Journal of Econometrics 164 (2011) 252–267

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Measuring correlations of integrated but not cointegrated variables: A semiparametric approach[☆]

Yiguo Sun[a], Cheng Hsiao[b,c,*], Qi Li[d]

[a] Department of Economics, University of Guelph, Guelph, ON, N1G 2W1, Canada
[b] Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, USA
[c] Department of Economics and Finance, City University of Hong Kong, Hong Kong
[d] Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA

ARTICLE INFO

Article history:
Received 22 July 2010
Received in revised form
18 April 2011
Accepted 17 May 2011
Available online 12 June 2011

JEL classification:
C13
C14
C20

Keywords:
Integrated time series
Non-cointegration
Semiparametric varying coefficient models

ABSTRACT

Many macroeconomic and financial variables are integrated of order one (or $I(1)$) processes and are correlated with each other but not necessarily cointegrated. In this paper, we propose to use a semiparametric varying coefficient approach to model/capture such correlations. We propose two consistent estimators to study the dependence relationship among some integrated but not cointegrated time series variables. Simulations are used to examine the finite sample performances of the proposed estimators.

1. Introduction

In this paper we consider a semiparametric varying coefficient model as follows:

$$Y_t = X_t^{\top}\theta(Z_t) + u_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where X_t (a $d \times 1$ vector) and u_t (a scalar) are integrated series; i.e., $X_t = \sum_{s=1}^{t} \zeta_s + X_0$ and $u_t = \sum_{s=1}^{t} \epsilon_s + u_0$, and ζ_t and ϵ_t are stationary processes with zero means and finite long-run variances and satisfy some memory and moment conditions to be given in Section 3. Also, X_0 and u_0 are $O_p(1)$ random variables with $X_t \equiv 0$ and $u_t \equiv 0$ for $t < 0$. And, $\theta(\cdot)$ is a $d \times 1$ vector of smooth measurable and squared integrable functions of a scalar stationary variable Z_t.[1] When Y_t and X_t are cointegrated in a model either with a (parametric) nonlinear or with a (semiparametric) varying cointegration vector of functions of time or of some other stationary covariates, its limiting theories have been studied by Juhl (2005), Cai et al. (2009), and Xiao (2009). In this paper, we extend this literature by allowing the dependent variable Y_t to be not cointegrated with the $I(1)$ regressors X_t in the above semiparametric framework; that is, this paper considers the case that u_t is an integrated process such that model (1.1) depicts a semiparametric regression model that can be used to measure the correlation of integrated but not cointegrated variables.

Model (1.1) is of interest because many macro and financial time series are $I(1)$ and correlated, but not necessarily cointegrated. For instance, let Y be the exchange rates between two countries, X be the price indices and Z be the interest rate differentials of these two countries. Should purchasing power parity (PPP) hypothesis holds, u_t will be $I(0)$. However, empirical studies often fail to support the PPP hypothesis (e.g., Taylor and Taylor

☆ We gratefully acknowledge the valuable comments from two anonymous referees, an associate editor and Peter Robinson that have greatly improved our paper. We would also like to thank K.S. Chan for providing several references cited in our paper. Yiguo Sun acknowledges the financial support from the Social Science and Humanities Research Council of Canada (SSHRC) grant 410-2009-0109. Li's research is partially supported by the NSF of China (Project #: 70773005).

* Corresponding author at: Department of Economics, University of Southern California, Los Angeles, CA 90089-0253, USA. Tel.: +1 213 740 8335; fax: +1 213 740 8543.
E-mail address: chsiao@usc.edu (C. Hsiao).

[1] It is straightforward to generalize the results to a vector Z_t case. For expositional simplicity, we will only consider the scalar Z_t case in this paper, as allowing Z_t to be a high-dimensional variable will not shed extra insight in to our theory.

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.05.013

(2004)), but this does not mean that exchange rates are not related to the inflation rates. Similarly, one can consider Y and X as stock market indices and Z be the interest rate differential. With perfect information and arbitrage, the law of one price would imply $Y_t - X_t\theta(Z_t) = u_t$, where u_t is an $I(0)$ process. However, imperfect information and many domestic factors could also lead to stock market indices to move in different directions so that u_t might be an $I(1)$ process. Nevertheless, $\theta(Z_t)$ remains of interest in the investigation of the spill-over effects from the variation of X_t to that of Y_t. Similar examples can be found in the investigation of interest rate parity (e.g. Mark (2001)) or equity parity (e.g. Gavin (1989)).[2]

Another example that model (1.1) could be of interest is that $(Y_t, X_t^\top, W_t)$ are individually $I(1)$ processes and cointegrated so that

$$Y_t = X_t^\top\theta(Z_t) + W_t\delta + \epsilon_t = X_t^\top\theta(Z_t) + u_t,$$

where ϵ_t is $I(0)$, and $\left[1, -\theta(Z_t)^\top, \delta\right]$ are the cointegrating vectors. However, if W_t is not observable, then the composite error in model (1.1), $u_t = W_t\delta + \epsilon_t$, is an $I(1)$ process, and Y_t and X_t are not cointegrated. Therefore, the $I(1)$ error term u_t may be due to omitted variables and/or measurement errors. Indeed, many macroeconomic and financial data are measured with errors; e.g., Barnett and Serletis (1990) and Blundell and Stoker (2005), where with integrated aggregate data, the measurement errors could be either $I(0)$ or $I(1)$ processes. Also, poor proxy could arise when the true data are not directly observable; e.g. consumer demand and consumer consumption may not be the same. When facing poor proxy problems, it seems plausible that $I(1)$ measurement errors are of concern; see the debate on the PPP hypothesis in Xu (2003) and Taylor and Taylor (2004), where the test and forecasting results of the PPP hypothesis heavily rely on the carefully-chosen price indices among several alternatives.

In this paper, we show that it is possible to obtain consistent estimator of the unknown smooth function $\theta(\cdot)$ even when u_t is an $I(1)$ process. Thus, the theory obtained in this paper provides a general method for estimating the dependence among non-cointegrated $I(1)$ processes in a flexible semiparametric framework not limited to model (1.1). It can be used by researchers to expand their repertoire to find meaningful relations among integrated variables even when the researchers suspect the possibility of persistent measurement errors in their data or of non-cointegration due to missing relevant integrated factors. For instance, stock market volatilities are highly persistent and can be viewed as $I(1)$, or near $I(1)$ processes (say, fractionally integrated processes).[3] Ray and Tsay (2000) found strong evidence of the existence of common factors generating persistent stock volatilities and the volatilities of stocks of companies in the same business sector sharing more in volatility persistency. Although the existence of common factors can lead to co-movement of stock volatilities, there are also many other (domestic) factors that affect a domestic market's volatility. It is rare that the domestic stock market volatility is cointegrated with a foreign stock market volatility even though the volatilities from different national stock markets are likely to be related with each other, especially during globally bearish market periods.[4] Furthermore, the relation between national stock market volatilities may well be time-varying depending on the varying risk premium. For instance, changes in exchange rate may lead to changes in risk premium, hence may enhance the volatility of stock markets; see Aloui (2007) and Bali and Wu (2010). Therefore, a varying coefficient model, with exchange rate changes as the (nonparametric) state variable entering the varying coefficient function, may provide a flexible approach to examine the spill-over effects of stock market volatility. Monte Carlo simulations reported in Section 4.2 show that our proposed estimators perform quite well when the data are near $I(1)$ processes.

The remaining parts of the paper is organized as follows. Section 2 presents the model and discusses the estimation methods. Section 3 provides asymptotic results of our proposed estimators. In Section 4, we report Monte Carlo simulation results to examine the finite sample performance of the proposed estimators. Finally, Section 5 concludes the paper. All the mathematical proofs are relegated to the Appendix.

2. The model and the proposed estimators

As explained in the introduction section, we aim to find a consistent nonparametric estimator of $\theta(\cdot)$ in the following semiparametric varying coefficient model,

$$Y_t = X_t^\top\theta(Z_t) + u_t, \tag{2.1}$$

where X_t (a $d \times 1$ vector) and u_t (a scalar) are all non-stationary $I(1)$ processes, Z_t is a scalar stationary process, and $\theta(\cdot)$ is a $d \times 1$ vector of unspecified smooth (three-time differentiable) functions to be estimated. The increments (or first differences) of all the $I(1)$ variables have zero means and finite long-run variances, and the unknown coefficient function $\theta(\cdot)$ has a constant finite mean and variance. We will delay more detailed assumptions regarding model (2.1) to Section 3.

Below we first examine the OLS estimator's behavior if one estimates model (2.1) by the least squares method. We assume that the multivariate functional central limit theorem applies to the $(d+1)$-dimensional vector of the partial sums $\left(n^{-1/2}X_{[nr]}^\top, n^{-1/2}u_{[nr]}\right)^\top \Longrightarrow \left(B_X(r)^\top, B_u(r)\right)^\top$ for $r \in [0, 1]$, where $\left(B_X(r)^\top, B_u(r)\right)^\top$ is the $(d+1)$-dimensional Brownian motion processes with a zero mean vector and finite and nonsingular variance–covariance matrix, $[a]$ is the integer part of a positive number a, and "$\Longrightarrow$" denotes the weak convergence with respect to the Skorohod metric.

The OLS estimator of model (2.1) is given by

$$\hat\theta_0 = \left(\sum_{t=1}^n X_tX_t^\top\right)^{-1}\sum_{t=1}^n X_t\left[X_t^\top\theta(Z_t) + u_t\right] = A_{n1} + A_{n2}, \tag{2.2}$$

where $A_{n1} = \left(\sum_{t=1}^n X_tX_t^\top\right)^{-1}\sum_{t=1}^n X_tX_t^\top\theta(Z_t)$ and $A_{n2} = \left(\sum_{t=1}^n X_tX_t^\top\right)^{-1}\sum_{t=1}^n X_tu_t$. It is well established that under some regularity conditions (as given in Section 3) and applying the multivariate functional central limit theorem and continuous mapping theorem, we have

$$A_{n2} = \left(n^{-2}\sum_{t=1}^n X_tX_t^\top\right)^{-1} n^{-2}\sum_{t=1}^n X_tu_t$$

$$\xrightarrow{d} \left[\int_0^1 B_X(r)B_X(r)^\top \mathrm{d}r\right]^{-1}\int_0^1 B_X(r)B_u(r)\mathrm{d}r \stackrel{def}{=} \bar\theta_1, \tag{2.3}$$

where $\bar\theta_1$ is an $O_e(1)$ random variable.[5] We show in the Appendix that

[2] We owe Chan for the references of Mark (2001) and Gavin (1989).

[3] Bollerslev and Mikkelsen (1996) and Baillie et al. (1996) treated conditional volatilities as fractionally integrated processes, and Christensen and Nielsen (2007) suggested to use changes of volatility as the M-term to estimate a GARCH-M type return model.

[4] Using a different approach, Koutmos and Booth (1995) and Karolyi (1995) studied volatility spillovers across international stock markets via multivariate GARCH-type models. GARCH setup and its various extensions are widely used in modeling stock volatility; see Davidson (2004) on some of the recent theoretical results on GARCH-type models.

[5] For any positive non-stochastic sequence a_n, we say that a random variable is $O_e(a_n)$, if it is $O_p(a_n)$ but not $o_p(a_n)$. It means that the random variable has an *exact* probability order of a_n.

$$A_{n1} = \left(n^{-2}\sum_{t=1}^{n} X_t X_t^{\top}\right)^{-1} n^{-2}\sum_{t=1}^{n} X_t X_t^{\top}\theta(Z_t)$$
$$= E\left[\theta(Z_t)\right] + \left(n^{-2}\sum_{t=1}^{n} X_t X_t^{\top}\right)^{-1} n^{-2}\sum_{t=1}^{n} X_t X_t^{\top}\alpha(Z_t)$$
$$\xrightarrow{p} E\left[\theta(Z_t)\right]. \quad (2.4)$$

Therefore, taking together (2.2)–(2.4), we obtain

$$\hat{\theta}_0 - E\left[\theta(Z_t)\right] \xrightarrow{d} \bar{\theta}_1. \quad (2.5)$$

The above result states that when u_t is an $I(1)$ process, the OLS estimator $\hat{\theta}_0$ deviates from the average value of the random coefficient by a variable of stochastic order of $O_e(1)$. However, in the Appendix (see the arguments given at the end of the proof of Lemma A.2), we show that one could consistently estimate $E\left[\theta(Z_t)\right]$ by the OLS estimator $\hat{\theta}_0$ if the error terms u_t were an $I(0)$ process.

Phillips (2009) showed that the kernel estimator is inconsistent estimator for nonparametric spurious regression models. One naturally expects that such inconsistency result will be carried over to our semiparametric regression model because of the $I(1)$ error term. This conjecture is supported by our theory. Consider the standard local constant kernel estimator of $\theta(z)$ given by

$$\check{\theta}(z) = \left(\sum_{t=1}^{n} X_t X_t^{\top} K_{tz}\right)^{-1} \sum_{t=1}^{n} X_t Y_t K_{tz}, \quad (2.6)$$

where $K_{tz} = K((Z_t - z)/h)$, $K(\cdot)$ is the kernel function satisfying conditions given in Section 3, and h is the smoothing parameter. In the Appendix, we establish the following result:

$$\check{\theta}(z) - \theta(z) \xrightarrow{d} \left[\int_0^1 B_X(r)B_X(r)^{\top}\mathrm{d}r\right]^{-1}\int_0^1 B_X(r)B_u(r)\mathrm{d}r \stackrel{def}{=} \bar{\theta}_2, \quad (2.7)$$

where $\bar{\theta}_1$ and $\bar{\theta}_2$ have the same distribution. As $\bar{\theta}_2 = O_e(1)$, $\check{\theta}(z)$ is not a consistent estimator of $\theta(z)$ due to the $I(1)$ error term u_t.

The OLS estimator inconsistently estimating the average value of the coefficient function and the kernel estimator inconsistently estimating the unknown coefficient function are both due to the same reason: the error term is an integrated process whose impact is non-ignorable even asymptotically. Comparing (2.5) with (2.7), it is natural to conjecture that one can asymptotically cancel out the persistency of the $I(1)$ error terms by taking the difference of the kernel estimator $\check{\theta}(z)$ and the OLS estimator $\hat{\theta}_0$. However, taking the difference between the two estimators not only asymptotically cancels out the impact of the non-stationary error terms but also removes the mean value of the unknown coefficient function. Therefore, in doing so, the best one can do is to estimate $\alpha(z) \stackrel{def}{=} \theta(z) - E[\theta(Z_t)]$, the zero mean random coefficient function. Therefore, we will aim to show

$$\check{\theta}(z) - \hat{\theta}_0 \xrightarrow{p} \theta(z) - c_0 = \alpha(z), \quad (2.8)$$

where $c_0 = E[\theta(Z_t)]$. This motivates our first estimator of $\alpha(z)$ given by

$$\tilde{\alpha}(z) = \check{\theta}(z) - \hat{\theta}_0, \quad (2.9)$$

where $\hat{\theta}_0$ and $\check{\theta}(z)$ are defined in (2.2) and (2.6), respectively.

Our second estimator of $\alpha(z)$ is given by

$$\hat{\alpha}(z) = \check{\theta}(z) - n^{-1}\sum_{t=1}^{n}\check{\theta}(Z_t)M_{nt}, \quad (2.10)$$

where $M_{nt} = M_n(Z_t)$ is the trimming function that trims away observations near the boundary of the support of Z_t. We will discuss more on the need of the trimming function in the Appendix.

The estimation of $\alpha(\cdot)$ may itself be of interest because it implies that one can consistently estimate the partial effects, $\frac{\partial}{\partial z}\theta(z) = \frac{\partial}{\partial z}\alpha(z)$. Indeed, this may be the main interest in many studies. However, naturally, one may also want to know the average value of the varying coefficient curve, c_0. We therefore illustrate our estimator of c_0 below, delaying to Section 3 its consistency result.

To obtain an estimator for c_0, we replace $\theta(Z_t)$ by its identity $\theta(Z_t) = c_0 + \alpha(Z_t)$ and rewrite (2.1) as

$$Y_t = X_t^{\top}c_0 + X_t^{\top}\alpha(Z_t) + u_t. \quad (2.11)$$

Then, adding/subtracting $X_t^{\top}\tilde{\alpha}(Z_t)$ in (2.11) gives

$$\tilde{Y}_t \stackrel{def}{=} Y_t - X_t^{\top}\tilde{\alpha}(Z_t) = X_t^{\top}c_0 + v_t, \quad (2.12)$$

where $v_t = X_t^{\top}\left[\alpha(Z_t) - \tilde{\alpha}(Z_t)\right] + u_t$. Because $\tilde{\alpha}(Z_t)$ is a consistent estimator of $\alpha(Z_t)$, $\tilde{Y}_t$ mimics the stochastic properties of $Y_t - X_t^{\top}\alpha(Z_t)$, and the stochastic property of v_t is dominated by that of u_t.

