

Econometrics Basics: Avoiding Spurious Regression

John E. Floyd
University of Toronto
July 24, 2013

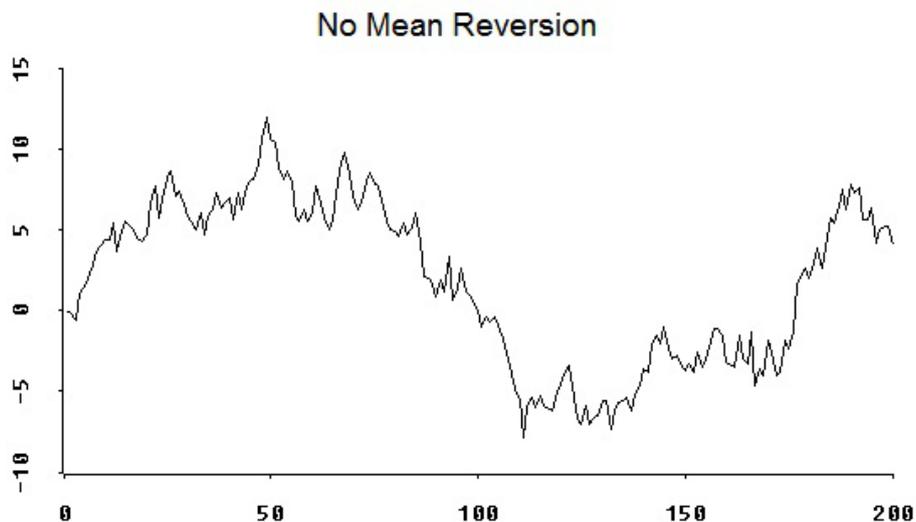
We deal here with the problem of spurious regression and the techniques for recognizing and avoiding it. The nature of this problem can be best understood by constructing a few purely random-walk variables and then regressing one of them on the others. The figure below plots a random walk or unit root variable that can be represented by the equation

$$y_t = \rho y_{t-1} + \epsilon_t \quad (1)$$

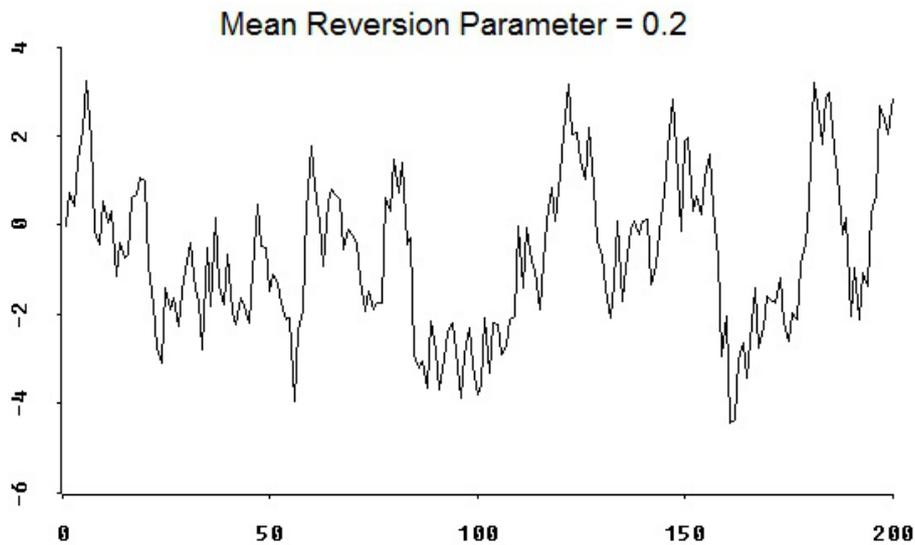
which can be written alternatively in Dickey-Fuller form as

$$\Delta y_t = -(1 - \rho) y_{t-1} + \epsilon_t \quad (2)$$

where y_t is the level of the series at time t , ϵ_t is a series of drawings of a zero-mean, constant-variance normal random variable, and $(1 - \rho)$ can be viewed as the mean-reversion parameter. If $\rho = 1$, there is no mean-reversion and y_t is a random walk.

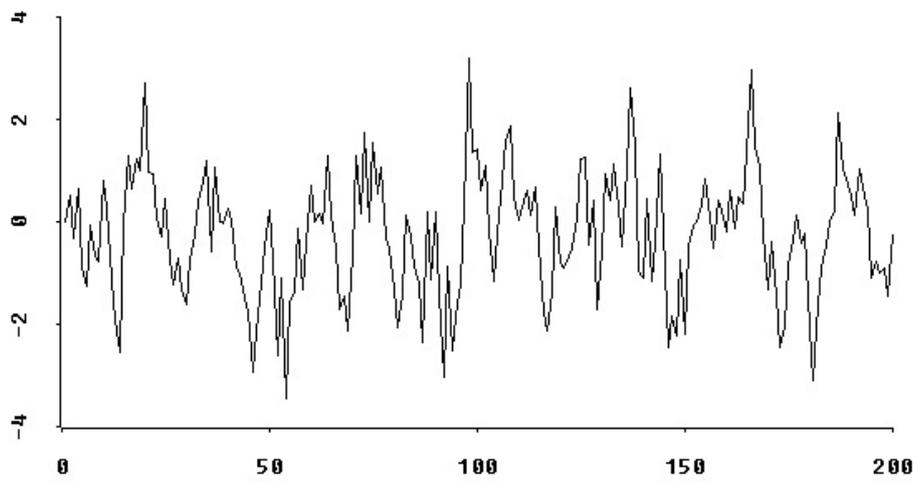


Notice that, apart from the short-term variations, the series trends upward for the first quarter of its length, then downward for a bit more than the next quarter and upward for the remainder of its length. This series will tend to be correlated with other series that move in either the same or the opposite directions during similar parts of their length. And if our series above is regressed on several other random-walk-series regressors, one can imagine that some or even all of those regressors will turn out to be statistically significant even though by construction there is no causal relationship between them—those parts of the dependent variable that are not correlated directly with a particular independent variable may well be correlated with it when the correlation with other independent variables is simultaneously taken into account.

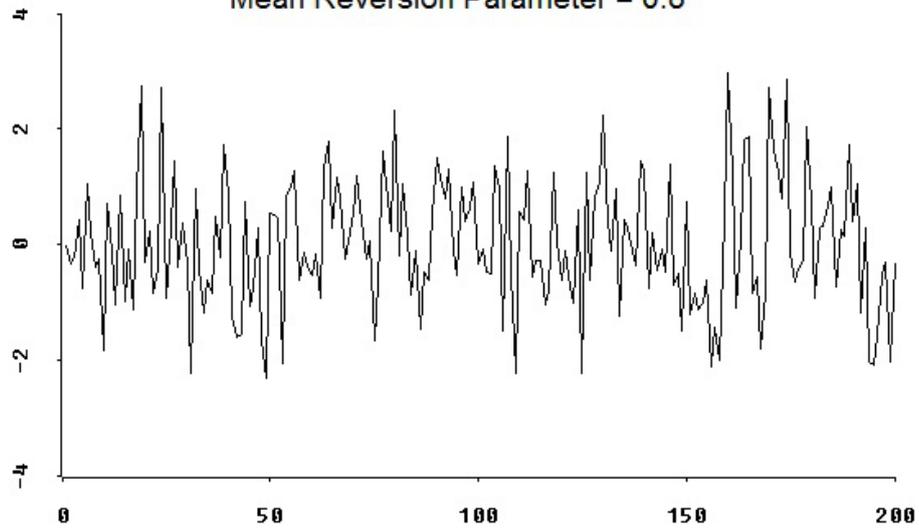


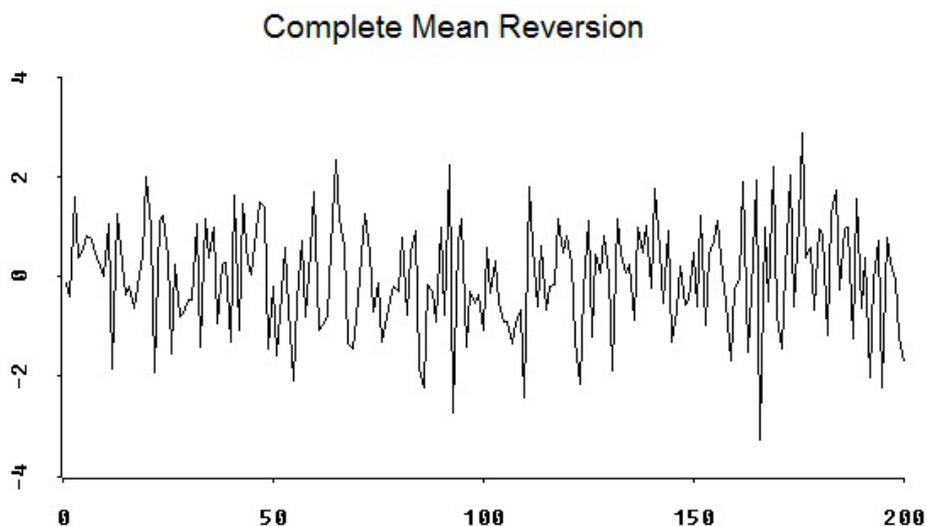
The figure above and the three below present plots, in turn, of constructed time-series variables that have increasing degrees of mean reversion, ending up with complete mean reversion in the case of the last figure.

Mean Reversion Parameter = 0.5



Mean Reversion Parameter = 0.8





Notice how the major fluctuations in the series become of smaller magnitude and closer together as the mean-reversion parameter $(1 - \rho)$ increases, with the result that in figure immediately above having complete mean reversion the variation of y_t is such that the variable deviates each period from its average level by a random amount.

Let us now construct 1000 groups of unit-root series and regress one series in each group on the other series in that group.¹ The middle column in the little table below gives the percentage of the 1000 regressions in which all the coefficients, apart from the constant term, are statistically significant. The first row indicates that when the 1000 spurious regressions have a single independent variable, other than the constant term, that independent variable is significant at the 5% level 85 percent of the time. When there are two independent variables in the group of 1000 regressions, both are statistically significant 70 percent of the time. As the number of regressors in subsequent groups of 1000 is increased further, the percentage of the time that all independent variables are significant falls to about 40 percent when there are four independent variables and to around 27 percent in the case of

¹These constructed data are generated and analyzed using the XLispStat batch file `spurregs.lsp`.

five independent variables. Of course, when we run additional corresponding groups of 1000 regressions the fractions of the regressions in which all independent variables are statistically significant will be slightly different from those below but of roughly the same magnitude. We can conclude that if we were to run a regression which, unknown to us, happens to be spurious, the probability of all variables being statistically significant is substantial, being greater than 0.25 even if there are five independent variables.

