

FUNCTIONAL COEFFICIENT NONSTATIONARY REGRESSION

By

Jiti Gao and Peter C.B. Phillips

September 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1911



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

<http://cowles.econ.yale.edu/>

Functional Coefficient Nonstationary Regression*

Jiti Gao[†] and Peter C. B. Phillips[‡]

August 3, 2013

Abstract

This paper studies a general class of nonlinear varying coefficient time series models with possible nonstationarity in both the regressors and the varying coefficient components. The model accommodates a cointegrating structure and allows for endogeneity with contemporaneous correlation among the regressors, the varying coefficient drivers, and the residuals. This framework allows for a mixture of stationary and nonstationary data and is well suited to a variety of models that are commonly used in applied econometric work. Nonparametric and semiparametric estimation methods are proposed to estimate the varying coefficient functions. The analytical findings reveal some important differences, including convergence rates, that can arise in the conduct of semiparametric regression with nonstationary data. The results include some new asymptotic theory for nonlinear functionals of nonstationary and stationary time series that are of wider interest and applicability and subsume much earlier research on such systems. The finite sample properties of the proposed econometric methods are analyzed in simulations. An empirical illustration examines nonlinear dependencies in aggregate consumption function behavior in the US over the period 1960 - 2009.

JEL Classifications: C13, C14, C23.

Keywords: Aggregate consumption, Asymptotic theory, cointegration, density, local time, nonlinear functional, nonparametric estimation, semiparametric, time series, varying coefficient model.

*The authors thank participants at many conferences and seminars for comments on earlier versions of this work. Gao acknowledges Australian Research Council Discovery Grants Program support under Grant number: DP1096374. Phillips acknowledges support from the NSF under Grant Nos. SES-0956687 and SES-1258258.

[†]*The University of Adelaide & Monash University.*

[‡]*Yale University, University of Auckland, University of Southampton, & Singapore Management University.*

1 Introduction

In recent years, there has been renewed econometric interest in time-varying coefficient models and their potential applications in economics. These models offer additional flexibility in practical work and often correspond more closely to underlying theory specifications than conventional fixed coefficient models. They are usually motivated by the potential instability in economic relations over time, as emphasized in the recent work of Müller and Petalas (2010). Allowing for time evolution in relationships can be particularly important when dealing with long time series where the stability of a relationship comes naturally into question. Examples abound in economics and finance, where institutional and technological changes may change the effects of fundamentals, the manner in which variables are related, or possible transitions over time (Phillips and Sul, 2007). Other examples arise in environmental science, resource depletion, and climate change regressions, where trends over long time frames are evidently endogenous to the sample period selected and considerable flexibility is needed in the specification of the trend mechanism (see Phillips, 2010). Trending time series models are also relevant in modeling macroeconomic activity and financial bubbles, where transition effects can alter cointegrating relations, the form of the nonstationarity, and introduce shifts in regimes (Phillips and Yu, 2011).

Early work by Robinson (1989) pioneered a nonparametric kernel regression approach to the estimation of time-varying coefficient models in a deterministic trend framework. Since then a vast empirical literature has emerged demonstrating that many economic variables manifest certain types of stochastic trend nonstationarity, primarily of the unit root or near unit root variety. Such variables include the nominal three month US Treasury bill rate (e.g., Hamilton 1994; Gao *et al* 2009a) as well as total consumption and income variables (e.g., Hall 1978; Campbell and Mankiw 1990; Muellbauer and Lattimore 1999). Some time series in economics, finance and climatology also show evidence of deterministic drift over various time periods, including consumer price index series in economics, stock price series in finance and temperature series in climatology. The nonstationarity in such series may then involve an unknown deterministic trend function in the systematic component of the model and there may be further complications arising from the superposition of deterministic and stochastic trends.

While it may be appealing in principle to address the complications of time varying coefficient variation and data nonstationarity in a nonparametric regression formulation, our experience suggests that a completely nonparametric time series approach may not be workable even for the bivariate case because of poor convergence characteristics arising from

multivariate nonstationarity. We are therefore motivated to consider a class of more specific but still rather general varying coefficient models of the form

$$\begin{aligned} Y_t &= X_t' \beta(Z_t, U_t) + e_t, \\ X_t &= X_{t-1} + \mu_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{1.1}$$

where (U_t, μ_t, ν_t) is a vector of weakly dependent linear processes generated by a sequence of independent and identically distributed (*iid*) variates $\{\xi_t\}$, $\{e_t\}$ is a martingale difference sequence that may be correlated with $\{\xi_t\}$, the functional coefficient $\beta(z, u)$ is assumed to be continuous in $(z, u) \in \mathbb{R}^{d_z+d_u} = (-\infty, \infty)^{d_z+d_u}$, and $1 \leq t \leq n$ with sample size n . Our specification of (1.1) allows U_t to be multivariate, while Z_t is univariate in the case where the unknown function $\beta(z, u)$ is estimated by a nonparametric method. The variable U_t may be endogenously correlated with the innovations (μ_t, ν_t) , a feature that allows for control function and augmented regression formulations. A strongly endogenous case of this type where $U_t = (\mu_t', \nu_t')'$ is considered in Example 2.6 below and then in the empirical example of Section 6. Model (1.1) imposes a unit root structure on X_t and Z_t to allow for possible cointegration in the relationship between X_t and Y_t and stochastic trend nonstationarity in one of the functional coefficient arguments (Z_t). As discussed in Remark A.2 of Appendix A, the discussion of this paper remains valid when U_t is generated by $U_t = \Lambda(\xi_{t-1}, \dots, \xi_{t-\tau}; \eta_t)$, where $\tau \geq 1$ is a positive integer, $\Lambda(\cdot, \dots, \cdot)$ is a measurable function and $\{\eta_t\}$ is another vector of *iid* random variables and independent of $\{\xi_t\}$. Extensions to near integrated processes seem also possible under some further conditions and with some modification of the methods but will not be considered in this paper.

In the parametric case, functional coefficient time series models have already been proposed to deal with various modelling problems involving explanatory variables in the coefficients of a time series model. For instance, in modelling the dependence between the return of a share and the market return, a short-term interest rate variable may naturally be involved in determining the betas of a CAPM structure (see, for example, Faff and Brooks 1998), thereby inducing variable dependence in the coefficients. Similar considerations apply to models of consumption behavior that allow for nonstationarity and varying coefficients (Hall, 1978). Some research has recently appeared that seeks to address this type of specification. Following the paper by Cai, Fan and Yao (2000) in the stationary nonparametric case, Cai, Li and Park (2009) consider a functional coefficient time series model where nonstationarity is involved in either the parametric regressors or the coefficients, Xiao (2009) discusses a functional coefficient cointegrating model where nonstationarity is only involved

in the parametric regressors, and Sun, Cai and Li (2013) further consider the case where both the parametric and nonparametric regressors are nonstationary. To cover a wider range of possible dependencies that may arise in practical work, model (1.1) allows for a varying coefficient structure in which potentially correlated stationary and nonstationary variables may both be accommodated within the same model while also permitting endogeneity in the regressors and covariates that drive the time varying coefficients.

We focus on the leading case where model (1.1) allows for a unit root stochastic trend in both X_t and Z_t . As discussed in Section 2, this model includes many existing nonparametric and semiparametric models. Some useful new models that allow for endogeneity and incorporate nonstationarity are also covered in this framework. This includes several nonlinear simultaneous equations systems that extend the models used in Newey, Powell and Vella (1999), Newey and Powell (2003), Su and Ullah (2008), and Florens, Johannes and van Bellegem (2012) to a nonstationary data setting. In the semiparametric case, (1.1) includes partially linear models with nonstationary regressors and therefore complements existing work on semiparametric regression for the stationary regressor case (e.g. Robinson 1988; Härdle, Liang and Gao 2000; Fan and Yao 2003; Gao 2007; Li and Racine 2007; Teräsvirta, Tjøstheim and Granger 2010).

Other related work for the nonstationary case has covered a variety of more specific models, including Park and Phillips (1999, 2001) for parameter estimation in parametric nonlinear regression, Phillips and Park (1998), Karlsen and Tjøstheim (2001) and Karlsen, Mykelbust and Tjøstheim (2007) for nonparametric nonstationary autoregression, Gao *et al* (2009a, 2009b) and Wang and Phillips (2012) for nonparametric specification testing in a nonstationary environment, Phillips (2009) for nonparametric kernel estimation involving integrated time series and potentially spurious regression, Wang and Phillips (2009a) for nonparametric cointegration estimation, Wang and Phillips (2009b) for structural nonparametric cointegration estimation by kernel regression, and Gao and Phillips (2010) for estimation in semiparametric simultaneous systems with time series data.

The present work based on model (1.1) has the following novel features:

- (i) The model allows for stochastic trends in both the regressor X_t and the varying coefficient covariate Z_t , while permitting another covariate U_t that is weakly dependent. A distinctive feature is that (1.1) involves both stationary and nonstationary covariate drivers in the functional form $\beta(\cdot, \cdot)$.
- (ii) The system is nonlinear, simultaneous and possibly cointegrating. There is contemporaneous correlation among the variates $\{(X_t, Z_t)\}$ and $\{U_t\}$, so endogeneity and some

heteroskedasticity may be accommodated (the latter by way of martingale difference sequence innovations and suitable functional limit theory for partial sum processes, as in Assumption A.1 (iii)-(iv).

- (iii) The models and proposed estimation methods provide a natural way of addressing the endogeneity, cointegration structure, and time varying coefficient behavior that is inherited from some commonly used modeling frameworks in econometrics and other disciplines, as discussed in the examples given in Section 2.
- (iv) The econometric methods and asymptotic theory involve new results for several non-parametric and semiparametric regression models in the presence of a mixture of stationary and integrated regressors. In consequence, both the model (1.1) and the limit theory considerably extend and complement existing work on asymptotics for nonstationary and varying coefficient models.

The rest of the paper is organized as follows. Several important examples of model (1.1) that are suited to econometric implementation are discussed in Section 2. Sections 3 and 4 propose estimation methods and establish asymptotics. The simulations reported in Section 5 assess the finite sample performance of the proposed estimation methods and limit theory in three illustrative examples. Section 6 provides an empirical application to US income, consumption and interest rate data, motivated by theory considerations that induce a varying coefficient structure (Hall, 1978). The framework of model (1.1) is used to examine whether more specific existing models, such as those in Campbell and Mankiw (1990), are empirically supported within this general framework. Section 7 concludes and discusses future research. Technical assumptions are given in Appendix A together with some useful preliminary lemmas. Proofs of the main theorems are in Appendix B.

2 Nonparametric and semiparametric cointegration

We discuss several special cases of model (1.1) to illustrate the range of potential application of our methods and results. Each of these models has its own particular features, which we consider separately below. Estimation and limit theory for the models along with model (1.1) are covered in Sections 3 and 4. For convenience, we assume $X_0 = Z_0 = O_P(1)$ throughout the paper. Extensions that allow for other initializations, including distant past initializations, may be considered following the lines of Phillips and Magdalinos (2009) but are not developed here. The dimension of the covariate Z_t is d_ν and when Z_t enters nonparametrically we will assume that $d_\nu = 1$.

Model 2.1 (*Multivariate nonparametric model*): Let $d_x = 1$ and $X_t \equiv 1$. Then (1.1) is a bivariate nonparametric model of the form

$$\begin{aligned} Y_t &= \beta(Z_t, U_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.1}$$

where both $\{\nu_t\}$ and $\{U_t\}$ are linear processes. This system (2.1) allows for the inclusion of both stationary and nonstationary regressors in the same model. The model may be interpreted as a nonparametric simultaneous equations model and complements existing models with only stationary regressors (e.g., Newey, Powell and Vella 1999; Newey and Powell 2003; Su and Ullah 2008) by permitting nonstationary components.

Model 2.2 (*Additive nonparametric regression*): Set $d_x = 2$, $X_t = (1, 1)'$ and $\beta(z, u) = (\beta_1(z), \beta_2(u))'$. In this case, (1.1) is a nonparametric additive model of the form

$$\begin{aligned} Y_t &= \beta_1(Z_t) + \beta_2(U_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.2}$$

where $\{U_t\}$ is a stationary time series, and both $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are unknown functions.

Model (2.2) includes nonparametric simultaneous equations such as

$$\begin{aligned} Y_t &= m(Z_t) + \varepsilon_t, \\ Z_t &= Z_{t-1} + \nu_t, \\ \mathbb{E}[\varepsilon_t | \nu_t, \dots, \nu_{t-\tau}] &= \lambda(\nu_t, \dots, \nu_{t-\tau}) \end{aligned} \tag{2.3}$$

for some $\tau \geq 0$, since system (2.3) can be written in augmented regression (control variate) form as

$$\begin{aligned} Y_t &= m(Z_t) + \lambda(\nu_t, \dots, \nu_{t-\tau}) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.4}$$

which is a special case of model (2.2) with $\beta_1(\cdot) = m(\cdot)$, $\beta_2(\cdot) = \lambda(\cdot)$, $U_t = (\nu_t, \dots, \nu_{t-\tau})'$ and $e_t = \varepsilon_t - \lambda(\nu_t, \dots, \nu_{t-\tau})$. We establish an estimation limit theory for model (2.4) in Corollary 4.2 below.

Model (2.2) complements existing additive models and nonparametric studies (such as Gao, 2007) where the regressors are all assumed to be stationary. Models (2.1) and (2.2) have various empirical applications where it is important to allow for nonstationarity, such as relationships between stock prices, long-term bond yields and the Treasury bill rate.

Model 2.3 (*Time varying and varying coefficient models*): Let $X_t = (X_{1t}, 1)'$ and $\beta(z, u) = (\beta_1(z), \beta_2(u))'$. In this case, (1.1) is a trending nonparametric model with varying coefficients of the form

$$\begin{aligned} Y_t &= X_{1t}'\beta_2(U_t) + \beta_1(Z_t) + e_t, \\ X_{1t} &= X_{1t-1} + \mu_t, \\ Z_t &= Z_{t-1} + \nu_t. \end{aligned} \tag{2.5}$$

Model (2.5) can incorporate both stationary and nonstationary regressors in the same additive system and therefore complements existing studies with restrictive classes of regressors and covariates (such as, Cai, Li and Park 2009; Xiao 2009). Model (2.5) has various empirical applications where multiple sources of nonstationarity are involved, such as the dependence of commodity group consumption expenditures on total consumption and disposable income.

Model 2.4 (*Partial linear model A*): Let $X_t = (X_{1t}, 1)'$ and $\beta(u, z) = (\beta, \beta_1(z))'$. In this case, (1.1) is a partial linear model of the form

$$\begin{aligned} Y_t &= X_{1t}'\beta + \beta_1(Z_t) + e_t, \\ X_{1t} &= X_{1t-1} + \mu_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.6}$$

where β is a $d_x - 1$ -dimensional vector of unknown parameters and $\beta_1(\cdot)$ is an unknown univariate function. Model (2.6) allows for the case where both the parametric and nonparametric regressors are nonstationary and may be generated by correlated linear processes. In a separate study, Gao and Phillips (2013) consider the case where X_{1t} is generated as $X_{1t} = H(Z_t) + \zeta_t$, in which $H(\cdot)$ is an unknown function of a trending regressor Z_t and ζ_t is a stationary time series independent of Z_t .

Model 2.5 (*Partial linear model B*): Consider the case where $d_x = d_z = 1$, $X_t \equiv 1$ and $\beta(u, z) = z\theta + \beta_1(z)$. Then (1.1) is a further partial linear model of the form

$$\begin{aligned} Y_t &= Z_t\theta + \beta_1(Z_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.7}$$

where θ is an unknown parameter measuring the linear impact of Z_t . When β_1 is an integrable function, the parameter θ and function $\beta_1(\cdot)$ are separable and identifiable in view of the stronger signal in the linear component of the model, as discussed in Section 4.2 below. A stationary counterpart of model (2.7) is $Y_t = U_t'\theta + \beta_2(U_t) + e_t$, for which a similar identification issue arises and additional conditions are required for identification.

Model 2.6 (*Partial linear model C*): Let $d_x = 1$, $X_t \equiv 1$ and $\beta(u, z) = z\gamma + \beta_2(u)$. In this case, model (1.1) takes the partial linear form

$$\begin{aligned} Y_t &= Z_t\gamma + \beta_2(U_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.8}$$

where γ is an unknown parameter and $\beta_2(\cdot)$ is an integrable function such that $\mathbb{E}[|\beta_2(U_t)|] < \infty$. This model (2.8) covers parametric linear specifications of the form

$$\begin{aligned} Y_t &= Z_t\gamma + \varepsilon_t, \\ Z_t &= Z_{t-1} + \nu_t, \\ \mathbb{E}[\varepsilon_t | \nu_t, \dots, \nu_{t-\tau}] &= \lambda(\nu_t, \dots, \nu_{t-\tau}), \end{aligned} \tag{2.9}$$

for some integer $\tau \geq 0$, since (2.9) can be written in the augmented regression form

$$\begin{aligned} Y_t &= Z_t\gamma + \lambda(\nu_t, \dots, \nu_{t-\tau}) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{2.10}$$

which is a special case of model (2.8) with $\beta_2(\cdot) = \lambda(\cdot)$, $U_t = (\nu_t, \dots, \nu_{t-\tau})'$ and $e_t = \varepsilon_t - \lambda(\nu_t, \dots, \nu_{t-\tau})$. The estimation limit theory for model (2.10) is established in Theorem 4.4 below.

As we discuss in the following section, the estimation methods and the asymptotic theory are different for the two partial linear models A and B given in equations (2.6) and (2.7). These models have their own particular empirical applications. For example, (2.6) can be used to deal with cases where nonstationary stochastic trends are involved in modelling the relationship among economic variables. On the other hand, (2.7) can be used to model the behavioral relationship between consumption and income while taking account of a potential nonlinear covariate impact that involves the short term interest rate, as in Hall (1978). When the nonlinear function $\beta_1(Z_t)$ in (2.7) is of smaller order than the linear component (for example when β_1 is an integrable function), the model components are effectively orthogonal asymptotically and treatment of the system is similar to that of a linear time series model, as we discuss in Section 4.2.

Sections 3 and 4 discuss identification, estimation, and the associated limit theory with a focus on models (1.1), (2.2), and (2.6)–(2.10), since the treatment of models (2.1) and (2.5) follows in a similar way to that of model (1.1). For other interesting subcases of model (1.1) the respective estimation procedures and asymptotics may also be established in related ways to (1.1) and the details are therefore omitted.

3 Nonparametric and coefficient varying cointegration

We start with some general results on kernel density estimation that are not included in the present literature and involve both nonstationary and stationary data. These results are needed in the development of nonparametric and semiparametric asymptotics for model (1.1) and its various special cases considered above. They will be useful in other applications of kernel methods where stationary and nonstationary data appear.

3.1 Density estimation

Our particular interest, given the nonparametric form of model (1.1), is a multivariate case of kernel density estimation where the component variables may be stationary and nonstationary. The limit theory for this case is presently unknown and is needed here for our regression applications.

We use $K_i(\cdot)$ and h_i ($i = 1, 2$) to denote kernel functions and bandwidth parameters. Typically K_1 is univariate (for the nonstationary component) and K_2 is a multivariate kernel (for the stationary component). Other notation and the assumptions used in the asymptotics here are laid out in section A.1 of Appendix A. Define the kernel density estimator

$$\hat{f}(z, u) = \frac{1}{\sqrt{n}h_1h_2^{d_u}} \sum_{t=1}^n K_1\left(\frac{Z_t - z}{h_1}\right) K_2\left(\frac{U_t - u}{h_2}\right), \quad (3.1)$$

where U_t is a d_u -dimensional stationary time series, and Z_t follows a unit root model of the form $Z_t = Z_{t-1} + \nu_t$. By virtue of Assumption A.1(iv), we have a functional law with limiting vector Brownian motion (B_μ, B_ν, B_u) such that $(\mu_n(r), \nu_n(r), u_n(r)) \Rightarrow_{\mathcal{D}} (B_\mu, B_\nu, B_u)$, where $\mu_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \mu_t$, $\nu_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \nu_t$, $u_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} U_t$, and $\lfloor nr \rfloor$ is the integer part of nr .

Our first result shows convergence of the kernel density estimate (3.1), involving a stationary and nonstationary pair, to the product of the marginal density of the stationary component and the local time of the limit of the (standardized) nonstationary component.

Theorem 3.1. *Let Assumptions A.1(i)(ii)(iv), A.2(ii)(iii), and A.3 hold. Let $X_t \equiv 1$ in (1.1) hold. Then as $n \rightarrow \infty$*

$$\hat{f}(z, u) = \frac{1}{\sqrt{n}h_1h_2^{d_u}} \sum_{t=1}^n K_1\left(\frac{Z_t - z}{h_1}\right) K_2\left(\frac{U_t - u}{h_2}\right) \rightarrow_D p(u) L_{B_\nu}(1, 0), \quad (3.2)$$

where $L_{B_\nu}(t, 0) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^t I[|B_\nu(s)| < \epsilon] ds$ is the local-time process of the Brownian $B_\nu(t)$ and $p(u)$ is the marginal density function of U_t .

Theorem 3.1 extends existing results on kernel density estimation for the nonstationary case (see, for example, Theorem 2.1 of Wang and Phillips 2009a) to the case where correlated nonstationary and stationary regressors are both involved in the same nonparametric form. Correspondingly, the limit involves both the probability density of U_t and the local time of the process B_v . The standardization in (3.2) is $\sqrt{n}h_1h_2^{d_u}$ and this affects the corresponding rate of convergence, in comparison with the stationary case where the usual standardization is $nh_1h_2^{d_u}$. This change arises because the amount of time spent by the time series Z_t around any particular spatial point (such as the origin) is of order \sqrt{n} rather than n . Importantly, even though the covariates Z_t and U_t are dependent, the limit in (3.2) is the product of the probability density $p(u)$ and the spatial local time $L_{B_v}(1, 0)$, reflecting independence in the limit.

3.2 Estimation in model (1.1)

For model (1.1) we propose to estimate the varying coefficient function $\beta(z, u)$ using local level kernel estimation. Since the first order biases are of order $\sqrt{\sqrt{n}h_1h_2^{d_u}} (O(h_1^2) + O(h_2^2))$, they are automatically removable under some mild bandwidth conditions which are listed in Assumption A.2(iii). Hence the use of local linear estimation is not necessary for bias reduction in this kind of integrated time series case where the amount of time spent by the nonstationary series around any particular spatial point is of order \sqrt{n} rather than n . The asymptotic equivalence of local level and local linear kernel estimation in nonstationary nonparametric regression was recently discovered in Wang and Phillips (2009a, 2009b) and is being considerably extended in the present multivariate case involving both stationary and nonstationary regressors.¹ In what follows we therefore propose using standard local level methods to estimate $\beta(z, u)$ at some fixed pair (z, u) .

