Introducing Dynamic Time Series Models

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: 81-112

Introducing Dynamic Time Series Models

In this chapter, we continue the discussion of modelling time series data with the introduction of dynamic models. These models include the autoregressive distributed lag (ADL), lagged dependent variable (LDV), autoregressive conditional heteroskedasticity (ARCH), and moving average (MA) models. As the conditions necessary for these dynamic models to meet the assumption of covariance stationarity differ from those of static models, these conditions are discussed.

4.1 Autoregressive Distributed Lag Models

In Chapter 2, we discussed the following process:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t, \qquad (4.1.1)$$

with $\varepsilon_t \sim \text{NID}(0,\sigma_\varepsilon^2)$. In other words, the ε_t values are independent and distributed normally with mean 0 and variance $\alpha^2{}_\varepsilon$. If $|\alpha_1| < 1$, this is an autoregressive process of order 1, AR(1). It contains one lag of itself on the right-hand side, y_{t-1}.

An autoregressive process is described by including lags (or just a single lag) of the dependent variable on the right-hand side as explanatory/independent variables. We may also include a constant, α_0, in Equation 4.1.1:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \qquad (4.1.2)$$

Note that we are using α_0 as the constant, rather than $ß_0$, in this model. This constant implies that the process y_t has a nonzero mean. In the static process, $y_t = ß_0 + \varepsilon_t$, the constant is the unconditional mean or expected value of y_t. This can also be interpreted as the equilibrium of the process—the value to which the process reverts if the stochastic component, ε_t, and the deterministic exogenous variables x_t (if there are any) are set to 0. It is the long-run mean (expected value) of the process. In the AR(1) process, the equilibrium of the process is:

$$\beta_0 = \frac{\alpha_0}{1-\alpha_1}. \qquad (4.1.3)$$

This again is the value to which the process eventually returns if the stochastic and exogenous variables are set to 0. This may not have a practical interpretation. It is the long-run expected value of y_t. More precisely, it is the unconditional expected value of y_t: $E(y_t)$. In a dynamic process, the *conditional* expected value, the mean of y_t conditioning on previous values of y_{t-1}, is

$$E(y_t|\, y_{t-1}) = \alpha_0 + \alpha_1 y_{t-1}. \qquad (4.1.4)$$

Note that it is the unconditional expected value that we assume to be constant in order to meet the conditions of covariance stationarity. Unless the dynamic process begins exactly in equilibrium and remains there, the conditional expected value will change over time.

With this in mind, we can define a dynamic time series model called an autoregressive distributed lag model (ADL). Such a model contains lags of the dependent and independent variables on the right-hand side. In the ADL model, it is the inclusion of a lag or lags of the dependent variable on the right-hand side that makes the model dynamic. The ADL model is a combination of an autoregressive process and a finite distributed lag (FDL) model. An ADL(1,1) model contains one lag of the dependent variable and one lag of the independent variable (it also includes a constant and the nonlagged independent variable):

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

$$\varepsilon_t \sim \mathrm{NID}\left(0, \sigma_\varepsilon^2\right). \qquad (4.1.5)$$

More generally, an ADL*(p,m)* model is an autoregressive process of order *p* (with *p* lags of the dependent variable included as independent variables), and it includes *m* lags of the other independent variable(s), in addition to the nonlagged independent variable(s):

$$y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \beta_1 x_t + \sum_{i=1}^{m} \beta_{i+1} x_{t-i} + \varepsilon_t. \qquad (4.1.6)$$

Interpreting the coefficients on independent variables in dynamic models is somewhat complicated by the inclusion of the lagged dependent variable. From an ADL*(p,m)* model, the short-run effect and the long-run effect of the independent variables, and the process equilibrium can be calculated. For the ADL(1,*m*) model, β_1 is the immediate (short-run) effect of a one-unit change in x_t. The long-run effect of a permanent one-unit change is

$$\frac{\beta_1 + \sum_{i=1}^{m} \beta_{i+1}}{1 - \alpha_1}. \qquad (4.1.7)$$

The reason why the short-run and long-run effects of an exogenous variable are different is not immediately obvious. This will be explored further in Section 4.6, but in the meantime, suppose we have an ADL(1,1) model, as written in Equation 4.1.5. Also, suppose we observe y_t up to time $t = \tau$. The conditional expected value of y_t is

$$E(y_t) = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1}. \qquad (4.1.8)$$

We can predict the expected value of y_t at time $t = \tau + 1$ using our observations of y_t up to time $t = \tau$ and our observation of x_t at $t = \tau + 1$. Such a forward prediction is a forecast of $y_{t=\tau+1}$. We denote such a forecast as

$$\hat{E}\left(y_{t=\tau+1|\tau}\right) = \alpha_0 + \alpha_1 y_\tau + \beta_1 x_{\tau+1} + \beta_2 x_\tau. \qquad (4.1.9)$$

The term on the left-hand side reads as the estimated (conditional) expected value of y_t at time $y_{t=\tau+1}$ using observations of y_t up to and including $t = \tau$. Suppose x_t increases by one unit at $\tau + 1$. Since the difference between x_τ and $x_{\tau+1}$ is 1, the effect of the one-unit increase in x_t at $\tau + 1$ on y_t at $\tau + 1$ is a β_1 increase. We can next forecast the value of y_t at time $t = \tau + 2$.

$$\hat{E}\left(y_{t=\tau+2|\tau}\right)=\alpha_0+\alpha_1 y_{\tau+1}+\beta_1 x_{\tau+2}+\beta_2 x_{\tau+1}.$$

We do not know $y_{\tau+1}$, but we can plug our forecast of $y_{\tau+1}$ into the above:

$$\begin{aligned}\hat{E}\left(y_{t=\tau+2|\tau}\right)&=\alpha_0+\alpha_1\left(\alpha_0+\alpha_1 y_\tau+\beta_1 x_{\tau+1}+\beta_2 x_\tau\right)+\beta_1 x_{\tau+2}+\beta_2 x_{\tau+1}\\&=\alpha_0+\alpha_1\alpha_0+\alpha_1^2 y_\tau+\beta_1 x_{\tau+2}+\left(\alpha_1\beta_1+\beta_2\right)x_{\tau+1}+\alpha_1\beta_2 x_\tau.\quad(4.1.10)\end{aligned}$$

The effect of the one-unit increase in x_t at τ + 1 (which remains at τ + 2) on y_t at τ + 2 is the sum of the coefficients on $x_{\tau+1}$ and $x_{\tau+2}$:

$$\beta_1+\alpha_1\beta_1+\beta_2=\beta_1\left(1+\alpha_1\right)+\beta_2.\quad(4.1.11)$$

Similarly, forecasting the value of y_t at time t = τ + 3,

$$\begin{aligned}\hat{E}\left(y_{t=\tau+3|\tau}\right)&=\alpha_0+\alpha_1\alpha_0+\alpha_1^2 y_{\tau+1}+\beta_1 x_{\tau+3}+\left(\alpha_1\beta_1+\beta_2\right)x_{\tau+2}+\alpha_1\beta_2 x_{\tau+1}\\&=\alpha_0+\alpha_1\alpha_0+\alpha_1^2\alpha_0+\alpha_1^3 y_\tau+\beta_1 x_{\tau+3}+\left(\alpha_1\beta_1+\beta_2\right)x_{\tau+2}\\&\quad+\left(\alpha_1^2\beta_1+\alpha_1\beta_2\right)x_{\tau+1}+\alpha_1^2\beta_2 x_\tau.\quad(4.1.12)\end{aligned}$$

The effect of the one-unit increase in x_t at τ + 1 on y_t at τ + 3 is the sum of the coefficients on $x_{\tau+1}$, $x_{\tau+2}$, and $x_{\tau+3}$:

$$\beta_1\left(1+\alpha_1+\alpha_1^2\right)+\beta_2\left(1+\alpha_1\right).\quad(4.1.13)$$

We can continue forecasting as long as we like, and we will find that the effect of the one-unit increase in x_t at τ + 1 on y_t at τ + m is

$$\beta_1\left(1+\alpha_1+\alpha_1^2+\cdots+\alpha_1^{t+m}\right)+\beta_2\left(1+\alpha_1+\alpha_1^2+\cdots+\alpha_1^{t+m-1}\right).\quad(4.1.14)$$

If we assume that $|\alpha_1|< 1$, an assumption we will discuss in Section 4.6, then the effect of the one-unit increase in x_t at τ + 1 on y_t increases at a diminishing rate as t increases. If m is a very large number of time points, Equation 4.1.14 converges to

$$\frac{\beta_1}{1-\alpha_1}+\frac{\beta_2}{1-\alpha_1}=\frac{\beta_1+\beta_2}{1-\alpha_1}.$$

This is the long-run effect of a one-unit increase in x_t, as expressed in Equation 4.1.7.

Let us now return to the example of the U.S. public responsiveness model from Chapter 3. We had found that we had a problem with autoregressive errors, which we addressed by estimating a static model with AR(1) errors. As an alternative, we can run an ADL model. For this model, we will not be able to make the assumption of strict exogeneity. A lagged dependent variable is by definition not strictly exogenous. Strict exogeneity here requires that

$$E\left(\varepsilon_t \mid y_{t-1+h}\right)=0, \forall h. \qquad (4.1.15)$$

But clearly, εt is not independent from yt: $E(\varepsilon t \mid yt-1+h) \neq 0$, $h = 1$. The best we can do with a lagged dependent variable is the following:

$$E\left(\varepsilon_t \mid y_{t-1+h}\right)=0,\ h \leq 0. \qquad (4.1.16)$$

This is called sequential exogeneity. This condition will be met if the values of εt are independent from current and past values of the exogenous variables and past values of the lagged dependent variable. Sometimes this condition will be expressed by saying that the variables in the model are *predetermined.* Sequential exogeneity meets the criteria for contemporaneous exogeneity but not strict exogeneity, and so an ADL model requires a sufficient number of time points to allow us to rely on the assumptions of ordinary least squares (OLS) estimator consistency (see Section 3.2 in Chapter 3).

Let us estimate an ADL(1,1) data model, as described in Equation 4.1.5 (Table 4.1). We include the trend we previously determined we needed in Chapter 3.[1]

Table 4.1 U.S. Public Responsiveness Model—ADL(1,1) Model

Preference	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
L1. Preference	0.57	0.14	4.15	<0.001
Spending	–0.151	0.067	–2.27	0.032
L1. Spending	–0.035	0.082	–0.42	0.676
Trend[2]	0.068	0.018	3.76	0.001
Constant	31.05	8.047	3.86	0.001

NOTE: $R^2 = 0.84$, $T = 32$; T = number of time points, ADL = autoregressive distributed lag, L1 = first lag.

Let us now test the residuals from this model with the Q test. The Q statistic is 11.97, with a P value of 0.61. We cannot reject the null hypothesis of a white noise process. It appears as though we have successfully taken care of the serial correlation found in the residuals of the static model using the ADL(1,1) model.

Based on these results, we can calculate the estimated short-run and long-run effects of social program spending and the equilibrium of relative public preference. The estimated immediate short-run effect of a one-unit change in social program spending is statistically significant (at the 0.05 significance level): -0.151. A billion-dollar increase in spending immediately (within the year) reduces the relative preference for a spending increase by 0.15 percentage points. The estimated long-run effect of a permanent one-unit change is

$$\frac{\hat{\beta}_1 + \hat{\beta}_2}{1 - \hat{\alpha}_1} = \frac{-0.151 - 0.035}{1 - 0.57} = -0.43.$$

A billion-dollar increase in spending has a long-run effect of decreasing the relative preference for a spending increase by 0.43 percentage points.

We can test the statistical significance of the estimated long-run effect using a Wald test (Wooldridge, 2006). We use a chi-squared statistic with 1 degree of freedom. It is 11.88, and the corresponding *P* value is 0.001. It appears that the long-run effect is statistically significant at the usual 0.05 significance level. The estimated equilibrium of the relative public spending preference is

$$\frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1} = \frac{31.05}{1 - 0.57} = 32.93.$$

In the static model with AR(1) errors from Chapter 3, the short-run effect was the long-run effect. A billion-dollar increase in spending reduced the relative preference for spending by 0.17 percentage points (Table 3.12). Substantively, this tells a different story from that of the ADL(1,1) model. In Section 4.3, we shall discuss in detail the criteria we should use in determining whether to use a dynamic model with a lagged dependent variable or a static model with autoregressive errors.

Next, let us also look at the example of economic popularity in West Germany, using the vote intention data introduced in Chapter 1. Consider the monthly vote intention and economic data for the Kohl government from October 1982 till September 1998; the economic variables are gross domestic product (GDP), unemployment, and inflation.

We estimate the ADL(1,1) model, regressing vote intention (percentage of respondents from a monthly poll indicating that they would vote for the Kohl government in an election) on a lag of vote intention, the three economic variables, lags of the three economic variables, and the trend variable (Table 4.2). It is not necessary to add a lag of *t* as it would just be a trend variable itself.

$$\begin{aligned} \text{vote}_t = \alpha_0 + \alpha_1 \text{vote}_{t-1} + \beta_1 \text{GDP}_t + \beta_2 \text{inf}_t + \beta_3 \text{unemp}_t \\ + \beta_4 \text{GDP}_{t-1} + \beta_5 \text{inf}_{t-1} + \beta_6 \text{unemp}_{t-1} + \beta_7 t + \varepsilon_t. \quad (4.1.17) \end{aligned}$$

Table 4.2 Economic Popularity in West Germany—ADL(1,1) Model

	Coefficient	*Standard Error*	*t Statistic*	*P Value*
L1. Vote	0.38	0.14	2.75	0.009
GDP	–0.53	0.44	–1.2	0.236
Unemployment	–5.08	5.56	–0.91	0.366
Inflation	0.32	1.27	0.25	0.802
L1. GDP	–0.010	0.43	–0.02	0.981
L1. Unemployment	5.92	5.08	1.17	0.25
L1. Inflation	–2.30	1.31	–1.75	0.087
Trend	–0.23	0.083	–2.77	0.008
Constant	35.22	18.38	1.92	0.062

NOTE: $R^2 = 0.54$, $T = 51$; T = number of time points, ADL = autoregressive distributed lag, GDP = gross domestic product, L1 = first lag.

We now test the residuals for serial correlation using the *Q* test for white noise. The *Q* statistic is 14.52. The corresponding *P* value is 0.91. We fail to reject the null hypothesis of white noise.

None of the coefficients on the economic variables on their own appear to be significant, but the long-run effect of the one-unit increase in inflation is

$$\frac{\beta_1 + \beta_2}{1 - \alpha_1} = \frac{0.320 - 2.30}{1 - 0.38} = -3.19.$$

The chi-squared statistic with 1 degree of freedom is 7.41, and the corresponding *P* value is 0.007. It appears as though inflation had a significant and negative effect on vote intention for the Kohl government.

To get an understanding of the relationship between dynamic and static models, let us examine what would happen if we have a time series with an ADL(1,1) data-generating process,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t \qquad (4.1.18)$$

with $\varepsilon_t \sim NID(0, \sigma^2_\varepsilon)$, but we estimated it with a static model:

$$y_t = \tilde{\alpha}_0 + \tilde{\beta}_1 x_t + \tilde{\mu}_t. \tag{4.1.19}$$

The data-generating process for y_t (Equation 4.1.18) could be reformulated as

$$y_t = \alpha_0 + \beta_1 x_t + \mu_t,$$

$$\mu_t = \alpha_1 y_{t-1} + \beta_2 x_{t-1} + \varepsilon_t \tag{4.1.20}$$

with $\varepsilon_t \sim \text{NID}(0,\sigma^2_\varepsilon)$. Therefore, our data model is

$$y_t = \tilde{\alpha}_0 + \tilde{\beta}_1 x_t + \tilde{\mu}_t,$$

$$\mu_t = \alpha_1 y_{t-1} + \beta_2 x_{t-1} + \varepsilon_t. \tag{4.1.21}$$

Note that we are suggesting that a misspecified model has resulted in a particular data-generating process for the residuals. From Equation 4.1.20,

$$y_{t-1} = \alpha_0 + \beta_1 x_{t-1} + \mu_{t-1}. \tag{4.1.22}$$

Substituting Equation 4.1.22 for y_{t-1} in the expression for the residuals of Equation 4.1.21,

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

$$\mu_t = \alpha_1 \alpha_0 + \left(\alpha_1 \beta_1 + \beta_2 \right) x_{t-1} + \alpha_1 \mu_{t-1} + \varepsilon_t. \tag{4.1.23}$$

This data-generating process for the residuals is first-order autoregressive with autoregressive parameter α_1. What this tells us is that if we estimate a static model and find that the residuals are autocorrelated, one potential reason for this is that we should have included a lag of the dependent variable in the model.

Also, note the consequence of excluding the lag of the independent variable for the assumption of exogeneity. We do not meet the conditions of sequential exogeneity. The first lag of the independent variable included in Equation 4.1.21, $x_t = x_{t-1}$, is clearly correlated with the error term, which is itself a function of x_{t-1} by Equation 4.1.23.

To further understand the relationship between the static model with AR(1) serially correlated errors from Chapter 3 and the ADL(1,1) model, consider the following. The static model with AR(1) errors is

$$y_t = \beta_0 + \beta_1 x_t + \mu_t. \tag{4.1.24}$$

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t. \tag{4.1.25}$$

From Equation 4.1.24,

$$\mu_t = y_t - \beta_0 - \beta_1 x_t.$$

And therefore,

$$\mu_{t-1} = y_{t-1} - \beta_0 - \beta_1 x_{t-1}. \qquad (4.1.26)$$

Substituting Equation 4.1.25 into Equation 4.1.24,

$$y_t = \beta_0 + \beta_1 x_t + \rho\mu_{t-1} + \varepsilon_t. \qquad (4.1.27)$$

Substituting Equation 4.1.26 into Equation 4.1.27,

$$y_t = \beta_0 + \beta_1 x_t + \rho\left(y_{t-1} - \beta_0 - \beta_1 x_{t-1}\right) + \varepsilon_t.$$

And rearranging terms,

$$y_t = \left(1-\rho\right)\beta_0 + \rho y_{t-1} + \beta_1 x_t - \rho\beta_1 x_{t-1} + \varepsilon_t. \qquad (4.1.28)$$

Define[2] $(1-\rho)\beta_0 \equiv \alpha_0$; $\rho \equiv \alpha_1$, and $-\rho\beta_1 \equiv \beta_2$, and our static model with AR(1) errors has now become an ADL(1,1) model:

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

with the restriction that $\beta_2 = -\alpha_1\beta_1$. We might ask, “What does this restriction mean?” Recall that the long-run effect for an ADL(1,1) model is

$$\frac{\beta_1 + \beta_2}{1-\alpha_1}.$$

Substituting our restriction $\beta_2 = -\alpha_1\beta_1$ into the calculation for the long-run effect,

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

The long-run effect is exactly the same as the short-run effect. This restricted ADL is equivalent to the static model with autoregressive errors, so it is not surprising that the short-run effect is the long-run effect.

We will now consider an alternative restriction to the ADL and in doing so introduce the LDV model.

4.2 Lagged Dependent Variable Models (or Partial Adjustment Models)

If we are willing to assume that $\beta_2 = 0$ in Equation 4.1.18, we have what is called an LDV model, also known as a *partial adjustment* model (Hendry, 2003):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \varepsilon_t. \qquad (4.2.1)$$

An LDV model has no lagged independent variables. Like the ADL model, the LDV model is a useful approach for dealing with serially correlated errors, but they both assume fundamentally different time series processes from the static model with autoregressive errors (unless the ADL is appropriately restricted). We could instead assume that $\beta_1 = 0$ in Equation 4.1.18. This model assumes that the independent variable takes one time period to have an effect. This is known as a *dead start* model (Hendry, 2003).

Returning to our U.S. public responsiveness model, we might note that the lag of the independent variable "social policy spending" is not statistically significant. Based on this information, we might try an LDV model—the difference is the exclusion of lags of the independent variables. Alternatively, we may have hypothesized from the beginning, based on theoretical expectations of the data-generating process, that the LDV model was the most appropriate. In either case, we would estimate Equation 4.2.1 (Table 4.3).

Table 4.3 U.S. Public Responsiveness Model, Social Programs—LDV Model

Preference	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
L1.Preference	0.61	0.11	5.77	<0.001
Spending	–0.17	0.043	–4.04	<0.001
Counter^2	0.064	0.015	4.32	<0.001
Constant	28.85	6.05	4.77	<0.001

NOTE: $R^2 = 0.84$, $T = 32$; T = number of time points, LDV = lagged dependent variable, L1 = first lag.

We can test whether or not the estimated errors (the residuals) from our model follow a white noise process. The *Q* statistic is 12.55 and is chi-squared distributed with 14 degrees of freedom. The corresponding *P* value is 0.562. We cannot reject the null hypothesis of white noise for the errors. It appears that we do not have serially correlated errors. Recall that we did reject the null hypothesis of white noise for the errors from the estimated static model. We can also test the errors for normality. The *P* value for the null hypothesis of no skewness is .11, and the *P* value for the null hypothesis of no kurtosis (relative to the normal) is 0.67. There is no evidence to suggest that we can reject the null hypothesis of skewness or kurtosis different from a normal distribution. Therefore, we cannot reject the null hypothesis of normally distributed errors.

We can estimate the short-run and long-run effects for the LDV model just as we did with the ADL(1,1) model, except that $ß2$ is now equal to 0. The short-run effect of increasing social policy spending by $1 billion is to reduce the relative preference for a spending increase by 0.17 of a percentage point. This is statistically significant at the 0.05 significance level. The long-run effect of increasing social policy spending is

$$\frac{\beta_1}{1-\alpha_1} = \frac{-0.17}{1-0.61} = -0.44.$$

The long-run effect of increasing social policy spending by $1 billion is to reduce the relative preference for a spending increase by 0.44 of a percentage point. The chi-squared test statistic is 10.50, and the corresponding *P* value is 0.001. This effect is significant at the 0.05 significance level.

4.3 Advice on Model Selection

We have now discussed enough modeling options to raise the issue of how to choose one model over another. We have already discussed the importance of including nonstationary elements, such as trends, periodicity, and structural breaks, in our models. But other issues will have come to your mind as you were introduced to the models of this and the previous chapter.

When we were discussing the ADL(1,1) and LDV models, you may have asked yourself, “Why did we settle on a model with one lag of the dependent variable instead of two or three?” The answer to this question can be found in the structure of the autocorrelations of the errors. We will talk about this further in Chapter 5 when we discuss model selection and testing. For the time being, it is worth noting that a time series model is considered a good fit to the data if the resulting residuals are a white noise process (Li, 2004). Therefore, the lag structure of a model is selected to produce such residuals.

As we will see in Chapter 5, there are often multiple possible lag structures that produce this result, and so additional requirements are applied in model selection. In Chapter 5, we will discuss the requirements used by the Box-Jenkins approach to model selection. However, another approach has emerged more recently within the field of time series analysis. This is the general-to-specific approach (Hendry, 2003).

We have seen that the LDV and the static (and FDL) models are restricted forms of the ADL model. In Chapter 6, we will see that the standard form of a model known as the error correction model is a transformation of the ADL model. Restrictions can also be placed on the error correction model to specify other models.

From the perspective of the general-to-specific approach, potential models are viewed as restricted versions of more general models. As the name of the approach suggests, model selection proceeds by estimating a more general model, maybe more general than is theoretically suggested, and testing whether restrictions can be placed on that model. This is the procedure we followed when testing whether an LDV, rather than an ADL(1,1), model suffices to model U.S. public responsiveness.

Continuing with the general-to-specific principle, we saw that the static model with serially correlated errors was a restricted form of the ADL(1,1) model. A reasonable question at this point is “How does one choose whether to address the problem of serially correlated errors with a more general LDV or ADL model or a more specific static model with autoregressive errors or serial correlation robust standard errors?” The decision to use one or the other depends on a number of criteria.

1. If we believe that the serial correlation is a nuisance produced by some problem, such as measurement error, we should use a static model and either include autoregressive errors or estimate serial correlation robust standard errors. The problem of measurement error resulting in serial correlation can occur when the measurement error in one period depends on the measurement error in another, such as when measurement errors accumulate over time.
2. If we believe that the serial correlation is a product of a very real dynamic in the data-generating process (e.g., vote intention today is a function of vote intention yesterday), we should model this dynamic with an LDV model.

