Journal of Econometrics 157 (2010) 165–178

Contents lists available at ScienceDirect

Journal of Econometrics

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

A semiparametric cointegrating regression: Investigating the effects of age distributions on consumption and saving☆

Joon Y. Park [a,b,*], Kwanho Shin [c], Yoon-Jae Whang [d]

[a] Department of Economics, Indiana University, United States
[b] School of Economics, Sungkyunkwan University, Republic of Korea
[c] Department of Economics, Korea University, Republic of Korea
[d] Department of Economics, Seoul National University, Republic of Korea

ARTICLE INFO

Article history:
Available online 31 October 2009

JEL classification:
C22

Keywords:
Cointegration
Semiparametric regression
Series estimation
Age distributions
Consumption and savings rate

ABSTRACT

We consider a semiparametric cointegrating regression model, for which the disequilibrium error is further explained nonparametrically by a functional of distributions changing over time. The paper develops the statistical theories of the model. We propose an efficient econometric estimator and obtain its asymptotic distribution. A specification test for the model is also investigated. The model and methodology are applied to analyze how an aging population in the US influences the consumption level and the savings rate. We find that the impact of age distribution on the consumption level and the savings rate is consistent with the life-cycle hypothesis.

1. Introduction

Nonlinear and nonparametric models have drawn much attention over the past decade, and it now seems generally agreed that many important economic relationships are intrinsically complex and cannot be modeled effectively using simple linear parametric models. See, for example, Pagan and Ullah (1999) and Granger and Teräsvirta (1993). Many empirical economic relationships, which were formerly represented by linear and parametric models, are indeed being modeled and estimated within nonlinear and nonparametric frameworks. Nonlinear and nonparametric approaches certainly provide more flexibility and accommodate a broader class of models. The generality, however, comes at a cost. It is well known that estimated nonlinear and nonparametric models often have relatively poor finite sample performance and/or slower rates of convergence asymptotically. The cost can be prohibitively high in nonstationary time series settings, as is well demonstrated in Park (2005), and this forces us to look for reasonable compromises. The class of partially nonlinear and nonparametric models certainly offers one such compromise.

We consider in this paper a semiparametric model that is partially nonlinear and nonparametric. The linear part of the model specifies a cointegrating relationship among a set of integrated variables in a parametric regression form, whereas the nonlinear part describes the effect of a functional regressor which we model nonparametrically by introducing a response function. In the paper, we consider a series estimation of the model and establish its statistical theories. In particular, we develop an efficient econometric estimator of the model and obtain its asymptotic distribution. A specification test, which we may use to check the adequacy of the model, is also introduced and analyzed. The efficient estimator is asymptotically Gaussian, and the proposed specification test has a limit chi-square distribution. Our statistical theories in the paper are therefore all Gaussian. The assumptions used in the paper are mild and allow for a wide class of integrated processes and functional regressors that may appear in practical applications.

The model and methodology developed in the paper can be used to analyze various time series macroeconomic models from a new perspective. In particular, our approach allows for modeling economic relationships over time that are also affected by the cross-sectional distributions in each period. Typically, individual specific variables such as age are averaged out when we investigate the relationships among a set of aggregate variables. Consequently,

☆ We are grateful to a Guest Editor and two anonymous referees for many useful comments. We also like to thank Yoosoon Chang, Oliver Linton and Akiko Tamura for helpful discussions and comments, and Qi Li and Kwan Koo Yun for their patience, encouragement and support. This research is supported financially by a Hosei University Research Grant, which we gratefully acknowledge.

* Corresponding author at: Department of Economics, Indiana University, United States.
E-mail address: jpark@tamu.edu (J.Y. Park).

0304-4076/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.jeconom.2009.10.032

we overlook the interaction between the levels of aggregate macroeconomic variables and their distributions since we expect that the distribution of age itself also matters in determining the levels of the aggregate macroeconomic variables. For example, the life cycle hypothesis suggests that the savings rate should vary along the stages of the life-cycle. Then the aggregate savings rate should depend on the age distribution, in addition to other typical aggregate macroeconomic variables such as the level of income or income growth.

When we apply our methodology to analyze the quarterly consumption level and savings rate from 1959:1 to 2002:3, we find that the impact of age distribution on the consumption level and the savings rate takes a U-shape and an inverted U-shape respectively, which is consistent with the life-cycle hypothesis. In other words, we find that consumption is lower for middle-agers and higher for both sides of young-agers and old-agers, while the savings rate responds in a mirrored fashion to the age distribution, being highest for middle-agers. This is in contrast with the previous studies based on parametric approaches, which repeatedly produced a U-shaped impact curve for the savings rate. We point out the reasons why our nonparametric approach outperforms the conventional parametric approaches. Of course, our nonparametric approach also has some well known costs: a slower rate of convergence for the estimates of model parameters and the dependency of inferential results on the choice of basis functions and truncation parameters.

The rest of the paper is organized as follows. Section 2 introduces the model and assumptions. The details of the model are given with the required assumptions, and the series estimation method is introduced to estimate the model. The basic statistical theories are developed in Section 3. Various approximation results are provided to facilitate the asymptotic analysis of our model. The limit distributions of the model estimators are also given, and the asymptotics of the long-run error variance estimator are developed. Section 4 provides the efficient estimation method based on the CCR transformation, and the specification test using the variable addition approach that suits our model conveniently well. Section 5 reports the empirical results for the application of the model; these analyze the effect of age distribution on the consumption level and savings rate. Section 6 concludes the paper, and the mathematical proofs are in the Appendix.

A word on notation. As usual, we use $|\cdot|$ to denote the modulus. If applied to vectors or matrices, the notation denotes the maximum of the moduli of their components. For a vector $x = (x_i)$, $\|\cdot\|$ signifies the standard Eucleadian norm, i.e., $\|x\|^2 = \sum x_i^2$. On the other hand, the notation is used to denote the operator norm for a matrix $A = (a_{ij})$. We therefore have $\|A\| = \sup \|Ax\|/\|x\|$. It is well known that $\|A\|^2$ is dominated by the maximum eigenvalue of $A'A$, and consequently, bounded in particular by $\|A\|^2 \leq \mathrm{tr}(A'A) = \sum a_{ij}^2$. We also use the same notation, $\|\cdot\|$, to signify the supremum norm for continuous functions defined on a compact interval. This should cause no confusion. For functions that are vector-valued, the notation denotes the maximum of the supremum norms of the component functions. Standard notations such as o_p and O_p for stochastic orders, and $\to_p$ and $\to_d$ for convergences of random sequences, are used without any reference. Moreover, equality in distribution is denoted by $=_d$, and $\mathbb{R}$ denotes the set of real numbers. Finally, we denote by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ respectively the smallest and largest eigenvalues of a matrix A.

2. The model and assumptions

We consider the regression model given by

$$y_t = \nu_t + x_t'\beta + u_t, \tag{1}$$

where

$$\nu_t = \int_{-\infty}^{\infty} f_t(s)g(s)\mathrm{d}s \tag{2}$$

for $t = 1, \ldots, n$. The model (1) is partially linear, consisting of both linear and nonlinear parts. The regressor (x_t) has a linear relationship with the regressand (y_t), which is specified parametrically. On the other hand, in the nonlinear part (ν_t), the regressor (f_t) is given as a set of functional observations, to which the regressand (y_t) responds nonlinearly as given in (2). The function g may be interpreted as a *response function*, which measures the effect of (f_t) on (y_t) in a nonparametric fashion. Throughout the paper, we assume that the functional regressor (f_t) is deterministic.[1] As usual, (u_t) denotes the regression error.

In what follows, we assume that (x_t) is a vector integrated process and (u_t) a stationary process, so that our model (1) represents a semiparametric cointegrating regression. Let (x_t) be m-dimensional. Further, we let $v_t = \Delta x_t$,

$$w_t = (u_t, v_t')', \tag{3}$$

and assume the following.

Assumption 1. Let

$$w_t = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i}$$

and we assume that $\Phi(1)$ is nonsingular and $\sum i|\Phi_i| < \infty$, and that (ε_t) are iid with $\Psi = \mathbb{E}\varepsilon_t\varepsilon_t' > 0$ and $\mathbb{E}|\varepsilon_t|^p < \infty$ for some $p > 4$.

The process (w_t) is therefore assumed to be a general linear process. We only require the standard summability condition for (Φ_i) and the moment condition for (ε_t), which are routinely imposed in the time series literature. The iid assumption on the innovation sequence (ε_t) is restrictive, since it does not allow for more general linear processes driven by ARCH-type innovations having conditional heteroskedasticity. The assumption does not appear to be crucial, and is imposed here to ease the proofs of our subsequent theoretical results.

Let

$$a = (p-2)/2p,$$

where p is the maximal order of the existing moment for (ε_t) introduced in Assumption 1. Under Assumption 1, we have by the result in Park and Hahn (1999):

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} w_t = \Phi(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t + O_p(n^{-a})$$

uniformly in $r \in [0, 1]$. Therefore, due to the strong approximation obtained by, e.g., Einmahl (1989), we have the following.

Lemma 1. *Let Assumptions 1 and 2 hold. Then there exists a stochastic process W_n defined on $[0, 1]$ such that*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} w_t =_d W_n(r) \tag{4}$$

and

$$\sup_{0\leq r\leq 1}|W_n(r) - W(r)| = O_p(n^{-a}),$$

where W is a vector Brownian motion with variance $\Omega = \Phi(1)\Psi\Phi(1)'$.

[1] We may of course let (f_t) be random and given exogenously.

The normalized partial sum of (w_t) may therefore be represented, up to the distributional equivalence, by a stochastic process that can be well approximated by a Brownian motion. Note that the approximation error approaches $\sqrt{n}$ as $p \to \infty$. For the development of our theory, we let by convention the distributional equality in (4) be the usual equality. Our main purpose is to obtain the asymptotic results for the model given by (1) and (2), and therefore the convention is innocuous. However, it allows us to avoid repetitious embeddings of the partial sum of (w_t) into the probability space where W_n and W are defined, and greatly simplifies the exposition of the paper.

For the functional regressor (f_t), we assume the following.

Assumption 2. The family of functions (f_t) has the common support T, which is a compact subset of $\mathbb{R}$. Moreover, we have

$$f_t(s) = K\left(\frac{t}{n}, s\right),$$

where K is Lipschitz on $[0, 1] \times T$.

The representation of (f_t) as in Assumption 2 is not absolutely necessary, but it greatly simplifies the proofs of our subsequent theoretical results. In particular, it guarantees that the mapping $t \mapsto f_t$ is continuous with respect to t, and (f_t) has a common compact support T. In the empirical application that we consider in the paper, (f_t) denotes the densities of age distributions that are changing over time. For the application, it seems reasonable to require the conditions in Assumption 2.

