Economics Letters 151 (2017) 71–74

Contents lists available at ScienceDirect

Economics Letters

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

A time series paradox: Unit root tests perform poorly when data are cointegrated

W. Robert Reed [a], Aaron Smith [b,*]

[a] Department of Economics and Finance, University of Canterbury, New Zealand
[b] Department of Agricultural and Resource Economics, University of California, Davis, United States

HIGHLIGHTS

- Presence of a unit root in individual time series is required for cointegration.
- Cointegration makes standard unit root tests more likely to reject a unit root null.
- The paradox arises because cointegration induces a moving average (MA) component.
- The cointegration-induced MA component causes unit root tests to be oversized.

ARTICLE INFO

Article history:
Received 8 September 2016
Received in revised form
30 November 2016
Accepted 3 December 2016
Available online 13 December 2016

JEL classification:
C32
C22
C18

Keywords:
Unit root testing
Cointegration
Augmented Dickey–Fuller test
Akaike Information Criterion (AIC)
Bayesian Information Criterion (BIC)
Modified Akaike Information Criterion (MAIC)

ABSTRACT

Cointegration among times series paradoxically makes it more likely that a unit test will reject the unit root null hypothesis on the individual series. This occurs because at least one series in the system has a negative moving average component.

1. Introduction

Standard practice when estimating relationships among time series variables is to first test the individual series for nonstationarity. If the individual series are concluded to have unit roots, one then tests for cointegration. This advice was first dispensed in Engle and Granger's (1987) seminal cointegration paper and has been repeated numerous times since.[1] This paper demonstrates that this practice is paradoxical. When data are cointegrated, unit root tests are more likely to reject the unit root hypothesis.

The paradox arises because cointegration generates a moving average (MA) component in the univariate representation of a time series. It is well known that two variables that are cointegrated can be rewritten as a linear combination of two series, one of which has a unit root and the other of which is stationary. Granger and Morris (1976) showed that such linear combinations typically have a moving average component even if the individual series have no MA structure. It is also well known that MA dependence causes over-rejection in unit root tests (Ng and Perron, 2001). The time series literature has not previously connected these results to unit root testing when data are cointegrated. That is the contribution of this paper.

* Corresponding author.
E-mail addresses: bob.reed@canterbury.ac.nz (W.R. Reed), adsmith@ucdavis.edu (A. Smith).

[1] For example, see Kennedy (2008, p. 303).

http://dx.doi.org/10.1016/j.econlet.2016.12.005

The most common unit root tests are of the augmented Dickey–Fuller (ADF) type. Elliott et al. (1996) show that, if generalized least squares is used to detrend the time series, then the ADF test has desirable asymptotic power properties.[2] This result gives rise to the DF-GLS test. In this paper, we focus on lag selection so we exclude deterministic components from the model. Elliott et al. (1996) write that the asymptotic power of this test "virtually equals the (upper) bound when power is one-half and is never far below".

Said and Dickey (1984) demonstrate that ADF tests have correct size if enough lags are included in the ADF specification. Several information criteria (IC) have been proposed to select the appropriate number of lags, including the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Ng and Perron (2001) demonstrate that AIC and BIC select too few lags, and they propose the Modified Akaike Information Criterion (MAIC). An alternative approach is to determine lag length using hypothesis testing, e.g., use only those lags with statistically significant coefficients. Our paper demonstrates that neither hypothesis testing nor any of the three IC are sufficient to eliminate size distortions when data are cointegrated.

2. Example: a triangular model

The triangular model is the prototypical example model in the time series literature (e.g., Phillips, 1991). Consider a pair of cointegrated random variables, y_t and x_t, defined such that

$$\begin{aligned} y_t &= x_t + z_t \\ x_t &= x_{t-1} + \varepsilon_{1t} \\ z_t &= \phi z_{t-1} + \varepsilon_{2t} \end{aligned} \tag{1}$$

where $|\phi| < 1$, $\varepsilon_{1t} \sim$ i.i.d. $N(0, 1)$, $\varepsilon_{2t} \sim$ i.i.d. $N\left(0, \sigma^2\left(1-\phi^2\right)\right)$, and $E[\varepsilon_{1s}\varepsilon_{2t}] = 0$ for all t, s. We impose a coefficient of one on x_t and a variance of one on ε_{1t} because the data can always be rescaled to achieve these restrictions. We parameterize the variance of ε_{2t} so that $\mathrm{var}[z_t] = \sigma^2$ regardless of ϕ. Note that y_t and x_t are not cointegrated when $\phi = 1$. The farther ϕ is from one, the less persistent are deviations from the cointegrating relationship and therefore the stronger is the cointegration.

Granger and Morris (1976) prove the following result for finite-order autoregressive moving average (ARMA) processes. If $x_t \sim ARMA(p_1, q_1)$ and $z_t \sim ARMA(p_2, q_2)$, then $(x_t + z_t) \sim ARMA(m, n)$, where $m \le p_1 + p_2$ and $n \le \max(p_1 + q_2, p_2 + q_1)$. The expressions for m and n generally hold as equalities. Only in special cases do some terms cancel out to make it an inequality. This occurs, for example, if the ARMA coefficients are the same in the two summands. Thus, if z_t has positive variance, then the Granger and Morris result implies that y_t has a moving average component in general.