Taking a first difference of (2.12), we obtain

$$\Delta\tilde{Y}_t = \zeta_t^{\top}c_0 + \Delta v_t, \quad (2.13)$$

where $\Delta\tilde{Y}_t = \tilde{Y}_t - \tilde{Y}_{t-1}$, $\zeta_t = X_t - X_{t-1}$, and $\Delta v_t = v_t - v_{t-1}$. Then, regressing $\Delta\tilde{Y}_t$ on ζ_t by the least squares method gives the OLS estimator of c_0 below,

$$\tilde{c}_0 = \left(\sum_{t=2}^{n}\zeta_t\zeta_t^{\top}\right)^{-1}\sum_{t=2}^{n}\zeta_t\Delta\tilde{Y}_t. \quad (2.14)$$

With $\tilde{\alpha}(z)$ and $\tilde{c}_0$, we obtain an estimator of $\theta(z)$ given by

$$\tilde{\theta}(z) = \tilde{\alpha}(z) + \tilde{c}_0. \quad (2.15)$$

The consistency of $\tilde{\theta}(z)$ follows directly from the consistencies of $\tilde{\alpha}(z)$ and $\tilde{c}_0$.

Alternatively, one can use $\hat{\alpha}(Z_t)$ to replace $\tilde{\alpha}(Z_t)$ in the above calculation; i.e., one can replace $\tilde{Y}_t$ by $\hat{Y}_t = Y_t - X_t^T\hat{\alpha}(Z_t)$ in (2.12). We will use $\hat{c}_0$ to denote the resulting estimator of c_0 and denote by $\hat{\theta}(z) = \hat{c}_0 + \hat{\alpha}(z)$ an alternative estimator of $\theta(z)$. The additional assumptions ensuring the consistency of $\tilde{c}_0$ and $\tilde{\theta}(\cdot)$ (or $\hat{c}_0$ and $\hat{\theta}(\cdot)$) as well as the asymptotic results of our proposed estimators of $\alpha(z)$ are the subject of the next section.

3. The asymptotic analysis

We first introduce some notation. For any $q > 0$ and an $m \times 1$ vector ξ_t, we denote by $\|\xi_t\|_q = \left(\sum_{j=1}^{m} E\left|\xi_{jt}\right|^q\right)^{1/q}$ the L^q-norm, and we use "$\|\cdot\|$" without any subscript to denote the Euclidean norm; I_m is the $m \times m$ identify matrix; "$\xrightarrow{d}$" and "$\xrightarrow{p}$" denote convergence in distribution and convergence in probability, respectively. For a differentiable function $g : R \to R^m$, we denote $g^{(s)}(z) = \mathrm{d}^s g(z)/\mathrm{d}z^s$. Also, we use M to denote a finite positive constant whose value may change from place to place.

We now list regularity conditions sufficient for establishing the consistency of $\tilde{\alpha}(z)$ and $\hat{\alpha}(z)$.

(A1) Let $\eta_t = \left(\zeta_t^{\top}, \epsilon_t, Z_t\right)^{\top}$ be a $(d+2)$-dimensional random vector, where $\zeta_t = X_t - X_{t-1}$ and $\epsilon_t = u_t - u_{t-1}$ have zero means. η_t is a strictly stationary, α-mixing sequence of size $-p/(p-2)$ with a finite, positive definite, long-run variance matrix, and $\|\eta_t\|_4 < M < \infty$, where $p = 2+\delta$ for some small $\delta > 0$. Also, both X_0 and u_0 are of order $O_p(1)$ with $X_t \equiv 0$ and $u_t \equiv 0$ for $t < 0$.

(A2) (i) The variable Z_t has a Lebesgue density $f(z)$ and $\inf_{z\in\mathcal{S}} f(z) > 0$, where $\mathcal{S}$, the support of Z_t, is a compact subset of $\mathcal{R}$. Also, (Z_s, Z_t) has a Lebesgue joint density function $f_{s,t}(z_1, z_2)$.

(ii) $f(z)$, $\theta(z)$, $E(\zeta_t|z)$, and $E\left(\|\zeta_t\|^2|z\right)$ are three times continuously differentiable in the vicinity of an interior point $z \in \mathcal{S}$. And, $f_{s,t}(z_1, z_2)$ are three times continuously differentiable in the vicinity of the interior point $(z, z) \in \mathcal{S} \times \mathcal{S}$.

(iii) $\left\|\zeta_t^\top \theta(Z_t)\right\|_{2+\tilde{\delta}} < M < \infty$ and $\|\theta(Z_t)\|_{2+\tilde{\delta}} < M < \infty$ for some $\tilde{\delta} > \delta > 0$.

(A3) The kernel function $K(u)$ is a bounded, symmetric (around zero) probability density function on interval $[-1, 1]$. Also, we denote $\mu_j = \int u^j K(u)\,du$ and $\nu_j = \int u^j K^2(u)\,du$.

(A4) $h \to 0$, $nh \to \infty$ and $nh^5 = O(1)$ as $n \to \infty$.

Assumption A1 ensures that the multivariate functional central limit theorem can be applied to $\left(X_{[nr]}^\top, u_{[nr]}\right)^\top$ for any $r \in [0, 1]$; see Lemma 3.1 below. It allows ζ_t and ϵ_t to be contemporaneously correlated such as $cov(\zeta_t, \epsilon_t) \neq 0$, provided that they are α-mixing processes with the mixing coefficients satisfying some rate of decaying condition, and that η_t satisfies some moment conditions. Assumption A2 requires that the density is bounded below by a positive constant in the compact support of Z_t. It also gives regularity conditions on smoothness and moment restrictions of the density function and some other functionals of Z_t. The three-time differentiable conditions ensure that these functions have Taylor expansions up to the third order so as to yield a leading asymptotic center term of the form $J(z)h^2 + O_p(h^3)$, where $J(z)$ is a leading asymptotic center defined in Theorem 3.1 below. The kernel function having a compact support in Assumption A3 is not essential and can be removed at the cost of lengthy proofs. Specifically, the Gaussian kernel is allowed. Assumption A4 ensures that the proposed estimators are consistent as sample size $n \to \infty$, and $nh^5 = O(1)$ allows for optimal smoothing, but it rules out the under-smoothing data case.

Denote $K_{tz} = K\left(h^{-1}(Z_t - z)\right)$ and $\breve{K}_{tz} = K_{tz} - EK_{tz}$. We need the following lemma for deriving the limiting results of the proposed estimators with the proof delayed to the Appendix.

Lemma 3.1. *Under Assumptions* A1–A4*, the following multivariate functional central limit theorem holds*

$$\begin{pmatrix} n^{-1/2}\sum_{t=1}^{[nr]}\zeta_t \\ n^{-1/2}\sum_{t=1}^{[nr]}\epsilon_t \\ [nh\nu_0 f(z)]^{-1/2}\sum_{t=1}^{[nr]}\breve{K}_{tz} \end{pmatrix} \Longrightarrow \begin{pmatrix} B_X(r) \\ B_u(r) \\ W_K(r) \end{pmatrix}, \tag{3.1}$$

where $\left(B_X(r)^\top, B_u(r), W_K(r)\right)^\top$ *is a* $(d+2)$*-dimensional Brownian motion with a zero mean vector and finite covariance matrix*

$$\begin{pmatrix} \Sigma & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.2}$$

Note that in Lemma 3.1, $W_K(\cdot)$ is a standard Brownian motion process independent of $\left(B_X(\cdot)^\top, B_u(\cdot)\right)^\top$.

Theorems 3.1 and 3.2 below give asymptotic results for $\tilde{\alpha}(z)$ and $\hat{\alpha}(z)$, respectively. The proofs are delayed to Appendix A.1.

Theorem 3.1. *Under Assumptions* A1–A4*, we have, for all interior point* $z \in \mathcal{S}$,

$$\sqrt{nh}[\tilde{\alpha}(z) - \alpha(z) - h^2 J(z)] \xrightarrow{d} \Lambda,$$

where

$$\Lambda = \sqrt{\frac{\nu_0}{f(z)}}\left\{B_{(X,2)}^{-1}\int_0^1 B_X(r)B_u(r)dW_K(r) - B_{(X,2)}^{-1}\int_0^1 B_X(r)B_X(r)^\top dW_K(r)B_{(X,2)}^{-1}B_{(X,u)}\right\}, \tag{3.3}$$

$B_{(X,2)} = \int_0^1 B_X(r)B_X(r)^\top dr$, $B_{(X,u)} = \int_0^1 B_X(r)B_u(r)dr$, *and* $J(z) = \mu_2[\theta^{(1)}(z)f^{(1)}(z)/f(z) + \theta^{(2)}(z)/2]$.

The derivation of Theorem 3.1 does not require zero contemporaneous correlation between the increments of the $I(1)$ processes, Δu_t and ΔX_t. Therefore, X_t is allowed to be endogenous.

In (2.10), $\hat{\alpha}(z) = \breve{\theta}(z) - n^{-1}\sum_{t=1}^n \breve{\theta}(Z_t)M_{nt}$ requires the calculation of $\breve{\theta}(Z_t)$ for all $t = 1, \ldots, n$. Therefore, the continuously differentiable condition in the vicinity of the interior point $z \in \mathcal{S}$ as imposed by Assumption A2 (i) will not be sufficient to ensure the consistency of $\hat{\alpha}(z)$, and we need to strengthen it to a uniform condition. In addition, to obtain the limiting distribution of $\hat{\alpha}(z)$, we need to strengthen our assumption further.

(A2*) (i) The variable Z_t has a Lebesgue density $f(z)$ and $\inf_{z\in\mathcal{S}} f(z) > 0$, where $\mathcal{S}$, the support of Z_t, is a compact subset of $\mathcal{R}$. Also, (Z_s, Z_t) has a Lebesgue joint density function $f_{s,t}(z_1, z_2)$.

(ii) $f(z)$, $\theta(z)$, $E(\zeta_t|z)$, and $E\left(\|\zeta_t\|^2|z\right)$ are continuous on $\mathcal{S}$ and three times continuously differentiable in the interior of $\mathcal{S}$. And, $f_{s,t}(z_1, z_2)$ is continuous on $\mathcal{S} \times \mathcal{S}$ and three times continuously differentiable along the diagonal line (or at all interior points (z, z)) in $\mathcal{S} \times \mathcal{S}$.

(A5) (i) The sequences $\{(\zeta_t, \epsilon_t)\}_{t=1}^n$ and $\{Z_t\}_{t=1}^n$ are independent of each other. (ii) $\{Z_t\}_{t=1}^n$ is a strictly stationary β-mixing process with the β-mixing coefficients satisfying $\beta(t) = O(t^{-\tau})$ for all t, where $\tau > 2\lambda^{-1}(1+\lambda)$ for some $\lambda \in (0, 1)$. (iii) $E(\epsilon_t|\zeta_t) = 0$ and $E\left(\zeta_t\zeta_t^\top\right)$ is nonsingular.

Assumption A2* strengthens Assumption A2 to the support of Z_t. Note that Assumption A5 implies $\beta(t)^{\lambda/(1+\lambda)} = O(t^{-\rho})$ for some $\lambda \in (0, 1)$ and $\rho > 2$. As a β-mixing condition implies an α-mixing condition, Assumption A5(ii) is stronger than Assumption A1.

Theorem 3.2. *Under Assumptions* A1–A5*, and* A2**, we have, for all interior point* $z \in \mathcal{S}$,

$$\sqrt{nh}\left\{\hat{\alpha}(z) - \alpha(z) - h^2[J(z) - E(J(Z_t))]\right\} = \mathring{\Lambda},$$

where $\mathring{\Lambda}$ *has the same distribution as* Λ *defined in* (3.3)*.*

Assumptions A1–A4 are sufficient to show the consistency of $\hat{\alpha}(z)$; i.e., $\hat{\alpha}(z) - \alpha(z) - h^2[J(z) - E(J(Z_t))] = o_p(1)$. The independence between $\{(\zeta_t, \epsilon_t)\}_{t=1}^n$ and $\{Z_t\}_{t=1}^n$ of Assumption A5(i) is a sufficient but not a necessary condition to show that $\mathring{\Lambda}$ and Λ have the same distribution, although this assumption significantly simplifies our proof of the limiting distribution in Theorem 3.2. Note that the asymptotic center of $\hat{\alpha}(z)$ is smaller than that of $\tilde{\alpha}(z)$ on average.

With the above additional assumptions, we are now ready to present the consistency results for $\tilde{c}_0$, $\tilde{\theta}(z)$, $\hat{c}_0$ and $\hat{\theta}(z)$.

Theorem 3.3. *Under Assumptions* A1–A5 *and* A2* *we have*

(i) $\tilde{c}_0 - c_0 = O_p(h^2 + (nh)^{-1/2})$,
$\tilde{\theta}(z) - \theta(z) = O_p(h^2 + (nh)^{-1/2})$;

(ii) $\hat{c}_0 - c_0 = O_p(h^2 + (nh)^{-1/2})$,
$\hat{\theta}(z) - \theta(z) = O_p(h^2 + (nh)^{-1/2})$.

The proof of Theorem 3.3 is given in Appendix A.2. Again, Assumption A5(i) is stronger than necessary, it can be relaxed to $E(\epsilon_t|\zeta_t, Z_t) = 0$ and $E[(\zeta_{t+1} - \zeta_t)J(Z_t)] = 0$. However, we would like to emphasize that Assumption A5 (iii) is necessary for Theorem 3.3 to hold. To see this, noting that $\tilde{c}_0$ is the OLS estimator of c_0 based on (2.13), where Δv_t contains a term equal to $\Delta u_t = \epsilon_t$, we need $E(\epsilon_t|\zeta_t) = 0$ (i.e., Assumption A5(iii)) such that the impact of this term vanishes asymptotically.

Theorems 3.1–3.3 show that our proposed semiparametric curve estimators have the same rate of convergence as in the case with stationary data. In fact, it can be shown that our estimators are robust to whether the error terms are an $I(0)$, a near $I(1)$ or an $I(1)$ process.

Assumption A5 requires that the innovations that generate X_t and u_t to be independent of the stationary covariate Z_t sequence. However, the regularity conditions on the consistency of $\tilde{\alpha}(\cdot)$ is quite weak and standard. If the main interest is to estimate the marginal effects $d\theta(z)/dz$, the stronger condition A5 is not needed as $d\alpha(z)/dz = d\theta(z)/dz$ and $\tilde{\alpha}(z)$ is a consistent estimator of $\alpha(z)$ without the need of Assumption A5.

Also, in some cases, Assumption A5 (i) may be relaxed to that (ζ_t, Z_t) is uncorrelated with ϵ_t or $E(\epsilon_t|\zeta_t, Z_t) = 0$; hence, conditional heteroskedastic error is allowed. For example, when $P(Z_t = Z_{t-1}) > 0$. To see this, we take the first difference to both sides of Eq. (1.1), which gives

$$\Delta Y_t = X_t^\top \theta(Z_t) - X_{t-1}^\top \theta(Z_{t-1}) + \epsilon_t. \tag{3.4}$$

Applying to the sub-data set satisfying $Z_t = Z_{t-1}$, we obtain

$$\Delta Y_t = \Delta X_t^\top \theta(Z_t) + \epsilon_t, \tag{3.5}$$

where Eq. (3.5) is a standard varying coefficient model with all the variables being stationary. The asymptotic results for various semiparametric estimators (local linear or local constant) can be found in Cai et al. (2000), Fan et al. (2000), Li et al. (2002), Fan et al. (2003), among others. The assumption that $P(Z_t = Z_{t-1}) > 0$ is not unrealistic even for time series data. For instance, if one has monthly observations of Z_t, the Federal Reserve fund rate, then for most months, $Z_t = Z_{t-1}$ as the Fed does not change the fund rate on a monthly basis.

