Econometric Theory, **31**, 2015, 753–777.
doi:10.1017/S0266466614000541

ESTIMATION AND INFERENCE FOR VARYING-COEFFICIENT MODELS WITH NONSTATIONARY REGRESSORS USING PENALIZED SPLINES

HAIQIANG CHEN, YING FANG, AND YINGXING LI
Xiamen University

This paper considers estimation and inference for varying-coefficient models with nonstationary regressors. We propose a nonparametric estimation method using penalized splines, which achieves the same optimal convergence rate as kernel-based methods, but enjoys computation advantages. Utilizing the mixed model representation of penalized splines, we develop a likelihood ratio test statistic for checking the stability of the regression coefficients. We derive both the exact and the asymptotic null distributions of this test statistic. We also demonstrate its optimality by examining its local power performance. These theoretical findings are well supported by simulation studies.

1. INTRODUCTION

Regression models with nonstationary regressors have received much attention in the literature of theoretical and applied econometrics since the seminal work by Nelson and Plosser (1982). Amongst the popular research in this field has been the study of cointegration. The traditional framework of Engle and Granger (1987), assuming constant cointegrating coefficients, provides an appealing analytical framework to characterize the long-run equilibrium relationship. However, very few empirical studies support the presence of cointegration with constant coefficients. Such an empirical frustration is due to the lack of flexibility of traditional cointegration models in accommodating the time-varying long-run equilibrium relationship.

The authors are grateful to Peter Phillips, Wolfgang Härdle, and anonymous referees for their helpful comments. We also thank Zongwu Cai, Jiti Gao, Yongmiao Hong, and all participants of the 3rd WISE-Humboldt Workshop in Nonparametric Nonstationary High-dimensional Econometrics in May 2012. We acknowledge the financial support from the National Science Foundation of China with grant numbers 71201137, 71271179, 71131008, and 11201390, and from the Natural Science Foundation of Fujian Province with grant number 2013J01024, as well as the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk". Address correspondence to Ying Fang, Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China; e-mail: yifst1@gmail.com; or to: Yingxing Li, Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China; e-mail: yxli@xmu.edu.cn.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

There are many empirical examples in economics and finance demonstrating time-varying features in cointegrating relationships. For example, Goldfajn and Baig (1998) argue that, during the Asian currency crisis, the cointegrating relation between the spot exchange rate and the interest rate differential is not fixed but depends on the level of the interest rate. Another example is in the literature of stock return predictability, where one of the theoretical and practical issues is to answer whether we could predict the asset return from fundamental variables such as the dividend-price ratio and the earning-price ratio, which are well known nonstationary time series variables (Campbell and Yogo, 2006). Although linear prediction models have been extensively used, Lettau and Ludvigsson (2001), Goyal and Welch (2003), and Paye and Timmermann (2006) find empirical evidence that the cointegrating stock return forecasting models might be unstable.

Many studies adopt nonparametric methods to capture the time-varying relation with nonstationary variables. The latest research include Wang and Phillips (2009a, 2009b) and Cai, Li, and Park (2009), among others. Wang and Phillips (2009b) construct asymptotic theories for the local time density estimation in nonparametric cointegrating regression. They find that the so called ill-posed inverse problem does not exist in nonparametric regression with nonstationary endogenous regressors. Cai et al. (2009) investigate the asymptotic property of local linear estimators in a varying-coefficient model when the smoothing variable is nonstationary but the covariates are either stationary or nonstationary.

Testing the stability of varying coefficients becomes another important issue in this literature. For example, Park and Hahn (1999) construct two residual-based statistics to test the constancy of the cointegrating coefficients based on the series estimation. Kasparis (2008) develops two residual-based statistics for testing the functional form misspecification in cointegrating relations. Bierens and Martins (2010) propose a vector error correction model with cointegration coefficients estimated by Chebyshev polynomials and conduct a likelihood ratio test on the stability by examining whether all Chebyshev polynomial coefficients are jointly zero. Juhl (2005) and Xiao (2009) further extend the studies to the case where cointegration coefficients are general smooth stochastic functions depending on some stationary covariates. Xiao (2009) considers both kernel and local polynomial estimators and establish their asymptotics. Moreover, he proposes a test statistic by comparing the functional-coefficient estimates to a fixed value estimated under the null.[1]

In this paper, we consider varying-coefficient regression models with nonstationary regressors. Our model setting is similar to Cai et al. (2009) and Xiao (2009). However, we propose to estimate the varying coefficients by penalized splines and construct a likelihood ratio test (LRT) for the stability of the varying coefficients. The basic idea of spline methods is to approximate the unknown regression function by splines, which are piecewise polynomials, and then estimate the spline coefficients by the least squares method. In order to maintain a good balance between the goodness of fit and the model variability, a large number of

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

basis functions are employed and a penalty term is imposed to avoid overfitting (Eilers and Marx, 1996). There are several prominent features of such a penalized splines approach. First, this method is simple and easy to implement. Its computation is usually less time-consuming compared to other nonparametric methods such as kernel or local polynomials. Second, it could easily incorporate correlation structure to improve the efficiency of estimator or to account for longitudinal and spatial effects (Ruppert, Wand, and Carroll, 2003). Third, it has close connections with Ridge regression, Bayesian methods, and mixed model representation, thus allowing fitting and testing to be conducted through the paradigm of likelihood (Crainiceanu, Ruppert, Claeskens, and Wand, 2005). However, theoretical explorations of penalized splines were less well developed until recently. Li and Ruppert (2008) establish the asymptotic normality of the penalized splines estimation. Claeskens, Krivobokova, and Opsomer (2009) systematically compare the asymptotics of penalized splines, regression splines, and smoothing splines. All these studies are under the univariate nonparametric model assumption $y_t = \theta(z_t) + v_t$ for stationary covariate z_t's.

Our studies contribute to the literature through the following aspects. First, we propose the penalized spline estimation method for varying-coefficient models with nonstationary regressors. We establish the consistency as well as the optimal convergence rate of the penalized splines estimators. In our study, the choice of the spline basis is not crucial, but the penalty parameter plays the key role of smoothing. To the best of our knowledge, this is the first work in establishing the asymptotics of penalized splines estimators for varying-coefficient models with nonstationary regressors. Second, we consider testing the stability of the regression coefficients. By utilizing the mixed model representation of penalized splines, we relate this problem to testing zero variance component of the spline coefficients. We then adopt the likelihood ratio test (LRT) procedure and derive the exact and the asymptotic null distribution. Since the exact null distribution is nonstandard, we provide a fast algorithm to simulate its critical values when the sample size is small. By assuming both the sample size and the number of spline functions grow to infinity, we, for the first time, show that the limiting null distribution of the proposed LRT statistic follows a simple χ^2 distribution rather than a mixture of χ^2 distributions. We also study the local power of the proposed LRT by deriving the asymptotic distribution under the local alternative. Simulations show that our method works very well.

The rest of the paper is organized as follows. In Section 2, we introduce the varying-coefficient regression model with nonstationary regressors and discuss some regularity assumptions. The penalized splines estimation of the varying coefficients and its asymptotics are presented in Section 3. In Section 4, we construct the LRT for the stability of the varying coefficients. Both the exact distribution and the asymptotic null distribution are derived. The local power property is examined as well. Simulations are conducted in Section 5 to demonstrate the finite sample performance of our procedure, while Section 6 concludes.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

In matters of notations, $\stackrel{D}{=}$ denotes equality in distribution, $=:$ denotes definition, $\Rightarrow$ denotes convergence in distribution, $\stackrel{a.s.}{\to}$ denotes almost sure convergence, $a \sim b$ denotes that a and b have the same order, $[Ts]$ denotes the integer part of Ts, and $\int$ denotes the integration from 0 to 1.

2. THE MODEL AND ASSUMPTIONS

Consider the following varying-coefficient regression model without intercept and time trends

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

where $\theta(\cdot)$ is a smooth function of unknown form, y_t can be either stationary or nonstationary, z_t and u_t are stationary, and x_t is an integrated process of order one, whose generating mechanism is given by

$$x_t = x_{t-1} + v_t, \qquad \text{for } t = 1, 2, \ldots T,$$

with v_t being stationary. We set $x_0 = 0$ to avoid some unnecessary complications in exposition, although $x_0 = o_{a.s.}(\sqrt{T})$ is sufficient for the asymptotic results. To save notations, we only consider the simple case when z_t is univariate.

Compared to traditional varying-coefficient models, which usually deal with independent and identically distributed (i.i.d.) or stationary time series, model (2.1) allows the regressors to be highly persistent variables, such as interest rate, GDP growth rate, and unemployment rates. On the other hand, compared to traditional linear cointegration models, which are widely used in the literature to capture the long term equilibrium among highly persistent economic variables, model (2.1) affords more flexibility as it allows the relationship to be varying according to some state variable z_t.

Before describing our estimating and testing procedures, we first discuss some regularity assumptions for our model (2.1).

Assumption 1. The sequence $\{v_t\}$ is stationary α-*mixing* with $\mathbb{E}(|v_t|^\rho) < \infty$ for some $\rho > 2 + r_1$ with $0 < r_1 \leq 2$ and the mixing coefficients α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{1/(2+r_1)-1/\rho} < \infty$.

Assumption 2.

(i) The error term u_t is a general linear process satisfying

$$u_t = \sum_{i=0}^{\infty} c_i e_{t-i} = \mathbb{C}_u(L)e_t,$$

where $\{e_t\}_{t=-\infty}^{\infty}$ are i.i.d $N\left(0, \sigma_e^2\right)$ with $\sigma_e^2 > 0$ and $\{c_i\}_{i=0}^{\infty}$ satisfies the summability conditions $\sum_{i=0}^{\infty} i|c_i| < \infty$ with $\mathbb{C}_u(1) \neq 0$.

(ii) u_t is independent of z_t and v_t.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Assumption 3.

(i) The sequence $\{z_t\}$ is strictly stationary, ergodic, and α-$mixing$ with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{1/2-1/r_2} < \infty$ for some $r_2 > 2$.

(ii) z_t has a marginal density $f_z(z)$ on a finite support $[0, 1]$. $f_z(z)$ is continuously differentiable and bounded away from 0.

(iii) $\theta(z)$ belongs to the Sobolev space of the m-th order $W^m[0, 1]$, i.e., $\theta(z)$ is $(m-1)$-th continuously differentiable and $\int\{\theta^{(m)}(z)\}^2 dz < \infty$.

Assumptions 1 and 2(i) provide sufficient conditions of strong approximations for the partial sum $(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} u_t, \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} v_t)$,

$$\sup_{s\in[0,1]} \left\| \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} u_t, \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} v_t \right) - \{B_u(s), B_v(s)\} \right\| \overset{a.s.}{\to} 0, \text{ as } T \to \infty,$$

where $\{B_u(s), B_v(s)\}$ are two Brownian motions defined on $D[0, 1]$, the space of cadlag functions defined in the unit interval $[0, 1]$. Note that the strong approximation is stronger than the multivariate invariance principle, but it is commonly used in the literature of nonlinear regression model with nonstationary regressors, including Park and Hahn (1999), Park and Phillips (2001), Kasparis (2008), Wang and Phillips (2009a, 2009b), Cai et al. (2009), and Shi and Phillips (2012), among others. Sufficient conditions to derive strong approximations for dependent random variables are also well established in the literature. For example, Lemma 1 in Park and Hahn (1999) establishes conditions of strong approximations for a general linear process u_t and Theorem 4.1 in Shao and Lu (1987) gives conditions of strong approximations for an α-$mixing$ process v_t.

Assumption 2(i) also defines u_t as an invertible Gaussian moving average process. The normality assumption is somewhat restrictive but it is for the purpose of employing likelihood principles. Define the vector $u = (u_1, \ldots, u_T)^T$ and denote $var(u) = \sigma^2 \sum_u$. At the current point, we assume that $\sum_u$ is known so that the full likelihood function could be constructed. In practice, $\sum_u$ can be posited to be of a particular form $\sum_u(\phi)$, where ϕ is a vector of parameters that could be estimated from the data. A simple example is to treat u_t as an AR(1) process. Then $\sum_u$ is a function of the first order coefficient ρ. One could apply a two-step procedure to obtain the estimate $\hat{\rho}$ and replace Σ_u by its estimate $\Sigma_u(\hat{\rho})$.

To simplify the derivation of the LRT, Assumption 2(ii) assumes the independence condition between the error term u_t and x_t, although this might be further relaxed. Following Saikkonen (1991) and Saikkonen and Choi (2004), we might remove the endogeneity between u_t and x_t by adding leads and lags of the term $\{v_t\}_{t=1}^{\infty}$ in the regression. On the other hand, we assume the independence between u_t and z_t, which rules out the ill-posed inverse problems in the nonparametric estimation with stationary smoothing variables.

Assumption 3(i) guarantees that z_t is strictly stationary and imposes some conditions on its dependency and moments. In this paper, we do not consider the case

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

when z_t is nonstationary. Assumption 3(ii) requires that z_t has a bounded support. In practice, one can always conduct some transformations, such as the probability integral transform, to satisfy this assumption. We also assume that the marginal density of z is continuously differentiable and bounded away from 0, thus ensuring there are enough observations for estimation. Finally, Assumption 3(iii) imposes some smoothness conditions on the unknown function $\theta(z)$, which is a standard assumption in nonparametric regression analysis.

3. PENALIZED SPLINES ESTIMATION

We employ the penalized splines approach to estimate the varying-coefficient regression model with nonstationary regressors. First, we approximate the unknown varying coefficient by splines basis. A popular choice is the uniform B-splines family defined by a set of equally spaced knots $\kappa_k = k/K$, for $k = 0, \ldots, K$. The simplest case is the zero degree B-splines, which are indicator functions between κ_{k-1} and κ_k. In general, we could use the iterative algorithm proposed by de Boor (1978) to calculate the p-th degree B-splines and express $\theta(z)$ as

$$\theta(z) = \sum_{k=1}^{K+p} \Psi_k^{[p]}(z)\beta_k + O(K^{-1}).$$

Following the idea of penalized least squares, we could estimate the spline coefficients β by minimizing the following criterion:

$$\sum_{t=1}^{T} \left(y_t - x_t \sum_{k=1}^{K+p} \beta_k \Psi_k^{[p]}(z) \right)^2 + \tilde{\lambda}^{-1} \int \left\{ \theta^{(m)}(z) \right\}^2 dz. \tag{3.1}$$

In a discrete version, this could be written as

$$\sum_{t=1}^{T} \left(y_t - x_t \sum_{k=1}^{K+p} \beta_k \Psi_k^{[p]}(z) \right)^2 + \tilde{\lambda}^{-1} K^{-1} \sum_{k=m+1}^{K} \left(\frac{\Delta^m \beta_k}{K^{-m}} \right)^2, \tag{3.2}$$

where Δ is the differencing operator such that $\Delta\beta_k = \beta_k - \beta_{k-1}$, and m is a positive integer indicating the order of differencing with $\Delta^m = \Delta(\Delta^{m-1})$. Let $Y = (y_1, \ldots, y_T)^T$, X be the diagonal matrix whose (i,i)th element is x_i, and Ψ be the matrix[2] whose (i,j)-th element is $\Psi_i^{[p]}(z_j)$. Define D_m as the differencing matrix such that the j-th element of $D_m\beta$ is $\Delta^m\beta_j$. Then the above minimization criterion could be written in a matrix form as

$$(Y - X\Psi\beta)^T (Y - X\Psi\beta) + \tilde{\lambda}^{-1} K^{2m-1} \beta^T D_m^T D_m \beta.$$

In general, if we take into account the correlation among u_t and the fact that $var(u) = \sigma^2 \Sigma_u$, we could incorporate the weighted penalized splines approach

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

and estimate β by minimizing

$$(Y - X\Psi\beta)^T \Sigma_u^{-1}(Y - X\Psi\beta) + \tilde{\lambda}^{-1} K^{2m-1} \beta^T D_m^T D_m \beta. \tag{3.3}$$

A direct calculation shows that the solution to (3.3) is

$$\hat{\beta} = (\Psi^T X \Sigma_u^{-1} X \Psi + \tilde{\lambda}^{-1} K^{2m-1} D_m^T D_m)^{-1} \Psi^T X \Sigma_u^{-1} Y. \tag{3.4}$$

Then the penalized spline estimator of $\theta(z)$ for model (2.1) is defined as

$$\hat{\theta}(z) = \sum_{k=1}^{K+p} \Psi_k^{[p]}(x) \hat{\beta}_k.$$

The methodology and applications of penalized splines are discussed extensively in Ruppert et al. (2003), but their theoretical studies had been largely absent until recently. For the univariate nonparametric model, Hall and Opsomer (2005) establish the consistency of the penalized splines estimators. Li and Ruppert (2008) derive the asymptotic normality and they were the first to obtain explicit formula for the asymptotic bias and variance. Claeskens et al. (2009) study the convergence rate of the penalized spline estimation and discuss the impact of the number of knots. However, all of these results are not directly applicable for varying-coefficient models with nonstationary regressors.

The following theorem establishes the consistency of the penalized spline estimator. Please note that all proofs of the theorems are relegated to the Appendix.

THEOREM 3.1. *Suppose that Assumptions 1–3 hold. In addition, assume that*

(i) *The pth degree uniform B-splines are used to model $\theta(z)$. The number of knots satisfies $K \sim T^{r_3}$ with $\frac{2m}{2m+1} < r_3 < 1$.*

(ii) *The mth order penalty is used and the penalty parameter $\tilde{\lambda}$ satisfies that $\tilde{\lambda} \to 0$ and $T^2\tilde{\lambda} \to \infty$.*

Then for $z \in (0, 1)$, the penalized spline estimator $\hat{\theta}(z)$ satisfies

$$\hat{\theta}(z) - \theta(z) = O_p\big(T^{-1}\tilde{\lambda}^{-1/2}\big) + O_p\big(T^{-1+1/(2m)}\tilde{\lambda}^{1/(4m)}\big) + O_p(K^{-1}). \tag{3.5}$$

Remark 3.1. Theorem 3.1 establishes the consistency of the penalized splines estimator for varying-coefficient models with nonstationary regressors. The term $O_p(T^{-1}\tilde{\lambda}^{-1/2})$ reflects the order of asymptotic bias due to smoothing. The term $O_p(K^{-1})$ is the design bias due to the use of splines in approximating the smooth functions $\theta(z)$. These results are standard. However, the order of the asymptotic variance becomes $O_p(T^{-2+1/m}\tilde{\lambda}^{1/(2m)})$ when x_t is an integrated process of order one, compared to a slower rate $O_p(T^{-1+1/2m}\tilde{\lambda}^{1/(2m)})$ for stationary x_t. Correspondingly, when $\tilde{\lambda}$ satisfies $\tilde{\lambda} \sim T^{-2/(2m+1)}$, our estimator achieves the optimal convergence rate $T^{\frac{2m}{2m+1}}$, faster than the $T^{m/(2m+1)}$ convergence rate with $\tilde{\lambda} \sim T^{-1/(2m+1)}$ for stationary x_t. In particular, when $\theta(z) \in W^2[0, 1]$, i.e., $m = 2$, the optimal convergence rate of $\hat{\theta}(z)$ is $T^{4/5}$ when x_t is integrated with order one. Such a result is consistent with Cai et al. (2009) and Xiao (2009).

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Remark 3.2. Penalized splines allow one to flexibly select the degree of splines p, the number of knots K, and the amount of penalty $\tilde{\lambda}^{-1}$. Our results have the following implications. First, the degree of splines p has no impact in the convergence rate of the estimator. Second, the number of knots K is not crucial as long as it exceeds a certain minimum bound. Third, the penalty parameter $\tilde{\lambda}$ could serve as the key smoother and it determines the convergence rate of the estimator. These three conclusions are consistent with the results obtained in Li and Ruppert (2008). However, when x_t is stationary, the term $(\tilde{\lambda}^{-1}/T)^{1/(2m)}$ serves as the equivalent bandwidth used in a Nadaraya–Watson kernel estimator. In contrast, when x_t is integrated with order 1, the term $(\tilde{\lambda}^{-1}/T^2)^{1/(2m)}$ serves as the equivalent bandwidth used in a Nadaraya–Watson kernel estimator.

