Chapter 12: Time Series Regression Models

• When data are gathered over time, the assumption that the error terms are uncorrelated across observations may be incorrect.
• In time series data, it is common for the errors to be positively correlated across time.

• Correlation of r.v.'s over time is called autocorrelation or serial correlation.

An Example:
Response = annual sales of a product, for a period of 30 years
Predictor = annual price of product, for the 30 years

• If population size affects sales, and if population size is omitted from the model, then the resulting linear regression will likely have errors that are positively correlated over time.

Problems Resulting from Autocorrelation

• The estimated regression coefficients are still unbiased, but their variances will be excessively large.
• MSE may severely underestimate the error variance σ^2.
• The standard error of the b_j's (as usually calculated) may seriously underestimate the true standard deviation of the estimated coefficients.
• The t-procedures and F-tests used for inference about the regression model will not be valid.

Effect of Autocorrelated Error Terms

Let

• Suppose ε_t consists of

where the disturbance terms u_t are distributed

• So any error term is a combination of the _____ and _____.

• This results in the ε_t's being _____ over time.

• See Figure 12.2, pg. 483 based on some simulated data:

Result:

• The variability of the data around the fitted regression line is _____ than the variability of the data around the true regression line.

• The effect of the first error term on the data (and the fit) is large (see Figure 12.2(c) for data having a different ε_0).

First-order Autoregressive Error SLR Model

• This model, which is common in time series analysis, assumes the errors follow an AR(1) process.

where the autocorrelation parameter ρ is such that
and

• If $\rho = 0$, then this is

• If we have multiple predictors $X_1, \ldots X_k$, then the model is

Properties of Error Terms under the Autoregressive Model

Note

Since all U_t have mean zero,

Writing the ε_t's in terms of the U_t's and using the independence of the U_t's, we can show

• For any $s \neq 0$, the covariance between error terms s time periods apart is

• This is called the

• The correlation coefficient between error terms s time periods apart is

• This is called the

Durbin-Watson Test for Autocorrelation

• This tests the null hypothesis
in the first-order autoregressive model

• The usual hypotheses are

• The test statistic D is based on the residuals from an ordinary least squares fit:

• A small value of D implies that ε_t and ε_{t-1} tend to be _____, and so a small D value leads us to conclude:

• Table B.7 gives values d_L and d_U such that if $D > d_U$ we conclude H_0, and if $D < d_L$ we conclude H_a.

• SAS gives an exact P-value for the D-W test with the DWPROB option in PROC REG. The R function `dwtest` in the `lmtest` package will also give an exact P-value.

• To test H_0: $\rho = 0$ vs. H_0: $\rho < 0$, we use the same test with as the test statistic.

• To test H_0: $\rho = 0$ vs. H_0: $\rho \neq 0$ at level α, we can do both of the one-sided tests and reject H_0 if <u>either</u> p-value is less than $\alpha/2$.

<u>Example</u> (Blaisdell sales data):
Response = the company's quarterly sales (seasonally adjusted) from 1998-2002.
Predictor = the industry's quarterly sales (seasonally adjusted) from 1998-2002.

D-W test (from SAS or R):

Remedies for Autocorrelation

• The easiest and best remedy when errors are autocorrelated is to add predictor variables to the model.
• Adding a key predictor that has time-ordered effects on the response can solve the problem.
• Including indicator variables for seasonal effects can be helpful when the data exhibit seasonality.

• If addition of predictor variables does not help, we can try transformations of the variables:

Let

Then under the SLR model with AR(1) errors:

• So using the transformations Y_t' and X_t' above yields a SLR model with independent errors, which can be fit with OLS.

Note: ρ is unknown, so we must estimate it by some statistic r.

Then

We regress Y' against X' to obtain

and back-transform via

to obtain

• There are three common procedures to find r: An iterative procedure called the Cochrane-Orcutt procedure, the Hildreth-Lu procedure, and the first-differences procedure.

• The Cochrane-Orcutt procedure chooses the estimate of ρ to be:

• After regressing Y' against X', we check with the D-W test whether autocorrelation of the errors still exists.
• If so, we can iterate this process one or two times, using the most recent residuals in the calculation of *r*.

• The Hildreth-Lu procedure uses a computer search to find the estimate of ρ that minimizes:

• The first differences procedure essentially just sets ρ = 1.

• PROC AUTOREG in SAS uses the Yule-Walker procedure, which is like the Cochrane-Orcutt method, except that it retains information in the first observation.

<u>Example</u> (Blaisdell data):

Forecasting with Time Series Models

• Forecasting refers to predicting a response value for some time in the future (after the most recent time period from the sample).

• For example, if our sample responses are (Y_1, Y_2, …, Y_n), then we may want to forecast the value of Y_{n+1} (using the predictor value(s) at time period n+1.

• Note that if the errors are autocorrelated, then information about the error at time n, ε_n, will be informative about the next error, ε_{n+1}.

• Based on our AR(1) model for SLR:

• We estimate $\beta_0 + \beta_1 X_{n+1}$ by

• We estimate $\rho\varepsilon_n$ by

• We estimate u_{n+1} by

• So our forecast is

• We can get an approximate 100(1 – α)% prediction interval for

where $s^2\{pred\}$ is the estimated <u>prediction variance</u> calculated based on the <u>transformed</u> variables:

• The error d.f. are (n – 3) since only (n – 1) transformed observations are used in the Cochrane-Orcutt procedure.

<u>Example</u> (Blaisdell data): Our sample data are for time periods 1, 2, …, 20.
• It is projected that the industry sales for period 21 will be
X_{21} **=**

• The forecast

• Forecasts for two or more time periods ahead can also be developed; for example: