Discussion Paper #2004-09

Efficient Estimation and Inference in Cointegrating Regressions with Structural Change

Eiji Kurozumi and Yoichi Arai

Efficient Estimation and Inference in Cointegrating Regressions with Structural Change*

EIJI KUROZUMI[†‡] YOICHI ARAI

Department of Economics
Hitotsubashi University

Faculty of Economics
University of Tokyo

January, 2005

Abstract

This paper investigates an efficient estimation method for a cointegrating regression model with structural change. Our proposal is that we first estimate the break point by minimizing the sum of squared residuals and then, by replacing the break fraction with the estimated one, we estimate the regression model by the canonical cointegrating regression (CCR) method proposed by Park (1992). We show that the estimator of the break fraction is consistent and of order faster than $T^{-1/2}$ and that the CCR estimator with the estimated break fraction has the same asymptotic property as the estimator with the known break point. Simulation experiments show how the finite sample distribution gets close to the limiting distribution as the magnitude of the break and/or the sample size increases.

*The discussion paper version of this paper might be updated occasionally. The latest version is available at http://www.econ.hit-u.ac.jp/~kurozumi/paper/efficient_jan05.pdf.

†This paper was written while Eiji Kurozumi was a visiting scholar of Boston University.

‡Correspondence: Eiji Kurozumi, Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA. E-mail: kurozumi@stat.hit-u.ac.jp

1. Introduction

Cointegration is among the primary interests of a researcher who investigates the long-run relationship between economic variables. Single equation methods for testing cointegration or no cointegration have been developed by Engle and Granger (1987), Phillips and Ouliaris (1990), and Shin (1994) among others, while a system equation model is considered by Johansen (1988, 1991), Ahn and Reinsel (1990), Lütkepohl and Saikkonen (2000), Saikkonen and Lütkepohl (2000a, b), and papers in references of Hubrich, Lütkepohl and Saikkonen (2001), who give a nice review of system methods.

It is often the case that data collected over relatively long time frames is used in the investigation of the long-run relationship, and the economic structure may change during the sample period. For a single or partial equation model, Campos, Ericsson and Hendry (1996) investigate the effect of structural change on cointegration tests, and Gregory and Hansen (1996a, b) propose tests for the null hypothesis of no cointegration with possibly one time structural break. Of importance is that the Gregory and Hansen's test is robust to the existence of structural change but the test is not helpful in determining whether structural change has occurred or not. Thus, once the cointegrating relation with or without structural change is observed by the test, we need to test for structural change. Hansen (1992) proposes various tests for the parameter stability and Quintos and Phillips (1993) investigate the LM test for structural change, while Hao and Inder (1996) regard testing for structural change as a diagnostic test and develop the CUSUM test. The finite sample properties of Hansen's (1992) tests are investigated by Gregory, Nason and Watt (1996) and Hao (1996), and they are generalized by Han (1996) to the exponential type tests as proposed by Andrews and Ploberger (1994) and Andrews, Lee and Ploberger (1996). Bai, Lumsdaine and Stock (1998) investigate testing for one time break and develop statistical inference about the estimator of the break point.

For a system equation model, tests of the cointegrating rank with a deterministic shift were developed by Saikkonen and Lütkepohl (2000c), Lütkepohl, Saikkonen and Trenkler (2003) for a known break point, while the unknown case is treated by Inoue (1999) and

Lütkepohl, Saikkonen and Trenkler (2004). On the other hand, change in cointegrating vectors is considered by Quintos (1995) for known break points, while tests for structural change for unknown break points are proposed by Quintos (1997), Seo (1998) and Hansen and Johansen (1999). Hansen (2003) derives the limiting distribution of the maximum likelihood estimator and proposes the likelihood ratio test for parameter restrictions when the break point and the cointegrating rank are known. Unfortunately, tests for the cointegrating rank with structural change assume the existence of structural change, whereas tests for structural change basically require the knowledge of the cointegrating rank. Additionally, to our best knowledge there is no test for the cointegrating rank with structural change in cointegrating vectors. Therefore, the system equation approach seems to be limited more or less when structural change in cointegrating vectors is incorporated in a model. For this reason, we consider a single equation model in this paper.

For a single equation model, suppose that we observe the cointegrating relation and structural change by using previously explained methods. If we know the date of structural change, we can estimate the model efficiently by the canonical cointegrating regression (CCR) method by Park (1992) and the fully modified regression (FMR) technique by Phillips and Hansen (1990) and Phillips (1995). However, we often encounter the case where we do not know the break date, and in this case the natural method for the estimation of the model is to first estimate the break point and then estimate the parameter in the model using the estimated break point. Bai, Lumsdaine and Stock (1998) show that the estimator of the parameter with the estimated break point has the same limiting distribution as the estimator with the known break point, assuming that the error term is independent of all leads and lags of the regressors. However, this assumption seems too restrictive for a cointegrating regression model because we often observe and commonly assume correlation between the error term and the regressors in a model. In this case, we cannot apply their result and are then required to find different methods to estimate the model with the unknown break point.

In this paper, we investigate the estimation method under the general assumptions that

first, the error term is correlated with leads and lags of the regressors, and second that the break point is unknown. We propose to estimate the break point at first by minimizing the sum of squared residuals (SSR) and then, using the estimated break fraction, to estimate the model by the CCR method of Park (1992). We show that the estimator of the break fraction is consistent and of order faster than $T^{-1/2}$. Although this order may not be sharp, this is enough for us to derive the asymptotic distribution of the estimator of the parameter. Since the limiting distribution is shown to be a mixed normal, we can test parameter restrictions by constructing the Wald type test statistic, which converges to a chi-square distribution.

The structure of this paper is as follows. Section 2 explains the model and assumptions. In Section 3 we first derive the CCR estimator with the known break point. We then investigate the asymptotic property of the estimator of the break fraction. Using this estimated break fraction, we estimate the regression model by the CCR method and show that the estimator of the parameter has the same limiting distribution as the CCR estimator with the known break point. Section 4 gives the finite sample property of the estimator. Section 5 concludes the paper.

2. A Model and Assumptions

Let us consider the following cointegrating regression model,

$$\begin{aligned} y_{1t} &= \mu_1 + \mu_2\varphi_{t\tau_o} + \beta_1' y_{2t} + \beta_2' y_{2t}\varphi_{t\tau_o} + v_{1t} \\ &= b' x_{t\tau_o} + v_{1t}, \end{aligned} \tag{1}$$

for $t = 1, \cdots, T$, where $\{y_{1t}\}$ and $\{y_{2t}\}$ are one and m dimensional stochastic sequences, $\varphi_{t\tau_o}$ is a step function such that $\varphi_{t\tau_o} = 0$ for $t \leq [T\tau_o]$ and $\varphi_{t\tau_o} = 1$ for $t > [T\tau_o]$, $b = [\mu_1, \mu_2, \beta_1', \beta_2']'$, and $x_{t\tau_o} = [1, \varphi_{t\tau_o}, y_{2t}', y_{2t}'\varphi_{t\tau_o}]'$. Let $v_t = [v_{1t}, v_{2t}']'$ where $v_{2t} = \triangle y_{2t}$, and define its long-run variance as $\Omega = \lim_T T^{-1} E[V_T V_T']$ where $V_T = \sum_{t=1}^{T} v_t$. We partition Ω conformably with v_t as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$

We also define

$$\Sigma = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E[v_t v_t'], \qquad \Lambda = \lim_{T\to\infty} T^{-1}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j} E[v_t v_{t+j}'],$$

$$\Gamma = \Sigma + \Lambda = \begin{bmatrix} \gamma_{11} & \gamma_{12}' \\ \gamma_{21} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} \gamma_1' \\ \Gamma_2' \end{bmatrix},$$

where Γ is partitioned conformably with v_t.

We employ the following set of assumptions throughout the paper.

Assumption 1 *(a) y_0 is a fixed or a random vector with $E[y_0] < \infty$ and independent of T.*
(b) $\{v_t\}$ is mean-zero and strong mixing with mixing coefficients of size $-p\alpha/(p-\alpha)$ and $E|v_t| < \infty$ for some $p > \alpha > 5/2$.
(c) The matrix Ω exists with finite elements, $\Omega > 0$, $\omega_{11} > 0$, and $\Omega_{22} > 0$.
(d) The break fraction τ_o is constant and $\tau_o \in \mathcal{T} = [\underline{\tau}, \bar{\tau}]$ for known $0 < \underline{\tau} < \bar{\tau} < 1$.
(e) $\beta_2 = \beta_{2T} = T^{-1/2}\beta_{2o}$ where β_{2o} is a fixed vector.

Assumption (a) gives the initial value condition such that y_0 does not affect the asymptotic theory derived in the following sections. Assumptions (b) and (c) ensure that the functional central limit theorem (FCLT) holds for the partial sum process of $\{v_t\}$, so that

$$T^{-1/2}\sum_{t=1}^{[Tr]} v_t \Rightarrow B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \begin{matrix} 1 \\ m \end{matrix},$$

where $B(r)$ is an $(m+1)$ dimensional Brownian motion with the variance matrix Ω and $\Rightarrow$ signifies weak convergence of the associated probability measures. The positive definiteness of Ω_{22} excludes the case where y_{2t} is cointegrated. Assumption (d) is standard for a structural break model. Assumption (e) is used to derive the convergence rate of the estimator of the break fraction. Strictly speaking, this assumption is not necessary for our asymptotic theory because we will not derive the limiting distribution of $\hat{\tau}$, the estimator of the break fraction. However, as will be discussed in the next section, if we assume that β_2 is fixed, the asymptotic property of $\hat{\tau}$ will be determined only by y_{2t} and a constant term will no longer play an important role for estimation of τ. (e) is assumed so that both a constant and the I(1) regressors are effective in the estimation of the break point.

3. The CCR Estimator with Structural Change

Our strategy for estimation is that we first obtain the estimate of the break point by minimizing the SSR and then estimate (1) by the CCR method using the estimated break point. We will show that the CCR estimator with the estimated break point has the same limiting distribution as the CCR estimator with the known break point. Note that, although the following explanation proceeds based on the CCR method, we can easily apply our result to the fully modified regression (FMR) technique by Phillips and Hansen (1990) and Phillips (1995). The difference between the CCR and the FMR methods resides in how to correct serial correlations. See Phillips and Hansen (1990) and Park (1992) for details.

3.1. The CCR method with a known break point

First, we briefly explain the CCR method for a known break point. It consists of two separate step estimations. The first step is to estimate (1) by OLS regression. Let $\hat{b}_{\tau_o} = [\hat{\mu}_{1\tau_o}, \hat{\mu}_{2\tau_o}, \hat{\beta}'_{1\tau_o}, \hat{\beta}'_{2\tau_o}]$ be the OLS estimator of b and $\hat{v}_{1t\tau_o}$ be the OLS residual. Using $\hat{b}_{\tau_o}$ and $\hat{v}_{1t\tau_o}$ we construct variables $y^*_{1t\tau_o}$ and $y^*_{2t\tau_o}$ as

$$y^*_{1t\tau_o} = y_{1t} - (\hat{\beta}'_{1\tau_o}\hat{\Gamma}'_{2\tau_o}\hat{\Sigma}^{-1}_{\tau_o} + \hat{\beta}'_{2\tau_o}\hat{\Gamma}'_{2\tau_o}\hat{\Sigma}^{-1}_{\tau_o}\varphi_{t\tau_o} + [0, \hat{\omega}'_{21\tau_o}\hat{\Omega}^{-1}_{22\tau_o}])\hat{v}_{t\tau_o}, \quad y^*_{2t\tau_o} = y_{2t} - \hat{\Gamma}'_{2\tau_o}\hat{\Sigma}^{-1}_{\tau_o}\hat{v}_{t\tau_o},$$

where $\hat{v}_{t\tau_o} = [\hat{v}_{1t\tau_o}, \triangle y'_{2t}]'$ and $\hat{\Gamma}_{2\tau_o}$, $\hat{\Sigma}_{\tau_o}$, $\hat{\omega}_{21\tau_o}$, and $\hat{\Omega}_{22\tau_o}$ are consistent estimators of Γ_2, Σ, ω_{21}, and Ω_{22} that are defined below. Then, the CCR estimator is obtained by regressing $y^*_{1t\tau_o}$ on $y^*_{2t\tau_o}$,

$$y^*_{1t\tau_o} = b^{*\prime}_{\tau_o} x^*_{t\tau_o} + e^*_{t\tau_o}, \tag{2}$$

where $x^*_{t\tau_o} = [1, \varphi_{t\tau_o}, y^{*\prime}_{2t\tau_o}, y^{*\prime}_{2t\tau_o}\varphi_{t\tau_o}]'$. We denote the CCR estimator and the estimated residual as $\hat{b}^*_{\tau_o}$ and $\hat{e}^*_{t\tau_o}$.

The long-run matrices are estimated by

$$\hat{\Sigma}_{\tau_o} = T^{-1}\sum_{t=1}^{T}\hat{v}_{t\tau_o}\hat{v}'_{t\tau_o}, \quad \hat{\Lambda}_{\tau_o} = T^{-1}\sum_{j=1}^{\ell}k(j/\ell)\sum_{t=1}^{T-j}\hat{v}_{t\tau_o}\hat{v}_{t+j\tau_o},$$

$$\hat{\Gamma}_{\tau_o} = \hat{\Sigma}_{\tau_o} + \hat{\Lambda}_{\tau_o}, \quad \hat{\Omega}_{\tau_o} = \hat{\Sigma}_{\tau_o} + \hat{\Lambda}_{\tau_o} + \hat{\Lambda}'_{\tau_o},$$

and $k(j/\ell)$ is a kernel function that satisfies the following restrictions.

Assumption 2 *(a) $k(\cdot)$ is a continuous and even faction with $|k(\cdot)| \leq 1$, $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(x)dx < \infty$.*
(b) ℓ goes to infinity as $n \to \infty$ and $\ell = o(T^{1/2})$.

Assumption 2 suffices to guarantee the consistency of $\hat{\Lambda}_{\tau_o}$, $\hat{\Gamma}_{\tau_o}$, and $\hat{\Omega}_{\tau_o}$, and many well known kernels such as the Bartlett and the quadratic spectral kernels satisfy this assumption. See, for example, Andrews (1991).

The asymptotic distribution of $\hat{b}^*_{\tau_o}$ is given by the following proposition, which can be proved in exactly the same way as Park (1992). Since this is the case, we omit the proof.

Proposition 1 *Let Assumptions 1 (a)-(d) and Assumption 2 hold. Then, as $T \to \infty$,*

$$D_T(\hat{b}^*_{\tau_o} - b) \xrightarrow{d} \left(\int_0^1 X_{\tau_o}(r) X_{\tau_o}(r)' dr \right)^{-1} \int_0^1 X_{\tau_o}(r) dB_{1\cdot 2}(r), \tag{3}$$

where $D_T = \text{diag}\{T^{1/2}, T^{1/2}, T\, I_m, T\, I_m\}$, $X_{\tau_o}(r) = [1, \varphi_{\tau_o}(r), B_2(r)', B_2(r)'\varphi_{\tau_o}(r)]'$, $\varphi_{\tau_o}(r)$ is a step function on $[0,1]$ such that $\varphi_{\tau_o}(r) = 1\{r \geq \tau_o\}$ with $1\{\cdot\}$ being an indicator function, and $B_{1\cdot 2}(r) = B_1(r) - \omega'_{21}\Omega_{22}^{-1} B_2(r)$.

As discussed in Park (1992), the Wald test statistic based on the CCR estimator has an asymptotic chi-square distribution because (3) is a mixed normal distribution. For example, let us consider the general hypothesis of the form

$$H_0 : \; g(b) = 0$$

where $g(\cdot)$ is a continuously differentiable q dimensional vector. Assume that $G(b) = \partial g(b)/\partial b'$ is of rank q. Then, from Proposition 1, we can see that

$$\begin{aligned} W_T(\hat{b}^*_{\tau_o}) \quad &= \quad g(\hat{b}^*_{\tau_o})' \left\{ \hat{\omega}^*_{1\cdot 2\tau_o} G(\hat{b}^*_{\tau_o}) \left(\sum_{t=1}^{T} x^*_{t\tau_o} x^{*\prime}_{t\tau_o} \right)^{-1} G(\hat{b}^{*\prime}_{\tau_o}) \right\}^{-1} g(\hat{b}^*_{\tau_o}) \\ &\xrightarrow{d} \quad \chi^2_q, \end{aligned} \tag{4}$$

where $\hat{\omega}^*_{1\cdot 2\tau_o}$ is a consistent estimator of the long-run variance $\omega_{1\cdot 2}$, which can be constructed using the CCR error $\hat{e}^*_{t\tau_o}$ in the same way as the nonparametric estimator of Ω.

3.2. The CCR method with an unknown break point

Regressions (1) and (2) are infeasible in practice because we do not know the true break point. A feasible method is that we first estimate the break point and then estimate the model using the estimated break point by the CCR method.

In the framework of cointegrating regressions, Bai, Lumsdaine and Stock (1998) investigated the quasi-maximum likelihood estimator of the break point and they derived the limiting distribution of the estimator. One of the important assumptions in their paper is that the disturbance v_{1t} must be independent of the regressors for all leads and lags (Assumption 3.2 in Bai, Lumsdaine and Stock, 1998). However, it is apparent that this assumption is not satisfied in our model, and we cannot apply their result. Therefore, we must investigate the asymptotic behavior of the estimator of the break point under general assumptions.

Let us consider a feasible version of the regression (1)

$$y_{1t} = b' x_{t\tau} + v_{1t\tau}, \tag{5}$$

where $\tau \in \mathcal{T}$ and $v_{1t\tau} = v_{1t} - b'(x_{t\tau} - x_{t\tau_o})$. Let $\hat{b}_\tau$ and $\hat{v}_{1t\tau}$ be the OLS estimator of b and the regression residual. The estimator of the break point is obtained by minimizing the sum of squared residuals in (5), or equivalently, the estimator of the break fraction is given by

$$\hat{\tau} = \arg \inf_{\tau \in \mathcal{T}} S_T(\tau),$$

where $S_T(\tau) = T^{-1} \sum_{t=1}^{T} \hat{v}_{1t\tau}^2$. We first give the consistency of $\hat{\tau}$ by the following proposition.

Proposition 2 *Let Assumptions 1 (a)-(e) hold. Then,* $\hat{\tau} \xrightarrow{p} \tau_o$.

Assumption 1 (e) implies that the magnitude of the break for y_{2t} shrinks to zero as T goes to infinity and is of order $T^{-1/2}$. It is not difficult to see that Proposition 2 holds without Assumption 1 (e). The reason we assume Assumption 1 (e) is that, if β_2 is supposed to be fixed, y_{2t} asymptotically dominates the other terms in the objective function and the asymptotic property of $\hat{\tau}$ will be determined only by the behavior of y_{2t}. Assumption 1 (e)

is supposed so that both a constant and y_{2t} have the same importance for determining the asymptotic behavior of the estimator. See also Bai, Lumsdaine and Stock (1998).

Once the consistency of $\hat{\tau}$ is obtained, we can restrict the parameter space of τ to only the vicinity of τ_o that shrinks to τ_o. Details are given in the appendix. By considering the shrinking parameter space, we can prove the next proposition.

Proposition 3 *Let Assumptions 1 (a)-(e) hold. Then,* $T^{1/2}(\hat{\tau} - \tau_o) \xrightarrow{p} 0$.

Proposition 3 implies that $\hat{\tau}$ converges in probability to τ_o of order faster than $T^{-1/2}$. This convergence rate is slower than that obtained by Bai, Lumsdaine and Stock (1998) and the result in Proposition 3 may not be sharp. However, our purpose is not to derive the limiting distribution of the estimator of the break fraction but to obtain a feasible method of statistical inference about regression coefficients when the break point is unknown. The convergence rate given by Proposition 3 is enough for us to obtain such a feasible method and we do not pursue a sharp rate of $\hat{\tau}$ under general assumptions.

To construct the CCR estimator we need the estimators of the long-run variances. In exactly the same way as the known break point case, we construct $\hat{\Sigma}_{\hat{\tau}}$, $\hat{\Lambda}_{\hat{\tau}}$, $\hat{\Gamma}_{\hat{\tau}}$, and $\hat{\Omega}_{\hat{\tau}}$ by replacing $\hat{v}_{t\tau_o}$ by $\hat{v}_{t\hat{\tau}} = [\hat{v}_{1t\hat{\tau}}, \triangle y_{2t}']'$. The following proposition shows that these estimators are consistent.

