

## NOTES AND PROBLEMS

# FIXED- $b$ ASYMPTOTICS IN SINGLE-EQUATION COINTEGRATION MODELS WITH ENDOGENOUS REGRESSORS

HELLE BUNZEL  
*Iowa State University*

This note uses fixed bandwidth (fixed- $b$ ) asymptotic theory to suggest a new approach to testing cointegration parameters in a single-equation cointegration environment. It is shown that the standard tests still have asymptotic distributions that are free of serial correlation nuisance parameters regardless of the bandwidth or kernel used, even if the regressors in the cointegration relationship are endogenous.

### 1. MOTIVATION

In regression models with heteroskedasticity and serial correlation of unknown form, the standard approach to testing hypotheses on the regression parameters involves estimation of the correlation structure using nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimators. These estimators have been thoroughly examined in the literature by, among many others, Andrews (1991), Andrews and Monahan (1992), Hansen (1992), de Jong and Davidson (2000), Newey and West (1987), Robinson (1991), and White (1984). They furnish consistent estimates of the correlation structure in single-equation models of cointegration also, allowing inference on the cointegrating vector to be carried out using conventional tests. Inference conducted in this manner leads to pivotal tests and is robust to heteroskedasticity and serial correlation of unknown form. Even though tests that use HAC estimators are valid *asymptotically*, they typically display substantial size distortions in finite samples (see, e.g., Andrews, 1991; Andrews and Monahan, 1992; den Haan and Levin, 1997).

I thank Tim Vogelsang for his guidance, Pentti Saikkonen for the promptness with which he provided me with an old working paper of his, Wayne Fuller for several discussions, and an anonymous referee for many helpful comments. The usual disclaimer applies. Address correspondence to Helle Bunzel, Department of Economics, 260 Heady Hall, Iowa State University, Ames, IA 50011; e-mail: hbunzel@econ.iastate.edu.

Recent efforts have been made to improve upon the HAC approach in standard (stationary) regression models using inconsistent covariance matrix estimates. The first paper in this literature was Kiefer, Vogelsang, and Bunzel (2000), where a new test based on the Bartlett kernel with bandwidth equal to sample size was developed. Continuing this line of research, Bunzel, Kiefer, and Vogelsang (2001) extended the theory to nonlinear stationary regression models, and Kiefer and Vogelsang (2002) developed the new fixed bandwidth (fixed-*b*) asymptotic theory. In the case of a cointegration relationship with exogenous regressors, applying fixed-*b* asymptotics would have been a straightforward extension of the theory in the standard regression model. However, when the regressors are allowed to be endogenous, the task is nontrivial. Here we provide the asymptotic distribution of Wald and *t*-type statistics in single-equation cointegration models with endogenous regressors.

The principle behind the fixed-*b* theory is to let  $b = M/T$  where *T* is the sample size and *M* is the truncation lag or bandwidth used to compute the HAC estimator. The standard assumptions would require that  $b \rightarrow 0$ , but fixed-*b* asymptotics instead assumes that the truncation lag is a fixed proportion of the sample, i.e., *b* is fixed. This approach has several advantages. First, it improves the asymptotic approximation, resulting in reduced size distortions. Second, it provides an asymptotic distribution that depends on the bandwidth and kernel, thus providing us with better tools for choosing these parameters.

**2. RESULTS**

Consider the following model containing a single cointegrating relationship in addition to some deterministic variables:

$$y_t = f(t)' \alpha + X_t' \beta + u_{1,t}, \quad t = 1, \dots, T, \tag{1}$$

$$X_t = X_{t-1} + u_{2,t},$$

where  $f(t)$  denotes a  $(k_1 \times 1)$  vector of trend functions,  $X_t$  is a  $(k \times 1)$  vector of regressors, and  $\alpha$  and  $\beta$  are  $(k_1 \times 1)$  and  $(k \times 1)$  vectors of parameters, respectively. Let ' denote the transpose, except when it is used in conjunction with the kernel function, where it will denote the derivative. It is assumed that the sequence  $\{u_t\} = \{(u_{1,t}, u'_{2,t})'\}$  does not contain unit roots but may exhibit serial correlation or heteroskedasticity.

At times, it will be useful to stack the first equation in (1) and rewrite it as

$$y = \mathbf{f}(T)\alpha + X\beta + u_1. \tag{2}$$

Here  $\mathbf{f}(T)$  is the  $(T \times k_1)$  stacked vector of trend functions, and  $X$  is the  $(T \times k)$  matrix of regressors. The following notation is required before we state the main assumptions of the note. Denote  $S_{ct} = \sum_{j=1}^t u_{c,j}$ ,  $c = 1, 2$ ,  $S_t = \sum_{j=1}^t u_j$ ,  $\Gamma(j) = E(u_t u'_{t+j})$ , and  $\Gamma_{22}(j) = E(u_{2,t} u'_{2,t+j})$ , let  $w_j(r)$  be a *j*-vector of inde-

pendent Wiener processes, and let  $[rT]$  denote the integer part of  $rT$ , where  $r \in [0, 1]$ . In the discussion that follows  $\Rightarrow$  is used to denote weak convergence.

The first assumption, which is similar to that of Vogelsang (1998), is made to rule out ill-behaved trend functions and to provide some useful notation for deriving and stating the asymptotic distributions.

Assumption 1. There exist a  $(k_1 \times k_1)$  diagonal matrix  $\tau_T$  and a vector of functions  $F$ , such that  $\|\tau_T f(t) - F(t/T)\|^2 = o(1)$  and  $\int_0^1 F(s)F(s)' ds$  exists and is nonsingular. In addition,  $f(t)$  includes a constant term.

Assumption 1 can be relaxed but, as it stands, is sufficiently general to cover most commonly used models. For later use, let  $\mathbf{F}(T)$  be the matrix of the stacked  $F(t/T)$  functions.

The next assumption provides us with the necessary invariance principles and ensures that we can estimate (1) consistently, even when the regressors are endogenous.

Assumption 2.  $\{u_t\}_{t=1}^\infty$  is weakly stationary and satisfies the following conditions:

- (a)  $E(u_t) = 0$  for all  $t$ .
- (b)  $E\|u_t\|^\kappa < \infty$  for all  $t$  and for some  $\kappa$  where  $2 < \kappa < \infty$ .
- (c)  $\{u_t\}_{t=1}^\infty$  is  $\alpha$ -mixing with coefficient  $-\kappa/(\kappa - 2)$ .
- (d)  $\Omega \equiv \lim_{T \rightarrow \infty} E(T^{-1}S_T S_T')$  is positive definite.
- (e)  $g_{uu}(\lambda) \geq \vartheta I_{\kappa+1}$ , where  $\vartheta > 0$ ,  $\lambda \in [0; \pi]$ , and  $g_{uu}(\lambda)$  is the spectral density matrix of  $u$ .
- (f)  $\sum_{j=-\infty}^\infty \|\Gamma(j)\| < \infty$
- (g)  $\sum_{m_1, m_2, m_3=-\infty}^\infty |kum_{ijkl}(m_1, m_2, m_3)| < \infty$ , where  $kum_{ijkl}(m_1, m_2, m_3)$  denotes the fourth-order cumulant of  $u_t$ .

Assumptions 2(a)–(d) have been used extensively in the literature on non-parametric covariance matrix estimation to ensure that the relevant multivariate invariance principles hold. These conditions are sufficient to provide the asymptotic distribution of the ordinary least squares (OLS) estimates of (1) if the regressors are exogenous. Assumptions 2(e)–(g) are made to allow us to deal with endogenous regressors in the manner suggested by Saikkonen (1991), Phillips and Loretan (1991), Stock and Watson (1993), and Wooldridge (1991). A direct implication (see Saikkonen, 1991) is that we can write  $u_{1,t}$  as

$$u_{1,t} = \sum_{j=-\infty}^\infty \gamma_j u_{2,t-j} + v_t, \tag{3}$$

where  $\sum_{j=-\infty}^\infty \|\gamma_j\| < \infty$  and  $v_t$  is a stationary process such that  $E(u_{2,t} v'_{t+l}) = E((X_t - X_{t-1})v'_{t+l}) = 0$ ,  $l = 0, \pm 1, \pm 2, \dots$ . Following standard procedure, we can thus estimate the model using dynamic ordinary least squares (DOLS); i.e., we estimate

$$y_t = f(t)' \alpha + X_t' \beta + \sum_{s=-p}^p \Delta X_{t-s}' \gamma_s + \dot{v}_t, \quad t = p + 1, \dots, T - p, \tag{4}$$

where  $\Delta X_t = X_t - X_{t-1}$  and  $\dot{v}_t = v_t + \sum_{|j|>p} \gamma_j u_{2t-j}$ . We are now ready to make the third and final assumption.

Assumption 3. Let  $p \rightarrow \infty$  such that  $p^3/T \rightarrow 0$  and  $T^{1/2} \sum_{|j|>p} \|\gamma_j\| \rightarrow 0$ .

Saikkonen (1991) shows that if Assumption 3 holds, then (4) is asymptotically equivalent to

$$y_t = f(t)' \alpha + X_t' \beta + v_t,$$

and under Assumptions 1–3, the asymptotic distributions of the least squares estimates of  $\alpha$  and  $\beta$  are well known (for the case where trends are included in the model, see Saikkonen, 1991; Phillips and Loretan, 1991; Stock and Watson, 1993; Wooldridge, 1991; Phillips and Hansen, 1990).

