FOURTH EDITION

APPLIED ECONOMETRIC TIME SERIES

WALTER ENDERS
University of Alabama

WILEY

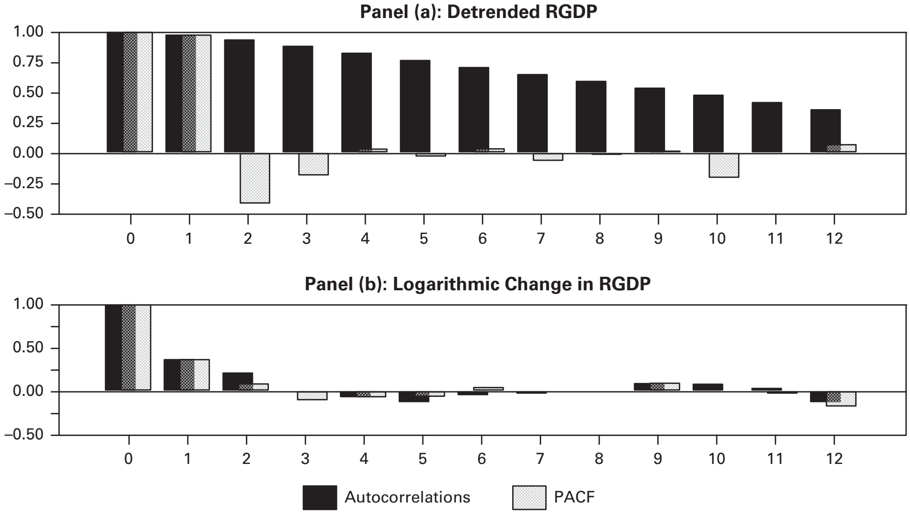


FIGURE 4.4 ACF and PACF

Rather than rely solely on an analysis of correlograms, it is possible to formally test whether a series is stationary. We examine such tests in the next several sections. The testing procedure is not as straightforward as it may seem. We cannot use the usual testing techniques because classical procedures all presume that the data are stationary. For now, it suffices to say that Nelson and Plosser are not able to reject the null hypothesis of a unit root. However, before we examine the tests for a unit root, it is important to note that the issue of nonstationarity also arises quite naturally in the context of the standard regression model.

3. UNIT ROOTS AND REGRESSION RESIDUALS

Consider the regression equation

$$y_t = a_0 + a_1 z_t + e_t \tag{4.12}$$

where the symbol e_t is used to indicate that the error term may be serially correlated.

The assumptions of the classical regression model necessitate that both the $\{y_t\}$ and $\{z_t\}$ sequences be stationary and that the errors have a zero mean and a finite variance. In the presence of nonstationary variables, there might be what Granger and Newbold (1974) call a **spurious regression**. A spurious regression has a high R^2- and t-statistics that appear to be significant, but the results are without any economic meaning. The regression output "looks good," but the least-squares estimates are not consistent and the customary tests of statistical inference do not hold. Granger and Newbold (1974) provide a detailed examination of the consequences of violating the stationarity assumption by generating two sequences, $\{y_t\}$ and $\{z_t\}$, as *independent* random walks using the formulas:

$$y_t = y_{t-1} + \varepsilon_{yt} \tag{4.13}$$

and

$$z_t = z_{t-1} + \varepsilon_{zt} \tag{4.14}$$

where ε_{yt} and ε_{zt} are white-noise processes that are independent of each other.

Granger and Newbold generated many such samples, and for each sample estimated, a regression in the form of (4.12). Since the $\{y_t\}$ and $\{z_t\}$ sequences are independent of each other, (4.12) is necessarily meaningless; any relationship between the two variables is spurious. Surprisingly, at the 5% significance level, they were able to reject the null hypothesis $a_1 = 0$ in approximately 75% of the cases. Of course, at the 5% level, a correctly sized test would yield rejections in only 5% of the regressions. Moreover, the regressions usually had very high R^2 values, and the estimated residuals exhibited a high degree of autocorrelation.

To explain the findings of Granger and Newbold, note that the regression equation (4.12) is necessarily meaningless if the residual series $\{e_t\}$ is nonstationary. Obviously, if the $\{e_t\}$ sequence has a stochastic trend, any error in period t never decays so that any deviation from the model is permanent. It is hard to imagine attaching any importance to an economic model having permanent errors. The simplest way to examine the properties of the $\{e_t\}$ sequence is to abstract from the intercept term a_0 and rewrite (4.12) as

$$e_t = y_t - a_1 z_t$$

If y_t and z_t are generated by (4.13) and (4.14), we can impose the initial conditions $y_0 = z_0 = 0$ so that

$$e_t = \sum_{i=1}^{t} \varepsilon_{yi} - a_1 \sum_{i=1}^{t} \varepsilon_{zi} \tag{4.15}$$

Clearly, the variance of the error becomes infinitely large as t increases. Moreover, the error has a permanent component in that $E_t e_{t+i} = e_t$ for all $i \geq 0$. Hence, the assumptions embedded in the usual hypothesis tests are violated so that any t-test, F-test, or R^2 values are unreliable. It is easy to see why the estimated residuals from a spurious regression will exhibit a high degree of autocorrelation. Updating (4.15), you should be able to demonstrate that the theoretical value of the correlation coefficient between e_t and e_{t+1} goes to unity as t increases.

