Chen, Haiqiang; Fang, Ying; Li, Yingxing

Working Paper

Estimation and inference for varying-coeffcient models with nonstationary regressors using penalized splines

SFB 649 Discussion Paper, No. 2013-033

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Chen, Haiqiang; Fang, Ying; Li, Yingxing (2013) : Estimation and inference for varying-coeffcient models with nonstationary regressors using penalized splines, SFB 649 Discussion Paper, No. 2013-033, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/79632

SFB 649 Discussion Paper 2013-033

Estimation and Inference for Varying-coefficient Models with Nonstationary Regressors using Penalized Splines

Haiqiang Chen*
Ying Fang*
Yingxing Li*

* Xiamen University, China

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Estimation and Inference for Varying-coefficient Models with Nonstationary Regressors using Penalized Splines *

Haiqiang Chen, Ying Fang and Yingxing Li

Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics, and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China

12th July 2013

Abstract

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.

JEL Classification: C12, C14, C22

Keywords: Nonstationary Time Series; Varying-coefficient Model; Likelihood Ratio Test; Penalized Splines

*A previous version of this paper was presented in the 3rd WISE-Humboldt Workshop in Nonparametric Nonstationary High-dimensional Econometrics in May 2012. The authors are grateful to Zongwu Cai, Jiti Gao, Wolfgang Härdle, Yongmiao Hong and all workshop participants for their valuable comments and suggestions. They also acknowledge the financial support received from Chinese National Science Foundation with grant numbers 71201137, 71271179, 71131008 and 11201390. This research was also partially supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

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.

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 instable.

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.

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 basis func-

[1] Other contributions to the literature on the stability tests in cointegrating regression includes Hansen (1992), Hao (1996), Quintos (1997), Kuo (1998), Hansen and Johansen (1999), Johansen, Mosconi and Nielsen (2000), Harris, McCabe and Leybourne (2002), and 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 a nonparametric time series model with nonstationary variables.

tions 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 our best 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 non-standard, 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.

In matters of notations, $\stackrel{D}{=}$ denotes equality in distribution, =: denotes definition, $\Rightarrow$ denotes convergence in distribution, $\stackrel{a.s.}{\rightarrow}$ denotes almost sure convergence, $a \sim b$ denotes that a and b has 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

$$y_t = x_t\theta(z_t) + u_t, \tag{1}$$

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

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

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.[2]

[2]To save notations, we only consider the simple case when z_t is univariate.

Compared to traditional varying-coefficient models, which usually deals with independent and identically distributed (iid) or stationary time series, Model (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 (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 (1).

Assumption 1: The sequence $\{v_t\}$ is stationary $\alpha - mixing$ with $E(|v_t|^\rho) < \infty$ 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 $N(0, \sigma_e^2)$ 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.

Assumption 3:

i) The sequence $\{z_t\}$ is strictly stationary, ergodic and $\alpha - mixing$ 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 $\int_0^1 \{\theta^{(m)}(z)\}^2 dz < \infty$.

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

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

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), 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) establish conditions of strong approximations for a general linear process u_t and Theorem 4.1 in Shao and Lu (1987) give conditions of strong approximations for an $\alpha - mixing$ 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 $u = (u_1, ..., u_T)^T$ 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, though 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 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 $k = 0, \cdots, K$. 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:

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

In a discrete version, this could be written as

$$\sum_{t=1}^{T} \left(y_t - x_t \sum_{k=1}^{K+p} \beta_k \Psi_k^{[p]}(z) \right)^2 + \tilde{\lambda}^{-1} K^{-1} \sum_{k=m+1}^{K} (\frac{\Delta^m \beta_k}{K^{-m}})^2, \tag{3}$$

where Δ is the differencing operator such that $\Delta\beta_k = \beta_k - \beta_{k-1}$, m is a positive integer indicating the order of differencing with $\Delta^m = \Delta(\Delta^{m-1})$. Let $Y = (y_1, \cdots, y_T)^T$ and X be the diagonal matrix whose (i,i)th element is x_i, and Ψ be the matrix[3] 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 and estimate $\hat{\beta}$ by minimizing

$$(Y - X\Psi\beta)^T\Sigma_u^{-1}(Y - X\Psi\beta) + \tilde{\lambda}^{-1}K^{2m-1}\beta^T D_m^T D_m\beta. \tag{4}$$

A direct calculation shows that the solution to (4) is

$$\hat{\beta} = (\Psi^T X\Sigma_u^{-1}X\Psi + \tilde{\lambda}^{-1}K^{2m-1}D_m^T D_m)^{-1}\Psi^T X\Sigma_u^{-1}Y. \tag{5}$$

Then the penalized spline estimator of $\theta(z)$ for model (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 its 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 discussed 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.

[3]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.

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* $K \sim T^{r_1}$ *with* $\frac{2m}{2m+1} < r_1 < 1$.

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*

$$\hat{\theta}(z) - \theta(z) = O_p(T^{-1}\tilde{\lambda}^{-1/2}) + O_p(T^{-1+1/(2m)}\tilde{\lambda}^{1/(4m)}) + O_p(K^{-1}). \tag{6}$$

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 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).

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

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

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 exists 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[4] 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 (4) for TPS as

$$(Y - XZ_1\gamma - XZ_2 b)^T \Sigma_u^{-1} (Y - XZ_1\gamma - XZ_2 b) + \tilde{\lambda}^{-1} K^{2m-2p-1} b^T \Lambda b, \tag{7}$$

where Z_1 and Z_2 are matrices whose i-th row are $(1, z, \cdots, z^p)$ and $\{(z-\kappa_1)_+^p, \cdots, (z-\kappa_K)_+^p\}$ respectively, and $\gamma = (\gamma_0, \cdots, \gamma_p)^T$. 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

[4]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.

I_K. If we denote $\lambda^{-1} = \tilde{\lambda}^{-1}K$, then the minimization criterion (7) could be written as

$$(Y - XZ_1\gamma - XZ_2b)^T\Sigma_u^{-1}(Y - XZ_1\gamma - XZ_2b) + \lambda^{-1}b^Tb, \tag{8}$$

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

4 Inference using Likelihood Ratio Tests

In this section, we consider testing whether the functional coefficients $\theta(z)$ is 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. Since 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.

Note that minimizing (7) 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 equations, which motivates us to utilize the mixed model representation to obtain $\hat{\gamma}$ and $\hat{b}$ as the best linear unbiased predictors (BLUP) 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 $E(Y) = A_1\gamma$ and $var(Y) = \sigma^2(\Sigma_u + \lambda A_2\Lambda^{-1}A_2^T) =: \sigma^2\Omega_\lambda$. Hence we could define

a twice of the log-likelihood of Y as

$$2l(\gamma, \lambda, \sigma^2) = -(Y - A_1\gamma)(\sigma^2\Omega_\lambda)^{-1}(Y - A_1\gamma) - \log|\sigma^2\Omega_\lambda| - T\log(2\pi). \tag{9}$$

By maximizing (9), 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 BLUP of γ and b are then obtained as

$$\hat{\gamma} = (A_1^T\hat{\Omega}_\lambda^{-1}A_1)^{-1}A_1^T\hat{\Omega}_\lambda^{-1}Y, \qquad \text{and} \qquad \hat{b} = \hat{\lambda}\Lambda^{-1}A_2^T\hat{\Omega}_\lambda^{-1}(Y - A_1\hat{\gamma}), \tag{10}$$

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 equation (7) yields the same solution as (10). 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

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

Note that the null distribution of the LRT statistic is not standard as the parameter λ is always non-negative 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 equi-spaced knots are used. Then under* $H_0 : \theta(z) \equiv \theta_0$,

$$LRT_T \stackrel{D}{=} \sup_{\lambda \geq 0}[T \log\{1 + \frac{N_T(\lambda)}{D_T(\lambda)}\} - \sum_{s=1}^{K} \log(1 + \lambda\xi_{s,T})] + T \log\{1 + \frac{w_{T-1}^2}{\sum_{s=1}^{T-2} w_s^2}\}, \tag{11}$$

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* $w_s \stackrel{D}{=} iidN(0,1)$.

Theorem 4.1 derives the exact null distribution of the LRT statistic when the sample size T is finite. Although equation (11) 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 $w_1^2, \cdots, w_K^2$.

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 equi-spaced knots are used. Then there exist* η_s *and* ξ_s, *for* $s = 1, \cdots, K$, *such that*

$$T^{-2}\eta_{s,T} \Rightarrow \eta_s, \qquad T^{-2}\xi_{s,T} \Rightarrow \xi_s, \qquad \text{as} \quad T \to \infty. \tag{12}$$

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

$$LRT_T \Rightarrow \sup_{d \geq 0}\{\sum_{s=1}^{K} \frac{d\eta_s}{1 + d\eta_s} w_s^2 - \sum_{s=1}^{K} \log(1 + d\xi_s)\} + \chi_1^2. \tag{13}$$

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 $\sup_{d\geq 0}\{\sum_{s=1}^{K} \frac{d\eta_s}{1+d\eta_s} w_s^2 - \sum_{s=1}^{K} \log(1+d\xi_s)\}$. 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 equi-spaced knots are 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*

$$K^{-1}T^{-2}\sum_{k=1}^{K}\eta_{k,T} \Rightarrow \bar{\eta}, \qquad K^{-1}T^{-2}\sum_{k=1}^{K}\xi_{k,T} \Rightarrow \bar{\xi}. \tag{14}$$

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

$$LRT_T \Rightarrow \chi_1^2. \tag{15}$$

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 $\bar{b} = (\bar{b}_1, \cdots, \bar{b}_K)^T$ 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 test 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 $H_{02} : \theta(z) = \theta_0 + T^{-4/5}\theta_1(z)$.

