CHAPTER 14

Time-Series Analysis

14.1 Introduction

There has been an enormous amount of research in time-series econometrics, and many economics departments have required a time-series econometrics course in their graduate sequence. Obviously, one chapter on this topic will not do it justice. Therefore, this chapter will focus on some of the basic concepts needed for such a course. Section 14.2 defines what is meant by a *stationary* time-series, while sections 14.3 and 14.4 briefly review the Box-Jenkins and *Vector Autoregression* (VAR) methods for time-series analysis. Section 14.5 considers a random walk model and various tests for the existence of a *unit root*. Section 14.6 studies *spurious regressions* and *trend stationary* versus *difference stationary* models. Section 14.7 gives a simple explanation of the concept of *cointegration* and illustrates it with an economic example. Finally, section 14.8 looks at *Autoregressive Conditionally Heteroskedastic* (ARCH) time-series.

14.2 Stationarity

Figure 14.1 plots the consumption and personal disposable income data considered in Chapter 5. This was done using EViews. This is annual data from 1959 to 2007 expressed in real terms. Both series seem to be trending upwards over time. This may be an indication that these time-series are non-stationary. Having a time-series x_t that is trending upwards over time may invalidate all the standard asymptotic theory we have been relying upon in this book. In fact, $\sum_{t=1}^{T} x_t^2/T$ may not tend to a finite limit as $T \to \infty$ and using regressors such as x_t means that $X'X/T$ does not tend in probability limits to a finite positive definite matrix, see problem 6.

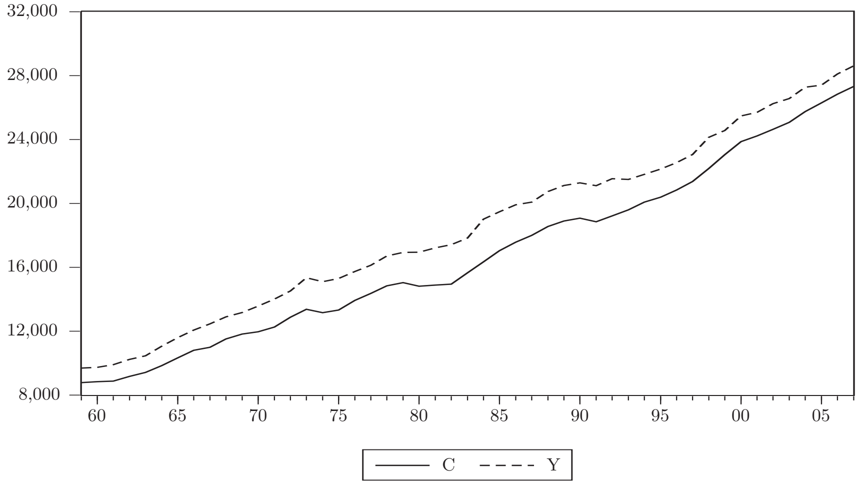

Figure 14.1 U.S. Consumption and Income, 1959–2007

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_14,

Non-standard asymptotic theory will have to be applied which is beyond the scope of this book, see problem 8.

Definition: A time-series process x_t is said to be covariance stationary (or weakly stationary) if its mean and variance are constant and independent of time and the covariances given by $\text{cov}(x_t, x_{t-s}) = \gamma_s$ depend only upon the distance between the two time periods, but not the time periods per se.

In order to check the time-series for weak stationarity one can compute its *autocorrelation function*. This is given by ρ_s= correlation $(x_t, x_{t-s}) = \gamma_s/\gamma_o$. These are correlation coefficients taking values between -1 and $+1$.

The sample counterparts of the variance and covariances are given by

$$\widehat{\gamma}_o = \textstyle\sum_{t=1}^{T}(x_t - \bar{x})^2/T$$

$$\widehat{\gamma}_s = \textstyle\sum_{t=1}^{T-s}(x_t - \bar{x})(x_{t+s} - \bar{x})/T$$

and the *sample autocorrelation function* is given by $\widehat{\rho}_s = \widehat{\gamma}_s/\widehat{\gamma}_o$. Figure 14.2 plots $\widehat{\rho}_s$ against s for the consumption series. This is called the *sample correlogram*. For a stationary process, ρ_s declines sharply as the number of lags s increase. This is not necessarily the case for a nonstationary series. In the next section, we briefly review a popular method for the analysis of time-series known as the Box and Jenkins (1970) technique. This method utilizes the sample autocorrelation function to establish whether a series is stationary or not.

Sample: 1959 2007
Included observations: 49

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>. |*******</td><td>. |*******</td><td>1</td><td>0.935</td><td>0.935</td><td>45.500</td><td>0.000</td></tr>
<tr><td>. |****** |</td><td>. |. |</td><td>2</td><td>0.868</td><td>–0.050</td><td>85.527</td><td>0.000</td></tr>
<tr><td>. |****** |</td><td>. |. |</td><td>3</td><td>0.800</td><td>–0.042</td><td>120.26</td><td>0.000</td></tr>
<tr><td>. |***** |</td><td>. |. |</td><td>4</td><td>0.733</td><td>–0.029</td><td>150.08</td><td>0.000</td></tr>
<tr><td>. |***** |</td><td>. |. |</td><td>5</td><td>0.668</td><td>–0.024</td><td>175.39</td><td>0.000</td></tr>
<tr><td>. |**** |</td><td>. |. |</td><td>6</td><td>0.604</td><td>–0.030</td><td>196.57</td><td>0.000</td></tr>
<tr><td>. |**** |</td><td>. |. |</td><td>7</td><td>0.541</td><td>–0.029</td><td>214.00</td><td>0.000</td></tr>
<tr><td>. |*** |</td><td>. |. |</td><td>8</td><td>0.480</td><td>–0.033</td><td>228.02</td><td>0.000</td></tr>
<tr><td>. |*** |</td><td>. |. |</td><td>9</td><td>0.421</td><td>–0.015</td><td>239.12</td><td>0.000</td></tr>
<tr><td>. |*** |</td><td>. |. |</td><td>10</td><td>0.369</td><td>0.004</td><td>247.83</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. |. |</td><td>11</td><td>0.320</td><td>–0.009</td><td>254.57</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. |. |</td><td>12</td><td>0.272</td><td>–0.033</td><td>259.57</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. |. |</td><td>13</td><td>0.226</td><td>–0.027</td><td>263.10</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. |. |</td><td>14</td><td>0.181</td><td>–0.021</td><td>265.45</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. |. |</td><td>15</td><td>0.140</td><td>–0.013</td><td>266.88</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. |. |</td><td>16</td><td>0.097</td><td>–0.052</td><td>267.60</td><td>0.000</td></tr>
<tr><td>. |. |</td><td>. |. |</td><td>17</td><td>0.055</td><td>–0.036</td><td>267.83</td><td>0.000</td></tr>
<tr><td>. |. |</td><td>. |. |</td><td>18</td><td>0.011</td><td>–0.052</td><td>267.84</td><td>0.000</td></tr>
<tr><td>. |. |</td><td>. |. |</td><td>19</td><td>–0.032</td><td>–0.034</td><td>267.93</td><td>0.000</td></tr>
<tr><td>.*|. |</td><td>. |. |</td><td>20</td><td>–0.073</td><td>–0.026</td><td>268.39</td><td>0.000</td></tr>
</table>

Figure 14.2 Correlogram of Consumption

14.3 The Box and Jenkins Method

This method fits *Autoregressive Integrated Moving Average* (ARIMA) type models. We have already considered simple AR and MA type models in Chapters 5 and 6. The Box-Jenkins methodology differences the series and looks at the sample correlogram of the differenced series to see whether stationarity is achieved. As will be clear shortly, if we have to difference the series once, twice or three times to make it stationary, this series is *integrated* of order 1, 2 or 3, respectively. Next, the Box-Jenkins method looks at the *autocorrelation function* and the *partial autocorrelation function* (synonymous with partial correlation coefficients) of the resulting stationary series to identify the order of the AR and MA process that is needed. The partial correlation between y_t and y_{t-s} is the correlation between those two variables holding constant the effect of all intervening lags, see Box and Jenkins (1970) for details. Figure 14.3 plots an AR(1) process of size $T = 250$ generated as $y_t = 0.7y_{t-1} + \epsilon_t$ with $\epsilon_t \sim \text{IIN}(0, 4)$. Figure 14.4 shows that the correlogram of this AR(1) process declines geometrically as s increases. Similarly, Figure 14.5 plots an MA(1) process of size $T = 250$ generated as $y_t = \epsilon_t + 0.4\epsilon_{t-1}$ with $\epsilon_t \sim \text{IIN}(0, 4)$. Figure 14.6 shows that the correlogram of this MA(1) process is zero after the first lag, see also problems 1 and 2 for further analysis. Identifying the right ARIMA model is not an exact science, but potential candidates emerge. These models are estimated using maximum likelihood techniques. Next, these models are subjected to some diagnostic checks. One commonly used check is to see whether the residuals are White noise. If they fail this test, these models are dropped from the list of viable candidates.

