Chapter 12

Autocorrelation

12.1 The Nature of the Problem

- The randomness of the sample implies that the error terms for different observations (households or firms) will be uncorrelated.
- When we have time-series data, where the observations follow a natural ordering through time, there is always a possibility that successive errors will be correlated with each other.
- In any one period, the current error term contains not only the effects of current shocks but also the carryover from previous shocks. This carryover will be related to, or *correlated with*, the effects of the earlier shocks. When circumstances such as these lead to error terms that are correlated, we say that *autocorrelation* exists.

- The possibility of autocorrelation should always be entertained when we are dealing with time-series data.
- Suppose we have a linear regression model with two explanatory variables. That is,

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t \qquad (12.1.1)$$

- The error term assumptions utilized in Chapters 3 through 9 are

$$E(e_t) = 0 \qquad \text{var}(e_t) = \sigma^2 \qquad (12.1.2a)$$

$$\text{cov}(e_t, e_s) = 0 \quad \text{for } t \neq s \qquad (12.1.2b)$$

- When (12.1.2*b*) does not hold, we say that the random errors e_t are autocorrelated.

12.1.1 An Area Response Model for Sugar Cane

- Letting *A* denote area planted, and *P* denote output price, and assuming a log-log (constant elasticity) functional form, an area response model of this type can be written as

$$\ln(A) = \beta_1 + \beta_2 \ln(P) \qquad (12.1.3)$$

- We use the model in (12.1.3) to explain the area of sugar cane planted in a region of the South-East Asian country of Bangladesh.
- The econometric model is

$$\ln(A_t) = \beta_1 + \beta_2 \ln(P_t) + e_t \qquad (12.1.4)$$

- We can write this equation as

$$y_t = \beta_1 + \beta_2 x_t + e_t \qquad (12.1.5)$$

where

$$y_t = \ln(A_t) \text{ and } x_t = \ln(P_t) \qquad (12.1.6)$$

12.1.1a Least Squares Estimation

- Application of least squares yields the following estimated equation

$$\hat{y}_t = 6.111 + 0.971\, x_t \quad R^2 = 0.706 \qquad \text{(R12.1)}$$

$$(0.169)(0.111) \qquad \text{(std. errors)}$$

- The least squares residuals appear in Table 12.2 and are plotted against time in Figure 12.1.

[Figure 12.1 Least squares residuals plotted against time]

- We can see that there is a tendency for negative residuals to follow negative residuals and for positive residuals to follow positive residuals. This kind of behavior is consistent with an assumption of positive correlation between successive residuals.

- With uncorrelated errors, we would not expect to see any particular pattern. If the errors are negatively autocorrelated, we would expect the residuals to show a tendency to oscillate in sign.

12.2 First-Order Autoregressive Errors

- If the assumption $\text{cov}(e_t, e_s) = 0$ is no longer valid, what alternative assumption can we use to replace it? Is there some way to describe how the e_t are correlated? If we are going to allow for autocorrelation then we need some way to represent it.
- The most common is model is a first-order autoregressive model or, more simply, an AR(1) model.

$$e_t = \rho e_{t-1} + v_t \tag{12.2.1}$$

$$E(v_t) = 0 \qquad \text{var}(v_t) = \sigma_v^2 \qquad \text{cov}(v_t, v_s) = 0 \;\; t \neq s \tag{12.2.2}$$

- The rationale for the AR(1) model is that the random component e_t in time period t is composed of two parts: (i) ρe_{t-1} is a carry over from the random error in the previous period (ii) v_t is a "new" shock to the level of the economic variable.
- The autoregressive model asserts that shocks to an economic variable do not work themselves out in one period.