THEOREM 4.4. *Suppose that Assumptions 1-3 hold and the linear TPS with equi-spaced knots are 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^2_1, 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 (52).

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 (53) and ϱ is defined right before equation (55).

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 the 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^2_1. 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^2_1 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 test. 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 iid $N(0,1)$, 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)* $\theta(z_t) = 0.25, \rho = 0$; *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 2000 replications.

Table 1 report 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 conclusions 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 the 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 increase 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 heated interests 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 coefficient. 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 non-standard, but could be simulated via a proposed algorithm using 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 be suffered from the mis-selection 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.

Appendix A: Proofs

Proof of Theorem 3.1: Note that our model could be written as $Y = X\Psi\beta + u$. If $\Sigma_u \neq I_T$, we could always multiply $\Sigma_u^{-1/2}$ and consider instead $\tilde{Y} = \tilde{X}\Psi\beta + \tilde{u}$, where $\tilde{Y} = \Sigma_u^{-1/2}Y$, $\tilde{X} = \Sigma_u^{-1/2}X$ is an integrated process and the elements of $\tilde{u} = \Sigma_u^{-1/2}u$ are uncorrelated. Hence without loss of generality, we only need to show that equation (6) holds when u_t's are uncorrelated, i.e. Σ_u is the identity matrix.

Recall that the penalized spline estimator $\hat{\theta}(z)$ could be written as in equation (5), i.e.

$$\hat{\theta}(z) \quad = \quad \Psi_z(\Psi^T X^2 \Psi + \tilde{\lambda}^{-1} K^{2m-1} D_m^T D_m)^{-1} \Psi^T XY, \tag{16}$$

where $\Psi_z = \{\Psi_1^{[p]}(z), \cdots, \Psi_{K+p}^{[p]}(z)\}$. First consider the (i,j)th element of the term $\Psi^T X^2 \Psi$. Define $R_0 =: T^{-2}\sum_{t=1}^T x_t^2[\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t) - E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}]$. By subtracting and adding the mean, we have,

$$T^{-2}(\Psi^T X^2 \Psi)_{i,j} = T^{-2}\sum_{t=1}^T x_t^2 \Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t) = R_0 + T^{-2}\sum_{t=1}^T E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}x_t^2.$$

Recall that $\Psi_i^{[p]}(z)$ is nonzero only in a small interval of length $(p+1)/K$. For example, when zero degree splines are used, $\Psi_i^{[0]}(z)$ is the indicator function $I_{(i-1)/K<z\leq i/K}$. Hence $E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\} = O(K^{-1})$ and $var\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\} = O(K^{-1})$. Moreover,

$$T^{-2}\sum_{t=1}^T E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}x_t^2 = E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}T^{-1}\sum_{t=1}^T(\frac{x_t}{\sqrt{T}})^2 = O_p(K^{-1}),$$

and hence

$$R_0 \quad = \quad T^{-1}\sum_{t=1}^T\left(\frac{x_t}{\sqrt{T}}\right)^2[\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t) - E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}] = O_p\{(TK)^{-1/2}\} = o_p(K^{-1}),$$

where the last equality holds as $K = o(T)$. Denote q_{ij} as the limit of $KE\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\}$. Then

$$KT^{-2}(\Psi^T X X \Psi)_{ij} = K[T^{-2}\sum_{t=1}^{T} x_t^2 E\{\Psi_i^{[p]}(z_t)\Psi_j^{[p]}(z_t)\} + R_0] \Rightarrow q_{ij}\int B_v^2(s)ds. \quad (17)$$

Therefore,

$$\begin{aligned}
\hat{\theta}(z) &= \Psi_z(\Psi^T X^2 \Psi + \tilde{\lambda}^{-1}K^{2m-1}D_m^T D_m)^{-1}\Psi^T XY, \\
&= \Psi_z(\Psi^T X^2 \Psi + \tilde{\lambda}^{-1}K^{2m-1}D_m^T D_m)^{-1}\Psi^T X^2\Psi\beta \\
&\quad +\Psi_z(\Psi^T X^2 \Psi + \tilde{\lambda}^{-1}K^{2m-1}D_m^T D_m)^{-1}\Psi^T Xu + O_p(K^{-1}) \\
&=: R_1 + R_2 + O_p(K^{-1}), \qquad (18)
\end{aligned}$$

where the term $O_p(K^{-1})$ comes from the bias due to splines approximation,

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

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

and $\bar{\lambda} = \tilde{\lambda}^{-1}/T^2$. By equation (17), the term $\frac{K}{T^2}\Psi^T X^2\Psi$ converges. Using a similar technique as in Li and Ruppert (2008), we can show that the term $\bar{\lambda}^{1/(2m)}$ serves equivalently as the bandwidth h used in a Nadaraya-Watson kernel estimator. Therefore,

$$R_1 - \Psi_z\beta = O_p(h^m) = O_p\{\bar{\lambda}^{m/(2m)}\} = O_p(T^{-1}\tilde{\lambda}^{-1/2}). \quad (19)$$

Now consider the second term R_2. Note that $ER_2 = 0$ and the i-th element of $\Psi^T Xu$ satisfies

$$\frac{\Psi^T(T^{-1/2}Xu)_i}{T/K} = \frac{\sum_{t=1}^{T}\{\Psi_i^{[p]}(z_t)\frac{x_t}{\sqrt{T}}u_t\}}{T/K} = O_p(\frac{1}{\sqrt{T/K}}). \quad (20)$$

By the fact that $\bar{\lambda} = \tilde{\lambda}^{-1}/T^2$ and $\bar{\lambda}^{1/(2m)}$ serves as the bandwidth used in Nadaraya-Watson kernel estimate, we have

$$var(\sqrt{T}R_2) = O(\frac{1}{Kh}\frac{1}{T/K}) = O(\frac{1}{Th}) = O(\frac{1}{T\bar{\lambda}^{1/(2m)}}) = O(T^{-1+1/m}\tilde{\lambda}^{1/(2m)}),$$

and hence

$$R_2 = \frac{1}{\sqrt{T}} O_p(T^{-1/2+1/(2m)}\tilde{\lambda}^{1/(4m)}) = O_p(T^{-1+1/(2m)}\tilde{\lambda}^{1/(4m)}). \tag{21}$$

Together with equation (18), (19) and (21),

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

and Theorem 3.1 holds. □

Note that the proofs of Theorem 4.1–4.4 share lots of similarity. Thus we first provide a general description and four useful propositions which could be applied to all these theorems.

Recall that twice of the log-likelihood of Y is defined as in equation (9), i.e.

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

where $\Omega_\lambda = \Sigma_u + \lambda A_2 \Lambda^{-1} A_2^T$ and $\sigma^2\Omega_\lambda = var(Y)$. Instead of maximizing $2l(\gamma, \lambda, \sigma^2)$ over the parameter space $(\gamma, \lambda, \sigma^2)$, we consider maximizing the profile log-likelihood function $2\log L(\lambda) = 2l(\hat{\gamma}_\lambda, \lambda, \hat{\sigma}_\lambda^2)$ over the parameter space $\lambda \geq 0$, where $\hat{\gamma}_\lambda$ and $\hat{\sigma}_\lambda^2$ are the profile maximum likelihood estimates (MLE) for γ and σ^2 when λ is given. Specifically, they satisfy

$$\hat{\gamma}_\lambda = (A_1^T\Omega_\lambda^{-1}A_1)^{-1}A_1^T\Omega_\lambda^{-1}Y \qquad \text{and} \qquad \hat{\sigma}_\lambda^2 = T^{-1}(Y - A_1\hat{\gamma}_\lambda)^T\Omega_\lambda^{-1}(Y - A_1\hat{\gamma}_\lambda).$$

By plugging $\hat{\gamma}_\lambda$ and $\hat{\sigma}_\lambda^2$ into equation (9), we could simplify the profile log-likelihood as

$$2\log L(\lambda) = -T - \log|\hat{\sigma}_\lambda^2\Omega_\lambda| - T\log(2\pi). \tag{22}$$

Denote $\log L^0$ as the maximum log-likelihood under the null hypothesis. Then we can decompose $2LRT_T$ into two parts by adding and subtracting $2\log L(0)$, i.e.

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

where the first part corresponds to testing $\lambda = 0$ and the second part corresponds to testing the linear coefficient $\gamma_1 = 0$ given that $\lambda = 0$. The following propositions establish some

useful preliminary results. In particular, Proposition 6.1–Proposition 6.3 studies the property related to the first part, while Proposition 6.4 studies the property of the second part.