The second criterion is related to the issue of exogeneity we discussed in Chapter 3. Recall the following example data model of vote intention:

$$\text{vote}_t = \beta_0 + \beta_1 \text{econ}_t + \varepsilon_t. \quad (4.3.1)$$

It is highly probable that the data-generating process for vote intention includes past intention, vote_{t-1}. If this is the case, then ε_t in our data model will be a function of (contain) vote_{t-1} (i.e., vote_{t-1} is omitted). If econ_t are subjective evaluations of the economy, it is unlikely that vote_{t-1} and econ_t are independent. An individual's past vote intention is likely to be predictive of current economic evaluations (Evans & Andersen, 2006), and so it is not likely that we have contemporaneous exogeniety: $e(\varepsilon_t \mid \text{econ}_t) \neq 0$. This may be resolved by explicitly including vote_{t-1} in our data model as follows:

$$\text{vote}_t = \beta_0 + \beta_1 \text{econ}_t + \beta_2 \text{vote}_{t-1} + \varepsilon_t. \quad (4.3.2)$$

In this data ε_t model, the values are not a function of vote_{t-1}. Including a lagged dependent variable can correct violations of endogeneity, while running a static model with AR(1) errors does not.

Continuing our discussion of when to choose a dynamic model with a lagged dependent variable or a static model with a correction for serial correlation, consider our LDV model of U.S. public policy responsiveness (Table 4.3). What if this LDV model exhibited serially correlated errors? This raises an important issue. Serial correlation when including a lagged dependent variable has serious consequences. In the context of our example, the LDV process with serially correlated errors looks as follows:

$$R_t = \alpha_0 + \alpha_1 R_{t-1} + \beta_1 P_t + \beta_2 W_t + \mu_t, \quad (4.3.3)$$

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t. \quad (4.3.4)$$

From Equation 4.3.3,

$$R_{t-1} = \alpha_0 + \alpha_1 R_{t-2} + \beta_1 P_{t-1} + \beta_2 W_{t-1} + \mu_{t-1}. \quad (4.3.5)$$

Note: R_{t-1} and μ_t are both functions of μ_{t-1}. Therefore, $E(\mu_t \mid R_{t-1}) \neq 0$, and we have not met the assumptions necessary for an unbiased OLS estimation. Note that this isn't just a violation of strict exogeneity. It is a violation of contemporaneous exogeneity. Therefore, we do not even have the conditions for an asymptotically unbiased OLS estimation.

The presence of serial correlation in such an LDV or ADL model is seen as evidence of model misspecification, which can be corrected by including additional lags of the independent or dependent variables. For example, consider a single realization of the following data-generating process:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_1 x_t + \varepsilon_t, \quad (4.3.6)$$

with $\alpha_0 = 10$, $\alpha_1 = 0.4$, $\alpha_2 = 0.3$, $ß_1 = 0.9$, and $\varepsilon_t \sim N(0, 1)$. In addition to including a lag of y_t as an independent variable, it also includes the second lag. Let us use a single realization, with 38 time points, to estimate a model without the second lag—an LDV model (Table 4.4).

Table 4.4 LDV Model With AR(2) Data

y_t	Coefficient	Standard Error	t Statistic	P Value
y_{t-1}	0.50	0.13	3.81	0.001
x_t	0.79	0.17	4.50	<0.001
Constant	17.57	4.75	3.70	0.001

NOTE: $R^2 = 0.45$, $T = 38$; T = number of time points, AR = autoregressive, LDV = lagged dependent variable.

If we test the residuals from this model, we will find that they are not a white noise process. The *Q* statistic is 33.29, and the corresponding *P* value is 0.010. The rejection of the null hypothesis of a white noise process is because the errors contain serial correlation due to the omission of the second lag of y_t. If we reestimate this model, including the second lag of y_t, we would find that we cannot reject the null hypothesis that the residuals are a white noise process. The *Q* statistic is 19.48, and the corresponding *P* value is 0.245. It is only by including the second lag of y_t that we meet one of the necessary conditions for an asymptotically unbiased OLS estimation.

The requirement for consistent OLS estimation with a lagged dependent variable is sometimes stated as *dynamic* completeness (Wooldridge, 1991, 2006, pp. 400–402). This simply means that there are enough lags of the dependent and independent variables in the model to ensure that $E(\mu_t \mid y_{t-1+h}, x_{t+h}) = 0$, for all $h \leq 0$. Recall that this is called *sequential exogeneity.* Dynamic completeness also implies that there is no serial correlation.

Some simulation work (Keele & Kelly, 2006) has shown that under some circumstances the bias from serial correlation with a lagged dependent variable is relatively minor and typically less than the bias caused by estimating a model without a lagged dependent variable when the data-generating process for the observed time series contains a lagged dependent variable. While this certainly suggests that the analyst should not shy away from models with lagged dependent variables, it is not a justification for ignoring dynamic completeness. It is relatively easy to detect residual serial correlation using the techniques discussed in Chapter 3. Such tests should always be conducted, and if serial correlation is detected, the model should be respecified with additional lags of the independent and/or the dependent variable to remove the serial correlation.[3]

An additional consideration when choosing between a dynamic and a static model is the following:

3. The dynamics for the effects of the independent variables is different between the two types of models. An LDV model has different short-run and long-run effects. For the static model with autoregressive errors (or not), the long-run effect is the short-run effect.

However, one can always add dynamics to the static model, allowing for different short-run and long-run effects, by including lags of the independent variables, as we did in the FDL model discussed in Chapter 3. Autoregressive errors can be added to an FDL model without any problem.

A fourth consideration is as follows:

4. A lagged dependent variable can *never* be strictly exogenous, as was demonstrated in Section 4.1. As strict exogeneity is a requirement for unbiasedness when we do not have enough time points to rely on the assumptions for asymptotic unbiasedness, an LDV model requires an adequate number of time points.

The problem of including a lagged dependent variable in a data model with only a few time points of data for its estimation was first discussed by Hurwicz (1950), and the resulting bias is sometimes called Hurwicz bias.

As in any modelling exercise, specification of the model must take into account these important primary considerations:

1. Is the model motivated/justified by the theoretical DGP?
2. Does it test the hypotheses in which we are interested?
    ° Will it tell us what we want to know?
3. Are the assumptions of the estimation technique met?

Chapter 3 provides an overview of the assumptions we have made in the estimation of the models we have considered so far. These assumptions include homoskedasticity. We now turn our attention to another type of time series model. This is a model that places the focus on the issue of heteroskedasticity.

4.4 Autoregressive Conditional Heteroskedasticity (ARCH) Models

For each model we have discussed, we have made the assumption of homoskedastic errors. Consider our LDV model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \varepsilon_t. \qquad (4.4.1)$$

Recall that ε_t is a white noise process. We assume that ε_t has a zero mean and a constant variance, σ^2. We might want to consider the possibility that the variance changes over time. As noted in Chapter 3, this has implications for the estimation of the standard errors of our model parameters and for subsequent inferential tests. The dynamics within the variance of the errors may also be of substantive interest—for example, a reduction in the variance in support for a U.S. president in response to critical events, such as economic crises and foreign conflict (Gronke & Brehm, 2002). Therefore, we may be interested in modelling the dynamics in the variance of ε_t. To do this, we must make the distinction between the unconditional variance of ε_t and the conditional variance of ε_t.

Considering again our LDV model, we assume that the unconditional variance of ε_t is constant and without serial correlation:

$$E\left(\varepsilon_t \varepsilon_{t-s}\right)=\begin{cases}\sigma^2 \text{for } s=0\\ 0 \text{ otherwise}\end{cases}. \qquad (4.4.2)$$

Note that the expected value of the squared errors is the variance of ε_t. Continuing to assume constant unconditional variance, we allow that the variance conditional on past values of the errors will change over time. Until now, we have assumed that the conditional variance of the errors is constant. Now we model the errors as a process with a conditional variance, $E(\varepsilon_t \varepsilon_t | \varepsilon_{t-s})$, that is a function of past values of the variance. A common way to do this is to model the squared errors as an autoregressive AR(m) process:

$$\varepsilon_t^2 = \zeta + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \varepsilon_{t-2}^2 + \cdots + \phi_m \varepsilon_{t-m}^2 + \omega_t, \qquad (4.4.3)$$

where ω_t is also a white noise process with a zero mean and constant unconditional variance. We assume that this process is covariance stationary and that $\varphi_i \geq 0$, $i = 1,\ldots, m$. As ε_t^2 cannot be negative, we also assume that $\zeta > 0$ and $\omega_t \geq -\zeta$ for all t. This is called an autoregressive conditional heteroskedastic process:

$$\varepsilon_t^2 \sim \text{ARCH}\left(m\right), \qquad (4.4.4)$$

with m denoting the order of the ARCH process (Engle, 1982). Time series models that include such a process in the modelling of the errors are called ARCH models. The logic of this model of the squared errors is that periods of high (low) variance tend to group together. If the variance is higher (lower) than average at a particular time point, it is also likely to be higher (lower) than average in the next time point. Such ARCH errors can be included in static, autoregressive, or ARMA (to be discussed in the next chapter) models.

