Economics Letters 111 (2011) 170–172

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

How useful is yet another data-driven bandwidth in long-run variance estimation?: A simulation study on cointegrating regressions

Masayuki Hirukawa *

Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka, 572-8508, Japan

ARTICLE INFO

Article history:
Received 4 November 2009
Received in revised form 12 January 2011
Accepted 2 February 2011
Available online 12 February 2011

JEL classification:
C13
C14
C22

Keywords:
Bandwidth
Cointegration
Kernel
Long-run variance
Simulation

ABSTRACT

This paper investigates how bandwidth choice rules in long-run variance estimation affect finite-sample performance of efficient estimators for cointegrating regression models. Monte Carlo results indicate that Hirukawa's (2010) bandwidth choice rule contributes bias reduction in the estimators.

1. Introduction

In time series econometrics, estimating the long-run variance ("LRV") matrix of a random vector process is essential for empirical research on estimation (e.g. generalized method of moments) and testing (e.g. standard error calculation and unit-root testing) problems. This paper focuses on a standard, kernel smoothing approach to LRV estimation, and investigates finite-sample performance of the bandwidth choice rule newly proposed by Hirukawa (2010) via Monte Carlo simulations, in comparison with two most popular ones by Andrews (1991) and Newey and West (1994). The simulation study is conducted in the context of cointegrating regressions. While there is rich literature in simulation studies on cointegrating regressions, this paper contributes the literature in the following two respects. First, while Andrews' (1991) and Newey and West's (1994) bandwidth choice rules ("A rule" and "NW rule") are frequently used in simulation studies, very little is known about finite-sample performance of Hirukawa's (2010) solve-the-equation plug-in bandwidth choice rule ("SP rule"). To the best of our knowledge, there are only two simulation results on the SP rule, both of which are available in Hirukawa (2010). Neither is on cointegrating regressions. Second, the SP rule is expected to contribute the bias correction method by Kurozumi and Hayakawa (2009) for the fully modified least squares ("FMLS"; Phillips and Hansen, 1990) and canonical cointegration regression ("CCR"; Park, 1992) estimators of cointegrating vectors. When the $I(1)$ regressors are endogenous and/or the regression errors are serially correlated in cointegrating regressions, the ordinary least squares ("OLS") estimators of cointegrating vectors suffer so-called the "second-order bias". Although FMLS and CCR are proposed as devices for correcting the second-order bias nonparametrically by means of LRV estimators, these methods work poorly in the presence of strong serial dependence in regression errors. Then, Kurozumi and Hayakawa (2009) develop a further bias reduction method for FMLS and CCR when regression errors obey an AR(1) model with the AR coefficient moderately close to unity. However, their simulation results indicate that average lengths of bandwidths from the A and NW rules are too long and too short for the purpose of bias reduction. Apparently, there is a need for a bandwidth choice rule that tends to yield intermediate lengths. The SP rule is expected to be a remedy, as suggested in the next section.

2. Three Bandwidth Formulae in LRV Estimation

To illustrate the difference in three bandwidth choice rules, consider a problem of estimating the LRV of a zero-mean scalar process h_t, where the LRV is defined as $\omega = \sum_{j=-\infty}^{\infty}\gamma(j) = \sum_{j=-\infty}^{\infty}E(h_t h_{t-j})$. Given T observations $\{h_t\}_{t=1}^{T}$, a kernel $k(\cdot)$ and a bandwidth M, the kernel estimator of ω is given by a weighted sum of sample autocovariances

* Tel.: +81 72 839 8095; fax: +81 72 839 8138.
E-mail address: hirukawa@econ.setsunan.ac.jp.

