Statistica Sinica Preprint No: SS-2017-0317	
Title	On a measure of lack of fit in nonlinear cointegrating regression with endogeneity
Manuscript ID	SS-2017-0317
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0317
Complete List of Authors	Qiying Wang and Ke Zhu
Corresponding Author	Ke Zhu
E-mail	mazhuke@hku.hk
Notice: Accepted version subject to English editing.	

On a measure of lack of fit in nonlinear cointegrating regression with endogeneity

Qiying Wang and Ke Zhu*

The University of Sydney and The University of Hong Kong

Abstract

This paper proposes the portmanteau test for the adequacy of nonlinear cointegrating regression models. This portmanteau test is applicable to a wide class of integrable and non-integrable regression functions with endogenous regressors derived by either short or long memory innovations, and its limiting distribution is shown to be approximated by a chi-squared distribution. Moreover, the applicability scope of this portmanteau test is generalized to include an additive nonlinear cointegrating regression model, whose consistency results are investigated as an independent interest. Finally, the importance of this portmanteau test is demonstrated by simulated and real data.

MSC2010 subject Classification: 62F03, 62F05, 62E20.

Key words and phrases: Additive model; cointegration; consistency; endogeneity; long memory regressor; nonlinear regression; nonstationarity; portmanteau test.

1 Introduction

Since the seminal work of Park and Phillips (1999, 2001), the last decade has witnessed great progresses on nonlinear cointegrating regression. As shown in Chang et al. (2001), Park and Phillips (2001), and Chan and Wang (2015), the asymptotics of the least squares estimator in parametric nonlinear cointegrating regression models highly depend on the specification of the function nonlinearity. Hence, a mis-specified or inadequate parametric model may lead to misleading statistical inferences and erroneous conclusions. This makes it desirable to propose a test for checking the adequacy of nonlinear cointegrating regression models.

*Address correspondence to Ke Zhu: Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong. E-mail: mazhuke@hku.hk

So far a growing literature has focused on testing the adequacy of parametric nonlinear cointegrating regression models. When the error term is a martingale difference sequence (m.d.s.), Kasparis (2010) constructed Bierens tests for the integrable regression function, Kasparis and Phillips (2012) proposed two robust tests for linearity, Wang and Phillips (2012) considered the kernel-smoothed U-test for integrable and non-integrable regression functions, and Wang et al. (2016) utilized the idea of the mark process to form a parametric specification test. See also Gao et al. (2009a, b) for testing linearity in autoregression and parametric time series regression. However, the m.d.s. assumption for the error term may be restrictive in applied works, since it rules out the endogenous regressors, which are expected in many potential applications but are cumbersome in the development of statistical inference methods; see, e.g., Wang and Phillips (2009a ,b) and Wang (2015). To take into account the endogenous regressor, Wang and Phillips (2016) studied the kernel-smoothed test by using the idea of Härdle and Mammen (1993); see also Gao et al. (2012). Their test is applicable in the case that the regressor is derived by short memory innovations, but it is not well-suited for practical implementation in the long memory case, resulting from the zero asymptotic size and substantial reductions in power. To our best knowledge, no attempt has been made to propose a useful test for examining the adequacy of nonlinear cointegrating regression model when the regressor is endogenous and derived by long memory innovations.

Utilizing the idea originated by Box and Pierce (1970) and Ljung and Box (1978), the present paper develops an easy-to-implement portmanteau test for checking the adequacy of parametric nonlinear cointegrating regression models. The limiting distribution of this portmanteau test is shown to be approximated by a chi-squared distribution under regular conditions, which cover a wide class of integrable and non-integrable regression functions with endogenous regressor derived by either short or long memory innovations. The implementation of this portmanteau test only requires a consistent preliminary estimator when the regression function is integrable, and a consistent preliminary estimator with certain convergence rate depending on the form of function non-linearity, when the regression function is non-integrable. In comparison with the portmanteau test for the stationary model, the estimation effect resulting from the nonlinear cointegrating regression model is not involved in the limiting distribution of this portmanteau test. In comparison with the kernel-smoothed test in Wang and Phillips (2016), this portmanteau test not only works for the endogenous regressor driven by long memory innovations, but also avoids the use of bandwidths. As we know, how to choose bandwidths is often difficult for practitioners. Furthermore, the applicability scope of this portmanteau test is generalized to include the additive nonlinear cointegrating regression model, whose consistency results developed in the present paper are interesting in their own rights.

The remainder of this paper is organized as follows. Section 2 proposes the portmanteau test for checking the adequacy of nonlinear cointegrating regression models, obtains its asymptotics, and generalizes its result to additive models. Section 3 gives the consistency results for the corresponding additive models. Simulation studies and applications are provided in Sections 4 and 5, respectively. All proofs are offered in the Appendix. Some additional simulation results are given in the supplementary material.

2 The model and main results

Consider a nonlinear cointegrating regression model

$$y_t = g(x_t, \theta) + u_t, \tag{2.1}$$

where $u_t = \rho u_{t-1} + \nu_t$ with $|\rho| < 1$, x_t is a nonstationary regressor, $g(x, \theta)$ is a given real function, and $\theta = (\theta_1, ..., \theta_m)'$ are unknown parameters which lie in the compact parameter space $\Omega_0 \subset R^m$. Model (2.1) allows the regressor x_t to be endogenous and to be derived by long memory innovations, which are two important features to meet the practical demand. However, so far no useful test for checking the adequacy of model (2.1) has taken these two features into account. This motivates us to propose a portmanteau test, which is compatible to these two features.

Assume that $\widehat{\theta}_n$ is a consistent estimator of θ_0 based on the observations $\{(x_t, y_t)\}_{t=1}^n$, where $\theta_0 = (\theta_{01}, ..., \theta_{0m})' \in \Omega_0$ is the true value of θ. Let $\widehat{u}_t = y_t - g(x_t, \widehat{\theta}_n)$ be the residual of u_t, and $\widehat{\nu}_t = \widehat{u}_t - \widehat{\rho}\widehat{u}_{t-1}$ be the residual of v_t, where

$$\widehat{\rho} = \frac{\sum_{s=2}^n \widehat{u}_s \widehat{u}_{s-1}}{\sum_{s=2}^n \widehat{u}_{s-1}^2}$$

is the least squares estimator (LSE) of ρ based on the autoregression $\widehat{u}_t = \rho \widehat{u}_{t-1} + \nu_t$. In particular, when $\rho = 0$, we set $\widehat{\nu}_t = \widehat{u}_t$ for all t. Based on $\{\widehat{\nu}_t\}_{t=1}^n$, our Portmanteau test statistic is defined by

$$\widehat{U}_n(M) := n(n+2) \sum_{k=1}^{M} \frac{\widehat{a}_k^2}{n-k}$$

for some integer $M \geq 1$, where

$$\widehat{a}_k = \frac{\sum_{t=k+1}^{n} \widehat{\nu}_t \widehat{\nu}_{t-k}}{\sum_{t=1}^{n} \widehat{\nu}_t^2}$$

is the sample autocorrelation of $\widehat{v}_t$ at lag k. Clearly, the portmanteau test $\widehat{U}_n(M)$ aims to detect the autocorrelation of the residual of v_t at the first M lags. This idea is initialled by Box and Pierce (1970) and Ljung and Box (1978), and its many variants for stationary models can be found in Romano and Thombs (1996), Francq et al. (2005), Escanciano and Lobato (2009), Delgado and Velasco (2011), Zhu (2016), and many others. As a parallel tool, the spectral test can be used to detect the residual autocorrelation at each valid lag; see, e.g., Hong (1996) and Zhu and Li (2015) for stationary models. A possible exploration on the spectral test for model (2.1) is an interesting topic for future study.

Throughout the section, let $\eta_i \equiv (\epsilon_i, \nu_i)', i \in \mathbb{Z}$, be a sequence of i.i.d. random vectors with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}(\eta_0 \eta_0') = \Sigma$, and $\mathbb{E}||\eta_0||^\alpha < \infty$ for some $\alpha > 2$; and assume that $\mathbb{E}\epsilon_0^2 = 1$ and the characteristic function $\varphi(t)$ of ϵ_0 satisfies the integrability condition $\int_{-\infty}^{\infty}(1 + |t|)\,|\varphi(t)|dt < \infty$, which assures smoothness in the corresponding density.

To establish the asymptotics of $\widehat{U}_n(M)$, we make use of the following assumptions.

ASSUMPTION 2.1. $x_t = \sum_{j=1}^{t} \xi_j$, where ξ_j, $j \geq 1$, is a linear process defined by $\xi_j = \sum_{k=0}^{\infty} \phi_k \, \epsilon_{j-k}$, with coefficients ϕ_k, $k \geq 0$, satisfying $\phi_0 \neq 0$ and one of the following conditions:

C1. $\phi_k \sim k^{-\mu}\, \pi(k)$, where $1/2 < \mu < 1$ and $\pi(k)$ is a function slowly varying at ∞;

C2. $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

ASSUMPTION 2.2. For each $\theta, \theta_0 \in \Omega_0$, there exists a bounded and integrable real function $T(x)$ such that

$$|g(x, \theta) - g(x, \theta_0)| \leq h(||\theta - \theta_0||)\, T(x), \tag{2.2}$$

where $h(x)$ is bounded real function satisfying $h(x) \to 0$ as $|x| \to 0$.

ASSUMPTION 2.3. For each $\theta, \theta_0 \in \Omega_0$, there exist positive real functions $T(x)$, $v(x)$ and $v_j(x), j = 1, ..., m$, such that, for any $\lambda > 0$,

(i) $T(\lambda x) \leq v(\lambda)(1 + |x|^\beta)$, $\left|\frac{\partial g(x,\theta_0)}{\partial \theta_j}\right| \leq T(x), j = 1, ..., m$, and

$$\left| g(x, \theta) - g(x, \theta_0) - \sum_{j=1}^{m} (\theta_j - \theta_{0j}) \frac{\partial g(x, \theta_0)}{\partial \theta_j} \right| \leq ||\theta - \theta_0||^{1+\alpha}\, T(x), \tag{2.3}$$

for some $\alpha > 0$ and $\beta > 0$;

(ii) whenever x and y are in a compact set, for each $1 \le j \le m$,

$$\Big|\frac{\partial g(\lambda x, \theta_0)}{\partial \theta_j} - \frac{\partial g(\lambda y, \theta_0)}{\partial \theta_j}\Big| \le v_j(\lambda)\big[|x-y| + R_{1j}(\lambda x) + R_{2j}(\lambda y)\big], \tag{2.4}$$

where $R_{1j}(z)$ and $R_{2j}(z)$ are bounded and integrable functions;

(iii) as $K \to \infty$, $\sup_{|x|\ge K} \max_{1\le j\le m} \frac{v(x)}{v_j(x)} < \infty$.