Number of Independent Variables	Percent of Coefficients Significant at 5% level	Plus Positive Signed
1	85.2	40.6
2	70.4	15.8
3	55.8	7.2
4	39.1	1.5
5	26.7	0.4

Each line based on 1000 regressions
with all variables non-stationary
and with constant terms added

As can be seen from the right-most column above, however, the fraction of the variables being not only statistically significant at the 5% level but also having signs predicted in advance, in this case all positive, gets very small as the number of regressors increases. When there are 4 independent variables, all significant, their signs are all positive less than 2 percent of the time, and this percentage drops to less than one-half of one percent when the number of independent variables is increased to 5. While we are predicting all signs to be positive in the case above, the same result would follow with any specific predicted sign pattern. Thus, if we observe the signs of all four independent variables to be what would be expected on the basis of economic theory, and all variables are statistically significant, we would expect the regression to be spurious less than 2% of the time. And, of course, this assumes that all variables in the regression are non-stationary.

Suppose now that the constructed variables in the groups of regressions are stationary but with some degree of mean reversion. In the table below, the degree of mean reversion is set, alternatively, to 0.2 and 0.5.

Number of Independent Variables	Mean Reversion – 0.2		Mean Reversion – 0.5	
	Percent of Coefficients Significant at 5% level	Plus Positive Signed	Percent of Coefficients Significant at 5% level	Plus Positive Signed
1	44.3	21.2	20.1	9.6
2	18.9	4.7	5	1.1
3	9.99	1.3	0.3	0
4	3.69	0.1	0.5	0.1
5	2.4	0.1	0.1	0

Each line based on 1000 regressions with constant terms added

As you can see, the chances of observing both of two independent variables, apart from the constant, significant at the 5% level with the expected signs would appear to be less than five percent when the mean reversion parameter equals 0.2 and slightly more than one percent when the mean reversion parameter is 0.5. When we increase the number of independent variables to 3, the chances of observing significant and correctly signed coefficients of all three variables is less than 1.5 percent when the mean reversion parameter is 0.2 and virtually zero when that parameter is 0.5.

Since the exact mean reversion parameter is difficult to measure precisely when mean reversion is not substantial, a useful rule is to conclude that a particular regression is unlikely to be spurious if all coefficients other than the constant are significant and correctly signed and at least four independent variables are present.

It is generally the case that when a stationary variable is regressed on a non-stationary one, the coefficient of that independent variable will be statistically insignificant. But substantial exceptions arise. For example, in 1000 simple regressions of independently constructed non-stationary variables on white-noise ones the coefficients of the white-noise variables were statistically significant about 9.3 percent of the time and both statistically significant and positively signed 4.7 percent of the time. Indeed, in the last of these 1000 regressions the totally mean reverting independent variable was significant at the 5% level as shown in the result below.

ORDINARY LEAST SQUARES REGRESSION

Dependent Variable: Var 0

	Coefficient	Std. Error	T-stat	P-Val
Const	-24.626	0.549	-44.863	0.000
Var 1	1.184	0.572	2.071	0.020

Number of Observations:	200
Degrees of Freedom:	198
R-Squared:	0.0212025228852057
Adjusted R-Squared:	0.016259101283615962
Sum of Squared Errors:	11882.762476888849
F-Statistic:	4.289037956703245
P-Value	0.015014651589257855

Not surprisingly, however, the R-Squared statistic was only .02 which says that the purely stationary independent variable explains only 2% of the variation of the non-stationary dependent variable. As the mean reversion of the single independent variable is reduced we obtain the results indicated in the table below.

Mean Reversion Parameter of Independent Variable	Percent of Coefficients Significant at 5% level	Plus Positive Signed
1.0	9.3	4.7
0.8	17.6	9.4
0.5	34.8	16.7
0.2	56.0	28.4
0.0	85.5	44.4

Each line based on 1000 regressions
with dependent variables non-stationary
and with constant terms added

As both variables become non-stationary, statistical significance of the spurious relationship between them occurs around 85% of the time and statistical significance along with a positive sign occurs almost 45% of the time. In the two-variable case, the fact that the independent variable is stationary as well as statistically significant cannot justify a conclusion that the regression is not spurious.

Consider now the case where we have a regression with a non-stationary dependent variable and four independent variables, one of which is stationary. Can we reasonably conclude that the regression is non-spurious if the stationary variable is statistically significant? Again we run sequences of 1000 regressions where all the non-stationary variables and the single stationary ones are independently constructed. The results are presented in the table below.

Mean Reversion Parameter of Stationary Variable	Stationary Variable		All Variables	
	Percent of Coefficients Significant at 5% level	Plus Positive Signed	Percent of Coefficients Significant at 5% level	Plus Positive Signed
1.0	11.3	6.1	5.2	0.3
0.8	17.7	9.0	8.3	0.3
0.5	31.4	14.8	15.9	1.1
0.2	53.2	25.0	27.2	2.0
0.0001	80.8	39.0	40.6	2.8

Each line based on 1000 regressions
with dependent variables non-stationary and
one of four independent variables stationary
and with constant terms added

Clearly, the fact that one of the independent variables is stationary reduces substantially the probability that all the coefficients will be statistically significant as well as the probability that they will be significant with the expected signs. And a situation where all of four independent variables are statistically significant and have the expected signs is likely to arise less than 5% of the time when the regression is spurious, regardless of whether one of the independent variables can be viewed as stationary. The more stationary a single variable is, the less likely that it will be statistically significant.

While we can roughly conclude that our regression is non-spurious if there are four or more independent variables, all of which are statistically significant and have the signs predicted by our economic theory, we need a clear statistical test of whether a regression is spurious. There are two such tests. The first is based on the fact that the residuals of a spurious regression are non-stationary while, if the residuals are stationary, the variables in a regression in which all are non-stationary are said to be cointegrated.² The residuals must be stationary if the variables are cointegrated because deviations of the dependent variable from the levels predicted by the independent variables must surely be temporary. A cointegration test in which we can reject the null-hypothesis that the regression residuals are non-stationary can therefore establish that the regression is not spurious—that the non-stationary variables in the regression are causally inter-related and simultaneously created by an underlying process, quite possibly represented by an economic theory we are testing.

The first step in regression-residual based cointegration test simply involves running a Dickey-Fuller test on the residuals of the regression. As an example, we conduct the test on the residuals of a spurious regression in which all the independent variables other than the constant term are statistically significant. We first present the regression results and plot the regression residuals.

²For a discussion of the basics of cointegration analysis, see Walter Enders, *Applied Econometric Time Series*, 1995, John Wiley and Sons, Chapter 6.

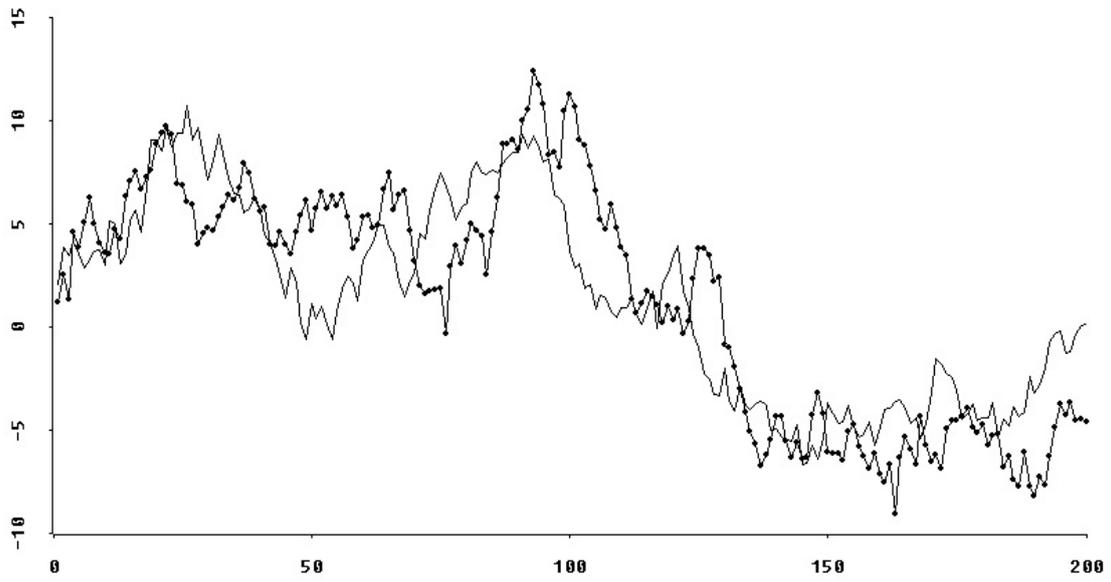
ORDINARY LEAST SQUARES REGRESSION

Dependent Variable: VAR0

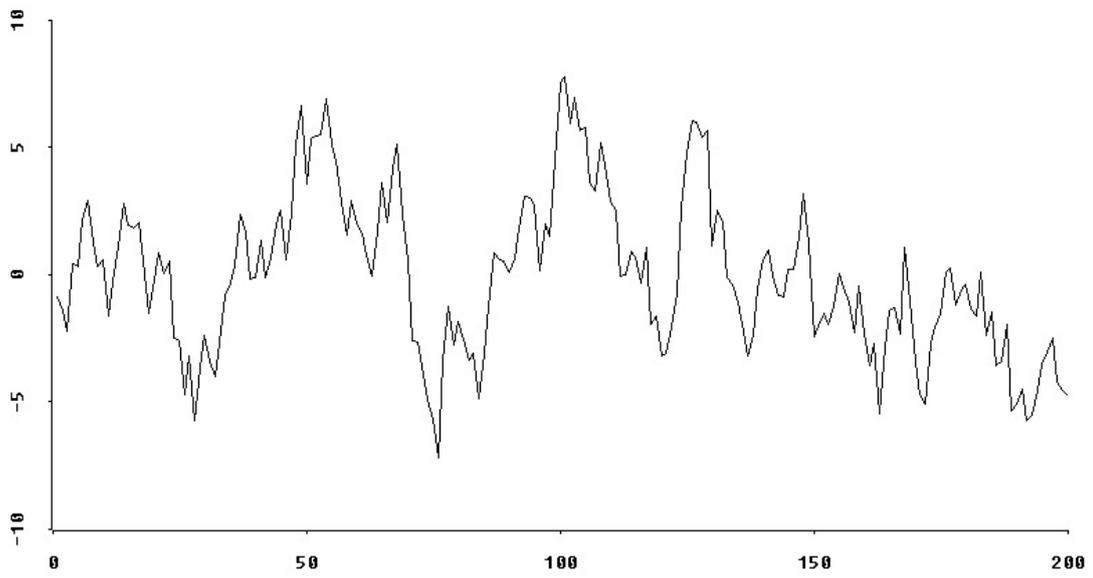
	Coefficient	Std. Error	T-stat	P-Val
Constant	0.853	0.678	1.259	0.105
VAR1	0.277	0.073	3.821	0.000
VAR2	0.730	0.056	13.072	0.000
VAR3	0.141	0.080	1.775	0.039
VAR4	0.550	0.069	7.972	0.000