12.2.1 Properties of an AR(1) Error

- Assume

$$-1 < \rho < 1 \qquad (12.2.3)$$

It can be shown that

$$E(e_t) = 0 \qquad (12.2.4)$$

$$\text{var}(e_t) = \sigma_e^2 = \frac{\sigma_v^2}{1-\rho^2} \tag{12.2.5}$$

Because σ_e^2 does not change over time, the error e_t is also homoskedastic.

$$\text{cov}\left(e_t, e_{t-k}\right) = \sigma_e^2 \rho^k \qquad k > 0 \tag{12.2.6}$$

The error correlation

$$\text{corr}(e_t, e_{t-k}) = \frac{\text{cov}(e_t, e_{t-k})}{\sqrt{\text{var}(e_t)\,\text{var}\left(e_{t-k}\right)}}$$

$$= \frac{\sigma_e^2 \rho^k}{\sqrt{\sigma_e^2 \sigma_e^2}} = \rho^k \tag{12.2.7}$$

- ρ is the correlation between two errors that are one period apart; it is sometimes called the autocorrelation coefficient.

12.3 Consequences for the Least Squares Estimator

- If we have an equation whose errors exhibit autocorrelation, but we ignore it, or are simply unaware of it, what does it have on the properties of least squares estimates?

1. The least squares estimator is still a linear unbiased estimator, but it is no longer best.
2. The formulas for the standard errors usually computed for the least squares estimator are no longer correct, and hence confidence intervals and hypothesis tests that use these standard errors may be misleading.

Proofs:

- For the simple regression model $y_t = \beta_1 + \beta_2 x_t + e_t$, we wrote the least squares estimator for β_2 as

$$b_2 = \beta_2 + \sum w_t e_t \tag{12.3.1}$$

where

$$w_t = \frac{\left(x_t - \overline{x}\right)}{\sum \left(x_t - \overline{x}\right)^2} \tag{12.3.2}$$

- We prove b_2 is still an unbiased estimator for β_2 under autocorrelation by showing that

$$E(b_2) = \beta_2 + \sum w_t E(e_t) = \beta_2 \tag{12.3.3}$$

- For the variance of b_2 we have

$$\text{var}(b_2) = \sum w_t^2 \,\text{var}\left(e_t\right) + \underset{i \neq j}{\Sigma\Sigma} w_i w_j \,\text{cov}(e_i, e_j)$$

$$= \sigma_e^2 \sum w_t^2 + \sigma_e^2 \sum\sum_{i \neq j} w_i w_j \rho^k \qquad \text{(where } k = |i-j|\text{)}$$

$$= \frac{\sigma_e^2}{\sum (x_t - \bar{x})^2} \left(1 + \frac{1}{\sum (x_t - \bar{x})^2} \right) \sum\sum_{i \neq j} (x_i - \bar{x})(x_j - \bar{x}) \rho^k$$

(12.3.4)

- When we were proving that $\text{var}(b_2) = \sigma_e^2 / \sum (x_t - \bar{x})^2$ in the absence of autocorrelation, the terms $\text{cov}(e_i, e_j)$ were all zero. This simplification no longer holds, however.

- Return to least squares estimation of the sugar cane example.
- Given estimates for ρ and σ_e^2, it is possible to use a computer to calculate an estimate for var(b_2) from equation 12.3.4. A similar estimate for var(b_1) can also be obtained.
- Suppose that we have estimates of ρ and σ_e^2, and that we have used them to estimate var(b_1) and var(b_2).
- The square roots of these quantities we can call *correct* standard errors, while those we calculated with our least squares estimates and reported in equation 12.1.7 we call *incorrect*. The two sets of standard errors, along with the estimated equation are:

$$\hat{y}_t = 6.111 + 0.971\, x_t$$

$$(0.169)(0.111) \quad \text{"incorrect" s.e.'s} \qquad (R12.2)$$

$$(0.226)(0.147) \quad \text{"correct" s.e.'s}$$

- Note that the correct standard errors are larger than the incorrect ones.