An equivalent, and common, representation of an ARCH process is as follows:

$$\varepsilon_t = \sqrt{h_t}\,\upsilon_t, \qquad (4.4.5)$$

where υ_t is a white noise process with a zero mean and unit variance and h_t is a function of past ε_t^2 (Hamilton, 1994). Specifically,

$$h_t = \zeta + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \varepsilon_{t-2}^2 + \cdots + \phi_m \varepsilon_{t-m}^2.$$

$$E\left(\upsilon_t\right)=0$$

$$E\left(\upsilon_t \upsilon_{t-s}\right)=\begin{cases}1 \text{ for } s=0\\ 0 \text{ otherwise}\end{cases}. \qquad (4.4.6)$$

We can see that in this representation, the unconditional expected value of ε_t is 0, as assumed in Equation 4.4.1:

$$E\left(\varepsilon_t\right)=E\left(\sqrt{h_t}\upsilon_t\right)=E\left(\sqrt{h_t}\right)E\left(\upsilon_t\right)=0. \qquad (4.4.7)$$

Note that the second equality utilizes the assumption that v*t* is a white noise process. The conditional expected value is also equal to 0 (Enders, 2004).

$$E\left(\varepsilon_t \mid \varepsilon_{t-1},\varepsilon_{t-2},\cdots,\varepsilon_{t-m}\right)=0. \qquad (4.4.8)$$

The unconditional variance of ε*t* is a constant, which we have defined as σ^2 in Equation 4.4.2. In particular,

$$\begin{aligned} E\left(\varepsilon_t^2\right) &= E\left(h_t\upsilon_t^2\right)=E\left(h_t\right), \\ &= E\left(\zeta+\phi_1\varepsilon_{t-1}^2+\phi_2\varepsilon_{t-2}^2+\cdots+\phi_m\varepsilon_{t-m}^2\right), \\ &= \frac{\zeta}{1-\phi_1-\phi_2-\cdots-\phi_m}. \qquad (4.4.9)\end{aligned}$$

Finally, the conditional variance of ε*t*, given past ε*t*, is a function of time. Specifically,

$$\begin{aligned} E\left(\varepsilon_t^2 \mid \varepsilon_{t-1},\ldots,\varepsilon_{t-m}\right) &= E\left(h_t \mid \varepsilon_{t-1},\ldots,\varepsilon_{t-m}\right)E\left(\upsilon_t^2\right), \\ &= h_t=\zeta+\phi_1\varepsilon_{t-1}^2+\phi_2\varepsilon_{t-2}^2+\cdots+\phi_m\varepsilon_{t-m}^2. \qquad (4.4.10)\end{aligned}$$

Note that the conditional variance is an autoregressive process but the errors themselves (ε*t*) are not.

We return to our example of economic popularity in West Germany, but now for a longer time period, 1977–1998, and in place of vote intention data, we use government approval data based on the following survey question: “Sind Sie mit dem was die jetzige CDU/CSU/FDP-Regierung in Bonn bisher geleistet hat eher zufrieden oder eher unzufrieden?” (Are you rather satisfied [or happy] or rather dissatisfied [or unhappy] with the performance of the current CDU/CSU/FDP government in Bonn thus far?)

During this period, we expect periodicity from one electoral cycle to the next, separate trends for the period in which the SPD (with the FDP) controlled government and for the period in which the CDU (with the FDP) controlled government, and heteroskedasticity. The latter is expected due to changes in the party system prompted by the reunification of East and West Germany. During this time, partisan identification transitioned from being quite strong for many West Germans to quite weak (Arzheimer, 2006; Pickup, 2010; Zelle, 1998).

We begin by partialling out the periodicity and trending in the government approval data. We regress government opposition on two periodicity terms and a separate trend for each of the governments.[4]

The results are presented in Table 4.5. Estimated residuals from this regression are the approval data with periodicity and trending partialled out. Using this data, we estimate an ADL(1,1) economic popularity model.

The independent variables included are GDP growth, unemployment, and inflation. The results from this estimation are presented in Table 4.6.

Table 4.5 West Germany Government Approval 1977–1998

Approval	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
SPD trend	0.035	0.031	1.13	0.258
CDU trend	–0.13	0.0095	–13.83	<0.001
Cycle 1	5.83	0.71	8.26	<0.001
Cycle 2	0.24	0.71	0.34	0.736
Constant	61.047	1.04	58.98	<0.001

NOTE: $R^2 = 0.59$, $T = 243$; T = number of time points.

Table 4.6 West Germany Government Approval 1977–1998, ADL(1,1)

Approval	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
L1. Approval	0.78	0.041	18.80	<0.001
Growth	0.41	0.23	1.75	0.081
Unemployment	0.23	1.85	0.13	0.900
Inflation	0.32	0.87	0.37	0.715
L1. Growth	–0.27	0.24	–1.13	0.259
L1. Unemployment	–0.77	1.88	–0.41	0.684
L1. Inflation	–0.79	0.85	–0.93	0.352
Constant	4.27	1.54	2.77	0.006

NOTE: $R^2 = 0.76$, $T = 242$; ADL = autoregressive distributed lag, T = number of time points, L1 = first lag.

None of the coefficients on the economic variables are statistically significant on their own. We can test the joint significance of each economic variable with its lag using an F statistic. For GDP growth, the $F(1, 234)$ statistic is 2.12, and the corresponding P value is 0.147. For inflation the $F(1, 234)$ statistic is 6.33, and the corresponding P value is 0.013. For unemployment, the $F(1, 234)$ statistic is 8.38, and the corresponding P value is 0.004. We can also calculate the long-run effects for inflation and unemployment and test their significance. For inflation, the long-run effect is (0.32-0.79)/(1-0.78) = -2.14. The chi-squared(1) statistic is 6.81 with a P value of 0.010. For unemployment, the long-run effect is (0.23-0.77)/(1-0.78) = -2.45. The chi-squared(1) statistic is 10.95 with a P value of 0.001. Inflation and unemployment have statistically significant long-run effects.[5]

Before settling on these results, we would want to test the estimated errors of the model for, among other things, heteroskedasticity. We do this here with Engle's Lagrange multiplier test for the presence of autoregressive conditional heteroscedasticity (Engle, 1982). Specifically, we test the null of no ARCH effects against the alternative of ARCH(m) effects. This test is conducted on the residuals of the model we just ran: $\hat{\varepsilon}_t$. We run the following auxiliary regression: $\hat{\varepsilon}_t^2 = \gamma_0 + \gamma_1 \hat{\varepsilon}_{t-1}^2 + \cdots + \gamma_p \hat{\varepsilon}_{t-m}^2$, for up to m lags of the squared residuals and test the null hypothesis that $\Gamma_1 = \ldots = \Gamma_m = 0$ using a Lagrange multiplier test (Greene, 2003, pp. 484–492). We select m on the basis of the order of the ARCH effects for which we wish to test. In our current example, we will test for an ARCH(1) process.

This test is applicable for both the alternatives of autoregressive and moving average processes within the variances. We will discuss the latter type of process in Chapter 5. For our example, we test for first-order ARCH effects. The resulting test statistic is 36.21 with a chi-squared distribution with 1 degree of freedom. The corresponding P value is <0.001. This means that we can reject the null hypothesis of no first-order ARCH effects. Based on this finding, we now estimate the West German economic approval model as an ADL(1,1) with an ARCH(1) term (Table 4.7). As with the static model with autoregressive errors, ARCH models are typically estimated using maximum likelihood.

Table 4.7 West Germany Government Approval 1977–1998, ADL(1,1) ARCH(1)

Approval	*Coefficient*	*Standard Error*	*z Statistic*	*P Value*
L1. Approval	0.84	0.044	19.17	<0.001
Growth	0.50	0.31	1.64	0.100
Unemployment	–0.11	3.62	–0.03	0.976
Inflation	0.036	0.93	0.04	0.969
L1. Growth	–0.41	0.29	–1.41	0.160
L1. Unemployment	–0.31	3.62	–0.09	0.931
L1. Inflation	–0.48	0.92	–0.52	0.602
Constant	3.65	1.44	2.53	0.011
ARCH				
L1.	0.21	0.061	3.37	0.001
Constant	11.29	0.70	16.04	<0.001

NOTE: Log likelihood = –658.07, T = 242; L1 = first lag, T = number of time points, ADL = autoregressive distributed lag, ARCH = autoregressive conditional heteroskedasticity.

