

Nonstationarity & unit root tests

by Kevin Kotzé

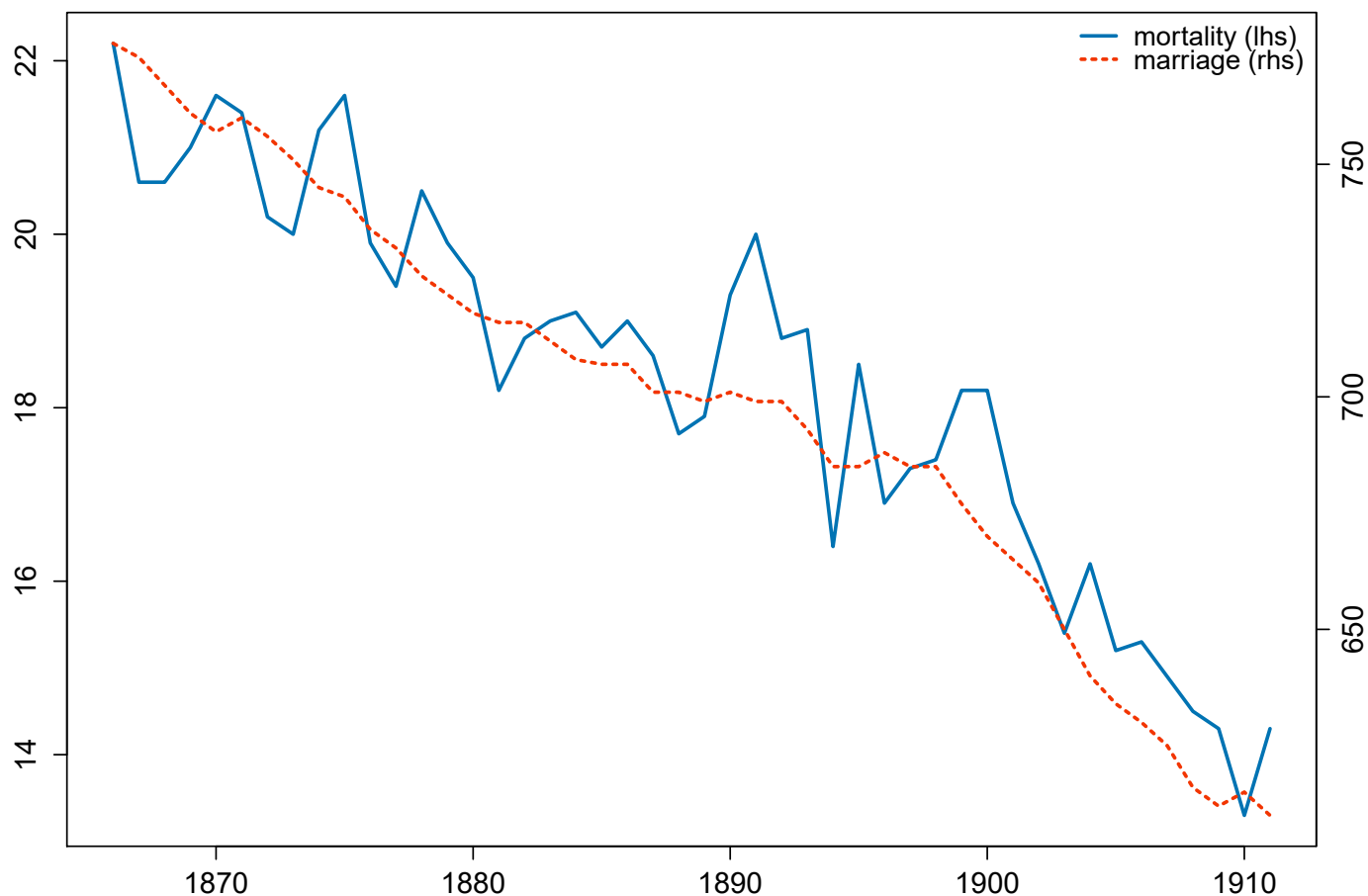
Many economic and financial time series exhibit some form of trending behaviour. Typical examples in economics include measures of output and employment, while in finance, examples consist of asset prices, dividends and financial market indices. The presence of trends, which be deterministic or stochastic, would induce nonstationary behaviour in a variable, which has important consequences for the construction and estimation of time series models. For example, a simple plot of the FTSE/JSE stock market index would suggest that the time series is not stationary as the mean would appear to depend on time. If either the data or the models are not conditioned to account for this phenomena, standard classical regression techniques (such as the use of ordinary least squares) would be inappropriate.

Many early approaches that sought to account for the effects of trends in macroeconomic data included a deterministic time trend in the specification of various models. However, Nelson and Plosser (1982) note that this strategy could represent a grave misspecification of the dynamics in a time series model and suggest that accounting for the stochastic trend in many macroeconomic variables may be more appropriate. As we will see, it is important to distinguish between those variables that may contain a deterministic or stochastic trend, as the transformations that are required to induce stationarity contain important differences.

The identification of a unit root process has also attracted much interest in the empirical literature, as a random walk process may be regarded as a prototype for various economic and financial hypotheses (i.e. it can be used to test the efficient market hypothesis or exchange rate overshooting). In addition, the use of Bayesian estimation techniques has also allowed for many interesting estimation strategies that could be used to describe the long-run behaviour of variables.

1 An intuitive example

Early studies that consider the effects of incorporating variables that contain a stochastic trend in a regression model include the work of Yule (1926), which studies the possible relationship between mortality and marriage. The data for this study is sampled at an annual frequency for England & Wales over the period 1866 to 1911 and is displayed in Figure 1. The trends in these variables would suggest that both mortality rates and the number of marriages have decreased over time. After regressing the data for marriages on mortality, we are able to produce the results in Table 1, where we note that the regressors have extremely large t -values. In addition, the joint explanatory power of the regression is relatively high, as the coefficient of determination is 0.9. This would suggest that there is a strong relationship between these two variables.¹

Figure 1: **Mortality and Marriage****Dependent Variable: Mortality**

	Coefficient	Std. Error	t-value	prob
constant	-13.88	1.57	-8.82	0
marriage	0.04	0	20.51	0
R^2	0.91			

Table 1: **Regression results for mortality and marriage - Levels**

However, after taking the first difference of the variables, which may be used to describe the change in the rate of mortality and the total number of marriages, the results from the regression would suggest that there is no relationship between these variables. Table 1, 2 includes the results of this regression, where we note that the measures for the significance of the regressor and the coefficient of determination are extremely small.²

Dependent Variable: Δ Mortality

	Coefficient	Std. Error	t-value	prob
constant	-0.133	0.21	-0.63	0.531
Δ marriage	0.011	0.043	0.27	0.788

Dependent Variable: Δ Mortality

	Coefficient	Std. Error	t-value	prob
R^2	0.001			

Table 2: **Regression results for mortality and marriage - First difference**

In what follows we describe the differences that may exist in variables that have either a deterministic or a stochastic trend. Thereafter, we consider some of the consequences that are incurred when these features of the data are ignored or misinterpreted. We also consider the use of autocorrelation functions that seek to describe the degree of persistence in time series data, before we consider more formal tests for the presence of a unit root. The tests that are described in this chapter are by no means exhaustive, and we refer the reader to Perron (2006) and Haldrup and Jansen (2006) for surveys.

2 Deterministic trend

As was noted in the introduction, many time series variables contain a trend, which may be either deterministic or stochastic. Hence, if we are to ignore the effect of a seasonal component, the variable y_t is comprised of the following dynamic components,

$$y_t = \text{trend} + \text{stationary component} + \text{irregular}$$

When the trend is deterministic, we would know the value of this component at each and every point in time with absolute certainty. To remove the deterministic trend from such a variable we would need to regress y_t on time, where $t = \{1, 2, 3, \dots, T\}$. This type of regression model could be structured as,

$$y_t = \alpha t + \varepsilon_t$$

Note that the residuals in this case, ε_t , would be free of the deterministic component and could be used for further analysis. To show that a variable with a deterministic trend is non-stationary, we note that a stationary univariate time series process could be written as the moving average,

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \dots$$

where $\varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, \sigma^2)$ and $t = \{1, 2, \dots, T\}$. As noted previously, such a variable has a constant mean and variance, which do not depend on time. After introducing a deterministic time trend, the variable y_t may be expressed as

$$y_t = \alpha t + \theta(L)\varepsilon_t \quad (2.1)$$

where the lag polynomial takes the form, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \dots$, and the deterministic trend is simply the time index, t , with a slope parameter, α . Since the expected value of all the white noise errors is zero, the expected mean of the variable would be,

$$\mathbb{E}[y_t] = \alpha t$$

which would clearly depend on time. However, if we remove the expected time-varying mean from y_t , then we are able to show that deviations from the expected mean are stationary

$$\begin{aligned} y_t - \mathbb{E}[y_t] &= \alpha t + \theta(L)\varepsilon_t - (\alpha t) \\ &= \theta(L)\varepsilon_t \end{aligned}$$

Therefore, this time series variable includes a stationary component and a deterministic trend. We could show that this variable would return to a point on the deterministic trend after a stochastic shock (i.e. an irregular innovation). For this reason we call these variables trend-stationary (TS).

In addition to linear trends, an economic process may include a nonlinear trend. For example, we may wish to include a quadratic polynomial for the trend to describe a variable that characterises increasing returns to scale. A model that takes various orders of nonlinear deterministic trends could then take the form,

$$y_t = \mu + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \dots + \alpha_n t^n + \varepsilon_t$$

To test for the inclusion of these nonlinear trends we would usually estimate a selection of different models and then compare the goodness-of-fit with the aid of various information criteria. Alternatively, we could estimate the above regression for some predetermined value of n , and if all the coefficients for $\alpha_2, \dots, \alpha_n$ are insignificantly different from zero, then it would suggest that a linear time trend would be most appropriate.

3 Stochastic trend

The counterpart to a deterministic trend is a stochastic trend, and as we will see below, the time series properties of a variable that has a deterministic trend are very different to those that have a stochastic trend. There are many examples of economic variables that have stochastic trends, where the values of variables are permanently affected by a shock (or innovation).³

3.1 Random walk

The simplest model of a variable with a stochastic trend is the random walk, which depends on past values of itself and Gaussian white noise errors,

$$y_t = y_{t-1} + \varepsilon_t \quad \text{where } \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, \sigma^2) \quad (3.1)$$

These random walk processes have a number of interesting features, where the best forecast of y_{t+1} at time t , is given by

$$\mathbb{E}[y_{t+1} | y_t] = y_t$$

The mean squared forecast error (MSFE) of this process, which describes the forecast error variance, would grow with the forecast horizon,

$$\mathbb{E}_t \left[\sigma_{t+h}^f \right] = \text{var}(y_{t+h} - \mathbb{E}[y_{t+h} | y_t]) = \sigma^2 h$$

Note that in this instance, the forecasting horizon may be used to denote the progression of time, where the difference between a two and a one step-ahead forecast is one period of time. With the aid of this expression, we would suggest that the variance depends on time, which would imply that it is nonstationary. This may be confirmed with the aid of a recursive substitution exercise.

For the random walk model, $y_t = y_{t-1} + \varepsilon_t$, we may use substitute recursive lag values of y_t to describe the evolution of the process,

$$\begin{aligned}
y_t &= y_{t-1} + \varepsilon_t \\
&= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\
&= y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
y_t &= \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0
\end{aligned}$$

Therefore each shock, ε_{t-j} , will influence subsequent values of y_t . This would imply that a shock to a random walk has a permanent effect on the time series variable. Alternatively, we may infer that these variables have infinite memory. If we assume that y_0 is equal to zero, without any loss of generality, we can write the random walk as,

$$y_t = \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

This would allow us to write the mean and variance of the random walk as

$$\mathbb{E}[y_t] = 0 \quad \text{and} \quad \text{var}(y_t) = \sigma^2 t$$

The covariance, γ_{t-j} , between y_t and y_{t-j} with $y_0 = 0$, would then yield the following result,

$$\begin{aligned}
\mathbb{E}[(y_t - y_0)(y_{t-j} - y_0)] &= \mathbb{E}[(\varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1) \dots \\
&\quad (\varepsilon_{t-j} + \varepsilon_{t-j-1} + \dots + \varepsilon_1)] \\
&= \mathbb{E}[(\varepsilon_{t-j})^2 + (\varepsilon_{t-j-1})^2 + \dots + (\varepsilon_1)^2] \\
&= (t - j)\sigma^2
\end{aligned}$$

which also depends on time. Hence, since the variance and covariance of the process depend on time, the random walk is certainly nonstationary. With such a process, the effect of a change in the error term in $t - j$ will continue to effect y_t , and the roots of the linear difference equation would contain a unitary element, which infers that such a process is a unit root.

These processes are also termed difference-stationary (DS), as the first difference of the random walk would may be expressed as,

$$\begin{aligned}
y_t &= y_{t-1} + \varepsilon_t \\
\Delta y_t &= \varepsilon_t
\end{aligned}$$

where Δy_t clearly is stationary as the expected mean and variance of the white noise error are stationary. If a variable, y_t , could be made stationary after differencing it once, it is integrated of the first order. We use the notation $I(1)$, to describe such a process. Stationary random variables, such as Δy_t are thus integrated of order zero (i.e. Δy_t is $I(0)$). If it is necessary to take the second difference to achieve stationarity, where $\Delta^2 y_t$ is $I(0)$. Such a process is integrated of the second order, where we would use the notation, $I(2)$.

3.2 Random walk with drift

Adding a constant term to the random walk model in equation (3.1) results in a random walk with drift, which may be expressed as,

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

Using recursive substitution we can show that the random walk with drift can be written as a function of a deterministic trend and stochastic term,

$$\begin{aligned} y_t &= \mu + y_{t-1} + \varepsilon_t \\ &= \mu + (y_{t-2} + \mu + \varepsilon_{t-1}) + \varepsilon_t \\ &= 2\mu + (y_{t-3} + \mu + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ y_t &= \mu \cdot t + \sum_{j=0}^{t-1} \varepsilon_{t-j} \end{aligned} \quad (3.2)$$

where we again assume that the starting value, y_0 , is equal to zero. In contrast with the random walk model in equation (3.1), a random walk with drift now also contains a deterministic trend, which results from the inclusion of the constant term, μ , that influences the slope of the deterministic trend. However, in contrast with the trend stationary model, the deviations from the deterministic trend are not stationary. This would imply that each ε_{t-j} will influence the value of y_t , even after removing the deterministic trend from the variable.

4 Implications of nonstationarity

When a time series process is stationary we noted that a shock to such a process would only have a temporary effect on future values of the process, where the expected mean, variance and covariance do not depend on time. In contrast with these properties, a shock to a nonstationary time series process would have a permanent effect on the future values of the process. In addition, it was also noted that the expected variance, covariance and/or mean, would depend on time. This finding may be substantiated with the results of the impulse response functions in Figure 2, where we show the effect of a shock on a random walk and two autoregressive processes.