Even when $P(Z_t = Z_{t-1}) = 0$, one may not need the complete independence between $\{(\zeta_t, \epsilon_t)\}_{t=1}^n$ and $\{Z_t\}_{t=1}^n$ at all the times. Let us rewrite Eq. (3.4) as

$$\Delta Y_t = \Delta X_t^\top \theta(Z_t) + X_{t-1}^\top [\theta(Z_t) - \theta(Z_{t-1})] + \epsilon_t. \tag{3.6}$$

Multiplying both sides of Eq. (3.6) by a kernel weight function $K((Z_t - Z_{t-1})/h)$, one may derive a consistent estimator of the smooth coefficient function $\theta(\cdot)$. That is, only data with $|Z_t - Z_{t-1}| \le h$ will be used in estimating the regression model. With Z_{t-1} close to Z_t, we have $\theta(Z_{t-1})$ close to $\theta(Z_t)$, and the term $X_{t-1}^T[\theta(Z_t) - \theta(Z_{t-1})]$ can be made negligible by choosing some small value of h. However, in order for the term associated with the $I(1)$ regressor X_{t-1} to be negligible, some under-smoothing condition such as $\sqrt{n}h^2 = o(1)$ may be needed because X_{t-1} needs to be divided by $\sqrt{n}$ to make it an $O_p(1)$ random variable. In addition to the undesired under-smoothing condition, this method also suffers more of the 'curse of dimensionality' problem because one effectively only uses the data satisfying both $|Z_t - z| \le h$ and $|Z_{t-1} - z| \le h$ when estimating $\theta(z)$, while our earlier estimator only requires the condition $|Z_t - z| \le h$. We leave it as a future research topic to find an alternative consistent estimator of $\theta(z)$ (or c_0) under conditions weaker than Assumption A5(i) and without triggering the 'curse of dimensionality' problem.

Up until now, we only consider the case that all the components of X_t are $I(1)$ processes, which rules out the case that X_t can contain a constant. Consider the following model,

$$Y_t = (1, X_t^\top)\begin{pmatrix}\gamma(Z_t)\\ \theta(Z_t)\end{pmatrix} = \gamma(Z_t) + X_t^\top \theta(Z_t) + u_t, \quad t = 1, 2, \ldots, n. \tag{3.7}$$

In this case, one can ignore the $\gamma(Z_t)$ term and treat $\gamma(Z_t) + u_t$ as the composite error term. It can be shown that our proposed estimators $\tilde{\alpha}(z)$ and $\hat{\alpha}(z)$ remain consistent with the same asymptotic distributions. This is similar to the linear regression model case, where the OLS estimator for the coefficients associated with $I(1)$ variables remains consistent even if one ignores the presence of $I(0)$ regressors.

We summarize the above analysis in a proposition below and give a brief proof in Appendix A.2.

Proposition 3.4. *The conclusions of Theorems 3.1–3.3 remain unchanged if Y_t is generated by (3.7) provided that $\gamma(z)$ is a uniformly bounded differentiable function over $z \in \mathcal{S}$.*

4. Monte Carlo simulations

In this section, we examine the finite sample performances of the semiparametric estimators $\hat{\alpha}(\cdot)$, $\tilde{\alpha}(\cdot)$ as well as $\hat{c}_0$ and $\hat{\theta}(\cdot)$.[6] In Section 4.1 we consider the case that both X_t and u_t are $I(1)$ processes, and in Section 4.2 we consider the case that X_t and u_t are near $I(1)$ processes.

4.1. X_t and u_t are $I(1)$ processes

We consider the following data generating process (DGP):

$$Y_t = X_t \theta(Z_t) + u_t, \quad (t = 1, \ldots, n),$$

where X_t and u_t are both $I(1)$ variables, and Z_t is stationary. Specifically, $X_t = X_{t-1} + \zeta_t$ with $X_0 = 0$ and ζ_t is i.i.d. $N(0, \sigma_\zeta^2)$; $u_t = u_{t-1} + \epsilon_t$ with $u_0 = 0$ and ϵ_t is i.i.d. $N(0, \sigma_\epsilon^2)$; $z_t = v_t + v_{t-1}$ and v_t is i.i.d. uniform [0, 2]. We consider two different $\alpha(\cdot)$ functions: (a) $\alpha(z) = \sin(\pi z) - E[\sin(\pi Z_t)]$ and (b) $\alpha(z) = z - .5z^2 - E\left(Z_t - .5Z_t^2\right)$ (so that $E[\alpha(Z_t)] = 0$). In both cases, we have $\theta(z) = c_0 + \alpha(z)$, and we choose $c_0 = 1$ and 2. It is easy to show that $\tilde{\alpha}(\cdot)$ and $\hat{\alpha}(\cdot)$ are invariant to different c_0 values as c_0 is canceled out by construction of the estimators. Hence, we only report the case of $c_0 = 1$ for brevity. We choose two different combinations for $(\sigma_\epsilon, \sigma_\zeta)$: (i) The case of $(\sigma_\epsilon, \sigma_\zeta) = (1, 2)$ means that the increment of u_t has a smaller variance relative to that of X_t (small noise to signal ratio); (ii) The case of $(\sigma_\epsilon, \sigma_\zeta) = (1, 1)$ corresponds to an equal variance case for the increments of X_t and u_t. The sample sizes are $n = 100$, 200 and 400. The number of replications is 10,000. We use a standard normal kernel function with the smoothing parameter equal to $h = \hat{\sigma}_z n^{-1/5}$, where $\hat{\sigma}_z$ is the sample standard deviation of $\{Z_t\}_{t=1}^n$.

We compute the sample average mean squared errors (or AMSEs) for $\hat{\alpha}(\cdot)$, $\tilde{\alpha}(\cdot)$, $\hat{\theta}(\cdot)$ and $\hat{c}_0$ as follows:

$$AMSE(\hat{\alpha}(\cdot)) = J^{-1}\sum_{j=1}^{J}\left\{n^{-1}\sum_{t=1}^{n}\left[\hat{\alpha}_j(Z_t) - \alpha(Z_t)\right]^2\right\},$$

$$AMSE(\tilde{\alpha}(\cdot)) = J^{-1}\sum_{j=1}^{J}\left\{n^{-1}\sum_{t=1}^{n}\left[\tilde{\alpha}_j(Z_t) - \alpha(Z_t)\right]^2\right\},$$

[6] The performances of $\tilde{c}_0$ and $\tilde{\theta}(\cdot)$ are very similar to those of $\hat{c}_0$ and $\hat{\theta}(\cdot)$; therefore, we omit these results to save space.

Table 1
Average mean squared errors when $\alpha(z) = \sin(\pi z) - E[\sin(\pi Z_t)]$.

n	$(\sigma_\epsilon = 1, \sigma_\zeta = 2)$				$(\sigma_\epsilon = 1, \sigma_\zeta = 1)$			
	$\tilde{\alpha}(\cdot)$	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$	$\tilde{\alpha}(\cdot)$	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$
100	.0821	.0799	.1537	.0830	.1388	.1356	.2403	.1083
200	.0428	.0422	.0848	.0443	.0767	.0749	.1306	.0577
400	.0234	.0225	.0436	.0224	.0428	.0418	.0703	.0297

Table 2
Average mean squared errors when $\alpha(z) = z - .5z^2 - E(Z_t - .5Z_t^2)$.

	$(\sigma_\epsilon = 1, \sigma_\zeta = 2)$				$(\sigma_\epsilon = 1, \sigma_\zeta = 1)$			
n	$\tilde{\alpha}(\cdot)$	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$	$\tilde{\alpha}(\cdot)$	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$
100	.00868	.00552	.00945	.00762	.0654	.0609	.0909	.0331
200	.00487	.00319	.00551	.00435	.0372	.0350	.0504	.0172
400	.00275	.00188	.00325	.00235	.0221	.0210	.0297	.0095

$$AMSE(\hat{\theta}(\cdot)) = J^{-1}\sum_{j=1}^{J}\left\{n^{-1}\sum_{t=1}^{n}\left[\hat{\theta}_j(Z_t) - \theta(Z_t)\right]^2\right\},$$

$$AMSE(\hat{c}_0) = J^{-1}\sum_{j=1}^{J}\left(\hat{c}_{0,j} - c_0\right)^2,$$

where J is the number of replications ($J = 10{,}000$), and the subscript j refers to the result from the jth replication. The estimation results are reported in Tables 1 and 2.

Both Tables 1 and 2 show the following features. First, all the AMSEs of $\hat{\alpha}(\cdot)$, $\tilde{\alpha}(\cdot)$, $\hat{\theta}(\cdot)$ and $\hat{c}_0$ are much smaller in magnitude for the case of $(\sigma_\epsilon, \sigma_\zeta) = (1, 2)$ than the case of $(\sigma_\epsilon, \sigma_\zeta) = (1, 1)$. This is expected since the former has a smaller noise to signal ratio. Second, the AMSE for each and every estimator decreases as the sample size increases, which shows the consistency of the estimator. When comparing the results in Table 1 with those in Table 2, we see the latter much smaller than the former. This is because quadratic function $\theta(z)$ (or $\alpha(z)$) in Table 2 is smoother than the sine function $\theta(z)$ (or $\alpha(z)$) in Table 1.

The above simulations results assume that Z_t follows a short memory MA(1) process. Below we report results on some different processes for Z_t. First we let Z_t follow an AR(1) process: (i) $Z_t = \rho_z Z_{t-1} + v_t$, where v_t is i.i.d. uniform $[-1, 1]$ and $\rho_z = 0.5$.[7] Also, Assumption A2 implies that Z_t has a bounded support. We now use Monte Carlo simulations to examine whether this assumption can be relaxed to an unbounded support case. We consider two more different data generating processes for Z_t: (ii) Z_t is i.i.d. $N(0, 1)$ which has an unbounded support but with a quite thin tail; (iii) Z_t is i.i.d. $t(3)$ (or a Student's t-distribution with 3 degrees of freedom), which is a fat-tailed distribution with no finite moment of order greater than two. X_t and u_t are generated the same way as above. To save space, we only consider the case of $(\sigma_\epsilon, \sigma_\zeta) = (1, 2)$ and $\theta(z) = 1+\alpha(z)$ with $\alpha(z) = \sin(\pi z) - E[\sin(\pi Z_t)]$. The simulation results are reported in Table 3.

First, we examine the results for case (i), where Z_t follows an AR(1) process with a density function bounded away from zero in its support. Comparing the results of Table 1 with those of Table 3, we observe that the sample AMSEs for an AR(1) Z_t case are very similar to those of an MA(1) Z_t case. This shows that our estimators are not sensitive to the serial correlation pattern of Z_t, provided that Z_t is a stationary process with a density function that is bounded away from zero in its support.

[7] As we set $Z_0 = 0$, and the i.i.d. innovations v_t are drawn independent of Z_0, the theorem in Athreya and Pantula (1986, p. 187), ensures that the stationary AR(1) process Z_t generated this way is a geometric strong mixing sequence.

Next, we investigate case (ii) that Z_t is i.i.d. $N(0, 1)$. First, we observe that the sample AMSEs decrease as sample size n increases for all the three estimators: $\hat{\alpha}(\cdot)$, $\hat{\theta}(\cdot)$ and $\hat{c}_0$. This suggests that our proposed estimators are likely to be consistent when Z_t has an unbounded support with a thin tail distribution such as a $N(0, 1)$ distribution. However, when comparing case (ii) with case (i), we see that the sample AMSEs for case (ii) are significantly larger than those of case (i) where the distribution of Z_t has a bounded support with the density function bounded away from zero in its support.

Finally, we examine the result that Z_t has a fat-tailed $t(3)$ distribution. In this case, Z_t has a finite second moment but does not have any finite moments of order higher than two. Table 3 clearly shows the inconsistency of $\hat{\theta}(\cdot)$ and $\hat{\alpha}(\cdot)$ in this case. The AMSEs of $\hat{\theta}(\cdot)$ and $\hat{\alpha}(\cdot)$ do not decrease, and in fact, they increase as sample increases.

Therefore, one needs to be cautious in practice if one is suspicious that Z_t may have a fat-tailed distribution. One way to avoid a fat-tail distributed Z_t is to trim out data with extreme Z_t values so that $f(z)$ is bounded away from zero on the trimmed set.

4.2. X_t and u_t are near $I(1)$ processes

In this section, we consider exactly the same data generating processes as discussed in Section 4.1 except that now X_t and u_t are near $I(1)$ rather than $I(1)$ processes. Specifically, we generate X_t and u_t by $X_t = \rho_x X_{t-1} + \zeta_t$ with ζ_t being drawn from an i.i.d. $N(0, \sigma_\zeta^2)$ and X_0 from $N(0, \sigma_\zeta^2/(1-\rho_x^2))$, and $u_t = \rho_u u_{t-1} + \epsilon_t$ with ϵ_t being drawn from an i.i.d. $N(0, \sigma_\epsilon^2)$ and u_0 from $N(0, \sigma_\epsilon^2/(1 - \rho_u^2))$. In addition, $Z_t = v_t + v_{t-1}$, and v_t is drawn from an i.i.d. uniform $[0, 2]$.

To save space, we only consider the case that $\alpha(z) = \sin(\pi z) - E[\sin(\pi Z_t)]$. Since all data are stationary now, $\check{\theta}(z)$ is a consistent estimator of $\theta(z)$, and $\bar{c}_0 \stackrel{def}{=} n^{-1}\sum_{t=1}^{n}\check{\theta}(Z_t)$ is a consistent estimator of c_0. We compare the finite sample performances of $\check{\theta}(\cdot)$ and $\bar{c}$ with our proposed estimator $\hat{\theta}(\cdot)$ and $\hat{c}_0$ for $(\rho_x, \rho_u) = (0.90, 0.90)$, $(0.95, 0.95)$, $(0.97, 0.97)$, $(0.99, 0.99)$. Usually the first lag autocorrelation coefficient of national stock volatility is around 0.95 to 0.98 (e.g., Sun et al. (2010)). Hence, our choice of ρ_x and ρ_u are consistent with the empirical stock volatility data. The simulation results are reported in Table 4.

From Table 4, we observe the followings. (i) Our proposed estimators $\hat{\theta}(\cdot)$ and $\hat{c}_0$ have smaller AMSEs than those of $\check{\theta}(\cdot)$ and $\bar{c}_0$ for all the different (ρ_x, ρ_u) combinations and all the different sample sizes considered. (ii) As the data becomes closer to $I(1)$ processes, the relative performances of our proposed estimators improve more over the conventional estimators. For example, for the case of $n = 400$, when $(\rho_x, \rho_u) = (.90, .90)$, the AMSE of $\check{\theta}(\cdot)$ is only about 20% higher than that of $\hat{\theta}(\cdot)$. However, when

Table 3
Three different distributions for Z_t.

n	(i) $Z_t \sim \mathrm{AR}(1)$			(ii) $Z_t \sim N(0, 1)$			(iii) $Z_t \sim t(3)$		
	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$	$\hat{\alpha}(\cdot)$	$\hat{\theta}(\cdot)$	$\hat{c}_0$
100	.0747	.1372	.0718	.1739	.3453	.1175	.4299	.7804	.3528
200	.0488	.0875	.0426	.1239	.2357	.1140	1.496	1.803	.3076
400	.0304	.0556	.0278	.0845	.1625	.0794	4.217	4.457	.2417

Table 4
X_t and u_t are near $I(1)$ processes.

n	$(\rho_x, \rho_u) = (.90,\ .90)$				$(\rho_x, \rho_u) = (.95,\ .95)$			
	$\check{\theta}(\cdot)$	$\hat{\theta}(\cdot)$	$\bar{c}_0$	$\hat{c}_0$	$\check{\theta}(\cdot)$	$\hat{\theta}(\cdot)$	$\bar{c}_0$	$\hat{c}_0$
100	.2987	.2445	.0977	.0366	.3883	.2517	.1872	.0535
200	.1987	.1613	.0502	.0149	.2472	.1665	.0998	.0210
400	.1286	.1094	.0249	.0067	.1534	.1114	.0499	.0090
	$(\rho_x, \rho_u) = (.97,\ .97)$				$(\rho_x, \rho_u) = (.99,\ .99)$			
100	.4870	.2603	.2858	.0721	.7515	.3251	.5490	.1255
200	.3088	.1734	.1621	.0285	.5472	.2025	.4009	.0575
400	.1855	.1143	.0821	.0119	.3250	.1275	.2219	.0252

$(\rho_x, \rho_u) = (.97,\ .97)$, the AMSE of $\check{\theta}(\cdot)$ is almost doubled that of $\hat{\theta}(\cdot)$. Moreover, the AMSE of $\check{\theta}(\cdot)$ becomes more than double that of $\hat{\theta}(\cdot)$ when $(\rho_x, \rho_u) = (.99, .99)$. The improvement for estimating c_0 is even more dramatic. For $n = 100$, the AMSE of $\bar{c}_0$ is about four times as large as the AMSE of $\hat{c}_0$ when $(\rho_x, \rho_u) = (.90, .90)$, and this AMSE ratio becomes almost 10 fold when $(\rho_x, \rho_u) = (.99, .99)$.

The simulation results reported in this section are consistent with our theoretical analysis. When the error term u_t is an $I(1)$ or a near $I(1)$ process, the conventional kernel-based estimator of $\theta(\cdot)$ becomes inconsistent or inaccurate, while our proposed estimator is consistent and accurate regardless of whether the error term u_t is an $I(1)$ or a near $I(1)$ process.

