

RESPONSE TO REFEREE REPORT NUMBER 2

Thank you for the comments. They are excellent. I greatly appreciate the effort the reviewer put into them. It will result in a much improved revision.

- 1) **COMMENT:** *“The paper shows by simulation that if data for $(x_t; y_t)$; $t = 1, \dots, T$ are generated by a cointegrated VAR in two dimensions, then various univariate tests for unit roots of x_t and y_t have rejection probabilities which deviate seriously from the nominal levels. This is of course a nice observation and should be followed up by an analysis of the two dimensional VAR to find out whether it helps to analyse the variables jointly.”*

RESPONSE: Thank you. I was surprised by the results -- as are most researchers when I share this finding with them. It is apparent from the applied literature that this problem is not widely known. I will address the last comment about finding a solution to the problem below.

- 2) **COMMENT:** *“As a check on the calculations it would also be interesting to see whether you get the correct rejection probability when you simulate a very long series. This would also be a safeguard against the possibility of programming errors.”*

RESPONSE: I followed the reviewer’s advice and simulated the DGP used in TABLE 1, except that the samples now had 1000 observations rather than 100 (and I increased the number of lagged dependent variables from 4 to 10 to accommodate serial correlation). The results are presented below. The ADF and DF-GLS Type I error rates are close to their nominal sizes, though the Phillips-Perron continues to perform poorly.

<i>TEST</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
<i>ADF</i>	0.051	0.088	0.050
<i>Phillips-Perron</i>	0.178	0.997	0.0657
<i>DF-GLS</i>	0.031	0.035	0.025
<i>BREUSCH-GODFREY TEST:</i>	0.057	0.068	0.070

As the possibility of programming errors is a serious issue, I have also gone much further than the reviewer asked to address his/her concerns. I am making available all the programs that I used to simulate the data and check whether the data are cointegrated. The respective programs are briefly described below. The Appendix to this response provides a much longer explanation of how one can use the programs to confirm that the simulations are done correctly:

- The Excel spreadsheet “Parameter Values”. This allows one to calculate β and γ values that produce cointegrated y and x time series.

- The Stata .do file “CharRoots”. This program solves for the characteristic roots of the CVAR system to check that $\lambda_1 = 1$, $|\lambda_2| < 1$.
- The Stata .do file “VECrnk”. This program checks the rank of $(A - I) = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix}$. A rank of 1 corresponds to the y and x variables being cointegrated.
- The Stata .do file “VECstable”. This program simulates 1000 observations of the data and estimates a VEC model to check that the estimated values of δ_y , δ_x , and θ are close to their analytical values.
- The Stata .do files “function_data”, “Table3A”, and “Table3B”. These programs are used to create the results for TABLE 3 in the paper.
- The Stata .do files “function_data”, “Table4A”, and “Table4B”. These programs are used to create the results for TABLE 4 in the paper.

Together, these programs allow one to not only check all the programming in the paper, but to derive other combinations of coefficient values for the DGP to check the performance of the corresponding unit root tests. The links to the respective data files are given below:

<http://www.economics-ejournal.org/dp-2015-57-charroots.do/view>

http://www.economics-ejournal.org/dp-2015-57-function_data.do/view

<http://www.economics-ejournal.org/dp-2015-57-parameter-values.xls/view>

<http://www.economics-ejournal.org/dp-2015-57-table3a.do/view>

<http://www.economics-ejournal.org/dp-2015-57-table3b.do/view>

<http://www.economics-ejournal.org/dp-2015-57-table4a.do/view>

<http://www.economics-ejournal.org/dp-2015-57-table4b.do/view>

<http://www.economics-ejournal.org/dp-2015-57-vecrnk.do/view>

<http://www.economics-ejournal.org/dp-2015-57-vecstable.do/view>

- 3) **COMMENT:** *“Some pages are spent calculating back and forth between the CVAR in (6) and the ARDL formulation in (1). This is of course well known and some space would be saved by sticking to only one representation.”*

RESPONSE: I would like to keep the current presentation the way that it is. The cost is approximately half a page of derivations and discussion (from the bottom of page 3 to the top of page 4). The benefit of the ARDL representation is that it allows one to connect the theory to the accompanying programs (more on the accompanying programs above). The benefit of the CVAR representation is that it provides some intuition about the long-run relationship between the variables and the dynamic return to equilibrium.

- 4) **COMMENT:** *"It is unclear what you mean when saying that a VEC model is well behaved."*

RESPONSE: I agree that the original manuscript needs to be clearer by what I mean by "well-behaved." δ_y , the speed of adjustment parameter in the Δy_t equation, identifies the change in y that accompanies a one-unit increase in y from its equilibrium value at time $t-1$. The condition $-1 < \delta_y < 0$ guarantees that y_t adjusts by returning to its equilibrium value by a fraction of the original deviation. Likewise, $\delta_x\theta$ identifies the change in x that accompanies a one-unit increase in x from its equilibrium value at time $t-1$. The condition $-1 < \delta_x\theta < 0$, guarantees that x_t adjusts by returning to its equilibrium value by a fraction of the original deviation. These conditions are more restrictive than they need to be, for example, they do not allow for overshooting, but they are sufficient to guarantee that deviations from long-run equilibrium cause the series to smoothly return to their long-run equilibrium values.

- 5) **COMMENT:** *For cointegrated I(1) data we know that $y_t + \theta x_t$ is an AR(1) process with coefficient $1 + \delta_y + \delta_x\theta$ which satisfies $-1 < 1 + \delta_y + \delta_x\theta < 1$. The right hand side is satisfied if $\delta_y + \delta_x\theta < 0$; so it is not sufficient that either δ_y or $\delta_x\theta$ is negative, but certainly sufficient that they both are.*

RESPONSE: I thank the reviewer for clarifying this. The revised manuscript will explicitly state this condition. I note that the simulations all satisfy this condition.

- 6) **COMMENT:** *"You write "The first two columns of TABLE 2 describe the model parameters." This is obviously not so, as the constant terms β_{10} and β_{20} in equation (1) do not appear in TABLE 2. However, based on the text of TABLE 1, and "Case 1" on page 4, it appear that the constant terms are set to zero. Presumably they are also zero in the other simulations. This is of course important as under the null of nonstationarity, a constant term would generate a trend in the process, and change the asymptotic behaviour of the test statistics, see (1).*

RESPONSE: The reviewer is correct. The constant terms β_{10} and β_{20} equal zero for all the simulations. The revised manuscript will make this clear.

- 7) **COMMENT:** *You write "The Z column is useful for illustrating the range of deviations that can be expected from a sampling error." This is clearly not correct. The Z column illustrates the deviation you can get from testing in a true random walk model, but you are simulating much more complex structures which have many more lags."*

RESPONSE: I don't think we disagree here. The reviewer is correct that the sampling error associated with unit root tests of z cannot be applied to the sampling error associated with the unit root tests of the cointegrated variables. Indeed, the main

point of the paper is that the distribution of sample statistics from unit root tests for cointegrated data is different than those for univariate random walk variables. The revised manuscript will avoid any confusion by dropping this statement.

- 8) **COMMENT:** *“...you estimate and calculate test statistics from models that are misspecified because of the lags chosen.”*

RESPONSE: I am not sure what the reviewer means by the last statement, “because of the lags chosen.” The revised manuscript will emphasize that the reason univariate unit root tests perform so poorly is because they are misspecified for cointegrated data.

- 9) **COMMENT:** *“It is unclear how from these equations you can derive that $\rho_y = \delta_y$ and $\rho_x = \theta\delta_x$.”*

RESPONSE: I agree with the reviewer that one cannot conclude from the derivations that $\rho_y = \delta_y$ and $\rho_x = \theta\delta_x$. However, I still think that the juxtaposition of the two sets of equations highlights the misspecification of the univariate unit root tests. The univariate unit root tests are specified as below:

$$\begin{aligned}\Delta y_t &= \gamma_y + \rho_y y_{t-1} + \text{potentially serially correlated error}_{yt} \\ \Delta x_t &= \gamma_x + \rho_x x_{t-1} + \text{potentially serially correlated error}_{xt}\end{aligned}$$

The test for unit root consists of testing that $\rho_y, \rho_x = 0$. In other words, changes in y and x are unrelated to their levels in the previous period.

In contrast, the true relationship between changes in y and x and their levels is given by the CVAR equations below:

$$\begin{aligned}\Delta y_t &= a_{10} + \delta_y y_{t-1} + \delta_y \theta x_{t-1} + \epsilon_{yt} \\ \Delta x_t &= a_{20} + \delta_x \theta x_{t-1} + \delta_x y_{t-1} + \epsilon_{xt}\end{aligned}$$

Clearly, changes in y and x ARE related to their levels in the previous period. In particular, deviations from their equilibrium values cause y and x to change in such a way as to return them to their equilibrium values.

- 10) **COMMENT:** *“In fact it follows from the Granger representation theorem that*

$$y_t = \frac{\theta}{\theta\delta_x + \delta_y} \sum_{i=1}^t \{\delta_x \epsilon_{yi} - \delta_y \epsilon_{xi}\} + \theta \frac{\delta_x a_{10} - \delta_y a_{20}}{\theta\delta_x + \delta_y} t + \text{stationary terms} \quad (1)$$

This shows that for the investigations to work you need the restriction $\delta_x a_{10} = \delta_y a_{20}$ to avoid a linear trend in the process, which otherwise would change the limit distributions of the statistics you work with.”

RESPONSE: I thank the reviewer for pointing this out. The revised manuscript will set $\beta_{10} = \beta_{20} = 0$ in the equations below, so that the DGP changes from this:

$$y_t = \beta_{10} + \beta_{12}x_t + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + \varepsilon_{yt} ,$$

$$x_t = \beta_{20} + \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}x_{t-1} + \varepsilon_{xt} ;$$

to this:

$$y_t = \beta_{12}x_t + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + \varepsilon_{yt} ,$$

$$x_t = \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}x_{t-1} + \varepsilon_{xt} .$$

As the reviewer has noted, the analysis already does this *de facto*, as all the simulations set $\beta_{10} = \beta_{20} = 0$. It follows that $a_{10} = a_{20} = 0$, so that the condition $\delta_x a_{10} = \delta_y a_{20}$ is met in all the simulations.

- 11) **COMMENT:** *““This suggests that researchers should..., opting for a more holistic approach.” It is a fact that although the CVAR methodology has been available for more than 30 years by now including the necessary software, many users still prefer to use unit root tests on individual variables. The standard (CVAR) approach suggests finding the rank first, and test for the stationarity of the individual variables by testing that a unit vector is a “cointegrating vector”. I assume that this is what the author calls a “holistic” approach, and I find that the follow up analysis of the 2 dimensional CVAR is needed to show how the problems of the univariate unit roots can be solved.*

RESPONSE: The implication of the reviewer’s comment is that I should not suggest a more “holistic” approach until I can confirm that a more “holistic” approach would produce satisfactory outcomes. I agree. The revised manuscript will drop this recommendation and suggest it instead as a topic for future research. With respect to his/her recommendation that my analysis be extended to show how the problems of univariate unit roots can be solved, I believe that this lies outside the purview of the current paper for two reasons.