Even though the true value of $a_1 = 0$, suppose that you estimate (4.12) and want to test the null hypothesis $a_1 = 0$. From (4.15), it should be clear that the error term is nonstationary. Yet, the assumption that the error term is a unit root process is inconsistent with the distributional theory underlying the use of OLS. This problem will not disappear in large samples. In fact, Phillips (1986) proves that the larger the sample, the more likely you are to falsely conclude that $a_1 \neq 0$.

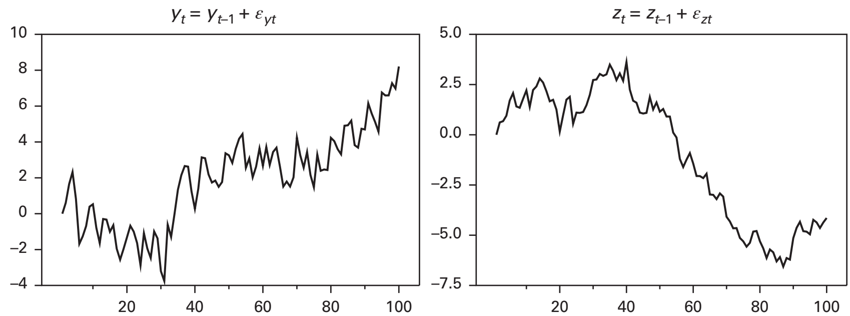
Worksheet 4.1 illustrates the problem of spurious regressions. The top two graphs show 100 realizations of the $\{y_t\}$ and $\{z_t\}$ sequences generated according to (4.13) and (4.14). Although $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ are drawn from white-noise distributions, the realizations of the two sequences are such that y_{100} is positive and z_{100} is negative.

In the lower left panel, you can see that the regression of y_t on z_t captures the *within-sample* tendency of the sequences to move in opposite directions. The straight

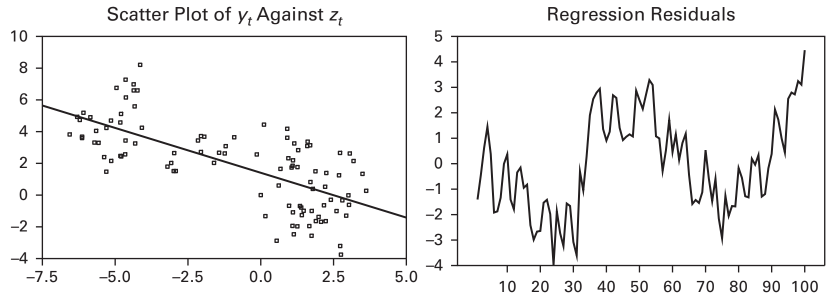
WORKSHEET 4.1

SPURIOUS REGRESSIONS: EXAMPLE 1

Consider the two random walk processes



Since both series are unit root processes with uncorrelated error terms, the regression of y_t on z_t is spurious. Given the realizations of $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$, it happens that y_t tends to increase as z_t tends to decrease. The regression line shown in the scatter plot of y_t on z_t captures this tendency. The correlation coefficient between y_t and z_t is -0.69 and a linear regression yields $y_t = 1.41 - 0.565z_t$. However, the residuals from the regression equation are nonstationary.

