Models for Integrated and Cointegrated Data

In: Introduction to Time Series Analysis

By: Mark Pickup

Pub. Date: 2015
Access Date: September 23, 2019
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781452282015
Online ISBN: 9781483390857
DOI: https://dx.doi.org/10.4135/9781483390857
Print pages: 165-194

Models for Integrated and Cointegrated Data

Chapter 6 extends the discussion of autoregressive moving average (ARMA) models, begun in Chapter 5, to the autoregressive integrated moving average (ARIMA) model and error correction model (ECM) by discussing the concept of differencing integrated data. The reader is also introduced to seasonal and fractional differencing. To that end, the chapter begins with an overview of identifying unit root processes and how one may distinguish these from other forms of trending. Differencing data and the differenced data model are then discussed, and this is used to motivate the ARIMA model and ECM.

6.1 Unit Root Processes and Differencing Data

A well-known example of a unit root process is the random walk (Pearson, 1905). A random walk is a process in which the current value of the series is equal to the last value plus a white noise term:

$$y_t = y_{t-1} + \varepsilon_t. \qquad (6.1.1)$$

With a random walk, $\mathrm{Var}(y_{t+h}) = \mathrm{Var}(y_t) + h\sigma^2_\varepsilon$, so it increases with time (h) and is not stationary. In fact, the variance is explosive. We say that a random walk is highly persistent since $E(y_{t+h}|y_t) = y_t$ for all $h \geq 1$, meaning that the series is not weakly dependent. The interpretation of this is that at time t_0 our best forecast of $y_{t=t_0+h}$ is $y_{t=t_0}$ (the current value), no matter how far in the future $t_0 + h$ might be. This is distinct from a stationary process, for which there is an equilibrium. In a stationary process, the equilibrium is our best forecast for the long-term future $(h \to \infty)$.

Note that trending and persistence are different things. A series can be trending but not persistent (e.g., a linear, deterministic trend), or a series can be highly persistent without any trend (e.g., a random walk). A random walk with drift is an example of a data-generating process that is both highly persistent and trending:

$$y_t = y_{t-1} + \alpha_0 + \varepsilon_t. \qquad (6.1.2)$$

In addition to the variance increasing with time, so does the mean:

$$E(y_{t+h} \mid y_t) = y_t + h\alpha_0. \qquad (6.1.3)$$

The series increases by α_0 with each time point. A series can also be a random walk with trend:

$$y_t = y_{t-1} + \alpha_1 t + \varepsilon_t. \qquad (6.1.4)$$

Note that although a random walk process without a drift or trend does not have a mean that increases (or decreases) with time, it is sometimes called a stochastic trend process.

A random walk is a special case of a unit root process. Unlike some nonstationary processes, unit root processes have the property that they can be transformed into a stationary process through differencing one (if they contain a single unit root) or more (if they contain multiple unit roots) times. These processes are called integrated of order d, and the degree of integration (d) is the number of times they must be differenced to become stationary. This is denoted as I(d).

A random walk is an I(1) process as it is made stationary—I(0)—by differencing once. We can easily see how a random walk or a random walk with drift can be made stationary through first differencing. Subtract y_{t-1} from both sides of Equation 6.1.2:

$$y_t - y_{t-1} = y_{t-1} - y_{t-1} + \alpha_0 + \varepsilon_t,$$

$$y_t - y_{t-1} = \Delta y_t = \alpha_0 + \varepsilon_t.$$

The resulting Δy_t is clearly an I(0) process. It is a constant plus white noise.

Not correcting for the fact that a data series comes from an I(1) data-generating process can lead to spurious regression problems (Granger & Newbold, 1974; Phillips, 1986). Consider running a simple regression of y_t on x_t where y_t and x_t have independent I(1) data-generating processes. The usual ordinary least squares (OLS) t statistic will often indicate a statistically significant relationship when there is none.

One solution is to simply difference the I(1) data series, producing I(0) series. Generally speaking, a series from a data-generating process that does not have a fixed equilibrium—such as the random walk—can be transformed through differencing (sometimes more than once[1]) to produce a stationary series. However, it is important not to overdifference data.

If y_t has a stationary or even a trend-stationary data-generating process, first differencing is inappropriate. To see why, suppose that y_t has an AR(1) (autoregressive process of order 1) with a trend (trend stationary) data-generating process:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 t + \varepsilon_t. \tag{6.1.5}$$

Then,

$$y_{t-1} = \alpha_0 + \alpha_1 y_{t-2} + \beta_1(t-1) + \varepsilon_{t-1}. \tag{6.1.6}$$

From Equations 6.1.5 and 6.1.6, we get

$$y_t - y_{t-1} = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 t + \varepsilon_t - (\alpha_0 + \alpha_1 y_{t-2} + \beta_1 (t-1) + \varepsilon_{t-1}),$$

$$\Delta y_t = \alpha_1 (y_{t-1} - y_{t-2}) + \beta_1 t - \beta_1 (t-1) + \varepsilon_t - \varepsilon_{t-1},$$

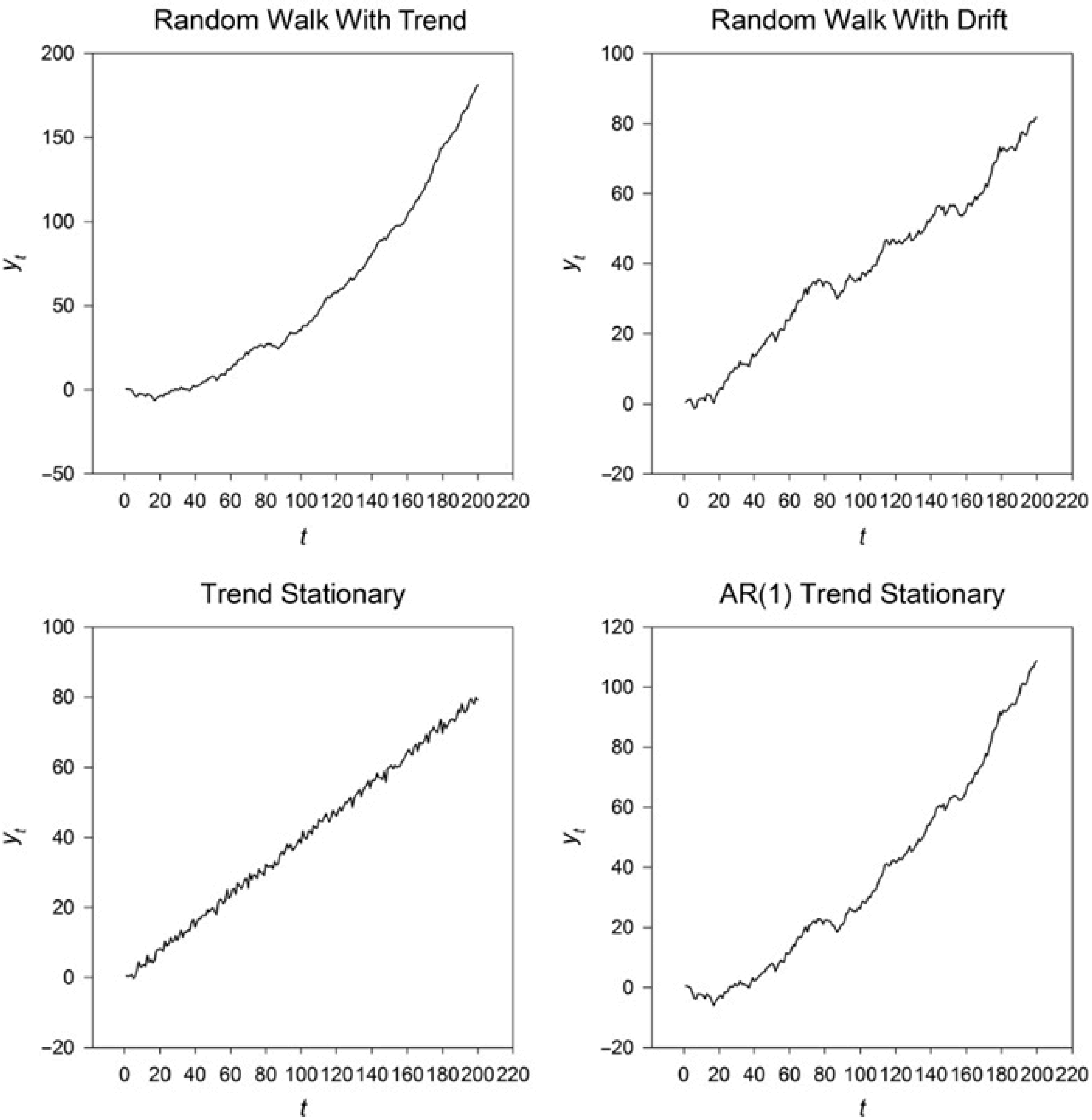
$$\Delta y_t = \alpha_1 (y_{t-1} - y_{t-2}) + \beta_1 + \varepsilon_t - \varepsilon_{t-1},$$

$$\Delta y_t = \alpha_1 \Delta y_{t-1} + \beta_1 + \varepsilon_t - \varepsilon_{t-1}. \tag{6.1.7}$$

We have eliminated the trend through differencing, but we have introduced an MA(1) random walk into the errors: $\varepsilon_t - \varepsilon_{t-1}$. This is an MA(1) process with $\varphi = 1$. Maximum likelihood estimation can run into problems when estimating data from such a data-generating process, particularly with a small sample size (Enders, 2004, p. 167). This is referred to as “overdifferencing.”

To know whether or not we need to first difference our data, we must know if we really have a unit root process, a unit root with drift process, a stationary process, or a trend-stationary process. However, this can be difficult. Figure 6.1 plots a realization from each of the following data-generating processes: random walk with drift, random walk with trend, trend stationary, and AR(1) trend stationary.

Figure 6.1 Trending Series



NOTE: AR = autoregressive.

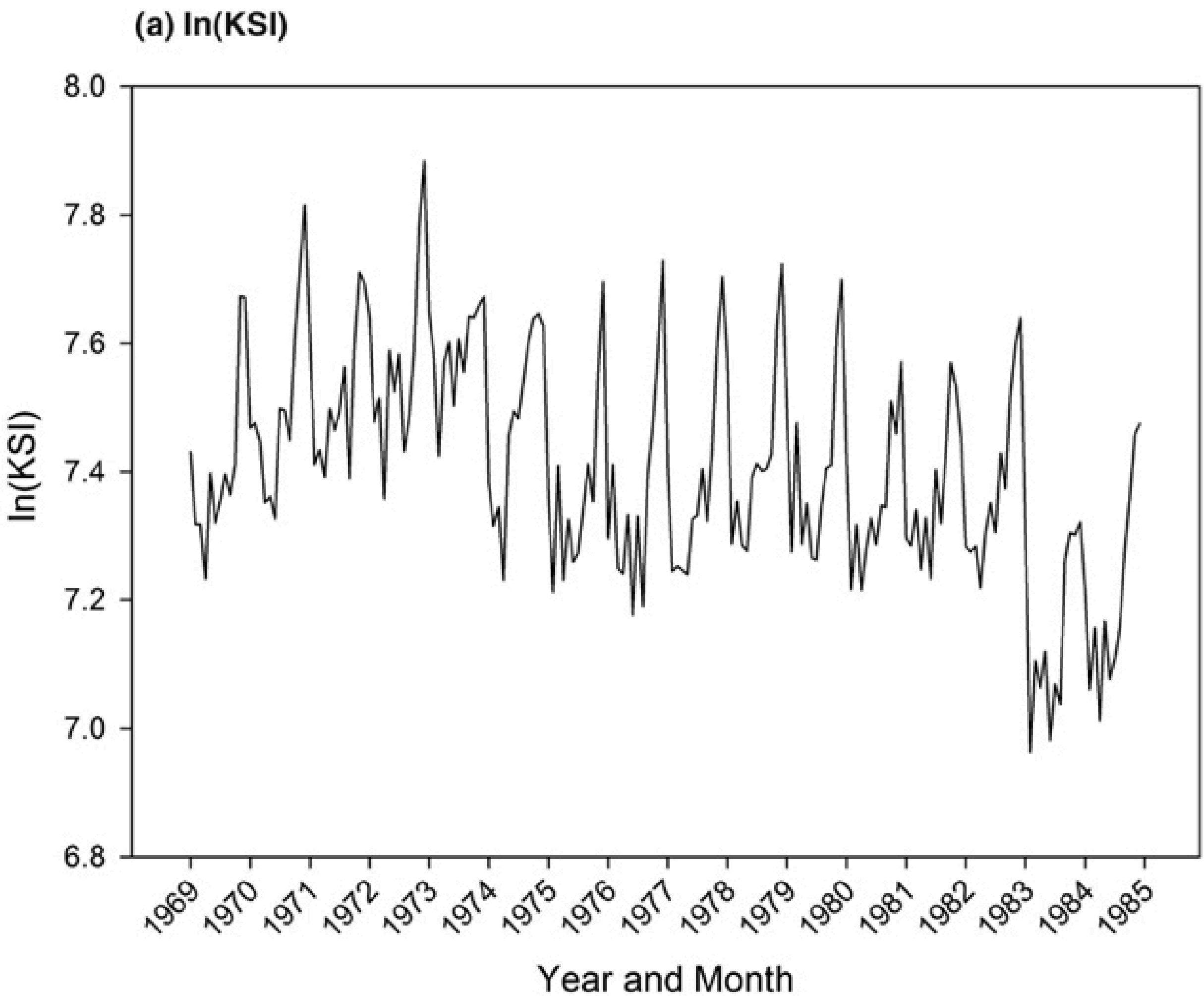
Random walk with drift	$y_t = y_{t-1} + 0.5 + \varepsilon_t$.
Random walk with trend	$y_t = y_{t-1} + 0.01t + \varepsilon_t$.

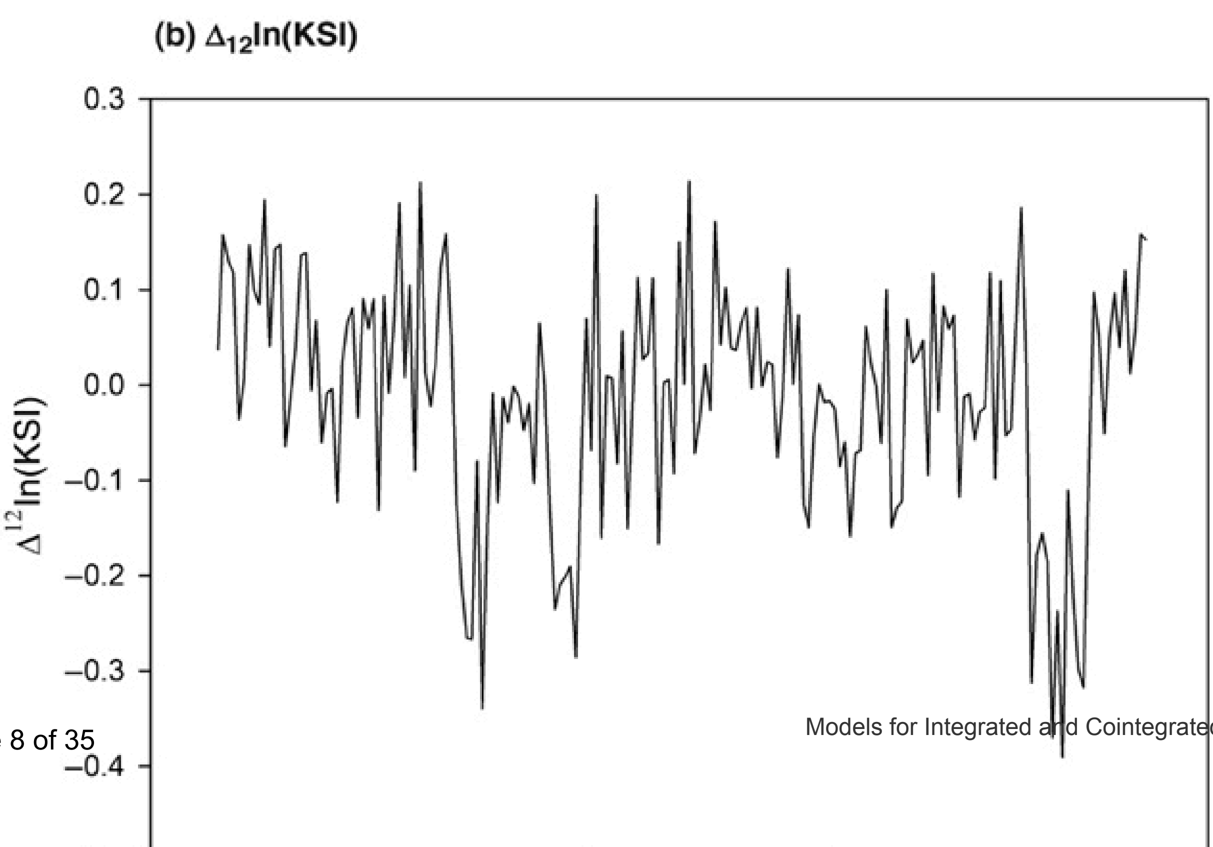
Trend stationary	$y_t = 0.4t + \varepsilon_t$.
AR(1) trend stationary	$y_t = 0.9y_{t-1} + 0.01t + \varepsilon_t$.

Note that plots of these processes are visually very similar. It is often difficult to distinguish a random walk (with drift or trend) process from a trend-stationary process, but there are test procedures to assist us. Before turning to how we might go about doing this, let us discuss seasonal differencing.

In Chapter 3, we first examined the KSI (killed or seriously injured) data and noted that they had a distinct seasonal pattern. We accounted for this by including a series of monthly indicator variables in our analysis. An alternative solution is to seasonally difference the data. In this instance, we have monthly data, and the pattern repeats itself each year. Seasonal differencing in this case would be to take the 12th difference: $\Delta_{12}y_t = y_t - y_{t-12}$. After seasonal differencing, the data represent the change since the same month the previous year. Figure 6.2 presents plots of the original ln(KSI) data and the 12th difference. The seasonal pattern that is so evident in the original data is removed from the seasonally differenced data.

Figure 6.2 ln(KSI) and the 12th Difference


(a) ln(KSI)
ln(KSI)
8.0
7.8
7.6
7.4
7.2
7.0
6.8
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
Year and Month


(b) Δ_{12}ln(KSI)
Δ^{12}ln(KSI)
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4

Data measured in other temporal units can also be seasonally differenced. For example, quarterly data can be seasonally differenced by taking the fourth difference. We will discuss the use of seasonally differenced data in ARIMA models in Section 6.3.

6.2 Testing for Unit Roots

We now consider tests that can help us distinguish between unit root (with or without drift or trend) and (trend-) stationary processes. Consider the following data-generating process:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t. \qquad (6.2.1)$$

Subtract y_{t-1} from both sides of Equation 6.2.1, and define θ = α_1 - 1 to obtain

$$y_t - y_{t-1} = \alpha_1 y_{t-1} - y_{t-1} + \varepsilon_t,$$

$$\Delta y_t = (\alpha_1 - 1) y_{t-1} + \varepsilon_t,$$

$$= \theta y_{t-1} + \varepsilon_t. \qquad (6.2.2)$$

Let H_0: α_1 = 1. That is, assume the null hypothesis that y_t is a unit root process. We can test that θ = 0 by regressing Δ y_t on y_{t-1}. Since θ = α_1 - 1, testing the null hypothesis that θ = 0 is equivalent to testing the null hypothesis that α_1 = 1. Rejecting the null θ = 0 means rejecting α_1 = 1. This means rejecting the null of a unit root data-generating process, with the alternative being an AR(1) process.

Unfortunately, a simple *t* test is inappropriate, since the data for y_{t-1} are from an I(1) data-generating process under the assumed null hypothesis. If it is a unit root process, then the assumptions necessary for OLS regression are not met, and we cannot use the *t* test.

A Dickey-Fuller test tests that a variable is from a unit root data-generating process by calculating the *t* statistic but using different critical values to test that θ = 0.[2] As above, the null hypothesis is that the data-generating process for the variable contains a unit root: θ = 0, which is to say, α_1 = 1.

Note that Equation 6.2.1 does not contain a constant, and so the alternative is a stationary process with the unconditional expected value of 0. If we wish the alternative to be a stationary process including a constant, we can, instead, use the following regression to conduct the Dickey-Fuller test:

$$\Delta y_t = \alpha_0 + \theta y_{t-1} + \varepsilon_t. \qquad (6.2.3)$$

For the null hypothesis that y_t is a unit root process, we test that θ = α_0 = 0 (Hamilton, 1994).

To get a better understanding of the properties of the Dickey-Fuller test, let us run it on data that we generate randomly and where we specify the data-generating process (e.g., I(1) or AR(1)). By doing so, we know the correct test results. To begin, we generate a single realization of the following data-generating process: $y_t = y_{t-1} + \varepsilon_t$. This is done by randomly generating 300 independent values of ε_t from a standard normal distribution: NID(0, 1). We next define $y_1 = \varepsilon_1$, $y_2 = y_1 + \varepsilon_2$, $y_3 = y_2 + \varepsilon_3$, and so on. The resulting series $\{y_1, y_2, \ldots, y_{300}\}$ is a single realization of a unit root process with $T = 300$.

We apply the Dickey-Fuller test for a unit root process to this realization. The test statistic is -1.57. The 5% critical value is -2.88. As we would hope, we are unable to reject the null hypothesis of a unit root process. If we did this many times over, we would expect to incorrectly reject the null of the unit root process 5% of the time (if using the standard 0.05 significance level).

We repeat this experiment 100 times, each time generating a new realization, applying the Dickey-Fuller test, and recording whether the null hypothesis of a unit root process is rejected at the 0.05 significance level. The null hypothesis of a unit root process is rejected 6% of the time. This is approximately correct. Next, we generate a realization of the following AR(1), $\alpha_1 = 0.95$, data-generating process:

$$y_t = 0.95y_{t-1} + \varepsilon_t.$$

This is done using the same generated values for ε_t, as used in the previous experiment. We define $y_1 = \varepsilon_1$, $y_2 = 0.95y_1 + \varepsilon_2$, $y_3 = 0.95y_2 + \varepsilon_3$, and so on. The resulting series is a single realization, $T = 300$, of an AR(1) process (not unit root), and ideally, the Dickey-Fuller test would indicate that we can reject the null hypothesis of a unit root process. We repeat this experiment 100 times, each time generating a new realization, applying the Dickey-Fuller test, and recording if the null hypothesis of a unit root process is rejected at the 0.05 significance level.

The result is that the null hypothesis of a unit root process is correctly rejected only 58% of the time. The remaining 42% of the time, the test is unable to reject the null hypothesis of a unit root process even though the process is not unit root. The Dickey-Fuller test appears to have little power to reject the null hypothesis of a unit root process under these conditions.

If we again repeat this 100 times but only use the last 100 time points, we find that the null hypothesis of a unit root process is correctly rejected only 12% of the time. The Dickey-Fuller test has even less power under these conditions. Things improve marginally if we again repeat this 100 times but this time generate data from an AR(1) data-generating process with $\alpha_1 = 0.90$ and T = 100. This time, the null hypothesis of a unit root process is correctly rejected 31% of the time. This is better, but clearly the Dickey-Fuller test has very little power to reject the null hypothesis of a unit root process when

1. the data-generating process is AR(1) with α_1 close to 1 and
2. the number of time points in the data are relatively small.

This must be kept in mind when interpreting the results of a Dickey-Fuller test.

One possible solution is to find or create a data set with a large T that does not necessarily include all the variables in your model but does contain the variables that may or may not be I(1). This can then be used for the purpose of testing if the variables of interest are I(1). With an inferential leap, this can be used to inform the model estimated using the shorter data set. Alternatively, we may have strong theoretical reasons to believe that a data-generating process is AR(1), as we often do with processes like public opinion—in which case, it may be wise to ignore the results of the Dickey-Fuller test unless we have a relatively large sample size.

It should also be noted that some time series are naturally bounded. For example, the proportion of U.S. survey respondents who identify with the Democratic Party is bounded by 0 and 1. Neither the AR(1) process, with or without a trend, nor the I(1), with or without drift or trend, is completely appropriate for bounded series, but the interpretations of the dynamics of each are quite different, and one might be more theoretically appropriate than the other. For example, the I(1) process assumes no equilibrium. This is not likely the case with partisanship over the short term (MacKuen, Erikson, & Stimson, 1989), assuming we control for structural breaks. One empirical approach to this issue is to transform the values of the time series using a logit transformation. If P_t is a time series of proportions, the logit-transformed series is $\ln[P_t / (1 - P_t)]$. Such a series is not bounded.

We could also apply a unit root test that has stationarity as the null hypothesis. One example is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for stationarity (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). The KPSS test can be conducted under the null of either trend stationarity or level (no trend) stationarity. However, the KPSS test also has little power to reject the null, which often leaves us unable to reject either the unit root or the stationarity hypothesis.

A more recent advancement is the Dickey-Fuller generalized least squares test (GLS). The Dickey-Fuller GLS test is the same as the Dickey-Fuller test except that the data are first transformed using a GLS regression. This is done just to give the test greater power (but only a little). We will come back to the Dickey-Fuller GLS test later.

It is also useful to test the null hypothesis that our data come from a random walk with drift process:

$$y_t = y_{t-1} + \alpha_0 + \varepsilon_t. \qquad (6.2.4)$$

If we do have such a data-generating process, first differencing produces a stationary, nontrending process: $\Delta y_t = \alpha_0 + \varepsilon_t$. While a random walk with drift process will appear to trend, it would be inappropriate to partial out a trend, like we would with a trend-stationary process. First differencing is the appropriate procedure. Therefore, it is necessary to distinguish between a trend-stationary and a random walk with drift process. We can begin by testing for a unit root process with drift using a variation on the Dickey-Fuller test. The test proceeds by estimating

$$\Delta y_t = \theta y_{t-1} + \alpha_0 + \varepsilon_t. \qquad (6.2.5)$$

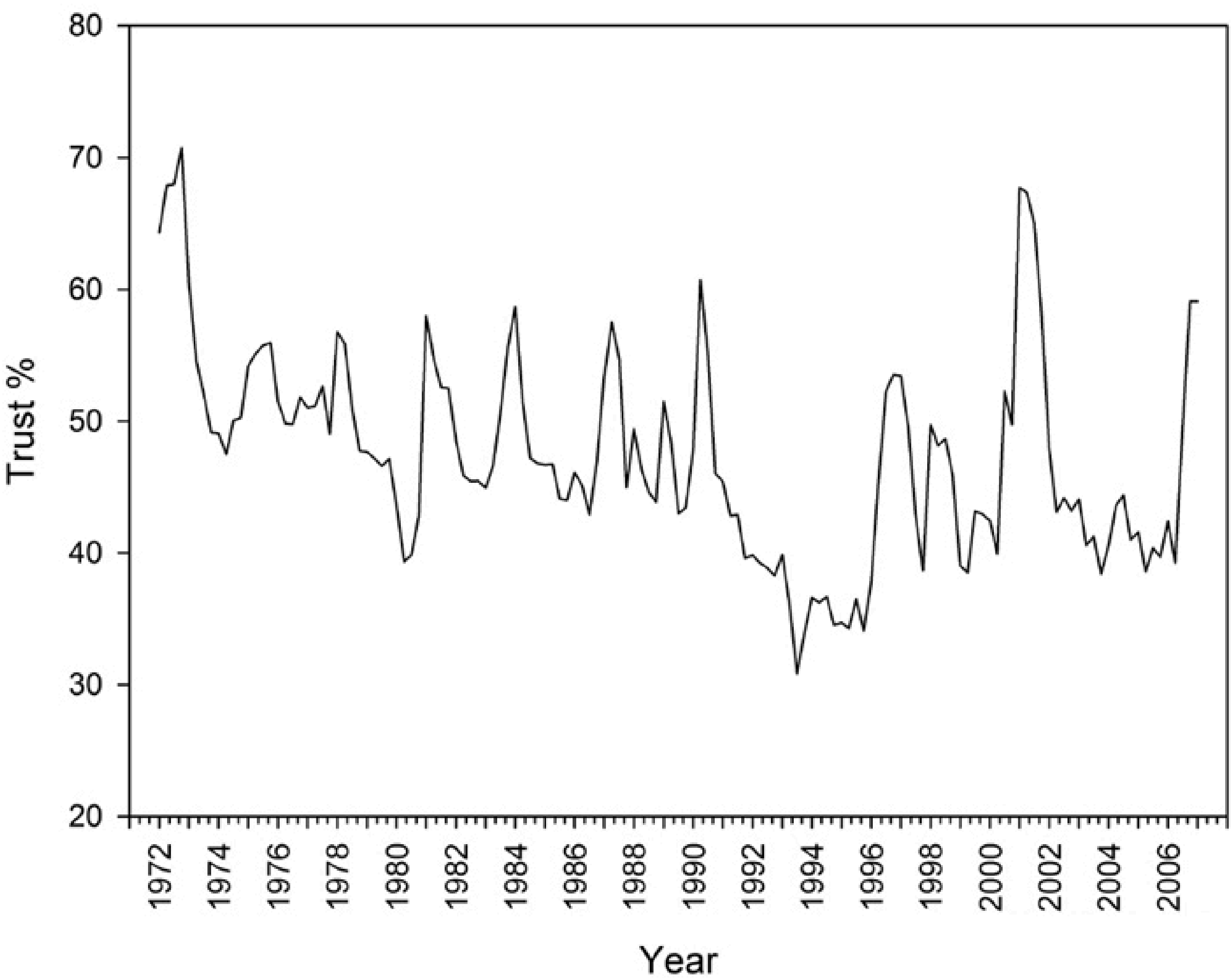
This is the same regression as Equation 6.2.3, but because the null is a unit root with drift, we test that $\theta = 0$, $\alpha_0 \neq 0$ (Hamilton, 1994). The process under the null hypothesis is a random walk with drift. The alternative is again a stationary process. This is called an augmented Dickey-Fuller test.[3] If we would like to test the null hypothesis of a unit root process (with or without drift) against the alternative of a trend-stationary process, we estimate

$$\Delta y_t = \theta y_{t-1} + \alpha_0 + \beta_1 t + \varepsilon_t$$

and test the null hypothesis: θ = 0 and $ß_1$ = 0. The null hypothesis is a unit root process (with or without drift).

For our next example, let's look at quarterly data on trust. Trust is operationalized as the percentage of respondents from quarterly surveys indicating that they felt “you can generally trust people.” Examining the trust variable in Figure 6.3, we can see that there does appear to be a trend or a drift.

Figure 6.3 Percentage of Respondents Indicating That You Can Generally Trust People



We can use the augmented Dickey-Fuller test to test the null hypothesis that the trust variable has a (a) unit root or (b) unit root with drift data generating process. Beginning with the first, the test statistic is -3.79, with a corresponding *P* value of 0.003. It appears that we can reject the null of a unit root process.

For the augmented Dickey-Fuller test of a unit root process with drift, the test statistic does not change; rather, it is the critical values that change. Now the corresponding *P* value is <0.001. It appears that we can also reject the null hypothesis of a unit root process with drift. In both cases, the alternative was a stationary

AR(1) process. For the augmented Dickey-Fuller test of a unit root process (with or without drift), against the alternative of a trend-stationary process, the *P* value is now 0.020. Allowing for the alternative of a trend-stationary process, we can once again reject the null hypothesis of a unit root process.

It is important to note that the Dickey-Fuller test (with stationarity as the alternative) is unlikely to reject the null hypothesis of a unit root if the series is trend stationary. Therefore, if a data series is clearly trending or drifting, we will want to test against the alternative of a trend-stationary data-generating process by using the *augmented* Dickey-Fuller test, or we might mistake a trend-stationary process for one with a unit root.

