

## 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  $\sigma^2$ .
- 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  $\varepsilon_t$  consists of

where the disturbance terms  $u_t$  are distributed

- So any error term is a combination of the \_\_\_\_\_  
and \_\_\_\_\_.

- This results in the  $\varepsilon_t$ '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  $\varepsilon_0$ ).

### 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  $\rho$  is such that  
and

- If  $\rho = 0$ , then this is
- If we have multiple predictors  $X_1, \dots, X_k$ , then the model is

### Properties of Error Terms under the Autoregressive Model

Note

Since all  $U_t$  have mean zero,

Writing the  $\varepsilon_t$ '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  $\varepsilon_t$  and  $\varepsilon_{t-1}$  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  $\alpha$ , we can do both of the one-sided tests and reject  $H_0$  if either p-value is less than  $\alpha/2$ .

**Example** (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:  $\rho$  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  $\rho$  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  $\rho$  that minimizes:**

- **The first differences procedure essentially just sets  $\rho = 1$ .**

- **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.**

**Example (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  $(Y_1, Y_2, \dots, Y_n)$ , 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  $n$ ,  $\varepsilon_n$ , will be informative about the next error,  $\varepsilon_{n+1}$ .
- 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  $100(1 - \alpha)\%$  prediction interval for

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

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

**Example** (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:

## Regression with Missing Data

- When some observations have missing values for some variables, we cannot use our usual regression or ANOVA formulas.

### Types of “Missingness”

- Suppose  $U$  is a variable that may be missing and the vector  $\underline{W}$  represents a set of variables whose values are completely observed.

- Let  $R$  be a “missingness” indicator such that

- Then the data are missing completely at random (MCAR) if

- MCAR implies that the probability of an observation being missing on  $U$  is unrelated to the values of ANY of the variables.

- The data are missing at random (MAR) if

- MAR implies that the probability of an observation being missing on  $U$  is unrelated to its true value of  $U$ , but this missingness probability could be related to the values of the other variables.

- MAR is a \_\_\_\_\_ assumption than MCAR, since if the data are \_\_\_\_\_, then they must be \_\_\_\_\_.

- There are several options for handling missing data in the linear model framework:

**Listwise deletion:** This method simply removes any observations that have missing values for any variable.

- Then the model is fit using only the observations that have no missing values for any variable.

**Disadvantages:** (1) If the number of missing values is not small, then this can result in a greatly reduced sample size and a lot of sample information gets “thrown away”.  
(2) The resulting regression estimates will be biased, unless the missing data are truly MCAR (which is a strong assumption).

**Multiple Imputation (MI):** This approach “fills in” (imputes) missing data several different times, creating several “imputed data sets”.

- Then the model is fit to all the imputed data sets separately.
- The results (parameter estimates, standard errors, test statistics, etc.) from each fit are combined into a single set of results.
- The MI estimates have nice (large-sample) properties:

## Creating the Imputed Data

- A common method of imputation is the linear regression method.
- Suppose  $U$  has missing values but  $W_1$  and  $W_2$  are completely observed for all observations.
- We fit a linear regression of  $U$  on  $W_1$  and  $W_2$  based on the observations having no missing data.
- Using the resulting regression equation, we predict the values of  $U$  that are missing, based on the observed  $W_1$  and  $W_2$  for those observations.
- To increase the variance of the imputed data, to account for the fact that it was predicted and not truly observed, we actually use as the imputed value of  $U$ :

where  $\varepsilon$  is a randomly generated standard normal r.v.

- This is done for each observation having a missing  $U$  value.
- We do this several times, each with different random draws of  $\varepsilon$  from the standard normal distribution, thus creating “multiple” imputed data sets.
- Formulas have been derived to estimate the standard error correctly under multiple imputation.
- Around  $M = 5$  imputed data sets are considered enough to estimate parameters accurately, although more may be needed if the amount of missing data is large.

- **One attractive implementation of MI is called “multiple imputation by chained equations” (MICE), which is capable of imputing both quantitative and categorical data.**
- **This can be implemented by PROC MI in SAS and by the mice package in R.**

**Maximum Likelihood: This works by:**

- (1) assuming a model for the data**
  - (2) factoring the likelihood into a part involving the observations with complete data and a part involving the observations with missing data**
  - (3) estimating the parameters by maximizing the likelihood**
- **A disadvantage of MI is that, since it relies on randomly generated data, you will get a different result each time you run it on the same data.**
  - **Also, there are a lot of modeling choices to make with MI (although in practice people tend to use the software default choices).**
  - **For example, you must choose both an “imputation model” and an “analysis model”, and problems can arise when the analysis model is more complicated than the imputation model.**
  - **With the maximum likelihood (ML) method for missing data, you assume one model and will get one result.**
  - **ML shares the same good large-sample properties as MI and is actually even more efficient.**
  - **But MI is somewhat more flexible (you can do any analysis you want on the imputed data sets, using familiar procedures and functions).**

- **With SAS, you can use ML on most linear models, and a few nonlinear models, but not all common models.**

**Example 1: (Regression Models with Missing Data on the Response Variable)**

- **If some observations are missing response values, but there is complete data for all the predictor variables, listwise deletion is recommended (ML will be equivalent to listwise deletion in this situation).**
- **If some observations are missing response values, and other observations are missing values of the predictor variables, it is recommended to use MI to impute the missing predictor values, but still delete the observations with missing responses.**

**Example 2: (Repeated Measures Models with Dropouts)**

- **This is a common situation in longitudinal studies when some subjects may drop out of the study before all the measurements across time have been taken.**

**Example:** (Dental data with some children dropping out before the end of the study)

- For GLMs (with non-normal responses) having missing data, PROC GLIMMIX can implement the ML approach.

**Example 3: (Regression Models with Missing Values on Predictor Variables)**

**Example: (1990 National Longitudinal Survey of Youth; data on 581 children)**

**Response:** ANTI (antisocial behavior, measured with a scale ranging from 0 to 6)

**Predictors:** SELF=self-esteem (measured on scale from 6 to 24)

**POV**=poverty status of family (1= in poverty, otherwise 0)

**BLACK** (1 if child is black, otherwise 0)

**HISPANIC** (1 if child is Hispanic, otherwise 0)

**CHILDAGE** (child's age in 1990)

**DIVORCE** (1 if mother was divorced in 1990, otherwise 0)

**GENDER** (1 if female, 0 if male)

**MOMAGE** (mother's age at birth of child)

**MOMWORK** (1 if mother was employed in 1990, otherwise 0)

- No missing response values here, but lots of missing predictor values.
- Compare results using listwise deletion to results using multiple imputation: