# Static Time Series Models and Ordinary Least Squares Estimation

## 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: 51-80

# Static Time Series Models and Ordinary Least Squares Estimation

Having covered many of the fundamental concepts of time series in Chapter 2, in this chapter we begin to explore basic models of time series data. This chapter examines the static and finite distributed lag (FDL) models estimated through ordinary least squares (OLS) regression. We examine the assumptions required for the estimation of such models to be unbiased. We also examine how to test for and correct violations of key assumptions—in particular covariance stationarity and no serial correlation. In the discussion of violations of covariance stationarity, we discuss the challenges of trending, periodicity, and structural breaks. In the discussion of correcting for serial correlation, we are introduced to time series models that are estimated by maximum likelihood.

---

## 3.1 Static and Finite Distributed Lag (FDL) Models

Let us consider two simple time series models. A *static* time series model relates contemporaneous variables. For example, with one independent variable,

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t. \qquad (3.1.1)$$

An FDL model allows one or more variables to affect $y_t$ with a lag. For example, with one independent variable and two of its lags,

$$y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \varepsilon_t. \qquad (3.1.2)$$

In the above, $\delta_0$, $\delta_1$, and $\delta_2$ are slope parameters just like $\beta_1$. More generally, an FDL model of order $q$ will include $q$ lags of $x_t$: ($x_{t-1}$, $x_{t-2}$, …, $x_{t-q}$). We call $\delta_0$ the impact propensity (effect)—it reflects the immediate change in $y_t$ due to a one-unit change in $x_t$. For a temporary, one-period change, $y_t$ returns to its original level $q$ periods after $x_t$ returns to its original value. To get a sense of what this means, consider the $q$ = 2 FDL process:

$$y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \varepsilon_t. \qquad (3.1.3)$$

Let us assume that $x_t = 0$ at $t = 1$ and $t = 2$, then, $x_t = 1$ at $t = 3$, and $x_t$ returns to 0 thereafter. From Equation 3.1.3, we can calculate the expected value of $y_t$, conditioning on $x_t$, $x_{t-1}$, $x_{t-2}$ for $t = 1$ to $t = 7$. The calculated values are given in Table 3.1.

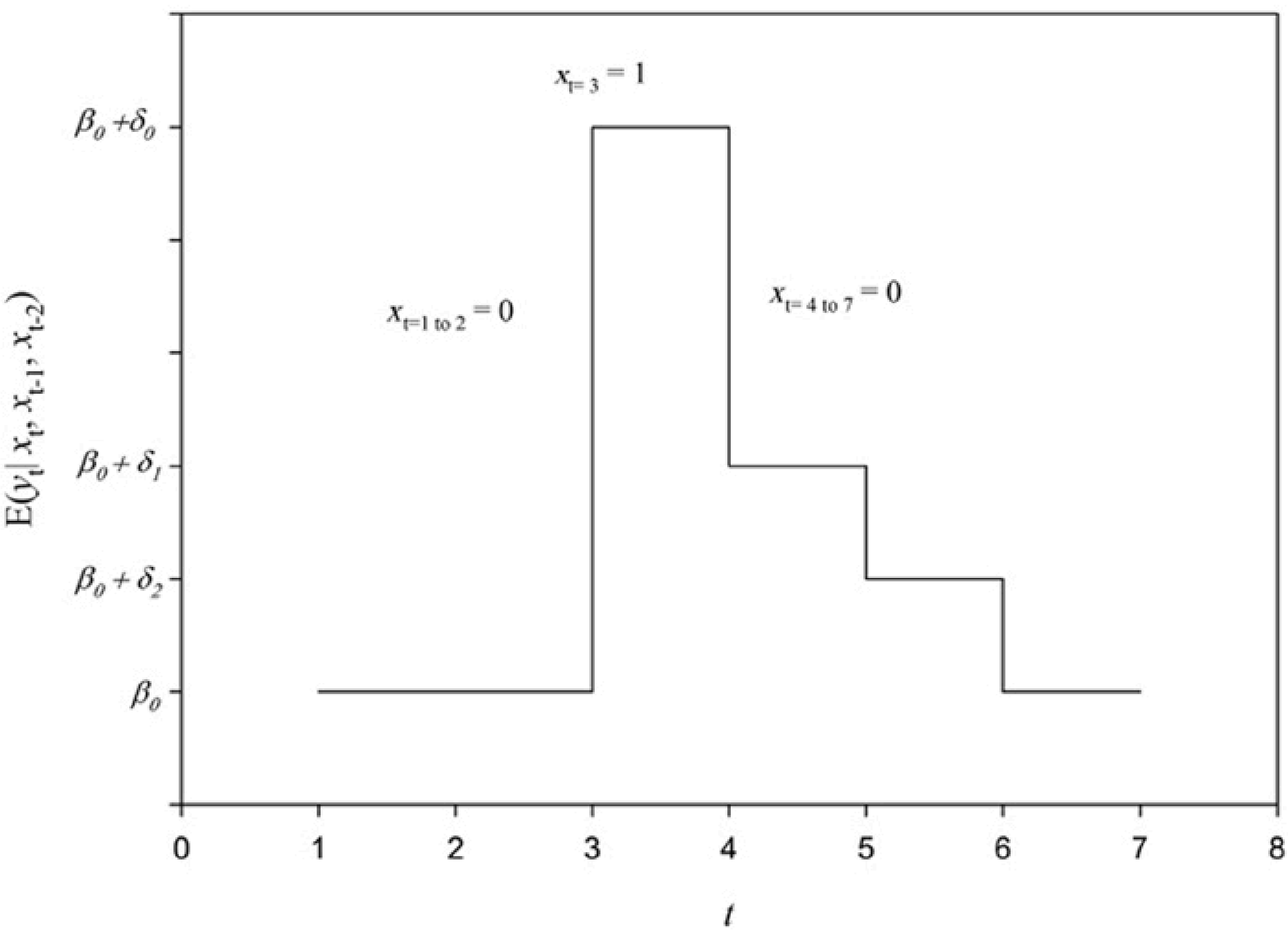
**Table 3.1 Expected Values From a *q* = 2 FDL Process**

| $t$ | $x_t$ | $E(y_t \mid x_t, x_{t-1}, x_{t-2})$ |
|---|---|---|
| 1 | 0 | $\beta_0$ |
| 2 | 0 | $\beta_0$ |
| 3 | 1 | $\beta_0 + \beta_0$ |
| 4 | 0 | $\beta_0 + \delta_1$ |

| | | |
|---|---|---|
| 5 | 0 | $\beta 0 + \delta 2$ |
| 6 | 0 | $\beta 0$ |
| 7 | 0 | $\beta 0$ |
| NOTE: FDL = finite distributed lag. | | |

We can graphically display the same results, as in Figure 3.1.

**Figure 3.1 Expected Values From a *q* = 2 FDL Process**




NOTE: FDL = finite distributed lag.

Both the table and the figure demonstrate the same thing. With $xt = 0$ for the first two time points, the expected value of $yt$ is the constant $\beta 0$. When $xt$ increases by one unit to 1 at $t = 3$, the expected value increases to $\beta 0 + \delta 0$. The change, $\delta 0$, is the coefficient on $xt$ and is the impact propensity. At $t = 4$, $xt$ returns to 0. However, $xt-1$ is the value of $xt$ at the previous time point, and so $xt-1 = 1$ at $t = 4$. This means that the expected value is now the constant plus the coefficient on $xt-1$: $\beta 0 + \delta 1$. Note that the graphical representation assumes that

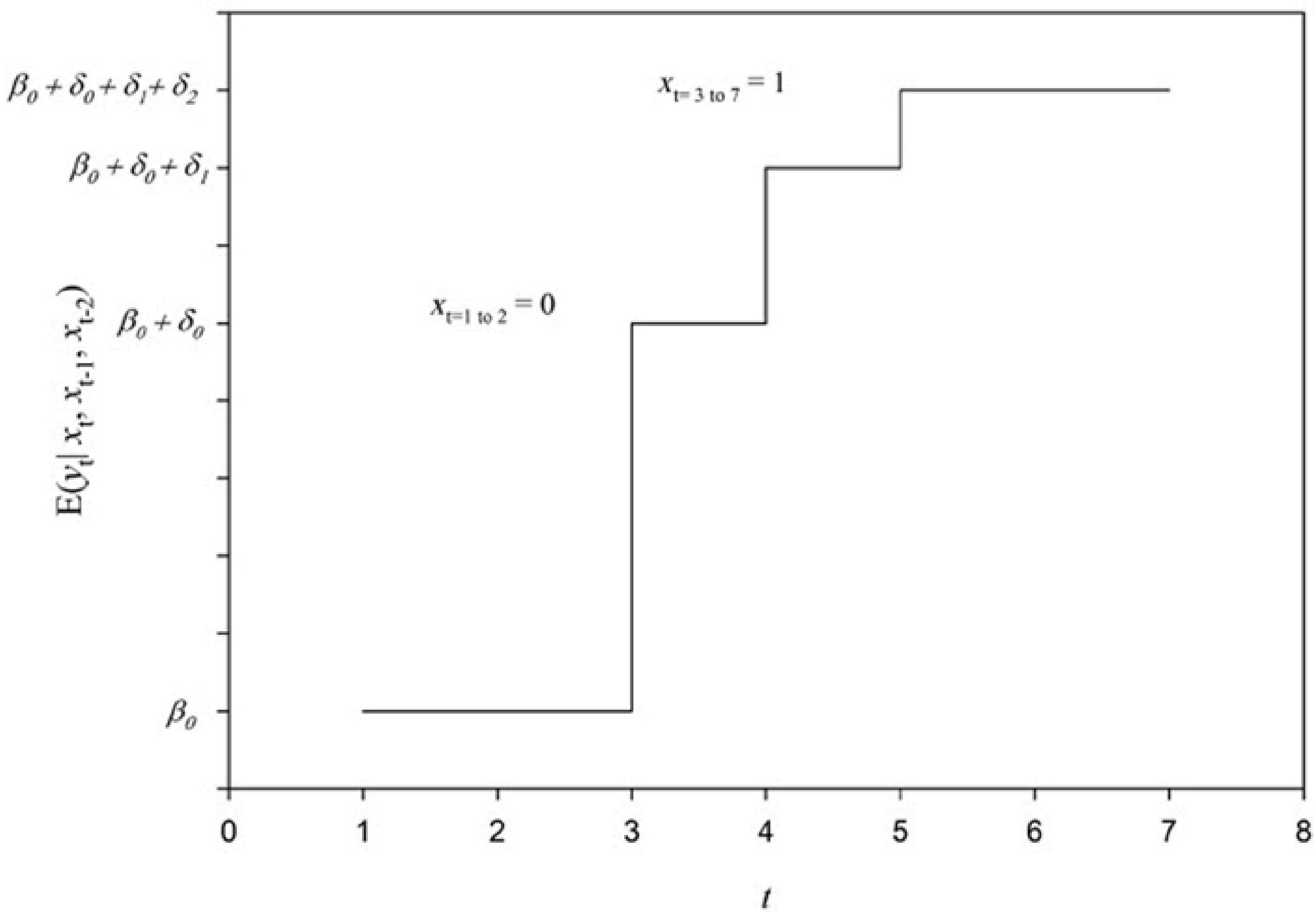
$\delta_1 < \delta_0$. This was just chosen for the purpose of illustration, and nothing requires it to be the case. At $t = 5$, $x_t$ remains 0, and $x_{t-1}$ now returns to 0. Since $x_{t-2}$ is the value of $x_t$ two times previous, $x_{t-2} = 1$ at $t = 5$. The expected value is now the constant plus the coefficient on $x_{t-2}$: $\beta_0 + \delta_2$. At $t = 6$, $x_t$ and $x_{t-1}$ remain 0, and $x_{t-2}$ returns to 0. The expected value is now simply back to the constant: $\beta_0$.

We call $\delta_0 + \delta_1 + \ldots + \delta_q$ the long-run propensity or long-run effect. It reflects the long-run change in $y_t$ after a permanent one-unit change in $x_t$. Let us assume that $x_t = 0$ at $t = 1$ and $t = 2$; then, $x_t = 1$ thereafter. Again from Equation 3.1.3, we can calculate the expected value of $y_t$, conditioning on $x_t$, $x_{t-1}$, $x_{t-2}$ for $t = 1$ to $t = 7$. The calculated values are given in Table 3.2 and graphically displayed in Figure 3.2.

**Table 3.2 Expected Values From a $q = 2$ FDL Process**

| $t$ | $x_t$ | $E(y_t \mid x_t, x_{t-1}, x_{t-2})$ |
|---|---|---|
| 1 | 0 | $\beta_0$ |
| 2 | 0 | $\beta_0$ |
| 3 | 1 | $\beta_0 + \delta_0$ |
| 4 | 1 | $\beta_0 + \delta_0 + \delta_1$ |
| 5 | 1 | $\beta_0 + \delta_0 + \delta_1 + \delta_2$ |
| 6 | 1 | $\beta_0 + \delta_0 + \delta_1 + \delta_2$ |
| 7 | 1 | $\beta_0 + \delta_0 + \delta_1 + \delta_2$ |
| NOTE: FDL = finite distributed lag. | | |

**Figure 3.2 Expected Values From a *q* = 2 FDL Process**




NOTE: FDL = finite distributed lag.

Again, the table and figure demonstrate the same thing. With $x_t = 0$ for the first two time points, the expected value of $y_t$ is the constant $\beta_0$. When $x_t$ increases by one unit to 1 at $t = 3$, the expected value increases to $\beta_0 + \delta_0$. At $t = 4$, $x_t$ remains 1 and $x_{t-1}$, the value of $x_t$ at the previous time point, is also 1. This means that the expected value is now the constant plus the coefficient on $x_t$ plus the coefficient on $x_{t-1}$: $\beta_0 + \delta_0 + \delta_1$. At $t = 5$, $x_t$ and $x_{t-1}$ remain 1, and $x_{t-2}$, the value of $x_t$ two time points prior, is 1. The expected value is now the constant plus the coefficient on $x_t$ plus the coefficient on $x_{t-1}$ plus the coefficient on $x_{t-2}$: $\beta_0 + \delta_0 + \delta_1 + \delta_2$. At $t = 6$, $x_t$, $x_{t-1}$, and $x_{t-2}$ remain 1, and the expected value remains as $\beta_0 + \delta_0 + \delta_1 + \delta_2$. As long as $x_t$ does not change, the expected value will not change. The difference between this expected value and the expected value before the one-unit change in $x_t$ at $t = 3$ is the full effect of this change: $\delta_0 + \delta_1 + \delta_2$. This is the long-run propensity.

We can compare the FDL process with the static process (no lags of the independent variables). In the static process (Equation 3.1.1), the estimated short-run and long-run effects are the same and equal: $\beta_1$. A permanent one-unit change in $x_t$ produces an immediate effect of $\beta_1$ on $y_t$, and this is the sum total of its estimated effect.

Both the static process and the FDL process can contain multiple independent variables. For example, consider an FDL process that has $k$ independent variables each with $q$ lags:

$$y_t = \beta_0 + \beta_{1,0}x_{1,t} \cdots + \beta_{k,0}x_{k,t} + \sum_{i=1}^{k}\sum_{m=1}^{q}\beta_{i,m}x_{i,t-m} + \varepsilon_t \qquad (3.1.4)$$

Note that if we do have more than one independent variable, these can each be included with a different number of lags. Both the static model and the FDL model can be estimated using OLS. Just as with the application of OLS to cross-sectional data, there are some assumptions that need to be met when applying OLS to time series data. It is to these assumptions that we now turn.

## 3.2 Assumptions for Unbiasedness with Time Series Analysis

Many of the assumptions necessary for OLS regression with cross-sectional data apply in the same way as when we are using time series data. We assume that the relationships between $Y_t$ and the $X_t$ s are linear. We assume that no $X_t$ is constant and there is no perfect collinearity. We will also need to make a zero conditional mean assumption, as defined for time series data in Section 2.5 in Chapter 2. But what about the assumption of a random sample?

As discussed in Chapter 2, time series data are not produced by repeatedly sampling randomly from a population or a data-generating process. Time series data are a single realization of a random data-generating process. We do not have a random sample of realizations of $Y_t$ with which we can estimate $E(y_t)$ for a particular time point $t$. We can, however, estimate $E(Y_t)$ based on the average value of a single realization of $Y_t$ across all $t$. This is useful if we can assume that this converges on $E(Y_t)$ when you have enough time points:

$$\text{As } T \rightarrow \infty\text{: } E(Y_t) \rightarrow E(y_t) \text{ for all } t,\ t = 1, \ldots, T \qquad (3.2.1)$$

As we saw in Chapter 2, this will be true if $y_t$ is covariance stationary. Recall that a stochastic process is covariance stationary if E($y_t$) is constant, Var($y_t$) is constant, and, for all $k$> 0 and $h \neq 0$, Cov($y_{t=k}$, $y_{t=k+h}$) depends only on $h$ and not on $k$. This allows us to estimate the means, variances, and covariances of the data-generating process from the observed data.

In summary, covariance stationarity makes it possible to use the estimated parameters from the data model based on the single observed realization to say something about the unobserved data-generating process. This is equivalent to the idea in cross-sectional data analysis that having a random sample of data means that the sample is representative of the larger population, and so we are able to say something about the unobserved population from the parameters estimated from the observed data model.

Assumptions about the representativeness of the means, variances, and covariances estimated from the data are the same in both cases, although with cross-sectional data this is sometimes expressed in terms of the population. For time series data, it is always expressed relative to the data- generating process. For cross-sectional data, this representativeness is often derived from the assumption that the data are the product of random, independent draws from the population or data-generating process. For time series data, this representativeness is derived from the assumption of stationarity.

Just as in the cross-sectional case, we need to add an assumption of homoskedasticity in order to be able to derive the correct standard errors from the usual OLS estimator. For time series data, we assume that $\mathrm{Var}(\varepsilon_t | x_s) = \sigma^2$, for all *s.* This means that the error variance is independent of the *x*s for all leads and lags and it is constant over time.

To derive the correct standard errors, when using time series data, we also need the assumption of no serial correlation, conditional on ***X***. As noted in Chapter 2, serial correlation is another name for autocorrelated errors:

$$\mathrm{Corr}\left(\varepsilon_t, \varepsilon_{t-h} \mid \boldsymbol{X}\right) = \frac{\mathrm{Cov}\left(\varepsilon_t, \varepsilon_{t-h} \mid \boldsymbol{X}\right)}{\mathrm{Var}\left(\varepsilon_t, \varepsilon_t \mid \boldsymbol{X}\right)}. \qquad (3.2.2)$$

The conditioning on ***X*** is often left implicit. For example, first-order serial correlation is the correlation of the error with the first lag of itself:

$$\mathrm{Corr}\left(\varepsilon_t, \varepsilon_{t-1}\right) = \frac{\mathrm{Cov}\left(\varepsilon_t, \varepsilon_{t-1}\right)}{\mathrm{Var}\left(\varepsilon_t, \varepsilon_{t-1}\right)}. \qquad (3.2.3)$$

The no serial correlation assumption is as follows:

$$\mathrm{Corr}(\varepsilon_t,\ \varepsilon_{t+h}) = 0 \text{ for } h \neq 0. \qquad (3.2.4)$$

In total, we have five assumptions:

A.1 Linear in parameters
A.2 Variance in all *X* and no perfect collinearity
A.3 Zero conditional mean (strict exogeneity)
A.4 Homoskedasticity
A.5 No serial correlation

Under these five assumptions and the assumption of covariance stationarity, the OLS estimator of the regression coefficients is unbiased, and the OLS estimator of $\sigma^2$ in the time series case is the same as in the cross-sectional case. These assumptions are called the Gauss-Markov assumptions. With the additional assumption of normal and independent errors, inference is the same as in the cross-sectional case.