Accordingly, $\beta(z, u)$ is estimated by

$$\widehat{\beta}(z, u) = \arg \min_{\beta} \sum_{t=1}^n (Y_t - X_t' \beta)^2 K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right), \quad (3.3)$$

giving the estimation error

$$\begin{aligned} \widehat{\beta}(z, u) - \beta(z, u) &= \left(\sum_{t=1}^n X_t X_t' K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \right)^{-1} \\ &\quad \times \left(\sum_{t=1}^n X_t e_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \right), \end{aligned} \quad (3.4)$$

¹In very recent work, Chan and Wang (2013) have discovered that local linear kernel estimation offers some advantages in terms of uniform consistency in nonstationary nonparametric regression.

provided the inverse exists. The asymptotic theory of $\widehat{\beta}(z, u)$ depends on the probabilistic structure of (X_t, Z_t, U_t) . Theorem 3.2 gives the relevant asymptotic theory for a self normalized version of $\widehat{\beta}(z, u)$.

Theorem 3.2. *Let Assumptions A.1–A.3 hold. Then as $n \rightarrow \infty$*

$$\sqrt{n\sqrt{nh_1h_2^{d_u}} \cdot \sqrt{A_{1n}(z, u)}} \left(\widehat{\beta}(z, u) - \beta(z, u) \right) \rightarrow_D N(0, \Sigma_1), \quad (3.5)$$

where $\Sigma_1 = I_{d_x} \cdot \sigma_e^2 C_{22}(K_1, K_2)$ with $C_{22}(K_1, K_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_1^2(x)K_2^2(y)dydx$, $\sigma_e^2 = \mathbb{E}[e_1^2]$, I_{d_x} the d_x -dimensional identity matrix, and

$$\begin{aligned} A_{1n}(z, u) &= \frac{1}{n\sqrt{nh_1h_2^{d_u}}} \sum_{t=1}^n X_t X_t' K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \\ &\rightarrow_D p(u) \int_0^1 B_\mu(r) B_\mu(r)' dL_{B_\nu}(r, 0). \end{aligned} \quad (3.6)$$

The rate of convergence of $\widehat{\beta}(z, u)$ to $\beta(z, u)$ is proportional to $O\left((n\sqrt{nh_1h_2^{d_u}})^{-\frac{1}{2}}\right)$ where the factors $(\sqrt{nh_1})^{\frac{1}{2}}$ and $(nh_2^{d_u})^{\frac{1}{2}}$ reflect the nonstationary and stationary components of $\beta(Z_t, U_t)$. The following corollary provides corresponding results for the multivariate nonparametric model (2.1) which is a special case of (1.1) with $X_t = 1$.

Corollary 3.1. *Let Assumptions A.1(i)–(iv), A.2, A.3 hold. Set $d_x = 1$ and $X_t \equiv 1$. Then as $n \rightarrow \infty$*

$$\sqrt{\sqrt{nh_1h_2^{d_u}} \cdot \sqrt{A_{2n}(z, u)}} \left(\widehat{\beta}(z, u) - \beta(z, u) \right) \rightarrow_D N(0, \Sigma_1), \quad (3.7)$$

where

$$A_{2n}(z, u) = \frac{1}{\sqrt{nh_1h_2^{d_u}}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \rightarrow_D p(u) L_{B_\nu}(1, 0). \quad (3.8)$$

Remark 3.1. (i) Theorem 3.2 shows that the rate of convergence is governed by the degree of nonstationarity involved in X_t , which affects the scaling of the moment matrix component $\sum_{t=1}^n X_t X_t'$ in $A_{1n}(z, u)$, and the nonstationarity of Z_t , which affects the convergence behavior of sums involving the kernel weights in (3.4). The rate of convergence differs by a factor of $O(n)$ from the simpler nonparametric regression case where X_t is stationary or a constant, as given in Corollary 3.1.

(ii) Corollary 3.1 extends results for univariate examples of nonparametric regression for integrated time series that have recently been studied in Karlsen and Tjøstheim (2001), Karlsen, Mykelbust and Tjøstheim (2007), Wang and Phillips (2009a, 2009b), and Gao and Phillips (2013) as well as standard kernel density results for the case where both Z_t and U_t are stationary (see, for example, Chapter 2 of Gao, 2007).

4 Additive nonlinear cointegration

4.1 Nonparametric additive cointegration

We start by considering a nonparametric additive model of the form suggested in model (2.2), viz.,

$$\begin{aligned} Y_t &= \beta_1(Z_t) + \beta_2(U_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{4.1}$$

giving a structure for Y_t that involves a nonstationary regressor, Z_t , and a stationary regressor, U_t . As mentioned earlier, this type of model seems appropriate in studying relationships among several economic variables whose stochastic orders may differ and where there may be determining factors of different orders in the nonparametric functions, some stationary and others nonstationary. Since nonlinear functions of an integrated regressor can change the order of integration, relationships such as (4.1) may be well-balanced in terms of stochastic order. Because this model mixes both stationary and nonstationary regressors, existing studies in the field of additive nonparametric modelling (such as Chapter 2 of Gao 2007) are not applicable.

Further specification of model (4.1) involves the following identification condition.

Assumption 4.1. *The function $m(z, u) \equiv \beta_1(z) + \beta_2(u)$ is uniquely defined. Both $\beta_1(z)$ and $\beta_2(u)$ are nonparametrically unknown functions and estimable up to an additive constant.*

We next discuss the consistent estimation of $\beta_1(z)$ and $\beta_2(u)$ using a marginal integration approach. Let $R_i(\cdot)$ be some known, nonnegative deterministic functions such that $\int_{-\infty}^{\infty} dR_1(u) = 1$ and $\int_{-\infty}^{\infty} dR_2(z) = 1$. Let $c_1(\beta) = \int \beta_2(u) dR_1(u)$ and $c_2(\beta) = \int \beta_1(z) dR_2(z)$. Before the choice of $R_i(\cdot)$ and the estimation of $c_i(\beta)$ are discussed, define the quantities

$$\bar{\beta}_1(z) = \int_{-\infty}^{\infty} (\beta_1(z) + \beta_2(u)) dR_1(u) = \beta_1(z) + c_1(\beta), \tag{4.2}$$

$$\bar{\beta}_2(u) = \int_{-\infty}^{\infty} (\beta_1(z) + \beta_2(u)) dR_2(z) = \beta_2(u) + c_2(\beta). \tag{4.3}$$

If $\bar{\beta}_1(z)$ and $\bar{\beta}_2(u)$ are estimated, then both $\beta_1(z)$ and $\beta_2(u)$ may be estimated up to an additive constant.

The systematic component $m(z, u) = \beta_1(z) + \beta_2(u)$ may be estimated by

$$\hat{m}(z, u) = \sum_{s=1}^n w_{ns}(z, u) Y_s, \quad \text{with } w_{ns}(z, u) = \frac{K_1\left(\frac{Z_s - z}{h_1}\right) K_2\left(\frac{U_s - u}{h_2}\right)}{\sum_{l=1}^n K_1\left(\frac{Z_l - z}{h_1}\right) K_2\left(\frac{U_l - u}{h_2}\right)} \tag{4.4}$$

in which $K_i(\cdot)$ and h_i are as in (3.3). We then estimate $\beta_1(z)$ and $\beta_2(u)$ by

$$\widehat{\beta}_1(z) = \int_{-\infty}^{\infty} \widehat{m}(z, u) dR_1(u) \quad \text{and} \quad \widehat{\beta}_2(u) = \int_{-\infty}^{\infty} \widehat{m}(z, u) dR_2(z). \quad (4.5)$$

The following asymptotic results for the estimation of (4.1) are established for the case where U_t is a stationary time series.

Theorem 4.1. *Let Assumptions 4.1, A.1(i)–(iv), A.2(ii)(iii), A.3 and A.4 hold. Let $d_x = 2$ and $X_t \equiv 1$ hold in (1.1) and the component functions of the model (4.1) be estimated as in (4.5).*

(i) *If, in addition, $\sqrt{n} h_1^3 = O(1)$ and $\sqrt{n} h_1 h_2^{2d_u} = O(1)$, then as $n \rightarrow \infty$*

$$B_{1n}^{-1}(z) \left(\widehat{\beta}_1(z) - \beta_1(z) - c_1(\beta) \right) \rightarrow_D N(0, \sigma_e^2), \quad (4.6)$$

where $B_{1n}^2(z) = \sum_{s=1}^n B_{1ns}^2(z)$ with $B_{1ns}(z) = \int_{-\infty}^{\infty} w_{ns}(z, u) dR_1(u)$. The asymptotic form of $B_{1n}^2(z)$ is given by

$$B_{1n}^2(z) = \frac{1}{\sqrt{n} h_1} \frac{C_1(K)}{L_B(1, 0)} \pi_1 (1 + o_D(1)), \quad (4.7)$$

where $C_1(K) = \int K_1^2(x) dx$ and $\pi_1 = \int \frac{r_1^2(u)}{p(u)} du$ is the same as in Assumption A.4(ii).

(ii) *If, in addition, $\sqrt{n} h_2^{3d_u} = O(1)$ and $\sqrt{n} h_2^{d_u} h_1^2 = O(1)$, then as $n \rightarrow \infty$*

$$B_{2n}^{-1}(u) \left(\widehat{\beta}_2(u) - \beta_2(u) - c_2(\beta) \right) \rightarrow_D N(0, \sigma_e^2), \quad (4.8)$$

where $B_{2n}^2(u) = \sum_{s=1}^n B_{2ns}^2(u)$ with $B_{2ns}(u) = \int_{-\infty}^{\infty} w_{ns}(z, u) dR_2(z)$. The asymptotic form of $B_{2n}^2(u)$ is given by

$$B_{2n}^2(u) = \frac{1}{\sqrt{n} h_2^{d_u}} \frac{C_2(K)}{L_B(1, 0)} \pi_2 (1 + o_P(1)), \quad (4.9)$$

where $C_2(K) = \int K_2^2(x) dx$ and $\pi_2 = \int r_2^2(z) dz$ is the same as in Assumption A.4(ii).

Corollary 4.1. (i) *Let the conditions of Theorem 4.1(i) hold. If, in addition, $c_1(\beta) = 0$, then as $n \rightarrow \infty$*

$$B_{1n}^{-1}(z) \left(\widehat{\beta}_1(z) - \beta_1(z) \right) \rightarrow_D N(0, \sigma_e^2). \quad (4.10)$$

(ii) *Let the conditions of Theorem 4.1(ii) hold. If, in addition, $c_2(\beta) = 0$, then as $n \rightarrow \infty$*

$$B_{2n}^{-1}(u) \left(\widehat{\beta}_2(u) - \beta_2(u) \right) \rightarrow_D N(0, \sigma_e^2). \quad (4.11)$$

Model (2.4) takes a special case of (4.1) of the form

$$\begin{aligned} Y_t &= \beta_1(Z_t) + \beta_2(U_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t \quad \text{with } U_t = \nu_t, \end{aligned} \quad (4.12)$$

where there is high correlation between the nonstationary regressor Z_t and stationary regressors $U_t = Z_t - Z_{t-1}$ that is induced by the specification $U_t = \nu_t$. Corollary 4.2 below shows that both functions $\beta_1(\cdot)$ and $\beta_2(\cdot)$ may still be consistently estimated. As discussed in the proof given in Appendix B, this is due to the fact that the standardized variate $\frac{1}{\sqrt{t}}Z_t$ and U_t are asymptotically uncorrelated when $t \rightarrow \infty$.

Corollary 4.2. *Let the conditions of Theorem 4.1 hold. Then the conclusions of Theorem 4.1 remain true for model (4.12).*

Remark 4.1. Theorem 4.1 and Corollaries 4.1–4.2 show that the functions $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are consistently estimable. In addition, Theorem 4.1(ii) shows that the convergence rate of $\widehat{\beta}_2(\cdot)$ is $\left(\sqrt{\sqrt{nh_2^{d_u}}}\right)^{-1}$, which is slower than the conventional rate $\left(\sqrt{nh_2^{d_u}}\right)^{-1}$ for the stationary case. As discussed following Theorem 3.1, this difference is due to the nonstationarity of Z_t which impacts the convergence rate of the joint nonparametric estimator (4.4) and thereby that of the estimator of $\beta_2(\cdot)$ even though the associated regressor U_t in this function is itself stationary.

4.2 Semiparametric additive cointegration

We now focus on models (2.6) and (2.7) and discuss the identifiability and estimability of the unknown parameters and functions in these semiparametric models. To open discussion and present the motivating ideas we consider a semiparametric time series model of the partial linear form

$$Y_t = V_t' \beta + \beta_1(Z_t) + e_t, \quad (4.13)$$

where $V_t = X_{1t}$ in model (2.6) and $V_t = Z_t$ in model (2.7).

To begin, suppose for the moment that V_t and Z_t are both stationary. Accordingly, let $H(z) = \mathbb{E}[V_t | Z_t = z]$. In the case where $H(z)$ is a smooth function and the covariance matrix $\Sigma = \mathbb{E}[(V_t - \mathbb{E}[V_t | Z_t])(V_t - \mathbb{E}[V_t | Z_t])']$ is positive definite, both β and $\beta_1(\cdot)$ can be semiparametrically estimated as in the existing literature by projecting out the nonparametrically fitted conditional expectations (see, for example, Robinson 1988; Härdle, Liang and Gao 2000; Gao 2007; Li and Racine 2007; Teräsvirta, Tjøstheim and Granger 2010). In the case where V_t and Z_t are highly correlated as they may be in models (2.6) and (2.7) with nonstationarity present, this conventional semiparametric approach can break down asymptotically because of potential singularity in the asymptotic moment matrix. We therefore examine a direct approach to fitting (4.13) that takes into account the effects of nonstationarity on the model components.

Suppose we treat β as the parameter of interest in (4.13) and $\beta_1(\cdot)$ as a nuisance parameter so that the nonparametric term is absorbed into the disturbance as random noise. If V_t and Z_t were both stationary and the model itself were a linear model, then minimizing $\mathbb{E}([Y_t - V_t' b]^2)$ leads to the regression coefficient $b = (\mathbb{E}[V_1 V_1'])^{-1} \mathbb{E}[V_1 Y_1]$ which equals β under

$$\mathbb{E}[V_1 \beta_1(Z_1)] = 0, \text{ and } \mathbb{E}[V_1 e_1] = 0 \quad (4.14)$$

the first part of which implies under ergodicity $\frac{1}{n} \sum_{t=1}^n V_t \beta_2(Z_t) \rightarrow_P 0$. So (4.14) is an orthogonality condition which ensures that β is identifiable and estimable in the stationary case. These considerations motivate the use of the orthogonality condition:

$$\frac{1}{n} \sum_{t=1}^n V_t \beta_1(Z_t) = O_P(1) \quad (4.15)$$

for the identification and estimability of β in the nonstationary case. To derive a more explicit condition that reduces to (4.15), we consider the explicit case where $X_t = X_{t-1} + \mu_t$ and $Z_t = Z_{t-1} + \nu_t$. Without loss of generality, we here assume that X_{1t} follows the same model as X_t . Let $d_n = \sqrt{n}$, $x_{tn} = \frac{X_t}{\sqrt{n}}$ and $z_{tn} = \frac{Z_t}{\sqrt{n}}$. Then, analogous to the proof of Lemma A.5 in Appendix A, when $d_z = 1$ we have the following limit behaviour as $n \rightarrow \infty$

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n X_t \beta_1(Z_t) &= \frac{d_n}{n} \sum_{t=1}^n (x_{tn}) \beta_1(d_n z_{tn}) \rightarrow_D \int_0^1 B_\mu(r) dL_{B_\nu}(r, 0) \cdot \int_{-\infty}^{\infty} \beta_1(z) dz, \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t \beta_1(Z_t) &= \frac{d_n}{n} \sum_{t=1}^n (d_n z_{tn}) \beta_1(d_n z_{tn}) \rightarrow_D L_{B_\nu}(1, 0) \cdot \int_{-\infty}^{\infty} z \beta_1(z) dz, \end{aligned} \quad (4.16)$$

where the limits exist under the integrability conditions of Assumption 4.2 below. The orthogonality conditions now follow directly from (4.16). As $n \rightarrow \infty$, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n X_t \beta_1(Z_t) &= \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t \beta_1(Z_t) = O_P(1), \\ \frac{1}{n} \sum_{t=1}^n Z_t \beta_1(Z_t) &= \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t \beta_1(Z_t) = o_P(1). \end{aligned} \quad (4.17)$$

Assumption 4.2. (i) Let $\beta_1(z)$ in model (2.6) be continuous in z and both $\beta_1(z)$ and $\beta_1^2(z)$ be integrable. In addition, $\int \beta_1(z) dz \neq 0$.

(ii) Let $\beta_1(z)$ in model (2.7) be continuous in z and both $z\beta_1(z)$ and $z^2\beta_1^2(z)$ be integrable. In addition, $\int z\beta_1(z) dz \neq 0$.

Under Assumption 4.2, both β in model (2.6) and θ in model (2.7) are uniquely identifiable and consistently estimable by

$$\hat{\beta} = \left(\sum_{t=1}^n X_{1t} X_{1t}' \right)^{-1} \sum_{t=1}^n X_{1t} Y_t \text{ and } \hat{\theta} = \left(\sum_{t=1}^n Z_t Z_t' \right)^{-1} \sum_{t=1}^n Z_t Y_t. \quad (4.18)$$

The nonparametric component $\beta_1(z)$ in these two models may respectively be estimated by

$$\begin{aligned}\widehat{\beta}_1(z) &= \sum_{t=1}^n w_{nt}(z) \left(Y_t - X'_{1t} \widehat{\beta} \right) \quad \text{for model (2.6),} \\ \widetilde{\beta}_1(z) &= \sum_{t=1}^n w_{nt}(z) \left(Y_t - Z'_t \widehat{\theta} \right) \quad \text{for model (2.7),}\end{aligned}\tag{4.19}$$

where $w_{nt}(z) = K_1\left(\frac{Z_t - z}{h_1}\right) / \sum_{s=1}^n K_1\left(\frac{Z_s - z}{h_1}\right)$, in which K_1 and h_1 are assumed to satisfy the following conditions.

Assumption 4.3. (i) *There exists a real function $L(u, v)$ such that*

$$|\beta_1(v + h_1 u) - \beta_1(v)| \leq hL(u, v)$$

for all $u \in \mathbb{R} = (-\infty, \infty)$ and $\int K_1(u)L(u, v)du < \infty$ for each given v .

(ii) *The probability kernel function $K_1(u)$ satisfies $\int uK_1(u)du = 0$, $\int u^2K_1(u)du < \infty$ and $\int K_1^2(u)du < \infty$.*

(iii) *The bandwidth h_1 satisfies $h_1 \rightarrow 0$, $\sqrt{nh_1} \rightarrow \infty$ and $\sqrt{nh_1^3} \rightarrow 0$.*

We are now ready to establish the following main results on the limit theory for the estimates of the parametric and nonparametric components in models (2.6) and (2.7). Proofs are in Appendix B.

Theorem 4.2. (i) *Let Assumptions 4.2(i) and A.1(i)(ii)(iii) hold. Then, as $n \rightarrow \infty$*

$$\begin{aligned}n \left(\widehat{\beta} - \beta \right) &\rightarrow_D \left(\int_0^1 B_\mu(r)B_\mu(r)'dr \right)^{-1} \times \\ &\left(\int_0^1 B_\mu(r)dB_e(r) + \int_0^1 B_\mu(r)dL_{B_\nu}(r, 0) \cdot \int_{-\infty}^{\infty} \beta_1(z)dz \right),\end{aligned}\tag{4.20}$$

where $B_e(r)$ is the Brownian motion limit of standardized partial sums $(n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} e_t)$ of e_t .

(ii) *If, in addition, Assumption 4.3 holds, then as $n \rightarrow \infty$*

$$\sqrt{\sum_{t=1}^n K_1\left(\frac{Z_t - z}{h_1}\right)} \left(\widehat{\beta}_1(z) - \beta_1(z) \right) \rightarrow_D N\left(0, \sigma_1^2(K_1)\right),\tag{4.21}$$

where $\sigma_1^2(K_1) = \sigma_e^2 \cdot \int K_1^2(u)du$.

Theorem 4.3. (i) *Let Assumptions 4.2(ii) and A.1(i)(ii)(iii) hold. Then, under $X_t \equiv 1$, we have as $n \rightarrow \infty$*

$$n \left(\widehat{\theta} - \theta \right) \rightarrow_D \left(\int_0^1 B_\nu(r)B_\nu(r)'dr \right)^{-1} \int_0^1 B_\nu(r)dB_e(r).\tag{4.22}$$

(ii) If, in addition, Assumption 4.3 holds, then we have as $n \rightarrow \infty$

$$\sqrt{\sum_{t=1}^n K_1\left(\frac{Z_t - z}{h_1}\right)} \left(\tilde{\beta}_1(z) - \beta_1(z)\right) \rightarrow_D N(0, \sigma_1^2(K_1)). \quad (4.23)$$

Theorem 4.2(i) shows that there is a kind of bias term involved in (4.20) when $\int \beta_1(z)dz \neq 0$. By contrast, Theorem 4.3(i) shows that such a bias term disappears in the case of $X_t = Z_t$ even when $\int z\beta_1(z)dz \neq 0$. This is basically because X_t and Z_t involved in the parametric and nonparametric components are generated by two different integrated time series in model (2.6), while model (2.7) involves only one integrated time series in both the parametric and nonparametric components. Note that the bias term involved in (4.20) disappears when $\int \beta_1(z)dz = 0$. Note also that Theorem 4.3(i) however remains the same when $\int z\beta_1(z)dz = 0$. We omit the discussion for the cases of $\int \beta_1(z)dz = 0$ and $\int z\beta_1(z)dz = 0$, since the detail is very similar.

We next consider augmented regression models of the type shown in (2.10) which involve a parametric nonstationary component combined with a nonparametric function of stationary variates, written in the form

$$\begin{aligned} Y_t &= Z_t\gamma + \beta_2(U_t) + e_t, \\ Z_t &= Z_{t-1} + \nu_t. \end{aligned} \quad (4.24)$$

The natural approach to (4.24), given the stationary nonparametric component $\beta_2(U_t)$, is to estimate γ and $\beta_2(\cdot)$ by semiparametric weighted least squares. However, since Z_t is nonstationary and is correlated with U_t , existing approaches and limit theory (e.g., Robinson 1988; Härdle, Liang and Gao 2000) are not directly applicable.