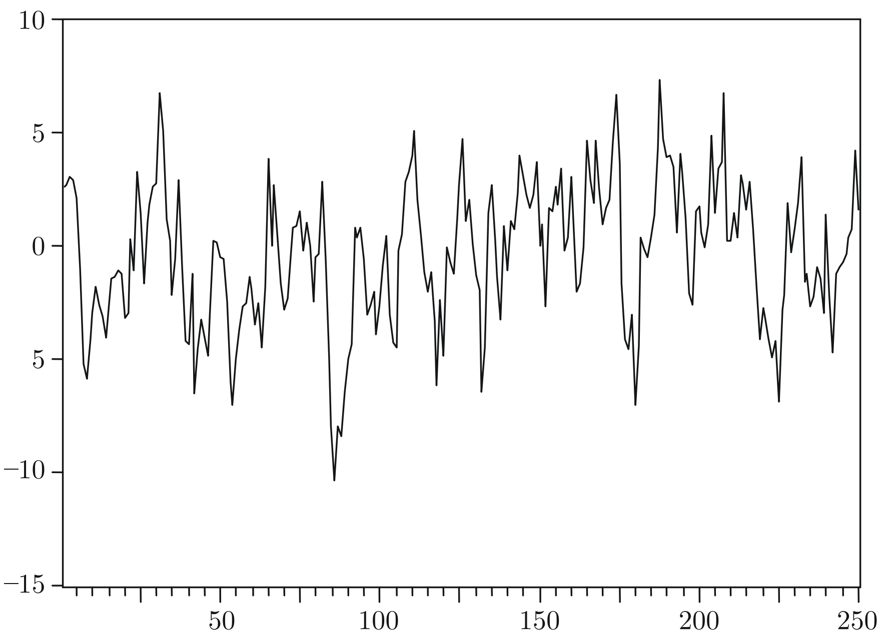

Figure 14.3 AR(1) Process, $\rho = 0.7$

If the time-series is *White noise*, i.e., purely random with constant mean and variance and zero autocorrelation, then $\rho_s = 0$ for $s > 0$. In fact, for a White noise series, if $T \to \infty$, $\sqrt{T}\widehat{\rho}_s$ will be asymptotically distributed $N(0, 1)$. A joint test for H_o; $\rho_s = 0$ for $s = 1, 2, \ldots, m$ lags, is given

Sample: 1 250
Included observations: 250

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |****** |	. |****** |	1	0.725	0.725	132.99	0.000
. |**** |	. |. |	2	0.503	–0.048	197.27	0.000
. |*** |	. |. |	3	0.330	–0.037	225.05	0.000
. |** |	. |. |	4	0.206	–0.016	235.92	0.000
. |* |	. |. |	5	0.115	–0.022	239.33	0.000
. |. |	. |. |	6	0.036	–0.050	239.67	0.000
. |. |	. |. |	7	–0.007	0.004	239.68	0.000
. |. |	. |. |	8	–0.003	0.050	239.68	0.000
. |. |	. |. |	9	–0.017	–0.041	239.75	0.000
* |. |	* |. |	10	–0.060	–0.083	240.71	0.000
* |. |	* |. |	11	–0.110	–0.063	243.91	0.000
. |. |	. |* |	12	–0.040	0.191	244.32	0.000

Figure 14.4 Correlogram of AR(1)

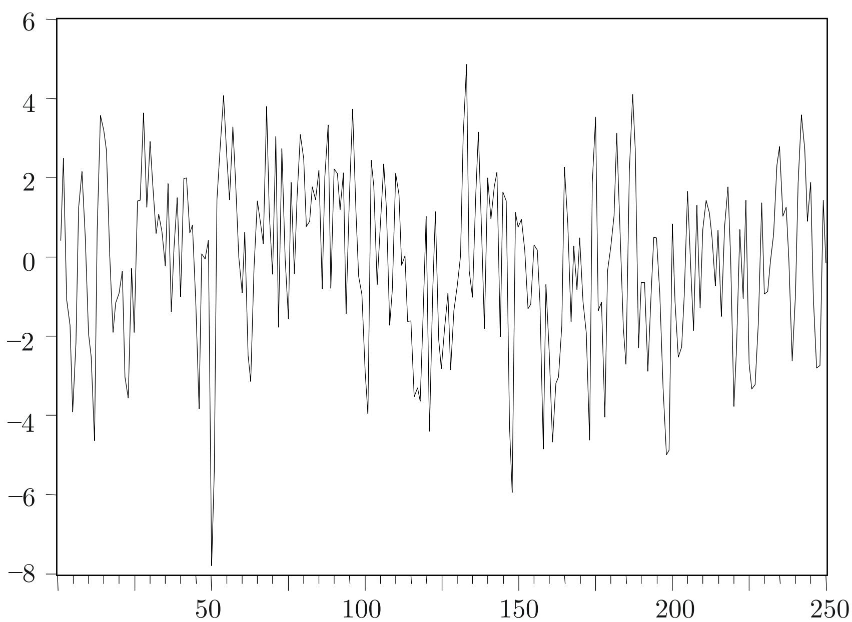

Figure 14.5 MA(1) Process, $\theta = 0.4$

by the Box and Pierce (1970) statistic

$$Q = T \sum_{s=1}^{m} \widehat{\rho}_s^2 \tag{14.1}$$

This is asymptotically distributed under the null as χ^2_m. A refinement of the Box-Pierce Q-statistic that performs better, i.e., have more power in small samples is the Ljung and Box (1978) Q_{LB} statistic

Sample: 1 250
Included observations: 250

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |*** |	. |*** |	1	0.399	0.399	40.240	0.000
. |. |	* |. |	2	0.033	–0.150	40.520	0.000
. |. |	. |* |	3	0.010	0.066	40.545	0.000
. |. |	. |. |	4	–0.002	–0.033	40.547	0.000
. |* |	. |* |	5	0.090	0.127	42.624	0.000
. |. |	. |. |	6	0.055	–0.045	43.417	0.000
. |. |	. |. |	7	0.028	0.042	43.625	0.000
. |. |	* |. |	8	–0.031	–0.075	43.881	0.000
. |. |	. |. |	9	–0.034	0.023	44.190	0.000
. |. |	. |. |	10	–0.027	–0.045	44.374	0.000
. |. |	. |. |	11	–0.013	0.020	44.421	0.000
. |* |	. |* |	12	0.082	0.086	46.190	0.000

Figure 14.6 Correlogram of MA(1)

$$Q_{LB} = T(T+2)\sum_{j=1}^{m}\widehat{\rho}_j^2/(T-j) \tag{14.2}$$

This is also distributed asymptotically as χ^2_m under the null hypothesis. Maddala (1992, p. 540) warns about the inappropriateness of the Q and Q_{LB} statistics for autoregressive models. The arguments against their use are the same as those for not using the Durbin-Watson statistic in autoregressive models. Maddala (1992) suggests the use of LM statistics of the type proposed by Godfrey (1979) to evaluate the adequacies of the ARMA model proposed.

For the consumption series, $T = 49$ and the 95% confidence interval for $\widehat{\rho}_s$ is $0 \pm 1.96\ (1/\sqrt{49})$ which is ± 0.28. Figure 14.2 plots this 95% confidence interval as two solid lines around zero. It is clear that the sample correlogram declines slowly as the number of lags s increase. Moreover, the Q_{LB} statistics which are reported for lags 1, 2, up to 13 are all statistically significant. These were computed using EViews. Based on the sample correlogram and the Ljung-Box statistic, the consumption series is not purely random white noise. Figure 14.7 plots the sample correlogram for $\Delta C_t = C_t - C_{t-1}$. Note that this sample correlogram dies out abruptly after the first lag. Also, the Q_{LB} statistics are not significant after the first lag. This indicates stationarity of the first-differenced consumption series. Problem 3 asks the reader to plot the sample correlogram for personal disposable income and its first difference, and to compute the Ljung-Box Q_{LB} statistic to test for purely White noise based on 13 lags.

A difficult question when modeling economic behavior is to decide on what lags should be in the ARIMA model, or the dynamic regression model. Granger et al. (1995) argue that there are disadvantages in using hypothesis testing to help make model specification decisions based on the data. They recommend instead the use of model selection criteria to make those decisions.

The Box-Jenkins methodology has been popular primarily among forecasters who claimed better performance than simultaneous equations models based upon economic theory. Box-Jenkins models are general enough to allow for nonstationarity and can handle seasonality. However, the Box-Jenkins models suffer from the fact that they are devoid of economic theory and as such they are not designed to test economic hypothesis, or provide estimates of key elasticity parameters. As a consequence, this method cannot be used for simulating the effects

Sample: 1959 2007
Included observations: 48

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>. |****</td><td>. |****</td><td>1</td><td>0.465</td><td>0.465</td><td>11.059</td><td>0.001</td></tr>
<tr><td>. |.</td><td>**|.</td><td>2</td><td>0.065</td><td>–0.194</td><td>11.277</td><td>0.004</td></tr>
<tr><td>.*|.</td><td>.*|.</td><td>3</td><td>–0.129</td><td>–0.099</td><td>12.160</td><td>0.007</td></tr>
<tr><td>. |.</td><td>. |*.</td><td>4</td><td>–0.048</td><td>0.101</td><td>12.288</td><td>0.015</td></tr>
<tr><td>. |.</td><td>. |.</td><td>5</td><td>–0.012</td><td>–0.053</td><td>12.296</td><td>0.031</td></tr>
<tr><td>. |.</td><td>. |.</td><td>6</td><td>0.015</td><td>0.015</td><td>12.309</td><td>0.055</td></tr>
<tr><td>. |.</td><td>. |.</td><td>7</td><td>–0.016</td><td>–0.027</td><td>12.324</td><td>0.090</td></tr>
<tr><td>.*|.</td><td>.*|.</td><td>8</td><td>–0.115</td><td>–0.133</td><td>13.123</td><td>0.108</td></tr>
<tr><td>.*|.</td><td>. |.</td><td>9</td><td>–0.081</td><td>0.056</td><td>13.528</td><td>0.140</td></tr>
<tr><td>. |*.</td><td>. |*.</td><td>10</td><td>0.124</td><td>0.194</td><td>14.501</td><td>0.151</td></tr>
<tr><td>. |*.</td><td>. |.</td><td>11</td><td>0.194</td><td>0.001</td><td>16.944</td><td>0.110</td></tr>
<tr><td>. |*.</td><td>.*|.</td><td>12</td><td>0.098</td><td>–0.027</td><td>17.589</td><td>0.129</td></tr>
<tr><td>. |.</td><td>. |*.</td><td>13</td><td>0.063</td><td>0.120</td><td>17.861</td><td>0.163</td></tr>
<tr><td>. |.</td><td>. |.</td><td>14</td><td>0.034</td><td>–0.017</td><td>17.941</td><td>0.209</td></tr>
<tr><td>. |.</td><td>. |.</td><td>15</td><td>0.027</td><td>0.016</td><td>17.994</td><td>0.263</td></tr>
<tr><td>. |.</td><td>. |.</td><td>16</td><td>0.028</td><td>0.034</td><td>18.051</td><td>0.321</td></tr>
<tr><td>. |.</td><td>. |.</td><td>17</td><td>0.026</td><td>–0.026</td><td>18.105</td><td>0.382</td></tr>
<tr><td>.*|.</td><td>**|.</td><td>18</td><td>–0.152</td><td>–0.193</td><td>19.959</td><td>0.335</td></tr>
<tr><td>. |.</td><td>. |**</td><td>19</td><td>–0.029</td><td>0.279</td><td>20.027</td><td>0.393</td></tr>
<tr><td>. |*.</td><td>. |.</td><td>20</td><td>0.101</td><td>0.016</td><td>20.902</td><td>0.403</td></tr>
</table>