We assume that the response function g is smooth. More precisely, we require the following.

Assumption 3. We assume that g is q-times differentiable with bounded derivative on T for some $q > 1$.

Assumption 4. $q > \max((3p - 2)/2(p - 2), 7/4)$.

Assumptions 3 and 4 imply that g needs to be three times differentiable and has bounded derivative, if $p \geq 4$ as assumed in Assumption 1. When higher moments of (ε_t) exist, we may allow for a less smooth function g. However, we must require g to be at least twice differentiable.

The response function g is well identified. To see this, first note that it is uniquely identified a.s. as a vector in the Hilbert space $L^2(T)$ of square integrable functions defined on the subset T of $\mathbb{R}$ with inner product

$$\langle f_t, g\rangle = \int_T f_t(s)g(s)\mathrm{d}s,$$

as long as (f_t) does not have support that is a proper subset of $L^2(T)$. This will be assumed throughout the paper. Given our assumption on the continuity, g is therefore uniquely identified pointwise.

For the estimation of g, we consider the estimand given by

$$\Pi(g) = (g(s_1), \ldots, g(s_d))', \tag{5}$$

where we denote by s_i's the numbers on T. Let $C(T)$ signify the class of continuous functions defined on T endowed with the uniform norm. It is easy to see that the functional $\Pi : C(T) \to \mathbb{R}^d$ is linear and bounded. Of course, linearity of the functional Π implies $\Pi(c_1g_1 + c_2g_2) = c_1\Pi(g_1) + c_2\Pi(g_2)$ for $c_1, c_2 \in \mathbb{R}$ and $g_1, g_2 \in C(T)$, and boundedness of Π implies

$$\|\Pi(g)\| \leq c\|g\|$$

for $g \in C(T)$ and $c \in \mathbb{R}$.[2] Though we primarily consider the estimand in (5), many other types of estimand may also be allowed. For this, we only need to make some obvious modifications in our technical assumptions.

If g is sufficiently smooth, we may approximate it by a series of polynomial and/or trigonometric functions on T. Of course, it is also possible to use their mixtures such as the Fourier flexible form (FFF) series of functions, which include the linear and quadratic functions to the sine and cosine trigonometric pairs. In the subsequent development of our theory, we just denote by (φ_i) an orthonomal basis of polynomial and/or trigonometric functions on T that we use to approximate g, and let g_κ be an approximation of g given by a finite sum of the series functions $\varphi_1, \ldots, \varphi_\kappa$. In estimating our empirical model, we use the trigonometric pairs

$$(\cos \lambda_i s, \sin \lambda_i s)_{i\geq 1}$$

with $\lambda_i = 2\pi i$, after we transform the functional regressor (f_t) and the response function g appropriately so that we may effectively assume $T = [0, 1]$.[3]

We have

Lemma 2. *Let Assumption 3 hold. Then for each $\kappa \geq 1$, there exists g_κ such that $\kappa^b\|g_\kappa - g\| \to 0$ as $\kappa \to \infty$ with any $b < q$.*

Due to the boundedness of the estimand Π introduced in (5), it follows directly from Lemma 2 that $\kappa^b\|\Pi(g_\kappa) - \Pi(g)\| \to 0$ as $\kappa \to \infty$. Therefore, we may use $\Pi(g_\kappa)$ to approximate $\Pi(g)$ for large κ.

For the subsequent development of our theory, we write g_κ more explicitly as

$$g_\kappa = \sum_{i=1}^{\kappa} \alpha_{\kappa i}\varphi_i \tag{6}$$

with $\alpha_{\kappa i} \in \mathbb{R}$ for $i = 1, \ldots, \kappa$. Moreover, we define

$$\alpha_\kappa = (\alpha_{\kappa 1}, \ldots, \alpha_{\kappa\kappa})'$$

and

$$\pi_\kappa = (\varphi_1, \ldots, \varphi_\kappa)',$$

so that we have $g_\kappa = \pi_\kappa'\alpha_\kappa$. Then we have

$$\Pi(g_\kappa) = (g_\kappa(s_1), \ldots, g_\kappa(s_d))' = P_\kappa\alpha_\kappa, \tag{7}$$

where $P_\kappa = (\pi_\kappa(s_1), \ldots, \pi_\kappa(s_d))'$. It is well known that

$$c_1 \leq \lambda_{\min}\left(\frac{P_\kappa P_\kappa'}{\kappa}\right) \leq \lambda_{\max}\left(\frac{P_\kappa P_\kappa'}{\kappa}\right) \leq c_2$$

for all $\kappa \geq d$, where $c_1, c_2 > 0$ are some constants.

Define

$$(K\varphi_i)(r) = \int_T K(r, s)\varphi_i(s)\mathrm{d}s,$$

and

$$K\pi_\kappa = (K\varphi_1, \ldots, K\varphi_\kappa)',$$

where K is the function introduced in Assumption 2. Moreover, we let

$$\varpi_{\kappa t} = (K\pi_\kappa)\left(\frac{t}{n}\right).$$

Then it follows from (6) that

$$\int_{-\infty}^{\infty} f_t(s)g_\kappa(s)\mathrm{d}s = \varpi_{\kappa t}'\alpha_\kappa,$$

[2] For the functional Π defined in (5), we have $\|\Pi(g)\| = (\sum_{i=1}^d g(s_i)^2)^{1/2} \leq \sqrt{d}\|g\|$. Therefore, we may take $c = \sqrt{d}$.

[3] If T is given by $T = [a, b]$, then we may transform (f_t) by $f_t^*(s) = f_t(a+(b-a)s)$, so that (f_t^*) has the common support $[0, 1]$. The original response function g may easily be obtained from the response function g^* with respect to (f_t^*) by $g(s) = g^*((s-a)/(b-a))$.

and we may write our model given in (1) and (2) as

$$y_t = \varpi_{\kappa t}'\alpha_\kappa + x_t'\beta + u_{\kappa t}, \tag{8}$$

where

$$u_{\kappa t} = u_t + \int_T f_t(s)(g - g_\kappa)(s)\mathrm{d}s.$$

This is the regression that we will use to estimate and test our model.

We may easily see that

$$\frac{1}{n}\sum_{t=1}^n \varpi_{\kappa t}\varpi_{\kappa t}' \approx \int_0^1 (K\pi_\kappa)(r)(K\pi_\kappa)(r)'\mathrm{d}r$$

as $n \to \infty$. Therefore, to avoid the asymptotic multicollinearity, it is necessary to assume the following.

Assumption 5. We have

$$\lambda_{\min}\left(\int_0^1 (K\pi_\kappa)(r)(K\pi_\kappa)(r)'\mathrm{d}r\right) \ge c$$

for all large κ, where $c > 0$ is some constant.

Assumption 5 requires that the $(K\varphi_i)$'s are linearly independent in the Hilbert space of functions defined on T. Note that the operator $K : \varphi \mapsto K\varphi$ may formally be regarded as the so-called kernel operator. For the condition in Assumption 5 to be met, the kernel operator should not be degenerate. If, for instance, the kernel function is given by $K(r, s) = K_1(r)K_2(s)$ for some functions K_1 and K_2 respectively on $[0, 1]$ and T, then the condition is clearly not satisfied. The kernel operator K is bounded, and therefore

$$\lambda_{\max}\left(\int_0^1 (K\pi_\kappa)(r)(K\pi_\kappa)(r)'\mathrm{d}r\right)$$

is bounded uniformly for all κ. This is well known. See, for example, Dunford and Schwartz (1988).

Needless to say, we should let κ increase as $n \to \infty$. The following assumption specifies the allowable range of κ given as a function of n.

Assumption 6. We let $\kappa = cn^r$ with $1/(2q-1) < r < \min((p-2)/2p, 2/5)$, where $c > 0$ is some constant and p and q are given respectively in Assumptions 1 and 2.

Note that Assumption 4 implies that $1/(2q-1) < \min((p-2)/2p, 2/5)$. Therefore, it is possible to choose κ as given by Assumption 6. Recall that p and q are the maximum orders of existing moments of (ε_t) and derivatives of g, respectively. The smoother the function g is, the smaller is the required rate of increase in the number of polynomial and/or trigonometric terms. This is well expected, since a relatively smaller number of basis functions would be needed to approximate g if it is smooth. On the other hand, we are allowed to include a larger number of basis functions to approximate g if (ε_t) has a higher moment. If $p = q = \infty$, then we may choose $\kappa = cn^r$ with any $0 < r < 2/5$. From now on, we let κ be given as in Assumption 6. For notational brevity, we will continue to denote by κ the number of the basis functions used to estimate the response function g, although it is obviously more appropriate to use the notation κ_n.

It may be of interest to test the null hypothesis $H_0 : \|g\| = 0$. If we assume that g is in the subspace of $L^2(T)$ spanned by the basis functions $\varphi_1, \ldots, \varphi_\kappa$ for some fixed $\kappa \ge 1$, we may simply test the hypothesis $H_0 : \alpha_i = 0$ for $i = 1, \ldots, \kappa$, where (α_i) are the coefficients in the representation $g = \sum_{i=1}^\kappa \alpha_i\varphi_i$. This can easily be done using methodology presented in the subsequent sections. Testing the hypothesis for the general case can also be developed as in Park and Qian (2008). For this, we may define g_κ as in (6) and standardize the Wald statistic for the hypothesis $H_0 : \alpha_{\kappa i} = 0$ for $i = 1, \ldots, \kappa$, using its asymptotic mean and variance that are given respectively by κ and 2κ. Under appropriate regularity conditions, we may naturally expect that the standardized Wald statistic converges in distribution to the standard normal as $\kappa \to \infty$, due to the central limit theorem. However, the test will not be formally developed in the paper, since the subject matter is beyond the scope of this paper.

3. Basic statistical theory

In this section, we develop the basic statistical theory for our model. As we mentioned above, the response function g and the parameter β in our model (1) and (2) can be estimated from the regression (8). Let $z_{\kappa t} = (\varpi_{\kappa t}', x_t')'$ and $\delta_\kappa = (\alpha_\kappa', \beta)'$, and write the regression (8) as

$$y_t = z_{\kappa t}'\delta_\kappa + u_{\kappa t}. \tag{9}$$

Moreover, we let

$$\hat\delta_{n\kappa} = (\hat\alpha_{n\kappa}', \hat\beta_n')'$$

be the OLS estimator of δ_κ, i.e.,

$$\hat\delta_{n\kappa} = (Z_{n\kappa}'Z_{n\kappa})^{-1}Z_{n\kappa}'y = \delta_\kappa + (Z_{n\kappa}'Z_{n\kappa})^{-1}Z_{n\kappa}'u_{n\kappa},$$

where $y = (y_1, \ldots, y_n)'$, $Z_{n\kappa} = (z_{\kappa 1}, \ldots, z_{\kappa n})'$, and $u_{n\kappa} = (u_{\kappa 1}, \ldots, u_{\kappa n})'$.

Define $D_n = \mathrm{diag}\,(I_\kappa, \sqrt{n}I_m)$ and

$$M_{n\kappa} = \frac{D_n^{-1}Z_{n\kappa}'Z_{n\kappa}D_n^{-1}}{n}$$

$$N_{n\kappa} = \frac{D_n^{-1}Z_{n\kappa}'u_{n\kappa}}{\sqrt{n}},$$

so that

$$\sqrt{n}D_n(\hat\delta_{n\kappa} - \delta_\kappa) = M_{n\kappa}^{-1}N_{n\kappa}.$$

Also, we introduce

$$M_\kappa = \int_0^1 \begin{pmatrix}(K\pi_\kappa)(r)\\ V(r)\end{pmatrix}\begin{pmatrix}(K\pi_\kappa)(r)\\ V(r)\end{pmatrix}'\mathrm{d}r$$

$$N_\kappa = \int_0^1 \begin{pmatrix}(K\pi_\kappa)(r)\\ V(r)\end{pmatrix}\mathrm{d}U(r) + \begin{pmatrix}0\\ \eta\end{pmatrix},$$

where $\eta = \sum_{i=0}^\infty \mathbb{E}v_{t-k}u_t$, to represent the limit distributions of $M_{n\kappa}$ and $N_{n\kappa}$, respectively. We have the following proposition.

Proposition 3. *Let Assumptions 1–5 hold. Then we have*

(a) $\|M_\kappa\|, \|M_\kappa^{-1}\| = O_p(1)$, $\|M_{n\kappa} - M_\kappa\| = O_p(n^{-a}\kappa^{1/2}) + O_p(n^{-1}\kappa^2)$,

(b) $\|N_\kappa\| = O_p(\kappa^{1/2})$, $\|N_{n\kappa} - N_\kappa\| = o(n^{1/2}\kappa^{1/2-b}) + O_p(n^{-a}\kappa^{1/2}) + O_p(n^{-1}\kappa^{3/2})$,

where a and b are introduced respectively in Lemmas 1 and 2.

Proposition 4. *Let Assumptions 1–5 hold. Then we have*

$$\sqrt{n}D_n(\hat\delta_{n\kappa} - \delta_\kappa) = M_\kappa^{-1}N_\kappa + O_p(n^{-a}\kappa) + O_p(n^{-1}\kappa^{5/2}) + o_p(n^{1/2}\kappa^{1/2-b}),$$

where a and b are introduced respectively in Lemmas 1 and 2.

Under Assumption 6, the result in Proposition 4 implies that

$$\sqrt{n}D_n(\hat\delta_{n\kappa} - \delta_\kappa) = M_\kappa^{-1}N_\kappa + o_p(1),$$

since, in particular,

$$n^{-a}\kappa,\ n^{-1}\kappa^{5/2},\ n^{1/2}\kappa^{1/2-b} = o(1) \quad (10)$$

if $\kappa = cn^r$ with $1/(2b-1) < r < \min(a, 2/5)$ and some constant $c > 0$.

We estimate the function g by

$$\hat{g}_{n\kappa} = \pi_\kappa' \hat{\alpha}_{n\kappa}. \quad (11)$$

The corresponding estimate for the estimand $\Pi(g)$ in (5) is then given by

$$\Pi(\hat{g}_{n\kappa}) = (\hat{g}_{n\kappa}(s_1), \ldots, \hat{g}_{n\kappa}(s_d))' = P_\kappa \hat{\alpha}_{n\kappa} \quad (12)$$

according to (7). In what follows, we let $Q_\kappa = \mathrm{diag}\,(P_\kappa, I_m)$ and

$$S_{n\kappa} = Q_\kappa \left(\frac{D_n^{-1} Z_{n\kappa}' Z_{n\kappa} D_n^{-1}}{n} \right)^{-1} Q_\kappa' = Q_\kappa M_{n\kappa}^{-1} Q_\kappa',$$

which will be used repeatedly in the rest of this section.

Now we may readily deduce the following.

Proposition 5. *Let Assumptions 1–6 hold. Then we have*

$$\begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g_\kappa)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = O_p(\kappa)$$

and

$$S_{n\kappa}^{-1/2} \begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g_\kappa)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = (Q_\kappa M_\kappa^{-1} Q_\kappa')^{-1/2} Q_\kappa M_\kappa^{-1} N_\kappa + o_p(1)$$

as $n \to \infty$.

Proposition 6. *Let Assumptions 1–6 hold. Then we have*

$$\begin{pmatrix} \sqrt{n}\left[\Pi(g_\kappa) - \Pi(g)\right] \\ 0 \end{pmatrix} = o(\kappa)$$

and

$$S_{n\kappa}^{-1/2} \begin{pmatrix} \sqrt{n}\left[\Pi(g_\kappa) - \Pi(g)\right] \\ 0 \end{pmatrix} = o_p(\kappa^{1/2})$$

as $n \to \infty$.

Now it is straightforward to deduce from Propositions 5 and 6 the following theorem.

Theorem 7. *Let Assumptions 1–6 hold. Then we have*

$$\begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = O_p(\kappa)$$

and

$$S_{n\kappa}^{-1/2} \begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = (Q_\kappa M_\kappa^{-1} Q_\kappa')^{-1/2} Q_\kappa M_\kappa^{-1} N_\kappa + o_p(1)$$

as $n \to \infty$.

As we have shown in Theorem 7, the estimators $\Pi(\hat{g}_{n\kappa})$ and $\hat{\beta}_n$ respectively for the nonparametric and parametric parts are both consistent. The convergence rate for the nonparametric part is given by $n^{-1/2}\kappa$. The parametric part can be estimated by a faster convergence rate $n^{-1}\kappa$. In both cases, the convergence rates are reduced by a factor of κ, i.e., the number of the basis functions used to estimate the response function, compared to those respectively in the standard stationary and cointegrating regressions. Naturally, the more basis functions that are used, the slower the convergence rates of the estimators become. Note that this is also true for the coefficient of the parametric component of the model. This is in contrast with the standard stationary regressions, where the parametric component can be estimated with the standard $n^{-1/2}$ rate.

The distributions of both $\Pi(\hat{g}_{n\kappa})$ and $\hat{\beta}_n$ are generally non-Gaussian and biased. They also depend in particular on the number κ of the included basis functions. There is, however, a special case where this is not so at least asymptotically. If the stochastic regressor (x_t) is strictly exogenous, then $\eta = 0$ and U and V become independent. In this case, the limit distributions of both estimators are reduced to being normal.

Corollary 8. *Let Assumptions 1–6 hold. Under strict exogeneity, we have*

$$S_{n\kappa}^{-1/2} \begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} \to_d \mathbb{N}(0, \omega^2 I)$$

as $n \to \infty$, where ω^2 is the long-run variance of the errors (u_t).

One important implication of Corollary 8 is that for practical purposes we may simply regard $\omega^2 S_{n\kappa}$ as the sampling variance of the estimator $\Pi(\hat{g}_{n\kappa})$ and $\hat{\beta}_n$, which are distributed as normal around their true values $\Pi(g)$ and β.

It is possible to consistently estimate the long-run variance ω^2 of the regression error (u_t) using the residuals from the truncated regression (9). To see this, we first let $(\hat{u}_{\kappa t})$ be the OLS residuals from the regression (9) and consider the kernel estimator that is given by

$$\hat{\omega}_{n\kappa}^2 = \frac{1}{n} \sum_{|i|<\ell_n} \tau\left(\frac{i}{\ell_n}\right) \sum_t \hat{u}_{\kappa t} \hat{u}_{\kappa, t-i}$$

with lag window τ and lag truncation number ℓ_n.

Assumption 7. We assume that

(a) τ is a bounded even function with $\tau(0) = 1$ such that

$$\int_{-1}^{1} \tau(x)^2 dx < \infty \quad \text{and} \quad \lim_{x\to 0} \frac{1-\tau(x)}{|x|} = \tau_0 < \infty,$$

(b) $\ell_n = cn^{1/3}$ for some constant $c > 0$.

The conditions required in Assumption 7 are quite mild. The conditions in (a) are satisfied for many commonly used lag windows including Bartlett and Parzen kernels and the rectangular kernel. The condition in (b) simply specifies the rate at which the lag truncation number should increase along with the sample size. This condition is not restrictive. For the Bartlett kernel, the rate is indeed optimal and balances off the sampling bias and variance, in the case when (u_t) is observed and ω^2 is estimated directly from (u_t). The reader is referred to Andrews (1991b) for more discussions on the choice of the lag window τ and lag truncation number ℓ_n.

Proposition 9. *Let Assumptions 1–7 hold. Then we have*

$$\hat{\omega}_{n\kappa}^2 = \omega^2 + O_p(n^{-1/3})$$

as $n \to \infty$.

We have thus shown that the long-run variance ω^2 of the regression error (u_t) in our semiparametric model (1) can be consistently estimated by the usual kernel method from the fitted residuals of the truncated regression (9). This is very important, since the estimation of ω^2 is crucial not only in conducting inference, but also in efficiently estimating the model. This will be shown clearly in Section 4.

4. Estimation and specification test

Though our model can be consistently estimated by the OLS regression for the truncated model, the resulting estimators for both the nonparametric and the parametric parts are in general asymptotically non-Gaussian. Just as in the case of the linear cointegrating regression model, non-Gaussianity implies asymptotic biasedness of the estimators and invalidity of the standard testing procedures. There is an exception where Gaussianity applies. However, the required strict exogeneity is too restrictive and is expected to hold rarely in practical applications. In this section, we develop asymptotically efficient estimators and a specification test. More specifically, we extend to our semiparametric cointegrating regression model the method of canonical cointegrating regression (CCR), which was developed by Park (1992) for the linear cointegrating regression. Moreover, to test the adequacy of our model specification, we consider the variable addition test proposed by Park (1990).

The CCR procedure requires a transformation of (y_t) and (x_t) in the regression (1) using the stationary components of the model. To make the required transformations explicit, we need to introduce some additional notation. Let $\Gamma(i) = \mathbb{E}w_t w_{t-i}'$ be the autocovariance function of (w_t) defined in (3). The variance matrix Ω of the limit Brownian motion W introduced in Lemma 1 is then given by

$$\Omega = \sum_{i=-\infty}^{\infty} \Gamma(i), \tag{13}$$

i.e., the long-run variance matrix of (w_t). We also define $\Sigma = \Gamma(0)$ and $\Lambda = \sum_{i=1}^{\infty} \Gamma(i)$, so that $\Omega = \Sigma + \Lambda + \Lambda'$. We also define

$$\Delta = \Sigma + \Lambda. \tag{14}$$

It will be useful to partition the parameter matrices Ω and Δ in (13) and (14) conformably with (w_t) in (3). Denote the partitioned submatrices by Ω_{ij} and Δ_{ij} for $i, j = 1, 2$. If the partitioned submatrices are scalars or vectors, then the lowercase letters ω_{ij} and δ_{ij} will be used. Finally, define $\Delta_2 = (\delta_{12}', \Delta_{22}')$, which will also be used to define the CCR transformation given below.

