

# Improved Estimation and Inference in Non-Cointegrated Functional-Coefficient Regression using Marginal Integration\*

Ying Wang<sup>b</sup> Peter C. B. Phillips<sup>a,b,c,d</sup>, Yundong Tu<sup>e</sup>,

<sup>a</sup>Yale University, <sup>b</sup>University of Auckland,  
<sup>c</sup>University of Southampton, <sup>d</sup>Singapore Management University  
<sup>e</sup>Peking University

January 29, 2019

## Abstract

To evaluate the relationship between two integrated processes that are correlated but not cointegrated in a functional-coefficient regression model, [Sun et al. \(2011\)](#) proposed a  $\sqrt{nh}$ -consistent estimator for the functional coefficient that serves as a correlation measure. The distributional properties of this estimator were not derived. The present paper proposes four nearly  $\sqrt{n}$ -consistent estimators for the functional coefficient using marginal integration and establishes their asymptotic normality, facilitating inference. Numerical studies reveal that the proposed estimators can achieve significant efficiency improvements compared to that of [Sun et al. \(2011\)](#). Among the estimators investigated a *three-step* estimator proposed in the paper is recommended for practical work as it demonstrates superior performance in simulations. This estimator leads to an easy-to-implement specification test to check constancy of the functional coefficient. This test shows significant power gains over the test proposed recently in [Gan et al. \(2014\)](#).

*JEL classification:* C14; C22.

*Keywords:* Cointegration; Correlation; Functional-coefficient regression; Marginal integration; Non-stationary time series.

---

\*Phillips and Wang acknowledge Marsden Fund support at the University of Auckland.

# 1 Introduction

This paper reconsiders the semiparametric functional-coefficient model of [Sun et al. \(2011\)](#) and [Gan et al. \(2014\)](#)

$$y_t = x_t' \beta(z_t) + u_t, t = 1, \dots, n, \quad (1.1)$$

where the  $k \times 1$  vector  $x_t$  is an integrated process of order 1,  $\beta(\cdot)$  is a  $k \times 1$  vector of smooth measurable and squared integrable functions of a scalar stationary variable  $z_t$ ,  $u_t$  is the unobserved unit root innovation.

The nonstationary error in (1.1) distinguishes this model from the functional-coefficient cointegration model studied by [Juhl \(2005\)](#), [Cai et al. \(2009\)](#), and [Xiao \(2009\)](#), where the innovation is assumed stationary. This formulation suggests that the variables  $x_t$ ,  $z_t$  and  $y_t$  are not cointegrated. However, the correlation between  $x_t$  and  $y_t$  is captured by the function  $\beta(\cdot)$  with the covariate  $z_t$ . The variables  $x_t$  and  $y_t$  move together, but not so closely as to ensure cointegration. This model is relevant in the study of economic relationships, such as the purchasing power parity in terms of characterizing the spill-over effect of price to exchange rate with imperfect information (missing integrated factors), or when there exists persistent measurement errors in data aggregation, as argued in [Sun et al. \(2011\)](#). It therefore serves as a semiparametric approach to measure the correlation between integrated but not cointegrated variables. [Sun et al. \(2011\)](#) provided a consistent estimator for  $\beta(z)$ , whose asymptotic distributional properties are yet available.

This paper aims to construct new estimators of  $\beta(z)$ , based on the marginal integration and backfitting techniques, which possess improved performance than those of [Sun et al. \(2011\)](#). Our proposed estimators are shown to follow an asymptotic normal distribution, with the variance converging at the rate  $1/n$  and the asymptotic bias diminishing at the usual rate  $h^2$ , where  $h$  is the smoothing bandwidth. This limit result is not standard in the semiparametric literature, where the asymptotic variance of the unknown function estimator often depends on the smoothing bandwidth. The new result is intrinsically due to the nonstationarity of the regressor  $x_t$  that alters the order of magnitude in the asymptotic decomposition when analyzing the properties of the estimators based on marginal integration. Consequently, the univariate standard nonparametric convergence rate, such as that obtained by [Fan et al. \(1998\)](#) in the stationary case, can be improved. Due to the additive form of the true coefficient, we finally get a near  $\sqrt{n}$ -convergence.

The second objective is to construct a test based on our proposed estimators for detecting linearity against the semiparametric specification in (1.1). This amounts to testing whether  $\beta(z)$  can be treated simply as a constant, an issue first studied by [Gan et al. \(2014\)](#). Of the two tests they proposed, the one based on [Sun et al. \(2011\)](#)'s estimator

of  $\alpha(z) \equiv \beta(z) - E\beta(z_t)$  shows better performance in simulations. Given the improved rate of convergence performance of our proposed estimator, a natural conjecture is the test statistic constructed based on our proposed estimators would enjoy better power than that of [Gan et al. \(2014\)](#). This is formally confirmed in our analysis. In particular, we show analytically that our tests diverge at a faster rate than that of [Gan et al. \(2014\)](#) under the alternative, leading to very good local power performance while the test of [Gan et al. \(2014\)](#) has no power. Finite sample simulation results corroborate the asymptotic analysis.

The remainder of the paper is organized as follows. [Section 2](#) presents our estimation approach, with the limiting properties of the estimators. [Section 3](#) discusses hypothesis testing on  $\beta(z)$ . Numerical studies reported in [Section 4](#) illustrate the finite sample performance of our estimators and test statistic. [Section 5](#) concludes the paper. All the proofs are relegated to the Appendix.

In matters of notations, we follow the tradition to use “ $\xrightarrow{d}$ ” to signify the weak convergence of the associated probability measures, “ $\xrightarrow{p}$ ” the convergence in probability, and  $\equiv$  the definitional equality. The Brownian motion  $B(r)$  on  $[0, 1]$  are usually written as  $B$  and integrals  $\int$  are understood to be taken over the interval  $[0, 1]$ , unless otherwise specified.

## 2 Estimators and Properties

### 2.1 Estimators

#### One-step estimation

Let  $\alpha(z) = \beta(z) - c_0$ , where  $c_0 = E\beta(z_t)$ . Then we have  $E\alpha(z_t) = 0$ . Model [\(1.1\)](#) is equivalent to

$$\Delta y_t = x'_t \beta(z_t) - x'_{t-1} \beta(z_{t-1}) + \Delta u_t \quad (2.1)$$

$$\begin{aligned} &= (\Delta x_t)' \beta(z_t) + x'_{t-1} [\beta(z_t) - \beta(z_{t-1})] + \Delta u_t \\ &= (\Delta x_t)' \beta(z_t) + x'_{t-1} [\alpha(z_t) - \alpha(z_{t-1})] + \Delta u_t. \end{aligned} \quad (2.2)$$

Rewrite [\(2.2\)](#) as

$$\Delta y_t = (\Delta x_t)' \beta_1(z_t, z_{t-1}) + x'_{t-1} \beta_2(z_t, z_{t-1}) + \Delta u_t, \quad (2.3)$$

where  $\beta_1(z, w) \equiv \beta(z)$ ,  $\beta_2(z, w) \equiv \alpha(z) - \alpha(w)$ . Then it is apparent that

$$\int \beta_1(z, w) \ell_1(w) dw = \beta(z), \quad \int \beta_2(z, w) \ell_2(w) dw = \alpha(z), \quad (2.4)$$

for any weight functions  $\ell_1(w)$  and  $\ell_2(w)$  satisfying  $\int \ell_1(w)dw = 1$  and  $\int \alpha(w)\ell_2(w)dw = 0$ . Noting that  $E\alpha(z_t) = 0$ , one choice could be  $\ell_1(w) = \ell_2(w) = f_Z(w)$ , the density function of  $z_t$ . Therefore, marginal integration estimators for  $\beta(z)$  and  $\alpha(z)$  can be constructed once we obtain estimators for  $\beta_1(z, w)$  and  $\beta_2(z, w)$ . To this end, for any interior point  $(z, w)$ , denote the local constant least square estimator for  $\beta_1(z, w)$  and  $\beta_2(z, w)$  in model (2.3), as  $\hat{\beta}_1(z, w)$  and  $\hat{\beta}_2(z, w)$ , respectively. To be precise, denote  $w_t = z_{t-1}$ , and  $X'_t = (\Delta x'_t, x'_{t-1})$ . Then (2.3) can be written as  $\Delta y_t = X'_t\theta(z_t, w_t) + \Delta u_t$ . Denote the bandwidth for  $z_t$  and  $w_t$  as  $h_1$  and  $h_2$ , respectively. The kernel weighted least square (KLS) estimator of  $\theta(z, w)$  is given by

$$(\hat{\beta}'_1(z, w), \hat{\beta}'_2(z, w))' = \left( \sum_t X_t X'_t K_{tz} K_{tw} \right)^{-1} \sum_t X_t \Delta y_t K_{tz} K_{tw}, \quad (2.5)$$

where  $K_{tz} = K((z_t - z)/h_1)$ ,  $K_{tw} = K((w_t - w)/h_2)$ .

Then, with the weighting function being the empirical density of  $z_t$ , the marginal integration based estimators become

$$\hat{\beta}_{1s}(z) = \frac{1}{n} \sum_{t=1}^n \hat{\beta}_1(z, z_t), \quad (2.6)$$

and

$$\hat{\alpha}(z) = \frac{1}{n} \sum_{t=1}^n \hat{\beta}_2(z, z_t), \quad (2.7)$$

where the subscript “1s” signifies that it is the one-step estimator of  $\beta(z)$ . The empirical density weight used above is mainly for technical convenience, and may be alternatively replaced by a kernel density estimator.

Since  $E\alpha(z_t) = 0$ , a re-centered estimator of  $\alpha(z)$  can be constructed as

$$\hat{\alpha}^*(z) = \hat{\alpha}(z) - \frac{1}{n} \sum_t \hat{\alpha}(z_t), \quad (2.8)$$

which satisfies  $E\hat{\alpha}^*(z) = 0$ . In addition, since  $\alpha(z) = \beta(z) - c_0$ , where  $c_0 = E\beta(z_t)$ , an alternative estimator of  $\beta(z)$  can be constructed based on that of  $\alpha(z)$  and  $c_0$ . With  $c_0$  estimated by

$$\tilde{c}_0 = \frac{1}{n} \sum_t \hat{\beta}_{1s}(z_t), \quad (2.9)$$

one such estimator is

$$\hat{\beta}_{1s}^*(z) = \hat{\alpha}(z) + \tilde{c}_0, \quad (2.10)$$

or, alternatively,

$$\hat{\beta}_{1s}^{**}(z) = \hat{\alpha}^*(z) + \tilde{c}_0. \quad (2.11)$$

### Two-step estimation

An alternative estimator for  $c_0$  can be constructed via another regression as follows. First, model (1.1) can be rewritten as  $y_t - x_t' \alpha(z_t) = x_t' c_0 + u_t$ . With the estimator  $\hat{\alpha}(z)$  replacing  $\alpha(z)$ , we get

$$y_t - x_t' \hat{\alpha}(z_t) = x_t' c_0 + u_t^*.$$

With  $y_t^* = y_t - x_t' \hat{\alpha}(z_t)$  and taking differences in the above equation, we get

$$\Delta y_t^* = (\Delta x_t)' c_0 + \Delta u_t^*.$$

This leads to the backfitted ordinary least squares estimator of  $c_0$

$$\hat{c}_0 = \left( \sum_t \Delta x_t \Delta x_t' \right)^{-1} \sum_t \Delta x_t \Delta y_t^* \quad (2.12)$$

Similar to the estimator in (2.10), we obtain the two-step estimator of  $\beta(z)$  as

$$\hat{\beta}_{2s}(z) = \hat{\alpha}(z) + \hat{c}_0. \quad (2.13)$$

Note that we can also plug in the re-centered estimator  $\hat{\alpha}^*(z)$ . Denote the corresponding estimator of  $c_0$  as  $\hat{c}_0^*$ . It can be shown that  $\hat{\beta}_{2s}(z) = \hat{c}_0^* + \hat{\alpha}^*(z)$  for a given sample, namely we get the same two-step estimator  $\hat{\beta}_{2s}(z)$ . The reason behind this equivalence is the same as the reason the two estimators of  $\beta(z)$  proposed in Sun et al. (2011) are equivalent with each other for a given sample. A proof is given in Appendix C to justify this equivalence.

### Three-step estimation

In (2.1), replacing the coefficient before  $x_{t-1}$  by the two-step estimator  $\hat{\beta}_{2s}(z)$  and rearranging<sup>1</sup>, we get

$$\Delta y_t + x_{t-1}' \hat{\beta}_{2s}(z_{t-1}) = x_t' \beta(z_t) + v_t.$$

---

<sup>1</sup>Alternatively, we can replace the coefficient before  $x_{t-1}$  by  $\hat{\beta}_{1s}^*(z)$  or  $\hat{\beta}_{1s}^{**}(z)$ . By doing this, it is not hard to justify that for a given sample, the resulted three-step estimator of  $\beta(z)$  is different with that based on  $\hat{\beta}_{2s}(z)$  by a constant. In the simulation, we find  $\hat{\beta}_{2s}(z)$  performs a bit better than  $\hat{\beta}_{1s}^*(z)$  and  $\hat{\beta}_{1s}^{**}(z)$ . Thus we use  $\hat{\beta}_{2s}(z)$  in the computation of the three-step estimator.

It is readily seen that  $\beta(z)$  in the above regression can be estimated by the conventional kernel smoothing method, with a bandwidth  $h_3$ . This leads to the backfitted three-step estimator of  $\beta(z)$ ,

$$\hat{\beta}_{3s}(z) = \left( \sum_t x_t x_t' K((z_t - z)/h_3) \right)^{-1} \sum_t x_t \tilde{y}_t K((z_t - z)/h_3), \quad (2.14)$$

where  $\tilde{y}_t = \Delta y_t + x_{t-1}' \hat{\beta}_{2s}(z_{t-1})$ .

## 2.2 Asymptotic properties

To study the asymptotic properties of the aforementioned estimators, we need the following assumptions.

**Assumption 1.** (a)  $\{\epsilon_{xt} = \Delta x_t, \Delta u_t, z_t\}$  is a strictly stationary  $\alpha$ -mixing process of size  $-p/(p-2)$  (for some  $p > 2$ ) with a finite, positive definite long-run variance matrix and fourth moments. And also,  $E(\exp|\epsilon_{xt}|^\delta) \leq C < \infty$ ;

(b)  $\{z_t\}$  is independent with  $\{\Delta u_t\}$ . Furthermore,  $E(\epsilon_{xt}|z_t, z_{t-1}) = 0$ ,  $E(\epsilon_{xt}\epsilon_{xt}'|z_t, z_{t-1}) = \Sigma_{\epsilon_x}$ ,  $E[\Delta u_t] = 0$ ,  $E[(\Delta u_t)^2|\epsilon_{xt}] = \sigma_{\Delta u}^2$ , and  $E(\Delta u_t \Delta u_s | \{\epsilon_{xk}\}) = 0$  for  $t \neq s$ ;

(c)  $z_t$  has Lebesgue density  $f(z)$ , and  $(z_t, z_{t-1})$  has joint Lebesgue density  $f(z, w)$ . And also,  $f(z) > 0$  on the support of  $z_t$  and  $f(z, w) > 0$  on the support of  $(z_t, z_{t-1})$ ;

(d)  $\beta(z)$ ,  $f(z)$ , and  $f(z, w)$  are three times continuously differentiable on the support of  $z_t$  (or  $(z_t, z_{t-1})$ ). Furthermore,  $\beta(z)$  is bounded on the support of  $z_t$ ;

(e) The kernel function  $K(\cdot)$  is a bounded symmetric probability density function with bounded support;

(f) As  $n \rightarrow \infty$ ,  $h_1, h_2, h_3 \rightarrow 0$ ,  $nh_1, nh_2, nh_3 \rightarrow \infty$ ,  $(\log n)^{1+\delta-1} \sqrt{h_1 h_2} \rightarrow 0$ ,  $\sqrt{nc_n^2} \rightarrow 0$ , where  $c_n = h_1^2 + h_2^2 + \sqrt{\frac{\log n}{nh_1 h_2}}$ .

**Remark 2.1.** In Assumption 1 (a), the requirement that the exponential moment of  $\epsilon_{xt}$  up to order  $\delta$  is bounded above from infinity is to ensure the uniform convergence of a random walk to the Brownian motion. See [Fraser \(1973\)](#) for more details about this result. Assumption 1 (b) assumes  $\{z_t\}$  is independent with  $\Delta u_t$ , and  $\Delta u_t$  is homogeneous and serially uncorrelated for the simplicity of the proof. These can be relaxed at the cost of a lengthier proof. Generally,  $\Delta u_t$  can be relaxed to be a quite general stationary process that satisfies the invariance principle and has absolutely summable long run variance. The conditions on the density functions and the functional coefficient given in Assumption 1 (c) (d) and that on kernel function stated in (e) are quite common in kernel smoothing.

**Remark 2.2. (Bandwidth)** *The condition  $(\log n)^{1+\delta^{-1}}\sqrt{h_1h_2} \rightarrow 0$  in Assumption 1 (f) is needed to ensure that the uniform convergence rate of a random walk to a Brownian motion is faster than that of stationary kernel density estimation. Unlike the condition that  $\sqrt{nh}c_n^2 \rightarrow 0$  imposed by Fan et al. (1998), we assume instead  $\sqrt{nc_n^2} \rightarrow 0$  to remove the error term obtained from the uniform convergence of kernel density estimator, as our estimators have a much faster rate of convergence ( $\sqrt{n}$ ) than theirs ( $\sqrt{nh}$ ). Other conditions on the bandwidths are standard. The range of bandwidth allowed in Assumption 1 (f) is actually quite wide. To see it clearly, suppose  $h_1 = O(n^{\gamma_1}), h_2 = O(n^{\gamma_2})$ , where  $\gamma_1$  and  $\gamma_2$  satisfy  $-1 < \gamma_1, \gamma_2 < 0$ . The condition  $\sqrt{nc_n^2} \rightarrow 0$  implies that*

$$1 + 8\gamma_1 < 0; 1 + 8\gamma_2 < 0; 1 + 2\gamma_1 + 2\gamma_2 > 0; 3\gamma_1 - \gamma_2 < 0; 3\gamma_2 - \gamma_1 < 0; 1 + 4\gamma_1 + 4\gamma_2 < 0.$$

*This leads to*

$$\gamma_1 < -1/8, \tag{2.15}$$

$$\gamma_2 < -1/8, \tag{2.16}$$

$$1 + 2\gamma_1 + 2\gamma_2 > 0. \tag{2.17}$$

*If we further adopt  $\gamma_1 = \gamma_2 = \gamma$ , this requires that  $-1/4 < \gamma < -1/8$ . Therefore, commonly used bandwidth orders are permitted. In our simulation, results are reported for  $\gamma = -1/5$  in Section 4.1 and those for  $\gamma = -1/6$  in Section 4.2.*

Now we present the asymptotic properties of the aforementioned estimators.

**Theorem 2.1.** *Under Assumption 1, for a given interior point  $z$ , it holds that, as  $n \rightarrow \infty$ ,*

*(i) for the one-step estimators (defined in (2.6) and (2.7)):*

$$\sqrt{nh_1} \left( \hat{\beta}_{1s}(z) - \beta(z) - h_1^2 B_1(z) \right) \xrightarrow{d} N(0, \Sigma_1), \tag{2.18}$$

$$\sqrt{n} (\hat{\alpha}(z) - \alpha(z) - B_2(z)) \xrightarrow{d} N(0, \Gamma(\alpha(z_t))), \tag{2.19}$$

*where  $B_1(z) = \mu_2(K) \{ \beta'(z) E[f'_z(z, w_t)/f(z, w_t)] + \frac{1}{2} \beta''(z) \}$ ,  $B_2(z) = h_1^2 B_1(z) - h_2^2 B_3(z)$ ,  $B_3(z) = \mu_2(K) E[\beta'(w_t) f'_w(z, w_t)/f(z, w_t) + \frac{1}{2} \beta''(w_t)]$ ,  $\Sigma_1 = \nu_0(K) f(z) \sigma_{\Delta u}^2 \cdot E[f^2(w_t)/f^2(z, w_t)] \Sigma_{\epsilon_x}^{-1}$ , and  $\Gamma(\alpha(z_t))$  denotes the long run variance of the process  $\{\alpha(z_t)\}$ ;*

*(ii) for the re-centered estimators (defined in (2.8), (2.9), (2.10) and (2.11)):*

$$\sqrt{n} (\hat{\alpha}^*(z) - \alpha(z) - B_4(z)) \xrightarrow{d} N(0, \Gamma(\alpha(z_t))), \tag{2.20}$$

$$\sqrt{n} (\tilde{c}_0 - c_0 - h_1^2 E B_1(z_t)) \xrightarrow{d} N(0, \Sigma_2 + \Gamma(\alpha(z_t))), \tag{2.21}$$

$$\sqrt{n} \left( \hat{\beta}_{1s}^*(z) - \beta(z) - B_5(z) \right) \xrightarrow{d} N(0, \Sigma_2), \tag{2.22}$$

$$\sqrt{n} \left( \hat{\beta}_{1s}^{**}(z) - \beta(z) - B_6(z) \right) \xrightarrow{d} N(0, \Sigma_2), \tag{2.23}$$

where  $B_4(z) = B_2(z) + EB_2(z_t)$ ,  $B_5(z) = B_2(z) + h_1^2 EB_1(z_t)$ ,  $B_6(z) = B_4(z) + h_1^2 EB_1(z_t)$  and  $\Sigma_2 = \sigma_{\Delta u}^2 E[f^2(w_s)f^2(z_s)/f^2(z_s, w_s)]\Sigma_{\epsilon_x}^{-1}$ ;

(iii) for the two-step estimators (defined in (2.12) and (2.13)):

$$\sqrt{n}(\hat{c}_0 - c_0 + \mathcal{B}_1) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1} + \Gamma(\alpha(z_t))), \quad (2.24)$$

$$\sqrt{n}(\hat{c}_0^* - c_0 + \mathcal{B}_1^*) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1} + \Gamma(\alpha(z_t))), \quad (2.25)$$

$$\sqrt{n}(\hat{\beta}_{2s}(z) - \beta(z) - \mathcal{B}_2(z)) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1}), \quad (2.26)$$

where  $\mathcal{B}_1 = \Sigma_{\epsilon_x}^{-1} E\Delta x_t \Delta[x_t' B_2(z_t)]$ ,  $\mathcal{B}_1^* = \Sigma_{\epsilon_x}^{-1} E\Delta x_t \Delta[x_t' B_4(z_t)]$ ,  $\mathcal{B}_2(z) = B_2(z) - \mathcal{B}_1$ ;

(iv) for the three-step estimator (defined in (2.14)):

$$\sqrt{n}(\hat{\beta}_{3s}(z) - \beta(z) - \mathcal{B}_3(z)) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1}), \quad (2.27)$$

where  $\mathcal{B}_3(z) = h_3^2 \mu_2(K) [f'(z)\beta'(z)/f(z) + \frac{1}{2}\beta''(z)] + E\mathcal{B}_2(z_t)$ .

**Remark 2.3. (Comments for  $\hat{\beta}_{1s}(z)$ )** The limiting distribution for  $\hat{\beta}_{1s}(z)$  given in (2.18) is standard in additive models, when the estimator is constructed based on marginal integration. It has the standard univariate nonparametric variance and bias order. See Fan et al. (1998) for similar results.

**Remark 2.4. (Comments for  $\hat{\alpha}(z)$ )** (i) The limiting results for  $\hat{\alpha}(z)$  are nonstandard in nonparametric smoothing. Although it has a standard asymptotic bias order, its asymptotic variance is of order  $O(1/n)$ , much smaller than the order of a typical nonparametric estimator which often relies on the smoothing bandwidth. This suggests that smaller bandwidths  $(h_1, h_2)$  are preferred to improve the rate of the convergence of this estimator. However, the bandwidths  $(h_1, h_2)$  needed to satisfy the restrictions given in (2.15)-(2.17). In addition, bandwidths that are too small cannot be used in finite samples, as this is likely to cause the zero denominator problem in local constant estimation. In Section 4.1, the sensitiveness of the estimators' performance to bandwidths is investigated numerically.

(ii) The overall convergence rate of  $\hat{\alpha}(z)$  relies on  $MSE(\hat{\alpha}(z)) = O(n^{-1} + h_1^4 + h_2^4 + h_1^2 h_2^2)$ . With (2.15)-(2.17), it's not hard to see that  $MSE(\hat{\alpha}(z))$  is approaching  $O(n^{-1})$  when  $\gamma_1$  and  $\gamma_2$  approach  $-1/4$  from the above. However, this lower bound can never be reached, due to the strict inequality in (2.17). Therefore, the rate of convergence for  $\hat{\alpha}(z)$  can be made arbitrarily close to  $n^{-1/2}$  as the bandwidth order is arbitrarily close to  $n^{-1/4}$ . In this sense, our estimator  $\hat{\alpha}(z)$  is nearly  $\sqrt{n}$ -consistent. We note that in nonparametric smoothing, a similar near  $\sqrt{n}$ -consistency may be obtained by using higher order kernels,



which requires that the unknown function is sufficiently smooth, a condition not needed for our result. The super convergence rate obtained here is mainly due to the fact that  $\beta_2(z, w) = \alpha(z) - \alpha(w)$  depends on the additive form of a single univariate function  $\alpha(\cdot)$ , and the nonstationarity of  $x_t$ .

(iii) The above comments on bandwidth conditions and MSE convergence rates also apply to the other estimators except  $\hat{\beta}_{1s}(z)$ .

(iv) If the functional coefficient  $\beta(z)$  is a constant, or equivalently,  $\alpha(z) = 0$ , the limiting distributions of  $\hat{\alpha}(z)$  and  $\hat{\alpha}^*(z)$  become degenerate. Further analysis for this case is provided in Section 3.

**Remark 2.5. (Comments on the asymptotic variance)** We note that the two estimators of  $\alpha(z)$ ,  $\hat{\alpha}(z)$  and  $\hat{\alpha}^*(z)$ , share the same asymptotic variance with each other. Out of the three estimators of  $c_0$ ,  $\hat{c}_0$  and  $\hat{c}_0^*$  share the same asymptotic variance. The asymptotic variance of  $\tilde{c}_0$  is different by a constant multiplier  $E[f^2(w_s)f^2(z_s)/f^2(z_s, w_s)] \equiv \lambda_f$ , which is unity when  $z_t$  is i.i.d. For the four nearly  $\sqrt{n}$ -consistent estimators of  $\beta(z)$ ,  $\hat{\beta}_{1s}^*(z)$  and  $\hat{\beta}_{1s}^{**}(z)$  share the same asymptotic variance, while  $\hat{\beta}_{2s}(z)$  and  $\hat{\beta}_{3s}(z)$  share the same asymptotical variance. The asymptotic variance of  $\hat{\beta}_{1s}^*(z)$  and  $\hat{\beta}_{1s}^{**}(z)$  has the additional constant multiplier  $\lambda_f$ . Analytical comparison of the efficiency of these estimators in the general case may not be possible. Numerical comparisons are explored in Section 4.1.