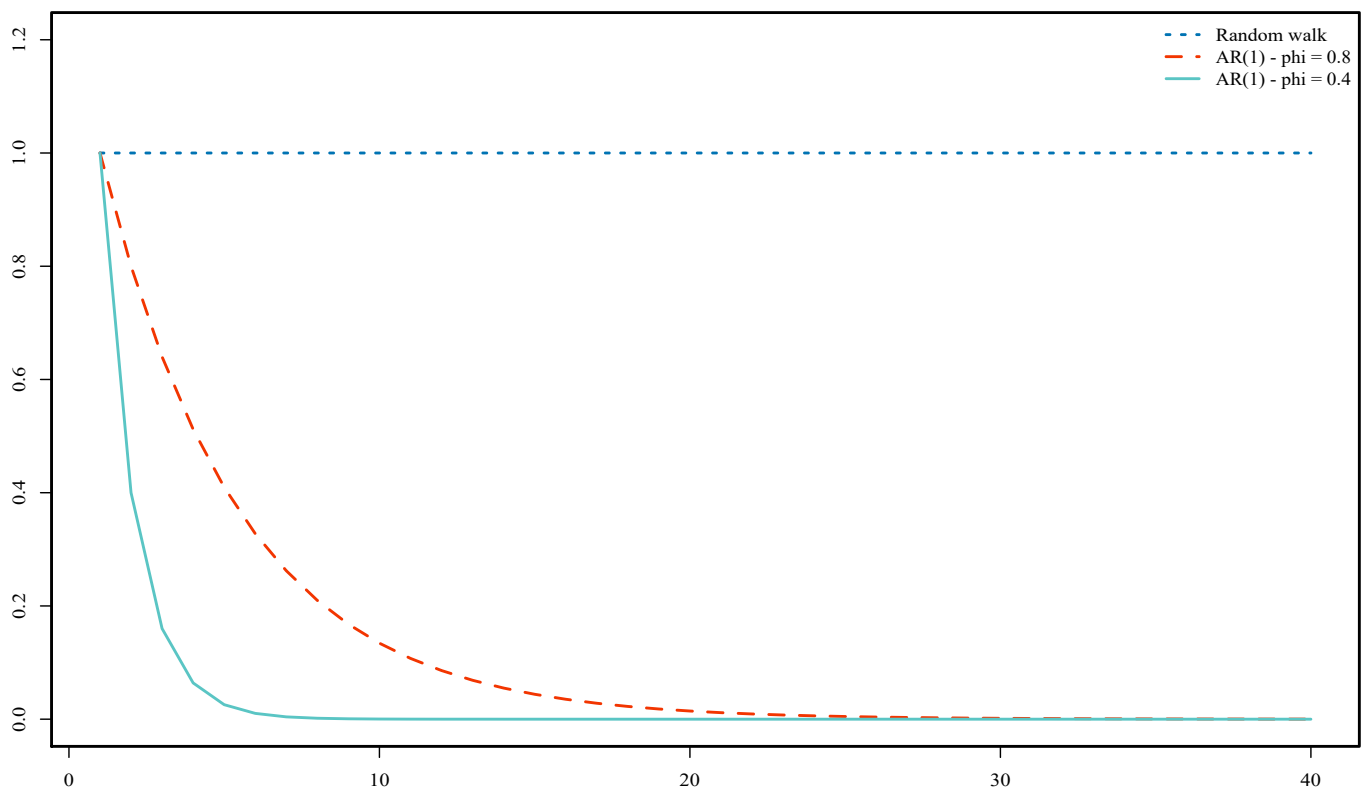
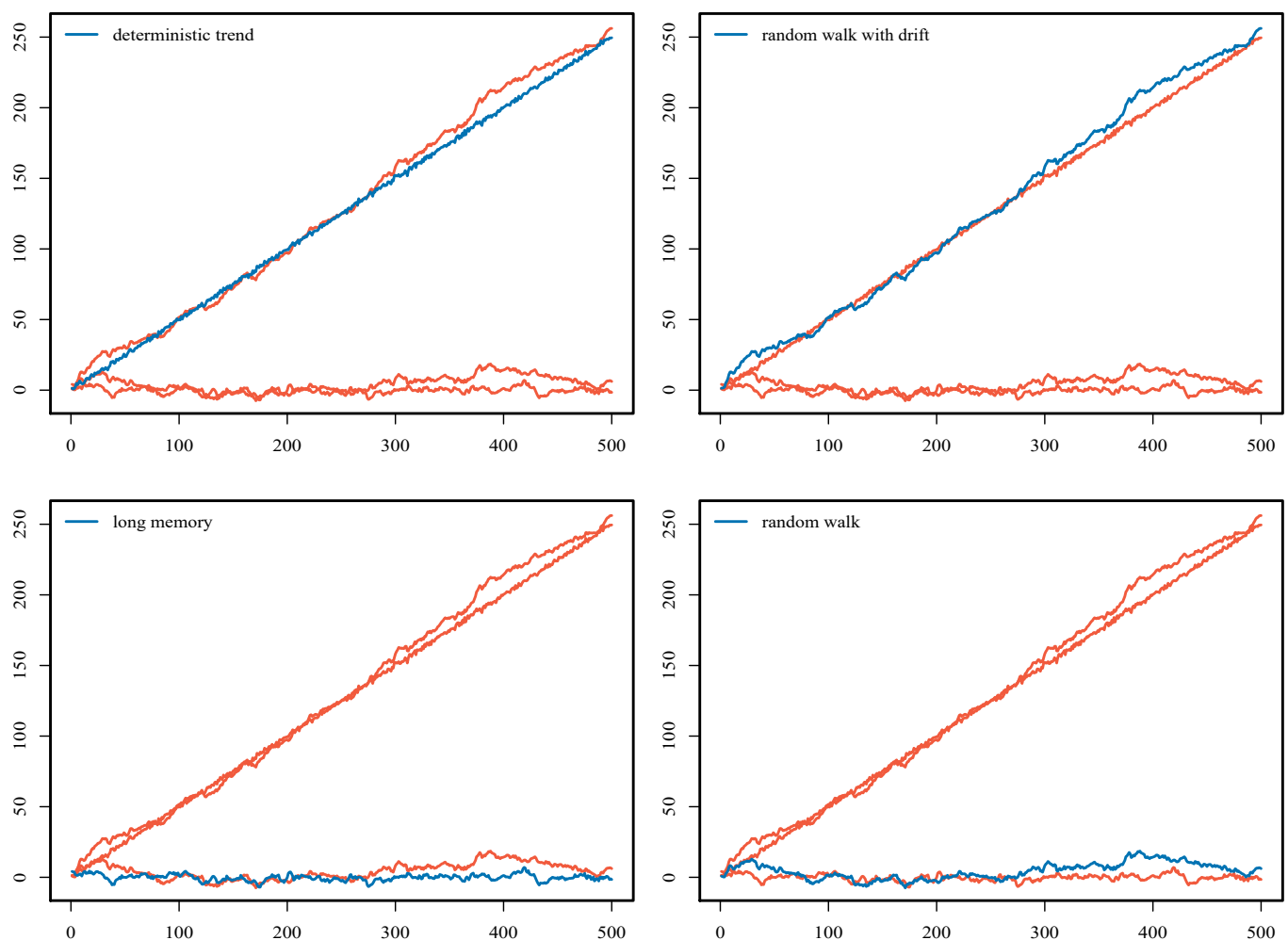


Figure 2: **Impulse response functions for autoregressive processes**

If a time series is trend-stationary, the expected mean value will depend on time, which would imply that it is nonstationary. As noted previously, we can simply remove the effects of this trend by regressing it on time, and as a result the residuals will be stationary. However, after removing a deterministic trend from a random walk with drift, we are left with a random walk process, which will continue to display nonstationary behaviour. Similarly, if we have a first-order autoregressive process that has a coefficient value that is slightly less than unity (i.e. a process that has long memory), then it is stationary. However, a first-order autoregressive process that has a coefficient of unity is a random-walk and is not stationary.

Examples of all of these processes are contained in Figure 3, where the top panels suggest that a visual inspection would not enable us to distinguish between a stationary variable that has deterministic trend and a variable that takes the form of a random-walk with drift. Similarly, as shown in the bottom panels we are also largely unable to distinguish between a stationary long memory process and a random walk, from a cursory visual inspection.

Figure 3: **Simulated time series processes**

To stress the importance of being able to correctly identify a particular time series variable, we now consider the potential effect of performing an incorrect transformation. As noted previously, time series variables that contain a stochastic trend can be transformed into a stationary variable, by taking the first difference of the data, while a variable that contains a deterministic trend should be

regressed on time to provide stationary residuals. In what follows, we consider the effect of taking the first difference of a process that has a deterministic trend. For example, consider the following trend stationary process,

$$y_t = \alpha t + \varepsilon_t$$

where the lag could be represented by, $y_{t-1} = \alpha(t-1) + \varepsilon_{t-1}$. The first difference of the above trend stationary process could then take the form,

$$\Delta y_t = \alpha + \varepsilon_t - \varepsilon_{t-1}$$

where the full effect of the previous shock is now incorporated in the solution. Hence, the process is nonstationary, as we have introduced a unit root in the moving average component, as the effects of previous shock in period $t-1$ does not dissipate with time.

It is worth noting that this result is very different to the one that would arise when the underlying time series has both a unit root and a deterministic trend. Consider, by way of example, the following process that has both deterministic and stochastic trends:

$$y_t = \alpha t + y_{t-1} + \varepsilon_t$$

To make this process stationary, we would need to subtract y_{t-1} from both sides, which would ensure that we are left with the following time series process,

$$\Delta y_t = \alpha + \varepsilon_t$$

To then remove the deterministic trend we could regress Δy_t on a variable that has a deterministic trend (i.e. $t = 1, 2, 3, \dots$), which would provide us with a stationary residual. In this case we would be left with a white noise process as there was no other stationary components in the original y_t process. We are then able to conclude that when a process has both a deterministic trend and a stochastic trend, then it would be appropriate to take the first difference if we are looking to transform the variable into a stationary process. However, if such a process only has a deterministic trend (and not a unit root) then we would induce an alternative form of nonstationarity, through the lag of the moving average term when taking the first difference of such a process.

4.1 The autocorrelation function

As has been noted previously, the autocorrelation function may be used to describe the persistence in a process. When we are modelling a stationary AR(1) process, the first correlation coefficient, ρ_1 , is equivalent to the coefficient in the AR(1) model, ϕ . Similarly, the second correlation coefficient, ρ_2 , is equivalent to ϕ^2 .

The subsequent values of the correlation coefficient, ρ_j , may be derived from the more general expression that considers the value of the covariance function, which is divided by the product of the standard deviation of y_t and the standard deviation of y_{t-j} . Hence, for a random walk the standard deviation of y_t may be derived from $\sqrt{\text{var}(y_t)} = \sqrt{T\sigma^2}$, where T is the sample size. In addition, the standard deviation of y_{t-j} may be similarly derived from, $\sqrt{\text{var}(y_{t-j})} = \sqrt{(T-j)\sigma^2}$. Given these conditions, the general form of the autocorrelation coefficient, for lag j , may then be derived from the following expression.

$$\begin{aligned}
 \rho_j &= (T-j)\sigma^2 / \sqrt{(T-j)\sigma^2} \sqrt{(T)\sigma^2} \\
 &= (T-j) / \sqrt{(T-j)T} \\
 &= \sqrt{(T-j)/T} < 1
 \end{aligned}$$

In most cases, where the sample size is large, when compared with the value for j , the ratio $(T-j)/T$ is approximately equal to unity. However, it will in all instances be less than 1. This is rather unfortunate as it would infer that we are not able to use the autocorrelation function to distinguish between a process that has a unit root and an AR(1) process that is stationary, but has a high degree of persistence. As such, a slowly decaying autocorrelation function indicates that the process has a large characteristic root, where the process may possibly include a unit root, a deterministic trend, or both of these features. In addition, such a slowly decaying autocorrelation function could also suggest that the process is stationary, but somewhat persistent. Furthermore, as we previously noted that the value of ρ_1 is equivalent to the $\hat{\phi}$ coefficient estimate in the AR(1) model, this would imply that the parameter estimate would be biased, as it will be less than the true value of unity.