We now consider the regression

$$\begin{aligned} y_t^* &= \varpi_{\kappa t}'\alpha_\kappa + x_t^{*\prime}\beta + u_{\kappa t}^* \\ &= z_{\kappa t}^{*\prime}\delta_\kappa + u_{\kappa t}^* \end{aligned} \tag{15}$$

corresponding respectively to (8) and (9), where

$$x_t^* = x_t - \Delta_2 \Sigma^{-1} v_t$$

and

$$y_t^* = y_t - \left(\Delta_2 \Sigma^{-1} v_t\right)' \beta - \left(0, \omega_{12}\Omega_{22}^{-1}\right) v_t.$$

Since the transformations above introduce only zero-mean stationary deviations to the regression (9), the same long-run relationship is sustained in the transformed regression (15). Note that only the parametric part is required to be transformed. The nonparametric part is non-stochastic and does not need to be transformed.

The stationary errors $(u_{\kappa t})$ in (9), however, go through some substantial changes. The transformed errors $(u_{\kappa t}^*)$ in (15) are now given by

$$u_{\kappa t}^* = u_t^* + \int_T f_t(s)(g - g_\kappa)(s)\mathrm{d}s$$

with the CCR error

$$u_t^* = u_t - \omega_{12}\Omega_{22}^{-1} v_t, \tag{16}$$

which constitutes the main non-perishing part. The long-run variance of the CCR error (u_t) is

$$\omega_*^2 = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21},$$

which is also the conditional long-run variance of the errors (u_t) given the innovations of the regressors (v_t) in the original regression (9).

The CCR transformations involve unknown parameters. However, we may easily get consistent estimates of these parameters from the first-step OLS estimation of the regression. In the standard linear cointegrating regression, the replacement of the unknown parameters by their consistent estimates in the CCR procedure does not affect the subsequent limit theory, as shown in Park (1992). This is also true for our semiparametric cointegrating regression. In what follows, we assume that the first-step OLS regression is run to obtain the estimates of the parameters in the CCR transformations and the CCR procedure is defined in terms of the estimated parameters.

We define the CCR estimator to be the OLS estimator for the transformed regression model (15). If we define $Z_{n\kappa}^*$ from $z_{\kappa t}^* = (\varpi_{\kappa t}', x_t^{*\prime})'$ by $Z_{n\kappa}^* = (z_{\kappa 1}^*, \ldots, z_{\kappa n}^*)'$, and similarly let $y^* = (y_1^*, \ldots, y_n^*)'$, then

$$\hat{\delta}_{n\kappa}^* = (\hat{\alpha}_{n\kappa}^{*\prime}, \hat{\beta}_n^{*\prime})' = (Z_{n\kappa}^{*\prime} Z_{n\kappa}^*)^{-1} Z_{n\kappa}^{*\prime} y^*.$$

Notice that the CCR error (u_t^*) in (16) is asymptotically independent of the innovations of the regressors (Δx_t^*). The OLS procedure in the transformed model thus yields an efficient and optimal estimator in the sense of Phillips (1991). Now we may estimate $\Pi(g)$ by $\Pi(\hat{g}_{n\kappa}^*)$, where

$$\hat{g}_{n\kappa}^* = \pi_\kappa' \hat{\alpha}_{n\kappa}^*$$

and

$$\Pi(\hat{g}_{n\kappa}^*) = (\hat{g}_{n\kappa}^*(s_1), \ldots, \hat{g}_{n\kappa}^*(s_d))' = P_\kappa \hat{\alpha}_{n\kappa}^*,$$

corresponding to (11) and (12), using the CCR estimate $\hat{\alpha}_{n\kappa}^*$.

Theorem 10. *Let Assumptions 1–7 hold. Then we have*

$$S_{n\kappa}^{*-1/2} \begin{pmatrix} \Pi(\hat{g}_{n\kappa}^*) - \Pi(g) \\ \hat{\beta}_n^* - \beta \end{pmatrix} \to_d \mathbb{N}(0, \omega_*^2 I)$$

as $n \to \infty$, where

$$S_{n\kappa}^* = Q_\kappa \left(\frac{D_n^{-1} Z_{n\kappa}^{*\prime} Z_{n\kappa}^* D_n^{-1}}{n} \right)^{-1} Q_\kappa$$

and ω_*^2 is the long-run variance of the CCR *errors (u_t^*).*

Theorem 10 implies that our previous result in Corollary 8 for the models with strictly exogenous regressors extends well to more general models with endogeneity if the models are transformed using the CCR method. It is therefore shown that the CCR methodology also works for the semiparametric cointegrating regression models we consider in this paper.

The CCR methodology also provides valid tests for our model. The presence of cointegration in the model (1) is testable by the variable addition test (VAT) in Park (1990).[4] For our model, it amounts to testing the null hypothesis $H_0 : \lambda = 0$ in the regression

$$y_t^* = z_{\kappa t}^{*\prime}\delta_\kappa + s_t'\lambda + u_{\kappa t}^*, \tag{17}$$

where (s_t) represents the k-dimensional regressors superfluously added to the original regression (15). In the presence of cointegration, the test should reveal the superfluousness of the added regressors, and therefore the hypothesis H_0 should not be rejected. When there is no cointegration, however, the regression (15) becomes spurious and the test would reject H_0. The test statistic VAT is given by

[4] It is also possible to consider other existing tests. Please read the discussion following Corollary 11. The VAT test is an omnibus test, which is expected to be consistent against many possible misspecifications. Its finite sample power would of course depend upon particular types of misspecifications.

$$\mathrm{VAT} = \frac{\mathrm{RSS} - \mathrm{RSS}^k}{\hat{\omega}^{*2}_{n\kappa}},$$

where RSS and RSS^k are the sums of squared residuals obtained, respectively, from (15) and (17), and $\hat{\omega}^{*2}_{n\kappa}$ is the long-run variance estimate for the CCR errors based on regression (17).

It is straightforward to show the following.

Corollary 11. *Let Assumptions 1–7 hold. Then we have*

$$\mathrm{VAT} \to_d \chi^2_k$$

as $n \to \infty$, where k is the number of superfluously added regressors.

The test based on the VAT has the usual chi-square limiting distribution, and is very easy to implement. Other existing tests, including residual based tests such as the ADF and Phillips tests, and the LM test by Shin (1994), are also expected to have some discriminatory power, if applied to the regression (9). All of them, however, have asymptotic distributions that are generally dependent upon the deterministic regressors included in the regression. For our model, it is indeed not difficult to show that their critical values would depend on the kernel function K introduced in Assumption 2. Therefore, they are not applicable for our model.[5]

The results in this section extend well, only with trivial modifications, to more general models with deterministic trends and other stationary covariates. The required modifications are indeed mostly notational. Here we will discuss some obvious extensions, which will be useful in implementing our model and methodology in practical applications. First, the inclusion of deterministic trends such as polynomial time trends possibly with structural breaks is allowed, and all our theories and results developed in the paper are applicable, at least qualitatively, as they stand. More precisely, for the models with deterministic regressors that have properly defined L^2-limits as those introduced in, for example, Park (1992), the CCR methodology works and yields Gaussian asymptotics. The convergence rates of the parameter estimates in such models are also reduced by the order of κ, exactly as in the case of our prototypical model. For instance, the coefficient of the linear time trend has the convergence rate $n^{-3/2}\kappa$.

Second, our basic statistical theories and results are also directly applicable for the model with additional stationary covariates. In general, stationary covariates are asymptotically orthogonal to the integrated regressors and deterministic trends, both of which are trending and distinguish themselves clearly from the stationary time series. This is well known and the reader is simply referred to, for example, Park and Phillips (1988) for more discussions on this. In the presence of additional stationary covariates, however, the CCR methodology has little motivation. In this case, it appears to be more reasonable to use the leads and lags method proposed by Saikkonen (1991). We may well expect that the asymptotic normality of the estimators for $\Pi(g)$ and β holds in this case, if we appropriately choose κ as an increasing function of the sample size n. The allowable range of κ, however, would certainly be changed, since the number of leads and lags is also expected to increase with n.

5. Empirical application

In this section, we use our methodology to investigate how changing age distribution influences the consumption level and the savings rate in the US. The effects of the age distribution on macroeconomic variables have previously been analyzed by Fair and Dominguez, 1991, hereafter FD, among others. The main differences between our approach and theirs are as follows. First, we explicitly allow for the nonstationarity of the variables and consider long-run cointegrating relationships between them.[6] We also use the knowledge of the presence of cointegration to estimate the model more efficiently. Second, we nonparametrically estimate the response of consumption and the savings rate to changing age distribution, whereas FD imposed a parametric restriction to estimate the response function.

The general form of our empirical regressions is as follows:

$$y_t = c_t'\gamma + \int_T f_t(s)g(s)\mathrm{d}s + x_t'\beta + u_t, \tag{18}$$

where (c_t) denotes the maintained deterministic trend, (f_t) is the density function for the age distribution, and g is the age response function. Throughout our applications, we maintain a linear time trend as our maintained deterministic trend, i.e., $c_t = t$ for $t \geq 1$. Note that the intercept term is not included in (18). We do not impose any restriction on g, and therefore the intercept term would not be identified, if included.

We consider two different specifications. First, we examine how changing the age distribution influences the consumption level by taking $y_t = \ln(C_t)$ and $x_t = \ln(I_t)$, where (C_t) and (I_t) represent the per capita consumption level and the per capita income level, respectively. In general, it is expected that per capita consumption and per capita income have a long-run economic relationship. However, if different age groups exhibit different propensity to consume, the per capita consumption cannot be explained solely by the per capita income. According to our specification, the remaining portion is explained by the age distribution of the population. Second, we analyze how changing the age distribution influences the savings rate by letting $y_t = S_t$ and $x_t = (\ln(I_t), \Delta\ln(I_t))'$, where S_t is the savings rate and $\Delta\ln(I_t) = \ln(I_t) - \ln(I_{t-1})$ is the per capita income growth rate. This specification has been popular since Leff (1969) pioneered examining how a high fertility rate influences the savings rate.[7]

The data, except for the population, were collected from the FRED (Federal Reserve Economic Data) maintained by the Federal Reserve Bank of St. Louis. We use the real non-durable consumption expenditure for consumption and the real GDP for income, and divide both series by the total population to obtain their per capita values. The savings rate is derived by dividing nominal gross saving by nominal GDP. The age-distribution data are obtained from the US Census Bureau. The Population Estimates Program at the US Census Bureau develops and releases national population estimates in 76 age groups. These data are available annually from 1900 to 1979, quarterly from 1980 to 2002 and monthly from 2000. We construct a sample of quarterly data from 1959:1 to 2002:3 by imputing them, when they are not available, through linear interpolation of the annual data. Age-group 1 consists of individuals less than one year, age-group 2 consists of individuals between one and two years of age, and so on through age-group 76, which consists of individuals of age 76 and older. By eliminating the age groups of individuals younger than 16, we consider 61 age groups in this paper (i.e., age-groups 16, 17, . . . , 76) and define the "total" population to be the sum of these groups only.

As a benchmark of our results, we also consider a regression model suggested by FD:

5 It may be possible to find their critical values using bootstrap or subsampling methods. The validity of such procedures in our model, however, has not been established.

6 See Attfield and Cannon (2003) for an extension of the FD approach to a cointegrating regression model using a UK data set.

7 Instead of considering the age distribution itself, Leff (1969) included two indices of dependency rates for child and old populations in order to investigate how increased dependency rates influence the savings rate.

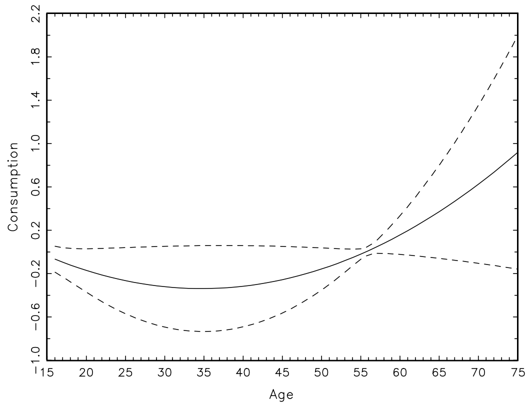

Fig. 1. Parametric consumption–age distribution coefficients.

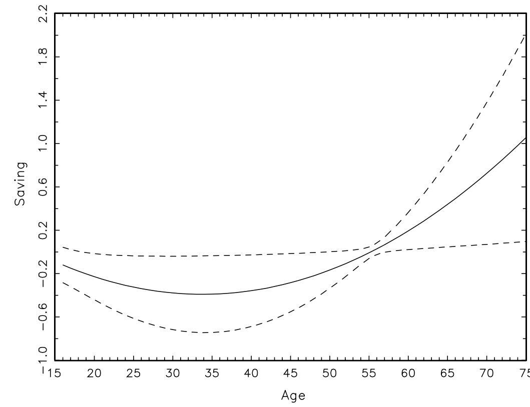

Fig. 2. Parametric saving–age distribution coefficients.

$$y_t = \mu_0 + c_t'\gamma + \sum_{j=1}^{J} \phi_j p_{jt} + x_t'\beta + u_t, \tag{19}$$

where $\{p_{jt}; j = 1, \ldots, J\}$ is the proportion of individuals in age group j in the total population at time t. To reduce the number of parameters to estimate in (19), FD suggested putting a parametric restriction on the ϕ coefficients, i.e.,

$$\phi_j = \delta_0 + \delta_1 j + \delta_2 j^2 \quad \text{and} \quad \sum_{j=1}^{J} \phi_j = 0. \tag{20}$$

Substituting (20) into (19) yields

$$y_t = \mu_0 + c_t'\gamma + z_{1t}\delta_1 + z_{2t}\delta_2 + x_t'\beta + u_t, \tag{21}$$

where $z_{1t} = \sum_{j=1}^{J} jp_{jt} - J^{-1}\sum_{j=1}^{J} j \sum_{j=1}^{J} p_{jt}$ and $z_{2t} = \sum_{j=1}^{J} j^2 p_{jt} - J^{-1}\sum_{j=1}^{J} j^2 \sum_{j=1}^{J} p_{jt}$; see Fair and Dominguez (1991, p. 1280).[8] To compare the two models (18) and (19), we partition T into J subintervals T_j (i.e., $T = \cup_{j=1}^{J} T_j$) and note $\int_{T_j} f_t(s)\mathrm{d}s = p_{tj}$. Then we may approximately represent the model (19) in the form of (18) with

$$g(s_j^*) \approx \phi_j \quad \text{for } j = 1, \ldots, J, \tag{22}$$

where s_j^* denotes a middle point in the interval T_j. Therefore, our model (18) has an advantage in that it avoids the polynomial restriction (20) on g, which can be arbitrary.

Estimation results for consumption and saving equations using the parametric specification (21) are given as in Box I, where the numbers in parentheses are robust t-statistics using the standard error estimates of Newey and West (1987). Figs. 1 and 2 show the estimates of the polynomial coefficients (ϕ_j), $j = 1, \ldots, J$, for consumption and the savings rate, respectively. The dotted lines are 95% pointwise confidence bands for these estimates. The implied quadratic age coefficient curve for consumption is U-shaped, as is predicted by the Life-Cycle Hypothesis with minimum point at around 34. For the savings rate, however, the estimation result given in Box I and Fig. 2 show that it is U-shaped, inconsistently with the hypothesis.[9] Since the savings rate is expected to be the highest for prime-aged people, the implied age coefficient curve should be inverted U-shaped. FD also reported a similar finding (see Fig. 8 of FD) and they attributed this problem to a cyclical factor in the data. However, we note that the latter finding might also result from the incorrect polynomial restriction (20) on g. As a preliminary step to examine this possibility, we estimate the model (19) with $J = 7$ different age groups (i.e., ages 16–25, ..., 66–75, 76+). This allows us to estimate the model with a reasonable number of degrees of freedom, but without having to impose the restriction (20).[10] The estimates of ϕ_j for $j = 1, \ldots, 7$ are plotted in Figs. 3 and 4 for the consumption and

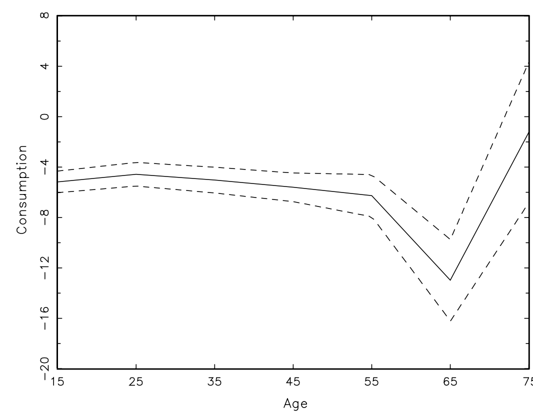

Fig. 3. Consumption–age distribution coefficients for seven age groups.

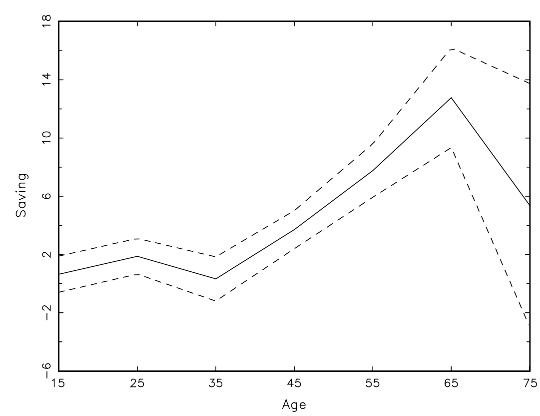

Fig. 4. Saving–age distribution coefficients for seven age groups.

8 Although we call it parametric, the FD model would have excessively many parameters and become overly nonparametric without restrictions in (20). On the other hand, the nonparametric approach introduced below effectively reduces the number of parameters to estimate by imposing "smoothness" conditions such as Assumption 3.

9 The regressors z_1 and z_2 are highly correlated with correlation coefficient 0.983, and therefore their coefficients are not individually significant. We find that the tendency for z_1 and z_2 to be highly correlated is pervasive for other countries as well (not reported), which suggests that it may be hard to estimate the age coefficient very accurately by restricting the coefficients in this way.

10 We set $\mu_0 = 0$ to avoid the perfect multicollinearity.

$$\widehat{\ln(C_t)} = \underset{(-11.776)}{-5.486} \underset{(1.187)}{+0.0003t} \underset{(-1.572)}{-0.030z_{1t}} \underset{(1.629)}{+0.0008z_{2t}} \underset{(14.559)}{+0.625\ln(I_t)}$$

$$\widehat{S_t} = \underset{(2.531)}{1.387} \underset{(-3.763)}{-0.001t} \underset{(-1.907)}{-0.032z_{1t}} \underset{(2.056)}{+0.0009z_{2t}} \underset{(2.107)}{+0.099\ln(I_t)} \underset{(0.126)}{+0.014\Delta\ln(I_t)}$$

Box I.

Table 1
Unit root tests.

	Lag	Augmented Dickey–Fuller		Phillips–Perron	
		Constant	Trend	Constant	Trend
Consumption	0	1.58	−0.109	1.58	−1.09
	1	1.24	−1.60	1.43	−1.32
	2	1.17	−1.62	1.36	−1.42
	3	0.86	−2.29	1.26	−1.58
	4	0.82	−2.06	1.20	−1.69
Income	0	1.13	−1.54	1.13	−1.54
	1	0.82	−2.36	0.93	−1.82
	2	0.59	−2.52	0.83	−1.99
	3	0.54	−2.50	0.77	−2.10
	4	0.61	−2.79	0.72	−2.19
Savings rate	0	−1.47	−3.05	−1.47	−3.06
	1	−1.82	−3.41	−1.57	−3.24
	2	−1.45	−3.34	−1.57	−3.27
	3	−2.09	−4.26**	−1.66	−3.41
	4	−2.09	−3.90*	−1.69	−3.49*
Log income	0	−1.51	−2.31	−1.51	−2.31
	1	−0.95	−2.41	−1.38	−2.44
	2	−1.02	−3.01	−1.30	−2.56
	3	−1.13	−3.15	−1.26	−2.63
	4	−0.85	−2.89	−1.24	−2.68

Notes: The sample period is 1959:1–1992:4. Based on the Dickey–Fuller and the Phillips–Perron statistics.
* denotes the rejection of a unit root at the 5% significance level.
** denotes the rejection of a unit root at the 1% significance level.

the saving rate, respectively. The result of Fig. 3 is more or less consistent with Fig. 1, but Fig. 4 suggests that the age response function might be inverted U-shaped if the restriction (20) is not imposed.[11]

We now consider the semiparametric cointegrating regression model. We first check whether the individual series are nonstationary. We apply to our data the unit root tests developed by Dickey and Fuller (1979) and Phillips and Perron (1988). The results are summarized in Table 1. As we allow the number of lags to change from 0 to 4, we find that most data except the savings rate show strong evidence of containing a unit root. Even for the savings rate, aside from the lag specifications of 3 or 4, most specifications do not reject the unit root null hypothesis.