### 3 Testing Linearity against Semiparametric Specification

In functional coefficient models, it is of both theoretical and empirical importance to know whether the functional coefficient  $\beta(z)$  is a constant. This is to test whether linearity or semiparametric specification fits better for a sample of observations. For the functional coefficient cointegration models, related tests were available in Xiao (2009) and Sun et al. (2016), to cite a few. Under the current non-cointegrated setting, Gan et al. (2014) studied such a testing issue and proposed two tests, of which the one based on the semiparametric estimator of  $\alpha(z)$  shows superior performance in simulation studies.

With more efficient estimators for  $\alpha(z)$  and  $\beta(z)$ , we propose improved tests for checking linearity against the semiparametric specification, following that of Gan et al. (2014). More specifically, our test statistic is based on  $\bar{I}_\alpha = n^{-1} \sum_{t=1}^n \hat{\alpha}(z_t)' \hat{\alpha}(z_t)$ . As we have pointed out in Remark 2.4 (iv),  $\sqrt{n}(\hat{\alpha}(z) - \alpha(z))$  has a degenerate distribution when  $\alpha(z) = 0$ . To figure out the order of  $\bar{I}_\alpha$ , we investigate below the limiting properties of

$\hat{\alpha}(z)$  when  $\alpha(z) = 0$ .

First we note that if  $\beta(z)$  is a constant, then  $B_1(z) = B_2(z) = B_3(z) = 0$ . Thus all the bias terms in Theorem 2.1 become zeros. In the study of the limit properties of  $\hat{\alpha}(z)$ , we obtain the following decomposition (see (A.18) for more details)

$$\sqrt{n}\{\hat{\alpha}(z) - \alpha(z) - B_2(z)\} = \left( \int B_x B'_x \right)^{-1} V_{n2}^*(z) + O_p(\sqrt{n}c_n^2) - \frac{1}{\sqrt{n}} \sum_t \alpha(z_t), \quad (3.1)$$

where  $V_{n2}^*(z) = O_p(1/\sqrt{nh_1}) = o_p(1)$  provided that  $nh_1 \rightarrow \infty$ . Note that Assumption 1 (f) ensures that the second term on the RHS of (3.1) is of order  $o_p(1)$ . When  $\alpha(z) \neq 0$ , the third term on the RHS of (3.1) is of order  $O_p(1)$ . Thus the third term dominates the RHS of (3.1) and determines the asymptotic distribution of  $\hat{\alpha}(z)$ . Under the null that  $\alpha(z) = 0$ , the third term and the bias  $B_2(z)$  disappear. Then we need to compare the first and second terms on the RHS of (3.1). Since  $\sqrt{nc_n^2}\sqrt{nh_1} = O(\log n/(\sqrt{h_1}h_2)) \rightarrow \infty$ , the second term now is the leading term. Therefore, we have  $\hat{\alpha}(z) - \alpha(z) = O_p(c_n^2)$  when  $\alpha(z) = 0$ .

To see the smallest order that  $c_n^2$  can achieve, we assume  $h_1 = h_2 = h = O(n^\gamma)$ , which leads to  $c_n = h^2 + \sqrt{\log n/nh^2}$ . As a result, it's easy to see that  $c_n^2$  has the smallest order  $O(n^{-2/3} \log n)$ , when  $\gamma = -1/6$ . Thus, the fastest convergence rate that  $\hat{\alpha}(z)$  can achieve under the null is  $O(n^{-2/3} \log n)$ . Therefore, we consider the test statistic

$$I_\alpha = \frac{n^{4/3}}{(\log n)^2} \bar{I}_\alpha = \frac{n^{1/3}}{(\log n)^2} \sum_{t=1}^n \hat{\alpha}(z_t)' \hat{\alpha}(z_t). \quad (3.2)$$

It is easy to see that  $I_\alpha = O_p(1)$  under the null.

Under the alternative, we have  $\hat{\alpha}(z) = O_p(1)$ . Thus  $I_\alpha = O_p(n^{4/3}/(\log n)^2)$ , which diverges to infinity. Therefore,  $I_\alpha$  is a valid statistic that can be used to test whether  $\alpha(z) = 0$ , or  $\beta(z)$  is a constant. However, the asymptotic distribution of  $\hat{\alpha}(z)$  under the null is quite involved to obtain, like that of the test studied by Gan et al. (2014). Furthermore, the asymptotic limit distribution may not provide accurate approximation for small or moderate samples, as found in other nonparametric testing problems, for example, Li and Wang (1998). We therefore follow Gan et al. (2014) to suggest a bootstrap procedure for the practical implementation of the test.

*Bootstrap steps for the  $I_\alpha$  test*

Step (i) First obtain the nonparametric estimator  $\hat{\alpha}(z_t)$  as discussed earlier for  $t = 1, \dots, n$ . With these estimates, we can compute the statistic  $I_\alpha$ . Then obtain the three-step estimator  $\hat{\beta}_{3s}(z_t)$  for  $t = 1, \dots, n$  and compute  $\hat{u}_t = y_t - x'_t \hat{\beta}_{3s}(z_t)$ . Let  $\hat{\epsilon}_t = \hat{u}_t - \hat{u}_{t-1}$  and  $\sigma^2 = \frac{1}{n-1} \sum_{t=2}^n \hat{\epsilon}_t^2$ .

Step (ii) Generate i.i.d.  $\epsilon_t^*$  from  $N(0, \sigma^2)$  and  $u_t^* = \sum_{s=1}^t \epsilon_s^*$  for  $t = 1, \dots, n$ . Compute  $y_t^* = x_t' \check{c}_0 + u_t^*$ , where  $\check{c}_0$  is the OLS estimate of  $c_0$  by regressing  $\Delta y_t$  on  $\Delta x_t$ . Using the bootstrap sample  $\{y_t^*, x_t, z_t\}_{t=1}^n$  to compute the bootstrap statistic  $I_\alpha^{(b)}$ .

Step (iii) Repeat Step (ii) a large number of times, say,  $B$  times, and use the upper  $a$ -percentile of  $\{I_\alpha^{(b)}\}_{b=1}^B$  to approximate the upper  $a$ -percentile critical value of the null distribution of  $I_\alpha$ .

Alternatively, we can use the re-centered estimator  $\hat{\alpha}^*(z)$  to replace  $\hat{\alpha}(z)$  in the calculation of  $I_\alpha$ . We denote the corresponding statistic as  $I_\alpha^*$ . We also note that test can be constructed based on the estimators of  $\beta(z)$ , following the same idea of the construction of  $I_\alpha$ . For example, based on the three-step estimator  $\hat{\beta}_{3s}(z)$ , we consider the quantity  $\bar{I}_\beta = n^{-1} \sum_t [\hat{\beta}_{3s}(z_t) - \check{c}_0]' [\hat{\beta}_{3s}(z_t) - \check{c}_0]$ , where  $\check{c}_0$  is the OLS estimate of the constant coefficient by regression  $\Delta y_t$  on  $\Delta x_t$ . Note that under the null,  $\hat{\beta}_{3s}(z) - \check{c}_0 = O_p(1/\sqrt{n})$ . Then we have  $\bar{I}_\beta = O_p(1/n)$  under the null. Therefore, we construct the test as  $I_\beta = n\bar{I}_\beta$ . It's not hard to see that  $I_\beta = O_p(n)$  under the alternative.

We summarize our findings in the following corollary.

**Corollary 3.1.** *Under the same conditions with Theorem 2.1,*

- (i) *with  $h_1 = h_2 = O(n^{-1/6})$ , we have  $I_\alpha = O_p(1)$  and  $I_\alpha^* = O_p(1)$  under the null of  $\alpha(z) = 0$ , and  $I_\alpha = O_p(n^{4/3}/(\log n)^2)$  and  $I_\alpha^* = O_p(n^{4/3}/(\log n)^2)$  under the alternative of  $\alpha(z) \neq 0$ ;*
- (ii) *under the null, we have  $I_\beta = O_p(1)$ , while under the alternative, we have  $I_\beta = O_p(n)$ ;*
- (iii) *under the null, we have  $\sqrt{nh_1}(\hat{\beta}_{1s}(z) - c_0) \xrightarrow{d} N(0, \Sigma_1)$ , for  $\hat{\theta} \in \{\check{c}_0, \hat{\beta}_{1s}^*(z), \hat{\beta}_{1s}^{**}(z)\}$ , we have  $\sqrt{n}(\hat{\theta} - c_0) \xrightarrow{d} N(0, \Sigma_2)$ , and for  $\tilde{\theta} \in \{\hat{c}_0, \hat{c}_0^*, \hat{\beta}_{2s}(z), \hat{\beta}_{3s}(z)\}$ , we have  $\sqrt{n}\{\tilde{\theta} - c_0\} \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1})$ .*

**Remark 3.1.** *We can see our test statistics  $I_\alpha$  and  $I_\alpha^*$  diverge at rate  $O_p(n^{4/3}/(\log n)^2)$  under the alternative, which is faster than that of  $\hat{J}_b$  proposed by Gan et al. (2014) ( $\hat{J}_b = O_p(nh)$  under the alternative). The test statistic  $I_\beta$  also diverges faster than  $\hat{J}_b$ , but slower than  $I_\alpha$  and  $I_\alpha^*$ . In Section 4.2, we compare the finite sample performance of these tests regarding size, power and local power.*

## 4 Simulation

Numerical studies in this section focus on finite sample estimation and testing performance. Section 4.1 compares the estimation accuracy of our proposed estimators with

those proposed by [Sun et al. \(2011\)](#). Section 4.2 compares our test statistics with that of [Gan et al. \(2014\)](#).

## 4.1 Compare the estimation accuracy

In this subsection, we consider the finite sample performance of all the available estimators, and examine the sensitiveness of their performance to the bandwidths. The estimators of [Sun et al. \(2011\)](#) (marked as SHL) and these proposed in Section 2.1 are included for comparison.

We consider the following data generating process:

$$y_t = x_t \beta(z_t) + u_t, \quad (4.1)$$

$$x_t = x_{t-1} + \epsilon_{xt}, \quad (4.2)$$

$$u_t = u_{t-1} + \epsilon_{ut}, \quad (4.3)$$

where  $\epsilon_{xt}$  and  $\epsilon_{ut}$  are independent  $N(0, 1)$ , and they are independent with each other. We consider three different processes for  $z_t$ :

- (1)  $z_t$  is iid uniform  $[-1, 1]$ ;
- (2)  $z_t = v_t + v_{t-1}$ , where  $v_t$  is iid uniform  $[0, 2]$ ;
- (3)  $z_t = 0.5z_{t-1} + v_t$ , where  $v_t$  is iid uniform  $[-1, 1]$ .

The first design assumes an independent process for  $z_t$  while the latter two allow serial correlation in  $z_t$ . The second design has been used by [Sun et al. \(2011\)](#). Following [Sun et al. \(2011\)](#), we consider two functions for  $\alpha(z)$ :

$$(1) \alpha(z) = z - 0.5z^2 - E(z - 0.5z^2)$$

$$(2) \alpha(z) = \sin(z) - E \sin(z).$$

We set  $\beta(z) = \alpha(z) + c_0$ . Without loss of generality, we set  $c_0 = 1$ .

We use Gaussian kernel in the computation. To investigate the role of bandwidth, we adopt the formula  $h = c \cdot \hat{\sigma}_z \cdot n^{-1/5}$  for all the bandwidths involved, where  $\hat{\sigma}_z$  is the sample standard deviation of  $z_t$ . To investigate the sensitivity of the performance with respect to bandwidth, the constant  $c$  varies from 0.2 to 2 with a step length of 0.2. From Remark 2.4 (ii), we know that for those nearly  $\sqrt{n}$ -consistent estimators, the unachievable optimal bandwidth order is  $O(n^{-1/4})$ . Thus we expect to see that those estimators perform better at a smaller value of  $c$ , compared to the other  $\sqrt{nh}$ -consistent estimators. Two sample sizes  $n = 50, 100$  are considered.

We adopt the Integrated MSE (IMSE) as a measure of the estimation accuracy for a functional coefficient estimator  $\hat{\theta}(z)$ . The IMSE is defined as

$$IMSE(\hat{\theta}(z)) = B^{-1} \sum_{b=1}^B \left\{ n^{-1} \sum_{t=1}^n [\hat{\theta}^{(b)}(z_t) - \theta(z_t)]^2 \right\}, \quad (4.4)$$

where  $\hat{\theta}^{(b)}(z)$  denotes the estimate obtained in the  $b$ -th replication. We use the conventional MSE to measure the estimation accuracy regarding  $c_0$ :

$$MSE(\hat{c}_0) = B^{-1} \sum_{b=1}^B [\hat{c}_0^{(b)} - c_0]^2.$$

The results are obtained for  $B = 200$  replications.

To demonstrate the sensitivity with respect to the bandwidth, we plot the IMSEs/MSEs against the constant  $c$ . We use  $c^*$  to denote the optimal value of  $c$  at which the smallest IMSE/MSE is achieved within the considered region. We only report the results for quadratic  $\alpha(z)$  since results for the *sin* function are similar. Plots are collected in Figures 1, 2 and 3 for three different designs of  $z_t$ , respectively.

The major findings from Figure 1 for independent  $z_t$  are summarized as follows. First, in terms of the estimation of  $\alpha(z)$ , our two estimators perform very close to each other, with  $\hat{\alpha}^*(z)$  a bit better than  $\hat{\alpha}(z)$ . They are both uniformly better than SHL's two estimators within the considered bandwidth region under both sample sizes. Our estimators achieve the best performance at a smaller bandwidth (both at  $c^* = 0.8$  when  $n = 50$  and  $c^* = 0.6$  when  $n = 100$ ) than SHL's two estimators (both at  $c^* = 1.4$  when  $n = 50, 100$ ). Second, in terms of the estimation of  $c_0$ , the estimator  $\hat{c}_0^*$  is the best out of our proposed three estimators. Its best performance (achieved at  $c^* = 0.8$  when  $n=50$  and  $c^* = 0.6$  when  $n=100$ ) is significantly better than that of the SHL estimator  $\hat{c}_0$  (achieved at  $c^* = 1.0$  when  $n=50$  and  $c^* = 0.8$  when  $n=100$ ). Third, in terms of the estimation of  $\beta(z)$ , we see that the three-step estimator  $\hat{\beta}_{3s}(z)$  is the best among all the estimators within the considered bandwidth range. The two-step estimator  $\hat{\beta}_{2s}(z)$  makes a very close second best, and then come the two re-centered estimators  $\hat{\beta}_{1s}^*(z)$  and  $\hat{\beta}_{1s}^{**}(z)$ . All the four nearly  $\sqrt{n}$ -consistent estimators are largely better than the SHL estimator when  $c$  is relatively small. For the three-step estimator  $\hat{\beta}_{3s}(z)$ , the optimal constant is  $c^* = 0.8$  for  $n = 50$  and  $c^* = 0.6$  for  $n = 100$ , while for SHL's estimator, it is  $c^* = 1.2$  for both  $n = 50, 100$ . This supports our conjecture that smaller bandwidth should be preferred for the nearly  $\sqrt{n}$ -consistent estimators.

Figures 2 and 3 for serially correlated  $z_t$  reveal similar findings as those from Figure 1. For the estimation of  $\alpha(z)$  and  $\beta(z)$ , our estimators outperform SHL's estimators at a

relatively small bandwidth. We further observe the IMSE curves will intersect with those of SHL’s estimators as the bandwidth increases. Compared with the case with independent  $z_t$ , the intersection happens at a smaller bandwidth. This suggests the range of bandwidth over which our marginal integration based estimators enjoy better performance than those of SHL become narrower for dependent covariate  $z_t$ . For the estimation of  $c_0$ , SHL’s  $\hat{c}_0$  is the best when the sample size is relatively large.

To better illustrate the efficiency gain, we report the ratios of the best IMSE of our estimators ( $\hat{\alpha}^*(z)$  and  $\hat{\beta}_{3s}(z)$ ) to that of SHL’s estimators (SHL $\hat{\alpha}(z)$  and SHL $\hat{\beta}(z)$ ) in Table 1. The levels of the best IMSE of SHL’s estimators are reported in the brackets. We also include the  $c$  values at which the best IMSEs are achieved,  $c^*$ , in the parentheses. Clearly, our estimators achieve significant efficiency gain compared to SHL’s estimators under all the considered scenarios and sample sizes. The improvement with independent  $z_t$  is the most profound. What’s more, we observe that our estimators tend to perform best at around  $c = 0.6$ , while SHL’s estimator at around  $c = 1.2$ . This again supports our claim that smaller bandwidth should be preferred for the nearly  $\sqrt{n}$ -consistent estimators.

Overall speaking, in terms of the estimation of  $\alpha(z)$  and  $\beta(z)$ , our nearly  $\sqrt{n}$ -consistent estimators can vastly outperform SHL’s estimators when small bandwidths are used. The improvement is more pronounced when  $z_t$  is iid. The three-step estimator  $\hat{\beta}_{3s}(z)$  performs the best among the four nearly  $\sqrt{n}$ -consistent estimators. We suggest using  $\hat{\beta}_{3s}(z)$  with a small bandwidth ( $c$  is around 0.6) to estimate  $\beta(z)$  in practice.

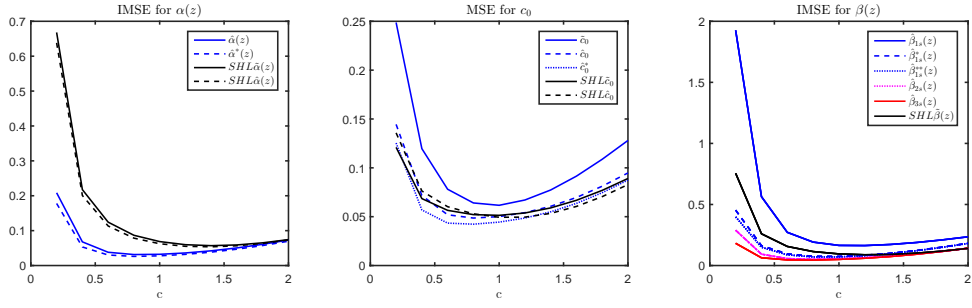
Table 1: IMSE ratios for quadratic  $\beta(z)$

	$z_t$ is iid				$z_t$ is MA(1)				$z_t$ is AR(1)			
	$\hat{\alpha}^*(z)$	SHL $\hat{\alpha}(z)$	$\hat{\beta}_{3s}(z)$	SHL $\hat{\beta}(z)$	$\hat{\alpha}^*(z)$	SHL $\hat{\alpha}(z)$	$\hat{\beta}_{3s}(z)$	SHL $\hat{\beta}(z)$	$\hat{\alpha}^*(z)$	SHL $\hat{\alpha}(z)$	$\hat{\beta}_{3s}(z)$	SHL $\hat{\beta}(z)$
$n = 50$	<b>0.489</b> (0.8)	[0.054] (1.4)	<b>0.515</b> (0.8)	[ 0.088] (1.2)	<b>0.821</b> (0.6)	[0.152] (1.2)	<b>0.611</b> (0.6)	[0.193] (1.0)	<b>0.764</b> (0.6)	[0.122] (1.2)	<b>0.627</b> (0.6)	[0.157] (1.2)
$n = 100$	<b>0.257</b> (0.6)	[0.040] (1.4)	<b>0.310</b> (0.6)	[ 0.063] (1.2)	<b>0.931</b> (0.6)	[0.084] (1.0)	<b>0.742</b> (0.6)	[0.102] (1.0)	<b>0.715</b> (0.6)	[0.089] (1.2)	<b>0.516</b> (0.6)	[0.104] (1.2)

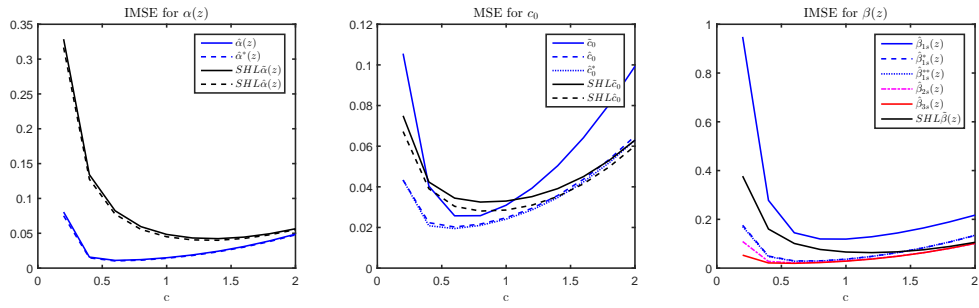
## 4.2 Compare the test performance

This subsection compares the performance of our proposed statistics ( $I_\alpha$ ,  $I_\alpha^*$  and  $I_\beta$ ) with that of the  $\hat{J}_b$  statistic proposed by Gan et al. (2014)<sup>2</sup>. The DGP is the same with Section 4.1. Following Gan et al. (2014), we consider  $\epsilon_{xt}$  is iid  $N(0, 2^2)$  with  $x_0 = 0$ , and  $\epsilon_{ut}$  is iid  $N(0, 1)$  with  $u_0 = 0$ . The stationary covariate  $z_t = v_t + v_{t-1} + \epsilon_{xt}$ , where  $v_t$  is iid uniform  $[0, 2]$ . We also include the iid case  $z_t = v_t$  into consideration.

<sup>2</sup>We compare our test with  $\hat{J}_b$  because it is the best according to the simulation study of Gan et al. (2014).

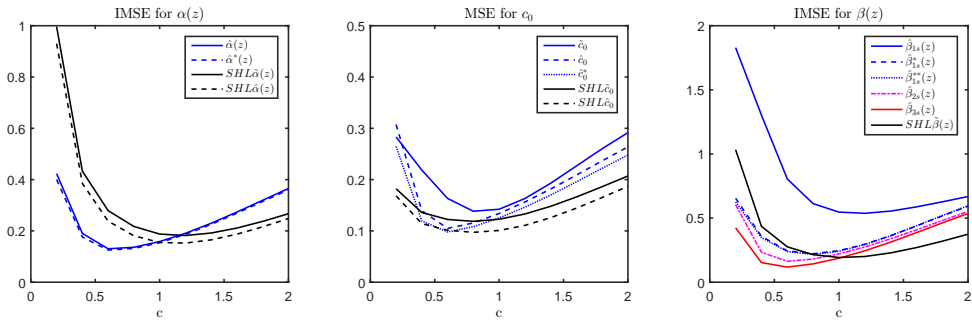


(a)  $n = 50$

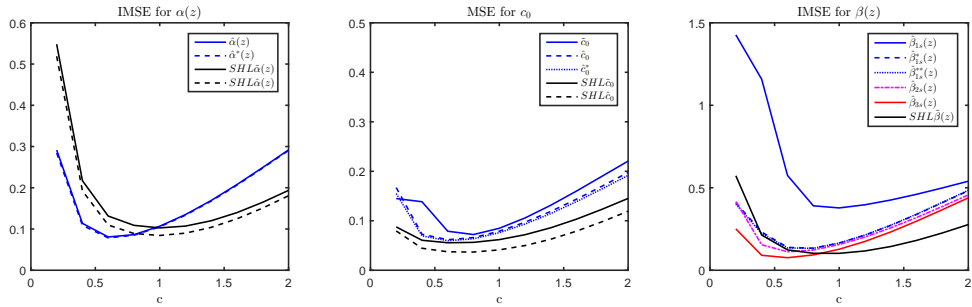


(b)  $n = 100$

Figure 1: Estimation accuracy when  $z_t$  is iid



(a)  $n = 50$



(b)  $n = 100$

Figure 2: Estimation accuracy when  $z_t$  is MA(1)

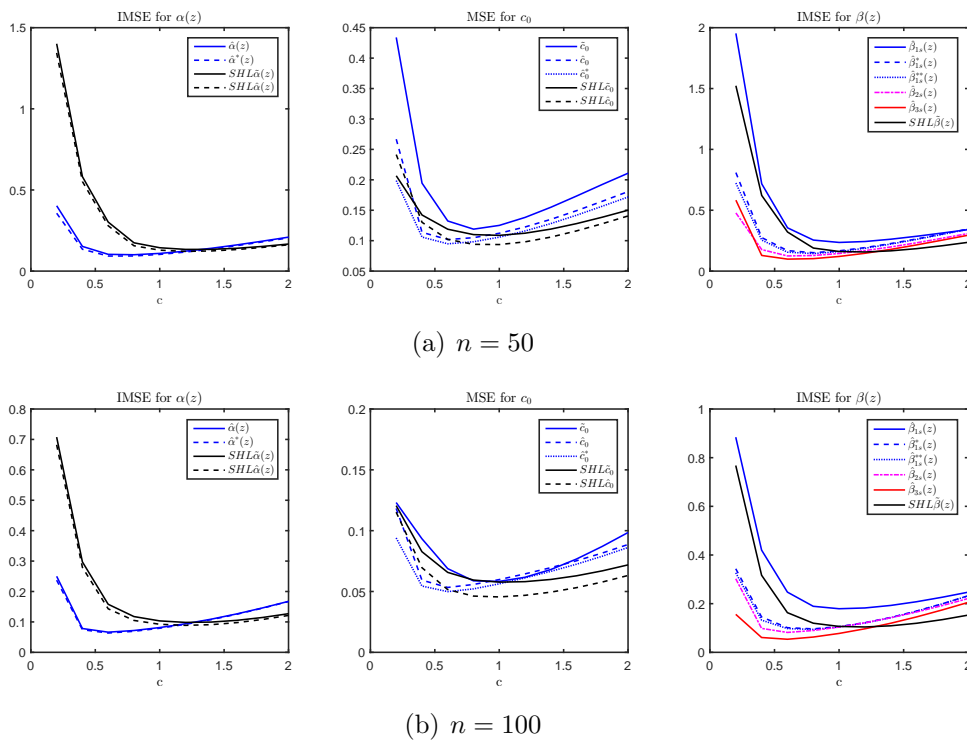


Figure 3: Estimation accuracy when  $z_t$  is AR(1)

Without loss of generality, we set  $\beta(z) = 1$  under the null. Under the alternative, we consider  $\beta(z) = z - 0.5z^2$  and  $\beta(z) = 1/(1 + e^{-z})$ , following Gan et al. (2014). Two sample sizes  $n = 50, 100$  are considered. The number of replications is 200, and within each replication, 200 bootstrap statistics are generated to yield 5% upper percentile values of the bootstrap statistic.

In the computation of  $I_\alpha$  and  $I_\alpha^*$ , the bandwidths are determined by  $h_1 = h_2 = 0.5 \cdot \hat{\sigma}_z \cdot n^{-1/6}$ . The order  $-1/6$  is to ensure that  $\hat{\alpha}(z)$  and  $\hat{\alpha}^*(z)$  achieve the fastest convergence rate under the null. The constant 0.5 is selected based on the observations in Section 4.1 that  $\hat{\alpha}(z)$  and  $\hat{\alpha}^*(z)$  have the best performance around  $c = 0.6$  when the bandwidth order is  $-1/5$ . In the computation of  $I_\beta$ , we use  $h_3 = 0.5 \cdot \hat{\sigma}_z \cdot n^{-1/5}$ . For the test of Gan et al. (2014), we follow their design and adopt  $h = \hat{\sigma}_z \cdot n^{-1/5}$ .