5. Conclusions

Most macroeconomic and finance variables show strong persistency, and many of them are measured with errors or even unobservable. This may render correlated but not cointegrated time series for actually observed or artificially constructed proxy variables, although the variables with true values could be cointegrated as predicted by economic or finance theory. In this paper we suggest using a flexible semiparametric varying coefficient model to capture the correlation among integrated but not cointegrated variables. We propose two consistent estimators to estimate the unknown smooth coefficient function and establish the consistency of the proposed estimators.

The current study is limited to the case that the nonparametric component Z_t has a density function that is bounded away from zero in its support. It would be useful to extend the asymptotic analysis to the case that Z_t has an unbounded support such as a normal distribution case. The simulation results reported in Section 4 show that our proposed estimators do not lead to consistent estimation of $\theta(\cdot)$ when Z_t has a fat-tailed distribution. It would be very useful if alternative estimation methods can be found that are robust to the fat-tailed distribution of Z_t. The theoretical results of this paper may also be useful in developing new cointegration tests based on semiparametric models with an $I(1)$ error term. Finally, we hope to be able to generalize the consistent model specification tests to our framework; i.e., testing the null hypothesis of a parametric functional form of $\theta(\cdot)$ by allowing the error term u_t to be an $I(1)$ process. These challenging problems are left as future research topics.

Appendix. Mathematical proofs

This Appendix contains two subsections. Appendix A.1 provides the proofs for Theorems 3.1 and 3.2, and Appendix A.2 gives proofs for Theorem 3.3 and Proposition 3.4.

A.1. Mathematical proofs of Theorems 3.1 and 3.2

Throughout this Appendix, we denote $B_{(X,2)} = \int_0^1 B_X(r)B_X(r)^\top \mathrm{d}r$, $B_{(X,u)} = \int_0^1 B_X(r)B_u(r)\mathrm{d}r$, $K_{tz} = K\left(h^{-1}(Z_t - z)\right)$, $\breve{K}_{tz} = K_{tz} - EK_{tz}$, $\theta_t = \theta(Z_t)$, and $\alpha_t = \theta_t - E(\theta_t)$. Also, we use the short-hand notation $\sum_t$ and $\sum_t \sum_{s\neq t}$ to denote $\sum_{t=1}^n$ and $\sum_{t=1}^n \sum_{s\neq t}^n$, respectively.

For readers' convenience, below we give modified versions of the strong mixing inequality of Lemma 2.1 of McLeish (1975) and Theorem 3.2 of de Jong and Davidson (2000). The presentation will be based on the assumptions imposed in this paper.

Denote by $\mathcal{F}_t = \sigma(\omega_t : i \le t, n \ge 1)$ the smallest σ-field containing the past history of $\{\omega_t\}$ for all $t \le n, n \in \mathcal{N}$, the set of natural number. For $1 \le p_1 \le p_2 < \infty$ and any $m > 0$, McLeish (1975) strong mixing inequality states that

$$\|E(\omega_{t+m}|\mathcal{F}_t) - E(\omega_t)\|_{p_1} \le 2\left(2^{1/p_1} + 1\right)[\alpha(m)]^{p_1^{-1} - p_2^{-1}} \|\omega_t\|_{p_2}. \tag{A.1}$$

de Jong and Davidson (2000) gave a different proof of the multivariate functional central theorem under quite weak regularity conditions.

Theorem 3.2 of de Jong and Davidson (2000): Let W_{nt} be an $d \times 1$ vector of array and assume that for every d-vector ξ of unit length, $\tilde{W}_{nt} = \xi^\top W_{nt}$, and $U_n(r) = \sum_{t=1}^{[nr]} W_{nt}$ for $r \in [0, 1]$. Then, $U_n \Longrightarrow U$, where U is a d-dimensional Gaussian process having almost sure continuous sample paths and independent increments, if there exists a positive constant array c_{nt} such that the following condition holds for $\tilde{W}_{nt}$:

(i) $E\left(\tilde{W}_{nt}\right) = 0$, $\left\|\sum_{t=1}^n \tilde{W}_{nt}\right\|_2 = 1$, and $\left\{\tilde{W}_{nt}/c_{nt}\right\}$ is L_p-bounded uniformly in t and n;

(ii) $\left\|\tilde{W}_{nt} - E\left(\tilde{W}_{nt}|\mathcal{F}_{n,t-m}^{t+m}\right)\right\|_2 \le d_{nt}\,\upsilon(m)$, where $\mathcal{F}_{n,t-m}^{t+m} = \sigma\left(W_{n,t-m}, \ldots, W_{n,t+m}\right)$ is the smallest σ-field containing $\left\{W_{n,t-m}, \ldots, W_{n,t+m}\right\}$ for all n, d_{nt}/c_{nt} is bounded uniformly in t and n, and $\upsilon(m) = O\left(m^{-1/2-\epsilon}\right)$ for some small $\epsilon > 0$;

(iii) For some sequence b_n such that $b_n = o(n)$ and $b_n^{-1} = o(1)$, letting $r_n = \left[nb_n^{-1}\right]$, $M_{ni} = \max_{(i-1)b_n+1\le t\le b_n} c_{nt}$ and $M_{n,r_n+1} = \max_{r_nb_n+1\le n} c_{nt}$, $\max_{1\le i\le r_n+1} M_{ni} = o\left(b_n^{-1/2}\right)$ and $\sum_{i=1}^{r_n} M_{nt}^2 = O\left(b_n^{-1}\right)$;
(iv) $\lim_{n\to\infty} E[U_n(r)^2]$ is finite for all $r \in [0,1]$, and $\lim_{\varphi\to 0}\sup_{r\in[0,1-\varphi]}\limsup_{n\to\infty}\sum_{[nr]+1}^{[n(r+\varphi)]} c_{nt}^2 = 0$.

Note that under Assumptions A1 and A2, the σ-field defined here, $\mathcal{F}_{n,t-m}^{t+m}$, only needs to depend on the information sets of $\left\{W_{n,t-m},\ldots,W_{n,t+m}\right\}$, while de Jong and Davidson (2000) considered more general case than cited here.

Proof of Lemma 3.1. By Assumptions A1 and A2, we can verify the conditions required by Corollary 4.2 of Wooldridge and White (1988) or Theorem 3.2 of de Jong and Davidson (2000) and obtain the following multivariate functional central limit theorem, for any $r\in[0,1]$,

$$\left(n^{-1/2}\sum_{t=1}^{[nr]}\zeta_t^\top, n^{-1/2}\sum_{t=1}^{[nr]}\epsilon_t, (nhv_0f(z))^{-1/2}\sum_{t=1}^{[nr]}\breve{K}_{tz}\right)^\top \Longrightarrow \left(B_X(r)^\top, B_u(r), W_K(r)\right)^\top, \tag{A.2}$$

where $\left(B_X(r)^\top, B_u(r)\right)^\top$ is a $(d+1)\times 1$ column vector of Brownian motion process with a zero mean and covariance matrix of $r\Sigma$ and is independent of the Standard Brownian motion $W_K(r)$. Here, Σ is a finite nonsingular matrix under Assumption A1.

The result that $(nhv_0f(z))^{-1/2}\sum_{t=1}^{[nr]}\breve{K}_{tz}$ is asymptotically independent of the other two terms and has an asymptotic variance r is obtained from the following results: (i) The asymptotic covariance between $(nh)^{-1/2}\sum_{t=1}^{[nr]}\breve{K}_{tz}$ and any one of the other two terms in the lemma is of order $O\left(\sqrt{h}\right)$; (ii) Under Assumptions A1–A4, the variance of $(nh)^{-1/2}\sum_{t=1}^{n}\breve{K}_{tz}$ converges to $v_0f(z)$. This completes the proof of Lemma 3.1. □

The following two lemmas are used to obtain the limiting property of the OLS estimator under model (2.1). Both the proof of Theorem 4.2 in Hansen (1992) and the proof of Theorem 4.1 in de Jong and Davidson (2000) can be used to show weak convergence to stochastic integral results, while the latter uses weaker assumptions than the former. As we imposed similar assumptions as in Hansen (1992), we will use Hansen's (1992) method to show the weak convergence to stochastic integral results below.

Lemma A.1. *If Assumptions* A1 *and* A2 *hold, we have*

$$n^{-3/2}\sum_{t=1}^{n}X_tX_t^\top\alpha(Z_t) = O_p(1). \tag{A.3}$$

Proof. If $\theta(Z_t)\equiv c_0$, a vector of constants, we have $n^{-3/2}\sum_{t=1}^{n}X_tX_t^\top\alpha(Z_t)\equiv 0$, and the lemma holds true for this trivial case. Below, we prove this lemma for the case that $\theta(\cdot)$ is a vector of non-constant measurable smooth functions.

Simple calculations give

$$\begin{aligned}\frac{1}{n^{3/2}}\sum_{t=1}^{n}X_tX_t^\top\alpha_t &= \frac{1}{n^{3/2}}\sum_{t=1}^{n}(X_{t-1}+\zeta_t)(X_{t-1}+\zeta_t)^\top\alpha_t\\ &= \frac{1}{n^{3/2}}\sum_{t=1}^{n}X_{t-1}X_{t-1}^\top\alpha_t + \frac{1}{n^{3/2}}\\ &\quad\times\sum_{t=1}^{n}X_{t-1}\zeta_t^\top\alpha_t + \frac{1}{n^{3/2}}\sum_{t=1}^{n}\zeta_tX_{t-1}^\top\alpha_t\\ &\quad+\frac{1}{n^{3/2}}\sum_{t=1}^{n}\zeta_t\zeta_t^\top\alpha_t\\ &= \Gamma_{n1}+\Gamma_{n2}+\Gamma_{n3}+\Gamma_{n4},\end{aligned} \tag{A.4}$$

where the definitions of Γ_{nj} for $j=1,2,3,4$ should be obvious from the context. Below, we will show that Γ_{n1}, Γ_{n2} and Γ_{n3} are all of order $O_p(1)$, but $\Gamma_{n4} = O_p\left(n^{-1/2}\right)$.

Consider Γ_{n4} first. Under Assumption A1 and applying Lemma 2.1 of White and Domowitz (1984), we obtain that $\zeta_t\zeta_t^\top\alpha_t$ is an α-mixing process of size $-p/(p-2)$. Then, applying Hölder's inequality, we have $E(\|\Gamma_{n4}\|)\le n^{-1/2}E\left(\|\zeta_t\|\left|\zeta_t^\top\alpha_t\right|\right)\le n^{-1/2}\|\zeta_t\|_2\left\|\zeta_t^\top\alpha_t\right\|_2 < M < \infty$ under Assumptions A1 and A2(ii). Therefore, we obtain $\Gamma_{n4} = O\left(n^{-1/2}\right)$.

Now, we consider Γ_{n2}. Decomposing Γ_{n2} into two components yields $\Gamma_{n2} = \sum_{t=1}^{n}X_{t-1}E\left(\zeta_t^\top\alpha_t\right) + n^{-3/2}\sum_{t=1}^{n}X_{t-1}e_t$, where $e_t = \zeta_t^\top\alpha_t - E\left(\zeta_t^\top\alpha_t\right)$ and $E(e_t) = 0$. Applying Hölder's inequality and McLeish's strong mixing inequality (A.1), we obtain $E[n^{-1}\sum_{t=1}^{n}|e_t|]\le 2E\left|\zeta_t^\top\alpha_t\right| < M < \infty$ and $Var\left(n^{-1}\sum_{t=1}^{n}e_t\right) = O_p\left(n^{-1}\right)$ by Assumptions A1 and A2(ii). Then, denoting $U_{nt} = n^{-1/2}X_t$ and closely following the proof of Theorem 3.3 of Hansen (1992), we obtain $n^{-3/2}\sum_{t=1}^{n}X_{t-1}e_t = o_p(1)$. It follows

$$\begin{aligned}\Gamma_{n2} &= n^{-3/2}\sum_{t=1}^{n}X_{t-1}E\left(\zeta_t^\top\alpha_t\right)+o_p(1)\\ &\stackrel{d}{\to}\int_0^1 B_X(r)\,\mathrm{d}rE\left(\zeta_t^\top\alpha_t\right).\end{aligned} \tag{A.5}$$

Similarly, we obtain

$$\begin{aligned}\Gamma_{n3} &= E\left(\zeta_t\alpha_t^\top\right)n^{-3/2}\sum_{t=1}^{n}X_{t-1}+o_p(1)\\ &\stackrel{d}{\to} E\left(\zeta_t\alpha_t^\top\right)\int_0^1 B_X(r)\,\mathrm{d}r.\end{aligned} \tag{A.6}$$

Finally, we will derive the limiting result of Γ_{n1}, applying the method used in the proof of Theorem 4.2 in Hansen (1992). For $r\in[0,1]$, let $U_n(r) = U_{n,[nr]} = n^{-1}X_{[nr]}X_{[nr]}^\top$ and $V_n(r) = n^{-1/2}\sum_{t=1}^{[nr]}\alpha_t$. Under Assumption A1, we actually can extend Lemma 3.1 such that $V_n(r)\Longrightarrow B_\alpha(r)$ holds jointly with the partial sums appearing in Lemma 3.1, where $(B_X(r)^\top, B_u(r), B_\alpha(r)^\top, W_K(r))^\top$ is a $(2d+2)$-dimensional Brownian motion with a zero mean and finite and nonsingular covariance matrix.

Then, applying Theorem 3.1 of Hansen (1992), we have $n^{-3/2}\sum_{t=1}^{[nr]}X_{t-1}X_{t-1}^\top\alpha_t = \int_0^r U_n(s)\,\mathrm{d}V_n(s) = \int_0^r U_n(s)\,\mathrm{d}Q_n(s) + \Lambda_{n,[nr]}^*$, where $\int_0^r U_n(s)\,\mathrm{d}Q_n(s)\stackrel{d}{\to}\int_0^r B_X(s)B_X(s)^\top\,\mathrm{d}B_\alpha(s) = O_p(1)$ and

$$\Lambda_{n,t}^* = n^{-1/2}\sum_{i=1}^{t}\left(U_{n,i}-U_{n,i-1}\right)w_i - n^{-1/2}U_{n,t}w_{t+1} \tag{A.7}$$

with $w_i = \sum_{k=1}^{\infty}E(\alpha_{i+k}|\mathcal{F}_i)$ and $\mathcal{F}_t = \sigma(U_{nt}, Z_t : i\le t, n\ge 1)$ being the smallest sigma-field containing the past history of $\{(U_{nt}, Z_t)\}$ for all n. It remains to show $\Lambda_{n,[nr]}^*\Longrightarrow r\Lambda$, a finite vector.

Applying Minkowski's inequality and McLeish's strong mixing inequality of (A.1), we obtain for some $\lambda_2>\lambda_1>2$,

$$\begin{aligned}\|w_i\|_{\lambda_1} &\le \sum_{k=1}^{\infty}\|E(\alpha_{i+k}|\mathcal{F}_i)\|_{\lambda_1}\\ &\le\sum_{k=1}^{\infty}6\alpha(k)^{1/\lambda_1-1/\lambda_2}\|\alpha_{i+k}\|_{\lambda_2}\\ &\le 6M\sum_{k=1}^{\infty}k^{-p/(p-2)(1/\lambda_1-1/\lambda_2)}\|\alpha_{i+k}\|_{\lambda_2} < M < \infty,\end{aligned} \tag{A.8}$$

if $\|\alpha_i\|_{\lambda_2} < \infty$ and $p/(p-2)(1/\lambda_1 - 1/\lambda_2) > 1$. Let $1/\lambda_1 - 1/\lambda_2 = \tilde{\delta}/\left(2+\tilde{\delta}\right)$ for some $\tilde{\delta} > \delta > 0$. Then, λ_1 and λ_2 s.t. $\lambda_2 > \lambda_1 = \left(2+\tilde{\delta}\right)\lambda_2/\left[2+(1+\lambda_2)\tilde{\delta}\right] > 2$ can be used to obtain (A.8). Applying Chebyshev's inequality, we can obtain for any given small $\xi > 0$, $\Pr\left(\sup_{1\le i\le n}\|w_i\| > \sqrt{n}\xi\right) \le \sum_{i=1}^n \Pr\left(\|w_i\| > \sqrt{n}\xi\right) \le \sum_{i=1}^n \left(\sqrt{n}\xi\right)^{-\lambda_1} E\|w_i\|^{\lambda_1} \le Mn^{1-\lambda_1/2}\xi^{-\lambda_1} \to 0$ as $n\to\infty$, it then follows that $\sup_{1\le i\le n}\|w_i\| = o_p\left(\sqrt{n}\right)$. Therefore, we obtain

$$\sup_{1\le t\le n}\left\|n^{-1/2}U_{n,t}w_{t+1}\right\| \le \sup_{1\le t\le n}\left\|n^{-1}X_tX_t^\top\right\| \sup_{1\le t\le n}\left\|n^{-1/2}w_{t+1}\right\| \xrightarrow{p} 0. \quad (A.9)$$

Next, we have

$$n^{-1/2}\sum_{i=1}^t\left(U_{n,i}-U_{n,i-1}\right)w_i = n^{-3/2}\sum_{i=1}^t\left(X_iX_i^\top - X_{i-1}X_{i-1}^\top\right)w_i$$
$$= n^{-3/2}\sum_{i=1}^t\left(X_{i-1}\zeta_i^\top + \zeta_iX_{i-1}^\top + \zeta_i\zeta_i^\top\right)w_i,$$

where $n^{-3/2}\sum_{i=1}^t \zeta_i\zeta_i^\top w_i = O_p\left(n^{-1/2}\right)$ as $E\sup_{1\le t\le n}\left\|n^{-3/2}\sum_{i=1}^t\zeta_i\zeta_i^\top w_i\right\| \le n^{-1/2}\|\zeta_i\|^2_{2\lambda_1/(\lambda_1-1)} \times \|w_i\|_{\lambda_1} = O_p\left(n^{-1/2}\right)$ by (A.8). Then following the proof of Γ_{n2}, we obtain $n^{-3/2}\sum_{i=1}^t\left(X_{i-1}\zeta_i^\top + \zeta_iX_{i-1}^\top\right)w_i = n^{-3/2}\sum_{i=1}^t X_{i-1}E\left(\zeta_i^\top w_i\right) + E\left(\zeta_i w_i^\top\right)n^{-3/2}\sum_{i=1}^t X_{i-1} + o_p(1) = O_p(1)$. Therefore, we have shown $\Lambda^*_{n,t} = O_p(1)$.