PROPOSITION 6.1. *Denote* $\tilde{Y} = \Sigma_u^{-1/2} Y$, $\tilde{A}_1 = \Sigma_u^{-1/2} A_1$, $\tilde{A}_2 = \Sigma_u^{-1/2} A_2$. *We have*

$$2\log L(\lambda) - 2\log L(0) = T\log(1 + \frac{\lambda R_3/\sigma^2}{T\hat{\sigma}_\lambda^2/\sigma^2}) - \sum_{k=1}^{K} \log(1 + \lambda\xi_{k,T}). \tag{23}$$

where R_3 *is given right below equation (29), and* $\xi_{k,T}$*'s are defined right before Theorem 4.1.*

Proof of Proposition 6.1: Define $V_\lambda =: I_T + \lambda\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T$. With the notations of $\tilde{Y}$, $\tilde{A}_1$ and $\tilde{A}_2$, we have $\Omega_\lambda = \Sigma_u^{1/2} V_\lambda \Sigma_u^{1/2}$. Correspondingly, the profile MLE can be rewritten as

$$\hat{\gamma}_\lambda = (\tilde{A}_1^T V_\lambda^{-1} \tilde{A}_1)^{-1} \tilde{A}_1 V_\lambda^{-1} \tilde{Y} \text{ and } \hat{\sigma}_\lambda^2 = T^{-1}(\tilde{Y} - \tilde{A}_1\hat{\gamma}_\lambda)^T V_\lambda^{-1}(\tilde{Y} - \tilde{A}_1\hat{\gamma}_\lambda).$$

By Patterson and Thompson (1971), it is well-known that there exists a matrix W satisfying

$$WW^T = P = I_T - \tilde{A}_1(\tilde{A}_1^T\tilde{A}_1)^{-1}\tilde{A}_1^T, \qquad W^T W = I_{T-2}, \tag{24}$$

and that

$$T\hat{\sigma}_\lambda^2 = (\tilde{Y} - \tilde{A}_1\hat{\gamma}_\lambda)^T V_\lambda^{-1}(\tilde{Y} - \tilde{A}_1\hat{\gamma}_\lambda) = \tilde{Y}^T W(W^T V_\lambda W)^{-1} W^T \tilde{Y}.$$

By equation (22) and that $|\Omega_\lambda| = |\Sigma_u^{1/2}||V_\lambda||\Sigma_u^{1/2}| = |V_\lambda||\Sigma_u|$, we have

$$\begin{aligned} -2\log L(\lambda) &= \log|\hat{\sigma}_\lambda^2\Omega_\lambda| + T\{1 + \log(2\pi)\} \\ &= T\log(\hat{\sigma}_\lambda^2) + \log|V_\lambda| + \log|\Sigma_u| + T\{1 + \log(2\pi)\} \\ &= T\log(\frac{T\hat{\sigma}_\lambda^2}{\sigma^2}) + \log|V_\lambda| + C_0. \end{aligned} \tag{25}$$

where $C_0 = \log|\Sigma_u| + T\{1 + \log(2\pi)\} - T\log(T/\sigma^2)$.

Note that for any $p_1 \times p_2$ matrices A and B, it holds that $|I_{p_1} + A^T B| = |I_{p_2} + B^T A|$. Moreover, $|A|$ equals the product of its eigenvalues. Hence

$$\log|V_\lambda| = \log|I_T + \lambda\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T| = \log|I_K + \lambda\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_2\Lambda^{-1/2}| = \sum_{k=1}^{K}\log(1 + \lambda\xi_{k,T}).$$

Together with equation (25), we have

$$-2\log L(\lambda) = T\log(\frac{T\hat{\sigma}_\lambda^2}{\sigma^2}) + \sum_{k=1}^{K}\log(1 + \lambda\xi_{k,T}) + C_0. \tag{26}$$

When $\lambda = 0$, we have $V_0 = I_T$, $W^T V_0 W = W^T W = I_{T-2}$ and $T\hat{\sigma}_0^2 = \tilde{Y}^T W W^T \tilde{Y}$. Thus,

$$-2\log L(0) = T\log(\frac{\tilde{Y}^T W W^T \tilde{Y}}{\sigma^2}) + C_0. \tag{27}$$

It follows that

$$2\log L(\lambda) - 2\log L(0) \quad = \quad T\log\{1 + \frac{(\tilde{Y}^T W W^T \tilde{Y} - T\hat{\sigma}_\lambda^2)/\sigma^2}{T\hat{\sigma}_\lambda^2/\sigma^2}\} - \sum_{k=1}^{K}(1 + \lambda\xi_{k,T}). \tag{28}$$

Note that $(W^T V_\lambda W)^{-1} = (I_{T-2} + \lambda W^T \tilde{A}_2 \tilde{\Lambda}^{-1} A_2^T W)^{-1} = I_{T-2} - \lambda\Delta_1$, where, by Woodbury Matrix Identity,

$$\Delta_1 =: W^T \tilde{A}_2 \tilde{\Lambda}^{-1/2}(I_K + \lambda\Lambda^{-1/2}\tilde{A}_2^T W W^T \tilde{A}_2 \Lambda^{-1/2})^{-1}\tilde{\Lambda}^{-1/2}\tilde{A}_2^T W. \tag{29}$$

Hence $T\hat{\sigma}_\lambda^2 = \tilde{Y}^T W(W^T V_\lambda W)^{-1} W^T \tilde{Y} = \tilde{Y}^T W W^T \tilde{Y} - \lambda R_3$, where $R_3 = \tilde{Y}^T W \Delta_1 W^T \tilde{Y}$. Together with equation (28), we conclude that equation (23) holds. □

PROPOSITION 6.2. *For any number of knots K, the following results hold.*

Under H_0 or H_{01}, $\lambda R_3/\sigma^2 = \sum_{k=1}^{K} \frac{\lambda\eta_{k,T}}{1+\lambda\eta_{k,T}} w_{k1}^2 \stackrel{D}{=} N_T(\lambda)$;

Under H_{02}, $\lambda R_3/\sigma^2 = \lambda\sum_{k=1}^{K} \frac{\eta_{k,T}+\bar{d}_0(KT^2)^{-1}\eta_{k,T}^2}{1+\lambda\eta_{k,T}} w_{k2}^2 \stackrel{D}{=} N_T'(\lambda)$;

where $N_T(\lambda)$ is defined in Theorem 4.1, $N_T'(\lambda) =: \lambda\sum_{k=1}^{K} \frac{\eta_{k,T}+\bar{d}_0(KT^2)^{-1}\eta_{k,T}^2}{1+\lambda\eta_{k,T}} w_k^2$ with $w_k \stackrel{D}{=} iidN(0,1)$, and $\eta_{k,T}$'s are defined right before Theorem 4.1.

Proof of Proposition 6.2: Under H_0 or H_{01}, the spline coefficients are all 0. Therefore, $\tilde{Y} \stackrel{D}{=} N(\tilde{A}_1\gamma, \sigma^2 I_T)$. Recall that the matrix W satisfies equation (24). Hence

$$W^T\tilde{A}_1 = (W^T W)W^T\tilde{A}_1 = W^T\{I_T - \tilde{A}_1(\tilde{A}_1^T\tilde{A}_1)^{-1}\tilde{A}_1^T\}\tilde{A}_1 = \mathbf{0}_{(T-2)\times 2},$$

and $W^T\tilde{Y}/\sigma \stackrel{D}{=} N(0_{T-2}, I_{T-2})$, where 0_{T-2} is the $(T-2)\times 1$ vector whose components are 0. Recall that the matrix Δ_1 is defined as in equation (29). Suppose its eigen decomposition is $U_1 S_1 U_1$, where S_1 is the diagonal matrix whose (i,i)th element ϕ_{i1} is also the ith eigenvalue of Δ_1. Then

$$\frac{R_3}{\sigma^2} = (\frac{W^T\tilde{Y}}{\sigma})^T\Delta_1(\frac{W^T\tilde{Y}}{\sigma}) = (\frac{U_1^T W^T\tilde{Y}}{\sigma})^T S_1(\frac{U_1^T W^T\tilde{Y}}{\sigma}) = \sum_{i=1}^{T-2}\phi_{i1}w_{i1}^2, \tag{30}$$

where w_{i1}'s are the elements of $U_1^T W^T \tilde{Y}/\sigma$.

Now we show that the eigenvalues ϕ_{i1}'s satisfy $\phi_{i1} = \eta_{i,T}/(1+\lambda\eta_{i,T})$ for $= 1, \cdots, K$ and $\phi_{i1} = 0$ for $i > K$. Note that for any $p_1 \times p_2$ matrices A and B, AB and BA shares the same nonzero eigenvalues. Hence the nonzero eigenvalues of Δ_1 are the same as those of the matrix $(I_K + \lambda\Lambda^{-1/2}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2})^{-1}\Lambda^{-1/2}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2}$. Suppose the eigen decomposition of the matrix $\Lambda^{-1/2}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2}$ is $U_2 S_2 U_2^T$, where S_2 is the $K \times K$ diagonal matrix whose (k,k)th element is the kth eigenvalue $\eta_{k,T}$. Then

$$\begin{aligned} & (I_K + \lambda\Lambda^{-1/2}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2})^{-1}\Lambda^{-1/2}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2} \\ = \; & (U_2U_2^T + \lambda U_2 S_2 U_2^T)^{-1} U_2 S_2 U_2^T \\ = \; & U_2\{(I_K + \lambda S_2)^{-1} S_2\} U_2^T. \end{aligned}$$

Note that the (k,k)th element in the diagonal matrix $(I_K + \lambda S_2)^{-1})S_2$ is $\eta_{k,T}/(1+\lambda\eta_{k,T})$, which equals the first K eigenvalues ϕ_{k1}'s. Since the rest $T-2-K$ eigenvalues are 0, we have $R_3/\sigma^2 = \sum_{k=1}^K \frac{\eta_{k,T}}{1+\lambda\eta_{k,T}} w_{k1}^2$. Note that $W^T\tilde{Y}/\sigma \overset{D}{=} N(0_{T-2}, I_{T-2})$. Moreover, U_1 is an orthogonal matrix. Hence we conclude that $U_1^T W^T\tilde{Y}/\sigma \overset{D}{=} N(0_{T-2}, I_{T-2})$, i.e. $w_{k1} \overset{D}{=} iidN(0,1)$. Therefore, $\lambda R_3/\sigma^2 \overset{D}{=} N_T(\lambda)$.