- If we ignored the autocorrelation, we would tend to overstate the reliability of the least squares estimates. The confidence intervals would be narrower than they should be. For example, using $t_c = 2.037$, we find the following 95% confidence interval for β_2:

  For β_2: (0.745, 1.197) (incorrect)

  (0.672, 1.269) (correct)

12.4 Generalized Least Squares

12.4.1A Transformation

- Our objective is to transform the model in equation 12.1.5

$$y_t = \beta_1 + \beta_2 x_t + e_t \tag{12.4.1}$$

- The relationship between e_t and v_t is given by

$$e_t = \rho e_{t-1} + v_t \tag{12.4.2}$$

Substituting (12.4.2) into (12.4.1) yields

$$y_t = \beta_1 + \beta_2 x_t + \rho e_{t-1} + v_t \tag{12.4.3}$$

- To substitute out e_{t-1}, we note that (12.4.1) holds for every single observation.

$$e_{t-1} = y_{t-1} - \beta_1 - \beta_2 x_{t-1} \tag{12.4.4}$$

- Multiplying (12.4.4) by ρ yields

$$\rho e_{t-1} = \rho y_{t-1} - \rho\beta_1 - \rho\beta_2 x_{t-1} \qquad (12.4.5)$$

- Substituting (12.4.5) into (12.4.3) yields

$$y_t = \beta_1 + \beta_2 x_t + \rho y_{t-1} - \rho\beta_1 - \rho\beta_2 x_{t-1} + v_t$$

or, after rearranging,

$$y_t - \rho y_{t-1} = \beta_1(1-\rho) + \beta_2(x_t - \rho x_{t-1}) + v_t \qquad (12.4.6)$$

- The transformed dependent variable is

$$y_t^* = y_t - \rho y_{t-1} \qquad t = 2,3,...,T \qquad (12.4.7a)$$

- The transformed explanatory variable is

$$x_{t2}^* = x_t - \rho x_{t-1} \qquad t = 2,3,...,T \qquad (12.4.7b)$$

- The new constant term is

$$x_{t1}^{*} = 1 - \rho \qquad t = 2,3,...,T \tag{12.4.7c}$$

- Making these substitutions we have

$$y_t^{*} = x_{t1}^{*}\beta_1 + x_{t2}^{*}\beta_2 + v_t \tag{12.4.8}$$

- Thus we have formed a new statistical model with transformed variables y_t^{*}, x_{t1}^{*} and x_{t2}^{*} and, *importantly*, with an error term that is *not* the correlated e_t, but the uncorrelated v_t that we have assumed to be distributed $(0, \sigma_v^2)$.

- There are two additional problems that we need to solve, however:

1. Because lagged values of y_t and x_t had to be formed, only $(T-1)$ new observations were created by the transformation in (12.4.7). We have values $(y_t^{*}, x_{t1}^{*}, x_{t2}^{*})$ for t = 2,3,...,T. But, we have no $(y_1^{*}, x_{11}^{*}, x_{12}^{*})$.

2. The value of the autoregressive parameter ρ.

12.4.1a Transforming the First Observation

The first observation in the regression model is

$$y_1 = \beta_1 + x_1\beta_2 + e_1 \tag{12.4.9}$$

with error variance $\text{var}(e_1) = \sigma_e^2 = \sigma_v^2/(1-\rho^2)$.

- The transformation that yields an error variance of σ_v^2 is multiplication by $\sqrt{1-\rho^2}$. The result is

$$\sqrt{1-\rho^2}\,y_1 = \sqrt{1-\rho^2}\,\beta_1 + \sqrt{1-\rho^2}\,x_1\beta_2 + \sqrt{1-\rho^2}\,e_1 \tag{12.4.10}$$

or

$$y_1^* = x_{11}^*\beta_1 + x_{12}^*\beta_2 + e_1^* \tag{12.4.11a}$$

where

$$y_1^* = \sqrt{1-\rho^2}\,y_1 \quad x_{11}^* = \sqrt{1-\rho^2}$$
$$x_{12}^* = \sqrt{1-\rho^2}\,x_1 \quad e_1^* = \sqrt{1-\rho^2}\,e_1 \qquad (12.4.11b)$$

- Note that

$$\text{var}(e_1^*) = (1-\rho^2)\,\text{var}(e_1) = (1-\rho^2)\frac{\sigma_v^2}{1-\rho^2} = \sigma_v^2$$