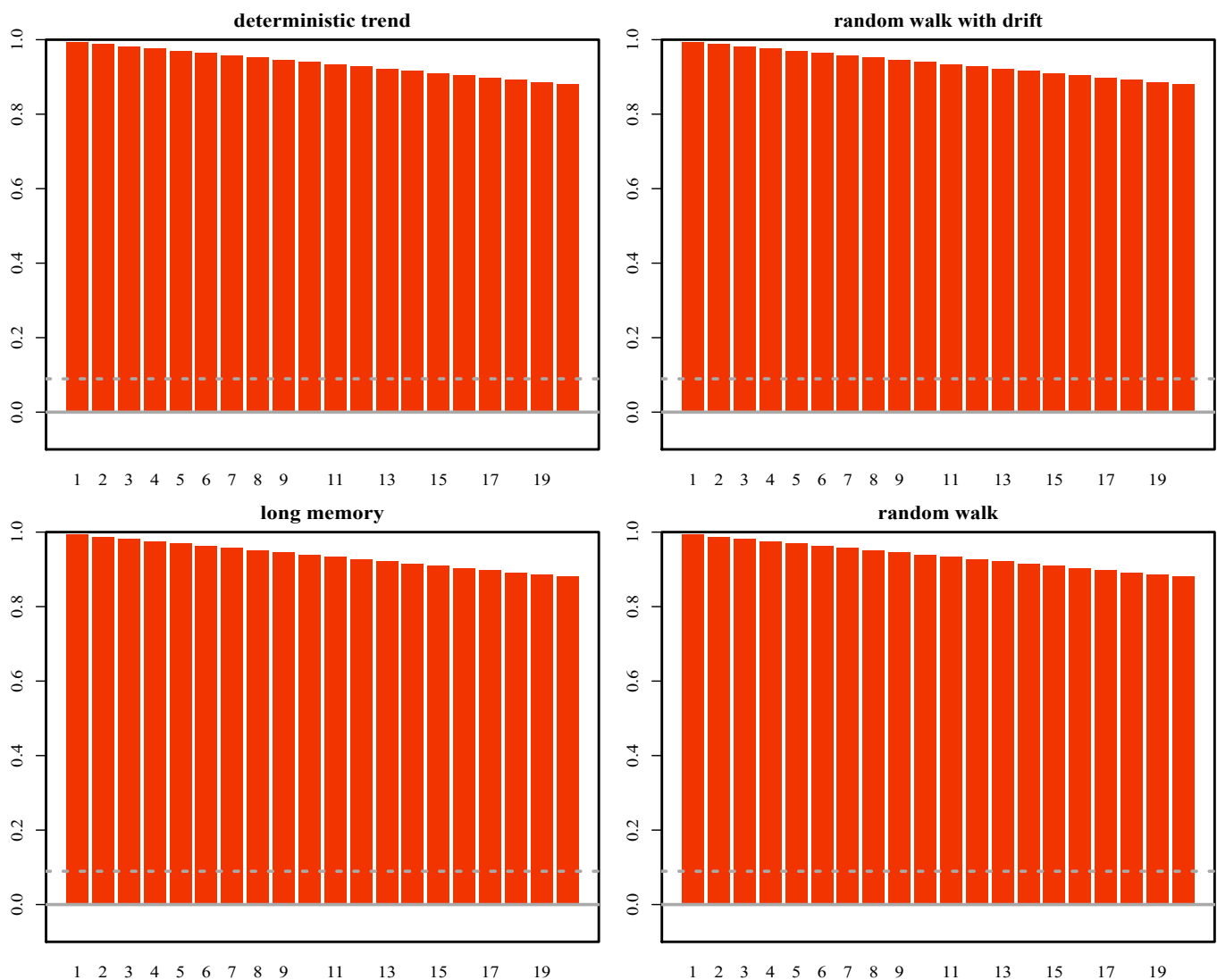


Figure 4: **Autocorrelation functions for simulated processes**

Examples of these processes are included in Figure 4, where we note that it would be difficult to use the autocorrelation function to distinguish between the various processes. This exercise should motivate for the use of formal tests that would allow for the identification of a variable that contains a

deterministic or a stochastic trend. In addition, we would also need to be able to identify those variables that contain both of these features, as well as those cases where the variable may indeed be stationary.

5 Tests for unit root

Several tests have been developed to test the order of integration of a time series variable. In what follows, these tests have been separated into three groups. The first group of tests investigate the null hypothesis of a unit root, against the alternative of stationarity, where the alternative could be stationarity in levels or around a deterministic trend (i.e. a trend-stationary process). The second group of tests, also consider the null hypothesis that there is a unit root, but allow for structural breaks that may prevail at a given point in time, or where the existence of such a break is unknown. The final group of test statistics investigate the null hypothesis that the process is stationary, against the alternative that the process has a unit root.

5.1 Dickey-Fuller test

The most widely used test for a presence of a unit root was originally proposed by Dickey and Fuller (1979), which tests the null hypothesis of whether a series is a random walk against the alternative that it is stationary. To perform this test, we assume that we have an AR(1) process,

$$y_t = \phi y_{t-1} + \varepsilon_t \quad \text{where } \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, \sigma^2)$$

With the use of this equation, we would want to determine whether $|\phi| = 1$, against the alternative that $|\phi| < 1$. If $|\phi| = 1$ then the above model would represent a random walk process, while if $|\phi| < 1$, the above process is stationary. As noted above, the value of the autocorrelation coefficient, ρ_1 and the estimated value of $\hat{\phi}$ would be biased towards a value that is less than one, when the underlying data generating process contains a unit root.⁴

To provide a formal test for the presence of a unit root, Dickey and Fuller (1979) make use of the following extension, where $\pi = \hat{\phi} - 1$. Therefore, the test for whether or not $\phi = 1$, would equivalent to the test for whether or not $\pi = 0$. In their model, Dickey and Fuller (1979) make use of the following transformation of the above AR(1) model,

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t \\ y_t - y_{t-1} &= \phi y_{t-1} - y_{t-1} + \varepsilon_t \\ \Delta y_t &= (\phi - 1)y_{t-1} + \varepsilon_t \\ \Delta y_t &= \pi y_{t-1} + \varepsilon_t \end{aligned} \tag{5.1}$$

Such a statistic would take the form of a traditional t -test, where we are primarily interested in determining the degree of certainty with which this coefficient has been estimated (i.e. is it significantly different from zero). Thus, using equation (5.1), the test for a unit root would make use of the following null hypothesis:

$$H_0 : \pi = 0$$

If we are unable to reject the null hypothesis, it would imply that y_t is integrated of order one, such that $y_t \sim I(1)$. The alternative hypothesis would then take the form,

$$H_1 : \pi < 0$$

which implies that y_t is stationary, such that $y_t \sim I(0)$. This procedure would imply that we would need to calculate an appropriate value for the t -statistic that is associated with the π parameter, which considers whether or not this parameter is significantly different from zero. Therefore, the test for the null hypothesis, H_0 may be expressed as,

$$\hat{t}_{DF} = \frac{\hat{\pi}}{SE(\hat{\pi})} = \frac{\phi - 1}{SE(\phi)}$$

where SE denotes the standard error that is associated with the coefficient estimate. The Dickey-Fuller test is a one-sided test, since the relative alternative to the null hypothesis is that y_t is stationary (i.e. $\phi \neq 1$).⁵ Note however, that the asymptotic distribution for this t -statistic is non-Gaussian, owing to the possible inclusion of bias in the parameter estimate, as was noted when we discussed the properties of the autocorrelation coefficient. This would imply that we cannot use the critical values from the standard t -distribution. The relevant critical values are included in the work of Dickey and Fuller (1979), Dickey and Fuller (1981) and MacKinnon (1991).⁶

5.2 Augmented Dickey-Fuller test

The above test describes the procedure for investigating whether the null hypothesis assumes a unit root, while the alternative hypothesis is that of stationarity. The use of such a null and alternative hypothesis would be appropriate for time series variables that do not drift systematically in any direction. However, if the time series contains evidence of a trend and is either increasing or decreasing over the sample, we would like to include a deterministic trend in the alternative hypothesis.⁷ Therefore, where there is evidence of a potential deterministic trend in the variable, such a testing procedure would consider the use of the regression model,

$$y_t = \beta_1 + \beta_2 t + \phi y_{t-1} + \varepsilon_t$$

which can be rewritten as,

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \varepsilon_t \quad (5.2)$$

where $\pi = \hat{\phi} - 1$, once again. This test for a unit root would still consider whether or not $\pi = 0$, with the aid of the null hypothesis,

$$H_0 : \pi = 0$$

which implies that $y_t \sim I(1)$, but in this case it would suggest that the variable is represented by a random walk with drift. The alternative hypothesis would then be given as,

$$H_1 : \pi < 0$$

which implies that $y_t \sim I(0)$, after the deterministic time trend has been removed (i.e. the process is trend-stationary). Note that the properties of the asymptotic distribution of the t -statistic will change if either a constant or a time trend are included in the estimated regression model. As such, the critical values would differ to those that are provided in the previous case.