First, we note that the ARCH(1) term is statistically significant. This suggests that the inclusion of the ARCH term is appropriate. As an alternative, we could have just estimated the ADL(1,1) and estimated heteroskedasticity robust standard errors. However, OLS is inefficient in the presence of heteroskedasticity, so there are efficiency gains (smaller standard errors) to be made by including the ARCH term. Furthermore, the statistical significance of the ARCH term may be of substantive interest.

As we have done previously, we can test the significance of the short-run effects using the coefficient on the contemporaneous values of each economic variable. In each case the P value is greater than 0.05 and we cannot reject the null hypothesis of no short-run effect. As for the estimated long-run effects for inflation and unemployment, these are as follows. For inflation, the long-run effect is (0.036-0.48)/(1-0.84) = -13.38. The chi-squared(1) statistic is 4.19 with a *P* value of 0.041. For unemployment, the long-run effect is (−0.11-0.32)/(1-0.84) = -15.31 The chi-squared(1) statistic is 5.58 with a *P* value of 0.018. Again, a change in inflation or unemployment has a statistically significant effect on approval.

We can also model the variance of ε_t as a function of exogenous variables. This is called multiplicative heteroskedasticity. For example, we can have an ARCH(1) process in which the variance is also a function of the variable x_t:

$$h_t = exp\left(\zeta_0 + \zeta_0 x_t\right) + \phi_1 \varepsilon_{t-1}^2.$$

When we discuss generalized autoregressive conditional heteroskedasticity (GARCH) models in Chapter 5, we will discuss how to test the adequacy of the specification of ARCH and GARCH models. We now turn to yet another type of time series model. This model assumes that the conditional variance is constant but allows for serial correlation in the errors of a form that differs from autoregressive errors.

4.5 The Moving Average Process and the Autocorrelated Error Model

A moving average data-generating process can be characterized as

$$y_t = \beta_0 + \varepsilon_t + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j}, \qquad (4.5.1)$$

where q determines the order of the moving average process, the φ_j values are coefficients with at least one $\varphi_j \neq 0$, and $\varepsilon_t \sim \text{NID}(0,\sigma^2_\varepsilon)$. Note that a moving average process differs from an autoregressive process in that the first includes lags of the errors while the second includes lags of the dependent variable. It also differs from a static process with autoregressive errors in that the moving average process includes lags of the errors in the equation for y_t, while the static model includes lags of the errors in the equation for the errors. As an example, a moving average process of order 1 includes one lag of the error:

$$y_t = \beta_0 + \varepsilon_t + \phi_1 \varepsilon_{t-1}. \qquad (4.5.2)$$

Such a process is denoted as MA(1). This is a weakly dependent sequence as y_t values one period apart are correlated but y_t values two periods apart are not. This can be demonstrated as follows: $y_t = \varepsilon_t + \varphi_1\, \varepsilon_{t-1}$ and $= y_{t-1} = \varepsilon_{t-1} + \varphi_1\, \varepsilon_{t-2}$ have ε_{t-1} in common and are therefore correlated, but $y_t = \varepsilon_t + \varphi_1\, \varepsilon_{t-1}$ and $y_{t-2} = \varepsilon_{t-2} + \varphi_1\, \varepsilon_{t-3}$ have nothing in common and are therefore not correlated.

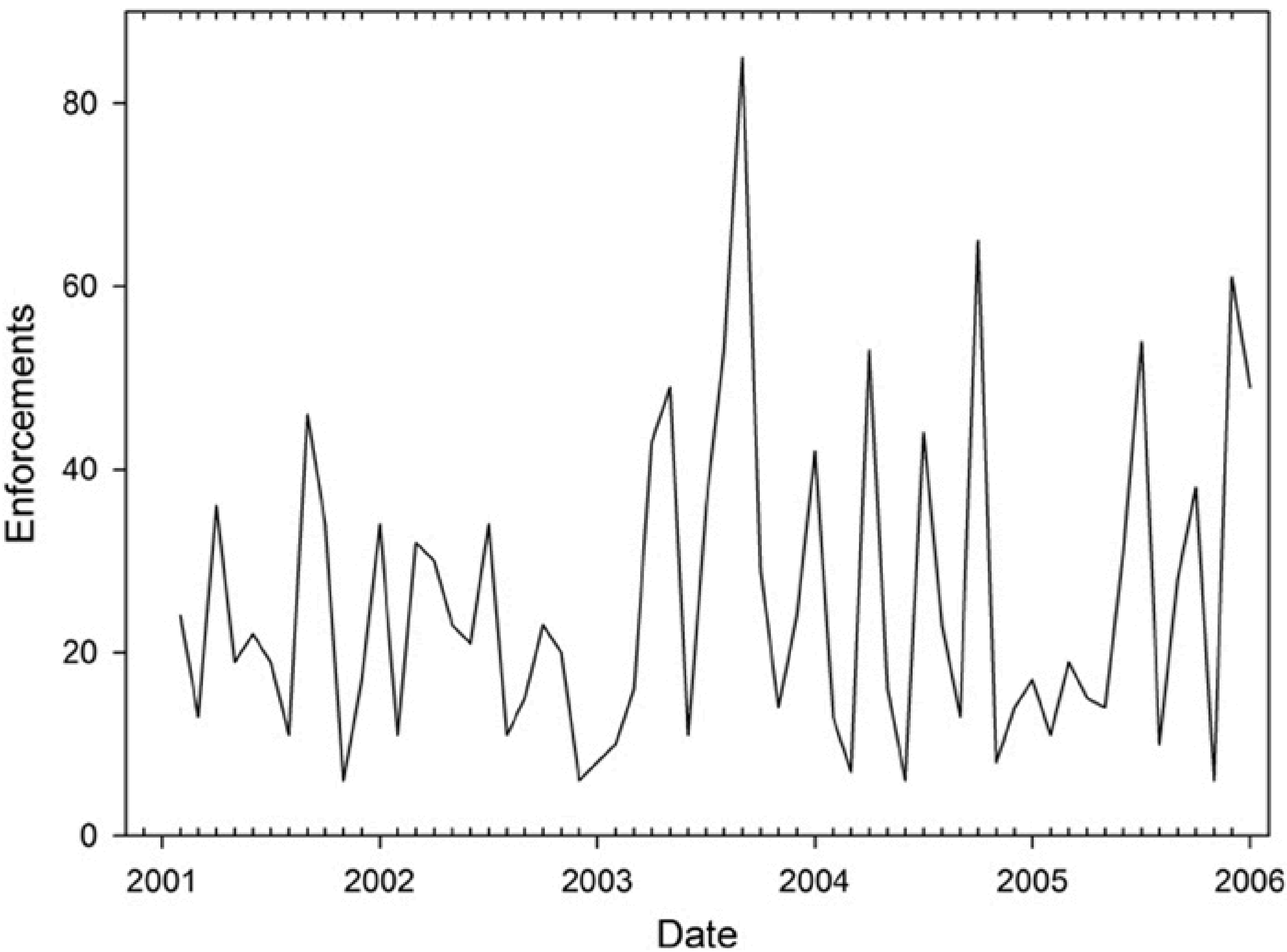
A moving average process can be a data model in itself. It is called the autocorrelated error model or just an MA model. Exogenous regressors (x_k) can also be included in such data models:

$$y_t = \beta_0 + \sum_{k=1}^{k} \beta_k x_k + \varepsilon_t + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j}. \qquad (4.5.3)$$

Like ARCH models and static models with autoregressive errors, MA models are estimated using maximum likelihood. If we were to estimate Equation 4.5.3 without the moving average component, we would have serial correlation in the errors. A moving average process is another source of violation of the assumption of no serial correlation. Therefore, an MA model can be useful to control for the moving average component of a data-generating process. As an example of an MA model, consider the following data on U.S. Environmental

Protection Agency (EPA) enforcements.[6] To begin, we look at a plot of the number of enforcements over time (Figure 4.1).