Besides the B-spline family, another popular choice of the basis is the p-th degree truncated power polynomial basis (TPS) defined as

$$\{1, z, \ldots, z^p, (z-\kappa_1)_+^p, \ldots, (z-\kappa_K)_+^p\},$$

where $(z-a)_+^p = \{\max(0, z-a)\}^p$. Note that the p-th degree TPS and the p-th degree B-splines span the same linear space. For any given β_k's, there exist γ_j's and b_k's such that

$$\sum_{k=1}^{K+p} \Psi_k(z)\beta_k = \sum_{j=0}^{p} \gamma_j z^j + \sum_{k=1}^{K} (z-\kappa_k)_+^p b_k.$$

Moreover, a direct calculation shows that the coefficients of TPS and B-splines satisfy $b_{k+1} = (-K)^p \Delta^{p+1}\beta_k$ (de Boor, 1978). Hence imposing penalty on $\sum_k b_k^2$ is equivalent to imposing the $(p+1)$-th order differencing penalty on the B-splines coefficients β_k's. In general, we could define the penalty matrix[3] Λ such that $\tilde{\lambda}^{-1}K^{2m-1}\beta^T D_m^T D_m \beta = \tilde{\lambda}^{-1}K^{2m-2p-1} b^T \Lambda b$. Equivalently, we could rewrite the minimization criterion (3.3) for TPS as

$$(Y - XZ_1\gamma - XZ_2 b)^T \Sigma_u^{-1}(Y - XZ_1\gamma - XZ_2 b) + \tilde{\lambda}^{-1}K^{2m-2p-1} b^T \Lambda b, \quad \textbf{(3.6)}$$

where Z_1 and Z_2 are matrices whose i-th row are $(1, z, \ldots, z^p)$ and $\{(z-\kappa_1)_+^p, \ldots, (z-\kappa_K)_+^p\}$ respectively, and $\gamma = (\gamma_0, \ldots, \gamma_p)^T$. Because of the equivalence between TPS and B-splines of the same degree, the penalized spline estimator based on TPS could achieve the same optimal convergence rate $T^{2m/(2m+1)}$ when the penalty parameter satisfies $\tilde{\lambda} \sim T^{-2/(2m+1)}$. Since the choice of p will not affect the convergence rate of the spline estimator, a conventional choice is to let $p = 1$ for $\theta(z) \in W^2[0,1]$, i.e., $m = 2$. In this case, Λ becomes the identity matrix I_K. If we denote $\lambda^{-1} = \tilde{\lambda}^{-1}K$, then the minimization criterion (3.6) could be written as

$$(Y - XZ_1\gamma - XZ_2 b)^T \Sigma_u^{-1}(Y - XZ_1\gamma - XZ_2 b) + \lambda^{-1} b^T b, \quad \textbf{(3.7)}$$

and the optimal rate of λ is of the order $T^{-2/5}K^{-1}$.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

4. INFERENCE USING LIKELIHOOD RATIO TESTS

In this section, we consider testing whether the functional coefficients $\theta(z)$ are time-invariant. The null hypothesis is H_0: $\theta(z) \equiv \theta_0$. Under the alternative, $\theta(z)$ is a smooth function of unknown form. Such a stability test is of both theoretical and empirical importance. For example, when a linear cointegration model is misspecified, the resulting estimation of $\hat{\theta}_0$ would not be consistent and neither of the equilibrium residuals. As a result, the traditional cointegration tests might fail to detect the cointegrating relationship.

In the literature of nonparametric regression, there are also lots of discussions on checking whether there is enough evidence to support the use of the general nonparametric method rather than a simple linear cointegration model. In general, traditional approaches often rely on (i) comparing the discrepancy measures between the estimates obtained under the null and the alternative (see Härdle and Mammen, 1993); or (ii) constructing the F-test statistic based on the sum of residuals (see Hong and White, 1995); or (iii) conducting the generalized likelihood ratio test using a reasonable smooth estimate under the alternative (see Fan, Zhang, and Zhang, 2001). In any of these methods, it is crucial to select the smoothing parameter under the nonparametric alternative. In practice, the power of the test is likely to be affected by the smoothing parameter, especially when it is chosen by some ad hoc methods.

In contrast, we are going to propose a likelihood ratio test procedure that could circumvent this difficulty as we use maximum likelihood principles for both estimation and inference. First, we model $\theta(z)$ by the p-th degree splines in order to define a general nonparametric alternative. As we show in the section above, there is not much difference to estimate $\theta(z)$ by using either the B-splines family or the TPS family. Moreover, the choice of p is not important. Therefore, we mainly focus on using the linear TPS family in this section. We could view the spline coefficients b_k's, associated with $(z - \kappa_k)_+$'s, as the deviations from the linear function. Hence testing the stability of $\theta(z)$ is equivalent to testing both the linear coefficient and the spline coefficients being 0, i.e.,

$$H_0 : \gamma_1 = 0 \qquad \text{and} \qquad b_1 = \cdots = b_K = 0$$

against

$$H_A : \gamma_1 \neq 0 \qquad \text{or} \qquad \exists\, k,\ s.t.\ b_k \neq 0.$$

Note that this is a multiple testing problem and the number of restrictions under H_0 grows as the sample size does. To circumvent this difficulty, a new idea is to utilize the mixed model representation for spline estimates based on TPS by treating b_k's as random coefficients with a common variance component, and then relate the null hypothesis above to the significance test of zero variance. More details are given below.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Note that minimizing (3.6) is equivalent to solving a system of equations

$$\begin{pmatrix} A_1^T \Sigma_u^{-1} A_1 & A_1^T \Sigma_u^{-1} A_2 \\ A_2^T \Sigma_u^{-1} A_1 & A_2^T \Sigma_u^{-1} A_2 + \lambda^{-1}\Lambda \end{pmatrix} \begin{pmatrix} \hat{\gamma} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} A_1^T \Sigma_u^{-1} Y \\ A_2^T \Sigma_u^{-1} Y \end{pmatrix},$$

where $A_1 = XZ_1$ and $A_2 = XZ_2$. The above equation is essentially Henderson's mixed model equation, which motivates us to utilize the mixed model representation to obtain $\hat{\gamma}$ and $\hat{b}$ as the best linear unbiased predictors (BLUPs) in the following model. To be specific, let $Y = A_1\gamma + A_2 b + u$, where γ is the 2×1 vector of fixed effect coefficients and b is the $K \times 1$ vector of random effect coefficients with mean 0 and variance $\lambda\sigma^2\Lambda^{-1}$, with $\Lambda = I_K$ when $m = 2$ and $p = 1$. The parameter λ controls the amount of smoothing and it could be viewed as the signal to noise ratio. Following Crainiceanu and Ruppert (2004), we could treat Y as

$$Y|b, A_1, A_2 \overset{D}{=} N(A_1\gamma + A_2 b, \sigma^2\Sigma_u), \quad b \overset{D}{=} N(0_{K\times 1}, \lambda\sigma^2\Lambda^{-1}).$$

Note that $\mathbb{E}(Y) = A_1\gamma$ and $var(Y) = \sigma^2(\Sigma_u + \lambda A_2\Lambda^{-1}A_2^T) =: \sigma^2\Omega_\lambda$. Hence we could define twice of the log-likelihood of Y as

$$2l(\gamma, \lambda, \sigma^2) = -(Y - A_1\gamma)(\sigma^2\Omega_\lambda)^{-1}(Y - A_1\gamma) - \log|\sigma^2\Omega_\lambda| - T\log(2\pi). \quad \textbf{(4.1)}$$

By maximizing (4.1), we could estimate the variance components by $\hat{\sigma}^2$ and $\hat{\lambda}$. Define $\hat{\Omega}_\lambda = \Sigma_u + \hat{\lambda}A_2\Lambda^{-1}A_2^T$. The BLUPs of γ and b are then obtained as

$$\hat{\gamma} = (A_1^T\hat{\Omega}_\lambda^{-1}A_1)^{-1}A_1^T\hat{\Omega}_\lambda^{-1}Y, \qquad \text{and} \qquad \hat{b} = \hat{\lambda}\Lambda^{-1}A_2^T\hat{\Omega}_\lambda^{-1}(Y - A_1\hat{\gamma}), \quad \textbf{(4.2)}$$

and we could estimate $\theta(z)$ by

$$\hat{\theta}(z) = \sum_{k=0}^{p}\hat{\gamma}_k z^k + \sum_{k=1}^{K}\widehat{b_k}(z - \kappa_k)_+^p.$$

For the same λ, minimizing the objective function (3.6) yields the same solution as (4.2). However, the use of the mixed model representation allows us to adopt the maximum likelihood principle to make estimation as well as inference on λ. In particular, $\lambda = 0$ implies $b_k = 0$ for all k. Hence the hypothesis test of $\theta(z)$ being constant is equivalent to testing

$$H_0 : \gamma_1 = 0 \qquad \text{and} \qquad \lambda = 0$$

against

$$H_A : \gamma_1 \neq 0 \qquad \text{or} \qquad \lambda \neq 0.$$

Then it is straightforward to rely on the LRT statistic for inference, where

$$LRT_T = \sup_{H_A} 2\log l(\gamma, \lambda, \sigma^2) - \sup_{H_0} 2\log l(\gamma, \lambda, \sigma^2).$$

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Note that the null distribution of the LRT statistic is not standard as the parameter λ is always nonnegative and it lies on the boundary of the parameter space under H_0. Therefore, we derive the exact and the limiting null distributions of our test statistic below.

First we consider the exact case, where both T and K are relatively small and could be treated as fixed. Let P be the projection matrix $P = I_T - \Sigma_u^{-1/2} A_1 (A_1^T \Sigma_u^{-1} A_1)^{-1} A_1^T \Sigma_u^{-1/2}$. Define $\xi_{s,T}$ and $\eta_{s,T}$ as the s-th eigenvalues of the $K \times K$ matrices $\Lambda^{-1/2} A_2^T \Sigma_u^{-1} A_2 \Lambda^{-1/2}$ and $\Lambda^{-1/2} A_2^T \Sigma_u^{-1/2} P \Sigma_u^{-1/2} A_2 \Lambda^{-1/2}$ respectively. We have the following results.

THEOREM 4.1. *Suppose that Assumptions 1–3 hold and the linear TPS with equispaced knots is used. Then under* $H_0 : \theta(z) \equiv \theta_0$,

$$LRT_T \stackrel{D}{=} \sup_{\lambda \geq 0} \left[T \log \left\{ 1 + \frac{N_T(\lambda)}{D_T(\lambda)} \right\} - \sum_{s=1}^{K} \log(1 + \lambda \xi_{s,T}) \right] + T \log \left\{ 1 + \frac{w_{T-1}^2}{\sum_{s=1}^{T-2} w_s^2} \right\}, \tag{4.3}$$

where $N_T(\lambda) = \sum_{s=1}^{K} \frac{\lambda \eta_{s,T}}{1+\lambda \eta_{s,T}} w_s^2$, $D_T = \sum_{s=1}^{K} \frac{w_s^2}{1+\lambda \eta_{s,T}} + \sum_{s=K+1}^{T-2} w_s^2$, *and* $w_s \stackrel{D}{=} i.i.d.N(0,1)$.

Theorem 4.1 derives the exact null distribution of the LRT statistic when the sample size T is finite. Although equation (4.3) does not have a close form, we could efficiently simulate this distribution using the following Algorithm A.

Step 1. define a grid $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_L$ of possible values for λ.
Step 2. simulate K independent χ_1^2 random variables $w_1^2, \ldots, w_K^2$.
Step 3. simulate a random variable ν_1 that follows χ_{T-2-K}^2.
Step 4. simulate a random variable ν_0 that follows χ_1^2.
Step 5. for every λ_i, compute $f_T(\lambda_i) = T \log\{1 + \frac{N_T(\lambda_i)}{D_T(\lambda_i)}\} - \sum_{s=1}^{K} (1 + \lambda \xi_{s,T})$.
Step 6. determine $\lambda_{\max}$ which maximizes $f_T(\lambda_i)$ over λ_i's.
Step 7. compute $f_T(\lambda_{\max}) + T \log\{1 + \frac{\nu_0}{\sum_{s=1}^{K} w_s^2 + \nu_1}\}$.
Step 8. repeat steps 2–7.