Number of Observations: 200
Degrees of Freedom: 195
R-Squared: 0.7007616828736924
Adjusted R-Squared: 0.694623460983922
Sum of Squared Errors: 1912.7399991593813
F-Statistic: 114.16362840215007
P-Value 0.0

Actual (Points) and Fitted



Residuals



DICKEY-FULLER TEST ON REGRESSION RESIDUALS

Lags = 0

Starting observation = 1

Number of observations = 199

Coefficients:

Constant	0.273748	-0.016721	
Trend	-0.002901		
Y(t-1)	-0.133168	-0.117676	-0.117718

t-Statistics:

Constant	1.247921	-0.157729	
Trend	-1.511015		
Y(t-1)	-3.721153	-3.421000	-3.430782

F-Statistics:

All Three Coefficients = 0	4.698931
Trend & Y(t-1) Coefs = 0	7.031318
Constant & Y(t-1) Coefs = 0	5.868588

While non-stationarity of these residuals can clearly be rejected using the standard Dickey-Fuller tables, it turns out that those critical values are inappropriate. Instead, we must use the critical values established by Engle and Yoo that are presented in the second page of the `statabs.pdf` file that is attached to the end of this document.³ We need look only at the t -statistics for the coefficients of the lagged residual. With 200 observations and 4 independent variables plus a constant term, those t -ratios must exceed 4.5 in absolute value for us to reject non-stationarity at the 5% level. Clearly, non-stationarity cannot be rejected.

³Robert F. Engle and Byung Sam Yoo, "Forecasting and Testing in Co-Integrated Systems," *Journal of Econometrics*, Vol. 35, 1987.

The problem with the Dickey-Fuller approach is that the test is has low power when there is significant serial correlation, and hence slow mean-reversion, in stationary regression residuals. An alternative test, developed by Soren Johansen, is capable of finding cointegration where residual-based tests cannot.⁴ It is also capable of finding more than one cointegrating vector and the test result does not hinge on which variable one chooses to be the dependent variable in the cointegrating regression. Johansen's approach allows a number of cointegrating relationships equal to as many as one less than the number of non-stationary variables being analyzed. The function developed here for performing Johansen tests is based on the explanation by James D. Hamilton in Chapter 20 of his book *Time Series Analysis*, Princeton University Press, 1994.⁵

The n variables being tested for cointegration are expressed in vector-autoregression form as a vector \mathbf{y}_t containing the n elements, each of which is itself a vector representing one of the n variables. This is then rearranged to express the first differences of the elements in \mathbf{y}_t as a linear function of all their previous-period values and lagged first differences

$$\begin{aligned} \Delta \mathbf{y}_t = & \mathbf{c} + \Phi_0 \mathbf{y}_{t-1} + \Phi_1 \Delta \mathbf{y}_{t-1} + \Phi_2 \Delta \mathbf{y}_{t-1} + \cdots \\ & \cdots + \Phi_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{e}_t \end{aligned} \quad (3)$$

where \mathbf{c} is a n -element column vector of vectors of constants, Φ_0 , Φ_1 , Φ_2 , etc., are $n \times n$ element matrices of coefficients, and \mathbf{e}_t is an n element column vector of vectors of residuals.

The variables are cointegrated if there are one or more positive eigenvalues of the matrix Φ_0 . Associated with each of these eigenvalues will be a cointegrating vector. The scalar λ will be an eigenvalue of the square matrix Φ_0 if there exists a nonzero vector of real numbers \mathbf{x} for which

$$\Phi_0 \mathbf{x} = \lambda \mathbf{x}.$$

This vector \mathbf{x} will be the eigenvector associated with λ and a cointegrating vector for the respective time series represented by the n equations.

⁴See Soren Johansen, "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, Vol. 12, 1988, pp. 231-54.

⁵A careful reading of the Enders reference in a previous footnote is an important prerequisite to working through Hamilton's exposition.

In performing a Johansen test, the first step is to estimate two sets of OLS regressions. The first regresses $\Delta \mathbf{y}_t$ on constant terms plus $p - 1$ of own lags.

$$\Delta \mathbf{y}_t = \mathbf{a} + \mathbf{A}_1 \Delta \mathbf{y}_{t-1} + \mathbf{A}_2 \Delta \mathbf{y}_{t-2} + \cdots + \mathbf{A}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{u}_t \quad (4)$$

where \mathbf{a} is a vector of constant-term vectors, the \mathbf{A}_i are $(n \times n)$ matrices of coefficients and \mathbf{u}_t is a column vector containing vectors of OLS residuals. The second set regresses \mathbf{y}_{t-1} on a constant and the same $(p - 1)$ lags as in the previous regression

$$\mathbf{y}_{t-1} = \mathbf{b} + \mathbf{B}_1 \Delta \mathbf{y}_{t-1} + \mathbf{B}_2 \Delta \mathbf{y}_{t-2} + \cdots + \mathbf{B}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{v}_t \quad (5)$$

where \mathbf{v}_t is the resulting n element vector of vectors of regression residuals.

We then calculate the sample variance-covariance matrices of the residuals \mathbf{u}_t and \mathbf{v}_t

$$\Sigma_{\mathbf{v}\mathbf{v}} = (1/T) \sum_{t=1}^T \mathbf{v}_t \mathbf{v}_t' \quad (6)$$

$$\Sigma_{\mathbf{u}\mathbf{u}} = (1/T) \sum_{t=1}^T \mathbf{u}_t \mathbf{u}_t' \quad (7)$$

$$\Sigma_{\mathbf{u}\mathbf{v}} = (1/T) \sum_{t=1}^T \mathbf{u}_t \mathbf{v}_t' \quad (8)$$

$$\Sigma_{\mathbf{v}\mathbf{u}} = \Sigma_{\mathbf{u}\mathbf{v}}' \quad (9)$$

and from these find the eigenvalues of the matrix

$$\Sigma_{\mathbf{v}\mathbf{v}}^{-1} \Sigma_{\mathbf{v}\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \Sigma_{\mathbf{u}\mathbf{v}} \quad (10)$$

ordered from large to small — $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. Let h be the number of cointegrating vectors. The test-statistic for the null-hypothesis of no cointegrating vectors ($h = 0$) as opposed to one or more, called the Trace Statistic, turns out to be

$$-T \sum_{i=1}^n \log(1 - \lambda_i) \quad (11)$$

while the test-statistic for the null-hypothesis of no cointegrating vector as opposed to one, the L-Max Statistic, is

$$-T \log(1 - \lambda_1). \quad (12)$$

The Trace Statistic for more than h cointegrating vectors as opposed to the null-hypothesis of only h is

$$-T \sum_{i=h+1}^n \log(1 - \lambda_i) \quad (13)$$

and the L-Max Statistic for $h + 1$ cointegrating vectors as opposed to the null-hypothesis of h is

$$-T \log(1 - \lambda_{h+1}). \quad (14)$$

The critical values for these statistics are given on the third page in the file `statabs.pdf` which, as noted previously, is attached to the end of this document. The estimation outlined above allows \mathbf{a} to be non-zero so that it is fully compatible with deterministic trend-drift in the data. In this case of unrestricted estimation, the critical values depend on whether or not there are really deterministic trends in the data—if there are, the critical values are given in the top panel of the table while if there are not, the critical values are given in the panel below.

We can restrict the estimation to exclude deterministic trends by repeating the above calculations with the constant term vectors \mathbf{a} and \mathbf{b} eliminated from the system of auxiliary regressions given by equations (4) and (5). The test statistics are calculated in the same way but, for reasons to be noted below, the critical values are not included in the above-noted file.

An alternative restricted estimation procedure is to exclude deterministic trends for all the variables but allow constant terms in the cointegrating relations. In this case we have to estimate three auxiliary equations instead of two in the previous cases. First we estimate equation (4) without the constant vector \mathbf{a} . Then we regress a constant term on the $(p - 1)$ lags of $\Delta \mathbf{y}_t$ as follows.

$$1 = \omega_1' \Delta \mathbf{y}_{t-1} + \omega_2' \Delta \mathbf{y}_{t-2} + \cdots \cdots + \omega_{p-1}' \Delta \mathbf{y}_{t-p+1} + \mathbf{w}_t \quad (15)$$

Finally we estimate equation (5) without the constant vector \mathbf{b} . We denote the residuals from the first and third regressions as $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ and construct a vector $\hat{\mathbf{w}}_t$ equal to the column vector $\hat{\mathbf{v}}_t$ with \mathbf{w}_t combined on top of it. We then calculate the variance-covariance matrices of the residuals

$$\hat{\Sigma}_{\mathbf{ww}} = (1/T) \sum_{t=1}^T \hat{\mathbf{w}}_t \hat{\mathbf{w}}_t' \quad (16)$$

$$\hat{\Sigma}_{\mathbf{uu}} = (1/T) \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \quad (17)$$

$$\hat{\Sigma}_{\mathbf{uw}} = (1/T) \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{w}}_t' \quad (18)$$

$$\hat{\Sigma}_{\mathbf{wu}} = \hat{\Sigma}_{\mathbf{uw}}' \quad (19)$$

and from these find the eigenvalues of the matrix

$$\hat{\Sigma}_{\mathbf{ww}}^{-1} \hat{\Sigma}_{\mathbf{wu}} \hat{\Sigma}_{\mathbf{uu}}^{-1} \hat{\Sigma}_{\mathbf{uw}} \quad (20)$$

ordered from large to small — $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_n$. The Trace and L-Max statistics are obtained in the same way as before. Again, for reasons that will be noted below, the appropriate critical values are not presented in our tables.