Let $\alpha = \mathbb{E}[\beta_2(U_t)]$ and $\eta_t = \beta_2(U_t) - \alpha + e_t$, so that (4.24) may be written as

$$\begin{aligned} Y_t &= \alpha + Z_t\gamma + \eta_t, \\ Z_t &= Z_{t-1} + \nu_t. \end{aligned} \quad (4.25)$$

This formulation suggests that in (4.25) the parameter γ can be estimated by least squares, giving

$$\hat{\gamma} = \left(\sum_{t=1}^n \tilde{Z}_t^2\right)^{-1} \sum_{t=1}^n \tilde{Z}_t \tilde{Y}_t, \quad (4.26)$$

where $\tilde{Z}_t = Z_t - n^{-1} \sum_{s=1}^n Z_s$ and $\tilde{Y}_t = Y_t - n^{-1} \sum_{s=1}^n Y_s$. The unknown function $\beta_2(u)$ is then estimated by kernel regression on the residuals, leading to

$$\hat{\beta}_2(u) = \sum_{t=1}^n W_{nt}(u) (Y_t - Z_t \hat{\gamma}), \quad (4.27)$$

where $W_{nt}(u) = K_2\left(\frac{U_t-u}{h_2}\right) / \sum_{s=1}^n K_2\left(\frac{U_s-u}{h_2}\right)$.

The limit theory for the estimates $(\widehat{\gamma}, \widehat{\beta}_2(u))$ obtained in this way is given in Theorem 4.4 below, for which we use the following conditions.

Assumption 4.4. (i) Suppose that $\beta_2(u)$ is twice differentiable and that the second derivative, $\beta_2^{(2)}(u)$, is continuous with $\int_{-\infty}^{\infty} |\beta_2^{(2)}(u)| p(u) < \infty$, where $p(u)$ is the marginal density of U_t .

(ii) $K_2(u)$ is a probability kernel with $\int u K_2(u) du = 0$, $\int u^2 K_2(u) du < \infty$ and $\int K_2^2(u) du < \infty$.

(iii) The bandwidth h_2 satisfies $h_2 \rightarrow 0$, $nh_2^{d_u} \rightarrow \infty$ and $\sqrt{nh_2^4} = O(1)$.

Theorem 4.4. (i) Let Assumptions A.1(i)(ii)(iii) hold. Then, for model (4.25), we have as $n \rightarrow \infty$

$$n (\widehat{\gamma} - \gamma) \rightarrow_D \left(\int_0^1 \tilde{B}_\nu(r)^2 \right)^{-1} \int_0^1 \tilde{B}_\nu(r) dB_\eta(r), \quad (4.28)$$

where $B_\eta(r)$ is defined as the weak limit of $\Pi_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \eta_t \Rightarrow_D B_\eta(r)$ on the Skorohod space $D[0, 1]$ and $\tilde{B}_\nu(r) = B_\nu(r) - \int_0^1 B_\nu(s) ds$.

(ii) If Assumption 4.4 also holds, then as $n \rightarrow \infty$

$$\sqrt{nh_2^{d_u}} \left(\widehat{\beta}_2(u) - \beta_2(u) \right) \rightarrow_D N(0, \sigma_2^2(K_2)), \quad (4.29)$$

where $\sigma_2^2(K_2) = \sigma_e^2 \cdot \int K_2^2(u) du$.

Remark 4.2. Theorems 4.2 and 4.3 show that the usual order n convergence rate of linear cointegration is achievable in the semiparametric case as long as the identification condition in Assumption 4.2 holds and direct least squares regression in (4.25) is used. Other methods, such as fully modified least squares (Phillips and Hansen, 1990) may be employed in this model to remove second order bias effects (arising from serial dependence in $\eta_t = \beta_2(U_t) - \alpha + e_t$ and endogeneity in Z_t) in the usual way and to conduct pivotal inference on γ . These methods are standard and do not need to be detailed here. However, as shown in Example 5.3 below, the more conventional semiparametric least squares estimator of γ in (4.24) can have a much slower rate of convergence than the order n rate in some nonstationary models when Assumption 4.2(ii) does not hold.

5 Examples of implementation

This section provides three examples of implementation with specific parameter settings. The first focuses on a varying-coefficient model. The second considers some particular cases

of a nonparametric additive model. Some versions of semiparametric cointegrating models are discussed in the third example.

Example 5.1. Consider a varying-coefficient model of the form

$$\begin{aligned} Y_t &= X_t \beta(Z_t, U_t) + e_t, \quad t = 1, 2, \dots, n, \\ X_t &= X_{t-1} + \mu_t \quad \text{with} \quad \mu_t = 0.5\mu_{t-1} + \varepsilon_t, \\ Z_t &= Z_{t-1} + \nu_t \quad \text{with} \quad \nu_t = -0.5\nu_{t-1} + \varepsilon_t, \\ \beta(z, u) &= \frac{z}{(1+z^2)^{\frac{3}{2}}} + u, \end{aligned} \tag{5.1}$$

where $X_0 = Z_0 = 0$, $\mu_0 = \nu_0 = 0$, $e_t = \varepsilon_{t+1}$ and $\{\varepsilon_t\} \stackrel{iid}{\sim} N[0, 1]$. We take the following two cases:

$$\text{Case A : } U_t = \nu_t \quad \text{Case B : } U_t \stackrel{iid}{\sim} U[0, 1],$$

where $U[0, 1]$ denotes the uniform distribution on $[0, 1]$.

The function $\beta(\cdot, \cdot)$ is estimated by local level regression as

$$\begin{aligned} \tilde{\beta}(z, u; h_1, h_2) &= \left(\sum_{t=1}^n X_t^2 K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \right)^{-1} \\ &\times \left(\sum_{t=1}^n X_t Y_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \right), \end{aligned} \tag{5.2}$$

for certain bandwidth choices (h_1, h_2) . To select (h_1, h_2) we first introduce the leave-one-out estimator of $\beta(z, u)$ which we write in the form $\tilde{\beta}_{(-t)}(z, u)$. We then define the leave-one-out cross-validation (CV) function as

$$\text{CV}(h_1, h_2) = \frac{1}{n} \sum_{t=1}^n \left(Y_t - X_t \tilde{\beta}_{(-t)}(Z_t, U_t) \right)^2. \tag{5.3}$$

An optimal bandwidth value for (h_1, h_2) is chosen according to the rule

$$\text{CV}(\hat{h}_1, \hat{h}_2) = \min_{h_1, h_2 \in \mathcal{H}_{2n}} \text{CV}(h_1, h_2), \tag{5.4}$$

where \mathcal{H}_{2n} is of the form $[n^{-1}, n^{-1+c_1}] \times [n^{-1}, n^{-1+c_2}]$, in which each $0 < c_i < 1$ is chosen such that each of (\hat{h}_1, \hat{h}_2) is achievable and locally unique. With this bandwidth choice, the function $\beta(\cdot, \cdot)$ is estimated by

$$\hat{\beta}(z, u) = \tilde{\beta}(z, u; \hat{h}_1, \hat{h}_2). \tag{5.5}$$

We perform a small simulation experiment to evaluate the performance characteristics of $\hat{\beta}(z, u)$ for sample sizes $n = 201, 551, 901$, and 1501 . Finite sample performance is assessed

in terms of average absolute bias

$$\text{MSE}_1 = \frac{1}{n} \sum_{t=1}^n \left| \widehat{\beta}(Z_t, U_t) - \beta(Z_t, U_t) \right|,$$

and cross validation via (5.3) is used for the bandwidth choice. The simulation results are based on $N = 5000$ replications and are tabulated in Table 5.1 below.

Table 5.1: Average Absolute Biases

n	Case A	Case B
$n = 201$	0.6213	0.4776
$n = 501$	0.3420	0.4617
$n = 901$	0.3108	0.2654
$n = 1501$	0.2014	0.2241

Table 5.1 shows that absolute biases decrease slowly as the sample size increases. The slow diminution of the bias conforms with asymptotic theory in the nonparametric nonstationary case.

Example 5.2. We consider a nonparametric additive model of the form

$$Y_t = \beta_1(Z_t) + \beta_2(U_t) + e_t, \quad (5.6)$$

$$Z_t = Z_{t-1} + \nu_t, \quad (5.7)$$

where $Z_0 = 0$, and $\{(Y_t, Z_t, U_t)\}$ is generated by one of the following models with $\beta_1(z) = \frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $\beta_2(u) = u^2$, and the respective error mechanisms

$$\text{Case C: } \nu_t = 0.5 \nu_{t-1} + \varepsilon_t \quad \text{with} \quad U_t = \varepsilon_{t-1} \quad \text{and} \quad e_t = \varepsilon_{t+1},$$

$$\text{Case D: } \nu_t = 0.5 \nu_{t-1} + \varepsilon_t \quad \text{with} \quad U_t = \nu_t \quad \text{and} \quad e_t = \varepsilon_{t+1},$$

$$\text{Case E: } \nu_t = \varepsilon_t \quad \text{with} \quad U_t = \varepsilon_t \quad \text{and} \quad e_t = \varepsilon_{t+1}.$$

where $\nu_0 = 0$ and $\{\varepsilon_t\} \stackrel{iid}{\sim} N(0, 1)$. Cases C–E cover some common simultaneous systems with strong endogeneity where a nonstationary regressor (Z_t), a stationary regressor (U_t), and the equation error e_t are all highly correlated and driven by ε_t .

The nonparametric systematic component of (5.6) is $m(z, u) = \beta_1(z) + \beta_2(u)$, which is estimated by $\widehat{m}(z, u) = \widehat{\beta}_1(z) + \widehat{\beta}_2(u)$. The finite sample performance of $\widehat{m}(z, u)$ is assessed by mean squared error using

$$\text{MSE}_2 = \sqrt{\frac{1}{n} \sum_{t=1}^n (\widehat{m}(Z_t, U_t) - m(Z_t, U_t))^2}.$$

Bandwidth selection is performed by cross validation as in (5.3) and the simulation results reported in Table 5.2 below are based on $N = 5000$ replications.

Table 5.2: Average Mean Squared Error

n	Case C	Case D	Case E	
$n = 201$	0.0175	0.0312	0.0169	
$n = 501$	0.0162	0.0284	0.0148	
$n = 901$	0.0148	0.0275	0.0137	

Table 5.2 shows that the proposed estimator performs adequately even in the presence of strong endogeneity among both stationary and nonstationary variables in this model. Again, the mean squared error of the nonparametric estimator declines slowly with n .

Example 5.3. We consider a regression model that involves a nonstationary regressor and takes the partially linear form

$$Y_t = Z_t \theta + \beta_1(Z_t) + e_t, \quad (5.8)$$

$$Z_t = Z_{t-1} + \nu_t, \quad (5.9)$$

where the initialization $Z_0 = 0$ and various functional forms for $\beta_1(\cdot)$ and probabilistic structures for $\{\nu_t\}$ are explored. This type of model might be used as a partially linear predictive regression model when Z_t is adapted to the past. We examine two approaches to estimating model (5.8). The first involves direct estimation and subsequent elimination of the linear cointegrating component. The second follows the conventional semiparametric approach of first eliminating the functional component by nonparametric regression. It will become apparent that these approaches have very different properties in a model such as (5.8) and (5.9) where the regressor is nonstationary.

The first approach proceeds on the presumption of a strong signal in the regressor Z_t from the stochastic trend (5.9) and the presence of a nonlinear integrable function $\beta_1 \in L_1$ that attenuates the effects of large Z_t . The coefficient θ of the linear cointegrating term in (5.8) may then be estimated directly by least squares as

$$\hat{\theta} = (Z'Z)^{-1}Z'Y, \quad (5.10)$$

where $Z' = (Z_1, \dots, Z_n)$ and $Y' = (Y_1, \dots, Y_n)$. The unknown function β_1 can subsequently be estimated using nonparametric regression on the parametric residuals to capture any

potential nonlinear effects as follows

$$\bar{\beta}_1(z, h_3) = \sum_{s=1}^n W_{ns}(z; h_3) \left(Y_s - Z_s \hat{\theta}_n \right), \quad (5.11)$$

where $W_{ns}(z; h_3) = K_1 \left(\frac{Z_s - z}{h_3} \right) / \sum_{l=1}^n K_1 \left(\frac{Z_l - z}{h_3} \right)$. The bandwidth h_3 is selected by minimizing the cross validation criterion

$$\text{CV}(h_3) = \frac{1}{n} \sum_{t=1}^n \left(Y_t - \hat{\theta}_n Z_t - \bar{\beta}_{1,(-t)}(Z_t; h_3) \right)^2, \quad (5.12)$$

with $\bar{\beta}_{1,(-t)}(Z_t, h_3) = \sum_{s=1, \neq t}^n K_1 \left(\frac{Z_s - Z_t}{h_3} \right) \left(Y_s - Z_s \hat{\theta} \right) / \sum_{k=1, \neq t}^n K_1 \left(\frac{Z_k - Z_t}{h_3} \right)$ such that

$$\text{CV}(\tilde{h}_3) = \min_{h \in \mathcal{H}_{3n}} \text{CV}(h_3), \quad (5.13)$$

in which \mathcal{H}_{3n} has the form $[n^{-1}, n^{-1+c_3}]$ with each $0 < c_3 < 1$ chosen so \tilde{h}_3 is achievable and locally unique. This process leads to the following estimate of β_1 as

$$\hat{\beta}_1(z) = \bar{\beta}_1(z; \tilde{h}_3). \quad (5.14)$$

The second approach follows conventional semiparametric practice of eliminating the nonparametric component and then proceeding with a direct regression for the linear component using the nonparametric residuals. Accordingly, we define the following leave-one-out nonparametric estimators

$$\tilde{\Psi}_t(Z_t) = \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(Z_t) Y_s \quad \text{and} \quad \tilde{\Gamma}_t(Z_t) = \sum_{s=1, \neq t}^n W_{ns}^{(-t)}(Z_t) X_s, \quad (5.15)$$

where $W_{ns}^{(-t)}(Z_t) = K_1 \left(\frac{Z_s - Z_t}{h_4} \right) / \sum_{k=1, \neq t}^n K_1 \left(\frac{Z_k - Z_t}{h_4} \right)$ and h_4 is a bandwidth parameter. Let the residuals from these regressions be $\tilde{Y}_t = \left(Y_t - \tilde{\Psi}(Z_t) \right)$ and $\tilde{Z}_t = \left(Z_t - \tilde{\Gamma}(Z_t) \right)$. We then consider the following approximate system that is induced after (semiparametric) elimination of the function β_1

$$\tilde{Y}_t = \theta \tilde{Z}_t + \text{error}. \quad (5.16)$$

The leave-one-out semiparametric least squares (SLS) estimator of θ is obtained by linear regression on this approximate system leading to

$$\tilde{\theta}(h_4) = (\tilde{Z}' \tilde{Z})^{-1} \tilde{Z}' \tilde{Y}, \quad (5.17)$$

where $\tilde{Z}' = (\tilde{Z}_1, \dots, \tilde{Z}_n)$ and $\tilde{Y}' = (\tilde{Y}_1, \dots, \tilde{Y}_n)$ for some bandwidth h_4 . Next we define the leave-one-out cross-validation (CV) function as

$$\text{CV}(h_4) = \frac{1}{n} \sum_{t=1}^n \left(\tilde{Y}_t - \tilde{\theta} \tilde{Z}_t \right)^2, \quad (5.18)$$

and an optimal bandwidth value for h_4 is chosen such that

$$\text{CV}(\tilde{h}_4) = \min_{h \in \mathcal{H}_{4n}} \text{CV}(h_4), \quad (5.19)$$

where \mathcal{H}_{4n} is of the form $[n^{-1}, n^{-1+c_4}]$, in which each $0 < c_4 < 1$ is chosen such that each of \tilde{h}_4 is achievable and locally unique. The parameter θ is then estimated by $\tilde{\theta} \equiv \tilde{\theta}(\tilde{h}_4)$. Estimation of β_1 follows by local level regression as

$$\tilde{\beta}_1(z) = \sum_{s=1}^n W_{ns}(z; \tilde{h}_4) \left(Y_s - Z_s \tilde{\theta} \right), \quad (5.20)$$

where $W_{ns}(z; h_4) = K_1 \left(\frac{Z_s - z}{h_4} \right) / \sum_{k=1}^n K_1 \left(\frac{Z_k - z}{h_4} \right)$.

We perform a small simulation exercise to assess the finite-sample performance of $\hat{\theta}$ and $\tilde{\theta}$. The relevant asymptotics for $\hat{\theta}$ are given in Theorem 4.3, where it is shown that $\hat{\theta}$ is asymptotically mixed normal with convergence rate n . By contrast, $\tilde{\theta}$ appears to have a slow rate of convergence, as evidenced by the large variances and mean squared errors reported in the simulations below. A rigorous asymptotic treatment of this case presents substantial challenges, and only the following heuristic analysis is attempted here.

The intuition is worth describing. In the first approach, the direct linear cointegrating regression (5.10) preserves the signal strength of the unit root process Z_t , thereby producing an $O(n)$ convergence rate because the effect of the misspecification from ignoring the nonlinear component is negligible when $\beta_1 \in L_1$ since $n^{-1} \sum_{k=1}^n Z_t \beta_1(Z_t) = o_p(1)$ under very general assumptions (see Phillips, 2009). In the second approach, the dependent variable Y_t and linear regressor Z_t are both effectively ‘detrended’ using nonparametric leave-one-out regression (5.15), producing residuals \tilde{Y}_t and \tilde{Z}_t . The effect of this detrending on \tilde{Z}_t is analogous to a nonparametric autoregression of Z_t on Z_{t-1} , which is consistent (Wang and Phillips, 2009a) and, in the present case, fits the trajectory of Z_t (as described in Phillips, 2009), and whose residuals therefore behave more like an $I(0)$ variate than an $I(1)$ variate. The resulting second stage estimator $\tilde{\theta}(h_4)$ of θ suffers from this semiparametric adjustment by using a regressor with a diminished signal, leading to an estimator with a slower convergence rate or possibly inconsistency. Hence, in the case of nonstationary semiparametric regression, conventional semiparametric estimation performs a preliminary nonparametric regressions on the dependent variable and regressor which acts as a form of stochastic detrending of those variables that reduces their signal strength. The secondary regression applies to the residuals from this first stage nonparametric regressions and therefore suffers the consequences of the reduced signal strength, thereby affecting the convergence of the estimates of the linear component.

Tables 5.3 and 5.4 below provide simulation findings that support this distinction between the two estimators. The simulation uses the following generating mechanisms involving two different L_1 functions β_1 :

$$\text{Case F : } \nu_t = 0.5 \nu_{t-1} + \varepsilon_t \quad \text{with} \quad e_t = \varepsilon_{t+1} \quad \text{and} \quad \beta_1(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$

$$\text{Case G : } \nu_t = 0.5 \nu_{t-1} + \varepsilon_t \quad \text{with} \quad e_t = \varepsilon_{t+1} \quad \text{and} \quad \beta_1(z) = \frac{1}{(1+z^2)^{\frac{3}{2}}},$$

where $\nu_0 = 0$ and $\{\varepsilon_t\} \stackrel{iid}{\sim} N[0, 1]$. We assess finite sample performance in terms of bias, standard deviation and mean square error computed as

$$\text{ABS}(\hat{\mu}) = \frac{1}{N} \sum_{j=1}^N |\hat{\mu}(j) - \theta| \quad \text{and} \quad \text{STD}(\hat{\mu}) = \sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{\mu}(j) - \bar{\mu})^2},$$

$$\text{MSE}_1(\hat{\beta}) = \frac{1}{n} \sum_{t=1}^n |\hat{\beta}(Z_t) - \beta(Z_t)| \quad \text{and} \quad \text{MSE}_2(\hat{\beta}) = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{\beta}(Z_t) - \beta(Z_t))^2},$$

where $\hat{\mu} = \hat{\theta}$ or $\tilde{\theta}$, $\hat{\beta}(\cdot) = \hat{\beta}_1(\cdot)$ or $\tilde{\beta}_1(\cdot)$, $\bar{\mu} = \frac{1}{N} \sum_{j=1}^N \hat{\mu}(j)$, and $\hat{\mu}(j)$ is the estimator at the j -th replication. The simulation results reported in Tables 5.3 and 5.4 below are based on $N = 5000$ replications and provide the mean ABS and MSE outcomes.

Table 5.3: ABS and MSE Outcomes for Case F

Case A	$n = 201$	$n = 501$	$n = 901$
$ABS(\hat{\theta})$	0.002723	0.001032	0.000587
$std(\hat{\theta})$	0.002541	0.000968	0.000521
$ABS(\tilde{\theta})$	0.029187	0.012981	0.0055897
$std(\tilde{\theta})$	0.098776	0.074876	0.035657
$MSE_1(\hat{\beta}_1)$	0.121563	0.093451	0.080748
$MSE_2(\hat{\beta}_1)$	0.141674	0.103564	0.090414
$MSE_1(\tilde{\beta}_1)$	0.654856	0.638715	0.604646
$MSE_2(\tilde{\beta}_1)$	0.914873	0.870187	0.843102

Tables 5.3 and 5.4 clearly show the superiority of the estimator $\hat{\theta}$ in terms of both bias and mean square error, at least under Assumption 4.2(ii). The semiparametric weighted least squares estimator $\tilde{\theta}$ has decidedly poor performance by comparison. Similarly, the nonparametric estimator $\hat{\beta}_1$ is superior to $\tilde{\beta}_1$ in all cases and by both criteria. The usual semiparametric estimator is therefore seen to be quite unreliable in this class of models

with a nonstationary regressor, even though it is commonly used in partially linear model estimation.

Table 5.4: ABS and MSE Outcomes for Case G

Case B	$n = 201$	$n = 501$	$n = 901$
$ABS(\hat{\theta})$	0.003576	0.001329	0.000658
$std(\hat{\theta})$	0.003109	0.001108	0.000529
$ABS(\tilde{\theta})$	0.034623	0.017328	0.008946
$std(\tilde{\theta})$	0.09826	0.07639	0.053879
$MSE_1(\hat{\beta}_1)$	0.114636	0.095783	0.081793
$MSE_2(\hat{\beta}_1)$	0.135723	0.10958	0.094764
$MSE_1(\tilde{\beta}_1)$	0.684102	0.668231	0.636728
$MSE_2(\tilde{\beta}_1)$	0.981439	0.947628	0.875682

6 Empirical application to consumption

This section provides an empirical application of the methods to explain aggregate consumption behavior in the US over the period 1960 - 2009. Our primary focus in this application is the identification and estimation of potential nonlinearities in relationships involving the nonstationary aggregate consumption data. A secondary focus is to explore the potential role of other macroeconomic variables like interest rates in influencing the form of the nonstationary relationships.