Under H_{02}, $\theta(z) = \theta_0 + T^{-4/5}\theta_1(z)$, where $\theta_1(z)$ has an associated $\bar{\lambda}_0$ satisfying $\bar{\lambda}_0 = \bar{d}_0 T^{-2/5}K^{-1}$ for some positive constant $\bar{d}_0$. Let $\bar{b}$ be the spline coefficients of $\theta_1(z)$. Then $\text{var}(T^{-4/5}\bar{b}) = T^{-8/5}\bar{d}_0 T^{-2/5}K^{-1}\sigma^2\Lambda^{-1} = \bar{d}_0(KT^2)^{-1}\sigma^2\Lambda^{-1}$. Since $W^T\tilde{A}_1 = \mathbf{0}_{(T-2)\times 2}$, we have $W^T\tilde{Y} = T^{-4/5}W^T\tilde{A}_2\bar{b} + W^T\tilde{u}$. Equivalently, we could write that $W^T\tilde{Y}/\sigma \overset{D}{=} N\left(0_{T-2}, I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T W\right)$. Denote

$$\Delta_2 = \{I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T W\}^{1/2}\Delta_1\{I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T W\}^{1/2}.$$

Similar as equation (30), we conclude that $\frac{R_3}{\sigma^2} = \sum_{i=1}^{T-2}\phi_{i2}w_{i2}^2$, where ϕ_{i2} is the ith eigenvalue of Δ_2 and w_{i2} is the ith element of $U_1^T\{I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T W\}^{-1/2}W^T\tilde{Y}/\sigma$. Note the nonzero eigenvalues of Δ_2 are the same as those of

$$(I_K + \lambda\Lambda^{-1}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2})^{-1}\Lambda^{-1/2}\tilde{A}_2^T W\{I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T W\}W^T\tilde{A}_2\Lambda^{-1/2}.$$

With the eigen decomposition $\Lambda^{-1/2}\tilde{A}_2^T WW^T\tilde{A}_2\Lambda^{-1/2} = U_2S_2U_2^T$, the above matrix equals

$$\begin{aligned} & (I_K + \lambda U_2S_2U_2^T)^{-1}\{U_2S_2U_2^T + \bar{d}_0(KT^2)^{-1}U_2S_2U_2^TU_2S_2U_2^T\} \\ = \; & U_2(I_K + \lambda S_2)^{-1}\{S_2 + \bar{d}_0(KT^2)^{-1}S_2^2\}U_2^T. \end{aligned}$$

Since the (k,k)th element in the diagonal matrix $(I_K + \lambda S_2)^{-1}\{S_2 + \bar{d}_0(KT^2)^{-1}S_2^2\}$ is $\frac{\eta_{k,T}+\bar{d}_0(KT^2)^{-1}\eta_{k,T}^2}{1+\eta_{k,T}}$, we have $\frac{R_3}{\sigma^2} = \sum_{k=1}^K \frac{\eta_{k,T}+\bar{d}_0(KT^2)^{-1}\eta_{k,T}^2}{1+\lambda\eta_{k,T}}w_{k2}^2$. Note that $w_{k2} \overset{D}{=} iidN(0,1)$ because $var[U_1^T\{I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^TW\}^{-1/2}W^T\tilde{Y}] = \sigma^2I_{T-2}$ under H_{02}. Therefore, Proposition 6.2 is proved. □

PROPOSITION 6.3. *Assume $K = o(T)$. Then under H_0, H_{01} or H_{02}, $\tilde{Y}^TWW^T\tilde{Y}/T \Rightarrow \sigma$.*

Proof of Proposition 6.3: Under H_0 or H_{01}, $W^T\tilde{Y}/\sigma \overset{D}{=} N(0_{T-2}, I_{T-2})$. Therefore,

$$\frac{\tilde{Y}^TWW^T\tilde{Y}}{T\sigma^2} = \frac{T-2}{T} + o_p(1) = 1 + o_p(1).$$

Under H_{02} $W^T\tilde{Y}/\sigma \overset{D}{=} N\Big(0_{T-2}, I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^TW\Big)$. Therefore,

$$\tilde{Y}^TWW^T\tilde{Y}/(T\sigma^2) = T^{-1}\sum_{i=1}^{T-2}\phi'_{i2}w_{i2}^2 = T^{-1}\sum_{i=1}^{T-2}\phi'_{i2} + o_p(1), \tag{31}$$

where ϕ'_{i2} is the ith eigenvalue of the matrix $I_{T-2} + \bar{d}_0(KT^2)^{-1}W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^TW$ and w_{i2} is defined as in the proof of Proposition 6.2 and satisfying $w_{i2} \overset{D}{=} iidN(0,1)$. Since $W^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^TW$ and $\Lambda^{-1/2}\tilde{A}_2^TWW^T\tilde{A}_2\Lambda^{-1/2}$ share the same nonzero eigenvalues,

$$T^{-1}\sum_{i=1}^{T-2}\phi'_{i2} = \frac{T-2}{T} + T^{-1}\sum_{i=1}^{K}\bar{d}_0(KT^2)^{-1}\eta_{i,T} = 1 + o_p(1).$$

With equation (31), $\tilde{Y}^TWW^T\tilde{Y}/T \Rightarrow \sigma^2$. Hence, Proposition 6.3 is proved. □

PROPOSITION 6.4. *Let $\tilde{A}_{01}^c$ be defined right before equation (33). It holds that*

$$2\log L(0) - 2\log L^0 = T\log\{1 + \frac{\tilde{Y}^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{Y}/\sigma^2}{\tilde{Y}^TWW^T\tilde{Y}/\sigma^2}\}. \tag{32}$$

Proof of Proposition 6.4: By equation (27), $2\log L(0) = T\log(\tilde{Y}^T WW^T\tilde{Y}/\sigma^2) + C_0$, where W is the $T \times (T-2)$ matrix satisfying equation (24). Now we show that there exists a $T \times (T-1)$ matrix W_0 such that $2\log L^0 = T\log(\tilde{Y}^T W_0 W_0^T \tilde{Y}/\sigma^2) + C_0$.

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 $x_t z_t$'s. Define $\tilde{A}_{01} = \Sigma_u^{-1/2} A_{01}$ and $\tilde{A}_{r1} = \Sigma_u^{-1/2} A_{r1}$. Note that the maximum likelihood estimates associated with $\log L^0$ satisfies that

$$\hat{\gamma}_0 = (\tilde{A}_{01}^T \tilde{A}_{01})^{-1} \tilde{A}_{01}^T \tilde{Y}, \qquad \hat{\sigma}_0^2 = T^{-1}(\tilde{Y} - \tilde{A}_{01}\hat{\gamma}_0)^T(\tilde{Y} - \tilde{A}_{01}\hat{\gamma}_0).$$

Similar as equation (24), there exists W_0 such that $(\tilde{Y} - \tilde{A}_{01}\hat{\gamma}_0)^T(\tilde{Y} - \tilde{A}_{01}\hat{\gamma}_0) = \tilde{Y}^T W_0 W_0^T \tilde{Y}$, $W_0^T W_0 = I_{T-1}$ and $W_0 W_0^T = P_0 =: I_T - \tilde{A}_{01}(\tilde{A}_{01}^T \tilde{A}_{01})^{-1} \tilde{A}_{01}^T$. Thus

$$-2\log L^0 = T\log(\hat{\sigma}_0^2) + (\tilde{Y} - \tilde{A}_{01}\hat{\gamma}_0)^T(\tilde{Y} - \tilde{A}_{01}\hat{\gamma}_0) + T\log(2\pi) = T\log(\frac{\tilde{Y}^T W_0 W_0^T \tilde{Y}}{\sigma^2}) + C_0,$$

and $2\log L(0) - 2\log L^0 = T\log\{1 + \frac{(\tilde{Y}^T W_0 W_0^T \tilde{Y} - \tilde{Y}^T WW^T \tilde{Y})/\sigma^2}{\tilde{Y}^T WW^T \tilde{Y}/\sigma^2}\}$. We could 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.

$$\tilde{A}_{r1} = \pi_1 \frac{\tilde{A}_{01}}{||\tilde{A}_{01}||} + \pi_2 \tilde{A}_{01}^c. \tag{33}$$

By standard linear algebra, $W_0 W_0^T - WW^T = \tilde{A}_{01}^c (\tilde{A}_{01}^c)^T$. Hence equation (32) holds. □

Now we study Theorem 4.1.