In order to see whether there exist cointegration relationships between the logarithms of consumption and income and between the savings rate and the logarithm of income, we employ Johansen's procedure. The results are reported in Table 2. We consider two specifications, with and without a linear deterministic trend in the data. The test statistics for a cointegration relationship between consumption and income demonstrate an ambiguous result. When we do not allow for a deterministic trend, there is evidence for a cointegration relationship at the 5% significance level. However, if a linear deterministic trend is introduced, the hypothesis of no cointegration is not rejected and the test does not support the existence of a cointegrating relationship. Since there is a strong theoretical reason to believe in the existence of a long-run relationship between consumption and income, the estimation procedure

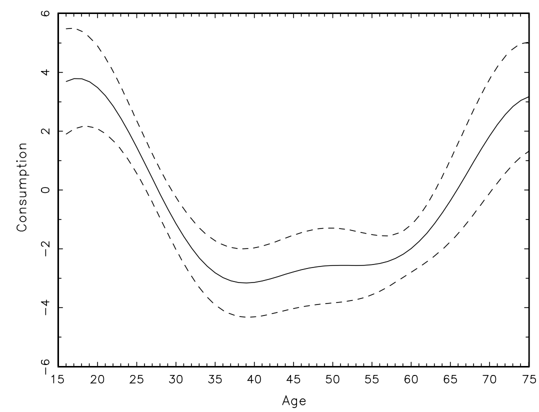

Fig. 5. Nonparametric consumption–age response function.

below assumes the presence of cointegration. In contrast, we obtain much stronger evidence for a cointegration relationship between the savings rate and the logarithm of income. Both specifications reject no cointegration either at the 5% or at the 1% levels.

We next report our estimation results from the semiparametric regression model (18) (see Box II).

Since the income variables are not usually strictly exogenous, we also estimate the models using the CCR method discussed in Section 4. The results are given in Box III.

Here we used the Parzen window along with a data-dependent lag truncation selection rule as in Andrews (1991b) in implementing the CCR procedure. The age response function $g(s)$ was estimated by the FFF which included a constant, s, s^2, and two trigonometric pairs (i.e., $(\cos(s), \sin(s))$ and $(\cos(2s), \sin(2s))$).[12] In order to confirm the cointegration relationship in this extended form of the equations, we also applied the variable addition test (VAT) using the quadratic and cubic trends (t^2 and t^3) as the superfluously added variables. The values of the test statistics are 0.673 (p-value: 0.714) and 1.635 (p-value: 0.441) for the consumption and saving equations, respectively. The test results imply that we can not reject the null of cointegration for both models. It seems, therefore, that the semiparametric cointegrating regression model is strongly consistent with the data.

Figs. 5 and 6 show the estimated age response functions $\widehat{g}_C$ and $\widehat{g}_S$ respectively for the consumption and saving equations. They reveal that the response functions for consumption and saving are U-shaped and inverted U-shaped, respectively, both of which are consistent with the implications of the Life-Cycle Hypothesis. In particular, unlike the results from the parametric model, our model exhibits that the savings rate is highest in the mid-aged range between 35 and 55, without relying on any other augmented arguments. Furthermore, the estimates were significant, implying that changing age distribution plays a significant role in explaining consumption and saving behavior. The effects of income were also

11 This result needs to be interpreted carefully because the estimates of ϕ_j might be sensitive to the choice of the number of age groups. Our semiparametric procedure suggested in this paper, however, does not rely on an ad hoc choice of age groups.

12 In the context of independent data, Andrews (1991c) suggests data-dependent procedures to select optimally the order k of the FFF expansion. To the best of our knowledge, however, there is yet no optimal data-dependent selection rule for k available in the nonstationary time series literature.

Table 2
Cointegration tests.

Johansen test					
Variables	Specification	λ_1	λ_2	Cointegrating vector	$LR = -T\ln(1-\lambda_1)$
C_t and I_t	No trend	0.0919	0.0280	$(1, -0.1603)$	21.840*
C_t and I_t	Trend	0.0818	0.0061	$(1, -0.1585)$	7.2336
S_t and $\ln(I_t)$	No trend	0.1204	0.0277	$(1, 0.00958)$	22.454**
S_t and $\ln(I_t)$	Trend	0.0827	0.0063	$(1, 0.0646)$	15.106*

Notes: This table shows the Johansen cointegration statistics to test the null hypothesis of no cointegration between consumption (C_t) and income (I_t) and between the savings rate (S_t) and the logarithm of income ($\ln(I_t)$). We consider two specifications allowing for no deterministic trend (No Trend) and linear deterministic trend (Trend) in the data.
* denotes the rejection of no cointegration hypothesis at the 5% significance level.
** denotes the rejection of no cointegration hypothesis at the 1% significance level.

$$\widehat{\ln(C_t)} = \int_T \widehat{g}_C(s) f_t(s)\mathrm{d}s \quad + 0.0009t \quad + 0.668\ln(I_t)$$
$$\qquad\qquad\qquad\qquad (1.913) \qquad (18.326)$$
$$\widehat{S}_t = \int_T \widehat{g}_S(s) f_t(s)\mathrm{d}s \quad - 0.005t \quad + 0.300\ln(I_t) \quad - 0.168\Delta\ln(I_t)$$
$$\qquad\qquad\qquad\qquad (-7.700) \qquad (7.696) \qquad (-2.054)$$

Box II.

$$\widehat{\ln(C_t^*)} = \int_T \widehat{g}_C^*(s) f_t(s)\mathrm{d}s \quad + 0.0008t \quad + 0.685\ln(I_t^*)$$
$$\qquad\qquad\qquad\qquad (1.405) \qquad (18.828)$$
$$\widehat{S_t^*} = \int_T \widehat{g}_S^*(s) f_t(s)\mathrm{d}s \quad - 0.005t \quad + 0.302\ln(I_t^*) \quad - 0.277\Delta\ln(I_t^*)$$
$$\qquad\qquad\qquad\qquad (-6.970) \qquad (6.983) \qquad (-1.634)$$

Box III.

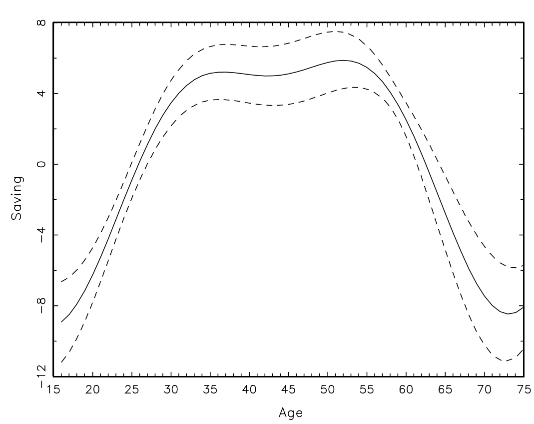

Fig. 6. Nonparametric saving–age response function.

highly significant, though the income growth rate was not very significant in explaining the savings rate.

Finally, to illustrate the robustness our empirical results, we plot the estimated saving–age response functions with different choices of the basis function (φ_i), $i = 1, \ldots, \kappa$, and truncation parameter κ.[13] Fig. 7 presents that the FFF series estimates with different numbers $k = 1, \ldots, 4$ of trigonometric pairs.[14] For $k = 1$, the response function is convex as in FD, but this might be due to the large bias resulting from an insufficient number of expansion terms. For $k \geq 2$, the function estimates are concave as we expected, though the estimates for $k \geq 3$ are less accurate due to large variance. Fig. 8 confirms that the results are also robust when the polynomial basis function (i.e., $\varphi_i(s) = s^{i-1}$ for $i = 1, \ldots, \kappa$) is used.

6. Conclusion

This paper studies a partially nonlinear and nonparametric cointegrating regression model. As usual, the linear part of the model specifies the long-run economic relationship as a cointegrating regression. The remaining error term, however, is further modeled nonlinearly and nonparametrically as being affected by other nonstationary covariate. More specifically, the nonlinear component of the model is assumed to be determined by the distribution, rather than the level, of a nonstationary covariate whose distribution changes over time. An efficient econometric estimator for this model is proposed, its asymptotic distribution obtained, and a specification test for the model is investigated. The framework is general enough to accommodate many practical time series models.

It is quite common that economic relationships over time are affected by the cross-sectional distributions in each period. The model and methodology developed in the paper can be used to effectively analyze such relationships in a new perspective. When we apply our model and methodology to analyze the long-run relationships between the consumption level and income, and the savings rate and income, we find that the impact of age distribution takes a U-shape and an inverted U-shape respectively, which

13 We found that consumption–age response function estimates are also robust in a similar sense, but we do not report the results, for brevity.

14 Therefore, since $1, s, s^2$ are always included in the FFF series estimation, the number of truncation parameters is $\kappa = 3 + 2k$.

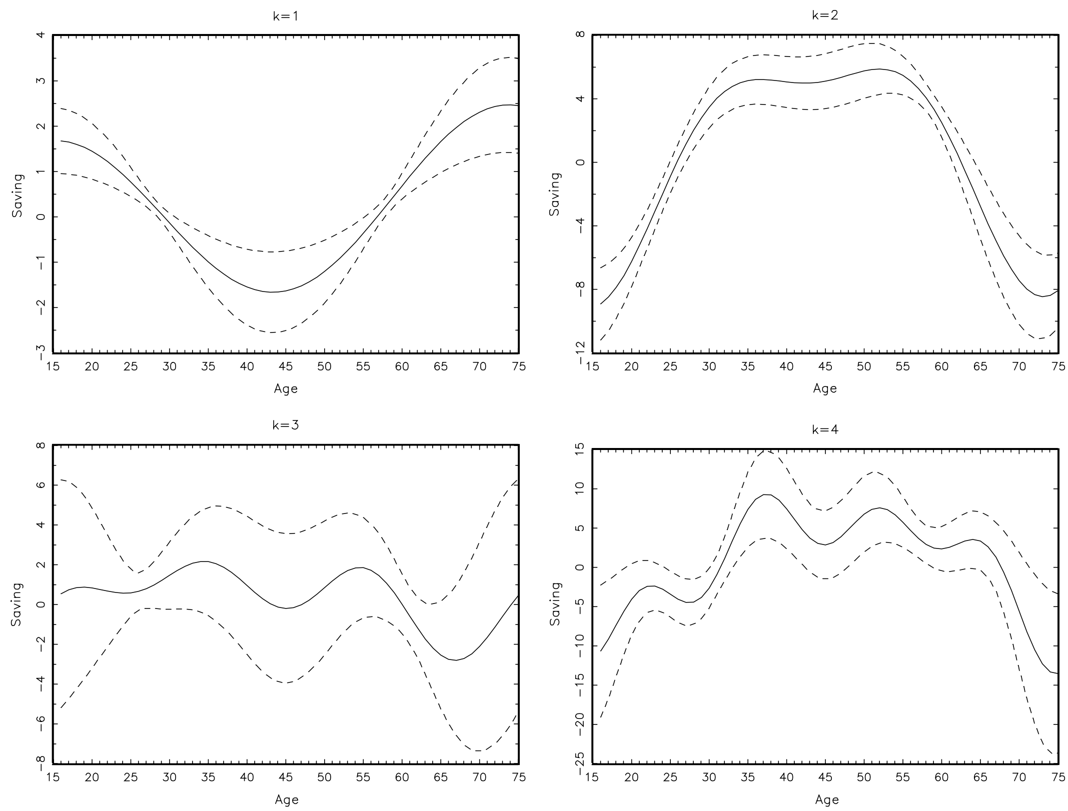

Fig. 7. FFF series estimates of the saving–age response function with different truncation parameters.