Figure 14.7 Correlogram of First Difference of Consumption

of a tax change or a Federal Reserve policy change. One lesson that economists learned from the Box-Jenkins methodology is that they have to take a hard look at the time-series properties of their variables and properly specify the dynamics of their economic models. Another popular forecasting technique in economics is the *Vector Autoregression* (VAR) methodology proposed by Sims (1980). This will be briefly discussed next.

14.4 Vector Autoregression

Sims (1980) criticized the simultaneous equation literature for the ad hoc restrictions needed for identification and for the ad hoc classification of exogenous and endogenous variables in the system, see Chapter 11. Instead, Sims (1980) suggested *Vector Autoregression* (VAR) models for forecasting macro time-series. VAR assumes that all the variables are endogenous. For example, consider the following three macro-series: money supply, interest rate, and output. VAR models this vector of three endogenous variables as an autoregressive function of their lagged values. VAR models can include some exogenous variables like trends and seasonal dummies, but the whole point is that it does not have to classify variables as endogenous or exogenous. If we allow 5 lags on each endogenous variable, each equation will have 16 parameters to estimate if we include a constant. For example, the money supply equation will be a function of 5 lags on money, 5 lags on the interest rate and 5 lags on output. Since the parameters are different for each equation the total number of parameters in this unrestricted VAR is $3 \times 16 = 48$

parameters. This degrees of freedom problem becomes more serious as the number of lags m and number of equations g increase. In fact, the number of parameters to be estimated becomes $g + mg^2$. With small samples, individual parameters may not be estimated precisely. So, only simple VAR models, can be considered for a short sample. Since this system of equations has the same set of variables in each equation SUR on the system is equivalent to OLS on each equation, see Chapter 10. Under normality of the disturbances, MLE as well as Likelihood Ratio tests can be performed. One important application of LR tests in the context of VAR is its use in determining the choice of lags to be used. In this case, one obtains the log-likelihood for the restricted model with m lags and the unrestricted model with $q > m$ lags. This LR test will be asymptotically distributed as $\chi^2_{(q-m)g^2}$. Once again, the sample size T should be large enough to estimate the large number of parameters $(qg^2 + g)$ for the unrestricted model.

One can of course impose restrictions to reduce the number of parameters to be estimated, but this reintroduces the problem of ad hoc restrictions which VAR was supposed to cure in the first place. Bayesian VAR procedures claim success with forecasting, see Litterman (1986), but again these models have been criticized because they are devoid of economic theory.

VAR models have also been used to test the hypothesis that some variables do not *Granger cause* some other variables.[1] For a two-equation VAR, as long as this VAR is correctly specified and no variables are omitted, one can test, for example, that y_1 does not Granger cause y_2. This hypothesis cannot be rejected if all the m lagged values of y_1 are insignificant in the equation for y_2. This is a simple F-test for the joint significance of the lagged coefficients of y_1 in the y_2 equation. This is asymptotically distributed as $F_{m,T-(2m+1)}$. The problem with the Granger test for non-causality is that it may be sensitive to the number of lags m, see Gujarati (1995). For an extensive analysis of nonstationary VAR models as well as testing and estimation of cointegrating relationships in VAR models, see Hamilton (1994) and Lütkepohl (2001).

14.5 Unit Roots

If $x_t = x_{t-1} + u_t$ where u_t is $\text{IID}(0, \sigma^2)$, then x_t is a *random walk*. Some stock market analysts believe that stock prices follow a random walk, i.e., the price of a stock today is equal to its price yesterday plus a random shock. This is a nonstationary time-series. Any shock to the price of this stock is *permanent* and does not die out like an AR(1) process. In fact, if the initial price of the stock is $x_o = \mu$, then

$$x_1 = \mu + u_1, \; x_2 = \mu + u_1 + u_2, \ldots, \quad \text{and} \quad x_t = \mu + \textstyle\sum_{j=1}^{t} u_j$$

with $E(x_t) = \mu$ and $\text{var}(x_t) = t\sigma^2$ since $u \sim \text{IID}(0, \sigma^2)$. Therefore, the variance of x_t is *dependent* on t and x_t is *not* covariance-stationary. In fact, as $t \to \infty$, so does $\text{var}(x_t)$. However, first differencing x_t we get u_t which is stationary. Figure 14.8 plots the graph of a random walk of size $T = 250$ generated as $x_t = x_{t-1} + \epsilon_t$ with $\epsilon_t \sim \text{IIN}(0, 4)$. Figure 14.9 shows that the autocorrelation function of this random walk process is persistent as s increases. Note that a random walk is an AR(1) model $x_t = \rho x_{t-1} + u_t$ with $\rho = 1$. Therefore, a test for nonstationarity is a test for $\rho = 1$ or a test for a *unit root*.

Using the lag operator L we can write the random walk as $(1 - L)x_t = u_t$ and in general, any autoregressive model in x_t can be written as $A(L)x_t = u_t$ where $A(L)$ is a polynomial in L. If $A(L)$ has $(1 - L)$ as one of its roots, then x_t has a unit root.

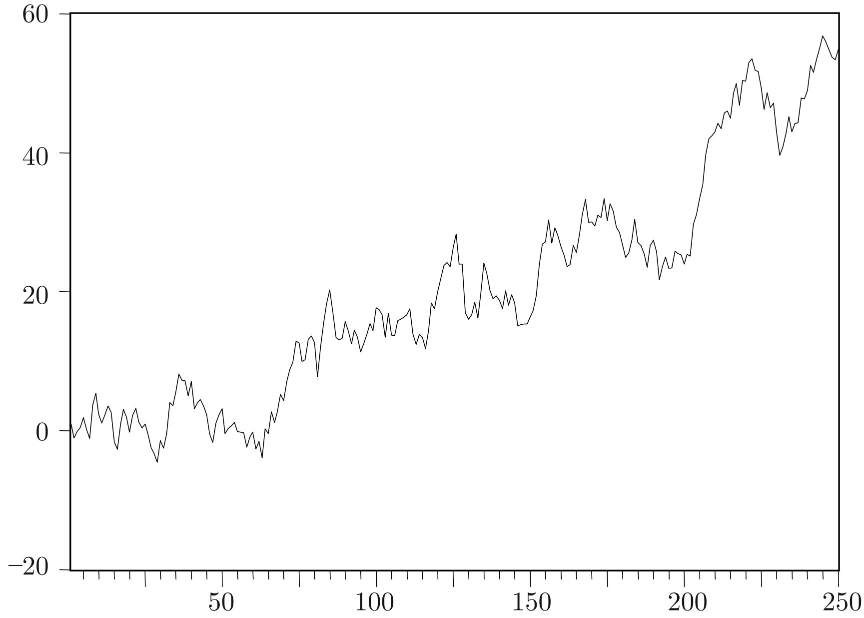

Figure 14.8 Random Walk Process

Subtracting x_{t-1} from both sides of the AR(1) model we get

$$\Delta x_t = (\rho - 1)x_{t-1} + u_t = \delta x_{t-1} + u_t \tag{14.3}$$

where $\delta = \rho - 1$ and $\Delta x_t = x_t - x_{t-1}$ is the first-difference of x_t. A test for H_o; $\rho = 1$ can be obtained by regressing Δx_t on x_{t-1} and testing that H_o; $\delta = 0$. Since u_t is stationary then if $\delta = 0$, $\Delta x_t = u_t$ and x_t is difference stationary, i.e., it becomes stationary after differencing it once. In this case, the original undifferenced series x_t is said to be integrated of order 1 or $I(1)$. If we need to difference x_t twice to make it stationary, then x_t is $I(2)$. A stationary process is by definition $I(0)$. Dickey and Fuller (1979) showed that the usual regression t-statistic for H_o; $\delta = 0$ from (14.3) does *not* have a t-distribution under H_o. In fact, this t-statistic has a non-standard distribution, see Bierens (2001) for a simple derivation of these results. Dickey and Fuller tabulated the critical values of the t-statistic $= (\widehat{\rho} - 1)/s.e.(\widehat{\rho}) = \widehat{\delta}/s.e.(\widehat{\delta})$ using Monte Carlo experiments. These tables have been extended by MacKinnon (1991). If $|t|$ exceeds the critical values, we reject H_o that $\rho = 1$ which also means that we do not reject the hypothesis of stationarity of the time-series. Non-rejection of H_o; $\rho = 1$ means that we do not reject the presence of a unit root and hence the nonstationarity of the time-series. Note that non-rejection of H_o may also be a non-rejection of $\rho = 0.99$. More formally stated, a weakness of unit root tests in general is that they have low power discriminating between a unit root process and a borderline stationary process. In practice, the *Dickey-Fuller* test has been applied in the following three forms:

$$\Delta x_t = \delta x_{t-1} + u_t \tag{14.4}$$

$$\Delta x_t = \alpha + \delta x_{t-1} + u_t \tag{14.5}$$

$$\Delta x_t = \alpha + \beta t + \delta x_{t-1} + u_t \tag{14.6}$$

where t is a time trend. The null hypothesis of the existence of a unit root is H_o; $\delta = 0$. This is the same for (14.4), (14.5) and (14.6), but the critical values for the corresponding t-statistics

Sample: 1 250
Included observations: 250

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |******** |	. |******** |	1	0.980	0.980	242.76	0.000
. |******* |	. |. |	2	0.959	–0.003	476.56	0.000
. |******* |	. |. |	3	0.940	0.004	701.83	0.000
. |******* |	. |. |	4	0.920	–0.013	918.61	0.000
. |******* |	. |. |	5	0.899	–0.044	1126.4	0.000
. |******* |	. |. |	6	0.876	–0.053	1324.6	0.000
. |******* |	. |. |	7	0.855	0.028	1514.2	0.000
. |****** |	. |* |	8	0.837	0.067	1696.7	0.000
. |****** |	. |. |	9	0.821	0.032	1872.8	0.000
. |****** |	. |. |	10	0.804	–0.006	2042.6	0.000
. |****** |	. |. |	11	0.788	–0.007	2206.3	0.000
. |****** |	. |. |	12	0.774	0.030	2364.8	0.000

Figure 14.9 Correlogram of a Random Walk Process

are different in each case. Standard time-series software like EViews give the proper critical values for the Dickey-Fuller statistic. For alternative unit root tests, see Phillips and Perron (1988) and Bierens and Guo (1993). In practice, one should run (14.6) if the series is trended with drift and (14.5) if it is trended without drift. Not including a constant or trend as in (14.4) is unlikely for economic data. The Box-Jenkins approach differences the series and looks at the sample correlogram of the differenced series. The Dickey-Fuller test is a more formal test for the existence of a unit root. Maddala (1992, p. 585) warns the reader to perform both the visual inspection and the unit root test before deciding on whether the time-series process is nonstationary.

If the disturbance term u_t follows a stationary AR(1) process, then the *augmented Dickey-Fuller* test runs the following modified version of (14.6) by including one additional regressor, Δx_{t-1}:

$$\Delta x_t = \alpha + \beta t + \delta x_{t-1} + \lambda \Delta x_{t-1} + \epsilon_t \tag{14.7}$$

In this case, the t-statistic for $\delta = 0$ is a unit root test allowing for first-order serial correlation. This augmented Dickey-Fuller test in (14.7) has the same asymptotic distribution as the corresponding Dickey-Fuller test in (14.6) and the same critical values can be used. Similarly, if u_t follows a stationary AR(p) process, this amounts to adding p extra regressors in (14.6) consisting of $\Delta x_{t-1}, \Delta x_{t-2}, \ldots, \Delta x_{t-p}$ and testing that the coefficient of x_{t-1} is zero. In practice, one does not know the process generating the serial correlation in u_t and the general practice is to include as many lags of Δx_t as is necessary to render the ϵ_t term in (14.7) serially uncorrelated. More lags may be needed if the disturbance term contains Moving Average terms, since a MA term can be thought of as an infinite autoregressive process, see Ng and Perron (1995) for an extensive Monte Carlo on the selection of the truncation lag. Two other important complications when doing unit root tests are: (i) structural breaks in the time-series, like the oil embargo of 1973, tend to bias the standard unit root tests against rejecting the null hypothesis of a unit root, see Perron (1989). (ii) Seasonally adjusted data also tend to bias the standard unit root tests against rejecting the null hypothesis of a unit root, see Ghysels and Perron (1992). For this reason, Davidson and MacKinnon (1993, p. 714) suggest using seasonally unadjusted data whenever available.

For the trended consumption series with drift, the Augmented Dickey-Fuller test using EViews yields the following regression:

$$\begin{aligned} \Delta C_t = &\, 665.60 + 30.57\, t - 0.072\, C_{t-1} + 0.449\, \Delta C_{t-1} + \text{residuals} \\ &\quad (1.80) \quad\;\; (1.60) \quad\;\; (1.42) \qquad\quad (3.17) \end{aligned} \tag{14.8}$$

where the numbers in parentheses are the usually reported t-statistics. The null hypothesis is that the coefficient of C_{t-1} in this regression is zero. Table 14.1 gives the Dickey-Fuller t-statistic (-1.42) and the corresponding 5% critical value (-3.508) tabulated by MacKinnon (1996). Note that @TREND(1959) is the time-trend starting at 1959. The Schwarz criterion found the optimal number of lags of ΔC_{t-1} to be included in this regression is one. Since the p-value is 0.84, we do not reject the null hypothesis of the existence of a unit root. We conclude that C_t is nonstationary. This confirms our finding from the sample correlogram of C_t given in Figure 14.2.

Table 14.1 Dickey-Fuller Test

Null Hypothesis: CONSUMP has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		–1.418937	0.8424
Test critical values:	1% level	–4.165756	
	5% level	–3.508508	
	10% level	–3.184230	

* MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable:	D(CONSUMP)
Method:	Least Squares
Sample (adjusted):	1961 2007
Included observations:	47 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
CONSUMP(–1)	–0.072427	0.051043	–1.418937	0.1631
D(CONSUMP(–1))	0.448898	0.141516	3.172064	0.0028
C	665.6031	370.5970	1.796030	0.0795
@TREND(1959)	30.56963	19.13925	1.597221	0.1175

R-squared	0.291770	Mean dependent var	393.2340
Adjusted R-squared	0.242359	S.D. dependent var	254.9362
S.E. of regression	221.9031	Akaike info criterion	13.72362
Sum squared resid	2117363.	Schwarz criterion	13.88108
Log likelihood	–318.5052	Hannah-Quinn criter.	13.78288
F-statistic	5.904911	Durbin-Watson stat	1.841198
Prob(F-statistic)	0.001816		

One can check whether the first-differenced series is stationary by performing a unit root test on the first-differenced model. Let $\widetilde{C}_t = \Delta C_t$, then run the following regression:

$$\begin{aligned}\Delta\widetilde{C}_t &= \underset{(3.59)}{213.77} - \underset{(4.13)}{0.533}\ \widetilde{C}_{t-1} + \text{residuals}\end{aligned} \tag{14.9}$$

the coefficient of $\widetilde{C}_{t-1}$ has a t-statistic of -4.13 which is smaller than the 5% critical value of -2.925. In other words, we reject the null hypothesis of unit root for the first-differenced series ΔC_t. The same conclusion would have been reached if a linear trend was included besides the constant. We conclude that C_t is $I(1)$.

So far, all tests for unit root have the hypothesis of nonstationarity as the null with the alternative being that the series is stationary. Two unit roots tests with stationarity as the null and nonstationarity as the alternative are given by Kwaitowski et al. (1992) and Leybourne and McCabe (1994). The first test known as KPSS is an analog of the Phillips-Perron test whereas the Leybourne-McCabe test is an analog of the augmented Dickey-Fuller test. Reversing the null may lead to confirmation of stationarity or nonstationarity or may yield conflicting decisions.

14.6 Trend Stationary Versus Difference Stationary

Many macroeconomic time-series that are trending upwards have been characterized as either

$$\text{Trend Stationary: } x_t = \alpha + \beta t + u_t \tag{14.10}$$

or

$$\text{Difference Stationary: } x_t = \gamma + x_{t-1} + u_t \tag{14.11}$$

where u_t is stationary. The first model (14.10) says that the macro-series is stationary except for a *deterministic* trend. $E(x_t) = \alpha + \beta t$ which varies with t. In contrast, the second model (14.11) says that the macro-series is a *random walk with drift*. The drift parameter γ in (14.11) plays the same role as the β parameter in (14.10), since both cause x_t to trend upwards over time. Model (14.10) is consistent with economists introducing a time trend in the regression. This has the same effect as detrending each variable in the regression rendering it stationary, see the Frisch-Waugh-Lovell Theorem in Chapter 7. This detrending is valid only if model (14.10) is true for every series in the regression. Model (14.11) on the other hand, requires differencing to obtain a stationary series. Detrending and differencing are two completely different remedies. What is valid for one model is not valid for the other. The choice between (14.10) and (14.11) is based on a test for the existence of a unit root. Essential reading on these two models are Nelson and Plosser (1982) and Stock and Watson (1988). Nelson and Plosser applied the Dickey-Fuller test to a wide range of historical macro time-series for the U.S. economy and found that all of these series were difference stationary, with the exception of the unemployment rate. Plosser and Schwert (1978) argued that for most economic macro time-series, it is best to difference the data rather than work with levels. The reasoning is that if these series are difference stationary and we run regressions in levels, the usual properties of our estimators as well as the distributions of the associated test statistics are invalid. On the other hand, if the true model is a regression in levels with the data series being trend stationary, differencing the model will produce a Moving

Average error term and at worst, ignoring it will lead to loss in efficiency. It is important to emphasize, that for nonstationary variables, the standard asymptotic theory does not apply, see problems 6 and 7, and that t and F-statistics obtained from regressions using these variables may have non-standard distributions, see Durlauf and Phillips (1988).