0165-1765/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2011.02.006

$\hat{\omega} = \sum_{j=-(T-1)}^{T-1} k(j/M)\,\hat{\gamma}(j)$, where $\hat{\gamma}(j) = T^{-1}\sum_{t=max\{1,1+j\}}^{min\{T+j,T\}} h_t h_{t-j}$. Each of the three rules is an estimator of the minimizer of the asymptotic mean squared error ("AMSE") of $\hat{\omega}$. The minimizer (= AMSE-optimal bandwidth) is given by

$$M^* = \left\{ \frac{q k_q^2 \left(R^{(q)}\right)^2}{\int_{-\infty}^{\infty} k^2(x)dx} \right\}^{1/(2q+1)} T^{1/(2q+1)},$$

where $q(>0)$ is the characteristic exponent of $k(\cdot)$ that satisfies $k_q = lim_{x\to 0}\{1-k(x)\}/|x|^q \in (0,\infty)$, $s^{(n)} = \sum_{j=-\infty}^{\infty} |j|^n \gamma(j)$, and $R^{(q)} = s^{(q)}/s^{(0)}$ is the only unknown quantity called the normalized curvature in Hirukawa (2010). The three rules differ in how to estimate the quantity. Andrews (1991) estimates it parametrically by fitting an AR(1) model to h_t as a reference. In contrast, to avoid the issue of misspecifying the process, Newey and West (1994) estimate the normalized curvature nonparametrically using the truncated kernel. Because the use of the truncated kernel prevents them from providing an optimal bandwidth for the normalized curvature estimator, they implement the bandwidth for the truncated estimator in an ad hoc manner.

The SP rule in Hirukawa (2010) is established as an analog to the bandwidth choice rule for probability density estimation by Sheather and Jones (1991). Similar to Newey and West (1994), the normalized curvature is estimated nonparametrically using the same kernel $k(\cdot)$ and a different bandwidth b. A remarkable difference from the approach in Newey and West (1994) is that Hirukawa (2010) has derived the AMSE-optimal bandwidth b^* for the kernel normalized curvature estimator $\hat{R}^{(q)}(b) = \sum_{j=-(T-1)}^{T-1} k(j/b)|j|^q \hat{\gamma}(j) / \sum_{j=-(T-1)}^{T-1} k(j/b)\hat{\gamma}(j)$. In spirit of Sheather and Jones (1991), the SP rule can be implemented by numerically solving the following fixed point problem for M^*:

$$M^* = \left[\frac{q k_q^2 \left\{ \hat{R}^{(q)}(b^*(M^*)) \right\}^2}{\int_{-\infty}^{\infty} k^2(x)dx} \right]^{1/(2q+1)} T^{1/(2q+1)},$$

$$b^*(M^*) = \left\{ \frac{\alpha^2(q)\int_{-\infty}^{\infty} k^2(x)dx}{(2q+1)\int_{-\infty}^{\infty} |x|^{2q} k^2(x)dx} \right\}^{1/(4q+1)} M^{*(2q+1)/(4q+1)},$$

where the unknown quantity $\alpha(q) = s^{(q)}/s^{(0)} - s^{(2q)}/s^{(q)}$ is estimated by fitting an AR(1) model to h_t. By construction, the SP rule incorporates both parametric and nonparametric approaches. While it relies on a parametric fitting to the unknown process h_t for implementation, it is still based on a kernel estimator of the normalized curvature, which limits the influence of the parametric reference. As a consequence, it is anticipated that when the A and NW rules tend to yield long and short bandwidth lengths (which is typical for moderately to highly persistent h_t), the SP rule is likely to pick up intermediate lengths.

3. Monte Carlo Simulations

This paper adopts the experimental design of Kurozumi and Hayakawa (2009). Let the data $\{(y_t, x_t)\}_{t=1}^{T} \in \mathbb{R}^2$ be generated by

$$y_t = \theta_0 + \theta_1 x_t + u_{1t} := \mathbf{z}_t'\theta + u_{1t}, \Delta x_t = u_{2t}, \tag{1}$$

where the error term $\mathbf{u}_t = (u_{1t}, u_{2t})'$ obeys

$$u_{1t} = \rho u_{1t-1} + \epsilon_{1t}, u_{2t} = \epsilon_{2t}, \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{iid}{\sim} N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{bmatrix} \right).$$

True parameter values of $\theta = (\theta_0, \theta_1)'$, ρ, and σ_{21} are $\theta_0 = \theta_1 = 1$, $\rho \in \{0.7, 0.8, 0.85, 0.9, 0.95\}$, and $\sigma_{21} \in \{0.4, 0.8\}$. For each combination of parameter values, 10,000 data sets of sample sizes $T = 100$ or $T = 300$ are simulated.

Because of the endogeneity and the serially correlated regression error in Eq. (1), $\hat{\theta}_{OLS}$ (= OLS estimator of θ) is consistent but inefficient due to the second-order bias. FMLS and CCR are developed as efficient estimation methods that eliminate the second-order bias. Let Ω and Λ be the two- and one-sided LRVs of $\mathbf{u}_t$. Specifically,

$$\Omega = \Sigma + \Gamma + \Gamma' = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \Lambda = \Sigma + \Gamma = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix},$$

where $\Sigma = E\left(\mathbf{u}_t\mathbf{u}_t'\right)$ and $\Gamma = \sum_{j=1}^{\infty} E\left(\mathbf{u}_t\mathbf{u}_{t+j}'\right)$. Also let $\hat{\Omega}$ and $\hat{\Lambda}$ be the kernel estimators of Ω and Λ with $\mathbf{u}_t$ replaced by $\hat{\mathbf{u}}_t = (\hat{u}_{1t}, \Delta x_t)'$, where $\hat{u}_{1t}$ is the OLS residual of Eq. (1). The FMLS estimator of θ is given by

$$\hat{\theta}_{FMLS} = \left(\sum_{t=1}^{T} \mathbf{z}_t\mathbf{z}_t' \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_t y_t^+ - T\hat{\mathbf{J}}^+ \right),$$

where $y_t^+ = y_t - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\Delta x_t$ and $\hat{\mathbf{J}}^+ = \left(0, \left(\hat{\Lambda}_{21} - \hat{\Lambda}_{22}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}\right)'\right)'$. CCR employs the transformed data

$$x_t^* = x_t - \left(\hat{\Sigma}^{-1}\hat{\Lambda}_2\right)'\hat{\mathbf{u}}_t, y_t^* = y_t - \left(\hat{\Sigma}^{-1}\hat{\Lambda}_2\hat{\theta}_{OLS} + \left(0, \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\right)'\right)'\hat{\mathbf{u}}_t,$$

where $\hat{\Sigma}$ is a consistent estimator of Σ. Defining $\mathbf{z}_t^* = (1, x_t^*)'$ yields the CCR estimator of θ as

$$\hat{\theta}_{CCR} = \left(\sum_{t=1}^{T} \mathbf{z}_t^*\mathbf{z}_t^{*\prime} \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_t^* y_t^* \right).$$

However, FMLS and CCR work poorly when the regression error u_{1t} exhibits strong serial dependence. Kurozumi and Hayakawa (2009) consider further bias reduction in FMLS and CCR for the N local-to-unity system such that ρ is modeled as $\rho = 1 - c/N$, where N satisfies $N \to \infty$ but $N = o(T)$. In addition, if $M/N \to d_M$ for a bandwidth of the LRV estimators M and if another bandwidth M_c satisfies $M_c = o(N)$, c can be consistently estimated by $\hat{c} = (N/2)(\hat{\omega}_{\Delta 11}/\hat{\sigma}_{11})$, where $\hat{\sigma}_{11} = T^{-1}\sum_{t=1}^{T} \hat{u}_{1t}^2$ and

$$\hat{\omega}_{\Delta 11} = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{M_c}\right)\left(\frac{1}{T}\sum_{t=max\{1,1+j\}}^{min\{T+j,T\}} \Delta\hat{u}_{1t}\Delta\hat{u}_{1t-j}\right).$$

The bias-corrected FMLS and CCR ("FMLS-BC" and "CCR-BC") estimators of θ can be obtained by replacing $\left(\hat{\Omega}_{21}, \hat{\Lambda}_{21}\right)$ in $\hat{\theta}_{FMLS}$ and $\hat{\theta}_{CCR}$ with $\left(\tilde{\Omega}_{21}, \tilde{\Lambda}_{21}\right) = \left(\hat{\Omega}_{21}/\hat{\kappa}, \hat{\Lambda}_{21}/\hat{\kappa}\right)$, where $\hat{\kappa} = \hat{c}\,d_M\int_0^\infty k(r) \times exp(-\hat{c}d_M r)dr$. To implement FMLS-BC and CCR-BC, Kurozumi and Hayakawa (2009) suggest that extra tuning parameters are set equal to $d_M = 1$, $N = M$, and $M_c = M^{2/3}$.

In this simulation study, θ_1 is estimated by OLS, FMLS, CCR, FMLS-BC, and CCR-BC for each data set. Finite-sample performance of each estimator is evaluated by its bias and mean squared error ("MSE"). The Parzen kernel is employed for all LRV estimators, and three bandwidths are computed via the A, NW, and SP rules. The lag length for the truncated estimator of the normalized curvature in the NW rule is set equal to $\lfloor 4(T/100)^{4/25} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. Bandwidth values are trimmed at the sample size whenever necessary. For FMLS-BC and CCR-BC, the SP rule sometimes picks up a zero bandwidth. In this case, $\hat{c} = 0$ by $N = M = 0$, and thus $\left(\tilde{\Omega}_{21}, \tilde{\Lambda}_{21}\right)$ are not well-defined. Since the zero bandwidth is an indication of weak serial dependence, no bias correction is made for FMLS or CCR whenever it occurs.

4. Conclusion

Table 1 reports biases and MSEs of the five estimators of θ_1. Because the results for $\sigma_{21} = 0.8$ are qualitatively similar, only those

Table 1
Biases and MSEs of estimators of θ_1 ($\sigma_{21} = 0.4$).

		$T = 100$					$T = 300$				
	ρ	0.70	0.80	0.85	0.90	0.95	0.70	0.80	0.85	0.90	0.95
Bias											
OLS		0.0623	0.0866	0.1077	0.1425	0.2114	0.0227	0.0330	0.0428	0.0609	0.1059
FMLS	A	0.0318	0.0527	0.0722	0.1062	0.1780	0.0068	0.0128	0.0192	0.0329	0.0727
	NW	0.0305	0.0510	0.0710	0.1070	0.1839	0.0070	0.0137	0.0213	0.0375	0.0830
	SP	0.0312	0.0512	0.0704	0.1043	0.1775	0.0071	0.0129	0.0192	0.0325	0.0715
CCR	A	0.0322	0.0538	0.0739	0.1087	0.1814	0.0070	0.0132	0.0200	0.0342	0.0751
	NW	0.0308	0.0516	0.0718	0.1079	0.1847	0.0072	0.0140	0.0217	0.0379	0.0833
	SP	0.0315	0.0520	0.0715	0.1059	0.1798	0.0072	0.0133	0.0199	0.0335	0.0734
FMLS-BC	A	0.0216	0.0464	0.0677	0.1033	0.1766	0.0027	0.0103	0.0175	0.0318	0.0723
	NW	0.0128	0.0370	0.0590	0.0973	0.1773	0.0007	0.0084	0.0166	0.0333	0.0797
	SP	0.0120	0.0373	0.0593	0.0962	0.1728	0.0008	0.0084	0.0157	0.0300	0.0702
CCR-BC	A	0.0239	0.0488	0.0703	0.1065	0.1804	0.0033	0.0110	0.0185	0.0333	0.0747
	NW	0.0164	0.0402	0.0619	0.0999	0.1794	0.0015	0.0092	0.0173	0.0340	0.0802
	SP	0.0159	0.0407	0.0625	0.0995	0.1762	0.0015	0.0093	0.0167	0.0314	0.0722
MSE											
OLS		0.0137	0.0261	0.0401	0.0697	0.1513	0.0019	0.0040	0.0067	0.0133	0.0392
FMLS	A	0.0123	0.0260	0.0423	0.0783	0.1828	0.0014	0.0032	0.0057	0.0125	0.0419
	NW	0.0115	0.0245	0.0404	0.0764	0.1822	0.0013	0.0030	0.0054	0.0118	0.0393
	SP	0.0114	0.0246	0.0408	0.0777	0.1882	0.0013	0.0030	0.0054	0.0119	0.0409
CCR	A	0.0122	0.0258	0.0419	0.0771	0.1786	0.0014	0.0032	0.0057	0.0124	0.0414
	NW	0.0114	0.0244	0.0401	0.0759	0.1811	0.0013	0.0030	0.0054	0.0117	0.0392
	SP	0.0114	0.0244	0.0405	0.0769	0.1852	0.0013	0.0030	0.0054	0.0118	0.0406
FMLS-BC	A	0.0129	0.0268	0.0433	0.0797	0.1850	0.0014	0.0032	0.0057	0.0126	0.0420
	NW	0.0121	0.0255	0.0421	0.0800	0.1919	0.0013	0.0030	0.0053	0.0117	0.0396
	SP	0.0120	0.0256	0.0425	0.0807	0.1945	0.0013	0.0030	0.0054	0.0119	0.0412
CCR-BC	A	0.0126	0.0264	0.0426	0.0781	0.1802	0.0014	0.0032	0.0057	0.0125	0.0415
	NW	0.0116	0.0249	0.0413	0.0786	0.1891	0.0013	0.0030	0.0053	0.0117	0.0395
	SP	0.0115	0.0250	0.0416	0.0791	0.1901	0.0013	0.0030	0.0054	0.0118	0.0407
Bandwidth											
A	Mean	18.33	25.62	31.58	40.88	56.53	24.21	34.86	44.28	60.74	98.10
	Std. Dev.	4.42	7.14	9.96	15.49	26.21	3.32	5.56	8.00	13.20	28.48
	#{Trimmed}	0	0	0	58	738	0	0	0	0	0
NW	Mean	9.94	11.23	11.88	12.47	12.96	13.48	15.07	15.77	16.40	16.98
	Std. Dev.	2.33	1.80	1.46	1.16	0.92	1.48	0.84	0.63	0.46	0.33
	#{Trimmed}	0	0	0	0	0	0	0	0	0	0
SP	Mean	9.24	12.19	14.36	17.45	21.86	13.99	19.17	23.35	29.99	43.19
	Std. Dev.	3.62	4.97	6.10	8.07	10.79	3.24	4.29	5.37	7.46	13.60
	#{Zero}	18	3	0	0	0	0	0	0	0	0