We use data from the Bureau of Economic Analysis² on the following three variables: $c_t = \log(\text{consumption expenditure})$, $i_t = \log(\text{disposable income})$, and $r_t = \text{real interest rate}$. Note that c_t , i_t and r_t are all real data. The data are quarterly and comprise 199 observations over the period from the first quarter of 1960 to the last quarter of 2009. The real interest rate variable is measured by subtracting the ex post inflation rate over the following quarter from the nominal interest rate. The data are plotted in Fig. 1. The histograms shown on the axis borders provide crude estimates of the local time spent at various levels by the series over the period 1960Q1-2009Q4. The strong trend component in c_t and i_t is reflected in the near uniform local time estimates for these series in comparison with r_t (c.f., Phillips, 2001b,

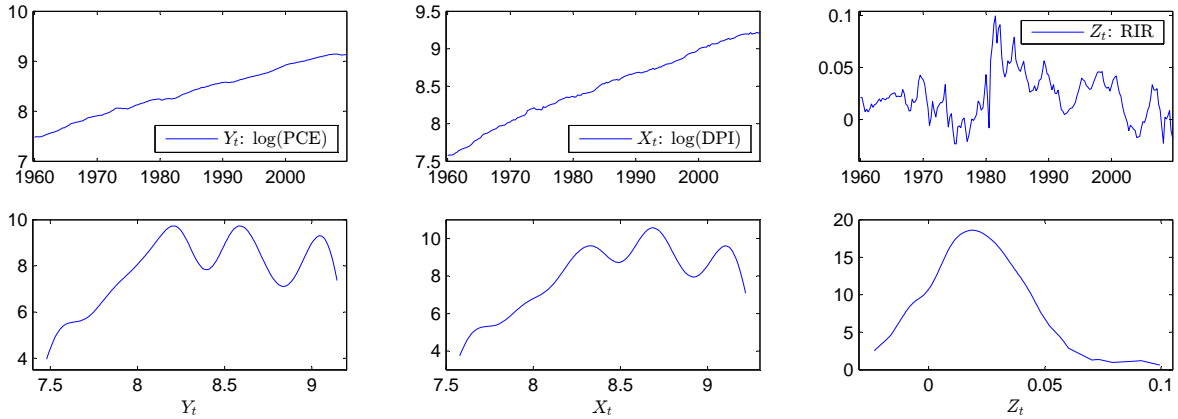


Figure 1. The top panel gives the plots of $Y_t = c_t$, $X_t = i_t$ and Z_t , respectively, and the bottom panel gives their corresponding local-time densities.

2005).

Transforming to the notation of the paper we set $Y_t = c_t$, $X_t = i_t$, and $Z_t = r_t$. A convenient starting point is the prototype fixed parameter consumption function

$$Y_t = \alpha + \beta X_t + e_t, \quad (6.1)$$

whose differenced form

$$Y_t - Y_{t-1} = \beta (X_t - X_{t-1}) + e_t - e_{t-1} \equiv \beta (X_t - X_{t-1}) + \varepsilon_t \quad (6.2)$$

is the basis for many empirical models that are common in the literature (e.g., Campbell and Mankiw 1990; Campbell, Lo and MacKinlay 1997).

The methods developed in the paper can be used to check whether a fixed parameter model such as (6.2) is supported empirically against more general functional specifications. There is a growing literature (see, for example, Gylfason, 1981; Faff and Brooks, 1998; Hahn and Steigerwald, 1999; Cai, Li and Park, 2009; Xiao, 2009) to support alternative formulations that treat the propensity to consume parameter β as a function of certain covariates. Among other possibilities, polynomial functions have often been suggested as flexible functional forms for variable coefficients such as $\beta(\cdot)$ – see, for example, Faff and Brooks (1998). Underlying such formulations is a Hall-type (1978) consumption model with varying coefficients of the form

$$Y_t = \alpha(Z_{t-1}) + \beta(Z_{t-1})Y_{t-1} + \zeta_t, \quad (6.3)$$

²The data are available at: <http://www.bea.gov>.

where both $\alpha(z)$ and $\beta(z)$ are unknown functions of z , and ζ_t is an error process. This model may be fitted and analyzed using the methods of the present paper. Specific functional forms may then be tested against (6.3).

Let $K(\cdot)$ be a probability kernel and h the bandwidth. Functions $\alpha(z)$ and $\beta(z)$ can be nonparametrically estimated by minimising

$$\frac{1}{n} \sum_{t=1}^n (Y_t - \alpha(z) - \beta(z)Y_{t-1})^2 K\left(\frac{Z_{t-1} - z}{h}\right) \quad (6.4)$$

over $\alpha = \alpha(z)$ and $\beta = \beta(z)$, leading to the nonparametric estimates

$$\hat{\beta}(z) = \frac{\sum_{s=1}^n \sum_{t=1}^n Y_s Y_{s-1} K_{st}(z) - \sum_{s=1}^n \sum_{t=1}^n Y_s Y_{t-1} K_{st}(z)}{\sum_{s=1}^n \sum_{t=1}^n Y_{s-1}^2 K_{st}(z) - \sum_{s=1}^n \sum_{t=1}^n Y_{s-1} Y_{t-1} K_{st}(z)}, \quad (6.5)$$

$$\hat{\alpha}(z) = \frac{\sum_{t=1}^n K\left(\frac{Z_{t-1}-z}{h}\right) (Y_t - \hat{\beta}(z)Y_{t-1})}{\sum_{t=1}^n K\left(\frac{Z_{t-1}-z}{h}\right)}, \quad (6.6)$$

where $K_{st}(z) = K\left(\frac{Z_{t-1}-z}{h}\right) K\left(\frac{Z_{s-1}-z}{h}\right)$.

An application of the nonparametric test proposed in Gao *et al* (2009a) to test the null hypothesis $H_0 : \alpha(z) = \alpha_0$ and $\beta(z) = \beta_0 \equiv 1$ produces a p -value of 0.1874. This test suggests that it may not be unreasonable to assume that Y_t follows a unit-root structure of the form $Y_t = \alpha_0 + Y_{t-1} + \zeta_t$, supporting the original analysis in Hall (1978).

To use the explicit framework of the present paper with potentially nonlinear nonstationary regressors, we propose a varying-coefficient model of the form

$$\begin{aligned} Y_t &= \beta(Z_t)X_t + e_t, \\ X_t &= L_1(X_{t-1}) + \mu_t, \\ Z_t &= L_2(Z_{t-1}) + \nu_t, \end{aligned} \quad (6.7)$$

where $\beta(\cdot)$ and $L_i(\cdot)$ are all unknown functions. Taking this general nonlinear framework as a starting point, we proceed to evaluate whether the data exhibit any unit root structure by using the nonparametric test proposed in Gao *et al* (2009a) for checking empirical support in the data for the null hypothesis $H_0 : P(L_1(X_{t-1}) = X_{t-1}) = P(L_2(Z_{t-1}) = Z_{t-1}) = 1$. The respective p -value outcomes of 0.2316 and 0.1092 imply that it is reasonable to assume that both X_t and Z_t follow the unit root structure given in model (1.1). This empirical simplification along with model (6.7) then suggests the simpler system

$$\begin{aligned} Y_t &= \beta(Z_t)X_t + e_t, \\ X_t &= X_{t-1} + \mu_t, \\ Z_t &= Z_{t-1} + \nu_t, \\ E[e_t] &= E[\mu_t] = E[\nu_t] = 0, \end{aligned} \quad (6.8)$$

To allow for endogeneity between e_t and (X_t, Z_t) , we decompose e_t as $e_t = \lambda(\mu_t, \nu_t) + \varepsilon_t$ so that $\mathbb{E}[\varepsilon_t | \mu_t, \nu_t] = 0$. Let $U_t = (\mu_t, \nu_t)'$, $\mathbf{X}_t = (X_t, 1)'$ and $\beta(Z_t, U_t) = (\beta(Z_t), \lambda(U_t))'$. Accordingly, we can rewrite (6.8) in augmented regression format as

$$\begin{aligned} Y_t &= \mathbf{X}_t' \beta(Z_t, U_t) + \varepsilon_t = X_t \beta(Z_t) + \lambda(U_t) + \varepsilon_t, \\ X_t &= X_{t-1} + \mu_t, \\ Z_t &= Z_{t-1} + \nu_t, \end{aligned} \tag{6.9}$$

which falls within the class of varying coefficient models studied in this paper. Accordingly, using (3.4) we have

$$\begin{aligned} \widehat{\beta}(z, u) &= \begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\lambda}(u) \end{pmatrix} = \left(\sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t' K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \right)^{-1} \\ &\times \begin{pmatrix} \sum_{t=1}^n \mathbf{X}_t Y_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{t=1}^n X_t^2 K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) & \sum_{t=1}^n X_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \\ \sum_{t=1}^n X_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) & \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \sum_{t=1}^n X_t Y_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \\ \sum_{t=1}^n Y_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \end{pmatrix}. \end{aligned} \tag{6.10}$$

The asymptotic theory developed in Theorem 3.2 is then applicable to the nonparametric component estimates $(\widehat{\beta}(z), \widehat{\lambda}(u))$ of $\beta(Z_t, U_t)$.

The plot of $\widehat{\beta}(z)$ shown in Fig. 2 over $z \in (-0.02, 0.1)$ suggests that the function $\beta(z)$ may be reasonably approximated by a second-order polynomial function of the following form, at least over this part of the sample space,

$$\beta_2(z) = \beta_0 + \beta_1 z + \beta_2 z^2. \tag{6.11}$$

The parameters β_0 , β_1 and β_2 may be estimated through (6.9) when $\beta(z)$ is replaced by the specification (6.11). Since $E[\lambda(\mu_t, \nu_t)] = 0$, we apply the approach given in (4.24)–(4.27) to estimate $(\beta_0, \beta_1, \beta_2)$ by $(\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\beta}_2)$ minimizing $\sum_{t=1}^n (Y_t - X_t \beta_2(Z_t))^2$. The unknown complementary function $\lambda(U_t)$ is then estimated by $\widetilde{\lambda}(u) = \frac{\sum_{s=1}^n K_2 \left(\frac{U_s - u}{h_2} \right) (Y_s - X_s \widetilde{\beta}_2(Z_s))}{\sum_{s=1}^n K_2 \left(\frac{U_s - u}{h_2} \right)}$, in which $K_2(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$ and h_2 is chosen by semiparametric cross-validation as in Section 5 above.

The fitted parametric model is

$$\widetilde{\beta}_2(z) = \widetilde{\beta}_0 + \widetilde{\beta}_1 z + \widetilde{\beta}_2 z^2, \tag{6.12}$$

where $\tilde{\beta}_0 = 0.6233$, $\tilde{\beta}_1 = -0.4293$ and $\tilde{\beta}_2 = 1.5725$. A t -test confirms that all these coefficients are significant with p -values close to zero³. An application of the specification test proposed in Gao, Tjøstheim and Yin (2012) to test this parametric specification against a general nonparametric alternative produces a p -value of 0.2317. This test outcome suggests that the quadratic approximation $\tilde{\beta}_2(z)$ of $\beta(z)$ may be reasonable over this particular region of the sample space. A simpler linear parametric form produced the fitted function

$$\tilde{\beta}_1(z) = 0.6165 - 1.901z. \quad (6.13)$$

As in the case of (6.12), a t -test shows that the coefficients are all significant.

The nonparametric estimate $\hat{\beta}(z)$ is shown in Figure 2 against plots of the parametric estimates $\tilde{\beta}_1(z)$ and $\tilde{\beta}_2(z)$. These plots corroborate the empirical test results, indicating that the functional slope coefficient $\beta(z)$ in (6.8) can be approximately treated as a second order polynomial function of Z_t rather than as a constant parameter for values of the real interest rate in the region $(-0.02, 0.1)$. Fixed coefficient models such as (6.2) do not seem to be supported against general nonlinear varying coefficient alternatives for these data.

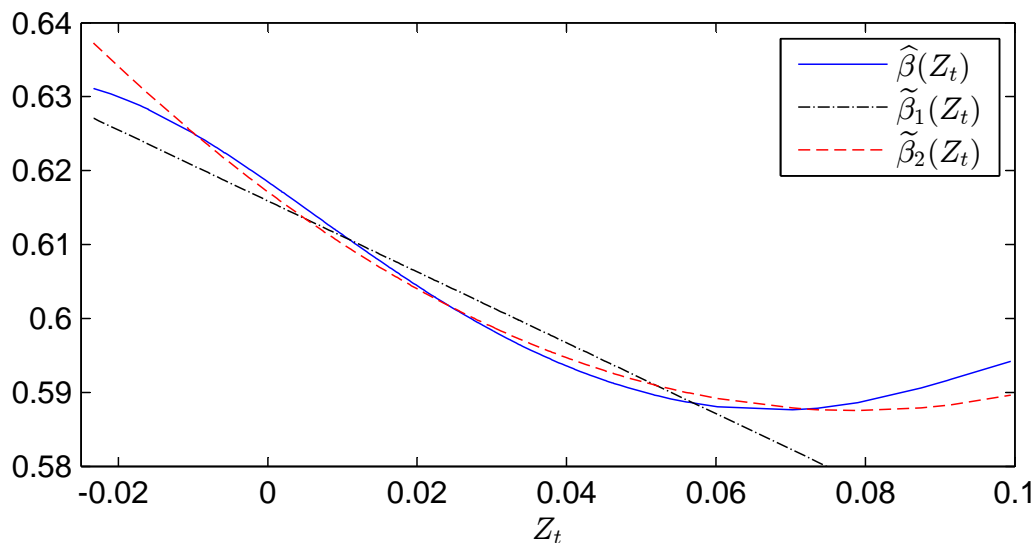


Figure 2. The curve in blue is the nonparametric estimate; the dot lines in red represent the second-order polynomial; and the dot lines in black denote the linear line.

This empirical implementation of kernel nonparametrics uses cross validation based choices of the bandwidths as in (5.3) and (5.4). Since the vector of regressors U_t involved in

³Conventional t -tests are robust to this type of parametric regression under nonstationarity, being equivalent to those from a standardised (weak trend) model of the form $Y_t = X_{d,t}\beta_2(Z_{d,t}) + e_t$, where $X_{d,t} = \frac{X_t}{\sqrt{n}}$ and $Z_{d,t} = \frac{Z_t}{\sqrt{n}}$, and giving the same p -values.

the nonparametric kernel estimation (6.9) is stationary, the asymptotic consistency of $\widehat{\theta}(z)$ follows in a similar way to Theorem 4.4, provided the parametric specification is correct. These asymptotic results therefore justify the use of the nonparametric and semiparametric methods in the empirical analysis in the presence of nonstationary time series regressors and varying coefficients.

7 Conclusion and further discussion

This paper has focused on varying coefficient models of the type (1.1) that include a wide range of related specifications such as multivariate nonparametric models of the form (2.1) and partial linear models such as those in (2.6) - (2.8). Many of these models are now used in empirical research with cross section data under independence assumptions. But the models are also relevant in time series contexts in econometrics where stochastic nonstationarity is a feature of much economic and financial data. The model constructions we have used here allow for intrinsically nonstationary specifications in the data generating mechanisms within a wide class of nonparametric and semiparametric regressions. Asymptotic theory for all these regressions is developed as well as some new methods of estimation for particular cases that take advantage of the data nonstationarity. These results will be of use to practitioners in time series econometrics who want to consider many different alternatives to linear specifications, including varying coefficient models and additive nonlinear nonstationary systems.

Some developments of the methods and results of the paper are possible and desirable. In an early draft of this paper (Gao and Phillips, 2012), we also considered models such as (1.1) in cases where $X_t = f_t + \mu_t$ and $Z_t = g_t + \nu_t$, with f_t and g_t being unknown deterministic functions of t . For such specifications and under certain conditions, the estimation theory corresponds to the case where X_t and Z_t are stationary. Rather more generally, the proposed procedures and limit theory may be extended to deal with models of the type

$$\begin{aligned} Y_t &= X_t' \beta(Z_t, U_t) + e_t, \\ X_t &= X_{t-1} + f_t + \mu_t, \\ Z_t &= Z_{t-1} + g_t + \nu_t, \end{aligned} \tag{7.1}$$

where f_t and g_t are unknown deterministic functions of t . In this case, the variables involve both stochastic trends and deterministic components. In such cases, it is realistic in practical work to expect that the functional forms of the deterministic components will be unknown. Then reparameterization, filtering, or linear regression extraction (as in Park and Phillips,

1988,1989) is generally not applicable for removing the trends involved in X_t and Z_t without risk of misspecification bias. Nonparametric estimation is needed to address these more general specifications. Establishment of a limit theory in such cases depends on how the deterministic trend components behave asymptotically. For instance, in the case where (f_t, g_t) are weak trends (i.e., representable in standardized form as $(f(\frac{t}{n}), g(\frac{t}{n}))$ for some continuous functions (f, g)), X_t has the following asymptotic form upon restandardization

$$\begin{aligned} \frac{1}{n}X_t &= \frac{1}{n} \sum_{s=1}^t \mu_s + \frac{1}{n} \sum_{s=1}^t f\left(\frac{s}{n}\right) = \frac{1}{\sqrt{n}}\mu_{t,n} + \int_0^{\frac{t}{n}} f(r)dr + o(1) \\ &= \frac{1}{\sqrt{n}}\mu_{t,n} + q\left(\frac{t}{n}\right) + o(1) = q\left(\frac{t}{n}\right) + o_P(1), \end{aligned} \quad (7.2)$$

where $\mu_{t,n} = \frac{1}{\sqrt{n}} \sum_{s=1}^t \mu_s = O_p(1)$ and $q(r) = \int_0^r f(s)ds$ is an accumulated trend function. A different array specification with $f_{t,n} = n^{-1/2}f(\frac{t}{n})$ and $g_{t,n} = n^{-1/2}g(\frac{t}{n})$, leads to an alternate asymptotic form in which both deterministic and stochastic trends remain relevant, viz.,

$$\frac{X_t}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{s=1}^t \mu_s + \frac{1}{n} \sum_{s=1}^t f\left(\frac{s}{n}\right) = \frac{1}{\sqrt{n}}\mu_{t,n} + \int_0^{\frac{t}{n}} f(r)dr + o(1) = B_\mu\left(\frac{t}{n}\right) + q\left(\frac{t}{n}\right) + o_P(1),$$

in a suitable probability space where $\mu_{[nr],n} = B_\mu(r) + o_p(1)$. For such a model, some results corresponding to those given here seem attainable and worthy of future investigation. Further extensions of our results to cases where the regressor X_t and functional coefficient argument Z_t have a local to unity rather than unit root structure also seem possible, following on from recent work on nonparametric asymptotics (Wang and Phillips, 2009a).

8 Appendix A: Assumptions and proofs

8.1 Assumptions

This paper considers the case where covariates X_t and Z_t are generated as integrated processes according to

$$X_t = X_{t-1} + \mu_t \quad \text{and} \quad Z_t = Z_{t-1} + \nu_t, \quad t = 1, 2, \dots, n, \quad (7.1)$$

where $X_0 = Z_0 = O_P(1)$. Let $d_\mu, d_\nu = 1$ and d_u be the dimensions of X_t, Z_t and U_t , respectively, $d_{u\mu\nu} = d_u + d_\mu + 1$ and $\{\xi_i : -\infty < i < \infty\}$ be a vector of $d_{u\mu\nu}$ -dimensional independent and identically distributed (*iid*) random variables.

Let $W_t = (\mu'_t, \nu'_t, U'_t)'$. Suppose that there is a real-valued matrix of lag coefficients of the form $C_j = (c_{j,kl} : 1 \leq k, l \leq d_{u\mu\nu})$ such that $W_t = \sum_{j=0}^{\infty} C_j \xi_{t-j}$. As discussed in Remark A.2 in Appendix A below, the main results of this paper remain true when μ_t and ν_t follow linear processes, and

U_t is a stationary time series generated by $U_t = \Lambda(\xi_{t-1}, \dots, \xi_{t-\tau}; \eta_t)$, where $\Lambda(\cdot, \dots, \cdot)$ is a vector function and η_t is another vector of independent and identically distributed random variables.

Throughout the paper we use $\|\cdot\|$ for the Euclidean norm, “ $\Rightarrow_{\mathcal{D}}$ ” for weak convergence, “ \rightarrow_D ” for convergence in distribution, and “ \rightarrow_P ” for convergence in probability.

Assumption A.1 (i) Let $\{\xi_i : -\infty < i < \infty\}$ be a sequence of *iid* continuous random vectors with $\mathbb{E}[\xi_1] = 0$ and positive definite matrix Σ_ξ and finite fourth order cumulants. Let $\varphi(u)$ be the characteristic function of ξ_1 and assume $\int_{-\infty}^{\infty} |u| |\varphi(u)| du < \infty$. Let the density, $p_\xi(\cdot)$, of ξ_1 satisfy $\int |p_\xi(x+y) - p_\xi(x)| dx \leq c_\xi |y|$ for each given y and some constant $c_\xi > 0$. Let $\mathbb{E} \left[\|\xi_1\|^{2+\delta} \right] < \infty$ for some $\delta > 0$ satisfying $2\delta^2 + 4\delta - 5 > 0$.

(ii) The coefficients $\{c_{j,kl}\}$ satisfy $\sum_{j=0}^{\infty} c_{j,kl} z^j \neq 0$ for $|z| \leq 1$ and $c_{j,kl} = O(j^{-\lambda})$ as $j \rightarrow \infty$, where $\lambda > 1$ is chosen such that $\lambda + \frac{1}{2} > 2 + \delta > \frac{2}{\lambda-1}$ with $\delta > 0$ as in (i).

(iii) Let $\mathcal{F}_t = \sigma(e_t, \dots, e_1; \xi_{t+1}, \xi_t, \dots, \xi_{-\infty})$ be a σ -field generated $\{(e_i, \xi_j) : 1 \leq i \leq t; -\infty < j \leq t+1\}$ with $\mathbb{E}[e_t | \mathcal{F}_{t-1}] = 0$ a.s., $\mathbb{E}[e_t^2 | \mathcal{F}_{t-1}] = \sigma_e^2$ a.s. and $\mathbb{E}[e_t^4 | \mathcal{F}_{t-1}] < \infty$ a.s. for all $t \geq 2$, where $\sigma_e^2 > 0$ is some constant.

(iv) Let $E_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} e_t$ and $W_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} W_t$. There is a vector Brownian motion (B_e, B_w) such that $(E_n(r), W_n(r)) \Rightarrow_{\mathcal{D}} (B_e(r), B_w(r))$ on the Skorohod space $D[0, 1]^{d_{uw}+1}$ as $n \rightarrow \infty$, where $B_w(r) = (B'_\mu(r), B_\nu(r), B'_u(r))'$.

(v) Let $\{\xi_t\}$ and $\{e_s\}$ be independent for all $t \geq s+2$.

Assumption A.2. (i) Suppose that $\beta(z, u)$ is continuously differentiable in (z, u) .

(ii) For $i = 1, 2$, let each $K_i(\cdot)$ be symmetric, continuous, non-negative and bounded probability densities with $\int \|u\|^2 K_i(u) du < \infty$ for $i = 1, 2$.

(iii) Let $\lim_{n \rightarrow \infty} h_1 = \lim_{n \rightarrow \infty} h_2 = 0$, $\lim_{n \rightarrow \infty} \sqrt{n} h_1 h_2^{d_u} = \infty$, $\lim_{n \rightarrow \infty} \sqrt{n} h_1^5 h_2^{d_u} = 0$ and $\lim_{n \rightarrow \infty} \sqrt{n} h_1 h_2^{4+d_u} = 0$.