Proof of Theorem 4.1: By Proposition 6.2, $\lambda R_3/\sigma^2 = \sum_{k=1}^K \frac{\lambda\eta_{k,T}}{1+\lambda\eta_{k,T}} w_{k1}^2 \stackrel{D}{=} N_T(\lambda)$. Note that $\tilde{Y}^T WW^T\tilde{Y}/\sigma^2 = \sum_{k=1}^{T-2} w_{k1}^2$. Hence

$$\frac{T\hat{\sigma}_\lambda^2}{\sigma^2} = \frac{\tilde{Y}^T WW^T\tilde{Y} - \lambda R_3}{\sigma^2} = \sum_{k=1}^K \frac{w_{k1}^2}{1+\lambda\eta_{k,T}} + \sum_{k=K+1}^{T-2} w_{k1}^2 \stackrel{D}{=} D_T(\lambda).$$

Together with Proposition 6.1, we have

$$\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{\lambda\geq 0}[T\log\{1 + N_T(\lambda)/D_T(\lambda)\} - \sum_{k=1}^K \log(1+\lambda\xi_{k,T})]. \tag{34}$$

Denote $(\tilde{A}_{01}^c)^T\tilde{Y}/\sigma = w_{T-1}$. Proposition 6.4 yields

$$2\log L(0) - 2\log L^0 = T\log\{1 + \frac{\tilde{Y}^T \tilde{A}_{01}^c (\tilde{A}_{01}^c)^T \tilde{Y}/\sigma^2}{\tilde{Y}^T WW^T\tilde{Y}/\sigma^2}\} = T\log\{1 + \frac{w_{T-1}^2}{\sum_{k=1}^{T-2} w_{k1}^2}\}.$$

It remains to show that under H_0, $w_{T-1} \stackrel{D}{=} N(0,1)$ and it is independent of w_{k1}'s, or equivalently, the vector $U_1^T W^T \tilde{Y}$. Under H_0, $\tilde{Y} = \tilde{A}_1\gamma = \tilde{A}_{01}\gamma_0 + \tilde{u}$. Recall that $\tilde{A}_{01}^c$ is the unit direction that is orthogonal to both $\tilde{A}_{01}$ and W. Since $(\tilde{A}_{01}^c)^T\tilde{Y} = (\tilde{A}_{01}^c)^T\tilde{u} \stackrel{D}{=} N(0,\sigma^2)$ and

$$cov\{(\tilde{A}_{01}^c)^T\tilde{Y}, U_1^T W^T\tilde{Y}\} = \sigma^2(\tilde{A}_{01}^c)^T W U_1 = \sigma^2 0_{T-2} U_1 = 0_{T-2},$$

we conclude Theorem 4.1 holds. □

Now we study Theorem 4.2.

Proof of Theorem 4.2: First we show equation (12) holds. Recall that $\xi_{s,T}$ and $\eta_{s,T}$ are the s-th eigenvalues of $\Lambda^{-1/2}A_2^T\Sigma_u^{-1}A_2\Lambda^{-1/2}$ and $\Lambda^{-1/2}A_2^T\Sigma_u^{-1/2}P\Sigma_u^{-1/2}A_2\Lambda^{-1/2}$ respectively. Note that $\Lambda^{-1/2}$ and $\Sigma_u^{-1/2}$ are bounded deterministic matrices. By continuous mapping theorem, it suffices to show that $T^{-2}A_2^TA_2$ and $T^{-2}A_2^TP'A_2$ converge, where $P' = I_T - A_2(A_2^TA_2)^{-1}A_2^T$.

Define that $\zeta_\iota = E(z_t^\iota)$, $\zeta_\iota(\kappa_i) = E\{z_t^\iota(z_t-\kappa_i)_+\}$, and $\zeta(\kappa_i,\kappa_j) = E\{(z_t-\kappa_i)_+(z_t-\kappa_j)_+\}$, for $l = 0$ or 1, $i = 1,\cdots,K$ and $j = 1,\cdots,K$. Let Π_1 be the 2×2 matrices with (i,j)-th element ζ_{i+j-2}, Π_2 be the $K\times K$ matrix with (i,j)-th element $\zeta(\kappa_i,\kappa_j)$ and Π_3 be the $2\times K$ matrix with (i,j)-th element $\zeta_{i-1}(\kappa_j)$. We first show that

$$T^{-2}A_1^TA_1 \Rightarrow \Pi_1\int B_v^2(s)ds, \tag{35}$$

$$T^{-2}A_2^TA_2 \Rightarrow \Pi_2\int B_v^2(s)ds, \tag{36}$$

$$T^{-2}A_1^TA_2 \Rightarrow \Pi_3\int B_v^2(s)ds. \tag{37}$$

Take the proof of equation (35) as an example. Note that the (i,j)th element of $A_1^TA_1$ satisfying $(A_1^TA_1)_{i,j} = \sum x_t^2 z_t^{i+j-2}$. By subtracting and adding the mean, we have

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

Hence $T^{-2}(A_1^TA_1)_{i,j} = T^{-2}\sum x_t^2 z_t^{i+j-2} \Rightarrow \zeta_{i+j-2}\int B_v^2(s)ds = (\Pi_1)_{i,j}\int B_v^2(s)ds$. Similarly,

we could show that equation (36) and (37) are true. Moreover,

$$\begin{aligned} T^{-2}A_2^T P' A_2 &= T^{-2}A_2^T A_2 - T^{-2}A_2^T A_1 (T^{-2}A_1^T A_1)^{-1} T^{-2} A_1^T A_2 \\ &\Rightarrow (\Pi_2 - \Pi_3^T \Pi_1^{-1} \Pi_3) \int B_v^2(s) ds. \end{aligned}$$

Therefore, we conclude that equation (12) is true.

Next we prove that equation (13) is valid. Recall that equation (34) holds for any T and K. Let $d = \lambda T^2$. Then we have

$$\sup_{\lambda \geq 0} \{2 \log L(\lambda) - 2 \log L(0)\} \stackrel{D}{=} \sup_{d \geq 0} [T \log\{1 + \frac{N_T(dT^{-2})}{D_T(dT^{-2})}\} - \sum_{k=1}^{K} \log(1 + dT^{-2}\xi_{k,T})]. \quad (38)$$

Define the right hand side in equation (38) as $\sup_{d \geq 0} f_T(d)$. We want to show that

$$\sup_{d \geq 0} f_T(d) \Rightarrow \sup_{d \geq 0} f(d) =: \sup_{d \geq 0} \{\sum_{s=1}^{K} \frac{d\eta_s}{1 + d\eta_s} w_s^2 - \sum_{s=1}^{K} \log(1 + d\xi_s)\}. \quad (39)$$

We first establish the finite dimensional convergence of $f_T(d)$. Since $T^{-2}\eta_{s,t} \Rightarrow \eta_s$, we have $N_T(dT^{-2})$ converges to $\sum_{s=1}^{K} \frac{d\eta_s}{1+d\eta_s} w_s^2$ for every fixed d. By Proposition 6.3,

$$T^{-1} D_T(dT^{-2}) = T^{-1}\{\tilde{Y}^T W W^T \tilde{Y}/\sigma^2 - N_T(dT^{-2})\} = 1 + o_p(1).$$

Therefore, $N_T(dT^{-2})/D_T(dT^{-2}) = O_p(T^{-1})$ and we have

$$T \log\{1 + \frac{N_T(dT^{-2})}{D_T(dT^{-2})}\} = T[\frac{N_T(dT^{-2})}{T\{1 + o_p(1)\}} + O_p(T^{-2})] = \sum_{s=1}^{K} \frac{d\eta_s}{1 + d\eta_s} w_s^2 + o_p(1).$$

Together with the fact $T^{-2}\xi_{s,T} \Rightarrow \xi_s$, we conclude $f_T(d)$ converges to $f(d)$ for every fixed d. Lemma 6.1 below shows that $f_T(d)$ form a tight sequence and hence $f_T(d)$ converges to $f(d)$ weakly. Lemma 6.2 further shows that $\sup_{d \geq 0} f_T(d) \Rightarrow \sup_{d \geq 0} f(d)$ by proving a continuous mapping theorem type results holds. Therefore, equation (38) holds.

By Proposition 6.4, $2 \log L(0) - 2 \log L^0 = T \log(1 + \frac{w_{T-1}^2}{\tilde{Y}^T W W^T \tilde{Y}/\sigma^2})$, where $w_{T-1} = (\tilde{A}_{01}^c)^T \tilde{Y}/\sigma$. Under H_0, $w_{T-1} \stackrel{D}{=} N(0,1)$. Recall that $W^T \tilde{Y}/\sigma \stackrel{D}{=} N(0_{T-2}, I_{T-2})$. Hence

$$\log(1 + \frac{w_{T-1}^2}{\tilde{Y}^T W W^T \tilde{Y}/\sigma^2}) = \frac{w_{T-1}^2}{T\{1 + o_p(1)\}} + O_p(T^{-2}) = \frac{w_{T-1}^2}{T} + o_p(T^{-1}).$$

Therefore, $2\log L(0) - 2\log L^0 = w_{T-1}^2 + o_p(1)$ and Theorem 4.2 is proved. □

LEMMA 6.1. *Under the assumptions of Theorem 4.2, $f_T(d) \Rightarrow f(d)$, where $f_T(d)$ and $f(d)$ are defined between equation (38) and equation (39).*

Proof of Lemma 6.1: We have already established the finite dimensional convergence of $f_T(d)$ to $f(d)$. It suffices to show that $f_T(d)$ form a tight sequence, i.e. for every ϵ and η strictly positive, there exist δ and T_0 such that for $T \geq T_0$,

$$\delta^{-1}\mathbb{P}\{\sup_{t\leq u\leq t+\delta} |f_T(u) - f_T(t)| \geq \epsilon\} \leq \eta.$$

By the definition of $f_T(\cdot)$, we have $|f_T(u) - f_T(t)| = T\log\{\frac{D_T(T^{-2}t)}{D_T(T^{-2}u)}\} + \sum_{s=1}^K \log\frac{1+tT^{-2}\xi_{s,T}}{1+uT^{-\alpha}\xi_{s,T}}$. Since $\log(1+x) < x$ for every $x > 0$, it holds that

$$\log\{\frac{D_T(T^{-2}t)}{D_T(T^{-2}u)}\} \leq \frac{D_T(T^{-2}t) - D_T(T^{-2}u)}{D_T(T^{-2}u)} \leq (u-t)\frac{\sum_{s=1}^K T^{-2}\eta_{s,T}w_s^2}{\sum_{s=K+1}^{T-2} w_s^2}.$$