If we treat K as fixed and let T grow to infinity, we have the following results.

THEOREM 4.2. *Suppose that Assumptions 1–3 hold and the linear TPS with equispaced knots is used. Then there exist* η_s *and* ξ_s, *for* $s = 1, \ldots, K$, *such that*

$$T^{-2} \eta_{s,T} \Rightarrow \eta_s, \qquad T^{-2} \xi_{s,T} \Rightarrow \xi_s, \qquad \text{as} \quad T \to \infty. \tag{4.4}$$

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Moreover, under $H_0 : \theta(z) \equiv \theta_0$,

$$LRT_T \Rightarrow \sup_{d\geq 0}\left\{\sum_{s=1}^{K}\frac{d\eta_s}{1+d\eta_s}\omega_s^2 - \sum_{s=1}^{K}\log(1+d\xi_s)\right\} + \chi_1^2. \tag{4.5}$$

Remark 4.1. In Theorem 4.2, we have explicitly derived that the convergence rate of the eigenvalues $\eta_{s,T}$ and $\xi_{s,T}$ is T^2, which is faster than the T convergence rate due to nonstationarity. When K is fixed, the part corresponding to testing $\lambda = 0$ converges to the term $\sup_{d\geq 0}\{\sum_{s=1}^{K}\frac{d\eta_s}{1+d\eta_s}\omega_s^2 - \sum_{s=1}^{K}\log(1+d\xi_s)\}$. Since this limiting distribution is nonstandard, one could simulate it by modifying Algorithm A described above.

Furthermore, if we assume K and T both grow to infinity, the null distribution of LRT_T approaches to a simple χ^2 distribution.

THEOREM 4.3. *Suppose that Assumptions 1–3 hold and the linear TPS with equispaced knots is used. Let the number of knots* K *satisfying* $K \sim T^r$ *with* $4/5 < r < 1$. *Then there exist* $\bar{\eta}$ *and* $\bar{\xi}$ *such that*

$$K^{-1}T^{-2}\sum_{k=1}^{K}\eta_{k,T} \Rightarrow \bar{\eta}, \qquad K^{-1}T^{-2}\sum_{k=1}^{K}\xi_{k,T} \Rightarrow \bar{\xi}. \tag{4.6}$$

Under $H_0 : \theta(z) \equiv \theta_0$, *we have*

$$LRT_T \Rightarrow \chi_1^2. \tag{4.7}$$

Remark 4.2. Theorem 4.3 assumes that K grows as T does. Compared to the fixed K case, the amount of penalty λ^{-1} is expected to be larger, and the probability of obtaining the maximum likelihood estimate (MLE) of λ at its actual value 0 approaches to 1 provided that H_0 is true. Therefore, the part corresponding to testing $\lambda = 0$ degenerates and we have a simple χ^2 distribution.

For the local alternatives, we assume that $\theta(z) = \theta_0 + T^{-\alpha}\theta_1(z)$, where $\theta_1(z)$ is a nonzero smooth function that belongs to $W^2[0, 1]$. Suppose we span $\theta_1(z)$ with the first degree TPS as $\bar{\gamma}_0 + \bar{\gamma}_1 z + \sum_{k=1}^{K}\bar{b}_k(z-\kappa_k)_+$. Utilizing the mixed model representation, we treat the spline coefficients $\bar{b} = (\bar{b}_1, \ldots, \bar{b}_K)^T$ as random with mean 0 and variance $\bar{\lambda}_0\sigma^2\Lambda^{-1}$. It has been shown in Section 3 that $\bar{\lambda}_0$ converges to 0 at the rate of $T^{2/5}K$. Therefore, we denote $\bar{\lambda}_0 = \bar{d}_0 T^{-2/5}K^{-1}$ for some constant $\bar{d}_0 \geq 0$. Recall that our LRT will examine both the fixed part $\bar{\gamma}_1$ and the variance part $\bar{d}_0$. Therefore, we will consider two different cases in the local alternatives. In Case 1, $\theta_1(z)$ is a linear function with nonzero slope, i.e., $\bar{\gamma}_1 \neq 0$ but $\bar{d}_0 = 0$. The local alternative is then set as $H_{01} : \theta(z) = \theta_0 + T^{-1}\theta_1(z)$. In Case 2, $\theta_1(z)$ has $\bar{\gamma}_1 = 0$ but $\bar{d}_0 \neq 0$. The local alternative is set as $H_{02} : \theta(z) = \theta_0 + T^{-4/5}\theta_1(z)$.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

THEOREM 4.4. *Suppose that Assumptions 1–3 hold and the linear TPS with equispaced knots is used. Let the number of knots K satisfying $K \sim T^r$ with $4/5 < r < 1$.*

Under the local alternative H_{01}, the LRT statistic converges to a noncentral χ_1^2, i.e.,

$$LRT_T \Rightarrow (w_1 + \bar{\gamma}_1 \bar{\pi}_2)^2,$$

where $w_1 \overset{D}{=} N(0,1)$ and $\bar{\pi}_2$ is defined right before equation (A.8).

Under the local alternative H_{02},

$$LRT_T \Rightarrow \max_{d \in [0, \bar{d}_0]} \{\bar{d}_0 \bar{\eta} + \bar{g}_3(d) - \bar{g}_2(d)\} + (1 + \bar{d}_0 \varrho) w_1^2,$$

where $w_1 \overset{D}{=} N(0,1)$, $\bar{\eta}$ and $\bar{g}_2(d)$ are defined as in Theorem 4.3, $\bar{g}_3(d)$ is defined right after equation (A.9), and ϱ is defined right before equation (A.10).

Remark 4.3. Strictly speaking, we should also consider Case 3 where neither $\bar{\gamma}_1$ or $\bar{d}_0$ equals 0. For this case, the local alternative could be set as $H_{03} : \theta(z) = \theta_0 + T^{-1}\theta_1(z)$. Notice that such a local alternative converges with a rate faster than $T^{4/5}$. The nonzero variance component will not affect the asymptotic distribution. Hence in Case 3, the LRT statistic still converges to a noncentral χ^2 distribution as in Case 1. To save the length of this paper, we omit detailed discussions of Case 3.

Remark 4.4. Under H_{02}, the asymptotic distribution of LRT_T has two components, where the first part is nonnegative and the second part is a scaled χ_1^2. In summary, our penalized spline estimator of $\theta(z)$ has the $T^{4/5}$ convergence rate, while our test statistics could detect an alternative whose convergence rate is not faster than $T^{4/5}$. On the other hand, for any sequence such that $H_{0A} : \theta(z) \equiv \theta_0 + T^{-\alpha}\theta_1(z)$ and $\alpha < 4/5$, LRT_T diverges and the power function satisfies

$$P(LRT_T > \chi^2_{1,\alpha/2}) \to 1,$$

where $\chi^2_{1,\alpha/2}$ is the upper $\alpha/2$ quantile of χ_1^2 distribution. Hence the proposed likelihood ratio test could achieve the optimality.

5. FINITE SAMPLE PERFORMANCE

Monte Carlo simulations are conducted in this section to examine the finite sample performance of the proposed LRT. The data generating process is

$$y_t = \theta(z_t)x_t + u_t,$$

where $x_t = x_{t-1} + v_t$, $u_t = \rho u_{t-1} + \varepsilon_t$, v_t's and ε_t's are i.i.d. $N(0,1)$, and they are independent of each other. The initial values are set to be zero. In particular, four cases for the parameter values are considered: (i) $\theta(z_t) = 0.25, \rho = 0$;

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

(ii) $\theta(z_t) = 0.25, \rho = 0.5$; (iii) $\theta(z_t) = (z_t - 0.5)^2, \rho = 0$; (iv) $\theta(z_t) = (z_t - 0.5)^2, \rho = 0.5$, where the first two cases are related to calculating the size of the test and the last two are related to calculating the power of the test. The simulation designs above are similar to those in Xiao (2009). The sample sizes we consider are $T = 100$ and 300. In particular, we would like to examine the impact of the number of knots K. Hence when $T = 100$, we consider three situations, $K = 10$, $K = 20$, and $K = 40$; when $T = 300$, we consider $K = 20$, $K = 40$, and $K = 80$. All reported results are based on 2,000 replications.

Table 1 reports the size of the proposed likelihood ratio test when ρ is given, i.e., the true covariance matrix Σ_u is known. The five columns on the left use the critical values based on the finite distribution derived in Theorem 4.1, while the five columns on the right use the critical values based on the asymptotic distribution χ^2_1 as indicated in Theorem 4.3. From Table 1, we find that both the finite distribution and the χ^2 limiting distribution work very well. For example, consider Panel A1 with K chosen as 10. Even though the sample size is just 100, the actual rejection rates based on the asymptotic χ^2 distribution are 0.1995, 0.1490, 0.0995, 0.0465, and 0.0100, very close to the nominal sizes 0.2, 0.15, 0.10, 0.05, and 0.01 respectively. Moreover, we find that the number of knots does not have much impact on the size performance. For a given nominal level in any reported panel, the absolute differences in the rejection rates associated with different K are not greater than 0.005. This is consistent with the empirical conclusion that the number of knots is not important, provided that it is above some minimum threshold (Ruppert, 2002).

Table 2 repeats all designs in Table 1, except that the covariance matrix Σ_u is treated as unknown and replaced by an estimate. We find that our LRT procedure still performs well and is less likely to be affected by the fact that the covariance is unknown.

Table 3 reports the power of our test statistic. Once again, we find that the choice of the number of knots K is not important and the procedure is robust against the use of an estimated covariance. When the sample size increases from 100 to 300, the rejection rates are all greater than 0.98, implying very good power performance of our testing procedure.

6. CONCLUSIONS

Varying-coefficient regression models with nonstationary regressors have received intense interest in recent years. This paper proposes a penalized splines approach to estimate the varying coefficients. Compared to kernel-based methods, penalized splines estimation not only achieves the same optimal convergence rate, but also enjoys the advantage of fast computation. Utilizing the mixed model representation of penalized splines, we construct a likelihood ratio test for the stability of the varying coefficients. We derive the exact and limiting distributions of the proposed test statistic. When the number of knots is treated as fixed, the null distribution is nonstandard, but could be simulated via a proposed algorithm using

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

spectral decomposition. When the number of knots grows as the sample size does, the limiting null distribution converges to a simple χ^2 distribution. Our test is less likely to suffer from the misselection of the smoothing parameters. Simulations show that the asymptotic distribution works very well even for the finite sample case.