Assumption A.3. Let $p(u)$ be the marginal density function of U_t and $p_\tau(u, v)$ be the joint density of $(U_t, U_{t+\tau})$. Suppose that $p(u)$ is continuous in u and that $p_\tau(u, v)$ is also continuous in (u, v) uniformly in $\tau \geq 1$.

Assumption A.4. (i) $\beta_1(z)$ and $\beta_2(u)$ are both continuously differentiable.

(ii) $R_1(\cdot)$ and $R_2(\cdot)$ are continuously differentiable. For $i = 1, 2$, there exist bounds $0 < c_{i \min} < c_{i \max} < \infty$ such that $c_{1 \min} \leq \pi_1 \leq c_{1 \max}$ and $c_{2 \min} \leq \pi_2 \leq c_{2 \max}$, where $\pi_1 = \int \frac{r_1^2(u)}{p(u)} du$ with $r_1(u) = \frac{dR_1(u)}{du}$ and $\pi_2 = \int r_2^2(z) dz$ with $r_2(z) = \frac{dR_2(z)}{dz}$.

(iii) Additionally, $\int |\beta_1^{(1)}(z)| r_2(z) dz < \infty$ and $\int |\beta_2^{(1)}(u)| r_1(u) du < \infty$.

Since the lag coefficient matrix C_j is not necessarily diagonal, Assumption A.1(i)(ii) allows for contemporaneous correlation between the regressors and the residuals. This joint dependence structure allows for the presence of endogeneity and nonstationarity.

Remark A.1(i) Assumption A.1(i)(ii) ensures that W_t is stationary and α -mixing (see, for example, Corollary 4 of Withers 1981; Theorem 2.1 of Pham and Tran 1985). Assumption A.1(i)(ii) allows for correlated $\{U_t\}$ and $\{\mu_t, \nu_t\}$, including the special case of strong endogeneity where $U_t = \nu_t$. As a result, we can have $Z_t - Z_{t-1} = U_t$ as mentioned in models (2.4) and (2.10) above. Note also that Assumption A.1(i)(ii) covers the case where $\{U_t\}$ is independent of $\{(\mu_t, \nu_t)\}$. Both the correlated and uncorrelated cases for $\{U_t\}$ and $\{(\mu_t, \nu_t)\}$ are considered in simulations. Instead of imposing a linear process structure, one may directly assume that W_t is a vector of stationary and α -mixing time series with mixing coefficient $\alpha_w(\cdot)$ satisfying $\sum_{k=1}^{\infty} \alpha_w^{\frac{\delta_1}{2+\delta_1}}(k) < \infty$, where $\delta_1 > 0$ is chosen such that $E \|W_t\|^{2+\delta_1} < \infty$. In order to validate the main theorems in this case (particularly the proof of Theorem 3.2), the extra condition is included in Assumption A.1(iv) that $B_\mu(r)$ and $B_\nu(r)$ are independent.

(ii) In applications, we may choose $\mathcal{F}_t = \sigma(e_t, \dots, e_1; \xi_{t+1}, \xi_t, \dots, \xi_{-\infty})$ generated by $\{(e_i, \xi_j) : 1 \leq i \leq t; -\infty < j \leq t\}$. In this case, Assumption A.1(v) holds if $\{\xi_s\}$ and $\{e_t\}$ are independent for all $s \geq t + 1$. Assumption A.1(iii)-(iv) also allows for heteroskedastic innovations e_t . Assumption A.2(i)(ii) imposes some mild conditions on the kernel functions and $\beta(z, u)$. Assumption A.2(iii) imposes some technical conditions on the bandwidth parameters. The last part of Assumption A.2(iii) links the mixing coefficient $\alpha_w(\cdot)$ with the bandwidths - it is verifiable and satisfied when $h_i = O(n^{-\lambda_i})$ for $i = 1, 2$, and $(\lambda_0, \lambda_1, \lambda_2)$ is chosen suitably such that $\lambda_1 \geq \frac{1}{2}(\lambda_2 - \lambda_0)$. Assumptions A.3 and A.4 are reasonable and may be justified under more primitive conditions.

8.2 Useful lemmas

The following lemmas are needed for us to establish some useful asymptotic properties and are of independent interest.

Let $q_t(u|x, z)$ and $q_t(u|z)$ be the conditional density functions of U_t given $\left(\frac{X_t}{\sqrt{t}}, \frac{Z_t}{\sqrt{t}}\right)$ and of U_t given $\frac{Z_t}{\sqrt{t}}$, respectively. For $i = 0, 1$, let $q_t^{(i)}(u|x, z)$ and $q_t^{(i)}(u|z)$ be the i -th partial derivative with respect to t . We then have the following lemma.

Lemma A.1. *Let Assumption A.1(i)(ii) hold. As $t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$, then $\left(\frac{X_t}{\sqrt{t}}, \frac{Z_t}{\sqrt{t}}\right), \left(\frac{X_s}{\sqrt{s}}, \frac{Z_s}{\sqrt{s}}\right), U_t$ and U_s are mutually independent.*

Meanwhile, we have as $t \rightarrow \infty$

$$q_t^{(i)}(u|x, z) \rightarrow p^{(i)}(u) \quad \text{and} \quad q_t^{(i)}(u|z) \rightarrow p^{(i)}(u) \quad (7.2)$$

for $i = 0, 1$, where $p(u)$ denotes the marginal density of U_t .

Proof: Let us introduce some notation. Recall the definitions of μ_t, ν_t and U_t as given in Assumption A.1. Let D_1 and D_2 be vectors of real numbers, and D_3 be a real number itself. Define

$u_t = D_1^\tau \mu_t + D_3 \nu_t$ and $v_t = D_2^\tau U_t$. It follows from Assumption A1(i)(ii) that there are coefficients $\{c_j : j \geq 0\}$ and $\{d_j : j \geq 0\}$ as well as a sequence of independent and identically distributed random variables $\{\varepsilon_i : -\infty < i < \infty\}$ such that

$$u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \quad \text{and} \quad v_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}, \quad (7.3)$$

where $\{\varepsilon_i\}$ satisfies the same conditions that are imposed on the components of ξ_i in Assumption A.1(i), and $\{c_j : j \geq 0\}$ and $\{d_j : j \geq 0\}$ satisfy such conditions that are imposed on $\{c_{j,kl}\}$ in Assumption A.1(ii). In order to prove Lemma A.1, it suffices to show that $\frac{u_t}{\sqrt{t}}$, $\frac{u_s}{\sqrt{s}}$, v_t and v_s are mutually independent when $t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$.

Before we prove this lemma, we point out that Lemma A.1 implies the mutual independence between $\frac{u_t}{\sqrt{t}}$ and $\left(\frac{u_s}{\sqrt{s}}, v_t, v_s\right)$ when $t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$. As a consequence, it implies the mutual independence between Z_t and (Z_s, v_t, v_s) when $t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$.

To simplify the notation, we assume without loss of generality that $\text{var}(u_t) = \text{var}(v_t) \equiv 1$. In addition, we avoid involving s and t as the indices. For $t > s$, let $Y_1 = \frac{\sum_{i=1}^t u_i}{\sqrt{t}}$, $Y_2 = \frac{\sum_{j=1}^s u_j}{\sqrt{s}}$, $Y_3 = v_t$ and $Y_4 = v_s$, and $\psi_{st}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \psi(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ be the characteristic function of (Y_1, Y_2, Y_3, Y_4) . Our aim in the following proof is to show that as $t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$,

$$\psi_{st}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \rightarrow \exp\left(-\frac{1}{2}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2)\right). \quad (7.4)$$

Using the structure of u_t , we have

$$u_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t = C(1)\varepsilon_t + \bar{u}_t, \quad (7.5)$$

where $C(1) = \sum_{j=0}^{\infty} c_j$ and $\bar{u}_t = \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t$ with $\tilde{\varepsilon}_t = \sum_{j=0}^{\infty} \tilde{c}_j \varepsilon_{t-j}$, in which $\tilde{c}_j = \sum_{k=j+1}^{\infty} c_k$.

Using the fact that $\max_i E[\varepsilon_i^2] < \infty$, we have

$$E\left[\frac{1}{\sqrt{t}} \sum_{i=1}^t \bar{u}_i\right]^2 = \frac{1}{t} E[\tilde{\varepsilon}_0 - \tilde{\varepsilon}_t]^2 = O\left(\frac{1}{t}\right), \quad (7.6)$$

which implies that $\frac{1}{\sqrt{t}} \sum_{i=1}^t \bar{u}_i = O_P\left(\frac{1}{\sqrt{t}}\right)$ as $t \rightarrow \infty$. This implies that we need only to approximate u_t by $C(1)\varepsilon_t$ in the following derivations. Without loss of generality, we set $C(1) \equiv 1$ in the rest of this proof without changing the notation of Y_1 and Y_2 . In other words, we now define $Y_{1t} = \frac{\sum_{i=1}^t \varepsilon_i}{\sqrt{t}}$ and $Y_{2s} = \frac{\sum_{j=1}^s \varepsilon_j}{\sqrt{s}}$.

In view of equations (7.3)–(7.6), we have the following decompositions: for $t > s$

$$\begin{aligned} Y_{1t} &= \frac{1}{\sqrt{t}} \sum_{i=1}^t \varepsilon_i = \frac{1}{\sqrt{t}} \sum_{i=1}^s \varepsilon_i + \frac{1}{\sqrt{t}} \sum_{i=s+1}^t \varepsilon_i, & Y_{2s} &= \frac{1}{\sqrt{s}} \sum_{i=1}^s \varepsilon_i, \\ Y_{3K} &= \sum_{j=0}^{\infty} d_j \varepsilon_{t-j} = \sum_{j=1}^s d_{t-j} \varepsilon_j + \sum_{j=s+1}^t d_{t-j} \varepsilon_j + \sum_{k=K}^{\infty} d_{t+k} \varepsilon_{-k}, \\ Y_{4K} &= \sum_{l=0}^{\infty} d_l \varepsilon_{s-l} = \sum_{l=1}^s d_{s-l} \varepsilon_l + \sum_{l=K}^{\infty} d_{s+l} \varepsilon_{-l}, \end{aligned} \quad (7.7)$$

where $K > 1$ is some positive integer.

For given $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, we have

$$\lambda_1 Y_{1t} + \lambda_2 Y_{2s} + \lambda_3 Y_{3K} + \lambda_4 Y_{4L} = \sum_{i=1}^s a_{st}(i) \varepsilon_i + \sum_{j=s+1}^t b_{st}(j) \varepsilon_j + \sum_{k=-K}^0 c_{st}(k) \varepsilon_k, \quad (7.8)$$

where $a_{st}(i) = \frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}} + \lambda_3 d_{t-i} + \lambda_4 d_{s-i}$, $b_{st}(j) = \frac{\lambda_1}{\sqrt{t}} + \lambda_3 d_{t-j}$ and $c_{st}(k) = \lambda_3 d_{t-k} + \lambda_4 d_{s-k}$.

Let $\psi(\cdot)$ be the characteristic function of ε_t . The logarithm of the characteristic function of $(Y_{1t}, Y_{2s}, Y_{3K}, Y_{4K})$ is then given by

$$\begin{aligned} \log(\Psi_{st,K}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)) &= \log \left(\prod_{i=1}^s \psi(a_{st}(i)) \prod_{j=s+1}^t \psi(b_{st}(j)) \prod_{k=-K}^0 \psi(c_{st}(k)) \right) \\ &= \sum_{i=1}^s \log(\psi(a_{st}(i))) + \sum_{j=s+1}^t \log(\psi(b_{st}(j))) + \sum_{k=-K}^0 \log(\psi(c_{st}(k))) \\ &= \frac{-(1+o(1))}{2} \left(\sum_{i=1}^s a_{st}^2(i) + \sum_{j=s+1}^t b_{st}^2(j) + \sum_{k=-K}^0 c_{st}^2(k) \right) \\ &= \frac{-(1+o(1))}{2} \sum_{i=1}^s \left(\frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}} + \lambda_3 d_{t-i} + \lambda_4 d_{s-i} \right)^2 \\ &\quad + \frac{-(1+o(1))}{2} \sum_{j=s+1}^t \left(\frac{\lambda_1}{\sqrt{t}} + \lambda_3 d_{t-j} \right)^2 + \frac{-(1+o(1))}{2} \sum_{k=-K}^0 (\lambda_3 d_{t-k} + \lambda_4 d_{s-k})^2 \\ &= -\frac{\lambda_1^2}{2} \sum_{j=s+1}^t \frac{1}{t} - \frac{\lambda_2^2}{2} \sum_{i=1}^s \frac{1}{s} - \frac{\lambda_3^2}{2} \left(\sum_{i=1}^s d_{t-i}^2 + \sum_{j=s+1}^t d_{t-j}^2 + \sum_{k=-K}^0 d_{t-k}^2 \right) \\ &\quad - \frac{\lambda_4^2}{2} \left(\sum_{i=1}^s d_{s-i}^2 + \sum_{j=s+1}^t d_{t-j}^2 + \sum_{k=-K}^0 d_{s-k}^2 \right) + o(1) \\ &\rightarrow -\frac{1}{2} (\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2) \end{aligned} \quad (7.9)$$

as $t, s \rightarrow \infty$, $\frac{s}{t} \rightarrow 0$ and $K \rightarrow \infty$, where we have used the following facts:

$$\sum_{i=1}^s d_{t-i}^2 = \sum_{j=t-s}^{t-1} d_j^2 = o(1) \quad \text{and} \quad \sum_{k=0}^{\infty} d_k^2 = E[v_1^2] = 1$$

$t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$. This completes equation (7.4) and thus the first part of Lemma A.1.

To prove the second part of Lemma A.1, we let $p_t(x, z, u)$ and $p_t(z, u)$ be the joint density functions of (X_t, Z_t, U_t) and (Z_t, U_t) , respectively. We also let $p_t(x, z)$ and $p_t(x, z|u)$ be the joint density of (X_t, Z_t) and the conditional density of (X_t, Z_t, U_t) given U_t . Let $q_t(x, z, u)$, $q_t(z, u)$, $q_t(x, z)$ and $q_t(x, z|u)$ denote the joint and conditional density functions of $\left(\frac{X_t}{\sqrt{t}}, \frac{Z_t}{\sqrt{t}}, U_t\right)$, $\left(\frac{Z_t}{\sqrt{t}}, U_t\right)$, $\left(\frac{X_t}{\sqrt{t}}, \frac{Z_t}{\sqrt{t}}\right)$ and $\left(\frac{X_t}{\sqrt{t}}, \frac{Z_t}{\sqrt{t}}, U_t\right)$ given U_t , respectively. Using the mutual independence between $\frac{X_t}{\sqrt{t}}$, $\frac{Z_t}{\sqrt{t}}$ and U_t , we then have

$$\begin{aligned} q_t(u|x, z) &= \frac{q_t(x, z, u)}{q_t(x, z)} = \frac{q_t(x, z|u)p(u)}{q_t(x, z)} \rightarrow p(u), \\ q_t(u|z) &= \frac{q_t(z, u)}{q_t(z)} = \frac{q_t(z|u)p(u)}{q_t(z)} \rightarrow p(u), \end{aligned} \quad (7.10)$$

where $q_t(z)$ denotes the density function of $\frac{Z_t}{\sqrt{t}}$. The proof of the second part is thus completed.

Remark A.2. As may be seen from the derivations in (7.5)–(7.9), U_t can still be correlated with (X_t, Z_t) , but it does not need to follow a linear process as assumed in Assumption A.1. As a matter of fact, U_t and (X_t, Z_t) can still be mutually independent when (X_t, Z_t) satisfies Assumption A.1, but $\{U_t\}$ is a stationary time series generated by a vector function of the form $U_t = \Lambda(\xi_{t-1}, \dots, \xi_{t-\tau}; \eta_t)$, where $\Lambda(\cdot, \dots, \cdot)$ is a measurable function such that $E[||U_t||^2] < \infty$, $\{\eta_j\}$ is a sequence of independent and identical distributed random variables and independent of $\{\xi_i\}$, and $\tau \geq 1$ is a positive integer.

Let us now give some details to show that the conclusions of Lemma A.1 remain true in the case where $\{U_t\}$ is generated by the functional form. Recall the notation of u_t and define $v_t = D_2^T U_t \equiv \lambda(\varepsilon_{t-1}, \dots, \varepsilon_{t-\tau}; \eta_t)$, in which $\lambda(\cdot, \dots, \cdot)$ is a measurable function such that $E[|\lambda(\varepsilon_{t-1}, \dots, \varepsilon_{t-\tau}; \eta_t)|^2] < \infty$. We then introduce the following notation: for $t > s + \tau - 1$

$$\begin{aligned} Y_{1t} &= \frac{1}{\sqrt{t}} \left(\sum_{i=1}^{s-\tau-1} \varepsilon_i + \sum_{j=s-\tau}^{s-1} \varepsilon_j + \varepsilon_s + \sum_{k=s+1}^{t-\tau-1} \varepsilon_k + \sum_{l=t-\tau}^{t-1} \varepsilon_l + \varepsilon_t \right), \\ Y_{2s} &= \frac{1}{\sqrt{t}} \left(\sum_{i=1}^{s-\tau-1} \varepsilon_i + \sum_{j=s-\tau}^{s-1} \varepsilon_j + \varepsilon_s \right), \\ Y_{3t} &= \lambda(\varepsilon_{t-1}, \dots, \varepsilon_{t-\tau}; \eta_t) \quad \text{and} \quad Y_{4s} = \lambda(\varepsilon_{s-1}, \dots, \varepsilon_{s-\tau}; \eta_s). \end{aligned} \quad (7.11)$$

For given $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, we have

$$\begin{aligned} &\lambda_1 Y_{1t} + \lambda_2 Y_{2s} + \lambda_3 Y_{3t} + \lambda_4 Y_{4s} \quad (7.12) \\ &= \left(\frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}} \right) \sum_{i=1}^{s-\tau-1} \varepsilon_i + \left(\frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}} \right) \sum_{j=s-\tau}^{s-1} \varepsilon_j + \lambda_4 Y_{4s} \\ &+ \frac{\lambda_1}{\sqrt{t}} (\varepsilon_s + \varepsilon_t) + \frac{\lambda_2}{\sqrt{s}} \varepsilon_s + \frac{\lambda_1}{\sqrt{t}} \sum_{k=s+1}^{t-\tau-1} \varepsilon_k + \frac{\lambda_1}{\sqrt{t}} \sum_{l=t-\tau}^{t-1} \varepsilon_l + \lambda_3 Y_{3t} \\ &= a_{st} \sum_{i=1}^{s-\tau-1} \varepsilon_i + a_{st} \varepsilon_s + b_t \sum_{k=s+1}^{t-\tau-1} \varepsilon_k + a_{st} \sum_{j=s-\tau}^{s-1} \varepsilon_j + \lambda_4 Y_{4s} + b_t \varepsilon_t + b_t \sum_{l=t-\tau}^{t-1} \varepsilon_l + \lambda_3 Y_{3t}, \end{aligned}$$

where $a_{st} = \frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}}$ and $b_t = \frac{\lambda_1}{\sqrt{t}}$.

Similarly to the derivations in (7.9), the characteristic function of $(Y_{1t}, Y_{2s}, Y_{3t}, Y_{4s})$ is given by

$$\begin{aligned} \Psi_{st}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) &= \mathbb{E}[\exp(i(\lambda_1 Y_{1t} + \lambda_2 Y_{2s} + \lambda_3 Y_{3t} + \lambda_4 Y_{4s}))] \\ &= \prod_{i=1}^{s-\tau-1} \psi(a_{st}) \psi(b_t) \psi(a_{st}) \prod_{k=s+1}^{t-\tau-1} \psi(b_t) \cdot \psi_{1st} \cdot \psi_{2st}, \end{aligned} \quad (7.13)$$

where $\psi_{1st} = \mathbb{E}[\exp(i(a_{st} \sum_{j=s-\tau}^{s-1} \varepsilon_j + \lambda_4 Y_{4s}))]$ and $\psi_{2st} = \mathbb{E}[\exp(i(b_t \sum_{l=t-\tau}^{t-1} \varepsilon_l + \lambda_3 Y_{3t}))]$.

Using the result that $e^x = 1 + x \left(1 + \frac{x}{2!} + \dots\right)$ as $x \rightarrow 0$, we have as $t, s \rightarrow \infty$

$$e^{i a_{st} \sum_{j=s-\tau}^{s-1} \varepsilon_j} = 1 + i a_{st} \sum_{j=s-\tau}^{s-1} \varepsilon_j - \frac{a_{st}^2}{2} \left(\sum_{j=s-\tau}^{s-1} \varepsilon_j \right)^2 + o(a_{st}^2) = 1 + O_P\left(\frac{1}{\sqrt{s}}\right) + O_P\left(\frac{1}{\sqrt{t}}\right), \quad (7.14)$$

where we have used $E\left[\left(\sum_{j=s-\tau}^{s-1} \varepsilon_j\right)^2\right] \leq C < \infty$.

Using the dominated convergence theorem, equations (7.13) and (7.14) therefore imply that as $t, s \rightarrow \infty$

$$\begin{aligned} \log(\Psi_{st}) &= (s - \tau - 1) \log(\psi(a_{st})) + \log(\psi(a_{st})) + \log(\psi(b_t)) + (t - s - \tau - 1) \log(\psi(b_t)) \\ &\quad + \log(\psi_{1st}) + \log(\psi_{2st}) \\ &= -\frac{1}{2}(s - \tau - 1) \left(\frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}}\right)^2 - \frac{1}{2} \left(\frac{\lambda_1}{\sqrt{t}}\right)^2 - \frac{1}{2} (t - s - \tau - 1) \left(\frac{\lambda_1}{\sqrt{t}}\right)^2 \\ &\quad - \frac{1}{2} \left(\frac{\lambda_1}{\sqrt{t}} + \frac{\lambda_2}{\sqrt{s}}\right)^2 - \left(\frac{\lambda_3^2}{2} + \frac{\lambda_4^2}{2}\right) (1 + o(1)) + o(1) \\ &\rightarrow -\frac{\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2}{2}, \end{aligned} \quad (7.15)$$

which shows that Y_{1t}, Y_{2s}, Y_{3t} and Y_{4s} are mutually independent when $t, s \rightarrow \infty$ and $\frac{s}{t} \rightarrow 0$.

Before proceeding we introduce a definition and some useful formulae.