Remark: We can summarize these results by saying that, *providing* ρ *is known*, we can find the best linear unbiased estimator for β_1 and β_2 by applying least squares to the transformed model

$$y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + v_t \qquad (12.4.12)$$

where the transformed variables are defined by

$$y_1^* = \sqrt{1-\rho^2}\, y_1 \qquad x_{11}^* = \sqrt{1-\rho^2} \qquad x_{12}^* = \sqrt{1-\rho^2}\, x_1$$

for the first observation, and

$$y_t^* = y_t - \rho y_{t-1} \qquad x_{t1}^* = 1-\rho \qquad x_{t2}^* = x_t - \rho x_{t-1}$$

for the remaining $t = 2,3,...,T$ observations.

12.5 Implementing Generalized Least Squares

The remaining problem is the fact that the transformed variables y_t^*, x_{t1}^* and x_{t2}^* cannot be calculated without knowledge of the parameter ρ

Consider the equation

$$e_t = \rho e_{t-1} + v_t \tag{12.5.1}$$

- If the e_t values were observable, we could treat this equation as a linear regression model and estimate ρ by least squares.
- However, the e_t are not observable because they depend on the unknown parameters β_1 and β_2 through the equation

$$e_t = y_t - \beta_1 - \beta_2 x_t \tag{12.5.2}$$

- As an approximation to the e_t we use instead the least squares residuals

$$\hat{e}_t = y_t - b_1 - b_2 x_t \tag{12.5.3}$$

where b_1 and b_2 are the least squares estimates from the untransformed model.

- Substituting the $\hat{e}_t$ for the e_t in (12.5.1) is justified providing the sample size T is large. Making this substitution yields the model

$$\hat{e}_t = \rho e_{t-1} + v_t \tag{12.5.4}$$

- The least squares estimator of ρ from (12.5.4) has good statistical properties if the sample size T is large; it is given by

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{T} \hat{e}_{t-1}^2} \tag{12.5.5}$$

12.5.1 The Sugar Cane Example Revisited

- We obtain

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{T} \hat{e}_{t-1}^2} = 0.342 \qquad (11.5.6)$$

- As examples, note that

$$\begin{aligned} y_1^* &= \sqrt{1-\hat{\rho}^2}\, y_1 \\ &= \sqrt{1-0.342^2}\,(3.3673) \\ &= 3.1642 \end{aligned} \qquad (R12.5)$$

and

$$x_{32}^{*} = x_{32} - \hat{\rho} x_{22}$$

$$= -2.2919 - (0.342)(-2.1637)$$

$$= -1.5519 \qquad \text{(R12.6)}$$

- Applying least squares to all transformed observations yields the generalized least squares estimated model

$$\ln(\hat{A}_t) = 6.164 + 1.007 \ln(P_t)$$

$$(0.213) \quad (0.137) \qquad \text{(R12.7)}$$

12.6 Testing for Autocorrelation

- Looking for runs in the least squares residuals gives some indication of whether autocorrelation is likely to be a problem.
- The Durbin-Watson test is by far the most important one for detecting AR(1) errors.
- Consider again the linear regression model

$$y_t = \beta_1 + \beta_2 x_t + e_t \tag{12.6.1}$$

where the errors may follow the first-order autoregressive model

$$e_t = \rho e_{t-1} + v_t \tag{12.6.2}$$

- It is assumed that the v_t are independent random errors with distribution $N(0, \sigma_v^2)$. The assumption of *normally* distributed random errors is needed to derive the probability distribution of the test statistic used in the Durbin-Watson test.

