Econometric Theory, **24**, 2008, 1425–1441. Printed in the United States of America.
doi:10.1017/S0266466608080560

DATA DEPENDENT RULES FOR SELECTION OF THE NUMBER OF LEADS AND LAGS IN THE DYNAMIC OLS COINTEGRATING REGRESSION

Mohitosh Kejriwal
Purdue University

Pierre Perron
Boston University

Saikkonen (1991, *Econometric Theory* 7, 1–21) developed an asymptotic optimality theory for the estimation of cointegrated regressions. He proposed the dynamic ordinary least squares (OLS) estimator obtained by augmenting the static cointegrating regression with leads and lags of the first differences of the I(1) regressors. However, the assumptions imposed preclude the use of information criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to select the number of leads and lags. We show that his results remain valid under weaker conditions that permit the use of such data dependent rules. Simulations show that, relative to sequential general to specific testing procedures, the use of such information criteria can indeed produce estimates with smaller mean squared errors and confidence intervals with better coverage rates.

1. INTRODUCTION

The estimation of cointegrated systems has received a great deal of attention in the econometrics literature. Several methods to estimate the cointegrating vectors have been proposed, which are asymptotically equivalent and yield estimates with some optimality properties as defined, e.g., by Saikkonen (1991). A popular method is based on estimating the full system by maximum likelihood assuming normal errors (Johansen, 1991). On the other hand, several single equation methods have been developed that yield estimates that are also optimal and lead to Wald tests with the usual chi-square limiting distribution. Phillips and Hansen (1990) developed a fully modified estimator based on a semiparametric two-step procedure. An alternative very simple method is the

We are grateful to Pentti Saikkonen (the co-editor) and two referees for useful comments. Address correspondence to Mohitosh Kejriwal, Krannert School of Management, Purdue University, 403 West State Street, West Lafayette, IN 47907, USA; e-mail: mkejriwa@purdue.edu.

 0266-4666/08 $15.00

so-called leads and lags, or dynamic ordinary least squares (OLS) estimator, proposed by Saikkonen (1991) and Stock and Watson (1993). This estimator is obtained by augmenting the static cointegrating regression with leads and lags of the first differences of the I(1) regressors. The idea is to remove the asymptotic inefficiency of the least squares estimate in the static regression by using the relevant information in the system to account for the correlation between the regressors and the dependent variable. Monte Carlo simulation studies (e.g., Carrion-i-Silvestre and Sansó-i-Rosselló, 2004) show that the dynamic OLS estimator performs better in finite samples compared to the fully modified estimation procedure.[1]

An issue that arises when using the dynamic OLS regression is how to choose the number of leads and lags. A possibility is to use an a priori fixed rule. Ng and Perron (1995) show, however, that the Dickey and Fuller (1979) and Said and Dickey (1984) augmented Dickey–Fuller (ADF) unit root test constructed with such an a priori rule for the truncation lag exhibits size distortions and/or power loss unless the value happens to be chosen appropriately. Similar conclusions are obtained for the ADF-based cointegration test in Haug (1996). These results point to the need for data dependent rules to select the number of leads and lags. A popular method is to use an information criterion such as the Akaike information criterion (AIC) (Akaike, 1973) or the Bayesian information criterion (BIC) of Schwarz (1978). This was also suggested by Saikkonen (1992, p. 10) for the dynamic OLS regression. The problem is that, for common processes, such as a finite-order autoregressive moving average (ARMA) process, the conditions imposed by Saikkonen (1991) preclude the use of such information criteria. This is because, in such cases, the information criteria select a lag order that increases with the sample size at a logarithmic rate. However, to derive the limit distribution of the estimates and test statistics, a lower bound that precludes this logarithmic rate is imposed.

We build on the work of Ng and Perron (1995), who analyzed the choice of the truncation lag in the context of the ADF unit root test in a general ARMA model. They show that the lower bound condition imposed by Said and Dickey (1984), a rate of increase at some polynomial rate, is not needed to obtain the usual limit distribution applicable in the fixed lag case (for a relaxation of the upper bound condition, see Chang and Park, 2002). A related paper is that by Lütkepohl and Saikkonen (1999), who study data dependent rules for choosing the truncation lag in the context of testing for cointegration.

We show that the lower bound condition in Saikkonen (1991) is not necessary to obtain asymptotically efficient estimates and perform hypothesis testing using standard Wald tests with a chi-square null limit distribution. In particular, we impose a weaker condition that does not preclude a logarithmic rate of increase (in the case of an ARMA process or a linear process with geometrically decaying weights) so that data dependent rules such as the AIC and the BIC can be used. This is an important practical result as empirical researchers often use such data dependent rules to select the order of the model. We further show that the upper bound condition can also be relaxed, with the implication that our results are valid even if the estimates of the nuisance parameters (the coefficients on the leads and lags of the first-differenced regressors) are inconsistent. Section 2 presents the model and assumptions, and Section 3 states the main results. Section 4 presents simulations showing that, relative to sequential general to specific testing procedures, the use of such information criteria can indeed produce estimates with smaller mean-squared errors (MSEs) and confidence intervals with better coverage rates. Section 5 offers some concluding remarks, and all technical derivations are included in an Appendix.

2. MODEL AND ASSUMPTIONS

We consider a scalar random variable y_t $(t = 1, \ldots, T)$ generated by the cointegrating relation

$$\begin{aligned} y_t &= z_t'\delta + u_{1t}, \\ \Delta z_t &= u_{2t} \end{aligned} \tag{1}$$

with z_t a $(q \times 1)$ vector of variables and $u_t = [u_{1t}', u_{2t}']'$.[2] We impose the same conditions as in Saikkonen (1991), which are stated in the following assumption.

Assumption DGP.