First, a proper analysis of how best to address the problems of univariate unit roots involves a substantial expansion of the scope of my paper. Such an investigation would involve a comparison of numerous cointegration tests. The reviewer mentions “finding the rank,” but there are many tests/procedures for identifying the rank of the system. There is the trace test, the maximum eigenvalue test, and numerous information criteria for rank selection. In addition, there are Engle Granger single equation tests for cointegration, and Pesaran and Shin (1999) and Pesaran et al. (2001) ARDL Bounds tests. In addition, there is a strand of research that suggests that univariate unit root tests can have their power increased by including correlated variables in the unit root specification (Hansen, *Econometric Theory*, Vol. 11, 1995: 1148-1171; Westerlund, *Journal of Time Series Analysis*, Vol. 34, 2013: 477-495). As a result, an analysis of how best to handle the problems of univariate unit root tests would constitute a separate paper on its own.

Second, I believe the contribution of the current paper is sufficient to stand on its own. My paper points out the problems associated with a very common procedure for estimating relationships between time series variables. How

common? The papers below were all recently published in *Economics: The Open-Access, Open-Assessment E-Journal*, and they all begin by testing for unit roots in the individual variables before going on to estimate a cointegrated relationship:

- Haertel and Lucke (2008). "Do News Shocks Drive Business Cycles? Evidence from German Data." Vol. 2, 2008-10.
- Schneider, Chen, and Frohn (2008). "A Long-Run Structural Macroeconometric Model for Germany: An Empirical Note." Vol. 2, 2008-16.
- Bogoev, Terzijan, Egert, and Petrovska (2008). "Real Exchange Rate Dynamics in Macedonia: Old Wisdoms and New Insights." Vol. 2, 2008-18.
- Juselius (2008). "Testing the New Keynesian Model on U.S. and Euro Area Data." Vol. 2, 2008-24.
- Bjornstad and Kalstad (2010). "Increased Price Markup from Union Coordination – OECD Panel Evidence." Vol. 4, 2010-30.
- Elgin and Kuzubas (2013). "Wage-Productivity Gap in OECD Economies." Vol. 7, 2013-21.
- Kleeman (2014). "Knowing Where Organic Markets Move Next – An Analysis of Developing Countries in the Pineapple Market." Vol. 8, 2014-4.
- Herzer and Nunnenkamp (2015). "Income Inequality and Health: Evidence from Developed and Developing Countries." Vol. 9, 2015-4.

I emphasize that these are just from the journal *Economics: The Open-Access, Open-Assessment E-Journal*.

Further evidence is given in Peter Kennedy's Guide to Econometrics, 6th edition, 2008, p. 303:

"These results suggest the following methodology for practitioners. First, use unit root tests to determine the order of integration of the raw data series. Second, run the cointegrating regression suggested by economic theory. Third, apply an appropriate unit test to the residuals from this regression to test for cointegration. Fourth, if cointegration is accepted, use the lagged residuals from the cointegrating regression as an error correction term in an ECM."

Kennedy goes on to discuss alternative estimation procedures, such as estimation of VEC models. But the key thing to note is that the first step is to use univariate unit root tests to determine the order of integration of the respective time series variables. And researchers are unaware of the problems with following this approach.

As a result, I believe that my paper's focus should be maintained on the inappropriateness of applying univariate unit root test to cointegrated data. To underscore the paper's main finding, I am thinking of changing the title of the paper to: "Univariate Unit Root Tests Perform Poorly When Applied to Cointegrated Variables". I would appreciate the reviewer's opinion about this title change.

APPENDIX

The Excel spreadsheet “Parameter Values”. The worksheet “ParameterValues” allows one to select parameter values that produce CI(1,1) series. The yellow-highlighted cells identify where the researcher enters values for a_{22} (STEP ONE), and a_{12} and a_{21} (STEP TWO) and gives instructions to ensure they satisfy the necessary conditions.

STEP ONE	Choose $0 < a_{22} < 1$								
a22 =	0.9								
STEP TWO	Choose a_{12} and a_{21} such that $0 < a_{12}a_{21} < 1 - a_{22}$								
a12 =	-0.1								
a21 =	-0.9								
a12a21 =	0.09	1-a22=	0.1						

From here, the spreadsheet calculates the corresponding value of a_{11} that guarantees that the largest eigenvalue = 1 (STEP THREE).

