Chapter 16

Regression With Time Series Data

In this chapter we extend our analysis of dynamic modeling using time series data. The analysis of time series data is of vital interest to many groups, such as macroeconomists studying the behavior of national and international economies, financial economists who study the stock market, agricultural economists who want to predict supplies and demands for agricultural products. We introduced the problem of autocorrelated errors when using time series data in Chapter 12. In Chapter 15 we considered distributed lag models. In both of these chapters we made implicit assumptions about the time series data, namely, that the time series we examined were **stationary**. In the context of the AR(1) model of autocorrelation, $e_t = \rho e_{t-1} + v_t$, we assumed that $|\rho| < 1$. In the infinite geometric lag model, $y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + e_t$, where $\beta_i = \beta\phi^i$, we assumed $|\phi| < 1$. These assumptions ensure that the time series variables in question are stationary time series. However, many of the variables studied in macroeconomics, monetary economics and finance are **nonstationary** time series. The econometric consequences of nonstationarity can be quite severe, leading to least squares estimators, test statistics and predictors that are unreliable. Moreover, the study of nonstationary time series is one of the fascinating recent developments in econometrics. In this chapter we examine these and related issues.

16.1 Stationary Time Series

Let y_t be an economic variable that we observe over time. Examples of such variables are interest rates, the inflation rate, the gross domestic product, disposable income, etc. The variable y_t is random, since we cannot perfectly predict it. We never know the values of these variables until they are observed. The economic model generating the time series variable y_t is called a **stochastic** or **random process**. We observe a sample of y_t values, which is called a particular **realization** of the stochastic process. It is one of many possible paths that the stochastic process could have taken.

The usual properties of the least squares estimator in a regression using time series data depend on the assumption that the time series variables involved are stationary stochastic processes. A stochastic process (time series) y_t is stationary if its mean and variance are constant over time, and the covariance between two values from the series depends only on the length of time separating the two values, and not on the actual times at which the variables are observed. That is, the time series y_t is stationary if for all values it is true that

$$E(y_t) = \mu \qquad \text{(constant mean)} \qquad (16.1.1a)$$

$$\text{var}(y_t) = \sigma^2 \qquad \text{(constant variance)} \qquad (16.1.1b)$$

$$\text{cov}(y_t, y_{t+s}) = \text{cov}(y_t, y_{t-s}) = \gamma_s \qquad \text{(covariance depends on } s\text{, not } t) \qquad (16.1.1c)$$

The conditions for stationarity may be difficult to grasp, but looking at some pictures may help. In Figure 16.1 (a)–(b) we plot some artificially generated, stationary time series. Note that the series vary randomly at a constant level (mean) and with constant dispersion (variance). In Figure 16.1 (c)–(d) are plots of series that are not stationary. These time series are called **random walks**, because they slowly wander upwards or downwards, but with no real pattern. In Figure 16.1 (e)–(f) are two more nonstationary series, but these show a definite trend either upwards or downwards. These are called **random walks with a drift**.

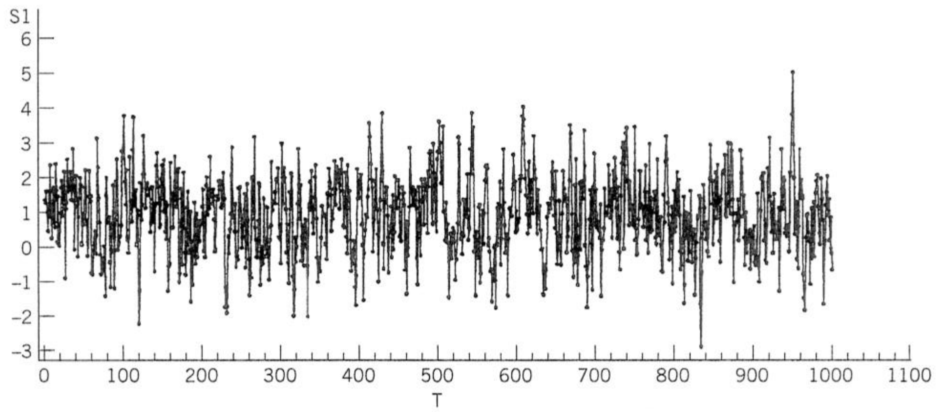


FIGURE 16.1 (a) $y(t) = .5 + .5y(t-1) + N(0, 1)$; stationary process.

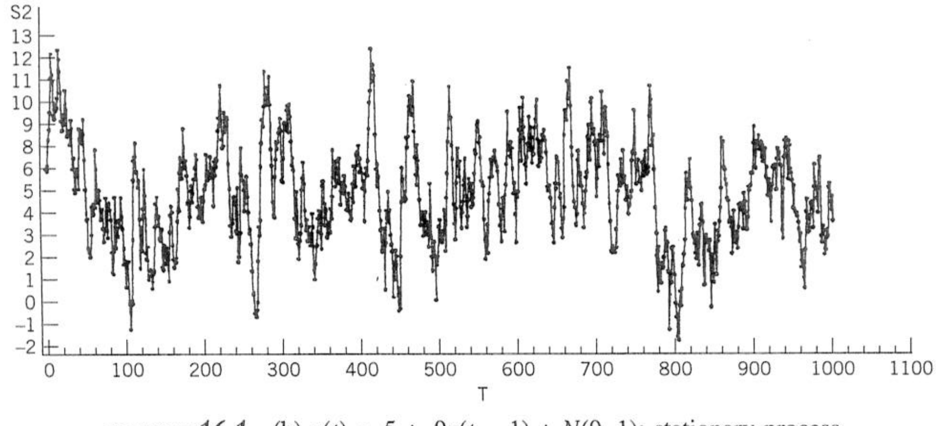


FIGURE 16.1 (b) $y(t) = .5 + .9y(t-1) + N(0, 1)$; stationary process.

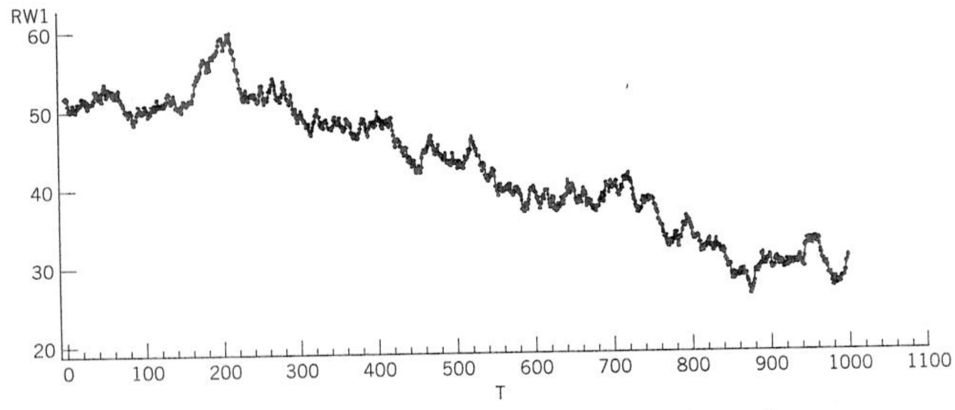


FIGURE **16.1** (c) $y(t) = y(t-1) + .5N(0, 1)$; random walk.

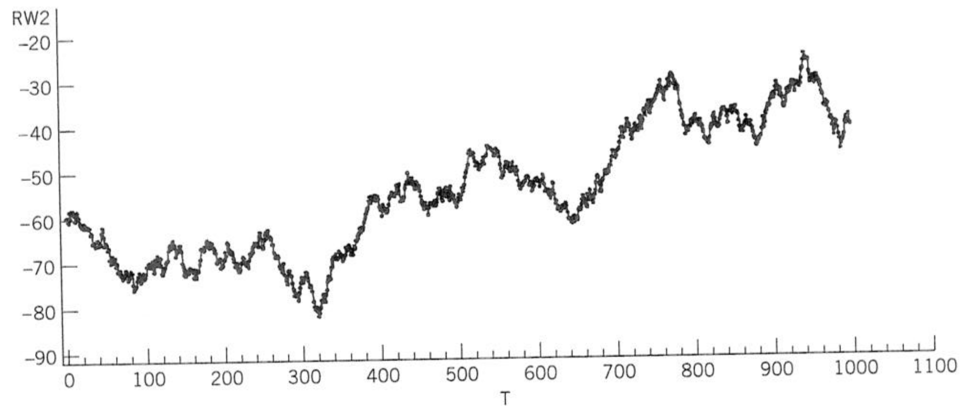


FIGURE **16.1** (d) $y(t) = y(t-1) + N(0, 1)$; random walk.

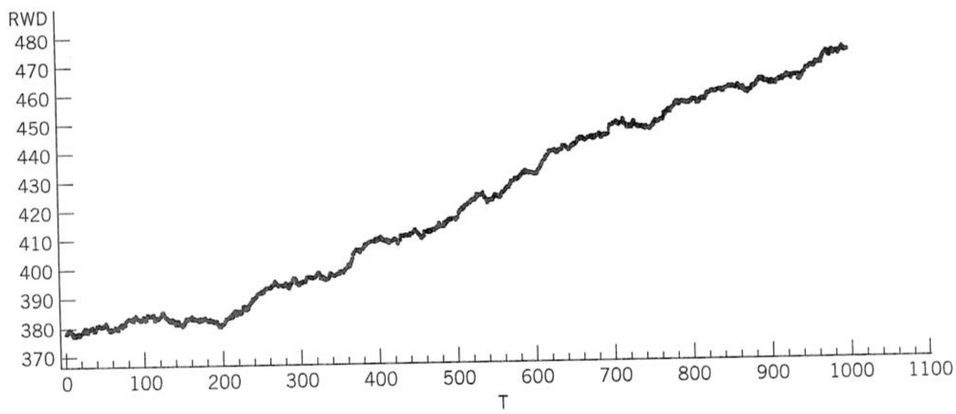


FIGURE **16.1** (e) $y(t) = .1 + y(t-1) + .5N(0, 1)$; random walk with drift.

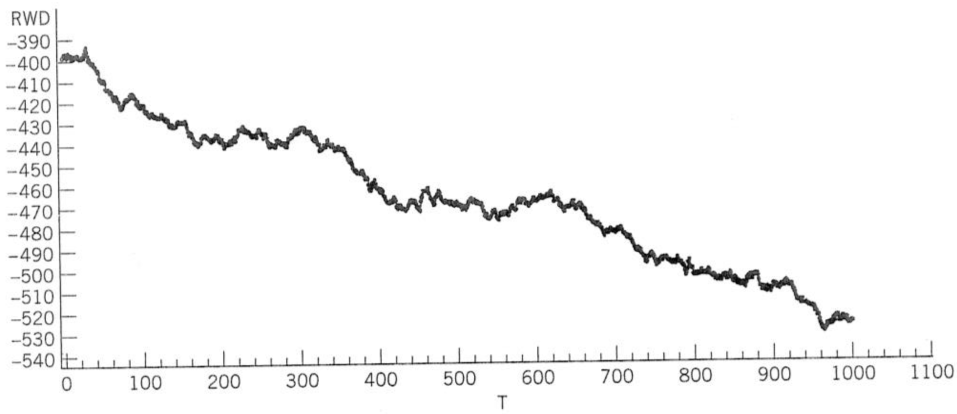


FIGURE ***16.1*** (f) $y(t) = -.1 + y(t - 1) + N(0, 1)$; random walk with drift.

