LIMIT THEORY AND INFERENCE IN NON-COINTEGRATED FUNCTIONAL COEFFICIENT REGRESSION

By

Ying Wang, Peter C. B. Phillips, Yundong Tu

April 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2399

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Limit Theory and Inference in Non-cointegrated Functional Coefficient Regression*

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

[a]Renmin University of China, [b]University of Auckland,
[c]Yale University, [d]Singapore Management University,
[e]Peking University

April 10, 2024

Abstract

Functional coefficient (FC) cointegrating regressions offer empirical investigators flexibility in modeling economic relationships by introducing covariates that influence the direction and intensity of comovement among nonstationary time series. FC regression models are also useful when formal cointegration is absent, in the sense that the equation errors may themselves be nonstationary, but where the nonstationary series display well-defined FC linkages that can be meaningfully interpreted as correlation measures involving the covariates. The present paper proposes new nonparametric estimators for such FC regression models where the nonstationary series display linkages that enable consistent estimation of the correlation measures between them. Specifically, we develop $\sqrt{n}$-consistent estimators for the functional coefficient and establish their asymptotic distributions, which involve mixed normal limits that facilitate inference. Two novel features that appear in the limit theory are (i) the need for non-diagonal matrix normalization due to the presence of stationary and nonstationary components in the regression; and (ii) random bias elements that appear in the asymptotic distribution of the kernel estimators, again resulting from the nonstationary regression components. Numerical studies reveal that the proposed estimators achieve significant efficiency improvements compared to the estimators suggested in earlier work by Sun et al. (2011). Easily implementable specification tests with standard chi-square asymptotics are suggested to check for constancy of the functional coefficient. These tests are shown to have faster divergence rate under local alternatives and enjoy superior performance in simulations than tests proposed recently in Gan et al. (2014). An empirical application based on the quantity theory of money illustrates the practical use of correlated but non-cointegrated regression relations.

JEL classification: C14; C22.

Keywords: Cointegration; Correlation measure; Functional coefficient regression; Marginal integration; Nonstationary time series.

*Wang and Phillips acknowledge Marsden Fund support at the University of Auckland. Wang acknowledges research support from the National Natural Science Foundation of China (Grant 72103197). Phillips acknowledges research support from the Kelly Fund at the University of Auckland. Tu acknowledges support from the National Natural Science Foundation of China (Grant 71532001, 71671002), China's National Key Research Special Program (2016YFC0207705), the Center for Statistical Science at Peking University, the Key Laboratory of Mathematical Economics and Quantitative Finance at Peking University, and the Ministry of Education for support. Email addresses: yingwang.econ@ruc.edu.cn; peter.phillips@yale.edu; yundong.tu@gsm.pku.edu.cn

1 Introduction

Cointegration has been an important tool in the empirical analysis of long run relationships among nonstationary time series since the seminal work of Granger (1981) and Engle and Granger (1987). While many economic and financial time series show strong evidence of comovement over time, such variables often fall short of being formally cointegrated given the strict criteria involved in models of cointegration. As an alternative approach to analyzing comovement in time series, a functional coefficient version of the standard cointegration model has attracted research attention in recent years. This model, which we call FC-cointegration, serves as a flexible semiparametric device for extending the standard model in a way that can capture more nuanced long run behavior in which the relationship may evolve over time in response to covariate influences that locally adjust comovement among the nonstationary variables.(Xiao, 2009; Cai et al., 2009; Li et al., 2015; Sun et al., 2016; Wang et al., 2016; Tu and Wang, 2019, 2020, 2022; Phillips and Wang, 2023).

In the absence of cointegration or FC-cointegration, purported linear or nonparametric nonlinear relationships among nonstationary variables may be completely spurious in the sense that no meaningful causal associations[1] exist, as when the variables are independently determined as integrated or near integrated processes (Phillips, 1986, 1998, 2009). FC models offer an intermediate possibility. Formal cointegrating linkages may be absent in a model where the equation errors are nonstationary. But nonstationary series may nonetheless display well-defined linkages in terms of explicit estimable functional coefficients that define specific correlation measures among the variables in terms of observable covariates. The present paper proposes new nonparametric estimators for such FC regression models where the nonstationary series display linkages or causal associations that enable consistent estimation of these correlation measures. Such models have been considered in recent research by Sun et al. (2011) and Gan et al. (2014). Specifically, this model has the form

$$y_t = x_t'\beta(z_t) + u_t, \quad t = 1, \cdots, n, \tag{1}$$

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, and u_t is an unobserved unit root innovation process.

The nonstationary error in (1) distinguishes this model from the FC-cointegration model studied by Juhl (2005), Cai et al. (2009), and Xiao (2009), where the innovations are assumed to be stationary. The formulation (1) suggests that the variables x_t and y_t are associated and may well move together over time but not so closely as to ensure the existence of cointegration or even FC-cointegration. In such cases, the association between x_t and y_t may be regarded as weak relative to that of FC-cointegration. The weak association is captured by the function $\beta(\cdot)$ with a covariate z_t that controls the nature and strength of this association.

Models such as (1) are relevant in the study of economic relationships in which the determining economic and financial forces are not strong enough to establish comovement that can be characterized as cointegration or FC-cointegration. One example is purchasing power parity (PPP), where the underlying theory of the impact of relative prices on exchange rates is not always supported by a manifest empirical association due to imperfect information or market frictions that lead to missing integrated factors in the relationships or where the observed adjustment towards PPP may be extremely slow. Other examples may arise in relationships

[1]Since trending time series are inevitably correlated (Phillips, 1986), we interpret causal association as being explicitly model-determined, as distinct from trend coordinate representations in the sense of Phillips (1998).

that sustain persistent measurement errors in data aggregation or through significant omitted variables, as argued in Sun et al. (2011). We call models that fall into the class described by (1) non-cointegrated FC (NC-FC) regressions. These models serve as a semiparametric interface in measuring associations among integrated variables that are neither cointegrated nor FC-cointegrated but not entirely spurious.

In studying models of the form (1) Sun et al. (2011) provided two consistent estimators for $\beta(z)$, but did not provide asymptotic distribution theory. So econometric machinery for conducting inference about the functional coefficients in such systems is presently unavailable. One of the goals of this paper is to address this issue so that a full limit distribution theory is available that reveals the asymptotic behavior of consistent estimators, thereby opening up an inferential framework for studying functional linkages that may exist among non-cointegrating nonstationary time series.

Specifically, this paper constructs new estimators of $\beta(\cdot)$ by means of direct non-parametric estimation of a differenced version of the model. We find that for the initial bivariate nonparametric estimators, the asymptotic distribution is delivered by the approximation error involved in capturing the functional coefficient of the nonstationary regressor x_t instead of from the error term and by central limit theory, as is conventional. This finding is new to the functional coefficient regression literature and is a second main contribution of the paper. Non-diagonal matrix normalization methods[2], and explicit partitioned regression methods are used to isolate the dominant components in this limit theory. Without employing these techniques, the nonparametric approximation error contribution to the limit distribution can be falsely ignored in derivations and can lead to incorrect limit theory, as will be explained later. For a full analysis of this latter phenomena readers are referred to other recent work (Phillips and Wang, 2023).

Based on this approach several new FC estimates are constructed using the initial bivariate nonparametric estimators by means of marginal integration and backfitting techniques. Marginal integration was proposed by Linton and Nielsen (1995), Newey (1994) and Tjøstheim and Auestad (1994) to estimate models with an additive structure. There is now a large literature, see Chen et al. (1996), Linton (1997, 2000), Fan et al. (1998), Yang et al. (2006), Cai and Masry (2000) and Cai and Xiao (2012), among others. Our preferred estimates of the functional quantities $\beta(z)$ and $\alpha(z) := \beta(z) - \mathbb{E}\beta(z_t)$ are consistent and asymptotically normal with variance converging at rate $1/n$ and asymptotic bias diminishing at the usual rate h^2, where h is the smoothing bandwidth used in functional coefficient estimation. The resulting $\sqrt{n}$ rate of convergence in the limit theory is not standard in the semiparametric literature. As a result, the proposed estimators possess improved performance over those in Sun et al. (2011) which have a standard $\sqrt{nh}$ rate.

A third objective of the paper is to construct a test based on our proposed estimators for detecting linearity against the semiparametric specification in (1). This procedure 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 $\alpha(z)$ has better performance in simulations. We show that when $\beta(\cdot)$ degenerates to a constant, our estimators for $\alpha(\cdot)$ become $n\sqrt{h}$-consistent. Given these improved rates of convergence, a natural expectation is that a test statistic constructed from the proposed estimators for $\alpha(\cdot)$ would enjoy better power performance than that of Gan et al. (2014). This conjecture is formally confirmed in our analysis. In particular, we show analytically that our tests diverge at a faster rate than

[2]Such methods have been used in vector autoregression asymptotics (Phillips, 1995) and in recent work on weak instrument and general limit theory (Magdalinos and Phillips, 2019) and other nonstationary asymptotics (Vogelsang and Wagner, 2014; Phillips and Kheifets, 2019)

that of Gan et al. (2014) under the alternative, leading to improved power performance. Finite sample simulation results corroborate the asymptotic analysis.

The remainder of the paper is organized as follows. Section 2 presents the estimation procedures, including one step and two step estimators. Section 3 derives their limit properties and discusses how the results differ from standard nonparametric estimation limit theory. Section 4 considers hypothesis testing on the functional coefficient $\beta(z)$. Numerical results are reported in Section 5 to illustrate the finite sample performance of the estimators and tests. Section 6 provides an empirical illustration of the methods in exploring a quantity theory of money relation where the variables are correlated but not cointegrated. Section 7 concludes. Proofs are given in the Appendix.

In matters of notation, we use '$\rightsquigarrow$' to signify weak convergence of the associated probability measures in the relevant probability space, '$\xrightarrow{p}$' to denote convergence in probability, and definitional equality is shown by ':=' and '=:'. Stochastic processes like Brownian motion $B(r)$ on $[0,1]$ are usually written as B and integrals such as $\int$ are understood to be taken over the interval $[0,1]$ unless otherwise indicated.

2 Estimation

Our approach to estimation starts by differencing the model (1), which leads to a version of the model with stationary errors and stationary and nonstationary regressors. Importantly, the coefficients in this regression equation can be interpreted as functions of two arguments (essentially, z_t and z_{t-1}), which may be estimated by kernel nonparametric estimation. Marginal integration is then used to deliver estimates of the coefficient function $\beta(z)$.

One-step estimation

Differencing (1) gives an equation with a stationary error and two sets of regressors with functional coefficients. To proceed, we start by defining $\alpha(z) = \beta(z) - \beta_0$, where $\beta_0 = \mathbb{E}\beta(z_t)$, so that $\mathbb{E}\alpha(z_t) = 0$. Model (1) is then equivalent up to initial conditions to the differenced equation

$$\begin{aligned}\Delta y_t &= x_t'\beta(z_t) - x_{t-1}'\beta(z_{t-1}) + \Delta u_t && (2)\\ &= (\Delta x_t)'\beta(z_t) + x_{t-1}'[\beta(z_t) - \beta(z_{t-1})] + \Delta u_t \\ &= (\Delta x_t)'\beta(z_t) + x_{t-1}'[\alpha(z_t) - \alpha(z_{t-1})] + \Delta u_t. && (3)\end{aligned}$$

An important aspect of (3) is that its regressors are both stationary (Δx_t) and nonstationary (x_{t-1}) when x_t is an integrated time series. This duality complicates the limit theory in ways that go beyond simple degeneracy of the asymptotic signal matrix, as will become clear later. Another feature of (3) is that the coefficient of x_{t-1} is the differenced function $\beta(z_t) - \beta(z_{t-1})$, which is zero when $\beta(\cdot)$ is constant.

The coefficients of (3) may be interpreted as functions of the two arguments z_t and z_{t-1}.

Specifically, we rewrite (3) as[3]

$$\Delta y_t = (\Delta x_t)'\beta_1(z_t, z_{t-1}) + x'_{t-1}\beta_2(z_t, z_{t-1}) + \Delta u_t, \tag{4}$$

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

$$\int \beta_1(z,w)\ell_1(w)dw = \beta(z), \quad \int \beta_2(z,w)\ell_2(w)dw = \alpha(z), \tag{5}$$

for any weight functions $\ell_1(w)$ and $\ell_2(w)$ satisfying $\int \ell_1(w)dw = 1$ and $\int \ell_2(w)dw = 1$. Noting that $\mathbb{E}\alpha(z_t) = 0$, one choice of such a weight function could be $\ell_1(w) = \ell_2(w) = f(w)$, the density function of z_t.

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 level least squares estimators of $\beta_1(z,w)$ and $\beta_2(z,w)$ in model (4) by $\hat{\beta}_1(z,w)$ and $\hat{\beta}_2(z,w)$. To be precise, let $w_t = z_{t-1}$, and $X'_t = ((\Delta x_t)', x'_{t-1})$. Then (4) can be written as $\Delta y_t = X'_t\beta(z_t, w_t) + \Delta u_t$. Let the bandwidth parameters in nonparametric estimation associated with the variables z_t and w_t be h_1 and h_2. The kernel weighted least squares (KLS) estimator of $\beta(z,w)$ is given by

$$(\hat{\beta}'_1(z,w), \hat{\beta}'_2(z,w))' = \left(\sum_{t=2}^{n} X_t X'_t K_{tz} K_{tw}\right)^{-1} \sum_{t=2}^{n} X_t \Delta y_t K_{tz} K_{tw}, \tag{6}$$

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

Using the empirical density of z_t as the weight function and adopting the trimming technique to avoid the boundary bias problem in local level estimation, the marginal integration estimators are

$$\hat{\beta}_{1s}(z) = \frac{1}{n}\sum_{t=1}^{n} \hat{\beta}_1(z, z_t) 1(z_t \in \mathcal{S}_n), \tag{7}$$

and

$$\hat{\alpha}(z) = \frac{1}{n}\sum_{t=1}^{n} \hat{\beta}_2(z, z_t) 1(z_t \in \mathcal{S}_n), \tag{8}$$

where the subscript "$1s$" in $\hat{\beta}_{1s}(z)$ signifies the one-step estimator of $\beta(z)$, and $\mathcal{S}_n$ is a compact subset of $\mathbb{R}$ that trims out the boundary region of the support of z_t so that weak uniform convergence results can be obtained for $\hat{\beta}_1(z,w)$ and $\hat{\beta}_2(z,w)$ over $(z,w) \in \mathcal{S}_n \times \mathcal{S}_n$. The set $\mathcal{S}_n$ needs to satisfy $P(z_t \in \mathcal{S}_n) \to 1$ as $n \to \infty$ uniformly over $t = 1, \cdots, n$ and then the use of a trimming function in this way will not affect the asymptotic theory. Trimming techniques of this type are now widely adopted in nonparametric regression; see, for example, Sun et al.

[3] As one referee pointed out, model (4) has appeared in the earlier literature, e.g., Cai et al. (2009) and Sun et al. (2016). However, valid asymptotic theory does not follow from these or other previous studies due to the error in the past functional coefficient literature, specifically the ignored random bias component which changes the limit theory in a materially important way, as pointed out in the Introduction of the present paper and analyzed by Phillips and Wang (2023).

(2011) and Sun et al. (2016) and, for the construction of such a trimming set, see Remark 6 in Appendix C of Sun et al. (2016).

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

$$\tilde{\alpha}(z) = \hat{\alpha}(z) - \frac{1}{n}\sum_{t=1}^{n}\hat{\alpha}(z_t), \tag{9}$$

which satisfies $\mathbb{E}\tilde{\alpha}(z) = 0$.

Two-step estimation

An estimator for β_0 can be constructed by means of an alternative regression as follows. First, model (1) can be rewritten as $y_t - x_t'\alpha(z_t) = x_t'\beta_0 + u_t$. Then, with the estimator $\hat{\alpha}(z)$ replacing $\alpha(z)$, we have the implied equation

$$y_t - x_t'\hat{\alpha}(z_t) = x_t'\beta_0 + \tilde{u}_t, \quad \text{with } \tilde{u}_t = u_t + x_t'(\alpha(z_t) - \hat{\alpha}(z_t)).$$

Setting $\tilde{y}_t = y_t - x_t'\hat{\alpha}(z_t)$ and taking differences in the above equation, we get

$$\Delta\tilde{y}_t = (\Delta x_t)'\beta_0 + \Delta\tilde{u}_t.$$

This formulation leads to the backfitted ordinary least squares estimator of β_0

$$\hat{\beta}_0 = \left(\sum_{t=2}^{n}\Delta x_t(\Delta x_t)'\right)^{-1}\sum_{t=2}^{n}\Delta x_t\Delta\tilde{y}_t. \tag{10}$$

In place of least squares, generalized least squares or frequency domain methods might be used in the regression to improve efficiency. Using $\hat{\beta}_0$ the two-step estimator of $\beta(z)$ can be constructed as

$$\hat{\beta}_{2s}(z) = \hat{\alpha}(z) + \hat{\beta}_0. \tag{11}$$

It can be shown that $\hat{\beta}_{2s}(z) = \tilde{\beta}_0 + \tilde{\alpha}(z)$, where $\tilde{\beta}_0$ is an estimator constructed as in (10), with $\hat{\alpha}(z_t)$ replaced by $\tilde{\alpha}(z_t)$ in computing $\tilde{y}_t$. This equivalence parallels that between the two estimators of $\beta(z)$ proposed in Sun et al. (2011), the proof of which is straightforward.

3 Asymptotic properties

To study the asymptotic properties of these estimators we make the following assumptions.

Assumption 1. *(i)* $\{\Delta x_t, z_t, \Delta u_t\}$ *are strictly stationary α-mixing processes of size* $-p/(p-2)$ *(where $p = 2+\delta$ for some small $\delta > 0$) with finite, positive definite long-run variance matrix and finite q moments for some $q > 2p > 4$;*

(ii) $\{\Delta x_t\}$ *and* $\{z_t\}$ *are independent and both are independent of* $\{\Delta u_t\}$*;*

(iii) z_t *has Lebesgue density $f(z)$, (z_t, z_{t-1}) has joint Lebesgue density $f(z,w)$, (z_{t-1}, z_t, z_{t+1}) has joint Lebesgue density $f(w,z,u)$, and $(z_{t-1}, z_t, z_{s-1}, z_s)$ $(s > t+1)$ has joint Lebesgue density $f(w,z,u,v)$. These densities are bounded above and away from zero and are three times continuously differentiable on their respective supports;*

(iv) $\beta(z)$ *is three times continuously differentiable and bounded on the support of* z_t*;*

(v) The kernel function $K(\cdot)$ *is a bounded probability density function symmetric around zero with bounded support. Define* $\mu_j(K) = \int u^j K(u)du$ *and* $\nu_j(K) = \int u^j K^2(u)du$*;*

(vi) As $n \to \infty$*,* $h_1, h_2 \to 0$*,* $nh_1h_2 \to \infty$ *and* $c_n/\sqrt{h_1} \to 0$*, where* $c_n = h_1^2 + h_2^2 + \sqrt{\frac{\log n}{nh_1h_2}}$.