STEP THREE	a11 is set equal to $1 - (a_{12}a_{21}/1 - a_{22})$ to guarantee that the second eigenvalue is < 1								
a11 =	0.1								

The next step works backwards to calculate the values of β and γ that will produce the values of a_{11} , a_{12} , a_{21} and a_{22} that one selected in STEPS ONE through THREE. It is straightforward to show that the values of a_{11} , a_{12} , a_{21} and a_{22} in Equation (2) are related to the values of β and γ in Equations (1) and (2) by the following:

$$(8a) \quad a_{11} = \frac{\gamma_{11} + \beta_{12}\gamma_{21}}{1 - \beta_{12}\beta_{21}}$$

$$(8b) \quad a_{12} = \frac{\gamma_{12} + \beta_{12}\gamma_{22}}{1 - \beta_{12}\beta_{21}}$$

$$(8c) \quad a_{21} = \frac{\gamma_{21} + \beta_{21}\gamma_{11}}{1 - \beta_{12}\beta_{21}}$$

$$(8d) \quad a_{22} = \frac{\gamma_{22} + \beta_{21}\gamma_{12}}{1 - \beta_{12}\beta_{21}}$$

(8a) – (8d) provide 4 equations with 6 unknowns, β_{12} , β_{21} , γ_{11} , γ_{12} , γ_{21} , and γ_{22} . Accordingly, the yellow-highlighted cells in STEP FOUR allow one to set β_{12} and β_{21} to any value so long as their product does not equal 1.

STEP FOUR	Let b12 and b21 take any values such that b12*b21 is not equal to 1	
b12 =	2	
b21 =	5	
b12b21 =	10	
X =	-9	

Once β_{12} and β_{21} have been set, one can use Equations (8a) – (8d) to solve for the corresponding values of γ_{11} , γ_{12} , γ_{21} , and γ_{22} (STEP FIVE).¹

STEP FIVE	This uses the definitions of a11, a12, a21, and a22 to solve for g11, g12, g21, and g22, given b12 and b22	
g21 =	-1.4000	NOTE: It is important that the last two digits be zero so that
g11 =	1.9000	these EXACT parameter values can be entered into Stata
g22 =	1.4000	
g12 =	-1.9000	

Note that these parameter values will be entered into the Stata .do files. The values need to be exact in order for the series to be cointegrated. Approximate values, such as 0.333 for 1/3, will not suffice. So this section warns that the user must be sure that the respective parameter values do not have trailing values.

The last part of the worksheet reports the corresponding values of θ , δ_y , and δ_x .

THIS REPORTS LR EQUILIBRIUM RELATIONSHIP (theta) AND SPEED OF ADJUSTMENT VALUES (deltay/deltax)		
theta =	0.111111	
deltay =	-0.9	NOTE: Need to confirm that this is between -1 and 0.
deltax =	-0.9	
deltax*theta =	-0.1	NOTE: Need to confirm that this is between -1 and 0.

This allows one to check that the speed of adjustment parameters satisfy the conditions (i) $-1 < \delta_y < 0$, and (ii) $-1 < \delta_x \theta < 0$, which are imposed to guarantee that the series return smoothly to their equilibrium values following shocks to the system.

¹ Note that g11, g12, g21, and g22 represent γ_{11} , γ_{12} , γ_{21} , and γ_{22} , respectively.

Also included in the Excel spreadsheet is a worksheet by the name of “PasteValues.” It summarizes all the parameter values that were calculated on the “ParameterValues” worksheet (see below).

local b10 =	0	b10 =	0
local b12 =	2	b12 =	2
local g11 =	1.9	g11 =	1.9
local g12 =	-1.9	g12 =	-1.9
local b20 =	0	b20 =	0
local b21 =	5	b21 =	5
local g21 =	-1.4	g21 =	-1.4
local g22 =	1.4	g22 =	1.4

These are formatted for easy copy-and-pasting into the subsequent .do files.