There are some issues worth of future studies. One potential analysis is to extend the current setting to the case allowing for dependence between u_t and v_t. Another natural extension is to consider a more general varying-coefficient cointegrating regression model which includes both the stochastic and the deterministic functional coefficients in the cointegrating relationship.

NOTES

1. Other contributions to the literature on the stability tests in cointegrating regression include Hansen (1992), Hao (1996), Quintos (1997), Kuo (1998), Hansen and Johansen (1999), Johansen, Mosconi, and Nielsen (2000), Harris, McCabe, and Leybourne (2002), among others. Furthermore, Hong and Phillips (2010) propose a modified RESET test for testing linearity in the cointegration model. Gao, King, Lu, and Tjøstheim (2009) consider a nonparametric specification test for the nonparametric time series model with nonstationary variables.

2. The dimensions of Ψ and β both depend on the degree of splines p and the number of knots K. For notation simplicity, we suppress the subscripts p and K.

3. The choice of Λ depends on both the degree of splines p and the order of penalty m. For notation simplicity, we suppress the subscripts p and m.

REFERENCES

Bierens, H.J. & L. Martins (2010) Time-varying cointegration. *Econometric Theory* 26, 1453–1490.

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

Campbell, J. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Econometrics* 81, 27–60.

Chen, H., Y. Fang, & Y. Li (2013) Estimation and inference for varying-coefficient models with nonstationary regressors using penalized splines. Available at http://ssrn.com/abstract=2287449.

Claeskens, G., T. Krivobokova, & J. Opsomer (2009) Asymptotical properties of penalized spline estimators. *Biometrika* 96, 529–544.

Crainiceanu, C.M. & D. Ruppert (2004) Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B* 66, 165–185.

Crainiceanu, C.M., D. Ruppert, G. Claeskens, & M.P. Wand (2005) Exact likelihood ratio tests for penalized splines. *Biometrika* 92, 91–103.

de Boor, C. (1978) *A Practical Guide to Splines.* Springer-Verlag.

Eilers, P.H.C. & B.D. Marx (1996) Flexible smoothing with B-splines and penalties. *Statistical Science* 11, 89–121.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.

Fan, J., C. Zhang, & J. Zhang (2001) Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153–193.

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Goldfajn, I. & T. Baig (1998) Monetary Policy in the Aftermath of Currency Crises: The Case of Asia. Working paper, IMF.
Goyal, A. & I. Welch (2003) Predicting the equity premium with dividend ratios. *Management Science* 49, 639–654.
Hall, P. & J.D. Opsomer (2005) Theory for penalised spline regression. *Biometrika* 92, 105–118.
Hansen, B.E. (1992) Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* 10, 321–335.
Hansen, H. & S. Johansen (1999) Some tests for parameter constancy in cointegrated VAR-models. *Econometrics Journal* 2, 306–333.
Hao, K. (1996) Testing for structural change in cointegrated regression models: Some comparisons and generalizations. *Econometric Reviews* 15, 401–429.
Härdle, W. & E. Mammen (1993) Comparison nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.
Harris, D., B. McCabe, & S. Leybourne (2002) Stochastic cointegration: Estimation and inference. *Journal of Econometrics* 111, 363–384.
Hong, S. & P.C.B. Phillips (2010) Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business and Economic Statistics* 28, 96–113.
Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.
Johansen, S., R. Mosconi, & B. Nielsen (2000) Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* 3, 216–249.
Juhl, T. (2005) Functional-coefficient models under unit root behavior. *Econometrics Journal* 8, 197–213.
Kasparis, I. (2008) Detection of functional form misspecification in cointegrating relations. *Econometric Theory* 24, 1373–1404.
Kuo, B.S. (1998) Test for partial parameter instability in regressions with I(1) processes. *Journal of Econometrics* 86, 337–368.
Lettau, M. & S. Ludvigsson (2001) Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56, 815–849.
Li, Y. & D. Ruppert (2008) On the asymptotics of penalized splines. *Biometrika* 95, 415–436.
Nelson, R. & C.I. Plosser (1982) Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10, 139–162.
Park, J. & S. Hahn (1999) Cointegrating regressions with time varying coefficients. *Econometric Theory* 15, 664–703.
Park, J. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.
Patterson, H.D. & R. Thompson (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
Paye, B.S. & A. Timmermann (2006) Instability of return prediction models. *Journal of Empirical Finance* 13, 274–315.
Quintos, C.E. (1997) Stability tests in error correction models. *Journal of Econometrics* 82, 289–315.
Ruppert, D. (2002) Selecting the number of knots for penalized splines. *Journal of the Computational and Graphical Statistics* 11, 735–757.
Ruppert, D., M.P. Wand, & R.J. Carroll (2003) *Semiparametric Regression.* Cambridge University Press.
Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–20.
Saikkonen, P. & I. Choi (2004) Cointegrating smooth transition regressions. *Econometric Theory* 20, 201–340.
Shao, Q. & C. Lu (1987) Strong approximation for partial sums of weakly dependent random variables. *Scientia Sinica* 15, 576–587.
Shi, X. & P.C.B. Phillips (2012) Nonlinear cointegrating regression under weak identification. *Econometric Theory* 28, 509–547.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Wang, Q. & P.C.B. Phillips (2009a) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
Wang, Q. & P.C.B. Phillips (2009b) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.
Xiao, Z. (2009) Functional-coefficient cointegration models. *Journal of Econometrics* 152, 81–92.

APPENDIX

A. Proofs

In this appendix, we provide a sketch of our main result. For more details please refer to the working paper by Chen, Fang, and Li (2013).

Proof of Theorem 3.1. Without loss of generality, we assume u_t's are uncorrelated. First we can decompose the penalized spline estimator as $R_1 + R_2 + O_p(K^{-1})$, where the term $O_p(K^{-1})$ comes from the bias due to splines approximation,

$$R_1 =: \Psi_z \left\{ \left(\frac{K}{T^2} \Psi^T X^2 \Psi \right) + \bar{\lambda} K^{2m} D_m^T D_m \right\}^{-1} \left(\frac{K}{T^2} \Psi^T X^2 \Psi \right) \beta,$$

$$R_2 =: \frac{1}{\sqrt{T}} \Psi_z \left\{ \left(\frac{K}{T^2} \Psi^T X^2 \Psi \right) + \bar{\lambda} K^{2m} D_m^T D_m \right\}^{-1} \left\{ \frac{\Psi^T (T^{-1/2} X u)}{T/K} \right\},$$

and $\bar{\lambda} = \tilde{\lambda}^{-1}/T^2$. Note that the (i, j)th element of $KT^{-2}\Psi^T X^2 \Psi$ satisfies

$$\begin{aligned} KT^{-2}(\Psi^T X^2 \Psi)_{i,j} &= KT^{-2} \sum_{t=1}^{T} \mathbb{E}\left\{ \Psi_i^{[p]}(z_t) \Psi_j^{[p]}(z_t) \right\} x_t^2 + o_p(1) \\ &= K\mathbb{E}\left\{ \Psi_i^{[p]}(z_t) \Psi_j^{[p]}(z_t) \right\} T^{-1} \sum_{t=1}^{T} \left(\frac{x_t}{\sqrt{T}} \right)^2 + o_p(1). \end{aligned}$$

Hence $KT^{-2}(\Psi^T X X \Psi)_{ij} \Rightarrow q_{ij} \int B_v^2(s) ds$, where q_{ij} is the limit of $K\mathbb{E}\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}$. Since $KT^{-2}\Psi^T X^2 \Psi$ converges, we could use a similar technique as in Li and Ruppert (2008) to conclude the term $\tilde{\lambda}^{1/(2m)}$ serves equivalently as the bandwidth h used in a Nadaraya–Watson kernel estimator. Hence $R_1 - \Psi_z \beta = O_p(h^m) = O_p\{\tilde{\lambda}^{m/(2m)}\} = O_p(T^{-1}\tilde{\lambda}^{-1/2})$. Note that the i-th element of $\Psi^T X u$ satisfies $\frac{\Psi^T (T^{-1/2} X u)_i}{T/K} = \frac{\sum_{t=1}^{T} \{\Psi_i^{[p]}(z_t) \frac{x_t}{\sqrt{T}} u_t\}}{T/K} = O_p\left(\frac{1}{\sqrt{T/K}} \right)$. Thus $\mathbb{E}(R_2) = 0$ and $var(\sqrt{T} R_2) = O(\frac{1}{Kh} \frac{1}{T/K}) = O(\frac{1}{Th}) = O(\frac{1}{T\tilde{\lambda}^{1/(2m)}}) = O(T^{-1+1/m} \tilde{\lambda}^{1/(2m)})$. Therefore, $R_2 = O_p(T^{-1+1/(2m)} \tilde{\lambda}^{1/(4m)})$ and Theorem 3.1 holds. ■

Proof of Theorem 4.1. For a given λ, define $L(\lambda)$ as the profile likelihood function. Let $\log L^0$ be the maximum log-likelihood under H_0. Then $2LRT_T$ can be decomposed as

$$2LRT_T = \sup_{\lambda \geq 0} \{2 \log L(\lambda) - 2 \log L(0)\} + \{2 \log L(0) - 2 \log L^0\}, \tag{A.1}$$

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

where the first part corresponds to testing $\lambda = 0$ and the second part corresponds to testing the linear coefficient $\gamma_1 = 0$ given that $\lambda = 0$. We first prove

$$\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{\lambda\geq 0}[T\log\{1 + N_T(\lambda)/D_T(\lambda)\} - \sum_{k=1}^{K}\log(1+\lambda\xi_{k,T})]. \quad \textbf{(A.2)}$$

Define $\tilde{Y} = \Sigma_u^{-1/2}Y$, $\tilde{A}_1 = \Sigma_u^{-1/2}A_1$, $\tilde{A}_2 = \Sigma_u^{-1/2}A_2$, and $V_\lambda =: I_T + \lambda\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T$. Then $-2\log L(\lambda) = T\log(T\hat{\sigma}_\lambda^2/\sigma^2) + \log|V_\lambda| + C_0$, where σ_λ^2 is the profile MLE and C_0 is a fixed constant. Note that $\log|V_\lambda| = \log|I_K + \lambda\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_2\Lambda^{-1/2}| = \sum_{k=1}^{K}\log(1+\lambda\xi_{k,T})$. According to Patterson and Thompson (1971), there exists a matrix W such that

$$WW^T = P = I_T - \tilde{A}_1\left(\tilde{A}_1^T\tilde{A}_1\right)^{-1}\tilde{A}_1^T, \qquad W^TW = I_{T-2}, \quad \textbf{(A.3)}$$

and $T\hat{\sigma}_\lambda^2 = (\tilde{Y} - \tilde{A}_1\hat{\gamma}_\lambda)^T V_\lambda^{-1}(\tilde{Y} - \tilde{A}_1\hat{\gamma}_\lambda) = \tilde{Y}^T W(W^TV_\lambda W)^{-1}W^T\tilde{Y}$. By eigendecomposition, $W^TV_\lambda W = U_1S_1U_1^T$, where S_1 is the diagonal matrix whose first K diagonal elements are $1+\lambda\eta_{k,T}$'s and the rest are 1. Under H_0, $W^T\tilde{Y}/\sigma \sim N(0_{I_{T-2}}, I_{T-2})$. Hence $-2\log L(\lambda) = T\log(\sum_{k=1}^{K}\{(1+\lambda\eta_{s,T})^{-1}w_k^2 + \sum_{k=K+1}^{T-2}w_k^2\} + \sum_{k=1}^{K}\log(1+\lambda\xi_{k,T}) + C_0$. In particular, $-2\log L(0) = T\log(\sum_{k=1}^{T-2}w_k^2) + C_0$. We conclude equation (A.2) holds. By standard results in linear regression, $-2\log L^0 = T\log(\sum_{k=1}^{T-1}w_k^2) + C_0$. Theorem 4.1 holds. ■

