DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

A Time Series Paradox:
Unit Root Tests Perform Poorly When Data Are Cointegrated

W. Robert Reed
Aaron Smith

WORKING PAPER

No. 19/2016

Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 19/2016

A Time Series Paradox:
Unit Root Tests Perform Poorly When Data Are Cointegrated

W. Robert Reed[1]
Aaron Smith[2†]

6 September 2016

Abstract: We show that cointegration among times series paradoxically makes it more likely that a unit test will reject the unit root null hypothesis on the individual series. If one time series is cointegrated with another, then it can be written as the sum of two processes, one with a unit root and one stationary. It follows that the series cannot be represented as a finite-order autoregressive process. Unit root tests use an autoregressive model to account for autocorrelation, so they perform poorly in this setting, even if standard methods are used to choose the number of lags. This finding implies that univariate unit root tests are of questionable use in cointegration analysis.

Keywords: Unit root testing, cointegration, Augmented Dickey-Fuller test, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Modified Akaike Information Criterion (MAIC)

JEL Classifications: C32, C22, C18

Acknowledgements: Earlier versions of this paper received valuable comments from David Giles; Andrew Harvey; Katarina Juselius, Morten Nielsen; Pierre Perron; Peter Phillips; Marco Reale; seminar participants at the "1st Conference on Recent Developments in Financial Econometrics and Applications," Deakin University, 4-5 December, 2014, and the "Advances in Empirical and Theoretical Econometrics" workshop, University of Queensland, 25-26 November, 2015. The authors are wholly responsible for any remaining errors.

[1] Department of Economics and Finance, University of Canterbury, NEW ZEALAND

[2] Department of Agricultural and Resource Economics, University of California, Davis, USA

† Corresponding author is Aaron Smith. Email: adsmith@ucdavis.edu ,

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 unreliable.

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.

The most common unit root tests are of the augmented Dickey-Fuller (ADF) type. Elliott, Rothenberg and Stock (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. Thus, we analyse the standard ADF test without a constant or trend. Elliott, Rothenberg and Stock (1996) write that the asymptotic power of this test "virtually equals the (upper) bound when power is one-half and is never far below."

---

[1] For example, see Kennedy (2008, p. 303).
[2] They write in their conclusion "these tests are essentially point optimal among tests based on second-order sample moments".

Said and Dickey (1984) demonstrate that ADF testing will still be valid if enough lags are included in the ADF specification. Towards this end, 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). However, Ng and Perron (2001) demonstrated that these criteria select too few lags. Instead, they propose the Modified Akaike Information Criterion (MAIC) and argue that it has good size and power properties. 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. We show rejection rates that substantially exceed their nominal significance levels under realistic conditions.

The paper proceeds as follows. In Section 2, we illustrate the size distortion problem using a two variable model. In Section 3, we analyse a general formulation of the problem. In Section 4, we conclude and offer recommendations.

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 $|\phi|<1$, $\varepsilon_{1t} \sim iidN(0,1)$, $\varepsilon_{2t} \sim iidN\left(0, \sigma^2(1-\phi^2)\right)$, 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 $\text{var}[z_t]=\sigma^2$ regardless of ϕ.

Note that y_t and x_t are not cointegrated when $\phi = 1$. The closer ϕ is to zero, 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 $m \leq p_1 + p_2$ and $n \leq max(p_1 + q_2,\ p_2 + q_1)$. 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 $\theta = \dfrac{1 - \sqrt{1 - 4\omega^2}}{2\omega} > \lambda$, $\omega = \dfrac{\phi + \sigma^2(1-\phi^2)}{1 + \phi^2 + 2\sigma^2(1-\phi^2)}$, and v_t is a white noise process. Augmented Dickey-Fuller tests rely on the autoregressive representation of the time series, which is

$$\Delta y_t = (\phi - \theta)\Delta y_{t-1} + \theta(\phi - \theta)\Delta y_{t-2} + \theta^2(\phi - \theta)\Delta y_{t-2} + \ldots + v_t\,. \tag{3}$$

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 ϕ implies that ϕ is close to 1.

Figure 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 1 (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

[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.

rates that result when the number of lags is chosen by three IC (AIC, BIC, MAIC) or by hypothesis testing. Our hypothesis testing procedure starts with 10 lags and repeatedly drops the last lag until it is statistically significant (details in figure notes).

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.