Granger and Newbold (1974) demonstrated some of the problems associated with regressing nonstationary time-series on each other. In fact, they showed that if x_t and y_t are independent random walks, then one should expect to find no evidence of a relationship when one regresses y_t on x_t. In other words, the estimate of β in the regression $y_t = \alpha + \beta x_t + u_t$ should be near zero and its associated t-statistic insignifiant. In fact, for a sample size of 50 and a 100 replications, Granger and Newbold found $|t| \leq 2$ on only 23 occasions, $2 < |t| \leq 4$ on 24 occasions, and $|t| > 4$ on 53 occasions. Granger and Newbold (1974) called this phenomenon *spurious regression* since it finds a significant relationship between the two time-series when none exists. Hence, one should be cautious when running time-series regressions involving unit root processes. High R^2 and significant t-statistics from OLS regressions may be hiding nonsense results. Phillips (1986) studied the asymptotic properties of the least squares spurious regression model and confirmed these simulation results. In fact, Phillips showed that the t-statistic for H_o; $\beta = 0$ converges in probability to ∞ as $T \to \infty$. This means that the t-statistic will reject H_o; $\beta = 0$ with probability 1 as $T \to \infty$. If both x_t and y_t are independent trend stationary series generated as described in (14.10), then the R^2 of the regression of y_t on x_t will tend to one as $T \to \infty$, see Davidson and MacKinnon (1993, p. 671). For a summary of several extensions of these results, see Granger (2001).

14.7 Cointegration

Let us continue with our consumption-income example. In Chapter 5, we regressed C_t on Y_t and obtained

$$\begin{aligned} C_t = \underset{(219.56)}{-1343.31} + \underset{(0.011)}{0.979}\; Y_t + \text{residuals} \end{aligned} \tag{14.12}$$

with $R^2 = 0.994$ and D.W. $= 0.18$. We have shown that C_t and Y_t are nonstationary series and that both are $I(1)$, see also problem 3. The regression in (14.12) could be a *spurious regression* owing to the fact that we regressed a nonstationary series on another. This invalidates the t and F-statistics of regression (14.12). Since both C_t and Y_t are integrated of the same order, and Figure 14.1 shows that they are trending upwards together, this random walk may be in unison. This is the idea behind *cointegration*. C_t and Y_t are cointegrated if there exists a linear combination of C_t and Y_t that yields a stationary series. More formally, if C_t and Y_t are both $I(1)$ but there exist a linear combination $C_t - \alpha - \beta Y_t = u_t$ which is $I(0)$, then C_t and Y_t are *cointegrated* and β is the cointegrating parameter. This idea can be extended to a vector of more than two time-series. This vector is cointegrated if the components of this vector have a unit root and there exists a linear combination of this vector that is stationary. Such a cointegrating relationship can be interpreted as a stable long-run relationship between the components of this time-series vector. Economic examples of long-run relationship include the quantity theory of money, purchasing power parity and the permanent income theory of consumption. The important point to emphasize here is that differencing these nonstationary time-series destroys potential valuable information about the long-run relationship between these economic variables. The theory of cointegration tries to estimate this long-run relationship using the non-stationary series themselves, rather than their first differences. In order to explain this, we state (without proof) one of the implications of the *Granger Representation Theorem*, namely, that a set of cointegrated variables will have an *Error-Correction Model* (ECM) representation. Let us illustrate with an example.

A Cointegration Example

This is based on Engle and Granger (1987). Assume that C_t and Y_t for $t = 1, 2, \ldots, T$ are $I(1)$ processes generated as follows:

$$C_t - \beta Y_t = u_t \quad \text{with} \quad u_t = \rho u_{t-1} + \epsilon_t \quad \text{and} \quad |\rho| < 1 \tag{14.13}$$

$$C_t - \alpha Y_t = \nu_t \quad \text{with} \quad \nu_t = \nu_{t-1} + \eta_t \quad \text{and} \quad \alpha \neq \beta \tag{14.14}$$

In other words, u_t follows a stationary AR(1) process, while ν_t follows a random walk. Suppose that $\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix}$ are independent bivariate normal random variables with mean zero and variance $\Sigma = [\sigma_{ij}]$ for $i, j = 1, 2$. First, we obtain the *reduced form* representation of Y_t and C_t in terms of u_t and ν_t. This is given by

$$C_t = \frac{\alpha}{(\alpha - \beta)} u_t + \frac{\beta}{(\alpha - \beta)} \nu_t \tag{14.15}$$

$$Y_t = \frac{1}{(\alpha - \beta)} u_t - \frac{1}{(\alpha - \beta)} \nu_t \tag{14.16}$$

Since u_t is $I(0)$ and ν_t is $I(1)$, we conclude from (14.15) and (14.16) that C_t and Y_t are in fact $I(1)$ series. In terms of the usual *order condition* for identification considered in Chapter 11, the system of equations given by (14.13) and (14.14) are not identified because there are no exclusion restrictions on either equation. However, if we take a linear combination of the two structural equations given in (14.13) and (14.14), the disturbance of the resulting linear combination is neither a stationary AR(1) process nor a random walk. Hence, both (14.13) and (14.14) are identified. Note that if $\rho = 1$, then u_t is a random walk and the linear combination of u_t and ν_t is also a random walk. In this case, neither (14.13) nor (14.14) are identified.

In the Engle-Granger terminology, $C_t - \beta Y_t$ is the cointegrating relationship and $(1, -\beta)$ is the cointegrating vector. This cointegrating relationship is unique. The proof is by contradiction. Assume there is another cointegrating relationship $C_t - \gamma Y_t$ that is $I(0)$, then the difference between the two cointegrating relationships yields $(\gamma - \beta)Y_t$. This is also $I(0)$. This can only happen for every value of Y_t, which is $I(1)$, if and only if $\beta = \gamma$.

Difference both equations in (14.13) and (14.14) and write both differenced equations as a system of two equations in $(\Delta C_t, \Delta Y_t)'$, one gets:

$$\begin{bmatrix} 1 & -\beta \\ 1 & -\alpha \end{bmatrix} \begin{bmatrix} \Delta C_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \Delta u_t \\ \Delta \nu_t \end{bmatrix} = \begin{bmatrix} \epsilon_t + (\rho - 1)C_{t-1} - \beta(\rho - 1)Y_{t-1} \\ \eta_t \end{bmatrix} \tag{14.17}$$

where the second equality is obtained by replacing $\Delta \nu_t$ by η_t, Δu_t by $(\rho - 1)u_{t-1} + \epsilon_t$, and substituting for u_{t-1} its value $(C_{t-1} - \beta Y_{t-1})$. Post-multiplying (14.17) by the inverse of the first matrix, one can show, see problem 9, that the resulting solution is the following VAR model:

$$\left[\begin{array}{c}\Delta C_t\\ \Delta Y_t\end{array}\right]=\frac{1}{(\beta-\alpha)}\left[\begin{array}{cc}-\alpha(\rho-1) & \alpha\beta(\rho-1)\\ -(\rho-1) & \beta(\rho-1)\end{array}\right]\left(\begin{array}{c}C_{t-1}\\ Y_{t-1}\end{array}\right)+\left(\begin{array}{c}h_t\\ g_t\end{array}\right) \tag{14.18}$$

where h_t and g_t are linear combinations of ϵ_t and η_t. Note that if $\rho = 1$, then the level variables C_{t-1} and Y_{t-1} drop from the VAR equations. Let $Z_t = C_t - \beta Y_t$ and define $\delta = (\rho-1)/(\beta-\alpha)$. Then the VAR representation in (14.18) can be written as follows:

$$\Delta C_t = -\alpha\delta Z_{t-1} + h_t \tag{14.19}$$

$$\Delta Y_t = -\delta Z_{t-1} + g_t \tag{14.20}$$

This is the *Error-Correction Model* (ECM) representation of the original model. Z_{t-1} is the error correction term. It represents a disequilibrium term showing the departure from long-run equilibrium, see section 6.4. Note that if $\rho = 1$, then $\delta = 0$ and Z_{t-1} drops from both ECM equations. As Banerjee et al. (1993, p. 139) explain, this ECM representation is a noteworthy "...contribution to resolving, or synthesizing, the debate between time-series analysts and those favoring econometric methods." The former considered only differenced time-series that can be legitimately assumed stationary, while the latter focused on equilibrium relationships expressed in levels. The former wiped out important long-run relationships by first differencing them, while the latter ignored the spurious regression problem. In contrast, the ECM allows the use of first differences and levels from the cointegrating relationship. For more details, see Banerjee et al. (1993). A simple two-step procedure for estimating cointegrating relationships is given by Engle and Granger (1987). In the first step, the OLS estimator of β is obtained by regressing C_t on Y_t. This can be shown to be *superconsistent*, i.e., plim $T(\widehat{\beta}_{OLS} - \beta) \to 0$ as $T \to \infty$. Using $\widehat{\beta}_{OLS}$ one obtains $\widehat{Z}_t = C_t - \widehat{\beta}_{OLS}Y_t$. In the second step, using $\widehat{Z}_{t-1}$ rather than Z_{t-1}, apply OLS to estimate the ECM in (14.19) and (14.20). Extensive Monte Carlo experiments have been conducted by Banerjee et al. (1993) to investigate the bias of β in small samples. This is pursued further in problem 9. An alternative estimation procedure is the maximum likelihood approach suggested by Johansen (1988). This is beyond the scope of this book. See Dolado et al. (2001) for a lucid summary of the cointegration literature.

A formal test for cointegration is given by Engle and Granger (1987) who suggest running regression (14.12) and testing that the residuals do not have a unit root. In other words, run a Dickey-Fuller test or its augmented version on the resulting residuals from (14.12). In fact, if C_t and Y_t are not cointegrated, then any linear combination of them would be nonstationary including the residuals of (14.12). Since these tests are based on residuals, their asymptotic distributions are not the same as those of the corresponding ordinary unit roots tests. Asymptotic critical values for these tests can be found in Davidson and MacKinnon (1993, p. 722). For our consumption regression the following Dickey-Fuller test is obtained on the residuals:

$$\begin{array}{l}\Delta\widehat{u}_t = -1.111 - \ 0.094\ \widehat{u}_t + \text{residuals}\\ \qquad\qquad (0.04) \quad\ (1.50)\end{array} \tag{14.21}$$

the Davidson and MacKinnon (1993) asymptotic 5% critical value for this t-statistic is -2.92 and its p-value is 0.52. Therefore, we cannot reject the hypothesis that $\widehat{u}_t$ is nonstationary. We have also included a trend and one lag of the first-differenced residuals. The resulting augmented Dickey-Fuller test did not reject the existence of a unit root. Therefore, C_t and Y_t are *not* cointegrated. This suggests that the relationship estimated in (14.12) is spurious.