As usual, inference on  $\beta$  is conducted using the DOLS estimates and a HAC estimate of the asymptotic covariance matrix to form Wald or  $t$ -type test statistics. Denote the Cholesky composition of  $\Omega$  by  $\Omega^{1/2} = \begin{bmatrix} \sigma & \sigma_{12} \\ \sigma_{21} & \Lambda \end{bmatrix}$ . Then HAC estimators of  $\sigma$  take the general form

$$\hat{\sigma}^2 = \sum_{j=-(T-p-1)}^{T-p-1} k(j/M) \hat{\Gamma}_j, \quad \text{where } \hat{\Gamma}_j = \begin{cases} \frac{1}{N} \sum_{t=j+1}^{T-p} \hat{v}_t \hat{v}_{t-j}, & \text{for } j \geq 0 \\ \frac{1}{N} \sum_{t=-j+1}^{T-p} \hat{v}_{t+j} \hat{v}_t, & \text{for } j < 0. \end{cases} \tag{5}$$

Here  $N = T - (2p + 1)$ ,  $\hat{v}_t$  are the residuals from (4),  $M$  is called the bandwidth or the truncation lag, and  $k(x)$  is a kernel function satisfying  $k(x) = k(-x)$ ,  $k(0) = 1$ ,  $|k(x)| \leq 1$ ,  $k(x)$  continuous at  $x = 0$ , and  $\int_0^1 k^2(x) dx < \infty$ . Following Kiefer and Vogelsang (2002), we assume that  $M$  is directly proportional to  $T$ , such that  $M = [bT]$  (in place of the usual assumption that  $M/T \rightarrow 0$  as  $T \rightarrow \infty$ ) and develop this asymptotic theory for the cointegration model. The limiting distribution of  $\hat{\sigma}^2$  will depend on the specific bandwidth (now fully determined by the parameter  $b$ ) and kernel used to construct the estimator. To proceed we provide the following definition, which describes two different types of kernels.

*Definition. A kernel is labeled type 1 if  $k(x)$  is twice continuously differentiable everywhere and is labeled type 2 if  $k(x)$  is continuous,  $k(x) = 0$  for  $|x| \geq 1$ , and  $k(x)$  is twice continuously differentiable everywhere except at  $|x| = 1$ .*

In addition, we will consider the Bartlett kernel separately. The following lemma provides the asymptotic distribution of  $\hat{\sigma}^2$  under fixed- $b$  asymptotics and for various choices of kernels. To state the asymptotic distributions, we define

$$V(r) = w_1(r) - \int_0^r F(s)' ds \left( \int_0^1 F^X(s)F^X(s)' ds \right)^{-1} \int_0^1 F^X(s) dw_1(s) - \int_0^r w_k(s)' ds \left( \int_0^1 w_k^F(s)w_k^F(s)' ds \right)^{-1} \int_0^1 w_k^F(s) dw_1,$$

where  $w_1(s)$  is independent of  $w_k(s)$  and where  $w_k^F(s)$  is defined as the residual from the projection of  $w_k(s)$  on the subspace generated by  $F(s)$  in the Hilbert space of square integrable functions on  $[0, 1]$  with the inner product  $(f, g) = \int_0^1 fg$ , such that  $w_k^F(s) = w_k(s) - \int_0^1 w_k(s)F'(s) ds (\int_0^1 F(s)F'(s) ds)^{-1}F(s)$ . Correspondingly,  $F(s)^X$  is the residual from the projection of  $F(s)$  onto the space generated by  $w_k(s)$ .

LEMMA 1. *If  $k$  is type 1,*

$$\hat{\sigma}^2 \Rightarrow -\sigma^2 \int_0^1 \int_0^1 k^{*''}(r-s)V(r)V(s) dr ds.$$

*If  $k$  is type 2*

$$\hat{\sigma}^2 \Rightarrow \sigma^2 \left( \iint_{|r-s|<b} -k^{*''}(r-s)V(r)V(s) dr ds + 2k_{-}'^*(b) \int_0^{1-b} V(r+b)V(r) dr \right),$$

where  $k^*(x) = k(x/b)$  and  $k_{-}'^*(b)$  is the derivative of  $k^*(x)$  from below at  $b$ . If  $k$  is the Bartlett kernel,

$$\hat{\sigma}^2 \Rightarrow \sigma^2 \cdot (2/b) \left[ \int_0^1 V(r)^2 dr - \int_0^{1-b} V(r)V(r+b) dr \right].$$

The proof of Lemma 1 follows that of Kiefer and Vogelsang (2002) but with the added complication that endogenous regressors are present, and therefore some additional work is required to determine the asymptotic distribution of the partial sums of the residuals. The asymptotic distribution of  $\hat{\sigma}^2$  is proportional to  $\sigma^2$  and depends on the bandwidth and kernel as expected.