- For a null hypothesis of no autocorrelation, we can use H_0: $\rho = 0$. For an alternative hypothesis we could use H_1: $\rho > 0$ or H_1: $\rho < 0$ or H_1: $\rho \neq 0$.
- We choose H_1: $\rho > 0$; in most empirical applications in economics, positive autocorrelation is the most likely form that autocorrelation will take.
- Thus, we consider testing

$$H_0: \ \rho = 0 \qquad \text{against} \qquad H_1: \ \rho > 0 \tag{12.6.3}$$

- The DW statistic is

$$d = \frac{\sum_{t=2}^{T} \left(\hat{e}_t - \hat{e}_{t-1}\right)^2}{\sum_{t=1}^{T} \hat{e}_t^2} \tag{12.6.4}$$

where the $\hat{e}_t$ are the least squares residuals $\hat{e}_t = y_t - b_1 - b_2 x_t$.

- To see why d is closely related to $\hat{\rho}$ expand (12.6.4) as

$$d = \frac{\sum_{t=2}^{T} \hat{e}_t^2 + \sum_{t=2}^{T} e_{t-1}^2 - 2\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2}$$

$$= \frac{\sum_{t=2}^{T} \hat{e}_t^2}{\sum_{t=1}^{T} \hat{e}_t^2} + \frac{\sum_{t=2}^{T} e_{t-1}^2}{\sum_{t=1}^{T} e_t^2} - 2\frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2}$$

$$\approx 1 + 1 - 2\hat{\rho} \qquad (12.6.5)$$

- Thus, we have

$$d \approx 2\left(1 - \hat{\rho}\right) \qquad (12.6.6)$$

- If $\hat{\rho} = 0$, then the Durbin-Watson statistic $d \approx 2$, which is taken as an indication that the model errors are not autocorrelated.

- If $\hat{\rho} = 1$ then $d \approx 0$, and thus a low value for the Durbin-Watson statistic implies that the model errors are correlated, and $\rho > 0$.
- What is a critical value d_C such that we reject H_0 when

$$d \leq d_c$$

- Determination of a critical value and a rejection region for the test requires knowledge of the probability distribution of the test statistic under the assumption that the null hypothesis, H_0: $\rho = 0$, is true.
- If a 5% significance level is required find d_C such that $P(d \leq d_c) = 0.05$.
- Then, as illustrated in Figure 12.2, we reject H_0 if $d \leq d_C$ and fail to reject H_0 if $d > d_c$.
- For this one-tail test, the *p*-value is given by the area under $f(d)$ to the left of the calculated value of *d*. Thus, if the *p*-value is less than or equal to 0.05, it follows that $d \leq d_C$ and H_0 is rejected. If the *p*-value is greater than 0.05, then $d > d_c$, and H_0 is accepted.

[Insert Figure 12.2 here]

- A difficulty associated with $f(d)$, and one that we have not previously encountered when using other test statistics, is that this probability distribution depends on the values of the explanatory variables. It is impossible to tabulate critical values that can be used for every possible problem.
- There are two ways to overcome this problem. The first way is to use software (SHAZAM is an example) that computes the *p*-value for the explanatory variables in the model under consideration. Instead of comparing the calculated *d* value with some tabulated values of d_c, we get our computer to calculate the *p*-value of the test. If this *p*-value is less than the specified significance level, $H_0 : \rho = 0$ is rejected and we conclude that autocorrelation does exist.

- In the sugar cane area response model the calculated value for the Durbin-Watson statistic is $d = 1.291$. Is this value sufficiently close to zero (or sufficiently less than 2), to reject H_0 and conclude that autocorrelation exists? Using SHAZAM, we find that

$$p\text{-value} = P(d \leq 1.291) = 0.0098$$

- This value is much less than a conventional 0.05 significance level; we conclude, therefore, that the equation's error is positively autocorrelated.

