

Problem Set 3

Due: November 21st, 5:00pm. Upload a pdf with your solutions on NYU Classes, late submissions will be penalized

Problem 1

The file DataExample.xls contains information about the regressors of the following model:

$$\begin{aligned} \log(G/Pop) = & \alpha + \beta_1 \log(P_g) + \beta_2 \log(Income/Pop) + \gamma_1 \log(P_{nc}) \\ & + \gamma_2 \log(P_{uc}) + \tau \cdot month + \delta_1 \log(P_d) + \varepsilon \end{aligned}$$

In this exercise you are going to generate artificial data and construct procedures to get an econometric estimation.

1. Generate a sequence of errors with normal standard distribution. Use this sequence and the coefficients in the second sheet to generate an artificial sequence for the dependent variable. Calculate the OLS estimator, the estimated residuals, the estimated variance of the residuals, the asymptotic variance of the OLS estimator, the t-statistics and the standard deviation of each coefficient. Present your results in a single table.
2. Estimate the Epanechnikov kernel and the Gaussian kernel for the previous case. In each case indicate the kernel regression estimator.
3. Generate now 100 sequences of errors standard normally distributed. We are trying to do repeated samples here. For each of the 100 sequences, generate the implied dependent variable. Reestimate the OLS for each case. Compute the sample variance of the OLS estimator across samples. Plot a histogram of the values of each coefficient. Does this distribution look normal? Should it look normal? Include in each histogram a normal curve with the same mean and variance than the coefficient.
4. Generate now 100 sequences of errors $\varepsilon \sim \sigma(e_t = 1)$ where e_t is a standard exponential variable with mean 1. Again, we are doing repeated samples here. Repeat the calculations in part 3. Does the distribution look normal in this case? Compare and comment.

Problem 2

The file datacrime.txt contains information to estimate a model on county level crime rates. You have to estimate a model using the year 1987 only.

1. Using logarithms of all variables, estimate a model relating the crime rate to the deterrent variables *prbarr*, *prbconv*, *prbpris* and *avgsen*. Include all the t-statistics and the standard errors of the coefficients.
2. Add $\log(\text{crm rte})$ for 1986 as an additional explanatory variable, and comment how the elasticities differ from the previous equation and why.
3. Compute the F statistic for joint significance of all of the wage variables (again in logs), using the restricted model from part b.

Problem 3

The file *invest.txt* contains data on 565 US firms extracted from Compustat for the year 1987. These variables (in order) are: investment to capital ratio (I_i), total market value to asset ratio (Tobin's Q, Q_i), cash flow to asset ratio (C_i) and long term debt to asset ratio (D_i). The flow variables are annual sums for 1987. The stock variables correspond to the beginning of the year.

1. Estimate a linear regression of I_i on the other variables. Calculate appropriate standard errors.
2. Calculate asymptotic confidence intervals for the coefficients.
3. This regression is related to Tobin's Q theory of investment, which suggests that investment should be predicted solely by Q_i . Thus, the coefficient on Q_i should be positive and the others should be zero. Test the joint hypothesis that the coefficients C_i and D_i are zero. Test the hypothesis that the coefficient on Q_i is zero. Are the results consistent with the predictions of the theory?
4. Now try a nonlinear specification. Regress I_i on Q_i , C_i , D_i , Q_i^2 , C_i^2 , D_i^2 , $Q_i C_i$, $Q_i D_i$, $D_i C_i$. Test the joint hypothesis that the six interaction and quadratic coefficients are zero.

Problem 4

(Harder) (Hausman-Wu statistic) This problem is about specification tests and asks you to derive the correct asymptotic variance for the Hausman-Wu statistic. Suppose you have the regression equation

$$y_i = x_i' \beta + \varepsilon_i$$

and $k = k_1 + k_2$ instrumental variables $z_i = [z_{i1}', z_{i2}']$, where we denote the partitioned matrix of IVs as $Z = [Z_1, Z_2]$. We consider the 2SLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ that use z_{i1} (the full vector z_i , respectively) as instruments. You can assume that the usual conditions for consistency and asymptotic normality hold in both cases, in particular the subset of IVs z_{i1} satisfies the rank and order conditions for identification of β .

1. Show that for the projection matrices $P_{z1} := Z_1(Z_1'Z_1)^{-1}Z_1'$ we have $P_z P_{z1} = P_{z1}$
2. Under the assumption of correct specification and homoskedasticity, $\mathbb{E}[\varepsilon_i|z_i] = 0$ and $\mathbb{E}[\varepsilon_i^2|z_i] = \sigma_0^2$, derive the limit of the joint distribution of $\hat{\beta}_1, \hat{\beta}_2$ as $n \rightarrow \infty$. Use $V_{ij} := \lim_n \mathbb{E}[(\hat{\beta}_i - \beta)(\hat{\beta}_j - \beta)']$ to denote the ij block of the asymptotic covariance matrix for $i, j = 1, 2$. Hint: use your result from (1) to simplify expressions.
3. Show that $V_{22} - V_{12} - V_{21} = -V_{22}$
4. Use your result to show that under the assumption of part (2),

$$\sqrt{n}(\hat{\beta}_1 - \hat{\beta}_2) \rightarrow_d N(0, V_{11} - V_{22})$$