A.6 Normal and independent errors

Together, these six assumptions are called the classical linear model assumptions of time series regression

(Wooldridge, 2009). Let us think about this more concretely by returning to our example of public responsiveness in Canada from Chapter 1. Our model is as follows:

$$R_t = \beta_0 + \beta_1 P_t + \beta_2 W_t + \varepsilon_t. \qquad (3.2.5)$$

Recall that $R_t$ is the public's relative preference for social policy spending in a given year. $P_t$ is the actual level of policy spending in a year. $W_t$ represents other, exogenous effects on the public's relative preferences.

The validity of any results from estimating this model by OLS is contingent on meeting the Gauss-Markov assumptions and the assumption of stationarity. In thinking about whether the necessary assumptions of OLS regression apply, we would ask ourselves a series of questions. First, is it reasonable to assume that an appropriate model of policy responsiveness is linear in its parameters (A.1)? For example, the public's relative preference for social policy spending may not be a linear function of actual policy spending. However, it may be a linear function of the log of actual policy spending, in which case we would want to use this as our independent variable.

If we have an independent variable in our model with no variation or if we have independent variables that are perfectly collinear (A.2), then any statistical package will pick up the problem and omit one or more variables to resolve the problem. However, we might want to ask ourselves if we have independent variables that are highly collinear. This would not be a concern of bias, but we might want to keep in mind that our standard errors are being increased by it.

We will also want to ask ourselves whether there are any obvious violations of strict exogeneity (A.3). This includes the usual concerns about spurious correlations and omitted-variable bias. It is also a possibility that the data-generating process includes a lag of the dependent variable, which we would then want to include in our data model. We will examine how to address this in Chapter 4. Not controlling for a trend that exists in the data-generating process can also be a violation of exogeneity. It is also a violation of covariance stationarity, which we should remember we are always assuming or must account for in the time series models discussed in this text. We will look at how to test and correct for trending in the next section.

Finally, we will ask ourselves whether or not our errors have constant variance—homoskedasticity—(A.4), no serial correlation (A.5), and are independently and normally distributed (A.6). We will examine how to test and correct for violations of these later in this chapter. We sometimes summarize Assumptions A.4 to A.6 by stating that the $\varepsilon_t$ values in an equation such as Equation 3.2.5 are independent and distributed normally with mean 0 and constant variance $\sigma^2_\varepsilon$, or $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$.

Next, we turn to a scenario in which we have a large number of time points. If we have such data, we can relax the strict-exogeneity assumption and still obtain asymptotic unbiasedness (and consistency). This also has a parallel with cross-sectional data analysis. For cross-sectional data with a large *N*, it is possible to relax the zero conditional mean assumption to no correlation between the exogenous variables and the errors. With time series data with a large *T*, it is possible to relax the strict-exogeneity assumption to contemporaneous exogeneity, *if* we can make an additional assumption. As discussed in Chapter 2, that assumption is weak dependence.

Weak dependence (with covariance stationarity) allows us to relax the strict-exogeneity assumption for large-*T* analysis in demonstrating that OLS is consistent. As *T* goes to infinity, the sample variances and covariances converge on the population variances and covariances. In this case, $\beta_1$ can be estimated as follows:

$$\hat{\beta}_1 = \frac{\widehat{\mathrm{Cov}}(Y_t, X_t)}{\widehat{\mathrm{Var}}(X_t)}. \qquad (3.2.6)$$

We can state the “asymptotic” Gauss-Markov assumptions for time series regression as follows:

AA.1 Linearity and weak dependence
AA.2 Variance in all *X* and no perfect collinearity
AA.3 Zero conditional mean (contemporaneous exogeneity)
AA.4 Homoskedasticity
AA.5 No serial correlation

Note that Assumptions AA.1 through AA.3 are required for the parameter estimates to be asymptotically unbiased. AA.4 and AA.5 are required to estimate the correct parameter standard errors.

## 3.3 Testing and Correcting for Trending, Periodicity, and Structural Breaks

The weaker form of stationarity that we assume (covariance stationarity) requires that the mean, variance, and covariances are constant across time. As discussed in Chapter 2, any trending, periodicity, or structural breaks will violate this assumption. It is therefore important to account for any of these processes. We begin with trends. One possibility is a deterministic linear trend, which can be modelled as follows:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \qquad t = 1, 2, \ldots \qquad (3.3.1)$$

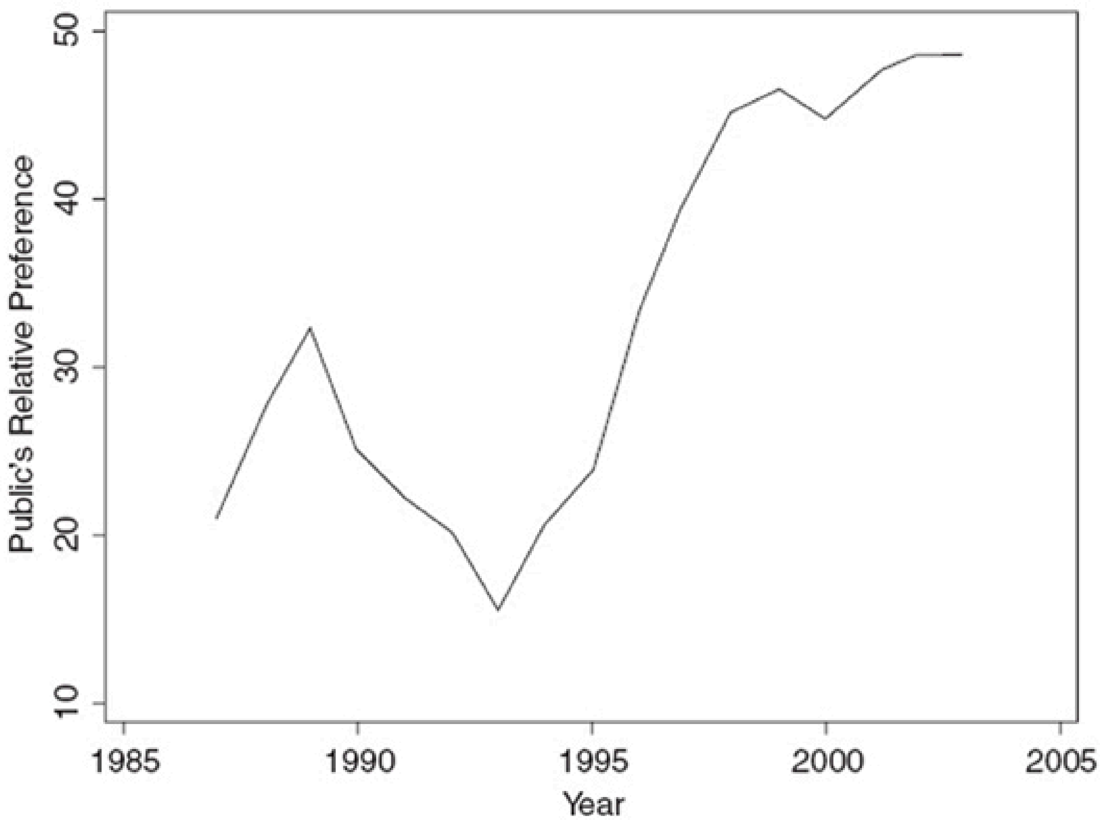
Another possibility is a deterministic quadratic trend, which can be modelled as follows:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t, \qquad t = 1, 2, \ldots \qquad (3.3.2)$$

We will discuss other types of trends in Chapter 6.

In Chapter 2, we found that we got very different results when we estimated the Canadian public responsiveness model with a trend from what we obtained when we estimated it without a trend. Before beginning that analysis, we might have plotted the dependent variable—the public's relative preference for social policy spending (Figure 3.3).

**Figure 3.3 Public's Relative Preference for Social Policy Spending—Canada**




Clearly, the public's relative policy spending preference has been going up over time. However, it has not been going up smoothly. It appears to be exhibiting trending but not linear trending. We may consider the possibility that the public's relative policy spending preference contains a quadratic trend.

$$R_t = \beta_0 + \beta_1 P_t + \beta_2 t + \beta_3 t^2 + \varepsilon_t. \qquad (3.3.3)$$

The results from estimating such a model by OLS are given in Table 3.3.

**Table 3.3 Canadian Public Responsiveness Model**

| *Preference* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Program spending | −0.57 | 0.10 | −5.41 | <0.001 |
| Counter | 1.47 | 0.84 | 1.75 | 0.106 |
| Counter$^2$ | 0.16 | 0.047 | 3.29 | 0.006 |
| Constant | 115.55 | 16.83 | 6.87 | <0.001 |

NOTE: $R^2 = 0.93$, $T = 16$; $T$ = number of time points.

We see that the squared time component in the quadratic trend is significant at the 0.05 significance level. In Chapter 2, we estimated this model with a linear trend. Which should we use? The significance of the squared time variable indicates that the linear trend is insufficient, which suggests that we should use the quadratic trend.[1] Ultimately, we will want to choose the model that ensures the errors are independent and normally distributed with mean 0 and constant variance $\sigma^2_\varepsilon$: $\varepsilon_t \sim NID(0, \sigma^2_\varepsilon)$. We shall test for this momentarily.