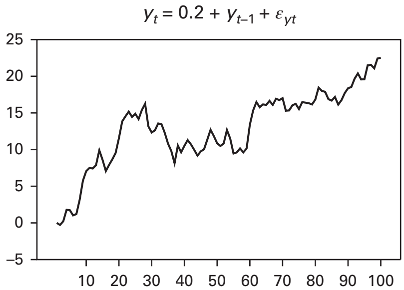


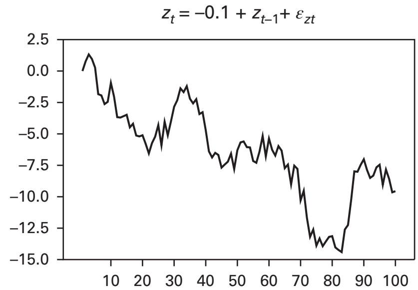
line shown in the scatter plot is the OLS regression line $y_t = 1.41 - 0.565z_t$. The correlation coefficient between $\{y_t\}$ and $\{z_t\}$ is -0.69. The residuals from this regression have a unit root; as such, the coefficients 1.41 and -0.565 are spurious. Worksheet 4.2 illustrates the same problem using two simulated random walk plus drift sequences: $y_t = 0.2 + y_{t-1} + \varepsilon_{yt}$ and $z_t = -0.1 + z_{t-1} + \varepsilon_{zt}$. The drift terms dominate so that for small values of t, it appears that $y_t = -2z_t$. As sample size increases, however, the cumulated sum of the errors (i.e., Σe_t) will pull the relationship further and further from -2.0. The scatter plot of the two sequences suggests that the R^2 statistic will be close to unity; in fact, R^2 is 0.93. However, as you can see in the last panel of

WORKSHEET 4.2

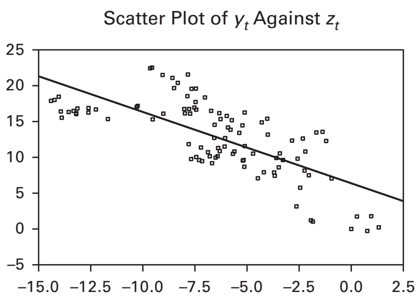
SPURIOUS REGRESSIONS: EXAMPLE 2

Consider the two random walk plus drift processes





Again, the $\{y_t\}$ and $\{z_t\}$ series are unit root processes with uncorrelated error terms so that the regression of y_t on z_t is spurious. Although it is the deterministic drift terms that cause the sustained increase in y_t and the overall decline in z_t, it appears that the two series are inversely related to each other. The residuals from the regression $y_t = 6.38 - 0.10z_t$ are nonstationary.



Worksheet 4.2, the residuals from the regression equation are nonstationary. All departures from this relationship are necessarily permanent.

The point is that the econometrician has to be very careful in working with nonstationary variables. In terms of (4.12), there are four cases to consider:

CASE 1

Both $\{y_t\}$ and $\{z_t\}$ are stationary. When both variables are stationary, the classical regression model is appropriate.

CASE 2

The $\{y_t\}$ and $\{z_t\}$ sequences are integrated of different orders. Regression equations using such variables are meaningless. For example, replace (4.14) with the stationary process $z_t = \rho z_{t-1} + \varepsilon_{zt}$ where $|\rho| < 1$. Now (4.15) is replaced by $e_t = \Sigma\varepsilon_{yi} - a_1\Sigma\rho^i\varepsilon_{zt-i}$. Although the expression $\Sigma\rho^i\varepsilon_{zt-i}$ is convergent, the $\{e_t\}$ sequence still contains a stochastic trend component.[2]

CASE 3

The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order, and the residual sequence contains a stochastic trend. This is the case in which the regression is spurious. The results from such spurious regressions are meaningless in that all errors are permanent. In this case, it is often recommended that the regression equation be estimated in first differences. Consider the first difference of (4.12):

$$\Delta y_t = a_1 \Delta z_t + \Delta e_t$$

Since y_t, z_t, and e_t each contain unit roots, the first difference of each is stationary. Hence, the usual asymptotic results apply. Of course, if one of the trends is deterministic and the other is stochastic, first differencing each is not appropriate.

CASE 4

The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order and the residual sequence is stationary. In this circumstance, $\{y_t\}$ and $\{z_t\}$ are **cointegrated**. A trivial example of a cointegrated system occurs if ε_{zt} and ε_{yt} are perfectly correlated. If $\varepsilon_{zt} = \varepsilon_{yt}$, then (4.15) can be set equal to zero (which is stationary) by setting $a_1 = 1$. To consider a more interesting example, suppose that both z_t and y_t are the random walk plus noise processes:

$$y_t = \mu_t + \varepsilon_{yt}$$
$$z_t = \mu_t + \varepsilon_{zt}$$

where ε_{yt} and ε_{zt} are white-noise processes and μ_t is the random walk process $\mu_t = \mu_{t-1} + \varepsilon_t$. Note that both $\{z_t\}$ and $\{y_t\}$ are $I(1)$ processes but that $y_t - z_t = \varepsilon_{yt} - \varepsilon_{zt}$ is stationary. The subtraction of z_t from y_t serves to nullify the stochastic trend.

All of Chapter 6 is devoted to the issue of cointegrated variables. For now, it is sufficient to note that pretesting the variables in a regression for nonstationarity is extremely important. Estimating a regression in the form of (4.12) is meaningless if cases 2 or 3 apply. If the variables are cointegrated, the results of Chapter 6 apply. The remainder of this chapter considers the formal test procedures for the presence of unit roots and/or deterministic time trends.