Since the alternative hypotheses in both of the above tests do not allow for any persistence in the underlying process, the residuals may be autocorrelated. This lead to the development of the augmented Dickey-Fuller (ADF) test, which is describe in Dickey and Fuller (1981). It controls for

residual autocorrelation by including lagged values of Δy_t , which are allowed to follow a higher order AR(p) process. To see how this works, consider an AR(2) representation,

$$y_t = \beta_1 + \beta_2 t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

which is the same as,

$$y_t = \beta_1 + \beta_2 t + (\phi_1 + \phi_2)y_{t-1} - \phi_2(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Subtracting y_{t-1} from both sides provides the expression

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \gamma_1 \Delta y_{t-1} + \varepsilon_t$$

where we have defined $\pi = \phi_1 + \phi_2 - 1$ and $\gamma_1 = -\phi_2$. Hence, if we allowed for p lags in the autoregressive process, we would have

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{j=1}^p \gamma_j \Delta y_{t-j} + \varepsilon_t$$

where $\pi = \sum_{j=1}^p \phi_j - 1$ and $\gamma_j = \sum_{k=j+1}^p \phi_k$, for $j = \{1, 2, 3, \dots, p\}$. To identify an appropriate value for p , the number of autoregressive lags, we can make use of information criteria, such as the AIC or BIC. This would allow us to isolate the persistence from other stationary components and this particular test may also be used to isolate the effects of intercepts and linear time trends, where we essentially make use of three test equations,

$$\Delta y_t = \pi y_{t-1} + \sum_{i=2}^p \gamma_i \Delta y_{t-i+1} + \varepsilon_t \quad (5.3)$$

$$\Delta y_t = \beta_1 + \pi y_{t-1} + \sum_{i=2}^p \gamma_i \Delta y_{t-i+1} + \varepsilon_t \quad (5.4)$$

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{i=2}^p \gamma_i \Delta y_{t-i+1} + \varepsilon_t \quad (5.5)$$

The difference between these test equations concerns the inclusion of β_1 and β_2 coefficients, where under the null hypothesis, equation (5.3) refers to a pure random walk model, equation (5.4) includes an intercept or drift term and equation (5.5) includes both a drift and linear time trend. In each case, the parameter of interest is π , where if $\pi = 0$ then the process y_t contains a unit root. Comparing the calculated t -statistic with the critical values from the Dickey-Fuller tables determines whether or not we should reject the null hypothesis, $H_0 : \pi = 0$.

Although the procedure is essentially the same, regardless of which test equation is used, the critical values of the t -statistics depend on whether the intercept or time trend is included, as noted previously. In addition, the critical values also depend upon the sample size.

Dickey and Fuller (1981) include three additional F -statistics, which we denote φ_1 , φ_2 and φ_3 . These statistics are used to test joint hypotheses relating the respective coefficient values and may be used to determine whether (5.3), (5.4) or (5.5) are appropriate for the underlying data generating process.

- The null hypothesis for equation (5.4), where $\pi = \beta_1 = 0$ is tested using φ_1 , to determine whether the process could possibly include a drift term.

- The most general null hypothesis for equation (5.5), where $\pi = \beta_1 = \beta_2 = 0$ is tested using φ_2 , to determine whether the process could possibly include a drift and a linear time trend.
- And lastly, the joint hypothesis $\pi = \beta_2 = 0$ may be tested with the aid of φ_3 , which seeks to determine whether the process has a deterministic time trend.

The values for the φ_1 , φ_2 and φ_3 statistics are constructed as if they were F -tests,

$$\varphi_i = \frac{[RSS(restricted) - RSS(unrestricted)]/r}{RSS(unrestricted)/(T - k)}$$

where $RSS(restricted)$ and $RSS(unrestricted)$ are the sum of the squared residuals for the two variants of the model, r refers to the number of restrictions, T is the number of usable observations, and k is determined by the number of estimated parameters in the unrestricted model.

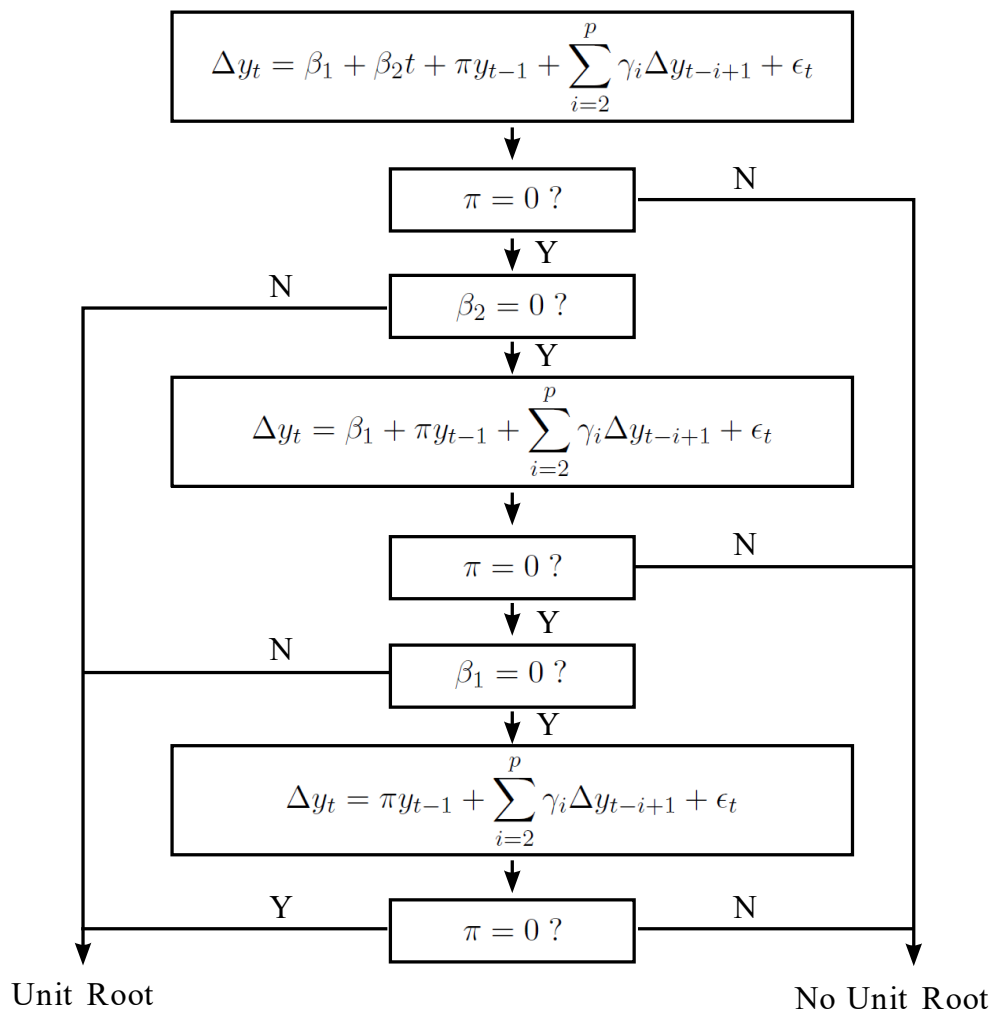
When comparing the calculated value of φ_i to the values in Augmented Dickey-Fuller tables, we need to determine the significance level at which the restriction is binding, to test the null hypothesis that the data is generated by the restricted model. In this case the alternative hypothesis is that the data is generated by the unrestricted model.

If the restriction is not binding, $RSS(restricted)$ should be close to the value for $RSS(unrestricted)$, and φ_i will be small. This would imply that large values of φ_i suggest that the restriction is binding, which would result in a rejection of the null hypothesis.

5.3 Implementing an ADF test

When implementing the augmented Dickey-Fuller test, it has been suggested that one should employ a general-to-specific approach, where the first step is to make use of the test equation that includes a constant and time trend, as provided in (5.5). If we find that we are unable to reject the null of a unit root (i.e. $\pi = 0$), we would then need to consider the value of φ_3 test statistic. If we are again unable to reject the joint null hypothesis, which in this case is given by $\pi = \beta_2 = 0$, then the process does not include a time trend and we would need to estimate the test equation that is provided in (5.4). To evaluate the results of this test equation, we would firstly consider whether or not we are able to reject the null hypothesis of a unit root. If we are unable to reject the null hypothesis, then we would consider the joint null hypothesis $\pi = \beta_1 = 0$ and compare it to the critical values for φ_1 test statistic. If we are once again unable to reject the joint null hypothesis then we would need to make use of the test equation in (5.3), where we are only interested in whether or not we are able to reject the null hypothesis of a unit root.

The full details of this testing procedure are provided in Figure 5.