Assumption 2.1 allows for long (under **C1**) and short (under **C2**) memory innovations ξ_j driving the regressor x_t, and allows the equation error u_t to be cross correlated with the regressor x_s for all $s \le t$, thereby inducing endogeneity and giving the structural model (2.1). Let $d_n^2 = var(x_n)$. Under Assumption 2.1(ii), it follows from Wang et al. (2003) that

$$d_n^2 \sim \begin{cases} c_\mu\, n^{3-2\mu}\, \pi^2(n), & \text{under } \mathbf{C1}, \\ \phi^2\, n, & \text{under } \mathbf{C2}, \end{cases} \tag{2.5}$$

where $c_\mu = \frac{1}{(1-\mu)(3-2\mu)} \int_0^\infty x^{-\mu}(x+1)^{-\mu}dx$ and $\max_{1\le k\le n} |x_k|/d_n = O_P(1)$. These facts will be used later without further explanation.

Assumption 2.2 essentially requires that $g(x,\theta)$ is bounded and integrable for each $\theta \in \Omega_0$. Typical examples for Assumption 2.2 include these integrable functions: $g(x,\theta) = \theta_1 |x|^{\theta_2} I(x \in [a,b])$ for finite constants a and b, the Gaussian function $g(x,\theta) = \theta_1 e^{-\theta_2 x^2}$, and the Laplacian function $g(x,\theta) = \theta_1 e^{-\theta_2|x|}$. Assumption 2.3 removes the boundedness and integrability conditions on $g(x,\theta)$, but imposing more smooth conditions for technical reasons. Typical examples for Assumption 2.3 include these asymptotically homogeneous functions: $g(x,\theta) = (x+\theta)^2, \theta e^x/(1+e^x), \theta \log|x|, \theta|x|^\alpha$ (α is fixed), and $\theta_1 + \theta_2|x| + \ldots + \theta_k|x|^k$. Both Assumptions 2.2 and 2.3 are weak and partially used in Wang and Phillips (2016) for the estimation of parameter θ in model (2.1). See also Section 3 below for more details.

We have the following main results for $\widehat{U}_n(M)$.

THEOREM 2.1. *Suppose that Assumptions 2.1 and 2.2 hold and there exists an estimator $\widehat{\theta}_n$ such that $\widehat{\theta}_n \in \Omega_0$ and $\widehat{\theta}_n \to_P \theta_0$. If model (2.1) is correctly specified, then the limiting distribution of $\widehat{U}_n(M)$ can be approximated by χ^2_{M-1} for large M.*

THEOREM 2.2. *Suppose that Assumptions 2.1 and 2.3 hold and there exists an estimator $\widehat{\theta}_n$ such that $\widehat{\theta}_n \in \Omega_0$ and $||D_n(\widehat{\theta}_n - \theta_0)|| = O_P(\log^\delta n)$ for some $\delta > 0$, where $D_n = diag\big(\sqrt{n}\, v_1(d_n), \ldots, \sqrt{n}\, v_m(d_n)\big)$. If model (2.1) is correctly specified, then the limiting distribution of $\widehat{U}_n(M)$ can be approximated by χ^2_{M-1} for large M.*

REMARK 2.1. The establishment of Theorems 2.1 and 2.2 depends only on the fact that, for any $k \geq 0$,

$$\frac{1}{\sqrt{n}} \sum_{s=k+1}^{n} \widehat{u}_s \, \widehat{u}_{s-k} = \frac{1}{\sqrt{n}} \sum_{s=k+1}^{n} u_s \, u_{s-k} + o_P(1), \tag{2.6}$$

which guarantees that the estimation effect on θ does not exist in the limiting distribution of $\widehat{U}_n(M)$. Indeed, due to (2.6), some standard calculations yield that

$$\sqrt{n}\widehat{a}_k = \sqrt{n}\overline{a}_k + o_P(1) := \sqrt{n}\left(\frac{\sum_{t=k+1}^{n} \overline{\nu}_t \overline{\nu}_{t-k}}{\sum_{t=1}^{n} \overline{\nu}_t^2}\right) + o_P(1),$$

where $\overline{\nu}_t = u_t - \overline{\rho} u_{t-1}$ and

$$\overline{\rho} = \frac{\sum_{s=2}^{n} u_s u_{s-1}}{\sum_{s=2}^{n} u_{s-1}^2}.$$

Hence, the limiting distribution of $\widehat{U}_n(M)$ is the same as the one of $\overline{U}_n(M)$, where

$$\overline{U}_n(M) = n(n+2) \sum_{k=1}^{M} \frac{\overline{a}_k^2}{n-k}.$$

Note that $\overline{\rho}$ is the LSE of ρ in the autoregressive model: $u_t = \rho u_{t-1} + \nu_t$, and $\overline{a}_k$ is exactly the lag-k autocorrelation of its model residuals. Therefore, the limiting distribution of $\overline{U}_n(M)$ (or $\widehat{U}_n(M)$), involving the estimation effect on ρ, has been given in Theorem 3 of Francq et al. (2005), and it can be approximated by χ^2_{M-1} for large M, when ν_t is i.i.d.

Under Assumption 2.1, the regressor x_t is nonstationary. If the regression function $g(x, \theta)$ is bounded and integrable, result (2.6) can be established under minimum conditions that $\widehat{\theta}_n \in \Omega_0$ and $\widehat{\theta}_n \to_P \theta_0$. The reason for this phenomenon is that nonstationarity weakens the signal and hence the restriction imposed on $\widehat{\theta}_n$ when $g(x, \theta)$ is integrable. This is quite different from the stationary regression and time series model. In the latter case, one usually requires $\sqrt{n}$-consistency of a preliminary estimator. If $g(x, \theta)$ is not bounded and integrable, result (2.6) requires certain convergence rate on $\widehat{\theta}_n$ to check the adequacy of model (2.1), and this is again different from the stationary situation, since the convergence rate depends on the form of $g(x, \theta)$. It should be mentioned that, both convergence conditions required for $\widehat{\theta}_n$ in Theorems 2.1 and 2.2 can be achieved under Assumption 2.1 and some additional smooth conditions on $g(x, \theta)$; see more details in Section 3.

REMARK 2.2. The portmanteau test $\widehat{U}_n(M)$ aims to check whether the form of $g(x, \theta)$ is correctly specified, but can not be used in the situation that $g(x, \theta)$ itself is unknown.

To see it clearly, we consider a simple nonparametric cointegrating regression model:

$$y_t = g(x_t, \theta_0) + u_t,$$

where θ_0 is given and $g(x, \theta_0)$ is an unknown real function. As investigated in Wang and Phillips (2009a, 2009b, 2016), the function $g(x, \theta_0)$ can be estimated by the conventional kernel estimator

$$\widehat{g}(x, \theta_0) = \frac{\sum_{t=1}^n y_t K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]},$$

where $K(x)$ is a positive kernel function and $h \to 0$ is a bandwidth. Define $\widehat{u}_t = y_t - \widehat{g}(x, \theta_0)$. As noticed in Lindon and Wang (2016), it is unrealistic to establish (2.6) even for $k = 0$ due to the slow convergence rate for $\widehat{g}(x, \theta_0) \to_P g(x, \theta_0)$, indicating that the portmanteau test $\widehat{U}_n(M)$ can not be used in nonparametric cointegrating regression models with nonstationarity.

REMARK 2.3. The condition that ν_t is i.i.d. is standard in the nonstationary time series literature; see, e.g., Chan and Wang (2015), Wang (2015), Wang and Phillips (2016), and many others. This technical condition is not necessary. Some simple algebra in part A.1 of Appendix shows that ν_t can be replaced by less restrictive linear process $\nu_t' = \sum_{k=0}^\infty \psi_k \nu_{t-k}$ with $\sum_{k=0}^\infty k^{1/4} |\psi_k| < \infty$. It is not clear, however, whether or not ν_t can be replaced by a nonlinear stationary process such as the autoregressive conditional heteroskedasticity (ARCH) type errors. Numerically, simulation studies in the supplementary material show that our portmanteau tests (with a slight modification to take into account the conditional heteroskedasticity and the estimation effect on ρ) have good finite-sample performance, when ν_t has the ARCH-type structure. Theoretically, new technique is required to modify Lemma A.1 in Appendix from a linear process ν_t to a nonlinear stationary process. This kind of modification seems to be very challengeable at the moment and hence the topic is left for future work.

REMARK 2.4. Consider model (2.1) with AR(p) errors, i.e., u_t is assumed to be strictly stationary satisfying

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \nu_t, \tag{2.7}$$

where $1 - \rho_1 z - \rho_2 z^2 - \cdots - \rho_p z^p \neq 0$ when $|z| \leq 1$. In this situation, we set $\widetilde{\nu}_t = \widehat{u}_t - \sum_{j=1}^p \widehat{\rho}_j \widehat{u}_{t-j}$, where $(\widehat{\rho}_1, \ldots, \widehat{\rho}_p)'$ is the LSE of $(\rho_1, \ldots, \rho_p)'$ based on the autoregression

$\widehat{u}_t = \rho_1 \widehat{u}_{t-1} + \rho_2 \widehat{u}_{t-2} + \cdots + \rho_p \widehat{u}_{t-p} + \nu_t$. As before, we construct the portmanteau test statistic by

$$\widetilde{U}_n(M) := n(n+2) \sum_{k=1}^{M} \frac{\widetilde{a}_k^2}{n-k}$$

for some integer $M \geq 1$, where

$$\widetilde{a}_k = \frac{\sum_{t=k+1}^{n} \widetilde{\nu}_t \widetilde{\nu}_{t-k}}{\sum_{t=1}^{n} \widetilde{\nu}_t^2}.$$

Under the conditions in Theorem 2.1 or 2.2, we can similarly show that the limiting distribution of $\widetilde{U}_n(M)$ can be approximated by χ^2_{M-p} for large M.

To end this section, we show that the results of our portmanteau tests can be generalized to the following additive nonlinear cointegrating regression model:

$$y_t = g(x_t, \theta) + f(z_t, \eta) + u_t, \tag{2.8}$$

where $u_t = \rho u_{t-1} + \nu_t$ with $|\rho| < 1$, x_t and z_t are nonstationary regressors, $g(x, \theta)$ and $f(x, \eta)$ are given real functions, and $\theta = (\theta_1, ..., \theta_m)'$ and $\eta = (\eta_1, ..., \eta_k)'$ are unknown parameters which lie in the compact parameter space $\Omega_0 \subset R^m$ and $\Omega_1 \subset R^k$, respectively.

Let θ_0 and η_0 be true values of θ and η in model (2.8). As x_t and $g(x, \theta)$ in Assumptions 2.1 and 2.3, we make the following two assumptions for z_t and $f(x, \eta)$, respectively.

ASSUMPTION 2.4. $z_t = \sum_{j=1}^{t} \zeta_j$, where $\{\zeta_j, j \geq 1\}$ is a linear process defined by $\zeta_j = \sum_{k=0}^{\infty} \varphi_k \, \epsilon_{j-k}$, with coefficients φ_k, $k \geq 0$, satisfying $\varphi_0 \neq 0$ and one of the following conditions:

C1'. $\varphi_k \sim k^{-\mu} \pi(k)$, where $1/2 < \mu < 1$ and $\pi(k)$ is a function slowly varying at ∞.

C2'. $\sum_{k=0}^{\infty} |\varphi_k| < \infty$ and $\varphi \equiv \sum_{k=0}^{\infty} \varphi_k \neq 0$.

ASSUMPTION 2.5. For each $\eta, \eta_0 \in \Omega_1$, there exists a bounded and integrable real function $T(x)$ such that

$$|f(x, \eta) - f(x, \eta_0)| \leq h(||\eta - \eta_0||) \, T(x), \tag{2.9}$$

where $h(x)$ is bounded real function satisfying $h(x) \to 0$ as $|x| \to 0$.

It is worth noting that the innovation ϵ_i in z_t can be replaced by the random sequence ϵ_i^* satisfying that $(\epsilon_i, \epsilon_i^*, \nu_i)'$, $i \in \mathbb{Z}$, are i.i.d. random vectors and ϵ_i^* has the same distributional properties as those of ϵ_i. Also, as discussed in Remark 2.2, the technical condition that ν_t is i.i.d. is not entirely necessary for our asymptotics below to hold.

As before, we define the portmanteau test statistic $\widehat{U}_n(M)$ for model (2.8), but with $\widehat{u}_t$ being replaced by

$$\widehat{u}_t = y_t - g(x_t, \widehat{\theta}_n) - f(z_t, \widehat{\eta}_n),$$

where $\widehat{\theta}_n$ and $\widehat{\eta}_n$ are consistent estimators of θ and η, respectively. We have the following result which extends Theorems 2.1 and 2.2.

THEOREM 2.3. *Suppose Assumptions 2.1 and 2.3-2.5 hold, and there exist estimators $\widehat{\theta}_n$ and $\widehat{\eta}_n$ such that (i) $\widehat{\theta}_n \in \Omega_0$ and $||D_n(\widehat{\theta}_n - \theta_0)|| = O_P(\log^\delta n)$ for some $\delta > 0$, where $D_n = diag\big(\sqrt{n}\, v_1(d_n), ..., \sqrt{n}\, v_m(d_n)\big)$, and (ii) $\widehat{\eta}_n \in \Omega_1$ and $\widehat{\eta}_n \to_P \eta_0$. If model (2.8) is correctly specified, then the limiting distribution of $\widehat{U}_n(M)$ can be approximated by χ^2_{M-1} for large M.*

REMARK 2.5. The estimators $\widehat{\theta}_n$ and $\widehat{\eta}_n$ of θ and η in model (2.8) that satisfy the conditions required in Theorem 2.3 will be constructed in the next section. In principle, there are no technical difficulties in extending model (2.8) to allow for the time trend and more integrable and non-integrable functions whenever the parameters appeared in the corresponding models can be estimated with certain convergence rates. When the regressors are endogenous and derived by long memory innovations, however, it seems to be quite challengeable in constructing the corresponding consistency estimators under general settings on the model. More details can be found in Remark 3.1.

3 Parametric consistency

The estimation of θ in model (2.1) has been considered in Wang and Phillips (2016). In this section, we provide primitive conditions for the verification of consistent parametric estimation of θ and η in model (2.8), which is required in Theorem 2.3 and seems to be new to literature.

Let $w_t = f(z_t, \eta) + u_t$. Then, model (2.8) can be rewritten as

$$y_t = g(x_t, \theta) + w_t. \tag{3.1}$$

Note that the behavior of w_t is similar to a stationary process due to the boundedness and integrability of $f(x, \eta)$. The unknown parameters θ_0 and η_0 in model (2.8) can be estimated by the two-step non-linear least squares estimation procedure as follows.

Step 1: Estimate θ_0 by

$$\widehat{\theta}_n = \arg\min_{\theta\in\Omega_0} \sum_{t=1}^{n} \big[y_t - g(x_t,\theta)\big]^2.$$

Step 2: Set $\widehat{w}_t = y_t - g(x_t,\widehat{\theta}_n)$. Estimate η_0 by

$$\widehat{\eta}_n = \arg\min_{\eta\in\Omega_1} \sum_{t=1}^{n} \big[\widehat{w}_t - f(z_t,\eta)\big]^2.$$

To establish the consistent properties of $\widehat{\theta}_n$ and $\widehat{\eta}_n$ as required in Theorem 2.3, we need some additional smooth conditions on $g(x,\theta)$ and $f(x,\eta)$. Let $\dot{g}$ and $\ddot{g}$ be the first and second derivatives of $g(x,\theta)$, so that $\dot{g} = \partial g/\partial\theta$ and $\ddot{g} = \partial^2 g/\partial\theta\partial\theta'$. Similar definitions are used for $\dot{f}$ and $\ddot{f}$.

ASSUMPTION 3.1. Let $p(x,\theta)$ be any of g, $\dot{g}_i$ or $\ddot{g}_{ij}$, $1\le i,j\le m$. There exists a positive real function $v_p(\lambda)$ which is bounded away from zero as $\lambda\to\infty$ and a constant $\beta\ge 0$ such that, for each $\theta,\theta_0\in\Omega_0$.

(i) $|p(x,\theta)-p(x,\theta_0)| \le C\,||\theta-\theta_0||\,T_{1p}(x)$, where $T_{1p}(\lambda x)\le C\,v_p(\lambda)\,(1+|x|^{\beta})$;

(ii) $p(\lambda x,\theta_0)\le C\,v_p(\lambda)\,(1+|x|^{\beta})$, and for $p(x,\theta_0)=\dot{g}_i(x,\theta_0)$ or $\ddot{g}_{ij}(x,\theta_0), 1\le i,j\le m$,

$$|p(\lambda x,\theta_0)-p(\lambda y,\theta_0)| \le C v_p(\lambda)\big[|x-y| + R_{1p}(\lambda x) + R_{2p}(\lambda x)\big],$$

whenever x and y are in a compact set, where $R_{1p}(z)$ and $R_{2p}(z)$ are bounded and integrable functions;

(iii) $\dot{g}_i(\lambda x,\theta_0) = v_{\dot{g}_i}(\lambda)\,h_i(x,\theta_0) + R_i(\lambda,x,\theta_0)$ for $1\le i\le m$, where $R_i(\lambda,x,\theta_0) = o\big[v_{\dot{g}_i}(\lambda)\,h_i(x,\theta_0)\big]$ *as* $|\lambda|\to\infty$, and $h_i(x,\theta_0)$ is a locally bounded function (i.e., bounded on any compact set) satisfying $\sum_\delta = \int_{|s|\le\delta} h(s,\theta_0)h(s,\theta_0)'ds > 0$ for all $\delta>0$, where $h(x,\theta_0) = \big(h_1(x,\theta_0),...,h_m(x,\theta_0)\big)'$;

(iv) $\sup_{1\le j\le m}|\frac{v(d_n)}{\dot{v}_j(d_n)}|<\infty$ and $\sup_{1\le i,j\le m}|\frac{v(d_n)\,\ddot{v}_{ij}(d_n)}{\dot{v}_i(d_n)\,\dot{v}_j(d_n)}|<\infty$, where $v(\lambda)=v_g(\lambda)$, $\dot{v}_i(\lambda)=v_{\dot{g}_i}(\lambda)$ and $\ddot{v}_{ij}(\lambda)=v_{\ddot{g}_{ij}}(\lambda)$.

ASSUMPTION 3.2. *Let $p(x,\eta)$ be any of f, $\dot{f}_i$* or $\ddot{f}_{ij}$, $1\le i,j\le k$.

(i) $p(x,\eta_0)$ is a bounded and integrable real function;

(ii) there exists a bounded and integrable function $T_p: R\to R$ such that $|p(x,\eta)-p(x,\eta_0)|\le C\,||\eta-\eta_0||\,T_p(x)$, for each $\eta,\eta_0\in\Omega_1$;

(iii) $\Sigma = \int_{-\infty}^{\infty} \dot{f}(s,\eta_0)\dot{f}(s,\eta_0)' ds > 0$ for each $\eta_0 \in \Omega_1$, where $\dot{f}(s,\eta_0) = \big(\dot{f}_1(s,\eta_0), ..., \dot{f}_k(s,\eta_0)\big)'$.

Both Assumptions 3.1 and 3.2 are used in Wang and Phillips (2016) for the consistency of θ in model (2.1). Assumption 3.1 allows for asymptotically homogeneous functions, and Assumption 3.2 holds for a wide range of integrable regression functions; see Section 2 above for more specific examples in each group.

We have the following result for the consistency of $\widehat{\theta}_n$ and $\widehat{\eta}_n$, and it indicates that $\widehat{\theta}_n$ and $\widehat{\eta}_n$ are applicable to construct $\widehat{U}_n(M)$.

THEOREM 3.1. *Suppose that Assumptions 2.1, 2.4-2.5 and 3.1 hold, and* $\tau = \int_{-\infty}^{\infty} [f(x,\eta) - f(x,\eta_0)]^2 dx \neq 0$ *for any* $\eta \neq \eta_0$. *Then, under model (2.8), we have*

$$||D_n(\widehat{\theta}_n - \theta_0)|| = O_P(1) \quad \textit{and} \quad \widehat{\eta}_n \to_P \eta_0, \tag{3.2}$$

where $D_n = diag(\sqrt{n} v_{\dot{g}_1}(d_n), ..., \sqrt{n} v_{\dot{g}_m}(d_n))$. *If in addition Assumption 3.2 holds, we further have*

$$||\widehat{\eta}_n - \eta_0|| = \left(\frac{d_{1n}}{n}\right)^{1/2} \begin{cases} O_P(1), & \textit{under } \textbf{\textit{C1'}}, \\ O_P(\log^{1/2} n), & \textit{under } \textbf{\textit{C2'}}, \end{cases} \tag{3.3}$$

where $d_{1n}^2 = var(z_n)$.

REMARK 3.1. When there is a martingale difference structure in the error term, Chang et al. (2001) considered the non-linear LSE in a general additive model, including the time trend and more integrable and non-integrable regression functions. The present model (2.8) is less general than that of Chang et al. (2001), but allowing for the endogenous regressors derived by the long memory innovations. From a view of nonlinear cointegrating regression, endogeneity seems to be more demands in practice. Moreover, unlike the LSE in Chang et al. (2001), the estimators $\widehat{\theta}_n$ and $\widehat{\eta}_n$ in present model (2.8) are constructed by using two-step least squares estimation procedure. For the usual LSE, it requires new technical development to establish the general limit distribution theory for $\widehat{\theta}_n$ and $\widehat{\eta}_n$; see, e.g., Wang and Phillips (2016). This remaining challenge in nonlinear nonstationary asymptotics will be left for future work. Although the theoretical development is absent, simulation studies in the supplementary material show that our portmanteau tests can have good finite-sample performance for the additive model in Chang et al. (2001) with the endogenous and long memory regressor. This implies that our portmanteau test shall have a very wide scope of applicability.

4 Simulation

In this section, we examine the finite sample performance of $\widehat{U}_n(M)$ for integrable regression functions, non-integrable regression functions, and additive regression functions. We consider the case that the error term u_t follows AR(1) model with an i.i.d. innovation in the sequel, and more simulation results can be found in the supplementary material when u_t follows AR(1) model with an ARCH-type innovation.

4.1 Integrable regression function

We generate 5000 replications of sample size $n = 100, 200$, or 500 from the following five different data-generating models:

$$y_t = \exp(-\theta_0|x_t|) + u_t; \tag{4.1}$$

$$y_t = \exp(-\theta_0|x_t|) + 0.5|x_t|^2 I(|x_t| \leq 10) + u_t; \tag{4.2}$$

$$y_t = \exp(-\theta_0|x_t|) + 20\exp(-|x_t|^2) + u_t; \tag{4.3}$$

$$y_t = \exp(-\theta_0|x_t|) + 0.1|x_t| + u_t; \tag{4.4}$$

$$y_t = \exp(-\theta_0|x_t|) + 0.1|x_t|^2 + u_t, \tag{4.5}$$

where $\theta_0 = 1$, $x_t = x_{t-1} + \xi_t$ with $(1-0.8B)(1-B)^d\xi_t = (1+0.3B)\epsilon_t$, $u_t = \rho u_{t-1} + \nu_t$ with $\rho = \pm 0.5$, and

$$(\epsilon_t, \nu_t) \sim \text{i.i.d. } N\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right).$$

Here, model (4.1) is used as the null model, and models (4.2)-(4.5) are used as alternative models, in which the first (or last) two models deviate from the null model by an integrable (or non-integrable) function. For each examined model, the regressor x_t is designed to be short memory ($d = 0$) or long memory ($d = 0.2$), and exogenous ($r = 0$) or endogenous ($r = 0.5$ or 0.8). In all calculations, we compute $\widehat{\theta}_n$ as the non-linear LSE of θ_0 based on model (4.1).

Table 1 reports the size and power of $\widehat{U}_n(M)$ for $M = 6, 12$ and 18 at the significance level 5%, and the size of $\widehat{U}_n(M)$ is corresponding to the case that $y_t \sim$ model (4.1), where the critical value of $\widehat{U}_n(M)$ is chosen to be the 5% upper percentile of χ^2_{M-1}. From this table, our findings are as follows.

(ai) The size of $\widehat{U}_n(M)$ is precise in general, although it seems to be slightly oversized when $M = 12$ (or 18) and n is small.

(aii) The power of $\widehat{U}_n(M)$ is less affected by the choice of M, and it increases with the value of n.

(aiii) Generally, the power of $\widehat{U}_n(M)$ under models (4.4)-(4.5) is larger than the one under models (4.2)-(4.3).

(avi) For each examined alternative with the same values of ρ and d, the power of $\widehat{U}_n(M)$ is almost unchanged to the choise of r, meaning that the endogeneity of x_t has litter impact on the performance of $\widehat{U}_n(M)$. For each examined alternative with the same value of ρ, the power of $\widehat{U}_n(M)$ is robust to the choices of d, especially when $M = 12$ and 18. For each examined alternative, the power of $\widehat{U}_n(M)$ for the case of $\rho = -0.5$ is larger than the corresponding one for the case of $\rho = 0.5$ in general.

Please insert Table 1 about here.

4.2 Non-integrable regression function

We generate 5000 replications of sample size $n = 100, 200$, or 500 from the following five different data-generating models:

$$y_t = \theta_{10} + \theta_{20}x_t + u_t; \tag{4.6}$$

$$y_t = \theta_{10} + \theta_{20}x_t + 0.5|x_t|^2 I(|x_t| \leq 10) + u_t; \tag{4.7}$$

$$y_t = \theta_{10} + \theta_{20}x_t + 20\exp(-|x_t|^2) + u_t; \tag{4.8}$$

$$y_t = \theta_{10} + \theta_{20}x_t + 0.1|x_t| + u_t; \tag{4.9}$$

$$y_t = \theta_{10} + \theta_{20}x_t + 0.1|x_t|^2 + u_t, \tag{4.10}$$

where $(\theta_{10}, \theta_{20}) = (0, 1)$ and the remaining set-ups inherit from models (4.1)-(4.5). In all calculations, we compute $(\widehat{\theta}_{0n}, \widehat{\theta}_{1n})$ as the non-linear LSE of $(\theta_{10}, \theta_{20})$ based on model (4.6).

Table 2 reports the size and power of $\widehat{U}_n(M)$ at the significance level 5%, and the size of $\widehat{U}_n(M)$ is corresponding to the case that $y_t \sim$ model (4.6), where the critical value of $\widehat{U}_n(M)$ is chosen to be the 5% upper percentile of χ^2_{M-1}. From this table, our findings are similar to those in Table 1, except that the power of $\widehat{U}_n(M)$ seems to less satisfactory when $y_t \sim$ model (4.9) with $\rho = 0.5$ and small n.

Please insert Table 2 about here.

4.3 Additive regression function

We generate 5000 replications of sample size $n = 100, 200$, or 500 from the following five different data-generating models:

$$y_t = \theta_{10} + \theta_{20} x_t + \exp(-\eta_0 |z_t|) + u_t; \tag{4.11}$$

$$y_t = \theta_{10} + \theta_{20} x_t + \exp(-\eta_0 |z_t|) + 0.5|\kappa_t|^2 I(|\kappa_t| \le 10) + u_t; \tag{4.12}$$

$$y_t = \theta_{10} + \theta_{20} x_t + \exp(-\eta_0 |z_t|) + 20 \exp(-|\kappa_t|^2) + u_t; \tag{4.13}$$

$$y_t = \theta_{10} + \theta_{20} x_t + \exp(-\eta_0 |z_t|) + 0.1|\kappa_t| + u_t; \tag{4.14}$$

$$y_t = \theta_{10} + \theta_{20} x_t + \exp(-\eta_0 |z_t|) + 0.1|\kappa_t|^2 + u_t, \tag{4.15}$$

where $\kappa_t = \max(x_t, z_t)$, $(\theta_{10}, \theta_{20}, \eta_0) = (0, 1, 1)$, $z_t = z_{t-1} + \zeta_t$ with $(1 - 0.8B)(1 - B)^d \zeta_t = (1 + 0.3B)\epsilon_t^*$,

$$(\epsilon_t, \epsilon_t^*, \nu_t) \sim \text{i.i.d. } N\left(0, \begin{pmatrix} 1 & 0.5 & r \\ 0.5 & 1 & 0.5 \\ r & 0.5 & 1 \end{pmatrix}\right),$$

and the remaining set-ups inherit from models (4.1)-(4.5). In all calculations, we compute $(\widehat{\theta}_{0n}, \widehat{\theta}_{1n}, \widehat{\eta}_n)$ as the two-step non-linear LSE of $(\theta_{10}, \theta_{20}, \eta_0)$ based on model (4.11).

Table 3 reports the size and power of $\widehat{U}_n(M)$ at the significance level 5%, and the size of $\widehat{U}_n(M)$ is corresponding to the case that $y_t \sim$ model (4.11), where the critical value of $\widehat{U}_n(M)$ is chosen to be the 5% upper percentile of χ^2_{M-1}. From this table, our findings are similar to those in Table 2. Meanwhile, we should highlight that the additional simulation studies in the supplementary material imply that our portmanteau test $\widehat{U}_n(M)$ also has a good finite sample performance for the additive model in Chang et al. (2001) with time trend and two integrable or non-integrable functions.

Please insert Table 3 about here.

In summary, regardless of the type of the regression function, our portmanteau test has a good finite sample performance in all examined cases. Particularly, our portmanteau test is not affected by the endogeneity of the regressor, and it works well for the regressor driven by either short or long memory innovations. These features shall be important for practitioners.

5 Application

In this section, we study the Carbon Kuznets Curve (CKC), which relates the per capita CO_2 emission of a country to its per capita GDP. As argued in Piaggio and Padilla

(2012) and Chan and Wang (2015), the CKC has the inverted-U shape (see, e.g., the right panel in Figure 1 below), where the upward slope of the CKC can be explained by the increase in natural resources depletion as economic activities grow, and the downward slope of the CKC results from a reduction in the emission of air pollutants as the country continues to develop technological advance and stricter regulatory policies. Suggested by the aforementioned two papers, we consider a quadratic polynomial formulation for the CKC to capture its inverted-U shape:

$$\begin{cases} \log(e_t) = \theta_1 + \theta_2 \log(x_t) + \theta_3[\log(x_t)]^2 + u_t, \\ u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \nu_t, \end{cases} \tag{5.1}$$

for $1 \leq t \leq n$, where e_t and x_t are the per capita emissions of CO_2 and GDP in period t, respectively. Here, we use an AR(p) model to fit u_t, since the specification test in Wang et al. (2016) indicates that most likely, u_t is not an m.d.s.

Now, we are interested in examining whether model (5.1) can fit the CKC adequately in 16 countries based on the annual data from 1951 to 2009 for each country (see Piaggio and Padilla (2012) and Chan and Wang (2015)). We first choose the order $p \in \{1, 2, \cdots, 6\}$ by the Bayesian information criterion (BIC) based on each data set, and it turns out that $p = 1$ is selected in all cases except CHN and JAP; see Table 4. Hence, we apply our portmanteau tests $\widetilde{U}_n(M)$ for CHN (or JAP) and $\widehat{U}_n(M)$ for the remaining countries to check the adequacy of model (5.1), and the corresponding results are given in Table 4. From this table, we find the strong evidence that model (5.1) can not fit the CKC adequately for JAP.

Please insert Table 4 about here.

It is worth noting that if we apply the Akaike information criterion (AIC) to select the order p, we get the same results as those based on BIC, except that the order $p = 5$ is selected for JAP. In this case, the p-value of $\widetilde{U}_n(M)$ (M=6, 12 or 18) for JAP is also close to zero, indicating the same conclusion made above.

To gain more evidence, Figure 1 plots the CKC for JAP and USA. From this figure, we can see that the CKC for JAP does not have the inverted-U shape as the one for USA, and hence this may result in the inadequacy of model (5.1) to fit the CKC for JAP.

Please insert Figure 1 about here.

6 Concluding remarks

In this paper, we propose the portmanteau test for the adequacy of nonlinear cointegrating regression models. This portmanteau test is based on a two-step estimation procedure. However, unlike the portmanteau test in stationary models, we find that the limiting distribution of this portmanteau test does not involve the estimation effect in the first step estimation of the nonlinear cointegrating regression model. Therefore, the limiting distribution of this portmanteau test is the same as the one for the stationary autoregressive model, and it can be approximated by a simple chi-squared distribution. Compared to the kernel-smoothed test in Gao et al. (2009b) and Wang and Phillips (2012, 2016), this portmanteau test has two advantages: first, it is valid for the endogenous regressor driven by long memory innovations; second, it is easy-to-implement without selecting bandwidths. Furthermore, we generalize the applicability scope of this portmanteau test to include the additive nonlinear cointegrating regression model, whose consistency results are established. Simulation studies reveal that this portmanteau test can have a very wide scope of applicability. Finally, this portmanteau test is used to study the CKC in 16 countries.

Acknowledgements

The authors thank two anonymous referees, the Associate Editor and the Co-Editor Ruey S. Tsay for very helpful comments and suggestions, leading to a substantial improvement in the presentation and the elimination of one error in the previous manuscript. Wang acknowledges research support from the Australian Research Council. Zhu acknowledges research support from NSFC (No.11571348, 11371354, 11690014, 11731015 and 71532013), Seed Fund for Basic Research (No. 201611159233), and Hung Hing Ying Physical Sciences Research Fund 2017-18.

Appendix: Proof of main results

A.1 Proofs of Theorems 2.1-2.3

In this appendix, we only prove Theorem 2.3, as others are similar except simpler. To facilitate the proof, the following lemma is needed, and its proof is referred to (7.2)-(7.3), (7.7) and (7.9) in Wang and Phillips (2016) with minor modifications due to the fact that, for the u_t appeared in one of model (2.1), Remark 2.3 and model (2.7), we may write $u_t = \sum_{k=0}^{\infty} \psi_{1k} \nu_{t-k}$ with the coefficients ψ_{1k} satisfying $\sum_{k=0}^{\infty} k^{1/4} |\psi_{1k}| < \infty$.

LEMMA A.1. *Suppose that Assumption 2.1 holds.*

(i) If $l(x)$ is a bounded function satisfying $\int_{-\infty}^{\infty} |l(x)| dx < \infty$, then

$$\frac{d_n}{n} \sum_{s=k+1}^{n} \big[|l(x_s)| \, (1 + |u_{s-k}|) + |l(x_{s-k})| \, |u_s| \big] = O_P(1), \tag{A.1}$$

$$\Big(\frac{d_n}{n}\Big)^{1/2} \sum_{k=1}^{n} l(x_k) u_k = \begin{cases} O_P(1), & \textit{under } \boldsymbol{C1}, \\ O_P(\log^{1/2} n), & \textit{under } \boldsymbol{C2}. \end{cases} \tag{A.2}$$

(ii) If $l(x)$ is a locally bounded function, then

$$\frac{1}{n} \sum_{s=k+1}^{n} \big[|l(x_s/d_n)| \, (1 + |u_{s-k}|) + |l(x_{s-k}/d_n)| \, |u_s| \big] = O_P(1). \tag{A.3}$$

(iii) Let $v(\lambda)$ be a positive real function which is bounded away from zero as $\lambda \to \infty$. For any real function $l(x)$ satisfying $|l(\lambda x)| \le C\, v(\lambda)(1 + |x|^{\beta})$ for some $\beta > 0$ and

$$|l(\lambda x) - l(\lambda y)| \le C\, v(\lambda) \big[|x - y| + R_1(\lambda x) + R_2(\lambda y) \big], \tag{A.4}$$

whenever x and y are in a compact set, where $R_1(z)$ and $R_2(z)$ are bounded and integrable functions, we have

$$\frac{1}{v(d_n)\sqrt{n}} \sum_{s=k+1}^{n} \big[l(x_s) u_{s-k} + l(x_{s-k}) u_s \big] = O_P(1). \tag{A.5}$$

(iv) Results in (i)–(iii) still hold if we replace x_t and $d_n^2 = var(x_n)$ by z_t and $d_{1n}^2 = var(z_n)$, respectively.

Proof of Theorem 2.3. As noticed in Remark 2.1, using some standard arguments, it suffices to show that, for any $k \ge 0$, we have

$$\frac{1}{\sqrt{n}} \sum_{s=k+1}^{n} \widehat{u}_s \, \widehat{u}_{s-k} = \frac{1}{\sqrt{n}} \sum_{s=k+1}^{n} u_s \, u_{s-k} + o_P(1), \tag{A.6}$$

where $\widehat{u}_t = y_t - g(x_t, \widehat{\theta}_n) - f(z_t, \widehat{\eta}_n)$. Let $\Delta_s = \Delta_{1,s} + \Delta_{2,s}$, where

$$\Delta_{1,s} = g(x_s, \widehat{\theta}_n) - g(x_s, \theta_0) \quad \text{and} \quad \Delta_{2,s} = f(z_s, \widehat{\eta}_n) - f(z_s, \eta_0).$$

For any $k \geq 0$, we may write that $\widehat{u}_s = u_s + \Delta_s$ and

$$\sum_{s=k+1}^{n} \widehat{u}_s \, \widehat{u}_{s-k} = \sum_{s=k+1}^{n} u_s u_{s-k} + R_{1n} + R_{2n} + R_{3n}, \tag{A.7}$$

where

$$R_{1n} = \sum_{s=k+1}^{n} u_s \, \Delta_{s-k}, \quad R_{2n} = \sum_{s=k+1}^{n} u_{s-k} \, \Delta_s \quad \text{and} \quad R_{3n} = \sum_{s=k+1}^{n} \Delta_s \, \Delta_{s-k}.$$

The result (A.6) will follow if we prove

$$R_{in} = o_P(\sqrt{n}), \quad i = 1, 2, 3. \tag{A.8}$$

We first prove (A.8) for $i = 2$. Since $f(x, \theta)$ satisfies (2.9), it follows from (A.1) with $l(x) = T(x)$ that

$$\begin{aligned}
\sum_{s=k+1}^{n} |u_{s-k}| \, |\Delta_{2,s}| &\leq h(||\widehat{\eta}_n - \eta_0||) \sum_{s=k+1}^{n} |u_{s-k}| \, T(x_s) \\
&= O_P(\sqrt{n}) \, h(||\widehat{\eta}_n - \eta_0||) = o_P(\sqrt{n}).
\end{aligned} \tag{A.9}$$

On the other hand, under Assumption 2.3, we have

$$\sum_{s=k+1}^{n} u_{s-k} \, \Delta_{1,s} := \sum_{j=1}^{m} (\widehat{\theta}_{nj} - \theta_{0j}) \sum_{s=k+1}^{n} u_{s-k} \, \frac{\partial g(x_s, \theta_0)}{\partial \theta_j} + R_{2n}^*,$$

where, by using (A.3),

$$\begin{aligned}
|R_{2n}^*| &\leq ||\widehat{\theta}_n - \theta_0||^{1+\alpha} \sum_{s=k+1}^{n} |u_{s-k}| T(x_t) \\
&\leq ||\widehat{\theta}_n - \theta_0||^{1+\alpha} \, v(d_n) \sum_{s=k+1}^{n} |u_{s-k}| \big[1 + (|x_t|/d_n)^{\beta}\big] \\
&= O_P(1) \, n v(d_n) \, ||\widehat{\theta}_n - \theta_0||^{1+\alpha},
\end{aligned}$$

for some $\alpha > 0$. This, together with (A.4) and Lemma A.1, yields that

$$\begin{aligned}
\Big| \sum_{s=k+1}^{n} u_{s-k} \, \Delta_{1,s} \Big| &\leq O_P(\sqrt{n}) \sum_{j=1}^{m} v_j(d_n) \, |\widehat{\theta}_{nj} - \theta_{0j}| + O_P(1) \, n v(d_n) \, ||\widehat{\theta}_n - \theta_0||^{1+\alpha} \\
&= O_P(1) \left[||D_n(\widehat{\theta}_n - \theta_0)|| + n^{(1-\alpha)/2} ||D_n(\widehat{\theta}_n - \theta_0)||^{1+\alpha} \right] \\
&= o_P(\sqrt{n}).
\end{aligned} \tag{A.10}$$

It follows from (A.9) and (A.10) that, for any $k \geq 0$,

$$|R_{2n}| \leq \sum_{s=k+1}^{n} |u_{s-k}| \, |\Delta_{2,s}| + \Big| \sum_{s=k+1}^{n} u_{s-k} \, \Delta_{1,s} \Big| = o_P(\sqrt{n}).$$

Similarly, we have $|R_{1n}| = o_P(\sqrt{n})$.

We next consider R_{3n}. By noting

$$|\Delta_{1,s}| = |g(x_s, \widehat{\theta}_n) - g(x_s, \theta_0)| \leq C\,||\widehat{\theta}_n - \theta_0||\,v(d_n)(1 + |x_s/d_n|^{\beta}),$$

due to Assumption 2.3(i), we have

$$\begin{aligned}\sum_{s=1}^{n} \Delta_{1,s}^2 &\leq C\,v^2(d_n)||\widehat{\theta}_n - \theta_0||^2 \sum_{s=1}^{n}(1 + |x_s/d_n|^{\beta})^2 \\ &\leq C||D_n(\widehat{\theta}_n - \theta_0)||^2 = o_P(\sqrt{n}).\end{aligned}$$

Similarly to (A.9), by using (2.9), we have

$$\sum_{s=1}^{n} \Delta_{2,s}^2 \leq h^2(||\widehat{\eta}_n - \eta_0||) \sum_{s=1}^{n} T^2(x_s) = o_P(\sqrt{n}).$$

It follows from these inequalities that

$$|R_{3n}| \leq \sum_{s=1}^{n} \Delta_s^2 \leq 2\left[\sum_{s=1}^{n} \Delta_{1,s}^2 + \sum_{s=1}^{n} \Delta_{2,s}^2\right] = o_P(\sqrt{n}).$$

We now establish (A.8), and hence completed the proof of Theorem 2.3. $\square$

A.2 Proof of Theorem 3.1

To prove Theorem 3.1, we first introduce two lemmas below. Let A_t and $B_t(\theta), \theta \in \Theta$, be well-defined sequences of random variables on some probability space, where $\Theta \subset R^m$ is a compact set. Let $\dot{B}_t$ and $\ddot{B}_t$ be the first and second derivatives of $B_t(\theta)$, so that $\dot{B}_t = \partial B_t/\partial\theta$ and $\ddot{B}_t = \partial^2 B_t/\partial\theta\partial\theta'$. We assume these quantities exist whenever they are introduced. Set $U_t = A_t - B_t(\theta_0)$, where θ_0 is a finite interior point of Θ, and define $\widehat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta)$, where $Q_n(\theta) = \sum_{t=1}^{n}\left[A_t - B_t(\theta)\right]^2$.

LEMMA A.2. *Suppose that there exists a sequence of random variables $T_j, j \geq 1$, such that*

(i) for each $\theta_1, \theta_2 \in \Theta$,

$$\big|B_j(\theta_1) - B_j(\theta_2)\big| \leq h(||\theta_1 - \theta_2||)\,T_j, \tag{A.11}$$

where $h(x)$ is a bounded real function such that $h(x) \downarrow h(0) = 0$, as $x \downarrow 0$;

(ii) for an increasing sequence $0 < \kappa_n \to \infty$,

(a) $\kappa_n^{-2} \sum_{j=1}^n T_j \left[1 + |U_j| + T_j\right] = O_{\mathrm{P}}(1)$,

(b) $\kappa_n^{-2} \sum_{t=1}^n [B_t(\theta) - B_t(\theta_0)]\, U_t = o_{\mathrm{P}}(1)$ *for each* $\theta \in \Theta$;

(iii) for any $\eta > 0$ *and* $\theta \neq \theta_0$, *where* $\theta, \theta_0 \in \Theta$, *there exist* $n_0 > 0$ *and* $M > 0$ *such that*

$$\mathrm{P}\Big(\kappa_n^{-2} \sum_{t=1}^n \left[B_t(\theta) - B_t(\theta_0)\right]^2 \geq 1/M\Big) \geq 1 - \eta, \tag{A.12}$$

for all $n > n_0$, *where* $0 < \kappa_n \to \infty$ *is given in (ii).*

Then, we have $\widehat{\theta}_n \to_{\mathrm{P}} \theta_0$.

Proof. The proof is similar to Theorem 5.8 of Wang (2015) with minor modifications. We omit the details. □

LEMMA A.3. *Suppose that there exist a sequence of constants* $\{k_n, n \geq 1\}$ *and a sequence of* $m \times m$ *nonrandom nonsingular matrices* $\{D_n, n \geq 1\}$ *satisfying* $k_n \to \infty$ *and* $k_n \parallel D_n^{-1} \parallel \to 0$, *as* $n \to \infty$, *such that the following conditions hold:*

(i) $\sup_{\theta: \|D_n(\theta-\theta_0)\| \leq k_n} \parallel (D_n^{-1})' \sum_{t=1}^n \left[\dot{B}_t(\theta)\dot{B}_t(\theta)' - \dot{B}_t(\theta_0)\dot{B}_t(\theta_0)'\right] D_n^{-1} \parallel = o_P(\delta_n^{-2})$;

(ii) $\sup_{\theta: \|D_n(\theta-\theta_0)\| \leq k_n} \parallel (D_n^{-1})' \sum_{t=1}^n \ddot{B}_t(\theta) \left[B_t(\theta) - B_t(\theta_0)\right] D_n^{-1} \parallel = o_P(\delta_n^{-2})$;

(iii) $\sup_{\theta: \|D_n(\theta-\theta_0)\| \leq k_n} \parallel (D_n^{-1})' \sum_{t=1}^n \ddot{B}_t(\theta)\, U_t\, D_n^{-1} \parallel = o_P(\delta_n^{-2})$;

(iv) $Y_n := (D_n^{-1})' \sum_{t=1}^n \dot{B}_t(\theta_0)\dot{B}_t(\theta_0)'\, D_n^{-1} \to_D M$, *where* $M > 0$ *(a.s.), and*

$$Z_n := (D_n^{-1})' \sum_{t=1}^n \dot{B}_t(\theta_0)\, U_t \; = O_P(\delta_n), \tag{A.13}$$

where $1 \leq \delta_n \leq k_n^{1-\epsilon_0}$ *for some* $\epsilon_0 > 0$. *Then, we have*

$$D_n(\widehat{\theta}_n - \theta_0) = Y_n^{-1}\, Z_n + o_P(1) = O_P(\delta_n). \tag{A.14}$$

Proof. The proof is similar to Theorem 4.1 of Wang and Phillips (2016) with minor modifications. We omit the details. □

Proof of Theorem 3.1. For the first part of (3.2), i.e., $||D_n(\hat{\theta}_n - \theta_0)|| = O_P(1)$, we make use of Lemma A.3 with $\delta_n = 1$, $k_n = \log n$,

$$A_t = y_t, \quad B_t(\theta) = g(x_t, \theta) \quad \text{and} \quad U_t = y_t - g(x_t, \theta_0).$$

By noting $U_t = f(z_t, \eta_0) + u_t$ under model (2.8), to verify conditions (i)-(iv) in Lemma A.3, it suffices to show that

$$I_{1n} := \sup_{\theta: \|D_n(\theta-\theta_0)\| \le \log n} \| \, (D_n^{-1})' \sum_{t=1}^n \ddot{g}(x_t, \theta) \, f(z_t, \eta_0) \, D_n^{-1} \, \| = o_P(1), \tag{A.15}$$

$$I_{2n} := (D_n^{-1})' \sum_{t=1}^n \dot{g}(x_t, \theta_0) \, f(z_t, \eta_0) \; = O_P(1), \tag{A.16}$$

where $D_n = \text{diag}(\sqrt{n} v_{\dot{g}_1}(d_n), ..., \sqrt{n} v_{\dot{g}_m}(d_n))$. In fact, it follows easily from Assumption 3.1(ii) with $p(x, \theta_0) = \dot{g}(x, \theta_0)$ and (iv) that

$$\begin{aligned} ||I_{2n}|| &\le \frac{C}{\sqrt{n}} \sum_{j=1}^n |f(z_t, \eta_0)| \left[1 + (|x_t|/d_n)^\beta\right] \max_{1 \le j \le m} v(d_n)/v_{\dot{g}_j}(d_n) \\ &= O_P(1) \left[1 + (\max_{1 \le k \le n} |x_k|/d_n)^\beta\right] \frac{1}{\sqrt{n}} \sum_{j=1}^n |f(z_t, \eta_0)| = O_P(1), \end{aligned}$$

due to Lemma A.1(iv) and $\max_{1 \le k \le n} |x_k|/d_n = O_P(1)$. This proves (A.15). Similarly, it follows from Assumption 3.1(i)–(ii) with $p(x, \theta_0) = \ddot{g}(x, \theta_0)$ and (iv) that

$$\begin{aligned} I_{1n} &\le C v^{-1}(d_n) \frac{C}{n} \sum_{j=1}^n |f(z_t, \eta_0)| \left[1 + (|x_t|/d_n)^\beta\right] \max_{1 \le i,j \le m} \frac{v(d_n) v_{\ddot{g}_{ij}}(d_n)}{v_{\dot{g}_i}(d_n) v_{\dot{g}_j}(d_n)} \\ &= O_P(n^{-1/2}) = o_P(1), \end{aligned}$$

which yields (A.16). So, the proof for the first part of (3.2) is completed.

We next consider the second part of (3.2), i.e., $\widehat{\eta}_n \to_P \eta_0$, by usng Lemma A.2 with

$$A_t = \widehat{w}_t, \quad B_t(\theta) = f(z_t, \eta) \quad \text{and} \quad U_t = \widehat{w}_t - f(z_t, \eta_0).$$

By noting $U_t = u_t + g(x_t, \widehat{\theta}_n) - g(x_t, \theta_0)$, to verify conditions (i)-(iii) in Lemma A.2, it suffices to show that

$$I_{3n} := n^{-1/2} \sum_{j=1}^n T(z_j) \, |g(x_t, \widehat{\theta}_n) - g(x_t, \theta_0)| = O_{\mathrm{P}}(1); \tag{A.17}$$

$$I_{4n} := n^{-1/2} \sum_{t=1}^n [f(z_t, \eta) - f(z_t, \eta_0)] \left[g(x_t, \widehat{\theta}_n) - g(x_t, \theta_0)\right] = o_{\mathrm{P}}(1), \tag{A.18}$$

for each $\theta \in \Theta$, where $T(x)$ is bounded and integrable. Note that

$$\begin{aligned} |g(x_t, \widehat{\theta}_n) - g(x_t, \theta_0)| &\le C \, ||\widehat{\theta}_n - \theta_0|| \, v_g(d_n) \left[1 + (|x_t|/d_n)^\beta\right] \\ &= O_P(n^{-1/2}) \, ||D_n(\widehat{\theta}_n - \theta_0)||, \end{aligned} \tag{A.19}$$

due to Assumption 3.1(i) with $p(x, \theta_0) = g(x, \theta_0)$ and (iv). The proofs of (A.17) and (A.18) are similar to those of the first part in (3.2). We omit the details.

We finally prove (3.3) by using Lemma A.3 with $\delta_n = \log^{1/2} n$, $k_n = \log n$,

$$A_t = \widehat{w}_t, \quad B_t(\theta) = f(z_t, \eta) \quad \text{and} \quad U_t = \widehat{w}_t - f(z_t, \eta_0).$$

Using the same ideas as in the proof of (3.2), together with Theorem 4.2 of Wang and Phillips (2016), it suffices to show that

$$\begin{aligned} I_{5n} &:= \sup_{\theta:\|D_n(\eta-\eta_0)\|\le \log n} \| (D_n^{-1})' \sum_{t=1}^{n} \ddot{f}(z_t, \eta) \left[g(x_t, \widehat{\theta}_n) - g(x_t, \theta_0)\right] D_n^{-1} \| \\ &= o_P(1), \qquad \text{(A.20)} \\ I_{6n} &:= (D_n^{-1})' \sum_{t=1}^{n} \dot{f}(z_t, \eta_0) \left[g(x_t, \widehat{\theta}_n) - g(x_t, \theta_0)\right] = O_P(1), \qquad \text{(A.21)} \end{aligned}$$

where $D_n = \text{diag}(\sqrt{n} v_{\dot{f}_1}(d_{1n}), ..., \sqrt{n} v_{\dot{f}_m}(d_{1n}))$. By recalling (A.19) and using Assumption 3.2, the proofs of (A.20) and (A.21) are the same as those of (A.15) and (A.16). We omit the details. The proof of Theorem 3.1 is now completed. $\square$

REFERENCES

Box, G.E.P. and Pierce, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, **65**, 1509–1526.

Chan, N. and Wang, Q. (2015). Nonlinear regression with nonstationary time series. *Journal of Econometrics*, **185**, 182–195.

Chang, Y., Park, J.Y. and Phillips, P.C.B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, **4**, 1–36.

Delgado, M.A. and Velasco, C. (2011). An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking. *Journal of the American Statistical Association*, **106**, 946–958.

Escanciano, J.C. and Lobato, I.N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, **151**, 140–149.

Francq, C., Roy, R., and Zakoïan, J.M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, **100**, 532–544.

Gao, J., Maxwell, K., Lu, Z. and Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory*, **25**, 1869–1892.

Gao, J., Maxwell, K., Lu, Z. and Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics*, **37**, 3893–3928.

Gao, J., Tjøstheim, D. and Yin, J. (2012). Model specification between parametric and nonparametric cointegration. Working paper.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, **21**, 1926–1947.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, **64**, 837–864.

Kasparis, I. (2010). The Bierens test for certain nonstationary models. *Journal of Econometrics*, **158**, 221–230.

Kasparis, I. and Phillips, P.C.B. (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics*, **168**, 270–284.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

Park, J.Y. and Phillips, P.C.B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, **15**, 269-298.

Park, J.Y. and Phillips, P.C.B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, **69**, 117-161.

Piaggio, M. and Padilla, E. (2012). CO_2 emissions and economic activity: Heterogeneity across countries and non-stationary series. *Energy Policy*, **46**, 370–381.

Romano, J.P. and Thombs, L.A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, **91**, 590–600.

Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression.* World Scientific.

Wang, Q., Lin, Y.X., and Gulati, C.M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, **19**, 143–164.

Wang, Q. and Phillips, P.C.B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, **25**, 710–738.

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

Wang, Q. and Phillips, P.C.B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, **40**, 727–758.

Wang, Q. and Phillips, P.C.B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory*, **32**, 235-259.

Wang, Q., Wu, D., and Zhu, K. (2016). Model checks for nonlinear cointegrating regression. Working paper.

Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society: Series B*, **78**, 463–485.

Zhu, K. and Li, W.K. (2015). A bootstrapped spectral test for adequacy in weak ARMA models. *Journal of Econometrics*, **187**, 113–130.