Proposition 4 *Let Assumptions 1 (a)-(e) and Assumption 2 hold. Then,* $\hat{\Sigma}_{\hat{\tau}}$, $\hat{\Lambda}_{\hat{\tau}}$, $\hat{\Gamma}_{\hat{\tau}}$, *and* $\hat{\Omega}_{\hat{\tau}}$ *converge in probability to* Σ, Λ, Γ, *and* Ω, *respectively.*

We are now in a position to construct the CCR estimator using the estimated break point, $\hat{\tau}$. Let $y^*_{1t\hat{\tau}}$ and $y^*_{2t\hat{\tau}}$ be defined in the same way as $y^*_{1t\tau_o}$ and $y^*_{2t\tau_o}$ using $\hat{\tau}$. The feasible CCR estimator, $\hat{b}_{\hat{\tau}}$, is obtained by regressing $y^*_{1t\hat{\tau}}$ and $y^*_{2t\hat{\tau}}$. The following is the main theorem in this paper.

Theorem 1 *Let Assumptions 1 (a)-(e) and Assumption 2 hold. Then,*

$$D_T(\hat{b}^*_{\hat{\tau}} - \hat{b}^*_{\tau_o}) \xrightarrow{p} 0.$$

Theorem 1 implies that the CCR estimator with the estimated break point has the same limiting distribution as the estimator with the known break point. Then, even if we construct the Wald test statistic for H_0 using the feasible CCR estimator, it converges in distribution to a chi-square distribution with q degrees of freedom, that is,

$$\begin{aligned} W_T(\hat{b}^*_{\hat{\tau}}) &= g(\hat{b}^*_{\hat{\tau}})' \left\{ \hat{\omega}^*_{1\cdot 2\hat{\tau}} G(\hat{b}^*_{\hat{\tau}}) \left(\sum_{t=1}^{T} x^*_{t\hat{\tau}} x^{*\prime}_{t\hat{\tau}} \right)^{-1} G(\hat{b}^{*\prime}_{\hat{\tau}}) \right\}^{-1} g(\hat{b}^*_{\hat{\tau}}) \\ &\xrightarrow{d} \chi^2_q. \end{aligned}$$

To conclude this section, we consider the extension of the model (1) in several directions. For example, we may be interested in a partial change of the parameters. In this case, we can easily see that all of the results in the paper are established in exactly the same manner. We may also want to include a linear trend as a regressor,

$$y_{1t} = \mu_1 + \mu_2 \varphi_{t\tau_o} + d_1 t + d_2 t \varphi_{t\tau_o} + \beta_1' y_{2t} + \beta_2' y_{2t} \varphi_{t\tau_o} + v_{1t}.$$

Again, the propositions and the theorem can be shown to hold for this model. In this case, D_T is defined as $D_T = \text{diag}\{T^{1/2}, T^{1/2}, T^{3/2}, T^{3/2}, T\ I_m, T\ I_m\}$ and the definition of X_{τ_o} in (3) should be changed appropriately. Seasonal constants may be of particular interest for some researchers and they may be included. In any case, although the expression (3) of the limiting distribution should be changed, the Wald statistic still has an asymptotic chi-square distribution and we can make statistical inferences about regression coefficients.

4. Finite sample evidence

In this section, we investigate finite sample properties of the feasible CCR estimator and the Wald test statistic proposed in the previous section. We consider the following data generating process:

$$\begin{aligned} y_{1t} &= \mu_1 + \mu_2 \varphi_{t\tau_o} + \beta_1 y_{2t} + \beta_2 y_{2t} \varphi_{t\tau_o} + v_{1t}, \qquad (6) \\ v_t &= A v_{t-1} + \varepsilon_t, \end{aligned}$$

where y_{2t} is a one dimensional unit root process, $v_t = [v_{1t}, \triangle y_{2t}]'$, $A = \text{diag}\{a, a\}$, and $\{\varepsilon_t\} \sim NID(0, I_2)$. We set $\mu_1 = 0$, $\beta_1 = 1$, $a = -0.6$, 0, or 6, and the sample size is 100,

300 or 500. The break fraction τ_o is set to be 0.5 for all experiments. The values of μ_2 and β_2 are selected as follows. First, we regard $d = \mu_2 + \beta_2 \times \{Var(y_{2t})\}^{1/2}$ as a measure of the magnitude of the change. We also note that variation in $y_{1,t+1}$ given $y_{1,t}$ is $v_{2,t+1} + v_{1,t+1}$ for all t if structural change does not occur at t and its standard deviation is given by $\{Var(v_{2,t+1} + v_{1,t+1})\}^{1/2} = 2^{1/2}$ when v_t is an i.i.d. sequence. We choose μ_2 and β_2 so that the magnitude of the break, d, becomes approximately equal to $s \times 2^{1/2}$ for $s = 0.5$, 1, 2, or 3 at $t = T\tau_o + 1$. According to this rule, we set $\{\mu_2, \beta_2\} = \{0.35, 0.05\}$, $\{0.7, 0.1\}$, $\{1.4, 0.2\}$, and $\{2.1, 0.3\}$, which correspond to the cases where $T = 100$ and $s = 0.5$, 1, 2, and 3. The same sets of values are also used for $T = 300$ and 500 to see the effect of the sample size on the finite sample property.

First, we see the finite sample distributions of the estimates of β_1 and β_2. Figure 1 shows the probability density functions (pdf) of $T(\hat{\beta}_1 - \beta_1)$ and $T(\hat{\beta}_2 - \beta_2)$ for $a = 0$, each of which is drawn based on 100,000 replications. We can see that the pdf has fatter tails for each case when the magnitude of the break is smaller. The finite sample distribution approaches the limiting distribution as the magnitude of the break becomes larger, and the pdf with the known break point is closest to the limiting distribution. As expected, the finite sample distribution approaches the limiting distribution when the sample size is large. We can also see that the pdf of $\hat{\beta}_2$ is not as close to the limiting distribution as the pdf of $\hat{\beta}_1$. This is because β_1 is estimated using the whole sample period, while β_2 is estimated using only the observations after the break point. As a whole, more than 300 observations are required to approximate the finite sample distribution by the limiting one when the magnitude of the change is very small ($s = 0.5$).

The above property is preserved when $a = 0.6$ and $a = -0.6$, but the finite sample distribution is slightly closer to the limiting one for $a = 0.6$ compared with the case when $a = 0$, while the difference between the finite and the limiting distributions is slightly larger when $a = -0.6$ than the case when $a = 0$ (we do not draw the pdfs when $a \neq 0$ to save space).

Next, we investigate the size and power of the Wald test statistic. We consider the null

hypothesis of $H_0 : \ \beta_1 = b$ and construct the test statistic. We set $b = 1$ to see the size of the test, while it is set to be 1.01, 1.05, and 1.1 to investigate the power of the test. The level of significance is 0.05 and the number of replications is 5,000 in all experiments.

Table 1 summarizes the results of the simulations. When the break point is known, the size of the test is close to the nominal one when $a = 0$, but the test suffers from size distortion when $a = 0.6$. As expected from Figure 1, the size of the Wald statistic with the estimated break point approaches the known break point case as the magnitude of the break is larger. Regarding power, the test becomes more powerful when $|b_{1o} - 1|$ increases. Although we must be cautious of the comparison of the power for different settings of parameters, the power property of the test does not seem to depend significantly on the value of a.

We also investigate the size and power of the Wald test for b_2. The performance of the test under the null hypothesis is similar to the test of b_1, but the test of b_2 is less powerful than that of b_1 (we do not report this result to save space).

5. Conclusion

In this paper we proposed to estimate the cointegrating regression model with structural change by the CCR estimation technique with the break point replaced by the estimated one. We first estimated the break fraction by minimizing the sum of squared residuals, and this estimator was shown to converge in probability to the true break fraction at a rate faster than $T^{1/2}$. We found that the feasible CCR estimator converges in distribution to a mixed normal distribution, so that the Wald test statistic based on it is asymptotically chi-square distributed. By Monte Carlo simulations, we showed that the finite sample distribution of the estimator approaches the limiting distribution as the magnitude of the break and/or the sample size becomes larger.

It might be possible to obtain an efficient estimator by other methods such as the dynamic OLS (DOLS) method used by Saikkonen (1991) and Stock and Watson (1993), which estimates the model by adding leads and lags of the first differences of the I(1) regressors, where the lag length goes to infinity as $T \to \infty$. Since the number of the regressors

changes depending on the sample size, much would be required to obtain the results given by Propositions 2 and 3.

References

[1] Ahn, S. K. and G. C. Reinsel (1990) Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* 85, 813-823.

[2] Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817-858.

[3] Andrews, D. W. K., I. Lee and W. Ploberger (1996) Optimal changepoint tests for normal linear regression. *Journal of Econometrics* 70, 9-38.

[4] Andrews, D. W. K. and W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383-1414.

[5] Bai, J., R. L. Lumsdaine and J. H. Stock (1998) Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65, 395-432.

[6] Campos, J., N. R. Ericsson and D. F. Hendry (1996) Cointegration tests in the presence of structural breaks. *Journal of Econometrics* 70, 187-220.

[7] Engle, R. F. and C. W. J. Granger (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251-276.

[8] Gregory, A. W. and B. E. Hansen (1996a) Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70, 99-126.

[9] Gregory, A. W. and B. E. Hansen (1996b) Tests for cointegration in models with regime and trend shifts. *Oxford Bulletin of Economics and Statistics* 58, 555-560.

[10] Gregory, A. W., J. M. Nason and D. G. Watt (1996) Testing for structural breaks in cointegrated relationships. *Journal of Econometrics* 71, 321-341.

[11] Hansen, B. E. (1992) Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* 10, 321-335.

[12] Hansen, P. R. (2003) Structural changes in the cointegrated vector autoregressive model. *Journal of Econometrics* 114, 261-295.

[13] Hansen, H. and S. Johansen (1999) Some tests for parameter constancy in cointegrated VAR-models. *Econometrics Journal* 2, 306-333.

[14] Hao, K. (1996) Testing for structural changes in cointegrated regression models: Some comparisons and generalizations. *Econometric Reviews* 15, 401-429.