Figure 5: **Augmented Dickey-Fuller: general-to-specific procedure**

Note that if we suspect that the process is integrated of the second order, we would need to perform Dickey-Fuller tests on successive differences of y_t . For example, if we want to test whether $y_t \sim I(2)$ then we would estimate the equation,

$$\Delta^2 y_t = \mu + \xi_1 \Delta y_{t-1} + \varepsilon_t$$

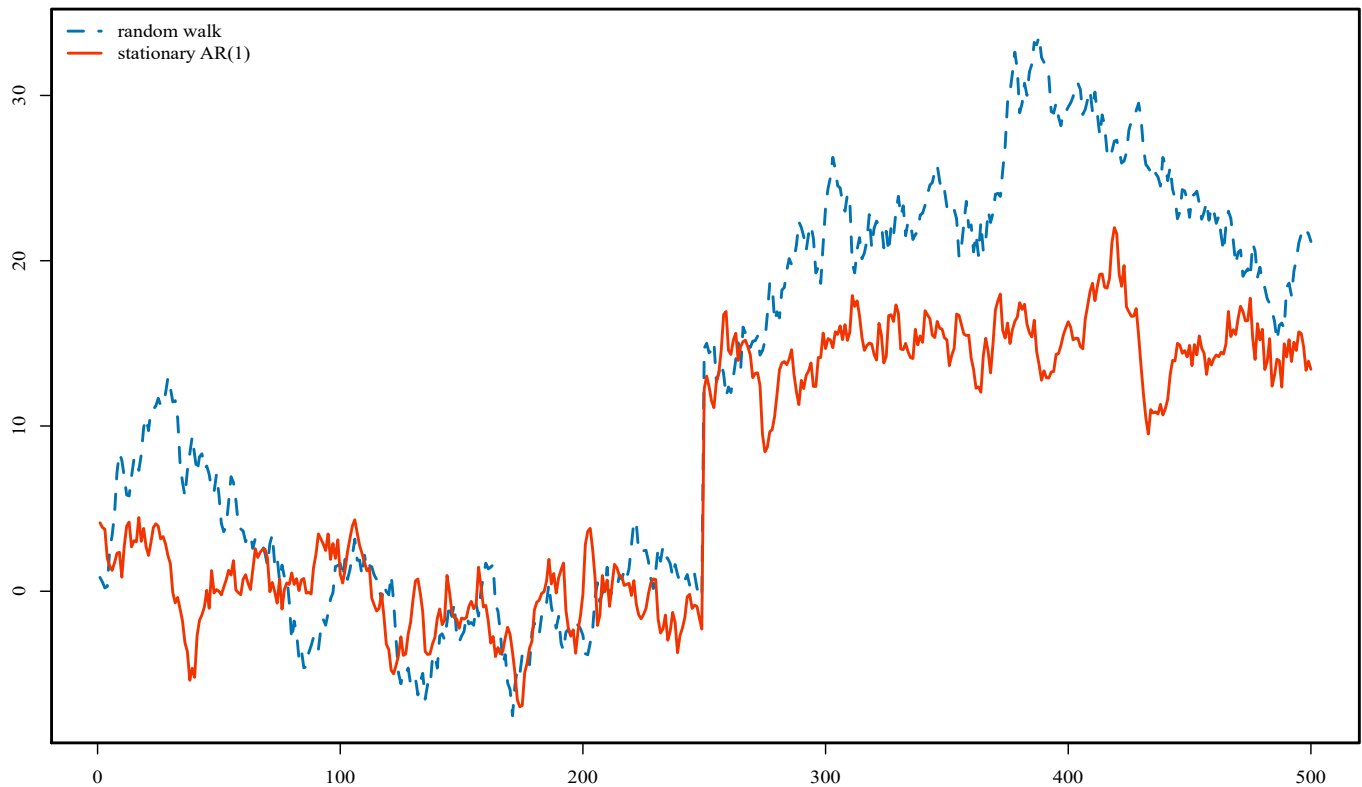
where we cannot reject the null that $\xi_1 = 0$, we would conclude that y_t is $I(2)$.

6 Testing for nonstationarity with structural breaks

6.1 Testing for a unit root with a known breakpoint

In a much cited paper, Perron (1989) showed that the ADF test has little power to discriminate between a stochastic and deterministic trend when the data is subject to structural break. This would imply that in the presence of a structural break, the various ADF tests are biased towards the non-rejection of a unit root.

For example, consider the moving average representation of an autoregressive model, $y_t = S_t + 0.5 \sum \varepsilon_t$. This time series has been simulated for 500 observations, where the level shift is described by S_t , where for the first half of the sample, $S_{1-249} = 0$ and for the second half, $S_{250-500} = 10$. This time series is depicted in Figure 6.

Figure 6: **Stationary time series plus structural break**

If we were to fit a AR(1) model to this process, the coefficient would be biased towards unity, since low values are followed by other low values during the first half of the sample, and high values are followed by other high values during the later part of the sample. In addition, since a unit root process has infinite memory, the effect of the structural break at the mid-point of the sample would be present in the remainder of the time series. Hence, if we misrepresent the structural break as a shock that does not dissipate, then the ADF tests may suggest that this process follows a random walk plus drift, where it is clearly just a stationary time series with a structural break.

Perron (1989) includes a formal procedure for testing unit roots in the presence of a structural change, where the parameter τ , is used to denote the position of the structural break, which in the above example would occur at position 250. This test could take one of the following three forms.

6.1.1 Change in level

If we assume that the null hypothesis includes a one-time jump (pulse) in the level of the unit root process, we could construct the hypothesis

$$H_0 : y_t = \mu + y_{t-1} + \beta_1 D_P + \varepsilon_t$$

where $D_P = 1$ if $t = \tau + 1$, and 0 otherwise. This specification would describe a random-walk plus drift with the addition of a structural break. Note that under the null hypothesis, the time series variable has an infinite memory. Hence, when we have a once-off change at a particular point in time (i.e. at observation 250), then the effect of this change would continue to influence what remains of the sample. For this reason, the pulse dummy would provide a shift in the level when it is incorporated within a unit root process. In addition, since we know that a random-walk plus drift would usually trend upwards or downwards, an appropriate alternative hypothesis would be to consider a (level shift) structural break in an equivalent stationary process that has a deterministic trend,⁸

$$H_1 : y_t = \mu + \alpha t + \beta_2 D_L + \varepsilon_t$$

where $D_L = 1$ if $t > \tau$, and 0 otherwise.

6.1.2 Change in slope

To consider a permanent change in the drift of a unit root process, we could construct the null hypothesis,

$$H_0 : y_t = \mu + y_{t-1} + \beta_1 D_L + \varepsilon_t$$

where $D_L = 1$ if $t > \tau$, and 0 otherwise. In this case the infinite memory of random-walk plus drift would ensure that the inclusion of the level shift dummy would provide behaviour that may be characterised by an increase (or decrease) in the drift. As such an appropriate alternative hypothesis would be to consider a trend-stationary process that has a dummy variable that would reflect a change in the slope,

$$H_1 : y_t = \mu + \alpha t + \beta_3 D_T + \varepsilon_t$$

where $D_T = t - \tau$ if $t > \tau$, and 0 otherwise.

6.1.3 Change in level and slope

To consider a change in both the level and drift, we could construct the null hypotheses that makes use of the previous two specifications,

$$H_0 : y_t = \mu + y_{t-1} + \beta_1 D_P + \beta_2 D_L + \varepsilon_t$$

For which the alternative would also make use of the previous two specifications, such that

$$H_1 : y_t = \mu + \alpha t + \beta_2 D_L + \beta_1 D_T + \varepsilon_t$$

To implement this procedure, one could estimate the model for the alternative hypothesis, which may contain the effects of the constant, time trend and structural break. The residuals from this model would then exclude the effects of these terms and could be tested using a simple ADF specification, as provided in equation (5.3). Alternatively, if we are testing the null of a one-time jump in a unit root process (against the alternative of level shift in a trend-stationary process), we could combine these steps by estimating the equation,

$$y_t = \mu + \phi_1 y_{t-1} + \alpha t + \beta_2 D_L + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$$

Appropriate critical values for this hypothesis test are contained in Perron (1989). After making use of this procedure Perron (1989) found that there was less evidence of unit roots in economic time series, than had been previously reported in the literature.

6.2 Testing for a unit root with an unknown breakpoint

While this technique is highly intuitive, Christiano (1992) and a number of other researchers criticised the Perron approach on the basis that it required prior knowledge about the exact date of the break point, which is not always available. This led to the development of a number of different

methods that treat the break point as unknown (prior to testing). Examples of these procedures are contained in the work of Perron and Vogelsang (1992), Banerjee, Lumsdaine, and Stock (1992), Perron (1997), and Vogelsang and Perron (1998).

While most of these studies provide interesting insights, the technique that is described in Zivot and Andrews (2002) is the most popular procedure for identifying a unit root with an unknown endogenous structural break. This procedure makes use of an optimisation routine that identifies the date of the most likely structural break, which is the point that gives the least favourable result for the null hypothesis of a random walk with drift.

Therefore, these test statistics are formulated as,

$$\begin{aligned}\Delta y_t &= \mu + \pi y_{t-1} + \alpha t + \beta_2 D_L \hat{\lambda} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t \\ \Delta y_t &= \mu + \pi y_{t-1} + \alpha t + \beta_3 D_T \hat{\lambda} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t \\ \Delta y_t &= \mu + \pi y_{t-1} + \alpha t + \beta_2 D_L \hat{\lambda} + \beta_3 D_T \hat{\lambda} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t\end{aligned}$$

where λ is the estimated date for the structural break and we are essentially interested in the value for $\pi = \phi - 1$. Critical values for this technique are provided in Zivot and Andrews (2002).