Table 14.2 Johansen Cointegration Test

Sample (adjusted): 1961 2007
Included observations: 47 after adjustments
Trend assumption: Linear deterministic trend
Series: CONSUMP Y
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.120482	6.322057	15.49471	0.6575
At most 1	0.006112	0.288141	3.841466	0.5914

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
** MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None	0.120482	6.033916	14.26460	0.6089
At most 1	0.006112	0.288141	3.841466	0.5914

Max-eigenvalue test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
** MacKinnon-Haug-Michelis (1999) p-values

Regressing an $I(1)$ series on another lead to spurious results unless they are cointegrated. Of course, other $I(1)$ series may have been erroneously excluded from (14.12) which when included may result in a cointegrating relationship among the resulting variables. In other words, C_t and Y_t may *not* be cointegrated because of an omitted variables problem. Table 14.2 gives the Johansen (1995) cointegration test reported by EViews which is beyond the scope of this book. The null hypothesis is that of no cointegration or at most one cointegration relationship. Both hypotheses are not rejected by the *trace* and *maximum eigenvalue* tests.

14.8 Autoregressive Conditional Heteroskedasticity

Financial time-series such as foreign exchange rates, inflation rates and stock prices may exhibit some volatility which varies over time. In the case of inflation or foreign exchange rates this could be due to changes in the Federal Reserve's policies. In the case of stock prices this could be due to rumors about a certain company's merger or takeover. This suggests that the variance of these time-series may be heteroskedastic. Engle (1982) modeled this heteroskedasticity by relating the conditional variance of the disturbance term at time t to the size of the squared disturbance terms in the recent past. A simple *Autoregressive Conditionally Heteroskedastic*

(ARCH) model is given by

$$\sigma_t^2 = E(u_t^2/\zeta_t) = \gamma_o + \gamma_1 u_{t-1}^2 + .. + \gamma_p u_{t-p}^2 \tag{14.22}$$

where ζ_t denotes the information set upon which the variance of u_t is to be conditioned. This typically includes all the information available prior to period t. In (14.22), the variance of u_t conditional on the information prior to period t is an autoregressive function of order p in squared lagged values of u_t. This is called an ARCH(p) process. Since (14.22) is a variance, this means that all the γ_i's for $i = 0, 1, \ldots, p$ have to be non-negative. Engle (1982) showed that a simple test for homoskedasticity, i.e., H_o; $\gamma_1 = \gamma_2 = .. = \gamma_p = 0$, can be based upon an ordinary F-test which regresses the squared OLS residuals (e_t^2) on their lagged values ($e_{t-1}^2, \ldots, e_{t-p}^2$) and a constant. The F-statistic tests the joint significance of the regressors and is reported by most regression packages. Alternatively, one can compute T times the centered R^2 of this regression and this is distributed as χ_p^2 under the null hypothesis H_o. This test resembles the usual homoskedasticity tests studied in Chapter 5 except that the squared OLS residuals are regressed upon their lagged values rather than some explanatory variables.

The simple ARCH(1) process

$$\sigma_t^2 = \gamma_o + \gamma_1 u_{t-1}^2 \tag{14.23}$$

can be generated as follows: $u_t = [\gamma_o + \gamma_1 u_{t-1}^2]^{1/2}\epsilon_t$ where $\epsilon_t \sim \text{IID}(0, 1)$. Note that the simplifying variance of unity for ϵ_t can be achieved by rescaling the parameters γ_o and γ_1. In this case, the conditional mean of u_t is given by

$$E(u_t/\zeta_t) = [\gamma_o + \gamma_1 u_{t-1}^2]^{1/2} E(\epsilon_t/\zeta_t) = 0$$

since u_{t-1}^2 is known at time t. Similarly, the conditional variance can be easily obtained from

$$E(u_t^2/\zeta_t) = [\gamma_o + \gamma_1 u_{t-1}^2] E(\epsilon_t^2/\zeta_t) = \gamma_o + \gamma_1 u_{t-1}^2$$

since $E(\epsilon_t^2) = 1$. Also, the conditional covariances can be easily shown to be zero since

$$E(u_t u_{t-s}/\zeta_t) = u_{t-s} E(u_t/\zeta_t) = 0 \quad \text{for} \quad s = 1, 2, \ldots, t.$$

The unconditional mean can be obtained by taking repeated conditional expectations period by period until we reach the initial period, see the Appendix to Chapter 2. For example, taking the conditional expectation of $E(u_t/\zeta_t)$ based on information prior to period $t-1$, we get

$$E[E(u_t/\zeta_t)/\zeta_{t-1}] = E(0/\zeta_{t-1}) = 0$$

It is clear that all prior conditional expectations of zero will be zero so that $E(u_t) = 0$. Similarly, taking the conditional expectations of $E(u_t^2/\zeta_t)$ based on information prior to period $t-1$, we get

$$E[E(u_t^2/\zeta_t)/\zeta_{t-1}] = \gamma_o + \gamma_1 E[u_{t-1}^2/\zeta_{t-1}] = \gamma_o + \gamma_1(\gamma_o + \gamma_1 u_{t-2}^2) = \gamma_o(1 + \gamma_1) + \gamma_1^2 u_{t-2}^2$$

By taking repeated conditional expectations one period at a time we finally get

$$E(u_t^2) = \gamma_o(1 + \gamma_1 + \gamma_1^2 + .. + \gamma_1^{t-1}) + \gamma_1^t u_o^2 \tag{14.24}$$

As $t \to \infty$, the unconditional variance of u_t is given by $\sigma^2 = \text{var}(u_t) = \gamma_o/(1-\gamma_1)$ for $|\gamma_1| < 1$ and $\gamma_o > 0$. Therefore, the ARCH(1) process is homoskedastic.

ARCH models can be estimated using feasible GLS or maximum likelihood methods. Alternatively, one can use a double-length regression procedure suggested by Davidson and MacKinnon (1993) to obtain (i) one-step efficient estimates starting from OLS estimates or (ii) the maximum likelihood estimates. Here we focus on the feasible GLS procedure suggested by Engle (1982). For the regression model

$$y = X\beta + u \tag{14.25}$$

where y is $T \times 1$ and X is $T \times k$. First, obtain the OLS estimates $\widehat{\beta}_{OLS}$ and the OLS residuals e. Second, perform the following regression: $e_t^2 = a_o + a_1 e_{t-1}^2 +$ residuals. This yields a test for homoskedasticity. Third, compute $\widehat{\sigma}_t^2 = a_o + a_1 e_{t-1}^2$ and regress $[(e_t^2/\widehat{\sigma}_t) - 1]$ on $(1/\widehat{\sigma}_t)$ and $(e_{t-1}^2/\widehat{\sigma}_t)$. Call the regression estimates d_a. One updates $a' = (a_o, a_1)$ by computing $\widehat{a} = a + d_a$. Fourth, recompute $\widehat{\sigma}_t^2$ using the updated $\widehat{a}$ from step 3, and form the set of regressors $x_{tj}r_t$ for $j = 1, \ldots, k$, where

$$r_t = \left[\frac{1}{\widehat{\sigma}_t} + 2\left(\frac{\widehat{a}_1 e_t}{\widehat{\sigma}_{t+1}}\right)^2\right]^{1/2} \tag{14.26}$$

Finally, regress $(e_t s_t/r_t)$ where

$$s_t = \frac{1}{\widehat{\sigma}_t} - \frac{\widehat{a}_1}{\widehat{\sigma}_{t+1}}\left(\frac{e_{t+1}^2}{\widehat{\sigma}_{t+1}} - 1\right)$$

on $x_{tj}r_t$ for $j = 1, \ldots, k$ and obtain the least squares coefficients d_β. Update the estimate of β by computing $\widehat{\beta} = \widehat{\beta}_{OLS} + d_\beta$. This procedure can run into problems if the $\widehat{\sigma}_t^2$ are not all positive, see Judge et al. (1985) and Engle (1982) for details.

The ARCH model has been generalized by Bollerslev (1986). The Generalized ARCH (GARCH (p, q)) model can be written as

$$\sigma_t^2 = \gamma_o + \textstyle\sum_{i=1}^p \gamma_i u_{t-i}^2 + \sum_{j=1}^q \delta_j \sigma_{t-j}^2 \tag{14.27}$$

In this case, the conditional variance of u_t depends upon q of its lagged values as well as p squared lagged values of u_t. The simple GARCH $(1,1)$ model is given by

$$\sigma_t^2 = \gamma_o + \gamma_1 u_{t-1}^2 + \delta_1 \sigma_{t-1}^2 \tag{14.28}$$

An LM test for GARCH (p, q) turns out to be equivalent to testing ARCH $(p+q)$. This simply regresses squared OLS residuals on $(p+q)$ of its squared lagged values. The test statistic is T times the uncentered R^2 and is asymptotically distributed as χ_{p+q}^2 under the null of homoskedasticity.

In conclusion, a lot of basic concepts have been introduced in this chapter and we barely scratched the surface. Hopefully, this will motivate the reader to take the next econometrics time series course.