[15] Hao, K. and B. Inder(1996) Diagnostic test for structural change in cointegrated regression models. *Economics Letters* 50, 179-187.

[16] Hubrich, K., H. Lütkepohl and P. Saikkonen (2001) A review of systems cointegrating tests. *Econometric Reviews* 20, 247-318.

[17] Inoue, A. (1999) Tests of cointegrating rank with a trend-break. *Journal of Econometrics* 90, 215-237.

[18] Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-254.

[19] Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.

[20] Lütkepohl, H. and P. Saikkonen (2000) Testing for the cointegrating rank of a VAR process with a time trend. *Journal of Econometrics* 95, 177-198.

[21] Lütkepohl, H., P. Saikkonen and C. Trenkler (2003) Comparison of tests for the cointegrating rank of a VAR process with a structural shift. *Journal of Econometrics* 113, 201-229.

[22] Lütkepohl, H., P. Saikkonen and C. Trenkler (2004) Testing for the cointegrating rank of a VAR process with level shift at unknown time. *Econometrica* 72, 647-662.

[23] Park, J. Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119-143.

[24] Park, J. Y., and P. C. B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468-497.

[25] Phillips, P. C. B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023-1078.

[26] Phillips, P. C. B., and B. E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99-125.

[27] Phillips, P. C. B. and S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165-193.

[28] Quintos, C. E. (1995) Sustainability of the deficit process with structural shifts. *Journal of Business and Economic Statistics* 13, 409-417.

[29] Quintos, C. E. (1997) Stability tests in error correction models. *Journal of Econometrics* 82, 289-315.

[30] Quintos, C. E. and P. C. B. Phillips (1993) Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675-706.

[31] Saikkonen, P. (1991) Asymptotically Efficient Estimation of Cointegration Regressions. *Econometric Theory* 7, 1-21.

[32] Saikkonen, P. and H. Lütkepohl (2000a) Testing for the cointegrating rank of a VAR process with a intercept. *Econometric Theory* 16, 373-406.

[33] Saikkonen, P. and H. Lütkepohl (2000b) Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. *Journal of Time Series Analysis* 21, 435-456.

[34] Saikkonen, P. and H. Lütkepohl (2000c) Testing for the cointegrating rank of a VAR process with structural shifts. *Journal of Business and Economic Statistics* 18, 451-464.

[35] Seo, B. (1998) Tests for structural change in cointegrated systems. *Econometric Theory* 14, 222-259.

[36] Shin, Y. (1994) A residual- based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91-115.

[37] Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[38] Stock, J. H. and M. W. Watson (1993) A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* 61, 783-820.

Appendix

Without loss of generality we shall assume that $T\tau_o$ and $T\hat{\tau}$ are integers in this appendix.

Proof of Proposition 2: We need to show that $P(|\hat{\tau} - \tau_o| > \varepsilon) \to 0$ for every $\varepsilon > 0$. Noting that

$$\begin{aligned} P\left(|\hat{\tau} - \tau_o| > \varepsilon\right) &= P\left(\inf_{\tau \in \mathcal{T} \setminus \delta(\varepsilon)} S_T(\tau) < \inf_{\tau \in \delta(\varepsilon)} S_T(\tau)\right) \\ &\leq P\left(\inf_{\tau \in \mathcal{T} \setminus \delta(\varepsilon)} S_T(\tau) < S_T(\tau_o)\right), \end{aligned} \tag{7}$$

where $\delta(\varepsilon) = \{\tau : |\tau - \tau_o| < \varepsilon\}$, it is sufficient to show that the right-hand side in (7) converges to zero.

The following lemma gives the limiting distribution of the OLS estimator of b.

Lemma 1 *When $\tau = \tau_o$,*

$$D_T(\hat{b}_{\tau_o} - b) \Rightarrow \left(\int_0^1 X_{\tau_o}(r) X'_{\tau_o}(r) dr\right)^{-1} \left(\int_0^1 X_{\tau_o}(r) dB_2(r) + [0, 0, \gamma'_{21}, (1-\tau_o)\gamma'_{21}]'\right) \equiv \eta_{\tau_o}, \tag{8}$$

while for $\tau \neq \tau_o$,

$$T^{-1/2} D_T(\hat{b}_\tau - b) \Rightarrow -\left(\int_0^1 X_\tau(r) X'_\tau(r) dr\right)^{-1} \int_0^1 X_\tau(r) \nabla X'_{2\tau}(r) dr \; b_2 \equiv \eta_\tau, \tag{9}$$

where $\nabla X_{2\tau}(r) = [\varphi_\tau(r) - \varphi_{\tau_o}(r), B'_2(r)(\varphi_\tau(r) - \varphi_{\tau_o}(r))]'$ and $b_2 = [\mu_2, \beta'_{2o}]'$.

Proof of Lemma 1: (8) is obtained in the same way as Park and Phillips (1988). To prove (9), note that

$$T^{-1/2} D_T(\hat{b}_\tau - b) = \left(D_T^{-1} \sum_{t=1}^T x_{t\tau} x'_{t\tau} D_T^{-1}\right)^{-1} \left(T^{-1/2} D_T^{-1} \sum_{t=1}^T x_{t\tau} v_{1t\tau}\right). \tag{10}$$

Using the FCLT and the continuous mapping theorem (CMT), we have $D_T^{-1} \sum_{t=1}^T x_{t\tau} x'_{t\tau} D_T^{-1} \Rightarrow \int_0^1 X_\tau(r) X'_\tau(r) dr$ uniformly over τ. On the other hand, the term in the last parentheses on the right hand side of (10) becomes

$$T^{-1/2} D_T^{-1} \sum_{t=1}^T x_{t\tau} v_{1t\tau}$$

$$= \left[T^{-1}\sum_{t=1}^{T} v_{1t\tau}, T^{-1}\sum_{t=1}^{T} v_{1t\tau}\varphi_{t\tau}, T^{-3/2}\sum_{t=1}^{T} y_{2t}' v_{1t\tau}, T^{-3/2}\sum_{t=1}^{T} y_{2t}' v_{1t\tau}\varphi_{t\tau}\right]'. \quad (11)$$

Since

$$v_{1t\tau} = v_{1t} - b_2'\nabla x_{2t\tau},$$

where $\nabla x_{2t\tau} = [\varphi_{t\tau} - \varphi_{t\tau_o}, T^{-1/2}y_{2t}'(\varphi_{t\tau} - \varphi_{t\tau_o})]'$, we have

$$\begin{aligned} T^{-1}\sum_{t=1}^{[Tr]} v_{1t\tau} &= T^{-1}\sum_{t=1}^{[Tr]} v_{1t} - b_2' T^{-1}\sum_{t=1}^{[Tr]} \nabla x_{2t\tau} \quad (12) \\ &\Rightarrow -b_2'\int_0^r \nabla X_{2\tau}(s)ds, \end{aligned}$$

$$\begin{aligned} T^{-3/2}\sum_{t=1}^{[Tr]} y_{2t} v_{1t\tau} &= T^{-3/2}\sum_{t=1}^{[Tr]} y_{2t} v_{1t} - T^{-3/2}\sum_{t=1}^{[Tr]} y_{2t}\nabla x_{2t\tau}' b_2 \quad (13) \\ &\Rightarrow -\int_0^r B_2(s)\nabla X_{2\tau}'(s)ds\; b_2, \end{aligned}$$

for $0 \le r \le 1$. Using these results, we obtain (9).□

Next, we investigate the asymptotic behavior of $S_T(\tau) - S_T(\tau_o)$ on $\tau \in \mathcal{T} \setminus \delta(\varepsilon)$. We expand $S_T(\tau)$ and $S_T(\tau_o)$ as

$$\begin{aligned} S_T(\tau) &= T^{-1}\sum_{t=1}^{T}(y_{1t} - \hat{b}_\tau' x_{t\tau})^2 \\ &= T^{-1}\sum_{t=1}^{T}(b' x_{t\tau_o} + v_{1t} - \hat{b}_\tau' x_{t\tau} + b' x_{t\tau} - b' x_{t\tau})^2 \\ &= T^{-1}\sum_{t=1}^{T}\{v_{1t} - (\hat{b}_\tau - b)' D_T D_T^{-1} x_{t\tau} - b_2'\nabla x_{2t\tau}\}^2 \\ &= T^{-1}\sum_{t=1}^{T} v_{1t}^2 + T^{-1/2}(\hat{b}_\tau - b)' D_T \left(D_T^{-1}\sum_{t=1}^{T} x_{t\tau}x_{t\tau}' D_T^{-1}\right) T^{-1/2} D_T(\hat{b}_\tau - b) \\ &\quad + T^{-1} b_2'\sum_{t=1}^{T}\nabla x_{2t\tau}\nabla x_{2t\tau}' b_2 + 2T^{-1/2}(\hat{b}_\tau - b)' D_T\left(T^{-1/2}D_T^{-1}\sum_{t=1}^{T} x_{t\tau}\nabla x_{2t\tau}' b_2\right) \\ &\quad - 2\left(T^{-1/2}D_T^{-1}\sum_{t=1}^{T} x_{t\tau}v_{1t}\right)' T^{-1/2} D_T(\hat{b}_\tau - b) - 2\left(T^{-1}\sum_{t=1}^{T}\nabla x_{2t\tau}v_{1t}\right)' b_2 \\ &\equiv S_{0T} + S_{1T} + S_{2T} + S_{3T} - S_{4T} - S_{5T}, \text{ say,} \end{aligned}$$

and

$$\begin{aligned} S_T(\tau_o) &= T^{-1}\sum_{t=1}^{T}(y_{1t}-\hat{b}'_{\tau_o}x_{t\tau_o})^2 \\ &= T^{-1}\sum_{t=1}^{T}\{v_{1t}-(\hat{b}_{\tau_o}-b)'D_TD_T^{-1}x_{t\tau_o}\}^2 \\ &= T^{-1}\sum_{t=1}^{T}v_{1t}^2-2T^{-1}(\hat{b}_{\tau_o}-b)'D_T\left(D_T^{-1}\sum_{t=1}^{T}x_{t\tau_o}v_{1t}\right) \\ &\quad +T^{-1}(\hat{b}_{\tau_o}-b)'D_T\left(D_T^{-1}\sum_{t=1}^{T}x_{t\tau_o}x'_{t\tau_o}D_T^{-1}\right)D_T(\hat{b}_{\tau_o}-b) \\ &\equiv S_{0T}-S_{6T}+S_{7T}, \text{ say.} \end{aligned}$$

We then have

$$S_T(\tau)-S_T(\tau_o)=S_{1T}+S_{2T}+S_{3T}-S_{4T}-S_{5T}+S_{6T}-S_{7T}. \tag{14}$$

In the following, we will show that $S_{1T}+S_{2T}+S_{3T}$ converges in distribution to a random variable that is positive almost surely (a.s.) while the rest of (14) converges to zero in probability.