Definition A.1. (i) *The local time process $\{L_S(t, s) : t \geq 0, s \in \mathbb{R}^1 = (-\infty, \infty)\}$ of a measurable stochastic process $\{S(t), t \geq 0\}$ is defined as*

$$L_S(t, s) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^t I\{|S(r) - s| < \epsilon\} dr. \quad (7.16)$$

(ii) *For any locally integrable functions $G(\cdot)$ and $G(\cdot, \cdot)$, the following occupation times formulae hold*

$$\begin{aligned} \int_0^t G(S(r)) dr &= \int_{-\infty}^{\infty} G(s) L_s(t, s) ds \quad \text{and} \\ \int_0^t G(r, S(r)) dr &= \int_{-\infty}^{\infty} \left(\int_0^t G(r, s) dL_s(r, s) \right) ds. \end{aligned} \quad (7.17)$$

See Revuz and Yor (1999) for more detail about the definition (7.16) and the two occupation time formulae (7.17).

Remark A.3. In order to deal with the case where both stationary and nonstationary regressors are involved in the unknown function $\beta(z, u)$, new limit results are established in Lemmas A.3 and A.4 below. These lemmas introduce some new technology for proving results of this kind. Under Gaussianity and the weak convergence $(\mathbb{E}_n(r), W_n(r)) \Rightarrow_{\mathcal{D}} (B_e(r), B_w(r))$ on $D[0, 1]^{d_{u\mu\nu}+1}$ imposed in Assumption A.1(iv), existing results such as Theorem 2.1 of Wang and Phillips (2009a) and

Theorem 1 of Phillips (2009) are directly applicable in the proofs of other lemmas and theorems in the rest of Appendix A and the whole part of Appendix B. Meanwhile, the present paper employs a new central limit argument in Lemma A.6, along with Lemma A.5, to assist in the proof of Theorem 3.2.

In the rest of Appendix A and the whole part of Appendix B below, we set $d_z = 1$ and let $d_u = 1$ without loss of generality to simplify notation and derivations in the proofs.

Lemma A.2. *Let $f(x)$ be a known function defined on $(-\infty, \infty)$. Suppose that $|f(x)|$ and $f^2(x)$ are integrable with respect to $x \in (-\infty, \infty)$. Let $C_1(f) = \int_{-\infty}^{\infty} f(x)dx \neq 0$. Suppose $\{W_t\}$ satisfies Assumption A.1(i)(ii)(iv). Then as $n \rightarrow \infty$*

$$\sup_{0 \leq r \leq 1} \left| \frac{d_n}{n} \sum_{t=1}^{\lfloor nr \rfloor} f(d_n z_{tn}) - \left(\int_{-\infty}^{\infty} f(x)dx \right) L_{B_\nu}(r, 0) \right| \rightarrow_D 0, \quad (7.18)$$

where d_n is a sequence of positive numbers satisfying $d_n \rightarrow \infty$ and $\frac{d_n}{n} \rightarrow 0$, $z_{tn} = \frac{Z_t}{d_n}$, and $\lfloor a \rfloor \leq a$ denotes the integer part of a .

Proof: Note that $d_n = \sqrt{n}$ or $d_n = \frac{\sqrt{n}}{h_2}$ in the rest of our discussion. The proof of Lemma A.2 follows immediately since the conditions of Theorem 2.1 of Wang and Phillips (2009a) are satisfied trivially.

Recall that $p(u)$ is the marginal density function of $\{U_t\}$. Let $p_t(u|z)$ be the conditional density of U_t given $Z_t = z$. We have the following lemma.

Lemma A.3. *Let Assumptions A.1(i)(ii), A.2(ii)(iii) and A.3 hold. Then, for $i = 1, 2$,*

$$\frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1^i \left(\frac{Z_t - z}{h_1} \right) p_t(u|z) \rightarrow_D p(u) \int K_1^i(v)dv \cdot L_{B_\nu}(1, 0), \quad (7.19)$$

$$\frac{1}{\sqrt{nh_1}} \sum_{t=1}^n \mathbb{E} \left[K_1^i \left(\frac{Z_t - z}{h_1} \right) \right] = O(1), \quad (7.20)$$

$$\frac{1}{nh_1^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[K_1^i \left(\frac{Z_t - z}{h_1} \right) K_1^i \left(\frac{Z_s - z}{h_1} \right) \right] = O(1), \quad (7.21)$$

$$\frac{1}{\sqrt{nh_1 h_2}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \pi_t(U_t, Z_t) \rightarrow_P 0, \quad (7.22)$$

where $\pi_t(U_t, Z_t) = K_2 \left(\frac{U_t - u}{h_2} \right) - \mathbb{E} \left[K_2 \left(\frac{U_t - u}{h_2} \right) | Z_t \right]$.

Proof: (i) We first prove (7.19). Let $p_t(\cdot)$ be the marginal density function of Z_t , $q_t(\cdot)$ and $q_t(\cdot|u)$ be the marginal density of $\frac{Z_t}{\sqrt{t}}$ and the conditional density of $\frac{Z_t}{\sqrt{t}}$ given $U_t = u$, respectively. Then

$$p_t(z) = \frac{1}{\sqrt{t}} q_t \left(\frac{z}{\sqrt{t}} \right) \quad \text{and} \quad p_t(z|u) = \frac{1}{\sqrt{t}} q_t \left(\frac{z}{\sqrt{t}} | u \right).$$

For each given z , Lemma A.1 then implies as $t \rightarrow \infty$

$$p_t(u|z) = \frac{p_t(z|u)p(u)}{p_t(z)} = \frac{q_t\left(\frac{z}{\sqrt{t}}|u\right)}{q_t\left(\frac{z}{\sqrt{t}}\right)} p(u) \rightarrow \frac{\phi(0)}{\phi(0)} p(u) = p(u). \quad (7.23)$$

Choose large enough $m \rightarrow \infty$ such that $\frac{m}{\sqrt{nh_1}} \rightarrow 0$. By (7.18) and (7.23) we have

$$\begin{aligned} & \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1^i\left(\frac{Z_t - z}{h_1}\right) p_t(u|z) = \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1^i\left(\frac{Z_t - z}{h_1}\right) p(u) \\ & + \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1^i\left(\frac{Z_t - z}{h_1}\right) (p_t(u|z) - p(u)) = \frac{p(u)}{\sqrt{nh_1}} \sum_{t=1}^n K_1^i\left(\frac{Z_t - z}{h_1}\right) \\ & + \frac{1}{\sqrt{nh_1}} \sum_{t=1}^m K_1^i\left(\frac{Z_t - z}{h_1}\right) (p_t(u|z) - p(u)) + \frac{1}{\sqrt{nh_1}} \sum_{t=m+1}^n K_1^i\left(\frac{Z_t - z}{h_1}\right) (p_t(u|z) - p(u)) \\ & = \int K_1^i(v) dv \cdot L_{B_\nu}(1, 0) \cdot p(u) + o_P(1) + \int K_1^i(v) dv \cdot L_{B_\nu}(1, 0) \cdot o_P(1) \\ & \rightarrow_D \int K_1^i(v) dv \cdot L_{B_\nu}(1, 0) p(u), \end{aligned} \quad (7.24)$$

where the first term follows from Lemma A.2 with $f(x) = K_1(x)$, the second term follows from the boundedness of $K_1(\cdot)$ as well as $p_t(u|z)$ and $p(u)$, and the third term follows from Lemma A.2(i) with $f(x) = K_1(x)$ and (7.18). We therefore complete the proof of (7.19).

(ii) We need only to prove (7.21) with $i = 1$. The case $i = 2$ follows similarly. Let $p_{st}(\cdot|Z_s)$ be the conditional density function of $Z_t - Z_s$ given Z_s . Let $q_{st}(\cdot|Z_s)$ be the conditional density function of $\frac{Z_t - Z_s}{\sqrt{t-s}}$ given Z_s . Then, we have

$$p_{st}(z|Z_s) = \frac{1}{\sqrt{t-s}} q_{st}\left(\frac{z}{\sqrt{t-s}}|Z_s\right). \quad (7.25)$$

To prove (7.21), we evaluate the order of

$$\begin{aligned} & \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[K_1\left(\frac{Z_t - z}{h_1}\right) K_1\left(\frac{Z_s - z}{h_1}\right) \right] = \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[K_1\left(\frac{Z_t - Z_s}{h_1} + \frac{Z_s - z}{h_1}\right) K_1\left(\frac{Z_s - z}{h_1}\right) \right] \\ & = \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[K_1\left(\frac{Z_s - z}{h_1}\right) \mathbb{E} \left(K_1\left(\frac{Z_t - Z_s}{h_1} + \frac{Z_s - z}{h_1}\right) | Z_s \right) \right] \\ & = \sum_{t=2}^n \sum_{s=1}^{t-1} \left[\int K_1\left(\frac{z_s - z}{h_1}\right) \left(\int K_1\left(\frac{z_{st}}{h_1} + \frac{z_s - z}{h_1}\right) p_{st}(z_{st}|Z_s) dz_{st} \right) p_s(z_s) dz_s \right] \\ & = \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{h_1^2}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \left(\int \int q_{st}\left(\frac{h_1 v}{\sqrt{t-s}}|Z_s\right) q_s\left(\frac{z + h_1 u}{\sqrt{s}}\right) \cdot K_1(u) K_1(u+v) dudv \right) \\ & = O(nh_1^2) \end{aligned} \quad (7.26)$$

using Lemma A.1. This therefore completes the proof of (7.21) for $i = 1$. The other case follows similarly.

(iii) We now prove (7.22). To simplify the notational expressions, we introduce the following notation:

$$\psi_1(Z_s) = K_1 \left(\frac{Z_s - z}{h_1} \right) \quad \text{and} \quad \psi_2(U_s) = K_2 \left(\frac{U_s - u}{h_2} \right).$$

Recall

$$\pi_s(U_s, Z_s) = \left(K_2 \left(\frac{U_s - u}{h_2} \right) - \mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_s \right] \right) = \psi_2(U_s) - \mathbb{E}[\psi_2(U_s) | Z_s]. \quad (7.27)$$

We then have

$$\begin{aligned} \hat{p}(z, u) &= \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n \psi_1(Z_s) \psi_2(U_s) = \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n \psi_1(Z_s) \mathbb{E}[\psi_2(U_s) | Z_s] \\ &+ \frac{1}{\sqrt{nh_1h_2}} \sum_{t=1}^n \psi_1(Z_t) \pi_s(U_s, Z_s) \equiv \frac{1}{\sqrt{nh_1h_2}} (J_{1n} + J_{2n}). \end{aligned} \quad (7.28)$$

Clearly

$$\frac{1}{h_2} \mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_t = z \right] = \int K_2 \left(\frac{y - u}{h_2} \right) p_t(y|z) dy = p_t(u|z) (1 + o(1)). \quad (7.29)$$

Meanwhile, note that $\mathbb{E}[\pi_s(U_s, Z_s) | Z_s] = 0$ and

$$\begin{aligned} \mathbb{E}[\pi_s^2(U_s, Z_s) | Z_s = z] &= \mathbb{E} \left[K_2^2 \left(\frac{U_s - u}{h_2} \right) | Z_s = z \right] \\ &- \left(\mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_s = z \right] \right)^2 = h_2 p_t(u|z) (1 + o(1)) \int K_2^2(y) dy. \end{aligned} \quad (7.30)$$

In order to deal with J_{2n} of equation (7.28), we consider the following partition such that $m \rightarrow \infty$ and $\frac{m}{n} \rightarrow 0$

$$\begin{aligned} |J_{2n}| &\leq \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} |\mathbb{E}[\psi_1(Z_t) \psi_1(Z_s) \pi_t(Z_t, U_t) \pi_s(Z_s, U_s)]| \\ &= \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} |\mathbb{E}[\psi_1(Z_t) \psi_1(Z_s) \mathbb{E}\{\pi_t(Z_t, U_t) \pi_s(Z_s, U_s) | (Z_t, Z_s)\}]| \\ &\leq \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} \mathbb{E}[|\psi_1(Z_t) \psi_1(Z_s)| \cdot |\mathbb{E}\{\pi_t(Z_t, U_t) \pi_s(Z_s, U_s) | (Z_t, Z_s)\}|] \\ &\equiv \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} \mathbb{E}[|\psi_1(Z_t) \psi_1(Z_s)| \cdot |\psi_2(Z_s, Z_t)|], \end{aligned} \quad (7.31)$$

where $\psi_2(Z_s, Z_t) = \mathbb{E}\{\pi_t(Z_t, U_t) \pi_s(Z_s, U_s) | (Z_t, Z_s)\}$.

Since $\mathbb{E} [|\pi_t(U_t, Z_t)|] \leq Ch_2$, we then apply some properties for the α -mixing condition on $\{U_t\}$ (see, for example, Lemma A.1 of Gao 2007 with $Q = 0$) to imply $|\psi_2(Z_s, Z_t)| \leq Ch_2^2 \cdot \alpha_u(t-s)$ for given (Z_t, Z_s) .

Equation (7.31) implies for $n \rightarrow \infty$

$$\begin{aligned} |J_{2n}| &\leq \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} \mathbb{E} [|\psi_1(Z_t)\psi_1(Z_s)| \cdot |\psi_2(Z_s, Z_t)|] \\ &\leq Ch_2^2 \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} \alpha_u(t-s) \cdot \mathbb{E} [|\phi_1(Z_s)\phi_1(Z_t)|] \\ &\leq Ch_2^2 h_1^2 \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} \alpha_u(t-s) \cdot \frac{1}{\sqrt{st}} \leq C \sum_{s=m+1}^n \sum_{k=1}^n \alpha_u(k) \cdot \frac{1}{\sqrt{s}} = o(nh_1^2 h_2^2), \end{aligned} \quad (7.32)$$

where we have used $E [|\psi_1(Z_s)\psi_1(Z_t)|] = O\left(\frac{h_1^2}{\sqrt{s}\sqrt{t}}\right)$ derived by Assumptions A.1(ii) and A.2(ii)(iii) as well as Lemma A.1, and we have also used Lemma A.1 and Assumption A.2(iii) in association with $\alpha_u(k) = O(k^{-\lambda_4})$ with $\lambda_4 = \lambda_0 + 1 = \frac{(2+\delta)(\lambda-1)-2}{3+\delta} > 0$ (by Corollary 4(a) of Withers 1981 and Assumptions A.1(i) and A.2(ii)).

By Assumption A.3 and Lemma A.1, in view of equations (7.31) and (7.32), we have as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{E} \left[\sum_{t=1}^n \psi_1(Z_t) \pi_t(U_t, Z_t) \right]^2 &= \sum_{t=1}^n \mathbb{E} [\psi_1^2(Z_t) \mathbb{E} (\pi_t^2(U_t, Z_t) | Z_t)] \\ &+ 2 \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} [\psi_1(Z_s) \Psi_1(Z_t) \mathbb{E} (\pi_t(U_t, Z_t) \pi_s(U_s, Z_s) | (Z_s, Z_t))] \\ &\leq 2(1 + o(1)) \sum_{s=1}^n \mathbb{E} [\psi_1^2(Z_s)] \cdot \mathbb{E} [\psi_2^2(U_s)] \\ &+ 2(1 + o(1)) \sum_{t=m+2}^n \sum_{s=m+1}^{t-1} \mathbb{E} [|\psi_1(Z_t)\psi_1(Z_s)| \cdot |\psi_2(Z_s, Z_t)|] \\ &\leq O(\sqrt{n}h_1h_2) + o(nh_1^2h_2^2) = o(nh_1^2h_2^2), \end{aligned} \quad (7.33)$$

where we have used $\mathbb{E} [(\psi_2(U_s) - \mathbb{E} [\psi_2(U_s) | Z_s])^2] \leq 2\mathbb{E} [\psi_2^2(U_s)]$ implied by Assumption A.3 and Lemma A.1. We therefore complete the proof of Lemma A.3.

Lemma A.4. *Let Assumptions A.1(i)(ii), A.2(ii)(iii) and A.3 hold. We have as $n \rightarrow \infty$*

$$\widehat{f}(z, u) \rightarrow_D p(u) L_{B_v}(1, 0), \quad (7.35)$$

where $\widehat{f}(z, u) = \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1\left(\frac{Z_s - z}{h_1}\right) K_2\left(\frac{U_s - u}{h_2}\right)$.

Proof: Equation (7.29) implies

$$\frac{1}{h_2} \mathbb{E} \left[K_2\left(\frac{U_s - u}{h_2}\right) | Z_t \right] = \frac{1}{h_2} \int K_2\left(\frac{y - u}{h_2}\right) p_t(y | Z_t) dy = p_t(u | Z_t) (1 + o_P(1)). \quad (7.36)$$

It then follows from Lemma A.3 that

$$\frac{1}{\sqrt{nh_1h_2}} \sum_{t=1}^n K_1\left(\frac{Z_t - z}{h_1}\right) K_2\left(\frac{U_t - u}{h_2}\right) - \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1\left(\frac{Z_t - z}{h_1}\right) p_t(u | Z_t) \rightarrow_P 0, \quad (7.37)$$

and to complete the proof of (7.35), in view of equation (7.19), it suffices to show that as $n \rightarrow \infty$

$$\frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) (p_t(u|Z_t) - p_t(u|z)) \rightarrow_P 0. \quad (7.38)$$

In a similar fashion to (7.23), we have $\frac{\partial p_t(u|z)}{\partial z} \rightarrow_P 0$ as $t \rightarrow \infty$ using

$$\begin{aligned} \frac{\partial p_t(u|z)}{\partial z} &= \frac{\partial p_t(z|u)}{\partial z} \frac{p(u)}{p_t(z)} - \frac{p_t(z|u)p(u)}{p_t^2(z)} \frac{dp_t(z)}{dz} \\ &= \frac{1}{\sqrt{t}} q_t^{(1)} \left(\frac{z}{\sqrt{t}} | u \right) \cdot \frac{p(u)}{q_t \left(\frac{z}{\sqrt{t}} \right)} - \frac{1}{\sqrt{t}} q_t^{(1)} \left(\frac{z}{\sqrt{t}} \right) \cdot \frac{q_t \left(\frac{z}{\sqrt{t}} | u \right) p(u)}{q_t^2 \left(\frac{z}{\sqrt{t}} \right)}, \end{aligned} \quad (7.39)$$

and Lemma A.1. This result, along with equations (7.28) and (7.29), implies as $n \rightarrow \infty$

$$\begin{aligned} \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) p_t(u|Z_t) &= \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) p_t(u|z) \\ &+ \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) (p_t(u|Z_t) - p_t(u|z)) = \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) p_t(u|z) \\ &+ \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \frac{\partial p_t(u|z_t^*)}{\partial z} (Z_t - z) = \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) p_t(u|z) \\ &+ h_1(1 + o_P(1)) \cdot \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \left(\frac{Z_t - z}{h_1} \right) \frac{\partial p_t(u|z_t^*)}{\partial z} \equiv A_{1n} + A_{2n}, \end{aligned} \quad (7.40)$$

where a Taylor expansion has been used in (7.40) with z_t^* on the line segment connecting z and Z_t , and

$$A_{1n} = \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) p_t(u|z) \rightarrow_D p(u) L_{B_\nu}(1, 0), \quad (7.41)$$

$$\begin{aligned} |A_{2n}| &= \left| h_1(1 + o_P(1)) \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n L_1 \left(\frac{Z_t - z}{h_1} \right) \frac{\partial p_t(u|z_t^*)}{\partial z} \right| \\ &\leq h_1(1 + o_P(1)) \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n \left| L_1 \left(\frac{Z_t - z}{h_1} \right) \right| \cdot \left| \frac{\partial p_t(u|z_t^*)}{\partial z} \right| \\ &= O \left(h_1 L_{B_\nu}(1, 0) \int |L_1(u)| du \right) = o_P(1), \end{aligned} \quad (7.42)$$

in which Lemma A.2 has been used in (7.41) with $K_1(u)$ and in (7.42) with $|L_1(v)| = |K_1(v)v|$, respectively. Note that equation (7.38) has been used in the line above (7.42) for the case where $t \rightarrow \infty$, and Assumption A.3(ii) has also been used to deal with the part where t is not large enough in the same way as in (7.24). The proof of (7.35) then follows from (7.41) and (7.42).

Define $x_{tn} = \frac{X_t}{\sqrt{n}}$, $c_n = \frac{\sqrt{n}}{h_1}$, $z_{tn} = \frac{Z_t}{\sqrt{n}}$ and

$$\widehat{L}(z) = \frac{c_n}{n} \sum_{s=1}^n x_{tn} x_{tn}^\tau K_1 \left(c_n \left(z_{tn} - \frac{z}{\sqrt{n}} \right) \right).$$

The following lemma extends an earlier result of Phillips (2009).

Lemma A.5. *Let Assumptions A.1(i)(ii) and A.2(ii)(iii) hold. Then, as $n \rightarrow \infty$*

$$\widehat{L}(z) \rightarrow_D \int_0^1 B_\mu(r) B_\mu(r)' dL_{B_z}(r, 0). \quad (7.43)$$

Proof: The proof of (7.43) follows that of the second conclusion of Theorem 1 of Phillips (2009), since Assumption A.1 ensures that Assumptions 2.2–2.4 of Phillips (2009) are satisfied. Note that $h(y)$ involved in Assumption 2.3(iii) of Phillips (2009) is proportional to $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, by which $\int_{-\infty}^{\infty} y^4 h(y) < \infty$ holds trivially. This completes the proof of Lemma A.5.

Lemma A.6. *Let $\{\varepsilon_{nk}\}$, $\{\eta_{nk}\}$ and $\{\xi_{nk}\}$ be sequences of random variables. Let $f_n(\dots)$ be a real function of its components and define*

$$u_{k,n} = f_n(\varepsilon_{n1}, \dots, \varepsilon_{nk}; \eta_{n1}, \dots, \eta_{nk}; \xi_{n1}, \dots, \xi_{nk}; \dots).$$

Let $\{\mathcal{F}_{n,k}; 1 \leq k \leq n\}$ be a sequence of increasing σ -fields such that $\{\varepsilon_{n,k+1}, \mathcal{F}_{n,k}; 1 \leq k \leq n\}$ forms a martingale difference and $\{u_{k,n}\}$ is adapted to $\mathcal{F}_{n,k}$ for all $1 \leq k \leq n$ and $n \geq 1$.

(i) Let $\{\eta_{n,k+1}, \varepsilon_{n,k+1}, \mathcal{F}_{n,k}; 1 \leq k \leq n\}$ form a martingale difference, and $\{\eta_{nk}\}$ and $\{\varepsilon_{nk}\}$ satisfy the following condition as $n \rightarrow \infty$ and then $m \rightarrow \infty$

$$\max_{m \leq k \leq n} |\mathbb{E}(\eta_{n,k+1}^2 | \mathcal{F}_{n,k}) - \sigma_\eta^2| \rightarrow 0 \quad \text{and} \quad \max_{m \leq k \leq n} |\mathbb{E}(\varepsilon_{n,k+1}^2 | \mathcal{F}_{n,k}) - \sigma_\varepsilon^2| \rightarrow 0 \quad a.s.,$$

for some $\sigma_\eta^2 > 0$ and $\sigma_\varepsilon^2 > 0$, and for some $\delta > 0$,

$$\max_{m \leq k \leq n} \left(\mathbb{E} \left(|\eta_{n,k+1}|^{2+\delta} | \mathcal{F}_{n,k} \right) + \mathbb{E} \left(|\varepsilon_{n,k+1}|^{2+\delta} | \mathcal{F}_{n,k} \right) \right) < \infty, \quad a.s.$$

(ii) Let $\{\xi_{nj}; j \geq 1\}$ be $\mathcal{F}_{n,1}$ -measurable for each $n \geq 1$, and there exist a sequence of positive constants $d_n \rightarrow \infty$ and a Gaussian process $G(r)$ such that $\frac{1}{d_n} \sum_{s=1}^{\lfloor nr \rfloor} \xi_{ns} \Rightarrow_D G(r)$ on $D[0, \infty)$.

