Communications of the Korean Statistical Society
2009, Vol. 16, No. 4, 707–711

Sign IV Cointegration Tests

Yujin Oh[1,a]

[a]Leeds School of Business, University of Colorado

Abstract

We propose new cointegration tests using signs of the regressors as instrumental variable. Our tests have the asymptotic standard normal distribution and are free from the dimension of regressors under the null hypothesis of no cointegration. A Monte-Carlo simulation shows that the proposed tests have a stable size and an improved power. Particulary, the tests have better power for small numbers of observations.

Keywords: Cointegration, test statistic, sign, instrumental variable, standard normal distribution.

1. Introduction

Since the seminal work of Engle and Granger (1987, hereafter: EG test), there have been a number of research and empirical works on cointegration. Most existing cointegration tests adopt ordinary least squares estimation and the limiting null distributions of the test statistics are not standard normal distributions. In addition, mostly the distributions depend on the dimension of the regressors and types of deterministic trends.

Motivated by the sign test for unit roots of So and Shin (1999a), we develop EG type tests based on sign as an instrumental variable(IV) in a single equation regression model. The asymptotic null distribution of the proposed tests, the sign IV cointegration tests, is the standard normal, and does not depend on the number of regressors. Thus there is no need to calculate separate critical values for the new tests in contrast to other tests based on ordinary least squares estimates, and the proposed tests are easy to implement.

The remainder of this paper is organized as follows. Section 2 proposes the test statistics, and investigates the asymptotic null distributions of the tests. Section 3 provides some Monte-Carlo simulation results, and Section 4 concludes.

2. Test Statistics

Let $y_{1,t}, \ldots, y_{p,t}$ be p integrated variables, $t = 1, \ldots, T$, and the first differences of $y_{i,t}$ be *i.i.d.* with mean 0 and variance σ_i^2. If there is a linear combination which is stationary, then we can conclude that the p variables are *cointegrated*. Let u_t be a linear combination of y_is such as $u_t = \beta' Y_t$ where $Y_t = (y_{1,t}, \ldots, y_{p,t})'$ and β is a $(p \times 1)$ coefficient vector. Without loss of generality we can let the first element of β be 1. In this sense, we are interested in testing whether u_t has a unit root or not. We consider the single equation cointegration model, $\Delta u_t = \phi u_t + \epsilon_t$, and this model can be represented as follows:

$$\Delta(y_{1,t} - \alpha' X_t) = \phi(y_{1,t-1} - \alpha' X_{t-1}) + \epsilon_t, \quad t = 1, \ldots, T, \tag{2.1}$$

This work was supported by the Korea Research Foundation Grant funded by the Korean Government[KRF-2008-356-B00002].

[1] Post Doctor, Department of Finance, Leeds School of Business, University of Colorado, Boulder, CO 80309-0419, USA. E-mail: ouj92@hotmail.com

where $X_t=(y_{2,t},\ldots,y_{p,t})$, $\alpha = (\alpha_2,\alpha_3,\ldots,\alpha_p)'$, ϕ is the AR(1) coefficient, ϵ_t are a white noise series, and Δ is the first-difference operator such as $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$. The null hypothesis of no cointegration is $H_0 : \phi = 0$, and the alternative hypothesis is $H_1 : \phi < 0$. Given a consistent estimator $\hat{\alpha}$ for α, the conventional EG cointegration test has been constructed from the estimated Dickey and Fuller (1979, hereafter: DF) equation,

$$\Delta(y_{1,t} - \hat{\alpha}'X_t) = \phi(y_{1,t-1} - \hat{\alpha}'X_{t-1}) + e_t, \tag{2.2}$$

where e_t are a white noise series. If we let $\hat{u}_t = y_{1,t} - \hat{\alpha}'X_t$, then the formula (2.2) can be represented by the following formula

$$\Delta\hat{u}_t = \phi\hat{u}_t + e_t \tag{2.3}$$

and the conventional DF test statistic is

$$t_{DF} = \frac{\hat{\phi}}{\mathrm{se}\left(\hat{\phi}\right)} = \frac{\sum \hat{u}_{t-1}\Delta\hat{u}_t / \sum \hat{u}_{t-1}^2}{\hat{\sigma}\left\{\sum \hat{u}_{t-1}^2\right\}^{-\frac{1}{2}}} = \frac{\sum \hat{u}_{t-1}\Delta\hat{u}_t}{\hat{\sigma}\left\{\sum \hat{u}_{t-1}^2\right\}^{\frac{1}{2}}}, \tag{2.4}$$

where $\hat{\phi} = \sum \hat{u}_{t-1}\Delta\hat{u}_t / \sum \hat{u}_{t-1}^2$, $\mathrm{se}(\hat{\phi}) = \hat{\sigma}(\sum \hat{u}_{t-1}^2)^{-1/2}$ and $\hat{\sigma}^2 = 1/(T-2)\sum(\hat{u}_t - \hat{\phi}\hat{u}_{t-1})^2$. All summations $\sum$ are from 2 to T unless otherwise noted.