Proof of Theorem 4.2. First we prove equation (4.4). By continuous mapping theorem, it suffices to show that $T^{-2}A_1^TA_1$, $T^{-2}A_2^TA_2$, and $T^{-2}A_1^TA_2$ converge. Take $T^{-2}A_1^TA_1$ as an example. Note that the (i,j)th element of $A_1^TA_1$ satisfying $(A_1^TA_1)_{i,j} = \sum x_t^2z_t^{i+j-2}$. Since

$$\begin{aligned} T^{-2}\sum x_t^2z_t^{i+j-2} &= T^{-2}\mathbb{E}\left(z_t^{i+j-2}\right)\sum x_t^2 + T^{-2}\sum x_t^2\left(z_t^{i+j-2} - \mathbb{E}\left(z_t^{i+j-2}\right)\right) \\ &= T^{-2}\mathbb{E}\left(z_t^{i+j-2}\right)\sum x_t^2 + o_p(1), \end{aligned}$$

we conclude $T^{-2}\left(A_1^TA_1\right)_{i,j} = T^{-2}\sum x_t^2z_t^{i+j-2} \Rightarrow \mathbb{E}\left(z_t^{i+j-2}\right)\int B_v^2(s)ds$. Applying the same technique, we could prove $T^{-2}A_2^TA_2$ and $T^{-2}A_1^TA_2$ converge, and thus equation (4.4) holds.

Next we prove equation (4.5). Let $d = \lambda T^2$. Then equation (A.2) yields

$$\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{d\geq 0}\left[T\log\left\{1 + \frac{N_T(dT^{-2})}{D_T(dT^{-2})}\right\} - \sum_{k=1}^{K}\log(1+dT^{-2}\xi_{k,T})\right]. \quad \textbf{(A.4)}$$

Define the right hand side in equation (A.4) as $\sup_{d\geq 0}f_T(d)$. We want to show that

$$\sup_{d\geq 0}f_T(d) \Rightarrow \sup_{d\geq 0}f(d) =: \sup_{d\geq 0}\left\{\sum_{s=1}^{K}\frac{d\eta_s}{1+d\eta_s}w_s^2 - \sum_{s=1}^{K}\log(1+d\xi_s)\right\}. \quad \textbf{(A.5)}$$

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

It is straightforward to establish the finite dimensional convergence of $f_T(d)$ to $f(d)$, as $N_T(dT^{-2})$ converges to $\sum_{s=1}^{K} \frac{d\eta_s}{1+d\eta_s} w_s^2$, $T^{-1} D_T(dT^{-2})$ converges to 1, and $T \log\{1 + \frac{N_T(dT^{-2})}{D_T(dT^{-2})}\}$ converges to $\sum_{s=1}^{K} \frac{d\eta_s}{1+d\eta_s} w_s^2$ for every fixed d. The weak convergence of $f_T(d)$ to $f(d)$ then follows as $f_T(d)$ forms a tight sequence. Following Crainiceanu and Ruppert (2004), we could establish a continuous mapping theorem type result. Hence equation (A.5) holds. By standard results in linear regression, $2\log L(0) - 2\log L^0 \Rightarrow \chi_1^2$. Thus Theorem 4.2 is proved. ■

Proof of Theorem 4.3. First we prove equation (4.6). By continuous mapping theorem, it suffices to show $(KT^2)^{-1} tr(A_2^T A_2)$, $(KT^2)^{-1} tr(A_2^T A_1)$, and $(KT^2)^{-1} tr(A_1^T A_2)$ converge. Take $(KT^2)^{-1} tr(A_2^T A_2)$ as an example. Note that the (i,i)th element in $T^{-2} A_2^T A_2$ converges to $\mathbb{E}(z - \kappa_i)_+^2 \int B_v^2(s) ds$. Since $\mathbb{E}(z - \kappa_i)_+^2$'s are bounded, $(KT^2)^{-1} tr(A_2^T A_2)$ converges. Similarly, $(KT^2)^{-1} tr(A_1^T A_1)$ and $(KT^2)^{-1} tr(A_1^T A_2)$ converge. Thus equation (4.6) is true.

Next we prove equation (4.7). Let $d = \lambda KT^2$. By equation (A.2), we could conclude

$$\sup_{\lambda \geq 0} \{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{d \geq 0} g_T(d),$$

where

$$g_T(d) =: T \log \left\{ 1 + \frac{N_T\left(d(KT^2)^{-1}\right)}{D_T\left(d(KT^2)^{-1}\right)} \right\} - \sum_{k=1}^{K} \log\{1 + d(KT^2)^{-1} \xi_{k,T}\}. \tag{A.6}$$

Let $g_1(d)$ and $g_2(d)$ be continuous functions defined respectively as the following limits:

$$\bar{g}_1(d) = \lim_{T \to \infty} \sum_{k=1}^{K} \frac{dK^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}},$$

$$\bar{g}_2(d) = \lim_{T \to \infty} \sum_{k=1}^{K} \log(1 + dK^{-1}T^{-2}\xi_{k,T}). \tag{A.7}$$

Lemma A.1 shows that $\bar{g}_1(d)$ and $\bar{g}_2(d)$ exist for every fixed d and established the finite dimensional convergence of $g_T(d)$ to $\bar{g}_1(d) - \bar{g}_2(d)$. Similar to Theorem 4.2, the weak convergence of $g_T(d)$ follows as $g_T(d)$ forms a tight sequence. By establishing a continuous mapping theorem type result, we have $\sup_{d \geq 0} g_T(d) \Rightarrow \sup_{d \geq 0}\{\bar{g}_1(d) - \bar{g}_2(d)\}$.

Next we prove $\sup_{d \geq 0}\{\bar{g}_1(d) - \bar{g}_2(d)\} = 0$ by considering its first derivative. Decompose the first derivative of the partial sum induced by $\bar{g}_1(d) - \bar{g}_2(d)$ as $Q_{1,T} + Q_{2,T}$ where we define $Q_{1,T} = \sum_{k=1}^{K} \frac{K^{-1}T^{-2}\eta_{k,T}}{(1+dK^{-1}T^{-2}\eta_{k,T})^2} - \sum_{k=1}^{K} \frac{K^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}$ and $Q_{2,T} = \sum_{k=1}^{K} \frac{K^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}} - \sum_{k=1}^{K} \frac{K^{-1}T^{-2}\xi_{k,T}}{1+dK^{-1}T^{-2}\xi_{k,T}}$. Since we have $(1 + dK^{-1}T^{-2}\eta_{k,T}) \geq 1$, it holds that $Q_{1,T} \leq 0$ for all T. Note that the difference between $\Lambda^{-1/2}\tilde{A}_2^T \tilde{A}_2 \Lambda^{-1/2}$ and $\Lambda^{-1/2}\tilde{A}_2^T P \tilde{A}_2 \Lambda^{-1/2}$ is a semipositive definite matrix. Hence the associated eigenvalues satisfy $\eta_{k,T} \leq \xi_{k,T}$ for all k and T. As $\frac{x}{1+dx}$ is an increasing function of x for $d \geq 0$, we conclude $Q_{2,T} \leq 0$. Because the first derivative satisfies $\bar{g}_1'(d) - \bar{g}_2'(d) = \lim_{T\to\infty}(Q_{1,T} + Q_{2,T}) \leq 0$ and $\bar{g}_1(0) - \bar{g}_2(0) = 0$, we conclude $\sup_{d \geq 0} \bar{g}_1(d) - \bar{g}_2(d) = 0$. Together with the fact that $2\log L(0) - 2\log L^0 \Rightarrow \chi_1^2$, Theorem 4.3 is proved. ■

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

LEMMA A.1. *Assume the conditions in Theorem 4.3 and define* $g_T(d)$ *as in equation (A.6).*

Then $g_T(d)$ *converges to* $\bar{g}_1(d) - \bar{g}_2(d)$ *for every fixed* d.

Proof of Lemma A.1. First, we prove $\bar{g}_1(d)$ and $\bar{g}_2(d)$ exist for any fixed $d \geq 0$.

Note that $1 + d\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T} \leq \prod_{k=1}^{K}(1 + dK^{-1}T^{-2}\xi_{k,T}) \leq \exp\{d\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T}\}$. If $d\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T}$ converges, so does $\log\{\prod_{k=1}^{K}(1 + dK^{-1}T^{-2}\xi_{k,T})\}$. By the fact that $\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T} \Rightarrow \bar{\xi}$, we conclude $\sum_{k=1}^{K}\log(1 + dK^{-1}T^{-2}\xi_{k,T})$ converges and its limit $\bar{g}_2(d)$ exists. Note that $0 \leq \frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}} \leq dK^{-1}T^{-2}\eta_{k,T} \leq dK^{-1}T^{-2}\xi_{k,T}$. As $\sum_{k=1}^{K} dK^{-1}T^{-2}\xi_{k,T}$ converges, the limit of $\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}$ exists and we could denote it as $\bar{g}_1(d)$.

Next we establish the finite dimensional convergence of $g_T(d)$ to $\bar{g}_1(d) - \bar{g}_2(d)$. Lemma A.2 shows $N_T(dK^{-1}T^{-2})$ converges to $\bar{g}_1(d)$ for every fixed d. Moreover, we have

$$T^{-1}D_T\left(d(KT^2)^{-1}\right) = T^{-1}\{\tilde{Y}^T WW^T\tilde{Y}/\sigma^2 - N_T\left(d(KT^2)^{-1}\right)\} = 1 + o_p(1),$$

and thus $T\log\{1 + \frac{N_T\left(d(KT^2)^{-1}\right)}{D_T\left(d(KT^2)^{-1}\right)}\}$ converges to $\bar{g}_1(d)$ for every fixed d. With the fact that $\sum_{k=1}^{K}\log\{1 + d(KT^2)^{-1}\xi_{k,T}\}$ converges to $\bar{g}_2(d)$ for every fixed d, Lemma A.1 holds. ■

LEMMA A.2. *Assume the conditions in Theorem 4.3.*

Then $N_T(dK^{-1}T^{-2})$ *converges to* $\bar{g}_1(d)$ *for every fixed* d.