Table 14.3 GARCH (1,1) model

Dependent Variable: CONSUMP
Method: ML – ARCH (Marquardt) – Normal distribution
Sample: 1959 2007
Included observations: 49
Convergence achieved after 19 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(–1)^2 + C(5)*GARCH(–1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	–1435.888	226.3933	–6.342449	0.0000
Y	0.986813	0.011923	82.76452	0.0000
	Variance Equation			
C	118728.2	87402.35	1.358410	0.1743
RESID(–1)^ 2	1.068561	0.326091	3.276885	0.0010
GARCH(–1)	–0.380949	0.187656	–2.030036	0.0424
R-squared	0.993542	Mean dependent var		16749.10
Adjusted R-squared	0.992955	S.D. dependent var		5447.060
S.E. of regression	457.1938	Akaike info criterion		14.86890
Sum squared resid	9197153.	Schwarz criterion		15.06194
Log likelihood	–359.2880	Hannah-Quinn criter.		14.94214
F-statistic	1692.353	Durbin-Watson stat		0.178409
Prob(F-statistic)	0.000000			

Note

1. Granger causality has been developed by Granger (1969). For another definition of causality, see Sims (1972). Also, Chamberlain (1982) for a discussion on when these two definitions are equivalent.

Problems

1. For the *AR(1) model*

$$y_t = \rho y_{t-1} + \epsilon_t \quad t = 1, 2, \ldots, T; \quad \text{with} \quad |\rho| < 1 \quad \text{and} \quad \epsilon_t \sim \text{IIN}(0, \sigma_\epsilon^2)$$

   (a) Show that if $y_o \sim N(0, \sigma_\epsilon^2/1 - \rho^2)$, then $E(y_t) = 0$ for all t and $\text{var}(y_t) = \sigma_\epsilon^2/(1 - \rho^2)$ so that the mean and variance are independent of t. Note that if $\rho = 1$ then $\text{var}(y_t)$ is ∞. If $|\rho| > 1$ then $\text{var}(y_t)$ is negative!

   (b) Show that $\text{cov}(y_t, y_{t-s}) = \rho^s \sigma^2$ which is only dependent on s, the distance between the two time periods. Conclude from parts (a) and (b) that this AR(1) model is *weakly stationary*.

   (c) Generate the above AR(1) series for $T = 250$, $\sigma_\epsilon^2 = 0.25$ and various values of $\rho = \pm 0.9, \pm 0.8$, ± 0.5, ± 0.3 and ± 0.1. Plot the AR(1) series and the autocorrelation function ρ_s versus s.

2. For the *MA(1) model*

$$y_t = \epsilon_t + \theta \epsilon_{t-1} \quad t = 1, 2, \ldots, T; \quad \text{with} \quad \epsilon_t \sim \text{IIN}(0, \sigma_\epsilon^2)$$

(a) Show that $E(y_t) = 0$ and $\text{var}(y_t) = \sigma_\epsilon^2(1+\theta^2)$ so that the mean and variance are independent of t.

(b) Show that $\text{cov}(y_t, y_{t-1}) = \theta\sigma_\epsilon^2$ and $\text{cov}(y_t, y_{t-s}) = 0$ for $s > 1$ which is only dependent on s, the distance between the two time periods. Conclude from parts (a) and (b) that this MA(1) model is *weakly stationary*.

(c) Generate the above MA(1) series for $T = 250$, $\sigma_\epsilon^2 = 0.25$ and various values of $\theta = 0.9$, 0.8, 0.5, 0.3 and 0.1. Plot the MA(1) series and the autocorrelation function versus s.

3. Using the *consumption-personal disposable income* data for the U.S. used in this chapter:

(a) Compute the sample autocorrelation function for personal disposable income (Y_t). Plot the sample correlogram. Repeat for the first-differenced series (ΔY_t). Compute the Ljung-Box Q_{LB} statistic, test that H_o; $\rho_s = 0$ for $s = 1, \ldots, 20$.

(b) Run the Augmented Dickey-Fuller test for the existence of a unit root in personal disposable income (Y_t).

(c) Define $\widetilde{Y}_t = \Delta Y_t$ and run $\Delta\widetilde{Y}_t$ on $\widetilde{Y}_{t-1}$ and a constant and trend. Test that the first-differenced series of personal disposable income is stationary. What do you conclude? Is Y_t an $I(1)$ process?

(d) Replicate the regression in (14.21) and verify the Engle-Granger (1987) test for cointegration.

(e) Replicate the GARCH(1,1) model given in Table 14.3.

(f) Repeat parts (a) through (e) using log$C$ and log$Y$. Are there any changes in the above results?

4. (a) Generate $T = 25$ observations on x_t and y_t as independent random walks with $\text{IIN}(0, 1)$ disturbances. Run the regression $y_t = \alpha + \beta x_t + u_t$ and test the null hypothesis H_o; $\beta = 0$ using the usual t-statistic at the 1%, 5% and 10% levels. Repeat this experiment 1000 times and report the frequency of rejections at each significance level. What do you conclude?

(b) Repeat part (a) for $T = 100$ and $T = 500$.

(c) Repeat parts (a) and (b) generating x_t and y_t as independent random walks with drift as described in (14.11), using $\text{IIN}(0, 1)$ disturbances. Let $\gamma = 0.2$ for both series.

(d) Repeat parts (a) and (b) generating x_t and y_t as independent trend stationary series as described in (14.10), using $\text{IIN}(0, 1)$ disturbances. Let $\alpha = 1$ and $\beta = 0.04$ for both series.

(e) Report the frequency distributions of the R^2 statistics obtained in parts (a) through (d) for each sample size and method of generating the time-series. What do you conclude? **Hint:** See the Monte Carlo experiments in Granger and Newbold (1974), Davidson and MacKinnon (1993) and Banerjee, Dolado, Galbraith and Hendry (1993).

5. For the *Money Supply, GNP and interest rate* series data for the U.S. given on the Springer web site as MACRO.ASC, fit a VAR three equation model using:

(a) Two lags on each variable.

(b) Three lags on each variable.

(c) Compute the Likelihood Ratio test for part (a) versus part (b).

(d) For the two-equation VAR of Money Supply and interest rate with three lags on each variable, test that the interest rate does not Granger cause the money supply?

(e) How sensitive are the tests in part (d) if we had used only two lags on each variable.

6. For the *simple Deterministic Time Trend Model*

$$y_t = \alpha + \beta t + u_t \qquad t = 1, .., T$$

where $u_t \sim \text{IIN}(0, \sigma^2)$.

(a) Show that

$$\begin{pmatrix} \widehat{\alpha}_{OLS} - \alpha \\ \widehat{\beta}_{OLS} - \beta \end{pmatrix} = (X'X)X'u = \begin{bmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} tu_t \end{bmatrix}$$

where the t-th observation of X, the matrix of regressors, is $[1, t]$.

(b) Use the results that $\sum_{t=1}^{T} t = T(T+1)/2$ and $\sum_{t=1}^{T} t^2 = T(T+1)(2T+1)/6$ to show that plim $(X'X/T)$ as $T \to \infty$ is not a positive definite matrix.

(c) Use the fact that

$$\begin{pmatrix} \sqrt{T}(\widehat{\alpha}_{OLS} - \alpha) \\ T\sqrt{T}(\widehat{\beta}_{OLS} - \beta) \end{pmatrix} = A(X'X)^{-1}AA^{-1}(X'u) = (A^{-1}(X'X)A^{-1})^{-1}A^{-1}(X'u)$$

$$\text{where} \qquad A = \begin{pmatrix} \sqrt{T} & 0 \\ 0 & T\sqrt{T} \end{pmatrix}$$

is the 2×2 nonsingular matrix, to show that plim $(A^{-1}(X'X)A^{-1})$ is the finite positive definite matrix

$$Q = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \quad \text{and} \qquad A^{-1}(X'u) = \begin{pmatrix} \sum_{t=1}^{T} u_t/\sqrt{T} \\ \sum_{t=1}^{T} tu_t/T\sqrt{T} \end{pmatrix}$$

(d) Show that $z_1 = \sum_{t=1}^{T} u_t/\sqrt{T}$ is $N(0, \sigma^2)$ and $z_2 = \sum_{t=1}^{T} tu_t/T\sqrt{T}$ is $N(0, \sigma^2(T+1)(2T+1)/6T^2)$ with $\text{cov}(z_1, z_2) = (T+1)\sigma^2/2T$, so that

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(0, \sigma^2 \begin{pmatrix} 1 & \frac{T+1}{2T} \\ \frac{T+1}{2T} & \frac{(T+1)(2T+1)}{6T^2} \end{pmatrix}\right).$$

Conclude that as $T \to \infty$, the asymptotic distribution of $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ is $N(0, \sigma^2 Q)$.

(e) Using the results in parts (c) and (d), conclude that the asymptotic distribution of $\begin{pmatrix} \sqrt{T}(\widehat{\alpha}_{OLS} - \alpha) \\ T\sqrt{T}(\widehat{\beta}_{OLS} - \beta) \end{pmatrix}$ is $N(0, \sigma^2 Q^{-1})$. Since $\widehat{\beta}_{OLS}$ has the factor $T\sqrt{T}$ rather than the usual $\sqrt{T}$, it is said to be *superconsistent*. This means that not only does $(\widehat{\beta}_{OLS} - \beta)$ converge to zero in probability limits, but so does $T(\widehat{\beta}_{OLS} - \beta)$. Note that the normality assumption is not needed for this result. Using the central limit theorem, all that is needed is that u_t is White noise with finite fourth moments, see Sims, Stock and Watson (1990) or Hamilton (1994).

7. *Test of Hypothesis with a Deterministic Time Trend Model.* This is based on Hamilton (1994). In problem 6, we showed that $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$ converged at different rates, $\sqrt{T}$ and $T\sqrt{T}$ respectively. Despite this fact, the usual least squares t and F-statistics are asymptotically valid even when the u_t's are not Normally distributed.

(a) Show that $s^2 = \sum_{t=1}^{T}(y_t - \widehat{\alpha}_{OLS} - \widehat{\beta}_{OLS}t)^2/(T-2)$ has plim $s^2 = \sigma^2$.