Remark 3.1. *Some of these conditions are stronger than necessary but are made so that model (1) has the main ingredients of a prototype non-cointegrated FC system, which is convenient in the development of the limit theory and ensures mixed normality which facilitates inference. Condition (ii) is restrictive, especially exogeneity, but matches conditions used earlier in this literature. In particular, strict exogeneity of* $(\Delta x_t, z_t)$ *was assumed in the FC cointegration model of Xiao (2009), Sun et al. (2013) and Sun et al. (2016). Sun et al. (2011) assumed independence between* z_t *and* $(\Delta x_t, \Delta u_t)$ *in their treatment of the non-cointegration case but permitted endogeneity of* x_t*, although they did not provide a limit theory for their functional coefficient estimator. A fully nonparametric regression model with endogeneity was considered in Wang and Phillips (2009) and standard normal limit theory was established for a self-normalized test statistic. But convergence rates are slow in the general nonparametric case and the functional coefficient regression model was not considered in that work. Recent work by Liang et al. (2022) allowed* (x_t, z_t) *to be endogenous in the functional coefficient cointegrating framework with either* x_t *or* z_t *nonstationary. In contrast to stationary nonparametric regression, they found that the conventional local level and local linear kernel estimators remain valid in the presence of endogeneity. However, treatment of the endogenous regressors case while allowing for a non-cointegrating nonstationary regression framework as in this paper will require substantial additional work and is left as a future project. The conditions on the density functions in Assumption 1 (iii), the functional coefficient in (iv) and those on the kernel function stated in (v) are all commonly used in kernel nonparametrics. Similar bandwidth conditions as those given in Assumption 1 (vi) can be found in Fan et al. (1998) and Cai and Masry (2000).*

With these conditions in hand, we obtain the following asymptotic properties of the aforementioned estimators. We proceed under the assumption that $h_1 = h_2 \equiv h$ to simplify the discussion. The restriction is not unreasonable because smoothing is performed with respect to the same variable z_t in this use of bivariate kernel estimation.

Theorem 3.1. *Under Assumption 1, for a given interior point* z *of the support of* z_t*, the following results hold as* $n \to \infty$*:*

(a) For the one-step estimator of $\beta(z)$ *defined in (7),*

$$\frac{1}{\sqrt{h}}(\hat{\beta}_{1s}(z) - \beta(z)) \rightsquigarrow \mathcal{MN}\left(0, \nu_2(K)\left(\int [\beta^{(1)}(z)' B_x]^2\right) \mathbb{E}\left[\frac{f(z_t)}{f(z, z_t)}\right] [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\right), \tag{12}$$

where B_x *is the limit Brownian Motion for which* $n^{-1/2}\sum_{t=1}^{[nr]} \Delta x_t \rightsquigarrow B_x(r)$.

(b) For the estimators of $\alpha(z)$ *defined in* (8) *and* (9),

$$\sqrt{n}(\hat{\alpha}(z) - \alpha(z) - h^2\mathcal{B}(z)) \rightsquigarrow \mathcal{N}\left(0, \Gamma(\alpha(z_t))\right), \tag{13}$$

$$\sqrt{n}[\tilde{\alpha}(z) - \alpha(z) - h^2\bar{\mathcal{B}}(z)] \rightsquigarrow \mathcal{N}(0, \Gamma(\alpha(z_t))), \tag{14}$$

where $\mathcal{B}(z) = \mu_2(K)\mathbb{E}\left(\frac{C_1(z,z_t) - C_2(z,z_t)}{f(z,z_t)}\right)$, $C_1(z,w) = \beta^{(1)}(z)f_z^{(1)}(z,w) + \frac{1}{2}\beta^{(2)}(z)f(z,w)$, $C_2(z,w) = \beta^{(1)}(w)f_w^{(1)}(z,w) + \frac{1}{2}\beta^{(2)}(w)f(z,w)$, $\bar{\mathcal{B}}(z) = \mathcal{B}(z) - \mathbb{E}\mathcal{B}(z_t)$ *and* $\Gamma(\alpha(z_t))$ *is the long run variance matrix of* $\{\alpha(z_t)\}$.

(c) For the estimator of β_0 *defined in* (10),

$$\sqrt{n}(\hat{\beta}_0 - \beta_0 + h^2\mathcal{B}) \rightsquigarrow \mathcal{N}\left(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t\Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1} + \Gamma(\alpha(z_t))\right), \tag{15}$$

$$\sqrt{n}(\tilde{\beta}_0 - \beta_0 + h^2\bar{\mathcal{B}}) \rightsquigarrow \mathcal{N}\left(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t\Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1} + \Gamma(\alpha(z_t))\right), \tag{16}$$

where $\Gamma(\Delta x_t\Delta u_t)$ *denotes the long run variance of* $\Delta x_t\Delta u_t$, $\mathcal{B} = \mathbb{E}\mathcal{B}(z_t) + [\mathbb{E}(\Delta x_t(\Delta x_t)')]^{-1}\int dB_{x\mathcal{B}}B_x$, $B_{x\mathcal{B}}(r)$ *is the matrix Brownian motion limit of the partial sum process* $\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\Delta x_t[\Delta\mathcal{B}(z_t)]'$. $\bar{\mathcal{B}}$ *is defined in the same way as* $\mathcal{B}$ *with* $\mathcal{B}(z)$ *replaced by* $\bar{\mathcal{B}}(z)$ *and* $B_{x\mathcal{B}}$ *replaced by* $B_{x\bar{\mathcal{B}}}$.

(d) For the two-step estimator of $\beta(z)$ *defined in* (11),

$$\sqrt{n}\left\{\hat{\beta}_{2s}(z) - \beta(z) - h^2[\mathcal{B}(z) - \mathcal{B}]\right\} \rightsquigarrow \mathcal{N}\left(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t\Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\right). \tag{17}$$

Remark 3.2. ***(Limit theory for*** $\hat{\beta}_{1s}(z)$***)*** *The limit distribution and convergence rate for* $\hat{\beta}_{1s}(z)$ *given in* (12) *is highly unusual in functional coefficient models and more generally in nonparametric regression. As shown in the proof of the theorem, the first step estimator* $\hat{\beta}_1(z,w)$ *in* (6) *of the bivariate kernel regression* (4) *is found to be an inconsistent estimator of the functional coefficient* $\beta(z)$. *More specifically, we find (see* (52) *in Appendix A) that*

$$\hat{\beta}_1(z,w) - \beta_1(z,w) \sim_a \left[\sum \Delta x_t(\Delta x_t)'K_{tz}K_{tw}\right]^{-1}\sum \Delta x_t x_{t-1}'[\beta(z_t) - \beta(z) - (\beta(w_t) - \beta(w))]K_{tz}K_{tw}, \tag{18}$$

so that part of the approximation error coming from the bivariate functional coefficient of x_{t-1} *in the kernel regression ends up dominating the asymptotic distribution of* $\hat{\beta}_1(z,w)$. *Detailed study of the proof reveals that the bias function associated with this coefficient dominates because it arises in that part of the partitioned regression associated with the integrated regressor* x_{t-1}, *whose order of magnitude dominates that of the regressor* Δx_t. *However, although the first step estimator* $\hat{\beta}_1(z,w)$ *is inconsistent, upon marginal averaging to obtain the estimator* $\hat{\beta}_{1s}(z)$, *we find that* $\hat{\beta}_{1s}(z)$ *is consistent but at the very slow rate* $\sqrt{h}$ *or* $n^{-\frac{1}{10}}$ *if* $h = O(n^{-\frac{1}{5}})$. *The mixed normal (*$\mathcal{MN}$*) limit theory for* $\hat{\beta}_{1s}(z)$ *in Theorem 3.1 (a) is the consequence of the limiting stochastic integral form of* (18) *and its conditional normality given the Brownian motion limit process* B_x.

Remark 3.3. ***(Limit theory for*** $\hat{\alpha}(z)$***)*** *The limit result for* $\hat{\alpha}(z)$ *is also nonstandard in nonparametric smoothing. Although the asymptotic bias has the usual order* h^2*, the asymptotic variance is of order* $O(1/n)$*, which is independent of the bandwidth* h*. This rate differs from the existing results in nonparametric smoothing involving nonstationary variables (Cai et al., 2009). Due to the additive form of the functional coefficient of* x_{t-1} *in (4), i.e.,* $\beta_2(z,w) = \alpha(z) - \alpha(w)$*, the marginal integration based estimator* $\hat{\alpha}(z)$ *involves the sample average* $n^{-1}\sum_{t=1}^{n}\alpha(z_t)$ *and can be decomposed as follows*

$$\hat{\alpha}(z) - \alpha(z) = \frac{1}{n}\sum_{t=1}^{n}[\hat{\beta}_2(z,z_t) - \beta_2(z,z_t)] - \frac{1}{n}\sum_{t=1}^{n}\alpha(z_t). \tag{19}$$

Here, the second term is clearly of order $O_p(1/\sqrt{n})$*. But for the bivariate estimator* $\hat{\beta}_2(z,w)$ *in the first term, we find in the proof that*

$$\hat{\beta}_2(z,w) - \beta_2(z,w) \sim_a \left[\sum_{t=2}^{n} x_{t-1}x'_{t-1}K_{tz}K_{tw}\right]^{-1}\sum_{t=2}^{n} x_{t-1}x'_{t-1}\zeta_t = O_p(h^2 + 1/\sqrt{n}) \tag{20}$$

where $\zeta_t = [\beta(z_t) - \beta(z) - (\beta(w_t) - \beta(w))]K_{tz}K_{tw}$*. We show in the proof that, after subsequent averaging with respect to* w*, the variance contribution from (20) is dominated by the second term of (19). Thus the first term of (19) only contributes an asymptotic bias term in determining the limit distribution and the second term of (19) delivers the limit normal distribution in Theorem 3.1 (b) and the rate of convergence* $\sqrt{n}$*. Similar considerations apply to the limit theory for* $\tilde{\alpha}(z)$*. The details are presented in the Appendix. These limit results again differ considerably from those of conventional nonparametric regression.*

Remark 3.4. ***(Limit theory for*** $\hat{\beta}_0$***)*** *We note that the bias function* $\mathcal{B}$ *is random and involves a stochastic integral, which is another highly unusual feature of this regression. In the proof we obtained the following decomposition*

$$\hat{\beta}_0 - \beta_0 = \left[\sum_{t=2}^{n}\Delta x_t(\Delta x_t)'\right]^{-1}\left[\sum_{t=2}^{n}\Delta x_t\Delta u_t + \sum_{t=2}^{n}\Delta x_t\Delta\{x'_t[\alpha(z_t) - \hat{\alpha}(z_t)]\}\right].$$

It turns out that the bias of $\hat{\alpha}(z_t)$*, which is* $\mathcal{B}(z_t)$*, contributes a term* $\sum_{t=2}^{n}\Delta x_t\Delta[x'_t\mathcal{B}(z_t)]$*. It is this term that leads to a stochastic integral in the bias of* $\hat{\beta}_0$ *(see (64) in Appendix A). Inference regarding* β_0 *can be conducted with the help of the limit theory presented here, for example by using the conventional Wald type statistic. Note that the limit theory here is different from that in Theorem 3.2 below. This suggests constancy of* $\beta(z)$ *should be tested first and then one can choose the right limit theory to conduct inference regarding* β_0*.*

Remark 3.5. ***(Bandwidth Roles)*** *The convergence rate and the order of the asymptotic variance of* $\hat{\alpha}(z)$ *are both independent of the bandwidth, in contrast to typical nonparametric estimators where the order depends on the smoothing bandwidth and the effective sample size. This result suggests that smaller bandwidths* (h_1, h_2) *might be preferred to improve bias properties in* *estimation. But the bandwidths (h_1, h_2) need to satisfy the restrictions given in Assumption 1 (vi) and bandwidths that are chosen too small lead to degeneracies in finite samples due to singular denominator problems in local level estimation. This comment on the role of bandwidths and their restictions applies as well to the estimators $\hat{\beta}_0$, $\tilde{\beta}_0$ and $\hat{\beta}_{2s}(z)$. Section 5.1 numerically investigates the sensitivity of estimator performance to bandwidth variation.*

If the functional coefficient $\beta(z)$ is a constant almost everywhere in the support of z_t, or equivalently, $\alpha(z) = 0$ a.e., it is easy to see that the asymptotic variances of $\hat{\beta}_{1s}(z)$, $\hat{\alpha}(z)$ and $\tilde{\alpha}(z)$ are all zero. These limit distributions are therefore degenerate. In this case, we have standard asymptotic theory presented below.

Theorem 3.2. *If $\beta(z) = \beta_0$, a constant, a.e. in the support of z_t, as $n \to \infty$, we have*

$$\sqrt{nh}(\hat{\beta}_{1s}(z) - \beta_0) \rightsquigarrow \mathcal{N}(0, \mathbb{E}(\Delta u_t)^2 \nu_0(K)\mathbb{E}[f(z_t)/f(z, z_t)][\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}), \tag{21}$$

$$n\sqrt{h}\hat{\alpha}(z) \rightsquigarrow \mathcal{MN}\left(0, \mathbb{E}(\Delta u_t)^2 \nu_0(K)\mathbb{E}\left[\frac{f(z_t)}{f(z, z_t)}\right]\left[\int B_x B_x'\right]^{-1}\right), \tag{22}$$

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \rightsquigarrow \mathcal{N}(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t \Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}), \tag{23}$$

$$\sqrt{n}(\hat{\beta}_{2s}(z) - \beta_0) \rightsquigarrow \mathcal{N}(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t \Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}). \tag{24}$$

Further, $\tilde{\alpha}(z)$ is asymptotically equivalent to $\hat{\alpha}(z)$ and $\tilde{\beta}_0$ is asymptotically equivalent to $\hat{\beta}_0$.

Remark 3.6. *When the functional coefficients are constant, there is no approximation error and the bias in kernel estimation is eliminated. This simplification means that more conventional results hold. In particular, the coefficient of the stationary regressor Δx_t now converges at the usual rate $\sqrt{nh}$ and the coefficient of the nonstationary variable x_{t-1} is superconsistent at the rate $n\sqrt{h}$.*

4 Testing Linear against Semiparametric Forms

Nonparametric functional coefficient models offer the potential to assess specific functional forms of dependence in $\beta(z)$, the most important of which is simple constancy, which leads to a linear regression specification. Theory and past empirical work often suggest linearity in both coefficients and variables and it is therefore of primary interest to test this specification against the more general semiparametric model which embodies the functional dependence $\beta(z)$.

For functional coefficient cointegration models, related tests of linearity were suggested in Xiao (2009) and Sun et al. (2016). Under the current non-cointegrated setting, Gan et al. (2014) also studied linearity testing and proposed two specific tests, showing in simulations that there was a performance advantage in using a test based on a semiparametric estimator of $\alpha(z)$.

With the more efficient nonparametric estimates of $\alpha(z)$ and $\beta(z)$ developed in the present paper, it might be expected that more efficient tests of linearity can be constructed. We employ the commonly used approach in tests of this kind which involves examining the discrepancy between the semiparametric estimate and the restricted estimate obtained under the null of linearity. Xiao (2009) used the maximum squared distance observed at a finite number of points, Sun et al. (2016) considered the integrated squared distance, and Gan et al. (2014) focused on the squared distance calculated over the observations $\{z_t\}$. For ease of implementation and comparison, we follow Xiao (2009) and consider the squared distance calculated at a finite number of distinct points $\{z_s^*\}_{s=1}^m$, where m is a fixed integer, in the support of z_t.

More specifically, our test statistic is based on a weighted distance measure of the form $\bar{I}_\alpha = \sum_{s=1}^m \hat{\alpha}(z_s^*)' W_n \hat{\alpha}(z_s^*)$, for some suitable weight matrix W_n. According to Theorem 3.2, $\hat{\alpha}(z)$ converges to zero at rate $n\sqrt{h}$ when $\alpha(\cdot) = 0$ and has a mixed normal limit distribution. Denote the conditional asymptotic variance of $n\sqrt{h}\hat{\alpha}(z)$ in (22) as $\Omega(z)$. We therefore consider the test statistic

$$\hat{I}_\alpha = n^2 h \sum_{s=1}^m \hat{\alpha}(z_s^*)' \hat{\Omega}(z_s^*)^{-1} \hat{\alpha}(z_s^*), \tag{25}$$

where $\hat{\Omega}(z)$ converges weakly to $\Omega(z)$[4]. It is easy to see that $\hat{I}_\alpha \rightsquigarrow \chi^2_{km}$ under the null, where χ^2_ℓ denotes a chi-squared distributed random variate with ℓ degree of freedom. Alternatively, if the re-centered estimator $\tilde{\alpha}(z)$ is used to replace $\hat{\alpha}(z)$ in the calculation of $\hat{I}_\alpha$, we obtain an analogous test statistic, denoted $\tilde{I}_\alpha$, that has the same asymptotic distribution.

We summarize the limit properties of these statistics under the null hypothesis as follows.

Theorem 4.1. *Assume that the conditions of Theorem 3.2 hold. Under $H_0 : \alpha(z) = 0$ a.e. and for fixed integer m, we have $\hat{I}_\alpha \rightsquigarrow \chi^2_{km}$, $\tilde{I}_\alpha \rightsquigarrow \chi^2_{km}$ as $n \to \infty$.*

According to Theorem 4.1, the two proposed test statistics are asymptotically chi-squared and the null is rejected if the test statistic exceeds the prescribed critical value. As is common in nonparametric testing, convergence to the null limit distribution is often slow and the chi-squared distribution may then be a poor approximation for the finite sample distribution of the statistic. We therefore present a bootstrap procedure for practical implementation of these tests, as in Gan et al. (2014).

Bootstrap steps for the $\hat{I}_\alpha$ test

Step (i). For observed sample $\{y_t, x_t, z_t\}_{t=1}^n$, obtain the nonparametric estimator $\hat{\alpha}(z_s^*)$ as in (8) at the selected grid points $\{z_s^*\}_{s=1}^m$. With these estimates, compute the statistic $\hat{I}_\alpha$ as in (25). Then obtain the two-step estimator $\hat{\beta}_{2s}(z_t)$ for $t = 1, \cdots, n$ and compute $\hat{u}_t = y_t - x_t'\hat{\beta}_{2s}(z_t)$. Let $\hat{\epsilon}_t = \hat{u}_t - \hat{u}_{t-1}$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{t=2}^n \hat{\epsilon}_t^2$.

Step (ii). Use the dependent wild bootstrap procedure of Shao (2010) to generate ϵ_t^* from $\hat{\epsilon}_t$, and compute $u_t^* = \sum_{s=1}^t \epsilon_s^*$ for $t = 1, \cdots, n$. Compute $y_t^* = x_t'\check{\beta}_0 + u_t^*$, where $\check{\beta}_0$ is the OLS estimate of β_0 obtained by regressing Δy_t on Δx_t. Compute the bootstrap statistic $\hat{I}_\alpha^*$ as in *Step (i)*, using the bootstrap sample $\{y_t^*, x_t, z_t\}_{t=1}^n$.

Step (iii). Repeat Step (ii) a large number of times, say $B = 400$, and use the upper δ-percentile of the bootstrap statistic empirical distribution $\{\hat{I}_\alpha^{*(b)}\}_{b=1}^B$, c_δ, to approximate the upper δ-percentile critical value of the null distribution of $\hat{I}_\alpha$.

[4] $\hat{\Omega}(z)$ can be constructed by replacing the unknown components of $\Omega(z)$ with consistent estimates. In particular, with $\hat{u}_t = y_t - x_t'\hat{\beta}_{2s}(z_t)$, $\mathbb{E}(\Delta u_t)^2$ can be estimated by $(n-1)^{-1}\sum_{t=2}^n (\Delta \hat{u}_t)^2$. The densities can be replaced by corresponding kernel density estimates and the expectation can be estimated by the sample average $n^{-1}\sum_{t=1}^n \hat{f}(z_t)/\hat{f}(z, z_t)$. Finally, the matrix quadratic functional $\int B_x B_x'$ can be estimated in the usual way by the standardized sample moment matrix $n^{-2}\sum_{t=1}^n x_t x_t'$. We then have

$$\hat{\Omega}(z) = \frac{1}{n-1}\sum_{t=2}^n (\Delta \hat{u}_t)^2 \times \nu_0(K) \times \frac{1}{n}\sum_{t=1}^n \frac{\hat{f}(z_t)}{\hat{f}(z, z_t)} \times \frac{1}{n^2}\sum_{t=1}^n x_t x_t' \sim_a \Omega(z).$$

Consistency of $\hat{\Omega}(z)$ is sufficient and follows easily from consistency of estimates for each component.