Proof of Lemma A.2. Notice that $w_k \overset{D}{=} i.i.d.\ N(0,1)$. It suffices to establish the finite dimensional convergence of $\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}w_k^2 - \bar{g}_1(d)$ to 0. Note that $\mathbb{E}\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}(w_k^2 - 1) = 0$ for every fixed d, and $var\{\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}(w_k^2 - 1)\} = 2\sum_{k=1}^{K}(\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}})^2 = O(K^{-1}) = o(1)$. Hence Lemma A.2 is proved. ■

Proof of Theorem 4.4. First we consider the local alternative H_{01}. Note that all spline coefficients are 0 under H_{01}. Therefore, it still holds $\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \Rightarrow 0$. It suffices to show $2\log L(0) - 2\log L^0$ converges to a noncentral χ_1^2 with parameter $\bar{\gamma}_1\bar{\pi}_2$. Define $\tilde{u} = \Sigma_u^{-1/2}u$. Note that $-2\log L(0) = T\log(\tilde{Y}^T WW^T\tilde{Y}/\sigma^2) + C_0 = T\log(\tilde{u}^T WW^T\tilde{u}/\sigma^2) + C_0$, where the last equality holds as W is orthogonal to $\tilde{A}_1$. Similarly, there exists a $(T-1)\times T$ matrix W_0 such that $-2\log L^0 = T\log(\tilde{Y}^T W_0W_0^T\tilde{Y}/\sigma^2) + C_0$. Partition A_1 as $A_1 = (A_{01}, A_{r1})$, where A_{01} is the $T\times 1$ vector whose elements are x_t's and A_{r1} is the $T\times 1$ vector whose elements are x_tz_t's. Define $\tilde{A}_{01} = \Sigma_u^{-1/2}A_{01}$ and $\tilde{A}_{r1} = \Sigma_u^{-1/2}A_{r1}$. Note that W_0 is orthogonal to $\tilde{A}_{01}$. Project $\tilde{A}_{r1}$ onto the unit direction $\frac{\tilde{A}_{01}}{||\tilde{A}_{01}||}$ and the unit direction orthogonal to $\tilde{A}_{01}$, i.e., $\tilde{A}_{r1} = \pi_1\frac{\tilde{A}_{01}}{||\tilde{A}_{01}||} + \pi_2\tilde{A}_{01}^c$. Since $W_0^T\tilde{A}_{r1} = \pi_2W_0^T\tilde{A}_{01}^c$, it holds that

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

$$\begin{aligned}
&\frac{\tilde{Y}^T W_0 W_0^T \tilde{Y} - \tilde{Y}^T W W^T \tilde{Y}}{\sigma^2} \\
&\quad = \frac{T^{-2}\bar{\gamma}_1^2\pi_2^2(\tilde{A}_{01}^c)^T W_0 W_0^T \tilde{A}_{01}^c}{\sigma^2} + \frac{2T^{-1}\bar{\gamma}_1\pi_2(\tilde{A}_{01}^c)^T W_0 W_0^T \tilde{u}}{\sigma^2} \\
&\qquad + \frac{\tilde{u}^T(W_0 W_0^T - W W^T)\tilde{u}}{\sigma^2} \\
&\quad = \frac{T^{-2}\bar{\gamma}_1^2\pi_2^2}{\sigma^2} + \frac{2T^{-1}\bar{\gamma}_1\pi_2(\tilde{A}_{01}^c)^T\tilde{u}}{\sigma^2} + \frac{\tilde{u}^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{u}}{\sigma^2} \\
&\quad = \{(\tilde{A}_{01}^c)^T\tilde{u}/\sigma + T^{-1}\pi_2\bar{\gamma}_1\}^2.
\end{aligned}$$

Note that $(\tilde{A}_{01}^c)^T\tilde{u}/\sigma \stackrel{D}{=} N(0,1)$ and we could denote it as w_1. Moreover,

$$(T^{-1}\pi_2)^2 = (T^{-1}\pi_2)^2(\tilde{A}_{01}^c)^T W_0 W_0^T \tilde{A}_{01}^c = T^{-2}\tilde{A}_{r1}^T W_0 W_0^T \tilde{A}_{r1} = T^{-2}\tilde{A}_{r1}^T P_0 \tilde{A}_{r1}.$$

Similar to Theorem 4.2, we conclude all elements in $T^{-2}A_{r1}^T A_{r1}$, $T^{-2}A_{r2}^T A_{r2}$, and $T^{-2}A_{r1}^T A_{r2}$ converge. Thus $T^{-1}\pi_2$ converges and we could denote its limit as $\bar{\pi}_2$. Then we have

$$LRT_T = 2\log L(0) - 2\log L^0 + o_p(1) = (w_1 + \bar{\gamma}_1\bar{\pi}_2)^2 + o_p(1), \tag{A.8}$$

i.e., the asymptotic distribution is a noncentral chi-square distribution with parameter $\bar{\gamma}_1\bar{\pi}_2$.

Next consider H_{02}. Let $d = \lambda KT^2$. Apply a similar technique as in the proof of Theorem 4.1, we could show that $\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{d\geq 0} h_T(d)$, where $h_T(d)$ is defined as $h_T(d) =: T\log\{1 + \frac{N_T'(d(KT^2)^{-1})}{D_T(d(KT^2)^{-1})}\} - \sum_{k=1}^K \log\{1 + d(KT^2)^{-1}\xi_{k,T}\}$ and $N_T'(\lambda) = \lambda\sum_{k=1}^K \frac{\eta_{k,T} + \bar{d}_0(KT^2)^{-1}\eta_{k,T}^2}{1+\lambda\eta_{k,T}} w_k^2$ with $w_k \stackrel{D}{=} i.i.d.N(0,1)$. We want to show that

$$\sup_{d\geq 0} h_T(d) \Rightarrow \sup_{d\geq 0} h(d) =: \sup_{d\geq 0}\{\bar{d}_0\bar{\eta} + \bar{g}_3(d) - \bar{g}_2(d)\}, \tag{A.9}$$

where $\bar{g}_2(d)$ and $\bar{\eta}$ are defined as in Theorem 4.3 and $\bar{g}_3(d)$ is the limit of $\sum_{k=1}^K \frac{(d-\bar{d}_0)(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}}$. Note that $\bar{g}_3(d)$ exists because the term $\sum_{k=1}^K \frac{(d-\bar{d}_0)(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}}$ is bounded by $-\bar{d}_0\bar{\eta}$ and $\bar{g}_1(d)$. To prove equation (A.9), we first establish the finite dimensional convergence of $h_T(d)$ to $h(d)$. For every fixed d, we could simplify $N_T'\left\{d(KT^2)^{-1}\right\}$ as

$$N_T'\left\{d(KT^2)^{-1}\right\} = \bar{d}_0\sum_{k=1}^K (KT^2)^{-1}\eta_{k,T}w_k^2 + \sum_{k=1}^K \frac{(d-\bar{d}_0)(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}} w_k^2.$$

Apply the same technique as Lemma A.2, we could show that $\sum_{k=1}^K (KT^2)^{-1}\eta_{k,T}(w_k^2 - 1) \Rightarrow 0$ and thus $\sum_{k=1}^K (KT^2)^{-1}\eta_{k,T}w_k^2 \Rightarrow \bar{\eta}$. Similarly, $\sum_{k=1}^K \frac{(d-\bar{d}_0)(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}} w_k^2$ converges to $\bar{g}_3(d)$ for every fixed d. Therefore, $N_T'\left(d(KT^2)^{-1}\right)$ does so to $\bar{d}_0\bar{\eta} + \bar{g}_3(d)$.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

Note that $T^{-1}D_T\left(d(KT^2)^{-1}\right)$ converges to 1 and thus $T\log\{1+\frac{N'_T(d(KT^2)^{-1})}{D_T(d(KT^2)^{-1})}\}$ converges to $\bar{d}_0\bar{\eta}+\bar{g}_3(d)$ for every fixed d. Since $\sum_{k=1}^K \log\{1+d(KT^2)^{-1}\xi_{k,T}\}$ converges to $\bar{g}_2(d)$ for every fixed d, so does $h_T(d)$ to $h(d)$. Similar to Theorem 4.3, we could further show $h_T(d)$ weakly converges to $h(d)$ and a continuous mapping theorem type result holds. Thus equation (A.9) holds.

Next we claim that $\sup_{d\geq 0}h(d)=\max_{d\in[0,\bar{d}_0]}\{\bar{d}_0\bar{\eta}+\bar{g}_3(d)-\bar{g}_2(d)\}$. Note that the first derivative of the partial sum induced by $\bar{g}_3(d)-\bar{g}_2(d)$ satisfies

$$\begin{aligned} l'_T(d) &= \sum_{k=1}^K \frac{\{1+d_0(KT^2)^{-1}\eta_{k,T}\}(KT^2)^{-1}\eta_{k,T}}{\{1+d(KT^2)^{-1}\eta_{k,T}\}^2} - \sum_{k=1}^K \frac{(KT^2)^{-1}\xi_{k,T}}{1+d(KT^2)^{-1}\xi_{k,T}} \\ &\leq \sum_{k=1}^K \frac{(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}} - \sum_{k=1}^K \frac{(KT^2)^{-1}\xi_{k,T}}{1+d(KT^2)^{-1}\xi_{k,T}} \\ &= \sum_{k=1}^K \frac{(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}} - \sum_{k=1}^K \frac{(KT^2)^{-1}\xi_{k,T}}{1+d(KT^2)^{-1}\xi_{k,T}}, \end{aligned}$$

where the inequality holds when $d\geq\bar{d}_0$. Note that $\sum_{k=1}^K \frac{(KT^2)^{-1}\eta_{k,T}}{1+d(KT^2)^{-1}\eta_{k,T}} \leq \sum_{k=1}^K \frac{(KT^2)^{-1}\xi_{k,T}}{1+d(KT^2)^{-1}\xi_{k,T}}$. When $d\geq\bar{d}_0$, $l'_T(d)\leq 0$ and hence $h(d)\leq h(\bar{d}_0)$ for $d\geq\bar{d}_0$. Thus our claim holds.