In addition, $G(r)$ is assumed to be independent of $W(r)$, which is the weak limit of $W_n(r) = \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor nr \rfloor} \eta_{n,s+1} \Rightarrow_D W(r)$ on $D[0, 1]$.

(iii) Let $\max_{1 \leq k \leq n} |u_{k,n}| = o_P(1)$ and $\frac{1}{\sqrt{n}} \sum_{k=1}^n |u_{k,n}| |\mathbb{E}[\eta_{n,k+1} \varepsilon_{n,k+1} | \mathcal{F}_{n,k}]| = o_P(1)$.

(iv) There exists a random variable $T(\xi, \eta) > 0$ such that $T_n^2 = \sum_{k=1}^n u_{kn}^2 \rightarrow_D T^2(\xi, \eta)$ as $n \rightarrow \infty$.

Then, we have $\frac{S_n}{T_n} \rightarrow_D N(0, 1)$ as $n \rightarrow \infty$, where $S_n = \sum_{k=1}^n u_{kn} \varepsilon_{n,k+1}$.

Proof: In a non-trivial extension of Theorems 3.2 and 3.4 of Hall and Heyde (1980) and under certain conditions, Wang (2012) established a new martingale central limit theorem in which the usual stability condition involving convergence in probability to the limiting conditional variance may be weakened to convergence in distribution, as in Assumption (iv) above. The proof of Lemma A.6 then follows from Theorem 2.1 of Wang (2012).

9 Appendix B: Proofs of Theorems 3.1–3.2 and 4.1–4.3

9.1 Proof of Theorem 3.1

The proof of Theorem 3.1 follows from Lemma A.4.

9.2 Proof of Theorem 3.2

We assume without loss of generality that the dimension of U_t is $d_u = 1$ throughout this proof. Let

$$\begin{aligned}\Lambda_{ij}(U_t; Z_t, X_t) &= K_1^i \left(\frac{Z_t - z}{h_1} \right) \left(K_2^j \left(\frac{U_t - u}{h_2} \right) - \mathbb{E} \left[K_2^j \left(\frac{U_t - u}{h_2} \right) \mid (Z_t, X_t) \right] \right) \\ &= K_1^i \left(\frac{Z_t - z}{h_1} \right) \cdot \Delta_j(U_t; Z_t, X_t).\end{aligned}\quad (7.1)$$

Let $p_t(\cdot, \cdot)$ be the joint density function of (Z_t, X_t) and $q_t(\cdot | \cdot)$ be the conditional density of $\frac{Z_t}{\sqrt{t}}$ given $\frac{X_t}{\sqrt{t}} = \frac{x}{\sqrt{t}}$. Then, in a similar derivation to (7.23) below, we have as $t \rightarrow \infty$

$$p_t(z|x) = \frac{1}{\sqrt{t}} q_t \left(\frac{z}{\sqrt{t}} \mid \frac{x}{\sqrt{t}} \right) = O \left(\frac{1}{\sqrt{t}} \right), \quad (7.2)$$

by Lemma A.1.

Let $J = (J_1, \dots, J_{d_\mu})'$ be any given vector of real numbers satisfying $J'J = 1$, and then define $\tilde{X}_t = J'X_t$. Thus, similarly to the derivations in equations (7.24)–(7.34), by (7.2) we have

$$\begin{aligned}& \frac{1}{(\sqrt{nh_1h_2})^2} \sum_{t=1}^n \mathbb{E} \left(\tilde{X}_t^4 K_1^{2i} \left(\frac{Z_t - z}{h_1} \right) \mathbb{E} [\Delta_j^2(U_t; Z_t, X_t)] \right) \\ &= C(1 + o(1)) \frac{h_2}{n^3 h_1^2 h_2^2} \sum_{t=1}^n \mathbb{E} \left(\tilde{X}_t^4 \mathbb{E} \left[K_1^{2i} \left(\frac{Z_t - z}{h_1} \right) \mid X_t \right] \right) \\ &= C(1 + o(1)) \frac{h_2}{n^3 h_1^2 h_2^2} \sum_{t=1}^n \mathbb{E} \left(\tilde{X}_t^4 \int K_1^{2i}(w) q_t \left(\frac{z + wh_1}{\sqrt{t}} \mid \frac{X_t}{\sqrt{t}} \right) dw \right) \\ &= C(1 + o(1)) \frac{h_2}{n^3 h_1^2 h_2^2} \sum_{t=1}^m \mathbb{E} \left(\tilde{X}_t^4 \int K_1^{2i}(w) q_t \left(\frac{z + wh_1}{\sqrt{t}} \mid \frac{X_t}{\sqrt{t}} \right) dw \right) \\ &\quad + C(1 + o(1)) \frac{h_2}{n^3 h_1^2 h_2^2} \sum_{t=m+1}^n \mathbb{E} \left(\tilde{X}_t^4 \int K_1^{2i}(w) q_t \left(\frac{z + wh_1}{\sqrt{t}} \mid \frac{X_t}{\sqrt{t}} \right) dw \right) \\ &\leq C(1 + o(1)) \frac{h_2 h_1}{n^3 h_1^2 h_2^2} \sum_{t=m+1}^n \frac{t^2}{\sqrt{t}} = O \left(\frac{1}{\sqrt{nh_1h_2}} \right) = o(1),\end{aligned}\quad (7.3)$$

using $\sqrt{nh_1h_2} \rightarrow \infty$, where m is a large integer chosen such that $\frac{m^{\frac{3}{2}}}{n^3 h_1 h_2} \rightarrow 0$.

Meanwhile, using the same arguments as in the derivations of (7.22), (7.27)–(7.34) and (7.3)

above, we have

$$\begin{aligned}
& \frac{1}{(\sqrt{nn}h_1h_2)^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[\tilde{X}_s^2 \tilde{X}_t^2 K_1^i \left(\frac{Z_s - z}{h_1} \right) K_1^i \left(\frac{Z_t - z}{h_1} \right) \right. \\
& \times \mathbb{E} (\Delta_j(U_t; Z_t, X_t) \Delta_j(U_s; Z_s, X_s) | (Z_s, X_s; Z_t, X_t)) \Big] \\
= & (1 + o(1)) \frac{1}{(\sqrt{nn}h_1h_2)^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[\tilde{X}_s^2 \tilde{X}_t^2 K_1^i \left(\frac{Z_s - z}{h_1} \right) K_1^i \left(\frac{Z_t - z}{h_1} \right) \right] \\
& \times \left(\mathbb{E} \left[K_2^j \left(\frac{U_s - u}{h_2} \right) K_2^j \left(\frac{U_t - u}{h_2} \right) \right] - \mathbb{E} \left[K_2^j \left(\frac{U_s - u}{h_2} \right) \right] \mathbb{E} \left[K_2^j \left(\frac{U_t - u}{h_2} \right) \right] \right) \\
\leq & \frac{C p(u) h_2^2}{n^3 h_1^2 h_2^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[\tilde{X}_s^2 \tilde{X}_t^2 K_1^i \left(\frac{Z_s - z}{h_1} \right) K_1^i \left(\frac{Z_t - z}{h_1} \right) \right] \alpha_u(t - s) \\
= & \frac{C p(u) h_2^2}{n^3 h_1^2 h_2^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[\tilde{X}_s^2 \tilde{X}_t^2 \mathbb{E} \left(K_1^i \left(\frac{Z_s - z}{h_1} \right) K_1^i \left(\frac{Z_t - z}{h_1} \right) | (X_s, X_t) \right) \right] \alpha_u(t - s) \\
\leq & \frac{C p(u) h_1^2 h_2^2}{n^3 h_1^2 h_2^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{s}{\sqrt{s}} \frac{t}{\sqrt{t-s}} \alpha_u(t - s) = o(1), \tag{7.4}
\end{aligned}$$

where we have used Assumptions A.1(i)(ii) and A.2(ii)(iii).

Equations (7.3) and (7.4) imply

$$\begin{aligned}
& \frac{1}{(\sqrt{nn}h_1h_2)^2} \mathbb{E} \left[\sum_{t=1}^n \tilde{X}_t^2 \Lambda_{ij}(U_t; Z_t, X_t) \right]^2 = \frac{1}{(\sqrt{nn}h_1h_2)^2} \sum_{s=1}^n \mathbb{E} \left[\tilde{X}_s^2 \Lambda_{ij}(U_s; Z_s, X_s) \right]^2 \\
& + 2 \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[\tilde{X}_s^2 \tilde{X}_t^2 \mathbb{E} (\Lambda_{ij}(U_s; Z_s, X_s) \Lambda_{ij}(U_t; Z_t, X_t) | (Z_s, X_s; Z_t, X_t)) \right] \\
= & \frac{1}{(\sqrt{nn}h_1h_2)^2} \sum_{t=1}^n \mathbb{E} \left(\tilde{X}_t^4 K_1^{2i} \left(\frac{Z_t - z}{h_1} \right) \mathbb{E} [\Delta_j^2(U_t; Z_t, X_t)] \right) \\
& + \frac{2}{(\sqrt{nn}h_1h_2)^2} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E} \left[\tilde{X}_s^2 \tilde{X}_t^2 K_1^i \left(\frac{Z_s - z}{h_1} \right) K_1^i \left(\frac{Z_t - z}{h_1} \right) \right. \\
& \times \mathbb{E} (\Delta_j(U_t; Z_t, X_t) \Delta_j(U_s; Z_s, X_s) | (Z_s, X_s; Z_t, X_t)) \Big] = o(1). \tag{7.5}
\end{aligned}$$

Lemma A.5, along with (7.1)–(7.5), implies that as $n \rightarrow \infty$

$$\begin{aligned}
& \frac{1}{\sqrt{nn}h_1h_2} \sum_{t=1}^n X_t X_t' K_1^2 \left(\frac{Z_t - z}{h_1} \right) K_2^2 \left(\frac{U_t - u}{h_2} \right) \\
\rightarrow & \int_{-\infty}^{\infty} K_1^2(u) du \cdot \left(\int K_2^2(v) dv \right) \cdot \int_0^1 B_\mu(r) B_\mu(r)' dL_{B_z}(r, 0), \tag{7.6}
\end{aligned}$$

where we have used the fact that $\frac{1}{h_2} \mathbb{E} \left[K_2^2 \left(\frac{U_t - u}{h_2} \right) | X_t = x, Z_t = z \right] \rightarrow p(u) \cdot \int K_2^2(v) dv$ as $t \rightarrow \infty$ in the same way as in the proof of Lemma A.4 above.

Recall that $J = (J_1, \dots, J_{d_\mu})'$ is any vector of real numbers satisfying $J'J = 1$, and define

$\tilde{X}_t = J'X_t$. Equation (7.6) then implies as $n \rightarrow \infty$

$$\begin{aligned} & \frac{1}{\sqrt{nn}h_1h_2} \sum_{t=1}^n \tilde{X}_t^2 K_1^2 \left(\frac{Z_t - z}{h_1} \right) K_2^2 \left(\frac{U_t - u}{h_2} \right) \\ \rightarrow & \int_{-\infty}^{\infty} p(u) \left(\int_{-\infty}^{\infty} K_1^2(u) du \right) \cdot \left(\int K_2^2(v) dv \right) \cdot \int_0^1 \tilde{B}_\mu^2(r) dL_{B_z}(r, 0), \end{aligned} \quad (7.7)$$

where $\tilde{B}_\mu(r)$ is defined in the same way as B_μ when μ_t is replaced by $J'\mu_t$. To complete the proof of Theorem 3.3, using the Cramér-Wold device, it suffices to show that as $n \rightarrow \infty$

$$\begin{aligned} & \left(\sum_{t=1}^n \tilde{X}_t^2 K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) \right)^{-\frac{1}{2}} \left(\sum_{t=1}^n \tilde{X}_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right) e_t \right) \\ \rightarrow & \int_{-\infty}^{\infty} N(0, \Sigma_{xz}), \end{aligned} \quad (7.8)$$

where $\Sigma_{xz} = I_d \cdot \sigma_e^2 \int_{-\infty}^{\infty} K_1^2(u) du \cdot \int_{-\infty}^{\infty} K_2^2(v) dv$. We can apply Lemma A.6 to derive (7.8). Let $\mathbf{l} = (l_1, \dots, l_{d_{u\mu\nu}})'$ be any vector of real numbers satisfying $\mathbf{l}'\mathbf{1} = 1$, and assume without loss of generality that $d_u = 1$. We then match the notation of Lemma A.6 in the following correspondence:

$$\begin{aligned} \eta_{n,t} &= \mathbf{l}'\xi_t, \quad \xi_{nt} = \mathbf{l}'\xi_{1-t}, \quad \varepsilon_{n,t+1} = e_t, \\ u_{t,n} &= f_n(e_1, \dots, e_n; \eta_{n1}, \dots, \eta_{nm}; \xi_{n1}, \xi_{n2}, \dots) \\ &= \frac{1}{\sqrt{\sqrt{n}h_1h_2} \sqrt{n}} \bar{X}_t K_1 \left(\frac{Z_t - z}{h_1} \right) K_2 \left(\frac{U_t - u}{h_2} \right), \end{aligned} \quad (7.9)$$

and let $\mathcal{F}_{n,t} = \sigma(e_t, \dots, e_1; \xi_t, \xi_{t-1}, \dots)$ be generated by $\{(e_i, \xi_j) : 1 \leq i \leq t, -\infty < j \leq t\}$, where $\bar{X}_t = \mathbf{l}'X_t$. Assumption A.1 then implies that Assumptions (i) and (ii) of Lemma A.6 are trivially satisfied. Equation (7.7) shows that Assumption (iv) of Lemma A.6 is also satisfied. Meanwhile, by the Kolmogorov inequality for any small $\delta_u > 0$

$$\begin{aligned} P \left(\max_{1 \leq t \leq n} |u_{t,n}| > \delta_u \right) &\leq \frac{1}{\sqrt{n}h_1h_2 \delta_u^2} \max_{1 \leq t \leq n} \mathbb{E} \left[\frac{\bar{X}_t^2}{n} \cdot K_1^2 \left(\frac{Z_t - z}{h_1} \right) K_2^2 \left(\frac{U_t - u}{h_2} \right) \right] \\ &\leq \frac{C}{\sqrt{n}h_1h_2} = o(1), \end{aligned} \quad (7.10)$$

which implies $\max_{1 \leq t \leq n} |u_{t,n}| = o_P(1)$. This, along with Assumption A.1(iii), implies that Assumption (iii) of Lemma A.6 is satisfied. Thus Lemma A.6 can be applied to deduce (7.8). This completes the proof of Theorem 3.2.

9.3 Proof of Theorem 4.1

Recall $\bar{\beta}_1(z)$ and $\bar{\beta}_2(z)$ as defined in (4.2) and (4.3). Note that

$$\begin{aligned}\widehat{\beta}_1(z) - \bar{\beta}_1(z) &= \sum_{s=1}^n \left(\int w_{ns}(z, u) dR_1(u) \right) e_s + \sum_{s=1}^n \left(\int w_{ns}(z, u) dR_1(u) \right) \beta_1(Z_s) - \beta_1(z) \\ &+ \int \left(\sum_{s=1}^n w_{ns}(z, u) (\beta_2(U_s) - \beta_2(u)) \right) dR_1(u),\end{aligned}\quad (7.11)$$

$$\begin{aligned}\widehat{\beta}_2(u) - \bar{\beta}_2(u) &= \sum_{s=1}^n \left(\int w_{ns}(z, u) dR_2(z) \right) e_s + \sum_{s=1}^n \left(\int w_{ns}(z, u) dR_2(z) \right) \beta_2(U_s) - \beta_2(u) \\ &+ \int \left(\sum_{s=1}^n w_{ns}(z, u) (\beta_1(Z_s) - \beta_1(z)) \right) dR_2(z).\end{aligned}\quad (7.12)$$

The proof of Theorem 4.1(i) follows in the same manner as that of Theorem 4.1(ii). So we provide the proof of Theorem 4.1(ii) here.

Observe that

$$\begin{aligned}w_{ns}(z, u) &= \frac{\frac{1}{\sqrt{nh_1h_2}} K_1 \left(\frac{Z_s - z}{h_1} \right) K_2 \left(\frac{U_s - u}{h_2} \right)}{\frac{1}{\sqrt{nh_1h_2}} \sum_{l=1}^n K_1 \left(\frac{Z_l - z}{h_1} \right) K_2 \left(\frac{U_l - u}{h_2} \right)} \\ &= \frac{\frac{1}{\sqrt{nh_1h_2}} K_1 \left(\frac{Z_s - z}{h_1} \right) \mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_s \right]}{\frac{1}{\sqrt{nh_1h_2}} \sum_{l=1}^n K_1 \left(\frac{Z_l - z}{h_1} \right) K_2 \left(\frac{U_l - u}{h_2} \right)} \\ &+ \frac{\frac{1}{\sqrt{nh_1h_2}} K_1 \left(\frac{Z_s - z}{h_1} \right) \left(K_2 \left(\frac{U_s - u}{h_2} \right) - \mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_s \right] \right)}{\frac{1}{\sqrt{nh_1h_2}} \sum_{l=1}^n K_1 \left(\frac{Z_l - z}{h_1} \right) K_2 \left(\frac{U_l - u}{h_2} \right)}.\end{aligned}\quad (7.13)$$

Similarly to the proofs of Lemmas A.3 and A.4, we have for $i, j = 1, 2$ and as $n \rightarrow \infty$

$$\begin{aligned}&\frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1 \left(\frac{Z_s - z}{h_1} \right) K_2 \left(\frac{U_s - u}{h_2} \right) e_s \\ &= \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1 \left(\frac{Z_s - z}{h_1} \right) \mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_s \right] e_s \\ &+ \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1 \left(\frac{Z_s - z}{h_1} \right) \left(K_2 \left(\frac{U_s - u}{h_2} \right) - \mathbb{E} \left[K_2 \left(\frac{U_s - u}{h_2} \right) | Z_s \right] \right) e_s \\ &= \frac{p(u)(1 + o_P(1))}{\sqrt{nh_1}} \sum_{s=1}^n K_1 \left(\frac{Z_s - z}{h_1} \right) e_s,\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1^i \left(\frac{Z_s - z}{h_1} \right) K_2^j \left(\frac{U_s - u}{h_2} \right) \\
&= \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1^i \left(\frac{Z_s - z}{h_1} \right) \mathbb{E} \left[K_2^j \left(\frac{U_s - u}{h_2} \right) | Z_s \right] \\
&+ \frac{1}{\sqrt{nh_1h_2}} \sum_{s=1}^n K_1^i \left(\frac{Z_s - z}{h_1} \right) \left(K_2^j \left(\frac{U_s - u}{h_2} \right) - \mathbb{E} \left[K_2^j \left(\frac{U_s - u}{h_2} \right) | Z_s \right] \right) \\
&= \frac{p(u)(1 + o_P(1))}{\sqrt{nh_1}} \sum_{s=1}^n K_1^i \left(\frac{Z_s - z}{h_1} \right) \cdot \int_{-\infty}^{\infty} K_2^j(v) dv. \tag{7.14}
\end{aligned}$$

In view of (7.13) and (7.14), in order to show Theorem 4.1(ii) with $\{w_{ns}(z, u)\}$, it suffices to show the result holds with $w_{ns}(z) = \frac{K_1\left(\frac{Z_s-z}{h_1}\right)}{\sum_{l=1}^n K_1\left(\frac{Z_l-z}{h_1}\right)}$. Let $b_{ns} = \int_{-\infty}^{\infty} w_{ns}(z) dR_2(z)$, $b_n = \sqrt{\sum_{s=1}^n b_{ns}^2}$ and $c_{ns} = \frac{b_{ns}}{b_n}$. Thus, $\sum_{s=1}^n c_{ns}^2 \equiv 1$ and, given (7.12)–(7.14), to prove Theorem 4.1(ii) it suffices to show that as $n \rightarrow \infty$

$$b_n^{-1} \Delta_n \rightarrow_D N(0, \sigma_e^2), \tag{7.15}$$

$$b_n^{-1} \delta_{in} \rightarrow_P 0 \text{ for } i = 1, 2, \tag{7.16}$$

where $\Delta_n = \sum_{s=1}^n \left(\int w_{ns}(z) dR_2(z) \right) e_s = \int \left(\sum_{s=1}^n w_{ns}(z) e_s \right) dR_2(z)$,

$$\begin{aligned}
\delta_{1n} &= \sum_{s=1}^n \left(\int w_{ns}(z) dR_2(z) \right) \beta_2(U_s) - \beta_2(u) \text{ and} \\
\delta_{2n} &= \int \left(\sum_{s=1}^n w_{ns}(z, u) (\beta_1(Z_s) - \beta_1(z)) \right) dR_2(z). \tag{7.17}
\end{aligned}$$

Since the partial sum involved in Δ_n is of the form $\sum_{s=1}^n K_1\left(\frac{Z_s-z}{h_1}\right) e_s$, under Assumptions A.1(i)–(iii), A.2 and A.3, the proof of (7.15) then follows similarly to that of (7.8) (with $X_t \equiv 1$). To prove (7.16), we first evaluate the order of b_n^2 . Observe that

$$\begin{aligned}
b_n^2 &= \sum_{s=1}^n b_{ns}^2 = \sum_{s=1}^n \int \int w_{ns}(x_1) w_{ns}(x_2) dR_2(x_1) dR_2(x_2) \\
&\equiv \int \int R_n(x_1, x_2) dR_2(x_1) dR_2(x_2), \tag{7.18}
\end{aligned}$$

where $R_n(x_1, x_2) = \frac{\sum_{s=1}^n K_1\left(\frac{Z_s-x_1}{h_1}\right) K_1\left(\frac{Z_s-x_2}{h_1}\right)}{nh_1^2 \hat{p}_1(x_1) \hat{p}_1(x_2)}$ with $\hat{p}_1(z) = \frac{1}{\sqrt{nh_1}} \sum_{s=1}^n K_1\left(\frac{Z_s-z}{h_1}\right)$.