Combining the results above completes the proof of Lemma A.1. □

Lemma A.2. *If Assumptions* A1 *and* A2 *hold, we have*

$$\hat{\theta}_0 - E\left[\theta\left(Z_t\right)\right] \xrightarrow{d} B^{-1}_{(X,2)}B_{(X,u)}. \quad (A.10)$$

Proof. As

$$n^{-2}\sum_{t=1}^n X_tY_t = n^{-2}\sum_{t=1}^n X_tX_t^\top\theta_t + n^{-2}\sum_{t=1}^n X_tu_t$$
$$= n^{-2}\sum_{t=1}^n X_tX_t^\top E(\theta_t) + n^{-2}\sum_{t=1}^n X_tu_t + n^{-2}\sum_{t=1}^n X_tX_t^\top\alpha_t,$$

we have

$$\hat{\theta}_0 = \left(\sum_t X_tX_t^\top\right)^{-1}\sum_t X_tY_t$$
$$= c_0 + \left(n^{-2}\sum_t X_tX_t^\top\right)^{-1} n^{-2}\sum_{t=1}^n X_tu_t + \left(n^{-2}\sum_t X_tX_t^\top\right)^{-1} n^{-2}\sum_{t=1}^n X_tX_t^\top\alpha_t$$
$$= c_0 + \left(n^{-2}\sum_t X_tX_t^\top\right)^{-1} n^{-2}\sum_{t=1}^n X_tu_t + O_p(n^{-1/2}), \quad (A.11)$$

where the last equality follows from Lemma A.1.

By Lemma 3.1 and the continuous mapping theorem, we have

$$n^{-2}\sum_{t=1}^n X_tX_t^\top = n^{-1}\sum_{t=1}^n\left(\frac{X_t}{\sqrt{n}}\right)\left(\frac{X_t}{\sqrt{n}}\right)^\top \xrightarrow{d} B_{(X,2)} \quad (A.12)$$

$$n^{-2}\sum_{t=1}^n X_tu_t = n^{-1}\sum_{t=1}^n \frac{X_t}{\sqrt{n}}\frac{u_t}{\sqrt{n}} \xrightarrow{d} B_{(X,u)}. \quad (A.13)$$

Combining the results above, we obtain

$\hat{\theta}_0 - c_0 \xrightarrow{d} B^{-1}_{(X,2)}B_{(X,u)}$.

This completes the proof of Lemma A.2. □

From the proof of Lemma A.2, it is obvious that, if u_t were an $I(0)$ process, $\hat{\theta}_0 - c_0 = O_p\left(n^{-1/2}\right)$; i.e., if Y_t and X_t were cointegrated with the varying cointegration vector $\theta(Z_t)$, then the OLS estimator, $\hat{\theta}_0$, would consistently estimate $c_0 = E[\theta(Z_t)]$. Therefore, substantial difference between the OLS estimator $\hat{\theta}_0$ and the semiparametric estimator $\hat{c}_0$ (or $\tilde{c}_0$) would suggest that Y_t and X_t are not cointegrated; i.e., u_t is not an $I(0)$ process.

The following three lemmas are used to derive the limiting properties of the proposed kernel estimators of model (2.1).

Lemma A.3. *If Assumptions* A1–A4 *hold, we have*

$$\left(n^3hv_0f(z)\right)^{-1/2}\sum_{t=1}^n X_tX_t^\top\breve{K}_{tz} \xrightarrow{d} \int_0^1 B_X(r)B_X(r)^\top dW_K(r). \quad (A.14)$$

Proof. Simple calculations give

$$\Gamma_n = \frac{1}{\sqrt{n^3h}}\sum_{t=1}^n X_tX_t^\top\breve{K}_{tz} = \frac{1}{\sqrt{n^3h}}\sum_{t=1}^n X_{t-1}X_{t-1}^\top\breve{K}_{tz} + \frac{1}{\sqrt{n^3h}}\sum_{t=1}^n X_{t-1}\zeta_t^\top\breve{K}_{tz} + \frac{1}{\sqrt{n^3h}}\sum_{t=1}^n \zeta_tX_{t-1}^\top\breve{K}_{tz} + \frac{1}{\sqrt{n^3h}}\sum_{t=1}^n \zeta_t\zeta_t^\top\breve{K}_{tz}$$
$$= \Gamma_{n1} + \Gamma_{n2} + \Gamma_{n2}^\top + \Gamma_{n3},$$

where we will show $\Gamma_{n1} = O_p(1)$, $\Gamma_{n2} = O_p\left(\sqrt{h}\right)$, and $\Gamma_{n3} = O_p\left(\sqrt{h/n}\right)$. Therefore, Γ_{n1} is the leading term of Γ_n.

Consider Γ_{n3} first. We have $E\left\|n^{-1}\sum_{t=1}^n\zeta_t\zeta_t^\top\breve{K}_{tz}\right\| \le E\left(\left\|\zeta_t\zeta_t^\top\right\|\left|\breve{K}_{tz}\right|\right) \le E\left[E\left(\left\|\zeta_t\zeta_t^\top\right\||Z_t\right)K_{tz}\right] + E\left(\left\|\zeta_t\zeta_t^\top\right\|\right)E\left(K_{tz}\right) = O(h)$ by Assumptions A1 and A2(i). It follows $n^{-1}\sum_{t=1}^n\zeta_t\zeta_t^\top\breve{K}_{tz} = O_p(h)$. Therefore, we have

$$\Gamma_{n3} = O_p\left(\sqrt{h/n}\right). \quad (A.15)$$

Consider $\Gamma_{n2} = \left(n^3h\right)^{-1/2}\sum_{t=1}^n X_{t-1}E\left(\zeta_t^\top\breve{K}_{tz}\right) + \left(n^3h\right)^{-1/2}\sum_{t=1}^n X_{t-1}e_{tz}$, where $e_{tz} = \zeta_t^\top\breve{K}_{tz} - E\left(\zeta_t^\top\breve{K}_{tz}\right)$ and $E\left(e_{tz}\right) = 0$. Following the Proof of Theorem 3.3 of Hansen (1992), we show that $(nh)^{-1}\sum_{t=1}^n n^{-1/2}X_{t-1}e_{tz} = o_p(1)$. Therefore,

$$h^{-1/2}\Gamma_{n2} = \frac{1}{n}\sum_{t=1}^n\frac{X_{t-1}}{\sqrt{n}}\frac{E\left(\zeta_t^\top\breve{K}_{tz}\right)}{h} + o_p(1) \xrightarrow{d} \int_0^1 B_X(r)\,drE\left(\zeta_t^\top|z\right)f(z),$$

which gives $\Gamma_{n2} = O_p\left(\sqrt{h}\right)$.

Now, consider $\Gamma_{n1} = \left(n^3h\right)^{-1/2}\sum_{t=1}^n X_{t-1}X_{t-1}^\top\breve{K}_{tz}$. For $r \in [0,1]$, let $U_n(r) = U_{n,[nr]} = n^{-1}X_{[nr]}X_{[nr]}^\top$ and $V_n(r) = (nhv_0f(z))^{-1/2}\sum_{t=1}^{[nr]}\breve{K}_{tz}$. Then, $\left(n^3hv_0f(z)\right)^{-1/2}\sum_{t=1}^{[nr]}X_{t-1}X_{t-1}^\top\breve{K}_{tz} = \int_0^r U_n(s)\,dV_n(s) = \int_0^r U_n(s)\,dQ_n(s) + \Lambda^*_{n,[nr]}$, where by Lemma 3.1

and Theorem 3.1 of Hansen (1992), we have $\int_0^r U_n(s)\,dQ_n(s) \xrightarrow{d} \int_0^r B_X(s) B_X(s)^\top dW_K(s)$ and

$$\Lambda_{n,t}^* = (nh)^{-1/2} \sum_{i=1}^{t} \left(U_{n,i} - U_{n,i-1}\right) w_i - n^{-1/2} U_{n,t} w_{t+1} \quad (A.16)$$

with $w_i = (v_0 f(z))^{-1/2} \sum_{k=1}^{\infty} E\left(\breve{K}_{i+k,z} | \mathcal{F}_i\right)$. It remains to show $\Lambda_{n,[nr]}^* = o_p(1)$ for any $r \in [0, 1]$.

Following the proof of Lemma A.1, we can show that $\|w_i\|_{\lambda_1} = O\left(h^{1/\lambda_2}\right)$ for all i, where $\lambda_2 > \lambda_1 > 2$. Applying Chebyshev's inequality, we obtain for any given small $\xi > 0$, $\Pr\left(\sup_{1\le i\le n} \|w_i\| > \sqrt{nh}\xi\right) \le \sum_{i=1}^n \Pr\left(\|w_i\| > \sqrt{nh}\xi\right) \le \sum_{i=1}^n \left(\sqrt{nh}\xi\right)^{-\lambda_1} E\|w_i\|^{\lambda_1} \le M(nh)^{1-\lambda_1/2}\xi^{-\lambda_1} \to 0$ as $n \to \infty$, it then follows that $\sup_{1\le i\le n}\|w_i\| = o_p\left(\sqrt{nh}\right)$. Consequently, we have

$$\sup_{1\le t\le n} \left\|(nh)^{-1/2} U_{n,t} w_{t+1}\right\| \le \sup_{1\le t\le n}\left\|n^{-1} X_t X_t^\top\right\| \sup_{1\le t\le n}\left\|(nh)^{-1/2} w_{t+1}\right\| \xrightarrow{p} 0.$$

Moreover,

$$\begin{aligned}(nh)^{-1/2} \sum_{i=1}^{t}\left(U_{n,i} - U_{n,i-1}\right) w_i &= \left(n^3 h\right)^{-1/2} \sum_{i=1}^{t}\left(X_{i-1}\zeta_i^\top + \zeta_i X_{i-1}^\top + \zeta_i\zeta_i^\top\right) w_i \\ &= \left(n^3 h\right)^{-1/2} \sum_{i=1}^{t}\left(X_{i-1}\zeta_i^\top + \zeta_i X_{i-1}^\top\right) w_i + o_p\left((nh)^{-1/2}\right),\end{aligned}$$

where $\left(n^3 h\right)^{-1/2} \sum_{i=1}^{t} \zeta_i\zeta_i^\top w_i = o_p\left((nh)^{-1/2}\right)$ as

$$E \sup_{1\le t\le n}\left\|\left(n^3 h\right)^{-1/2} \sum_{i=1}^{t}\zeta_i\zeta_i^\top w_i\right\| \le M(nh)^{-1/2}\|\zeta_i\|^2_{2\lambda_1/(\lambda_1-1)}\|w_i\|_{\lambda_1} = O_p\left((nh)^{-1/2} h^{1/\lambda_2}\right).$$

Following the proof of Γ_{n2} in Lemma A.1, we can show that $\left(n^3 h\right)^{-1/2}\sum_{i=1}^t \left(X_{i-1}\zeta_i^\top + \zeta_i X_{i-1}^\top\right) w_i = O_p\left(\sqrt{h}\right)$. Therefore, taking together the results above gives $\Lambda_{n,t}^* = o_p(1)$.

Hence, we have

$$\begin{aligned}(v_0 f(z))^{-1/2}\,\Gamma_{n1} &= \int_0^1 U_n(s)\,dQ_n(s) + o_p(1) \\ &\xrightarrow{d} \int_0^1 B_X(s) B_X(s)^\top dW_K(s) = O_p(1). \end{aligned} \quad (A.17)$$

This completes the proof of Lemma A.3. □

Lemma A.4. *If Assumptions A1–A4 hold, and z is an interior point of $\mathcal{S}$, then we have*

$$(nh)^{-3/2}\sum_{t=1}^{n} X_t X_t^\top \breve{\theta}_{tz} = O_p(1), \quad (A.18)$$

where $\breve{\theta}_{tz} = [\theta_t - \theta(z)]K_{tz} - E\{[\theta_t - \theta(z)]K_{tz}\}$.

Proof. Let $V_n(r) = \left(nh^3\right)^{-1/2}\sum_{t=1}^{[nr]}\breve{\theta}_{tz}$ and $U_n(r) = n^{-1}X_{[nr]}X_{[nr]}^\top$ for any $r \in [0, 1]$. The proof will closely follow that of Lemma A.3 with some tedious calculations. Therefore, we omit the proof here. □

Lemma A.5. *If Assumption A1–A4 hold, and z is an interior point of $\mathcal{S}$, we have*

$$\left[n^3 h v_0 f(z)\right]^{-1/2}\sum_{t=1}^{n} X_t u_t \breve{K}_{tz} \xrightarrow{d} \int_0^1 B_X(r) B_u(r)\,dW_K(r). \quad (A.19)$$

Proof. As both X_t and u_t are $I(1)$ processes, applying the Proof of Lemma A.3 with $X_t X_t^\top$ replaced by $X_t u_t$ proves Lemma A.5. □

Lemma A.6. *If Assumption A1–A4 hold, we have for all interior points $z \in \mathcal{S}$,*

$$\breve{\theta}(z) - \theta(z) - h^2 J(z) \xrightarrow{d} B_{(X,2)}^{-1} B_{(X,u)}, \quad (A.20)$$

where $J(z) = \mu_2[\theta^{(1)}(z) f^{(1)}(z)/f(z) + \theta^{(2)}(z)/2]$.

Proof. By adding and subtracting terms, we have

$$\begin{aligned}\breve{\theta}(z) &= \left(\sum_t X_t X_t^\top K_{tz}\right)^{-1}\sum_t X_t Y_t K_{tz} \\ &= \left(\sum_t X_t X_t^\top K_{tz}\right)^{-1}\sum_t X_t (X_t^\top\theta_t + u_t) K_{tz} \\ &= \theta(z) + \left(\sum_t X_t X_t^\top K_{tz}\right)^{-1} \\ &\quad\times \sum_t X_t\left\{X_t^\top[\theta_t - \theta(z)] + u_t\right\}K_{tz}.\end{aligned}$$

It follows that

$$\begin{aligned}\breve{\theta}(z) - \theta(z) &= \left[\sum_t X_t X_t^\top K_{tz}\right]^{-1}\sum_t X_t X_t^\top[\theta_t - \theta(z)]K_{tz} \\ &\quad + \left[\sum_t X_t X_t^\top K_{tz}\right]^{-1}\sum_t X_t u_t K_{tz} \\ &\equiv B_{1n}(z) + B_{2n}(z),\end{aligned} \quad (A.21)$$

where the definitions of $B_{1n}(z)$ and $B_{2n}(z)$ should be apparent.