(b) In order to test H_o; $\alpha = \alpha_o$, the usual least squares package computes

$$t_\alpha = (\widehat{\alpha}_{OLS} - \alpha_o)/[s^2(1,0)(X'X)^{-1}(1,0)']^{1/2}$$

where $(X'X)$ is given in problem 6. Multiply the numerator and denominator by $\sqrt{T}$ and use the results of part (c) of problem 6 to show that this t-statistic has the same asymptotic distribution as $t_\alpha^* = \sqrt{T}(\widehat{\alpha}_{OLS} - \alpha_o)/\sigma\sqrt{q^{11}}$ where q^{11} is the $(1,1)$ element of Q^{-1} defined in problem 6. t_α^* has an asymptotic $N(0,1)$ distribution using the results of part (e) in problem 6.

(c) Similarly, to test H_o; $\beta = \beta_o$, the usual least squares package computes

$$t_\beta = (\widehat{\beta}_{OLS} - \beta)/[s^2(0,1)(X'X)^{-1}(0,1)']^{1/2}.$$

Multiply the numerator and denominator by $T\sqrt{T}$ and use the results of part (c) of problem 6 to show that this t-statistic has the same asymptotic distribution as $t_\beta^* = T\sqrt{T}(\widehat{\beta}_{OLS} - \beta)/\sigma\sqrt{q^{22}}$ where q^{22} is the $(2,2)$ element of Q^{-1} defined in problem 6. t_β^* has an asymptotic $N(0,1)$ distribution using the results of part (e) in problem 6.

8. *A Random Walk Model.* This is based on Fuller (1976) and Hamilton (1994). Consider the following random walk model

$$y_t = y_{t-1} + u_t \quad t = 0, 1, \ldots, T \quad \text{where} \quad u_t \sim \text{IIN}(0, \sigma^2) \quad \text{and} \quad y_o = 0.$$

(a) Show that y_t can be written as $y_t = u_1 + u_2 + .. + u_t$ with $E(y_t) = 0$ and $\text{var}(y_t) = t\sigma^2$ so that $y_t \sim N(0, t\sigma^2)$.

(b) Square the random walk equation $y_t^2 = (y_{t-1} + u_t)^2$ and solve for $y_{t-1}u_t$. Sum this over $t = 1, 2, \ldots, T$ and show that

$$\sum_{t=1}^{T} y_{t-1}u_t = (y_T^2/2) - \sum_{t=1}^{T} u_t^2/2$$

Divide by $T\sigma^2$ and show that $\sum_{t=1}^{T} y_{t-1}u_t/T\sigma^2$ is asymptotically distributed as $(\chi_1^2 - 1)/2$. **Hint:** Use the fact that $y_T \sim N(0, T\sigma^2)$.

(c) Using the fact that $y_{t-1} \sim N(0, (t-1)\sigma^2)$ show that $E\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \sigma^2 T(T-1)/2$. **Hint:** Use the expression for $\sum_{t=1}^{T} t$ in problem 6.

(d) Suppose we had estimated an AR(1) model rather than a random walk, i.e., $y_t = \rho y_{t-1} + u_t$ when the true $\rho = 1$. The OLS estimate is

$$\widehat{\rho} = \sum_{t=1}^{T} y_{t-1}y_t / \sum_{t=1}^{T} y_{t-1}^2 = \rho + \sum_{t=1}^{T} y_{t-1}u_t / \sum_{t=1}^{T} y_{t-1}^2$$

Show that

$$\text{plim } T(\widehat{\rho} - \rho) = \text{plim} \frac{\sum_{t=1}^{T} y_{t-1}u_t/T\sigma^2}{\sum_{t=1}^{T} y_{t-1}^2/T^2\sigma^2} = 0$$

Note that the numerator was considered in part (b), while the denominator was considered in part (c). One can see that the asymptotic distribution of $\widehat{\rho}$ when $\rho = 1$ is a ratio of $(\chi_1^2 - 1)/2$ random variable to a non-standard distribution in the denominator which is beyond the scope of this book, see Hamilton (1994) or Fuller (1976) for further details. The object of this exercise is to show that if $\rho = 1$, $\sqrt{T}(\widehat{\rho} - \rho)$ is no longer normal as in the standard stationary least squares regression with $|\rho| < 1$. Also, to show that for the nonstationary (random walk) model, $\widehat{\rho}$ converges at a faster rate (T) than for the stationary case $(\sqrt{T})$. From part (c) it is clear that one has to divide the denominator of $\widehat{\rho}$ by T^2 rather than T to get a convergent distribution.

9. Consider the *cointegration example* given in (14.13) and (14.14).

   (a) Verify equations (14.15)-(14.20).

   (b) Show that the OLS estimator of β obtained by regressing C_t on Y_t is *superconsistent*, i.e., show that plim $T(\widehat{\beta}_{OLS} - \beta) \to 0$ as $T \to \infty$.

References

This chapter draws on the material in Davidson and MacKinnon (1993), Maddala (1992), Hamilton (1994), Banerjee et al. (1993) and Gujarati (1995). Advanced readings include Fuller (1976) and Hamilton (1994). Easier readings include Mills (1990) and Enders (1995).

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.F. Hendry (1993), *Co-Integration, Error-Correction, and The Econometric Analysis of Non-stationary Data* (Oxford University Press: Oxford).

Bierens, H.J. (2001), " Unit Roots," Chapter 29 in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Bierens, H.J. and S. Guo (1993), "Testing for Stationarity and Trend Stationarity Against the Unit Root Hypothesis," *Econometric Reviews*, 12: 1–32.

Bollerslev, T. (1986), "Generalized Autoregressive Heteroskedasticity," *Journal of Econometrics*, 31: 307–327.

Box, G.E.P. and G.M. Jenkins (1970), *Time Series Analysis, Forecasting and Control* (Holden Day: San Francisco).

Box, G.E.P. and D.A. Pierce (1970), "The Distribution of Residual Autocorrelations in Auto-regressive-Integrated Moving Average Time Series Models," *Journal of American Statistical Association*, 65: 1509–1526.

Chamberlain, G. (1982), "The General Equivalence of Granger and Sims Causality," *Econometrica*, 50: 569–582.

Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics* (Oxford University Press: Oxford).

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with A Unit Root," *Journal of the American Statistical Association*, 74: 427–431.

Dolado, J.J., J. Gonzalo and F. Marmol (2001), "Cointegration," Chapter 30 in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Durlauf, S.N. and P.C.B. Phillips (1988), "Trends versus Random Walks in Time Series Analysis," *Econometrica*, 56: 1333–1354.

Enders, W. (1995), *Applied Econometric Time Series* (Wiley: New York).

Engle, R.F. (1982), "Autogressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50: 987–1007.

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

Fuller, W.A. (1976), *Introduction to Statistical Time Series* (John Wiley and Sons: New York).

Geweke, J., R. Meese and W. Dent (1983), "Comparing Alternative Tests of Causality in Temporal Systems: Analytic Results and Experimental Evidence," *Journal of Econometrics*, 21: 161–194.

Ghysels, E. and P. Perron (1993), "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root," *Journal of Econometrics*, 55: 57–98.

Godfrey, L.G. (1979), "Testing the Adequacy of a Time Series Model," *Biometrika*, 66: 67–72.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37: 424–438.

Granger, C.W.J. (2001), "Spurious Regressions in Econometrics," Chapter 26 in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Granger, C.W.J., M.L. King and H. White (1995), "Comments on Testing Economic Theories and the Use of Model Selection Criteria," *Journal of Econometrics*, 67: 173–187.

Granger, C.W.J. and P. Newbold (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2: 111–120.

Gujarati, D.N. (1995), *Basic Econometrics* (McGraw Hill: New York).

Hamilton, J.D. (1994), *Time Series Analysis* (Princeton University Press: Princeton, New Jersey).

Johansen, S. (1988), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12: 231–254.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl and T.C. Lee (1985), *The Theory and Practice of Econometrics* (John Wiley and Sons: New York).

Kwaitowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54: 159–178.

Leybourne, S.J. and B.P.M. McCabe (1994), "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12: 157–166.

Litterman, R.B. (1986), "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience," *Journal of Business and Economic Statistics*, 4: 25–38.

Ljung, G.M. and G.E.P. Box (1978), "On a Measure of Lack of Fit in Time-Series Models," *Biometrika*, 65: 297–303.

Lütkepohl, H. (2001), "Vector Autoregressions," Chapter 32 in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

MacKinnon, J.G. (1991), "Critical Values for Cointegration Tests," Ch. 13 in *Long-Run Economic Relationships: Readings in Cointegration*, eds. R.F. Engle and C.W.J. Granger (Oxford University Press: Oxford).

Maddala, G.S. (1992), *Introduction to Econometrics* (Macmillan: New York).

Mills, T.C. (1990), *Time Series Techniques for Economists* (Cambridge University Press: Cambridge).

Nelson, C.R. and C.I. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10: 139–162.

Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90: 268–281.

Perron, P. (1989), "The Great Cash, The Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57: 1361–1401.

Phillips, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33: 311–340.

Phillips, P.C.B. and P. Perron (1988), "Testing for A Unit Root in Time Series Regression," *Biometrika*, 75: 335–346.

Plosser, C.I. and G.W. Shwert (1978), "Money, Income and Sunspots: Measuring Economic Relationships and the Effects of Differencing," *Journal of Monetary Economics*, 4: 637–660.

Sims, C.A. (1972), "Money, Income and Causality," *American Economic Review*, 62: 540–552.

Sims, C.A. (1980), "Macroeconomics and Reality," *Econometrica*, 48: 1–48.

Sims, C.A., J.H. Stock and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58: 113–144.

Stock, J.H. and M.W. Watson (1988), "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, 2: 147–174.