To obtain a clear idea of which one of the above Johansen procedures to draw conclusions from in any particular case, we can perform some tests. To test for the presence of an intercept in the cointegrating vector but no trends in the data as opposed to complete trend-drift, we can calculate the statistic

$$-T \left[\log|\hat{\Sigma}_{\mathbf{uu}}| - \log|\Sigma_{\mathbf{uu}}| + \sum_{i=1}^h \log(1 - \hat{\lambda}_i) - \sum_{i=1}^h \log(1 - \lambda_i) \right] \quad (21)$$

which is asymptotically χ^2 with $(n - h)$ degrees of freedom. The terms $|\hat{\Sigma}_{\mathbf{uu}}|$ and $\hat{\lambda}_i$ are from the estimation that includes a constant but no trend drift while the terms $|\Sigma_{\mathbf{uu}}|$ and λ_i are from the estimation that includes trend drift. If the magnitude of the expression is large enough and the test-statistic is therefore sufficiently large, we can reject the null-hypothesis of no deterministic trends in the series but intercepts in the cointegrating vectors in favour of the inclusion of trend drift—a test statistic that large is unlikely to have occurred on the basis of pure chance. Similarly, we can test the null-hypothesis of no trends in any of the series and no constants in the cointegrating vector as opposed to the other two alternatives by replacing $\hat{\Sigma}_{\mathbf{uu}}$ and $\hat{\lambda}_i$ above with the corresponding terms for the calculations that exclude deterministic trends and allow no constant in the cointegrating vector and the terms $\Sigma_{\mathbf{uu}}$ and λ_i with either the corresponding terms when a constant but no deterministic trends are included or the corresponding terms when unrestricted deterministic trends are allowed.

One further issue must be faced before we can proceed to do Johansen tests—we must figure out how many lags to include in the basic system (3). The common procedure is to estimate the system in the levels (using undifferenced data) with various lag lengths

$$\mathbf{y}_t = \mathbf{a} + \mathbf{b}_1 \mathbf{y}_{t-1} + \mathbf{b}_2 \mathbf{y}_{t-2} + \mathbf{b}_3 \mathbf{y}_{t-2} + \cdots + \mathbf{e}_t \quad (22)$$

and then calculate test-statistics for a Likelihood Ratio test of the form suggested by Sims⁶

$$(T - c)(\log|\boldsymbol{\Sigma}_R| - \log|\boldsymbol{\Sigma}_U|)$$

where T is the number of observations, c is the number of parameters in the unrestricted system (the one including the most lags) and $\log|\boldsymbol{\Sigma}_R|$ and $\log|\boldsymbol{\Sigma}_U|$ are the logarithms of determinants of the variance-covariance matrices of the residuals in the respective restricted (by the inclusion of fewer lags) and unrestricted systems. This statistic is distributed as χ^2 with degrees of freedom equal to the number of coefficient restrictions. In this case, each lag has n^2 coefficients (n lags of n variables) so the number of restrictions will equal $m \times n^2$ where m is the difference in the number of lags of the restricted as opposed to the unrestricted system. Alternatively, we can select the lag length that minimises the following multivariate generalisation of AIC or SBC.

$$\begin{aligned} \text{AIC} &= T \log|\boldsymbol{\Sigma}| + N \\ \text{SBC} &= T \log|\boldsymbol{\Sigma}| + N \log(T) \end{aligned}$$

where $|\boldsymbol{\Sigma}|$ is the variance-covariance matrix of the residuals and N is the number of parameters estimated in all equations. The same sample period must be used for all systems compared.⁷

Before writing a function to perform the Johansen test we need one to enable us to decide on the appropriate lag length. This function, which we call `Var-lag-length`, is added to the `ourfuncs.lsp` file. It uses no functions other than the ones used in earlier work and the underlying programming procedures are described in annotations to the coding that creates it. The new function takes five arguments: the list of variables, a list of strings giving

⁶Christopher Sims, “Macroeconomics and Reality,” *Econometrica*, Vol.48, 1980, pp. 1-49.

⁷For a discussion of the above tests, see pages 312 through 314, and pages 396 and 397 of the Enders book cited previously.

names to the variables, the maximum lag length, the datelist for the variables, the starting date (which must be consistent with the maximum selected number of lags) and the ending date. Essentially, matrices are constructed for each variable for the maximum number of lags and then bound together to obtain the matrix of independent variables (other than the constant term) for the first set of regressions. After running those regression, the number of lags in each of these matrices is reduced by one and they are bound together for the second set of regressions. The variance-covariance matrices of error terms for both these sets of regressions are calculated and the likelihood-ratio statistic above is then calculated to test the null-hypothesis that the eliminated lag makes no contribution. The AIC and SBC statistics for that lower lag level are also calculated. The number of lags is then successively reduced by one lag and the same calculations are made. Finally, the results are printed to the screen in tabular form.

As an example, let us consider the three cointegrated variables in the regression of the logarithm Canada's real exchange rate with respect to the United States (`LREXCAUS`) on the logarithm of world energy prices relative to U.S. traded goods prices (`LRPENENERGY`), and the excess of the net capital inflow into Canada plus the debt service balance as a fraction of GDP over the net capital inflow into the United States plus the debt service balance as a fraction of that country's GDP (`DNCICAUS`).⁸ Here we use an expanded data set that includes the years 2008 through 2010. First we must establish that for the longer data set the logarithms of the prices of commodities excluding energy divided by U.S. traded goods prices is stationary and the other three variables are not. The code for setting up the variables, checking them for stationarity and then running the regression noted above is layed out below.

We start using `XLispStat`⁹ The data are in the Excel file `causqdat.xls` and the `Gretl` file `causqdat.gdt`, where they and their sources are described in detail, and in a tabular text file `causqdat.tab` with a `;` character placed in front of the first line so that the `XLispStat` function `read-data-columns`, which reads columns of data and puts them in lists, will not read that line. In the `johcoint.lsp` input file we load in the data and then save them in a

⁸These variables were used in a Canada/U.S. real exchange rate regression and tested for cointegration in chapter 10 of my book, John E. Floyd, *Interest Rates, Exchange Rates and World Monetary Policy*, Springer-Verlag, 2010.

⁹This code is for doing these calculations in `XLispStat` is in the file `johcoint.lsp` and the output generated by that file is in the file `johcoint.lou`.

more accessible form in the file `causqdat.lsp`. The code for setting up the variables and doing the test is as follows. All the data we will be using here conform to the datelist `dates72` and run from 1972Q1 through 2010Q4.

```

; JOHANSEN COINTEGRATION TESTS
;
(load "ourfuncs.lsp")
(load "causqdat.lsp")
;
; Set up the real exchange rate, nominal exchange rate and price level
; ratio variables on a 1975 base
;
(def rexcaus (base (/ cpica (* exrcaus cpius)) dates72 1975.00 4)) ; real
; exchange rate
(def nexcaus (base (/ 1 exrcaus) dates72 1975.00 4)) ; nominal exchange
; rate index
(def nprcaus (base (/ cpica cpius) dates72 1975.00 4)) ; price level ratio
;
; Set up variables for the real exchange rate regression
;
(def rexcaus (base rexcaus dates72 1985.00 4)) ; shift the base of the
; real exchange rate
(def ptgdsus (+ (* .5 pexpus)(* .5 pimpus))) ; price of U.S. traded
; goods in U.S. dollars
(def rpcomm (* 100 (/ pcomm ptgdsus))) ; real commodity prices
;
(def rpcomxen (* 100 (/ pcomxen ptgdsus))) ; real prices of commodities
; excludeing energy
(def rpenergy (* 100 (/ penergy ptgdsus))) ; real energy prices
;
(def nciyca (* 100 (/ (- impgsca expgsca) gdpca))) ; net capital inflows
; of Canada and the
(def nciyus (* 100 (/ (- impgsus expgsus) gdpus))) ; U.S. as percentages
; of their GDPs
(def dncicaus (- nciyca nciyus)) ; Canadian minus U.S. net capital inflows
; as percentages of GDP

```

```

(def lrexcaus (log rexcaus)) ; take logarithms of real exchange rate
(def lrpenenergy (log rpenenergy)) ; and energy prices and
(def lrpcmxen (log rpcmxen)) ; commodity prices excluding energy
;
; Run the main regression on page 159 of my cited book for the longer period
; using the remove-first function to start at 1974Q1 rather than 1972Q1
;
(def regressand "LREXCAUS")
(def regressors (list "Constant" "LRPCOMXEN" "LRPENENERGY" "DNCICAUS"))
(OLSreg (bind-columns (remove-first 8 lrexcaus))
(bind-columns (remove-first 8 lrpcmxen)
(remove-first 8 lrpenenergy)(remove-first 8 dncicaus)) 1)
;
; Test above variables for stationarity using the DFlags function to
; choose the appropriate lag lengths
;
(DFlags lrexcaus 8)
(dfunit lrexcaus "LREXCAUS" 3 4)
;
(DFlags dncicaus 8)
(dfunit dncicaus "DNCICAUS" 0 1)
;
(DFlags lrpenenergy 8)
(dfunit lrpenenergy "LRPENENERGY" 4 5)
;
(DFlags lrpcmxen 8)
(dfunit lrpcmxen "LRPCOMXEN" 4 5)
;

```

```

; Run the above regression without the commodity price variable, which
;   is stationary, and check for the appropriate lag length for the
;   three remaining variables for the Johansen Cointegration Test.
;
(def regressand "LREXCAUS")
(def regressors (list "Constant" "LRPENERGY" "DNCICAUS"))
(OLSreg (bind-columns (remove-first 8 lrexcaus))
(bind-columns (remove-first 8 lrpenergy)(remove-first 8 dncicaus)) 1)
;
(VAR-lag-length (list lrexcaus lrpenergy dncicaus)
(list "LREXCAUS" "LRPENERGY" "DNCICAUS") 12 dates72 1975.0 2010.75)
;

```

An appropriate selection of the relevant output is as follows.

