B_x'\right)$. (ii) If $nh \to c \in [0,\infty)$, then $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n\lfloor \cdot \rfloor} u_t K_{tz} = O_p(1)$ as in Lemma B.1 (a)(ii), and then $\frac{1}{\sqrt{n^2 h}} \sum_{t=1}^{n} x_t u_t K_{tz} = O_p(1)$ but no invariance principle holds. □

C. ADDITIONAL COMPUTATIONAL DETAILS

The following paragraphs provide further details of how the three statistics in Figures 4, 5, 6 and Table 2 were computed.

(i) Computation of the naive t-ratio $\hat{T}(z; L = 1)$ follows the definition (3.1). Since the use of $\hat{T}(z; L = 1)$ implies belief that $L = 1$, the optimal bandwidth order for that case is employed in the computation. For the computation of $\hat{\beta}(z)$ and K_{tz} the bandwidth $h = \hat{\sigma}_z n^\gamma$ was used with $\gamma = -1/2$ for nonstationary x_t and $\gamma = -1/5$ for stationary x_t. For the other unknown components $\beta^{(1)}, \beta^{(2)}(z), f(z)$ and $f^{(1)}(z)$ involved in (3.2) and (3.3), different

---

[10]Since x_t is exogenous the covariance of the two processes is

$$\mathbb{E}\left(\frac{x_{n\lfloor \cdot \rfloor}}{\sqrt{n}} \frac{1}{\sqrt{nh}} \sum_{t=1}^{n\lfloor \cdot \rfloor} u_t K_{tz} \right) = \frac{\sqrt{h}}{n} \sum_{t,s=1}^{n\lfloor \cdot \rfloor} \mathbb{E}\left(u_{xt} u_s \frac{1}{h} K\left(\frac{z_s - z}{h} \right) \right) = \frac{\sqrt{h}}{n} \sum_{t,s=1}^{n\lfloor \cdot \rfloor} \mathbb{E}\left(u_{xt} \mathbb{E}\left[u_s \frac{1}{h} K\left(\frac{z_s - z}{h} \right) | \mathcal{F}_x \right] \right)$$

$$= \frac{\sqrt{h}}{n} \sum_{t,s=1}^{n\lfloor \cdot \rfloor} \mathbb{E}\left(u_{xt} \mathbb{E}\left[\mathbb{E}(u_s | \mathcal{F}_x) \frac{1}{h} K\left(\frac{z_s - z}{h} \right) \right] \right) = 0, \tag{B.23}$$

leading directly to the stated independence of B_x and B_{uK}. When exogeneity is relaxed, it can be shown that (B.23) holds asymptotically as $n \to \infty$ and $h \to 0$ under a weak additional summability condition.

bandwidth orders were used. Specifically, $\hat{f}(z)$ used $\gamma = -1/5$, $\hat{f}^{(1)}(z)$ used $\gamma = -1/7$, $\hat{\beta}^{(1)}$ was estimated using local linear estimation with $\gamma = -1/7$ for stationary x_t and $\gamma = -2/7$ for nonstationary x_t, and $\hat{\beta}^{(2)}$ was estimated by local quadratic estimation with $\gamma = -1/9$ for stationary x_t and $\gamma = -2/9$ for nonstationary x_t. Those orders were selected based on optimal bandwidth order rules in the case of local p-th order polynomial estimation to estimate $\beta^{(p)}(z)$ for $p = 1, 2$ and they utilize ongoing work by the authors for local p-th order polynomial estimation in functional coefficient regression.

(ii) For the infeasible statistic $\hat{T}(z; true\ L)$, the true L is used in the computation. For $L = 1$, it is identical to the naive choice. For $L = 4$, we need to estimate $\beta^{(4)}(z)$. We use local 4-th order estimation with bandwidth order $\gamma = -1/13$ for stationary x_t and $\gamma = -2/13$ for nonstationary x_t. For the computation of $\hat{\beta}(z)$ and K_{tz} the optimal order $\gamma = -1/(2L^* + 1)$ for stationary x_t and $\gamma = -2/(2L^* + 1)$ for nonstationary x_t were used.

(iii) For computation of the oracle t-ratio $T(z; true\ L)$, the quantities L, $\beta^{(L)}(z)$, $\beta^{(L+1)}(z)$, $f(z)$, $f^{(1)}(z)$ and σ_u^2 were assumed known. Given a known L, the optimal order was used in the estimation of $\hat{\beta}(z)$ and computation of K_{tz}. More specifically, for L greater than 1, the optimal order used is $\gamma = -1/(2L^* + 1)$ for stationary x_t and $\gamma = -2/(2L^* + 1)$ for nonstationary x_t. For $L = 1$, the optimal order is $\gamma = -1/5$ for stationary x_t and $\gamma = -1/2$ for nonstationary x_t.

(iv) Computation of the pre-test based statistic $\hat{T}(z; \hat{L}^{pre})$ follows similar lines as that of the infeasible statistic, except that L is obtained from the two-step pre-test procedure.

(v) For the residual variance σ_u^2 the same estimate was used in the naive choice and the infeasible statistics. We used $\gamma = -1/2$ to compute $\hat{\beta}(z_t)$ and hence the residual estimates $\hat{u}_t = y_t - x_t\hat{\beta}(z_t)$. Then σ_u^2 was estimated by $\hat{\sigma}_u^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2$.

D. ASYMPTOTICS USED IN TESTING OF $\mathbb{H}_{0,\beta^{(2)}} : \beta^{(2)}(z) = 0$

For the convenience of readers, we outline the limit theory of estimator $\hat{\beta}^{(2)}(z)$, which is obtained by local quadratic estimation ($p = 2$). Detailed derivations are available in our ongoing work Wang and Phillips (2022). We start by noting that

$$\frac{\hat{\beta}^{(2)}(z) - \beta^{(2)}(z)}{2} - h^2 f^{-1}(z)[a_{20}\mu_4(K) + a_{22}\mu_6(K)]B_{2p}(z) \sim_a \sqrt{\frac{h}{n}}\mathcal{MN}(0, \Omega_{\beta,2,2}(z))$$

$$+ \frac{1}{n\sqrt{h^5}}\mathcal{MN}(0, \Omega_{u,2,2}), \tag{D.1}$$

where a_{ij}, $i, j = 0, 1, 2$ denotes the $(i + 1, j + 1)$-element of the 3×3 matrix $M_p(K)^{-1}$, $B_{2p}(z) = \frac{\beta^{(p+1)}(z)}{(p+1)!}f^{(1)}(z) + \frac{\beta^{(p+2)}(z)}{(p+2)!}f(z)$, $\Omega_{\beta,2,2}(z) = f^{-1}(z)\omega_{2,2}^*(K)B_{(x,2)}^{-1}$ $\int\left[B_xB_x'\frac{\beta^{(3)}(z)}{3!}\frac{\beta^{(3)}(z)'}{3!}B_xB_x'\right]B_{(x,2)}^{-1}$, $\Omega_{u,2,2}(z) = \sigma_u^2 f^{-1}(z)\omega_{2,2}(K)B_{(x,2)}^{-1}$, with $\omega_{2,2}(K) = m_2'R_p(K)m_2$, $\omega_{2,2}^*(K) = m_2'R_p^*(K)m_2$, row vector m_2' denotes the third row of $M_p(K)^{-1}$,

$$M_p(K) = \begin{pmatrix} 1 & \mu_1(K) & \mu_2(K) \\ \mu_1(K) & \mu_2(K) & \mu_3(K) \\ \mu_2(K) & \mu_3(K) & \mu_4(K) \end{pmatrix}, R_p(K) = \begin{pmatrix} v_0(K) & v_1(K) & v_2(K) \\ v_1(K) & v_2(K) & v_3(K) \\ v_2(K) & v_3(K) & v_4(K) \end{pmatrix}, \text{and } R_p^*(K) =$$

$$\begin{pmatrix} v_6(K) & v_7(K) & v_8(K) \\ v_7(K) & v_8(K) & v_9(K) \\ v_8(K) & v_9(K) & v_{10}(K) \end{pmatrix}.$$ Similarly, it can be verified that the MSE optimal bandwidth

order is $n^{-2/9}$, and with this order the first term in the RHS of (D.1) is ignorable. Therefore, (D.1) is also asymptotically valid if local flatness exists. Construction of the t-ratio based on (D.1) again follows in the same way as the t-ratio given in (3.1). Estimation of the asymptotic variances in (D.1) can be constructed as in (3.2) since the asymptotic variances $\Omega_{\beta,2,2}(z)$ and $\Omega_{u,2,2}(z)$ in (D.1) take similar forms to those in (2.11) and (2.12) and are not reported.

SUPPLEMENTARY MATERIAL

Peter C. B. Phillips and Ying Wang (2022). Supplement to "Limit Theory for Locally Flat Functional Coefficient Regressions," *Econometric Theory* Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466622000287

REFERENCES

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

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

Fan, J. and Q. Yao (2008) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Science & Business Media.

Gao, J. and P.C.B. Phillips (2013) Functional coefficient nonstationary regression with non-and semi-parametric cointegration. *Working Paper*.

Hall, P. and J.S. Marron (1987) Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields* 74(4), 567–581.

Hall, P., S.J. Sheather, M. Jones, and J.S. Marron (1991) On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 78(2), 263–269.

Ibragimov, R. and P.C.B. Phillips (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, D., P.C.B. Phillips, and J. Gao (2016) Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression. *Econometric Theory* 32(3), 655–685.

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

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

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

Phillips, P.C.B. and T. Magdalinos (2009) Econometric inference in the vicinity of unity. *CoFie Working Paper* 7, Singapore Management University.

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

Phillips, P. C. B. and Y. Wang (2021) When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Journal of Econometrics*, published online 20 October 2021. doi: 10.1016/j.jeconom.2021.09.007.

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

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

Tu, Y. and Y. Wang (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.

Tu, Y. and Y. Wang (2020) Adaptive estimation of heteroskedastic functional-coefficient regressions with an application to fiscal policy evaluation on asset markets. *Econometric Reviews* 39(3), 299–318.

Tu, Y. and Y. Wang (202 2) Spurious functional-coefficient regression models and robust inference with marginal integration. *Journal of Econometrics* 229(2), 396–421. doi: 10.1016/j.jeconom.2020.12.010.

Wang, Y. and P.C.B. Phillips (2022) Limit theory of local p-th order polynomial estimator in functional coefficient cointegrating regression. *Working Paper,* Renmin University of China.

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

Wang, Y., Y. Tu, and S.X. Chen (2016) Improving inflation prediction with the quantity theory. *Economics Letters* 149, 112–115.

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