The univariate ARMA representation of y_t is[3]

$$\Delta y_t = \phi \Delta y_{t-1} + v_t - \theta v_{t-1} \tag{2}$$

where $\theta = \frac{1-\sqrt{1-4\omega^2}}{2\omega} > \phi$, $\omega = \frac{\phi+\sigma^2(1-\phi^2)}{1+\phi^2+2\sigma^2(1-\phi^2)}$, and v_t is a white noise process. Augmented Dickey–Fuller tests rely on the autoregressive representation of the time series, which is

$$\begin{aligned} \Delta y_t = {} & (\phi-\theta)\Delta y_{t-1} + \theta(\phi-\theta)\Delta y_{t-2} + \theta^2(\phi-\theta)\Delta y_{t-3} \\ & + \cdots + v_t. \end{aligned} \tag{3}$$

The number of autoregressive lags required to fit this process will be large if θ is large. Because θ is increasing in ϕ, a large θ implies that ϕ is close to 1.

Fig. 1 shows the size of ADF tests for a unit root for various values of the parameters ϕ and σ and for two different sample sizes. The left column reports rejection rates for parameter settings corresponding to those in Figure A1 in the Supplementary Appendix ($T = 100$). The right column shows how rejection rates change as the sample size increases ($T = 500$). The four lines in each figure represent the rejection rates that result when the number of lags is chosen by AIC, BIC, MAIC, or hypothesis testing.

There is a clear rank order to the different IC: MAIC is better than AIC, and AIC is better than BIC. The t-test procedure performs similarly to AIC for $T = 500$ and slightly better than AIC for $T = 100$. However, the ADF test is oversized in all four cases. The worst size distortion occurs for mid to high values of ϕ, i.e., when the cointegration is relatively weak.[4] Recall from (2) that θ is increasing in ϕ and from (3) that the number of autoregressive lags required to fit this process will be large if θ is large. These facts indicate that all lag selection methods choose too few lags to control for the MA term.[5]

There is no size distortion when there is no cointegration ($\phi = 1$), which demonstrates that cointegration is the source of the problem. Cointegration causes the size distortion.

3. The general case

Consider the $n \times 1$ vector X_t, which follows the cointegrated vector autoregression

$$\Phi(L)X_t = \varepsilon_t, \quad \varepsilon_t = [\varepsilon_{1t}, \ldots, \varepsilon_{nt}]' \tag{4}$$

where $E[\varepsilon_t] = 0$, $E[\varepsilon_t\varepsilon_t'] = \Omega$, and $E[\varepsilon_{is}\varepsilon_{jt}] = 0$ for each $i, j = 1, 2, \ldots, n$ and all $s \neq t$.

Cointegration implies $|\Phi(L)| = (1-L)^d \prod_{j=1}^{n-d}(1-\phi_j L)$ for some $0 < d < n$, where $|\phi_j| < 1$ for all j (Lütkepohl, 2005, pg. 243). Defining $\Phi(L)^+$ as the adjoint of $\Phi(L)$, we can write $|\Phi(L)|\Phi(L)^{-1} = \Phi(L)^+$, which implies

$$\prod_{j=1}^{n-d}(1-\phi_j L)\Delta X_t = C(L)\varepsilon_t \tag{5}$$

where $C(L) \equiv (1-L)^{1-d}\Phi(L)^+$. This representation implies that each series is an ARMA process because both $\prod_{j=1}^{n-d}(1-\phi_j L)$ and $C(L)$ are finite polynomials. To eliminate the MA component, we need the scalar polynomial on the left hand side to cancel out the autocorrelations implied by each right hand side polynomial. It is immediately apparent that only in special cases will the scalar factors in $\prod_{j=1}^{n-d}(1-\phi_j L)$ be able to cancel the autocorrelations in each row of $C(L)\varepsilon_t$.

We illustrate with a two variable system that has a single lag. This case preserves clarity, but extends readily to cases with more lags or variables. We write the model in error correction form as

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} -\alpha_1 \\ \alpha_2 \end{bmatrix} (x_{1,t-1} - x_{2,t-1}) + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{6}$$

---

2 They write in their conclusion "these tests are essentially point optimal among tests based on second-order sample moments".

3 Note that $(1-\phi L)\Delta y_t = u_t$, where $u_t = (1-\phi L)\varepsilon_{1t} + (1-L)\varepsilon_{2t}$. Now u_t is an MA (1) process with $E[u_t^2] = (1+\phi^2) + 2\sigma^2(1-\phi^2)$ and $E[u_t u_{t-1}] = -\phi - \sigma^2(1-\phi^2)$. The expression for θ can then be derived from method of moments.

4 The expected R^2 from a regression of y_t on x_t is another indicator of the strength of cointegration. For $\phi = 0.8$, the expected R^2 values are (a) 0.92, (b) 0.98, (c) 0.39, (d) 0.69, (e) 0.18, (f) 0.40. To calculate the expected R^2, we averaged the R^2 values obtained from 100,000 random samples. The expected R^2 in cointegrating regressions varies somewhat with ϕ. Figure A1 in the online appendix plots sample time series.

5 We repeated these simulations using the M_α statistic of Ng and Perron (2001), which can reduce size distortion in the presence of MA components. The results were very similar to Fig. 1. We did not include the Phillips–Perron test which is known to perform very poorly in these settings.