ORDINARY LEAST SQUARES REGRESSION

Dependent Variable: LREXCAUS

	Coefficient	Std. Error	T-stat	P-Val
Constant	2.108	0.252	8.375	0.000
LRPCOMXEN	0.311	0.045	6.949	0.000
LRPENERGY	0.138	0.016	8.786	0.000
DNCICAUS	0.027	0.002	13.747	0.000

Number of Observations:	148
Degrees of Freedom:	144
R-Squared:	0.7206317588499688
Adjusted R-Squared:	0.7148115871593431
Sum of Squared Errors:	0.6039005016567894
F-Statistic:	123.81623724445541
P-Value	0.0

Output from DFlags function applied to REXCAUS
 AIC and SBC (first-two rows) and P-Values of up to 8 Dickey-Fuller Lags

AIC	1.000	1.005	1.005	1.005	1.010	1.008	1.003	1.002
SBC	1.000	1.015	1.024	1.033	1.048	1.055	1.058	1.067
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.021	0.021	0.018	0.010	0.010	0.021	0.074	
3	0.011	0.011	0.017	0.008	0.009	0.028		
4	0.066	0.069	0.102	0.052	0.053			
5	0.424	0.448	0.421	0.358				
6	0.232	0.204	0.078					
7	0.077	0.079						
8	0.372							

DICKEY-FULLER TEST --- LREXCAUS

Lags = 3

Starting observation = 4

Number of observations = 152

Coefficients:

Constant	0.108714	0.134332	
Trend	0.000027		
Y(t-1)	-0.023695	-0.028740	0.000011

t-Statistics:

Constant	1.198506	1.772832	
Trend	0.516741		
Y(t-1)	-1.249208	-1.771558	0.027299
Lagged (Y(t)-Y(t-1))	4.922507	5.071908	4.943823
.....	-2.136288	-2.097807	-2.255214
.....	1.814015	1.936582	1.714592

F-Statistics:

All Three Coefficients = 0	1.131679
Trend & Y(t-1) Coefs = 0	1.694895

Constant & Y(t-1) Coefs = 0 1.571845

Output from DFlags function applied to DNCICAUS
AIC and SBC (first-two rows) and P-Values of up to 8 Dickey-Fuller Lags

AIC	1.000	0.998	0.995	1.000	0.997	0.995	0.993	0.990
SBC	1.000	0.994	0.987	0.988	0.981	0.975	0.969	0.962
1	0.031	0.033	0.034	0.046	0.042	0.036	0.035	0.033
2	0.407	0.381	0.373	0.447	0.434	0.433	0.463	
3	0.337	0.320	0.323	0.244	0.243	0.225		
4	0.303	0.293	0.298	0.292	0.279			
5	0.302	0.281	0.281	0.365				
6	0.013	0.012	0.012					
7	0.409	0.426						
8	0.309							

DICKEY-FULLER TEST --- DNCICAUS

Lags = 0

Starting observation = 1

Number of observations = 155

Coefficients:

Constant	-0.122961	-0.141409		
Trend	-0.000408			
Y(t-1)	-0.041737	-0.038214	-0.016980	

t-Statistics:

Constant	-0.824736	-1.256080		
Trend	-0.189528			
Y(t-1)	-1.430637	-1.704823	-1.151411	

F-Statistics:

All Three Coefficients	= 0	0.975350
Trend & Y(t-1) Coefs	= 0	1.462014

Constant & Y(t-1) Coefs = 0 1.454229

Output from DFlags function applied to LRPENERGY
AIC and SBC (first-two rows) and P-Values of up to 8 Dickey-Fuller Lags

AIC	1.000	0.974	0.936	0.938	0.939	0.981	0.953	1.021
SBC	1.000	0.946	0.885	0.849	0.812	0.801	0.747	0.752
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.008	0.006	0.007	0.004	0.002	0.008	0.011	
3	0.236	0.184	0.172	0.154	0.080	0.221		
4	0.140	0.159	0.154	0.082	0.024			
5	0.042	0.036	0.038	0.083				
6	0.121	0.076	0.080					
7	0.494	0.363						
8	0.213							

DICKEY-FULLER TEST --- LRPENERGY

Lags = 4

Starting observation = 5

Number of observations = 151

Coefficients:

Constant	0.265272	0.233433	
Trend	0.000127		
Y(t-1)	-0.041133	-0.034772	0.000912

t-Statistics:

Constant	1.862016	1.749953	
Trend	0.645817		
Y(t-1)	-1.810946	-1.702371	0.762887
Lagged (Y(t)-Y(t-1))	4.956490	4.940432	4.771843
.....	-2.757524	-2.817849	-3.086849
.....	1.436420	1.416750	1.262485
.....	-1.970646	-2.027902	-2.305547

F-Statistics:

All Three Coefficients = 0	1.351650
Trend & Y(t-1) Coefs = 0	1.651747
Constant & Y(t-1) Coefs = 0	1.826278

Output from DFlags function applied to LRPCOMXEN

AIC and SBC (first-two rows) and P-Values of up to 8 Dickey-Fuller Lags

AIC	1.000	1.003	1.012	0.998	0.994	0.987	0.964	0.948
SBC	1.000	1.019	1.045	1.044	1.054	1.062	1.050	1.048
1	0.057	0.057	0.054	0.021	0.033	0.013	0.044	0.079
2	0.089	0.056	0.052	0.026	0.012	0.004	0.011	
3	0.008	0.012	0.011	0.023	0.010	0.004		
4	0.065	0.050	0.046	0.021	0.036			
5	0.067	0.089	0.085	0.043				
6	0.021	0.015	0.014					
7	0.466	0.477						
8	0.142							

DICKEY-FULLER TEST --- LRPCOMXEN

Lags = 4

Starting observation = 5

Number of observations = 151

Coefficients:

Constant	0.380220	0.395127	
Trend	0.000019		
Y(t-1)	-0.067776	-0.070158	-0.000320

t-Statistics:

Constant	2.764946	3.300247	
Trend	0.222691		
Y(t-1)	-2.850050	-3.314167	-0.547734
Lagged (Y(t)-Y(t-1))	2.137948	2.201759	2.003901
.....	-1.965614	-1.963361	-2.254152
.....	2.390607	2.452181	2.228327
.....	1.753438	1.809000	1.430369

F-Statistics:

All Three Coefficients = 0	3.729357
Trend & Y(t-1) Coefs = 0	5.480650
Constant & Y(t-1) Coefs = 0	5.605984

ORDINARY LEAST SQUARES REGRESSION

Dependent Variable: LREXCAUS

	Coefficient	Std. Error	T-stat	P-Val
Constant	3.720	0.113	33.029	0.000
LRPENRGY	0.163	0.018	9.231	0.000
DNCICAUS	0.032	0.002	15.401	0.000

Number of Observations:	148
Degrees of Freedom:	145
R-Squared:	0.6269536046006158
Adjusted R-Squared:	0.6218081370778656
Sum of Squared Errors:	0.8064012730851527
F-Statistic:	121.84579959519874
P-Value	0.0

LAG SELECTION INFORMATION

Variables are (LREXCAUS LRPENERGY DNCICAUS)

Lag	Degrees of Freedom Per Equation	Data Points Less All Parameters	LR-Test PVal	AIC	BIC
12	107	321		-1738.77	-1409.12
11	110	330	0.065	-1747.04	-1444.12
10	113	339	0.066	-1755.63	-1479.43
9	116	348	0.260	-1768.51	-1519.05
8	119	357	0.100	-1778.74	-1556.00
7	122	366	0.006	-1781.72	-1585.71
6	125	375	0.001	-1779.45	-1610.17
5	128	384	0.000	-1775.54	-1632.99
4	131	393	0.003	-1778.12	-1662.30
3	134	402	0.001	-1778.83	-1689.74
2	137	411	0.005	-1783.10	-1720.73
1	140	420	0.000	-1758.34	-1722.70

The P-Values denote the statistical significance of the next higher lag.

When doing Johansen cointegration tests 2 additional degrees of freedom per equation will be lost, reducing the data points minus total parameters estimated by 6

As can be seen by comparing the relevant t-statistic with the critical values in the first of the tables at the end of this document, the commodity price less energy variable `RPCOMXEN` is stationary around constant drift at almost the 1% level. And the appropriate lags for our Johansen cointegration test for the three non-stationary variables are either the 8 suggested by the likelihood ratio test or the 2 suggested by the AIC or no lags at all as suggested by the BIC. Actually, the AIC comes within a hair of also suggesting 8 lags and it therefore would seem reasonable to run the Johansen test using 8 lags.

The next task is to set up our function in `XLispStat` to perform Johansen cointegration tests. We call the function `Johansen-coint` and set it up to take five arguments in the following order: the list of variables being tested, a

list giving their names for the printout, the number of of lags to be used, the datelist for the variables, the beginning date of the test (which must be sufficiently far along in the datelist to allow for the number of lags chosen) and, finally, the ending date for the test. The function is included in `ourfuncs.lsp` with appropriate comments added to make it understandable—little would be gained by working through the code here. The coding simply follows the procedure outlined by Hamilton described above for the case where deterministic trends are allowed. After setting up the variables, the two auxiliary regressions (4) and (5) are run and the variance-covariance matrices are calculated and used to create the matrix (10) for which the eigenvalues and eigenvectors are then obtained. At this point, the function uses two `XLispStat` functions that we have not yet used—`eigenvalues` and `eigenvectors`. Both these functions are fed the matrix (10) with the first function returning a vector of eigenvalues, in decreasing order, and the second a list of the corresponding eigenvectors. A problem here is that these two functions are specified as returning eigenvalues and eigenvectors for symmetric matrices while the matrix we apply them to here is not symmetric.¹⁰ Although the functions work and produce results very similar, as will be shown below, to those produced by `Gretl` and `R`, this is nevertheless a matter of concern.

Given this concern and the resulting programming complexities together the fact that our purpose is simply to determine whether the variables in a regression are cointegrated and the regression is therefore not spurious, and not to conduct sophisticated cointegration analysis involving short-run adjustments to long-run equilibrium, our function calculates Trace and L-Max statistics only for the case in which no restrictions are placed on the presence of deterministic trends. It turns out that if deterministic trends are indeed present, the critical values are those in the top panel of the third page of the table at the end of this document and if deterministic trends are not actually present, the critical values are in the bottom table on that page. It turns out that the critical values in the two tables are quite similar—indeed, a conservative approach is to simply compare the Trace and L-Max statistics to the critical values in the lower table, which tend to be very slightly higher than the critical values in the upper table. It should also be noted that deterministic trends here refer to the presence of constants in the auxiliary regressions. The procedure available in `Gretl` also enables a trend that is

¹⁰See Luke Tierney’s book, *LISP-STAT*, Wiley Series in Probability and Mathematical Statistics, 1990, page 169.

increasing through time to be incorporated by including a time variable in the auxiliary regression equations. If cointegration, or lack of it, is not clearly established by the application of our function in the fashion suggested above, a more sophisticated analysis can be undertaken using `Gretl`. Indeed, all important results should be verified by using two statistical programs in any case, thereby substantially reducing the possibility of mechanical errors by requiring that any surviving error be programmed twice under somewhat different frameworks.

Our `XLispStat` code for calling our Johansen-coint function using eight lags is simply

```
(Johansen-coint (list lrexcaus lrpenergy dncicaus)
(list "LREXCAUS" "LRPENERGY" "DNCICAUS") 8 dates72 1974.25 2010.75)
```

and the output is as follows.