How do we interpret the significance of the quadratic trend terms? In Chapter 2, we noted that the interpretation of a linear trend is that $y_t$ increases or decreases on average by the magnitude of $\beta_2$ each time period. A quadratic term means that $y_t$ increases or decreases on average by a decreasing or increasing magnitude, of $\beta_2 + \beta_3(2t)$ each time period. Also noted in Chapter 2, trend terms (linear or quadratic) are controls for violations of covariance stationarity and should only be interpreted as controls. Normally, any further substantive interpretation of these components of the model should be avoided.

Interpreting the coefficient on program spending, for each billion-dollar increase in government spending, the public's relative preference for spending decreases by 0.57 of a percentage point. Recall that relative preference is the difference between the percentage of respondents who want more spending and the percentage of respondents who want less.

An alternative to adding a linear or quadratic trend to a regression is to use “detrended” data in the regression. Detrending a series involves regressing each variable in the model on $t$ (or a more complex function of time) and predicting the errors. The predicted errors from each of these regressions form the detrended series for each variable; the trend has been partialled out. Let us consider an example.

In this example, we estimate a public responsiveness model for the United Kingdom. We begin by plotting the dependent variable (Figure 3.4).

**Figure 3.4 Relative Preference for Social Program Spending in the United Kingdom**


As with the Canadian public's relative spending preference, the upward trend in the U.K. public demand for an increase in major social program spending is self-evident. Accordingly, we could include a quadratic time trend in our U.K. model. The results are presented in Table 3.4.

**Table 3.4 U.K. Public Responsiveness Model**

| *Preference* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Program spending | −0.98 | 0.37 | −2.62 | 0.020 |
| Counter | 4.09 | 0.37 | 10.95 | <0.001 |
| Counter$^2$ | −0.05 | 0.04 | −1.17 | 0.261 |
| Constant | 111.52 | 22.45 | 4.97 | <0.001 |

NOTE: $R^2 = 0.96$, $T = 18$; $T$ = number of time points.

The results indicate that an increase in major social program spending results in a decline in the public's relative preference—this means relatively less demand for spending. For each billion-pound increase in government spending, the public's relative preference for spending decreases by 0.98 of a percentage point.

Alternatively, we could detrend the public's relative spending preference and government policy spending variables. We regress the public's relative spending preference on the time trend variables. The results are presented in Table 3.5.

**Table 3.5 U.K. Public's Relative Spending Preference**

| *Preference* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Counter | 4.03 | 0.44 | 9.14 | <0.001 |
| Counter$^2$ | −0.14 | 0.02 | −5.78 | <0.001 |
| Constant | 52.73 | 1.61 | 32.67 | <0.001 |

NOTE: $R^2 = 0.94$, $T = 18$; $T$ = number of time points.

Using these results, we can estimate the errors (the residuals) for this model. This is our detrended public's relative spending preference variable. We then do the same for the government policy spending variable (Table 3.6).

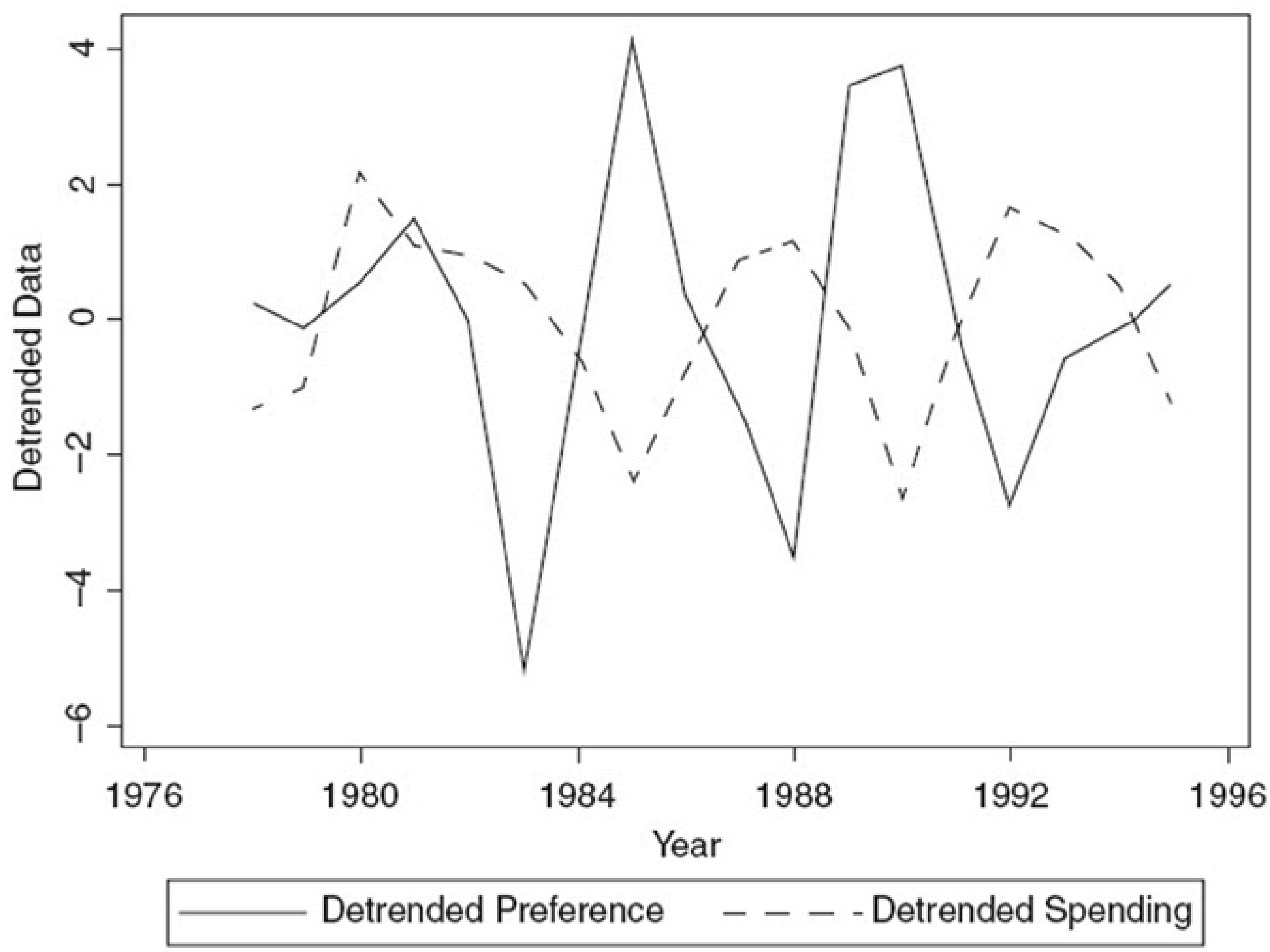
**Table 3.6 U.K. Government Policy Spending**

| *Spending* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Counter | 0.071 | 0.26 | 0.28 | 0.787 |
| Counter$^2$ | 0.097 | 0.015 | 6.63 | <0.001 |
| Constant | 60.12 | 0.95 | 63.44 | <0.001 |

NOTE: $R^2 = 0.98$, $T = 22$; $T$ = number of time points.

The residuals from these two estimated models provide us with our two detrended variables. These are plotted in Figure 3.5.

**Figure 3.5 Detrended Government Policy Spending and Public Relative Spending Preference**




It is clear that the trend is removed from the public's relative spending preference variable, as is the case with the government policy spending variable.

We can now regress the detrended public's relative spending preference variable on the detrended government policy spending variable. We do not include a constant in this model as it has been partialled out along with the trend (Table 3.7).

**Table 3.7 U.K. Public Responsiveness Model—Detrended**

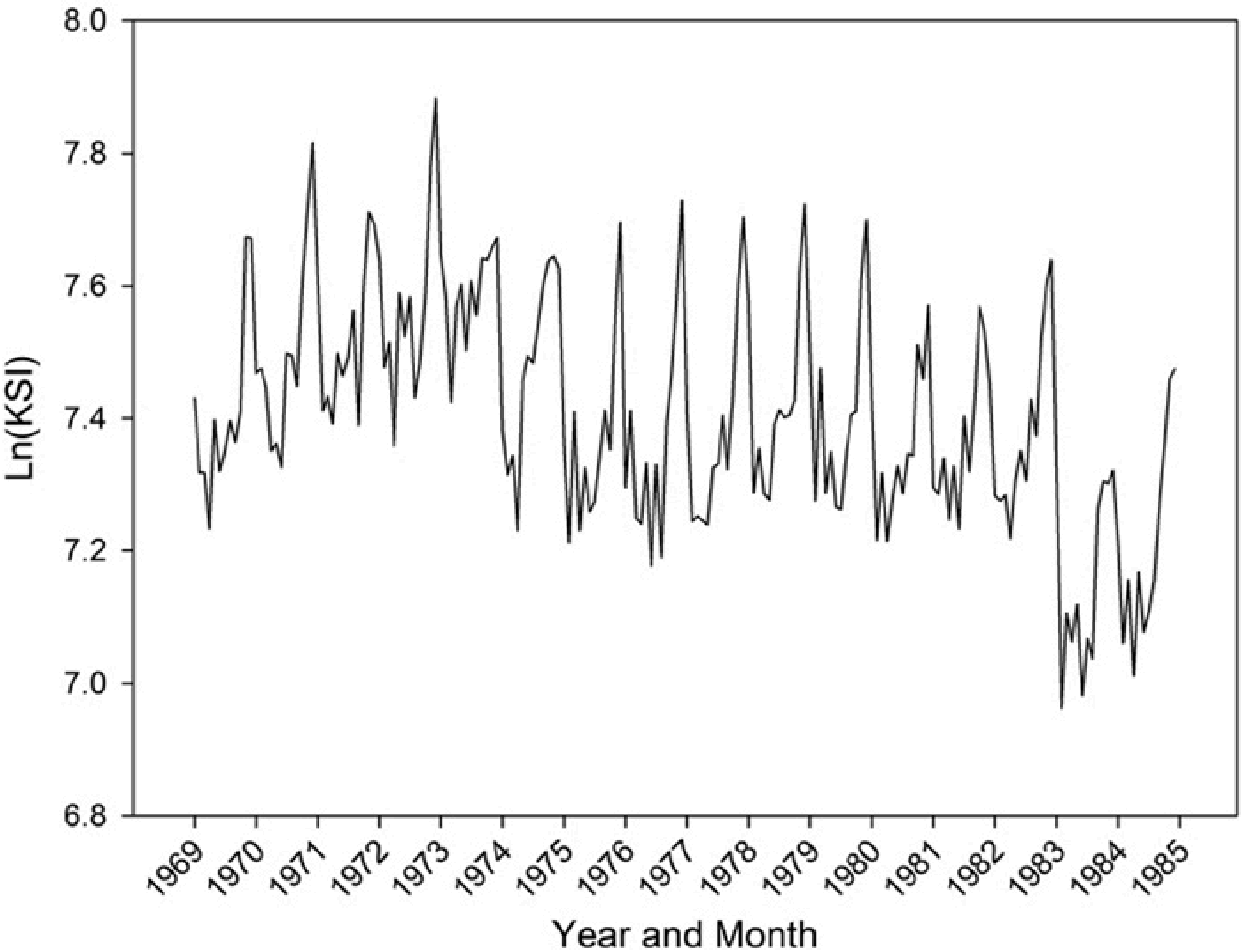
| *Preference* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Detrended spending | −0.98 | 0.34 | −2.89 | 0.010 |