Using (9) we have

$$\begin{aligned} S_{1T}+S_{2T}+S_{3T} \;\Rightarrow\; & \eta'_\tau\int_0^1 X_\tau(r)X'_\tau(r)dr\;\eta_\tau+b'_2\int_0^1\nabla X_2(r)\nabla X'_2(r)dr\;b_2 \\ & +2\eta'_\tau\int_0^1 X_\tau(r)\nabla X'_2(r)dr\;b_2 \\ = & \int_0^1\left(\eta'_\tau X_\tau(r)+b'_2\nabla X_2(r)\right)^2dr, \end{aligned}$$

while we can see that S_{6T} and S_{7T} are $O_p(T^{-1})$ since $D_T(\hat{b}_{\tau_o}-b)=O_p(1)$ as shown in Lemma 1. On the other hand, since S_{4T} and S_{5T} are expressed as

$$S_{4T}=2\left[T^{-1}\sum_{t=1}^{T}v_{1t},T^{-1}\sum_{t=1}^{T}v_{1t}\varphi_{t\tau},T^{-3/2}\sum_{t=1}^{T}v_{1t}y'_{2t},T^{-3/2}\sum_{t=1}^{T}v_{1t}y'_{2t}\varphi_{t\tau}\right]T^{-1/2}D_T(\hat{b}_\tau-b),$$

$$S_{5T}=2\left[T^{-1}\sum_{t=1}^{T}v_{1t}(\varphi_{t\tau}-\varphi_{t\tau_o}),T^{-3/2}\sum_{t=1}^{T}y'_{2t}v_{1t}(\varphi_{t\tau}-\varphi_{t\tau_o})\right]b_2,$$

we can see that both terms are $O_p(T^{-1/2})$. Since these convergences hold uniformly over τ, we have

$$\inf_{\tau\in\mathcal{T}\setminus\delta(\varepsilon)} S_T(\tau) - S_T(\tau_o) \Rightarrow \inf_{\tau\in\mathcal{T}\setminus\delta(\varepsilon)} \int_0^1 \left(\eta_\tau' X_\tau(r) + b_2'\nabla X_2(r)\right)^2 dr > 0 \ \text{(a.s.)}, \tag{15}$$

which implies that (7) converges to zero as T goes to infinity.□

Proof of Proposition 3: Since the consistency of $\hat{\tau}$ is obtained in Proposition 2, we can restrict the range of τ only to the vicinity of τ_o that shrinks to τ_o. More precisely, for a given $\epsilon > 0$, we define a sequence of positive real numbers, $r_T(\varepsilon)$, such that

$$r_T(\varepsilon) = \inf_r \left\{r : P(|\hat{\tau} - \tau_o| \le r) \ge 1 - \varepsilon\right\},$$

and consider only τ that satisfies $|\tau - \tau_o| \le r_T(\varepsilon)$. Since $\hat{\tau}$ is a consistent estimator, $r_T(\varepsilon)$ goes to zero as $T \to \infty$. Without loss of generality, we assume that $Tr_T(\varepsilon)$ goes to infinity as $T \to \infty$. This property, in fact, holds if we redefine $r_T(\varepsilon)$ as $\max(r_T(\varepsilon), T^{-a})$ for some $0 < a < 1$. We abbreviate $r_T(\varepsilon)$ as r_T for simplicity. We also reparameterize the break fraction as $\tau = \tau_o + cT^{-1/2}$. Since we are considering only the vicinity of τ_o, the possible range of c is $\mathcal{C} = \{c : |c| \le T^{1/2} r_T\}$.

In the following, we will show that, for every $c_o > 0$, $T^{1/2}(S_T(\tau) - S_T(\tau_o))$ is asymptotically positive (a.s.) uniformly over $c \in \mathcal{C} \setminus \delta(c_o)$ where $\delta(c_o) = \{c : |c| < c_o\}$. This implies that $T^{1/2}(S_T(\tau) - S_T(\tau_o))$ does not take its minimum on $\mathcal{C} \setminus \delta(c_o)$, so that $\hat{c} = T^{1/2}(\hat{\tau} - \tau_o)$ converges to zero in probability.

Lemma 2 *The following results hold uniformly over $c \in \mathcal{C} \setminus \delta(c_o)$.*

$$T^{-3/2} \sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t} \ \stackrel{d}{=}\ |c| T^{-1/2} \left(B_2(\tau_o) + o_p(1)\right), \tag{16}$$

$$T^{-2} \sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t} y_{2t}' \ \stackrel{d}{=}\ |c| T^{-1/2} \left(B_2(\tau_o) B_2(\tau_o)' + o_p(1)\right), \tag{17}$$

$$T^{-1/2} \sum_{t=[T\tau_o]+1}^{[T\tau]} v_{1t} \ =\ o_p(1). \tag{18}$$

$$T^{-1}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t}v_{1t} \;=\; o_p(1). \tag{19}$$

Proof of Lemma 2: We proceed with the proof for $\tau > \tau_o$. The case where $\tau < \tau_o$ is treated in the same way. Since $y_{2[T\tau_o]+t} = y_{2[T\tau_o]} + \sum_{j=1}^{t} v_{2[T\tau_o]+j}$ from the definition of y_{2t}, we have

$$\begin{aligned}
T^{-3/2}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t} &= T^{-3/2}\sum_{t=1}^{[T\tau]-[T\tau_o]}\left(y_{2[T\tau_o]} + \sum_{j=1}^{t} v_{2[T\tau_o]+j}\right) \\
&= (\tau-\tau_o)T^{-1/2}y_{2[T\tau_o]} + T^{-3/2}(Tr_T)^{1/2}\sum_{t=1}^{[T\tau]-[T\tau_o]}(Tr_T)^{-1/2}\sum_{j=1}^{t} v_{2[T\tau_o]+j} \\
&= (\tau-\tau_o)T^{-1/2}y_{2[T\tau_o]} + T^{-1}r_T^{1/2}([T\tau]-[T\tau_o])O_p(1) \\
&\overset{d}{=} (\tau-\tau_o)(B_2(\tau_o)+o_p(1)).
\end{aligned}$$

The second last equality is established because $(Tr_T)^{-1/2}\sum_{j=1}^{t} v_{2[T\tau_o]+j}$ is $O_p(1)$ uniformly over $\mathcal{C}$ since $|[T\tau]-[T\tau_o]| \le Tr_T$ on $\mathcal{C}$. We also used the fact that $r_T \to 0$ so that $r_T \times O_p(1) = o_p(1)$. Since $\tau - \tau_o = cT^{-1/2}$, (16) is obtained.

Similarly, we can see that

$$\begin{aligned}
T^{-2}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t}y_{2t}' &= T^{-2}\sum_{t=1}^{[T\tau]-[T\tau_o]}\left(y_{2[T\tau_o]} + \sum_{j=1}^{t} v_{2[T\tau_o]+j}\right)\left(y_{2[T\tau_o]} + \sum_{j=1}^{t} v_{2[T\tau_o]+j}\right)' \\
&= (\tau-\tau_o)T^{-1}y_{2[T\tau_o]}y_{2[T\tau_o]}' \\
&\quad + T^{-1/2}y_{2[T\tau_o]}T^{-3/2}(Tr_T)^{1/2}\sum_{t=1}^{[T\tau]-[T\tau_o]}(Tr_T)^{-1/2}\sum_{j=1}^{t} v_{2[T\tau_o]+j}' \\
&\quad + T^{-3/2}(Tr_T)^{1/2}\sum_{t=1}^{[T\tau]-[T\tau_o]}(Tr_T)^{-1/2}\sum_{j=1}^{t} v_{2[T\tau_o]+j}(T^{-1/2}y_{2[T\tau_o]})' \\
&\quad + T^{-2}(Tr_T)\sum_{t=1}^{[T\tau]-[T\tau_o]}\left((Tr_T)^{-1/2}\sum_{j=1}^{t} v_{2[T\tau_o]+j}\right)\left((Tr_T)^{-1/2}\sum_{j=1}^{t} v_{2[T\tau_o]+j}\right)' \\
&= (\tau-\tau_o)(T^{-1/2}y_{2[T\tau_o]})(T^{-1/2}y_{2[T\tau_o]}') + T^{-1/2}y_{2[T\tau_o]}r_T^{1/2}(\tau-\tau_o)O_p(1) \\
&\quad + r_T^{1/2}(\tau-\tau_o)O_p(1)(T^{-1/2}y_{2[T\tau_o]})' + (\tau-\tau_o)r_TO_p(1) \\
&\overset{d}{=} (\tau-\tau_o)(B_{2\tau_o}B_{2\tau_o}' + o_p(1)),
\end{aligned}$$

from which (17) is obtained.