JOHANSEN COINTEGRATION TEST

Variables are (LREXCAUS LRPENERGY DNCICAUS)

Number of Lags = 8

Period of Test --- 1974.0 to 2010.75

Deterministic trends are included

Coint Vectors Under Null	Eigen- Values	L-max	Trace
0.000	0.143	22.815	36.623
1.000	0.079	12.167	13.808
2.000	0.011	1.641	1.641

Eigenvectors together with the associated variables

LREXCAUS	1.000	1.000	1.000
LRPENRGY	-0.138	-3.689	6.946
DNCICAUS	-0.050	30.445	0.809

As can be seen from the bottom table of the last page of our last page of tables at the end of this document, we can conclude that the three variables are cointegrated at the 1% level according to both the L-Max and Trace tests.

The code for doing an equivalent analysis in Gretl is presented below. The input file is `johcoint.inp`, the data file is `causqdat.gdt` as noted above and the output file is `johcoint.gou`.

```
# JOHANSEN COINTEGRATION TESTS IN GRETl
#
open E:\DSLHTML\SPURREG\causqdat.gdt
#
# Set up the real exchange rate, nominal price ratio and nominal
# exchange rate series for plotting
#
genr REXCAUS = 100*CPICA/(EXRCAUS*CPIUS)
#
# Now put the real exchange rate on a 1985 base
#
smpl 1985:1 1985:4
genr REX1985 = sum(REXCAUS)/4
#
smpl 1957:1 2010:4
genr REXCAUS = 100*REXCAUS/REX1985
#
# Calculate real commodity less energy and energy prices
#
genr PTGDSUS = 0.5*PEXPUS + 0.5*PIMPUS
genr RPCOMM = 100*PCOMM/PTGDSUS
genr RPCOMXEN = 100*PCOMXEN/PTGDSUS
genr RPENERGY = 100*PENRGY/PTGDSUS
#
```

```

# Take logarithms of the relevant variables
#
logs REXCAUS RPCOMXEN RPENERGY
#
# Calculate the net capital inflow variable
#
genr NCIYCA = 100*(IMPGSCA-EXPGSCA)/GDPCA
genr NCIYUS = 100*(IMPGSUS-EXPGSUS)/GDPUS
genr DNCICAUS = NCIYCA - NCIYUS
#
# Run the basic regression and then test the variables
#   for stationarity
#
smpl 1974:1 2010:4
#
ols l_REXCAUS const l_RPCOMXEN l_RPENERGY DNCICAUS
#
adf 8 l_REXCAUS --nc --c --ct --verbose --test-down
adf 8 l_RPENERGY --nc --c --ct --verbose --test-down
adf 8 l_RPCOMXEN --nc --c --ct --verbose --test-down
adf 8 DNCICAUS --nc --c --ct --verbose --test-down
#
var 12 l_REXCAUS l_RPENERGY DNCICAUS --lagsselect
#
# Run the regression with non-stationary variables only
#   and then perform the Johansen cointegration tests
#
ols l_REXCAUS const l_RPENERGY DNCICAUS
#
coint2 8 l_REXCAUS l_RPENERGY DNCICAUS --ct
coint2 8 l_REXCAUS l_RPENERGY DNCICAUS
coint2 8 l_REXCAUS l_RPENERGY DNCICAUS --rc
coint2 8 l_REXCAUS l_RPENERGY DNCICAUS --nc

```

Notice that the Johansen cointegration test, which is performed by the function `coint2` in `Gretl`, has a complete set of options, unlike the test we programmed in `XLispStat`. The default option is to incorporate an unrestricted constant, the procedure we followed in `XLispStat`. By adding `--rc` to the end of the code line we can instruct it to include no deterministic trend but put a constant in the cointegrating vector. By adding `--nc` we can instruct it to include no constants in the auxiliary equations and thereby to include no trend. Alternatively, we can add `--ct` to the code line and have `Gretl` include both a constant and trend in the auxiliary equations to allow for increasing or decreasing trend, an option that is not discussed by Hamilton. Obviously, `Gretl` is the best program to use to cover all options. In addition, it produces P-Values for all the tests so we do not need to look things up in statistical tables. An edited version of the Johansen test results is presented below following the relevant preliminaries, with the sophisticated results used in fancy Johansen analyses removed.

```
#
? ols l_REXCAUS const l_RPCOMXEN l_RPENERGY DNCICAUS
```

Model 1: OLS, using observations 1974:1-2010:4 (T = 148)
 Dependent variable: l_REXCAUS

	coefficient	std. error	t-ratio	p-value	
const	2.10812	0.251710	8.375	4.46e-014	***
l_RPCOMXEN	0.311019	0.0447585	6.949	1.18e-010	***
l_RPENERGY	0.138481	0.0157618	8.786	4.21e-015	***
DNCICAUS	0.0266766	0.00194054	13.75	5.38e-028	***
Mean dependent var	4.665032	S.D. dependent var		0.121265	
Sum squared resid	0.603901	S.E. of regression		0.064759	
R-squared	0.720632	Adjusted R-squared		0.714812	
F(3, 144)	123.8162	P-value(F)		1.09e-39	
Log-likelihood	197.1124	Akaike criterion		-386.2248	
Schwarz criterion	-374.2359	Hannan-Quinn		-381.3538	
rho	0.896679	Durbin-Watson		0.207332	

Log-likelihood for REXCAUS = -493.312

```
#  
? adf 8 l_REXCAUS --nc --c --ct --verbose --test-down
```

```
Augmented Dickey-Fuller test for l_REXCAUS  
including 4 lags of (1-L)l_REXCAUS  
sample size 148  
unit-root null hypothesis: a = 1
```

```
test without constant  
model: (1-L)y = (a-1)*y(-1) + ... + e  
1st-order autocorrelation coeff. for e: 0.003  
lagged differences: F(4, 143) = 7.171 [0.0000]  
estimated value of (a - 1): 1.39806e-005  
test statistic: tau_nc(1) = 0.0329822  
asymptotic p-value 0.6934
```

```
test with constant  
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e  
1st-order autocorrelation coeff. for e: 0.004  
lagged differences: F(4, 142) = 7.360 [0.0000]  
estimated value of (a - 1): -0.025602  
test statistic: tau_c(1) = -1.51885  
asymptotic p-value 0.5242
```

```
with constant and trend  
model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e  
1st-order autocorrelation coeff. for e: 0.001  
lagged differences: F(4, 141) = 7.002 [0.0000]  
estimated value of (a - 1): -0.01697  
test statistic: tau_ct(1) = -0.868132  
asymptotic p-value 0.9579
```

```
? adf 8 l_RPENERGY --nc --c --ct --verbose --test-down
```

```
Augmented Dickey-Fuller test for l_RPENERGY  
including 4 lags of (1-L)l_RPENERGY  
sample size 148  
unit-root null hypothesis: a = 1
```

```
test without constant  
model:  $(1-L)y = (a-1)*y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: -0.022  
lagged differences:  $F(4, 143) = 7.160$  [0.0000]  
estimated value of (a - 1): 0.000932219  
test statistic:  $\tau_{nc}(1) = 0.765346$   
asymptotic p-value 0.879
```

```
test with constant  
model:  $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: -0.015  
lagged differences:  $F(4, 142) = 7.111$  [0.0000]  
estimated value of (a - 1): -0.0370146  
test statistic:  $\tau_c(1) = -1.75586$   
asymptotic p-value 0.403
```

```
with constant and trend  
model:  $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: -0.014  
lagged differences:  $F(4, 141) = 7.077$  [0.0000]  
estimated value of (a - 1): -0.0422319  
test statistic:  $\tau_{ct}(1) = -1.82833$   
asymptotic p-value 0.6911
```

```
? adf 8 l_RPCOMXEN --nc --c --ct --verbose --test-down
```

```
Augmented Dickey-Fuller test for l_RPCOMXEN  
including 6 lags of (1-L)l_RPCOMXEN  
sample size 148  
unit-root null hypothesis: a = 1
```

```
test without constant  
model:  $(1-L)y = (a-1)y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: -0.005  
lagged differences:  $F(6, 141) = 4.807$  [0.0002]  
estimated value of  $(a - 1)$ : -0.00055061  
test statistic:  $\tau_{nc}(1) = -0.968449$   
asymptotic p-value 0.298
```

```
test with constant  
model:  $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: 0.002  
lagged differences:  $F(6, 140) = 4.589$  [0.0003]  
estimated value of  $(a - 1)$ : -0.0859545  
test statistic:  $\tau_c(1) = -3.75159$   
asymptotic p-value 0.003462
```

```
with constant and trend  
model:  $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: -0.001  
lagged differences:  $F(6, 139) = 4.573$  [0.0003]  
estimated value of  $(a - 1)$ : -0.0752582  
test statistic:  $\tau_{ct}(1) = -3.03535$   
asymptotic p-value 0.1225
```

```
? adf 8 DNCICAUS --nc --c --ct --verbose --test-down
```

```
Augmented Dickey-Fuller test for DNCICAUS  
including 6 lags of (1-L)DNCICAUS  
sample size 148  
unit-root null hypothesis: a = 1
```

```
test without constant  
model:  $(1-L)y = (a-1)*y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: 0.005  
lagged differences:  $F(6, 141) = 1.530$  [0.1726]  
estimated value of  $(a - 1)$ : -0.0180835  
test statistic:  $\tau_{nc}(1) = -1.2001$   
asymptotic p-value 0.2112
```

```
test with constant  
model:  $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: 0.003  
lagged differences:  $F(6, 140) = 1.633$  [0.1424]  
estimated value of  $(a - 1)$ : -0.0458741  
test statistic:  $\tau_c(1) = -1.90806$   
asymptotic p-value 0.3289
```

```
with constant and trend  
model:  $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$   
1st-order autocorrelation coeff. for e: -0.000  
lagged differences:  $F(6, 139) = 1.693$  [0.1271]  
estimated value of  $(a - 1)$ : -0.0623213  
test statistic:  $\tau_{ct}(1) = -1.76106$   
asymptotic p-value 0.7237
```

```
? var 12 l_REXCAUS l_RPENERGY DNCICAUS --lagselect
VAR system, maximum lag order 12
```

The asterisks below indicate the best (that is, minimized) values of the respective information criteria, AIC = Akaike criterion, BIC = Schwartz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags	loglik	p(LR)	AIC	BIC	HQC
1	278.18879		-3.697067	-3.449582*	-3.596503
2	299.56717	0.00000	-3.868988*	-3.435891	-3.693002*
3	306.43594	0.13197	-3.839388	-3.220677	-3.587979
4	315.07956	0.04440	-3.834438	-3.030114	-3.507606
5	322.78924	0.08004	-3.816517	-2.826579	-3.414262
6	333.74163	0.00919	-3.843634	-2.668083	-3.365956
7	343.87810	0.01630	-3.859418	-2.498254	-3.306318
8	351.38738	0.09043	-3.838714	-2.291936	-3.210190
9	355.27354	0.55724	-3.767688	-2.035297	-3.063742
10	357.83178	0.82404	-3.678219	-1.760215	-2.898850
11	362.53832	0.40005	-3.618588	-1.514970	-2.763796
12	367.40508	0.37249	-3.561182	-1.271951	-2.630967

```
#
# Run the regression with non-stationary variables only
# and then perform the Johansen cointegration tests
#
? ols l_REXCAUS const l_RPENERGY DNCICAUS
```