Step (iv). Reject the null hypothesis if $\hat{I}_\alpha > c_\delta$, or if the p-value$=\frac{1}{B}\sum_{b=1}^{B} 1(\hat{I}_\alpha^{*(b)} > \hat{I}_\alpha) < \delta$. Otherwise, the null is not rejected.

The bootstrap procedure for $\tilde{I}_\alpha$ is similar. Bootstrap consistency is established below.

Theorem 4.2. *Let $\mathcal{W}_n$ denote the sample observations $\{y_t, x_t, z_t\}_{t=1}^n$. Under Assumption 1, we have*

$$G_n(x) \xrightarrow{p} \Phi_{\chi,km}(x), \tag{26}$$

for all x, as $n \to \infty$, where $G_n(x)$ is the conditional bootstrap distribution of $\hat{I}_\alpha^*|\mathcal{W}_n$, given the original sample $\mathcal{W}_n$, and $\Phi_{\chi,km}(x)$ is the χ^2_{km} cumulative distribution function.

Simulations reported in Section 5.2 are used to examine size performance of the tests. The asymptotic tests are found to be satisfactory provided m is not too large. But the bootstrap tests typically have improved size performance and are generally recommended for practical work.

Next we examine the local power properties of the tests. We consider the local alternative $H_1^L : \alpha(z) = \rho_n \alpha^0(z)$, where ρ_n is a deterministic sequence converging to zero, and $\alpha^0(z)$ is a nonzero function. First, in view of Theorem 3.1 (b), it is not hard to see that the bias $\mathcal{B}(z)$ of $\hat{\alpha}(z)$ is of order $\rho_n h^2$ in this case and the asymptotic variance $\Gamma(\alpha(z_t))$ is of order ρ_n^2. Then from (13), we have

$$\hat{\alpha}(z) = O_p(\rho_n) + O_p(\rho_n h^2) + O_p(\rho_n^2/\sqrt{n}) = O_p(\rho_n). \tag{27}$$

It follows that $\hat{I}_\alpha = O_p(\rho_n^2 n^2 h)$ under the local alternative H_1^L. In a similar fashion, $\tilde{I}_\alpha = O_p(\rho_n^2 n^2 h)$ for fixed m. The findings are summarized as follows.

Theorem 4.3. *Under Assumption 1, the local alternative H_1^L, and for fixed integer m as $n \to \infty$, we have $\hat{I}_\alpha = O_p(\rho_n^2 n^2 h), \tilde{I}_\alpha = O_p(\rho_n^2 n^2 h)$.*

Remark 4.1. *Under local alternatives of the form H_1^L, tests based on $\hat{I}_\alpha$ and $\tilde{I}_\alpha$ are consistent when $\rho_n^2 n^2 h \to \infty$, i.e., when the localizing rate ρ_n diminishes slower than $1/\sqrt{n^2 h}$. When $\rho_n = O(1/\sqrt{n^2 h})$, the tests have nontrivial power whose magnitude depends specifically on the form of the sequence ρ_n.*

Remark 4.2. *For fixed alternatives with ρ_n a constant sequence the tests diverge at the rate $O_p(n^2 h)$, which is much faster than the test $\hat{J}_b$ proposed by Gan et al. (2014) for which $\hat{J}_b = O_p(nh)$ under fixed alternatives.*

Tests can also be constructed using estimates of $\beta(z)$. For example, based on the two-step estimator $\hat{\beta}_{2s}(z)$ we may consider test statistics that are constructed in the form of distance measures such as $\bar{I}_\beta = \sum_{s=1}^m [\hat{\beta}_{2s}(z_s^*) - \check{\beta}_0]'[\hat{\beta}_{2s}(z_s^*) - \check{\beta}_0]$ or analogous weighted measures, where $\check{\beta}_0$ is the OLS estimator obtained by regressing Δy_t on Δx_t. Note that under the null, $\hat{\beta}_{2s}(z) - \check{\beta}_0 = O_p(1/\sqrt{n})$, which suggests that $\bar{I}_\beta = O_p(1/n)$. Therefore, the test statistic $\hat{I}_\beta = n\bar{I}_\beta = O_p(n\rho_n^2)$ under the local alternative H_1^L. Since $\hat{I}_\beta$ is less powerful than $\hat{I}_\alpha$ under local alternatives, we do not investigate its properties any further.

5 Simulations

This section investigates the properties of the proposed estimates and tests in finite samples and makes comparisons with procedures suggested in earlier work.

5.1 Estimation accuracy

We compare the finite sample performance of the estimates proposed in Section 2 with those of Sun et al. (2011) (SHL thereafter), and conduct simulations to assess sensitivity in performance to bandwidth choice.

The following model is used as the data generating process:

$$y_t = x_t \beta(z_t) + u_t, \tag{28}$$

$$x_t = x_{t-1} + \epsilon_{xt}, \tag{29}$$

$$u_t = u_{t-1} + \epsilon_{ut}, \tag{30}$$

where ϵ_{xt} and ϵ_{ut} are independent $\mathcal{N}(0,1)$, and independent of each other[5]. Three different design processes are used for z_t in the simulations:

(IID) z_t is iid uniform [-1,1];

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

(AR) $z_t = 0.5z_{t-1} + v_t$, where v_t is iid uniform [-1,1].

The first design assumes an independent process for z_t while the latter two allow for serial correlation in z_t. To model the functional coefficient $\beta(z_t)$ we consider three specifications for $\alpha(z)$:

(i) Quadratic: $\alpha(z) = z - 0.5z^2 - \mathbb{E}(z_t - 0.5z_t^2)$;

(ii) Trigonometric: $\alpha(z) = \sin(z) - \mathbb{E}\sin(z_t)$;

(iii) Constant: $\alpha(z) = 0$ *a.e.*.

The first two function forms have been considered by Sun et al. (2011). We set $\beta(z) = \alpha(z) + \beta_0$ and, without loss of generality, set $\beta_0 = 1$.

A Gaussian kernel is used throughout the simulations with bandwidth $h = c \cdot \hat{\sigma}_z n^{-1/5}$, where $\hat{\sigma}_z$ is the sample standard deviation of z_t. To check the sensitivity of performance to bandwidth choice, the constant c is allowed to take values on a grid with step length 0.2 on the interval

[5] The DGP here follows Sun et al. (2011) for fair comparisons. We also looked at the case where both ϵ_{xt} and ϵ_{ut} are serially correlated. The results remain qualitatively the same and are available from the authors on request.

[0.2, 2.0]. We adopt the Integrated Mean Squared Error (IMSE) as a measure of estimation accuracy for a functional coefficient estimator $\hat{\theta}(z)$, viz.,

$$IMSE(\hat{\theta}(z)) \quad = \quad R^{-1}\sum_{r=1}^{R}\left\{n^{-1}\sum_{t=1}^{n}[\hat{\theta}^{(r)}(z_t)-\theta(z_t)]^2\right\}, \tag{31}$$

where $\hat{\theta}^{(r)}(z)$ denotes the estimate obtained in the r-th replication. We use the conventional MSE to measure estimation accuracy for β_0, viz.,

$$MSE(\hat{\beta}_0) = R^{-1}\sum_{r=1}^{R}[\hat{\beta}_0^{(r)}-\beta_0]^2.$$

The results are reported for $R = 400$ replications. Two sample sizes $n = 50, 100$ are considered.

To demonstrate sensitivity with respect to bandwidth, we plot the IMSEs and MSEs against the constant c employed in the bandwidth formula. We use c^* to denote the optimal value of c at which the smallest IMSE or MSE is achieved. From Remark 3.5 we know that for the $\sqrt{n}$-consistent estimates, the asymptotic variance is independent of the bandwidth. Thus we expect these estimates to perform better at smaller values of c compared to the $\sqrt{nh}$-consistent estimates. To save space, the results are reported only for quadratic $\alpha(z)$ and $\alpha(z) = 0$ *a.e.*, since those for the trigonometric function (ii) are similar. Figures 1, 2 and 3 present the results for quadratic $\alpha(z)$ for the three different designs of z_t. Results for $\alpha(z) = 0$ *a.e.* are collected in Figures 4, 5 and 6.

We start by considering the case of quadratic $\alpha(z)$ in detail. When z_t has no serial correlation, the main findings from Figure 1 can be summarized as follows. First, the two estimates of $\alpha(z)$ perform similarly, with the re-centered estimate $\tilde{\alpha}(z)$ being slightly better than $\hat{\alpha}(z)$. Both uniformly outperform the two SHL estimates for all bandwidth choices and all sample sizes. Our estimates achieve best performance at smaller bandwidth choices (both at $c^* = 0.8$ when $n = 50$ and both at $c^* = 0.6$ when $n = 100$) in comparison to the SHL estimates where the best bandwidth choice is $c^* = 1.4$ for $n = 50$ and $n = 100$. Second, in estimating β_0, the estimator $\tilde{\beta}_0$ is the best of our proposed two estimators. The best performance is attained at $c^* = 0.8$ when $n = 50$ and $c^* = 0.6$ when $n = 100$, and significantly improves that of the SHL estimator $\hat{\beta}_0$, whose best performance occurs for $c^* = 1.0$ when $n = 50$ and $c^* = 0.8$ when $n = 100$. Third, in estimating $\beta(z)$, the first step estimator $\hat{\beta}_{1s}(z)$ is inferior to the SHL estimate while the two-step estimator $\hat{\beta}_{2s}(z)$ is superior to the SHL estimates especially when c is relatively small. For the two-step estimator $\hat{\beta}_{2s}(z)$, the optimal constant is $c^* = 0.8$ for $n = 50$ and $c^* = 0.6$ for $n = 100$, while for the SHL estimator the optimal constant is $c^* = 1.2$ for both $n = 50$ and 100. These results support the asymptotic theory, which indicates smaller bandwidths within this permissable range should be preferred for all the $\sqrt{n}$-consistent estimators.

Figures 2 and 3 show results for serially correlated z_t. For the estimation of $\alpha(z)$ and $\beta(z)$, the proposed $\sqrt{n}$-consistent estimators all outperform the SHL estimators at a relatively small bandwidth. We further observe that the IMSE curves intersect those of the SHL estimates as the bandwidth increases. Compared to the case of independent z_t, the cross-over point happens at a smaller bandwidth. This finding suggests that the range of bandwidths over which our marginal integration based estimators enjoy better performance than those of SHL become narrower for dependent covariates z_t. For the estimation of β_0, the SHL estimate $\hat{\beta}_0$ enjoys a large advantage when $n = 100$.

Next, when $\alpha(z) = 0$ *a.e.*, Figures 4-6 indicate that all the estimates enjoy much smaller IMSE and MSE than when $\alpha(z) \neq 0$ irrespective of the presence or absence of serial dependence in z_t. The relative improvement of the proposed estimates over the SHL estimates is more significant in this case. The improvement is especially marked for the IMSE of $\alpha(z)$, an outcome that reflects the super consistency of our estimates of $\alpha(z)$. In this case, the best performing estimates of $\alpha(z)$, β_0 and $\beta(z)$ are uniformly superior to the SHL estimates[6]. All the IMSEs and MSEs decrease as c increases, leading to the best performance at $c = 2$. This is well expected as the optimal bandwidth tends to infinity when the function to be estimated is a constant.

To better illustrate the efficiency gains of the proposed estimates, we report the ratios of the best IMSE estimates ($\tilde{\alpha}(z)$ and $\hat{\beta}_{2s}(z)$) to those of the SHL estimates (SHL$\hat{\alpha}(z)$ and SHL$\tilde{\beta}(z)$). Table 1 presents the results for quadratic $\alpha(z)$ and Table 2 presents those for $\alpha(z) = 0$ *a.e.*. The levels of the best IMSE of the SHL estimators are reported in the square brackets. We include the values of c for which the best IMSEs are achieved, c^*, in parentheses. In both Table 1 and 2, our estimates achieve significant efficiency gains compared to SHL under almost all the considered scenarios and sample sizes. In Table 1 for the quadratic function, the improvement with independent z_t is the most marked. Furthermore, we observe that our estimates tend to perform best around $c = 0.6$, while the SHL estimates achieve their best performance around $c = 1.2$. This outcome again supports the asymptotic theory, which indicates that smaller bandwidths within the allowable range should be preferred for the $\sqrt{n}$-consistent estimators. It is apparent in Table 2 that the improvement regarding $\alpha(z)$ is more marked than that of $\alpha(z)$ in Table 1. This finding again supports the super-consistency of our estimates of $\alpha(z)$ under the null $\alpha(z) = 0$ *a.e.*.

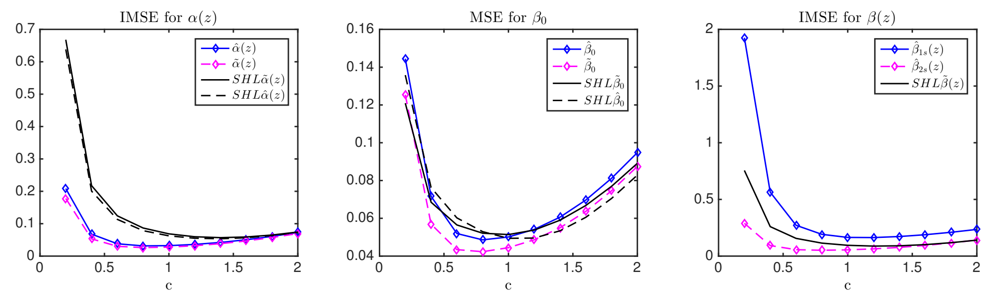
In sum, for estimation of $\alpha(z)$ and $\beta(z)$, the $\sqrt{n}$-consistent estimates vastly outperform the SHL estimates when small bandwidths are employed. The improvement is even more pronounced when $\alpha(z) = 0$ *a.e.*.

5.2 Test performance

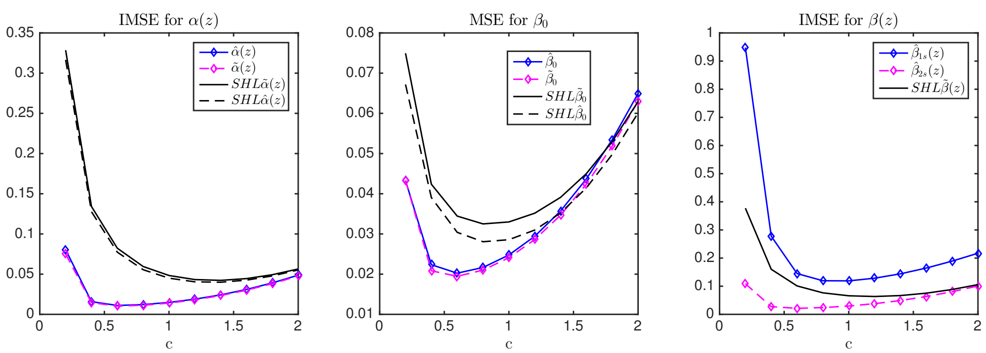
We compare the performance of the proposed tests ($\hat{I}_\alpha$ and $\tilde{I}_\alpha$) with that of the $\hat{J}_b$ statistic proposed by Gan et al. (2014)[7] (GHX, hereafter). The DGP is the same as (28)-(30). Following

[6]The results for the IMSE and MSE at $c = 0.2$ in Figures 4-6 are excluded in the plots due to the extremely large values.

[7]Only $\hat{J}_b$ is included because it has superior performance according to the simulations in Gan et al. (2014).

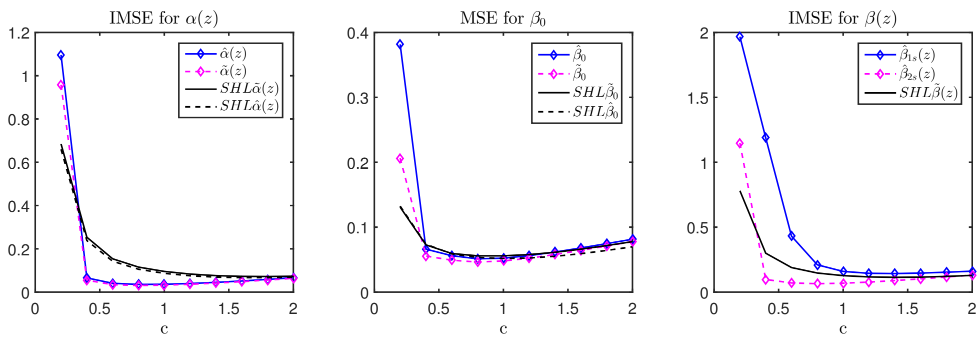


(a) $n = 50$

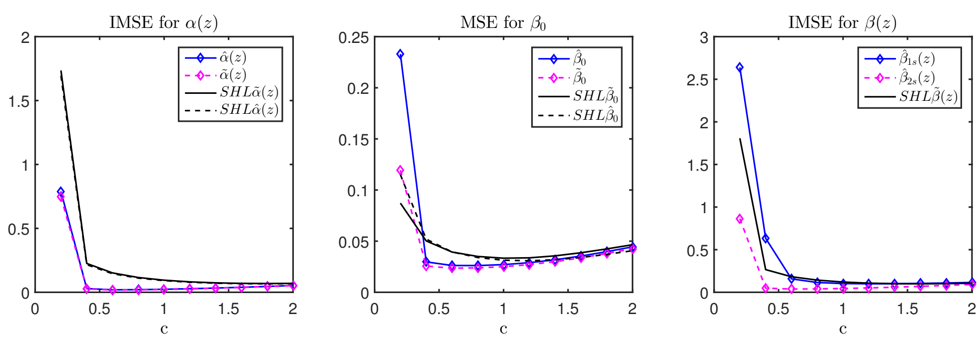


(b) $n = 100$

Figure 1: Plots of IMSE and MSE, where $\alpha(z)$ is quadratic, z_t is iid with $h = c \cdot \hat{\sigma}_z n^{-1/5}$