NOTE: $R^2 = 0.33$, $T = 18$; $T$ = number of time points.

The results are equivalent to those we get from including the trend variable in the model (c.f. Table 3.4). An advantage of detrending the data (vs. adding a trend) involves the calculation of goodness of fit. Time series regressions with trend variables tend to have very high $R^2$ as the trending data are well explained by the trend variable. The $R^2$ from a regression of the detrended dependent variable on the detrended independent variables better reflects how well the $X_t$s explain $Y_t$. Compare the $R^2$ value in Table 3.7 with that in Table 3.4. The disadvantage of detrending is that the uncertainty in the estimation of the trend is not reflected in subsequent analyses using the detrended data; so the standard errors are a bit too small. This is generally not a big concern.

We turn now to periodicity and structural breaks. As we have seen, even if those factors causing trending are unobserved, we can control for them by directly controlling for the trend. This applies equally to structural breaks and periodicity. We examine this using an example that looks at traffic injuries in the United Kingdom from 1969 to 1984. Consider the monthly data on the number of drivers in the United Kingdom who were killed or seriously injured (KSI) in a traffic accident (Harvey & Durbin, 1986) between January 1969 and December 1984. Figure 3.6 provides a plot of the natural log of KSI.

**Figure 3.6 ln(KSI), United Kingdom, From 1969 to 1984**




The data appear to exhibit periodicity. Specifically, the data appear to go through a complete cycle each year, with KSIs hitting a peak in winter and a nadir in summer. Periodicity that follows a seasonal pattern is called seasonality. There also appears to have been a sudden downward shift in KSIs in February 1983, when a new seatbelt law was introduced.

We begin by testing for seasonality and a structural break at $t$ = 170 (February 1983). We include a time variable to capture any trending and a series of dummy variables for all but one month. We exclude December, so this will be the reference month for the coefficients on the included dummy variables. We also include a variable called "seatbelt law," which is coded "0" before $t$ = 170 and "1" on and after $t$ = 170. We can use this to test for a structural break in the equilibrium of ln(KSI) at the introduction of the new seatbelt law. Finally, we test for the effect of (the log of) petrol prices (Table 3.8).

**Table 3.8 Static Model of ln(KSI), United Kingdom, From 1969 to 1984**

| *ln(KSI)* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Counter | −0.001 | 0.0001 | −5.10 | <0.001 |
| January | −0.24 | 0.029 | −8.43 | <0.001 |
| February | −0.35 | 0.029 | −12.51 | <0.001 |
| March | −0.31 | 0.029 | −11.09 | <0.001 |
| April | −0.39 | 0.029 | −13.74 | <0.001 |
| May | −0.30 | 0.028 | −10.62 | <0.001 |
| June | −0.33 | 0.028 | −11.80 | <0.001 |
| July | −0.28 | 0.028 | −10.01 | <0.001 |
| August | −0.27 | 0.028 | −9.71 | <0.001 |
| September | −0.24 | 0.028 | −8.45 | <0.001 |
| October | −0.16 | 0.028 | −5.74 | <0.001 |
| November | −0.056 | 0.028 | −1.99 | 0.049 |
| Seatbelt law | −0.15 | 0.022 | −6.63 | <0.001 |
| Petrol—ln(£) | −0.34 | 0.057 | −6.06 | <0.001 |
| Constant | 6.95 | 0.14 | 50.81 | <0.001 |

NOTE: $R^2 = 0.80$, $T = 192$; $T$ = number of time points, KSI = killed or seriously injured.

On examining the results in Table 3.8, there appears to be significant seasonality with fewer KSIs in all months compared with December (especially February through August). This is apparent in the statistical significance (at the 0.05 level) of the monthly dummy variables. Furthermore, an *F* test of the joint significance of the 11 monthly dummy variables is $F(11, 177) = 34.66$, with a *P* value <0.001.[2] It also appears that the price of petrol, controlling for seasonality and the new seatbelt law, has a negative effect on the number of KSIs. Finally, the average number of KSIs seems to have dropped considerably following the introduction of the new seatbelt law.

The estimated coefficient for the seatbelt law variable indicates that, on average and holding log of the petrol prices and seasonality constant, there were 15% fewer KSIs per month.[3] The variable captures a level change in the number of KSIs (logged). The seatbelt law is an intervention, and testing for its effect in this manner is a form of intervention analysis. We will return to this more formally in Chapter 5.

## 3.4 Testing for Serial Correlation, Heteroskedasticity, and Nonnormally Distributed Errors

Before moving on to the topic of testing for serial correlation, let us examine the consequences of violating the assumption of no serial correlation. Consider two separate OLS estimations of a data model

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\mu}_t. \qquad (3.4.1)$$

1. On data that have a data-generating process of

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t. \qquad (3.4.2)$$

2. On data that have a data-generating process of

$$y_t = \beta_0 + \beta_1 x_t + \mu_t, \qquad (3.4.3)$$

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t,$$

with $\varepsilon_t$ in both cases being independent and normally distributed with mean = 0 and constant variance = $\sigma^2_\varepsilon$: $\varepsilon_t \sim NID(0, \sigma^2_\varepsilon)$.

Note that $\mu_t$ in (Equation 3.4.3) is AR(1), an autoregressive process of order 1, as it contains one lag of itself on the right-hand side, $\mu_{t-1}$. In this equation, we are describing the data-generating process of the error term and so use ρ in place of $\alpha_1$. This type of process is one possible source of serial correlation in error terms. We will discuss others in this section.

When using the usual estimator of the variance of $\hat{\beta}_1$, we assume that the residuals (errors) are independent—no serial correlation:

$$\mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum \left(x_i - \bar{x}\right)^2}. \qquad (3.4.4)$$

With serial correlation, we cannot make this assumption, and so the calculation must account for this . If serial correlation is the product of first-order autoregression, the correct calculation is

$$\operatorname{Var}\left(\hat{\beta}_1\right)=\left(\frac{\sigma^2}{\sum\left(x_t-\bar{x}\right)^2}\right)+2\sigma^2\left(\frac{\sum_{t=2}^{T}\left(x_t-\bar{x}\right)\left(x_{t-1}-\bar{x}\right)}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}\right)\rho. \quad (3.4.5)$$

Note that

$$\left(\frac{\sum_{t=2}^{T}\left(x_t-\bar{x}\right)\left(x_{t-1}-\bar{x}\right)}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2}\right)$$

is the estimated autocorrelation in $x_t$. This autocorrelation is commonly positive, and the serial correlation (ρ) is usually positive. Therefore, the second term in Equation 3.4.5 will very often be positive. The consequence is that the serial correlation in the errors usually results in larger standard errors, which are not accounted for by the usual estimator. In other words, the usual estimator underestimates the standard errors. Note though, if the autocorrelation in the independent variable is 0, the standard errors will not be misestimated even in the presence of serially correlated errors.

As an example, let us return to the Canadian public responsiveness model. Let us first check whether or not there is autocorrelation in the independent variable $P_t$ (preferences). Specifically, let us check whether or not it is an autoregressive AR(1) process. This is achieved by regressing $P_t$ on a lag of itself, $P_{t-1}$. As we previously noted that the series might contain a trend, we use detrended data for our test. Accordingly, the regression does not include a constant. This was partialled out in the detrending. The regression results are presented in Table 3.9.

**Table 3.9 Canadian Social Program Spending as an AR(1) Process**

| $P_t$ | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| $P_{t-1}$ | 0.76 | 0.17 | 4.46 | 0.001 |

NOTE: $R^2 = 0.59$, $T = 15$; AR = autoregressive, $T$ = number of time points.