We consider the new tests based on sign instrumental variable, $\mathrm{sign}(u_{t-1})$. The sign IV estimator for ϕ is

$$\hat{\phi}_{IV} = \frac{\sum \mathrm{sign}(\hat{u}_{t-1})\Delta\hat{u}_t}{\sum \mathrm{sign}(\hat{u}_{t-1})\hat{u}_{t-1}} = \frac{\sum \mathrm{sign}(\hat{u}_{t-1})\Delta\hat{u}_t}{\sum |\hat{u}_{t-1}|} \tag{2.5}$$

and the standard deviation of $\hat{\phi}_{IV}$ is

$$\mathrm{se}\left(\hat{\phi}_{IV}\right) = \frac{\hat{\sigma}\sqrt{T}}{\sum |\hat{u}_{t-1}|}, \tag{2.6}$$

where $\hat{\sigma} = 1/(T-2)\sum(\hat{u}_t - \hat{\alpha}\hat{u}_{t-1})^2$ and $\hat{\alpha} = \sum \hat{u}_{t-1}\hat{u}_t / \sum \hat{u}_{t-1}^2$. Therefore the new test statistic can be constructed as follows

$$t_{IV}^{*} = \frac{\hat{\phi}_{IV}}{\mathrm{se}\left(\hat{\phi}_{IV}\right)} = \frac{1}{\hat{\sigma}\sqrt{T}} \sum \mathrm{sign}(\hat{u}_{t-1})\Delta\hat{u}_t. \tag{2.7}$$

Since $\Delta\hat{u}_t$ in the formula (2.7) is not a t-measurable, the asymptotic of t_{IV}^{*} can not have the standard normal distribution. In order to reach the t-measurable property, we employ the recursive estimator for α of So and Shin (1999b). They introduce the recursive mean adjustment, show several nice merits, and apply to unit root tests. Also, they note that the recursive estimator not only brings a nice property in asymptotic null distribution for the test statistic but also improves the power performance of a unit root test. We replace $\Delta\hat{u}_t = \Delta(y_{1,t} - \hat{\alpha}'X_t)$ by

$$\Delta\tilde{u}_{t|t-1} \equiv \Delta y_{1,t} - \tilde{\alpha}_{t-1}'\Delta X_t, \tag{2.8}$$

where $\tilde{\alpha}_t = (\mathcal{X}_t'\mathcal{X}_t)^{-1}\mathcal{X}_t'\mathcal{Y}_t$, $\mathcal{X}_t = (X_1, X_2, \ldots, X_t)'$ and $\mathcal{Y}_t = (y_{1,1}, y_{1,2}, \ldots, y_{1,t})'$. The recursive estimator, $\tilde{\alpha}_t$, just uses data up to time t, and this allows the test statistic to become t-measurable. In this context the sign IV cointegration test statistic is

$$t_{IV} = \frac{1}{\tilde{\sigma}\sqrt{T}} \sum_{t=p-1}^{T} \text{sign}(\tilde{u}_{t-1})\Delta\tilde{u}_{t|t-1} \tag{2.9}$$

$$= \frac{1}{\tilde{\sigma}\sqrt{T}} \sum_{t=p-1}^{T} \text{sign}(\tilde{u}_{t-1})\left\{\Delta y_{1,t} - \tilde{\alpha}'_{t-1}\Delta X_t\right\}. \tag{2.10}$$

The proposed test statistic, t_{IV}, has the standard normal limiting distribution that is free from the number of variables.

Theorem 1. *Consider model (2.1) then we have* $t_{IV} \xrightarrow{d} N(0,1)$ *as* $T \to \infty$, *under the null hypothesis* $H_0 : \phi = 0$.