(a) The errors u_t are a stationary process with zero mean and continuous spectral density matrix $f_{uu}(\lambda)$.
(b) The spectral density matrix $f_{uu}(\lambda)$ is bounded away from zero so that $f_{uu}(\lambda) \geq \alpha I_n$, $n = q + 1, \lambda \in [0, \pi]$, and $\alpha > 0$.
(c) The covariance function of u_t is absolutely summable; i.e., $\sum_{k=-\infty}^{\infty} \|\Gamma(k)\| < \infty$ where $E(u_t u_{t+k}') = \Gamma(k)$ and $\|\cdot\|$ is the standard euclidian norm.
(d) The fourth-order cumulants of u_t, $\text{cum}(m_1, m_2, m_3)$, satisfy

$$\sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \sum_{m_3=-\infty}^{\infty} |\text{cum}(m_1, m_2, m_3)| < \infty.$$

First note that part (a) implies that u_t satisfy a multivariate invariance principle (see Hall and Heyde, 1980, p. 146) such that $T^{-1/2} \sum_{t=1}^{[\text{Tr}]} u_t \Rightarrow B(r)$, $0 < r \leq 1$, where $\Rightarrow$ denotes weak convergence in distribution under the Skorohod topology and $B(r)$ is a vector Brownian motion with a positive definite covariance matrix $\Omega = 2\pi f_{uu}(0)$, which we partition as

$$\Omega = \begin{bmatrix} \sigma_{11}^2 & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \qquad f_{uu}(\lambda) = \begin{bmatrix} f_{u_1 u_1}(\lambda) & f_{u_1 u_2}(\lambda) \\ f_{u_2 u_1}(\lambda) & f_{u_2 u_2}(\lambda) \end{bmatrix}.$$

When parts (b) and (c) of Assumption DGP hold,

$$u_{1t} = \sum_{j=-\infty}^{\infty} u'_{2,t-j}\Pi_j + v_t,$$

where $\sum_{j=-\infty}^{\infty}\|\Pi_j\| < \infty$ and v_t is a stationary process such that $E(u_{2t}v_{t+k}) = 0$ for all k and $f_{vv}(\lambda) = f_{u_1u_1}(\lambda) - f_{u_1u_2}(\lambda)f_{u_2u_2}(\lambda)^{-1}f_{u_2u_1}(\lambda)$. Also, $T^{-1}\sum_{t=1}^{T} v_t^2 \xrightarrow{p} \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(v_t^2) \equiv \sigma_v^2$, say. We thus have $2\pi f_{vv}(0) = \sigma_{1.2}^2 = \sigma_{11}^2 - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ (see, e.g., Brillinger, 1975, p. 296). Hence, the cointegrating relation specified by (1) can be represented by[3]

$$y_t = z_t'\delta + \sum_{j=-k_T}^{k_T} \Delta z'_{t-j}\Pi_j + v_t^*, \qquad t = k_T + 1, \ldots, T - k_T, \tag{2}$$

where $v_t^* = v_t + e_t$ with $e_t = \sum_{|j|>k_T} \Delta z'_{t-j}\Pi_j$. We now state the conditions on the rate of increase of k_T as the sample size increases.

Assumption K. As $T \to \infty$, $k_T \to \infty$ in such a way that

(a) (Upper bound condition) $k_T^2/T \to 0$ and
(b) (Lower bound condition) $k_T \sum_{|j|>k_T}\|\Pi_j\| \to 0$.

Saikkonen (1991) assumed $k_T^3/T \to 0$, which is stronger than our upper bound. More importantly, he specified $T^{1/2}\sum_{|j|>k_T}\|\Pi_j\| \to 0$, which is much more restrictive than our lower bound. For instance, when the data are generated by a finite-order ARMA process, it precludes using popular information criteria such as the AIC or BIC because these data dependent methods to select k_T imply a logarithmic rate of increase that his lower bound condition does not permit. In contrast, Assumption K merely requires that $k_T \to \infty$ at any rate in the case of ARMA processes, thereby allowing the use of information criteria.

In this framework, the estimate of the cointegrating vector is simply the OLS estimate of δ from regression (2), and inference can be carried using the standard normal or chi-square distribution provided the standard errors are adjusted for potential serial correlation in v_t^*. We show in the next section that all results derived by Saikkonen (1991) remain valid under our less restrictive assumptions on k_T.

3. MAIN RESULTS

As a matter of notation, we let $B_T^* = T^{-1/2}\sum_{j=1}^{[Tr]} w_j$, where $w_t = (v_t, u'_{2t})'$. Note that B_T^* converges weakly to B^* as $T \to \infty$, where B^* is a vector Brownian motion with block diagonal covariance matrix $\Omega^* = \text{diag}(\sigma_{1.2}^2, \Omega_{22})$. We partition B^* as $B^* \equiv (B_{1.2}, B_2')'$ conformably with $(v_t, u'_{2t})'$, where $B_{1.2} = B_1 - \Omega_{12}\Omega_{22}^{-1}B_2$. Note that $B_{1.2}$ is independent of B_2.

We first state a proposition about the limiting distribution of the OLS estimate of δ, denoted $\hat{\delta}$, obtained from regression (2), and also the limit of the sample variance of the errors $\hat{\sigma}_v^{*2} = T^{-1}\sum_{t=1}^{T}\hat{v}_t^{*2}$, where $\hat{v}_t^*$ are the OLS residuals from (2).

PROPOSITION 1. *Under Assumptions DGP and K,*

(i) $T(\hat{\delta} - \delta) \Rightarrow (\int_0^1 B_2 B_2')^{-1}(\int_0^1 B_2\, dB_{1.2})$ *and*
(ii) $\hat{\sigma}_v^{*2} \xrightarrow{p} \sigma_v^2$.

Hence, $\hat{\delta}$ has the usual mixed normal distribution, and, because B_2 and $B_{1.2}$ are independent, inference can be carried out using the standard normal or chi-square distributions when the errors v_t^* are martingale differences. In the general case where the errors v_t^* are serially correlated, we need to replace $\hat{\sigma}_v^{*2}$ by an estimate of (2π times) the spectral density function at frequency zero of v_t, namely, $\sigma_{1.2}^2$ (the so-called long-run variance). We consider a class of estimates based on a weighted sum of the sample autocovariances of $\hat{v}_t^*$ defined by

$$\hat{\sigma}_{1.2}^2 = \hat{\sigma}_v^{*2} + 2\sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T}\hat{v}_t^*\hat{v}_{t-j}^*. \quad \textbf{(3)}$$

We impose the following conditions on the kernel function $w(\cdot)$ and the bandwidth m_T, which are satisfied for well-known kernels such as the Bartlett and the quadratic spectral, and data dependent methods to select the bandwidth (e.g., Andrews, 1991; Newey and West, 1994).

Assumption W. The kernel function $w(\cdot)$ is a continuous and even function with $|w(\cdot)| \leq 1$, $w(0) = 1$, and $\int_{-\infty}^{\infty} w^2(x)\, dx < \infty$; and the bandwidth is such that $m_T \to \infty$ and $m_T = o(T^{1/2})$ as $T \to \infty$.

We then have the following result.

PROPOSITION 2. *Under Assumptions DGP, K, and W,* $\hat{\sigma}_{1.2}^2 \xrightarrow{p} \sigma_{1.2}^2$.

Now that we have shown the consistency of the long-run variance estimator, Wald tests of hypotheses on the cointegrating vectors can be constructed straightforwardly and have the usual chi-square asymptotic distribution (see, e.g., Phillips and Hansen, 1990).