Since $T^{-2}\eta_{s,T} \Rightarrow \eta_s$, there exists a constant C_1 such that $T^{-2}\eta_{s,T}w_s^2 \leq w_s^2$ for all s and T. Denote $R_{K,T} = \frac{\sum_{s=1}^K w_s^2/K}{\sum_{s=K+1}^{T-2} w_s^2/(T-K-2)}$. Then $T\log\{\frac{D_T(T^{-2}t)}{D_T(T^{-2}u)}\} \leq (u-t)C_1KR_{K,T}$. Since $T^{-2}\xi_{s,T} \Rightarrow \xi_s$, there exists a constant C_2 such that

$$\sum_{s=1}^K \log(\frac{1+uT^{-2}\xi_{s,T}}{1+tT^{-2}\xi_{s,T}}) \leq (u-t)\sum_{s=1}^K \xi_{s,T}T^{-2} \leq (u-t)C_2K.$$

Let $C_3 = \max(C_1, C_2)$. Then $\mathbb{P}\{\sup_{t\leq u\leq t+\delta} |f_T(u) - f_T(t)| \geq \epsilon\} \leq \mathbb{P}(R_{K,T} \geq \frac{\epsilon}{C_3K\delta} - 1)$, and it reduces to show the cumulative distribution function (c.d.f.) $H_{K,T}$ of $R_{K,T}$ satisfies

$$1 - H_{K,T}(\frac{\epsilon}{C_3K\delta} - 1) \leq \eta\delta. \tag{40}$$

Note that $R_{K,T}$ follows the F-distribution with degrees of freedom K and $T-2-K$. For every x, $\lim_{T\to\infty} H_{K,T}(x) = H_K(Kx)$, where H_K is the c.d.f of χ_K^2 random variables. Using L'Hospital rules, we have $\lim_{\delta\to 0+}\{1 - H_K(\frac{\epsilon}{C_3\delta} - K)\}/\{\frac{\eta\delta}{2}\} = 0$. Therefore, we could find $\delta = \delta(\epsilon, \eta)$, with $\delta < 1$ and $\frac{\epsilon}{C_\delta} - K > 0$, such that $1 - H_K(\frac{\epsilon}{C_3\delta} - K) \leq \frac{\eta\delta}{2}$. Because of the convergence of $H_{K,T}(x)$ to $H_K(Kx)$, we could find T_0 such that for $T \geq T_0$, it holds that

$|H_{K,T}(\frac{\epsilon}{C_3 K\delta} - 1) - H_K(\frac{\epsilon}{C_3\delta} - K)| \leq \frac{\eta\delta}{2}$. Thus equation (40) holds and we conclude that it converges to $f(d)$ weakly. □

LEMMA 6.2. *Under the assumptions of Theorem 4.2,* $\sup_{d\geq 0} f_T(d) \Rightarrow \sup_{d\geq 0} f(d)$.

Proof of Lemma 6.2: Lemma 6.1 shows that $f_T(d)$ weakly converges to $f(d)$. Similar as Crainiceanu and Ruppert (2004), we first find a random variable $F_{K,T}$ such that

$$\sup_{d\geq 0} f_T(d) = \max_{d\in[0,F_{K,T}]} f_T(d). \tag{41}$$

Note that $f_T(0) = 0$ for all T. It suffices to find $F_{K,T}$ such that $f_T(d) < 0$ when $d > F_{K,T}$. Recall that $\log(1 + x) \leq x$ when $x \geq 0$. By definition, $D_T(dT^{-2}) \geq \sum_{s=K+1}^{T-2} w_s^2$, and $N_T(dT^{-2}) \leq \sum_{s=1}^{K} w_s^2$ for all d. Hence

$$T\log\{1 + \frac{N_T(dT^{-2})}{D_T(dT^{-2})}\} \leq T\frac{N_T(dT^{-2})}{D_T(dT^{-2})} \leq T\frac{\sum_{s=1}^{K} w_s^2}{\sum_{s=K+1}^{T-2} w_s^2}. \tag{42}$$

Let m_0 be the positive constant such that all nonzero $\xi_{s,T}$'s satisfy $T^{-2}\xi_{s,T} \geq m_0$. Then

$$-\sum_{s=1}^{K} \log(1 + T^{-2}\xi_{s,T}) \leq -K\log(1 + dm_0). \tag{43}$$

With equations (42) and (43), we establish that

$$f_T(d) \leq T\frac{\sum_{s=1}^{K} w_s^2}{\sum_{s=K+1}^{T-2} w_s^2} - K\log(1 + dm_0). \tag{44}$$

Let $F_{K,T} =: m_0^{-1}\{\exp(\frac{T}{T-2-K}KR_{K,T}) - 1\}$. Mind that $f_T(F_{K,T}) = 0$. Since the right hand side of equation (44) is monotonic decreasing in d, $F_{K,T}$ has the desired property (41).

For any fixed $M > 0$ and $t \geq 0$, we have $\max_{d\in[0,M]} f_T(d) \leq \sup_{d\geq 0} f_T(d)$. Hence

$$\mathbb{P}\{\sup_{d\geq 0} f_T(d) \leq t\} \leq \mathbb{P}\{\max_{d\in[0,M]} f_T(d) \leq t\}.$$

Taking lim sup for $T \to \infty$ and applying the Continuous Mapping Theorem,

$$\limsup_{T\to\infty} \mathbb{P}(\sup_{d\geq 0} f_T(d) \leq t) \leq \limsup_{T\to\infty} \mathbb{P}(\max_{d\in[0,M]} f_T(d) \leq t) = \mathbb{P}(\max_{d\in[0,M]} f(d) \leq t).$$

Using the fact that

$$\lim_{M\to\infty} \mathbb{P}\{\max_{d\in[0,M]} f(d) \le t\} = \mathbb{P}\{\sup_{d\ge 0} f(d) \le t\}, \tag{45}$$

we have

$$\limsup_{T\to\infty} \mathbb{P}(\sup_{d\ge 0} f_T(d) \le t) \le \mathbb{P}\{\sup_{d\ge 0} f(d) \le t\}. \tag{46}$$

Since $\mathbb{P}(AB) \ge \mathbb{P}(A) - \mathbb{P}(B^c)$,

$$\begin{aligned}
\mathbb{P}(\sup_{d\ge 0} f_T(d) \le t) &\ge \mathbb{P}(\sup_{d\ge 0} f_T(d) \le t, F_{K,T} < M) \\
&= \mathbb{P}(\max_{d\in[0,M]} f_T(d) \le t, F_{K,T} < M) \\
&\ge \mathbb{P}(\max_{d\in[0,M]} f_T(d) \le t) - \mathbb{P}(F_{K,T} > M).
\end{aligned}$$

Note that $\mathbb{P}(F_{K,T} > M) \to \mathbb{P}(F_K > KM)$, where F_K is a χ^2_K distributed random variable. Taking $\liminf$, $\liminf_{T\to\infty} \mathbb{P}(\sup_{d\ge 0} f_T(d) \le t) \ge \mathbb{P}(\max_{d\in[0,M]} f(d) \le t) - \mathbb{P}(F_K \ge KM)$. Using equation (45) and that $\lim_{M\to\infty} \mathbb{P}(F_K \ge KM) = 0$, we conclude

$$\liminf_{T\to\infty} \mathbb{P}(\sup_{d\ge 0} f_T(d) \le t) \ge \mathbb{P}\{\sup_{d\ge 0} f(d) \le t\}.$$

Together with equation (46), the limit of $\mathbb{P}\{\sup_{d\ge 0} f_T(d)\}$ exists and satisfying

$$\lim_{T\to\infty} \mathbb{P}\{\sup_{d\ge 0} f_T(d) \le t\} = \mathbb{P}\{\sup_{d\ge 0} f(d) \le t\}.$$

Therefore, $\sup_{d\ge 0} f_T(d) \Rightarrow \sup_{d\ge 0} f(d)$. □

Now we study Theorem 4.3.

Proof of Theorem 4.3: First we show that equation (14) holds. Note that for any matrix A, its trace equals the sum of its eigenvalues. Since $\Lambda^{-1/2}$ and $\Sigma_u^{-1/2}$ are bounded deterministic matrices, it suffices to show both $(KT^2)^{-1}tr(A_2^T A_2)$ and $(KT^2)^{-1}tr(A_2^T P' A_2)$ converge, where P' is defined the same as in the proof of Theorem 4.2.

Take the term $(KT^2)^{-1}tr(A_2^T A_2)$ as an example. Let $\prod_2^* = \prod_2 \int B_v^2(s)ds$. From Equation (36), the (i,j)-th element of $T^{-2}A_2^T A_2$ satisfies $T^{-2}(A_2^T A_2)_{ij} \Rightarrow (\prod_2^*)_{ij}$. Note that

$$K^{-1}\sum_{i=1}^{K}(\Pi_2^*)_{ii} = K^{-1}\sum_{i=1}^{K}[E\{(z_t - \kappa_i)_+^2\}\int B_v^2(s)ds] \le C\int B_v^2(s)ds,$$

where the last inequality holds as $E\{(z_t - \kappa_i)_+^2\}$ is bounded. Therefore, $(KT^2)^{-1}tr(A_2^T A_2)$ converges. Similarly, $(KT^2)^{-1}tr(A_2^T P' A_2)$ converges and thus equation (14) is true.