It is also important to note that unit root tests are less likely to reject the null hypothesis of a unit root, when they should, if the series contains a structural break (Campos, Ericsson, & Hendry, 1996). Therefore, if the series of interest clearly contains an equilibrium shift, this should be partialled out before unit root testing is conducted.

There is a possible alternative to the unit root/stationary dichotomy. It is possible that the process is fractionally integrated. This concept has been gaining popularity since it was originally introduced (Granger & Joyeux, 1980). A fractionally integrated time series process is one that is neither I(0) stationary nor I(*d*) integrated for any integer value of *d.* It is an integrated process of a noninteger value of *d.* This means that it can be transformed into a stationary series through a form of differencing called fractional differencing, just as an integrated series can be transformed into a stationary series through differencing. This is an advanced topic, and the interested reader is directed to the following advanced readings on fractional integration: Granger (1980), Granger and Joyeux (1980), Box-Steffensmeier and Smith (1996, 1998), Lebo and Moore (2003), and Pickup (2009).

Note that we have been using an AR(1) process (with or without a trend) as the alternative when testing for a unit root (with or without drift). The augmented Dickey-Fuller test and KPSS test can also be applied with other stationary processes as the alternative. We can add *p* lags of Δy_t to allow for more dynamics in the process. The lags are intended to account for additional dynamics. If we use too few, the test will not be right. Returning to our trust example, we apply the augmented Dickey-Fuller test, with a trend-stationary process as the alternative hypothesis, and add three lags of Δy_t. Table 6.1 provides the results from the regression used to produce the augmented Dickey-Fuller test statistic.

Table 6.1 Regression Table for Dickey-Fuller Test

Trust	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
L1. Trust	−0.29	0.065	−4.46	0.00
LD. Trust	0.23	0.086	2.65	0.009
L2D. Trust	0.095	0.087	1.09	0.276
L3D. Trust	−0.081	0.088	−0.92	0.361
Trend	−0.011	0.011	−1.07	0.286
Constant	14.34	3.52	4.07	0.00

NOTE: L1 = first lag, L2 = second lag, L3 = third lag, D = difference.

The statistical significance of the first lag (of the first difference) suggests that it is a good idea to include it and that the alternative to a unit root (with or without drift) data-generating process is an AR(2) process. The Dickey-Fuller test statistic is -4.46, with a corresponding *P* value of 0.002. We can still reject the null hypothesis of a unit root process against the alternative of a trend-stationary process.

The only difficulty with including additional lags of the first difference is that it reduces the number of time points available; remember that the Dickey-Fuller (and KPSS) test is already problematic without a long time series of data. A solution to this problem is the Phillips-Perron test for a unit root (Phillips & Perron, 1988). This is the same as the Dickey-Fuller test but uses Newey-West heteroskedasticity—and autocorrelation—consistence standard errors, instead of including lags of the first difference dependent variable.

As with the Dickey-Fuller test, we can also specify the alternative hypothesis of a trend-stationary process. Specifying standard errors that are corrected for up to four lags of serial correlation, the Phillips-Perron tau test statistic is -3.89, with a corresponding *P* value of 0.013. Based on these results, we can again reject the null hypothesis of a unit root process against the alternative of a trend-stationary process.

The fact is that we often do not have enough data to reject a unit root process, in which case we may need to rely on evidence from other similar but longer time series to decide. If we truly do believe that we have a unit root data-generating process, then we will want to take account of this. We have seen that first differencing our data is one approach. We turn to models that use differenced data next, but first we consider higher-order unit root processes.

It is possible to test for higher-order unit root processes with the tests we have discussed, such as the Dickey-Fuller test. For example, y_t is an I(2) process if it is stationary after being differenced twice. If we apply the

Dickey-Fuller test to the first difference of y_t, we are testing the null hypothesis of a second order unit root process I(2) against the alternatives of a first order unit root process I(1) or a stationary process—see Enders (204, pp. 194–195) for details. There are more powerful (and complex) ways of testing for higher-order unit root processes (e.g., Dickey & Pantula, 2002), but it is rare within the social sciences to go beyond testing for an I(1) process.

6.3 Differenced Data and Autoregressive Integrated Moving Average (ARIMA) Models

This brings us next to two new data models. If the dependent variable has an I(1) data-generating process, we can use what is called a differenced data model. Starting with the autoregressive distributed lag model, ADL(1,1),

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t \mp \beta_2 x_{t-1} + \varepsilon_t. \qquad (6.3.1)$$

If we believe that y_t is I(1), $\alpha_1 = 1$, this is an inappropriate model, but Equation 6.3.1 with $\alpha_1 = 1$ can be transformed into a differenced data model. Subtract y_{t-1} from each side, and assume that $\beta_2 = -\beta_1$; then,

$$y_t - y_{t-1} = \alpha_0 + y_{t-1} - y_{t-1} + \beta_1 \left(x_t - x_{t-1}\right) + \varepsilon_t,$$

$$\Delta y_t = \alpha_0 + \beta_1 \Delta x_t + \varepsilon_t. \qquad (6.3.2)$$

This is a differenced data model (Hendry, 2003). If y_t is I(1), Δy_t is I(0); and this model is appropriate for estimation by OLS or maximum likelihood estimation. Note that x_t is also first differenced.[4] What does it mean to assume that $\beta_2 = -\beta_1$ in this context? It assumes that a permanent one-unit change in x_t at t has an effect at t, and that is the sum total of the effect—like a static model. This type of approach has been suggested as appropriate for consumption models (Campbell & Mankiw, 1991).

After estimating a differenced data model, we might find that the residuals are serially correlated. This motivates our next model: An ARIMA model is simply an ARMA model with the dependent and independent variables differenced one or more times.

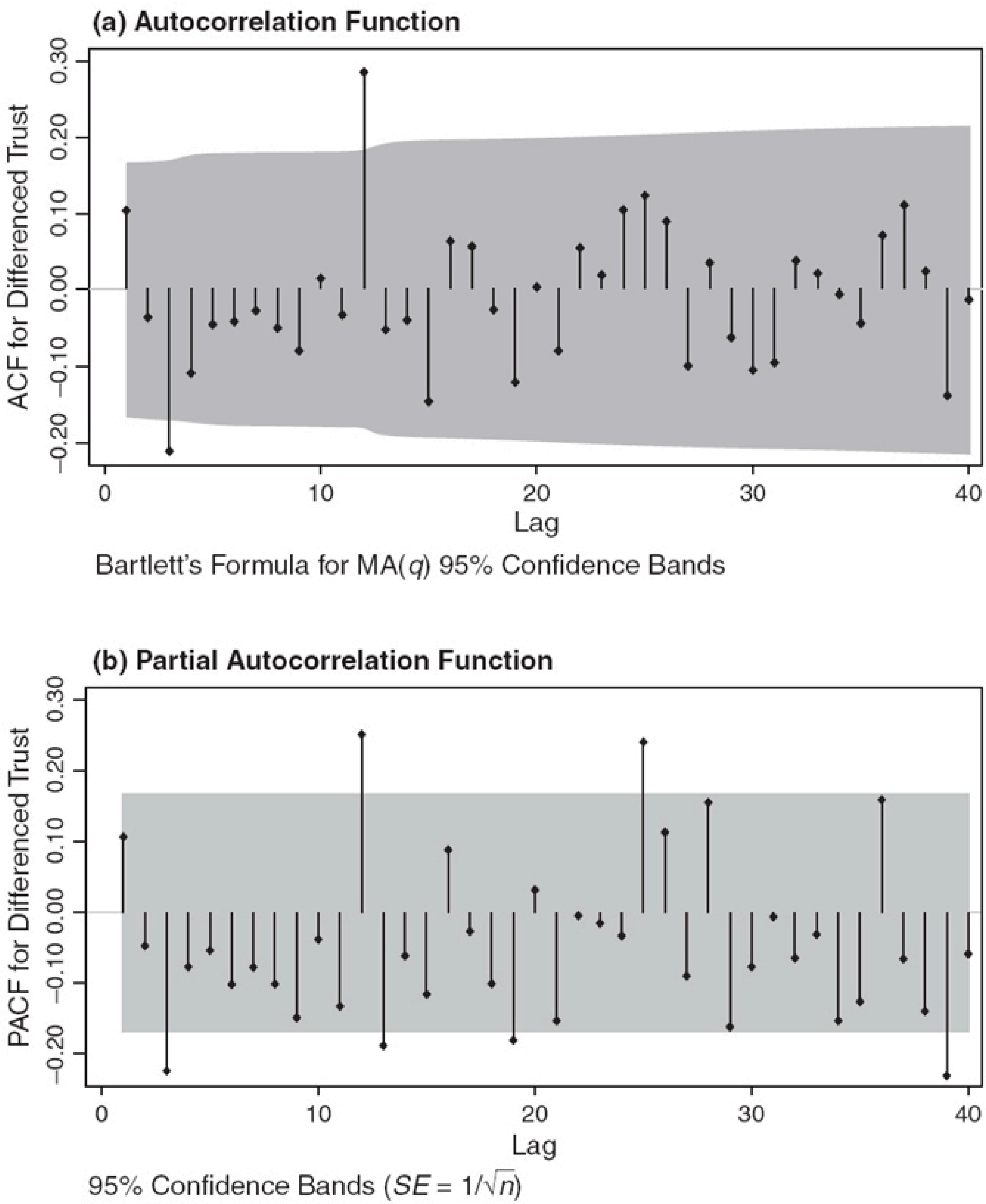
The purpose of an ARIMA model is simply to difference transform a variable with an I*(d)* data-generating process into one with an I(0) data-generating process and then apply the usual approach to building an ARMA model—including autoregressive and moving average components. If it does not contain any autoregressive or moving average components, then it is a differenced data model. In an ARIMA model, we may also difference the data to account for seasonality. Sometimes such seasonal ARIMA models are denoted by the acronym SARIMA.

Let us say that we wish to run a difference data model on the trust variable from our previous example, with civic engagement as an independent variable.[5] We specify an ARIMA with zero autoregressive components, zero moving average components, and once differencing of the variables as follows:

$$\Delta \text{trust}_t = \alpha_0 + \beta_1 \Delta \text{engage}_t + \varepsilon_t.$$

In other words, we have a differenced data model. Before running this model, we examine the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for Δtrust_t. These are presented in Figure 6.4.

Figure 6.4 ACF and PACF for Differenced Trust



NOTE: ACF = autocorrelation function, PACF = partial autocorrelation function, MA = moving average, *SE* = standard error.

Now let us say we estimate an ARIMA, based on Figure 6.4, with the independent and dependent variables differenced once, zero autoregressive components, and an *MA*(q = 3) component. The results are presented in Table 6.2.