Model 2: OLS, using observations 1974:1-2010:4 (T = 148)
Dependent variable: l_REXCAUS

	coefficient	std. error	t-ratio	p-value
const	3.71980	0.112621	33.03	2.38e-069 ***
l_RPENERGY	0.163216	0.0176819	9.231	3.00e-016 ***
DNCICAUS	0.0318174	0.00206589	15.40	2.54e-032 ***

Mean dependent var	4.665032	S.D. dependent var	0.121265
Sum squared resid	0.806401	S.E. of regression	0.074575
R-squared	0.626954	Adjusted R-squared	0.621808
F(2, 145)	121.8458	P-value(F)	8.97e-32
Log-likelihood	175.7137	Akaike criterion	-345.4273
Schwarz criterion	-336.4357	Hannan-Quinn	-341.7741
rho	0.876785	Durbin-Watson	0.224324

Log-likelihood for REXCAUS = -514.711

? coint2 8 1_REXCAUS 1_RPENERGY DNCICAUS --ct

Johansen test:

Number of equations = 3

Lag order = 8

Estimation period: 1974:1 - 2010:4 (T = 148)

Case 5: Unrestricted trend and constant

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.15832	35.125	[0.0488]	25.508	[0.0316]
1	0.042581	9.6168	[0.5254]	6.4401	[0.7721]
2	0.021236	3.1768	[0.0747]	3.1768	[0.0747]

eigenvalue	0.15832	0.042581	0.021236
------------	---------	----------	----------

? coint2 8 1_REXCAUS 1_RPENERGY DNCICAUS

Johansen test:

Number of equations = 3

Lag order = 8

Estimation period: 1974:1 - 2007:4 (T = 136)

Case 3: Unrestricted constant

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.14876	34.827	[0.0113]	21.904	[0.0368]
1	0.079606	12.924	[0.1179]	11.282	[0.1420]
2	0.012001	1.6420	[0.2001]	1.6420	[0.2001]

eigenvalue 0.14876 0.079606 0.012001

? coint2 8 1_REXCAUS 1_RPENERGY DNCICAUS --rc

Johansen test:

Number of equations = 3

Lag order = 8

Estimation period: 1974:1 - 2007:4 (T = 136)

Case 2: Restricted constant

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.14885	37.056	[0.0294]	21.918	[0.0546]
1	0.079833	15.137	[0.2233]	11.315	[0.2374]
2	0.027713	3.8221	[0.4512]	3.8221	[0.4502]

eigenvalue 0.14885 0.079833 0.027713

```
? coint2 8 1_REXCAUS 1_RPENERGY DNCICAUS --nc
Johansen test:
Number of equations = 3
Lag order = 8
Estimation period: 1974:1 - 2007:4 (T = 136)
Case 1: No constant
```

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.080391	18.777	[0.2143]	11.398	[0.3592]
1	0.042155	7.3796	[0.2906]	5.8574	[0.3740]
2	0.011130	1.5222	[0.2547]	1.5222	[0.2539]

```
eigenvalue    0.080391    0.042155    0.011130
```

The three stationary variables are cointegrated at the 5% level or better in all cases except the one with no constant.

Finally, we need to have a go at performing Johansen cointegration tests in the statistical program R. We use the same tabular data file as was used in the case of `XLispStat`. Although sophisticated Johansen cointegration analysis of the form incorporated in `Gretl` must surely be present somewhere in the myriad of available applications of R, the only one we really need, given what is available in `Gretl`, is the function `ca.jo` that is included in the library `urca`, the package containing which can easily be downloaded off the Web and installed. The appropriate R code, which is in the file `johcoint.R` and produces the output `johcoint.Rou`, is as follows.

```
# R SCRIPT FOR INVESTIGATING THE STATIONARITY OF THE REAL EXCHANGE RATES
#
causdata <- read.table("causqdat.tab",header=TRUE)
names(causdata)
#
attach(causdata)
```

```

#Set up the relevant variables in the object rexdata as time-series
#
CPIUS <- ts(causdata$CPIUS,start=c(1972,1),end=c(2010,4),frequency=4)
CPICA <- ts(causdata$CPICA,start=c(1972,1),end=c(2010,4),frequency=4)
EXRCAUS <- ts(causdata$EXRCAUS,start=c(1972,1),end=c(2010,4),frequency=4)
PEXPUS <- ts(causdata$PEXPUS,start=c(1972,1),end=c(2010,4),frequency=4)
PIMPUS <- ts(causdata$PIMPUS,start=c(1972,1),end=c(2010,4),frequency=4)
PENEFY <- ts(causdata$PENERGY,start=c(1972,1),end=c(2010,4),frequency=4)
PCOMXEN <- ts(causdata$PCOMXEN,start=c(1972,1),end=c(2010,4),frequency=4)
PCOMM <- ts(causdata$PCOMM,start=c(1972,1),end=c(2010,4),frequency=4)
PCROIL <- ts(causdata$PCROIL,start=c(1972,1),end=c(2010,4),frequency=4)
EXPGSUS <- ts(causdata$EXPGSUS,start=c(1972,1),end=c(2010,4),frequency=4)
IMPGSUS <- ts(causdata$IMPGSUS,start=c(1972,1),end=c(2010,4),frequency=4)
EXPGSCA <- ts(causdata$EXPGSCA,start=c(1972,1),end=c(2010,4),frequency=4)
IMPGSCA <- ts(causdata$IMPGSCA,start=c(1972,1),end=c(2010,4),frequency=4)
GDPUS <- ts(causdata$GDPUS,start=c(1972,1),end=c(2010,4),frequency=4)
GDPCA <- ts(causdata$GDPCA,start=c(1972,1),end=c(2010,4),frequency=4)
IPDUS <- ts(causdata$IPDUS,start=c(1972,1),end=c(2010,4),frequency=4)
IPDCA <- ts(causdata$IPDCA,start=c(1972,1),end=c(2010,4),frequency=4)
#
# Set up the real exchange rate and put it on a 1985 base
#
REXCAUS <- 100*CPICA/(EXRCAUS*CPIUS)
BASEREX85 <- sum(REXCAUS[53:56])/4
REXCAUS <- 100*REXCAUS/BASEREX85
#
# Set up the commodity price excluding energy and energy price variables
#
PTGDSUS = 0.5*PEXPUS + 0.5*PIMPUS
RPENERGY = 100*PENERGY/PTGDSUS
RPCOMXEN = 100*PCOMXEN/PTGDSUS
RPCOMM = 100*PCOMM/PTGDSUS

```

```

# Take the logarithms of the relevant variables
#
LREXCAUS <- log(REXCAUS)
LRPCOMXEN <- log(RPCOMXEN)
LRPENERGY <- log(RPENERGY)
#
# Set up the real net capital inflow variable
#
NCIYCA <- 100*(IMPGSCA-EXPGSCA)/GDPCA
NCIYUS <- 100*(IMPGSUS-EXPGSUS)/GDPUS
DNCICAUS <- NCIYCA-NCIYUS
#
library(urca)
#
# 1974Q1 is observation 9
# 2010Q4 is observation 156
#
reg1 <- lm(LREXCAUS[9:156] ~ LRPCOMXEN[9:156] + LRPENERGY[9:156]
+ DNCICAUS[9:156])
summary(reg1)
#
# Do unit root tests on the variables
#
summary(ur.df(LREXCAUS,lags=3,type="trend"))
summary(ur.df(LREXCAUS,lags=3,type="drift"))
summary(ur.df(LREXCAUS,lags=3,type="none"))
#
summary(ur.df(LRPCOMXEN,lags=4,type="trend"))
summary(ur.df(LRPCOMXEN,lags=4,type="drift"))
summary(ur.df(LRPCOMXEN,lags=4,type="none"))
#
summary(ur.df(LRPENERGY,lags=4,type="trend"))
summary(ur.df(LRPENERGY,lags=4,type="drift"))
summary(ur.df(LRPENERGY,lags=4,type="none"))
#
summary(ur.df(DNCICAUS,lags=0,type="trend"))
summary(ur.df(DNCICAUS,lags=0,type="drift"))
summary(ur.df(DNCICAUS,lags=0,"AIC",type="none"))

```

```

#
summary(ur.df(DNCICAUS,selectlags = "AIC",type="trend"))
summary(ur.df(DNCICAUS,selectlags = "AIC",type="drift"))
summary(ur.df(DNCICAUS,selectlags = "AIC",type="none"))
#
#
# Run the Johansen Cointegration Test
#
LREX <- LREXCAUS[9:156]
DNCI <- DNCICAUS[9:156]
LPCM <- LRPCOMXEN[9:156]
LPEN <- LRPENERGY[9:156]
#
TESTMAT <- cbind(LREX,DNCI,LPEN)
#
JOHRES1 <- ca.jo(TESTMAT,type ="eigen",K=8,
spec="longrun",season=NULL,dumvar=NULL)
summary(JOHRES1)
#
JOHRES2 <- ca.jo(TESTMAT,type ="trace",K=8,
spec="longrun",season=NULL,dumvar=NULL)
summary(JOHRES2)
#
#q()

```

The relevant results are as follows.

```

> # 1974Q1 is observation 9
> # 2007Q4 is observation 144
> # 2010Q4 is observation 156
> #

```

```
> reg1 <- lm(LREXCAUS[9:156] ~ LRPCOMXEN[9:156] + LRPENERGY[9:156]
+ DNCICAUS[9:156])
> summary(reg1)
```

Call:

```
lm(formula = LREXCAUS[9:156] ~ LRPCOMXEN[9:156] + LRPENERGY[9:156] +
    DNCICAUS[9:156])
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.141107	-0.054579	0.002012	0.046212	0.135159

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.108121	0.251710	8.375	4.46e-14	***
LRPCOMXEN[9:156]	0.311019	0.044759	6.949	1.18e-10	***
LRPENERGY[9:156]	0.138481	0.015762	8.786	4.21e-15	***
DNCICAUS[9:156]	0.026677	0.001941	13.747	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06476 on 144 degrees of freedom

Multiple R-squared: 0.7206, Adjusted R-squared: 0.7148

F-statistic: 123.8 on 3 and 144 DF, p-value: < 2.2e-16