Figure 4.1 U.S. Environmental Protection Agency Enforcements, 2001 to 2005



The data are monthly, and the first time point (t = 1) is January 2001. The last data point is December 2005. These data were used by Provost, Gerber, and Pickup (2009) to test the impact of the various tools President Bush had at his disposal to influence EPA environmental regulation. They considered the impact of budgetary changes, the appointment and resignation of various EPA administrators, and the introduction of two major rule changes. The first major change was to the New Source Review (NSR) rules, and the second was the issuance of the Clean Air Interstate Rule (CAIR).[7]

With “enforcements” as the dependent variable, we estimate the following data model with exogenous regressors (independent variables) and a single second-order moving average term:

$$y_t = \beta_0 + \sum_{k=1}^{K} \beta_k x_k + \varepsilon_t + \phi_j \varepsilon_{t-2}. \tag{4.5.4}$$

The K regressors included in this model are (a) increases in the budget and decreases in the budget—increases and decreases are measured in millions of U.S. dollars and separated, as they are expected to have different effects; (b) the appointments of Whitman, Holmstead, Leavitt, Johnson, and Sansonetti as EPA administrators; (c) Whitman's resignation; (d) the introduction of the NSR; and (e) the introduction of the CAIR. Note that we have not included a first-order moving average term. This is a perfectly acceptable modelling choice, but it is fair to ask at this point how it was decided to use a second-order MA model for these data. This topic is covered in Chapter 5, along with that of determining the number of lags of the dependent variable to include in a model. The results of the maximum likelihood estimation of the model are presented in Table 4.8.

Table 4.8 MA Model of U.S. Environmental Protection Agency Enforcements, 2001 to 2005

Enforcements	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
Budget increase	–8.84	27.15	–0.33	0.745
Budget decrease	86.45	29.19	2.96	0.003
Whitman	–30.64	6.06	–5.05	<0.001
Whitman quit	9.65	17.17	0.56	0.574
Holmstead	1.00	11.85	0.08	0.932
Leavitt	–22.29	5.53	–4.03	<0.001
Johnson	–15.40	25.09	–0.61	0.539
Sansonetti	–5.08	8.95	–0.57	0.571
NSR	–17.68	8.95	–1.97	0.048
CAIR	5.80	17.92	0.32	0.746
Constant	56.15	10.40	5.40	<0.001
L2. MA	–0.66	0.12	–5.45	<0.001

NOTE: Log likelihood = –235.22, T = 59; T = number of time points, L2 = second lag, CAIR = Clean Air Interstate Rule, MA = moving average, NSR = New Source Review.

Certain policy initiatives of the Bush administration did have a clear impact on EPA enforcements. A million-dollar decrease in the budget resulted, on average, in 86 less enforcements. Budget increases, on the other hand, do not seem to have had any effect. Changes to the NSR rules did have a reducing effect on enforcements, which arguably was the intended effect (Provost et al., 2009).

It is also clear that the coefficient for the second-order moving average process is statistically significant at the usual 0.05 significance level. Interestingly, the coefficient is negative. It is generally difficult to give a substantive interpretation of moving average processes. In this particular instance, it is probably best to just treat it as a nuisance that needs to be dealt with in order to get the correctly estimated values for the other parameters (and their standard errors) in the model.

We can now use the *Q* test to test the residuals from the EPA enforcement model for deviation from a white noise process. The *Q* statistic is 31.21. It is chi-squared distributed with 27 degrees of freedom, giving us a *P* value of 0.26. Based on this test, we cannot reject the null hypothesis of a white noise process at the 0.05 significance level. Therefore, there is no evidence that any serial correlation remains within the residuals. For illustration, we can also run the same model, but without the moving average component, and test the residuals for deviation from a white noise process. The results are presented in Table 4.9.

Table 4.9 Model of U.S. Environmental Protection Agency Enforcements, 2001 to 2005

Enforcements	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
Budget increase	22.53	51.66	0.44	0.663
Budget decrease	75.63	53.48	1.41	0.157
Whitman	–32.38	8.22	–3.94	<0.001
Whitman quit	–11.27	19.62	–0.57	0.566
Holmstead	2.75	15.82	0.17	0.862
Leavitt	–22.61	8.19	–2.76	0.006
Johnson	–18.13	25.75	–0.70	0.481
Sansonetti	–1.96	12.12	–0.16	0.871
NSR	–24.54	10.29	–2.39	0.017
CAIR	7.37	13.13	0.56	0.575
Constant	49.79	19.01	2.62	0.009

NOTE: Log likelihood = –242.71, T = 59; P = probability, T = number of time points, CAIR = Clean Air Interstate Rule, NSR = New Source Review.

The Q statistic is now 49.65, giving us a P value of 0.005. We reject the null hypothesis of a white noise process at the 0.05 significance level. This indicates that there is serial correlation in the residuals. Also, note that when we do not control for the moving average process, the effect of budget decreases is no longer statistically significant. As in this example, there is often no intuitive interpretation for a moving average process or model, but it is important to include it in order to control for serial correlation in the data model residuals.

Now let's consider the necessary conditions for stationarity for the time series processes we have discussed in this chapter.

4.6 Stability and Stationarity Conditions

In Chapter 2, we discussed the importance of meeting the assumption of covariance stationarity. This would be a good time to discuss the conditions that are necessary for a dynamic model to meet this assumption. We start with the stationarity conditions for an AR(1) process. We do this by revisiting the topic of stability, first discussed in Chapter 2. Stability is a necessary condition for stationarity.

Time series analysis is concerned with describing and modelling time series sequences $Y_t = \{y_1, y_2, \ldots, y_T\}$ with equations that contain both nonstochastic and stochastic components. The LDV model without independent variables is an example of such an equation. It contains ε_t (the stochastic component), one lag of y_t, and a constant:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \qquad (4.6.1)$$

This assumes an AR(1) data-generating process with a constant α_0 (and therefore a nonzero equilibrium). Statisticians often talk of a "solution" to the equation describing a data-generating process. The solution expresses the value of y_t as a function of the elements of the $\{x_t\}$ sequence (if there is any x_t in the model), ε_t, the data-generating process parameters, and t (time). The solution does not contain lags of the dependent variable, even though the equation for the data-generating process does. The solution may possibly contain some initial conditions for the sequence $\{y_t\}$. The initial condition is the value of y_t when the time series process began. It is sometimes denoted as y_0, but it may not necessarily correspond with the first data point in our data set; we may not observe the data-generating process when it first began.

The solution allows us to do something we cannot do with the equation itself. It allows us to talk in the abstract about what values we would expect y_t to take on average; that is, we can calculate the mean value of y_t over the long run. This is the unconditional expected value.

Earlier in this chapter, we learned that the estimated long-run equilibrium of the LDV model is $\alpha_0/(1-\alpha_1)$. This is the unconditional expected value. We will now see how this was derived from the solution to the LDV data-generating process. As the solution allows us to calculate the value we expect y_t to take (on average) over the long run, we can use it to determine whether a time series has a stable or an explosive time path. If the time series is stable, the unconditional expected value (long-term mean) will remain constant over the long run—not change over time. We cannot determine this with an equation containing a lag of y_t. Why not?

Take, for example, $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$, and say we know α_0 and α_1 (which we normally would not). We know the expected value of $\{\varepsilon_t\}$: $E(\varepsilon_t) = 0$; but to calculate the expected value of y_t, we need to know the expected value *of* y_{t-1}. But this is the same as knowing the expected value of y_t, which is what we are trying to determine; that is, we do not know the expected value of y_{t-1} any more than we know the expected value of y_t. We can plug $y_{t-1} = \alpha_0 + \alpha_1 y_{t-2} + \varepsilon_{t-1}$ into $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ and get

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

$$= \alpha_0 + \alpha_0\alpha_1 + \alpha_1\alpha_1 y_{t-2} + \varepsilon_t + \alpha_1\varepsilon_{t-1}.$$

But what is the expected value of y_{t-2}? Well, $y_{t-2} = \alpha_0 + \alpha_1 y_{t-3} + \varepsilon_{t-2}$, and we can plug this in, but we are going to need to plug in y_{t-3}, and so on. If we repeated this $t + m$ times, we would end up with the following:

$$y_t = \alpha_0 \sum_{i=0}^{t+m} \alpha_1^i + \alpha_1^{t+m+1} y_{-(m+1)} + \sum_{i=0}^{t+m} \alpha_1^i \varepsilon_{t-i} \text{ for all } t > 0. \qquad (4.6.2)$$