The series in Figure 16.1 are generated from an AR(1) process, much like the AR(1) error process we discussed in Chapter 12. The AR(1) process we consider is

$$\textbf{AR(1) process} \qquad y_t = \alpha + \rho y_{t-1} + v_t \tag{16.1.2}$$

The AR(1) process is stationary if $|\rho| < 1$, as is the case in Figure 16.1 (a)–(b). If $\alpha = 0$ and $\rho = 1$ the AR(1) process reduces to a nonstationary random walk series, depicted in Figure 16.1 (c)–(d), in which the value of y_t this period is equal to the value y_{t-1} from the previous period plus a disturbance v_t.

$$\textbf{Random walk} \qquad y_t = y_{t-1} + v_t \tag{16.1.3}$$

A random walk series shows no definite trend, and slowly turns one way or the other.

If $\alpha \neq 0$ and $\rho = 1$ the series produced is also nonstationary and is called a random walk with a drift.

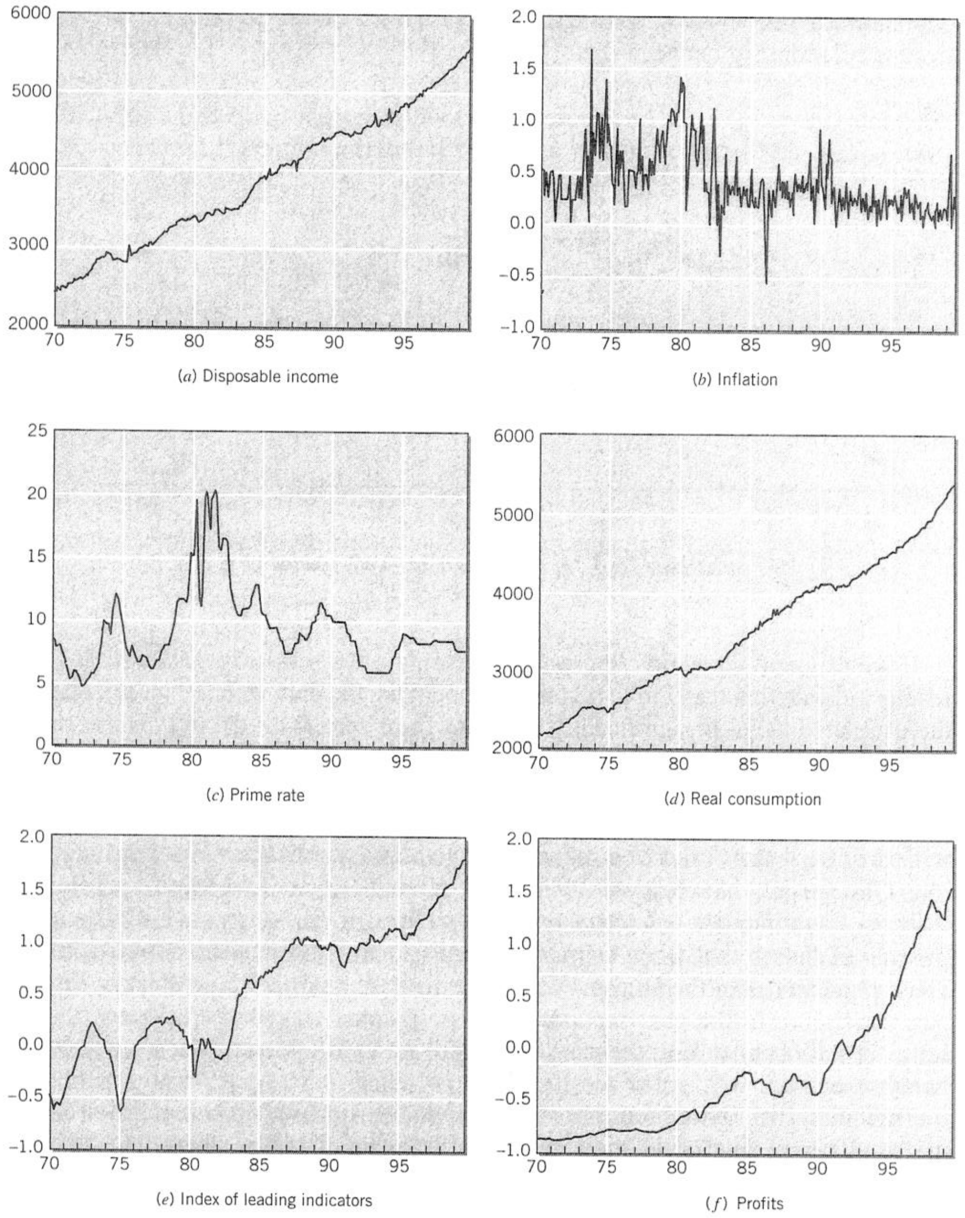
$$\textbf{Random walk with drift} \qquad y_t = \alpha + y_{t-1} + v_t \tag{16.1.4}$$

Such series do show a trend, as illustrated in Figure 16.1 (e)–(f).

Many macroeconomic and financial time series are nonstationary. In Figure 16.2 we plot time series of some important economic variables. Compare these plots to those in Figure 16.1. Which ones look stationary? The ability to distinguish stationary series from nonstationary series is important because, as we noted earlier, using nonstationary variables in regression can lead to least squares estimators, test statistics and predictors that are unreliable and misleading, as we illustrate in the next section.