7 Testing the assumption of stationarity

An alternative testing procedure has been proposed by Kiawatkowski et al. (1992), who consider the null hypothesis that a time series variable is stationary, where the alternate hypothesis is obviously that the variable is nonstationary. This procedure is usually referred to as the KPSS test.

To consider the intuitive appeal of this procedure, assume that the data generating process has the form,

$$y_t = \mu + x_t + v_t \quad (7.1)$$

where μ is a constant, v_t is a stationary component, and x_t takes the form of a random walk, such that

$$x_t = x_{t-1} + \varepsilon_t \quad \text{where } \varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, \sigma^2) \quad (7.2)$$

It can then be shown that if the variance of ε_t is zero, then $x_t = x_0$ for all t . This would imply that if there is no variation in the error term, ε_t , then x_t must be constant. Note that if what was perceived to be a potential random walk is a constant, then y_t would be stationary, since it would include two constants and the stationary process, v_t . Such a test statistic could be formulated with the null hypothesis that y_t is stationary, where we specify,

$$H_0 : \sigma^2 = 0$$

which implies that x_t is a constant, against the alternative hypothesis,

$$H_0 : \sigma^2 > 0$$

which implies that x_t varies over time and y_t will be nonstationary. To derive an appropriate test statistic for this procedure, we regress y_t on a constant, μ , to obtain the residuals, which we call \hat{v}_t . Thereafter, we calculate $S_t = \sum_{s=1}^t \hat{v}_s$ and $\hat{\sigma}_\infty^2$, which describes the long-run variance of the process. The KPSS test statistic could then be derived from the following calculation,

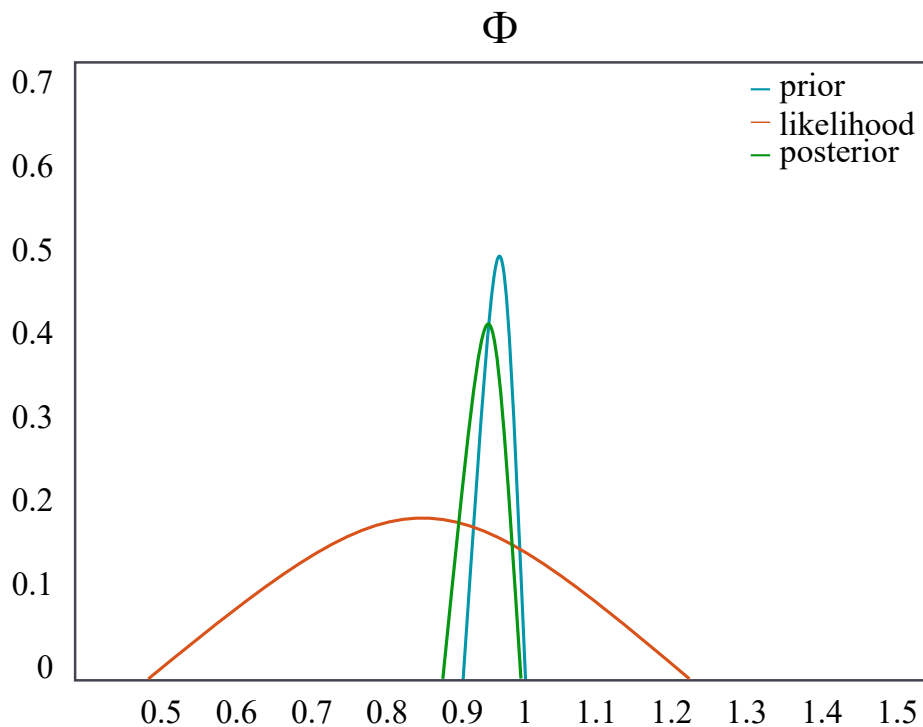
$$KPSS = \frac{1}{T^2} \frac{\sum_{t=1}^T \hat{S}_t^2}{\hat{\sigma}_\infty^2} \quad (7.3)$$

This test statistic may be augmented to allow for additional deterministic components, such as a deterministic trend. As with the Augmented Dickey-Fuller test, any change to the test equation would obviously require a different set of critical values. Note also that all these test have relatively low power, where the Augmented Dickey-Fuller test is potentially biased against the non-rejection of the null hypothesis of nonstationarity, while the KPSS test is potentially biased against the non-rejection of the null hypothesis of stationarity. Therefore, if both of these statistics provide the same result, (i.e. stationary or nonstationary) then we can be pretty sure that the order of integration that is provided by the respective test statistics is relatively robust.

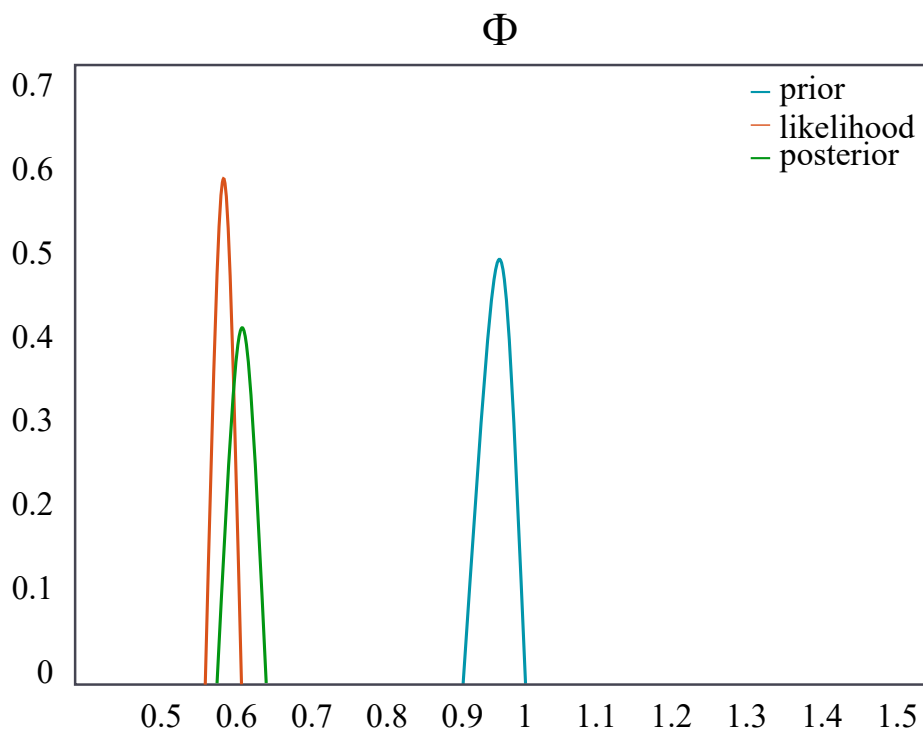
8 Bayesian analysis and unit roots

Up to this point we have adopted the classical statistical perspective, where we estimate the value of ϕ in an autoregressive model. When using these classical techniques, the Dickey-Fuller testing procedure suggested that if the uncertainty with which we estimate of the coefficient value is relatively high, and that coefficient is relatively close to one, then we would be unable to reject the null hypothesis of a unit root.

When using Bayesian estimation techniques, all the parameters are treated as random variables and we need to specify the moments for the prior distribution of all the parameters in the model. Information from the prior densities is then combined with the likelihood function (which would provide a summary of the parameter estimates, conditional on the observed values of the data) to provide the posterior parameter estimates. To combine information from the prior densities and likelihood function in an effective manner we make use of Bayes rule, which ensures that if the distribution for the likelihood function is relatively flat (i.e. when the data suggests that there is a great deal of uncertainty about the parameter estimates), then the posterior density would converge on the prior density, as is shown in Figure 7.

Figure 7: **Bayesian parameter estimation - uninformed likelihood**

Similarly, when the likelihood function is relatively narrow and there is a great deal of certainty relating to the estimated parameter estimates, then the posterior would converge on the values that are provided by the likelihood function. This case is displayed in Figure 8, where the likelihood function is notably narrow.

Figure 8: **Bayesian parameter estimation - informed likelihood**

Hence, if we suspect that the time series contains a unit root, then we would make use of a prior distribution that has a mean value of unity. If the data strongly suggests that this is not a unit root process then the posterior would converge on the value that is provided by the likelihood function to

provide a parameter estimate that is less than one. Similarly, if the data suggests that there is a great deal of uncertainty about the possible value of the parameter, then the final parameter estimate would be unity (or closely related to unity). In this way the final parameter estimate would not be biased and it could be used for subsequent inference.

For further use of Bayesian techniques in the presence of a unit root, see Sims (1988) and Sims and Uhlig (1991).