The Stata .do file “CharRoots”. The .do file “CharRoots” takes the parameter values from the spreadsheet and calculates the associated eigenvalues. It also provides a plot of some of the y_t and x_t values so that one can visualize how they relate. All one has to do is copy the appropriate cells from the “PasteValues” worksheet, paste them into the corresponding section of the .do file, and then run the program (see below).

```
// This section fixes the parameter values of the DGP
b10 = 0
b12 = 2
g11 = 1.9
g12 = -1.9
b20 = 0
b21 = 5
g21 = -1.4
g22 = 1.4
```

The associated output confirms that the two eigenvalues satisfy the conditions $\lambda_1 = 1$, $|\lambda_2| <$

1:

```

      1   2
1   1   0

abs(lambda)
      1   2
1   1   0
```

The Stata .do file “VECrnk”. The .do file “VECrnk” takes the parameters from the spreadsheet and uses the respective testing procedures in Stata to test for the rank of the matrix $(A - I) = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix}$. A rank of 1 is consistent with the variables Y and X being cointegrated. To run this program, follow the same procedure as above and copy and paste the parameter values from the “PasteValues” worksheet into the appropriate section of the .do file, as discussed above.

This program produces output in two parts. The first part calculates the rank of $(A - I)$ directly from the population parameter values of the DGP:

```
A1I = A1-I(2)
rank(A1I)
1
```

The second part simulates 1000 observations of X and Y and performs (i) the trace test, (ii) the maximum eigenvalue test, and (iii) presents a series of information criterion values using the Schwarz Bayesian information criterion (SBIC), the Hannan and Quinn information criterion (HQIC), and the Akaike information criterion (AIC). All the results from the table below indicate that $(A - I) = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix}$ has a rank of 1, consistent with the data being cointegrated.

Johansen tests for cointegration						
Trend: constant			Number of obs =		998	
Sample: 3 - 1000			Lags =		2	
<hr/>						
maximum				trace	5%	
rank	parms	LL	eigenvalue	statistic	critical value	
0	6	-832.09042	.	430.6955	15.41	
1	9	-616.91072	0.35029	0.3362*	3.76	
2	10	-616.74265	0.00034			
<hr/>						
maximum				max	5%	
rank	parms	LL	eigenvalue	statistic	critical value	
0	6	-832.09042	.	430.3594	14.07	
1	9	-616.91072	0.35029	0.3362	3.76	
2	10	-616.74265	0.00034			
<hr/>						
maximum						
rank	parms	LL	eigenvalue	SBIC	HQIC	AIC
0	6	-832.09042		1.709033	1.690751	1.67954
1	9	-616.91072	0.35029	1.29857*	1.271146*	1.25433
2	10	-616.74265	0.00034	1.305153	1.274682	1.255997

The Stata .do file “VECstable”. The .do file “VECstable” takes the parameters from the spreadsheet, simulates 1000 observations using those parameter values, estimates the VEC model, and then reports the estimated eigenvalues of the system. This program also has output in two parts. The first part estimates of the VEC model and allows one to compare the estimated values of θ , δ_y , and δ_x with their true values, as reported at the bottom of the “ParameterValues” worksheet.

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
D_y	_ce1						
	L1.	-.9075725	.0403104	-22.51	0.000	-.9865793	-.8285656
	y						
	LD.	-.0016972	.0332669	-0.05	0.959	-.0668991	.0635047
	x						
	LD.	-.0019733	.0196572	-0.10	0.920	-.0405007	.0365541
	_cons	.0124837	.0079948	1.56	0.118	-.0031858	.0281533
D_x	_ce1						
	L1.	-.935039	.0906559	-10.31	0.000	-1.112721	-.7573566
	y						
	LD.	.0216589	.0748156	0.29	0.772	-.1249769	.1682947
	x						
	LD.	-.0001907	.044208	-0.00	0.997	-.0868369	.0864554
	_cons	-.012117	.0179799	-0.67	0.500	-.047357	.0231229