16.2 Spurious Regressions

There is a danger of obtaining apparently significant regression results from unrelated data when using nonstationary series in regression analysis. Such regressions



[Note: data obtained from www.economagic.com]
(*a*) Real disposable personal income; billions of 1992 dollars, monthly
(*b*) Inflation in consumer prices: percent, monthly
(*c*) Bank prime loan rate, monthly
(*d*) Real personal consumption expenditures; billions of 1992 dollars, monthly
(*e*) U.S. index of leading indicators, 1987 = 100, monthly
(*f*) Corporate profits after tax; billions of dollars, quarterly

FIGURE **16.2** Economic time series.

are said to be **spurious**. To illustrate the problem, let us take the random walk data from Figure 16.1 (c)–(d) and estimate a regression of series one ($y = rw1$) on series two ($x = rw2$). These series were generated independently and have no relation to one another. Yet, when we plot them, as we have done in Figure 16.3, we see an inverse relationship between them.

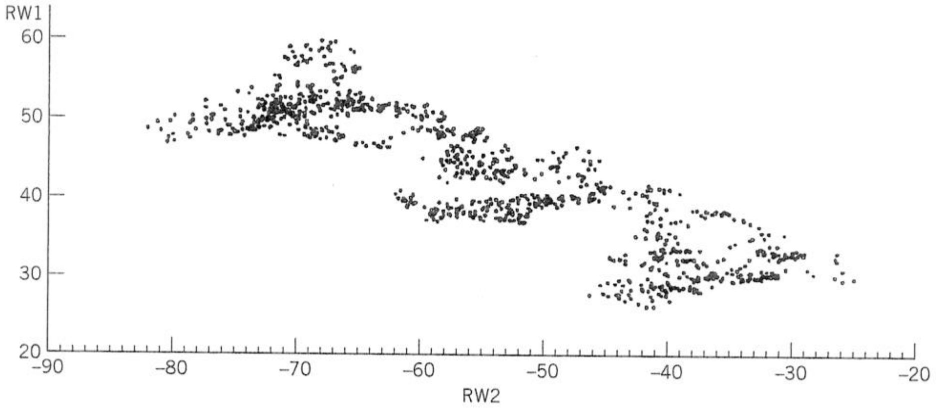


FIGURE **16.3** $y = rw1$ and $x = rw2$ scatter plot.

If we estimate the simple regression, we obtain the results in Table 16.1. These results indicate that the simple regression model fits the data well ($R^2 = .75$), and that the estimated slope is significantly different from zero ($t = -54.67$). These results are completely meaningless, or spurious. The apparent significance of the relationship is false, resulting from the fact that we have related one slowly turning series to another. Similar and more dramatic results are obtained when the random walk with drift series are used in a regression. Note that the Durbin–Watson statistic is low. Granger and Newbold (C. W. J. Granger and P. Newbold, "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 1974, pp. 111–120) suggest the rule of thumb that when estimating regressions with time series data, if the R^2 value is greater than the Durbin–Watson statistic, then one should suspect a spurious regression. In the next section we propose some ways to test whether a time series is stationary or not; these tests can also be used to detect when a problem is likely to occur.

To summarize, when nonstationary time series are used in a regression model the results may spuriously indicate a significant relationship when there is none. In these cases the least squares estimator and least squares predictor do not have their usual properties, and t-statistics are not reliable. Since many macroeconomic time series are nonstationary, it is very important that we take care when estimating regressions with macro-variables.

Table **16.1** **Spurious Regression Results**

Reg Rsq 0.7495 Durbin-Watson 0.0305

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	14.204040	0.5429	26.162	0.0001
RW2	1	-0.526263	0.00963	-54.667	0.0001

16.3 Checking Stationarity Using the Autocorrelation Function

In (16.1.1c) we defined the covariance between y_t and y_{t+s}. Using this definition we can construct the **autocorrelation function**, ρ_s, of the series as

$$\rho_s = \frac{\operatorname{cov}(y_t, y_{t+s})}{\operatorname{var}(y_t)} = \frac{\gamma_s}{\gamma_0} \tag{16.3.1}$$

The value of $\rho_0 = 1$, and for $s > 1$ the correlations ρ_s are pure numbers (unitless) between -1 and 1. The estimated sample correlations are

$$\hat{\rho}_s = \frac{\widehat{\operatorname{cov}}(y_t, y_{t+s})}{\widehat{\operatorname{var}}(y_t)} = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} \tag{16.3.2}$$

where the sample variance and covariance are estimated from a sample of size T as

$$\hat{\gamma}_s = \frac{\sum (y_t - \bar{y})(y_{t+s} - \bar{y})}{T}$$

$$\hat{\gamma}_0 = \frac{\sum (y_t - \bar{y})^2}{T} \tag{16.3.3}$$

If we plot the sample correlations $\hat{\rho}_s$ against s we obtain what is called a **correlogram**. Econometric software will compute the sample correlations. In Tables 16.2 and 16.3 we show the first 10 correlations (AC) for the stationary series *s*2 and the nonstationary series *rw*1.