12.6.1a The Bounds Test

- In the absence of software that computes a *p*-value, a test known as the bounds test can be used to partially overcome the problem of not having general critical values. Durbin and Watson considered two other statistics d_L and d_U whose probability distributions do not depend on the explanatory variables and which have the property that

$$d_L < d < d_U \qquad (12.6.7)$$

- That is, irrespective of the explanatory variables in the model under consideration, d will be bounded by an upper bound d_U and a lower bound d_L. The relationship between the probability distributions $f(d_L)$, $f(d)$, and $f(d_U)$ is depicted in Figure 12.3.
- Let d_{Lc} be the 5% critical value from the probability distribution for d_L. Similarly, let d_{Uc} be such that $P(d_U < d_{Uc}) = .05$. Since the probability distributions $f(d_L)$ and $f(d_U)$ do not depend on the explanatory variables, it is possible to tabulate the critical values d_{Lc} and d_{Uc}. Table 5 at the end of this book.
  - In Figure 12.3 we have three critical values.
  - If the calculated value d is such that $d < d_{Lc}$, then it must follow that $d < d_c$, and H_0 is rejected.
  - If $d > d_{Uc}$, then it follows that $d > d_c$, and H_0 is accepted.
  - If $d_{Lc} < d < d_{Uc}$, then, we cannot be sure whether to accept or reject.

- These considerations led Durbin and Watson to suggest the following decision rules, which are known collectively as the Durbin-Watson *bounds test*.

If $d < d_{Lc}$, reject $H_0 : \rho = 0$ and accept $H_1 : \rho > 0$;

if $d > d_{Uc}$, do not reject $H_0 : \rho = 0$;

if $d_{Lc} < d < d_{Uc}$, the test is inconclusive.

To find the critical bounds for the sugar cane example we consult Table 5 at the end of the book for $T = 34$ and $K = 2$. The values are

$$d_{Lc} = 1.393 \quad d_{Uc} = 1.514$$

Since $d = 1.291 < d_{Lc}$, we conclude that $d < d_c$, and hence we reject H_0; there is evidence to suggest that autocorrelation exists.

12.6.2 A Lagrange Multiplier Test

- To introduce this test, return to equation (12.4.3) which was written as

$$y_t = \beta_1 + \beta_2 x_t + \rho e_{t-1} + v_t \qquad (12.6.8)$$

- If e_{t-1} was observable, an obvious way to test the null hypothesis $H_0 : \rho = 0$ would be to regress y_t on x_t and e_{t-1} and to use a *t*- or *F*-test to test the significance of the coefficient of e_{t-1}. Because e_{t-1} is not observable, we replace it by the lagged least squares residuals $\hat{e}_{t-1}$, and then perform the test in the usual way.
- Proceeding in this way for the sugar can example yields

$$t = 2.006 \qquad F = 4.022 \;\; p\text{-value} = 0.054$$

- Obtaining a *p*-value greater than 0.05 means that, at a 5% significance level, the LM test does not reject a null hypothesis of no autocorrelation. This test outcome conflicts with that obtained earlier using the Durbin-Watson test. Such conflicts are a fact of life.

You should note the following 4 points:

1. When estimating the regression in (12.6.8), using the first observation, $t = 1$, requires knowledge of $\hat{e}_0$. Two ways of overcoming this lack of knowledge are often employed. One is to set $e_0 = 0.$ The other is to omit the first observation. In our calculations we set $e_0 = 0.$ The results change very little if the first observation is omitted instead.
2. The Durbin-Watson test is an exact test valid in finite samples. The LM test is an approximate large-sample test, the approximation occurring because e_{t-1} is replaced by $\hat{e}_{t-1}$.

3. The Durbin-Watson test is not valid if one of the explanatory variables is the lagged dependent variable y_{t-1}. The LM test can still be used in these circumstances. This fact is particularly relevant for a distributed lag model studied in Chapter 15.
4. We have only been concerned with testing for autocorrelation involving one lagged error e_{t-1}. To test for more complicated autocorrelation structures, involving higher order lags such as e_{t-2}, e_{t-3}, etc, the LM test can be used by including the additional lagged errors in (12.6.8), and using an F test to test the relevance of their inclusion.

12.7 Prediction With AR(1) Errors

- For the problem of forecasting or predicting a future observation $y0$ that we assume

$$y_0 = \beta_1 + x_0\beta_2 + e_0 \qquad (12.7.1)$$

where x_0 is a given future value of an explanatory variable and e_0 is a future error term.

- When the errors are uncorrelated, the best linear unbiased predictor for y_0 is the least squares predictor

$$\hat{y}_0 = b_1 + b_2 x_0 \qquad (12.7.2)$$

- When the errors are autocorrelated, the generalized least squares estimators, denoted by $\hat{\beta}_1$ and $\hat{\beta}_2$, are more precise than their least squares counterparts b_1 and b_2. A better predictor is obtained, therefore, if we replace b_1 and b_2 by $\hat{\beta}_1$ and $\hat{\beta}_2$.

- When e_0 is correlated with past errors we can use information contained in the past errors to improve upon zero as a forecast for e_0.
- For example, if the last error e_T is positive, then it is likely that the next error e_{T+1} will also be positive.
  - When we are predicting one period into the future, the model with an AR(1) error can be written as

$$y_{T+1} = \beta_1 + \beta_2 x_{T+1} + e_{T+1}$$

$$= \beta_1 + \beta_2 x_{T+1} + \rho e_T + v_{T+1} \qquad (12.7.3)$$

where we have used $e_{T+1} = \rho e_T + v_{T+1}$.

Equation 12.7.3 has three distinct components:

1. Given the explanatory variable x_{T+1}, the best linear unbiased predictor for $\beta_1 + \beta_2 x_{T+1}$ is $\tilde{\beta}_1 + \beta_2 x_{T+1}$ where $(\tilde{\beta}_1, \beta_2)$ are generalized least squares estimates.
2. To predict the component ρe_T, we need estimates for both ρ and e_T. For ρ we can use the estimator $\hat{\rho}$ specified in (12.5.5). To estimate e_T we use the generalized least squares residual, defined as

$$\tilde{e}_T = y_T - \tilde{\beta}_1 - \beta_2 x_T \qquad (12.7.4)$$

3. The best forecast of the third component v_{T+1} is zero because this component is uncorrelated with past values $v_1, v_2, ..., v_T$.

- Collecting all these results, *our predictor for* y_{T+1} *is given by*

$$\hat{y}_{T+1} = \tilde{\beta}_1 + \beta_2 x_{T+1} + \hat{\rho}\tilde{e}_T \qquad (12.7.5)$$

- What about predicting more than one period into the future? For *h* periods ahead, it can be shown that the best predictor is

$$\hat{y}_{T+h} = \hat{\beta}_1 + \beta_2 x_{T+h} + \hat{\rho}^h \tilde{e}_T \qquad (12.7.6)$$

- Assuming $|\hat{\rho}| < 1$, the influence of the term $\hat{\rho}^h \tilde{e}_T$ diminishes the further we go into the future (the larger *h* becomes).
- In the Bangladesh sugar cane example

$$\hat{\beta}_1 = 6.1641 \qquad \hat{\beta}_2 = 1.0066 \qquad \hat{\rho} = 0.342$$

and

$$\begin{aligned} \tilde{e}_T &= y_T - \hat{\beta}_1 - \beta_2 x_T \\ &= \ln(A_T) - \hat{\beta}_1 - \beta_2 \ln(P_T) \\ &= 5.4596 - 6.1641 - 1.0066(-0.9374) \\ &= 0.239 \end{aligned} \qquad (R12.10)$$

- To predict y_{T+1} and y_{T+2} for a sugar cane price of 0.4, in both periods (T + 1) and (T + 2), we have

$$\hat{y}_{T+1} = \hat{\beta}_1 + \beta_2 x_{T+1} + \hat{\rho}\tilde{e}_T$$

$$= 6.1641 + 1.0066 \ln(0.4) + (0.342)(0.239)$$

$$= 5.3235 \qquad \text{(R12.11)}$$

$$\hat{y}_{T+2} = \hat{\beta}_1 + \beta_2 x_{T+1} + \hat{\rho}^2\tilde{e}_T$$

$$= 6.1641 + 1.0066 \ln(0.4) + (0.342)^2(0.239) \qquad \text{(R12.12)}$$

$$= 5.2697$$

- Note that these predictions are for the logarithm of area; they correspond to areas of 205 and 194, respectively.