By Lemma A.3, we have $(n^2 h)^{-1}\sum_t X_t X_t^\top K_{tz} = (n^2 h)^{-1}\sum_t X_t X_t^\top E(K_{tz}) + O_p\left((nh)^{-1/2}\right)$. By Lemma A.4, we have $\left(n^2 h^3\right)^{-1}\sum_t X_t X_t^\top[\theta_t - \theta(z)]K_{tz} = \left(n^2 h^3\right)^{-1}\sum_t X_t X_t^\top E\{[\theta_t - \theta(z)]K_{tz}\} + O_p\left(\left(nh^3\right)^{-1/2}\right)$. It follows that

$$\begin{aligned}B_{1n}(z) &= \left(\sum_t X_t X_t^\top K_{tz}\right)^{-1}\sum_t X_t X_t^\top[\theta_t - \theta(z)]K_{tz} \\ &= h^2\left\{\left[\left(n^2 h\right)^{-1}\sum_t X_t X_t^\top E(K_{tz})\right]^{-1} + O_p((nh)^{-1/2})\right\} \\ &\quad\times\left\{\left(n^2 h^3\right)^{-1}\sum_t X_t X_t^\top E[(\theta_t - \theta(z))K_{tz}] + O_p\left(\left(nh^3\right)^{-1/2}\right)\right\} \\ &= E[(\theta_t - \theta(z))K_{tz}]/E(K_{tz}) + O_p(\sqrt{h/n}) \\ &\equiv h^2 J(z) + O_p(h^4) + O_p(\sqrt{h/n}).\end{aligned} \quad (A.22)$$

By Lemmas A.3 and A.5, we have

$$B_{2n}(z) = \left(\sum_t X_t X_t^\top K_{tz}\right)^{-1} \sum_t X_t u_t K_{tz}$$
$$= \left\{\left[\left(n^2 h\right)^{-1} \sum_t X_t X_t^\top E\left(K_{tz}\right)\right]^{-1} + O_p((nh)^{-1/2})\right\}$$
$$\times \left[\left(n^2 h\right)^{-1} \sum_t X_t u_t E\left(K_{tz}\right) + O_p((nh)^{-1/2})\right]$$
$$= \left(n^{-2}\sum_t X_t X_t^T\right)^{-1} n^{-2}\sum_t X_t u_t + O_p((nh)^{-1/2})$$
$$\xrightarrow{d} B_{(X,2)}^{-1} B_{(X,u)}. \quad (A.23)$$

This completes the proof of Lemma A.6. □

Remark. We see that because u_t is $I(1)$, $\check{\theta}(z) - \theta(z)$ does not converge to zero.

Proof of Theorem 3.1. Combining (A.11) and (A.21)–(A.23), if $nh^5 = O(1)$, $h \to 0$ and $nh \to \infty$ as $n \to \infty$, we have

$$\sqrt{nh}(\tilde{\alpha}(z) - \alpha(z) - h^2 J(z))$$
$$= \sqrt{nh}\left[\left(\sum_t X_t X_t^\top K_{tz}\right)^{-1} \sum_t X_t u_t K_{tz} - \left(\sum_t X_t X_t^\top\right)^{-1}\sum_t X_t u_t\right] + O_p(\sqrt{h})$$
$$= \sqrt{nh}\left[\left(n^{-2}\sum_t X_t X_t^\top \frac{K_{tz}}{EK_{tz}}\right)^{-1} - \left(n^{-2}\sum_t X_t X_t^\top\right)^{-1}\right] n^{-2}\sum_t X_t u_t$$
$$+ \left(\frac{1}{n^2 h}\sum_t X_t X_t^\top K_{tz}\right)^{-1} \sum_t \frac{X_t u_t}{n}\frac{\check{K}_{tz}}{\sqrt{nh}} + O_p(\sqrt{h})$$
$$= \Pi_{n1} + \Pi_{n2} + O_p(\sqrt{h}), \quad (A.24)$$

where applying (A.14) and (A.19) and the continuous mapping theorem, we have

$$\Pi_{n2} = \left(\frac{1}{n^2 h}\sum_t X_t X_t^\top K_{tz}\right)^{-1} \sum_t \frac{X_t u_t}{n}\frac{\check{K}_{tz}}{\sqrt{nh}}$$
$$= \left[\left(\frac{EK_{tz}}{n^2 h}\sum_t X_t X_t^\top\right)^{-1} + O_p\left((nh)^{-1/2}\right)\right] \times \sum_t \frac{X_t u_t}{n}\frac{\check{K}_{tz}}{\sqrt{nh}}$$
$$\xrightarrow{d} \sqrt{\frac{\nu_0}{f(z)}} B_{(X,2)}^{-1} \int_0^1 B_X(r) B_u(r) \mathrm{d}W_K(r). \quad (A.25)$$

Next, let $\Delta_n = n^{-2}\sum_t X_t X_t^\top K_{tz}/EK_{tz} = \Delta_{n1} + \Delta_{n2}$, where $\Delta_{n1} = n^{-2}\sum_t X_t X_t^\top$ and $\Delta_{n2} = n^{-2}\sum_t X_t X_t^\top \check{K}_{tz}/E(K_{tz})$. By Theorem 4.3 in Poirier (1995, p. 627), we have $\Delta_n^{-1} = \Delta_{n1}^{-1} - \Delta_{n1}^{-1}\Delta_{n2}\left(\Delta_{n1}^{-1}\Delta_{n2} + I_d\right)^{-1}\Delta_{n1}^{-1}$. Then, we have

$$\Pi_{n1} = \sqrt{nh}\left(\Delta_n^{-1} - \Delta_{n1}^{-1}\right) n^{-2}\sum_t X_t u_t$$
$$= -\sqrt{nh}\Delta_{n1}^{-1}\Delta_{n2}\Delta_{n1}^{-1} n^{-2}\sum_t X_t u_t\left[1 + o_p(1)\right]$$

as $\Delta_{n1} = O_p(1)$ and $\Delta_{n2} = O_p\left((nh)^{-1/2}\right)$ by (A.14). Applying Lemmas 3.1 and A.3 and the continuous mapping theorem, we obtain

$$\Pi_{n1} \xrightarrow{d} -\sqrt{\frac{\nu_0}{f(z)}} B_{(X,2)}^{-1} \int_0^1 B_X(r) B_X(r)^\top \mathrm{d}W_K(r) B_{(X,2)}^{-1} B_{(X,u)}. \quad (A.26)$$

Taking all the results above completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Before proving Theorem 3.2, we first briefly discuss the need of the trimming function $M_{nt} = M_n(Z_t)$. Without loss of generality, we assume that $\mathcal{S} = [a, b]$, where $-\infty < a < b < \infty$, a and b are constants. To avoid boundary bias problem, we choose $M_{nt} = \mathbf{1}(a + \delta_n \le Z_t \le b - \delta_n)$, where $\mathbf{1}(A)$ is an indicator function taking value 1 if A holds true and 0 otherwise. δ_n is a non-stochastic sequence of real numbers that satisfies the conditions: $\delta_n \to 0$, $h/\delta_n \to 0$ as $n \to \infty$. For example, $\delta_n = \sqrt{h}$ is allowed. The use of the trimming function is needed (theoretically) to avoid the slow convergence rate of $\check{\theta}(z)$ when z falls into the boundary region; i.e., $[a, a+h] \cup [b-h, b]$.

For notational simplicity, we will omit the trimming function M_{nt} below. By (A.21), the proposed estimator is given by

$$\hat{\alpha}(z) = \check{\theta}(z) - n^{-1}\sum_{t=1}^n \check{\theta}(Z_t)$$
$$= \left[\theta(z) - n^{-1}\sum_{t=1}^n \theta(Z_t)\right] + \left[B_{1n}(z) - n^{-1}\sum_{t=1}^n B_{1n}(Z_t)\right] + \left[B_{2n}(z) - n^{-1}\sum_{t=1}^n B_{2n}(Z_t)\right]$$
$$= \Lambda_{n1}(z) + \Lambda_{n2}(z) + \Lambda_{n3}(z), \quad (A.27)$$

where $\Lambda_{n1}(z) = \theta(z) - n^{-1}\sum_{t=1}^n \theta(Z_t)$, $\Lambda_{n2}(z) = B_{1n}(z) - n^{-1}\sum_{t=1}^n B_{1n}(Z_t)$ and $\Lambda_{n1}(z) = B_{2n}(z) - n^{-1}\sum_{t=1}^n B_{2n}(Z_t)$. $\Lambda_{n1}(z) = \alpha(z) + O_p\left(n^{-1/2}\right)$ since $n^{-1}\sum_{t=1}^n \theta(Z_t) = E[\theta(Z_t)] + O_p(n^{-1/2}) = c_0 + O_p(n^{-1/2})$. Below, we will show $\Lambda_{n2}(z) = O_p\left(h^2\right)$ and $\Lambda_{n3}(z) = O_p\left((nh)^{-1/2}\right)$.

We first denote $\check{E}_t[g(Z_s, Z_t)] = \int g(Z_s, Z_t) f(Z_s)\, \mathrm{d}Z_s$, where $g(Z_s, Z_t)$ is a measurable function of (Z_s, Z_t). Then, let $K_{st} = K((Z_s - Z_t)/h)$,

$$\check{K}_{st} = K_{st} - \check{E}_t(K_{st}), \quad e_{st} = \check{K}_{st}/\check{E}_t(K_{st}),$$
$$\text{and} \quad \check{\theta}_{st} = (\theta_s - \theta_t) K_{ts} - \check{E}_t[(\theta_s - \theta_t) K_{st}]. \quad (A.28)$$

Obviously, $K_{tt} = \check{E}_t(K_{tt}) = K(0)$, but $e_{tt} = \check{\theta}_{tt} = 0$. Denote

$$\Delta_{n1} = n^{-2}\sum_{t=1}^n X_t X_t^\top \quad \text{and}$$
$$\Delta_{n2}(Z_t) = n^{-2}\sum_{s=1}^n X_s X_s^\top \check{K}_{st}/\check{E}_t(K_{st}). \quad (A.29)$$

By (A.22) and noting that $K_{st}/\check{E}_t(K_{st}) = 1 + \check{K}_{st}/\check{E}_t(K_{st})$, we have

$$n^{-1}\sum_{t=1}^{n}B_{1n}(Z_t)$$
$$=n^{-1}\sum_{t=1}^{n}\left(\sum_s X_sX_s^{\top}K_{st}\right)^{-1}\sum_s X_sX_s^{\top}(\theta_s-\theta_t)K_{st}$$
$$=n^{-1}\sum_{t=1}^{n}\left(n^{-2}\sum_s X_sX_s^{\top}+n^{-2}\sum_s X_sX_s^{\top}\frac{\check{K}_{st}}{\check{E}_t(K_{st})}\right)^{-1}n^{-2}$$
$$\times\sum_s X_sX_s^{\top}\frac{(\theta_s-\theta_t)K_{st}}{\check{E}_t(K_{st})}$$
$$=n^{-1}\sum_{t=1}^{n}\left\{\Delta_{n1}^{-1}-\Delta_{n1}^{-1}\Delta_{n2}(Z_t)\left[\Delta_{n1}^{-1}\Delta_{n2}(Z_t)+I_d\right]^{-1}\Delta_{n1}^{-1}\right\}n^{-2}$$
$$\times\sum_s X_sX_s^{\top}\frac{(\theta_s-\theta_t)K_{st}}{\check{E}_t(K_{st})}$$
$$=\Delta_{n1}^{-1}n^{-3}\sum_{t=1}^{n}\sum_s X_sX_s^{\top}\frac{(\theta_s-\theta_t)K_{st}}{\check{E}_t(K_{st})}-\Delta_{n1}^{-1}n^{-3}$$
$$\times\sum_{t=1}^{n}\Delta_{n2}(Z_t)\Delta_{n1}^{-1}\sum_s X_sX_s^{\top}\frac{(\theta_s-\theta_t)K_{st}}{\check{E}_t(K_{st})}\left[1+o_p(1)\right]$$
$$=n^{-1}\sum_{t=1}^{n}\frac{\check{E}_t\left[(\theta_s-\theta_t)K_{st}\right]}{\check{E}_t(K_{st})}-\Delta_{n1}^{-1}n^{-1}\sum_{t=1}^{n}\Delta_{n2}(Z_t)$$
$$\times\frac{\check{E}_t\left[(\theta_s-\theta_t)K_{st}\right]}{\check{E}_t(K_{st})}\left[1+o_p(1)\right]+o_p\left(h^2\right)$$
$$=n^{-1}\sum_{t=1}^{n}\left[I_d-\Delta_{n1}^{-1}\Delta_{n2}(Z_t)\right]\frac{\check{E}\left[(\theta_s-\theta_t)K_{st}\right]}{\check{E}_t(K_{st})}+o_p\left(h^2\right)$$
$$=h^2n^{-1}\sum_{t=1}^{n}J(Z_t)+o_p\left(h^2\right),$$

where Lemmas A.3 and A.9 given in Appendix A.2 are used. Therefore, taking this result with (A.22), we have

$$\Lambda_{n2}(z)=B_{1n}(z)-n^{-1}\sum_{t=1}^{n}B_{1n}(Z_t)$$
$$=h^2\left\{B(z)-E\left[J(Z_t)\right]\right\}+o_p\left(h^2\right)+O_p\left(\sqrt{h/n}\right). \quad \text{(A.30)}$$

Now, denoting

$$\pi_{n1}=n^{-2}\sum_{s=1}^{n}X_su_s \quad\text{and}$$
$$\pi_{n2}(Z_t)=n^{-2}\sum_{s=1}^{n}X_su_s\check{K}_{st}/\check{E}_t(K_{st}), \quad \text{(A.31)}$$

by (A.23), we have

$$n^{-1}\sum_{t=1}^{n}B_{2n}(Z_t)=n^{-1}\sum_{t=1}^{n}\left(\sum_s X_sX_s^{\top}K_{st}\right)^{-1}\sum_s X_su_sK_{st}$$
$$=n^{-1}\sum_{t=1}^{n}\left(\sum_s X_sX_s^{\top}+\sum_s X_sX_s^{\top}\frac{\check{K}_{st}}{\check{E}_t(K_{st})}\right)^{-1}$$
$$\times\sum_s X_su_s\frac{K_{st}}{\check{E}_t(K_{st})}$$
$$=n^{-1}\sum_{t=1}^{n}\left\{\Delta_{n1}^{-1}-\Delta_{n1}^{-1}\Delta_{n2}(Z_t)\left[\Delta_{n1}^{-1}\Delta_{n2}(Z_t)\right.\right.$$
$$\left.\left.+I_d\right]^{-1}\Delta_{n1}^{-1}\right\}n^{-2}\sum_s X_su_s\frac{K_{st}}{\check{E}_t(K_{st})}$$
$$=\Delta_{n1}^{-1}n^{-3}\sum_{t=1}^{n}\sum_s X_su_s\frac{K_{st}}{\check{E}_t(K_{st})}$$
$$-\Delta_{n1}^{-1}n^{-3}\sum_{t=1}^{n}\Delta_{n2}(Z_t)\Delta_{n1}^{-1}\sum_s X_su_s\frac{K_{st}}{\check{E}_t(K_{st})}\left[1+o_p(1)\right]$$
$$=\Delta_{n1}^{-1}\pi_{n1}+\Delta_{n1}^{-1}n^{-1}\sum_{t=1}^{n}\pi_{n2}(Z_t)$$
$$-\Delta_{n1}^{-1}n^{-1}\sum_{t=1}^{n}\Delta_{n2}(Z_t)\Delta_{n1}^{-1}\pi_{n1}\left[1+o_p(1)\right]$$
$$-\Delta_{n1}^{-1}n^{-1}\sum_{t=1}^{n}\Delta_{n2}(Z_t)\Delta_{n1}^{-1}\pi_{n2}(Z_t)\left[1+o_p(1)\right]$$
$$=\Delta_{n1}^{-1}\pi_{n1}+o_p\left(n^{-1}h^{-1}\right),$$

where the last equation results from Lemma A.9.

Taking this result with (A.24), we have

$$\sqrt{nh}\Lambda_{n3}(z)=\sqrt{nh}\left[B_{2n}(z)-n^{-1}\sum_{t=1}^{n}B_{2n}(Z_t)\right]$$
$$=\Pi_{n1}+\Pi_{n2}+o_p\left((nh)^{-1/2}\right)+O_p\left(\sqrt{h}\right), \quad \text{(A.32)}$$

which completes the proof of Theorem 3.2. □

A.2. Mathematical proof for Theorem 3.3 and Proposition 3.4

As we will frequently use Lemma 3.1 of Yoshihara (1976) in our proofs, we present Yoshihara's (1976) Lemma 3.1 below for readers' convenience.