Using Lemma 1, hypotheses of the form  $H_0: R\beta = \beta_0$  can be tested using the standard Wald test. In what follows  $R$  is a nonstochastic restriction matrix of dimension  $q \times k$  and rank  $q$ . The Wald test for  $H_0$  is defined as

$$W = T(R\hat{\beta} - \beta_0)' [\hat{\sigma}^2 R Q_{fX}^{-1} R']^{-1} (R\hat{\beta} - \beta_0), \quad \text{where}$$

$$Q_{fX} = \left( \frac{1}{T} \begin{bmatrix} \mathbf{f}(T)' \mathbf{f}(T) & \mathbf{f}(T)' X \\ X' \mathbf{f}(T) & X' X \end{bmatrix} \right).$$

The corresponding one-dimensional  $t$ -test can be obtained in the usual manner. Theorem 1, which follows, states the asymptotic distribution of  $W$  under fixed- $b$  asymptotics.

**THEOREM 1.** *Suppose Assumptions 1–3 hold. Then, under  $H_0$ ,*

$$W \Rightarrow \begin{cases} \left( - \int_0^1 \int_0^1 k^{**}(r-s)V^F(r)V^F(s) dr ds \right)^{-1} \widetilde{W}_X & \text{if } k \text{ is type 1} \\ U^F(k,b)^{-1} \widetilde{W}_X & \text{if } k \text{ is type 2} \\ \left( \frac{2}{b} \left[ \int_0^1 V^F(r)^2 dr - \int_0^{1-b} V^F(r)V^F(r+b) dr \right] \right)^{-1} \widetilde{W}_X & \text{if } k \text{ is Bartlett,} \end{cases}$$

where

$$\begin{aligned} \widetilde{W}_X &= \int_0^1 \hat{w}_q^F(s)' d\hat{w}_1^F(s) \left( \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds \right)^{-1} \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s), \\ V^F(r) &= \hat{w}_1^F(r) - \int_0^r \hat{w}_q^F(s)' ds \left( \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds \right)^{-1} \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s), \\ U^F(k,b) &= \iint_{|r-s|<b} -k^{**}(r-s)V^F(r)V^F(s) dr ds \\ &\quad + 2k_{-}'(b) \int_0^{1-b} V^F(r+b)V^F(r) dr, \end{aligned}$$

and  $\hat{w}_q^F(s)$  is the residual from the projection of the first  $q$  coordinates of  $w_k^F(s)$  onto the last  $(k - q)$  coordinates of  $w_k^F(s)$ .

This theorem demonstrates that it is possible to obtain pivotal test statistics with the fixed- $b$  assumption. The asymptotic distribution of the Wald test depends on the kernel and bandwidth, and through  $\hat{w}_q^F(s)$  it also depends on the number of restrictions being tested, the number of regressors in the model, and the trends included, where standard  $b \rightarrow 0$  asymptotics would have resulted in an asymptotic  $\chi^2$  distribution. Note that although the asymptotic distribution is nonstandard, it is simple to obtain critical values through simulations because the distribution is simply a function of independent Wiener processes. (Simulations guiding the choice of  $b$  and also some critical values are available from the author upon request.) With the critical values in hand, it is possible to use the fixed- $b$  approach for testing in single-equation models of cointegration. Although the choice of  $b$  and the simulations demonstrating the efficacy of the test are treated in a companion paper (Bunzel, 2006), the fixed- $b$  approach is likely to improve the asymptotic approximation, resulting in reduced size dis-

tortions. It also provides an asymptotic distribution that depends on the bandwidth and kernel, thus providing us with better tools for choosing these parameters.

## REFERENCES

- Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.
- Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Bunzel, H. (2006) Practical Use of Fixed- $b$  Asymptotics in Cointegration Models. Mimeo, Iowa State University.
- Bunzel, H., N.M. Kiefer, & T.J. Vogelsang (2001) Simple robust testing of hypotheses in non-linear models. *Journal of the American Statistical Association* 96, 1088–1098.
- de Jong, R.M. & J. Davidson (2000) Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–424.
- den Haan, W.J. & A. Levin (1997) A practitioner's guide to robust covariance matrix estimation. In G. Maddala & C. Rao (eds.), *Handbook of Statistics: Robust Inference*, vol. 15, pp. 291–341. New York: Elsevier.
- Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.
- Kiefer, N.M. & T.J. Vogelsang (2002) Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.
- Kiefer, N.M., T.J. Vogelsang, & H. Bunzel (2000) Simple robust testing of regression hypotheses. *Econometrica* 68, 695–714.
- Newey, W.K. & K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
- Phillips, P.C.B. & M. Loretan (1991) Estimating long run economic equilibria. *Review of Economic Studies* 58, 407–436.
- Robinson, P. (1991) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59, 1329–1363.
- Saikkonen, P. (1991) Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.
- Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.
- Vogelsang, T.J. (1998) Trend function hypothesis testing in the presence of serial correlation parameters. *Econometrica* 65, 123–148.
- White, H. (1984) *Asymptotic Theory for Econometricians*. Academic Press.
- Wooldridge, J.M. (1991) Notes on Regression with Difference-Stationary Data. Mimeo, Michigan State University.

## APPENDIX

**A.1. Proof of Lemma 1.** Following Kiefer and Vogelsang (2002), we define

$$\Delta^2 \kappa_{ij} = \left\{ k \left( \frac{i-j}{[bN]} \right) - k \left( \frac{i-j-1}{[bN]} \right) \right\} - \left\{ k \left( \frac{i-j+1}{[bN]} \right) - k \left( \frac{i-j}{[bN]} \right) \right\}$$

and use this expression to rewrite  $\hat{\sigma}^2$  as

$$\hat{\sigma}^2 = N^{-1} \sum_{t=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2 \Delta^2 \kappa_{it} (N^{-1/2} \hat{\xi}_i) (N^{-1/2} \hat{\xi}_t), \tag{A.1}$$

where  $\hat{S}_{[rN]} = \sum_{t=1}^{[rN]} \hat{v}_{t+p}$  and  $\{\hat{v}_t\}_{t=p+1}^{T-p}$  are the residuals from (4). Note that for (A.1) to be valid it must be the case that the residuals sum to zero. Therefore, for the results that follow to be valid  $f(t)$  must include a constant term as assumed in Assumption 1. To establish the asymptotic distribution of  $\hat{\sigma}^2$ , it is necessary first to determine the asymptotic distribution of  $\hat{S}_{[rN]}$ .

LEMMA 2. Under Assumptions 1–3,  $N^{-1/2} \hat{S}_{[rN]} \Rightarrow \sigma V(r)$ .

**Proof.** First define  $\gamma = [\gamma'_p, \dots, \gamma'_p]'$ , which is a  $(k(2p + 1) \times 1)$  vector of parameters and let  $\Delta Z_{t+p} = [\Delta X'_{t-p}, \Delta X'_{t-p+1}, \dots, \Delta X'_{t+p}]'$  be the corresponding vector of regressors. Simple matrix manipulations yield

$$\begin{aligned} N^{-1/2} \hat{S}_{[rN]} &= N^{-1/2} \sum_{t=1}^{[rN]} (\hat{v}_{t+p} - f(t+p))' (\hat{\alpha} - \alpha) - X'_{t+p} (\hat{\beta} - \beta) - \Delta Z'_{t+p} (\hat{\gamma} - \gamma) \\ &= N^{-1/2} \sum_{t=1}^{[rN]} \hat{v}_{t+p} - \begin{bmatrix} N^{-1} \sum_{t=1}^{[rN]} f(t+p) \\ N^{-3/2} \sum_{t=1}^{[rN]} X_{t+p} \end{bmatrix}' \begin{bmatrix} N^{1/2} (\hat{\alpha} - \alpha) \\ N (\hat{\beta} - \beta) \end{bmatrix} \\ &\quad - N^{-1/2} \sum_{t=1}^{[rN]} \Delta Z'_{t+p} (\hat{\gamma} - \gamma). \end{aligned} \tag{A.2}$$

In what follows, we will show that the last term in the expression for  $\hat{S}_{[rN]}$ , (A.2), is  $O_p(p/\sqrt{T})$  and therefore does not affect the asymptotic distribution of  $\hat{S}_{[rN]}$ . First note that

$$\begin{aligned} E \left\| N^{-1/2} \sum_{t=1}^{[rN]} \Delta Z'_{t+p} \right\|^2 &= N^{-1} \sum_{t=1}^{[rN]} \sum_{s=1}^{[rN]} E(\Delta Z'_{t+p} \Delta Z_{s+p}) \\ &= (2p + 1) N^{-1} \sum_{t=1}^{[rN]} \sum_{s=1}^{[rN]} \text{tr} \Gamma_{22}(s-t) \\ &\leq (2p + 1) N^{-1} \sum_{s=-N}^N \text{tr} \Gamma_{22}(s) = O_p(p) \end{aligned}$$

by Assumption 2(f). Thus  $\|N^{-1/2} \sum_{t=1}^{[rN]} \Delta Z'_{t+p}\| = O_p(\sqrt{p})$ . We then obtain

$$\left\| N^{-1/2} \sum_{t=1}^{[rN]} \Delta Z'_{t+p} (\hat{\gamma} - \gamma) \right\| \leq \left\| N^{-1/2} \sum_{t=1}^{[rN]} \Delta Z'_{t+p} \right\| \|(\hat{\gamma} - \gamma)\| = O_p\left(\frac{p}{\sqrt{T}}\right)$$

by equation (23) in Saikkonen (1991), which states that  $\|\hat{\gamma} - \gamma\|^2 = \sum_{s=-p}^p \|\hat{\gamma}_s - \gamma_s\|^2 = O_p(p/T)$ .