Note: "Mean", "Std. Dev.", "#{Trimmed}", and "#{Zero}" for bandwidths are averages, standard deviations, numbers of trimmed bandwidths at the sample size, and numbers of zero bandwidths, respectively.

for $\sigma_{21} = 0.4$ are presented. Descriptive statistics on bandwidths are also provided for convenience. Frequencies of trimmed and zero bandwidths are relatively small, and thus their influences over the results appear to be negligible.

The results are summarized as follows:

- FMLS and CCR are less biased than OLS, and FMLS-BC and CCR-BC are less biased than FMLS and CCR. However, the advantage in bias of FMLS-BC and CCR-BC over FMLS and CCR appears to be diminishing as ρ increases. In addition, FMLS and FMLS-BC tend to be less biased than CCR and CCR-BC.
- Results on MSEs are mixed. While four efficient estimators do not always reduce MSEs from OLS for $\sigma_{21} = 0.4$, these estimators yield smaller MSEs than OLS for $\sigma_{21} = 0.8$ (unreported).
- For each of FMLS and CCR, when ρ is close to unity, the SP rule tends to yield the smallest bias among three bandwidth choice rules.
- In the presence of moderate to high persistence in regression errors, average lengths of bandwidths tend to be in the order of the NW, SP, and A rules from shortest to longest, as anticipated. For each of FMLS-BC and CCR-BC, as ρ gets closer to unity, the SP rule is more likely to give the smallest bias among three bandwidth choice rules.
- Due to the structure of the error term, the experimental design is favorable to the A rule. Nonetheless, estimators based on the A rule are often most biased.

References

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

Hirukawa, M., 2010. A two-stage plug-in bandwidth selection and its implementation for covariance estimation. Econometric Theory 26, 710–743.

Kurozumi, E., Hayakawa, K., 2009. Asymptotic properties of the efficient estimators for cointegrating regression models with serially dependent errors. Journal of Econometrics 149, 118–135.

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

Park, J.Y., 1992. Canonical cointegrating regressions. Econometrica 60, 119–143.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. The Review of Economic Studies 57, 99–125.

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B 53, 683–690.