There is no size distortion when there is no cointegration (ϕ=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

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

where $E[\varepsilon_t]=0$, $E[\varepsilon_t\varepsilon_t']=\Omega$, and $E[\varepsilon_{is}\varepsilon_{jt}]=0$ for each i,j=1,2,…, 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

[4] The expected R^2 from a regression of y_t on x_t is another indicator of the strength of cointegration. For ϕ=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 vales 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.

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

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 each one extends readily to cases with more lags or variables. We write the model in error correction form as

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

where $\text{var}[\varepsilon_{1t}]=1$, $\text{var}[\varepsilon_{2t}]=\sigma^2$, and $E[\varepsilon_{1t}\varepsilon_{2t}]=\rho\sigma$. As in the previous section, we set the variance of ε_{1t} and the cointegration parameter to one without loss of generality. For each variable j=1,2, the univariate representation is $\Delta x_{jt}=\phi\Delta x_{j,t-1}+v_{jt}-\theta_j v_{j,t-1}$, where $\phi\equiv 1-\alpha_1-\alpha_2$ and v_{jt} are white noise processes. We derive analytic expressions for θ_j in the online appendix. To cancel the MA component from both equations, we need $\theta_1=\theta_2=\phi$. There are only a few special cases for which this holds.

The relationship between $\{\theta_1, \theta_2\}$ and ϕ is multi-dimensional; it depends on all four parameters (α_1, α_2, ρ, and σ), so it cannot be displayed on a single graph. Panels (a)-(d) of Figure 2 plot the implied values of $\theta_1-\phi$ and $\theta_2-\phi$ against ϕ for 4 parameter combinations. For both MA terms to cancel, and therefore have no size distortion in unit root tests, we require $\theta_1-\phi$ and $\theta_2-\phi$ to equal zero at the same point. A large value of ϕ implies that neither variable responds strongly to deviations from the cointegrating relationship and therefore that the cointegrating relationship is weak. When $\phi = 1$ there is no cointegration, and when $\phi = 0$, the variables correct any cointegrating errors in one period.

Panel (a) of Figure 2 sets α_1=0, which implies that x_{1t} does not respond to deviations from the cointegrating relationship. In this case, x_{1t} is a simple random walk, so it has no MA component; we have $\theta_1 = \phi$ for all values. This is a generalized version of the model in (1) that allows nonzero correlation in the errors. However, we see $\theta_2 = \phi$ only in the limit as cointegration disappears. Otherwise x_{2t} always has an MA component and unit root tests on this variable will be oversized. Again, we see how cointegration causes size distortion. Panels (b)-(d) of Figure 2 allows x_{1t} to also respond to deviations from the cointegrating relationship. In no cases do the two lines cross zero at the same point.

Are there any combinations of parameters for which $\theta_1 = \theta_2 = \phi$? To answer this question, we note that for every combination of α_1, α_2, and ρ, it is possible to find a value of σ such that θ_1= θ_2. Call the resulting MA parameter $\theta(\sigma)$. Panels (e) and (f) of Figure 2 show $\theta(\sigma) - \phi$ as a function of ϕ for various values of α_1 and ρ.

There are only two settings in which $\theta(\sigma) - \phi = 0$. First, if ρ=1, there always exists a σ such that there is no size distortion. This case implies that the two variables are linear combinations of each other because the errors have correlation equal to one. This is not an empirically interesting case – no-one would estimate a cointegrating relationship between a price measured in dollars and the same price measured in cents. The second case is when α_2=0, i.e., when x_{2t} does not respond to deviations from the cointegrating relationship. In this case, it is possible to find a value of σ

such that there would be no size distortion. If σ were to take any other value, the size distortion would be present.

4. CONCLUSION AND RECOMMENDATIONS

Our analysis demonstrates that unit root tests are unreliable in the presence of cointegration. In general, when series are cointegrated, an MA term is introduced into the univariate time series representations. Although it is theoretically possible to eliminate the associated size distortion by including sufficient lags in the ADF test specification, conventional information criteria will select too few lags. The resulting size distortions can be quite substantial for realistic cases.

Our analysis leads us to two recommendations. First, we suggest that researchers dispense with univariate unit root pre-tests. If a researcher aims to fit a model of multiple time series in the presence of potential cointegration, there is no need to pre-test the data. One should directly test for cointegration. Second, if a researcher feels compelled to conduct unit root pre-tests, and if there is reason to believe that the data are cointegrated, we recommend using Ng and Perron's (2001) Modified Akaike Information Criterion. Although the associated test may still be oversized, possibly substantially so, it will perform better than the familiar AIC, BIC, and hypothesis testing methods.

REFERENCES

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996) "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4): 813-836.

Engle, R.F. and C.W.J. Granger (1987) "Co-integration and Error Corection: Representation, Estimation, and Testing," *Econometrica,* 55(2): 251-276.

Granger, C.W.J., and M.J. Morris (1976) "Time Series Modelling and Interpretation," *Journal of the Royal Statistical Society. Series A (General),* 139(2): 246-257.

Kennedy, P. (2008), *A Guide to Econometrics,* Wiley.

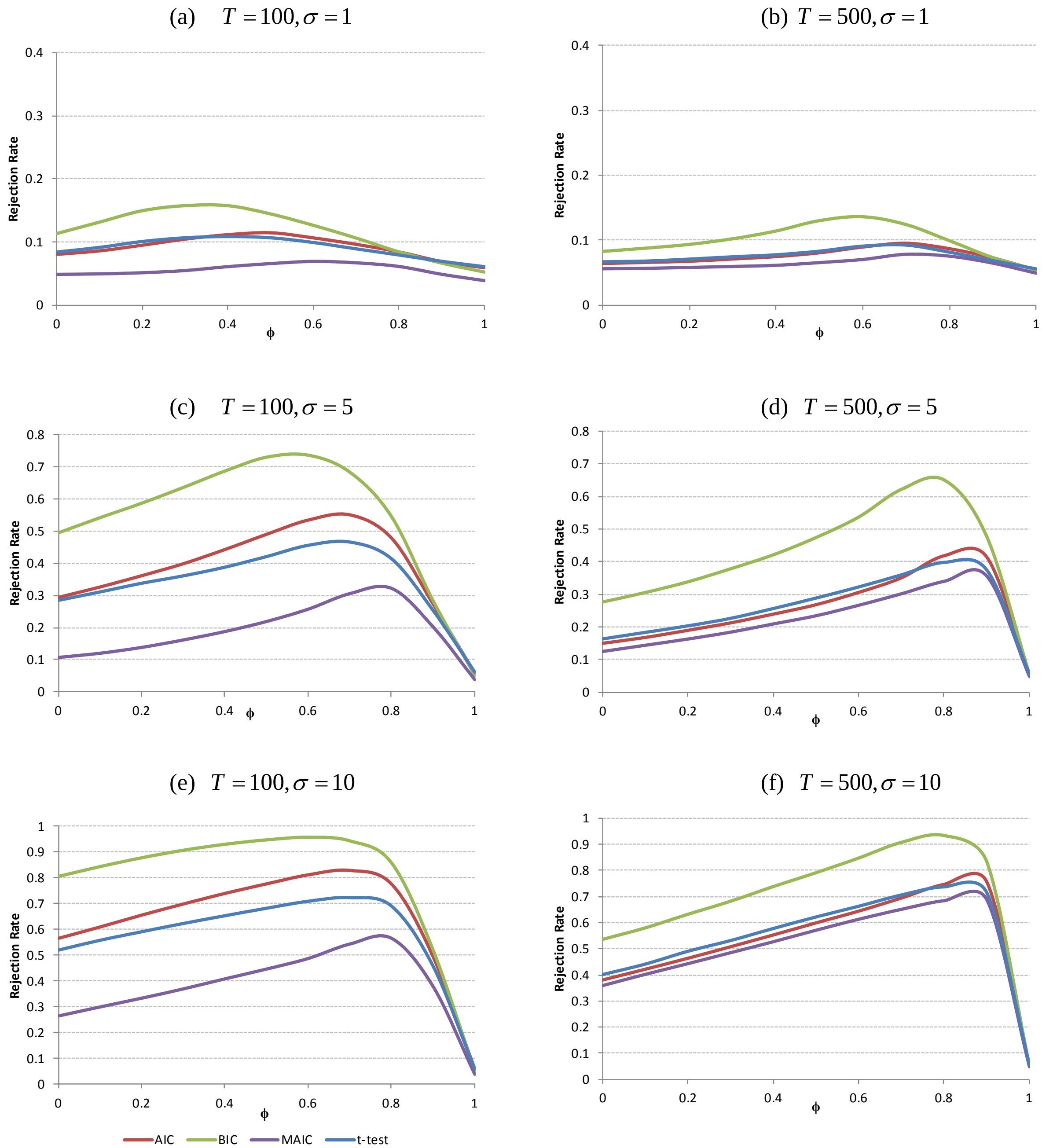
Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis,* Springer-Verlag, Berlin.

Ng, S. and P. Perron (2001) "Lag length selection and the construction of unit root tests with good size and power," *Econometrica,* 69(6): 1519-1554.

Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica,* 59:283-306.

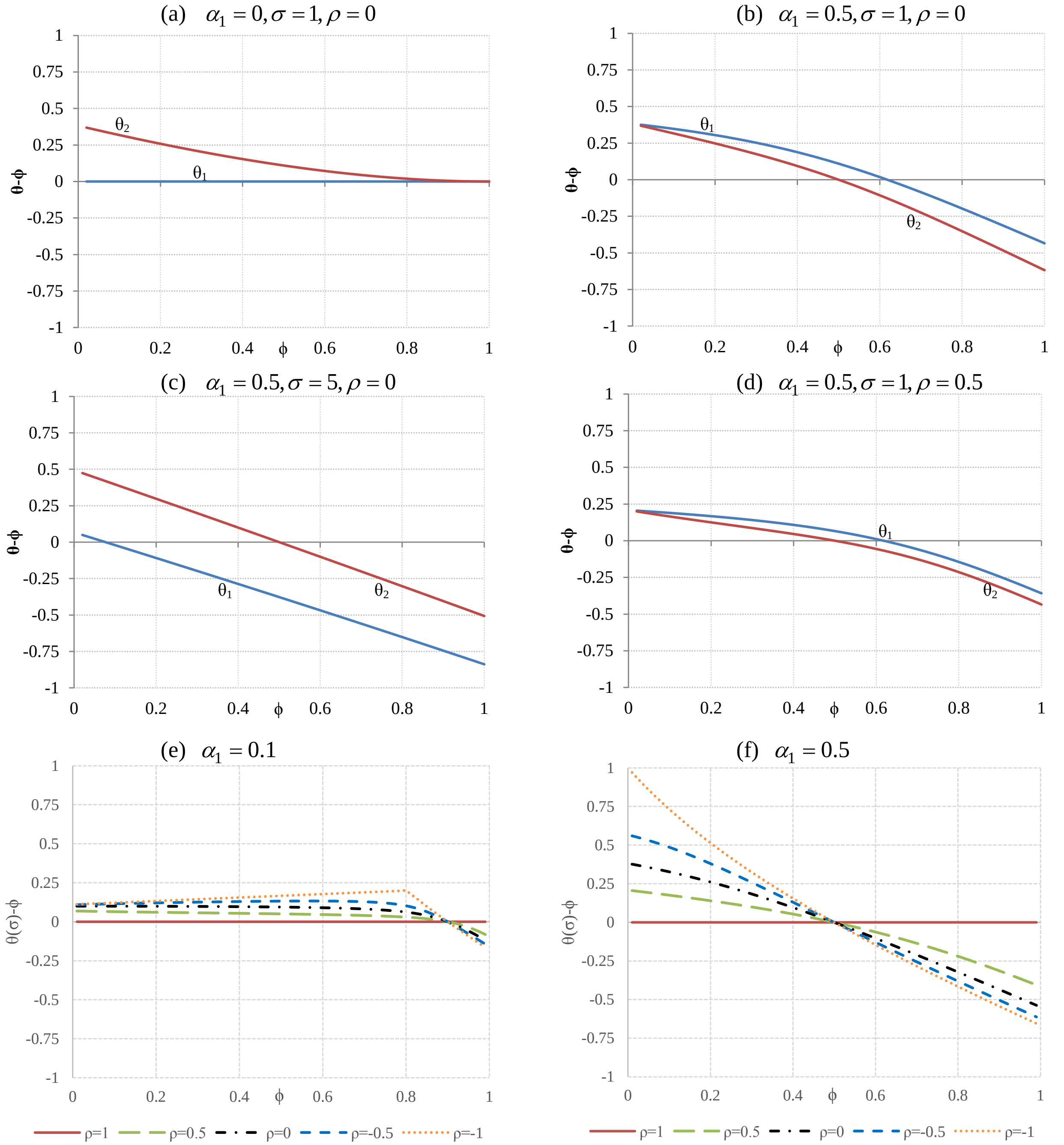
Said, S. E. and D. A. Dickey (1984). "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.

Figure 1: Rejection Rate of ADF Test for Triangular Model



Notes: Generated from Monte Carlo simulations of the process in (1) with 10,000 replications. The maximum number of lags is 10. The *t*-test procedure starts with the largest number of lags. If the *t*-statistic on the last lag coefficient is less than 1.96 in absolute value, then re-estimate with one fewer lag. Repeat until the last lag coefficient is significant.

Figure 2: Difference Between MA and AR Parameters



Notes: In Panels (a)-(d), the blue line applies to the first equation (θ_1) and the red line applies to the second equation (θ_2). The MA terms in both equations cancel only if the red and blue lines cross zero at the same point. Panels (a)-(d) generated analytically. For Panels (e) and (f), we first solved for the σ that sets $\theta_1 = \theta_2$. Call this θ(σ). We then plot θ(σ)-ϕ against ϕ for various ρ and two different α_1. The MA terms cancel only when the lines cross zero.

ONLINE APPENDIX – NOT FOR PUBLICATION

Derivations for the General Case in Section 3.

We write the model in error correction form as

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} -\alpha_1 \\ \alpha_2 \end{bmatrix} (x_{1,t-1} - x_{2,t-1}) + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{A1}$$

where $\text{var}[\varepsilon_{1t}] = 1$, $\text{var}[\varepsilon_{2t}] = \sigma^2$, and $E[\varepsilon_{1t}\varepsilon_{2t}] = \rho\sigma$. As in the previous section, we set the variance of ε_{1t} and the cointegration parameter to one without loss of generality. The determinant of the lag polynomial matrix is

$$\begin{vmatrix} 1-(1-\alpha_1)L & -\alpha_1 L \\ -\alpha_2 L & 1-(1-\alpha_2)L \end{vmatrix} = \big(1-(1-\alpha_1-\alpha_2)L\big)(1-L). \tag{A2}$$

From above, cointegration implies $\phi = 1-\alpha_1-\alpha_2 < 1$. The matrix $C(L)$ is

$$C(L) = \begin{bmatrix} 1-(1-\alpha_2)L & \alpha_1 L \\ \alpha_2 L & 1-(1-\alpha_1)L \end{bmatrix}, \tag{A3}$$

which implies that we can write the model in (A1) as

$$\begin{bmatrix} \big(1-(1-\alpha_1-\alpha_2)L\big)\Delta x_{1t} \\ \big(1-(1-\alpha_1-\alpha_2)L\big)\Delta x_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} - (1-\alpha_2)\varepsilon_{1,t-1} + \alpha_1\varepsilon_{2,t-1} \\ \varepsilon_{2t} - (1-\alpha_1)\varepsilon_{2,t-1} + \alpha_2\varepsilon_{1,t-1} \end{bmatrix}. \tag{A4}$$

Thus, the two error components on the right hand side need an MA coefficient of $\alpha_1+\alpha_2-1$ to match the left hand side and thereby cancel out the MA component.

Define the error in the first equation of (A4) as $u_{1t} \equiv \varepsilon_{1t} - (1-\alpha_2)\varepsilon_{1,t-1} + \alpha_1\varepsilon_{2,t-1}$. The MA representation of u_{1t} is