Table 1: Size and power ($\times 100$) of $\widehat{U}_n(M)$ for models (4.1)-(4.5)

				$M=6$			$M=12$			$M=18$		
Model	ρ	d	r \ n	100	200	500	100	200	500	100	200	500
(4.1)	0.5	0.0	0.0	5.4	4.7	5.4	6.1	5.1	5.3	6.4	5.4	5.9
			0.5	4.7	4.4	4.9	5.6	5.3	4.8	6.2	5.8	5.5
			0.8	5.2	4.3	4.9	5.4	5.0	4.5	6.8	5.9	4.8
		0.2	0.0	5.1	4.8	4.7	5.4	5.2	4.9	6.1	5.5	5.5
			0.5	5.3	5.5	4.3	5.7	5.5	4.9	6.2	5.8	5.4
			0.8	5.0	4.9	5.2	5.1	5.1	5.5	5.9	5.2	5.6
	-0.5	0.0	0.0	5.1	5.3	5.1	5.8	5.6	5.7	6.3	5.7	5.6
			0.5	5.2	4.8	5.0	6.0	5.3	5.1	6.6	6.1	4.7
			0.8	5.4	5.2	5.2	6.3	5.6	4.9	6.7	6.1	4.9
		0.2	0.0	5.2	5.4	5.2	5.6	5.4	5.2	6.2	5.7	5.0
			0.5	4.9	5.2	4.7	5.5	5.2	4.4	6.1	5.8	4.8
			0.8	5.5	5.3	4.6	5.6	5.3	5.2	5.9	5.7	5.3
(4.2)	0.5	0.0	0.0	12.9	28.5	53.5	13.7	35.4	70.6	13.0	35.0	74.2
			0.5	11.9	28.2	52.9	13.2	34.1	70.5	12.9	32.8	74.7
			0.8	13.2	27.1	54.6	13.3	34.2	72.2	12.6	34.1	74.7
		0.2	0.0	13.8	33.6	63.2	12.9	35.4	70.3	12.0	32.8	68.8
			0.5	14.0	33.7	63.4	12.6	34.6	71.4	12.0	31.8	70.6
			0.8	13.5	33.7	62.8	13.5	34.9	70.3	12.0	31.7	69.0
	-0.5	0.0	0.0	13.0	30.5	66.1	12.6	36.4	78.8	12.0	36.5	80.2
			0.5	12.4	29.9	65.1	12.6	37.0	78.0	12.4	36.0	79.6
			0.8	12.4	30.0	64.8	13.6	36.8	78.2	12.5	36.5	80.1
		0.2	0.0	14.4	37.6	75.3	13.1	38.5	78.0	12.0	35.1	75.2
			0.5	14.6	37.2	74.4	12.9	38.2	77.9	11.7	34.7	75.3
			0.8	15.0	36.3	75.2	14.2	38.0	78.5	12.4	34.7	75.8
(4.3)	0.5	0.0	0.0	17.1	24.1	36.4	12.4	24.1	38.7	8.4	20.9	38.0
			0.5	16.5	24.9	35.1	12.1	23.5	38.0	8.5	20.2	37.9
			0.8	16.0	24.2	36.5	11.8	23.0	38.8	7.9	20.0	38.2
		0.2	0.0	14.1	20.7	27.9	9.9	18.8	28.8	6.2	15.7	27.4
			0.5	13.9	21.3	26.1	9.9	19.5	27.6	6.1	16.2	27.1
			0.8	13.9	19.5	26.9	10.0	18.5	28.1	6.6	15.4	26.7
	-0.5	0.0	0.0	14.0	28.1	55.5	11.4	26.8	55.0	7.6	23.3	53.9
			0.5	14.5	28.5	56.6	11.0	27.7	55.8	8.1	24.5	54.9
			0.8	14.1	27.3	56.6	11.0	26.2	55.3	8.0	23.9	54.1
		0.2	0.0	12.4	22.4	40.2	9.7	21.5	39.0	6.2	18.9	37.9
			0.5	11.5	21.9	41.9	8.9	21.4	40.8	6.1	19.4	39.6
			0.8	11.3	22.5	40.9	8.6	21.4	39.6	5.8	18.9	39.4
(4.4)	0.5	0.0	0.0	16.1	39.1	83.3	14.2	33.7	85.0	13.1	29.5	83.4
			0.5	15.8	36.9	83.4	13.9	31.3	83.8	13.2	27.8	80.9
			0.8	14.7	33.6	80.8	12.9	28.7	81.6	12.1	25.3	79.0
		0.2	0.0	20.1	43.2	88.3	17.0	35.5	84.2	15.8	31.6	77.5
			0.5	19.0	40.8	86.9	16.8	33.8	81.9	15.4	29.7	75.7
			0.8	17.9	37.3	82.2	15.3	30.6	75.9	14.5	27.9	69.9
	-0.5	0.0	0.0	83.9	98.4	100	81.0	97.9	100	78.9	97.2	100
			0.5	85.5	98.5	100	82.4	97.9	100	80.4	97.2	100
			0.8	86.5	98.6	100	83.9	98.3	100	80.9	97.7	100
		0.2	0.0	94.2	99.9	100	92.7	99.7	100	91.6	99.6	100
			0.5	94.6	99.8	100	93.0	99.7	100	92.1	99.7	100
			0.8	94.1	99.8	100	92.5	99.6	100	91.3	99.6	100
(4.5)	0.5	0.0	0.0	98.8	100	100	98.3	99.9	100	97.6	99.9	100
			0.5	98.6	99.9	100	98.0	99.9	100	97.1	99.9	100
			0.8	98.8	100	100	98.2	100	100	97.5	100	100
		0.2	0.0	99.9	100	100	99.8	100	100	99.5	100	100
			0.5	99.8	100	100	99.7	100	100	99.4	100	100
			0.8	99.8	100	100	99.5	100	100	99.3	100	100
	-0.5	0.0	0.0	97.7	99.9	100	96.5	99.9	100	95.1	99.8	100
			0.5	98.0	100	100	96.6	99.8	100	95.2	99.8	100
			0.8	97.1	100	100	96.0	100	100	95.0	99.9	100
		0.2	0.0	99.7	100	100	99.4	100	100	98.8	100	100
			0.5	99.7	100	100	99.4	100	100	99.0	100	100
			0.8	99.8	100	100	99.4	100	100	99.1	100	100

Table 2: Size and power ($\times$100) of $\widehat{U}_n(M)$ for models (4.6)-(4.10)

Model	ρ	d	r \ n	$M=6$ 100	$M=6$ 200	$M=6$ 500	$M=12$ 100	$M=12$ 200	$M=12$ 500	$M=18$ 100	$M=18$ 200	$M=18$ 500
(4.6)	0.5	0.0	0.0	5.7	5.2	4.5	6.4	6.3	5.2	7.0	6.5	5.4
			0.5	5.1	5.2	4.8	6.1	4.9	5.5	6.8	5.6	5.8
			0.8	5.7	5.9	5.0	6.3	6.2	5.3	6.9	6.4	5.6
		0.2	0.0	5.1	4.9	4.9	6.1	5.5	5.3	6.4	6.1	5.4
			0.5	5.7	5.1	5.4	7.0	5.3	5.2	7.3	5.5	5.3
			0.8	5.8	4.9	5.6	6.3	6.1	5.3	6.7	6.5	5.5
	-0.5	0.0	0.0	5.5	5.6	5.1	6.4	5.9	4.9	6.8	6.2	5.5
			0.5	5.3	5.6	5.1	6.0	5.7	5.2	7.1	5.6	5.1
			0.8	5.1	5.4	5.0	5.7	5.8	5.0	5.9	6.0	5.4
		0.2	0.0	6.0	5.2	5.6	7.0	5.8	5.3	7.4	6.4	5.7
			0.5	5.9	5.3	4.8	6.2	5.8	5.1	7.1	6.6	5.1
			0.8	5.6	5.9	5.3	5.9	6.1	5.2	6.1	6.6	5.0
(4.7)	0.5	0.0	0.0	14.7	27.8	50.3	14.5	32.7	64.6	14.3	31.5	67.3
			0.5	13.6	27.6	49.4	13.7	31.6	65.0	13.0	31.3	67.9
			0.8	14.1	26.8	49.5	14.3	31.7	64.3	13.7	31.3	67.0
		0.2	0.0	13.8	32.2	61.6	13.4	34.1	68.7	12.5	31.8	68.1
			0.5	14.7	32.1	62.5	14.7	33.2	70.4	13.0	31.2	68.9
			0.8	14.7	31.6	61.1	13.9	32.4	68.8	12.6	30.1	67.1
	-0.5	0.0	0.0	13.1	26.6	59.7	13.2	30.9	70.4	13.2	30.5	72.5
			0.5	12.7	26.3	59.0	13.1	30.8	70.2	12.3	30.8	71.5
			0.8	12.5	26.9	59.2	13.2	31.1	69.6	12.8	30.3	71.5
		0.2	0.0	13.5	33.5	74.5	13.2	33.3	77.3	12.5	30.6	74.4
			0.5	14.0	35.2	75.6	13.2	36.0	77.0	11.7	32.5	74.7
			0.8	14.4	33.7	73.8	14.0	34.2	76.4	12.1	31.0	73.7
(4.8)	0.5	0.0	0.0	17.2	23.9	35.9	13.3	23.2	38.9	9.3	20.3	37.8
			0.5	15.8	24.0	34.5	12.6	23.0	37.5	8.5	20.5	36.7
			0.8	16.6	23.3	34.0	13.2	22.6	36.4	8.6	19.7	36.7
		0.2	0.0	14.7	20.8	26.7	10.8	18.8	28.1	7.1	16.6	27.3
			0.5	13.8	20.3	25.7	9.5	19.0	26.2	5.7	17.3	25.9
			0.8	14.0	19.6	26.1	10.1	18.4	27.6	6.6	15.7	26.4
	-0.5	0.0	0.0	19.7	30.7	59.8	16.5	29.7	58.4	12.4	27.4	56.2
			0.5	20.0	30.9	60.9	16.9	29.1	58.5	12.8	25.7	57.6
			0.8	19.9	30.9	59.8	16.4	29.3	57.9	13.2	26.9	56.9
		0.2	0.0	19.3	25.8	44.0	16.5	25.0	42.1	13.1	22.8	41.7
			0.5	20.1	27.0	44.8	16.6	25.1	43.2	13.3	22.7	41.5
			0.8	20.5	26.1	45.2	17.7	26.2	43.7	13.9	23.9	41.9
(4.9)	0.5	0.0	0.0	5.5	12.1	48.5	6.3	11.3	50.2	6.6	10.4	47.7
			0.5	5.7	11.0	47.8	6.8	10.5	49.7	7.5	10.0	47.6
			0.8	6.1	12.0	46.8	6.6	11.0	49.1	7.3	10.6	47.1
		0.2	0.0	5.6	14.7	52.5	6.0	11.6	49.3	6.3	10.3	43.5
			0.5	6.4	13.8	51.2	6.5	10.9	47.4	7.2	9.9	42.7
			0.8	6.6	13.4	51.4	6.4	11.6	48.6	7.3	10.8	43.8
	-0.5	0.0	0.0	42.5	61.6	74.7	40.6	60.2	74.5	40.0	60.0	74.1
			0.5	42.2	60.5	76.0	40.1	59.9	75.8	39.4	59.5	75.3
			0.8	41.8	61.3	75.4	39.8	60.2	74.7	39.3	59.4	74.7
		0.2	0.0	45.3	61.2	74.4	44.7	60.2	74.1	44.1	60.1	74.0
			0.5	45.6	60.6	75.1	44.7	60.7	74.6	44.2	60.4	74.7
			0.8	45.9	60.9	74.7	44.7	60.2	74.2	43.6	59.6	73.9
(4.10)	0.5	0.0	0.0	86.2	92.8	97.6	82.3	89.5	96.1	80.0	87.7	94.8
			0.5	86.5	92.4	97.0	82.6	89.3	95.1	80.5	87.7	93.9
			0.8	86.1	93.8	97.7	82.6	91.0	95.7	80.4	89.3	94.3
		0.2	0.0	83.1	87.3	92.9	78.1	83.0	89.0	75.4	80.3	86.7
			0.5	82.9	87.4	93.1	77.8	83.0	90.2	75.0	80.4	88.1
			0.8	82.0	88.0	92.6	76.8	83.2	89.4	74.3	80.4	87.2
	-0.5	0.0	0.0	85.4	92.5	97.5	81.0	89.3	95.7	78.2	87.4	94.5
			0.5	85.1	92.4	97.5	81.0	89.3	96.2	78.6	87.5	95.1
			0.8	84.1	92.7	97.3	80.3	89.5	95.6	78.2	87.7	94.4
		0.2	0.0	83.0	87.5	92.3	77.5	82.6	89.2	75.1	79.7	87.0
			0.5	81.6	87.7	92.8	76.5	83.2	89.8	73.8	80.3	88.1
			0.8	82.2	87.2	92.7	76.3	82.4	89.9	73.1	79.9	87.8