Proof: First we need to show that the $\sum \text{sign}(\tilde{u}_{t-1})\Delta\tilde{u}_{t|t-1}$ is a martingale. Let $z_t = \sum \text{sign}(\tilde{u}_{t-1})\Delta\tilde{u}_{t|t-1}$ and $\mathcal{L}_t$ be the σ-field generated by the data up to time point t. Then we have

$$\begin{aligned}
E(z_{T+1}|\mathcal{L}_T) &= E\left(\sum_{t=p-1}^{T+1} \text{sign}(\tilde{u}_{t-1})\Delta\tilde{u}_{t|t-1}|\mathcal{L}_T\right) \\
&= E\left(\sum_{t=p-1}^{T} \text{sign}(\tilde{u}_{t-1})\Delta\tilde{u}_{t|t-1} + \text{sign}(\tilde{u}_T)\Delta\tilde{u}_{T+1|T}|\mathcal{L}_T\right) \\
&= \sum_{t=p-1}^{T} \text{sign}(\tilde{u}_{t-1})\Delta\tilde{u}_{t|t-1} + \text{sign}(\tilde{u}_T)E\left(\Delta\tilde{u}_{T+1|T}|\mathcal{L}_T\right) \\
&= z_T
\end{aligned}$$

The last equation comes from $E(\Delta\tilde{u}_{T+1|T}|\mathcal{L}_T) = E(\Delta y_{1,T+1} - \tilde{\alpha}'_T\Delta X_{T+1}|\mathcal{L}_T) = E(\Delta y_{1,T+1}|\mathcal{L}_T) + \tilde{\alpha}'_T E(\Delta X_{T+1}|\mathcal{L}_T) = 0$. The $E(\Delta y_{1,T+1}|\mathcal{L}_T)$ and $E(\Delta X_{T+1}|\mathcal{L}_T)$ are 0 because Δy_{T+1} and ΔX_{T+1} are *i.i.d.* with mean 0. Since we have $E(z_{T+1}|\mathcal{L}_T) = z_T$, the z_t is a martingale. Now the martingale central limit theorem can be applied to get the limiting distribution. Therefore we complete the proof to show $t_{IV} \xrightarrow{d} N(0,1)$. Q.E.D. □

3. Simulation

We generate the $\{y_t\}$ from the following model as in Kremers *et al.* (1992)

$$\begin{aligned}
\Delta y_{1t} &= \sum_{i=2}^{p} \alpha_i \Delta y_{i,t} + \phi(y_1 - y_2 - \cdots - y_p)_{t-1} + \epsilon_t, \quad t = 1, \ldots, T, \\
\Delta y_{it} &= \xi_{it}, \quad i = 2, \ldots, p,
\end{aligned}$$

where X_t is a vector with p exogenous variables, $(\epsilon_t, \xi_{2,t}, \ldots, \xi_{p,t})' \sim N_p(0_p, \Sigma)$ where the $0_p = (0, \ldots, 0)'$ is a $p \times 1$ vector and $\Sigma = \text{Diag}(\sigma_\epsilon^2, \sigma_2^2, \ldots, \sigma_p^2)$ is a $p \times p$ diagonal matrix. We are interested in testing the hypotheses that $H_0 : \phi = 0$ versus $H_1 : \phi < 0$. If we reject the null hypothesis, then the data has strong evidence of cointegration.

Table 1: Rejection Frequencies(%), $\phi = 0.0$

		(α, s)					
		(1.0, 1)		(0.5, 6)		(0.5, 16)	
T	p	Sign IV	EG	Sign IV	EG	Sign IV	EG
A. Rejection frequency at the 5 percent critical value							
	1	4.9	5.3	5.2	5.5	5.1	5.2
20	2	5.0	5.2	5.3	5.4	5.0	5.1
	3	4.8	5.3	5.2	5.3	4.9	4.9
	1	5.1	5.2	5.1	5.4	5.0	5.1
100	2	5.2	5.0	4.9	5.6	5.1	5.2
	3	5.1	5.1	4.9	5.3	5.2	5.1
B. Rejection frequency at the 1 percent critical value							
	1	1.2	1.1	1.1	1.3	1.2	1.2
20	2	1.1	1.1	1.2	1.2	0.9	1.3
	3	1.0	1.0	1.1	1.3	1.0	1.1
	1	1.0	1.0	1.3	1.1	1.2	1.1
100	2	0.9	1.1	1.1	1.2	1.3	1.2
	3	1.0	1.1	1.1	1.2	1.1	1.3

Note: Tests for $H_0 : \phi = 0$ for model $\Delta(y_{1t} - \hat{\alpha}'X_t) = \phi(y_{1,t-1} - \hat{\alpha}'X_{t-1}) + e_t$; $\alpha = (\alpha_2, ..., \alpha_p)'$ and $X_t = (y_{2t}, ..., y_{pt})'$; Number of replications = 10,000.