We have thus shown that standard asymptotic inference on the I(1) regression coefficients is still valid even if we relax both the lower and upper bound conditions.[4] It is important to stress that our condition $k_T = o(T^{1/2})$ is not sufficient for the consistency of the estimates of the nuisance parameters Π_j. To guarantee the consistency of such estimates, we would need to strengthen Assumption K to $k_T = o(T^{1/3})$ as in Berk (1974). Also, to obtain estimates that are $\sqrt{T}$ consistent (more precisely that $\sum_{j=-k_T}^{j=k_T}\|\hat{\Pi}_j - \Pi_j\|^2 = O_p(k_T/T)$) the lower bound used by Saikkonen (1991) is needed. But as shown previously these are not needed insofar as inference on the cointegrating vector is concerned.

Remark 1. For models with a constant and/or time trend included in the regression, we simply replace the Brownian motion by its demeaned or detrended counterpart. With only a constant we have $\bar{B}_{2t} = B_{2t} - \int_0^1 B_{2s}\,ds$, whereas including a constant and a time trend we have $\tilde{B}_{2t} = B_{2t} + (6t-4)\int_0^1 B_{2s}\,ds - (12t-6)\int_0^1 sW_s\,ds$.

4. SIMULATIONS

In this section, we assess the adequacy of different information criteria to select the number of leads and lags in finite samples via a small-scale simulation experiment. Our goal is not to provide a comprehensive treatment or to come up with definite conclusions about the best method to select the number of leads and lags. Our aim is more modest in that we wish to show that, using a class of data generating processes (DGPs) that have been used in other studies, the use of selected information criteria can indeed lead to estimates with smaller MSEs and confidence intervals with better coverage rates relative to sequential general to specific testing procedures.

We consider three information criteria: the AIC, the BIC, and the PIC (posterior information criterion) suggested by Phillips and Ploberger (1994). We compare their performance to a general to specific sequential testing procedure as described in Ng and Perron (1995) for the selection of the autoregressive order in the context of testing for a unit root. The latter starts by imposing some maximal value $k_{\max}$ to the number of leads and lags of the first differences of the regressors in regression (2). Usually, $k_{\max}$ is some function of the sample size that increases at some polynomial rate, e.g., $T^{1/4}$. Then for each of the leads and lags, it assesses whether the last lag is significant using a two-sided test on the last coefficient. We start with the lags. If a rejection occurs, the number of lags selected is $k_{\max}$, and we assess the number of leads sequentially. Otherwise, the model is reestimated with $k_{\max}$ lags and the test is applied to the last coefficient on the leads. Upon a nonrejection, we then return to assess if the $(k_{\max}-1)$th coefficient on the leads is significant in a regression estimated with $k_{\max}-1$ leads and lags. And so on, until a rejection occurs for the leads and lags. We consider two variations of this procedure: one based on 10% two-sided tests, labeled *tsig*10, and one based on 5% two-sided tests, labeled *tsig*05.

The simulation design closely follows Haug (1996). The DGP considered is

$$y_t = x_{1t} + v_t,$$

$$a_1 y_t - a_2 x_{1t} = \psi_t,$$

where $v_t = \rho v_{t-1} + w_t$, $\psi_t = \psi_{t-1} + \rho_t$, and $\rho_t = \varphi_t + \theta\varphi_{t-1}$, with

$$\begin{bmatrix} w_t \\ \varphi_t \end{bmatrix} = i.i.d.\, N\left[\begin{bmatrix} w_t \\ \varphi_t \end{bmatrix}, \begin{bmatrix} 1 & \eta\sigma \\ \eta\sigma & \sigma^2 \end{bmatrix}\right].$$

We set $v_0 = 0$ and $\psi_0 = 0$. The parameters that are held fixed throughout are $a_1 = 1$, $a_2 = -1$, and $\sigma = 4$. We also consider a single sample size $T = 200$, and 1,000 replications are used. We use three values for the correlation coefficient of the disequilibrium errors, namely, $\rho = (0, 0.5, 0.85)$, and three values for the correlation between w_t and φ_t, $\eta = (-0.5, 0, 0.5)$. The base case sets $\theta = 0$, but we also consider $\theta = 0.8$ when $\rho = 0.85$. The estimate of the long-run variance $\hat{\sigma}^2_{1.2}$ as defined by (3) is calculated based on a quadratic spectral kernel and an AR(1) approximation to select the bandwidth (see Andrews, 1991). We also use AR(1) prewhitening as suggested by Andrews and Monahan (1992). Finally, for all methods we consider two maximal values for the number of leads and lags, $k_{\max} = 5, 10$.

Table 1 reports the MSE of the estimate of the cointegrating coefficient obtained using the different procedures and also the coverage rates of the 95% confidence intervals constructed using the estimate $\hat{\sigma}^2_{1.2}$.

Consider first the case with $\rho = \theta = 0$. Here, all methods lead to confidence intervals with a coverage rate close to the 95% nominal level. Using an information criterion generally leads to results with smaller MSE than using the sequential *tsig* method. However, none of the three considered dominates in all cases. When $\eta = 0$ or 0.5, the BIC works relatively well, and when $\eta = -0.5$, using the PIC is as good as using the BIC. Consider now keeping $\theta = 0$ but setting $\rho = 0.5$. The coverage rates become somewhat liberal when $\eta \geq 0$. However, it is the same across methods to select the number of leads and lags, and it is accordingly due to the fact that $\hat{\sigma}^2_{1.2}$ is a less precise estimate of $\sigma^2_{1.2}$. Things are similar when $\rho = 0.85$, except that using the AIC or BIC generally leads to better coverage rates. For the MSE, with $\rho = 0.5$ or 0.85, using the AIC is clearly best when $\eta = 0.5$. When $\eta = 0$, the AIC and BIC are comparable with $k_{\max} = 5$ but using the BIC is better when $k_{\max} = 10$. When $\eta = -0.5$, the PIC and the two *tsig* methods yield lower MSE than the AIC or BIC. Consider now the case with $\rho = 0.85$ and $\theta = 0.8$. Using the AIC is clearly preferable when $\eta = 0.5$. When $\eta \leq 0$, using the *tsig*05 leads to smallest MSE though the gains over what can be achieved using the BIC or the PIC are marginal. The coverage rates are a bit off the 95% nominal level but are generally closer using the AIC.

In summary, when the errors do not contain a moving average (MA) component, at least one of the information criteria performs at least as well as *tsig*10/*tsig*05 for all values of η and both choices for the maximum number of leads and lags. With an MA component in ρ_t, the information criteria outperform the sequential procedures for $\eta > 0$ but not otherwise. However, although the sequential procedures are only slightly better than the information criteria when

Table 1. Mean-squared errors of the estimates and coverage rates of the confidence intervals

	$\rho = 0, \theta = 0$						$\rho = 0.5, \theta = 0$					
	$k_{\max} = 5$			$k_{\max} = 10$			$k_{\max} = 5$			$k_{\max} = 10$		
	η			η			η			η		
	−0.5	0	0.5	−0.5	0	0.5	−0.5	0	0.5	−0.5	0	0.5
(a) MSE of the estimates ($\times 10^5$)												
AIC	6.39	8.27	6.35	7.57	9.91	7.40	25.3	33.2	25.9	30.1	39.9	30.4
BIC	5.94	7.76	6.40	6.54	8.63	7.16	23.3	33.5	28.6	25.9	37.1	32.2
PIC	5.93	7.87	6.68	6.55	8.69	7.39	22.6	34.5	33.3	25.1	38.5	36.7
*tsig*10	6.88	9.46	6.63	9.38	12.20	8.91	22.6	35.0	30.4	28.8	42.2	33.8
*tsig*05	6.62	8.52	6.57	9.23	11.77	8.60	22.2	35.2	34.3	27.2	39.4	38.9
(b) Coverage rates of 95% confidence intervals												
AIC	0.94	0.94	0.94	0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.92	0.92
BIC	0.95	0.95	0.94	0.95	0.95	0.93	0.94	0.92	0.91	0.94	0.92	0.92
PIC	0.95	0.95	0.93	0.95	0.95	0.93	0.94	0.92	0.90	0.95	0.92	0.91
*tsig*10	0.93	0.94	0.93	0.94	0.94	0.94	0.94	0.92	0.91	0.94	0.91	0.90
*tsig*05	0.94	0.93	0.93	0.94	0.95	0.94	0.94	0.92	0.91	0.94	0.92	0.90

	$\rho = 0.85, \theta = 0$						$\rho = 0.85, \theta = 0.8$					
	$k_{\max} = 5$			$k_{\max} = 10$			$k_{\max} = 5$			$k_{\max} = 10$		
	η			η			η			η		
	−0.5	0	0.5	−0.5	0	0.5	−0.5	0	0.5	−0.5	0	0.5
(a) MSE of the estimates ($\times 10^5$)												
AIC	252	313	357	330	382	372	84.5	94.7	97.3	106.6	123.1	109.0
BIC	248	303	405	287	338	448	83.0	86.8	105.0	96.9	97.0	115.2
PIC	229	299	510	260	331	557	79.6	82.9	124.4	89.8	92.7	136.7
*tsig*10	230	299	526	261	332	576	74.0	82.5	124.3	83.2	91.7	137.1
*tsig*05	230	299	530	259	331	580	73.3	82.3	124.6	82.6	91.5	137.0
(b) Coverage rates of 95% confidence intervals												
AIC	0.92	0.89	0.86	0.90	0.88	0.86	0.92	0.93	0.91	0.91	0.92	0.91
BIC	0.92	0.90	0.84	0.91	0.89	0.84	0.91	0.92	0.90	0.90	0.92	0.89
PIC	0.92	0.90	0.81	0.92	0.89	0.81	0.90	0.92	0.88	0.89	0.92	0.86
*tsig*10	0.92	0.90	0.81	0.92	0.89	0.80	0.89	0.92	0.88	0.88	0.92	0.86
*tsig*05	0.92	0.90	0.81	0.91	0.89	0.80	0.89	0.91	0.88	0.88	0.92	0.87

$\eta \leq 0$, the gain from using the information criteria can be quite substantial when $\eta > 0$. With respect to the coverage rates of the 95% asymptotic confidence intervals, at least one of the information criteria performs at least as well as *tsig*10/*tsig*05 for all values of η and both choices for the maximum number of leads and lags. They offer a substantial improvement when $\eta > 0$. Hence, overall our limited simulations show that using an information criterion can lead to estimates with smaller MSE and confidence intervals with better coverage rates compared to sequential testing procedures. Finally, a word on the ranking among the three information criteria in terms of their performance in finite samples. The PIC performs best when $\eta \leq 0$, but its performance deteriorates when $\eta > 0$, in which case AIC/BIC perform relatively better. This is true whether the errors contain an MA component or not.

5. CONCLUSION

We have shown that it is possible to relax both the upper and lower bound conditions on the number of leads and lags in the dynamic cointegrating regression estimated by OLS. The assumptions on the errors and regressors are fairly general, thus making our result valid for a wide range of empirical applications. The result on the lower bound condition is especially important because it allows the use of data dependent rules such as information criteria, which are widely used in applied work. Simulations showed that, relative to sequential general to specific testing procedures, the use of such information criteria can indeed produce estimates with smaller MSEs and confidence intervals with better coverage rates.

NOTES

1. A related procedure is that of Phillips and Loretan (1991), which introduces lags of the disequilibrium error as regressors to make the residuals approximately uncorrelated. This has the advantage of not having to correct the standard errors for correlation in the residuals at the cost of dealing with a nonlinear regression.

2. Deterministic components could be included at the expense of additional complexities in the proofs. All results stated in the text remain valid with minor changes related to the exact nature of the distributions stated.

3. Note that the number of leads and lags of Δz_t need not be the same. We specify the same value for simplicity. Alternatively, one can interpret k_T as the maximum of the number of leads and lags.

4. Chang, Park, and Song (2006) claim that the lower bound condition is not needed. The DGP they consider is, however, different and more restrictive than that considered by Saikkonen (1991). Hence, their conclusions are of less practical interest. They also do not consider the conditions under which $\sigma_{1.2}^2$ can be consistently estimated, and their results are therefore also not applicable to the problem of inference.

REFERENCES

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (eds.), *2nd International Symposium on Information Theory*, pp. 267–281. Akademia Kiado.

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

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston.

Carrion-i-Silvestre, J.L. & A.S. Sansó-i-Rosselló (2004) Testing the Null Hypothesis of Cointegration with Structural Breaks. Manuscript, Departament d'Econometria, Estadística i Economia Espanyola, Universitat de Barcelona.

Chang, Y. & J.Y. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431–447.

Chang, Y., J.Y. Park, & K. Song (2006) Bootstrapping cointegrating regressions. *Journal of Econometrics* 133, 703–739.

Dickey, D.A. & W.A. Fuller (1979) Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.

Haug, A.A. (1996) Tests for cointegration: A Monte Carlo comparison. *Journal of Econometrics* 71, 89–115.

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

Lütkepohl, H. & P. Saikkonen (1999) Order selection in testing for the cointegrating rank of a VAR process. In R.F. Engle & H. White (eds.), *Cointegration, Causality and Forecasting*, pp. 168–199. Oxford University Press.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.

Ng, S. & P. Perron (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268–281.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regressions 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.

Phillips, P.C.B. & W. Ploberger (1994) Posterior odds testing for a unit root with data-based model selection. *Econometric Theory* 10, 774–808.

Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.

Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–21.

Saikkonen, P. (1992) Estimation and testing of cointegrating systems by an autoregressive approximation. *Econometric Theory* 8, 1–27.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.

Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

APPENDIX

As a matter of notation, throughout we use the matrix norm $\|B\|_1 = \sup_{\|x\|\le 1} \|Bx\|$, with $\|\cdot\|$ the standard euclidean norm. Note that $\|B\|_1$ equals the square root of the largest eigenvalue of $B'B$ and that $\|Bx\| \le \|B\|_1\|x\|$. Also, we use the usual norm $\|B\|^2 = \text{tr}(B'B)$, such that $\|B\|_1^2 \le \|B\|^2$. Finally, for any conformable matrices B_1 and B_2, $\|B_1B_2\| \le \|B_1\|\,\|B_2\|_1$.

Proof of Proposition 1. Let $\eta_t = (\Delta z'_{t-k_T}, \ldots, \Delta z'_{t+k_T})'$, $\Pi = (\Pi'_{-k_T}, \ldots, \Pi'_{k_T})'$. We can write (2) as

$$y_t = z_t'\delta + \eta_t'\Pi + v_t^*, \qquad t = k_T + 1, \ldots, T - k_T$$

or in matrix form as

$$Y = Z\delta + \eta\Pi + V^*,$$

where $Z = (z_1, \ldots, z_T)'$, $\eta = (\eta_1, \ldots, \eta_T)'$, $V^* = (v_1^*, \ldots, v_T^*)' = V + \mathcal{E}$ with $V = (v_1, \ldots, v_T)'$ and $\mathcal{E} = (e_1, \ldots e_T)'$. Also let $M_\eta = I - \eta(\eta'\eta)^{-1}\eta'$. We have

$$\hat{\delta} - \delta = (Z'M_\eta Z)^{-1}Z'M_\eta V^* = (Z'M_\eta Z)^{-1}Z'M_\eta(V + \mathcal{E})$$

so that

$$T(\hat{\delta} - \delta) = (A_{1T})^{-1}(A_{2T} + A_{3T}), \tag{A.1}$$

where

$$A_{1T} = T^{-2}Z'Z - T^{-2}Z'\eta(\eta'\eta)^{-1}\eta'Z,$$

$$A_{2T} = T^{-1}Z'V - T^{-1}Z'\eta(\eta'\eta)^{-1}\eta'V,$$

$$A_{3T} = T^{-1}Z'\mathcal{E} - T^{-1}Z'\eta(\eta'\eta)^{-1}\eta'\mathcal{E}.$$

We first prove the following lemma.

LEMMA A.1. *Under Assumptions DGP and K: (i)* $\|(\eta'\eta)^{-1}\|_1 = O_p(T^{-1})$, *(ii)* $\|Z'\eta\| = O_p(Tk_T^{1/2})$, *(iii)* $\|\eta'V\| = O_p(T^{1/2}k_T^{1/2})$, *(iv)* $\|\eta'\mathcal{E}\| = o_p(Tk_T^{-1/2})$, *(v)* $\|Z'\mathcal{E}\| = o_p(Tk_T^{-1})$, *(vi)* $\|\mathcal{E}'\mathcal{E}\| = o_p(T)$, *(vii)* $\|\mathcal{E}'V\| = o_p(T)$, *(viii)* $\|\eta'V^*\| = o_p(Tk_T^{-1/2})$.

Proof of Lemma A.1.

(i) Let $\Gamma_2 = (\Gamma_{2,i-j})_{i,j=-k_T}^{k_T}$, where $\Gamma_{2,h} = E(u_{2t}u_{2t-h})$. From Berk (1974, Lem. 3), it follows that $E\|(T^{-1}\eta'\eta)^{-1} - \Gamma_2^{-1}\|_1^2 \le C_1 T^{-1}k_T^2$ for some constant C_1. Hence, $\|(T^{-1}\eta'\eta)^{-1} - \Gamma_2^{-1}\|_1 = O_p(T^{-1/2}k_T)$. Because $\|\Gamma_2^{-1}\|_1 = O(1)$ uniformly in k_T for sequences such that $T^{-1/2}k_T \to 0$, we have

$$|\,\|(T^{-1}\eta'\eta)^{-1}\|_1 - \|\Gamma_2^{-1}\|_1| \le \|(T^{-1}\eta'\eta)^{-1} - \Gamma_2^{-1}\|_1 = o_p(1),$$

and the result follows.

(ii) The result follows because each element $T^{-1}Z'\eta$ is $O_p(1)$ and the number of elements is of order $O(k_T)$.

(iii) Because the elements of η and V are uncorrelated, the elements of $T^{-1/2}\eta'V$ are each $O_p(1)$, and the result follows because the number of elements is of order $O(k_T)$.

(iv) We have

$$\begin{aligned} E\|T^{-1}\eta'\mathcal{E}\| &\le T^{-1}\sum_{t=k_T+1}^{T-k_T} E(\|e_t\eta_t\|) \le \{E(\|\eta_t\|^2)E(e_t^2)\}^{1/2} \\ &= C_2 k_T^{1/2}\left\{E\left(\sum_{|j|>k_T}\Delta z_{t-j}'\Pi_j\right)^2\right\}^{1/2} \\ &\le C_2 k_T^{1/2}\left\{\sum_{|i|>k_T}\sum_{|j|>k_T}\|\Gamma_{2,i-j}\|\,\|\Pi_i\|\,\|\Pi_j\|\right\}^{1/2} \\ &\le C_3 k_T^{1/2}\sum_{|j|>k_T}\|\Pi_j\| = o(k_T^{-1/2}) \end{aligned}$$

using the fact that $\|\Gamma_{2,i-j}\|$ is uniformly bounded by Assumption DGP.

(v) The result follows from Lemma A.2(a) of Lütkepohl and Saikkonen (1999).

(vi) We have

$$E\|T^{-1}\mathcal{E}'\mathcal{E}\| = T^{-1}\sum_{t=1}^{T} E(e_t^2) = T^{-1}\sum_{t=1}^{T}\sum_{|i|>k_T}\sum_{|a|>k_T}\Pi_i' E(\Delta z_{t-i}\Delta z_{t-a}')\Pi_a$$

$$\leq T^{-1}\sum_{|i|>k_T}\sum_{|a|>k_T}\sum_{t=1}^{T}\|\Pi_i'\|\,\|\Gamma_2(a-i)\|\,\|\Pi_a\|$$

$$\leq o(k_T^{-2}) = o(1)$$

because $\|\Gamma_2(j)\|$ is bounded uniformly in j.

(vii) We have $T^{-1}\sum_{t=1}^{T} v_t e_t = T^{-1}\sum_{|i|>k_T}\Pi_i'\sum_{t=1}^{T}\Delta z_{t-i}v_t$, so that

$$\left\| T^{-1}\sum_{t=1}^{T} v_t e_t \right\| \leq T^{-1}\sum_{|i|>k_T}\|\Pi_i\| \left\| \sum_{t=1}^{T}\Delta z_{t-i}v_t \right\| = o_p(k_T^{-1}T^{-1/2}) = o_p(1),$$

where we used the fact that $T^{-1/2}\sum_{t=1}^{T}\Delta z_{t-i}v_t = O_p(1)$.

(viii) Because $V^* = V + \mathcal{E}$, $\|\eta' V^*\| \leq \|\eta' V\| + \|\eta'\mathcal{E}\| = O_p(T^{1/2}k_T^{1/2}) + o_p(Tk_T^{-1/2}) = o_p(Tk_T^{-1/2})$. ■

Using Lemma A.1, we have

$$\|T^{-2}Z'\eta(\eta'\eta)^{-1}\eta' Z\| \leq T^{-2}\|Z'\eta\|^2\|(\eta'\eta)^{-1}\|_1 = O_p(k_T T^{-1}) = o_p(1),$$

$$\|T^{-1}Z'\eta(\eta'\eta)^{-1}\eta' V\| \leq T^{-1}\|Z'\eta\|\,\|(\eta'\eta)^{-1}\|_1\|\eta' V\| = O_p(k_T T^{-1/2}) = o_p(1),$$

$$\|T^{-1}Z'\eta(\eta'\eta)^{-1}\eta'\mathcal{E}\| \leq T^{-1}\|Z'\eta\|\,\|(\eta'\eta)^{-1}\|_1\|\eta'\mathcal{E}\| = o_p(1),$$

which implies that

$$A_{1T} = T^{-2}Z'Z + o_p(1), \tag{A.2}$$

$$A_{2T} = T^{-1}Z'V + o_p(1), \tag{A.3}$$

$$A_{3T} = o_p(1). \tag{A.4}$$

Because $T^{-2}Z'Z \Rightarrow \int_0^1 B_2 B_2'$ and $T^{-1}Z'V \Rightarrow \int_0^1 B_2\, dB_{1.2}$, part (i) follows upon substitution in (A.1). For part (ii), we have

$$\begin{aligned}
T^{-1}\hat{V}'^{*}\hat{V}^{*} &= T^{-1}(M_\eta Y - M_\eta Z\hat{\delta})'(M_\eta Y - M_\eta Z\hat{\delta}) \\
&= T^{-1}(M_\eta Z(\delta-\hat{\delta}) + M_\eta\mathcal{E} + M_\eta V)'(M_\eta Z(\delta-\hat{\delta}) + M_\eta\mathcal{E} + M_\eta V) \\
&= T^{-1}(T(\delta-\hat{\delta}))'(T^{-2}Z'M_\eta Z)T(\delta-\hat{\delta}) + 2T^{-1}(T(\delta-\hat{\delta}))'T^{-1}Z'M_\eta\mathcal{E} \\
&\quad + 2T^{-1}(T(\delta-\hat{\delta}))'T^{-1}Z'M_\eta V + T^{-1}\mathcal{E}'M_\eta\mathcal{E} + 2T^{-1}\mathcal{E}'M_\eta V + T^{-1}V'M_\eta V \\
&\equiv T1 + T2 + T3 + T4 + T5 + T6.
\end{aligned}$$

We consider the limit of each term. Because $T(\delta - \hat{\delta}) = O_p(1)$, $T^{-2}Z'M_\eta Z = O_p(1)$ from (A.2), $T1 = O_p(T^{-1})$. From (A.4), $T^{-1}Z'M_\eta \mathcal{E} = o_p(1)$ so that $T2 = o_p(T^{-1})$. From (A.3), $T^{-1}Z'M_\eta V = O_p(1)$, which gives $T3 = O_p(T^{-1})$. For $T4$, $T^{-1}\mathcal{E}'M_\eta \mathcal{E} \le T^{-1}\mathcal{E}'\mathcal{E}$ because M_η is an orthogonal projection matrix, and $T4 = o_p(1)$ follows from Lemma A.1(vi). For $T5$,

$$\begin{aligned}\|T^{-1}\mathcal{E}'M_\eta V\| &\le T^{-1}\|\mathcal{E}'V\| + T^{-1}\|\mathcal{E}'\eta\|\,\|(\eta'\eta)^{-1}\|\,\|\eta'V\| \\ &= o_p(k_T^{-1/2}) + T^{-1}o_p(Tk_T^{-1/2})O_p(T^{-1})O_p(T^{1/2}k_T^{1/2}) = o_p(1)\end{aligned}$$

using Lemma A.1(i), (iii), (iv), and (vii). For $T6$,

$$\begin{aligned}T^{-1}V'M_\eta V &= T^{-1}V'V - T^{-1}(V'\eta)(\eta'\eta)^{-1}\eta'V = T^{-1}V'V + O_p(k_T T^{-1}) \\ &= T^{-1}V'V + o_p(1).\end{aligned}$$

Thus, we have $T^{-1}\hat{V}'^*\hat{V}^* = T^{-1}V'V + o_p(1) \xrightarrow{p} \sigma_v^2$, which proves the proposition. ■

Proof of Proposition 2. As a matter of notation, we let $O_p^{j*}(1)$ denote a variable indexed by j that is $O_p(1)$ with bounded second moments uniformly in $0 \le j < T$. We can write

$$\hat{v}_t^* = z_t'(\delta - \hat{\delta}) - \eta_t'(\eta'\eta)^{-1}\eta'Z(\delta - \hat{\delta}) + v_t^* - \eta_t'(\eta'\eta)^{-1}\eta'V^*$$

so that

$$\begin{aligned}&\sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T}\hat{v}_t^*\hat{v}_{t-j}^* \\ &\quad= \sum_{j=1}^{T-1} w(j/m_T)T^{-1}\Bigg\{(\delta-\hat{\delta})'\sum_{t=j+1}^{T} z_t z_{t-j}'(\delta-\hat{\delta}) - (\delta-\hat{\delta})' \\ &\qquad\times\sum_{t=j+1}^{T} z_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'Z(\delta-\hat{\delta}) \\ &\qquad+ (\delta-\hat{\delta})'\sum_{t=j+1}^{T} z_t v_{t-j}^* - (\delta-\hat{\delta})'\sum_{t=j+1}^{T} z_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^* \\ &\qquad+ (\delta-\hat{\delta})'Z'\eta(\eta'\eta)^{-1}\sum_{t=j+1}^{T}\eta_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'Z(\delta-\hat{\delta}) \\ &\qquad- (\delta-\hat{\delta})'Z'\eta(\eta'\eta)^{-1}\sum_{t=j+1}^{T}\eta_t v_{t-j}^* \\ &\qquad+ (\delta-\hat{\delta})'Z'\eta(\eta'\eta)^{-1}\sum_{t=j+1}^{T}\eta_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^* \\ &\qquad+ \sum_{t=j+1}^{T} v_t^*\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^* \\ &\qquad+ V^{*\prime}\eta(\eta'\eta)^{-1}\sum_{t=j+1}^{T}\eta_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^* + \sum_{t=j+1}^{T} v_t^* v_{t-j}^*\Bigg\} \\ &\quad\equiv \sum_{i=1}^{10} Si, \quad \text{say.}\end{aligned}$$