```

> summary(ur.df(LRPCOMXEN,lags=4,type="drift"))

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
      Min       1Q   Median       3Q      Max
-0.1180298 -0.0245403 -0.0002287  0.0225490  0.1170614

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.39513    0.11973   3.300  0.00122 **
z.lag.1     -0.07016    0.02117  -3.314  0.00116 **
z.diff.lag1  0.17451    0.07926   2.202  0.02926 *
z.diff.lag2 -0.15649    0.07971  -1.963  0.05152 .
z.diff.lag3  0.19514    0.07958   2.452  0.01539 *
z.diff.lag4  0.14572    0.08055   1.809  0.07252 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.03906 on 145 degrees of freedom
Multiple R-squared:  0.1623,    Adjusted R-squared:  0.1335
F-statistic:  5.62 on 5 and 145 DF,  p-value: 9.138e-05

Value of test-statistic is: -3.3142 5.606

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81

```

```

> #
> LREX <- LREXCAUS[9:156]
> DNCI <- DNCICAUS[9:156]
> LPCM <- LRPCOMXEN[9:156]
> LPEN <- LRPENERGY[9:156]
> #
> TESTMAT <- cbind(LREX,DNCI,LPEN)
> #
> JOHRES1 <- ca.jo(TESTMAT,type ="eigen",K=8,
+ spec="longrun",season=NULL,dumvar=NULL)
> summary(JOHRES1)

#####
# Johansen-Procedure #
#####

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
[1] 0.14338390 0.05809768 0.01373048

Values of teststatistic and critical values of test:

      test 10pct  5pct  1pct
r <= 2 |  1.94  6.50  8.18 11.65
r <= 1 |  8.38 12.91 14.90 19.19
r = 0  | 21.67 18.90 21.07 25.75

```

```

> #
> JOHRES2 <- ca.jo(TESTMAT,type ="trace",K=8,
+ spec="longrun",season=NULL,dumvar=NULL)
> summary(JOHRES2)

#####
# Johansen-Procedure #
#####

```

```
Test type: trace statistic , with linear trend
```

```
Eigenvalues (lambda):
[1] 0.14338390 0.05809768 0.01373048
```

```
Values of teststatistic and critical values of test:
```

	test	10pct	5pct	1pct
r <= 2	1.94	6.50	8.18	11.65
r <= 1	10.32	15.66	17.95	23.52
r = 0	31.98	28.71	31.52	37.22

It turns out that non-stationarity of the price variable for commodities excluding energy can only be rejected at more than the 10% level, given the lags R chooses using the AIC criteria. But when we assign it the 8 lags used in `XLispStat` and `Gretl` we can conclude that it is stationary and omit it from the Johansen test. In doing the Johansen test it is necessary to instruct R to calculate the relevant Trace statistics and L-Max (`lambda`) statistics separately. Rather than presenting P-Values, R prints out the statistic followed by the relevant critical values. It is clear from a look at our tables that R is using the critical values in the bottom table even though it is incorporating deterministic trends in its calculations. This is precisely what we did in our `XLispStat` analysis. Using eight lags only, we verify cointegration at the 5% level in both the Trace and L-Max tests.

In concluding this discussion, it should be noted that non-stationarity of the variables in the cointegration tests above would seem to be very unlikely and stationarity should therefore be easily established if we have data for sufficiently long time periods. Since non-stationary variables have no upper

or lower limits when the time period is long enough, it is difficult to imagine a country's real exchange rate heading towards zero or infinity in the long-run. The same should be true of real income and price levels for countries other than ones like Zimbabwe (at the time of this writing), although stationarity would be expected to occur around trend for these variables. The reason why we cannot establish statistically that many series are stationary is that we do not have available sufficiently long time periods of data. If our theoretical analysis gives us a clear expectation of the signs of the variables in a regression, and we are therefore not on a fishing expedition that engages in post-hoc theorizing, and four or more independent variables are significant and have the expected signs, it would seem highly improbable that the regression is spurious. Yet even under these circumstances, it is desirable to conduct appropriate Johansen cointegration tests to cover ourselves against the possibility that a segment of the profession will reject our results on the grounds that such tests were not performed.

Statistical Tables

While the simplest way to calculate P-Values is to calculate the relevant cumulative densities using Gretl, R, XLispStat, or other statistical software, the test-statistics for unit root and cointegration tests do not follow standard distributions. Accordingly, the next three pages contain the relevant statistical tables for Dickey-Fuller and Phillips-Perron unit root tests, for cointegration tests based on unit root tests of regression residuals, and for Johansen cointegration tests.

The critical values for the unit root tests in the table that follows were calculated using Monte Carlo methods by David Dickey and Wayne A. Fuller and were obtained from their paper “Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root,” *Econometrica*, Vol. 49, July 1981, pages 1062 and 1063, and from Walter Enders, *Applied Econometric Time Series*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, 1995, pages 223, 419 and 421, and from James D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994, page 763. These critical values are based on sample sizes of 100 and remain unchanged when the Dickey-Fuller estimating equations are augmented by inclusion of lagged values of Δy_t to improve the fit as indicated by the AIC and SBC. Larger sample sizes will result in critical values that are slightly smaller in absolute value and smaller sample sizes will result in somewhat larger critical values.

STATISTICAL TABLES FOR UNIT ROOT TESTS

True Model Used to Generate the Data: $y_t = y_{t-1} + \epsilon_t$

1. Model Estimated:		Dickey-Fuller	$\Delta y_t = a_0 + a_1 y_{t-1} + a_2 t + \epsilon_t$		
		Phillips-Perron	$y_t = \tilde{a}_0 + \tilde{a}_1 y_{t-1} + \tilde{a}_2 (t - n/2) + \tilde{\epsilon}_t$		
Hypothesis		Test Statistic	Critical Values		
			10%	5%	1%
$a_1 = 0, \tilde{a}_1 = 1$		t-based	-3.15	-3.45	-4.04
$a_0 = 0, \tilde{a}_0 = 0$		t-based	2.73	3.11	3.78
$a_2 = 0, \tilde{a}_2 = 0$		t-based	2.38	2.79	3.53
$a_1 = a_2 = 0, \tilde{a}_1 = 1 \text{ \& } \tilde{a}_2 = 0$		F-based	5.47	6.49	8.73
$a_0 = a_1 = a_2 = 0, \tilde{a}_0 = \tilde{a}_2 = 0 \text{ \& } \tilde{a}_1 = 1$		F-based	4.16	4.88	6.50
2. Model Estimated:		Dickey-Fuller	$\Delta y_t = a_0 + a_1 y_{t-1} + \epsilon_t$		
		Phillips-Perron	$y_t = \tilde{a}_0 + \tilde{a}_1 y_{t-1} + \tilde{\epsilon}_t$		
Hypothesis		Test Statistic	Critical Values		
			10%	5%	1%
$a_1 = 0, \tilde{a}_1 = 1$		t-based	-2.58	-2.89	-3.51
$a_0 = 0, \tilde{a}_0 = 0$		t-based	2.17	2.54	3.22
$a_0 = a_1 = 0, \tilde{a}_0 = 0 \text{ \& } \tilde{a}_1 = 1$		F-based	3.86	4.71	6.70
3. Model Estimated:		Dickey-Fuller	$\Delta y_t = a_1 y_{t-1} + \epsilon_t$		
		Phillips-Perron	$y_t = \tilde{a}_1 y_{t-1} + \tilde{\epsilon}_t$		
Hypothesis		Test Statistic	Critical Values		
			10%	5%	1%
$a_1 = 0, \tilde{a}_1 = 1$		t-based	-1.61	-1.95	-2.60

CRITICAL VALUES FOR REGRESSION-RESIDUAL BASED
COINTEGRATION TESTS

Estimated Cointegrating Regression Residual:

$$z_t = y_t - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t} - \beta_3 x_{3t} - \dots - \beta_N x_{Nt}$$

Number of Variables $N + 1$	Sample Size	Critical Values		
		10%	5%	1%
2	50	3.28	3.67	4.32
	100	3.03	3.37	4.07
	200	3.02	3.37	4.00
3	50	3.73	4.11	4.84
	100	3.59	3.93	4.45
	200	3.47	3.78	4.35
4	50	4.02	4.35	4.94
	100	3.89	4.22	4.75
	200	3.89	4.18	4.70
5	50	4.42	4.76	5.41
	100	4.26	4.58	5.18
	200	4.18	4.48	5.02
6	500	4.43	4.71	5.28

Notes and Sources: Standard Dickey-Fuller and Phillips Perron unit-root tests are applied to the regression residuals using the critical values above instead of those on the previous page, focussing entirely on the coefficients of the lagged residual. Except for the case of 6 variables, these critical values were calculated using Monte Carlo methods by Robert F. Engle and Byung Sam Yoo and obtained from their paper "Forecasting and Testing in Co-Integrated Systems," *Journal of Econometrics*, Vol. 35, 1987, page 157. The critical values for the case of 6 variables using 500 observations were calculated by Peter C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, Vol. 58, 1990, 165-93, and were obtained from James D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994, page 766, Case 2. The complete set of Phillips-Ouliaris critical values distinguish between whether or not a constant and trend are included in the cointegrating regression. These values are so similar in the three cases to the ones calculated by Engle and Yoo, based on the inclusion of a constant but not trend, that the complexities of including them here are avoided.

CRITICAL VALUES FOR JOHANSEN COINTEGRATION TESTS

	Probability that Statistic Exceeds Entry					
$n - h$	0.10	0.05	0.01	0.10	0.05	0.01
	Unrestricted Estimation: Trend Drift in Data					
	L-max			Trace		
1	2.816	3.962	6.936	2.816	3.962	6.936
2	12.099	14.036	17.936	13.338	15.197	19.310
3	18.697	20.778	25.521	26.791	29.509	35.397
4	24.712	27.169	31.943	43.964	47.181	53.792
5	30.774	33.178	38.341	65.063	68.905	76.955
	Unrestricted Estimation: No Trend Drift in Data					
	L-max			Trace		
1	6.691	8.083	11.576	6.691	8.083	11.576
2	12.783	14.595	18.782	15.583	17.844	21.962
3	18.959	21.279	26.154	28.436	31.256	37.291
4	24.917	27.341	32.616	45.245	48.419	55.551
5	30.818	33.262	38.858	69.956	69.977	77.911

Notes and Sources: n is the number of variables and h is the number of cointegrating vectors under the null hypothesis. The critical values in the table are copied from Walter Enders, *Applied Economic Time Series*, Wiley Series in Probability and Statistics, 1995, page 420, and are identical to those found in James D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994, pages 767 and 768, Cases 2 and 3.