Quant II - Problem Set V
Problems with Predictors and Errors

Kenneth Benoit

Assigned: Wednesday, February 11, 2009
Due: Wednesday, February 25, 2009

1. Revision: Quadratic functional forms
   Consider the following model including a quadratic term:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \tag{1}$$

   The marginal effect of x on y, i.e. the change in y if x changes by one unit is given by the first partial derivative with respect to x:

$$\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x \tag{2}$$

   In order to find out when the marginal effect is equal to zero, we set eq. (2) to zero, which yields:

$$x = -\frac{\beta_1}{2\beta_2} \tag{3}$$

   The second partial derivative

$$\frac{\partial^2 y}{\partial x^2} = 2\beta_2 \tag{4}$$

   indicates whether the extreme point given by eq. (3) is a minimum (eq. $4 > 0$) or a maximum (eq. $4 < 0$).

   (a) What does eq. (2) tell you about the marginal effect of x on y in any model that looks like eq. (1)?
   (b) Re-run the original main-effects model from Tavits (2005), which appeared at the beginning of the last homework. The variable yearsdem2 is the square of yearsdem, which measures years since the first democratic election.
   Plot the expected value of electoral volatility (compare eq. 1) against yearsdem. (You may choose whatever intercept you like.

Choosing zero, for example, has the advantage that you effectively look at first differences of the expected value of the dependent variable, i.e. how the dependent variable changes depending on yearsdem, holding everything else constant, regardless of the values you actually hold everything else constant at.) Add the marginal effect of yearsdem to the plot. It may also be helpful to include the lines $y = 0$ and the one given in eq. 3. Discuss the findings concerning the variable yearsdem with the help of your plot and the relationships expressed in the above equations. Make sure you refer to the substantive question at hand.

2. Collinearity

   Load the INES 2002 data set that we used previously. We are interested in analyzing B55_1, the probability of ever giving one's first preference vote to Fianna Fáil, which is measured on a scale from 1 (not at all probable) to 10 (very probable). As independent variables we use the following:
   DO44_1, the rating whether Bertie Ahern is in touch with ordinary people, which is measured on a scale from 0 (not in touch) to 10 (in touch).
   DO45_1, the rating whether Bertie Ahern would be good in running the country, which is measured on a scale from 0 (would not be very good) to 10 (would be very good).

   (a) If necessary, set observations with values of 11 and 99 to missing. For the sample that contains only complete cases with regard to all three variables, regress B55_1 on DO44_1 and DO45_1. Showing commands and output is sufficient, there is no need for interpretation.
   (b) Next, run a regression where the dependent variable is the residual from the regression of B55_1 on DO44_1 and the independent variable is the residual from the regression of DO45_1 on DO44_1. How do the findings of the second regression compare to those of the main regression and why?
   (c) Calculate the variance inflation factor of DO45_1 (referring to the main regression) "manually" and briefly interpret.

3. Kennedy (pp. 157-158, though this may vary depending on the edition) illustrates the problem of structural equations using a Keynesian system.

(a) Show step by step how the structural equations can be expressed in the reduced forms given on p. 158.

(b) Show why the structural parameters are overidentified.

(c) Could you use indirect least squares to estimate the structural parameters?

(d) Could you use an instrumental variables approach to estimate the structural parameters? How would you go about doing this?

4. Since you might already miss our favourite data set on campaign spending, we will use it now.

   (a) Try to replicate Model 3 from Table 3 in Benoit and Marsh (2008). (R command and output is sufficient.)

   (b) Imagine you are consulting a candidate who knows nothing about statistics. Explain the endogeneity problem in the above model in one or two sentences to him/her.

   (c) Try to replicate Model 1 from Table 3 by proceeding as follows: Run Model 1 from Table 2. Obtain the predicted values of this regression. Replace the original regular spending variable with the fitted values, both for the variable as such and in the interaction effect. Then estimate Model 1 from Table 3 with the new variables. (R command and output is sufficient.)

   (d) Explain to the candidate why the new model yields a better estimator of the effectiveness of spending (a few sentences are enough).

   (e) Perform a Hausman test of the exogeneity of regular spending, according to the description of this procedure in Kennedy p. 174. Interpret.