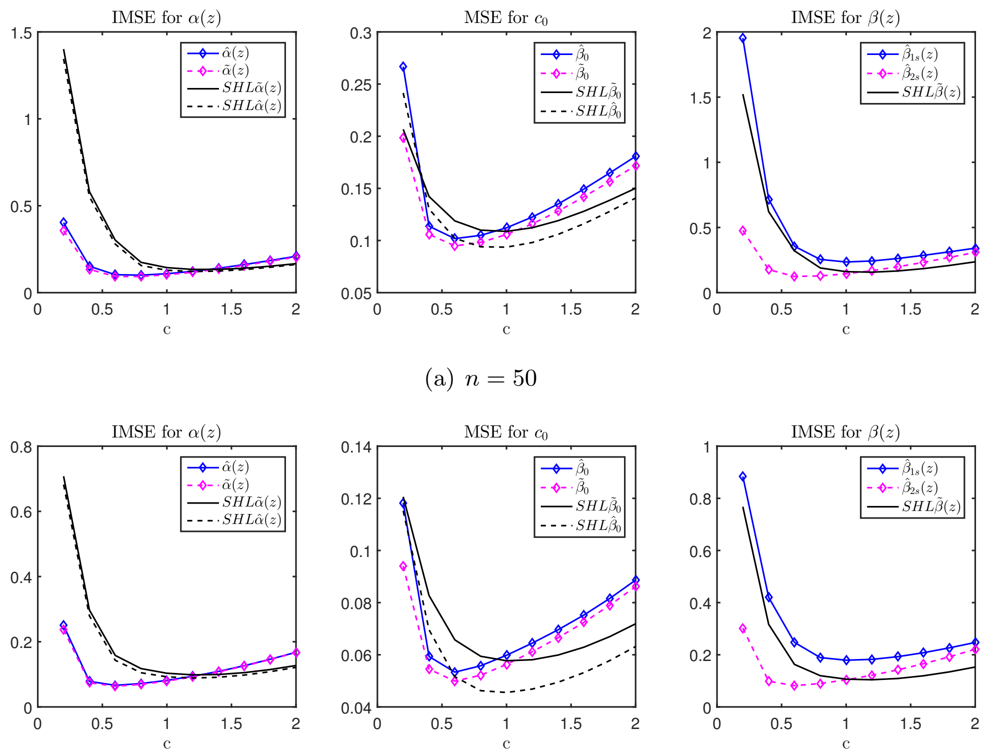


(a) $n = 50$



(b) $n = 100$

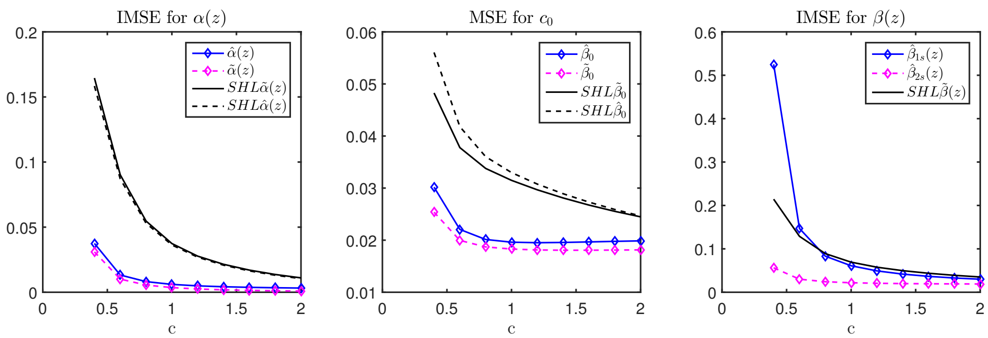
Figure 2: Plots of IMSE and MSE, where $\alpha(z)$ is quadratic, z_t is MA(1) with $h = c \cdot \hat{\sigma}_z n^{-1/5}$



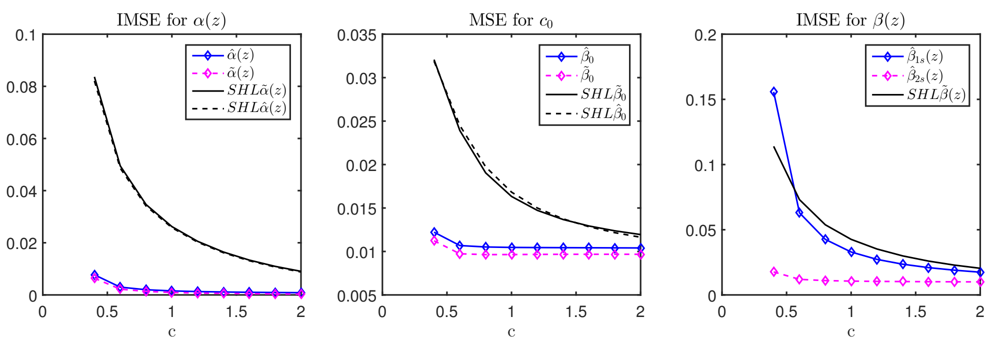
(a) $n = 50$

(b) $n = 100$

Figure 3: Plots of IMSE and MSE, where $\alpha(z)$ is quadratic, z_t is AR(1) with $h = c \cdot \hat{\sigma}_z n^{-1/5}$

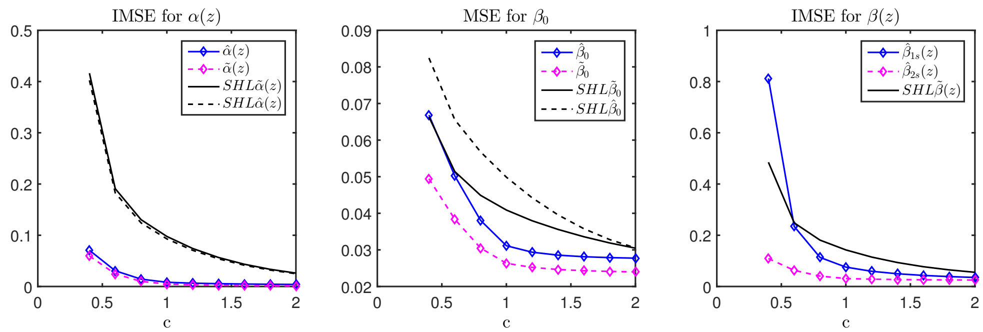


(a) $n = 50$

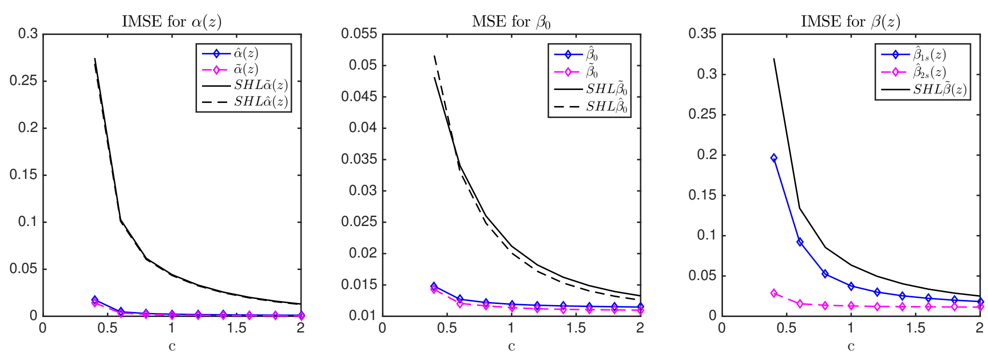


(b) $n = 100$

Figure 4: Plots of IMSE and MSE, where $\alpha(z) = 0$ $a.e.$, z_t is iid with $h = c \cdot \hat{\sigma}_z n^{-1/5}$

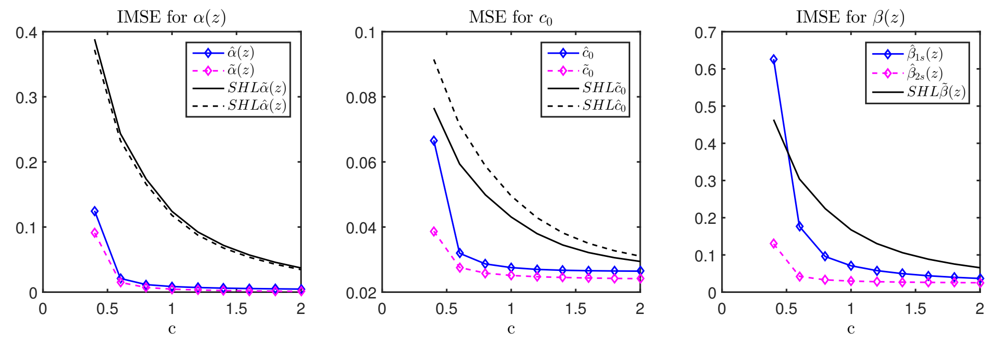


(a) $n = 50$

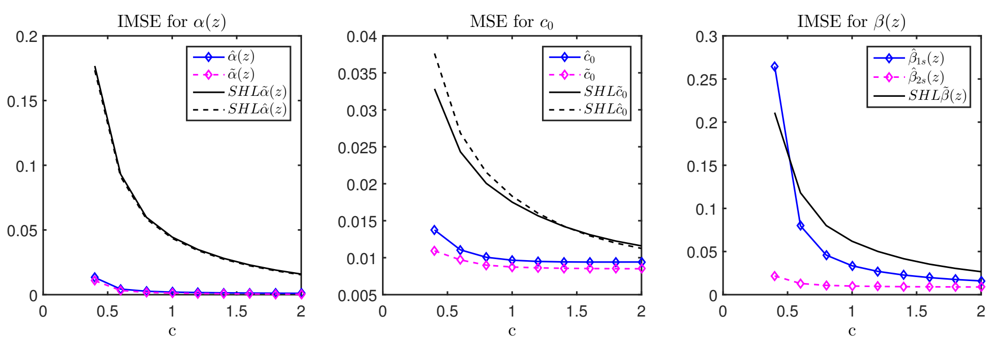


(b) $n = 100$

Figure 5: Plots of IMSE and MSE, where $\alpha(z) = 0$ $a.e.$, z_t is MA(1) with $h = c \cdot \hat{\sigma}_z n^{-1/5}$



(a) $n = 50$



(b) $n = 100$

Figure 6: Plots of IMSE and MSE, where $\alpha(z) = 0$ *a.e.*, z_t is AR(1) with $h = c \cdot \hat{\sigma}_z n^{-1/5}$

Table 1: IMSE ratios (in bold) when $\alpha(z)$ is quadratic

	z_t is iid				z_t is MA(1)				z_t is AR(1)			
	$\tilde{\alpha}(z)$	SHL$\hat{\alpha}(z)$	$\hat{\beta}_{2s}(z)$	SHL$\tilde{\beta}(z)$	$\tilde{\alpha}(z)$	SHL$\hat{\alpha}(z)$	$\hat{\beta}_{2s}(z)$	SHL$\tilde{\beta}(z)$	$\tilde{\alpha}(z)$	SHL$\hat{\alpha}(z)$	$\hat{\beta}_{2s}(z)$	SHL$\tilde{\beta}(z)$
$n = 50$	**0.489**	[0.054]	**0.581**	[0.088]	**0.459**	[0.067]	**0.568**	[0.114]	**0.764**	[0.122]	**0.793**	[0.157]
	(0.8)	(1.4)	(0.8)	(1.2)	(0.8)	(1.6)	(0.8)	(1.4)	(0.6)	(1.2)	(0.6)	(1.2)
$n = 100$	**0.257**	[0.040]	**0.338**	[0.063]	**0.271**	[0.066]	**0.281**	[0.099]	**0.715**	[0.089]	**0.784**	[0.104]
	(0.6)	(1.4)	(0.6)	(1.2)	(0.6)	(1.8)	(0.6)	(1.6)	(0.6)	(1.2)	(0.6)	(1.2)

Table 2: IMSE ratios (in bold) when $\alpha(z) = 0$ *a.e.*

	z_t is iid				z_t is MA(1)				z_t is AR(1)			
	$\tilde{\alpha}(z)$	SHL$\hat{\alpha}(z)$	$\hat{\beta}_{2s}(z)$	SHL$\tilde{\beta}(z)$	$\tilde{\alpha}(z)$	SHL$\hat{\alpha}(z)$	$\hat{\beta}_{2s}(z)$	SHL$\tilde{\beta}(z)$	$\tilde{\alpha}(z)$	SHL$\hat{\alpha}(z)$	$\hat{\beta}_{2s}(z)$	SHL$\tilde{\beta}(z)$
$n = 50$	**0.083**	[0.011]	**0.541**	[0.035]	**0.047**	[0.025]	**0.454**	[0.055]	**0.032**	[0.035]	**0.385**	[0.066]
	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)
$n = 100$	**0.036**	[0.009]	**0.489**	[0.020]	**0.033**	[0.013]	**0.455**	[0.025]	**0.023**	[0.015]	**0.333**	[0.027]
	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)

Gan et al. (2014), let ϵ_{xt} be i.i.d. $\mathcal{N}(0, 2^2)$, $x_0 = 0$, and ϵ_{ut} be i.i.d. $\mathcal{N}(0, 1)$ with $u_0 = 0$[8]. For the covariate z_t, we consider both the dependent case $z_t = v_t + v_{t-1} + \epsilon_{xt}$, and the independent case $z_t = v_t$, where v_t is i.i.d. uniform [0,2]. The bandwidths are determined by the rule $h = \hat{\sigma}_z n^{-1/5}$, for two sample sizes $n = 50, 100$.

Table 3: Test size based on asymptotic critical values (in percentages, nominal size=5%)

		$z_t = v_t \sim iid\ U[0,2]$		$z_t = v_t + v_{t-1} + \epsilon_{xt}$	
		$\hat{I}_\alpha$	$\tilde{I}_\alpha$	$\hat{I}_\alpha$	$\tilde{I}_\alpha$
$m = 3$	$n = 50$	8.75	0.75	12.25	1.75
	$n = 100$	7.25	1.75	9.50	1.25
$m = 5$	$n = 50$	11.75	2.50	13.25	3.00
	$n = 100$	11.25	3.00	15.50	4.00
$m = 9$	$n = 50$	20.00	6.00	22.25	5.25
	$n = 100$	17.00	5.00	18.00	6.75
$m = 20$	$n = 50$	25.00	7.50	25.50	8.25
	$n = 100$	23.25	7.75	20.25	7.00

We first consider size performance. Without loss of generality, we set $\beta(z) = 1$ under the null. We consider $m \in \{3, 5, 9, 20\}$ and use the $\{j/(m+1)\}_{j=1}^{m}$ percentiles of $\{z_t\}$ for a grid choice of m distinct points. The number of replications is 400. Table 3 reports empirical size (in

[8]Serial correlation in Δx_t and Δu_t does not qualitatively change the findings. Detailed results are available from the authors on request.

percentages) using asymptotic critical values with nominal size 5%. The $\hat{I}_\alpha$ test is found to be slightly oversized when m is small and oversizing worsens as m increases. On the other hand, the $\tilde{I}_\alpha$ test is undersized when m is small, increases as m increases, and when $m = 9$ the size of $\tilde{I}_\alpha$ is close to nominal. Based on size performance, the $\tilde{I}_\alpha$ test is preferred. But choice of the number of grid points m has an important bearing on size control, as is clear from Table 3.

To ameliorate size dependence on m we explore the bootstrap tests following the procedure of Section 4. Table 4 reports the size of our bootstrap tests (denoted as $\hat{I}^*_\alpha$ and $\tilde{I}^*_\alpha$) and that of GHX test[9] ($\hat{J}_b$) with bootstrap size $B = 200$. We see that the GHX test is severely undersized for all specifications, consistent with their findings. In addition, the size distortion is not corrected by increasing the sample size. The two proposed tests on the other hand have size close to nominal, which seems also to be insensitive to the choice of m.

Table 4: Size of the bootstrap tests (in percentage, nominal size=5%)

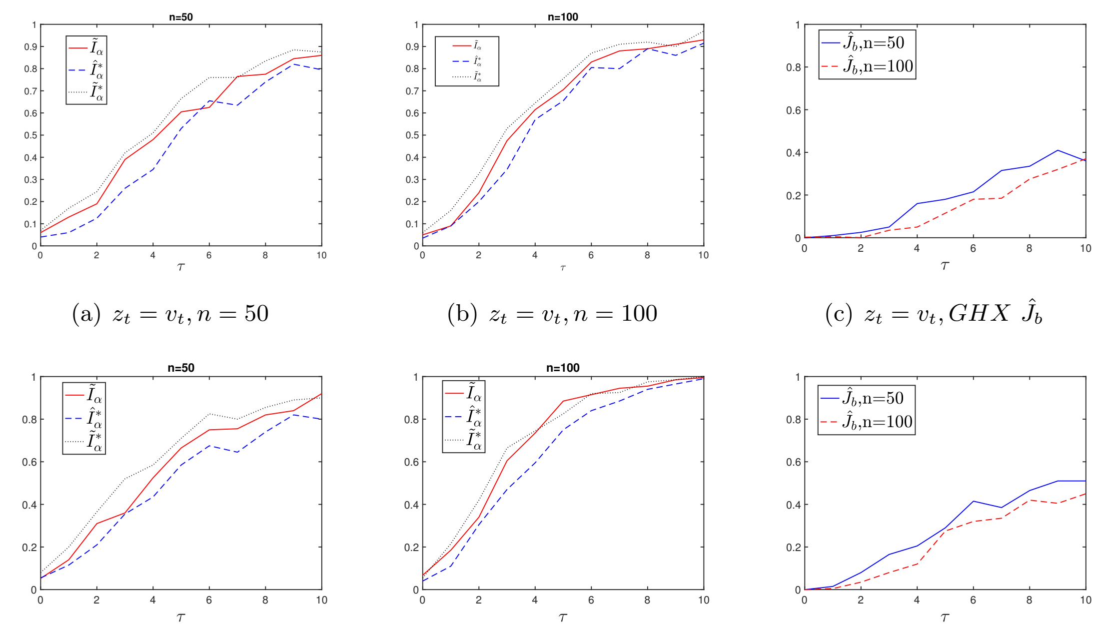
		$z_t = v_t \sim iid\ U[0,2]$			$z_t = v_t + v_{t-1} + \epsilon_{xt}$		
		$\hat{I}^*_\alpha$	$\tilde{I}^*_\alpha$	$\hat{J}_b$	$\hat{I}^*_\alpha$	$\tilde{I}^*_\alpha$	$\hat{J}_b$
$m = 3$	$n = 50$	4.00	8.00	0.00	3.00	6.00	0.00
	$n = 100$	4.50	5.50	0.00	7.00	7.00	0.00
$m = 5$	$n = 50$	3.00	7.00	0.75	4.00	5.50	0.25
	$n = 100$	3.00	7.00	0.00	4.00	6.50	0.00
$m = 9$	$n = 50$	4.00	7.00	0.00	5.50	8.00	0.00
	$n = 100$	3.50	6.00	0.25	4.00	5.50	0.00
$m = 20$	$n = 50$	4.00	7.00	0.50	4.50	6.50	0.75
	$n = 100$	4.00	7.50	0.00	5.00	6.00	0.00

Next we consider the power performance of $\tilde{I}_\alpha$ with $m = 9$ and the bootstrapped tests $\hat{I}^*_\alpha$ and $\tilde{I}^*_\alpha$ as these three tests have satisfactory size performance. For ease of reporting and without loss of generality we use $m = 9$ also for the bootstrapped tests. We consider two local alternatives $H_1^L : \beta(z) = \tau(z - 0.5z^2)n^{-3/5} + 1$ and $H_2^L : \beta(z) = \tau n^{-3/5}/(1 + e^{-z}) + 1$, for $\tau = 0, 1, 2, \ldots, 10$. In view of the divergence rates discussed in Remark 4.1 and 4.2, we know that the proposed tests have non-trivial asymptotic power whereas the GHX test has trivial asymptotic power for such local alternatives.

We plot power curves in Figure 7 for H_2^L. Results for H_1^L are similar and are unreported here. The power curves with independent z_t are collected in the upper panel with (a) and (b) for our proposed tests and (c) for the test GHX $\hat{J}_b$. Results with serially correlated z_t are summarized in

[9] For the test of Gan et al. (2014) we follow their suggestion and use $h = \hat{\sigma}_z \cdot n^{-1/5}$.

the lower panel in the same fashion. We first observe that the power performance of $\tilde{I}_\alpha$ and that of its bootstrapped version $\tilde{I}^*_\alpha$ are close for all cases considered. The powers of $\hat{I}^*_\alpha$ are slightly lower than those of $\tilde{I}_\alpha$ and $\tilde{I}^*_\alpha$. Second, the power of our proposed tests evidently rises quickly towards unity as τ increases and tends to increase as the sample size increases. This behavior holds for all specifications, showing that the proposed tests have power under the current local alternative. The GHX $\hat{J}_b$ test has substantially lower finite sample power and the power curves actually decline as n increases, with power curves for $n = 100$ lying below the curves for $n = 50$, as is clear from sub-figures (c) and (f). These results corroborate the asymptotic theory that our tests diverge at a faster rate than that of Gan et al. (2014). In short, the simulation findings support the use of $\tilde{I}_\alpha$ for an appropriate choice of m that controls size and, more generally, its bootstrapped version $\tilde{I}^*_\alpha$.