There is a dramatic difference between the correlograms for the stationary series *s*2 and the nonstationary series *rw*1. For the stationary series the autocorrelations, the column labeled AC in Table 16.2, gradually die out, indicating that values further in the past are less correlated with the current value. For the nonstationary series *rw*1, the autocorrelations in Table 16.3 do not die out rapidly at all. The cor-

Table 16.2 **Correlogram for *s*2**

Autocorrelation	s	AC	Q-Stat.	Prob.
.|******* |	1	0.900	813.42	0.000
.|****** |	2	0.803	1461.0	0.000
.|****** |	3	0.718	1979.1	0.000
.|***** |	4	0.629	2377.9	0.000
.|**** |	5	0.545	2677.4	0.000
.|**** |	6	0.470	2900.7	0.000
.|*** |	7	0.408	3068.7	0.000
.|*** |	8	0.348	3191.2	0.000
.|** |	9	0.299	3281.8	0.000
.|** |	10	0.266	3353.2	0.000

Table 16.3 **Correlogram for *rw*1**

Autocorrelation	s	AC	Q-Stat	Prob
.|********	1	0.997	997.31	0.000
.|********	2	0.993	1988.8	0.000
.|********	3	0.990	2973.9	0.000
.|********	4	0.986	3953.2	0.000
.|********	5	0.983	4926.3	0.000
.|********	6	0.979	5893.4	0.000
.|********	7	0.975	6854.4	0.000
.|*******|	8	0.972	7809.4	0.000
.|*******|	9	0.968	8758.3	0.000
.|*******|	10	0.965	9701.0	0.000

relation between $rw1_t$, and $rw1_{t-10}$ is .965. Thus visual inspection of these functions can be a first indicator of nonstationarity.

Are the autocorrelations statistically different from zero? In large samples, if the autocorrelation is zero, then the estimated autocorrelations $\hat{\rho}_s$ are approximately normally distributed with mean 0 and variance $1/T$. Thus for our sample, of size $T = 1001$, the approximate standard error is $\sqrt{1/T} = 0.0316$. A 95% confidence interval is $\pm 1.96(0.0316) = \pm 0.062$. Thus, if a value of $\hat{\rho}_s$ falls outside the interval $(-0.062, 0.062)$, we conclude that it is significantly different from zero. Given our large sample, and correspondingly narrow confidence interval, the autocorrelations in Tables 16.2 and 16.3 are statistically different from zero.

When the autocorrelations are computed they are customarily accompanied by one or more test statistics for the null hypothesis that all the autocorrelations ρ_s, up to some lag m, are zero. Two commonly reported statistics are the Box-Pierce statistic

$$Q = T \sum_{s=1}^{m} \hat{\rho}_s^2 \tag{16.3.4}$$

and a variation of it developed by Ljung and Box,

$$Q' = T(T+2) \sum_{s=1}^{m} \frac{\hat{\rho}_s^2}{T-s} \tag{16.3.5}$$

Under the null hypothesis that all autocorrelations up to lag m are zero, the statistics Q and Q' are distributed in large samples as $\chi^2_{(m)}$ random variables. If the value of either test statistic is greater than the critical value from the appropriate chi-square distribution, then we reject the null hypothesis that all the autocorrelations are zero and accept the alternative that one or more of them are not zero. In Tables 16.2 and 16.3 the column labeled Q-Stat is the Ljung-Box statistic Q'. The reported p-values indicate that for both series we can reject the null hypothesis that all the autocorrelations are zero.

Testing for zero autocorrelations is, of course, not actually a test for stationarity. The series $s2$ is a stationary series, with statistically significant autocorrelations, as shown in Table 16.2. However, if we fail to reject the null hypothesis of zero autocorrelations, then we conclude that the series is a purely random, or **white noise**, process, which is a special kind of stationary process. In the next section we provide direct tests for nonstationarity.

16.4 Unit Root Tests for Stationarity

The stationarity of a time series can be tested directly with a **unit root test**. The AR(1) model for the time series variable y_t is

$$y_t = \rho y_{t-1} + \nu_t \tag{16.4.1}$$

Assume that ν_t is a random disturbance with zero mean and constant variance σ_ν^2. In this model, if $\rho = 1$ then y_t is the nonstationary random walk, $y_t = y_{t-1} + \nu_t$, and is said to have a **unit root**, because the coefficient $\rho = 1$.

By computing its variance, we can show that the random walk process $y_t = y_{t-1} + \nu_t$ is nonstationary. Suppose that $y_0 = 0$, then, by repeated substitution,

$$\begin{aligned}
y_1 &= \nu_1 \\
y_2 &= y_1 + \nu_2 = \nu_1 + \nu_2 \\
y_3 &= y_2 + \nu_3 = \nu_1 + \nu_2 + \nu_3 \\
&\vdots \\
y_t &= \sum_{j=1}^{t} \nu_j
\end{aligned} \tag{16.4.2}$$

Therefore,

$$\text{var}(y_t) = t\sigma_\nu^2 \tag{16.4.3}$$

Since the variance of y_t changes over time, it is a nonstationary series. In fact, as $t \rightarrow \infty$ the variance of y_t becomes infinitely large.

Recall that if $|\rho| < 1$, then the AR(1) process is stationary. Thus we can test for nonstationarity by testing the null hypothesis that $\rho = 1$ against the alternative that $|\rho| < 1$, or simply $\rho < 1$. The test is put into a convenient form by subtracting y_{t-1} from both sides of (16.4.1), to obtain

$$\begin{aligned}
y_t - y_{t-1} &= \rho y_{t-1} - y_{t-1} + \nu_t \\
\Delta y_t &= (\rho - 1)y_{t-1} + \nu_t \\
&= \gamma y_{t-1} + \nu_t
\end{aligned} \tag{16.4.4}$$

where $\Delta y_t = y_t - y_{t-1}$ and $\gamma = \rho - 1$. Then

$$\begin{aligned}
H_0 : \rho = 1 &\leftrightarrow H_0 : \gamma = 0 \\
H_1 : \rho < 1 &\leftrightarrow H_1 : \gamma < 0
\end{aligned} \tag{16.4.5}$$

The variable $\Delta y_t = y_t - y_{t-1}$ is called the **first difference** of the series y_t. If y_t follows a random walk, then $\gamma = 0$ and

$$\Delta y_t = y_t - y_{t-1} = v_t \qquad (16.4.6)$$

An interesting feature of the series $\Delta y_t = y_t - y_{t-1}$ is that it is stationary if, as we have assumed, the random error v_t is purely random. Series like y_t, which can be made stationary by taking the first difference, are said to be **integrated of order 1**, and denoted I(1). Stationary series are said to be integrated of order zero, I(0). In general, if a series must be differenced d times to be made stationary it is **integrated of order *d***, or **I(*d*)**.