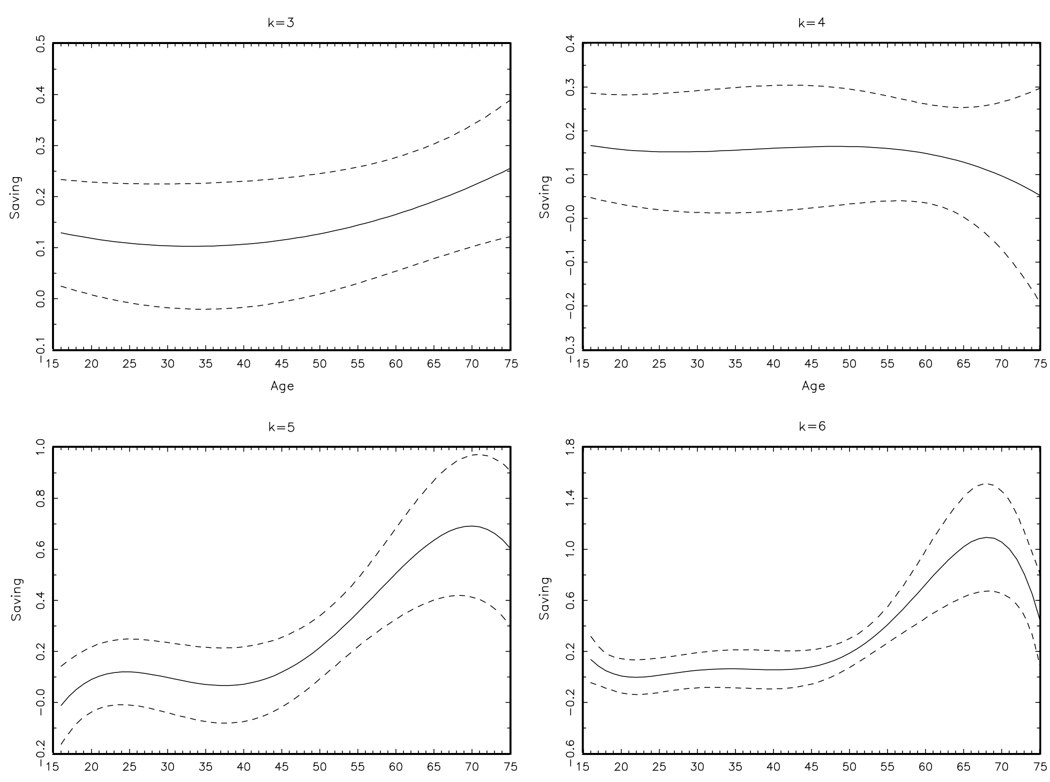

Fig. 8. Polynomial series estimates of the saving–age response function with different truncation parameters.

is consistent with the life-cycle hypothesis. This shows superiority of our approach over the previous studies based on parametric approaches, which repeatedly produced the U-shaped impact curve for the savings rate.

Appendix. Mathematical proofs

Proof of Lemma 1. See Andrews (1991a) and the proof of Lemma 1 in Park and Hahn (1999). □

Proof of Lemma 2. The stated result follows as a special case of Edmunds and Moscatelli (1977). □

Proof of Proposition 3. If we let

$$\varphi_{ni}(r) = \varphi_i\left(\frac{[nr]}{n}\right),$$

then we have

$$\sup_{1\le i\le\kappa} \|\varphi_{ni} - \varphi_i\| = O(n^{-1}\kappa). \tag{23}$$

Furthermore, if we let

$$(K_n\varphi_i)(r) = (K\varphi_i)\left(\frac{[nr]}{n}\right),$$

then it follows from (23) that

$$\sup_{1\le i\le\kappa} \|K_n\varphi_i - K\varphi_i\| = O(n^{-1}\kappa) \tag{24}$$

since $K(\cdot, s)$ is Lipschitz uniformly in $s \in T$.

We define

$$A_{n\kappa} = \frac{1}{n}\sum_{t=1}^{n} \varpi_{\kappa t}\varpi_{\kappa t}'$$

$$B_{n\kappa} = \frac{1}{n^{3/2}}\sum_{t=1}^{n} \varpi_{\kappa t}x_t'$$

$$C_{n\kappa} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \varpi_{\kappa t}u_t$$

$$D_{n\kappa} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \varpi_{\kappa t}(u_{\kappa t} - u_t)$$

$$E_{n\kappa} = \frac{1}{n}\sum_{t=1}^{n} x_t(u_{\kappa t} - u_t),$$

each of which we will subsequently consider in what follows.

To analyze $A_{n\kappa}$, we first introduce

$$A_\kappa = \int_0^1 (K\pi_\kappa)(r)(K\pi_\kappa)(r)'\mathrm{d}r.$$

Then we have

$$\|A_{n\kappa} - A_\kappa\| = O(n^{-1}\kappa^2). \tag{25}$$

Note that every element of $A_{n\kappa} - A_\kappa$ is of order $O(n^{-1}\kappa)$ uniformly, due to (24), and that for matrices the operator norm is bounded by the Euclidian norm. Moreover, we have

$$\|A_\kappa\|,\ \|A_\kappa^{-1}\| = O(1) \tag{26}$$

uniformly in κ, as follows from Assumption 5 and the subsequent discussion.

For $B_{n\kappa}$, we define

$$B_\kappa = \int_0^1 (K\pi_\kappa)(r)V(r)'\mathrm{d}r.$$

Due to Lemma 1 and (24), every element of $B_{n\kappa} - B_\kappa$ is of order $O_p(n^{-a}) + O_p(n^{-1}\kappa)$ uniformly. Therefore, we have

$$\|B_{n\kappa} - B_\kappa\| = O_p(n^{-a}\kappa^{1/2}) + O_p(n^{-1}\kappa^{3/2}). \tag{27}$$

We may also easily deduce that

$$\|B_\kappa\| = O_p(\kappa^{1/2}) \tag{28}$$

uniformly in κ.

To obtain the corresponding results for $C_{n\kappa}$, we let

$$C_\kappa = \int_0^1 (K\pi_\kappa)(r)\mathrm{d}U(r).$$

Similarly as before, we may readily show that every element of $C_{n\kappa} - C_\kappa$ is of order $O_p(n^{-a}) + O_p(n^{-1}\kappa)$, due to Lemma 1 and (24). Therefore, it follows that

$$\|C_{n\kappa} - C_\kappa\| = O_p(n^{-a}\kappa^{1/2}) + O_p(n^{-1}\kappa^{3/2}). \tag{29}$$

We also have

$$\|C_\kappa\| = O_p(\kappa^{1/2}) \tag{30}$$

uniformly in κ.

For $D_{n\kappa}$ and $E_{n\kappa}$, we note that

$$\sup_{t\ge1} |u_{\kappa t} - u_t| \le c\|g_\kappa - g\| \left(\sup_{r,s} |K(r,s)|\right) = o(\kappa^{-b}),$$

where $c > 0$ is some constant. Therefore, we may easily see that every element of $D_{n\kappa}$ is of order $o(n^{1/2}\kappa^{-b})$, and consequently,

$$D_{n\kappa} = o(n^{1/2}\kappa^{1/2-b}). \tag{31}$$

Moreover, we have

$$E_{n\kappa} = o_p(n^{1/2}\kappa^{-b}), \tag{32}$$

since

$$\sup_{1\le t\le n} n^{-1/2}|x_t| \le 1 + \sup_{0\le r\le 1} |V(r)| = O_p(1)$$

for all large n.

Finally, we let

$$F_n = \int_0^1 V_n(r)V_n(r)'\mathrm{d}r \quad \text{and} \quad F = \int_0^1 V(r)V(r)'\mathrm{d}r.$$

It follows immediately that

$$\|F_n - F\| = O_p(n^{-a}), \tag{33}$$

due to Lemma 1. Moreover, if we let

$$G_n = \int_0^1 V_n(r)\mathrm{d}U_n(r) \quad \text{and} \quad G = \int_0^1 V(r)\mathrm{d}U(r) + \eta,$$

then we also have

$$\|G_n - G\| = O_p(n^{-a}). \tag{34}$$

This is shown in the proof of Lemma 4 in Park and Hahn (1999).

Now we may easily deduce the results in (a) and (b). We first prove the results in (a). Clearly, $\|M_\kappa\| = O_p(1)$ follows from (26), Lemma A1 of Park and Hahn (1999) and that $K\pi_\kappa$ is linearly independent of V a.s. for all κ. Similarly, $\|M_\kappa^{-1}\| = O_p(1)$ may easily be obtained from (26), Lemma A1 of Park and Hahn (1999) and that $K\pi_\kappa$ is linearly independent of V a.s. for all κ. Furthermore, we have

$$\|M_{n\kappa} - M_\kappa\| \le \|A_{n\kappa} - A_\kappa\| + 2\|B_{n\kappa} - B_\kappa\| + \|F_n - F\|,$$

from which and (25), (27) and (33) the stated result for $\|M_{n\kappa} - M_\kappa\|$ follows immediately.