Next we prove equation (15). Let $d = \lambda KT^2$. By equation (34), we could conclude that $\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{d\geq 0} g_T(d)$, where

$$\sup_{d\geq 0} g_T(d) \quad =: \quad \sup_{d\geq 0}[T\log\{1 + \frac{N_T\left(d(KT^2)^{-1}\right)}{D_T\left(d(KT^2)^{-1}\right)}\} - \sum_{k=1}^{K}\log\{1 + d(KT^2)^{-1}\xi_{k,T}\}]. \quad (47)$$

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

$$\bar{g}_1(d) = \lim_{T\to\infty}\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}}, \qquad \bar{g}_2(d) = \lim_{T\to\infty}\sum_{k=1}^{K}\log(1 + dK^{-1}T^{-2}\xi_{k,T}). \quad (48)$$

Lemma 6.3 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 as Lemma 6.1, we could show that $g_T(d)$ form a tight sequence and hence $g_T(d)$ converges to $\bar{g}_1(d) - \bar{g}_2(d)$ weakly. Similar as Lemma 6.2, we could establish a continuous mapping theorem type result and conclude that

$$\sup_{d\geq 0} g_T(d) \Rightarrow \sup_{d\geq 0}\{\bar{g}_1(d) - \bar{g}_2(d)\}, \quad (49)$$

Next we want to prove that

$$\sup_{d\geq 0}\{\bar{g}_1(d) - \bar{g}_2(d)\} = 0, \quad (50)$$

To prove equation (50), we will show that $\bar{g}_1(d) - \bar{g}_2(d) \leq 0$ for all d. Note that the first derivative of the partial sum induced by $\bar{g}_1(d) - \bar{g}_2(d)$ satisfies

$$\begin{aligned} & \left\{\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}} - \sum_{k=1}^{K}\log(1 + dK^{-1}T^{-2}\xi_{k,T})\right\}^{(1)} \\ = & \sum_{k=1}^{K}\frac{K^{-1}T^{-2}\eta_{k,T}}{(1 + dK^{-1}T^{-2}\eta_{k,T})^2} - \sum_{k=1}^{K}\frac{K^{-1}T^{-2}\xi_{k,T}}{1 + dK^{-1}T^{-2}\xi_{k,T}} \\ =: & \; Q_{1,T} + Q_{2,T}, \end{aligned}$$

where

$$Q_{1,T} = \sum_{k=1}^{K}\frac{K^{-1}T^{-2}\eta_{k,T}}{(1 + dK^{-1}T^{-2}\eta_{k,T})^2} - \sum_{k=1}^{K}\frac{K^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}},$$

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 $(1 + dK^{-1}T^{-2}\eta_{k,T}) \geq 1$, we have $Q_{1,T} \leq 0$ for all T. Moreover, $Q_{2,T} \leq 0$ as we explained below. Recall that $\xi_{k,T}$ and $\eta_{k,T}$ are the k-th eigenvalues of $\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_2\Lambda^{-1/2}$ and $\Lambda^{-1/2}\tilde{A}_2^T P\tilde{A}_2\Lambda^{-1/2}$ respectively. Moreover,

$$\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_2\Lambda^{-1/2} - \Lambda^{-1/2}\tilde{A}_2^T P\tilde{A}_2\Lambda^{-1/2} = \Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_1(\tilde{A}_1^T\tilde{A}_1)^{-1}\tilde{A}_1^T\tilde{A}_2\Lambda^{-1/2},$$

which is a semi-positive definite matrix. Hence $\eta_{k,T} \leq \xi_{k,T}$ for all k and T. Since $\frac{x}{1+dx}$ is an increasing function of x, we have

$$Q_{2,T} =: \sum_{k=1}^{K} \frac{K^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}} - \sum_{k=1}^{K} \frac{K^{-1}T^{-2}\xi_{k,T}}{1 + dK^{-1}T^{-2}\xi_{k,T}} \leq 0. \tag{51}$$

Because $\bar{g}_1(d)$ and $\bar{g}_2(d)$ are both absolutely summable, we could change the order between summation and derivative. Since $Q_{1,T} \leq 0$ and $Q_{2,T} \leq 0$, we conclude that the first derivative of $\bar{g}_1(d) - \bar{g}_2(d)$ satisfies

$$\bar{g}_1'(d) - \bar{g}_2'(d) = \lim_{T\to\infty}(Q_{1,T} + Q_{2,T}) \leq 0.$$

Recall that $\bar{g}_1(0) - \bar{g}_2(0) = 0$. For $d \geq 0$, $\bar{g}_1(d) - \bar{g}_2(d) = 0 + \int_0^d \{\bar{g}_1'(x) - \bar{g}_2'(x)\}dx \leq 0$. Therefore, equation (50) holds and thus $\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} = \sup_{d\geq 0} g_T(d) \Rightarrow 0$.

Similarly as in Theorem 4.2, we conclude that

$$2\log L(0) - 2\log L^0 = w_{T-1}^2 + o_p(1),$$

where $w_{T-1} \stackrel{D}{=} N(0,1)$ under H_0. Therefore, $LRT_T \Rightarrow \chi_1^2$ under H_0. □

LEMMA 6.3. *Assume the conditions in Theorem 4.3 and define $g_T(d)$ as in equation (47).*

Then $g_T(d)$ converges to $\bar{g}_1(d) - \bar{g}_2(d)$ for every fixed d, where $\bar{g}_1(d)$ and $\bar{g}_2(d)$ are defined in equation (48).

Proof of Lemma 6.3: First, we prove that $\bar{g}_1(d)$ and $\bar{g}_2(d)$ exist for any fixed $d \geq 0$. Since $1 + x \leq \exp(x)$ for $x \geq 0$, we have

$$1 + d\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T} \leq \prod_{k=1}^{K}(1 + dK^{-1}T^{-2}\xi_{k,T}) \leq \exp\{d\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T}\}.$$

If $d\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T}$ converges, so does $\log\{\prod_{k=1}^{K}(1 + dK^{-1}T^{-2}\xi_{k,T})\}$. We have already proved equation (14), i.e. $\sum_{k=1}^{K} K^{-1}T^{-2}\xi_{k,T} \Rightarrow \bar{\xi}$. Hence $\sum_{k=1}^{K}\log(1 + dK^{-1}T^{-2}\xi_{k,T})$ converges and its limit $\bar{g}_2(d)$ exists. Note that

$$0 \leq \frac{dK^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}} \leq dK^{-1}T^{-2}\eta_{k,T} \leq dK^{-1}T^{-2}\xi_{k,T}.$$

Since $\sum_{k=1}^{K} dK^{-1}T^{-2}\xi_{k,T}$ 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 6.4 below shows $N_T(dK^{-1}T^{-2})$ converges to $\bar{g}_1(d)$ for every fixed d. By Proposition 6.3,

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

Correspondingly $N_T\left(d(KT^2)^{-1}\right)/D_T\left(d(KT^2)^{-1}\right) = O_p(T^{-1})$ and

$$T\log\{1 + \frac{N_T\left(d(KT^2)^{-1}\right)}{D_T\left(d(KT^2)^{-1}\right)}\} = T[\frac{N_T\left(d(KT^2)^{-1}\right)}{T\{1 + o_p(1)\}} + O_p(T^{-2})] = \bar{g}_1(d) + o_p(1).$$

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 6.3 holds. □

LEMMA 6.4. *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 6.4: Notice that $w_k \overset{D}{=} iidN(0,1)$. It suffices to consider show that $\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}w_k^2 - \bar{g}_1(d) = \sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}(w_k^2 - 1)$ converges to 0 in finite dimension. For every fixed d,

$$E\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1 + dK^{-1}T^{-2}\eta_{k,T}}(w_k^2 - 1) = 0.$$

Moreover,

$$var\{\sum_{k=1}^{K}\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}}(w_k^2-1)\}=2\sum_{k=1}^{K}(\frac{dK^{-1}T^{-2}\eta_{k,T}}{1+dK^{-1}T^{-2}\eta_{k,T}})^2=O(K^{-1})=o(1).$$

Hence Lemma 6.4 is valid. □

Now we consider Theorem 4.4.

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 $\bar{\gamma}_1\bar{\pi}_2$. As $W^T\tilde{A}_1=\mathbf{0}_{(T-2)\times 2}$, we have $W^T\tilde{Y}=W^T\tilde{u}$. By the fact that $W_0^T\tilde{A}_{r1}=W_0^T(\pi_1\frac{\tilde{A}_{01}}{||\tilde{A}_{01}||}+\pi_2\tilde{A}_{01}^c)=\pi_2W_0^T\tilde{A}_{01}^c$, it holds $W_0^T\tilde{Y}=(T^{-1}\bar{\gamma}_1)\pi_2W_0^T\tilde{A}_{01}^c+W_0^T\tilde{u}$. Hence

$$\begin{aligned}
&\frac{\tilde{Y}^TW_0W_0^T\tilde{Y}-\tilde{Y}^TWW^T\tilde{Y}}{\sigma^2}\\
=\;&\frac{T^{-2}\bar{\gamma}_1^2\pi_2^2(\tilde{A}_{01}^c)^TW_0W_0^T\tilde{A}_{01}^c}{\sigma^2}+\frac{2T^{-1}\bar{\gamma}_1\pi_2(\tilde{A}_{01}^c)^TW_0W_0^T\tilde{u}}{\sigma^2}+\frac{\tilde{u}^T(W_0W_0^T-WW^T)\tilde{u}}{\sigma^2}\\
=\;&\frac{T^{-2}\bar{\gamma}_1^2\pi_2^2}{\sigma^2}+\frac{2T^{-1}\bar{\gamma}_1\pi_2(\tilde{A}_{01}^c)^T\tilde{u}}{\sigma^2}+\frac{\tilde{u}^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{u}}{\sigma^2}\\
\overset{D}{=}\;&(w_1+T^{-1}\pi_2\bar{\gamma}_1)^2,
\end{aligned}$$

where $w_1=(\tilde{A}_{01}^c)^T\tilde{u}/\sigma\overset{D}{=}N(0,1)$, and the second equation holds as any unit direction orthogonal to $\tilde{A}_{01}$ is the eigenvector of the projection matrix $W_0W_0^T$, i.e. $(\tilde{A}_{01}^c)^TW_0W_0^T=(\tilde{A}_{01}^c)^T$. Now we show the limit of $T^{-1}\pi_2$ exists. Note that

$$(T^{-1}\pi_2)^2=(T^{-1}\pi_2)^2(\tilde{A}_{01}^c)^TW_0W_0^T\tilde{A}_{01}^c=T^{-2}\tilde{A}_{r1}^TW_0W_0^T\tilde{A}_{r1}=T^{-2}\tilde{A}_{r1}^TP_0\tilde{A}_{r1}.$$