Finally, we consider the term $2\log L(0)-2\log L^0$. Under H_{02}, $\theta(z)=\theta_0+T^{-4/5}\theta_1(z)$. Recall that $\bar{b}$ is the spline coefficient vector of $\theta_1(z)$ and $\mathrm{var}(T^{-4/5}\tilde{b})=\bar{d}_0(KT^2)^{-1}\sigma^2\Lambda^{-1}$. Since $W^T\tilde{Y}=T^{-4/5}W^T\tilde{A}_2\bar{b}+W^T\tilde{u}$, we have $W^T\tilde{Y}/\sigma\overset{D}{=}N\left(0_{T-2},I_{T-2}+\bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^TW\right)$. Note that

$$\begin{aligned} &\frac{\tilde{Y}^TW_0W_0^T\tilde{Y}-\tilde{Y}^TWW^T\tilde{Y}}{\sigma^2} \\ &\quad=\frac{(T^{-4/5}\tilde{A}_2\bar{b}+\tilde{u})^T\tilde{A}^c_{01}(\tilde{A}^c_{01})^T(T^{-4/5}\tilde{A}_2\bar{b}+\tilde{u})}{\sigma^2}\overset{D}{=}\sum_{i=1}^T\phi_{i4}w_i^2, \end{aligned}$$

where $w_i\overset{D}{=}i.i.d.N(0,1)$ and ϕ_{i4} is the ith eigenvalue of

$$\Delta_4=\{I_T+\bar{d}_0(KT^2)^{-1}\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\}^{1/2}\tilde{A}^c_{01}(\tilde{A}^c_{01})^T\{I_T+\bar{d}_0(KT^2)^{-1}\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\}^{1/2}.$$

Note that Δ_4 shares the same nonzero eigenvalue as

$$(\tilde{A}^c_{01})^T\{I_T+\bar{d}_0(KT^2)^{-1}\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\}\tilde{A}^c_{01}=1+\bar{d}_0(KT^2)^{-1}(\tilde{A}^c_{01})^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\tilde{A}^c_{01}.$$

Hence only the largest eigenvalue of Δ_4 is positive. It reduces to show that the limit of the term $(KT^2)^{-1}(\tilde{A}^c_{01})^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\tilde{A}^c_{01}$ exists. Note that each element in $T^{-2}A_2^TA_{01}(A^c_{01})^TA_2$ converges. Since $\Sigma_u^{-1/2}$ and $\Lambda^{-1/2}$ are bounded deterministic matrices, each element of $T^{-2}\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}^c_{01}(\tilde{A}^c_{01})^T\tilde{A}_2\Lambda^{-1/2}$ converges and $K^{-1}T^{-2}tr$

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

$\{\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1/2}\}$ converges. By the fact that $(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^c\tilde{A}_{01}^c = tr\{\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1/2}\}$, the limit of $(KT^2)^{-1}(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\tilde{A}_{01}^c$ exists and we could denote it as ϱ. Therefore,

$$\sigma^{-2}(\tilde{Y}^T W_0 W_0^T\tilde{Y} - \tilde{Y}^T W W^T\tilde{Y}) \stackrel{D}{=} (1+\bar{d}_0\varrho)w_1^2, \tag{A.10}$$

with $w_1 \stackrel{D}{=} N(0,1)$. It yields $2\log L(0) - 2\log L^0 \Rightarrow (1+\bar{d}_0\varrho)w_1^2$ and hence Theorem 4.4 is proved. ■

B. Tables

TABLE 1. The empirical size of the proposed LRT when Σ_u is known

	Finite dist.					Asymptotic dist.				
	20%	15%	10%	5%	1%	20%	15%	10%	5%	1%
Panel A1: $T = 100$, $\rho = 0$										
$K = 10$	0.1915	0.1445	0.0930	0.0430	0.0095	0.1995	0.1490	0.0995	0.0465	0.0100
$K = 20$	0.1940	0.1425	0.0935	0.0425	0.0100	0.1985	0.1480	0.0990	0.0465	0.0100
$K = 40$	0.1920	0.1425	0.0925	0.0415	0.0085	0.1985	0.1480	0.0990	0.0465	0.0100
Panel A2: $T = 100$, $\rho = 0.5$										
$K = 10$	0.1940	0.1435	0.0975	0.0550	0.0095	0.1975	0.1495	0.1025	0.0585	0.0105
$K = 20$	0.1930	0.1440	0.0990	0.0550	0.0095	0.1970	0.1495	0.1035	0.0580	0.0105
$K = 40$	0.1945	0.1435	0.0975	0.0555	0.0090	0.1985	0.1500	0.1035	0.0585	0.0105
Panel B1: $T = 300$, $\rho = 0$										
$K = 20$	0.2035	0.1560	0.0975	0.0540	0.0160	0.2050	0.1565	0.0995	0.0535	0.0165
$K = 40$	0.2035	0.1550	0.0990	0.0530	0.0160	0.2050	0.1565	0.0995	0.0535	0.0165
$K = 80$	0.2025	0.1560	0.0980	0.0530	0.0160	0.2045	0.1560	0.0995	0.0535	0.0165
Panel B2: $T = 300$, $\rho = 0.5$										
$K = 20$	0.1985	0.1550	0.0975	0.0485	0.0105	0.1995	0.1555	0.1010	0.0495	0.0110
$K = 40$	0.1985	0.1520	0.1000	0.0485	0.0120	0.1980	0.1550	0.1005	0.0490	0.0110
$K = 80$	0.1975	0.1540	0.0975	0.0485	0.0105	0.1975	0.1550	0.1005	0.0495	0.0110

Note: The model is $y_t = \theta(z_t)x_t + u_t$ with $x_t = x_{t-1} + v_t$ and $u_t = \rho u_{t-1} + \varepsilon_t$, where v_t's and ε_t's are *i.i.d.* $N(0,1)$ and are independent with each other. The initial values are set to be zero. In particular, $\theta(z_t) = 0.25$ and the true covariance matrix is used. The rejection frequencies are calculated based on 2,000 replications.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

TABLE 2. The empirical size of the proposed LRT when Σ_u is unknown

	Finite dist.					Asymptotic dist.				
	20%	15%	10%	5%	1%	20%	15%	10%	5%	1%
Panel A1: $T = 100$, $\rho = 0$										
$K = 10$	0.2170	0.1595	0.1065	0.0500	0.0150	0.2245	0.1635	0.1120	0.0565	0.0155
$K = 20$	0.2190	0.1610	0.1075	0.0530	0.0125	0.2245	0.1630	0.1115	0.0565	0.0150
$K = 40$	0.2190	0.1585	0.1060	0.0530	0.0135	0.2245	0.1630	0.1115	0.0565	0.0150
Panel A2: $T = 100$, $\rho = 0.5$										
$K = 10$	0.2065	0.1565	0.1120	0.0635	0.0125	0.2110	0.1610	0.1155	0.0650	0.0145
$K = 20$	0.2045	0.1545	0.1080	0.0630	0.0130	0.2105	0.1610	0.1135	0.0650	0.0145
$K = 40$	0.2060	0.1555	0.1105	0.0625	0.0125	0.2105	0.1605	0.1135	0.0650	0.0145
Panel B1: $T = 300$, $\rho = 0$										
$K = 20$	0.2110	0.1640	0.1075	0.0575	0.0165	0.2125	0.1645	0.1075	0.0585	0.0175
$K = 40$	0.2110	0.1635	0.1065	0.0570	0.0170	0.2125	0.1645	0.1075	0.0585	0.0175
$K = 80$	0.2105	0.1635	0.1060	0.0565	0.0165	0.2120	0.1640	0.1075	0.0585	0.0175
Panel B2: $T = 300$, $\rho = 0.5$										
$K = 20$	0.2075	0.1550	0.1000	0.0500	0.0115	0.2090	0.1595	0.1025	0.0500	0.0115
$K = 40$	0.2065	0.1555	0.0995	0.0480	0.0115	0.2075	0.1590	0.1020	0.0500	0.0110
$K = 80$	0.2065	0.1565	0.1000	0.0485	0.0105	0.2070	0.1590	0.1020	0.0500	0.0110

Note: The model is $y_t = \theta(z_t)x_t + u_t$ with $x_t = x_{t-1} + v_t$ and $u_t = \rho u_{t-1} + \varepsilon_t$, where v_t's and ε_t's are $i.i.d.$ $N(0, 1)$ and are independent with each other. The initial values are set to be zero. In particular, $\theta(z_t) = 0.25$. The true covariance matrix is unknown and is replaced by its estimate. The rejection frequencies are calculated based on 2,000 replications.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press

TABLE 3. The power of the proposed LRT

	Finite dist.		Asymptotic dist.	
	Σ_u is known	Σ_u is unknown	Σ_u is known	Σ_u is unknown
Panel A1: $T = 100$, $\rho = 0$				
$K = 10$	0.6220 (0.5255)	0.6330 (0.5405)	0.6255 (0.5300)	0.6365 (0.5455)
$K = 20$	0.6175 (0.5195)	0.6275 (0.5325)	0.6210 (0.5240)	0.6330 (0.5390)
$K = 40$	0.6140 (0.5150)	0.6265 (0.5290)	0.6195 (0.5205)	0.6295 (0.5350)
Panel A2: $T = 100$, $\rho = 0.5$				
$K = 10$	0.7085 (0.6050)	0.6930 (0.5905)	0.7125 (0.6110)	0.6975 (0.5960)
$K = 20$	0.7050 (0.6010)	0.6855 (0.5890)	0.7080 (0.6055)	0.6890 (0.5925)
$K = 40$	0.7025 (0.5960)	0.6855 (0.5875)	0.7055 (0.6025)	0.6875 (0.5900)
Panel B1: $T = 300$, $\rho = 0$				
$K = 20$	0.9920 (0.9855)	0.9915 (0.9855)	0.9920 (0.9855)	0.9915 (0.9850)
$K = 40$	0.9915 (0.9855)	0.9905 (0.9850)	0.9915 (0.9855)	0.9910 (0.9850)
$K = 80$	0.9910 (0.9855)	0.9910 (0.9850)	0.9910 (0.9855)	0.9910 (0.9850)
Panel B2: $T = 300$, $\rho = 0$				
$K = 20$	0.994 (0.99)	0.9935 (0.9910)	0.994 (0.99)	0.9940 (0.9910)
$K = 40$	0.994 (0.99)	0.9940 (0.9905)	0.994 (0.99)	0.9935 (0.9910)
$K = 80$	0.994 (0.99)	0.9935 (0.9905)	0.994 (0.99)	0.9935 (0.9910)

Note: The model is $y_t = \theta(z_t)x_t + u_t$ with $x_t = x_{t-1} + v_t$ and $u_t = \rho u_{t-1} + \varepsilon_t$, where v_t's and ε_t's are *i.i.d.* $N(0, 1)$ and are independent with each other. The initial values are set to be zero. In particular, $\theta(z_t) = (z_t - 0.5)^2$. The rejection frequencies are calculated using the critical values associated with $\alpha = 0.05$ or $\alpha = 0.01$ as indicated inside the parenthesis. All results are based on 2,000 replications.

https://doi.org/10.1017/S0266466614000541 Published online by Cambridge University Press