We consider each term in turn, and in the derivations we shall repeatedly use the facts that $m_T^{-1}\sum_{j=1}^{T-1}|w(j/m_T)| \to \int_0^\infty |w(s)|\,ds$ and $T(\delta-\hat{\delta}) = O_p(1)$. For $S1$,

$$T^{-1}\left\| \sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})' \sum_{t=j+1}^{T} z_t z_{t-j}'(\delta-\hat{\delta}) \right\|$$

$$\leq T^{-1}\sum_{j=1}^{T-1} |w(j/m_T)|\,\|T(\delta-\hat{\delta})\|^2 \left\| T^{-2}\sum_{t=j+1}^{T} z_t z_{t-j}' \right\| = O_p(m_T T^{-1}) = o_p(1)$$

because each element of the matrix $T^{-2}\sum_{t=j+1}^{T} z_t z_{t-j}'$ is $O_p^{j*}(1)$. For $S2$,

$$T^{-1}\left\| \sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})' \sum_{t=j+1}^{T} z_t \eta_{t-j}'(\eta'\eta)^{-1}\eta' Z(\delta-\hat{\delta}) \right\|$$

$$\leq T^{-2}\sum_{j=1}^{T-1} |w(j/m_T)|\,\|T(\delta-\hat{\delta})\|^2 \left\| T^{-1}\sum_{t=j+1}^{T} z_t \eta_{t-j}' \right\| \|(\eta'\eta)^{-1}\|_1 \|\eta' Z\|$$

$$= O_p(m_T k_T T^{-2}) = o_p(1)$$

because each element of the matrix $T^{-1}\sum_{t=j+1}^{T} z_t \eta_{t-j}'$ is $O_p^{j*}(1)$ and the number of elements is of order k_T. For $S3$,

$$T^{-1}\left\| \sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})' \sum_{t=j+1}^{T} z_t v_{t-j}^{*} \right\|$$

$$\leq T^{-1}\sum_{j=1}^{T-1} |w(j/m_T)|\,\|T(\delta-\hat{\delta})\| \left\| T^{-1}\sum_{t=j+1}^{T} z_t v_{t-j}^{*} \right\| = O_p(m_T T^{-1}) = o_p(1)$$

because each element of the vector $T^{-1}\sum_{t=j+1}^{T} z_t v_{t-j}^{*}$ is $O_p^{j*}(1)$. For $S4$,

$$T^{-1}\left\| \sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})' \sum_{t=j+1}^{T} z_t \eta_{t-j}'(\eta'\eta)^{-1}\eta' V^{*} \right\|$$

$$\leq T^{-1}\sum_{j=1}^{T-1} |w(j/m_T)|\,\|T(\delta-\hat{\delta})\| \left\| T^{-1}\sum_{t=j+1}^{T} z_t \eta_{t-j}' \right\| \|(\eta'\eta)^{-1}\|_1 \|\eta' V^{*}\|$$

$$= O_p(m_T T^{-1}) = o_p(1).$$

For $S5$,

$$T^{-1}\left\| \sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})' Z'\eta(\eta'\eta)^{-1} \sum_{t=j+1}^{T} \eta_t \eta_{t-j}'(\eta'\eta)^{-1}\eta' Z(\delta-\hat{\delta}) \right\|$$

$$\leq T^{-2}\sum_{j=1}^{T-1} |w(j/m_T)|\,\|T(\delta-\hat{\delta})\|^2 \|Z'\eta\|^2 \|(\eta'\eta)^{-1}\|_1^2 \left\| T^{-1}\sum_{t=j+1}^{T} \eta_t \eta_{t-j}' \right\|_1$$

$$= O_p(m_T k_T T^{-2}) = o_p(1).$$

This follows since $\|T^{-1}\sum_{t=j+1}^{T}\eta_t\eta_{t-j}' - \Gamma_{2,j}\|_1 = o_p(1)$ (e.g., Berk, 1974) and $\|\Gamma_{2,j}\|_1 = O_p^{j*}(1)$. For $S6$,

$$\begin{aligned}
&T^{-1}\left\|\sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})'Z'\eta(\eta'\eta)^{-1}\sum_{t=j+1}^{T}\eta_t v_{t-j}^*\right\| \\
&\quad\le T^{-1}\sum_{j=1}^{T-1}|w(j/m_T)|\,\|T(\delta-\hat{\delta})\|\,\|Z'\eta\|\,\|(\eta'\eta)^{-1}\|_1\left\|T^{-1}\sum_{t=j+1}^{T}\eta_t v_{t-j}^*\right\| \\
&\quad= o_p(m_T T^{-1}) = o_p(1)
\end{aligned}$$

because $\|T^{-1}\sum_{t=j+1}^{T}\eta_t v_{t-j}^*\| = o_p(k_T^{-1/2})$ uniformly in j because the result of Lemma A.1(viii) continues to hold uniformly in j with v_t^* replaced by v_{t-j}^*. For $S7$,

$$\begin{aligned}
&T^{-1}\left\|\sum_{j=1}^{T-1} w(j/m_T)(\delta-\hat{\delta})'Z'\eta(\eta'\eta)^{-1}\sum_{t=j+1}^{T}\eta_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^*\right\| \\
&\quad\le T^{-1}\sum_{j=1}^{T-1}|w(j/m_T)|\,\|T(\delta-\hat{\delta})\|\,\|Z'\eta\|\,\|(\eta'\eta)^{-1}\|_1^2\left\|T^{-1}\sum_{t=j+1}^{T}\eta_t\eta_{t-j}'\right\|_1\|\eta'V^*\| \\
&\quad= O_p(m_T T^{-1}) = o_p(1).
\end{aligned}$$