We can now determine the asymptotic distribution of  $T^{-1/2}\hat{S}_{[rN]}$  from the first two terms of (A.2). By Assumptions 1–3 we know from Saikkonen (1991) and Phillips and Hansen (1990) that

$$\begin{bmatrix} T^{1/2}\tau_T^{-1}(\hat{\alpha} - \alpha) \\ T(\hat{\beta} - \beta) \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma \left( \int_0^1 F^X(s)F^X(s)' ds \right)^{-1} \int_0^1 F^X(s) dw_1(s) \\ \sigma(\Lambda)^{-1} \left( \int_0^1 w_k^F(s)w_k^F(s)' ds \right)^{-1} \int_0^1 w_k^F(s) dw_1(s) \end{bmatrix}$$

and

$$\begin{bmatrix} T^{-1} \sum_{t=1}^{[rN]} \tau_T f(t+p) \\ T^{-3/2} \sum_{t=1}^{[rN]} X_{t+p} \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^r F^X(s) ds \\ \Lambda \int_0^r w_k^F(s) ds \end{bmatrix}.$$

It also follows directly from Assumption 2 that  $N^{-1/2} \sum_{t=p+1}^{[r(T-p)]} \dot{v}_t \Rightarrow \sigma w_1(r)$ . Because  $N/T \rightarrow 1$ , it will also be the case that  $T^{-1/2} \sum_{t=p+1}^{[r(T-p)]} \dot{v}_t \Rightarrow \sigma w_1(r)$ . So it is now established that

$$\begin{aligned} T^{-1/2}\hat{S}_{[rN]} &\Rightarrow \sigma w_1(r) - \sigma \int_0^r F(s)' ds \left( \int_0^1 F^X(s)F^X(s)' ds \right)^{-1} \int_0^1 F^X(s) dw_1(s) \\ &\quad - \sigma \int_0^r w_k(s) ds \left( \int_0^1 w_k^F(s)w_k^F(s)' ds \right)^{-1} \int_0^1 w_k^F(s) dw_1(s) \\ &= \sigma V(r). \end{aligned}$$

The rest of the proof is split into three cases, corresponding to type 1, type 2, and the Bartlett kernels. It follows directly from Kiefer and Vogelsang (2002) and Lemma 2.

**Case 1.**  $k(x)$  is a type 1 kernel. By definition of the second derivative,  $T^2\Delta^2\kappa_{il} - (-k^{**}((i-l)/N)) \rightarrow 0$ , and using Lemma 2 it follows easily that

$$\begin{aligned} \hat{\sigma}^2 &= N^{-1} \sum_{i=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2\Delta^2\kappa_{ij} N^{-1/2}\hat{S}_i N^{-1/2}\hat{S}_j \\ &\Rightarrow \sigma^2 \int_0^1 \int_0^1 -k^{**}(r-s)V(s)V(r) dr ds. \end{aligned}$$

**Case 2.**  $k(x)$  is a type 2 kernel. Following Kiefer and Vogelsang (2002), we use simple algebra and the definition of  $\Delta^2\kappa_{ij}$  to establish that when  $|i-j| > [bN]$ ,  $\Delta^2\kappa_{ij} = 0$ , and when  $|i-j| = [bN]$ ,  $\Delta^2\kappa_{ij} = -k(([bN]-1)/[bN])$ . Also recall that when  $|i-j| < [bN]$   $k(x)$  is twice continuously differentiable. We split up the expression of  $\hat{\sigma}^2$  as follows:

$$\begin{aligned} \hat{\sigma}^2 &= N^{-1} \sum_{i=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} 1_{\{|i-j| < [bN]\}} N^2 \Delta^2 \kappa_{ij} N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_j \\ &\quad + 2N^{-2} \sum_{i=1}^{N-1-[bN]} N^2 k\left(\frac{[bN]-1}{[bN]}\right) N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_{i+[bN]} \\ &= N^{-1} \sum_{i=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} 1_{\{|i-j| < [bN]\}} N^2 \Delta^2 \kappa_{ij} N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_j \\ &\quad - 2k\left(1 - \frac{1}{[bN]}\right) \sum_{i=1}^{N-1-[bN]} N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_{i+[bN]} \\ &\Rightarrow \sigma^2 \left( \iint_{|r-s| < b} -k^{**}(r-s)V(r)V(s) dr ds + 2k_{*}'(b) \int_0^{1-b} V(r+b)V(r) dr \right), \end{aligned}$$

where the asymptotic distribution follows directly from Lemma 2 and Kiefer and Vogelsang (2002).