For (18) and (19) we have

$$\begin{aligned} T^{-1/2}\sum_{t=[T\tau_o]+1}^{[T\tau]} v_{1t} &= r_T^{1/2}(Tr_T)^{-1/2}\sum_{t=1}^{[T\tau]-[T\tau_o]} v_{1[T\tau_o]+t} \\ &= r_T^{1/2}O_p(1), \end{aligned}$$

$$\begin{aligned} T^{-1}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t}v_{1t} &= T^{-1}\sum_{t=1}^{[T\tau]-[T\tau_o]}\left(y_{2[T\tau_o]}+\sum_{j=1}^{t} v_{2[T\tau_o]+j}\right)v_{1[T\tau_o]+t} \\ &= r_T^{1/2}(T^{-1/2}y_{2[T\tau_o]})(Tr_T)^{-1/2}\sum_{t=1}^{[T\tau]-[T\tau_o]} v_{1[T\tau_o]+t} \\ &\quad +r_T(Tr_T)^{-1}\sum_{t=1}^{[T\tau]-[T\tau_o]}\left(\sum_{j=1}^{t} v_{2[T\tau_o]+j}\right)v_{1[T\tau_o]+t} \\ &= r_T^{1/2}O_p(1)+r_TO_p(1). \end{aligned}$$

Again, since $r_T \to 0$, these equations imply (18) and (19).□

Next, we investigate the asymptotic property of $\hat{b}_\tau$ on $\mathcal{C}\setminus\delta(c_o)$. Again, we proceed with the proof for $\tau > \tau_o$. We first investigate the asymptotic behavior of $T^{-1/2}D_T^{-1}\sum_{t=1}^{T} x_{t\tau}v_{1t\tau}$, which is expressed as (11). Here note that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\nabla x'_{2t\tau} &= -\left[\tau-\tau_o,\ T^{-3/2}\sum_{t=[T\tau_o]+1}^{[T\tau]} y'_{2t}\right] \\ &= -cT^{-1/2}[1, O_p(1)] \end{aligned}$$

from (16), and then, using expression (12), we have

$$T^{-1}\sum_{t=1}^{T} v_{1t\tau} = O_p(T^{-1/2}) + cO_p(T^{-1/2}). \tag{20}$$

Similarly, from (16) and (17) we have

$$\begin{aligned} T^{-3/2}\sum_{t=1}^{T} y_{2t}\nabla x'_{2t\tau} &= -\left[T^{-3/2}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t},\ T^{-2}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t}y'_{2t}\right] \\ &= -cT^{-1/2}[O_p(1), O_p(1)], \end{aligned}$$

and then, using expression (13),

$$T^{-3/2}\sum_{t=1}^{T} y_{2t}v_{1t\tau} = O_p(T^{-1/2}) + cO_p(T^{-1/2}). \tag{21}$$

We can easily see that $T^{-1}\sum_{t=1}^{T} v_{1t\tau}\varphi_{t\tau}$ and $T^{-3/2}\sum_{t=1}^{T} y_{2t}v_{1t\tau}$ have the same orders as (20) and (21). Then, from expression (11), we have

$$T^{-1/2}D_T^{-1}\sum_{t=1}^{T} x_{t\tau}v_{1t\tau} = O_p(T^{-1/2}) + cO_p(T^{-1/2}).$$

Since the term in the first parentheses on the right hand side of (10) converges in distribution uniformly as was seen in the proof of Lemma 1, we have

$$T^{1/4}T^{-1/2}D_T(\hat{b}_\tau - b) = O_p(T^{-1/4}) + cO_p(T^{-1/4}). \tag{22}$$

Using this result, we get

$$\begin{aligned} T^{1/2}S_{1T} &= c^2O_p(T^{-1/2}) + cO_p(T^{-1/2}) + O_p(T^{-1/2}) \\ &= cO_p(r_T) + O_p(r_T) + O_p(T^{-1/2}), \end{aligned}$$

where the second equality holds because $|c|T^{-1/2} = |\tau - \tau_o| \leq r_T$ so that $cT^{-1/2} = O(r_T)$.

Similarly, using Lemma 2 and (22) we have

$$\begin{aligned} T^{1/2}S_{2T} &= T^{1/2}b_2' \begin{bmatrix} |\tau-\tau_o| & -T^{-3/2}\sum_{[T\tau_o]+1}^{[T\tau]} y_{2t}' \\ -T^{-3/2}\sum_{[T\tau_o]+1}^{[T\tau]} y_{2t} & T^{-2}\sum_{[T\tau_o]+1}^{[T\tau]} y_{2t}y_{2t}' \end{bmatrix} b_2 \\ &\stackrel{d}{=} T^{1/2}b_2' \begin{bmatrix} cT^{-1/2} & -cT^{-1/2}(B_2'(\tau_o) + o_p(1)) \\ -cT^{-1/2}(B_2(\tau_o) + o_p(1)) & cT^{-1/2}(B_2(\tau_o)B_2'(\tau_o) + o_p(1)) \end{bmatrix} b_2 \\ &= c\left\{(\mu_2 - \beta_2'B_{2\tau_o})^2 + o_p(1)\right\}, \end{aligned}$$

$$\begin{aligned} T^{1/2}S_{3T} &= 2T^{1/4}T^{-1/2}(\hat{b}_\tau - b)'D_T\left(T^{1/4}T^{-1/2}D_T^{-1}\sum_{t=1}^{T} x_{t\tau}\nabla x_{2t\tau}'b_2\right) \\ &= (O_p(T^{-1/4}) + cO_p(T^{-1/4})) \times (cO_p(T^{-1/4})) \\ &= O_p(r_T) + cO_p(r_T), \end{aligned}$$

$$
\begin{aligned}
T^{1/2}S_{4T} &= 2T^{1/4}T^{-1/2}(\hat{b}_\tau - b)'D_T\left(T^{1/4}T^{-1/2}D_T^{-1}\sum_{t=1}^{T}x_{t\tau}v_{1t}\right) \\
&= (O_p(T^{-1/4}) + cO_p(T^{-1/4})) \times O_p(T^{-1/4}) \\
&= O_p(T^{-1/2}) + O_p(r_T),
\end{aligned}
$$

$$
\begin{aligned}
T^{1/2}S_{5T} &= 2\left[T^{-1/2}\sum_{t=[T\tau_o]+1}^{[T\tau]} v_{1t}, T^{-1}\sum_{t=[T\tau_o]+1}^{[T\tau]} y_{2t}'v_{1t}\right]b_2 \\
&= o_p(1).
\end{aligned}
$$

Since S_{6T} and S_{7T} do not depend on τ, they are $o_p(T^{-1})$ uniformly over τ. Then, by combining these results, we get

$$
T^{1/2}(S_T(\tau) - S_T(\tau_o)) \stackrel{d}{=} c\left\{(\mu_2 - \beta_2'B_{2\tau_o})^2 + o_p(1)\right\} + o_p(1).
$$

Note that $c > 0$ because $\tau > \tau_o$. Since

$$
c(\mu_2 - \beta_2'B_{2\tau_o})^2 \geq c_o(\mu_2 - \beta_2'B_{2\tau_o})^2 > 0 \ (a.s.)
$$

and the middle term in the above inequality does not depend on c, we can see that $S_T(\tau) - S_T(\tau_o)$ is asymptotically positive (a.s.) over $\mathcal{C} \setminus \delta(c_o)$. This implies $T^{1/2}(\hat{\tau} - \tau_o)$ converges to zero in probability.□

Proof of Proposition 4: We first prove the following lemma.

Lemma 3 *Assume that* $T^{1/2}(\hat{\tau} - \tau_o) \xrightarrow{p} 0$. *Then, for* $0 \leq r \leq 1$,
(i) $T^{-1}\sum_{t=1}^{[Tr]} y_{2t}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) \xrightarrow{p} 0$.
(ii) $T^{-3/2}\sum_{t=1}^{[Tr]} y_{2t}y_{2t}'(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) \xrightarrow{p} 0$.
(iii) $T^{-1/2}\sum_{t=1}^{[Tr]} v_t(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) \xrightarrow{p} 0$.
(iv) $T^{-1}\sum_{t=1}^{[Tr]} y_{2t}v_t'(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) \xrightarrow{p} 0$.
(v) $D_T(\hat{b}_{\hat{\tau}} - \hat{b}_{\tau_o}) \xrightarrow{p} 0$.
Convergences (i)-(iv) hold uniformly over $0 \leq r \leq 1$.

Proof of Lemma 3: (i) We have

$$\left| T^{-1} \sum_{t=1}^{[Tr]} y_{2t}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) \right| \leq \sup_{0 \leq r \leq 1} \left| T^{-1/2} y_{2[Tr]} \right| \left| T^{1/2}(\hat{\tau} - \tau_o) \right| \xrightarrow{p} 0.$$

(ii)-(iv) can be proved in the same way as (i).

(v) is proved if we show that

$$D_T^{-1} \left(\sum_{t=1}^{T} x_{t\hat{\tau}} x'_{t\hat{\tau}} - \sum_{t=1}^{T} x_{t\tau_o} x'_{t\tau_o} \right) D_T^{-1} \xrightarrow{p} 0, \tag{23}$$

and

$$D_T^{-1} \left(\sum_{t=1}^{T} x_{t\hat{\tau}} v_{1t\hat{\tau}} - \sum_{t=1}^{T} x_{t\tau_o} v_{1t\tau_o} \right) \xrightarrow{p} 0. \tag{24}$$

Noting that

$$x_{t\hat{\tau}} x'_{t\hat{\tau}} - x_{t\tau_o} x'_{t\tau_o} = (x_{t\hat{\tau}} - x_{t\tau_o}) x'_{t\hat{\tau}} + x_{t\tau_o}(x_{t\hat{\tau}} - x_{t\tau_o})'$$

and

$$(x_{t\hat{\tau}} - x_{t\tau_o})' = [0, (\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}), 0, y'_{2t}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o})], \tag{25}$$

we can show (23) using Lemma 3 (i) and (ii). Similarly, we have

$$D_T^{-1} \sum_{t=1}^{T} (x_{t\hat{\tau}} v_{1t\hat{\tau}} - x_{t\tau_o} v_{1t\tau_o}) = D_T^{-1} \sum_{t=1}^{T} x_{t\hat{\tau}} (v_{1t\hat{\tau}} - v_{1t\tau_o}) + D_T^{-1} \sum_{t=1}^{T} (x_{t\hat{\tau}} - x_{t\tau_o}) v_{1t\tau_o}.$$

Since $v_{1t\hat{\tau}} - v_{1t\tau_o} = -b'_2 \nabla x_{2t\hat{\tau}}$, the first term converges to zero in probability by using Lemma 3 (i) and (ii). Similarly, the second term is shown to be $o_p(1)$ using expression (25) and Lemma 3 (iii) and (iv). Then, convergence (24) is established.□