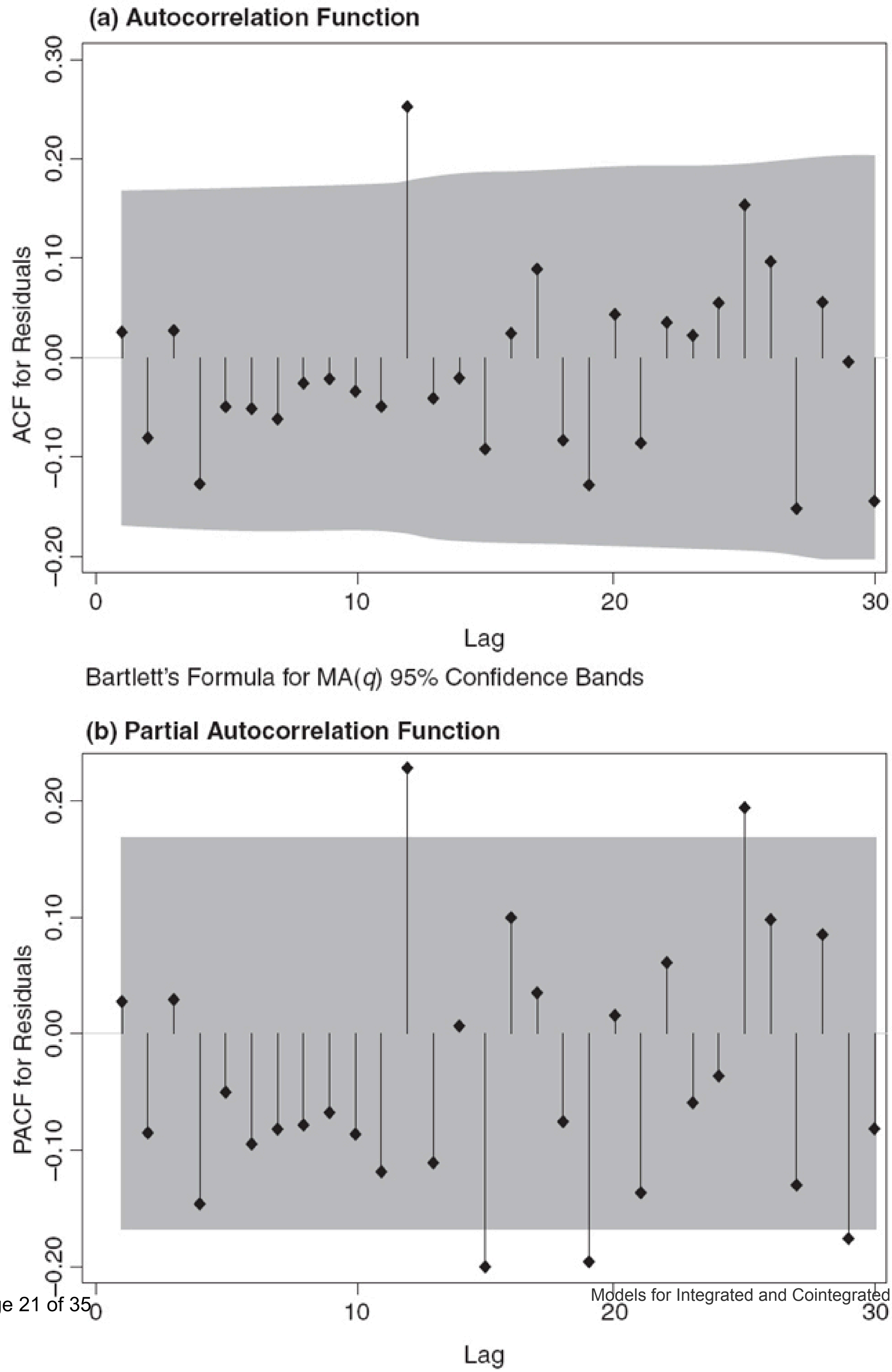
Table 6.2 ARIMA (d = 1, q = 3) Model of Trust

D1. Trust	*Coefficient*	*Standard Error*	*z Statistic*	*P Value*
D1. Engage	0.70	0.29	2.43	0.015
Constant	−0.087	0.31	−0.28	0.780
L3. MA	−0.29	0.11	−2.56	0.010

NOTE: Log likelihood = −393.439, T = 136; D1 = first difference, L3 = third lag, ARIMA = autoregressive integrated moving average model, MA = moving average, P = probability, T = number of time points.

For the residuals, we estimate the ACF and PACF (Figure 6.5) and calculate the Q statistic.

Figure 6.5 ACF and PACF for Residuals From ARIMA *(d* = 1, *q* = 3) Model of Trust


(a) Autocorrelation Function
ACF for Residuals
Lag
Bartlett's Formula for MA(*q*) 95% Confidence Bands
(b) Partial Autocorrelation Function
PACF for Residuals
Lag

The portmanteau *(Q)* statistic is 47.81 and is chi-squared distributed with 40 degrees of freedom. The corresponding *P* value is 0.185. Therefore, we cannot reject the null hypothesis of a white noise process for the residuals from this model.

This model assumes that there is no long-run effect of civic engagement on trust, beyond the short-run effect, in the data-generating process (or at very least, we do not model it). If this is not true, the data model is misspecified, and the inference regarding the significance of the parameters will be incorrect or incomplete. We will also not properly understand the relationship between trust and civic engagement. These issues will motivate the next model: the ECM.

6.4 Cointegration and the Error Correction Model (ECM)

Before discussing ECMs, we need to understand the concept of cointegration. Say for two I(1) data-generating processes, y_t and x_t, there is a *ß* such that $s_t = y_t - ßx_t$ is a stationary process (i.e., an I(0) process). If so, we say that y_t and x_t are cointegrated and call *ß* the cointegration parameter (Engle & Granger, 1987). We denote this relationship C(1,1), meaning that y_t and x_t are I(1) and cointegrated to produce a series that is one order of integration lower—that is, I(0). If we know *ß*, testing for cointegration is straightforward.

1. First confirm that y_t and x_t are both I(1), using the Dickey-Fuller or equivalent test.
2. Second, define $s_t = y_t - ßx_t$. We could also add a constant, trend, periodicity, or structural break to this if we believed that it was part of the cointegrating process. For example, $s_t = y_t - α- ßx_t$.
3. Third, run the Dickey-Fuller or equivalent test on s_t and if we reject a unit root, then y_t and x_t are cointegrated.

Of course, *ß* is generally unknown, as is any constant, trend, periodicity, or structural break, so we first have to estimate these, which adds a complication. We can estimate $y_t = ßx_t + u_t$, and calculate

$$\hat{s}_t = \hat{\mu}_t = y_t - \hat{\beta}x_t. \qquad (6.4.1)$$

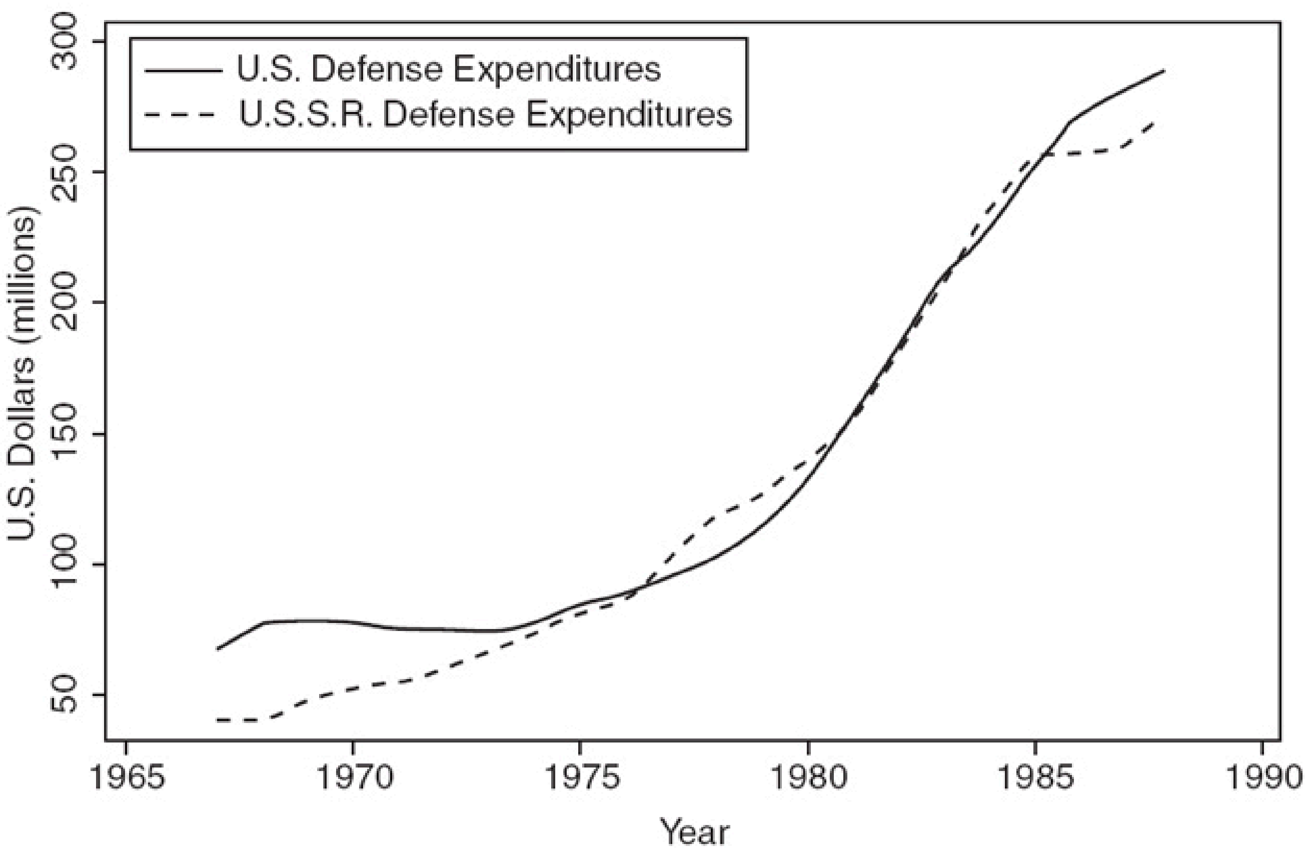
And we can apply the Dickey-Fuller test to $\hat{s}_t$, in order to determine if it is I(0), but this requires different critical values from those for the usual Dickey-Fuller test. A higher threshold is needed to take into account that the *ß* is estimated and not known ahead of time, and therefore, the Dickey-Fuller test is applied to estimates of s_t. The appropriate critical values can be looked up (MacKinnon, 2010).

If we believe that it is part of the cointegrating relationship, we can include a constant or trend to the initial regression that estimates *ß*, and use different critical values again when testing whether $\hat{s}_t$ is I(0). The common practice is to include a constant: $\hat{s}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$. We shall revisit the issue of including constants or trends in cointegrating relationships shortly.

If the two I(1) processes are cointegrated, *β* tells us the long-run relationship between y_t and x_t. We will return to this later. If the two I(1) processes are not cointegrated, then $y_t = \hat{\alpha} - \hat{\beta}x_t + \hat{\mu}_t$ is a spurious regression, the function $s_t = y_t - \alpha - \beta x_t$ has no meaning, and $\hat{\beta}$ tells us nothing.

Let us introduce a new example: the relative defense expenditures of the United States and U.S.S.R. These are yearly defense expenditures in U.S. dollars from 1967 to 1988 (Ostrom, 1990; Ostrom & Marra, 1986). They are plotted in Figure 6.6.

Figure 6.6 U.S. and U.S.S.R. Defense Expenditures in Millions of U.S. Dollars, 1967 to 1988



As expenditures are drifting/trending upward, we begin by testing the U.S. defense expenditures against the null hypothesis of a unit root process with or without drift, including a trend in the alternative hypothesis. The test statistic for the augmented Dickey-Fuller test is -1.097, with a corresponding *P* value of 0.93. It would seem that we cannot reject the null hypothesis of a unit root process against the alternative of a trend-stationary process, but let us also test the null hypothesis of a unit root process with drift against the alternative of a stationary process without the trend.

The test statistic for the augmented Dickey-Fuller test is 2.87, with a corresponding *P* value of 0.995. It appears that we cannot reject the null hypothesis of a unit root process with drift. We do the same thing for U.S.S.R. defense expenditures.

Testing the null hypothesis of a unit root process against the alternative of a trend-stationary process, the test statistic for the augmented Dickey-Fuller test is -1.85, with a corresponding *P* value of 0.68. Testing the null hypothesis of a unit root process with drift, the test statistic for the augmented Dickey-Fuller test is 1.82, with a corresponding *P* value of 0.958. Again, it would seem that we cannot reject the null hypothesis of a unit root process with drift and we cannot reject the null hypothesis of a unit root process against the alternative of a trend-stationary process.

On the basis that we believe that both U.S. and U.S.S.R. defense expenditures are unit root processes (with drift), we regress U.S. expenditures on U.S.S.R. expenditures and request the residuals. But first we reexamine the data in Figure 6.6. As noted, both series seem to contain a drift or trend. When testing each series against the hypothesis of a unit root process, it was necessary to control for this drift/trend. This does not necessarily mean that there is a trend within the cointegrating relationship. In fact, a strict definition of cointegration requires the cointegrating relationship to produce a stationary series, ruling out the possibility of a trend. However, this can be relaxed to allow the cointegrating relationship to produce a trend-stationary series.

For our current example, we include a constant in the cointegrating relationship (as is the convention) and choose to be agnostic about the trend by also including it.

$$\mathrm{US}_t = \lambda + \kappa_1 \mathrm{USSR}_t + \alpha_1 \mathrm{Year}_t + \mu_t . \qquad (6.4.2)$$

The results from the OLS estimation of this model are presented in Table 6.3.

Table 6.3 Cointegrating Relationship Between U.S. and U.S.S.R. Defense Expenditures

U.S. Defense Expenditures	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
U.S.S.R. Defense Expenditures	1.42	0.11	13.33	0.000
Year	−6.22	1.35	−4.60	<0.001
Constant	12,261.35	2,662.69	4.60	<0.001

NOTE: $R^2 = 0.99$, $T = 22$; T = number of time points.

This gives us the potential cointegrating relationship. From this cointegrating relationship, we predict residuals and test for cointegration.

$$\hat{s}_t = \hat{\mu}_t = \mathrm{US}_t - \hat{\lambda} - \hat{\kappa}_1 \mathrm{USSR}_t - \hat{\alpha}_1 \mathrm{Year}_t \qquad (6.4.3)$$

The Dickey-Fuller test statistic for the residuals is -2.35. The critical value for cointegration, including a deterministic trend, at the 0.05 significance level is -4.24 (MacKinnon, 2010), as the test statistic is not lower than the critical value (left-tailed test). It appears that we cannot reject the null of a unit root process in the residuals, and therefore, we cannot reject the hypothesis that the variables are not cointegrated. In other words, this test does not provide evidence of cointegration. However, we shall find later in this chapter that this test is no longer the preferred test for cointegration and the preferred test does find evidence of cointegration.[6]

Error Correction Models

Cointegration allows us to expand the type of dynamic models that can be estimated for I(1) data beyond using first differences. Since $\hat{s}_t$ is stationary, it can be included in time series models. Such models include ECMs. They are also known as *equilibrium correction* models (Hendry, 2003) . Unlike differenced data models, ECMs allow us to examine both long-run and short-run relationships.

A standard ECM is just an ADL(1,1) model that has been transformed as follows. Starting with the ADL(1,1) data-generating process,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t. \qquad (6.4.4)$$

Subtracting y_{t-1} from each side of Equation 6.4.4, we get

$$y_t - y_{t-1} = \alpha_0 + \left(\alpha_1 - 1\right) y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t,$$

$$\Delta y_t = \alpha_0 + \left(\alpha_1 - 1\right) y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t. \qquad (6.4.5)$$

Adding and subtracting $\beta_1 x_{t-1}$, on the right-hand side of Equation 6.4.5, we get

$$\Delta y_t = \alpha_0 + \left(\alpha_1 - 1\right) y_{t-1} + \beta_1 x - \beta_1 x_{t-1} + \beta_1 x_{t-1} + \beta_2 x_{t-1} + \varepsilon_t.$$

Note that

$$\beta_1 x_t - \beta_1 x_{t-1} = \beta_1 \Delta x_t \text{ and } \beta_1 x_{t-1} + \beta_2 x_{t-1} = \left(\beta_1 + \beta_2\right) x_{t-1},$$

$$\Delta y_t = \alpha_0 + \left(\alpha_1 - 1\right) y_{t-1} + \left(\beta_1 + \beta_2\right) x_{t-1} + \beta_1 \Delta x_t + \varepsilon_t. \qquad (6.4.6)$$

Next, use

$$\left(\beta_1 + \beta_2\right) = \frac{\left(\alpha_1 - 1\right)\left(\beta_1 + \beta_2\right)}{\left(\alpha_1 - 1\right)} = -\frac{\left(\alpha_1 - 1\right)\left(\beta_1 + \beta_2\right)}{\left(1 - \alpha_1\right)}.$$

And insert into Equation 6.4.6

$$\Delta y_t = \alpha_0 + \left(\alpha_1 - 1\right) y_{t-1} - \frac{\left(\alpha_1 - 1\right)\left(\beta_1 + \beta_2\right)}{\left(1 - \alpha_1\right)} x_{t-1} + \beta_1 \Delta x_t + \varepsilon_t. \qquad (6.4.7)$$

Collect the terms multiplied by $(\alpha_1 - 1)$:

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1)\left[y_{t-1} - \frac{(\beta_1 + \beta_2)}{(1-\alpha_1)} x_{t-1} \right] + \beta_1 \Delta x_t + \varepsilon_t. \qquad (6.4.8)$$

This transformation motivates the standard ECM:

$$\Delta y_t = \alpha_0 + \gamma\left(y_{t-1} - \kappa_1 x_{t-1} \right) + \kappa_0 \Delta x_t + \varepsilon_t, \qquad (6.4.9)$$

where $\Gamma \equiv (\alpha_1 - 1)$, $\kappa_0 \equiv \beta_1$, and $\kappa_1 \equiv (ß_1 + ß_2)/(1 - \alpha_1)$.

As we will discuss later in the chapter, ECMs can be applied to stationary data—in which case, κ_0 is equivalent to the ADL short-run impact of a unit change in x_t, and κ_1 is equivalent to the ADL long-run impact; and this is exactly how they are interpreted. This standard ECM models an ADL(1,1) process where the short- and long-run effects are explicitly modelled as parameters. However, if y_t and x_t are C(1,1), the long-run effect term is the cointegrating relationship between them:

$$y_{t-1} - \kappa_1 x_{t-1} = s_{t-1},$$

which is a different type of long-run relationship. In this long-run relationship, two I(1) variables track each other in such a way that the (proportional) difference is a stationary process. If y_t and x_t are I(1), the remaining terms in the ECM (Δy_{t-1} and Δx_{t-1}) are stationary processes. Therefore, all constituent processes in the ECM are stationary (as long as y_t and x_t are cointegrated). The constant in Equation 6.4.9 can be moved into the cointegrating relationship, and the ECM can be rewritten as follows:

$$\Delta y_t = \gamma\left(y_{t-1} - \lambda - \kappa_1 x_{t-1} \right) + \kappa_0 \Delta x_t + \varepsilon_t, \qquad (6.4.10)$$

where $\lambda \equiv -(\alpha_0 / \Gamma) = \alpha_0 /(1- \alpha_1)$, which is in the estimate of the ADL(1,1) equilibrium if the data are stationary.

A nonzero constant outside the cointegrating relationship, as in Equation 6.4.9, can imply a linear trend in the levels of y_t, just as it does in the random walk with drift process. It may, however, imply that a constant should be included in the cointegrating relationship (Equation 6.4.10). In fact, the data-generating process may contain both. This possibility regarding the data-generating process is true regardless of whether or not the constant is included inside (Equation 6.4.10) or outside (Equation 6.4.9), the cointegrating relationship in the data model. It makes theoretical sense to include both if you believe that the data-generating process includes both. However, such a model is not identified, and some identifying strategy is required. Such a strategy is left to a more advanced text (Johansen, 1988; Maddala & Kim, 1998,), but a general guide is to include the constant (drift term) outside the cointegrating relationship if y_t exhibits a trend and to include it within the cointegrating relationship otherwise.

Just as a constant can be included inside or outside the cointegrating relationship, a trend term can be included inside or outside the cointegrating relationship. However, they imply different data-generating processes. A trend within the cointegrating relationship is appropriate if y_t and x_t are cointegrated, controlling for the linear trend. A trend outside the cointegrating relationship is appropriate if Δy_t contains a trend that would exhibit itself as a quadratic trend in y_t.