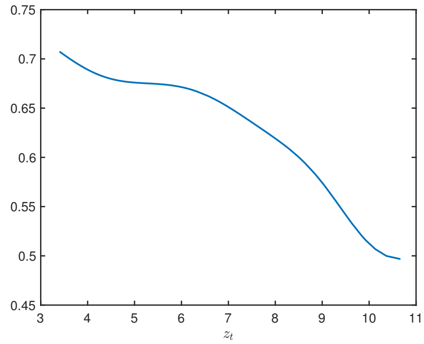
(a) $z_t = v_t, n = 50$ (b) $z_t = v_t, n = 100$ (c) $z_t = v_t, GHX\ \hat{J}_b$

(d) $z_t = v_t + v_{t-1} + \epsilon_{xt}, n = 50$ (e) $z_t = v_t + v_{t-1} + \epsilon_{xt}, n = 100$ (f) $z_t = v_t + v_{t-1} + \epsilon_{xt}, GHX\ \hat{J}_b$

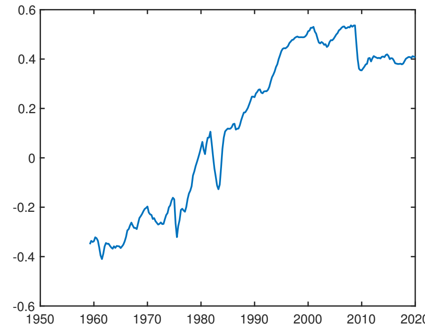
Figure 7: Power curves for H_2^L (the calculation used $m = 9$)

6 Empirical application

This section provides an empirical application based on the quantity theory of money to illustrate the uses of correlated but non-cointegrated regression relations. The classical quantity theory of money suggests $M_t V_t = P_t Y_t$, where M stands for the money stock, V for the velocity of circulation, P for the price level and Y for real output. Upon appropriate transformation we have $p_t = m_t + v_t$ where $p_t \equiv \log P_t$, and $m_t \equiv \log(M_t/Y_t)$ is referred to as excess money supply. The velocity $v_t = \log V_t$ has often been argued to be stationary. A general cointegration model



(a) functional coefficient estimates



(b) residual estimates

Figure 8: Estimation results of model (32) using Nadaraya-Watson estimator

considered by Bachmeier and Swanson (2005) is $p_t = \beta m_t + v'_t$ where β may be different from 1 and v'_t is the regression residual. To incorporate possible structural breaks and parameter instability, Wang et al. (2016) considered a functional-coefficient cointegration model which is given by

$$p_t = m_t \beta(z_t) + u_t, \tag{32}$$

where z_t is taken as the unemployment rate.

We follow Wang et al. (2016) and start from model (32). We are interested in the property of u_t in model (32). Specifically, we use the Consumer Price Index (2017=100) for P, the Gross Domestic Product in chained 2017 dollars as a measure of Y , and M2 for measure of M . All data are seasonally adjusted quarterly US figures ranging from 1959Q1 to 2019Q4. The m_t and p_t are found to be unit root processes. We first estimate model (32) by the Nadaraya-Watson estimator. The coefficient estimates and correponding residuals are displayed in Figure 8. From Figure 8 the residuals are evidently nonstationary and this is confirmed by the ADF test. This suggests model (32) coincides with the correlated but not cointegrated framework studied in the present paper. Therefore we re-estimate the coefficient $\beta(\cdot)$ using our proposed method. The estimation results are displayed in Figure 9 showing that the coefficient estimate differs substantially from that of Figure 8. The findings in Figure 9 are supported by the theory in the present paper and are reliable whereas those in Figure 8 are not, due to the evident nonstationarity of u_t. Further, using our proposed constancy test, we cannot reject the null that $\beta(\cdot)$ can be treated as constant. So, one might consider simplifying model (32) to a linear model. Nonetheless, we find clear evidence of a correlated but not cointegrated functional coefficient relation between the price level p_t and excess money supply m_t.

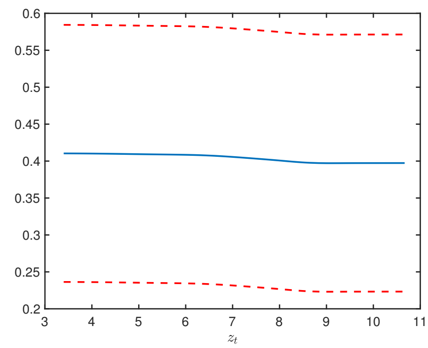


Figure 9: Estimation of $\beta(\cdot)$ in model (32) using our proposed estimator $\hat{\beta}_{2s}(z)$

7 Concluding Remarks

Functional coefficient regression models have attracted econometric interest because of their flexibility in modeling comovement among economic and financial variables where fixed linear and nonlinear relations fail to capture relations that may co-vary with other variables. Such models are useful in both stationary and nonstationary time series settings. This paper studies models in which the comovement among variables is not so close as to be characterized as cointegrating or even cointegrating with functional coefficients, leading to a framework that is intermediate between cointegration and completely spurious regression among independent variables. Our results complement earlier research (Sun et al., 2011; Gan et al., 2014) showing that $\sqrt{n}$ consistent estimation of the functional coefficient and asymptotically valid inference about the form of this function are both possible in such non-cointegrated functional coefficient models. The methods use marginal integration and back fitting techniques, and numerical studies show that substantial efficiency gains are possible in estimation and bootstrap tests can be constructed with standard asymptotic chi-squared distributions, stable size, and good local power properties.

The present paper suggests several useful directions for future study. The model in the present study is a prototypical non-cointegrated system with a functional coefficient of general form. However, the model is limited by the assumption of exogenous regressors, as in the functional cointegration analysis of Xiao (2009). A primary task in future research is to provide a methodology of estimation that allows for endogenous regressors in the non-cointegrated framework with functional coefficients and a mixed normal limit theory that facilitates inference. A further goal is to extend the current research to encompass the cointegration case, thereby providing empirical researchers with tools of estimation and inference that are robust to a wide degree of comovement among integrated variables. Another direction of research is to investigate the properties of the proposed estimates when the smoothing covariate is itself nonstationary, a substantial additional complication that raises technical issues considered in other recent work Wang et al. (2023). A further line of research is to provide for a wider range of nonstationary variables in the system, including near integrated and mildly integrated time series (Phillips, 1987, 1988; Phillips and Magdalinos, 2007) with the use of suitable instrumental variable methods such as functional coefficient versions of the IVX method of Phillips et al. (2009). The multivariate z_t case also needs investigation since the marginal integration approach has very poor performance when z_t is multivariate. New estimation methods are needed to deal with this 'curse of dimensionality' problem. Furthermore, testing procedures appropriate for a general form of the functional coefficient also requires study. Some of these extensions are currently under investigation.

Appendix

This appendix has two parts. Appendix A contains proofs of the main theorems of the paper. Appendix B provides useful lemmas and intermediate results used in the proofs of Appendix A. Some of the results provide functional limit theory for partial sum matrix processes that involve kernel functions in various levels of complexity, which may be of wider interest and application.

In terms of notation, we use := or =: to denote definitional equivalence. When we write $S_n = O_p(nh) \times S$, we mean that the stochastic order of S_n is $O_p(nh)$ and $(nh)^{-1}S_n$ converges to S either in probability or in distribution. The notation $\sim_a$ means equivalence in asymptotic distribution.

A Proofs of the Main Theorems

As mentioned, the use of the trimming function $1(z_t \in \mathcal{S}_n)$ in (7) and (8) does not affect the asymptotic theory and is therefore omitted in the proofs that follow to simplify notation.

Proof of Theorem 3.1. With some algebra we have the decomposition

$$\begin{aligned}\hat{\beta}(z,w) - \beta(z,w) &= \left[\sum_t X_t X_t' K_{tz} K_{tw}\right]^{-1} \sum_t X_t X_t' (\beta(z_t, w_t) - \beta(z,w)) K_{tz} K_{tw} \\ &\quad + \left[\sum_t X_t X_t' K_{tz} K_{tw}\right]^{-1} \sum_t X_t \Delta u_t K_{tz} K_{tw} \\ &=: [\Omega_n(z,w)]^{-1} B_n(z,w) + [\Omega_n(z,w)]^{-1} V_n(z,w), \end{aligned} \tag{33}$$

As explained in the Introduction use of standard diagonal matrix normalization fails in developing the correct asymptotic behavior of the component products in (33). We therefore examine below the component elements of $\Omega_n(z,w)$, $B_n(z,w)$ and $V_n(z,w)$ in sequence in order to isolate leading terms that contribute to the limit theory. We start with the signal matrix

$$\Omega_n(z,w) = \begin{pmatrix} \sum \Delta x_t (\Delta x_t)' K_{tz} K_{tw} & \sum \Delta x_t x_{t-1}' K_{tz} K_{tw} \\ \sum x_{t-1} (\Delta x_t)' K_{tz} K_{tw} & \sum x_{t-1} x_{t-1}' K_{tz} K_{tw} \end{pmatrix} =: \begin{pmatrix} \Omega_{n11}(z,w) & \Omega_{n12}(z,w) \\ \Omega_{n21}(z,w) & \Omega_{n22}(z,w) \end{pmatrix}.$$

We take each element in turn and, for notational simplicity, omit dependence on (z,w). Starting with Ω_{n11} we have the decomposition

$$\Omega_{n11} = \sum \Delta x_t (\Delta x_t)' K_{tz} K_{tw} =: \sum \Delta x_t (\Delta x_t)' \mathbb{E}[\xi_t] + \sum \Delta x_t (\Delta x_t)' \bar{\xi}_t,$$

where $\xi_t = K_{tz} K_{tw}$, $\bar{\xi}_t = \xi_t - \mathbb{E}\xi_t$. In view of Lemma B.1 (a), we have $\mathbb{E}\xi_t = O(h_1 h_2) \times f(z,w)$, $\bar{\xi}_t = O_p(\sqrt{h_1 h_2})$, and then $\sum \Delta x_t (\Delta x_t)' \bar{\xi}_t = O_p(\sqrt{n h_1 h_2})$. With $n h_1 h_2 \to \infty$, we can write

$$\begin{aligned}\Omega_{n11} &\sim_a O(n h_1 h_2) \times \mathbb{E}[\Delta x_t (\Delta x_t)'] f(z,w) + O_p(\sqrt{n h_1 h_2}) \\ &= O(n h_1 h_2) \times \mathbb{E}[\Delta x_t (\Delta x_t)'] f(z,w). \end{aligned} \tag{34}$$

Similarly, for the (2,2) element Ω_{n22} we have

$$\begin{aligned}
\Omega_{n22} &= \sum x_{t-1}x'_{t-1}K_{tz}K_{tw} \\
&= \sum x_{t-1}x'_{t-1}\mathbb{E}[\xi_t] + \sum x_{t-1}x'_{t-1}\bar{\xi}_t \\
&\sim_a O_p(n^2h_1h_2) \times \int B_xB'_xf(z,w) + O_p(n\sqrt{nh_1h_2}) \\
&= O_p(n^2h_1h_2) \times \int B_xB'_xf(z,w). \qquad (35)
\end{aligned}$$

For the (1,2) element Ω_{n12}, in view of Lemma B.1 (d), we can write

$$\Omega_{n12} = \sum \Delta x_t x'_{t-1}K_{tz}K_{tw} \sim_a O_p(n\sqrt{h_1h_2}) \times \int dB_{x\xi}B'_x. \qquad (36)$$

To analyze the inverse matrix $[\Omega_n(z,w)]^{-1}$, we use partitioned inversion. This approach ensures that all key elements are retained in deriving the limit theory, whereas diagonal matrix normalization eliminates elements that are important in the matrix product (33). Writing

$$[\Omega_n(z,w)]^{-1} = \begin{pmatrix} \Omega_n^{11} & \Omega_n^{12} \\ \Omega_n^{21} & \Omega_n^{22} \end{pmatrix},$$

we have

$$\begin{aligned}
\Omega_n^{11} &= \Omega_{n11}^{-1} + \Omega_{n11}^{-1}\Omega_{n12}[\Omega^*_{n22}]^{-1}\Omega'_{n12}\Omega_{n11}^{-1}, \qquad (37)\\
\Omega_n^{12} &= -\Omega_{n11}^{-1}\Omega_{n12}[\Omega^*_{n22}]^{-1}, \\
\Omega_n^{21} &= -\Omega_{n22}^{-1}\Omega'_{n12}[\Omega^*_{n11}]^{-1}, \\
\Omega_n^{22} &= \Omega_{n22}^{-1} + \Omega_{n22}^{-1}\Omega'_{n12}[\Omega^*_{n11}]^{-1}\Omega_{n12}\Omega_{n22}^{-1}, \qquad (38)
\end{aligned}$$

where $\Omega^*_{n22} = \Omega_{n22} - \Omega'_{n12}\Omega_{n11}^{-1}\Omega_{n12}$, $\Omega^*_{n11} = \Omega_{n11} - \Omega_{n12}\Omega_{n22}^{-1}\Omega'_{n12}$. In view of the previous analysis of the orders of $\Omega_{n11}, \Omega_{n12}$ and Ω_{n22}, it is clear that Ω_{n22} is the leading term of Ω^*_{n22} and Ω_{n11} is the leading term of Ω^*_{n11}. Further inspection reveals that the second term on the right hand side of (37) is negligible compared to the first term Ω_{n11}^{-1}, and this is also true for (38). Therefore we have $\Omega_n^{11} \sim_a \Omega_{n11}^{-1}$, $\Omega_n^{12} \sim_a -\Omega_{n11}^{-1}\Omega_{n12}[\Omega_{n22}]^{-1}$, $\Omega_n^{21} \sim_a -\Omega_{n22}^{-1}\Omega'_{n12}[\Omega_{n11}]^{-1}$ and $\Omega_n^{22} \sim_a \Omega_{n22}^{-1}$.

Next we analyze $B_n(z,w)$. For the same reason as for $[\Omega_n(z,w)]^{-1}$ we write this matrix in partitioned components form as

$$\begin{aligned}
B_n(z,w) &= \sum_t X_tX'_t(\beta(z_t,w_t) - \beta(z,w))K_{tz}K_{tw} \\
&= \begin{pmatrix} \sum \Delta x_t(\Delta x_t)'[\beta_1(z_t,w_t) - \beta_1(z,w)]K_{tz}K_{tw} + \sum \Delta x_t x'_{t-1}[\beta_2(z_t,w_t) - \beta_2(z,w)]K_{tz}K_{tw} \\ \sum x_{t-1}(\Delta x_t)'[\beta_1(z_t,w_t) - \beta_1(z,w)]K_{tz}K_{tw} + \sum x_{t-1}x'_{t-1}[\beta_2(z_t,w_t) - \beta_2(z,w)]K_{tz}K_{tw} \end{pmatrix} \\
&=: \begin{pmatrix} B_{n1}(z,w) + B_{n2}(z,w) \\ B_{n3}(z,w) + B_{n4}(z,w) \end{pmatrix}. \qquad (39)
\end{aligned}$$

Omitting dependence on (z, w), we take these elements in turn. Starting with B_{n1}, we apply the same decomposition as before, giving

$$\begin{aligned} B_{n1} &= \sum(\Delta x_t)(\Delta x_t)'(\beta_1(z_t, w_t) - \beta_1(z, w))K_{tz}K_{tw} \\ &= \sum(\Delta x_t)(\Delta x_t)'(\beta(z_t) - \beta(z))K_{tz}K_{tw} \\ &= \sum(\Delta x_t)(\Delta x_t)'\mathbb{E}[\eta_t] + \sum(\Delta x_t)(\Delta x_t)'\bar{\eta}_t, \end{aligned} \tag{40}$$

where $\eta_t = (\beta(z_t) - \beta(z))K_{tz}K_{tw}$, $\bar{\eta}_t = \eta_t - \mathbb{E}\eta_t$. In view of Lemma B.1 (b), we know that $\mathbb{E}[\eta_t] = O(h_1 h_2 h_1^2) \times \mu_2(K)C_1(z, w)$, where

$$C_1(z, w) = \beta^{(1)}(z)f_z^{(1)}(z, w) + \frac{1}{2}\beta^{(2)}(z)f(z, w),$$

and in view of Lemma B.1 (g) we can write

$$B_{n1} \sim_a O_p(nh_1h_2h_1^2) \times \mu_2(K)\mathbb{E}[\Delta x_t(\Delta x_t)']C_1(z, w) + O_p(\sqrt{nh_1h_2h_1^2}) \times \int dB_{xx\eta}. \tag{41}$$

Next, decomposing B_{n4} in the same way, we have

$$\begin{aligned} B_{n4} &= \sum x_{t-1}x_{t-1}'[\beta_2(z_t, w_t) - \beta_2(z, w)]K_{tz}K_{tw} \\ &= \sum x_{t-1}x_{t-1}'[\beta(z_t) - \beta(z) - (\beta(w_t) - \beta(w))]K_{tz}K_{tw} \\ &= \sum x_{t-1}x_{t-1}'\mathbb{E}[\zeta_t] + \sum x_{t-1}x_{t-1}'\bar{\zeta}_t, \end{aligned} \tag{42}$$

where $\zeta_t = [\beta(z_t) - \beta(z) - (\beta(w_t) - \beta(w))]K_{tz}K_{tw}$ and $\bar{\zeta}_t = \zeta_t - \mathbb{E}\zeta_t$. In view of Lemma B.1 (c), we have

$$\mathbb{E}[\zeta_t] = O(h_1h_2) \times \mu_2(K)[h_1^2C_1(z, w) - h_2^2C_2(z, w)],$$

where

$$C_2(z, w) = \beta^{(1)}(w)f_w^{(1)}(z, w) + \frac{1}{2}\beta^{(2)}(w)f(z, w).$$

By virtue of the independence of x_t and z_t we have $\bar{\zeta}_t = O_p(\sqrt{h_1h_2(h_1^2 + h_2^2)})$. We then deduce that

$$B_{n4} \sim_a O_p(n^2h_1h_2) \times \int B_xB_x'\mu_2(K)[h_1^2C_1(z, w) - h_2^2C_2(z, w)] + O_p(n\sqrt{nh_1h_2(h_1^2 + h_2^2)}) \times \int B_xB_x'dB_\zeta. \tag{43}$$

Next consider B_{n2}, which we write in the form

$$B_{n2} = \sum \Delta x_t x_{t-1}'[\beta_2(z_t, w_t) - \beta_2(z, w)]K_{tz}K_{tw} = \sum \Delta x_t\zeta_t'x_{t-1}. \tag{44}$$

In view of Lemma B.1 (f), we find that

$$B_{n2} \sim_a O_p\left(n\sqrt{h_1h_2(h_1^2 + h_2^2)}\right) \times \int dB_{x\zeta}B_x. \tag{45}$$

Similarly, for B_{n3} we have

$$B_{n3} = \sum x_{t-1}(\Delta x_t)'[\beta_1(z_t, w_t) - \beta_1(z, w)]K_{tz}K_{tw} = \sum x_{t-1}(\Delta x_t)'\eta_t, \tag{46}$$

and in view of Lemma B.1 (e),

$$B_{n3} = O_p\left(n\sqrt{h_1 h_2 h_1^2}\right) \times \int B_x dB_{x\eta}. \tag{47}$$

Turning to $V_n(z, w)$ we again use the partitioned form

$$V_n(z, w) = \sum_t X_t \Delta u_t K_{tz}K_{tw} = \begin{pmatrix} \Delta x_t \Delta u_t K_{tz}K_{tw} \\ x_{t-1}\Delta u_t K_{tz}K_{tw} \end{pmatrix} =: \begin{pmatrix} V_{n1} \\ V_{n2} \end{pmatrix}. \tag{48}$$

Since $\{u_t\}$ is independent of $\{x_t, z_t\}$, we have $\mathbb{E}V_{n1} = 0$ and

$$V_{n1} = O_p\left(\sqrt{nh_1h_2}\right) \times \int dB_{xuK}, \tag{49}$$

where $B_{xuK}(r)$ is the limit Brownian motion of the partial sum process $\frac{1}{\sqrt{nh_1h_2}}\sum_{t=1}^{[nr]} \Delta x_t \Delta u_t K_{tz}K_{tw}$, and $\int dB_{xuK} = B_{xuK}(1)$ has variance matrix $\mathbb{E}[(\Delta x_t)(\Delta x_t)']\mathbb{E}[(\Delta u_t)^2]f(z, w)\nu_0^2(K)$. Similarly, for V_{n2} we have

$$V_{n2} = O_p\left(n\sqrt{h_1h_2}\right) \times \int B_x dB_{uK}, \tag{50}$$

where $B_{uK}(r)$ is the limit Brownian motion of the partial sum process $\frac{1}{\sqrt{nh_1h_2}}\sum_{t=1}^{[nr]} \Delta u_t K_{tz}K_{tw}$.

