Econ 300/QAC 201: Quantitative Methods in Economics/Applied Data Analysis

17th Class **7/1/10**

"The only function of economic forecasting is to make astrology look respectable."
--John Kenneth Galbraith

show the class examples of regressions from Wesleyan faculty members
[handout]

going through problems with regression; differentiating problems with data from problems with technique

mention multicollinearity and omitted variables as the problems we last discussed

a way of dealing with the problem of omitted variables if you have panel data (a time-series on a cross-section): first-differencing
--assume for now that the error terms are serially uncorrelated
If the true model is:

$$Y = b_0 + b_1X + b_2Z + e$$

then this holds true at all points in time:

$$Y_t = b_0 + b_1X_t + b_2Z_t + e_t$$

$$Y_{t-1} = b_0 + b_1X_{t-1} + b_2Z_{t-1} + e_{t-1}$$

if data are not available on Z, then we are forced to estimate (for either time):

$$Y = b_0 + b_1X$$

and if X and Z are correlated, then the estimate of b_1 will be biased

but can take first differences between the true regression equations:

$$(Y_t - Y_{t-1}) = b_0 - b_0 + b_1(X_t - X_{t-1}) + b_2(Z_t - Z_{t-1}) + (e_t - e_{t-1})$$

using the Δ sign to signify a change and getting rid of zeros, this reduces to:

$$\Delta Y = b_1 \Delta X + b_2 \Delta Z$$

and if Z does not change over time (e.g., one's sex or race), then this reduces to running the regression:

$$\Delta Y = b_1 \Delta X$$

and the coefficient from this regression on ΔX will be an unbiased estimate of b_1

note in order to have a constant term in this regression there must be a time trend T, which increments by one, and the coefficient on this term will be the constant in the new equation (e.g., years of experience, or age):

$$Y_t = b_0 + b_1 X_t + b_2 Z_t + b_3 t + e_t$$

$$Y_{t-1} = b_0 + b_1 X_{t-1} + b_2 Z_{t-1} + b_3(t-1) + e_t$$

first differences yields:

$$\Delta Y = b_1 \Delta X + b_3$$

so far we have justified OLS as a generally sound estimation technique, and have only discussed problems (multicollinearity, omitted variables) related to the data rather than the estimation technique

but consider the basic assumptions underlying regression:

$E(e_i) = 0$ error terms have mean 0

$Var(e_i) = s^2$ all error terms have the same variance

$E(e_i e_j) = 0$ error terms are not correlated across observations (they are independent)

when these assumptions are met, OLS yields the BLUE (best linear unbiased estimator) as evaluated by minimized MSE (mean squared error) but when one or more of these assumptions are violated, OLS needs to be adjusted or a different method used

a check for patterns in the residuals will show violations

problem of nonconstant variance is called heteroscedasticity; then OLS estimates of the coefficients are still nonbiased, but do not have minimum variance
--then we use a form of computation of the variance/standard errors that is robust to heteroscedasticity (White standard errors) so as to calculate the correct significance tests for the coefficients

problem of correlation across residuals is called autocorrelation (refers to the commonness of this phenomenon in time-series data); then alternative estimation techniques that exploit this correlation can be used to improve prediction (Ch. 24)

Ch. 24

two major categories of statistical information: cross-section and time series (also panel data)

consider special characteristics of time series data on top of their having a trend (i.e., a regression line with nonzero slope):

  --they tend to exhibit serial correlation: each observation is statistically dependent on the previous ones (so assumption of independence of observations does not hold)
  consider Figure 24-1: the long-term trend does not do a good job in predicting year to year changes; a better predictor would be to forecast next year's value as being the same as this year's, but we can do even better on average than that prediction
  --they often have seasonal variation too
    consider examples for quarterly, monthly, weekly, and daily data series
    --retail sales would have a spike in December that would affect quarterly and monthly data
    --stock data: prices drop on common x-dividend days
    --money demand: spikes on Fridays--common paydays, particularly at middle and end of months

the general technique of regression (not necessarily the particular estimation technique of OLS) can be used to decompose the time series into its different systematic and pure noise components and to forecast future values

elements include:

--trend (not necessarily linear; many economic time series have exponential trend, growing at a constant rate over time, so they are linear on a logarithmic scale)

--seasonal fluctuation around the trend (can handle this often with a simple set of dummy variables)

--serially correlated residual, e.g. $e_t = \rho e_{t-1} + v_t$, where ρ represents the strength of the serial correlation (and is typically less than 1 in absolute value—a correlation coefficient between residual and lagged residual); the general form of this is $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_k e_{t-k} + v_t$, which would be called a <u>kth order autoregression</u>

--white, or pure noise: v_t in the above equation

in general, when forecasts further into the future than next period are dependent upon previous periods' residuals, they will need to be updated as soon as actual figures become available to improve their accuracy

Durbin-Watson test for serial correlation:

tests H_0: $\rho = 0$

DW = 2 - 2r

so roughly DW = 0 means positive serial correlation, 2 is no correlation, and 4 is negative serial correlation (can look up critical values in Table IX for the exact test values)

What if we want to decompose a time series into 4 components:

--trend T (same as above)

--seasonal variation S (same as above)

--cyclical variation C (modeled as part of the error term structure)

--disturbance D (white noise)

pretty easy to get rid of T and S (use a time trend t in the regression, and dummies for the seasons), then just need to model C to get the error down to D

What if we want to forecast a series that has serial correlation (assume no trend or seasonal variation for simplicity)? Several possible techniques can be used. Consider the <u>autoregressive model</u> which predicts Y_t as a linear combination of previous values of Y:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t$$

$$\text{so } \hat{Y}_t = b_1 Y_{t-1} + b_2 Y_{t-2} + \ldots + b_p Y_{t-p}$$

1) can use the simple technique of exponential smoothing, which assumes that the autoregression extends back infinitely with smaller and smaller weights that decline exponentially, so:

$$b_k = b\lambda^k \text{ where } k = 1, 2, \ldots \text{ and } 0 < \lambda < 1$$

substituting this into the autoregressive model, we get:

$$\hat{Y}_t = b\lambda Y_{t-1} + b\lambda^2 Y_{t-2} + b\lambda^3 Y_{t-3} + \ldots$$

since this can be applied to any time period, we can also write:

$$\hat{Y}_{t-1} = b\lambda Y_{t-2} + b\lambda^2 Y_{t-3} + b\lambda^3 Y_{t-4} + \ldots$$

note we can simplify the model by writing it as:

$$\hat{Y}_t = b\lambda Y_{t-1} + \lambda(b\lambda Y_{t-2} + b\lambda^2 Y_{t-3} + \ldots)$$

but now the term in parentheses is $\hat{Y}_{t-1}$, so we can write the model as:

$$\hat{Y}_t = b\lambda Y_{t-1} + \lambda \hat{Y}_{t-1}$$

and if we require the weights to add to 1, then this must equal:

$$\hat{Y}_t = (1 - \lambda)Y_{t-1} + \lambda\hat{Y}_{t-1}$$

thereby allowing us to estimate what is an infinite series as a simple equation which can be estimated by least squares and used to forecast the current Y using a weighted average of the last Y and the prediction of the last Y that we had made last period

2) can use the Box-Jenkins model, which is a freer fit of the parameters than exponential smoothing, but does not require all the parameters of a completely free fit of a long autoregression; it cuts the autoregressive model off at a small number of lags, including only the most important recent lagged values. Therefore, it does not capture all of the serial correlation, some of which remains in u_t. u_t could therefore be modeled as an autoregressive structure as well, but is instead modeled as a moving average of serially independent terms v_{t-i} (white noise terms):

$$u_t = \gamma_0 v_t + \gamma_1 v_{t-1} + \ldots + \gamma_q v_{t-q}$$

the coefficients are not required to sum to 1 or to be positive. It is generally restricted slightly however, and written as:

$$u_t = v_t - \alpha_1 v_{t-1} - \ldots - \alpha_q v_{t-q}$$

Upon substituting this into the autoregressive model, we obtain a combination of an autoregression (AR) of length p and a moving average (MA) of length q, or the ARMA (p,q) model:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + v_t - \alpha_1 v_{t-1} - \ldots - \alpha_q v_{t-q}$$

the β and α parameters can be estimated using EViews and this model used for forecasting Y

both of these techniques can be revised to incorporate seasonal and trend components

now consider how we might not only forecast a time series, but also understand how it is related to one or more other time series problem with serial correlation: successive observations tend to resemble previous observations (are dependent) and hence provide less new information about trend than a set of independent observations would. Therefore, estimates are less reliable, which should be reflected in wider confidence intervals around coefficients.

Consider our basic regression model:

$$Y_t = \alpha + \beta X_t + e_t$$

where X and Y are two time series

assume that:

$$e_t = e_{t-1} + v_t$$

so for simplicity we are assuming that residuals are perfectly correlated; i.e., $\rho = 1$

then b will not be biased (it could either overestimate or underestimate β, depending on the particular pattern of the residuals), but will have a large standard error [see Figure 24-10]. Note that the true line will actually look less reliable than the OLS estimate because the observed/estimated residuals will be smaller with OLS (autocorrelated data is smooth) than the true residuals

remedy for serial correlation: transform the data to a form that conforms to the assumptions underlying OLS, namely estimate the equation in changes rather than levels. Our model is:

$$Y_t = \alpha + \beta X_t + e_t$$

and therefore:

$$Y_{t-1} = \alpha + \beta X_{t-1} + e_{t-1}$$

subtracting the second equation from the first, we get:

$$Y_t - Y_{t-1} = \beta(X_t - X_{t-1}) + (e_t - e_{t-1})$$

define the differences:

$$\Delta Y_t = Y_t - Y_{t-1}$$
$$\Delta X_t = X_t - X_{t-1}$$

then this can be written as:

$$\Delta Y_t = \beta \Delta X_t + v_t$$

so b may be estimated validly using OLS on this equation

let us now generalize this model to allow $|\rho| < 1$, so:

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

we can lag the model and multiply through by r:

$$\rho Y_{t-1} = \rho\alpha + \rho\beta X_{t-1} + \rho e_{t-1}$$

then, upon first-differencing, the model is:

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + (e_t - \rho e_{t-1})$$

define the <u>generalized differences</u>:

$$\Delta Y_t = Y_t - \rho Y_{t-1}$$
$$\Delta X_t = X_t - \rho X_{t-1}$$

then the model can be written:

$$\Delta Y_t = \alpha(1 - \rho) + \beta \Delta X_t + v_t$$

so after this transformation, regress ΔY_t on ΔX_t to estimate β. This technique is an example of <u>generalized least squares</u> (GLS).

Note the problem here: we don't know what ρ is. We have to estimate it somehow:

1) We can estimate it by first using OLS to fit a line to the data and generate estimates of the residuals, and then using OLS to estimate the equation:

$$\hat{e}_t = r\hat{e}_{t-1}$$

however, this technique causes r to be a downward-biased estimate of r: the estimated residuals are less serially correlated than are the true residuals (as can be seen in Figure 24-10)

2) alternatively, we can estimate ρ by first rearranging our model:

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + (e_t - \rho e_{t-1})$$

into the following form:

$$Y_t = \alpha(1 - \rho) + \rho Y_{t-1} + \beta X_t - \beta\rho X_{t-1} + v_t$$

which can then be estimated using OLS. Then the coefficient of Y_{t-1} provides us with an estimate of ρ.

3) the same as (2), except estimate ρ by dividing the coefficient of X_{t-1} by the coefficient of X_t (and change the sign)

4) the same as (2) and (3), except take an average of the two estimates of ρ to get an even better estimate of r. Note that if the two estimates are very different, this is a warning that the model may be poorly specified.

[consider problem 24-11 if there is time]