Looking at Table 3.9, we can see that $P_t$ does contain autocorrelation. The coefficient on $P_{t-1}$ is statistically significant. This means that if the residuals in our public responsiveness model contained serial correlation, the usual estimate of the standard errors of our model coefficients will be incorrect—quite possibly too small. Given this, we want to be able to test for whether the estimated errors from our data model (Table 3.3 for our relative preference model) are serially correlated or not. If we believe that the potential serial correlation is AR(1), we might proceed by testing the null hypothesis, $H_0$: $\rho = 0$, in $\mu_t = \rho\mu_{t-1} + \varepsilon_t$, where $\mu_t$ is the error from our data model and it is assumed that $\varepsilon_t \sim \text{NID}(0, \sigma^2_\varepsilon)$.

If we can assume that we have only strictly exogenous independent variables in the data model, the test is very straightforward. We simply calculate and regress the residuals from the estimated data model on the lagged residuals and use a *t* test to test the significance of ρ. We can test for AR(*q*) serial correlation in the same basic manner as for AR(1). AR(1) errors are an example of serial correlation, but not all serial correlation is of this form. For example, $\mu_t$ could be a function of $\mu_{t-2}$ and $\mu_{t-1}$. This would be an AR(2) process:

$$\mu_t = \rho_1 \mu_{t-1} + \rho_2 \mu_{t-2} + \varepsilon_t. \quad (3.4.6)$$

We will discuss higher-order autoregressive processes further in Chapter 5. To test for higher-order serial correlation, we can include *q* lags of the residuals in the test regression and test for joint significance of the coefficients on the lags of the residuals using an *F* test (Agresti & Finlay, 2009, chap. 11). However, this approach will only be valid for a relatively large sample size *T*, because these regressions include one or more lags of the dependent variable ($\mu_t$), which you will remember from Chapter 2 can never be strictly exogenous. We can have contemporaneous exogeneity, $E(\varepsilon_t|\mu_{t-1}) = 0$, but contemporaneous exogeneity is only sufficient with a large sample size, *T.*

Another test of AR(1) serial correlation is the Durbin-Watson statistic (Bryman, Liao, & Lewis-Beck, 2004; Ostrom, 1990; Durbin & Watson, 1950). The logic of this test is the same as regressing the residuals on a lag of themselves and calculating a *t* statistic for the resulting parameter. The validity of the Durbin-Watson statistic does not require a large sample size. However, it does still assume strictly exogenous independent variables, and the Durbin-Watson test can produce inconclusive results. Consequently, this test is rarely used anymore.

Alternatively, we can use Durbin's alternative test, also known as Durbin's *h* (Durbin, 1970). The logic of this test is conceptually equivalent to regressing the residuals on the lagged residuals and all of the independent variables, $X_t$s. The inclusion of the $X_t$s controls for the possibility that an included dependent variable is correlated with $\mu_{t-1}$, so we don't need the assumption of strict exogeneity—contemporaneous endogeneity will do as long as we have a large *T.* Furthermore, the Durbin's *h* statistic is chi-squared distributed, and so we can calculate a *P* value to test the null hypothesis of no serial correlation, $H_0$: ρ = 0. Tests for serial correlation assume homoskedasticity of errors; however, Durbin's alternative test can be made robust to heteroskedasticity in most statistical software packages. The downside is that this test is valid asymptotically. Therefore, if we do have a relatively small *T*, the test is problematic. Increasingly, the next two tests we will discuss are being used in place of the Durbin-Watson or Durbin's alternative test.

The common practice is to use one of two tests. The first is the Portmanteau *(Q)* test to test the null hypothesis that the residuals form a white noise process (Ljung & Box, 1978). Such a process contains no serial correlation. If $\mu_t = \varepsilon_t$ and $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, $\mu_t$ is a white noise process. So we can apply the *Q* test to the residuals of our estimated data model as a test of serial correlation. This tests that the error autocorrelations are jointly zero, based on the first *p* autocorrelations, as described in Chapter 2.

$$Q = T(T+2)\sum_{s=1}^{p}\frac{\hat{\rho}_s^2}{T-s},$$

$$H_0: \rho_1, \rho_2, \ldots, \rho_p = 0.$$

Recall that we specify $p$ when conducting the test. The $Q$ test tests for more than just autoregressive errors. The white noise test will also test for moving average errors. We have not yet learned about moving average errors, but we will do so in Chapter 4. For now, it suffices to say that it is another source of serial correlation. It should also be noted that we may reject the null hypothesis of a white noise process because of factors other than serial correlation, such as trending errors. We will also address this possibility in Chapter 4.

Importantly, the $Q$ test does not assume strictly exogenous independent variables and does not require a large $T$.[4]

The alternative method is to use a test called the Breusch-Godfrey Lagrange Multiplier test (Breusch, 1978; Godfrey, 1978; Wooldridge, 1991, 2006). This tests the null hypothesis of no serial correlation against the alternatives of autoregressive errors up to an order of $q$ and/or moving average errors up to an order of $q$. We specify $q$ when conducting the test. This test also does not require strictly exogenous independent variables. The Breusch-Godfrey test uses the $R^2$ value from the regression of the residuals on $q$ lags of themselves and the independent variables. The test statistic is computed as $TR^2$. It is chi-squared distributed with degrees of freedom equal to $q$. The Breusch-Godfrey test will have somewhat more power to reject the null hypothesis than the $Q$ test when the null hypothesis is not true (Greene, 2003).

Returning to the Canadian public responsiveness model, we apply the $Q$ test to the residuals. The $Q$ statistic is 15.83 and is chi-squared distributed with degrees of freedom equal to $p$. In this instance, $p$ = 6. We can calculate the corresponding $P$ value for the test: 0.015. The null hypothesis is that the errors are a white noise process. The $Q$ statistic indicates that we can reject the null hypothesis of the errors being a white noise process, suggesting that we may have serial correlation.

We next apply the Breusch-Godfrey test, testing for serial correlation up to an order of 1, then 2, then 3, and all the way up to 6. The results are presented in Table 3.10.

**Table 3.10 Breusch-Godfrey Lagrange Multiplier Test for Autocorrelation—Canada**

| *lags(p)* | $\chi^2$ | *P Value* |
|---|---|---|
| 1 | 1.76 | 0.184 |
| 2 | 2.08 | 0.354 |
| 3 | 5.14 | 0.162 |
| 4 | 8.98 | 0.062 |
| 5 | 10.73 | 0.057 |

| | | |
|---|---|---|
| 6 | 13.78 | 0.032 |
| NOTE: $\chi^2$ =chi-squared statistic. | | |

We can reject the null hypothesis of no serial correlation in up to six lags of the errors. Note that this test does not test serial correlation at lag *q* controlling for serial correlation at lower lags. This means, for example, that if the residuals are AR(1) serially correlated and we test for second-order serial correlation, we will most likely find that we can reject the null hypothesis of no second-order serial correlation. This is a bit of a problem as it can make it a challenge to figure out at what order the serial correlation exists. In Chapter 5, we will look at how to address this challenge using autocorrelation and partial autocorrelation functions.

As a general rule, it is good practice to test residuals for serial correlation by testing the residuals against the null of a white noise process using the *Q* test. If higher-order serial correlation of a specific order is suspected, the Breusch-Godfrey test can be used. Higher-order serial correlation of this sort may be present in seasonal data; for example, quarterly data may exhibit 4th-order serial correlation, and monthly data may exhibit 12th-order serial correlation.

In the next section we look at how to correct for serial correlation but first let us look at how to test for heteroskedasticity. As in the cross- sectional case, heteroskedasticity does not cause bias or inconsistency. It may, however, invalidate the standard errors (and therefore the *t* statistics and *F* statistics). To test for heteroskedasticity, we can use the Breusch-Pagan test, which involves estimating $\hat{\varepsilon}_t^2 = \delta_0 + \delta_1 x_{1,t} + \ldots + \delta_k x_{k,t} + \delta_k x_{k+1,t} + v_t$ and testing $H_0$: $\delta_1 = \delta_2 = \ldots = \delta_{k+1} = 0$ with an *F* statistic (Wooldridge, 2006, pp. 278–281).[5] In other words, we regress the squared residuals on our independent variables and test the null hypothesis that the resulting slope coefficients are jointly equal to 0.[6] The Breusch-Pagan test assumes that the $\hat{\varepsilon}_t$ values are not serially correlated, so serial correlation should be accounted for first. We will see how to do that in the next section of this chapter. The Breusch-Pagan test also assumes that exogeneity is not violated and the functional form of the model is correct (e.g., the relationships that are assumed to be linear really are linear).

We now apply the Breusch-Pagan test to the U.K. public responsiveness model from Table 3.4 to test for heteroskedasticity with respect to the government spending variable. The test statistic is 3.80 and is chi-squared distributed with 3 degrees of freedom (because we are testing for heteroskedasticity with respect to three variables). The *P* value is 0.28. The null hypothesis is that the errors are homoskedastic. The Breusch-Pagan test indicates that we cannot reject the null hypothesis of constant variance (homoskedasticity).

Finally, we can also test for the skewness and kurtosis of the errors to determine if there is evidence that they are not normally distributed. This is a necessary assumption for the usual methods of inference (again, *t* and *F* statistics) when we have a small number of time points on which to estimate our models.

Again, testing the estimated errors from the U.K. public responsiveness model from Table 3.4, the *P* value for the null hypothesis of no skewness is 0.48, and the *P* value for the null hypothesis of no kurtosis (relative to the normal) is 0.33. The results indicate that we cannot reject the null of normally distributed errors.

Note that there is no straightforward test for exogeneity. This is an assumption about the data-generating process that we never observe. We can only ever test the exogeneity assumption indirectly by estimating models with different specifications. For example, if we believe that we have an omitted-variable bias problem, we may estimate a model with that variable, or a proxy for it, included and observe the consequences.

## 3.5 Correcting for Serial Correlation and Heteroskedasticity

We now turn to what we would do if we believed that we have serial correlation. We motivate this by running a U.S. public responsiveness model, including a quadratic trend just as we did in the Canadian and U.K. models.