With these preliminaries in hand, we now proceed to deduce the limit behavior of the estimators given in Theorem 3.1.

(a) We start with the bivariate kernel estimator $\hat{\beta}_1(z, w)$. From the partitioned regression, we have the following decomposition

$$\begin{aligned} &\hat{\beta}_1(z, w) - \beta_1(z, w) \\ &= \Omega_n^{11}[B_{n1}(z, w) + B_{n2}(z, w)] + \Omega_n^{12}[B_{n3}(z, w) + B_{n4}(z, w)] + \Omega_n^{11}V_{n1} + \Omega_n^{12}V_{n2} \\ &\sim_a \Omega_{n11}^{-1}[B_{n1}(z, w) + B_{n2}(z, w)] - \Omega_{n11}^{-1}\Omega_{n12}[\Omega_{n22}]^{-1}[B_{n3}(z, w) + B_{n4}(z, w)] \\ &\quad + \Omega_{n11}^{-1}V_{n1} - \Omega_{n11}^{-1}\Omega_{n12}[\Omega_{n22}]^{-1}V_{n2}. \end{aligned} \tag{51}$$

In view of the analysis of the orders of the components given above and using $h_1 = h_2 = h$, the leading term on the right hand side of (51) is evidently $\Omega_{n11}^{-1}B_{n2}(z, w)$, which is of order $O_p(1)$. Specifically,

$$\begin{aligned} \hat{\beta}_1(z, w) - \beta_1(z, w) &= \Omega_{n11}^{-1}B_{n2} + O_p(h + 1/\sqrt{nh^2}) \\ &\sim_a O_p(1) \times \frac{1}{f(z, w)}\{\mathbb{E}[\Delta x_t(\Delta x_t)']\}^{-1}\int dB_{x\zeta}B_x, \end{aligned} \tag{52}$$

so that the bivariate estimator $\hat{\beta}_1(z,w)$ is actually inconsistent. The inconsistency originates in the nonstationarity of x_t which leads to the nonzero stochastic integral component that is present in (52).

Next consider the time averaged estimator $\hat{\beta}_{1s}(z)$. Noting that $\beta_1(z,w)=\beta(z)$, we have

$$\hat{\beta}_{1s}(z)-\beta(z)=\frac{1}{n}\sum_{t=1}^{n}[\hat{\beta}_1(z,z_t)-\beta_1(z,z_t)]. \tag{53}$$

In view of (52), to obtain the asymptotic distribution of $\hat{\beta}_{1s}(z)$, we need analyze the matrix ratio $\Omega_{n11}^{-1}(z,w)B_{n2}(z,w)$ time averaged with respect to w. For the component Ω_{n11}, we have the uniform result $(nh_1h_2)^{-1}\Omega_{n11}=f(z,w)\mathbb{E}[\Delta x_t(\Delta x_t)']+O_p(c_n)$, where $c_n=h_1^2+h_2^2+\sqrt{\frac{\log n}{nh_1h_2}}$, uniformly for all (z,w) (see, for example, Li and Racine (2007, p.79)). Hence, we need to average of $f^{-1}(z,w)(nh_1h_2)^{-1}B_{n2}$ with respect to w. We have

$$\begin{aligned}
&\frac{1}{n}\sum_s f^{-1}(z,z_s)\frac{1}{nh_1h_2}\sum_t \Delta x_t x_{t-1}'[\beta(z_t)-\beta(z)-(\beta(w_t)-\beta(z_s))]K_{tz}K(\frac{w_t-z_s}{h_2})\\
&=\frac{1}{nh_1}\sum_t \Delta x_t x_{t-1}'K_{tz}\frac{1}{nh_2}\sum_s f^{-1}(z,z_s)[\beta(z_t)-\beta(z)-(\beta(w_t)-\beta(z_s))]K(\frac{w_t-z_s}{h_2})\\
&\sim_a \frac{1}{nh_1}\sum_t \Delta x_t K_{tz}f^{-1}(z,w_t)[\beta(z_t)-\beta(z)]'f(w_t)x_{t-1}\\
&=:\frac{1}{nh_1}\sum_t \Delta x_t\epsilon_{\beta t}'x_{t-1}\\
&\sim_a O_p(\sqrt{h_1})\times\int dB_{x\epsilon}B_x,
\end{aligned} \tag{54}$$

where $\epsilon_{\beta t}=K_{tz}f^{-1}(z,w_t)[\beta(z_t)-\beta(z)]f(w_t)$ and $B_{x\epsilon}$ is the matrix Brownian motion limit of the partial sum process $\frac{1}{\sqrt{nh_1^3}}\sum_t \Delta x_t\epsilon_{\beta t}'$. The third line above follows because

$$\begin{aligned}
&\frac{1}{nh_2}\sum_s f^{-1}(z,z_s)[\beta(z_t)-\beta(z)-(\beta(w_t)-\beta(z_s))]K(\frac{w_t-z_s}{h_2})\\
&\xrightarrow{p} f^{-1}(z,w_t)[\beta(z_t)-\beta(z)-(\beta(w_t)-\beta(w_t))]f(w_t)\\
&=f^{-1}(z,w_t)[\beta(z_t)-\beta(z)]f(w_t).
\end{aligned} \tag{55}$$

The final line (54) follows from Lemma B.1 (h) by virtue of the independence of x_t and z_t. Therefore, with $c_n/\sqrt{h}\to 0$, as assumed in Assumption 1 (vi), we have

$$\frac{1}{\sqrt{h}}[\hat{\beta}_{1s}(z)-\beta(z)]\sim_a\{\mathbb{E}[\Delta x_t(\Delta x_t)']\}^{-1}\int dB_{x\epsilon}B_x. \tag{56}$$

The result follows because the average of the smaller order terms in (51) are all at most of order $o_p(\sqrt{h})$. As pointed out in (52), these terms are originally of order $O_p(h+1/\sqrt{nh^2})$, which is $o_p(\sqrt{h})$ uniformly in w since $nh^3\to\infty$ as implied by $c_n/\sqrt{h}\to 0$, leading to order at most $o_p(\sqrt{h})$ upon averaging.

We now analyze the limit distribution on the right side of (56). In view of Assumption 1 (ii), the Brownian motions B_x and $B_{x\epsilon}$ are independent, as can be verified element-wise. For example, when x_t is univariate we have the sample covariance

$$\begin{aligned}
&\mathbb{E}[\frac{1}{\sqrt{n}}\sum_t \Delta x_t][\frac{1}{\sqrt{nh_1^3}}\sum_t \Delta x_t \epsilon_{\beta t}] = \frac{1}{n\sqrt{h_1^3}}\sum_{t,s}\mathbb{E}[\Delta x_t \Delta x_s]\mathbb{E}\epsilon_{\beta s} \\
=&\frac{\mathbb{E}\epsilon_{\beta s}}{n\sqrt{h_1^3}}\sum_{1\le t,t-\ell\le n}\mathbb{E}[\Delta x_t \Delta x_{t-\ell}] = \frac{\mathbb{E}\epsilon_{\beta s}}{n\sqrt{h_1^3}}\sum_{\ell=1-n}^{n-1}[n-|\ell|]\gamma_{\Delta x}(\ell) \\
=&O(h_1^3)\times\frac{1}{n\sqrt{h_1^3}}\times O(n) = O(h^{3/2}) = o(1),
\end{aligned}$$

where $\gamma_{\Delta x}(\ell) = \mathbb{E}[\Delta x_t \Delta x_{t-\ell}]$ and the second to last equality follows because Δx_t has finite long run variance and $\mathbb{E}\epsilon_{\beta s} = O(h_1^3)$, as shown in (93) in the proof of Lemma B.1 (h) below. Hence conditional on B_x, $\int dB_{x\epsilon}B_x$ is a normal random vector with variance

$$\nu_2(K)\mathbb{E}[f(w_t)/f(z,w_t)]\int(\beta^{(1)}(z)'B_x)^2\mathbb{E}[\Delta x_t(\Delta x_t)'].$$

The asymptotic variance of $\hat{\beta}_{1s}(z)$ is then $\nu_2(K)\mathbb{E}[f(w_t)/f(z,w_t)]\int(\beta^{(1)}(z)'B_x)^2[\mathbb{E}(\Delta x_t)(\Delta x_t)']^{-1}$ and we deduce that

$$\frac{1}{\sqrt{h}}(\hat{\beta}_{1s}(z)-\beta(z)) \rightsquigarrow \mathcal{MN}\left(0, \nu_2(K)\mathbb{E}[f(w_t)/f(z,w_t)]\int(\beta^{(1)}(z)'B_x)^2[\mathbb{E}(\Delta x_t)(\Delta x_t)']^{-1}\right),$$

giving the stated result.

(b) We first consider $\hat{\beta}_2(z,w)$. The estimation error is

$$\begin{aligned}
\hat{\beta}_2(z,w)-\beta_2(z,w) &= \Omega_n^{21}[B_{n1}+B_{n2}] + \Omega_n^{22}[B_{n3}+B_{n4}] + \Omega_n^{21}V_{n1} + \Omega_n^{22}V_{n2} \\
&\sim_a -\Omega_{n22}^{-1}\Omega_{n12}'\Omega_{n11}^{-1}[B_{n1}+B_{n2}] + \Omega_{n22}^{-1}[B_{n3}+B_{n4}] \\
&\quad -\Omega_{n22}^{-1}\Omega_{n12}'\Omega_{n11}^{-1}V_{n1} + \Omega_{n22}^{-1}V_{n2}. \qquad (57)
\end{aligned}$$

In the same way as before for (a), it can be verified that the leading term on the right side of (57) is $\Omega_{n22}^{-1}B_{n4}$, which is of order $O_p(h^2+1/\sqrt{n})$. More specifically, we have

$$\hat{\beta}_2(z,w)-\beta_2(z,w) = \Omega_{n22}^{-1}B_{n4} + O_p(1/(nh)) = o_p(1). \qquad (58)$$

Thus, $\hat{\beta}_2(z,w)$ is consistent. In view of (35) and (43), we have

$$\hat{\beta}_2(z,w)-\beta_2(z,w) \sim_a h^2\mathcal{B}_2(z,w) + \frac{1}{\sqrt{n}}f^{-1}(z,w)\left[\int B_xB_x'\right]^{-1}\int B_xB_x'dB_\zeta, \qquad (59)$$

where $\mathcal{B}_2(z,w) = f^{-1}(z,w)\mu_2(K)[C_1(z,w)-C_2(z,w)]$ is deterministic and $\int B_xB_x'dB_\zeta$ is a $k\times 1$ random vector stochastic integral with zero mean.

The marginal integration estimator $\hat{\alpha}(z)$ is defined as $\hat{\alpha}(z) = \frac{1}{n}\sum_{t=1}^{n}\hat{\beta}_2(z, z_t)$. Since $\beta_2(z, w) = \beta(z) - \beta(w) = \alpha(z) - \alpha(w)$, we have

$$\begin{aligned}\hat{\alpha}(z) - \alpha(z) &= \frac{1}{n}\sum_{t=1}^{n}\hat{\beta}_2(z, z_t) - \alpha(z) \\ &= \frac{1}{n}\sum_{t=1}^{n}[\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] + \frac{1}{n}\sum_{t=1}^{n}\beta_2(z, z_t) - \alpha(z) \\ &= \frac{1}{n}\sum_{t=1}^{n}[\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] - \frac{1}{n}\sum_{t=1}^{n}\alpha(z_t). \end{aligned} \tag{60}$$

Evidently, $\frac{1}{n}\sum_{t=1}^{n}\alpha(z_t) = O_p(1/\sqrt{n})$. We need to analyze the order of the first term in (60) for comparative purposes. Since $\frac{1}{n^2}\sum x_{t-1}x_{t-1}' \rightsquigarrow \int B_x B_x'$ independently of z_t and $\mathbb{E}[\xi_t] = f(z, w)$ we can focus on the sample average of $f^{-1}(z, w)B_{n4}$ with respect to w, or equivalently, the sample average of the two terms in (59). The second term of (59) has zero mean because $\mathbb{E}\int B_x B_x' dB_\zeta = 0$. Similar to the analysis in (54), we can show that after averaging over w, the second term of (59) contributes a term of order $o_p(1/\sqrt{n})$ and is therefore negligible compared with the sample average $\frac{1}{n}\sum_{t=1}^{n}\alpha(z_t)$. Moreover, the average of $\mathcal{B}_2(z, w)$ with respect to w contributes a bias term, which will converge to the expectation $\mathbb{E}\mathcal{B}_2(z, z_t)$. At the same time, the average of the terms of smaller order in (57) is negligible because it is at most of order $O_p(1/(nh)) = o_p(1/\sqrt{n})$ in view of (58). It follows that the limit distribution of $\hat{\alpha}(z)$ is determined by the sample average $\frac{1}{n}\sum_{t=1}^{n}\alpha(z_t)$. Therefore we have

$$\sqrt{n}\left(\hat{\alpha}(z) - \alpha(z) - h^2\mathcal{B}(z)\right) \rightsquigarrow \mathcal{N}(0, \Gamma(\alpha(z_t))), \tag{61}$$

where $\mathcal{B}(z) = \mathbb{E}\mathcal{B}_2(z, z_t) = \mu_2(K)\mathbb{E}\frac{C_1(z,z_t)-C_2(z,z_t)}{f(z,z_t)}$ and $\Gamma(\alpha(z_t))$ denotes the long run variance matrix of $\{\alpha(z_t)\}$.

For the estimator $\tilde{\alpha}(z)$, we have

$$\begin{aligned}&\tilde{\alpha}(z) - \alpha(z) = \hat{\alpha}(z) - \alpha(z) - n^{-1}\sum_{t=1}^{n}\hat{\alpha}(z_t) \\ &= \frac{1}{n}\sum_{t=1}^{n}[\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)] - \frac{1}{n}\sum_{t=1}^{n}[\hat{\alpha}(z_t) - \alpha(z_t)] - \frac{2}{n}\sum_{t=1}^{n}\alpha(z_t) \\ &= -\frac{1}{n}\sum_{t=1}^{n}\alpha(z_t) + \frac{1}{n}\sum_{t=1}^{n}\left[\hat{\beta}_2(z, z_t) - \beta_2(z, z_t)\right] - \frac{1}{n}\sum_{t=1}^{n}\frac{1}{n}\sum_{s=1}^{n}[\hat{\beta}_2(z_t, z_s) - \beta_2(z_t, z_s)]. \end{aligned} \tag{62}$$

Similar to the analysis in (60), we find that (62) is dominated by the term $-\frac{1}{n}\sum_{t=1}^{n}\alpha(z_t)$. The biases coming from the second and third term in (62) are $h^2\mathcal{B}(z)$ and $-h^2\mathbb{E}\mathcal{B}(z_t)$, respectively. Then we have

$$\sqrt{n}[\tilde{\alpha}(z) - \alpha(z) - h^2(\mathcal{B}(z) - \mathbb{E}\mathcal{B}(z_t))] \rightsquigarrow \mathcal{N}(0, \Gamma(\alpha(z_t))). \tag{63}$$

(c) Least squares gives

$$\hat{\beta}_0 = \left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \Delta \tilde{y}_t = \beta_0 + \left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \Delta \tilde{u}_t,$$

where $\tilde{y}_t = y_t - x_t'\hat{\alpha}(z_t)$, $\tilde{u}_t = y_t - x_t'\hat{\alpha}(z_t) - x_t'\beta_0$, and

$$\begin{aligned}\Delta \tilde{u}_t &= y_t - x_t'\hat{\alpha}(z_t) - x_t'\beta_0 - [y_{t-1} - x_{t-1}'\hat{\alpha}(z_{t-1}) - x_{t-1}'\beta_0] \\ &= \Delta y_t - x_t'\hat{\alpha}(z_t) + x_{t-1}'\hat{\alpha}(z_{t-1}) - (\Delta x_t)'\beta_0 \\ &= \Delta u_t + x_t'[\alpha(z_t) - \hat{\alpha}(z_t)] - x_{t-1}'[\alpha(z_{t-1}) - \hat{\alpha}(z_{t-1})].\end{aligned}$$

Thus

$$\begin{aligned}\hat{\beta}_0 = \beta_0 &+ \left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \Delta u_t \\ &+ \left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \{x_t'[\alpha(z_t) - \hat{\alpha}(z_t)] - x_{t-1}'[\alpha(z_{t-1}) - \hat{\alpha}(z_{t-1})]\} \\ =:&\ \beta_0 + A_{1n} + \left[n^{-1} \sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} A_{2n},\end{aligned}$$

defining A_{1n} and A_{2n}. It is easy to see that $A_{1n} = O_p(n^{-1/2})$. Recall that $\hat{\alpha}(z) - \alpha(z) - h^2\mathcal{B}(z) \sim_a -\frac{1}{n}\sum_{s=1}^{n} \alpha(z_s)$. Then

$$\begin{aligned}A_{2n} =& \frac{1}{n}\sum_{t=2}^{n} \Delta x_t \{x_t'[\alpha(z_t) - \hat{\alpha}(z_t)] - x_{t-1}'[\alpha(z_{t-1}) - \hat{\alpha}(z_{t-1})]\} \\ =& \frac{1}{n}\sum_{t=2}^{n} \Delta x_t (\Delta x_t)' \frac{1}{n}\sum_{s=1}^{n} \alpha(z_s) - h^2 \frac{1}{n}\sum_{t=2}^{n} \Delta x_t \Delta[x_t'\mathcal{B}(z_t)],\end{aligned}$$

where

$$\frac{1}{n}\sum_{t=2}^{n} \Delta x_t \Delta[x_t'\mathcal{B}(z_t)] = \frac{1}{n}\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\mathcal{B}(z_t) + \frac{1}{n}\sum_{t=2}^{n} \Delta x_t x_t' \Delta\mathcal{B}(z_t) \rightsquigarrow \mathbb{E}[\Delta x_t (\Delta x_t)']\mathbb{E}\mathcal{B}(z_t) + \int dB_{x\mathcal{B}} B_x, \tag{64}$$

and $B_{x\mathcal{B}}$ is defined as the matrix Brownian motion limit of the associated partial sum process, viz., $\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \Delta x_t [\Delta\mathcal{B}(z_t)]' \rightsquigarrow B_{x\mathcal{B}}(r)$. Then we can write

$$\begin{aligned}&\hat{\beta}_0 - \beta_0 + h^2 \left\{\mathbb{E}\mathcal{B}(z_t) + [\mathbb{E}\Delta x_t (\Delta x_t)']^{-1} \int dB_{x\mathcal{B}} B_x\right\} \\ \sim_a &\left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \Delta u_t + \frac{1}{n}\sum_{s=1}^{n} \alpha(z_s) \qquad (65) \\ \sim_a &\ O_p(1/\sqrt{n}) \times \{\mathcal{N}(0, [\mathbb{E}\Delta x_t (\Delta x_t)']^{-1} \Gamma(\Delta x_t \Delta u_t) [\mathbb{E}\Delta x_t (\Delta x_t)']^{-1}) + \mathcal{N}(0, \Gamma(\alpha(z_t)))\} \\ = &\ O_p(1/\sqrt{n}) \times \mathcal{N}(0, [\mathbb{E}\Delta x_t (\Delta x_t)']^{-1} \Gamma(\Delta x_t \Delta u_t) [\mathbb{E}\Delta x_t (\Delta x_t)']^{-1} + \Gamma(\alpha(z_t))),\end{aligned}$$

where the last step is due to the independence of $\{u_t\}$ and $\{z_t\}$. We therefore have

$$\sqrt{n}(\hat{\beta}_0 - \beta_0 + h^2\mathcal{B}) \rightsquigarrow \mathcal{N}(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t \Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1} + \Gamma(\alpha(z_t))), \tag{66}$$

where the bias $\mathcal{B} = \mathbb{E}\mathcal{B}(z_t) + [\mathbb{E}\Delta x_t(\Delta x)_t']^{-1}\int dB_{x\mathcal{B}}B_x$ is a random vector.