16.4.1 The Dickey–Fuller Tests

To test the hypothesis in (16.4.5) we estimate (16.4.4) by least squares as usual, and examine the t-statistic for the hypothesis that $\gamma = 0$ as usual. Unfortunately this t-statistic no longer has a t-distribution, since, if the null hypothesis is true, y_t follows a random walk. Consequently this statistic, which is often called the **τ (tau) statistic,** must be compared to specially constructed critical values. Originally these critical values were tabulated by statisticians Dickey and Fuller (D. A. Dickey and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 1979, 427–431). The test using these critical values has become known as the **Dickey–Fuller test**.

In addition to testing if a series is a random walk, Dickey and Fuller also developed critical values for the presence of a unit root (a random walk process) in the presence of a **drift**.

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + v_t \qquad (16.4.7)$$

Such series display a definite trend, as we have illustrated with simulated data in Figure 16.1 (e)–(f). This is an extremely important case, because as you can see in Figure 16.2, macroeconomic variables often exhibit a strong trend.

It is also possible to allow explicitly for a nonstochastic trend. To do so, the model is further modified to include a time trend, or time, t

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + v_t \qquad (16.4.8)$$

Critical values for the *tau* (τ) statistic, which are valid in large samples for a one-tailed test, are given in Table 16.4. Comparing these values to the standard values in the last row, you see that the τ-statistic must take larger (negative) values than usual in order for the null hypothesis $\gamma = 0$, a unit root-nonstationary process, to be rejected in favor of the alternative that $\gamma < 0$, a stationary process. To control for the possibility that the error term in one of the equations, for example (16.4.7), is autocorrelated, additional terms are included. The modified model is

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=1}^{m} a_i \Delta y_{t-i} + v_t \qquad (16.4.9)$$

where

$$\Delta y_{t-1} = (y_{t-1} - y_{t-2}), \qquad \Delta y_{t-2} = (y_{t-2} - y_{t-3}), \ldots$$

Table 16.4 **Critical Values for the Dickey–Fuller Test**

Model	1%	5%	10%
$\Delta y_t = \gamma y_{t-1} + v_t$	−2.56	−1.94	−1.62
$\Delta y_t = \alpha_0 + \gamma y_{t-1} + v_t$	−3.43	−2.86	−2.57
$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + v_t$	−3.96	−3.41	−3.13
Standard critical values	−2.33	−1.65	−1.28

Note: These critical values are taken from R. Davidson and J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York: Oxford University Press, p. 708.

Testing the null hypothesis that $\gamma = 0$ in the context of this model is called the **augmented Dickey–Fuller test**. The test critical values are the same as for the Dickey–Fuller test, as shown in Table 16.4.

16.4.2 The Dickey–Fuller Tests: An Example

As an example, consider real personal consumption expenditures (y_t) as plotted in Figure 16.2(d). This variable is strongly trended, and we suspect that it is nonstationary. Inspection of the correlogram shows very slowly declining autocorrelations, a first indicator of nonstationarity. We estimate (16.4.7) and (16.4.8) with and without additional terms to control for autocorrelation. These results are reported in (16.4.10). In each case the estimated value of γ (the coefficient of PCE_{t-1}) is positive, as are the associated *tau* statistics. Clearly we do not reject the null hypothesis that personal consumption expenditures have a unit root.

$$\Delta \hat{PCE}_t = -1.5144 + .0030 PCE_{t-1}$$
$$\text{(tau)} \quad (-0.349) \quad (2.557) \qquad (16.4.10a)$$

$$\Delta \hat{PCE}_t = 2.0239 + 0.0152t + 0.0013 PCE_{t-1}$$
$$\text{(tau)} \quad (0.1068) \quad (0.1917) \quad (0.1377) \qquad (16.4.10b)$$

$$\Delta \hat{PCE}_t = -2.111 + 0.00397 PCE_{t-1} - 0.2503 \Delta PCE_{t-1} - 0.0412 \Delta PCE_{t-2}$$
$$\text{(tau)} \quad (-0.4951) \quad (3.3068) \quad (-4.6594) \quad (-0.7679) \qquad (16.4.10c)$$

The question then becomes: Is the first difference ($\Delta PCE_t = PCE_t - PCE_{t-1}$) of the personal consumption series stationary?