Since $\hat{v}_{1t\hat{\tau}} = v_{1t} - (\hat{b}_{\hat{\tau}} - b)' D_T D_T^{-1} x_{t\hat{\tau}} - b'_2 \nabla x_{2t\hat{\tau}}$, we have

$$\begin{aligned}
T^{-1} \sum_{t=1}^{T} \hat{v}_{1t\hat{\tau}}^2 &= T^{-1} \sum_{t=1}^{T} v_{1t}^2 + (\hat{b}_{\hat{\tau}} - b)' D_T \left(T^{-1} D_T^{-1} \sum_{t=1}^{T} x_{t\hat{\tau}} x'_{t\hat{\tau}} D_T^{-1} \right) D_T (\hat{b}_{\hat{\tau}} - b) \\
&\quad + b'_2 \left(T^{-1} \sum_{t=1}^{T} \nabla x_{2t\hat{\tau}} \nabla x'_{2t\hat{\tau}} \right) b_2 - 2(\hat{b}_{\hat{\tau}} - b)' D_T T^{-1} D_T^{-1} \sum_{t=1}^{T} x_{t\hat{\tau}} v_{1t} \\
&\quad - 2b'_2 T^{-1} \sum_{t=1}^{T} \nabla x_{2t\hat{\tau}} v_{1t} + 2(\hat{b}_{\hat{\tau}} - b)' D_T T^{-1} D_T^{-1} \sum_{t=1}^{T} x_{t\hat{\tau}} \nabla x'_{2t\hat{\tau}} b_2 \\
&= T^{-1} \sum_{t=1}^{n} v_{1t}^2 + R_{1T} + R_{2T} + R_{3T} + R_{4T} + R_{5T}, \text{ say.}
\end{aligned}$$

We can see that R_{2T}, R_{4T}, and R_{5T} are $o_p(T^{-1/2})$ from Lemma 3. On the other hand, it is seen that $D_T^{-1}\sum_{t=1}^T x_{t\hat{\tau}}x_{t\hat{\tau}}'D_T^{-1}$ is bounded in probability, so that R_{1T} is of order T^{-1} since $D_T(\hat{b}_{\hat{\tau}}-b) \Rightarrow \eta_{\tau_o}$ from (8) and Lemma 3 (v). Similarly, since $D_T^{-1}\sum_{t=1}^T x_{t\hat{\tau}}v_{1t} = O_p(1)$, we can see that $R_{3T} = O_p(T^{-1})$. Then, we showed that $T^{-1}\sum_{t=1}^T \hat{v}_{1t\hat{\tau}}^2 = T^{-1}\sum_{t=1}^T v_{1t}^2 + o_p(T^{-1/2})$. Since $T^{-1}\sum_{t=1}^T v_{1t}^2 \xrightarrow{p} \sigma_{11}$ by the weak law of large numbers, we have $T^{-1}\sum_{t=1}^T \hat{v}_{1t\hat{\tau}}^2 \xrightarrow{p} \sigma_{11}$. Similarly we have $T^{-1}\sum_{t=1}^T v_{2t}\hat{v}_{1t\hat{\tau}} = T^{-1}\sum_{t=1}^T v_{2t}v_{1t} + o_p(T^{-1/2}) \xrightarrow{p} \sigma_{21}$, so that we obtain $T^{-1}\sum_{t=1}^T \hat{v}_{t\hat{\tau}}\hat{v}_{t\hat{\tau}}' \xrightarrow{p} \Sigma$.

In exactly the same way, we see that

$$T^{-1}\sum_{t=1}^{T-j}\hat{v}_{t\hat{\tau}}\hat{v}_{t+j\hat{\tau}} = T^{-1}\sum_{t=1}^{T-j} v_t v_{t+j}' + o_p(T^{-1/2})$$

for a given j. This implies

$$\begin{aligned}\hat{\Lambda}_{\hat{\tau}} - T^{-1}\sum_{j=1}^{\ell}k(j/\ell)\sum_{t=1}^{T-j}v_t v_{t+j}' &= \sum_{j=1}^{\ell}k(j/\ell)\times o_p(T^{-1/2}) \\ &= o_p(\ell/T^{1/2}).\end{aligned}$$

Since $\ell = o(T^{1/2})$ by Assumption 2 (b), the last term converges to zero in probability, which implies $\hat{\Lambda}_{\hat{\tau}} \xrightarrow{p} \Lambda$.

In the same way, we can show the consistency of $\hat{\Gamma}_{\hat{\tau}}$ and $\hat{\Omega}_{\hat{\tau}}$.□

Proof of Theorem 1: In the following, we replace the estimators of the long-run variances by the true ones without loss of generality because they are consistent estimators from Proposition 4. Similar to the proof of Lemma 3 (v), it is enough to show that

$$\begin{aligned}&D_T^{-1}\left(\sum_{t=1}^T x_{t\hat{\tau}}^* x_{t\hat{\tau}}^{*\prime} - \sum_{t=1}^T x_{t\tau_o}^* x_{t\tau_o}^{*\prime}\right)D_T^{-1} \\ &= D_T^{-1}\sum_{t=1}^T(x_{t\hat{\tau}}^* - x_{t\tau_o}^*)x_{t\hat{\tau}}^{*\prime}D_T^{-1} + D_T^{-1}\sum_{t=1}^T x_{t\tau_o}^*(x_{t\hat{\tau}}^* - x_{t\tau_o}^{*\prime})D_T^{-1} \xrightarrow{p} 0, \qquad (26)\end{aligned}$$

and

$$\begin{aligned}&D_T^{-1}\left(\sum_{t=1}^T x_{t\hat{\tau}}^* e_{t\hat{\tau}}^* - \sum_{t=1}^T x_{t\tau_o}^* e_{t\tau_o}^*\right) \\ &= D_T^{-1}\sum_{t=1}^T x_{t\hat{\tau}}^*(e_{t\hat{\tau}}^* - e_{t\tau_o}^*) + D_T^{-1}\sum_{t=1}^T(x_{t\hat{\tau}}^* - x_{t\tau_o}^*)e_{t\tau_o}^* \xrightarrow{p} 0. \qquad (27)\end{aligned}$$

From the definitions of $x^*_{t\hat{\tau}}$ and $x^*_{t\tau_o}$ we see that

$$(x^*_{t\hat{\tau}} - x^*_{t\tau_o})' = [0, (\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}), y^{*\prime}_{2t\hat{\tau}} - y^{*\prime}_{2t\tau_o}, y^{*\prime}_{2t\hat{\tau}}\varphi_{t\hat{\tau}} - y^{*\prime}_{2t\tau_o}\varphi_{t\tau_o}]. \tag{28}$$

The third element of (28) is expressed as

$$y^*_{2t\hat{\tau}} - y^*_{2t\tau_o} = -\Gamma_2'\Sigma^{-1}(\hat{v}_{t\hat{\tau}} - \hat{v}_{t\tau_o}), \tag{29}$$

where the last m rows of $\hat{v}_{t\hat{\tau}} - \hat{v}_{t\tau_o}$ are apparently zero while the first element of it can be expressed as

$$\hat{v}_{1t\hat{\tau}} - \hat{v}_{1t\tau_o} = -(\hat{b}_{\hat{\tau}} - b)'D_T D_T^{-1}(x_{t\hat{\tau}} - x_{t\tau_o}) - (\hat{b}_{\hat{\tau}} - \hat{b}_{\tau_o})'D_T D_T^{-1}x_{t\tau_o} - b_2'\nabla x_{2t\hat{\tau}} \tag{30}$$

because $\hat{v}_{1t\hat{\tau}} = v_{1t} - (\hat{b}_{\hat{\tau}} - b)'x_{t\hat{\tau}} - b_2'\nabla x_{2t\hat{\tau}}$ and $\hat{v}_{1t\tau_o} = v_{1t} - (\hat{b}_{\tau_o} - b)'x_{t\tau_o}$. Similarly, the fourth element of (28) is expressed as

$$y^*_{2t\hat{\tau}}\varphi_{t\hat{\tau}} - y^*_{2t\tau_o}\varphi_{t\tau_o} = y_{2t}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) - \Gamma_2'\Sigma^{-1}(\hat{v}_{t\hat{\tau}}\varphi_{t\hat{\tau}} - \hat{v}_{t\tau_o}\varphi_{t\tau_o}), \tag{31}$$

and the first element of $(\hat{v}_{t\hat{\tau}}\varphi_{t\hat{\tau}} - \hat{v}_{t\tau_o}\varphi_{t\tau_o})$ becomes

$$\begin{aligned}
&\hat{v}_{1t\hat{\tau}}\varphi_{t\hat{\tau}} - \hat{v}_{1t\tau_o}\varphi_{t\tau_o} \\
&= v_{1t}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) - (\hat{b}_{\hat{\tau}} - b)'D_T D_T^{-1}(x_{t\hat{\tau}} - x_{t\tau_o})\varphi_{t\hat{\tau}} - (\hat{b}_{\hat{\tau}} - \hat{b}_{\tau_o})'D_T D_T^{-1}x_{t\tau_o}\varphi_{t\hat{\tau}} \\
&\quad -(\hat{b}_{\tau_o} - b)'D_T D_T^{-1}x_{t\tau_o}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) - b_2'\nabla x_{2t\hat{\tau}}\varphi_{t\hat{\tau}}.
\end{aligned} \tag{32}$$

By carefully checking each element of (26) using expressions (28)–(32), we can show by Lemma 3 that the convergence of (26) is established.