Lemma A.7. *Let $z_t\in\mathcal{R}^p$ be a strictly stationary β-mixing process, $i_1<i_2<\cdots<i_k$ be arbitrary integers, $F(z_{i_1},\ldots,z_{i_k})$ the distribution function for $(z_{i_1},\ldots,z_{i_k})$. For any j $(0\le j\le k-1)$, $dF_j=dF(z_{i_1},\ldots,z_{i_j})dF(z_{i_{j+1}},\ldots,z_{i_k})$ and $dF_0=dF(z_{i_1},\ldots,z_{i_k})$. Let $h_n(z_{i_1},\ldots,z_{i_k})$ be a Borel function such that, for some $\lambda>0$, $\int\ldots\int_{R^{kp}}|h_n(z_{i1},\ldots,z_{i_k})|^{1+\lambda}dF(z_{i_1},\ldots,z_{i_j})dF(z_{i_{j+1}},\ldots,z_{i_k})\le M_n$. Then*

$$\left|\int\cdots\int_{R^{kp}}h_n(z_{i_1},\ldots,z_{i_k})dF_0-\int\cdots\int_{R^{kp}}h_n(z_{i_1},\ldots,z_{i_k})dF_j\right|$$
$$\le 4M_n^{1/(1+\lambda)}[\beta(i_{j+1}-i_j)]^{\lambda/(1+\lambda)}.$$

In this section, $\check{K}_{st}$, $\check{E}_t(K_{st})$, e_{st}, Δ_{n1}, $\Delta_{n2}(Z_t)$, π_{n1}, and $\pi_{n2}(Z_t)$ are defined the same way as in (A.28), (A.29) and (A.31). Also, we denote $\chi_n=\{X_1,\ldots,X_n,u_1,\ldots,u_n\}$. The following lemma is used by the Proofs of Theorem 3.3 and Lemma A.9.

Lemma A.8. *If Assumptions* A1–A5 *and* A2* *hold, we have*

$$\mathcal{D}_n\overset{def}{=}n^{-4}h^{-2}\sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{s=1}^{n}\sum_{s'=1}^{n}|cov(e_{st},e_{s't'})|=O\left(n^{-2}h^{-1}\right). \quad \text{(A.33)}$$

Proof. Let $\mathcal{D}_n=\mathcal{D}_{n1}+\mathcal{D}_{n2}+\mathcal{D}_{n3}+\mathcal{D}_{n4}$, where D_{n1} is for the case that $t=t'$ and $s=s'$, D_{n2} is for the case that $t=t'$ but $s\neq s'$, D_{n3} is for the case that $t\neq t'$ but $s=s'$, and D_{n4} is for the case that $t\neq t'$ and $s\neq s'$.

Firstly, as $\sup_{s,t}Var(e_{st})=O(h)$, we have

$$\mathcal{D}_{n1}=n^{-4}h^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n}Var(e_{st})=O\left(n^{-2}h^{-1}\right). \quad \text{(A.34)}$$

Secondly, we will show that $\mathcal{D}_{n2}$ and $\mathcal{D}_{n3}$ are both of $o\left(\left(n^2h\right)^{-1}\right)$. We will give the detailed proof for $\mathcal{D}_{n2}$ and omit the proof for $\mathcal{D}_{n3}$ due to similarity of the proofs. In this case, $\mathcal{D}_{n2} = n^{-4}h^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{s'\neq s}|cov(e_{st}, e_{s't})|$. Applying Berbee (1979) coupling method, we rewrite $cov(e_{st}, e_{s't})$ as follows

$$cov(e_{st}, e_{s't}) = \left[cov(e_{st}, e_{s't}) - cov^*(e_{st}, e_{s't})\right] + \left[cov^*(e_{st}, e_{s't}) - \widetilde{cov}(e_{st}, e_{s't})\right] + \widetilde{cov}(e_{st}, e_{s't}), \quad (A.35)$$

where the expectation for $cov^*(e_{st}, e_{s't})$ is taken with respect to the joint pdf $f_{s',s}(Z_{s'}, Z_s) f(Z_t)$, and the expectation for $\widetilde{cov}(e_{st}, e_{s't})$ is taken with respect to the joint pdf $f(Z_{s'}) f(Z_s) f(Z_t)$. Hence, $\widetilde{cov}(e_{st}, e_{s't}) = 0$. Therefore, applying Lemma A.7 gives

$$|cov(e_{st}, e_{s't})| \leq Mh^{2/(1+\lambda)} \times \left[\beta^{\lambda/(1+\lambda)}\left(\min\{t-s, t-s'\}\right) + \beta^{\lambda/(1+\lambda)}\left(s-s'\right)\right],$$

and it follows that

$$\begin{aligned}\mathcal{D}_{n2} &\leq Mn^{-4}h^{-2}\sum_{t}\sum_{s<t}\sum_{s'<s} h^{2/(1+\lambda)} \times \left[\beta^{\lambda/(1+\lambda)}(t-s) + \beta^{\lambda/(1+\lambda)}\left(s-s'\right)\right] \\ &= O\left(n^{-2}h^{(-2\lambda)/(1+\lambda)}\right) = o\left((n^2h)^{-1}\right), \end{aligned} \quad (A.36)$$

under Assumption A5(ii) with $\lambda \in (0, 1)$. Hence, we obtain $\mathcal{D}_{n2} = o\left((n^2h)^{-1}\right)$.

Finally, we will show $\mathcal{D}_{n4} \leq Mn^{-4}h^{-2}\sum_{t=4}^{n}\sum_{s=3}^{t-1}\sum_{t'=2}^{s-1}\sum_{s'=1}^{t'-1}|cov(e_{st}, e_{s't'})| = o\left((n^2h)^{-1}\right)$. Let $m = [1/h^{1/3}]$ be the integer part of $1/h^{1/3}$. We now bound $\mathcal{D}_{n4}$ separately for three cases: (i) $t-s' \leq m$; (ii) $t-s' > m$ and $t-t' \leq m$; (iii) $t-t' > m$. Conformably, we rewrite $\mathcal{D}_{n4} = \mathcal{D}_{n4,(i)} + \mathcal{D}_{n4,(ii)} + \mathcal{D}_{n4,(iii)}$.

For cases (i) and (ii), we have $|cov(e_{st}, e_{s't'})| \leq E|e_{st}e_{s't'}| + |E(e_{st})E(e_{s't'})| \leq Mh^2$. As the number of summations is of order nm^3 for case (i) and n^2m^2 for case (ii), we have

$$\begin{aligned}\mathcal{D}_{n4,(i)} + \mathcal{D}_{n4,(ii)} &= O\left(nm^3h^2 + n^2m^2h^2\right)n^{-4}h^{-2} \\ &= O\left(n^{-3}h^{-1}\right) + O\left(n^{-2}h^{-2/3}\right) \\ &= o\left(\left(n^2h\right)^{-1}\right). \end{aligned} \quad (A.37)$$

For case (iii), applying Lemma A.7, we have $|cov(e_{st}, e_{s't'})| \leq Mh^{2/(1+\lambda)}\beta^{\lambda/(1+\lambda)}\left(s-t'\right)$ for $t > s > t' > s'$. The leading term of $n^3h^2\mathcal{D}_{n4,(iii)}$ is given by

$$\begin{aligned}& h^{2/(1+\lambda)}\sum_{t=4}^{n}\sum_{s=3}^{t-1}\sum_{t'=2}^{t-m-1}\sum_{s'=1}^{t'-1}\left(s-t'\right)^{-\tau\lambda/(1+\lambda)} \\ &\sim h^{2/(1+\lambda)}\int_4^n\int_3^{x-1}\int_2^{x-m-1}\int_1^{z-1}(y-z)^{-\tau\lambda/(1+\lambda)} \times \mathrm{d}s\mathrm{d}z\mathrm{d}y\mathrm{d}x \\ &\sim n^{4-\tau\lambda/(1+\lambda)}h^{2/(1+\lambda)}. \end{aligned} \quad (A.38)$$

Taking (A.37) and (A.38) together, we obtain

$$\begin{aligned}\mathcal{D}_{n4} &= n^{-4}h^{-2}O\left(nm^3h^2 + n^2m^2h^2 + n^{4-\tau\lambda/(1+\lambda)}h^{2/(1+\lambda)}\right) \\ &= o\left(\left(n^2h\right)^{-1}\right) + O\left(\left(n^2h\right)^{-1}n^{2-\tau\lambda/(1+\lambda)}h^{(1-\lambda)/(1+\lambda)}\right) \\ &= o\left(\left(n^2h\right)^{-1}\right), \end{aligned}$$

if $\lambda \in (0, 1)$ and $\tau > 2\lambda^{-1}(1+\lambda)$. This completes the proof of Lemma A.8. □

Lemma A.9. *If Assumptions* A1–A5 *and* A2* *hold, we have*

$$\Gamma_{n1} = n^{-1}\sum_{t=1}^{n}\pi_{n2}(Z_t) = o_p\left(n^{-1}h^{-1}\right), \quad (A.39)$$

$$\Gamma_{n2} = n^{-1}\sum_{t=1}^{n}\Delta_{n2}(Z_t) = o_p\left(n^{-1}h^{-1}\right), \quad (A.40)$$

$$\Gamma_{n3} = n^{-3}\sum_{t=1}^{n}\sum_{s}X_sX_s^{\top}\check{\theta}_{st}/\check{E}_t(K_{st}) = o_p\left(n^{-1}\right). \quad (A.41)$$

Proof. Evidently, we only need to give the proofs for (A.39) and (A.41). To simplify notation, without loss of generality, we treat X_t as a scalar; otherwise the result holds for each element of Γ_{n1} (and Γ_{n3}).

Consider Γ_{n1} first. Under Assumption A2, we have $\sup_{1\leq t\leq n}\left|h^{-1}\check{E}_t(K_{st}) - f(Z_t)\right| = O_p\left(h^2\right)$. Hence,

$$\begin{aligned}\Gamma_{n1} &= \left(n^3h\right)^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}X_su_s\check{K}_{st}/f(Z_t)\left[1+o_p(1)\right] \\ &\stackrel{def}{=} \tilde{\Gamma}_{n1}\left[1+o_p(1)\right]. \end{aligned}$$

By Assumption A5(i), we have $E\left(\tilde{\Gamma}_{n1}|\chi_n\right) = \left(n^3h\right)^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}X_su_sE(e_{st})$. As $\check{E}(e_{st}) = 0$ and for $\lambda \in (0, 1)$, $\int\int\left[\check{K}_{st}/f(Z_t)\right]^{1+\lambda}f(Z_t)f(Z_s)\,\mathrm{d}Z_t\mathrm{d}Z_s \leq Mh$, applying Lemma A.7 gives

$$|E(e_{st})| \leq Mh^{1/(1+\lambda)}\beta(|t-s|)^{\lambda/(1+\lambda)}.$$

Hence, by

$$\sup_{1\leq t\leq n}|X_t| = O_p\left(\sqrt{n}\right) \quad \text{and} \quad \sup_{1\leq t\leq n}|u_t| = O_p\left(\sqrt{n}\right), \quad (A.42)$$

we obtain

$$\begin{aligned}E\left(\tilde{\Gamma}_{n1}|\chi_n\right) &= O_p\left(n^{-2}h^{-1}\right)\sum_{t=1}^{n}\sum_{s=1}^{n}h^{1/(1+\lambda)}\beta^{\lambda/(1+\lambda)}(|s-t|) \\ &= O_p\left(n^{-1}h^{-1}h^{1/(1+\lambda)}\right) = o_p\left((nh)^{-1}\right) \end{aligned} \quad (A.43)$$

because $n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}\beta^{\lambda/(1+\lambda)}(|s-t|) \leq M$ under Assumption A5(ii).

Now, we consider the conditional variance of $\tilde{\Gamma}_{n1}$ given χ_n. Again, by Assumption A5(i), we have

$$\begin{aligned}var(\tilde{\Gamma}_{n1}|\chi_n) &= n^{-6}h^{-2}\sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{s=1}^{n}\sum_{s'=1}^{n}X_su_su_{s'}X_{s'}cov(e_{st}, e_{s't'}) \\ &= O_p\left(n^{-4}h^{-2}\right)\sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{s=1}^{n}\sum_{s'=1}^{n}|cov(e_{st}, e_{s't'})|. \end{aligned}$$

Applying Lemma A.8 gives $var(\tilde{\Gamma}_{n1}|\chi_n) = O_p\left(n^{-2}h^{-1}\right)$. By Markov's inequality, we then obtain $\tilde{\Gamma}_{n1} - E\left(\tilde{\Gamma}_{n1}|\chi_n\right) = O_p\left(n^{-1}h^{-1/2}\right)$. By (A.43), we have $\tilde{\Gamma}_{n1} = o_p\left((nh)^{-1}\right)$. This completes the proof of (A.39).

Consider Γ_{n3}. Again, we have $\Gamma_{n3} = \left(n^3h\right)^{-1}\sum_{t=1}^{n}\sum_{s}X_s^2\check{\theta}_{st}/f(Z_t)\left[1+o_p(1)\right] \stackrel{def}{=} \tilde{\Gamma}_{n3}\left[1+o_p(1)\right]$. As $\int\int\left[\check{\theta}_{st}/f(Z_t)\right]^{1+\lambda}f(Z_t)f(Z_s)\,\mathrm{d}Z_t\mathrm{d}Z_s \leq Mh^{2+\lambda}$ for $\lambda \in (0, 1)$ and $\check{E}(\xi_{st}) = 0$, where $\xi_{st} = \check{\theta}_{st}/f(Z_t)$, applying Lemma A.7 gives

$$|E(\xi_{st})| \leq Mh^{(2+\lambda)/(1+\lambda)}\beta(|t-s|)^{\lambda/(1+\lambda)}.$$

Hence, by (A.42), we obtain

$$E\left(\tilde{\Gamma}_{n3}|\chi_n\right) = O_p\left(n^{-2}h^{-1}\right)\sum_{t=1}^{n}\sum_{s=1}^{n}h^{(2+\lambda)/(1+\lambda)}\beta^{\lambda/(1+\lambda)}\left(|s-t|\right)$$
$$= O_p\left(n^{-1}h^{1/(1+\lambda)}\right) = o_p\left(n^{-1}\right). \quad \text{(A.44)}$$

Moreover, we have

$$var(\tilde{\Gamma}_{n3}|\chi_n) = n^{-6}h^{-2}\sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{s=1}^{n}\sum_{s'=1}^{n}X_s^2X_{s'}^2cov(\xi_{st},\xi_{s't'})$$
$$= O_p\left(n^{-4}h^{-2}\right)\sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{s=1}^{n}\sum_{s'=1}^{n}|cov(\xi_{st},\xi_{s't'})|.$$

Following the proof of Lemma A.8, we show $var(\tilde{\Gamma}_{n3}|\chi_n) = O_p\left(n^{-2}h\right)$. Again, by Markov's inequality, we obtain $\tilde{\Gamma}_{n3} - E\left(\tilde{\Gamma}_{n3}|\chi_n\right) = O_p(n^{-1}h^{1/2})$. Therefore, by (A.44), we have $\tilde{\Gamma}_{n3} = o_p(n^{-1})$. This proves (A.41). Hence, the proof of Lemma A.9 is complete. □

Proof of Theorem 3.3. To simplify notation, we denote $\tilde{\alpha}_t = \tilde{\alpha}(Z_t)$ and $\check{\theta}_t = \check{\theta}(Z_t)$. By (2.14), we have $\tilde{c}_0 = \left(\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)^{-1}\sum_{t=2}^{n}\zeta_t\Delta\tilde{Y}_t$, where $\Delta\tilde{Y}_t = Y_t - Y_{t-1} - X_t^\top\tilde{\alpha}_t + X_{t-1}^\top\tilde{\alpha}_{t-1} = X_t^\top(\alpha_t-\tilde{\alpha}_t) - X_{t-1}^\top(\alpha_{t-1}-\tilde{\alpha}_{t-1}) + \zeta_t^\top c_0 + \epsilon_t$. Therefore, we have

$$\tilde{c}_0 = c_0 + \left(\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)^{-1}\sum_{t=2}^{n}\zeta_t\epsilon_t + \left(\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)^{-1}$$
$$\times\sum_{t=2}^{n}\zeta_t\left[X_t^\top(\alpha_t-\tilde{\alpha}_t) - X_{t-1}^\top(\alpha_{t-1}-\tilde{\alpha}_{t-1})\right]$$
$$\stackrel{def}{=} c_0 + A_{n1} + \left(n^{-1}\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)^{-1}A_{n2},$$

where $A_{n1} = \left[\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right]^{-1}\sum_{t=2}^{n}\zeta_t\epsilon_t$ and $A_{n2} = n^{-1}\sum_{t=2}^{n}\zeta_t\left[X_t^\top(\alpha_t-\tilde{\alpha}_t) - X_{t-1}^\top(\alpha_{t-1}-\tilde{\alpha}_{t-1})\right]$. $A_{n1} = O_p\left(n^{-1/2}\right)$ given $E(\epsilon_t|\zeta_t) = 0$ and $Var\left(n^{-1}\sum_{t=2}^{n}\zeta_t\epsilon_t\right) = O\left(n^{-1}\right)$ by McLeish's strong mixing inequality. Below, we will show $\left(n^{-1}\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)^{-1}A_{n2} = O_p\left(h^2\right) + O_p\left((nh)^{-1/2}\right)$.