We summarize the size and power in Table 2. From Table 2, we can see  $\hat{J}_b$  is severely under-sized. Our three tests are a bit under-sized when  $z_t$  is serially correlated and correlated with the innovations of  $x_t$ . When  $z_t$  is independent, our tests is a bit over-sized when  $n=50$ , but has satisfying size when  $n=100$ . In terms of power, our three tests outperform the test of Gan et al. (2014), especially when  $z_t$  is iid. This is consistent with the observations in Section 4.1 that our estimators can achieve more efficiency gain when  $z_t$  is independent. The power improvement is more profound under the alternative that



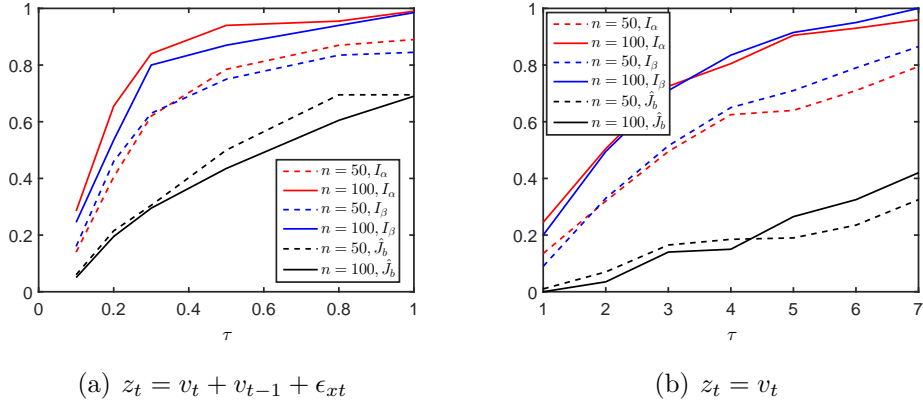


Figure 4: Local power against  $\beta(z) = \tau(z - 0.5z^2)/n^{1/2}$

$$\beta(z) = 1/(1 + e^{-z}).$$

Since the power improvement against  $\beta(z) = z - 0.5z^2$  is less profound, we further examine the local power performance against the local alternative  $\beta(z; \tau) = \tau(z - 0.5z^2)/n^{1/2}$ , where  $\tau$  controls the signal strength. In view of the divergence rates in Remark 3.1, we expect that  $I_\beta$  and  $\hat{J}_b$  have no power under this case, while  $I_\alpha$  and  $I_\alpha^*$  still have good power. From the local power plots in Figure 4, we observe that the rejection rate of  $\hat{J}_b$  is decreasing as sample size increases, suggesting no power. However, our test  $I_\alpha$  has very good local power as  $n$  increases. The performance of  $I_\alpha^*$  is very close to that of  $I_\alpha$ , and thus is omitted. We further observe that  $I_\beta$  still has power, which might be the finite sample phenomenon.

Overall speaking, all the tests have satisfying size, while our tests  $I_\alpha$  and  $I_\alpha^*$  have significantly better (local) power than the test of Gan et al. (2014). Thus, we recommend using  $I_\alpha$  or  $I_\alpha^*$  to check the constancy of the functional coefficient.

Table 2: Size and Power

	size: $\beta(z) = 1$				power: $\beta(z) = z - z^2/2$				power: $\beta(z) = 1/(1 + e^{-z})$			
	$I_\alpha$	$I_\alpha^*$	$I_\beta$	$\hat{J}_b$	$I_\alpha$	$I_\alpha^*$	$I_\beta$	$\hat{J}_b$	$I_\alpha$	$I_\alpha^*$	$I_\beta$	$\hat{J}_b$
$z_t = v_t + v_{t-1} + \epsilon_{xt}$												
$n = 50$	0.025	0.030	0.020	0.000	0.915	0.895	0.870	0.860	0.860	0.860	0.900	0.615
$n = 100$	0.035	0.020	0.035	0.000	0.960	0.960	0.945	0.925	0.975	0.975	0.990	0.635
$z_t = v_t$												
$n = 50$	0.085	0.090	0.060	0.005	0.805	0.855	0.895	0.325	0.820	0.815	0.790	0.425
$n = 100$	0.065	0.080	0.050	0.000	1.000	1.000	0.990	0.485	0.975	0.980	1.000	0.530

## 5 Conclusions

This paper focuses on the case where cointegration does not exist in the functional-coefficient regression model. Since the conventional estimator is inconsistent under this case, [Sun et al. \(2011\)](#) proposed a  $\sqrt{nh}$ -consistent estimator for the functional coefficient. However, the asymptotic distribution of their estimator is unavailable. In this paper, we provide four nearly  $\sqrt{n}$ -consistent estimators for the functional coefficient based on marginal integration and back fitting technique. Our estimators are found to share standard asymptotic distributions. Numerical studies show that the three-step estimator is the most efficient one and can achieve great efficiency gain compared to that of [Sun et al. \(2011\)](#).

Based on the improved estimators, we further propose three test statistics that can be used to test the constancy of the functional coefficient. Compared to the test of [Gan et al. \(2014\)](#), which is based on the estimator of [Sun et al. \(2011\)](#), our tests enjoy faster divergence rates under the alternative. Simulations show that our tests have profound power gain compared to that of [Gan et al. \(2014\)](#), and can detect the local alternatives when the test of [Gan et al. \(2014\)](#) has no power.

One important direction to extend this study is to examine the performance of our method under the cointegration case. We believe our estimators remain consistent when cointegration exists. With this being said, our estimators are robust to the true relation between two integrated variables. As the asymptotic theories under the cointegration case require a lot effort, we leave this as a separate study.

## Acknowledgements

Tu would like to thank support from National Natural Science Foundation of China (Grant 71532001, 71671002) and that from the Center for Statistical Science at Peking University.

## Appendix

This appendix contains three parts. Appendix [A](#) proves [Theorem 2.1](#). Appendix [B](#) proves [Corollary 3.1](#). Appendix [C](#) is to justify that the two estimators of the functional coefficient  $\beta(z)$  proposed by [Sun et al. \(2011\)](#) are actually equivalent to each other.

### A Proof of [Theorem 2.1](#)

**First step estimation:** We first derive the asymptotic distribution of the first step estimator  $\hat{\beta}_{1s}(z)$  given in [\(2.6\)](#) and that of  $\hat{\alpha}(z)$  given in [\(2.7\)](#).

Denote  $w_t = z_{t-1}$ ,  $\theta(z_t, w_t) = (\beta'_1(z_t, w_t), \beta'_2(z_t, w_t))'$ , and  $X'_t = [(\Delta x_t)', x'_{t-1}]$ . Then [\(2.3\)](#) can be written as  $\Delta y_t = X'_t \theta(z_t, w_t) + \Delta u_t$ . Denote the bandwidth for  $z_t$  and  $w_t$  as  $h_1$  and  $h_2$ , respectively. Denote  $D_n = \text{diag}\{I_k, \sqrt{n}I_k\}$ . The kernel weighted least square (KLS) estimator of  $\theta(z, w)$  is given by

$$\hat{\theta}(z, w) = \left( \sum_t X_t X'_t K_{tz} K_{tw} \right)^{-1} \sum_t X_t \Delta y_t K_{tz} K_{tw},$$

where  $K_{tz} = K((z_t - z)/h_1)$ ,  $K_{tw} = K((w_t - w)/h_2)$ . By some algebra, we have

$$\begin{aligned} & D_n(\hat{\theta}(z, w) - \theta(z, w)) \\ &= \left[ \frac{1}{nh_1 h_2} D_n^{-1} \sum_t X_t X'_t K_{tz} K_{tw} D_n^{-1} \right]^{-1} \frac{1}{nh_1 h_2} D_n^{-1} \sum_t X_t X'_t (\theta(z_t, w_t) - \theta(z, w)) K_{tz} K_{tw} \\ &+ \left[ \frac{1}{nh_1 h_2} D_n^{-1} \sum_t X_t X'_t K_{tz} K_{tw} D_n^{-1} \right]^{-1} \frac{1}{nh_1 h_2} D_n^{-1} \sum_t X_t \Delta u_t K_{tz} K_{tw} \\ &\equiv [\Omega_n(z, w)]^{-1} B_n(z, w) + [\Omega_n(z, w)]^{-1} V_n(z, w), \end{aligned} \tag{A.1}$$

where the definitions of  $\Omega_n(z, w)$ ,  $B_n(z, w)$  and  $V_n(z, w)$  should be obvious.

Let's first look at  $\Omega_n(z, w)$ . Under [Assumption 1 \(a\)-\(c\)](#), we have

$$\begin{aligned} \Omega_n(z, w) &= \frac{1}{nh_1 h_2} \begin{pmatrix} \sum_t \Delta x_t \Delta x'_t & \sum_t (\Delta x_t) x'_{t-1} / \sqrt{n} \\ \sum_t (x_{t-1} / \sqrt{n}) \Delta x'_t & \sum_t (x_{t-1} / \sqrt{n}) (x'_{t-1} / \sqrt{n}) \end{pmatrix} K_{tz} K_{tw} \\ &\xrightarrow{d} f(z, w) \begin{pmatrix} E[\Delta x_t \Delta x'_t | z_t = z, w_t = w] & E[\Delta x_t | z_t = z, w_t = w] \int_0^1 B'_x(r) dr \\ \int_0^1 B_x(r) dr E[\Delta x'_t | z_t = z, w_t = w] & \int_0^1 B_x(r) B'_x(r) dr \end{pmatrix} \\ &= f(z, w) \begin{pmatrix} \Sigma_{\epsilon_x} & 0 \\ 0 & \int B_x B'_x \end{pmatrix} \\ &\equiv f(z, w) M, \end{aligned}$$

where  $x_{t-1}/\sqrt{n} \xrightarrow{d} B_x(r)$ ,  $r = (t-1)/n$ , as  $n$  goes to infinity. The joint convergence in distribution is based on the Cramér's Theorem (Davidson, 1994, p.355). The definition of  $M$  should be obvious. In view of Assumption 1 (f), the uniform convergence rate of the kernel density estimator (Stone, 1983) and the rate of convergence of a random walk to Brownian motion (Fraser, 1973)<sup>3</sup>, we have

$$\Omega_n(z, w) = f(z, w)M + O_p(c_n),$$

where  $c_n = h_1^2 + h_2^2 + a_n$ ,  $a_n = \sqrt{\frac{\log n}{nh_1h_2}}$ . Then we have

$$[\Omega_n(z, w)]^{-1} = f^{-1}(z, w)W + O_p(c_n), \quad (\text{A.2})$$

where

$$W \equiv M^{-1} = \begin{pmatrix} \Sigma_{\epsilon_x}^{-1} & 0 \\ 0 & (\int B_x B_x')^{-1} \end{pmatrix}.$$

Then we look at  $B_n(z, w)$ . We have

$$\begin{aligned} & B_n(z, w) \\ &= \frac{1}{nh_1h_2} D_n^{-1} \sum_t X_t X_t' (\theta(z_t, w_t) - \theta(z, w)) K_{tz} K_{tw} \\ &= \left( \begin{array}{l} \frac{1}{nh_1h_2} \sum (\Delta x_t) (\Delta x_t)' (\beta_1(z_t, w_t) - \beta_1(z, w)) K_{tz} K_{tw} + \frac{1}{nh_1h_2} \sum (\Delta x_t) (x'_{t-1}) (\beta_2(z_t, w_t) - \beta_2(z, w)) K_{tz} K_{tw} \\ \frac{1}{n^{3/2}h_1h_2} \sum (x_{t-1}) (\Delta x_t)' (\beta_1(z_t, w_t) - \beta_1(z, w)) K_{tz} K_{tw} + \frac{1}{n^{3/2}h_1h_2} \sum (x_{t-1}) (x'_{t-1}) (\beta_2(z_t, w_t) - \beta_2(z, w)) K_{tz} K_{tw} \end{array} \right) \\ &\equiv \begin{pmatrix} B_{n1}(z, w) + B_{n2}(z, w) \\ B_{n3}(z, w) + B_{n4}(z, w) \end{pmatrix}, \end{aligned} \quad (\text{A.3})$$

where the definitions of  $B_{ni}(z, w)$ ,  $i = 1, 2, 3, 4$ , should be obvious. Below we analyze these terms in sequence.

$$\begin{aligned} B_{n1}(z, w) &= \frac{1}{nh_1h_2} \sum (\Delta x_t) (\Delta x_t)' (\beta_1(z_t, w_t) - \beta_1(z, w)) K_{tz} K_{tw} \\ &= \frac{1}{nh_1h_2} \sum (\Delta x_t) (\Delta x_t)' (\beta(z_t) - \beta(z)) K_{tz} K_{tw} \\ &= h_1^2 \mu_2(K) E[(\Delta x_t) (\Delta x_t)' | z_t = z, w_t = w] [\beta'(z) f'_z(z, w) + \frac{1}{2} \beta''(z) f(z, w)] \{1 + o_p(1)\} \\ &= h_1^2 \mu_2(K) \Sigma_{\epsilon_x} [\beta'(z) f'_z(z, w) + \frac{1}{2} \beta''(z) f(z, w)] \{1 + o_p(1)\}. \end{aligned}$$

---

<sup>3</sup>(Fraser, 1973) assumed independent innovation in the derivation. We believe the uniform convergence rate will not be altered by the serial correlation. Under Assumption 1 (f), the rate obtained by (Fraser, 1973) is dominated by the uniform convergence rate of the kernel density estimator  $O_p(c_n)$ .