We next test for serial correlation of the errors by using the *Q* test to test if the residuals form a white noise process. The *Q* statistic is 38.36, with a *P* value of <0.001. We can reject the null hypothesis of a white noise process and suspect that we have serial correlation. We have a number of options available to us for the purpose of correcting the violation of this assumption. As with the tests of serial correlation, we maintain that all Gauss-Markov assumptions hold except no serial correlation. Let us assume that the data-generating process for the errors is an AR(1) process, so it is as follows:

$$y_t = \beta_0 + \beta_1 x_t + \mu_t,$$

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t. \qquad (3.5.1)$$

We proceed by estimating Equation 3.5.1 by maximum likelihood. This is unlike the models we have estimated so far, for which we have used OLS. Good introductions to the maximum likelihood approach to model estimation can be found in Fox (2008), Gill (2000), and King (1998).

The results for the maximum likelihood estimation of our U.S. public responsiveness model with AR(1) errors are presented in Table 3.12.

**Table 3.11 U.S. Public Responsiveness Model**

| *Preference* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Program spending | −0.23 | 0.058 | −3.93 | <0.001 |
| Counter | 0.67 | 0.37 | 1.79 | 0.084 |
| Counter$^2$ | 0.07 | 0.022 | 3.36 | 0.002 |
| Constant | 44.04 | 7.06 | 6.24 | <0.001 |

NOTE: $R^2 = 0.67$, $T = 33$; $T$ = number of time points.

**Table 3.12 Maximum Likelihood Estimation of the U.S. Public Responsiveness Model With AR(1) Errors**

| *Preferences* | *Coefficient* | *Standard Error* | *z Statistic* | *P Value* |
|---|---|---|---|---|
| Program spending | −0.17 | 0.087 | −1.98 | 0.047 |
| Counter | 0.65 | 0.89 | 0.73 | 0.47 |
| Counter$^2$ | 0.059 | 0.43 | 1.38 | 0.167 |
| Constant | 37.03 | 10.97 | 3.37 | 0.001 |
| $\rho$ | 0.66 | 0.18 | 5.10 | <0.001 |

NOTE: Log likelihood = −91.352, $T = 33$; $T$ = number of time points, AR = autoregressive.

The results include the estimated coefficient of the AR(1) serial correlation. This parameter is statistically significant at the 0.05 significance level. 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.

Another option for accounting for serial correlation is to calculate serial correlation robust standard errors. The idea is to scale the standard errors from an OLS estimation to take serial correlation into account. Newey-West standard errors correct the standard errors estimated by OLS regression (Newey & West, 1987). The error structure is assumed to be heteroskedastic and possibly autocorrelated up to some lag.[7] Most statistical software packages have a default procedure to determine how many lags to correct for, accounting for the number of time points in the data. Estimating Newey-West standard errors that correct for first-order serial correlation in the results presented in Table 3.11 produces the results shown in Table 3.13.

**Table 3.13 Newey-West Standard Errors for the U.S. Public Responsiveness Model**

| *Preference* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Spending | –0.23 | 0.062 | –3.67 | <0.001 |
| Counter | 0.67 | 0.45 | 1.50 | 0.144 |
| Counter$^2$ | 0.074 | 0.021 | 3.61 | <0.001 |
| Constant | 44.04 | 7.06 | 6.24 | <0.001 |

NOTE: $R^2 = 0.36$, $T = 33$; $T$ = number of time points.

Examining Table 3.13, note that the estimated coefficients are just those estimated by OLS (c.f. Table 3.11). It should be noted though that Newey-West standard errors can be poorly behaved with a low *T* (Wooldridge, 2006). Also, note that Newey-West standard errors correct for heteroskedasticity. If we are not using Newey-West standard errors and we believe we have homoskedasticity, we must calculate heteroskedasticity robust standard errors. Heteroskedastic robust errors are easily computed by most statistical packages.

Let us finish this chapter with an example that puts together everything we have learned in this chapter. In this example, we examine the relative legislative success of minority and majority parliamentary governments. We will examine the Canadian federal parliament. Legislative success is the ability of a government to pass the legislation that it deems important. The expectations are that the legislative success of a minority government will generally be lower than that of a majority government and the legislative success of a minority government will increase as its popularity in the polls published in the media increases.

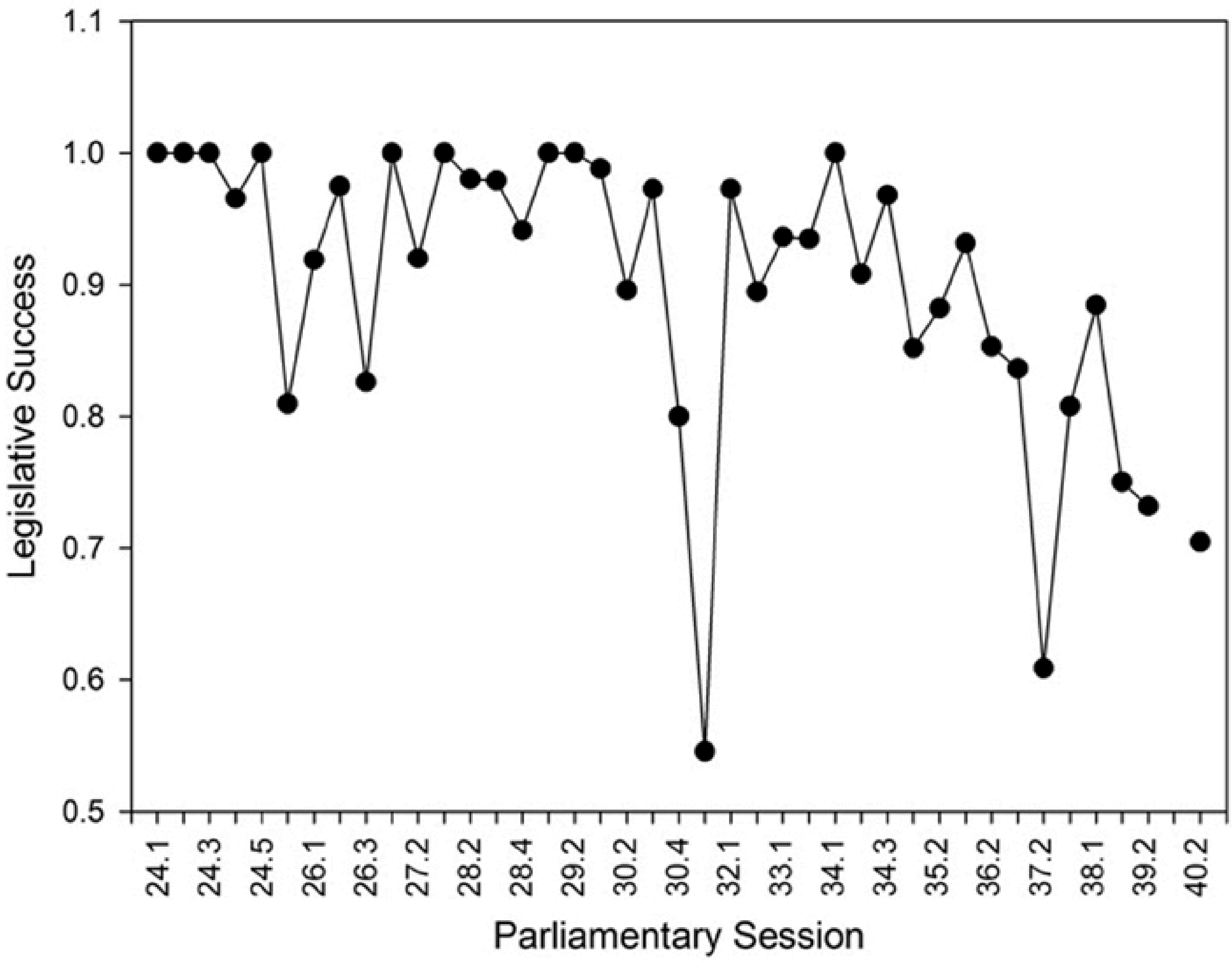
The measure of legislative success used is the proportion of bills the government moved past second reading in the House of Commons that received Royal Assent in each session. Bills that do not pass second reading are not considered.[8] The measure of the popularity of the government is based on all published results from polls asking respondents their vote intention. A typical vote intention question is "If a federal election were held tomorrow, which party would you vote for?" The measure of popularity is the average share of vote intention the governing party received in all polls published over each session.

As for whether a government is a minority or a majority government, a minority government is simply defined as any government in which the governing party has less than 50% + 1 of the seats in the House of Commons. The minority variable is coded "1" for minority governments and "0" for majority governments. We also control for whether the governing party was 1 = *the Liberals* or 0 = *the (Progressive) Conservatives.*

Our data include these measures for the period extending from the beginning of the 24th to the end of the

40th Parliament. This includes 42 sessions of Parliament and spans the temporal period 1958 to 2008. The temporal unit of analysis is the parliamentary session. Using these measures and a "government popularity × minority" interaction variable, we use OLS regression to estimate a model of legislative success testing the effects of minority versus majority government, government popularity, and their interaction and a control for the party in government. Before doing so, we plot the dependent variable, "legislative success" (Figure 3.7).

**Figure 3.7 Legislative Success of the Canadian Federal Government**




The plot of legislative success does not reveal any periodicity or structural breaks. If it did contain either of these processes, they would be a violation of stationarity, and we would have to control for them. There are possibly two outliers. One of which (the 31st Parliament) was a result of a quickly defeated minority government. The inclusion of a minority variable in the model may account for this outlier. The other outlier (the third session of the 37th Parliament) reflects the last session of a long-standing prime minister (Jean Chrétien) and a high degree of infighting within the governing Liberal party over whether or not Chrétien should resign. These outliers might cause issues with homoskedasticity and with the distribution of the errors.

It is also possible that there is a systematic decline in legislative success in the later parliamentary sessions. This would be a nonlinear trend that could be modelled with a quadratic function of time. It is possible that this is just an end effect; that is, there appears to be a trend at the end, but if we were able to continue observing this series further in time, we would see that it is just a temporary downward shift (Tufte, 2001). To test the possibility that the series exhibits a nonlinear trend, we regress legislative success on time and time-squared variables. The results (Table 3.14) suggest that no such trending exists as the coefficients on both variables are not statistically significant at the 0.05 significance level.

**Table 3.14 Test for Trending in Legislative Success**

| *Legislative Success* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Trend | 0.0026 | 0.0051 | 0.50 | 0.622 |
| $t^2$ | –0.00019 | 0.00012 | –1.56 | 0.128 |
| Constant | 0.95 | 0.046 | 20.69 | <0.001 |

NOTE: $R^2 = 0.34$, $T = 40$; $T$ = number of time points.

Satisfied that our results are not going to be biased by trending, periodicity, or structural breaks, we estimate our legislative success model. The results are presented in Table 3.15.

**Table 3.15 Canadian Legislative Success Model**

| *Legislative Success* | *Coefficient* | *Standard Error* | *t Statistic* | *P Value* |
|---|---|---|---|---|
| Governing party | –0.02 | 0.0384 | –0.52 | 0.603 |
| Government popularity | –0.00016 | 0.0021 | –0.08 | 0.939 |
| Minority government | –0.74 | 0.21 | –3.46 | 0.001 |
| Govpop × Min | 0.017 | 0.0052 | 3.25 | 0.003 |
| Constant | 0.94 | 0.089 | 10.56 | <0.001 |

NOTE: $R^2 = 0.30$, $T = 40$; $T$ = number of time points.

The estimated coefficient for government popularity is the effect of popularity on the legislative success of majority governments. It is not statistically significant at a 0.05 significance level. This suggests that the legislative success of majority governments is not affected by their standing in the polls. The effect of government popularity on the success of minority governments is the sum of the coefficients on government popularity and the government popularity × minority interaction. The sum of these coefficients is 0.017, and an $F$ test of the hypothesis that the sum of these coefficients is 0 gives us a statistic of $F(1, 35) = 10.57$. The corresponding $P$ value is 0.0025. We can reject the null hypothesis that the sum of the coefficients is 0. It would appear that the legislative success of a minority government is improved as its popularity in the polls increases. However, as this chapter has discussed, these results are contingent on certain assumptions.

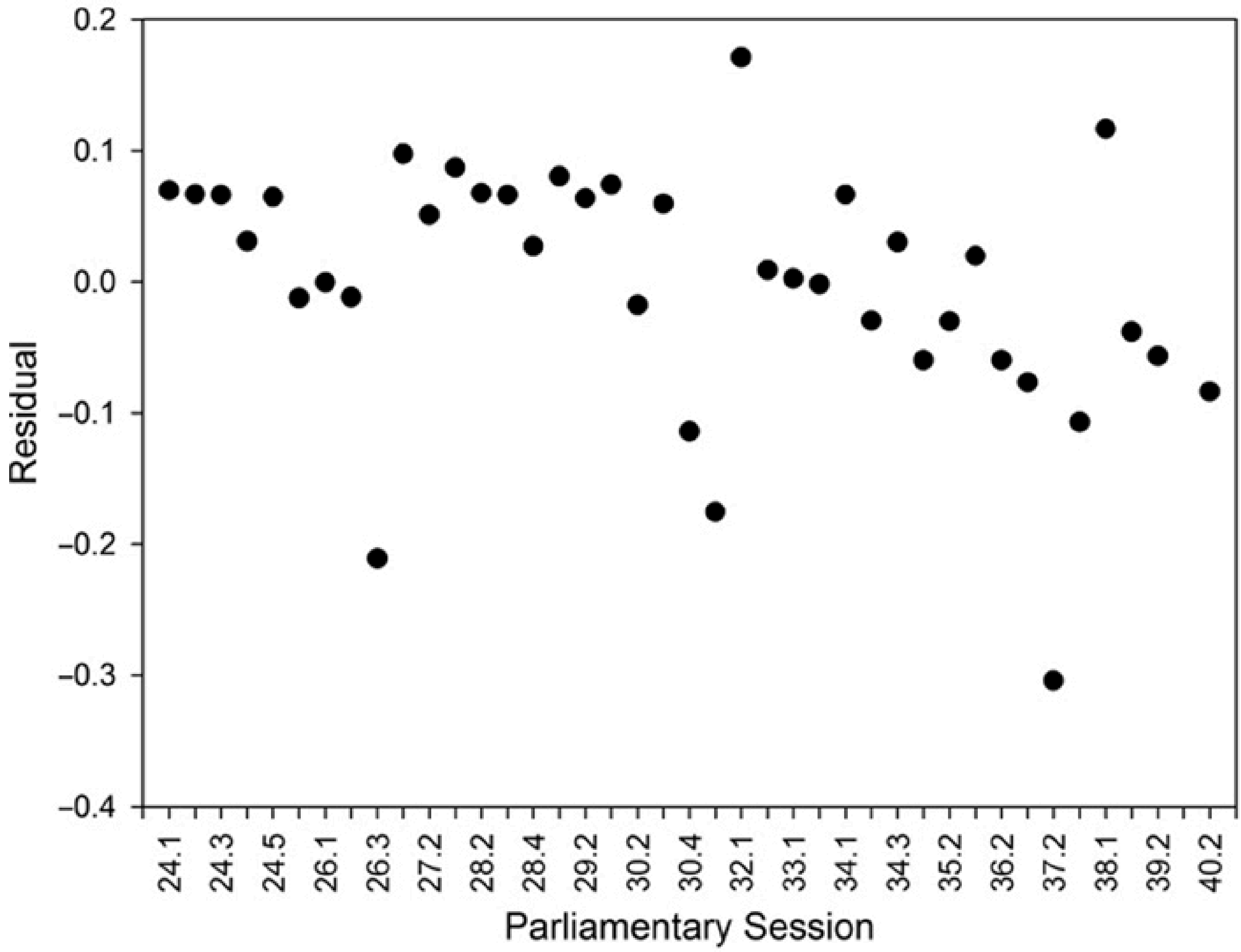
At this point, we might want to test the residuals for serial correlation. We do this here by conducting an overall test of white noise on the residuals. We use the $Q$ test for white noise. It gives us a $Q$ statistic of 14.63. This is chi-squared distributed with 18 degrees of freedom and a $P$ value of 0.69. We cannot reject the null hypothesis that the residuals are a white noise process.

Next, we might want to test for heteroskedasticity. The presence of heteroskedasticity would not bias the coefficients but would result in the estimation of incorrect standard errors. We test for heteroskedasticity with respect to the minority, government popularity, and governing party variables. The Breusch-Pagan test for heteroskedasticity gives us a statistic of 3.40, which is chi-squared distributed with 4 degrees of freedom. The corresponding $P$ value is 0.49, which means that we cannot reject the null hypothesis of homoskedasticity.

Finally, we can test whether the residuals deviate from a normal distribution. Again, this is not a necessary condition for unbiased coefficients, but it is necessary for the use of $t$ and $F$ statistics as tests of inference when we have a small number of time points. The tests of the null hypothesis of no skewness and of the null hypothesis of no kurtosis (relative to the normal) have $P$ values of 0.004 and 0.032, respectively. This suggests that we can reject the null hypothesis of normally distributed residuals. With 40 time points, we may not be concerned about the deviations from normality. If we were, we would want to investigate the source of the violation.

Regardless of the outcomes of these tests, it is always a good idea to plot the residuals (Figure 3.8). If we observe any systematic pattern in the plot of the residuals, it could be an indication of a violation of an important assumption, such as no trending, no serial correlation, or normally distributed errors. In this way, the residuals plot can be a good visual confirmation of the statistical tests.

**Figure 3.8 Residuals From the Canadian Legislative Success Model**




The plot reveals that the primary reason why the residuals are not normally distributed is the large negative residual for the previously identified third session of the 37th Parliament. At this point, we might want to think about whether we have excluded an important variable that might explain outliers (e.g., the level of infighting within the governing party) and/or test whether or not these outliers have much leverage on the results. Procedures to do this are exactly as they are in a cross-sectional analysis (e.g., see Andersen, 2008).

## Summary

In this chapter, you have been introduced to static and FDL models, and the assumptions necessary to estimate them. In the next chapter, we move on to dynamic time series models and discuss the complications these introduce for unbiased estimation.

[1] A more advanced method for controlling for complex time trends is the use of regression splines (Durlauf &

Blume, 2010; Keele, 2008).

[2] For a overview of the *F* test, see Foster and Christian (2008).

[3] This is based on the following approximation, appropriate when the log of $y_t$ is linearly related to $x_t$: $\%\Delta y_t \approx (100ß_1)\Delta x_t$.

[4] The *Q* test is valid asymptotically $(T \rightarrow \infty)$ but it has good small *T* properties (Harvey, 1993).

[5] Alternatively, a Lagrange multiplier statistic could be calculated (Koenker, 1981).

[6] The squared errors are proportional to the error variances.

[7] See Wooldridge (2012) for a straightforward exposition.

[8] Bills were selected in this way because governing parties will sometimes introduce bills that they do not intend to pursue. There are a number of reasons why a governing party may do this. However, if the government pursues a bill past second reading in the House of Commons, one can be relatively certain that it is serious about passing it into law.

http://dx.doi.org/10.4135/9781483390857.n3