If the data are cointegrated, the ECM can be estimated using the Engle-Granger two-step procedure (Engle & Granger, 1987). For example, if we were applying this to the economic popularity model we have seen a number of times already, including for the moment only gross domestic product (GDP) as a covariate, the ECM would be as follows:

$$\Delta \text{Pop}_t = \gamma\left(\text{Pop}_{t-1} - \lambda - \kappa_1 \text{GDP}_{t-1}\right) + \kappa_0 \Delta \text{GDP}_t + \varepsilon_t .$$

The Engle-Granger two-step procedure would proceed as follows:

In Step 1, estimate by OLS	$\text{Pop}_t = \lambda + \kappa_1 \text{GDP}_t + \mu_t$.
In Step 2, calculate	$\hat{s}_t = \hat{\mu}_t = \text{Pop}_t - \hat{\lambda} - \hat{\kappa}_1 \text{GDP}_t$.
And by OLS, estimate	$\Delta\text{Pop}_t = \gamma \hat{s}_{t-1} + \kappa_0 \Delta\text{GDP}_t + \varepsilon_t$

Unfortunately, we cannot use the standard errors from Step 1 for hypothesis testing (e.g., Is the coefficient statistically different from 0?) for κ_1 because the variables are unit root. This is not a problem when estimating the coefficient of cointegration, but it is when testing the statistical significance of κ_1. The solution is to use the regression results from Step 1 to estimate $\hat{s}_t$, but to use a separate procedure to test the significance of κ_1.

One available technique is the leads and lags estimate. This technique involves reestimating the regression from Step 1 with the addition of the first difference of Δx_t (in this case ΔGDP_t) and both leads and lags of Δx_t: Δx_{t+1}, Δx_{t-1}, Δx_{t+2}, Δx_{t-2}, …. Leaving the rationale for an advanced text (Stock & Watson, 1993), the inclusion of enough leads and lags produces unbiased estimates of the standard errors for $\hat{\kappa}_1$. The number of leads and lags to include is going to be restricted by the data. If we have only a few data points, each additional lead and lag increases our standard errors; however, a greater number of leads and lags means greater confidence in our results. A single lead-lag estimate would proceed as follows. (1) Estimate

$$\text{Pop}_t = \lambda + \kappa_1 \text{GDP}_t + \psi_0 \Delta\text{GDP}_t + \psi_1 \Delta\text{GDP}_{t+1} + \psi_2 \Delta\text{GDP}_{t-1} + \mu_t .$$

The ψ_i values are just parameters to be estimated. (2) Use the new estimate of κ_1 and the new estimate of its standard errors to test the null hypothesis $\kappa_1 = 0$.

We return to our U.S. and U.S.S.R defense expenditures example.[7] As noted previously, we will learn later in this chapter that there is evidence that the U.S. and U.S.S.R. defense expenditures are cointegrated C(1,1). Anticipating that evidence, we use the Engle-Granger two-step procedure to estimate an ECM using the U.S. and U.S.S.R. data. The results presented in Table 6.3 represent Step 1, giving us

$$\text{US}_t = \hat{\lambda} + \hat{\kappa}_1 \text{USSR}_t + \hat{\alpha}_1 \text{Year}_t + \hat{\mu}_t,$$

from which we can estimate the residuals:

$\hat{s}_{t-1} = \text{US}_t - \hat{\lambda} - \hat{\kappa}_1 \text{USSR}_t - \hat{\alpha}_t \text{Year}_t$. In Step 2, we estimate the ECM:

$$\Delta \mathrm{US}_t = \gamma \hat{s}_{t-1} + \kappa_0 \Delta \mathrm{USSR}_t + \varepsilon_t.$$

Note that we have not included a constant in this step, as we have included it in the cointegrating relationship.

The results in Table 6.4 give us the immediate effect of U.S.S.R. expenditures on U.S. expenditures: $\hat{\kappa}_0 = 0.86$. This effect is significant at the 95% confidence level. We can also test the residuals from the ECM against the null hypothesis of a white noise process. The portmanteau *Q* statistic is 15.10 and is chi-squared distributed with 8 degrees of freedom. The corresponding *P* value is 0.057. Therefore, we cannot reject the null hypothesis of a white noise process for the residuals from this model at the 0.05 significance level.

Table 6.4 U.S. and U.S.S.R. Defense Expenditure Error Correction Model

D1. U.S. Defense Expenditures	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
D1. U.S.S.R. defense expenditures	0.86	0.096	8.89	<0.001
L1. Cointegrating relationship	−0.48	0.15	−3.22	0.005

NOTE: $R^2 = 0.59$, $T = 21$; T = number of time points, D1 = first difference, L1 = first lag.

Note that the estimated coefficient of cointegration for U.S.S.R. and U.S. expenditures is $\hat{\kappa}_1 = 1.42$ (Table 6.3), but we cannot use the cointegrating regression to test the significance of this coefficient. To do this, we estimate the lead-lag estimator of the cointegrating relationship, including one lead and one lag.

The result (Table 6.5) is that the coefficient of cointegration for U.S.S.R. and U.S. expenditures does appear to be significant at the 0.05 significance level. Note, however, that the estimated coefficient from the lead-lag estimator is different from that from the original OLS regression: $\hat{\kappa}_1 = 1.61$. This is a problem with the Engle-Granger two-step procedure. It has been demonstrated that over repeated samples, there will be no difference between the coefficient from the lead-lag estimator and that from the OLS estimator (Engle & Granger, 1987), but of course, this is not true in any given sample.

Table 6.5 U.S. and U.S.S.R. Defense Expenditure Error Correction Model

U.S. Defense Expenditures	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
Year	−8.36	0.95	−8.78	<0.001
U.S.S.R. defense expenditures	1.61	0.068	23.62	<0.001
D	−0.96	0.22	−4.44	<0.001
FD	0.76	0.20	3.86	0.002
LD	−0.50	0.21	−2.35	0.035
Constant	16,469.93	1,874.57	8.79	<0.001

NOTE: $R^2 = 0.996$, $T = 19$; P = probability, T = number of time points, D = first difference, FD = first lead of the first difference, LD = first lag of the first difference.

The results from Table 6.5 show us that U.S.S.R. defense expenditures have a statistically significant effect within the cointegrating relationship. For the U.S.S.R. defense expenditures to have a statistically significant effect on U.S. defense expenditures, it is also necessary that $\hat{s}_{t-1}$ has a statistically significant coefficient within the main ECM. The results presented in Table 6.4 suggest that this is the case at the 0.05 significance level. The cointegrating relationship is a type of long-run equilibrium between x_t and y_t. Therefore, the effect of the cointegrating relationship on US expenditures indicates that US expenditures respond to US and USSR expenditures being out of their equilibrium, so as to bring the two back into equilibrium—see Enders (2010) for a further discussion of interpreting ECMs.

The simple ECM examined so far can be extended in a number of ways. First, additional variables can be added to the right-hand side. For example, if $y_{t-1} - \lambda - \kappa_{1,1}x_{t-1} - \kappa_{1,2}z_{t-1}$ represents a cointegrating relationship, our ECM could be as follows:

$$\Delta y_t = \gamma\left(y_{t-1} - \lambda - \kappa_{1,1}x_{t-1} - \kappa_{1,2}z_{t-1}\right) + \kappa_{0,1}\Delta x_t + \kappa_{0,2}\Delta z_t + \varepsilon_t. \qquad (6.4.11)$$

We could also add additional lags of Δx_t and/or Δy_t to the ECM. For example,

$$\Delta y_t = \gamma\left(y_{t-1} - \lambda - \kappa_1 x_{t-1}\right) + \kappa_0\Delta x_t + \kappa_2\Delta x_{t-1} + \kappa_3\Delta y_{t-1} + \varepsilon_t. \quad (6.4.12)$$

In doing so, the ECM is no longer isomorphic to the ADL(1,1) model and would not be appropriate for the stationarity ADL(1,1) data-generating process, but it is a perfectly valid model for the corresponding data-generating process.

We are now in a position to discuss more recent developments in time series analysis involving cointegration. Such developments have occurred in the context of vector error correction models (VECMs). This topic is beyond the scope of this book, but a brief introduction will allow us to understand modern approaches to testing cointegration. Start with the following two error correction processes:

$$\Delta y_t = \gamma_1\left(y_{t-1} - \lambda_1 - \kappa_{1,1}x_{t-1}\right) + \varepsilon_{1t},$$

$$\Delta x_t = \gamma_2\left(x_{t-1} - \lambda_2 - \kappa_{2,1}y_{t-1}\right) + \varepsilon_{2t}. \qquad (6.4.13)$$

Note that we have had to add an extra index on the cointegrating coefficients to keep them clear. In combination, the above two equations (6.4.13) are known as a VECM.

Just as the standard ECM is isomorphic to the ADL(1,1), the VECM in Equation 6.4.13 is isomorphic to the following model, called a vector autoregression model:

$$y_t = \alpha_{1,0} + \alpha_{1,1}y_{t-1} + \beta_{1,1}x_{t-1} + \omega_{1t},$$

$$x_t = \alpha_{2,0} + \alpha_{2,1}x_{t-1} + \beta_{2,1}y_{t-1} + \omega_{2t}. \qquad (6.4.14)$$

The transformation from Equation 6.4.14 back to Equation 6.4.13 is analogous to the transformation from the ADL(1,1) to the standard ECM. Just as in that transformation, it is required that we define $\Gamma_1 = (\alpha_{1,1} - 1)$ and $\Gamma_2 = (\alpha_{2,1} - 1)$. Considering just the equation for y_t in Equation 6.4.14, recall that if y_t is unit-root, then $\alpha_{1,1} - 1 = \Gamma_1 = 0$. If we reject the null hypothesis, $\Gamma_1 = 0$, we conclude that y_t is stationary. If we fail to reject the null hypothesis, we conclude that y_t is unit root.

Returning to the multi-equation case (Equation 6.4.14), we can conduct a test of cointegration by examining the following matrix formed by the parameters in the equation:

$$\begin{bmatrix} \alpha_{1,1} - 1 = \gamma_1 & \beta_{1,1} \\ \beta_{2,2} & \alpha_{2,1} - 1 = \gamma_2 \end{bmatrix}. \qquad (6.4.15)$$

We proceed by estimating the rank of this matrix. The rank is the maximum number of linearly independent rows. Johansen (1995, 1988) demonstrated that if y_t and x_t are cointegrated, the two rows will not be linearly independent and the rank is equal to the number of cointegrating vectors. Also, the maximum number of possible cointegrating equations is the number of rows minus 1. With only two variables, an estimated rank of 2 means that the two rows are independent, which implies that y_t and x_t are both stationary. If the estimated rank is 0 and the matrix is all zeros, we fail to reject both the null hypothesis that $\Gamma_1 = 0$ and the null hypothesis that $\Gamma_2 = 0$, and by Equation 6.4.13,

$$\Delta y_t = \varepsilon_{1t,}$$

$$\Delta x_t = \varepsilon_{2t}. \qquad (6.4.16)$$

We conclude that both *yt* are *xt* are unit root and are not cointegrated. An estimated rank of 1 indicates that a single cointegrating relationship exists. Note that with only two variables, the maximum number of cointegrating vectors is 1.