Johansen normalization restriction imposed							
beta		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_ce1	y	1
	x	.1106794	.0012192	90.78	0.000	.1082897	.113069
	_cons	.0069962

From the output above, we see that the VEC estimates of δ_y , δ_x , and θ are -0.908, -0.935, and 0.111. The true values are reported on the bottom of the “Parameter Values” worksheet:

THIS REPORTS LR EQUILIBRIUM RELATIONSHIP (theta) AND SPEED OF ADJUSTMENT VALUES (deltay/deltax)							
theta =	0.111111						
deltay =	-0.9	NOTE: Need to confirm that this is between -1 and 0.					
deltax =	-0.9						
deltax*theta=	-0.1	NOTE: Need to confirm that this is between -1 and 0.					

The last part of the program estimates the value of the second eigenvalue, assuming that the larger of the two eigenvalues equals one:

Eigenvalue stability condition	
Eigenvalue	Modulus
1	1
.04103287	.041033
-.0269914 + .01791487i	.032396
-.0269914 - .01791487i	.032396

The VECM specification imposes a unit modulus.

$|\lambda_2|$ is estimated to be 0.041. We know from “CharRoots” program above that the true value of the second eigenvalue is zero. While standard errors are not calculated, the estimated value of $|\lambda_2|$ is well below 1, confirming again that the series are cointegrated.

The Stata .do files “function_data”, “Table3A”, and “Table3B”. The .do files “function_data”, “Table3A”, and “Table3B” are a suite of three programs that are designed to be used together. These programs perform the simulation exercises that test each of the series for unit root using the (i) ADF, (ii) Phillips-Perron, and (iii) DF-GLS tests. The programs simulate 10000 data sets of 100 observations each of X, Y, and a third variable Z. X and Y are simulated using the parameter values pasted into the program “Table3B”:

```
// This section fixes the parameter values of the DGP
local b10 = 0
local b12 = 2
local g11 = 1.9
local g12 = -1.9
local b20 = 0
local b21 = 5
local g21 = -1.4
local g22 = 1.4
```

The variable Z is a classic random walk variable:

$$(9) \quad z_t = z_{t-1} + \varepsilon_{zt}, \quad \varepsilon_{zt} \sim NID(0,1) .$$

It is included in the simulations for comparison’s sake.

The programs must be run in order (first “function_data”, then “Table3A”, and then “Table3B”). The output consists of two parts. The first part are the Type I error rates associated with the three unit root tests:

```

RESULTS[3,3]
      X      Y      Z
DFuller  .035  .325  .053
PFerron  .291   1    .066
  DFGLS  .075  .831  .039

```

While the Phillips-Perron and DF-GLS tests automatically select lag lengths, the default option in Stata is to include no lagged differenced terms in the Dickey-Fuller specification. As serial correlation in the error term distort test results, it is advisable to add lagged differenced terms.

My simulations add 4 lags to the Dickey-Fuller test. To ensure that this is sufficient to account for serial correlation, I perform a Breusch-Godfrey test following each Dickey-Fuller test. The associated rejection rates are reported in the second part of the output:

```

BGODFREY[1,3]
      X      Y      Z
BGodfrey  .056  .056  .055

. display as text "Number of lags = `lagz'"
Number of lags = 5

```

This provides a check to ensure that I have added sufficient lagged differenced terms to ensure that the Dickey-Fuller test results are not distorted by the presence of serial correlation.

The Stata `.do` files “`function_data`”, “`Table4A`”, and “`Table4B`”. The last set of programs are designed to illustrate the effect of increasing the number of lagged differenced terms to the Dickey-Fuller unit root specification. As with the TABLE 3 programs, they must be run in order (first “`function_data`”, then “`Table4A`”, and then “`Table4B`”).