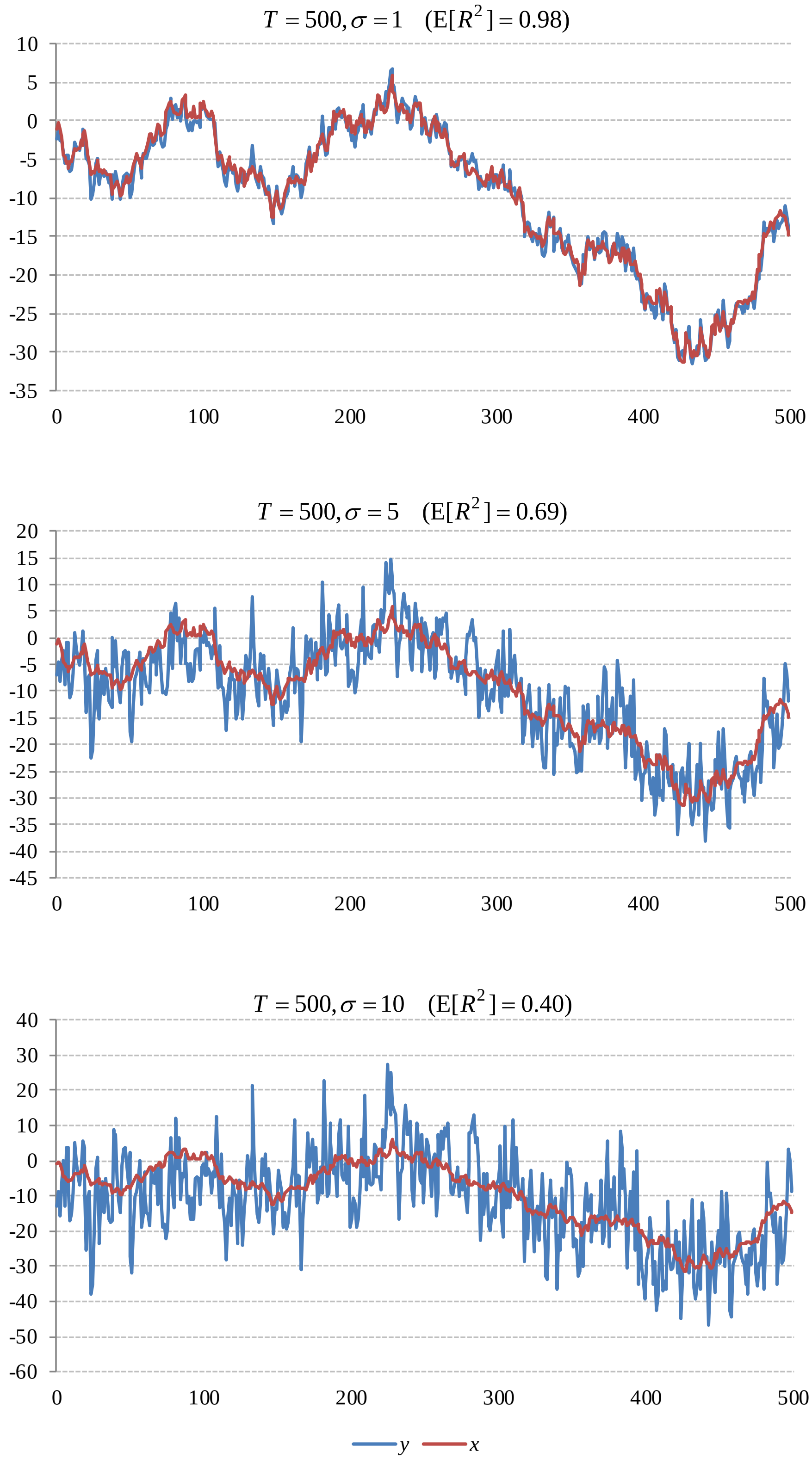
$$u_{1t} = v_{1t} - \theta_1 v_{1,t-1}, \tag{A5}$$

where $\theta_1 = \frac{1-\sqrt{1-4\omega_1^2}}{2\omega_1}$, $\omega_1 = \frac{1-\alpha_1\rho\sigma-\alpha_2}{1+(1-\alpha_2)^2+\alpha_1^2\sigma^2-2\alpha_1(1-\alpha_2)\rho\sigma}$, and v_{1t} is a white noise process (see footnote 3in the main text). Similarly, define the error in the second equation of (A4) as $u_{2t} \equiv \varepsilon_{2t} - (1-\alpha_1)\varepsilon_{2,t-1} + \alpha_2\varepsilon_{1,t-1}$ and write

$$u_{2t} = v_{2t} - \theta_2 v_{2,t-1}, \qquad \text{(A6)}$$

where $\theta_2 = \frac{1-\sqrt{1-4\omega_2^2}}{2\omega_2}$ and $\omega_2 = \frac{(1-\alpha_1)\sigma^2-\alpha_2\rho\sigma}{\sigma^2+(1-\alpha_1)^2\sigma^2+\alpha_2^2-2\alpha_2(1-\alpha_1)\rho\sigma}$, and v_{2t} is a white noise process. To cancel the MA component from both equations, we need $\theta_1 = \phi$ and $\theta_2 = \phi$, where we define $\phi = 1-\alpha_1-\alpha_2$ as above. There are only a few special cases for which this holds, as shown in Figure 2.

Figure A1: Sample Time Series for Various σ



Notes: Sample data generated from (1) with ϕ=0.5. We use the same realization of the shocks for each panel to enable easy comparison. These are examples of the time series in Figure 1 of the manuscript.