9 Conclusion

Standard regressions that are performed on nonstationary data may provide spurious results. This is important since many time series variables have deterministic or stochastic trends, which would infer that they are nonstationary. If a process returns to its (non-zero) trend value after a shock we say that it has a deterministic trend and is trend-stationary. These variables can be made stationary by removing the deterministic time trend. Time series variables that are integrated of order one, $I(1)$, can be made stationary by differencing. Such variables are often termed difference-stationary, or we say that they have a unit root. The most widely used unit root test is the Augmented Dickey-Fuller test, which is usually employed with aid of a general-to-specific strategy. The Perron test could be used to test for the presence of a unit root, when we know the date of a potential structural break, while the Zivot-Andrews test could be used in those cases where we do not have any information about the date of a potential structural break. An alternative method that tests the null hypothesis of stationarity is the KPSS test and it can be used in conjunction with Augmented Dickey-Fuller test to confirm the order of integration. Bayesian methods may also be used to make inference in those cases where we may suspect that a time series variable may contain a unit root.

10 Appendix

10.1 Monte Carlo simulations for the bias in a unit root

In the tutorial we constructed a number of simulation exercises, where we noted that after generating a random walk process 10,000 times, the estimated coefficients for $\hat{\phi}$, were biased to values below 1. The results of this simulation exercise are contained in Figure 9.

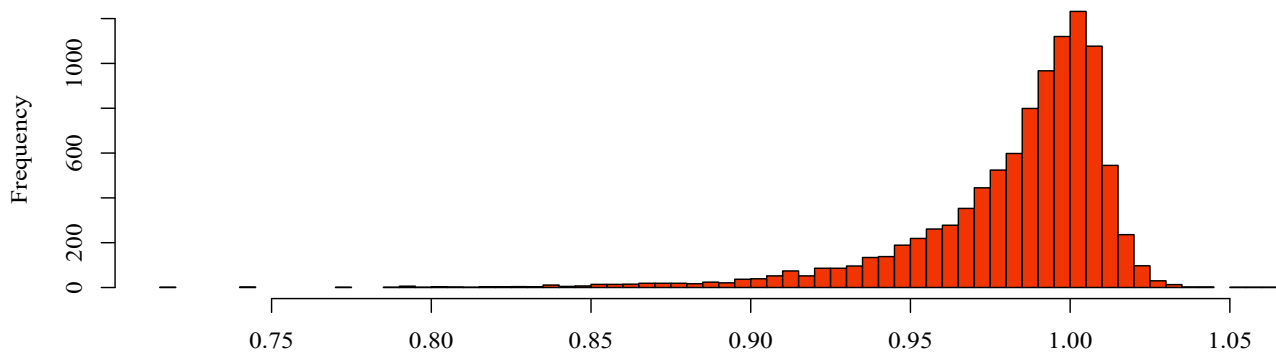


Figure 9: **Bias in unit root process when $\phi = 1$**

To make use of a Monte Carlo simulation for such a data generating process (DGP) that may have been generated for a particular model, we need to specify information relating to:

- Model structure
- Error term's distribution
- Parameter values
- Sample size

Therefore, if we assume that the DGP is generated by an AR(1) model that does not have a constant, such as:

$$y_t = \phi y_{t-1} + \epsilon_t, \quad \text{for } t = 1, \dots, T \quad \text{and } \epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

Then we would need to specify values for the follow terms, where by way of example,

$$y_0 = 0, \phi = 1, \sigma = 1 \text{ and } T = 100$$

.

We would then be able to generate values for the variables with the aid a some form of simulatio, where the number of simulations would need to take on a defined value, e.g. $N = 10,000$. Thereafter, we could estimate an AR(1) model for each of these simulated time series, which could be used to investigate the bias in the estimated value of $\hat{\phi}$. Hence,

- For $i = 1, \dots, N$:
 - Generate y_t for $t = 1, \dots, T$ from the AR(1) model for given ϕ and σ

- Estimate an AR(1) model for the simulated y_t to obtain estimated values for $\hat{\phi}$ and $\hat{\sigma}$
- Compute the t -statistics that is associated with $\hat{\phi}$
- Store the $\hat{\phi}$ estimates and the associated t -statistics in a vector that has a length of N
- To compute the average bias calculate:

$$\text{Average Bias} = \frac{1}{N} \sum_{i=1}^N (\hat{\phi}_i - \phi)$$

- One could then consider different values of T , ϕ , and σ to investigate the relation between the bias and T , ϕ , and σ
- To obtain critical values from the sampled t -statistics, sort the absolute values of the t -statistics from smallest to largest. Then as we are seeking to perform a one-sided test, where you want to apply a 95% confidence interval, sort the sample from smallest to largest and take the $(0.95 \times N)$ t -statistic from the sorted sample

10.2 Power studies

The power of a test is the probability of rejecting the null hypothesis given that the null hypothesis is not true (that is, one minus type II error).

For example consider the power of the Dickey-Fuller test, where we assume that you know the 5% critical value of the one-sided t -test for $\phi = 1$ denoted by $\tau_{0.05}$. To ascertain the power of the Dickey-Fuller test for $\phi \neq 1$, which is where the test suggests that the series contains a unit root (when you know it doesn't).

- Fix T , N , and assume the errors have, $\sigma = 1$
- Fix ϕ at a different value to the null hypothesis, for example $\phi = 0.95$
- Then setup the loop, where for $j = 1, \dots, N$:
 - Generate y_t for $t = 1, \dots, T$ from the AR(1) model for $\phi = 0.95$ and $\sigma = 1$
 - Estimate an AR(1) model for the simulated y_t to obtain $\hat{\phi}$
 - Compute the t -statistic that is associated with $\hat{\phi}$
 - Store the $\hat{\phi}$ estimates and the associated t -statistics in a vector that has a length of N

To obtain sample of estimated t -statistics:

- Count the relative number of times that the t -statistics for $\hat{\phi}$ are larger than $\tau_{0.05}$, where you have incorrectly assumed that y_t contains a unit root
- This is the power of the Dickey-Fuller test for $\phi = 0.95$

We could then consider different values of ϕ to investigate the relation between power and ϕ , which may be used to draw a power function. These studies suggest that the power of the Dickey-Fuller test is relatively low. For example, when making use of a simulation exercise for a stationary time series process that has a long memory, where $\phi = 0.95$, we noted that the Dickey-Fuller test was only able to reject the null of a unit root 4.3% of the time (when using the critical values at the 95% level).

- Banerjee, A., R. L. Lumsdaine, and J. H. Stock. 1992. "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence." *Journal of Business and Economic Statistics* 10(3): 271–87.
- Christiano, Lawrence J. 1992. "Searching for a Break in GNP." *Journal of Business and Economic Statistics* 10(3): 237–50.
- Dickey, D. A., and W. A. Fuller. 1979. "Distribution of the Estimates for Autoregressive Time Series with a Unit Root." *Journal of American Statistical Association* 74(366): 427–31.
- . 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49: 1057–72.
- Haldrup, N., and W. Jansen. 2006. "Palgrave Handbook of Econometrics: Vol 1 Economic Theory." In, edited by T. Mills and K. Patterson. Pgrave MacMillan.
- Kiawatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin. 1992. "Testing the Null Hypothesis of Staionarity Aganist the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54(1): 159–78.
- MacKinnon, J. 1991. "Long-Run Economic Relationships: Readings in Cointegration." In, edited by R. F. Engle and C. W. J. Granger. Advanced Texts in Econometrics. Oxford: Oxford University Press.
- Nelson, C. R., and C. I. Plosser. 1982. "Trends and Random Walks in Macroeconmic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10: 139–62.
- Perron, Pierre. 1989. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica*, 1361–1401.
- . 1997. "Further Evidence on Breaking Trend Functions in Macroeconomic Variables." *Journal of Econometrics* 80: 355–85.
- . 2006. "Palgrave Handbook of Econometrics, Volume 1." In, 278–352. Pgrave MacMillan.
- Perron, Pierre, and Timothy Vogelsang. 1992. "Nonstationary and Level Shifts with an Application to Purchasing Power Parity." *Journal of Business and Economic Statistics* 10: 301–20.
- Sims, Christopher A. 1988. "Bayesian Skepticism on Unit Root Econometrics." *Journal of Economic Dynamics and Control* 12 (2-3): 463–74.
- Sims, Christopher A., and Harald Uhlig. 1991. "Understanding Unit Rooters: A Helicopter Tour." *Econometrica* 59(6): 1591–9.
- Vogelsang, Timothy, and Pierre Perron. 1998. "Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time." *International Economic Review* 39: 1073–1100.
- Yule, G. U. 1926. "Why Do We Sometimes Get Nonsense-Correlations Between Time Series." *Journal of Statistical Society* 89: 1–64.
- Zivot, Eric, and Donald Andrews. 2002. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit Root Hypothesis." *Journal of Business and Economic Statistics* 20: 25–44.

1. Similar results were obtained after including a deterministic time trend in the model.↵

2. The earlier results for the regression in levels may be due to the dramatic improvements to medicine and a change in preferences (to get married) that may have occurred over this period of time.↵
3. Such an example may allow for instances where a change in technology permanently affects the level of output.↵
4. A simulation study that is used to illustrate this property of integrated data is provided in the appendix to this chapter.↵
5. This would imply that we are able to reject the null of a unit root, when the calculate value of the test statistic is smaller (or more negative) than the critical value.↵
6. The values of Dickey and Fuller (1979) have been included in the `urca` package.↵
7. For example, the level of economic output is usually increasing over time.↵
8. Figure 3 contains an example of a random walk plus drift.↵