Table 2: Rejection Frequencies(%), $\phi = -0.1$

		(α, s)					
		(1.0, 1)		(0.5, 6)		(0.5, 16)	
T	p	Sign IV	EG	Sign IV	EG	Sign IV	EG
A. Rejection frequency at the 5 percent critical value							
	1	12.01	11.55	17.55	12.31	14.32	11.98
20	2	10.77	9.72	14.67	10.65	12.34	9.61
	3	8.90	7.64	10.22	8.33	10.57	7.91
	1	39.10	37.00	38.96	38.51	36.54	37.52
100	2	25.11	23.70	29.51	25.15	27.89	25.10
	3	18.95	17.85	12.50	19.83	15.91	18.77
B. Rejection frequency at the 1 percent critical value							
	1	3.71	2.44	3.62	2.78	3.31	2.58
20	2	3.10	2.21	3.38	2.99	3.13	2.37
	3	2.87	2.07	2.71	2.13	2.93	2.84
	1	9.16	8.55	9.83	9.13	9.02	8.22
100	2	9.01	8.32	9.52	9.17	8.90	8.50
	3	8.53	7.88	9.11	8.16	8.19	8.10

Note: Tests for $H_0 : \phi = 0$ for model $\Delta(y_{1t} - \hat{\alpha}'X_t) = \phi(y_{1,t-1} - \hat{\alpha}'X_{t-1}) + e_t$; $\alpha = (\alpha_2, ..., \alpha_p)'$ and $X_t = (y_{2t}, ..., y_{pt})'$; Number of replications = 10,000.

Although the σ_is can have different values, we let σ_i have identical value and let σ_i represent the variances. And also we let the coefficients α_i are all identical and let α_i represent the coefficients. And therefore, we have six parameters to set which are ϕ, α_i, σ_ϵ, σ_i, p and T. Without loss of generality, $\sigma_\epsilon = 1$ and let $s = \sigma_i/\sigma_\epsilon$ that s=σ_i in this case. Kremers *et al.* (1992) note that in many empirical studies, $\alpha \approx 0.5$ and $\phi \approx -0.1$, with $\sigma_i^2 > \sigma_\epsilon^2$ for $i = 2, \ldots, p$. In considering their findings, we set $\phi = 0.0$ for the size and $\phi = -0.1$ for the power. We also set $(\alpha_i, s) = [(1.0, 1), (0.5, 6), (0.5, 16)]$. We experiment with cases of $T = 20$ and 100. We set $p = 2$, 3 and 4 to examine whether the tests are sensitive to the number of variables. Let the initial values of each y_{it} be fixed. The number of replications per experiment is $N = 10{,}000$, the first 20 observations of each replication is discarded in order to attenuate the effect of initial values, and the $y_2, \ldots, y_p$'s are generated for each replication.

Table 1 and 2 show rejection probabilities(%) under the null and the alternative hypothesis, respectively.

From the table 1, we can see that the sizes of both tests are stable under all cases. From the table 2, the less the number of variables, the more the power for both tests. Our tests are locally more powerful than the EG test except some cases with $T = 100$ under the 5% significance level. When the time span is only 20, the proposed tests show better property in powers. This is a superior aspect of our test because data having small numbers of observations is more likely to assess in practice.

4. Conclusion

We propose sign IV cointegration tests. Our tests are based on the instrumental variables, signs of the regressors. The limiting null distributions of the test statistics are the standard normal distribution and are free from the number of variables. According to the Monte-Carlo simulation results, the proposed tests are locally more powerful than those based on the usual t-type tests especially for the data with the short time span such as $T = 20$.

References

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

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

Kremers, J. J. M., Ericsson, N. R. and Dolado, J. J. (1992). The power of cointegration tests, *Oxford Bulletin of Economics and Statistics*, **54**, 325–348.

So, B. S. and Shin, D. W. (1999a). Cauchy Estimators for autoregressive processes with applications to unit root tests and confidence intervals, *Econometric Theory*, **15**, 165–176.

So, B. S. and Shin, D. W. (1999b). Recursive mean adjustment in time series inferences, *Statistics and Probability Letters*, **43**, 65–73.

Received May 2009; Accepted June 2009