Johansen (1995, 1988) provides a maximum likelihood estimate of the parameters in Equation 6.4.14 and two test statistics for determining the rank of the matrix in Equation 6.4.15. We will discuss the maximum eigenvalue statistic (λ_{max}). This statistic can be calculated for each possible rank value. It can first be calculated to test the null hypothesis that the rank is 0 against the alternative that the rank is 1. If we reject the null that the rank is 0 against this alternative, we can calculate the statistic to test the null hypothesis that the rank is 1 against the alternative that the rank is 2. Using this procedure, we now reexamine the potential cointegrating relationship between U.S. and U.S.S.R. defense expenditures. As before, we include a constant and a trend within the cointegrating relationship of the VECM.

In Table 6.6, the row that starts with maximum rank 0 tests the null hypothesis that the rank is 0 against the alternative that the rank is 1. In other words, the null hypothesis is that there is no cointegration. This can be rejected. The row that starts with maximum rank 1 tests the null hypothesis that the rank is 1 against the null hypothesis that the rank is 2. The null in this case is that there is one cointegrating equation, and the alternative is that both variables are stationary. This null hypothesis cannot be rejected, meaning that we conclude that there is a cointegrating equation. If we had rejected the null in this case, we would have concluded that the two series are stationary. Based on our test results, we can use the cointegrating relationship regression (Equation 6.4.2) to estimate the cointegrating relationship between USSR and US expenditures.

Table 6.6 Johansen Maximum Eigenvalue Statistics

Maximum Rank	*Parms*	*LL*	*Eigenvalue*	*Maximum Statistic*	*5% Critical Value*
0	6	−124.14	—	21.11	18.96
1	10	−113.59	0.65	9.08	12.52
2	12	−109.05	0.36	—	—

NOTE: LL= Loglikelihood.

Given that both *xt* and *yt* can be seen as dependent variables in a cointegrating relationship, we need to ask when it is valid to estimate the single-equation model for Δ *yt*—as we did in the case of the U.S., U.S.S.R.

defense expenditures example. This can be answered by returning to Equation 6.4.13. The requirement for estimating the single-equation ECM for Δy_t is that Γ_2 in Equation 6.4.13 is equal to 0. If this is the case, then x_t does not respond to x_t and y_t being out of equilibrium. It is the response of y_t that brings them back into equilibrium. Under these circumstances, we can say that x_t is weakly exogenous for the estimation of the single equation for Δy_t in Equation 6.4.13. Recall that we introduced the concept of weak exogeneity in Chapter 2.

Error Correction Models with Stationary Data

As mentioned previously, ECMs can also be used with data produced by stationary data-generating processes. Consider again government popularity and GDP data with stationary data-generating processes. The standard error correction model is as follows:

$$\Delta \mathrm{Pop}_t = \gamma\left(\mathrm{Pop}_{t-1} - \lambda - \kappa_1 \mathrm{GDP}_{t-1}\right) + \kappa_0 \Delta\mathrm{GDP}_t + \varepsilon_t.$$

Multiplying out the cointegrating relationship, this can be rewritten in the “general form”:

$$\Delta \mathrm{Pop}_t = \gamma \mathrm{Pop}_{t-1} - \gamma\lambda - \gamma\kappa_1 \mathrm{GDP}_{t-1} + \kappa_0 \Delta\mathrm{GDP}_t + \varepsilon_t.$$

This can be rewritten as

$$\Delta \mathrm{Pop}_t = \gamma \mathrm{Pop}_{t-1} + \delta_0 + \delta_1 \Delta\mathrm{GDP}_t + \delta_2 \mathrm{GDP}_{t-1} + \varepsilon_t, \qquad (6.4.17)$$

where $\delta_0 = -\Gamma\lambda$, $\delta_1 = \kappa_0$, $\delta_2 = -\Gamma\kappa_1$.

The expression of the standard ECM is easier to interpret because κ_1 is estimated directly, but the general form with stationary data can be estimated using OLS in a single step—this cannot be done with unit root data because of the inclusion of GDP_{t-1}.[8] If Pop_t and GDP_t are stationary, so are $\Delta\mathrm{Pop}_t$ and $\Delta\mathrm{GDP}_t$; therefore, we can use OLS to estimate Equation 6.4.17. After estimating Γ, δ_0, δ_1, and δ_2, we can calculate the estimated long-run effect $\hat{\kappa}_1 = -(\hat{\delta}_2 / \hat{\gamma})$, and the estimated equilibrium, $\hat{\lambda} = -(\hat{\delta}_0 / \hat{\gamma})$.

Using the German government popularity and economic data that we have used before, we estimate the following ECM (we have three independent variables—GDP, inflation, and unemployment) (Table 6.7):

$$\Delta \mathrm{Pop}_t = \gamma \mathrm{Pop}_{t-1} + \delta_0 + \delta_{1a} \Delta\mathrm{GDP}_t + \delta_{2a} \mathrm{GDP}_{t-1} + \delta_{1b} \Delta \mathrm{inf}_t + \delta_{2b} \mathrm{inf}_{t-1}$$
$$+ \delta_{1c} \Delta\mathrm{unemp}_t + \delta_{2c} \mathrm{unemp}_{t-1} + \varepsilon_t.$$

Table 6.7 General Form Error Correction Model for Economic Popularity

D. Vote	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
L1. Vote	−0.62	0.14	−4.58	<0.001
D1. GDP	−0.53	0.44	−1.20	0.236
L1. GDP	−0.54	0.37	−1.45	0.155
D1. Inflation	0.32	1.27	0.25	0.802
L1. Inflation	−1.98	0.84	−2.34	0.024
D1. Unemployment	−5.08	5.56	−0.91	0.366
L1.Unemployment	0.84	2.33	0.36	0.719
Trend	−0.23	0.083	−2.77	0.008
Constant	35.22	18.38	1.92	0.062

NOTE: $R^2 = 0.39$, $T = 51$; T = number of time points, D1 = first difference, L1 = first lag.

Note that we also included a trend as we previously determined that the German government popularity variable is a trend-stationary data-generating process. The trend does not need to be entered as its first difference or lag because the first difference of a trend is simply a constant and the lag of a trend is just a trend.

The estimated coefficient on each of the first-differenced economic variables is the corresponding $\hat{\kappa}_0$, which is the estimated short-run effect. There does not appear to be a significant short-run effect for any of the three economic variables.

The coefficient on each of the lagged economic variables is the corresponding $\widehat{\gamma\kappa}_1$, where $\hat{\kappa}$ is the estimated long-run effect. We also have the estimate of Γ, so we can calculate $\hat{\kappa}$ for each of the independent variables.

For GDP, $$\hat{\kappa}_{1a} = -\left(\frac{\hat{\delta}_{2a}}{\hat{\gamma}}\right) = -\left(\frac{-0.54}{-0.62}\right) = -0.87.$$

For inflation, $$\hat{\kappa}_{1b} = -\left(\frac{\hat{\delta}_{2b}}{\hat{\gamma}}\right) = -\left(\frac{-1.98}{-0.62}\right) = -3.19.$$

For unemployment, $$\hat{\kappa}_{1c} = -\left(\frac{\hat{\delta}_{2c}}{\hat{\gamma}}\right) = -\left(\frac{-5.08}{-0.62}\right) = -8.19.$$

We can test the hypothesis that $\kappa_1 = 0$ with an *F* statistic. For GDP $\hat{\kappa}_{1a}$, the test statistic is 1.96 with an $F(1, 42)$ distribution and a corresponding *P* value of 0.169. For GDP, we cannot reject, at the 0.05 significance level, the null hypothesis that the long-run effect is zero. For inflation $\hat{\kappa}_{1b}$, the test statistic is 7.41 with an $F(1, 42)$ distribution and a corresponding *P* value of 0.009. For inflation, we can reject the null hypothesis, at the 0.05 significance level, that the long-run effect is zero. An increase in inflation by 1 percentage point is estimated to produce a long-run decline in approval of 3.19 percentage points.

For unemployment $\hat{\kappa}_{1c}$, the test statistic is 0.13 with an $F(1, 42)$ distribution and a corresponding *P* value of 0.716. For unemployment, we cannot reject, at the 0.05 significance level, the null hypothesis that the long-run effect is zero.

As always, we want to make sure that the residuals are free from serial correlation and follow a white noise process. The *Q* statistic is 14.52 and is chi-squared distributed with 23 degrees of freedom. The corresponding *P* value is 0.911. We cannot reject the null hypothesis, at the 0.05 significance level, that the residuals follow a white noise process. If the residuals do show signs of serial correlation, we can include lags of the (differenced) dependent variable.

We would also like to test that the residuals are homoskedastic. The Breusch-Pagan test of the null hypothesis of constant variance gives us a test statistic of 9.90. This has a chi-squared distribution with 8 degrees of freedom and a corresponding *P* value of 0.272. We cannot reject the null hypothesis, at the 0.05 significance level, that the residuals are homoskedastic (have constant variance).

Summary

If your data are stationary, you may use the ECM, but the ADL will produce the same results. If your data are integrated and you are willing to assume that there is no long-run relationship between your variables, you can use the differenced data or ARIMA models. If you believe that there is a long-run relationship and that relationship can be represented as a cointegrating relationship, the ECM is of great value. Keep in mind, though, that the single-equation ECM assumes that x_t is not a function of the cointegrating relationship—that is, x_t is weakly exogenous. If this is not an appropriate assumption, you may want to consider the multi-equation VECM (Brandt & Williams, 2007; Enders, 2004).

[1] First differencing a time series that has already been first differenced is called second differencing.

[2] The test used is one-sided (left-tailed), as we are assuming α_1 is not greater than 1 and so θ is not greater than 0.

[3] For a more detailed procedure to follow when testing the null hypothesis of a unit root process, see Enders (2004).

[4] Modelling Δy_t as a function of x_t would be to model a process in which any nonchanging, nonzero value for x_t would cause y_t to steadily and perpetually increase or decrease.

[5] Data are from Keele (2007). The civic engagement variable is based on four indicators: (1) participation in community organizations, (2) participation in politics and public affairs, (3) volunteering, and (4) informal socializing. See Keele (2005, 2007) for further details.

[6] It has been noted that a series may be fractionally integrated. It follows that two or more series may be fractionally cointegrated. Again, this is an advanced topic. See Davidson (2002) and Cheung and Lai (1993) for a review of the issues encountered when testing for fractional cointegration.

[7] For an extensive discussion of the value of error correction models within the discipline of political science, see Durr (1992) and the other articles in the same issue of *Political Analysis.*

[8] The general form ECM is sometimes used with I(1) data as part of a cointegration test. See Ericsson and MacKinnon 2002.

http://dx.doi.org/10.4135/9781483390857.n6