Define $\Delta_{n3} = n^{-2}\sum_t X_tX_t^\top\alpha_t$. Then (A.11) can be written as $\hat{\theta}_0 = c_0 + \Delta_{n1}^{-1}(\pi_{n1}+\Delta_{n3})$. Combining this with (A.21), we have

$$X_t^\top(\alpha_t-\tilde{\alpha}_t) - X_{t-1}^\top(\alpha_{t-1}-\tilde{\alpha}_{t-1})$$
$$= X_t^\top\left(-B_{1n}(Z_t) - B_{2n}(Z_t) + \Delta_{n1}^{-1}\pi_{n1} + \Delta_{n1}^{-1}\Delta_{n3}\right)$$
$$- X_{t-1}^\top\left(-B_{1n}(Z_{t-1}) - B_{2n}(Z_{t-1}) + \Delta_{n1}^{-1}\pi_{n1} + \Delta_{n1}^{-1}\Delta_{n3}\right)$$
$$= \zeta_t^\top\Delta_{n1}^{-1}(\pi_{n1}+\Delta_{n3}) - X_t^\top\left[B_{1n}(Z_t)+B_{2n}(Z_t)\right]$$
$$+ X_{t-1}^\top\left[B_{1n}(Z_{t-1})+B_{2n}(Z_{t-1})\right].$$

Substituting the above result into A_{n2}, we can write $A_{n2} = A_{n2,1} + A_{n2,2} + A_{n2,3}$, where $A_{n2,1} = n^{-1}\sum_t\zeta_t\Delta_{n1}^{-1}(\pi_{n1}+\Delta_{n3})$, $A_{n2,2} = -n^{-1}\sum_{t=2}^{n}\zeta_t\left[X_t^\top B_{1n}(Z_t) - X_{t-1}^\top B_{1n}(Z_{t-1})\right]$, and $A_{n2,3} = -n^{-1}\sum_{t=2}^{n}\zeta_t\left[X_t^\top B_{2n}(Z_t) - X_{t-1}^\top B_{2n}(Z_{t-1})\right]$.

Firstly, applying Lemma A.1 yields

$$\left(n^{-1}\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)^{-1}A_{n2,1} = \Delta_{n1}^{-1}\pi_{n1} + \Delta_{n1}^{-1}\Delta_{n3}$$
$$= \Delta_{n1}^{-1}\pi_{n1} + O_p\left(n^{-1/2}\right). \quad \text{(A.45)}$$

Secondly, we obtain

$$A_{n2,2} = -n^{-1}\sum_{t=2}^{n}\zeta_t\left[X_t^\top B_{1n}(Z_t) - X_{t-1}^\top B_{1n}(Z_{t-1})\right]$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top B_{1n}(Z_t) - n^{-1}\zeta_nX_n^\top B_{1n}(Z_n)$$
$$+ n^{-1}\zeta_2X_1^\top B_{1n}(Z_1)$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top B_{1n}(Z_t)$$
$$+ O_p\left(n^{-1/2}h^2\right) \quad \text{by (A.22)}$$
$$= n^{-1}h^2\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top B(Z_t)$$
$$+ o_p\left(h^2\right) + O_p\left(n^{-1/2}h^2\right) = O_p\left(h^2\right), \quad \text{(A.46)}$$

where following the proof of Lemma A.3, we can show $n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top B(Z_t) = O_p(1)$ given the fact that $E[(\zeta_{t+1}-\zeta_t)B(Z_t)^\top] = 0$ under Assumption A5(i).

Similar manipulations give

$$A_{n2,3} = -n^{-1}\sum_{t=2}^{n}\zeta_t\left[X_t^\top B_{2n}(Z_t) - X_{t-1}^\top B_{2n}(Z_{t-1})\right]$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top B_{2n}(Z_t) - n^{-1}\zeta_nX_n^\top B_{2n}(Z_n)$$
$$+ n^{-1}\zeta_2X_1^\top B_{2n}(Z_1)$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top B_{2n}(Z_t) + O_p\left(n^{-1/2}\right) \quad \text{by (A.23)}$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top\left(\sum_{s=1}^{n}X_sX_s^\top K_{st}\right)^{-1}\sum_{s=1}^{n}X_su_sK_{st}$$
$$+ O_p\left(n^{-1/2}\right)$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top\left[\left(n^2h\right)^{-1}\sum_{s=1}^{n}X_sX_s^\top\check{E}_t(K_{st})\right]^{-1}$$
$$\times\left(n^2h\right)^{-1}\sum_{s=1}^{n}X_su_s\left[\check{K}_{st} + \check{E}_t(K_{st})\right]\left[1+o_p(1)\right]$$
$$+ O_p\left(n^{-1/2}\right)$$
$$= \left(A_{n2,3,1} + A_{n2,3,2}\right)\left[1+o_p(1)\right] + O_p\left(n^{-1/2}\right),$$

where

$$A_{n2,3,1} = n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top$$
$$\times\left[\left(n^2h\right)^{-1}\sum_{s=1}^{n}X_sX_s^\top\check{E}_t(K_{st})\right]^{-1}$$
$$\times\left(n^2h\right)^{-1}\sum_{s=1}^{n}X_su_s\check{E}_t(K_{st})$$
$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top\Delta_{n1}^{-1}\pi_{n1}$$
$$= -\left(n^{-1}\sum_{t=2}^{n}\zeta_t\zeta_t^\top\right)\Delta_{n1}^{-1}\pi_{n1} + O_p\left(n^{-1/2}\right), \quad \text{(A.47)}$$

and

$$A_{n2,3,2} = n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top \times \left[\left(n^2h\right)^{-1}\sum_{s=1}^{n}X_sX_s^\top \check{E}_t(K_{st})\right]^{-1} \times \left(n^2h\right)^{-1}\sum_{s=1}^{n}X_su_s\check{K}_{st}$$

$$= n^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t^\top\Delta_{n1}^{-1}\left(n^2h\right)^{-1} \times \sum_{s=1}^{n}X_su_s\check{K}_{st}/f(Z_t)\left[1+o_p(1)\right] \stackrel{def}{=} \Gamma_n\left[1+o_p(1)\right]$$

as $\sup_{1\le t\le n}\left|h^{-1}\check{E}_t(K_{st})-f(Z_t)\right| = O\left(h^2\right)$ by Assumptions A2 and A3. Below, we will show $\Gamma_n = O_p\left((nh)^{-1/2}\right)$.

In the following proof, without loss of generality, we again treat X_t as a scalar; otherwise the result holds for each element of Γ_n. By Assumption A5(i), we have

$$E\left(\Gamma_n|\chi_n\right) = \left(n^3h\right)^{-1}\sum_{t=2}^{n-1}(\zeta_{t+1}-\zeta_t)X_t\Delta_{n1}^{-1}\sum_{s=1}^{n}X_su_sE\left(e_{st}\right),$$

where applying Lemma A.7 gives

$$|E(e_{st})| \le Mh^{1/(1+\lambda)}\beta\left(|t-s|\right)^{\lambda/(1+\lambda)}.$$

Hence, applying (A.42), we obtain

$$|E(\Gamma_n|\chi_n)| = O_p\left(n^{-3/2}h^{-1}\right)\sum_{t=2}^{n-1}|\zeta_{t+1}-\zeta_t| \times \sum_{s=1}^{n}h^{1/(1+\lambda)}\beta^{\lambda/(1+\lambda)}\left(|s-t|\right)$$

$$= O_p\left(n^{-3/2}h^{-1}\right)h^{1/(1+\lambda)}\sum_{t=2}^{n-1}|\zeta_{t+1}-\zeta_t| \times \sum_{s=1}^{n}\beta^{\lambda/(1+\lambda)}\left(|s-t|\right)$$

$$= O_p\left(n^{-1/2}h^{-1}h^{1/(1+\lambda)}\right) = o_p\left((nh)^{-1/2}\right) \quad (A.48)$$

because $E\,|\zeta_{t+1}-\zeta_t| < M$ by Assumption A1 and $n^{-1}\sum_{t=2}^{n-1}\sum_{s=1}^{n}\beta^{\lambda/(1+\lambda)}\left(|s-t|\right) \le M$ under Assumption A5(ii).

Now, we consider the conditional variance of Γ_n given χ_n. Again, by Assumption A5(i), we have

$$D_n \stackrel{def}{=} var(\Gamma_n|\chi_n) = n^{-6}h^{-2}\sum_{t=2}^{n-1}\sum_{t'=2}^{n-1}\sum_{s=1}^{n}\sum_{s'=1}^{n}(\zeta_{t+1}-\zeta_t) \times X_tX_su_su_{s'}X_{s'}X_{t'}\left(\zeta_{t'+1}-\zeta_{t'}\right)cov(e_{st},e_{s't'})$$

$$= O_p\left(n^{-3}h^{-2}\right)\sum_{t=2}^{n-1}\sum_{t'=2}^{n-1}\sum_{s=1}^{n}\sum_{s'=1}^{n}|cov(e_{st},e_{s't'})|, \quad (A.49)$$

where the last equation is obtained by (A.42) and $\sup_{1\le t\le n} E\,\|\zeta_{t+1}-\zeta_t\|_2 \le M < \infty$. Applying Lemma A.8 gives $D_n = O_p(n^{-1}h^{-1})$. By Markov's inequality, we obtain $\Gamma_n - E\left(\Gamma_n|\chi_n\right) = O_p((nh)^{-1/2})$. By (A.48), we have $\Gamma_n = O_p((nh)^{-1/2})$. Hence, we have $A_{n2,3,2} = O_p((nh)^{-1/2})$.

Taking together this result with (A.45)–(A.47) yields

$$\tilde{c}_0 - c_0 = O_p(h^2 + (nh)^{-1/2}). \quad (A.50)$$

Also, Theorem 3.1 indicates $\tilde{\alpha}(z) = \alpha(z) + O_p(h^2+(nh)^{-1/2})$. We therefore obtain

$$\tilde{\theta}(z) = \tilde{\alpha}(z) + \tilde{c}_0 = \theta(z) + O_p(h^2+(nh)^{-1/2}),$$

which shows that $\tilde{\theta}(z)$ is a consistent estimator of $\theta(z)$ under Assumption A4.

The proof for the consistency of $\hat{c}_0$ and $\hat{\theta}(z)$ are similar, and thus are omitted here. This completes the proof of Theorem 3.3. □

Proof of Proposition 3.4. $\check{\theta}(z)$ is still defined as in (2.6). When there is a (varying) intercept term $\gamma(Z_t)$, $\check{\theta}(z)$ contains the following extra term:

$$\mathcal{M}_n \stackrel{def}{=} \left(\sum_t X_tX_t^TK_{tz}\right)^{-1}\sum_t X_tK_{tz}\gamma(Z_t).$$

Now,

$$\sqrt{n}\mathcal{M}_n = \sqrt{n}\left[\sum_t X_tX_t^\top E(K_{tz})\right]^{-1} \times \sum_t X_tE\left[K_{tz}\gamma(Z_t)\right]\left[1+o_p(1)\right]$$

$$= \left(n^{-1}\sum_t\frac{X_t}{\sqrt{n}}\frac{X_t^\top}{\sqrt{n}}\right)^{-1} \times n^{-1}\sum_t\frac{X_t}{\sqrt{n}}[E(K_{tz})]^{-1}E\left[K_{tz}\gamma(Z_t)\right]\left[1+o_p(1)\right]$$

$$\stackrel{d}{\to} \left\{\left[\int_0^1 B_X(r)B_X(r)^\top\,\mathrm{d}r\right]^{-1}\int_0^1 B_X(r)\mathrm{d}r\right\}\gamma(z) = O_p(1).$$

It follows that $\mathcal{M}_n = \left(\sum_t X_tX_t^\top K_{tz}\right)^{-1}\sum_t X_tK_{tz}\gamma(Z_t) = O_p\left(n^{-1/2}\right)$ uniformly over $z\in\mathcal{S}$ since $\sup_{z\in\mathcal{S}}|\gamma(z)| \le M < \infty$. Hence, we have shown that $\mathcal{M}_n = O_p(n^{-1/2}) = o_p(h^2+(nh)^{-1/2})$.

Therefore, the inclusion of an additive bounded function $\gamma(\cdot)$ in the regression model is asymptotically negligible as far as the estimation of $\theta(\cdot)$ is concerned. This completes the proof of Proposition 3.4. □

References

Aloui, C., 2007. Price and volatility spillovers between exchange rates and stock indexes for the pre- and post-euro period. Quantitative Finance 7, 669–685.

Athreya, K.B., Pantula, S.G., 1986. A note on strong mixing of ARMA processes. Statistics & Probability Letters 4, 187–190.

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3–30.

Bali, T.G., Wu, L., 2010. The role of exchange rates in intertemporal risk-return relations. Journal of International Money and Finance 29, 1670–1686.

Barnett, W.A., Serletis, A., 1990. A dispersion-dependency diagnostic test for aggregation error: with applications to monetary economics and income distribution. Journal of Econometrics 43, 5–34.

Berbee, H.C.P., 1979. Random walks with stationary increments and renewal Theory. Math. Cent. Tracts, Amstardam.

Blundell, R., Stoker, T.M., 2005. Heterogeneity and aggregation. Journal of Economic Literature 43, 347–391.

Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long-memory in stock market volatility. Journal of Econometrics 73, 151–184.

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

Cai, Z., Li, Q., Park, J., 2009. Functional-Coefficient Models for Nonstationary Time Series Data. Journal of Econometrics 148, 101–113.

Christensen, B.J., Nielsen, M.O., 2007. The effect of long memory in volatility on stock market fluctuations. Review of Economics and Statistics 89, 684–700.

Davidson, J., 2004. Moment and memory properties of linear conditional heteroscedastic models, and a new model. Journal of Business and Economic Statistics 22, 16–29.

de Jong, R.M., Davidson, J., 2000. The functional central limit theorem and weak convergence to stochastic integrals I. Econometric Theory 16, 621–642.
Fan, J., Cai, Z., Li, R., 2000. Efficient estimation and inferences of varying-coefficient models. Journal of the American Statistical Association 95, 465–501.
Fan, J., Yao, Q., Cai, Z., 2003. Adaptive varying coefficient linear models. Journal of the Royal Statistical Society (Series B) 65, 57–80.
Gavin, M., 1989. The stock market and exchange rate dynamics. Journal of International Money and Finance 8, 181–200.
Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.
Juhl, T., 2005. Functional coefficient models under unit root behavior. Econometrics Journal 8, 197–213.
Karolyi, G., 1995. A multivariate GARCH model of international transmissions of stock returns and volatility: the case of the United States and Canada. Journal of Business & Economic Statistics 13, 11–25.
Koutmos, G., Booth, G.G., 1995. Asymmetric volatility transmission in international stock markets. Journal of International Money and Finance 14, 747–762.
Li, Q., Huang, C., Li, D., Fu, T., 2002. Semiparametric smooth coefficient models. Journal of Business and Economic Statistics 20, 412–422.
Mark, N.C., 2001. International Marcoeconomics and Finance. Blackwell, New York.
McLeish, D.L., 1975. A maximal inequality and dependent strong laws. The Annals of Probability 3, 829–839.
Phillips, P.C.B., 2009. Local limit theory and spurious nonparametric regression. Econometric Theory 25, 1466–1497.
Poirier, D.J., 1995. Intermediate Statistics and Econometrics: a Comparative Approach. The MIT Press.
Ray, B.K., Tsay, R.S., 2000. Long-range dependence in daily stock volatilities. Journal of Business & Economic Statistics 18, 254–262.
Sun, Y., Hsiao, C., Li, Q., 2010. Volatility spill over effects: empirical evidence based semiparametric analysis (manuscript).
Taylor, A.M., Taylor, M.P., 2004. The purchasing power parity debate. Journal of Economic Perspective 18, 135–158.
White, H., Domowitz, I., 1984. Nonlinear regression with dependent observations. Econometrica 52, 143–162.
Wooldridge, J.M., White, H., 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. Econometric Theory 4, 210–230.
Xiao, Z., 2009. Functional coefficient cointegration models. Journal of Econometrics 152, 81–92.
Xu, Z.H., 2003. Purchasing power parity, price indices, and exchange rate forecasts. Journal of International Money and Finance 22, 105–130.
Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 35, 237–252.