The results in (b) can also be obtained quite easily. We have

$$\|N_\kappa\| \le \|C_\kappa\| + \|G\| = O_p(\kappa^{1/2}),$$

due to (30). Moreover, it follows that

$$\|N_{n\kappa} - N_\kappa\| \le \|C_{n\kappa} - C_\kappa\| + \|D_{n\kappa}\| + \|E_{n\kappa}\| + \|G_n - G\|,$$

and the stated result for $\|N_{n\kappa} - N_\kappa\|$ can be obtained readily from (29), (31), (32) and (34). □

Proof of Proposition 4. Note that

$$\|M_{n\kappa}^{-1}N_{n\kappa} - M_\kappa^{-1}N_\kappa\| \le \|M_{n\kappa}^{-1}(N_{n\kappa} - N_\kappa)\| + \|(M_{n\kappa}^{-1} - M_\kappa^{-1})N_\kappa\| \le \|M_{n\kappa}^{-1}\|\,\|N_{n\kappa} - N_\kappa\| + \|M_{n\kappa}^{-1} - M_\kappa^{-1}\|\,\|N_\kappa\| \quad (35)$$

and that

$$\|M_{n\kappa}^{-1} - M_\kappa^{-1}\| \le \|M_{n\kappa}^{-1}\|\,\|M_{n\kappa} - M_\kappa\|\,\|M_\kappa^{-1}\|. \quad (36)$$

The stated result now follows readily from (35) and (36), due to Proposition 3 and the fact that $\|M_{n\kappa} - M_\kappa\| = o_p(1)$ together with $\|M_\kappa^{-1}\| = O_p(1)$ implies $\|M_{n\kappa}^{-1}\| = O_p(1)$. □

Proof of Proposition 5. The first part follows straightforwardly from Proposition 4. To prove the first part, we first note that

$$\begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g_\kappa)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = \begin{pmatrix} \sqrt{n}P_\kappa(\hat{\alpha}_{n\kappa} - \alpha_\kappa) \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = Q_\kappa\left[\sqrt{n}D_n(\hat{\delta}_{n\kappa} - \delta_\kappa)\right]. \quad (37)$$

However, we have

$$Q_\kappa\left[\sqrt{n}D_n(\hat{\delta}_{n\kappa} - \delta_\kappa)\right] = Q_\kappa M_\kappa^{-1}N_\kappa + o_p(\|Q_\kappa\|), \quad (38)$$

due to Proposition 4 and (10). The stated result therefore follows immediately from (37) and (38), Proposition 3 and that $Q_\kappa = o(\kappa^{1/2})$.

The proof for the second part is more involved. First, we note that

$$\lambda_{\min}(Q_\kappa M_{n\kappa}^{-1}Q_\kappa') \ge \frac{\lambda_{\min}(Q_\kappa Q_\kappa')}{\lambda_{\max}(M_{n\kappa})} \ge \frac{c_1\kappa}{\lambda_{\max}(M_{n\kappa})}$$

$$\lambda_{\min}(Q_\kappa M_\kappa^{-1}Q_\kappa') \ge \frac{\lambda_{\min}(Q_\kappa Q_\kappa')}{\lambda_{\max}(M_\kappa)} \ge \frac{c_2\kappa}{\lambda_{\max}(M_\kappa)}$$

for some constants $c_1, c_2 > 0$, and therefore we have

$$(Q_\kappa M_{n\kappa}^{-1}Q_\kappa')^{-1/2},\ (Q_\kappa M_\kappa^{-1}Q_\kappa')^{-1/2} = O_p(\kappa^{-1/2}). \quad (39)$$

Second, we also have

$$\|Q_\kappa(M_{n\kappa}^{-1} - M_\kappa^{-1})Q_\kappa'\| \le \|Q_\kappa\|^2\|M_{n\kappa}^{-1} - M_\kappa^{-1}\| = O_p(n^{-a}\kappa^{3/2}) + O_p(n^{-1}\kappa^{5/2}), \quad (40)$$

due to Proposition 3, (36) and that $\|Q_\kappa\| = O(\kappa^{1/2})$.

Consequently, it follows from (39) and (40) that

$$\begin{aligned}(Q_\kappa M_{n\kappa}^{-1}Q_\kappa')^{-1/2} &- (Q_\kappa M_\kappa^{-1}Q_\kappa')^{-1/2} \\ &= (Q_\kappa M_{n\kappa}^{-1}Q_\kappa')^{-1/2}\left[Q_\kappa(M_\kappa^{-1} - M_{n\kappa}^{-1})Q_\kappa'\right] \\ &\quad \times \left[(Q_\kappa M_\kappa^{-1}Q_\kappa')^{1/2} + (Q_\kappa M_{n\kappa}^{-1}Q_\kappa')^{1/2}\right]^{-1}(Q_\kappa M_\kappa^{-1}Q_\kappa')^{-1/2} \\ &= O_p(n^{-a}) = o_p(\kappa^{-1/2}),\end{aligned} \quad (41)$$

since in particular $\kappa = o(n^a)$. Now we have

$$S_{n\kappa}^{-1/2}\begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_{n\kappa}) - \Pi(g_\kappa)\right] \\ n(\hat{\beta}_n - \beta) \end{pmatrix} = (Q_\kappa M_\kappa^{-1}Q_\kappa')^{-1/2}Q_\kappa M_\kappa^{-1}N_\kappa + o_p(1),$$

due to (37)–(39) and (41). This was to be shown. The proof is therefore complete. □

Proof of Proposition 6. Note that

$$\begin{pmatrix} \sqrt{n}\left[\Pi(\hat{g}_\kappa) - \Pi(g)\right] \\ 0 \end{pmatrix} = o(n^{1/2}\kappa^{-b}).$$

The proof of the first part is trivial, since $n^{1/2}\kappa^{1/2-b} = o(1)$, as shown in (10). For the proof of the second part, note that

$$S_{n\kappa}^{-1/2} = O_p(\kappa^{-1/2}),$$

from which the stated result follows immediately. □

Proof of Theorem 7. The stated result follows immediately from Propositions 5 and 6, upon noticing

$$\|\Pi(\hat{g}_{n\kappa}) - \Pi(g)\| \le \|\Pi(\hat{g}_{n\kappa}) - \Pi(g_\kappa)\| + \|\Pi(g_\kappa) - \Pi(g)\|$$

and

$$(Q_\kappa M_\kappa^{-1}Q_\kappa')^{-1/2}Q_\kappa M_\kappa^{-1}N_\kappa = O_p(\kappa^{1/2}).$$

The proof is therefore complete. □

Proof of Corollary 8. Under strict exogeneity, we have

$$Q_\kappa M_\kappa^{-1}N_\kappa \sim \mathbb{MN}\left(0, \omega^2 Q_\kappa M_\kappa^{-1}Q_\kappa'\right),$$

where $\mathbb{MN}$ signifies the mixed normal distribution. Consequently, it follows that

$$(Q_\kappa M_\kappa^{-1}Q_\kappa')^{-1/2}Q_\kappa M_\kappa^{-1}N_\kappa \sim \mathbb{N}(0, \omega^2 I).$$

We may now easily deduce the stated result, since this is true for all κ. □

Proof of Proposition 9. Given our previous results, the proof is essentially identical to that of Proposition 9 in Park and Hahn (1999). The details are therefore omitted to save space. □

Proof of Theorem 10. It is obvious that the stated result holds if the true parameters are known. Therefore, it suffices to show that the replacement of the true parameters by their estimated values in the CCR transformations does not affect the asymptotics of the CCR methodology for our model. For this, we denote by $(\hat{x}_t^*)$ and $(\hat{y}_t^*)$ respectively the CCR transformations of (x_t) and (y_t) defined with the estimated parameters. They correspond to (x_t^*) and (y_t^*) defined with the true parameters.

Define $\hat{\Sigma}$, $\hat{\Delta}_2$, $\hat{\Omega}_{22}$ and $\hat{\omega}_{12}$ to be the estimates of the parameters used in the CCR transformations. Though they depend upon the sample size n and the number of basis functions κ, we suppress the subscript n and κ in them for notational brevity. It follows from Proposition 9 that

$$\hat{\Delta}_2 = \Delta_2 + O_p(n^{-1/3}), \qquad \hat{\Omega}_{22} = \Omega_{22} + O_p(n^{-1/3}),$$
$$\hat{\omega}_{12} = \omega_{12} + O_p(n^{-1/3}).$$

Therefore, we may readily show that the CCR methodology based on $(\hat{x}_t^*)$ and $(\hat{y}_t^*)$ is asymptotically equivalent to that based on (x_t^*) and (y_t^*). The details of the proof will, however, not be given, since it is essentially the same as the proof of Theorem 10 in Park and Hahn (1999). □

Proof of Corollary 11. The stated result follows immediately, since our theory applies only with some obvious and trivial modifications, mostly notational, to the models with additional deterministic and/or stochastic trends. See the discussions below Corollary 11. □

References

Attfield, L.L.F., Cannon, E., 2003. The impact of age distribution variables on the long run consumption function. Discussion Paper No. 03/546, Department of Economics, University of Bristol.

Andrews, D.W.K., 1991a. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. Econometrica 59, 307–345.

Andrews, D.W.K., 1991b. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D.W.K., 1991c. Asymptotic optimality of generalized C_L, cross-validation and generalized cross-validation in regression with heteroskedastic errors. Journal of Econometrics 47, 359–377.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Dunford, N., Schwartz, J., 1988. Linear Operators, Part II: Spectral Theory. Wiley, New York.

Edmunds, D.E., Moscatelli, V.B., 1977. Fourier approximation and embeddings in Sobolev space. Dissertationes Mathematicae 145, 1–46.

Einmahl, U., 1989. Extensions of results of Komlós, major and Tusnády to the multivariate case. Journal of Multivariate Analysis 28, 20–68.

Fair, R.C., Dominguez, K.M., 1991. Effects of the changing US age distribution on macroeconomic equations. American Economic Review 81, 1276–1294.

Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, New York.

Leff, N.H., 1969. Dependency rates and savings rates. American Economic Review 59, 886–896.

Newey, W.K., West, K.D., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press, Cambridge, England.

Park, J.Y., 1990. Testing for unit roots and cointegration by variable addition. In: Rhodes, G.F., Fomby, T.B. (Eds.), Advances in Econometrics. JAI Press, Greenwich.

Park, J.Y., 1992. Canonical cointegrating regressions. Econometrica 60, 119–143.

Park, J.Y., 2005. Nonstationary nonlinearity: An outlook for new opportunities. In: Corbae, D., Durlauf, S.N., Hansen, B.E. (Eds.), Econometric Theory and Practice. Cambridge University Press, New York.

Park, J.Y., Hahn, S.B., 1999. Cointegrating regressions with time varying coefficients. Econometric Theory 15, 664–703.

Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part I. Econometric Theory 4, 468–497.

Park, J.Y., Qian, J., 2008. Functional regression of continuous state distributions. Mimeograph, Department of Economics, Texas A&M University.

Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. Econometrica 59, 283–306.

Phillips, P., Perron, P., 1988. Testing for a unit root in time series regression. Biometrica 75, 335–346.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegrating regressions. Econometric Theory 7, 1–21.

Shin, Y.C., 1994. A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.