$t + m$ is an arbitrarily large number of time points backward in time. It is as many times as we may wish to replace y_{t-j} with $y_{t-j} = \alpha_0 + \alpha_1 y_{t-j-1} + \varepsilon_{t-j}$. There is a lot of mathematical notation involved here, but it can be simplified. If $|\alpha_1| < 1$, the term α_1^{t+m+1} approaches 0 as m approaches infinity, and the infinite sum $\sum_{i=0}^{t+m} \alpha_1^i = 1 + \alpha_1 + \alpha_1^2 + \cdots)$ converges to $1/(1 - \alpha_1)$. Therefore, as long as $|\alpha_1| < 1$ and the time series began a long time ago—allowing us to replace y_{t-j} many, many times—we can rewrite Equation 4.6.2 as

$$y_t = \frac{\alpha_0}{\left(1 - \alpha_1\right)} + \sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}. \qquad (4.6.3)$$

This is a solution to the original equation: We have expressed y_t as a function of the parameters of the data-generating process, ε_t, and time. Since $E(\varepsilon_t) = 0$, the unconditional expected value of y_t is

$$E\left(y_t\right) = \frac{\alpha_0}{\left(1 - \alpha_1\right)}. \qquad (4.6.4)$$

As we noted before, $\alpha_0/(1-\alpha_1)$ is the equilibrium to which the process will converge in the absence of external shocks. Hence, we use $\hat{\alpha}_0/(1 - \hat{\alpha}_1)$ to estimate the long-run equilibrium. As the process does have a constant value to which it will converge, it is stable.

If our lagged dependent variable process contains an exogenous variable, $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \varepsilon_t$, Equation 4.6.2 would be

$$y_t = \alpha_0 \sum_{i=0}^{t+m} \alpha_1^i + \alpha_1^{t+m+1} y_{-(m+1)} + \sum_{i=0}^{t+m} \alpha_1^i \beta_1 x_{t-i}$$

$$+ \sum_{i=0}^{t+m} \alpha_1^i \varepsilon_{t-i} \text{ for all } t > 0. \qquad (4.6.5)$$

If we assume that x_t has been constant since before $t + m$ time periods (a very long time ago), $x_t = c$, then Equation 4.6.3 would be

$$y_t = \frac{\alpha_0}{\left(1 - \alpha_1\right)} + \frac{\beta_1}{\left(1 - \alpha_1\right)} c + \sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}. \qquad (4.6.6)$$

If we instead assume that x_t increased by one unit $t + m$ time periods ago and has not changed since, $x_t = c + 1$, then Equation 4.6.3 would be

$$y_t = \frac{\alpha_0}{(1-\alpha_1)} + \frac{\beta_1}{(1-\alpha_1)}(c+1) + \sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}. \qquad (4.6.7)$$

The difference between Equations 4.6.6 and 4.6.7 is $\beta_1/(1-\alpha_1)$. Hence, we use $\hat{\beta}_1/(1-\hat{\alpha}_1)$ to estimate the long-run effect of a one-unit change in x_t, as we saw earlier.

Returning to the issue of stability, if $\alpha_1 = 1$, the term $\Sigma_{i=0}^{t+m} \alpha_1^i = 1 + \alpha_1 + \alpha_1^2 + \cdots$ in Equation 4.6.2 does not converge, and the equilibrium is undefined: $E(y_t) = ?$ If $|\alpha_1| > 1$, the term α_1^{t+m+1} in Equation 4.6.2 will tend to infinity over time: $E(y_t) = \pm\infty$.

Therefore, the data-generating process $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ is stable if $|\alpha_1| < 1$, but not otherwise. As stability is a necessary condition for stationarity, the data-generating process $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ cannot be stationary unless $|\alpha_1| < 1$.

This is true for any first-order autoregressive process: An AR(1) process is stable if $|\alpha_1| < 1$. In addition to stability, stationarity requires that the data-generating process contain no trending, periodicity, or structural breaks. In addition to this, we require one further condition for stationarity. When a time series process begins, it may not begin in equilibrium. In the convergent examples we looked at in Chapter 2, they each started at 1 and eventually converged on the equilibrium of 0. If we happen to be measuring this time series during this period of equilibration, our observed time series will not be stationary (the mean, variance, and covariances will be changing during this period), even though the series is stable in the long run. Therefore, we need to assume that the sequence started long enough ago so that it has had a chance to equilibrate, or that the process started immediately in equilibrium.

For the AR(1) process, note how important the value of α_1 is to stability. For the AR(1) equation, α_1 is called the characteristic root (denoted as m_1). For an AR(1) process to be stable, the absolute value of its characteristic root must be less than 1.

We now consider higher-order autoregressive processes. A second-order autoregressive process is written as

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t. \qquad (4.6.8)$$

The stability conditions for such an AR(2) process are as follows (Harvey, 1993, pp. 18–19):

$$\alpha_1 + \alpha_2 < 1.$$

$$-\alpha_1 + \alpha_2 < 1.$$

$$\alpha_2 > -1. \qquad (4.8.9)$$

For higher-order autoregressive processes,

$$y_t + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t. \qquad (4.6.10)$$

The characteristic roots (m_1, m_2, …) are a function of the parameters on the lagged dependent variables (α_1, α_2, …). The condition for stationarity is that the moduli (absolute value of a complex number) of the characteristic roots are less than 1. We do not need to worry about the exact functions. For higher-order autoregressive processes the characteristic roots are difficult to calculate and may take on imaginary values. Fortunately, social scientists rarely deal with anything higher than a second- or third-order autoregressive process. When we do, software packages will usually calculate the moduli of the characteristic roots for us, so that we can test stability.

There are some general rules regarding the equation parameters (α_i) that can be followed:[8]

1. In the equation describing a *p*th-order autoregressive process, a necessary condition for stability is $\sum_{i=1}^{p} \alpha_i < 1$.
2. A sufficient condition for stability is $\sum_{i=1}^{p} |\alpha_i| < 1$.
3. At least one characteristic root equals unity if $\sum_{i=1}^{p} \alpha_i = 1$.

Any sequence that contains one or more characteristic roots that equal unity is called a unit root process. Such a process is not stable. For example, $y_t = \alpha_1 y_{t-1} + \varepsilon_t$, with $\alpha_1 = 1$, is a unit root process. It is often called a random walk and has very interesting properties, which we shall discuss further in Chapter 6.

For the stationarity conditions for ARCH processes, we can apply the same stationarity conditions listed above for autoregressive processes to the squared errors in the data-generating process of the ARCH model (Equation 4.4.3). For a moving average process to be stationary, it simply needs to be of a finite order ($q \neq \infty$). This is not an overly restrictive condition.

Summary

You have now been introduced to both static and dynamic models and the idea of selecting between these models using the general-to-specific approach, as well as a set of criteria to choose between using an LDV model and a static model that corrects for serial correlation in the errors.

In the next chapter, you will be introduced to another set of models called autoregressive moving average (ARMA) models. These lend themselves to another form of model selection. This is the Box-Jenkins approach, to which you will also be introduced in the next chapter.

[1] Note that we include the squared trend term without the linear component. This follows the modelling choice made by Soroka and Wlezien (2010).

[2]Note that we are are assuming that $E(\varepsilon_t \mid x_t) = 0$.

[3] As we will discuss in Chapter 5, the serial correlation may also be due to an unmodelled moving average process.

[4] The two periodicity variables allow us to model a flexible election cycle, where not only is the amplitude of the cycle estimated but so is the timing of the peaks and troughs (phase).

The cycling is captured by including two terms in the regression: $\beta_1 \sin(\lambda\theta)$ and $\beta_2 \cos(\lambda\theta)$, where β_1 and β_2 are the parameters to be estimated and λ is the frequency of the election cycle, so that $\lambda\theta$ is defined by the length of the interelection period. If you plot the sine and cosine functions, they represent waves. As $\lambda\theta$ increases over a range of 2π, both the sine and the cosine functions cycle through a complete peak and trough. The sine and cosine functions are out of phase with each other, in that when one is at its maximum the other is halfway between its maximum and minimum. By combining a sine and a cosine function, weighted by β_1 and β_2, respectively, we can create a wave of any amplitude (and phase). The estimated parameters β_1 and β_2 can be used to calculate the amplitude of the interelection cycle.

[5] If we tested the long-run effect of GDP, we would find it is not statistically significant at the 0.05 significance level.

[6] Enforcement activities consist of notices of violation, consent decrees, and administrative orders.

[7] Each is coded as “1” for the month after which the change occurred and for the 2 months following that; otherwise, they are coded as “0.”

[8] Enders (2004).

http://dx.doi.org/10.4135/9781483390857.n4