The analysis of $\tilde{\beta}_0$ is similar. We only need to replace $\mathcal{B}(z)$ with $\mathcal{B}(z) - \mathbb{E}\mathcal{B}(z_t) =: \bar{\mathcal{B}}(z)$. So (66) continues to hold for $\tilde{\beta}_0$, with $\mathcal{B}$ replaced by $\bar{\mathcal{B}} = \mathbb{E}\bar{\mathcal{B}}(z_t) + [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\int dB_{x\bar{\mathcal{B}}}B_x$, where $B_{x\bar{\mathcal{B}}}$ is defined as $\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\Delta x_t[\Delta\bar{\mathcal{B}}(z_t)]' \rightsquigarrow B_{x\bar{\mathcal{B}}}(r)$.

(d) In view of (65), we have $\hat{\beta}_0 - \beta_0 \sim_a -h^2\mathcal{B} + [\sum_{t=2}^n \Delta x_t(\Delta x_t)']^{-1}\sum_{t=2}^n \Delta x_t \Delta u_t + \frac{1}{n}\sum_{s=1}^n \alpha(z_s)$, and in view of (60), we can write $\hat{\alpha}(z) - \alpha(z) \sim_a h^2\mathcal{B}(z) - \frac{1}{n}\sum_{s=1}^n \alpha(z_s)$. Then we have

$$\begin{aligned}\hat{\beta}_{2s}(z) - \beta(z) &= \hat{\beta}_0 - \beta_0 + \hat{\alpha}(z) - \alpha(z) \\ &\sim_a h^2(\mathcal{B}(z) - \mathcal{B}) + \left[\sum_{t=2}^n \Delta x_t(\Delta x_t)'\right]^{-1}\sum_{t=2}^n \Delta x_t \Delta u_t,\end{aligned} \tag{67}$$

from which we deduce

$$\sqrt{n}[\hat{\beta}_{2s}(z) - \beta(z) - h^2(\mathcal{B}(z) - \mathcal{B})] \rightsquigarrow \mathcal{N}\left(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t\Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\right), \tag{68}$$

as stated. ■

Proof of Theorem 3.2

When $\beta(z)$ is a constant function, say $\beta(z) = \beta_0$ for all z, different asymptotics apply. First of all, we have

$$\hat{\beta}_1(z, w) - \beta_0 = \Omega_n^{11}V_{n1} + \Omega_n^{12}V_{n2} \sim_a \Omega_{n11}^{-1}V_{n1} - \Omega_{n11}^{-1}\Omega_{n12}[\Omega_{n22}]^{-1}V_{n2}. \tag{69}$$

Based on the order analysis in the proof of Theorem 3.1, we know the first term in the right hand side of (69) is the leading term and the second term can be ignored. Thus we can write

$$\hat{\beta}_1(z, w) - \beta_0 \sim_a \Omega_{n11}^{-1}V_{n1}, \tag{70}$$

from which we deduce the following limit theory

$$\sqrt{nh_1h_2}(\hat{\beta}_1(z, w) - \beta_0) \rightsquigarrow \mathcal{N}(0, \nu_0^2(K)f^{-1}(z, w)\mathbb{E}(\Delta u_t)^2[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}). \tag{71}$$

To analyze the marginal integration estimator $\hat{\beta}_{1s}(z)$, we need to look at the average of $\frac{1}{nh_1h_2}f^{-1}(z, w)V_{n1}$ with respect to w. We have

$$\begin{aligned}&\frac{1}{n}\sum_{s=1}^n f^{-1}(z, z_s)\frac{1}{nh_1h_2}\sum_{t=2}^n \Delta x_t\Delta u_t K_{tz}K(\frac{w_t - z_s}{h_2}) \\ &= \frac{1}{nh_1}\sum_{t=2}^n \Delta x_t\Delta u_t K_{tz}\frac{1}{nh_2}\sum_{s=1}^n f^{-1}(z, z_s)K(\frac{w_t - z_s}{h_2}) \\ &\sim_a \frac{1}{nh_1}\sum_{t=2}^n \Delta x_t\Delta u_t K_{tz}f^{-1}(z, w_t)f(w_t) \\ &= O_p(\frac{1}{\sqrt{nh_1}}) \times \mathcal{N}(0, \mathbb{E}(\Delta u_t)^2\nu_0(K)\mathbb{E}[f(w_t)/f(z, w_t)]\mathbb{E}[\Delta x_t(\Delta x_t)']),\end{aligned} \tag{72}$$

where last equality follows by direct evaluation as

$$
\begin{aligned}
\mathbb{E}K_{tz}^2 f^{-2}(z,w_t)f^2(w_t) &= \int K_{tz}^2 f^{-2}(z,w_t)f^2(w_t)f(z_t,w_t)dz_t dw_t \\
&= \int K^2(u)[f(z,w_t)+f_z^{(1)}(z,w_t)h_1u+\cdots]h_1 du f^{-2}(z,w_t)f^2(w_t)dw_t \\
&= h_1\nu_0(K)\int f(z,w_t)f^{-2}(z,w_t)f^2(w_t)dw_t + o(h_1) \\
&= h_1\nu_0(K)\mathbb{E}[f(w_t)/f(z,w_t)]\{1+o(1)\}. \qquad (73)
\end{aligned}
$$

Thus we have

$$
\sqrt{nh_1}(\hat{\beta}_{1s}(z)-\beta_0) \rightsquigarrow \mathcal{N}(0,\mathbb{E}(\Delta u_t)^2\nu_0(K)\mathbb{E}[f(w_t)/f(z,w_t)][\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}). \qquad (74)
$$

In a similar way for the coefficient of x_{t-1}, we have

$$
\hat{\beta}_2(z,w) = \Omega_n^{21}V_{n1}+\Omega_n^{22}V_{n2} \sim_a -\Omega_{n22}^{-1}\Omega_{n12}'\Omega_{n11}^{-1}V_{n1}+\Omega_{n22}^{-1}V_{n2} \sim_a \Omega_{n22}^{-1}V_{n2}. \qquad (75)
$$

Proceeding as before, we then find that

$$
n\sqrt{h_1h_2}\hat{\beta}_2(z,w) \rightsquigarrow \mathcal{MN}\left(0,\nu_0^2(K)f^{-1}(z,w)\mathbb{E}(\Delta u_t)^2\left[\int B_xB_x'\right]^{-1}\right). \qquad (76)
$$

Then, averaging $\frac{1}{n^2h_1h_2}f^{-1}(z,w)V_{n2}$ with respect to w, we obtain

$$
\begin{aligned}
&\frac{1}{n}\sum_{s=1}^n f^{-1}(z,z_s)\frac{1}{n^2h_1h_2}\sum_{t=2}^n x_{t-1}\Delta u_t K_{tz}K(\frac{w_t-z_s}{h_2}) \\
=\;& \frac{1}{n^2h_1}\sum_{t=2}^n x_{t-1}\Delta u_t K_{tz}\frac{1}{nh_2}\sum_{s=1}^n f^{-1}(z,z_s)K(\frac{w_t-z_s}{h_2}) \\
\sim_a\;& \frac{1}{n^2h_1}\sum_{t=2}^n x_{t-1}\Delta u_t K_{tz}f^{-1}(z,w_t)f(w_t) \\
=\;& O_p(\frac{1}{n\sqrt{h_1}})\times\mathcal{MN}\left(0,\int B_xB_x'\mathbb{E}(\Delta u_t)^2\nu_0(K)\mathbb{E}[f(w_t)/f(z,w_t)]\right). \qquad (77)
\end{aligned}
$$

Note that (60) now becomes $\hat{\alpha}(z)=\frac{1}{n}\sum_{s=1}^n\hat{\beta}_2(z,z_s)$. Therefore we have

$$
n\sqrt{h_1}\hat{\alpha}(z) \rightsquigarrow \mathcal{MN}\left(0,\mathbb{E}(\Delta u_t)^2\nu_0(K)\mathbb{E}[f(w_t)/f(z,w_t)]\left[\int B_xB_x'\right]^{-1}\right). \qquad (78)
$$

For the estimator $\tilde{\alpha}(z)$, we have

$$
\tilde{\alpha}(z) = \hat{\alpha}(z)-n^{-1}\sum_{t=1}^n\hat{\alpha}(z_t) = \frac{1}{n}\sum_{s=1}^n\hat{\beta}_2(z,z_s)+\frac{1}{n}\sum_{t=1}^n\frac{1}{n}\sum_{s=1}^n\hat{\beta}_2(z_t,z_s), \qquad (79)
$$

and following a similar argument as in (77) we can show that $\frac{1}{n}\sum_{t=1}^n\frac{1}{n}\sum_{s=1}^n\hat{\beta}_2(z_t,z_s) = O_p(1/n)$. Thus, $\tilde{\alpha}(z)\sim_a\frac{1}{n}\sum_{s=1}^n\hat{\beta}_2(z,z_s)=\hat{\alpha}(z)$.

For $\hat{\beta}_0$, we first have by substitution

$$\hat{\beta}_0 = \beta_0 + \left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \{\Delta u_t - x_t' \hat{\alpha}(z_t) + x_{t-1}' \hat{\alpha}(z_{t-1})\}.$$

Since $\hat{\alpha}(z) = O_p(1/n\sqrt{h_1})$, we have $x_t'\hat{\alpha}(z_t) = O_p(1/\sqrt{nh_1}) = o_p(1)$. Then

$$\hat{\beta}_0 - \beta_0 \sim_a \left[\sum_{t=2}^{n} \Delta x_t (\Delta x_t)'\right]^{-1} \sum_{t=2}^{n} \Delta x_t \Delta u_t,$$

from which we deduce that

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \rightsquigarrow \mathcal{N}\left(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t \Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\right). \tag{80}$$

The analysis for $\tilde{\beta}_0$ follows the same lines and is omitted.

Finally, for the two-step estimator $\hat{\beta}_{2s}(z)$, we have $\hat{\beta}_{2s}(z) - \beta_0 = \hat{\beta}_0 - \beta_0 + \hat{\alpha}(z) \sim_a \hat{\beta}_0 - \beta_0$ because $\hat{\alpha}(z) = O_p(1/n\sqrt{h_1}) = o_p(1/\sqrt{n})$. Therefore

$$\sqrt{n}(\hat{\beta}_{2s}(z) - \beta_0) \rightsquigarrow \mathcal{N}\left(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t \Delta u_t)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\right).$$

■

Proof of Theorem 4.2.

The proof is similar to the proof of Theorem 3.2 and is simply sketched here. First, noting that for the bootstrapped statistic $\hat{I}_\alpha^*$, the corresponding true model is $y_t^* = x_t'\check{\beta}_0 + u_t^*$, where u_t^* is I(1) according to its generating process. The bootstrap sequence $\{\epsilon_t^* = \Delta u_t^*\}$ evidently satisfies the restrictions imposed on $\{\Delta u_t\}$ given in Assumption 1 (i) (ii). Then following Theorem 3.2, we have $n\sqrt{h}\hat{\alpha}^*(z) \rightsquigarrow \mathcal{MN}(0, \Omega^*(z))$, where $\hat{\alpha}^*(z)$ is the bootstrap analogue of $\hat{\alpha}(z)$, $\Omega^*(z) = \nu_0(K)\mathbb{E}(\Delta u_t^*)^2\mathbb{E}[f(w_t)/f(z, w_t)]\left[\int B_x B_x'\right]^{-1}$ is the bootstrap analogue of $\Omega(z)$. Denote the bootstrap version of $\hat{\beta}_{2s}(z)$ as $\hat{\beta}_{2s}^*(z)$. Then, from Theorem 3.2 we know that $\sqrt{n}(\hat{\beta}_{2s}^*(z) - \check{\beta}_0) \rightsquigarrow \mathcal{N}(0, [\mathbb{E}\Delta x_t(\Delta x_t)']^{-1}\Gamma(\Delta x_t \Delta u_t^*)[\mathbb{E}\Delta x_t(\Delta x_t)']^{-1})$. Therefore, $\Delta\hat{u}_t^* = \Delta u_t^* + o_p(1)$, where $\hat{u}_t^* = y_t^* - x_t'\hat{\beta}_{2s}^*(z_t)$. Thus $\mathbb{E}(\Delta u_t^*)^2$ can be consistently estimated by $(n-1)^{-1}\sum_{t=2}^n(\Delta\hat{u}_t^*)^2$. Then the consistent estimator $\hat{\Omega}^*(z)$ of $\Omega^*(z)$ can be constructed in the same way as $\hat{\Omega}(z)$. Following the same argument as in the proof of Theorem 4.1, we have $(\hat{I}_\alpha^*|\mathcal{W}_n) \rightsquigarrow \chi^2_{km}$. ■

B Useful Lemmas

Lemma B.1. *(a)* $\frac{1}{\sqrt{nh_1h_2}}\sum_{t=1}^{[nr]}\bar{\xi}_t \rightsquigarrow B_\xi(r)$, *where* $\bar{\xi}_t = \xi_t - \mathbb{E}\xi_t$, $\xi_t = K_{tz}K_{tw}$, *and* $B_\xi(r)$ *is scalar Brownian motion with variance given by* (83) *;*

(b) $\frac{1}{\sqrt{nh_1h_2h_1^2}}\sum_{t=1}^{[nr]}\bar{\eta}_t \rightsquigarrow B_\eta(r)$ *where* $\bar{\eta}_t = \eta_t - \mathbb{E}\eta_t$, $\eta_t = (\beta(z_t) - \beta(z))K_{tz}K_{tw}$, *and* $B_\xi(r)$ *is vector Brownian motion with variance matrix given by* (86) *;*

(c) $\frac{1}{\sqrt{nh_1h_2(h_1^2+h_2^2)}}\sum_{t=1}^{[nr]}\bar{\zeta}_t \rightsquigarrow B_\zeta(r)$ *where* $\bar{\zeta}_t=\zeta_t-\mathbb{E}\zeta_t$, $\zeta_t=[\beta(z_t)-\beta(z)-(\beta(w_t)-\beta(w))]K_{tz}K_{tw}$ *and* $B_\zeta(r)$ *is vector Brownian motion with variance matrix given by* (87);

(d) $\frac{1}{\sqrt{nh_1h_2}}\sum_{t=1}^{[nr]}\Delta x_t\xi_t \rightsquigarrow B_{x\xi}(r)$, *where* $B_{x\xi}(r)$ *is vector Brownian motion with variance matrix given by* (89);

(e) $\frac{1}{\sqrt{nh_1h_2h_1^2}}\sum_{t=1}^{[nr]}(\Delta x_t)'\eta_t \rightsquigarrow B_{x\eta}(r)$ *where* $B_{x\eta}(r)$ *is scalar Brownian motion with variance given by* (90);

(f) $\frac{1}{\sqrt{nh_1h_2(h_1^2+h_2^2)}}\sum\Delta x_t\zeta_t' \rightsquigarrow B_{x\zeta}(r)$ *where* $B_{x\zeta}(r)$ *is matrix Brownian motion with variance matrix given in tensor form by* (91);

(g) $\frac{1}{\sqrt{nh_1h_2h_1^2}}\sum_{t=1}^{[nr]}\Delta x_t(\Delta x_t)'\bar{\eta}_t \rightsquigarrow B_{xx\eta}(r)$ *where* $B_{xx\eta}(r)$ *is vector Brownian motion with variance matrix given by* (92);

(h) $\frac{1}{\sqrt{nh_1^3}}\sum_{t=1}^{[nr]}\Delta x_tK_{tz}f^{-1}(z,w_t)[\beta(z_t)-\beta(z)]'f(w_t) \rightsquigarrow B_{x\epsilon}(r)$ *where* $B_{x\epsilon}(r)$ *is matrix Brownian motion with variance matrix given in tensor form by* (95).