**Case 3.**  $k(x)$  is the Bartlett kernel. Here again following Kiefer and Vogelsang (2002), it can be verified that when  $|i - j| = 0$ ,  $\Delta^2 \kappa_{ij} = 2/[bN]$ , and when  $|i - j| = [bN]$ ,  $\Delta^2 \kappa_{ij} = -(1/[bN])$ . Using these expressions and Lemma 2 in (A.1), we obtain the following limiting distribution:

$$\begin{aligned} \hat{\sigma}^2 &= N^{-1} \sum_{i=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2 \Delta^2 \kappa_{ij} N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_j \\ &= \frac{2}{[bN]} \sum_{i=1}^{N-1} (N^{-1/2} \hat{S}_i)^2 - \frac{2}{[bN]} \sum_{i=1}^{N-1-[bN]} N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_{i+[bN]} \\ &\Rightarrow \sigma^2 \left( \frac{2}{b} \int_0^1 V(r)^2 dr - \frac{2}{b} \int_0^{1-b} V(r+b)V(r) dr \right). \quad \blacksquare \end{aligned}$$

**A.2. Proof of Theorem 1.** The initial step of the proof will be to rewrite the model, projecting out all regressors that are not related to the hypothesis in question. Then we will prove that the statistic is numerically unchanged if it is calculated from the rewritten model. Finally the expression of  $W$  obtained from the rewritten model will be used to derive the asymptotic distribution of the statistic.

To rewrite the model, let  $L = \begin{bmatrix} R \\ D \end{bmatrix}$ , where  $D$  is chosen such that  $L$  has full rank  $(k)$ , and define  $[\tilde{X}_1 \tilde{X}_2] = XL^{-1}$  and  $\tilde{Z} = Z \cdot (L^{-1} \otimes I_{2p+1})$ . Using these definitions, model (4) can be rewritten in the following manner:

$$\begin{aligned} y &= f(T)\alpha + (XL^{-1})(L\beta) + (\Delta Z \cdot (L^{-1} \otimes I_{2p+1}))(L \otimes I_{2p+1})\gamma + \dot{v} \\ &= f(T)\alpha + [\tilde{X}_1 \tilde{X}_2] \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + \Delta \tilde{Z} \tilde{\gamma} + \dot{v} \\ &= f(T)\alpha + \tilde{X}_1 \beta_1^* + \tilde{X}_2 \beta_2^* + \Delta \tilde{Z} \tilde{\gamma} + \dot{v}. \end{aligned}$$

Because  $\tilde{X}_1$  and  $\tilde{X}_2$  are linear combinations of  $X$ , they too contain unit root processes if the original assumption of just one cointegration relationship is maintained. Furthermore  $\Delta\tilde{Z}$  contains the leads and lags of the differenced  $\tilde{X}$  variables. We will now show that testing  $H_0: R\beta = \beta_0$  is equivalent to testing the hypothesis  $\tilde{H}_0: \beta_1^* = \beta_0$  in the model

$$y^* = X_1^* \beta_1^* + v^*, \tag{A.3}$$

where for any matrix  $G$ ,  $M_G = I - G(G'G)^{-1}G$ ,  $\tilde{X}_2^F = M_f M_{\Delta\tilde{Z}} \tilde{X}_2$ ,  $X_1^* = M_{\tilde{X}_2^F} M_f M_{\Delta\tilde{Z}} \tilde{X}_1$ ,  $v^* = M_{\tilde{X}_2^F} M_f M_{\Delta\tilde{Z}} v$ , and  $y^* = M_{\tilde{X}_2^F} M_f M_{\Delta\tilde{Z}} y$ .

LEMMA 3. *The statistic for testing  $\tilde{H}_0: \beta_1^* = \beta_0$  from (A.3) is numerically identical to the statistic for testing  $H_0: R\beta = \beta_0$  in the model given by (4).*

The proof follows from extensive but simple matrix algebra. To complete the proof of the theorem we thus need to determine the asymptotic distribution of

$$\begin{aligned} W^* &= T(\beta_1^* - \beta_0)' [\hat{\sigma}^{2*} (T^{-1}(X_1^*)' X_1^*)^{-1}]^{-1} (\beta_1^* - \beta_0) \\ &= T(\beta_1^* - \beta_0)' [\hat{\sigma}^{2*} (T^{-2}(X_1^*)' X_1^*)^{-1}]^{-1} T(\beta_1^* - \beta_0). \end{aligned}$$