For (27), noting that

$$e^*_{t\hat{\tau}} = v_{1t} - \omega_{21}'\Omega_{22}^{-1}v_{2t} - (\hat{\beta}_{1\hat{\tau}} - \beta_1)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\hat{\tau}} - (\hat{\beta}_{2\hat{\tau}} - \beta_2)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\hat{\tau}}\varphi_{t\hat{\tau}} - b_2'\nabla x_{2t\hat{\tau}}$$

and

$$e^*_{t\tau_o} = v_{1t} - \omega_{21}'\Omega_{22}^{-1}v_{2t} - (\hat{\beta}_{1\tau_o} - \beta_1)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\tau_o} - (\hat{\beta}_{2\tau_o} - \beta_2)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\tau_o}\varphi_{t\tau_o},$$

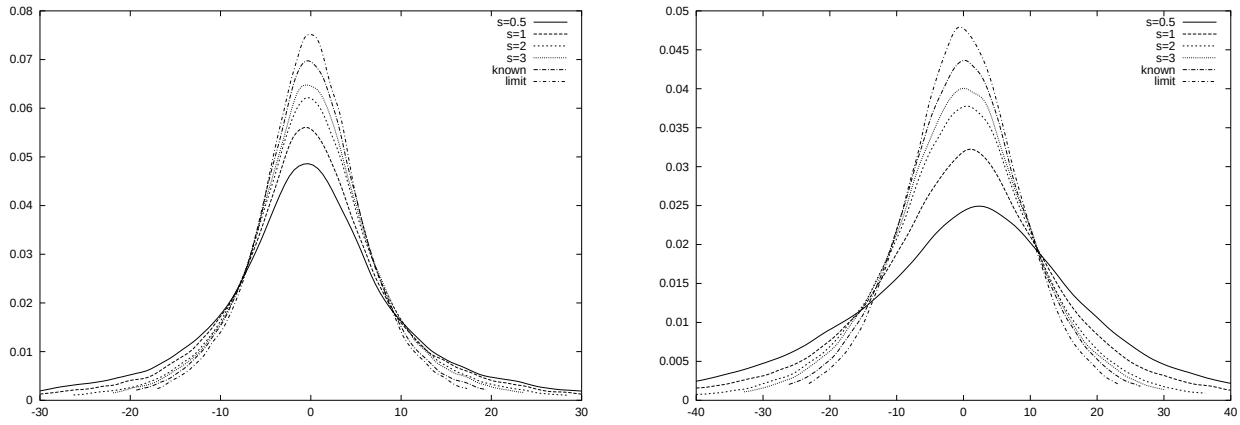
we can see that

$$\begin{aligned} e^*_{t\hat{\tau}} - e^*_{t\tau_o} &= -(\hat{\beta}_{1\hat{\tau}} - \beta_1)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\hat{\tau}} - (\hat{\beta}_{2\hat{\tau}} - \beta_2)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\hat{\tau}}\varphi_{t\hat{\tau}} - b_2'\nabla x_{2t\hat{\tau}} \\ &\quad +(\hat{\beta}_{1\tau_o} - \beta_1)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\tau_o} + (\hat{\beta}_{2\tau_o} - \beta_2)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\tau_o}\varphi_{t\tau_o} \\ &= -(\hat{\beta}_{1\hat{\tau}} - \beta_1)'\Gamma_2'\Sigma^{-1}(\hat{v}_{t\hat{\tau}} - \hat{v}_{t\tau_o}) - (\hat{\beta}_{1\hat{\tau}} - \hat{\beta}_{1\tau_o})'\Gamma_2'\Sigma^{-1}\hat{v}_{t\tau_o} \\ &\quad -(\hat{\beta}_{2\hat{\tau}} - \beta_2)'\Gamma_2'\Sigma^{-1}\hat{v}_{t\hat{\tau}}(\varphi_{t\hat{\tau}} - \varphi_{t\tau_o}) - (\hat{\beta}_{2\hat{\tau}} - \beta_2)'\Gamma_2'\Sigma^{-1}(\hat{v}_{t\hat{\tau}} - \hat{v}_{t\tau_o})\varphi_{t\tau_o} \\ &\quad -(\hat{\beta}_{2\hat{\tau}} - \hat{\beta}_{2\tau_o})'\Gamma_2'\Sigma^{-1}\hat{v}_{t\tau_o}\varphi_{t\tau_o} - b_2'\nabla x_{2t\hat{\tau}}. \end{aligned} \tag{33}$$

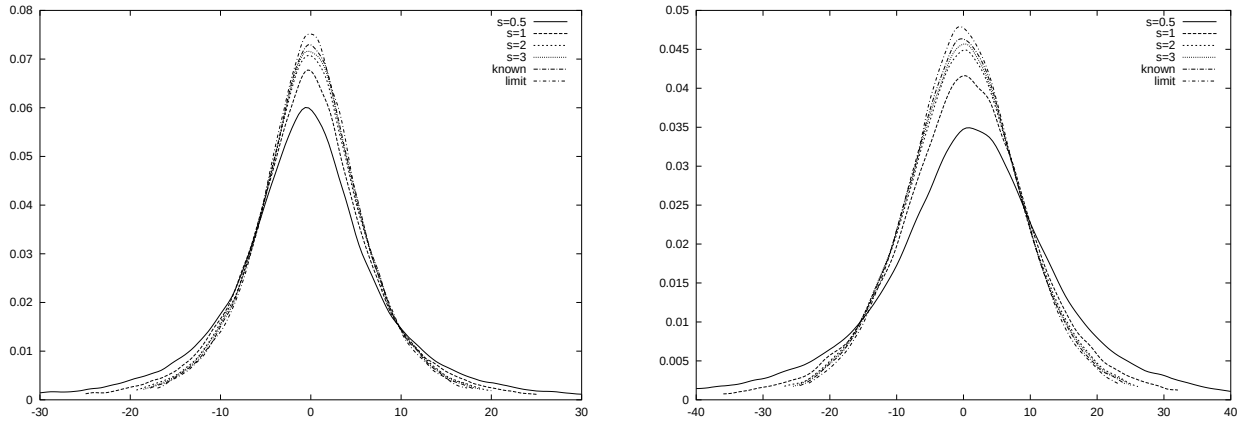
Again, by checking each element of (27) using expressions (28)–(33), the convergence of (27) is proved.□

Table 1. Size and power of the Wald test statistic

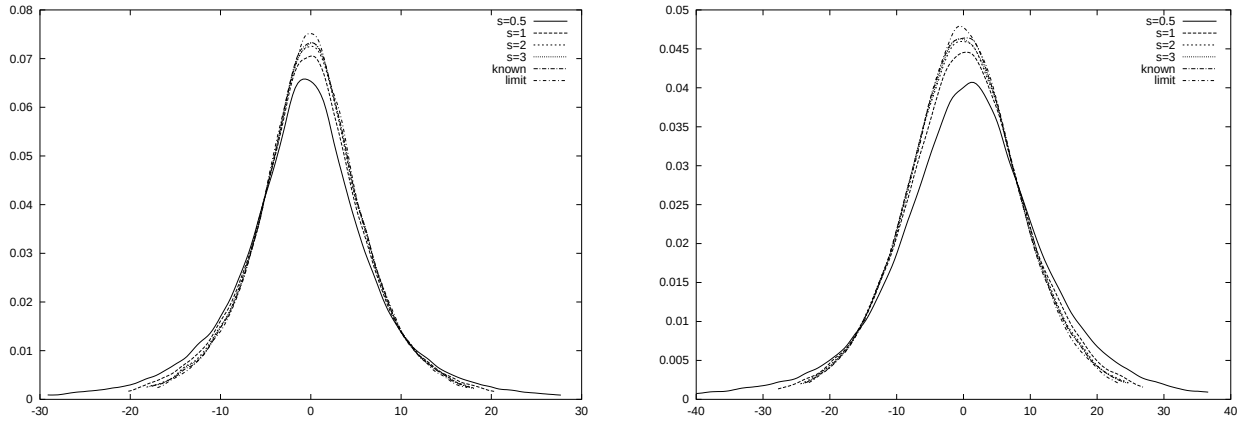
T	b	$W_T(\hat{b}^*_{\tau_o})$	$W_T(\hat{b}^*_{\hat{\tau}_o})$			
			$s=0.5$	$s=1$	$s=2$	$s=3$
		$\alpha=-0.6$				
100	1.00	0.119	0.269	0.220	0.166	0.151
	1.01	0.123	0.266	0.222	0.165	0.152
	1.05	0.247	0.352	0.324	0.272	0.257
	1.10	0.491	0.505	0.512	0.481	0.473
300	1.00	0.070	0.157	0.110	0.084	0.074
	1.01	0.115	0.214	0.162	0.133	0.125
	1.05	0.697	0.689	0.689	0.689	0.692
	1.10	0.961	0.900	0.931	0.951	0.957
500	1.00	0.066	0.119	0.090	0.073	0.070
	1.01	0.185	0.251	0.211	0.190	0.189
	1.05	0.926	0.888	0.910	0.921	0.922
	1.10	0.999	0.978	0.994	0.997	0.998
		$\alpha=0$				
100	1.00	0.089	0.243	0.196	0.140	0.118
	1.01	0.088	0.247	0.195	0.141	0.118
	1.05	0.225	0.362	0.316	0.261	0.249
	1.10	0.504	0.557	0.539	0.514	0.507
300	1.00	0.059	0.146	0.094	0.072	0.063
	1.01	0.111	0.205	0.145	0.123	0.116
	1.05	0.710	0.722	0.713	0.712	0.711
	1.10	0.967	0.931	0.954	0.963	0.967
500	1.00	0.055	0.109	0.073	0.061	0.058
	1.01	0.183	0.242	0.203	0.190	0.186
	1.05	0.931	0.916	0.922	0.928	0.930
	1.10	0.998	0.987	0.996	0.998	0.998
		$\alpha=0.6$				
100	1.00	0.213	0.411	0.355	0.286	0.250
	1.01	0.223	0.418	0.366	0.299	0.260
	1.05	0.389	0.536	0.506	0.462	0.430
	1.10	0.658	0.700	0.703	0.696	0.677
300	1.00	0.129	0.238	0.176	0.145	0.136
	1.01	0.201	0.311	0.252	0.215	0.208
	1.05	0.779	0.801	0.791	0.786	0.781
	1.10	0.976	0.954	0.968	0.974	0.975
500	1.00	0.108	0.165	0.129	0.114	0.108
	1.01	0.264	0.338	0.291	0.272	0.268
	1.05	0.949	0.945	0.948	0.947	0.949
	1.10	0.999	0.991	0.997	0.998	0.999



(i-a) $T = 100$ (ii-a) $T = 100$



(i-b) $T = 300$ (ii-b) $T = 300$



(i-c) $T = 500$ (ii-c) $T = 500$

Figure 1: Probability density functions of $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$

Note: (i-a)–(i-c) are the probability density functions of $T(\hat{\beta}_1 - \beta_1)$ while (ii-a)–(ii-c) are those of $T(\hat{\beta}_2 - \beta_2)$. In each figure 'limit' corresponds to the probability density function of the limiting distribution that is approximated by $T = 2,000$ observations, while 'known' and '$s = 0.5, \cdots, 3$' correspond to the finite sample distributions for the cases where the break point is known and unknown. These densities are drawn by the kernel method with a Gaussian kernel. The smoothing parameter, h, is decided by equation (3.31) in Silverman (1986): $h = 0.9AT^{-1/5}$ where $A = \min(\text{standard deviation, interquartile range}/1.34)$.