Let $L_1(x, y) = K_1(x)K_1(x+y)$. By the same reasoning as in the proof of Lemma A.2 with $f_y(x) = L_1(x, y)$, we have as $n \rightarrow \infty$

$$\begin{aligned}
& \frac{1}{\sqrt{nh_1}} \sum_{s=1}^n K_1\left(\frac{Z_s - x_1}{h_1}\right) K_1\left(\frac{Z_s - x_1}{h_1} + \frac{x_1 - x_2}{h_1}\right) = \frac{1}{\sqrt{nh_1}} \sum_{s=1}^n L_1\left(\frac{Z_s - x_1}{h_1}, \frac{x_1 - x_2}{h_1}\right) \\
&= (1 + o_P(1)) L_{B_\nu}(1, 0) \int_{-\infty}^{\infty} L_1\left(u, \frac{x_1 - x_2}{h_1}\right) du \\
&= (1 + o_P(1)) L_{B_\nu}(1, 0) K_3\left(\frac{x_1 - x_2}{h_1}\right), \tag{7.19}
\end{aligned}$$

where $K_3(u) = \int K_1(u+v)K_1(v)dv$. Equation (7.19) implies as $n \rightarrow \infty$

$$\begin{aligned}
R_n(x_1, x_2) &= \frac{\sum_{s=1}^n K_1\left(\frac{Z_s-x_1}{h_1}\right) K_1\left(\frac{Z_s-x_2}{h_1}\right)}{nh_1^2 \hat{p}(x_1) \hat{p}(x_2)} \\
&= \frac{1}{\sqrt{nh_1}} \frac{\frac{1}{\sqrt{nh_1}} \sum_{s=1}^n K_1\left(\frac{Z_s-x_1}{h_1}\right) K_1\left(\frac{Z_s-x_2}{h_1}\right)}{\hat{p}_1(x_1) \hat{p}_1(x_2)} \\
&= \frac{1}{\sqrt{nh_1}} \frac{\frac{1}{\sqrt{nh_1}} \sum_{s=1}^n K_1\left(\frac{Z_s-x_1}{h_1}\right) K_1\left(\frac{Z_s-x_1}{h_1} + \frac{x_1-x_2}{h_1}\right)}{\hat{p}_1(x_1) \hat{p}_1(x_2)} \\
&= \frac{1}{\sqrt{nh_1}} \frac{(1 + o_P(1)) K_3\left(\frac{x_1-x_2}{h_1}\right)}{L_{B_\nu}(1, 0)}. \tag{7.20}
\end{aligned}$$

Equations (7.18) and (7.20) imply that

$$\begin{aligned}
b_n^2 &= \frac{(1 + o_P(1))}{\sqrt{nh_1}} \frac{1}{L_B(1, 0)} \int \int K_3\left(\frac{x_1-x_2}{h_1}\right) r_2(x_1) r_2(x_2) dx_1 dx_2 \\
&= \frac{(1 + o_P(1))}{\sqrt{nh_1}} \frac{1}{L_{B_\nu}(1, 0)} \int r_2^2(x) dx. \tag{7.21}
\end{aligned}$$

Meanwhile, using Lemma A.2 again and applying Taylor expansions, we have

$$\begin{aligned}
|d_{2n}| &= \left| \int \left(\sum_{s=1}^n w_{ns}(z) (\beta_1(Z_s) - \beta_1(z)) \right) dR_2(z) \right| \\
&\leq \left| (1 + o_P(1)) \int \left(\sum_{s=1}^n w_{ns}(x) \beta^{(1)}(z) (Z_s - z) \right) dR_2(z) \right| \\
&= (1 + o_P(1)) \cdot h_1 \\
&\quad \times \int \left(\frac{\sum_{s=1}^n \left| K_1\left(\frac{Z_s-z}{h_1}\right) \left(\frac{Z_s-z}{h_1}\right) \right|}{\sum_{l=1}^n K_1\left(\frac{Z_l-z}{h_1}\right)} \right) |\beta_1^{(1)}(z)| dR_2(z) \\
&= (1 + o_P(1)) \cdot h_1 \cdot \int \left(\int |u K_1(u)| du \right) |\beta_1^{(1)}(z)| dR_2(z) = o_P(h_1), \tag{7.22}
\end{aligned}$$

where Assumptions A.2(iii) and A.4(iii) have been used. Similarly, under Assumptions A.2(iii) and A.4(iii) we have $d_{1n} = o_P(h_2)$. The proof of Theorem 4.1(ii) is therefore completed.

9.4 Proof of Theorem 4.2

(i) By definition

$$\hat{\beta} - \beta = \left(\sum_{t=1}^n X_{1t} X'_{1t} \right)^{-1} \cdot \sum_{t=1}^n X_{1t} e_t + \left(\sum_{t=1}^n X_{1t} X'_{1t} \right)^{-1} \cdot \sum_{t=1}^n X_{1t} \beta_2(Z_t). \tag{7.23}$$

In the same spirit as the proof of Lemma A.5, we have as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n X_{1t} \beta_1(Z_t) = \frac{d_n}{n} \sum_{t=1}^n (x_{t1n}) \beta_1(d_n z_{tn}) \rightarrow_D \int_0^1 B_\mu(r) dL_{B_\nu}(r, 0) \cdot \int_{-\infty}^{\infty} \beta_1(z) dz, \tag{7.24}$$

where $x_{t1n} = \frac{X_{1t}}{\sqrt{n}}$, $z_{tn} = \frac{Z_t}{\sqrt{n}}$ and $d_n = \sqrt{n}$.

To complete the proof of Theorem 4.2(i), in view of (7.11), it suffices to show that as $n \rightarrow \infty$

$$\frac{1}{n^2} \sum_{t=1}^n X_{1t} X'_{1t} = \frac{1}{n} \sum_{t=1}^n x_{t1n} x'_{t1n} \Rightarrow_{\mathcal{D}} \int_0^1 B_\mu(r) B_\mu(r)' dr, \quad (7.25)$$

$$\frac{1}{n} \sum_{t=1}^n X_{1t} e_t = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_{t1n} e_t \Rightarrow_{\mathcal{D}} \int_0^1 B_\mu(r) dB_e(r), \quad (7.26)$$

which follow by standard weak convergence arguments. Results (7.23)–(7.26) complete the proof of Theorem 4.2(i).

(ii) By definition of $\widehat{\beta}_1(z)$ we have

$$\begin{aligned} \widehat{\beta}_1(z) - \beta_1(z) &= \sum_{t=1}^n w_{nt}(z) e_t + \sum_{t=1}^n w_{nt}(z) (\beta_1(Z_t) - \beta_1(z)) \\ &+ \sum_{t=1}^n w_{nt}(z) X'_{1t} (\beta - \widehat{\beta}). \end{aligned} \quad (7.27)$$

In view of the proof of Theorem 3.1 of Wang and Phillips (2009a), in order to complete the proof of Theorem 4.3(ii), it suffices to evaluate the last term of (7.27). Similar to the derivation of (7.24), we have as $n \rightarrow \infty$

$$\begin{aligned} \frac{1}{nh_1} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) X_{1t} &= \frac{c_n}{n} \sum_{t=1}^n K_1 \left(c_n \left(z_{tn} - \frac{z}{\sqrt{n}} \right) \right) x_{t1n} \\ &\rightarrow_D \int_0^1 B_\mu(r) dL_{B_\nu}(r, 0), \end{aligned} \quad (7.28)$$

where $c_n = \frac{\sqrt{n}}{h_1}$. This result, along with the conclusion of Theorem 4.2(i), leads to

$$\sqrt{\sqrt{n}h_1} \left(\sum_{t=1}^n w_{nt}(z) X'_{1t} (\beta - \widehat{\beta}) \right) = O_P \left(\sqrt{\sqrt{n}h_1} \sqrt{n} n^{-1} \right) = o_P(1), \quad (7.29)$$

which completes the proof of Theorem 4.2(ii).

9.5 Proof of Theorem 4.3

(i) From the definition of $\widehat{\theta}$

$$\widehat{\theta} - \theta = \left(\sum_{t=1}^n Z_t Z'_t \right)^{-1} \cdot \sum_{t=1}^n Z_t e_t + \left(\sum_{t=1}^n Z_t Z'_t \right)^{-1} \cdot \sum_{t=1}^n Z_t \beta_1(Z_t). \quad (7.30)$$

Standard derivations and limit theory imply the following as $n \rightarrow \infty$

$$\frac{1}{n^2} \sum_{t=1}^n Z_t Z_t' = \frac{1}{n} \sum_{t=1}^n z_{tn} z_{tn}' \Rightarrow_D \int_0^1 B_\nu(r) B_\nu(r)' dr, \quad (7.31)$$

$$\frac{1}{n} \sum_{t=1}^n Z_t e_t = \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{tn} e_t \Rightarrow_D \int_0^1 B_\nu(r) dB_e(r),$$

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t \beta_1(Z_t) &= \frac{d_n}{n} \sum_{t=1}^n (d_n z_{tn}) \beta_1(d_n z_{tn}) \\ &\rightarrow_D L_{B_\nu}(1, 0) \cdot \int_{-\infty}^{\infty} z \beta_1(z) dz, \end{aligned} \quad (7.32)$$

which in turn imply that as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n Z_t \beta_1(Z_t) = \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t \beta_1(Z_t) \rightarrow_P 0, \quad (7.33)$$

where $z_{tn} = \frac{Z_t}{\sqrt{n}}$ and $d_n = \sqrt{n}$. The proof of Theorem 4.3(i) follows immediately from (7.30) and (7.31).

(ii) By definition of $\tilde{\beta}_1(z)$ we have the decomposition

$$\begin{aligned} \tilde{\beta}_1(z) - \beta_1(z) &= \sum_{t=1}^n w_{nt}(z) e_t + \sum_{t=1}^n w_{nt}(z) (\beta_1(Z_t) - \beta_1(z)) \\ &+ \sum_{t=1}^n w_{nt}(z) Z_t (\theta - \hat{\theta}). \end{aligned} \quad (7.34)$$

Similarly to the proof of Theorem 4.2(ii), in order to complete the proof of Theorem 4.3(ii), it suffices to deal with the last term of (7.34). In view of the proof of (7.26), we have as $n \rightarrow \infty$

$$\begin{aligned} \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) Z_t &= \frac{h_1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \left(\frac{Z_t - z}{h_1} \right) \\ &+ \frac{z}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \equiv h_1 J_{1n} + J_{2n}, \end{aligned} \quad (7.35)$$

where by Lemma A.2 as $n \rightarrow \infty$

$$\begin{aligned} |J_{1n}| &= \left| \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \left(\frac{Z_t - z}{h_1} \right) \right| \leq \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n \left| K_1 \left(\frac{Z_t - z}{h_1} \right) \left(\frac{Z_t - z}{h_1} \right) \right| \\ &\rightarrow_D \int |u K_1(u)| du \cdot L_{B_\nu}(1, 0), \end{aligned} \quad (7.36)$$

$$J_{2n} = \frac{1}{\sqrt{nh_1}} \sum_{t=1}^n K_1 \left(\frac{Z_t - z}{h_1} \right) \rightarrow_D L_{B_\nu}(1, 0), \quad (7.37)$$

which, along with the conclusion of Theorem 4.3(i), leads to

$$\sqrt{\sqrt{nh_1}} \left(\sum_{t=1}^n w_{nt}(z) Z_t (\theta - \hat{\theta}) \right) = O_P \left(\sqrt{\sqrt{nh_1}} n^{-1} \right) = o_P(1), \quad (7.38)$$

which completes the proof of Theorem 4.3(ii).

9.6 Proof of Theorem 4.4

Equations (4.25) and (4.26) imply

$$\hat{\gamma} - \gamma = \left(\sum_{t=1}^n \tilde{Z}_t^2 \right)^{-1} \sum_{t=1}^n \tilde{Z}_t \tilde{\eta}_t, \quad (7.39)$$

where $\tilde{Z}_t = Z_t - \frac{1}{n} \sum_{s=1}^n Z_s$ and $\tilde{\eta}_t = \eta_t - \frac{1}{n} \sum_{s=1}^n \eta_s$.

By standard weak convergence arguments, using Assumption A.1(iv), we have as $n \rightarrow \infty$

$$\begin{aligned} \frac{1}{n^2} \sum_{t=1}^n \tilde{Z}_t^2 &= \frac{1}{n} \sum_{t=1}^n \left(\frac{Z_t}{\sqrt{n}} - \frac{1}{n} \sum_{s=1}^n \frac{Z_s}{\sqrt{n}} \right)^2 \Rightarrow_{\mathcal{D}} \int_0^1 \left(B_\nu(r) - \int_0^1 B_{\nu u}(s) ds \right)^2 dr, \quad (7.40) \\ \frac{1}{n} \sum_{t=1}^n \tilde{Z}_t \tilde{\eta}_t &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{Z_t}{\sqrt{n}} - \frac{1}{n} \sum_{s=1}^n \frac{Z_s}{\sqrt{n}} \right) \left(\eta_t - \frac{1}{n} \sum_{s=1}^n \eta_s \right) \\ &\Rightarrow_{\mathcal{D}} \int_0^1 \left(B_\nu(r) - \int_0^1 B_{\nu u}(s) ds \right) dB_\eta(r). \quad (7.41) \end{aligned}$$

This completes the proof of Theorem 4.4(i). In consequence, as $n \rightarrow \infty$

$$\hat{\gamma} - \gamma = O_P(n^{-1}). \quad (7.42)$$

The proof of Theorem 4.4(ii) follows from (7.42), Assumption 4.4(ii) and the decomposition

$$\begin{aligned} \hat{\beta}_2(u) - \beta_2(u) &= \sum_{t=1}^n W_{nt}(u) (Y_t - Z_t \hat{\gamma}) - \beta_2(u) = \sum_{t=1}^n W_{nt}(u) e_t \\ &+ \sum_{t=1}^n W_{nt}(u) Z_t (\gamma - \hat{\gamma}) + \sum_{t=1}^n W_{nt}(u) (\beta_2(U_t) - \beta_2(u)), \quad (7.43) \end{aligned}$$

where we may use the following order arguments:

$$\begin{aligned} \left| \sum_{t=2}^n K_2 \left(\frac{U_t - u}{h_2} \right) Z_t \right| &= \sqrt{\sum_{t=2}^n K_2^2 \left(\frac{U_t - u}{h_2} \right)} \cdot \sqrt{\sum_{t=2}^n Z_t^2} = O_P(\sqrt{n^3 h_2}), \\ \sum_{t=1}^n W_{nt}(u) (\beta_2(U_t) - \beta_2(u)) &= \frac{h_2^2 \cdot \beta_2^{(2)}(u)}{2} (1 + o_P(1)), \quad (7.44) \end{aligned}$$

which, along with (7.42), complete the proof of Theorem 4.4(ii).

References

- CAI, Z., J. FAN, AND Q. YAO (2000): “Functional–Coefficient Regression Models for Nonlinear Time Series,” *Journal of the American Statistical Association*, 95, 941–956.
- CAI, Z., Q. LI, AND J. PARK (2009): “Functional–coefficient Models for Nonstationary Time Series Data,” *Journal of Econometrics*, 148, 101–113.

- CAMPBELL, J. Y., A. LO AND A. W. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press, New Jersey.
- CAMPBELL, J. Y., AND N. G. MANKIW (1990): “Permanent Income, Current Income and Consumption,” *Journal of Business and Economic Statistics*, 8, 265–279.
- CHAN, N., AND Q. WANG (2013): “Uniform Convergence for Non-parametric Estimators with Nonstationary data”, *Econometric Theory*, (forthcoming)
- FAFF, R., AND R. BROOKS (1998): “Time-Varying Beta Risk for Australian Industry Portfolios: an Exploratory Analysis,” *Journal of Business Finance & Accounting*, 25, 721–745.
- FAN, J., AND Q. YAO (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.
- FLORENS, J. P., J. JOHANNES, AND S. VAN BELLEGEM (2012): “Instrumental Regression in Partially Linear Models,” *Econometrics Journal*, 15, 304–324.
- GAO, J. (2007): *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. London: Chapman & Hall/CRC.
- GAO, J., M. L. KING, Z. LU, AND D. TJØSTHEIM (2009a): “Specification Testing in Nonstationary Time Series Autoregression,” *Annals of Statistics*, 37, 3893–3928.
- GAO, J., M. L. KING, Z. LU, AND D. TJØSTHEIM (2009b): “Nonparametric Specification Testing for Nonlinear Time Series with Nonstationarity,” *Econometric Theory*, 25, 1869–1892.
- GAO, J., AND P. C. B. PHILLIPS (2012): “Trending Time Series Models with Nonparametric and Semiparametric Cointegration,” Working paper at <http://jitigao.com/working-papers.html>.
- GAO, J., AND P. C. B. PHILLIPS (2013): “Semiparametric Estimation in Triangular Simultaneous Equations with Nonstationarity,” *Journal of Econometrics*, 176, 59–79.
- GAO, J., TJØSTHEIM, D., AND YIN, J. (2012): “Specification between Parametric and Nonparametric Cointegration,” Working paper at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2012/wp18-12.pdf>.
- GYLFASON, T. (1981): “Interest Rates, Inflation and the Aggregate Consumption Function,” *Review of Economics and Statistics*, 63, 233–245.
- HAHM, J. AND D. G. STEIGERWALD (1999): “Consumption Adjustment Under Time-Varying Income Uncertainty,” *Review of Economics and Statistics* 81, 32–40.
- HALL, P. AND C. C. HEDYE (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press.
- HALL, R. E. (1978): “Stochastic Implications of the Life Cycle-permanent Income Hypothesis: Theory and Evidence,” *Journal Political Economy*, 86, 971–987.
- HAMILTON, J. D. (1994): *Time Series Analysis*. New Jersey: Princeton University.
- KARLSEN, H. A., AND D. TJØSTHEIM (2001): “Nonparametric Estimation in Null Recurrent Time Series,” *Annals of Statistics*, 29, 372–416.

- KARLSEN, H. A., T. MYKELBUST, AND D. TJØSTHEIM (2007): “Nonparametric Estimation in a Nonlinear Cointegration type Model,” *Annals of Statistics*, 35, 252–299.
- LI, Q., AND J. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.
- MUELLBAUER, J., AND R. LATTIMORE (1999): “The Consumption Function: a Theoretical and Empirical Overview,” In *Handbook of Applied Econometrics Volume I: Macroeconomics*, ed. by M. Hashem Pesaran and M. R. Wickens. Oxford: Blackwell Publishers Ltd, pp. 221–311.
- MÜLLER, U. AND P-E PETALAS (2010): “Efficient Estimation of the Parameter Path in Unstable Time Series Models”, *Review of Economic Studies*, 77, 1508-1539.
- NEWKEY, W. K., AND J. POWELL (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71, 1565–1578.
- NEWKEY, W. K., J. POWELL, AND F. VELLA (1999): “Nonparametric Estimation of Triangular Simultaneous Equations Models,” *Econometrica*, 67, 565–603.
- PARK, J., AND P. C. B. PHILLIPS (1988): “Statistical Inference in Regressions With Integrated Processes: Part 1,” *Econometric Theory*, 4, 468–497.
- PARK, J., AND P. C. B. PHILLIPS (1989): “Statistical Inference in Regressions With Integrated Processes: Part 2,” *Econometric Theory*, 5, 95-131.
- PARK, J., AND P. C. B. PHILLIPS (1999): “Asymptotics for Nonlinear Transformations of Integrated Time Series,” *Econometric Theory*, 15, 269–298.
- PARK, J., AND P. C. B. PHILLIPS (2001): “Nonlinear Regressions with Integrated Time Series,” *Econometrica*, 69, 117-162.
- PHAM, T. D., AND L. T. TRAN (1985): “Some Mixing Properties of Time Series Models,” *Stochastic Processes and Their Applications*, 19, 297–303.
- PHILLIPS, P. C. B. (2001): “Trending Time Series and Macroeconomic Activity: Some Present and Future Challengers,” *Journal of Econometrics*, 100, 21–27.
- PHILLIPS, P. C. B. (2009): “Local Limit Theory and Spurious Regressions,” *Econometric Theory*, 25, 1466–1497.
- PHILLIPS, P. C. B. (2010): “The Mysteries of Trend”, *Macroeconomic Review*, Special Issue C, 82–89.
- PHILLIPS, P. C. B. AND T. MAGDALINOS (2009): “Unit Root and Cointegrating Limit Theory when the Initialization is in the Infinite Past,” *Econometric Theory*, 25, 1682–1715.
- PHILLIPS, P. C. B., AND P. PERRON (1988): “Testing for a Unit–Root in Time Series Regression,” *Biometrika*, 75, 335–346.
- PHILLIPS, P. C. B., AND D. SUL (2007): “Transition Modeling and Econometric Convergence Tests”, *Econometrica*, 75, 1771-1855.
- PHILLIPS, P. C. B., AND J. YU (2011): “Dating and Timeline of Financial Bubbles during the Subprime Crisis”, *Quantitative Economics*, 2, 455–491.

- REVUZ, D., AND M. YOR (1999): *Continuous Martingale and Brownian Motion*. 3rd Edition. New York: Springer-Verlag.
- ROBINSON, P. M. (1988): “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 56, 931–954.
- ROBINSON, P. M. (1989): “Nonparametric Estimation of Time-varying Parameters,” *Statistical Analysis and Forecasting of Economic Structural Change*. Hackl, P. (Ed.). Springer, Berlin, pp. 164–253.
- SU, L., AND A. ULLAH (2008): “Local Polynomial Estimation of Nonparametric Simultaneous Equations Models,” *Journal of Econometrics*, 144, 193–218.
- SUN, Y., CAI, Z., AND LI, Q. (2013): “Semiparametric Functional Coefficient Models with Integrated Covariates,” *Econometric Theory*, available on CJO2013. doi:10.1017/S0266466612000710.
- TERÄSVIRTA, T., D. TJØSTHEIM, AND C. W. J. GRANGER (2010): *Modelling Nonlinear Economic Time Series*. Oxford: Oxford University Press.
- WANG, Q. Y. (2012): “Martingale Limit Theorems Revisited and Nonlinear Cointegrating Regression,” Available at <http://www.maths.usyd.edu.au/u/pubs/publist/preprints/2011/wang-3.html>.
- WANG, Q. Y., AND P. C. B. PHILLIPS (2009a): “Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression,” *Econometric Theory*, 25, 710–738.
- WANG, Q. Y., AND P. C. B. PHILLIPS (2009b): “Structural Nonparametric Cointegrating Regression,” *Econometrica*, 77, 1901–1948.
- WANG, Q. Y., AND P. C. B. PHILLIPS (2012): “A Specification Test for Nonlinear Nonstationary Models”, *Annals of Statistics*, 40, 727–758.
- WITHERS, C. S (1981): “Conditions for Linear Processes to be Strong Mixing,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 57, 509–534.
- XIAO, Z. (2009): “Functional-coefficient Cointegration Models,” *Journal of Econometrics*, 152, 81–92.