For $S8$,

$$\begin{aligned}
&\sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T} v_t^*\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^* \\
&\quad= \sum_{j=1}^{T-1} w(j/m_T)\left(T^{-1}\sum_{t=j+1}^{T} v_t\eta_{t-j}' + T^{-1}\sum_{t=j+1}^{T} e_t\eta_{t-j}'\right)(\eta'\eta)^{-1}\eta'V^* \\
&\quad= \sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T} e_t\eta_{t-j}'(\eta'\eta)^{-1}\eta'V^* + O_p(m_T T^{-1/2}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T} e_t\eta_{t-j}' \\
&\quad= \sum_{|i|>k_T}\Pi_i'\sum_{j=1}^{T-1} w(j/m_T)\left\{\left(T^{-1}\sum_{t=j+1}^{T}\Delta z_{t-i}\eta_{t-j}' - \Gamma_2^i(j)\right) + \Gamma_2^i(j)\right\},
\end{aligned}$$

where $\Gamma_2^i(j) = (\Gamma_2(k_T+j-i),\ldots,\Gamma_2(j-k_T-i))$ so that

$$\begin{aligned}
&\left\|\sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T} e_t\eta_{t-j}'\right\|_1 \\
&\quad\le \left\|\sum_{|i|>k_T}\Pi_i'\right\|\left\{\sum_{j=1}^{T-1}|w(j/m_T)|\left\|\left(T^{-1}\sum_{t=j+1}^{T}\Delta z_{t-i}\eta_{t-j}' - \Gamma_2^i(j)\right)\right\|_1\right. \\
&\qquad\qquad\left. + \left\|\sum_{j=1}^{T-1} w(j/m_T)\Gamma_2^i(j)\right\|\right\} \\
&\quad\le o_p(m_T k_T^{-1/2}T^{-1/2}) + o(k_T^{-1})\left\|\sum_{j=1}^{T-1} w(j/m_T)\Gamma_2^i(j)\right\|. \qquad \textbf{(A.5)}
\end{aligned}$$

Finally,

$$\left\| \sum_{j=1}^{T-1} w(j/m_T)\Gamma_2^i(j) \right\|^2 = \sum_{l=-k_T}^{k_T} \left\| \sum_{j=1}^{T-1} w(j/m_T)\Gamma_2(j-i-l) \right\|^2$$

$$\leq \sum_{l=-k_T}^{k_T} \left\{ \sum_{j=1}^{T-1} \|\Gamma_2(j-i-l)\| \right\}^2 = O(k_T).$$

Hence, from (A.5), $\|\sum_{j=1}^{T-1} w(j/m_T)T^{-1}\sum_{t=j+1}^{T} e_t \eta'_{t-j}\|_1 = o_p(1)$. For $S9$, where $\Gamma_\eta(j) = E(\eta_t \eta'_{t-j})$,

$$\left\| \sum_{j=1}^{T-1} w(j/m_T)T^{-1} \sum_{t=j+1}^{T} \eta_t \eta'_{t-j} \right\|_1$$

$$\leq \sum_{j=1}^{T-1} |w(j/m_T)| \left\| T^{-1} \sum_{t=j+1}^{T} \eta_t \eta'_{t-j} - \Gamma_\eta(j) \right\|_1 + \left\| \sum_{j=1}^{T-1} w(j/m_T)\Gamma_\eta(j) \right\|$$

$$\leq O_p(m_T k_T T^{-1/2}) + O(k_T) = O_p(k_T)$$

using similar arguments as in $S8$, which implies that

$$\left\| V^{*\prime}\eta(\eta'\eta)^{-1} \sum_{j=1}^{T-1} w(j/m_T)T^{-1} \sum_{t=j+1}^{T} \eta_t \eta'_{t-j}(\eta'\eta)^{-1}\eta' V^* \right\|$$

$$\leq \|V^{*\prime}\eta\|^2 \|(\eta'\eta)^{-1}\|_1^2 \left\| \sum_{j=1}^{T-1} w(j/m_T)T^{-1} \sum_{t=j+1}^{T} \eta_t \eta'_{t-j} \right\|_1 = o_p(1).$$

For $S10$,

$$\sum_{j=1}^{T-1} w(j/m_T)T^{-1} \sum_{t=j+1}^{T} v_t^* v_{t-j}^*$$

$$= \sum_{j=1}^{T-1} w(j/m_T)$$

$$\times \left\{ T^{-1} \sum_{t=j+1}^{T} v_t v_{t-j} + T^{-1} \sum_{t=j+1}^{T} v_t e_{t-j} + T^{-1} \sum_{t=j+1}^{T} v_{t-j} e_t + T^{-1} \sum_{t=j+1}^{T} e_t e_{t-j} \right\}.$$

First, note that

$$\sum_{j=1}^{T-1} w(j/m_T)E(e_t e_{t-j}) = \sum_{j=1}^{T-1} w(j/m_T) \sum_{|i|>k_T} \sum_{|a|>k_T} \Pi_i' E(\Delta z_{t-i} \Delta z'_{t-a-j})\Pi_a$$

$$\leq \sum_{|i|>k_T} \sum_{|a|>k_T} \sum_{j=1}^{T-1} |w(j/m_T)| \|\Pi_i'\| \|\Gamma_2(j+a-i)\| \|\Pi_a\|$$

$$\leq o(k_T^{-2}) \sum_{j=-\infty}^{\infty} \|\Gamma_2(j)\| = o(k_T^{-2}).$$

Finally, we have $T^{-1}\sum_{t=j+1}^{T} v_t e_{t-j} = \sum_{|i|>k_T} \Pi_i' \sum_{t=j+1}^{T} \Delta z_{t-i-j} v_t$, so that

$$\left\| T^{-1} \sum_{t=j+1}^{T} v_t e_{t-j} \right\| \leq \sum_{|i|>k_T}^{T-1} \|\Pi_i\| \left\| T^{-1} \sum_{t=j+1}^{T} \Delta z_{t-i-j} v_t \right\| = o_p(k_T^{-1} T^{-1/2}) = o_p(1),$$

where we used the fact that $\sup_{0<r\leq 1} \|T^{-1/2} \sum_{t=j+1}^{[\mathrm{Tr}]} \Delta z_{t-i-j} v_t\| = O_p(1)$. Similar arguments hold for $T^{-1}\sum_{t=j+1}^{T} v_{t-j} e_t$. Hence, combining the results for terms $S1$–$S10$, we have

$$\sum_{j=1}^{T-1} w(j/m_T) T^{-1} \sum_{t=j+1}^{T} \hat{v}_t^* \hat{v}_{t-j}^* = \sum_{j=1}^{T-1} w(j/m_T) T^{-1} \sum_{t=j+1}^{T} v_t v_{t-j} + o_p(1),$$

which proves the proposition. ∎