In Figure 16.4 we plot the first differences, which certainly look like the plots of stationary processes in Figure 16.1 (a)–(b). The correlogram shows small correlations at all lags, suggesting stationarity. The result of the Dickey–Fuller test for a random walk (since there is no trend) applied to the series ΔPCE_t, which we denote as D_t, is given in (16.4.11):

$$\Delta \hat{D}_t = -0.9969 D_{t-1}$$
$$\text{(tau)} \quad (-18.668) \qquad (16.4.11)$$

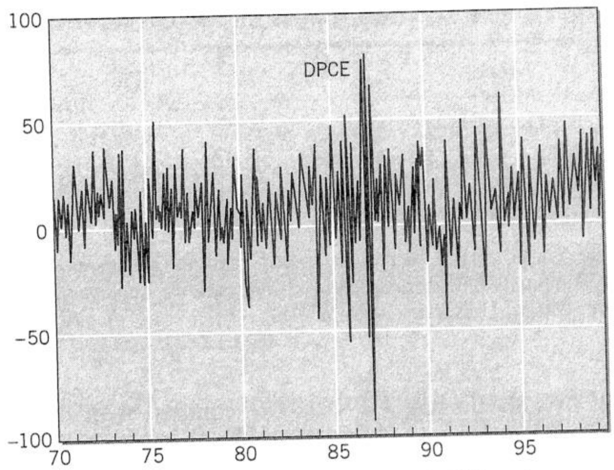


FIGURE ***16.4*** First differences of PCE series.

Based on the large negative value of the *tau* statistic we reject the null hypothesis that ΔPCE_t has a unit root and accept the alternative that it is stationary. Collecting the results from the unit root tests on PCE_t and ΔPCE_t, we can say that the series PCE is I(1). Had the null hypothesis of a unit root not been rejected in (16.4.11), we would have concluded that PCE is I(2) or integrated of an order higher than 2.

16.5 Cointegration

As a general rule nonstationary time series variables should not be used in regression models, in order to avoid the problem of spurious regression. However, there is an exception to this rule. If y_t and x_t are nonstationary I(1) variables, then we would expect that their difference, or any linear combination of them, such as $e_t = y_t - \beta_1 - \beta_2 x_t$, to be I(1) as well. However, there are important cases when $e_t = y_t - \beta_1 - \beta_2 x_t$ is a stationary I(0) process. In this case y_t and x_t are said to be **cointegrated**. Cointegration implies that y_t and x_t share similar stochastic trends, and in fact, since their difference e_t is stationary, they never diverge too far from each other. The cointegrated variables y_t and x_t exhibit a *long-term equilibrium* relationship defined by $y_t = \beta_1 + \beta_2 x_t$, and e_t is the *equilibrium error*, which represents short-term deviations from the long-term relationship.

We can test whether y_t and x_t are cointegrated by testing whether the errors $e_t = y_t - \beta_1 - \beta_2 x_t$ are stationary. Since we cannot observe e_t, we instead test the stationarity of the least squares residuals, $\hat{e}_t = y_t - b_1 - b_2 x_t$, using a Dickey–Fuller test. We estimate the regression

$$\Delta \hat{e}_t = \alpha_0 + \gamma \hat{e}_{t-1} + \nu_t \tag{16.5.1}$$

where $\Delta \hat{e}_t = \hat{e}_t - \hat{e}_{t-1}$, and examine the t (or *tau*) statistic for the estimated slope. Because we are basing this test upon estimated values, the critical values given in Table 16.5 are somewhat different than those in Table 16.4.

Table 16.5 **Critical Values for the Co-integration Test**

Model	1%	5%	10%
$\Delta\hat{e}_t = \alpha_0 + \gamma\hat{e}_{t-1} + v_t$	−3.90	−3.34	−3.04

Note: These critical values are taken from R. Davidson and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press, p. 722.

16.5.1 An Example of a Cointegration Test

To illustrate, let us test whether $y_t = PCE_t$ and $x_t = PDI_t$, where PDI_t is real personal disposable income (monthly), as plotted in Figure 16.2(a), are cointegrated. You may confirm that PDI_t is nonstationary. The estimated least squares regression between these variables is

$$\begin{aligned}\widehat{PCE}_t &= -390.7848 + 1.0160DPI_t \\ \text{(t-stats)} &\quad (-24.50) \qquad (252.97)\end{aligned} \tag{16.5.2}$$

Estimating (16.5.1) we obtain

$$\begin{aligned}\Delta\hat{e}_t &= 0.188250 - 0.120344\hat{e}_{t-1} \\ \text{(tau)} &\quad (0.1107) \qquad (-4.5642)\end{aligned} \tag{16.5.3}$$

The *tau* statistic is less than the critical value −3.90 for the 1% level of significance, thus we reject the null hypothesis that the least squares residuals are nonstationary, and conclude that they are stationary. Thus we conclude that personal consumption expenditures and personal disposable income are cointegrated, indicating that there is a long-run, equilibrium relationship between these variables.

16.6 Summarizing Estimation Strategies When Using Time Series Data

We have mentioned in this chapter some of the complexities that arise when using time series data. In doing so we have glossed over many issues that are discussed in books devoted to the study of time series data. However, we have made some important points. Let us summarize what we have discovered so far in this chapter.

- A regression between two nonstationary variables can produce spurious results.
- Nonstationarity of variables can be assessed using the autocorrelation function, and through unit root tests.
- Spurious regressions exhibit a low value of the Durbin–Watson statistic and a high R^2.
- If two nonstationary variables are cointegrated, their long-run relationship can be estimated via a least squares regression.
- Cointegration can be assessed via a unit root test on the residuals of the regression.

There are still some unanswered questions.

1. First, if the variables are nonstationary, and not cointegrated, is there any relationship that can be estimated? In these circumstances one can investigate whether there is a relationship between the variables after they have been differenced to achieve stationarity. For example, suppose that the two variables y_t and x_t are I(1) variables, and that they are not cointegrated. Since the changes Δy_t and Δx_t are stationary, we can run regressions of the form

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_t + e_t \tag{16.6.1}$$

Estimating equations like this one gives information on any relationship between the *changes* in the variables.
2. A second case is the one in which y_t and x_t are stationary, the implicit assumption maintained for most of the text. In this case least squares or generalized least squares, whichever is more appropriate, can be used to estimate a relationship between y and x.
3. Finally, there is a third relationship that is of interest, called an **error correction model**, that can be estimated when y_t and x_t are nonstationary, but cointegrated.