Proof (a) First note that $\mathbb{E}[\bar{\xi}_t]=0$. To find out the order of $\bar{\xi}_t$, we need to compute $\mathbb{E}\bar{\xi}_t^2=\mathbb{E}\xi_t^2-[\mathbb{E}\xi_t]^2$, which requires the computation of $\mathbb{E}[\xi_t]$ and $\mathbb{E}[\xi_t^2]$. We have

$$
\begin{aligned}
\mathbb{E}[K_{tz}K_{tw}] &= \int K(\frac{z_t-z}{h_1})K(\frac{w_t-w}{h_2})f(z_t,w_t)dz_tdw_t \\
&= \int K(u)K(v)f(z+h_1u,w+h_2v)h_1duh_2dv \\
&= h_1h_2\int K(u)K(v)[f(z,w)+f_z^{(1)}(z,w)h_1u+f_w^{(1)}(z,w)h_2v+\cdots]dudv \\
&= h_1h_2[f(z,w)+o(1)] \\
&= O(h_1h_2)\times f(z,w),
\end{aligned} \tag{81}
$$

and

$$
\begin{aligned}
\mathbb{E}[K_{tz}^2K_{tw}^2] &= \int K^2(\frac{z_t-z}{h_1})K^2(\frac{w_t-w}{h_2})f(z_t,w_t)dz_tdw_t \\
&= \int K^2(u)K^2(v)f(z+h_1u,w+h_2v)h_1duh_2dv \\
&= h_1h_2\int K^2(u)K^2(v)[f(z,w)+f_z^{(1)}(z,w)h_1u+f_w^{(1)}(z,w)h_2v+\cdots]dudv \\
&= h_1h_2[f(z,w)(\nu_0(K))^2+o(1)] \\
&= O(h_1h_2)\times f(z,w)(\nu_0(K))^2.
\end{aligned} \tag{82}
$$

Then $\mathbb{E}\bar{\xi}_t^2=\mathbb{E}\xi_t^2-[\mathbb{E}\xi_t]^2=O(h_1h_2)+[O(h_1h_2)]^2=O(h_1h_2)\times f(z,w)\nu_0^2(K)$. Therefore $\bar{\xi}_t=O_p(\sqrt{h_1h_2})$. Then to obtain the desired result, we only need to show that the long run variance of $\bar{\xi}_t$, which is $lim_{n\to\infty}\mathbb{E}(\frac{1}{\sqrt{nh_1h_2}}\sum_{t=1}^n(\xi_t-\mathbb{E}\xi_t))^2$, is finite. This is a standard result in the literature, and can be found in, for example, Robinson (1983), Masry (1996). The long run variance is dominated by the variance and so the variance of $B_\xi(1)$ is

$$Var(B_\xi(1)) = f(z,w)\nu_0^2(K). \tag{83}$$

(b) First note that $\mathbb{E}[\bar{\eta}_t] = 0$. To find the order of $\bar{\eta}_t$, we need to compute $\mathbb{E}(\bar{\eta}_t\bar{\eta}_t') = \mathbb{E}(\eta_t\eta_t') - (\mathbb{E}\eta_t)(\mathbb{E}\eta_t)'$, which requires the computation of $\mathbb{E}[\eta_t]$ and $\mathbb{E}(\eta_t\eta_t')$. We have

$$\begin{aligned}
\mathbb{E}[\eta_t] &= \mathbb{E}[(\beta(z_t)-\beta(z))K_{tz}K_{tw}] \\
&= \int(\beta(z_t)-\beta(z))K(\frac{z_t-z}{h_1})K(\frac{w_t-w}{h_2})f(z_t,w_t)dz_tdw_t \\
&= \int[\beta^{(1)}(z)h_1u+\frac{1}{2}\beta^{(2)}(z)h_1^2u^2+\cdots]K(u)K(v) \\
&\quad\times[f(z,w)+f_z^{(1)}(z,w)h_1u+f_w^{(1)}(z,w)h_2v+\cdots]h_1duh_2dv \\
&= h_1h_2\{h_1^2[\beta^{(1)}(z)f_z^{(1)}(z,w)+\frac{1}{2}\beta^{(2)}(z)f(z,w)]\mu_2(K)+o(h_1^2)\} \\
&= O(h_1h_2h_1^2)\times[\beta^{(1)}(z)f_z^{(1)}(z,w)+\frac{1}{2}\beta^{(2)}(z)f(z,w)]\mu_2(K) \\
&=: O(h_1h_2h_1^2)\times C_1(z,w)\mu_2(K),
\end{aligned} \tag{84}$$

and

$$\begin{aligned}
\mathbb{E}(\eta_t\eta_t') =&\mathbb{E}[(\beta(z_t)-\beta(z))(\beta(z_t)-\beta(z))'K_{tz}^2K_{tw}^2] \\
=&\int(\beta(z_t)-\beta(z))(\beta(z_t)-\beta(z))'K^2(\frac{z_t-z}{h_1})K^2(\frac{w_t-w}{h_2})f(z_t,w_t)dz_tdw_t \\
=&\int[\beta^{(1)}(z)h_1u+\frac{1}{2}\beta^{(2)}(z)h_1^2u^2+\cdots][\beta^{(1)}(z)h_1u+\frac{1}{2}\beta^{(2)}(z)h_1^2u^2+\cdots]'K^2(u)K^2(v) \\
&\times[f(z,w)+f_z^{(1)}(z,w)h_1u+f_w^{(1)}(z,w)h_2v+\cdots]h_1duh_2dv \\
=&h_1h_2\{h_1^2[\beta^{(1)}(z)][\beta^{(1)}(z)]'f(z,w)\nu_2(K)\nu_0(K)+o(h_1^2)\} \\
=&O(h_1h_2h_1^2)\times[\beta^{(1)}(z)][\beta^{(1)}(z)]'f(z,w)\nu_2(K)\nu_0(K).
\end{aligned} \tag{85}$$

It follows that $\mathbb{E}(\bar{\eta}_t\bar{\eta}_t') = O(h_1h_2h_1^2)$ and thus $\bar{\eta}_t = O_p(\sqrt{h_1h_2h_1^2})$. Using the same argument as in (a), the long run variance of $\bar{\eta}_t$ is dominated by the variance and thus the claimed convergence holds. The variance matrix of the vector Brownian motion B_η is

$$Var(B_\eta(1)) = [\beta^{(1)}(z)][\beta^{(1)}(z)]'f(z,w)\nu_0(K)\nu_2(K), \tag{86}$$

which is singular and lies in the range space of the vector $\beta^{(1)}(z)$.

(c) Note that $\mathbb{E}\bar{\zeta}_t = 0$. Proceeding in a similar fashion to the proof of (b), we get $\mathbb{E}[\zeta_t] = h_1h_2\mu_2(K)[h_1^2(\beta^{(1)}(z)f_z^{(1)}(z,w)+\frac{1}{2}\beta^{(2)}(z)f(z,w))-h_2^2(\beta^{(1)}(w)f_w^{(1)}(z,w)+\frac{1}{2}\beta^{(2)}(w)f(z,w))]\{1+o(1)\} =: h_1h_2\mu_2(K)[h_1^2C_1(z,w)-h_2^2C_2(z,w)]\{1+o(1)\}$. Further,

$$\mathbb{E}[\zeta_t\zeta_t'] = h_1h_2f(z,w)\nu_0(K)\nu_2(K)[h_1^2(\beta^{(1)}(z))(\beta^{(1)}(z))'+h_2^2(\beta^{(1)}(w))(\beta^{(1)}(w))']\{1+o(1)\}.$$

Then $\mathbb{E}(\bar{\zeta}_t\bar{\zeta}_t') = O(h_1h_2(h_1^2+h_2^2))$ and $\bar{\zeta}_t = O_p(\sqrt{h_1h_2(h_1^2+h_2^2)})$. With $h_1 = h_2 = h$, we can write $Var(\bar{\zeta}_t) = h^4 f(z,w)\nu_0(K)\nu_2(K)[(\beta^{(1)}(z))(\beta^{(1)}(z))' + (\beta^{(1)}(w))(\beta^{(1)}(w))']\{1+o(1)\}$. The variance matrix of the vector Brownian motion B_ζ is

$$Var(B_\zeta(1)) = f(z,w)\nu_0(K)\nu_2(K)[\beta^{(1)}(z)(\beta^{(1)}(z))' + \beta^{(1)}(w)(\beta^{(1)}(w))'], \tag{87}$$

which is singular if the dimension $k > 2$.

(d) Note that $\mathbb{E}[\Delta x_t\xi_t] = 0$ by Assumption 1 (ii). In view of the proof of (a), we have $\xi_t = \bar{\xi}_t + \mathbb{E}\xi_t = O_p(\sqrt{h_1h_2}) + O(h_1h_2) = O_p(\sqrt{h_1h_2})$. Therefore by a functional law for mixing random variables we have $\frac{1}{\sqrt{nh_1h_2}}\sum_{t=1}^{[nr]}(\Delta x_t)\xi_t \rightsquigarrow B_{x\xi}(r)$ for a vector Brownian motion $B_{x\xi}$ whose variance matrix is given by the long run variance of $\Delta x_t\xi_t$. We have

$$\begin{aligned}
lrvar(\Delta x_t\xi_t) &= \frac{1}{nh_1h_2}\mathbb{E}\left[\sum_{t=2}^{n}\Delta x_t\xi_t\right]\left[\sum_{t=2}^{n}(\Delta x_t)'\xi_t\right] = \frac{1}{nh_1h_2}\sum_{t,s=2}^{n}\mathbb{E}\Delta x_t(\Delta x_s)'\xi_t\xi_s \\
&= \frac{1}{nh_1h_2}[n\mathbb{E}\Delta x_t(\Delta x_t)'\mathbb{E}\xi_t^2 + 2\sum_{j=1}^{n-1}(n-j)\mathbb{E}\Delta x_t(\Delta x_{t-j})'\mathbb{E}\xi_t\xi_{t-j}] \\
&= \mathbb{E}\Delta x_t(\Delta x_t)'[(h_1h_2)^{-1}\mathbb{E}\xi_t^2] + 2h_1h_2\sum_{j=1}^{n-1}(n-j)\mathbb{E}\Delta x_t(\Delta x_{t-j})'\{(h_1^2h_2^2)^{-1}\mathbb{E}\xi_t\xi_{t-j}]\}.
\end{aligned} \tag{88}$$

Note that $\mathbb{E}\xi_t^2 = O(h_1h_2)$ and $\mathbb{E}\xi_t\xi_{t-j} = O(h_1^2h_2^2)$ uniformly for $j > 0$. Since Δx_t has finite long run variance by Assumption 1 (i), the second term in (88) is of order $O(h_1h_2)$ and can be ignored. Hence, $lrvar(\Delta x_t\xi_t) = \mathbb{E}\Delta x_t(\Delta x_t)'f(z,w)\nu_0^2(K)$. It follows that the variance matrix of the vector Brownian motion $B_{x\xi}$ is

$$Var(B_{x\xi}(1)) = \mathbb{E}[\Delta x_t(\Delta x_t)']f(z,w)\nu_0^2(K). \tag{89}$$

(e) Note that $\mathbb{E}[(\Delta x_t)'\eta_t] = 0$. In the analysis of (b), we have shown that $\eta_t = \bar{\eta}_t + \mathbb{E}\eta_t = O_p(\sqrt{h_1h_2h_1^2})$. Therefore the claimed convergence holds. Similar to the analysis in (88), we can verify that the long run variance of $(\Delta x_t)'\eta_t$ is dominated by its variance. Thus the variance of the scalar Brownian motion $B_{x\eta}(1)$ is

$$Var(B_{x\eta}(1)) = \mathbb{E}\{[(\Delta x_t)'\eta_t]^2\} = f(z,w)\nu_2(K)\nu_0(K)\mathbb{E}\{[(\Delta x_t)'\beta^{(1)}(z)]^2\}. \tag{90}$$

(f) Note that $B_{x\zeta}$ is a $k\times k$ Brownian motion matrix. It is sufficient to verify the elementwise convergence and focus on the scalar x_t case. First $\mathbb{E}\{\Delta x_t\zeta_t'\} = 0$ by Assumption 1 (ii). In the proof of (c), we have seen that $\zeta_t = \bar{\zeta}_t + \mathbb{E}[\zeta_t] = O_p(\sqrt{h_1h_2(h_1^2+h_2^2)})$. Then the claimed convergence holds by functional limit theory. Similar to the analysis in (88), the long run variance of $\Delta x_t \zeta_t'$ is dominated by its variance. When x_t is a scalar, the variance of $B_{x\zeta}(1)$ is $Var(B_{x\zeta}(1)) = \mathbb{E}[\Delta x_t \zeta_t' \zeta_t (\Delta x_t)'] = \mathbb{E}[(\Delta x_t)^2] \mathbb{E}\zeta_t^2 = \mathbb{E}[(\Delta x_t)^2] f(z,w)\nu_0(K)\nu_2(K)[(\beta^{(1)}(z))^2 + (\beta^{(1)}(w))^2]$. In the vector x_t case we work with the vector $\Delta x_t \otimes \zeta_t$. The variance matrix of the matrix Brownian motion is then written as the variance matrix of the vector Brownian motion $B_{x\otimes\zeta}$, which is just

$$
\begin{aligned}
&Var(B_{x\otimes\zeta}(1)) = \mathbb{E}[\Delta x_t \Delta x_t' \otimes \zeta_t \zeta_t'] = \mathbb{E}\Delta x_t \Delta x_t' \otimes \mathbb{E}\zeta_t\zeta_t' \\
=& f(z,w)\nu_0(K)\nu_2(K)\mathbb{E}\Delta x_t \Delta x_t' \otimes [\beta^{(1)}(z)(\beta^{(1)}(z))' + \beta^{(1)}(w)(\beta^{(1)}(w))']. \qquad (91)
\end{aligned}
$$

(g) First we have $\mathbb{E}\Delta x_t(\Delta x_t)'\bar{\eta}_t = 0$. In the proof of (b), we have shown that $\bar{\eta}_t = O_p(\sqrt{h_1 h_2 h_1^2})$. So the claimed convergence holds by functional limit theory. Similar to (88), we can verify that the long run variance of $\Delta x_t (\Delta x_t)' \bar{\eta}_t$ is dominated by its variance. The variance matrix of $B_{xx\eta}(1)$ is therefore given by

$$
Var(B_{xx\eta}(1)) = \mathbb{E}[\Delta x_t (\Delta x_t)' \bar{\eta}_t \bar{\eta}_t' \Delta x_t (\Delta x_t)'] = \mathbb{E}\{(\Delta x_t)(\Delta x_t)'[(\Delta x_t)'\beta^{(1)}(z)]^2\} f(z,w)\nu_2(K)\nu_0(K). \qquad (92)
$$

(h) Note that $B_{x\epsilon}$ is a $k \times k$ matrix Brownian motion. To prove the desired convergence it is sufficient to prove the element-wise convergence and focus on the case where x_t is scalar. Let $\epsilon_{\beta t} = K_{tz} f^{-1}(z,w_t)[\beta(z_t) - \beta(z)] f(w_t)$. We first show $\epsilon_{\beta t} = O_p(\sqrt{h_1^3})$. This can be seen from the fact that

$$
\begin{aligned}
\mathbb{E}\epsilon_{\beta t} =& \int K_{tz} f^{-1}(z,w_t)[\beta(z_t) - \beta(z)] f(w_t) f(z_t,w_t) dz_t dw_t \\
=& \int K(u) f^{-1}(z,w_t)[\beta^{(1)}(z)h_1 u + \frac{1}{2}\beta^{(2)}(z)h_1^2 u^2 + \cdots] f(w_t) \\
& \times [f(z,w_t) + f_z^{(1)}(z,w_t)h_1 u + \cdots] h_1 du dw_t \\
=& h_1 \int f^{-1}(z,w_t)\{h_1^2[\beta^{(1)}(z) f_z^{(1)}(z,w_t) + \frac{1}{2}\beta^{(2)}(z) f(z,w_t)]\mu_2(K) + o(h_1^2)\} f(w_t) dw_t \\
=& O(h_1^3) \times \mu_2(K) \int f^{-1}(z,w_t)[\beta^{(1)}(z) f_z^{(1)}(z,w_t) + \frac{1}{2}\beta^{(2)}(z) f(z,w_t)] f(w_t) dw_t, \qquad (93)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\epsilon_{\beta t}\epsilon_{\beta t}' =& \int K_{tz}^2 f^{-2}(z,w_t)[\beta(z_t) - \beta(z)][\beta(z_t) - \beta(z)]' f^2(w_t) f(z_t,w_t) dz_t dw_t \\
=& \int K^2(u) f^{-2}(z,w_t)[\beta^{(1)}(z)h_1 u + \frac{1}{2}\beta^{(2)}(z)h_1^2 u^2 + \cdots][\beta^{(1)}(z)h_1 u + \frac{1}{2}\beta^{(2)}(z)h_1^2 u^2 + \cdots]' \\
& \times f^2(w_t)[f(z,w_t) + f_z^{(1)}(z,w_t)h_1 u + \cdots] h_1 du dw_t \\
=& h_1 \int f^{-2}(z,w_t)\{h_1^2(\beta^{(1)}(z))(\beta^{(1)}(z))' f(z,w_t)\nu_2(K) + o(h_1^2)\} f^2(w_t) dw_t \\
=& O(h_1^3) \times (\beta^{(1)}(z))(\beta^{(1)}(z))'\nu_2(K) \int f^{-1}(z,w_t) f^2(w_t) dw_t \\
=& O(h_1^3) \times (\beta^{(1)}(z))(\beta^{(1)}(z))'\nu_2(K)\mathbb{E}[f(w_t)/f(z,w_t)]. \qquad (94)
\end{aligned}
$$

Then the claimed convergence holds. Similar to the analysis in (88), we can show that the long run variance of $\Delta x_t \epsilon'_{\beta t}$ is dominated by its variance. Therefore, when x_t is a scalar, the variance matrix of $B_{x\epsilon}(1)$ is $Var(B_{x\epsilon}(1)) = \nu_2(K)\mathbb{E}[f(z_t)/f(z, z_t)]\mathbb{E}[(\Delta x_t)^2][\beta^{(1)}(z)]^2$. In the matrix case, we work with the vector $\Delta x_t \otimes \epsilon_{\beta t}$. The variance matrix of the matrix Brownian motion is then written as the variance matrix of the vector Brownian motion $B_{x\otimes\epsilon}$, which is just

$$\begin{aligned} &Var(B_{x\otimes\epsilon}(1)) = \mathbb{E}[\Delta x_t(\Delta x_t)' \otimes \epsilon_{\beta t}\epsilon'_{\beta t}] = \mathbb{E}\Delta x_t(\Delta x_t)' \otimes \mathbb{E}\epsilon_{\beta t}\epsilon'_{\beta t} \\ =&\nu_2(K)\mathbb{E}[f(z_t)/f(z, z_t)]\mathbb{E}\Delta x_t(\Delta x_t)' \otimes (\beta^{(1)}(z))(\beta^{(1)}(z))', \end{aligned} \tag{95}$$

which is singular. ■

References

Bachmeier, L. J. and Swanson, N. R. (2005). Predicting inflation: does the quantity theory help? *Economic Inquiry*, 43(3):570–585.

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

Cai, Z. and Masry, E. (2000). Nonparametric estimation of additive nonlinear arx time series: local linear fitting and projections. *Econometric Theory*, 16(4):465–501.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167(2):413–425.

Chen, R., Härdle, W., Linton, O. B., and Severance-Lossin, E. (1996). Nonparametric estimation of additive separable regression models. In *Statistical Theory and Computational Aspects of Smoothing: Proceedings of the COMPSTAT'94 Satellite Meeting held in Semmering, Austria, 27–28 August 1994*, pages 247–265. Springer.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55:251–276.

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.

Gan, L., Hsiao, C., and Xu, S. (2014). Model specification test with correlated but not cointegrated variables. *Journal of Econometrics*, 178:80–85.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121–130.

Juhl, T. (2005). Functional-coefficient models under unit root behaviour. *The Econometrics Journal*, 8(2):197–213.

Li, H., Lin, Z., and Hsiao, C. (2015). Testing purchasing power parity hypothesis: A semiparametric varying coefficient approach. *Empirical Economics*, 48(1):427–438.

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

Liang, H., Shen, Y., and Wang, Q. (2022). Functional-coefficient cointegrating regression with endogeneity. *Advances in Econometrics, forthcoming*.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, pages 93–100.

Linton, O. B. (1997). Miscellanea efficient estimation of additive nonparametric regression models. *Biometrika*, 84(2):469–473.

Linton, O. B. (2000). Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory*, 16(4):502–523.

Magdalinos, T. and Phillips, P. C. B. (2019). Hypothesis testing under matrix normalization. *Working Paper, University of Southampton*.

Masry, E. (1996). Multivariate regression estimation local polynomial fitting for time series. *Stochastic Processes and their Applications*, 65(1):81–101.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(2):1–21.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3):311–340.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica: Journal of the Econometric Society*, pages 1021–1043.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica*, 63:1023–1023.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica*, pages 1299–1325.

Phillips, P. C. B. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory*, 25(6):1466–1497.

Phillips, P. C. B. and Kheifets, I. (2019). On multicointegration. *Working Paper, Yale University*.

Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.

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

Phillips, P. C. B. and Wang, Y. (2023). When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Journal of Econometrics*, 232(2):469–489.

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

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235.

Sun, Y., Cai, Z., and Li, Q. (2013). Semiparametric functional coefficient models with integrated covariates. *Econometric Theory*, 29(03):659–672.

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.

Tjøstheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association*, 89(428):1398–1409.

Tu, Y. and Wang, Y. (2019). Functional coefficient cointegration models subject to time–varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, 8(6):1401–1423.

Tu, Y. and Wang, Y. (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 Wang, Y. (2022). Spurious functional-coefficient regression models and robust inference with marginal integration. *Journal of Econometrics*, 229(2):396–421.

Vogelsang, T. J. and Wagner, M. (2014). Integrated modified ols estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics*, 178(2):741–760.

Wang, Q. and Phillips, P. C. B. (2009). Structural nonparametric cointegrating regression. *Econometrica*, 77(6):1901–1948.

Wang, Q., Phillips, P. C. B., and Wang, Y. (2023). New asymptotics applied to functional coefficient regression and climate sensitivity analysis. *Working Paper, Yale University*, pages 1–45.

Wang, Y., Tu, Y., and Chen, S. X. (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.

Yang, L., Park, B. U., Xue, L., and Härdle, W. (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association*, 101(475):1212–1227.