Because  $\hat{\sigma}^{2*} = N^{-1} \sum_{i=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2 \Delta^2 \kappa_{il} (N^{-1/2} \hat{S}_i^*) (N^{-1/2} \hat{S}_i^*)'$ , where  $\hat{S}_i^*$  is defined as  $\hat{S}_i^*$ , but for the model in (A.3), we know

$$\begin{aligned} N^{-1/2} \hat{S}_{[rN]}^* &= N^{-1/2} \sum_{t=1}^{[rN]} v_{t+p}^* - N^{-3/2} \sum_{t=1}^{[rN]} X_{1,t+p}^* (N(\hat{\beta}_1^* - \beta_1^*)) \\ &\Rightarrow \sigma^* \hat{w}_1^F(r) - \sigma^* \int_0^r \hat{w}_q^F(s)' ds \left( \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds \right)^{-1} \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s) \\ &= \sigma^* V^F(r), \end{aligned}$$

and therefore,

$$\hat{\sigma}^{*2} \Rightarrow -\sigma^{*2} \int_0^1 \int_0^1 k^{**}(r-s) V^F(r) V^F(s) dr ds \quad \text{if } k \text{ is type 1,}$$

$$\begin{aligned} \hat{\sigma}^{*2} &\Rightarrow \sigma^{*2} \left( \iint_{|r-s|<b} -k^{**}(r-s) V^F(r) V^F(s) dr ds \right. \\ &\quad \left. + 2k_{-}'(b) \int_0^{1-b} V^F(r+b) V^F(r) dr \right) \quad \text{if } k \text{ is type 2,} \end{aligned}$$

$$\hat{\sigma}^{*2} \Rightarrow \sigma^{*2} \frac{2}{b} \left[ \int_0^1 V^F(r)^2 dr - \int_0^{1-b} V^F(r) V^F(r+b) dr \right] \quad \text{if } k \text{ is Bartlett.}$$

By the definition of  $X_1^*$ ,

$$(T^{-2}(X_1^*)'X_1^*)^{-1} \Rightarrow \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1}.$$

The distribution of  $W^*$  can now be obtained.

If  $k$  is type 1,

$$\begin{aligned} W^* &= T(\beta_1^* - \beta_0)' [\hat{\sigma}^{2*} (T^{-2}(X_1^*)'X_1^*)^{-1}]^{-1} T(\beta_1^* - \beta_0) \\ &\Rightarrow \left( - \int_0^1 \int_0^1 k^{*''}(r-s) V^F(s) V^F(r) dr ds \right)^{-1} \\ &\quad \times \int_0^1 \hat{w}_q^F(s)' d\hat{w}_1^F(s) \left( \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds \right)^{-1} \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s). \end{aligned}$$

If  $k$  is type 2,

$$\begin{aligned} W^* &= T(\beta_1^* - \beta_0)' [\hat{\sigma}^{2*} (T^{-2}(X_1^*)'X_1^*)^{-1}]^{-1} T(\beta_1^* - \beta_0) \\ &\Rightarrow \sigma^* \int_0^1 \hat{w}_q^F(s)' d\hat{w}_1^F(s) (\Lambda^*)' \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1} \\ &\quad \times \left\{ \sigma^{*2} \left( \iint_{|r-s|<b} -k^{*''}(r-s) V^F(r) V^F(s) dr ds \right. \right. \\ &\quad \left. \left. + 2k_{-}'^*(b) \int_0^{1-b} V^F(r+b) V^F(r) dr \right) \right. \\ &\quad \left. \times \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1} \right\}^{-1} \\ &\quad \times \sigma^* \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1} \Lambda^* \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s) \\ &= \left( \iint_{|r-s|<b} -k^{*''}(r-s) V^F(r) V^F(s) dr ds \right. \\ &\quad \left. + 2k_{-}'^*(b) \int_0^{1-b} V^F(r+b) V^F(r) dr \right)^{-1} \\ &\quad \times \int_0^1 \hat{w}_q^F(s)' d\hat{w}_1^F(s) \left( \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds \right)^{-1} \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s). \end{aligned}$$

If  $k$  is Bartlett,

$$\begin{aligned}
 W^* &= T(\beta_1^* - \beta_0)' [\hat{\sigma}^{2*} (T^{-2} X_1^{*'} X_1^*)^{-1}]^{-1} T(\beta_1^* - \beta_0) \\
 &\Rightarrow \sigma^* \int_0^1 \hat{w}_q^F(s)' d\hat{w}_1^F(s) (\Lambda^*)' \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1} \\
 &\quad \times \left\{ \frac{2}{b} (\sigma^*)^2 \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1} \right. \\
 &\quad \left. \times \left[ \int_0^1 V^F(r)^2 dr - \int_0^{1-b} V^F(r) V^F(r+b) dr \right] \right\}^{-1} \\
 &\quad \times \sigma^* \left( \Lambda^* \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds (\Lambda^*)' \right)^{-1} \Lambda^* \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s) \\
 &= \frac{b}{2} \left( \int_0^1 V^F(r)^2 dr - \int_0^{1-b} V^F(r) V^F(r+b) dr \right)^{-1} \\
 &\quad \times \int_0^1 \hat{w}_q^F(s)' d\hat{w}_1^F(s) \left( \int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)' ds \right)^{-1} \int_0^1 \hat{w}_q^F(s) d\hat{w}_1^F(s). \quad \blacksquare
 \end{aligned}$$