Using the same technique as in Theorem 4.2, we conclude all elements in $T^{-2}A_{r1}^TA_{r1}$ and $T^{-2}A_{r1}^TP_0'A_{r1}$ converge, where $P_0'=I_T-A_{01}(A_{01}^TA_{01})^{-1}A_{01}^T$. Since $\Sigma_u^{-1/2}$ and $\Lambda^{-1/2}$ are bounded deterministic matrices, $T^{-2}\tilde{A}_{r1}^T\tilde{A}_{r1}$ and $T^{-2}\tilde{A}_{r1}^TP_0\tilde{A}_{r1}$ converge. Equivalently, $T^{-1}\pi_2$ converges and we could denote its limit as $\bar{\pi}_2$. Hence

$$LRT_T=2\log L(0)-2\log L^0+o_p(1)=(w_1+\bar{\gamma}_1\bar{\pi}_2)^2+o_p(1),\tag{52}$$

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

Next consider H_{02}. Let $d = \lambda KT^2$. By Proposition 6.1 and Proposition 6.2, we have $\sup_{\lambda\geq 0}\{2\log L(\lambda) - 2\log L(0)\} \stackrel{D}{=} \sup_{d\geq 0} h_T(d)$, where $h_T(d) =: T\log\{1 + \frac{N_T'\left(d(KT^2)^{-1}\right)}{D_T(d(KT^2)^{-1})}\} - \sum_{k=1}^K \log\{1 + d(KT^2)^{-1}\xi_{k,T}\}$. We want to show

$$\sup_{d\geq 0} h_T(d) \Rightarrow \sup_{d\geq 0} h(d) =: \sup_{d\geq 0}\{\bar{d}_0\bar{\eta} + \bar{g}_3(d) - \bar{g}_2(d)\}, \tag{53}$$

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 (53), 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

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

Apply the same technique as Lemma 6.4, 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 $\bar{d}_0\bar{\eta} + \bar{g}_3(d)$. Using Proposition 6.3, $T^{-1}D_T\left(d(KT^2)^{-1}\right) = T^{-1}\{\tilde{Y}^TWW^T\tilde{Y}/\sigma^2 - N_T'\left(d(KT^2)^{-1}\right)\} = 1 + o_p(1)$. Correspondingly, $N_T'\left(d(KT^2)^{-1}\right)/D_T\left(d(KT^2)^{-1}\right) = O_p(T^{-1})$. Hence

$$T\log\{1 + \frac{N_T'\left(d(KT^2)^{-1}\right)}{D_T\left(d(KT^2)^{-1}\right)}\} = T[\frac{N_T'\left(d(KT^2)^{-1}\right)}{T\{1+o_p(1)\}} + O_p(T^{-2})] = \bar{d}_0\bar{\eta} + \bar{g}_3(d) + o_p(1).$$

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 as Lemma 6.1 and Lemma 6.2, we could further show that $h_T(d)$ weakly converges to $h(d)$ and a continuous mapping theorem type results holds. Thus equation (53) holds.

Next we want to show

$$\sup_{d\geq 0} h(d) = \max_{d\in[0,\bar{d}_0]}\{\bar{d}_0\bar{\eta} + \bar{g}_3(d) - \bar{g}_2(d)\}. \tag{54}$$

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 $d \geq \bar{d}_0$. By equation (51),

$$\sum_{k=1}^{K} \frac{(KT^2)^{-1}\eta_{k,T}}{1 + d(KT^2)^{-1}\eta_{k,T}} \leq \sum_{k=1}^{K} \frac{(KT^2)^{-1}\xi_{k,T}}{1 + d(KT^2)^{-1}\xi_{k,T}}.$$

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 equation (54) holds.

Finally, we will show that $2\log L(0) - 2\log L^0 \Rightarrow (1 + \bar{d}_0\varrho)w_1^2$. Under H_{02}, we have that $W^T\tilde{Y} = W^T(T^{-4/5}\tilde{A}_2\bar{b} + \tilde{u})$ and $W_0^T\tilde{Y} = W_0^T(T^{-4/5}\tilde{A}_2\bar{b} + \tilde{u})$. Hence

$$\frac{\tilde{Y}^T W_0 W_0^T \tilde{Y} - \tilde{Y}^T W W^T \tilde{Y}}{\sigma^2} = \frac{(T^{-4/5}\tilde{A}_2\bar{b} + \tilde{u})^T \tilde{A}_{01}^c (\tilde{A}_{01}^c)^T (T^{-4/5}\tilde{A}_2\bar{b} + \tilde{u})}{\sigma^2} \overset{D}{=} \sum_{i=1}^{T} \phi_{i4} w_i^2,$$

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

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

Note that Δ_4 share the same nonzero eigenvalue as

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

Moreover, we could show that the limit of $\frac{(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1}\tilde{A}_2^T\tilde{A}_{01}^c}{KT^2}$ exists. Using the same technique as in Theorem 4.2, we conclude that each element in $T^{-2}A_2^TA_{01}(A_{01}^c)^TA_2$ converges. Since $\Sigma_u^{-1/2}$ and $\Lambda^{-1/2}$ are bounded deterministic matrices, each element of the matrix $T^{-2}\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1/2}$ converges and $K^{-1}T^{-2}tr\{\Lambda^{-1/2}\tilde{A}_2^T\tilde{A}_{01}^c(\tilde{A}_{01}^c)^T\tilde{A}_2\Lambda^{-1/2}\}$ 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,

$$\frac{\tilde{Y}^T W_0 W_0^T \tilde{Y} - \tilde{Y}^T W W^T \tilde{Y}}{\sigma^2} \overset{D}{=} (1 + \bar{d}_0\varrho)w_1^2, \tag{55}$$

and we could further conclude $2\log L(0) - 2\log L^0 \Rightarrow (1 + \bar{d}_0\varrho)w_1^2$. Theorem 4.4 is proved. □

Appendix 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 *iid* $N(0, 1)$ 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 2000 replications.

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 *iid* $N(0, 1)$ 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 2000 replications.

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 *iid* $N(0, 1)$ 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 2000 replications.

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.

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, New York.

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.

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, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133-1159.

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.

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, UK: 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.

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.

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.
007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.
008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.
009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.
010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.
011 "The Real Consequences of Financial Stress" by Stefan Mittnik and Willi Semmler, February 2013.
012 "Are There Bubbles in the Sterling-dollar Exchange Rate? New Evidence from Sequential ADF Tests" by Timo Bettendorf and Wenjuan Chen, February 2013.
013 "A Transfer Mechanism for a Monetary Union" by Philipp Engler and Simon Voigts, March 2013.
014 "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, March 2013.
015 "Cyclical Variation in Labor Hours and Productivity Using the ATUS" by Michael C. Burda, Daniel S. Hamermesh and Jay Stewart, March 2013.
016 "Quantitative forward guidance and the predictability of monetary policy – A wavelet based jump detection approach –" by Lars Winkelmann, April 2013.
017 "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency" by Markus Bibinger, Nikolaus Hautsch, Peter Malec and Markus Reiss, April 2013.
018 "Fair re-valuation of wine as an investment" by Fabian Y.R.P. Bocart and Christian M. Hafner, April 2013.
019 "The European Debt Crisis: How did we get into this mess? How can we get out of it?" by Michael C. Burda, April 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

020 "Disaster Risk in a New Keynesian Model" by Maren Brede, April 2013.
021 "Econometrics of co-jumps in high-frequency data with noise" by Markus Bibinger and Lars Winkelmann, May 2013.
022 "Decomposing Risk in Dynamic Stochastic General Equilibrium" by Hong Lan and Alexander Meyer-Gohde, May 2013.
023 "Reference Dependent Preferences and the EPK Puzzle" by Maria Grith, Wolfgang Karl Härdle and Volker Krätschmer, May 2013.
024 "Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations" by Hong Lan and Alexander Meyer-Gohde, May 2013.
025 "The 'Celtic Crisis': Guarantees, transparency, and systemic liquidity risk" by Philipp König, Kartik Anand and Frank Heinemann, May 2013.
026 "State Price Densities implied from weather derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Huei-Wen Teng, May 2013.
027 "Bank Lending Relationships and the Use of Performance-Sensitive Debt" by Tim R. Adam and Daniel Streitz, May 2013.
028 "Analysis of Deviance in Generalized Partial Linear Models" by Wolfgang Karl Härdle and Li-Shan Huang, May 2013.
029 "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps" by Markus Bibinger and Mathias Vetter, May 2013.
030 "Can expert knowledge compensate for data scarcity in crop insurance pricing?" by Zhiwei Shen, Martin Odening and Ostap Okhrin, May 2013.
031 "Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, May 2013.
032 "CDO Surfaces Dynamics" by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ostap Okhrin, July 2013.
033 "Estimation and Inference for Varying-coefficient Models with Nonstationary Regressors using Penalized Splines" by Haiqiang Chen, Ying Fang and Yingxing Li, July 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".