For I(1) variables, the error-correction model relates changes in a variable, say Δy_t, to departures from the long-run equilibrium in the previous period $(y_{t-1} - \beta_1 - \beta_2 x_{t-1})$. It can be written as

$$\Delta y_t = \alpha_1 + \alpha_2 (y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + v_t \tag{16.6.2}$$

The changes or *corrections* Δy_t depend on the departure of the system from its long-run equilibrium in the previous period. The shock v_t leads to a short-term departure from the cointegrating equilibrium path; then, there is a tendency to correct back toward the equilibrium. The coefficient α_2 governs the speed of adjustment back toward the long-run equilibrium. We usually expect the sign of α_2 to be negative, so that a positive (negative) departure from equilibrium in the previous period will be corrected by a negative (positive) amount in the current period.

One way to estimate the error correction model is to use least squares to estimate the cointegrating relationship $y_t = \beta_1 + \beta_2 x_t$, and to then use the lagged residuals $\hat{e}_{t-1} = y_{t-1} - b_1 - b_2 x_{t-1}$ as the right-hand-side variable in the error-correction model, estimating it with a second least squares regression.

16.7 Learning Objectives

Based on the material in this chapter, you should be able to explain:

1. In intuitive terms, what the difference is between stationary and nonstationary time series processes?
2. What is the general behavior of a random walk time series?
3. What is a "spurious regression."

4. The general characteristics of the correlogram for stationary and nonstationary processes.
5. What is a "unit root" test, and state implications of the null and alternative hypotheses?
6. The meaning of a series being "integrated of order 1," or I(1).
7. The Dickey–Fuller and Augmented Dickey–Fuller tests for unit roots, and carry them out using your computer software.
8. The concept of cointegration, and how to test whether two series are cointegrated.
9. The formulation of an error-correction model.

16.8 Exercises

16.1 The data for Figure 16.1 are in the file *fig16-1.dat*. There are 1001 observations on 6 series, *s1*, *s2*, *rw1*, *rw2*, *rwd1*, *rwd2*.
(a) Plot each of these series using your computer software.
(b) Obtain the correlogram for each series and comment.
(c) Estimate a simple regression of *rwd1* on *rwd2*. Comment on the results.
(d) Estimate a simple regression of *s1* on *s2*. Comment on the results.
(e) Explain how the regressions in (c) and (d) are related to the problem of spurious regressions.

16.2 The data for Figure 16.2 (a)–(e) are in the file *fig16-2.dat*. These series are monthly from 1970.01 to 1999.08. The quarterly data for Figure 16.2 (f), covering the period 1959.I to 1999.III, are in the file *profits.dat*.
(a) Plot each of these series using your computer software.
(b) Obtain the correlogram for each series and comment.
(c) Test each of these series for a unit root. (Include an intercept, and perform the test with and without a trend. Use four augmentation terms and the 10% significance level.)
(d) For each series in (c) that is nonstationary, take the first difference. Plot this series and test it for stationarity. What do you conclude about the "order of integration" of each series?

16.3 In the file *oil.dat* are 88 annual observations on the price of oil (in 1967 constant dollars) for the period 1883–1970. These data are from Pindyck and Rubinfeld, *Econometric Models and Economic Forecasts*, 3rd Ed. (1991, p. 463).
(a) Plot the data and obtain the correlogram. Based on this information, do oil prices appear stationary or not?
(b) Use a unit root test to demonstrate that the series is stationary. Use the instructions in Exercise 16.2(c).

16.4 In the file *bond.dat* there are 102 monthly observations on AA railroad bond yields for the period January 1968 to June 1976. [They are taken from Cryer, J. D., *Time Series Analysis* (1986)].
(a) Plot the data and obtain the correlogram. Based on this information, do railroad bond yields appear stationary?
(b) Use a unit root test to demonstrate that the series is nonstationary. Use the instructions in Exercise 16.2(c).
(c) Find the first difference of the bond yield series and check it for non-

stationarity. What do you conclude about the order of integration of this series?

16.5 Test whether the quarterly observations for nondurable consumption and disposable income (for 1947:I to 1980:IV) stored in the file *macrovar.dat* are cointegrated.

16.6 In the file *texas.dat* there are 57 quarterly observations on the real price of oil (*RPO*), Texas nonagricultural employment (*TXNAG*), and nonagricultural employment in the rest of the U.S. (*USNAG*). The data cover the period 1974:Q1 through 1988:Q1 and were used in a study by Fomby and Hirschberg (T. B. Fomby and J. G. Hirschberg, "Texas in Transition: Dependence on Oil and the National Economy," *Federal Reserve Bank of Dallas Economic Review*, January 1989, 11–28). Let $x_t = \ln(RPO_t/RPO_{t-1})$, $y_t = \ln(TXNAG_t/TXNAG_{t-1})$, and $z_t = \ln(USNAG_t/USNAG_{t-1})$.
(a) Test *RPO*, *TXNAG*, and *USNAG* for unit roots.
(b) Test x_t, y_t, and z_t for the existence of unit roots, using the instructions in Exercise 16.2(c).

16.7 Visit one of the web sites at which economic data can be downloaded. Several of these are listed in Chapter 19. Download time series data on five macroeconomic variables.
(a) Plot these variables and examine them for stationarity.
(b) Are any of these series cointegrated?