Table 3: Size and power ($\times 100$) of $\widehat{U}_n(M)$ for models (4.11)-(4.15)

				$M=6$			$M=12$			$M=18$		
Model	ρ	d	r \ n	100	200	500	100	200	500	100	200	500
(4.11)	0.5	0.0	0.0	6.0	5.1	4.9	6.1	5.6	5.6	7.0	6.2	5.2
			0.5	5.4	5.0	5.4	6.4	5.6	5.4	6.8	5.8	5.8
			0.8	5.8	5.9	5.2	6.1	5.7	5.7	6.6	6.2	5.8
		0.2	0.0	6.2	5.2	5.2	6.2	5.7	5.6	7.0	6.6	5.7
			0.5	5.8	5.5	5.7	6.5	5.4	5.6	6.9	6.2	5.7
			0.8	5.5	5.2	5.1	6.4	5.3	5.5	7.0	5.4	5.9
	-0.5	0.0	0.0	5.6	5.0	4.7	6.5	5.1	5.2	7.2	5.6	5.8
			0.5	5.5	4.8	4.8	5.8	5.0	5.3	6.5	5.1	5.6
			0.8	5.0	4.3	5.3	5.5	4.6	4.8	6.5	5.3	5.2
		0.2	0.0	5.4	4.7	5.0	5.9	5.4	5.7	6.8	6.1	5.2
			0.5	5.4	5.3	5.3	5.9	5.7	5.1	6.3	5.9	5.8
			0.8	5.4	5.4	5.0	6.2	5.6	5.3	6.9	6.1	5.6
(4.12)	0.5	0.0	0.0	13.2	26.6	51.0	13.3	30.4	65.6	12.6	29.9	67.9
			0.5	13.5	26.0	50.8	13.5	31.0	63.7	13.2	30.6	66.1
			0.8	13.4	26.7	51.4	13.1	32.4	64.0	12.4	31.8	66.7
		0.2	0.0	13.6	31.2	59.4	12.9	33.9	66.0	12.0	31.3	64.7
			0.5	14.2	32.0	59.9	13.6	33.0	66.8	12.6	30.4	65.4
			0.8	13.7	31.0	60.0	13.6	32.6	66.8	12.6	30.4	65.2
	-0.5	0.0	0.0	12.4	26.7	58.9	12.8	62.1	68.9	11.6	30.4	70.8
			0.5	12.5	28.4	60.2	12.3	32.6	69.9	11.7	30.8	71.3
			0.8	13.7	28.2	60.1	13.6	31.8	70.0	12.1	30.3	71.1
		0.2	0.0	13.5	33.6	71.9	12.6	33.7	74.5	11.5	30.9	73.0
			0.5	12.7	33.3	72.0	12.3	34.0	73.7	10.8	31.7	70.8
			0.8	12.6	33.3	72.2	12.1	34.0	74.6	10.8	31.2	72.9
(4.13)	0.5	0.0	0.0	16.9	25.5	35.1	12.9	23.9	37.9	8.8	21.1	36.7
			0.5	16.3	24.8	35.7	12.4	23.3	38.0	8.6	20.7	37.0
			0.8	17.1	24.3	35.6	13.1	23.5	37.9	9.1	21.0	37.1
		0.2	0.0	14.0	20.3	26.5	10.2	18.0	28.7	6.9	15.6	27.2
			0.5	14.5	19.7	26.6	10.4	18.2	27.7	6.9	15.7	26.2
			0.8	14.3	20.9	26.4	10.0	18.6	28.0	7.0	16.2	27.0
	-0.5	0.0	0.0	21.0	31.3	57.3	18.2	30.0	56.6	14.4	28.0	54.8
			0.5	20.4	31.7	57.9	16.5	30.9	56.7	13.7	28.4	55.1
			0.8	20.9	31.9	58.8	17.9	30.8	57.9	14.8	28.5	56.4
		0.2	0.0	21.4	26.6	43.0	18.5	26.6	42.6	14.8	24.2	41.4
			0.5	21.3	26.0	42.1	19.0	26.0	40.8	15.9	23.1	40.1
			0.8	21.4	26.5	43.5	17.7	25.4	42.5	14.2	23.8	41.8
(4.14)	0.5	0.0	0.0	7.0	22.4	76.7	6.7	20.7	77.9	6.7	18.1	74.6
			0.5	7.6	22.0	77.2	8.0	20.0	77.5	8.1	18.3	74.5
			0.8	7.6	22.2	75.4	7.8	19.1	76.5	8.1	16.9	73.5
		0.2	0.0	8.4	23.9	71.5	7.7	19.2	64.7	7.5	16.4	57.7
			0.5	8.6	24.1	71.4	8.3	18.5	64.9	8.2	15.4	58.3
			0.8	9.5	25.8	69.8	9.0	20.6	64.7	8.5	17.0	57.3
	-0.5	0.0	0.0	76.5	88.9	95.2	74.3	88.5	94.9	73.7	88.1	94.8
			0.5	77.5	88.9	95.1	75.8	88.5	95.0	74.9	88.2	94.9
			0.8	76.3	88.8	94.7	74.5	88.7	94.7	73.2	88.3	94.6
		0.2	0.0	80.7	89.0	94.8	79.9	88.6	94.7	79.4	88.4	94.5
			0.5	80.4	89.4	94.8	78.3	89.1	94.7	77.7	88.8	94.6
			0.8	79.7	88.2	95.2	78.6	88.0	95.1	77.8	87.5	95.0
(4.15)	0.5	0.0	0.0	91.9	96.7	99.1	88.5	94.9	98.3	86.5	93.3	97.8
			0.5	91.9	96.5	98.9	88.1	94.4	98.1	86.0	92.9	97.6
			0.8	92.9	96.2	99.0	89.3	94.4	98.4	87.0	92.8	97.8
		0.2	0.0	88.0	93.6	96.5	83.4	90.3	94.9	81.4	88.0	93.7
			0.5	87.8	92.2	96.7	83.6	89.1	94.6	81.2	86.8	92.9
			0.8	88.4	92.4	96.7	83.7	88.9	95.0	81.4	86.6	93.5
	-0.5	0.0	0.0	91.1	97.0	99.1	87.2	94.9	98.1	84.9	93.3	97.4
			0.5	91.5	96.8	99.1	87.1	94.7	98.2	84.9	93.1	97.5
			0.8	91.3	96.4	99.1	87.2	94.6	98.3	84.6	93.4	97.7
		0.2	0.0	88.6	92.7	96.6	84.3	89.3	94.7	81.8	87.0	93.4
			0.5	88.7	92.6	97.0	84.3	89.1	95.2	81.7	86.6	93.5
			0.8	88.3	93.2	96.4	83.4	89.5	94.0	80.6	87.3	92.5

Table 4: The choice of p and the p-value of $\widehat{U}_n(M)$ or $\widetilde{U}_n(M)$ for 16 different countries when $u_t \sim \mathrm{AR}(p)$

	Countries							
	AUS	AUT	BEL	CAN	CHN	DEN	FIN	FRA
p	1	1	1	1	2	1	1	1
$M = 6$	0.8470	0.1001	0.7905	0.9973	0.9473	1.0000	1.0000	0.9529
$M = 12$	0.9690	0.0713	0.9399	1.0000	0.9990	1.0000	1.0000	0.9975
$M = 18$	0.9929	0.0485	0.9805	1.0000	1.0000	1.0000	1.0000	0.9999

	Countries							
	HOL	IND	IRE	ITA	JAP	NOR	SWI	USA
p	1	1	1	1	2	1	1	1
$M = 6$	0.7338	0.9997	0.9492	0.9645	0.0000	0.5927	0.6042	0.2175
$M = 12$	0.9010	1.0000	0.9971	0.9987	0.0000	0.7650	0.7781	0.2284
$M = 18$	0.9587	1.0000	0.9998	0.9999	0.0000	0.8506	0.8626	0.2207

[1] The value of p is selected by BIC.
[2] When $p = 1$, the reported p-values are for $\widehat{U}_n(M)$, and when $p = 2$, the reported p-value are for $\widetilde{U}_n(M)$.

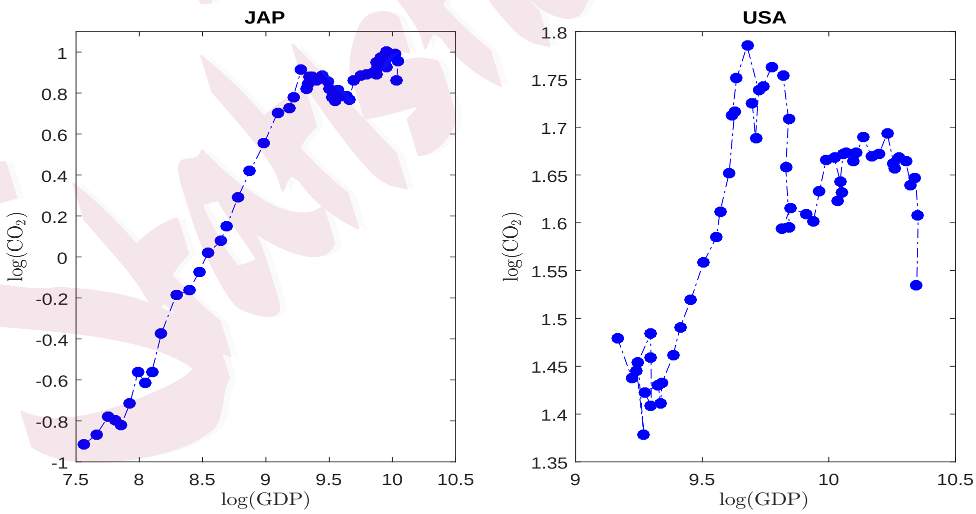

Figure 1: Plots of $\log(CO_2)$ against $\log(\mathrm{GDP})$ for JAP and USA