$$\begin{aligned}
n^{-1/2}B_{n2}(z, w) &= \frac{n^{-1/2}}{nh_1h_2} \sum (\Delta x_t)(x'_{t-1})(\beta_2(z_t, w_t) - \beta_2(z, w))K_{tz}K_{tw} \\
&\xrightarrow{d} \mu_2(K)E[\Delta x_t|z_t = z, w_t = w] \int B'_x(r)dr \\
&\quad \cdot \{h_1^2[\beta'(z)f'_z(z, w) + \frac{1}{2}\beta''(z)f(z, w)] - h_2^2[\beta'(w)f'_w(z, w) + \frac{1}{2}\beta''(w)f(z, w)]\} \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
B_{n3}(z, w) &= \frac{1}{n^{3/2}h_1h_2} \sum (x_{t-1})(\Delta x_t)'(\beta_1(z_t, w_t) - \beta_1(z, w))K_{tz}K_{tw} \\
&\xrightarrow{d} h_1^2\mu_2(K) \int B_x(r)dr E[(\Delta x_t)'|z_t = z, w_t = w] \left[ \beta'(z)f'_z(z, w) + \frac{1}{2}\beta''(z)f(z, w) \right] \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
n^{-1/2}B_{n4}(z, w) &= \frac{1}{n^2h_1h_2} \sum x_{t-1}x'_{t-1}(\beta_2(z_t, w_t) - \beta_2(z, w))K_{tz}K_{tw} \\
&\xrightarrow{d} \mu_2(K) \int B_x(r)B'_x(r)dr \\
&\quad \cdot \{h_1^2[\beta'(z)f'_z(z, w) + \frac{1}{2}\beta''(z)f(z, w)] - h_2^2[\beta'(w)f'_w(z, w) + \frac{1}{2}\beta''(w)f(z, w)]\}.
\end{aligned}$$

Combining all the above results, we get

$$D_n^{-1}B_n(z, w) \xrightarrow{d} \mu_2(K)MB(h_1, h_2, z, w), \quad (\text{A.4})$$

where

$$B(h_1, h_2, z, w) = \begin{pmatrix} h_1^2[\beta'(z)f'_z(z, w) + \frac{1}{2}\beta''(z)f(z, w)] \\ h_1^2[\beta'(z)f'_z(z, w) + \frac{1}{2}\beta''(z)f(z, w)] - h_2^2[\beta'(w)f'_w(z, w) + \frac{1}{2}\beta''(w)f(z, w)] \end{pmatrix}.$$

Plugging (A.2) into (A.1), we get

$$\begin{aligned}
&D_n(\hat{\theta}(z, w) - \theta(z, w)) \\
&= [f^{-1}(z, w)W + O_p(c_n)][B_n(z, w) + V_n(z, w)] \\
&= f^{-1}(z, w)WB_n(z, w) + f^{-1}(z, w)WV_n(z, w) + O_p(c_n^2) \begin{pmatrix} \mathbf{1}_{k \times 1} \\ \sqrt{n}\mathbf{1}_{k \times 1} \end{pmatrix}, \quad (\text{A.5})
\end{aligned}$$

where  $\mathbf{1}_{k \times 1}$  denotes a  $k \times 1$  vector of unity. Now we prove (A.5). From (A.4), owing to the uniform convergence rate of kernel density estimation and the fact that  $B(h_1, h_2, z, w) = O(h_1^2 + h_2^2)$ , it's not hard to see that  $D_n^{-1}B_n(z, w) = O_p(h_1^2 + h_2^2) + O_p(a_n) = O_p(c_n)$ , for any  $(z, w)$ . Therefore,  $O_p(c_n)B_n(z, w) = O_p(c_n^2)D_n\mathbf{1}_{2k \times 1} = O_p(c_n^2)(\mathbf{1}'_{k \times 1}, \sqrt{n}\mathbf{1}'_{k \times 1})'$ . Similarly, since  $EV_n(z, w) = 0$ , we have  $V_n(z, w) = O_p(a_n)$ , for any  $(z, w)$ . Then we have  $O_p(c_n)[B_n(z, w) + V_n(z, w)] = O_p(c_n^2)(\mathbf{1}'_{k \times 1}, \sqrt{n}\mathbf{1}'_{k \times 1})'$ . This proves (A.5).

Let's further rewrite (A.5) as

$$D_n[\hat{\theta}(z, w) - \theta(z, w) - D_n^{-1}f^{-1}(z, w)WB_n(z, w)] = f^{-1}(z, w)WV_n(z, w) + O_p(c_n^2) \begin{pmatrix} 1_{k \times 1} \\ \sqrt{n}1_{k \times 1} \end{pmatrix}. \quad (\text{A.6})$$

In view of (A.4), we have

$$D_n^{-1}f^{-1}(z, w)WB_n(z, w) \xrightarrow{p} \mu_2(K)f^{-1}(z, w)B(h_1, h_2, z, w). \quad (\text{A.7})$$

Plugging (A.7) into (A.6), we get

$$D_n(\hat{\theta}(z, w) - \theta(z, w) - \mu_2(K)f^{-1}(z, w)B(h_1, h_2, z, w)) = f^{-1}(z, w)WV_n(z, w) + O_p(c_n^2) \begin{pmatrix} 1_{k \times 1} \\ \sqrt{n}1_{k \times 1} \end{pmatrix}. \quad (\text{A.8})$$

Note that (A.8) is true for any  $(z, w)$ . Now we evaluate (A.8) at  $\{(z, z_t)\}_{t=1}^n$  and then take average, we get

$$\begin{aligned} & D_n \left[ \begin{pmatrix} \frac{1}{n} \sum_t [\hat{\beta}_1(z, z_t) - \beta_1(z, z_t)] \\ \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] \end{pmatrix} - \mu_2(K)n^{-1} \sum_t f^{-1}(z, z_t)B(h_1, h_2, z, z_t) \right] \\ &= \begin{pmatrix} \Sigma_{\epsilon_x}^{-1} \frac{1}{n^2 h_1 h_2} \sum_t \sum_s \Delta x_s \Delta u_s K((z_s - z)/h_1) K((w_s - z_t)/h_2) f^{-1}(z, z_t) \\ (\int B_x B'_x)^{-1} \frac{1}{n^2 h_1 h_2} \sum_t \sum_s \frac{x_s - 1}{\sqrt{n}} \Delta u_s K((z_s - z)/h_1) K((w_s - z_t)/h_2) f^{-1}(z, z_t) \end{pmatrix} + O_p(c_n^2) \begin{pmatrix} 1_{k \times 1} \\ \sqrt{n}1_{k \times 1} \end{pmatrix} \\ &\equiv \begin{pmatrix} \Sigma_{\epsilon_x}^{-1} V_{n1}^*(z) \\ (\int B_x B'_x)^{-1} V_{n2}^*(z) \end{pmatrix} + O_p(c_n^2) \begin{pmatrix} 1_{k \times 1} \\ \sqrt{n}1_{k \times 1} \end{pmatrix}, \end{aligned} \quad (\text{A.9}) \quad (\text{A.10})$$

where the definitions of  $V_{n1}^*(z)$  and  $V_{n2}^*(z)$  should be obvious.

For the bias term in (A.9), we have

$$\begin{aligned} & \mu_2(K)n^{-1} \sum_t f^{-1}(z, z_t)B(h_1, h_2, z, z_t) \\ & \xrightarrow{p} \mu_2(K)E[f^{-1}(z, z_t)B(h_1, h_2, z, z_t)] \\ &= \mu_2(K) \begin{pmatrix} h_1^2 \{ \beta'(z) E[f'_z(z, z_t)/f(z, z_t)] + \frac{1}{2} \beta''(z) \} \\ h_1^2 \{ \beta'(z) E[f'_z(z, z_t)/f(z, z_t)] + \frac{1}{2} \beta''(z) \} - h_2^2 E[\beta'(z_t) f'_w(z, z_t)/f(z, z_t) + \frac{1}{2} \beta''(z_t)] \end{pmatrix} \\ &\equiv \begin{pmatrix} h_1^2 B_1(z) \\ h_1^2 B_1(z) - h_2^2 B_3(z) \end{pmatrix} \end{aligned} \quad (\text{A.11})$$

$$\equiv \begin{pmatrix} h_1^2 B_1(z) \\ B_2(z) \end{pmatrix}, \quad (\text{A.12})$$

where the definitions of  $B_1(z)$ ,  $B_2(z)$  and  $B_3(z)$  should be obvious.

Then we analyze the first term in (A.10). Below we will show

$$\begin{aligned} V_{n1}^*(z) &= \frac{1}{nh_1} \sum_s K((z_s - z)/h_1) \epsilon_s^* + o_p((nh_1)^{-1/2}) \\ &\equiv \bar{V}_{n1}(z) + o_p((nh_1)^{-1/2}), \end{aligned} \quad (\text{A.13})$$

where  $\epsilon_s^* = \Delta x_s \Delta u_s f(w_s) f^{-1}(z, w_s)$ , and

$$\begin{aligned} V_{n2}^*(z) &= \frac{1}{nh_1} \sum_s K((z_s - z)/h_1) \epsilon_s^* + o_p((nh_1)^{-1/2}) \\ &\equiv \bar{V}_{n2}(z) + o_p((nh_1)^{-1/2}), \end{aligned} \quad (\text{A.14})$$

where  $\epsilon_s^* = \frac{x_{s-1}}{\sqrt{n}} \Delta u_s f(w_s) f^{-1}(z, w_s)$ .

The proof of (A.13) and (A.14) follows the idea of Fan et al. (1998). To save space, we only prove (A.13) here. The proof of (A.14) is quite similar. Let's look at the difference between  $V_{n1}^*(z)$  and  $\bar{V}_{n1}(z)$ , which is

$$\begin{aligned} &V_{n1}^*(z) - \bar{V}_{n1}(z) \\ &= \frac{1}{n^2 h_1 h_2} \sum_{t \neq s-1} \sum_s \Delta x_s \Delta u_s K((z_s - z)/h_1) K((w_s - z_t)/h_2) f^{-1}(z, z_t) \\ &\quad - \frac{1}{nh_1} \sum_s K((z_s - z)/h_1) \epsilon_s^* + o_p((nh_1)^{-1/2}) \\ &= \frac{1}{n^2} \sum_{t \neq s-1} \sum_s \epsilon_{xs} \Delta u_s h_1^{-1} K((z_s - z)/h_1) [h_2^{-1} K((w_s - z_t)/h_2) f^{-1}(z, z_t) - f(w_s) f^{-1}(z, w_s)] + o_p((nh_1)^{-1/2}) \\ &\equiv \frac{1}{n^2} \sum_{t \neq s-1} \sum_s \epsilon_{xs} \Delta u_s h_1^{-1} K((z_s - z)/h_1) V_{s,t} + o_p((nh_1)^{-1/2}) \\ &\equiv v(z) + o_p((nh_1)^{-1/2}), \end{aligned} \quad (\text{A.15})$$

where the first argument is because the sum with  $t = s - 1$  is of order  $O_p(n^{-3/2} h_1^{-1/2} h_2^{-1}) = o_p((nh_1)^{-1/2})$ , as  $K((w_s - z_t)/h_2) = K(0)$  when  $t = s - 1$ . It's easy to see that

$$V_{s,t} = h_2^{-1} K\left(\frac{w_s - z_t}{h_2}\right) f^{-1}(z, z_t) - f(w_s) f^{-1}(z, w_s),$$

and the definition of  $v(z)$  is obvious. To show (A.13), we only need to show  $v(z) = o_p((nh_1)^{-1/2})$ .

Note that for  $t \neq s - 1$ , we have

$$\begin{aligned} E_t V_{s,t} &= \int h_2^{-1} K\left(\frac{w_s - z_t}{h_2}\right) f^{-1}(z, z_t) f(z_t) dz_t - f(w_s) f^{-1}(z, w_s) \\ &= \int K(u) f^{-1}(z, w_s + h_2 u) f(w_s + h_2 u) du - f(w_s) f^{-1}(z, w_s) \\ &= o(1). \end{aligned}$$

It's easy to see that  $E v(z) = 0$ . So to prove  $v(z) = o_p((nh_1)^{-1/2})$ , we only need to show  $E[v(z)]^2 = o((nh_1)^{-1})$ . Note that

$$E[v(z)]^2 = \frac{1}{n^4} \sum_{t \neq s-1} \sum_s \sum_{k \neq \ell-1} \sum_\ell E[\epsilon_{xs} \Delta u_s h_1^{-1} K((z_s - z)/h_1) V_{s,t} \epsilon_{x\ell} \Delta u_\ell h_1^{-1} K((z_\ell - z)/h_1) V_{\ell,k}].$$

Since  $E(\Delta u_s \Delta u_\ell | \{\epsilon_{xt}\}) = 0$  if  $s \neq \ell$  as assumed in Assumption (1) (b), we have

$$E[v(z)]^2 = \frac{1}{n^4} \sum_{t \neq s-1} \sum_s \sum_{k \neq s-1} E[\epsilon_{xs}^2 (\Delta u_s)^2 h_1^{-2} K_{sz}^2 V_{s,t} V_{s,k}].$$

There are  $n^2$  terms with  $t = k$ , which can be ignored as the sum of the terms with  $t = k$  is of order  $O((n^2 h_1)^{-1})$ . Then we get

$$\begin{aligned}
E[v(z)]^2 &= \frac{1}{n^4} \sum_{t \neq s-1, k \neq s-1, t \neq k} E[\epsilon_{xs}^2 (\Delta u_s)^2 h_1^{-2} K_{sz}^2 V_{s,t} V_{s,k}] \\
&= \frac{1}{n^4} \sum_{t \neq s-1, k \neq s-1, t \neq k} E \epsilon_{xs}^2 (\Delta u_s)^2 h_1^{-2} K_{sz}^2 E_t[V_{s,t}] E_k[V_{s,k}] \\
&= O(n^{-1} h_1^{-1}) \cdot o(1) \\
&= o(n^{-1} h_1^{-1}).
\end{aligned}$$

By now we have proved (A.13).

