Econometrics 2 – Regression

Wednesday, March 16th, 2016

1. (Greene) **Beta coefficients.** In other social sciences (typically not in economics), researchers often use linear least squares to compute beta coefficients. Starting with a data set consisting of n observations on K variables in $\boldsymbol{X}$ and on the regressand $\boldsymbol{y}$, beta coefficients are obtained by the linear regression of $\boldsymbol{y}^*$ on $\boldsymbol{X}^*$, where $\boldsymbol{y}^*$ and $\boldsymbol{X}^*$ are obtained by standardizing the data. That is, every observation on each variable in the model is transformed by subtracting the respective sample variable mean, then dividing by the respective sample variable standard deviation. Assume that the data were already in the form of deviations from means, so all variables in the model have precisely zero means to begin with. Thus, standardizing amounts only to dividing each variable by its respective sample standard deviation. Note that $\boldsymbol{X}$ does not contain a constant term (and $\boldsymbol{y}$ has zero mean). (The assumption about the means simplifies the algebra, but does not change any results.)
   a) Could you obtain the coefficients in the transformed regression from the untransformed regression? How? Or, why not?
   b) How does the sum of squared residuals in the transformed regression compare to the sum of squared residuals in the regression using untransformed data?

2. (Greene) **Partitioned regression.** Suppose a data set consists of n observations on $\boldsymbol{y}$, K_1 variables in $\boldsymbol{X}_1$ and K_2 variables in $\boldsymbol{X}_2$. Do the following three procedures produce the same value for the least squares coefficients on $\boldsymbol{X}_2$?
   a) Regress $\boldsymbol{y}$ on both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.
   b) Regress the residuals from a regression of $\boldsymbol{y}$ on $\boldsymbol{X}_1$ on the residuals (column by column) of regressions of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$.
   c) Same as b), but do not transform $\boldsymbol{y}$.
   d) Same as b), but do not transform $\boldsymbol{X}_2$.

3. (Greene) **Multicollinearity.** The regression model of interest is $\boldsymbol{y} = \boldsymbol{X}_{1\ 1} + \boldsymbol{X}_{2\ 2} +$, where $\boldsymbol{X}_1$ is K_1 variables, including a constant and $\boldsymbol{X}_2$ is K_2 variables not including a constant. It is believed that multicollinearity between the columns of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is adversely affecting the regression. Consider the following »cure«. We will first regress each variable in $\boldsymbol{X}_2$ on all of the variables in $\boldsymbol{X}_1$. By construction, the residuals in these regressions, call them $\boldsymbol{Z} = (\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{K2})$, are orthogonal to every variable in $\boldsymbol{X}_1$. So, instead of regressing $\boldsymbol{y}$ on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, we linearly regress $\boldsymbol{y}$ on $\boldsymbol{X}_1$ and $\boldsymbol{Z}$. Denote by $\boldsymbol{b} = (\boldsymbol{b}_1,\boldsymbol{b}_2)$ the least squares coefficients in the original regression, and by $\boldsymbol{c} = (\boldsymbol{c}_1,\boldsymbol{c}_2)$ the least squares coefficients in the regression of $\boldsymbol{y}$ on $\boldsymbol{X}_1$ and $\boldsymbol{Z}$. Show the algebraic relation between $\boldsymbol{b}$ and $\boldsymbol{c}$. Is $\boldsymbol{c}$ unbiased? Using the gasoline data, let the dependent variable be *GasQ/Pop*, let $\boldsymbol{X}_2$ denote the three macroeconomic price indexes, P_d, P_n, and P_s, and let $\boldsymbol{X}_1$ denote the other independent variables, *constant*, *GasP*, and *PCIncome*. Carry out the computations listed above and verify that the algebraic results you obtained do appear in the empirical results.