It's easy to see that

$$\sqrt{nh_1} \bar{V}_{n1}(z) \xrightarrow{d} N(0, \nu_0(K) f(z) \Sigma_{\epsilon^*}),$$

where

$$\Sigma_{\epsilon^*} = E[\epsilon_t^* (\epsilon_t^*)'] = E \left[ \frac{f^2(w_t)}{f^2(z, w_t)} \epsilon_{xt} \epsilon_{xt}' (\Delta u_t)^2 \right] = \sigma_{\Delta_u}^2 E [f^2(w_t) / f^2(z, w_t)] \Sigma_{\epsilon_x}.$$

Similarly,

$$\sqrt{nh_1} \bar{V}_{n2}(z) \xrightarrow{d} N(0, \nu_0(K) f(z) \Sigma_{\epsilon^*}),$$

where

$$\Sigma_{\epsilon^*} = \sigma_{\Delta_u}^2 E [f^2(w_t) / f^2(z, w_t)] \int B_x(r) B_x'(r) dr.$$

Therefore, we have  $V_{n1}^*(z) = O_p(1/\sqrt{nh_1})$  and  $V_{n2}^*(z) = O_p(1/\sqrt{nh_1})$ .

Note that  $\beta_1(z, w) = \beta(z)$ , thus  $\hat{\beta}_{1s}(z) - \beta(z) = \frac{1}{n} \sum_t [\hat{\beta}_1(z, z_t) - \beta_1(z, z_t)]$ . For the first element in (A.9), since  $\sqrt{nh_1} c_n^2 \rightarrow 0$ , we have

$$\sqrt{nh_1} (\hat{\beta}_{1s}(z) - \beta(z) - h_1^2 B_1(z)) = \Sigma_{\epsilon_x}^{-1} \sqrt{nh_1} V_{n1}^*(z) + O_p(\sqrt{nh_1} c_n^2) \xrightarrow{d} N(0, \Sigma_1), \quad (\text{A.16})$$

where  $\Sigma_1 = \nu_0(K) f(z) \sigma_{\Delta_u}^2 E [f^2(w_t) / f^2(z, w_t)] \Sigma_{\epsilon_x}^{-1}$ . Since  $\beta_2(z, w) = \alpha(z) - \alpha(w)$ , then

$$\begin{aligned}
\hat{\alpha}(z) - \alpha(z) &= \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \alpha(z)] \\
&= \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] - \frac{1}{n} \sum_t \alpha(z_t).
\end{aligned} \quad (\text{A.17})$$

In view of the second element in (A.9), we have

$$\begin{aligned}
&\sqrt{n} \{ \hat{\alpha}(z) - \alpha(z) - B_2(z) \} \\
&= \sqrt{n} \left\{ \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] - B_2(z) \right\} - \frac{1}{\sqrt{n}} \sum_t \alpha(z_t) \\
&= \left( \int B_x B_x' \right)^{-1} V_{n2}^*(z) + O_p(\sqrt{n} c_n^2) - \frac{1}{\sqrt{n}} \sum_t \alpha(z_t).
\end{aligned} \quad (\text{A.18})$$



Under Assumption 1 (f), we have  $V_{n2}^*(z) = o_p(1)$ ,  $\sqrt{n}c_n^2 \rightarrow 0$ . Then the dominant term in (A.18) is  $\frac{1}{\sqrt{n}} \sum_t \alpha(z_t)$ , which is  $O_p(1)$ . Therefore, we have

$$\sqrt{n}\{\hat{\alpha}(z) - \alpha(z) - B_2(z)\} \xrightarrow{d} N(0, \Gamma(\alpha(z_t))),$$

where  $\Gamma(\alpha(z_t))$  denotes the long run variance of the series  $\{\alpha(z_t)\}$ .

**The re-centered estimators:** For  $\hat{\alpha}^*(z)$ , we have

$$\begin{aligned} \hat{\alpha}^*(z) - \alpha(z) &= \hat{\alpha}(z) - n^{-1} \sum_t \hat{\alpha}(z_t) - \alpha(z) \\ &= \frac{1}{n} \sum_t \hat{\beta}_2(z, z_t) - \frac{1}{n} \sum_t \beta_2(z, z_t) + \frac{1}{n} \sum_t \beta_2(z, z_t) - \frac{1}{n} \sum_t \hat{\alpha}(z_t) - \alpha(z) \\ &= \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] + \frac{1}{n} \sum_t [\hat{\alpha}(z_t) - \alpha(z_t)] \\ &= -n^{-1} \sum_t \alpha(z_t) + \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] \\ &\quad - \frac{1}{n} \sum_t \frac{1}{n} \sum_s [\hat{\beta}_2(z_t, z_s) - \beta_2(z_t, z_s)], \end{aligned} \tag{A.19}$$

where the second last line is due to the fact that  $\beta_2(z, z_t) = \alpha(z) - \alpha(z_t)$ , and the last line is based on equation (A.17).

We know the first term in (A.19) is of order  $O_p(1/\sqrt{n})$ . Except for the bias, the second term in (A.19) is of order  $O_p\left(\frac{1}{n\sqrt{h_1}} + c_n^2\right) = O_p(c_n^2)$  (because  $n\sqrt{h_1}a_n^2 \rightarrow \infty$ ), and the third term is also  $O_p(c_n^2)$ . So under the condition that  $\sqrt{n}c_n^2 \rightarrow 0$ , (A.19) is dominated by the first term. Thus asymptotic variance of  $\hat{\alpha}^*(z)$  is the same with that of  $\hat{\alpha}(z)$ . The bias that comes from the second term in (A.19) is  $B_2(z)$ , while that comes from the third term in (A.19) is  $EB_2(z_t)$  (the derivation for this result is very similar with (A.12) and thus is omitted). Finally, we have

$$\sqrt{n}\{\hat{\alpha}^*(z) - \alpha(z) - B_4(z)\} \xrightarrow{d} N(0, \Gamma(\alpha(z_t))),$$

where  $B_4(z) = B_2(z) - EB_2(z_t)$ .

Then we look at  $\tilde{c}_0$ . We have

$$\begin{aligned} \tilde{c}_0 - c_0 &= \frac{1}{n} \sum_t \hat{\beta}_{1s}(z_t) - c_0 \\ &= \frac{1}{n} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)] + \frac{1}{n} \sum_t \alpha(z_t). \end{aligned} \tag{A.20}$$

Except for the bias, the first term in (A.20) is of order  $O_p(1/\sqrt{n} + c_n^2) = O_p(1/\sqrt{n})$ , and the second term is also of order  $O_p(1/\sqrt{n})$ . Let's first analyse the first term. We have obtained the asymptotics of  $\hat{\beta}_{1s}(z) - \beta(z)$ . In view of (A.10) and (A.12), we have

$$\hat{\beta}_{1s}(z_t) - \beta(z_t) - h_1^2 B_1(z_t) = \Sigma_{\epsilon_x}^{-1} V_{n1}^*(z_t) + O_p(c_n^2),$$

and  $V_{n1}^*(z) = \bar{V}_{n1}(z) + o_p((nh_1)^{-1/2}) = \bar{V}_{n1}(z)\{1 + o_p(1)\}$ . So the leading term of the sample average  $\frac{1}{n} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)]$  is  $\frac{1}{n} \sum_t \bar{V}_{n1}(z_t)$ , which admits

$$\begin{aligned}
\frac{1}{n} \sum_t \bar{V}_{n1}(z_t) &= \frac{1}{n} \sum_t \frac{1}{nh_1} \sum_s K\left(\frac{z_s - z_t}{h_1}\right) \epsilon_s^* \\
&= \frac{1}{n^2 h_1} \sum_t \sum_s \Delta x_s \Delta u_s f(w_s) f^{-1}(z_t, w_s) K\left(\frac{z_s - z_t}{h_1}\right) \\
&= \frac{1}{n} \sum_s \Delta x_s \Delta u_s f(w_s) f^{-1}(z_s, w_s) f(z_s) + o_p(n^{-1/2}) \\
&\equiv \frac{1}{n} \sum_s \xi_s + o_p(n^{-1/2}), \tag{A.21}
\end{aligned}$$

where the definition of  $\xi_s$  should be obvious. Below we prove (A.21). First we have

$$\begin{aligned}
&\frac{1}{n} \sum_t \bar{V}_{n1}(z_t) - \frac{1}{n} \sum_s \xi_s \\
&= \frac{1}{n^2} \sum_t \sum_s \Delta x_s \Delta u_s f(w_s) [h_1^{-1} f^{-1}(z_t, w_s) K\left(\frac{z_s - z_t}{h_1}\right) - f^{-1}(z_s, w_s) f(z_s)] \\
&= \frac{1}{n^2} \sum_{t \neq s} \sum_s \Delta x_s \Delta u_s f(w_s) U_{st} + o_p(n^{-1/2}) \\
&\equiv U + o_p(n^{-1/2}),
\end{aligned}$$

where the definitions of  $U_{st}$  and  $U$  are obvious. To prove (A.21), we only need to show  $U = o_p(n^{-1/2})$ . Note that

$$\begin{aligned}
E_t U_{st} &= \int h_1^{-1} f^{-1}(z_t, w_s) K\left(\frac{z_s - z_t}{h_1}\right) f(z_t) dz_t - f^{-1}(z_s, w_s) f(z_s) \\
&= \int h_1^{-1} f^{-1}(z_s + h_1 u, w_s) K(u) f(z_s + h_1 u) h_1 du - f^{-1}(z_s, w_s) f(z_s) \\
&= o(1).
\end{aligned}$$

It is easy to see that  $EU = 0$ . Below we will show  $EUU' = o(n^{-1})$ .

$$\begin{aligned}
EUU' &= E\left[\frac{1}{n^4} \sum_{t \neq s} \sum_s \sum_{\ell \neq k} \sum_k \Delta x_s \Delta u_s f(w_s) U_{st} (\Delta x_k)' \Delta u_k f(w_k) U_{k\ell}\right] \\
&= \frac{1}{n^4} \sum_{t \neq s} \sum_s \sum_{\ell \neq s} E[(\Delta x_s)(\Delta x_s)' (\Delta u_s)^2 f^2(w_s) U_{st} U_{s\ell}] \\
&= \frac{1}{n^4} \sum_{t \neq s, t \neq \ell} \sum_s \sum_{\ell \neq s} E[(\Delta x_s)(\Delta x_s)' (\Delta u_s)^2 f^2(w_s) U_{st} U_{s\ell}] + o(n^{-1}) \\
&= O(n^{-1}) \cdot o(1) + o(n^{-1}) \\
&= o(n^{-1}),
\end{aligned}$$

where the second line is due to the fact that  $E(\Delta u_s \Delta u_k | \{\Delta x_\ell\}) = 0$  for  $s \neq k$  (Assumption 1 (b)), the third line is because there are  $n^2$  terms with  $t = \ell$ , and their summation is of order

$O(n^{-2}) = o(n^{-1})$ , the second last argument is due to  $E_t U_{st} = o(1)$ . Up to now we have proved (A.21).

It's easy to see that

$$\frac{1}{\sqrt{n}} \sum_s \xi_s \xrightarrow{d} N(0, \Sigma_2),$$

where  $\Sigma_2 = \sigma_{\Delta u}^2 E[f^2(w_s) f^2(z_s) / f^2(z_s, w_s)] \Sigma_{\epsilon_x}$ , since there is no serial correlation in  $\Delta u_t$  (Assumption (1) (b)). Then we have

$$\sqrt{n} \frac{1}{n} \sum_t \{\hat{\beta}_{1s}(z_t) - \beta(z_t) - h_1^2 E[B_1(z_t)]\} \xrightarrow{d} N(0, \Sigma_2).$$

Since  $Cov(\xi_s, \alpha(z_t)) = 0$ , because of the independence assumption given in Assumption (1) (b). Thus we have

$$\sqrt{n} \{\tilde{c}_0 - c_0 - h_1^2 E B_1(z_t)\} \xrightarrow{d} N(0, \Sigma_2 + \Gamma(\alpha(z_t))),$$

Then we look at  $\hat{\beta}_{1s}^*(z)$ . We have

$$\begin{aligned} \hat{\beta}_{1s}^*(z) - \beta(z) &= \hat{\alpha}(z) - \alpha(z) + n^{-1} \sum_t \hat{\beta}_{1s}(z_t) - c_0 \\ &= \hat{\alpha}(z) - \alpha(z) + n^{-1} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)] + \frac{1}{n} \sum_t \beta(z_t) - c_0 \\ &= \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] - \frac{1}{n} \sum_t \alpha(z_t) \\ &\quad + n^{-1} \sum_t n^{-1} \sum_s [\hat{\beta}_1(z_t, z_s) - \beta_1(z_t, z_s)] + \frac{1}{n} \sum_t \alpha(z_t) \\ &= \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] + \frac{1}{n} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)]. \end{aligned} \quad (\text{A.22})$$

Except for the bias term, the first term in (A.22) is of order  $O_p(1/n\sqrt{h_1} + c_n^2)$ , and the second term is  $O_p(1/\sqrt{n} + c_n^2) = O_p(1/\sqrt{n})$ . So the second term dominates (A.22). Therefore, we have

$$\sqrt{n} \{\hat{\beta}_{1s}^*(z) - \beta(z) - B_5(z)\} \xrightarrow{d} N(0, \Sigma_2),$$

where  $B_5(z) = B_2(z) + h_1^2 E B_1(z_t)$ .

For the estimator  $\hat{\beta}_{1s}^{**}(z)$ , we have  $\hat{\beta}_{1s}^{**}(z) - \beta(z) = \hat{\alpha}^*(z) - \alpha(z) + \tilde{c}_0 - c_0$ . In view of (A.19) and (A.20), we have

$$\begin{aligned} \hat{\beta}_{1s}^{**}(z) - \beta(z) &= \frac{1}{n} \sum_t [\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] - \frac{1}{n} \sum_t \frac{1}{n} \sum_s [\hat{\beta}_2(z_t, z_s) - \beta_2(z_t, z_s)] \\ &\quad + \frac{1}{n} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)]. \end{aligned}$$

Thus the asymptotic distribution is dominated by the term  $\frac{1}{n} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)]$ . Therefore, we have

$$\sqrt{n} \left\{ \hat{\beta}_{1s}^{**}(z) - \beta(z) - B_6(z) \right\} \xrightarrow{d} N(0, \Sigma_2),$$

where  $B_6(z) = B_4(z) + h_1^2 EB_1(z)$ .

**Two-step estimation:** Let's first analyze  $\hat{c}_0$ . By OLS, we have

$$\hat{c}_0 = \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \Delta y_t^* = c_0 + \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \Delta u_t^*,$$

where  $u_t^* = y_t - x_t' \hat{\alpha}(z_t) - x_t' c_0$ , and

$$\begin{aligned} \Delta u_t^* &= y_t - x_t' \hat{\alpha}(z_t) - x_t' c_0 - [y_{t-1} - x_{t-1}' \hat{\alpha}(z_{t-1}) - x_{t-1}' c_0] \\ &= \Delta y_t - x_t' \hat{\alpha}(z_t) + x_{t-1}' \hat{\alpha}(z_{t-1}) - (\Delta x_t)' c_0 \\ &= \Delta u_t + x_t' [\alpha(z_t) - \hat{\alpha}(z_t)] - x_{t-1}' [\alpha(z_{t-1}) - \hat{\alpha}(z_{t-1})]. \end{aligned}$$

Then we have

$$\begin{aligned} \hat{c}_0 &= c_0 + \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \Delta u_t \\ &\quad + \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \{x_t' [\alpha(z_t) - \hat{\alpha}(z_t)] - x_{t-1}' [\alpha(z_{t-1}) - \hat{\alpha}(z_{t-1})]\} \\ &\equiv c_0 + A_{1n} + \left[ n^{-1} \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} A_{2n}, \end{aligned}$$

where the definitions of  $A_{1n}$  and  $A_{2n}$  should be obvious. It is easy to see that  $A_{1n} = O_p(n^{-1/2})$ .

Below we analyze the order of  $A_{2n}$ .

Recall that  $\hat{\alpha}(z) - \alpha(z) = B_2(z) - \frac{1}{n} \sum_s \alpha(w_s) + O_p(1/n\sqrt{h_1}) + O_p(c_n^2) = B_2(z) - \frac{1}{n} \sum_s \alpha(w_s) + o_p(1/\sqrt{n})$ . Then we have

$$\begin{aligned} A_{2n} &= \frac{1}{n} \sum_t \Delta x_t \{x_t' [\alpha(z_t) - \hat{\alpha}(z_t)] - x_{t-1}' [\alpha(z_{t-1}) - \hat{\alpha}(z_{t-1})]\} \\ &= \frac{1}{n} \sum_t \Delta x_t (\Delta x_t)' \frac{1}{n} \sum_s \alpha(w_s) - \frac{1}{n} \sum_t \Delta x_t \Delta [x_t' B_2(z_t)] + O_p(c_n^2). \end{aligned}$$

Therefore,

$$\begin{aligned} &\hat{c}_0 - c_0 + [n^{-1} \sum_t \Delta x_t (\Delta x_t)']^{-1} \frac{1}{n} \sum_t \Delta x_t \Delta [x_t' B_2(z_t)] \\ &= \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \Delta u_t + \frac{1}{n} \sum_s \alpha(w_s) + O_p(c_n^2) \tag{A.23} \\ &= \frac{1}{n} \sum_t \left[ [n^{-1} \sum_s \epsilon_{xs} \epsilon_{xs}']^{-1} \epsilon_{xt} \Delta u_t + \alpha(z_t) \right] + O_p(c_n^2) \\ &= O_p(1/\sqrt{n}). \end{aligned}$$

Note that the bias term on the left hand side is of order  $O_p(h_1^2 + h_2^2)$ , and it converges in probability to  $\mathcal{B}_1 = \Sigma_{\epsilon_x}^{-1} E \Delta x_t \Delta [x_t' B_2(z_t)]$ .

Denote  $e_t = [n^{-1} \sum_s \epsilon_{xs} \epsilon'_{xs}]^{-1} \epsilon_{xt} \Delta u_t + \alpha(z_t)$ . It's easy to see that  $Ee_t = 0$ . Now we compute the long run variance of  $e_t$  and get

$$\Gamma(e_t) = \frac{1}{n} \sum_t \sum_s Ee_t e'_s = \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1} + \Gamma(\alpha(z_t)), \quad (\text{A.24})$$

because of Assumption 1 (b). Then we get

$$\sqrt{n}(\hat{c}_0 - c_0 + \mathcal{B}_1) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1} + \Gamma(\alpha(z_t))).$$

The analysis for  $\hat{c}_0^*$  is quite similar. Note that  $\hat{\alpha}^*(z) - \alpha(z) = B_4(z) - n^{-1} \sum_t \alpha(z_t) + o_p(1/\sqrt{n})$ . We finally can obtain

$$\sqrt{n}(\hat{c}_0^* - c_0 + \mathcal{B}_1^*) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1} + \Gamma(\alpha(z_t))),$$

where  $\mathcal{B}_1^* = \Sigma_{\epsilon_x}^{-1} E \Delta x_t \Delta [x'_t B_4(z_t)]$ .

Now we analyze the two-step estimator  $\hat{\beta}_{2s}(z)$ . Note that  $\hat{\beta}_{2s}(z) = \hat{c}_0 + \hat{\alpha}(z)$ . Then we have

$$\begin{aligned} \hat{\beta}_{2s}(z) - \beta(z) &= \hat{c}_0 - c_0 + \hat{\alpha}(z) - \alpha(z) \\ &= B_2(z) - \mathcal{B}_1 + \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \Delta u_t + o_p(1/\sqrt{n}), \end{aligned} \quad (\text{A.25})$$

where the second argument is due to (A.17) and (A.23). So we have

$$\sqrt{n}(\hat{\beta}_{2s}(z) - \beta(z) - \mathcal{B}_2(z)) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1}),$$

where the bias is  $\mathcal{B}_2(z) = B_2(z) - \mathcal{B}_1$ .

**Three-step estimation:** Now we analyze the three-step estimator  $\hat{\beta}_{3s}(z)$ . Note that

$$\hat{\beta}_{3s}(z) = \left[ \sum_t x_t x'_t K((z_t - z)/h_3) \right]^{-1} \sum_t x_t \tilde{y}_t K((z_t - z)/h_3),$$

where  $\tilde{y}_t = \Delta y_t + x'_{t-1} \hat{\beta}_{2s}(z_{t-1}) = x'_t \beta(z_t) + x'_{t-1} [\hat{\beta}_{2s}(z_{t-1}) - \beta(z_{t-1})] + \Delta u_t$ . Then we have

$$\begin{aligned} \hat{\beta}_{3s}(z) - \beta(z) &= \left[ \sum_t x_t x'_t K((z_t - z)/h_3) \right]^{-1} \sum_t x_t x'_t [\beta(z_t) - \beta(z)] K((z_t - z)/h_3) \\ &\quad + \left[ \sum_t x_t x'_t K((z_t - z)/h_3) \right]^{-1} \sum_t x_t x'_{t-1} [\hat{\beta}_{2s}(z_{t-1}) - \beta(z_{t-1})] K((z_t - z)/h_3) \\ &\quad + \left[ \sum_t x_t x'_t K((z_t - z)/h_3) \right]^{-1} \sum_t x_t \Delta u_t K((z_t - z)/h_3). \end{aligned} \quad (\text{A.26})$$

The first term in (A.26) will produce a bias given by  $h_3^2 \mu_2(K) [f'_z(z) \beta'(z) / f(z) + \frac{1}{2} \beta''(z)]$ . The third term in (A.26) is of order  $O_p(1/n\sqrt{h_3})$ . In view of (A.25) and the fact that  $\hat{\beta}_{2s}(z) - \beta(z) = O_p(1/\sqrt{n} + h_1^2 + h_2^2)$ , the leading term of the second term in (A.26) is

$$\left[ \sum_t x_t x'_t K((z_t - z)/h_3) \right]^{-1} \sum_t x_t x'_{t-1} K((z_t - z)/h_3) \left[ \sum_t \Delta x_t (\Delta x_t)' \right]^{-1} \sum_t \Delta x_t \Delta u_t = O_p(n^{-1/2}).$$

Thus the asymptotic variance of  $\hat{\beta}_{3s}(z)$  would be the same as that of  $\hat{\beta}_{2s}(z)$ , namely

$$\sqrt{n}(\hat{\beta}_{3s}(z) - \beta(z) - \mathcal{B}_3(z)) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1}),$$

where  $\mathcal{B}_3(z) = h_3^2 \mu_2(K) [f'_z(z) \beta'(z) / f(z) + \frac{1}{2} \beta''(z)] + E \mathcal{B}_2(z_t)$ . ■

## B Proof of Corollary 3.1

We only provide a sketch proof.

For the re-centered estimator  $\hat{\alpha}^*(z)$ , from (A.19), we know that  $\hat{\alpha}^*(z) = O_p(c_n^2)$ . Then the order is the same with that of  $\hat{\alpha}(z)$ . Thus we can conclude that the fastest convergence rate it can reach under the null is also  $O_p(n^{-2/3} \log n)$ , and this is achieved with  $h_1 = h_2 = O(n^{-1/6})$ .

For the one-step estimator  $\hat{\beta}_{1s}(z)$ , from the derivation of (A.16), we see that its asymptotic variance is not affected by the true value of  $\alpha(z)$ .

For  $\tilde{c}_0$ , from (A.20), we have  $\tilde{c}_0 - c_0 = \frac{1}{n} \sum_t [\hat{\beta}_{1s}(z_t) - \beta(z_t)]$ . Therefore, we have  $\sqrt{n}(\tilde{c}_0 - c_0) \xrightarrow{d} N(0, \Sigma_2)$ .

For  $\hat{\beta}_{1s}^*(z)$ , from (A.22), we can see its asymptotic distribution is still determined by the second term. So under the null, we still have (2.22), except that the bias term is zero. Analysis for  $\hat{\beta}_{1s}^{**}(z)$  are the same.

For  $\hat{c}_0$ , from (A.23), we know that the asymptotic distribution will be determined by the first term. Thus we have  $\sqrt{n}(\hat{c}_0 - c_0) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1})$ . Analysis for  $\hat{c}_0^*$  are the same.

For the two-step estimator  $\hat{\beta}_{2s}(z)$ , from (A.25), we get  $\sqrt{n}(\hat{\beta}_{2s}(z) - c_0) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1})$ .

For the three-step estimator  $\hat{\beta}_{3s}(z)$ , from (A.26), the fact  $\alpha(z) = 0$  will not change the orders there. The asymptotic distribution is still dominated by the second term. Thus we have  $\sqrt{n}(\hat{\beta}_{3s}(z) - c_0) \xrightarrow{d} N(0, \sigma_{\Delta u}^2 \Sigma_{\epsilon_x}^{-1})$ . ■

## C Proof of the equivalence of the two estimators of Sun et al. (2011)

Following their notations, the estimator  $\hat{\alpha}(z)$  defined in (2.10) can be obtained by shifting a constant vector toward the estimator  $\tilde{\alpha}(z)$  defined in (2.9), namely,  $\hat{\alpha}(z) = \tilde{\alpha}(z) + c$ , where  $c$  is a  $d \times 1$  constant vector<sup>4</sup>. This is because the summation term in (2.10) is indeed a constant for a given sample. Then based on their definitions, we have  $\hat{Y}_t = Y_t - X_t' \hat{\alpha}(Z_t) = Y_t - X_t' [\tilde{\alpha}(Z_t) + c] = \tilde{Y}_t - X_t' c$ . Consequently,  $\Delta \hat{Y}_t = \Delta \tilde{Y}_t - (\Delta X_t)' c$ . Note that  $\hat{c}_0$  is the OLS estimator by regressing  $\Delta \hat{Y}_t$  on  $\Delta X_t$ , and  $\tilde{c}_0$  is the OLS estimator by regressing  $\Delta \tilde{Y}_t$  on  $\Delta X_t$ , it's not hard to get  $\hat{c}_0 = \tilde{c}_0 - c$ . Therefore,  $\hat{\theta}(z) = \hat{c}_0 + \hat{\alpha}(z) = \tilde{c}_0 - c + \tilde{\alpha}(z) + c = \tilde{\theta}(z)$ . This completes the proof. ■

---

<sup>4</sup>Strictly speaking, we should write  $\hat{\alpha}(z) = \tilde{\alpha}(z) + \text{random vector}$ . But for a given sample, the random vector is just a constant vector.

## References

- Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.
- Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.
- Fan, J., Härdle, W., and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26(3):943–971.
- Fraser, D. F. (1973). The rate of convergence of a random walk to brownian motion. *The Annals of Probability*, pages 699–701.
- Gan, L., Hsiao, C., and Xu, S. (2014). Model specification test with correlated but not cointegrated variables. *Journal of Econometrics*, 178:80–85.
- Juhl, T. (2005). Functional-coefficient models under unit root behaviour. *The Econometrics Journal*, 8(2):197–213.
- Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, 87:45–165.
- Phillips, P. C. (2007). Unit root log periodogram regression. *Journal of econometrics*, 138(1):104–124.
- Stone, C. J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. In *Recent advances in statistics*, pages 393–406. Elsevier.
- Sun, Y., Cai, Z., and Li, Q. (2016). A consistent nonparametric test on semiparametric smooth coefficient models with integrated time series. *Econometric Theory*, 32(4):988–1022.
- Sun, Y., Hsiao, C., and Li, Q. (2011). Measuring correlations of integrated but not cointegrated variables: A semiparametric approach. *Journal of econometrics*, 164(2):252–267.
- Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.