**Advanced Microeconometrics**
**Homework Assignment 1**

## 1. Estimating Equations (15 marks)

(a) Show that the least squares estimating equations, sometimes also called "normal equations" for $\beta_1$ and $\beta_2$ in following specifications of the bivariate regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i, \quad i = 1, ..., N, \quad N > 2$$

are:

$$\sum_{i=1}^{N} Y_i = N\widehat{\beta}_1 + \widehat{\beta}_2 \sum_{i=1}^{N} X_{2i}, \tag{1}$$



$$\sum_{i=1}^{N} Y_i X_{2i} = \widehat{\beta}_1 \sum_{i=1}^{N} X_{2i} + \widehat{\beta}_2 \sum_{i=1}^{N} X_{2i}^2. \tag{2}$$

whose solution is given by:



$$\widehat{\beta}_2 = \frac{N \sum_{i=1}^{N} Y_i X_{2i} - \sum_{i=1}^{N} X_{2i} \sum_{i=1}^{N} Y_i}{N \sum_{i=1}^{N} X_{2i}^2 - (\sum_{i=1}^{N} X_{2i})^2}, \tag{3}$$



$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X}_2, \tag{4}$$

where $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ and $\overline{X}_2 = \frac{1}{N}\sum_{i=1}^{N} X_{2i}$. (6 marks)

(b) Express equations (1)-(2) and the solution of these linear equations in matrix notation. (4 marks)

(c) Write the usual moment condition of the linear regression model in matrix notation. (2 marks)

(d) Show that moment conditions imply the least squares equations (1)-(2). (6 marks)

**2. Generalised Least Squares (15 marks)**

Consider the linear regression model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \quad \mathbb{E}[u_i|\mathbf{x}_i] = 0$$

and suppose that the errors $u_i$ exhibit the following correlation structure:

$$\mathbb{E}[u_i^2|\mathbf{x}_i] = \sigma^2, \quad \mathbb{E}[u_i u_j|\mathbf{x}_i, \mathbf{x}_j] = \begin{cases} \rho\sigma^2 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

This implies that the errors of immediately adjacent observations are correlated whereas errors are otherwise uncorrelated. In matrix form we have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

(a) Verify that $\boldsymbol{\Omega} = \mathbb{E}[\mathbf{u}\mathbf{u}']$ is a band matrix with non-zero terms only on the diagonal and the first off-diagonal; and give these nonzero terms (2 marks). (Hint: $\mathbb{E}[u_i^2] = \mathbb{E}[\mathbb{E}[u_i^2|\mathbf{x}_i]]$) (3 marks)



(b) Show that $V[\hat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ where $\hat{\boldsymbol{\beta}}$ is the OLS estimator. (2 marks)



(c) Is the usual OLS estimate $s^2(\mathbf{X}'\mathbf{X})^{-1}$ a consistent estimator of $\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]$? Justify your answer. (2 marks)

(d) Is White's heteroskedasticity robust estimator of $\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]$ consistent? Justify your answer. (2 marks)



(e) State how to obtain a consistent estimate of $\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]$ that does not depend on unknown parameters. (6 marks)

**3. Minimizing a Quadratic Form (20 marks)**

Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

(a) Obtain the formula for $\hat{\boldsymbol{\beta}}$ which maximizes the objective function

$$Q_N(\boldsymbol{\beta}) = -\mathbf{u}'\mathbf{W}\mathbf{u}$$

where $\mathbf{W}$ has full rank. (9 marks)

(b) For which $\mathbf{W}$ does your answer to (a) equal the OLS estimator? (3 marks)

(c) For which $\mathbf{W}$ does your answer to (a) equal the GLS estimator? (4 marks)

(d) Use your answer to (c) to explain how you would obtain the Feasible GLS estimator if $\mathbf{\Omega} = \mathbb{E}[\mathbf{uu}']$ is that of question 2 above. (4 marks)

**4. Data Analysis (50 marks)**

You may use STATA or any other statistical software to answer this question. The datafile is **nerlove63.csv** available on Blackboard. These very old data were used by Marc Nerlove in a 1963 classic paper, "Returns to Scale in Electricity Supply," Chapter 7 in C.F. Christ, ed., *Measurement in Economics: Studies in Honor of Yehuda Grunfeld.* They are also used in a number of text books. You will use these data on 145 electricity generating plants to study the relationship between costs and output (i.e. a cost function) and to make inferences about returns to scale in electricity generation. The variables in this file are:



**ORDER**: The number of the observation, ascending in order from smallest in size to largest



**COSTS**: Total Production Costs in Millions of Dollars (dependent variable)
**KWH**: Kilowatt hours of output, in billions



**PL**: The wage rate per hour
**PF**: The price of fuels in cents per million BTU's
**PK**: The rental price index of capital

The regression model for cost function, obtained from the theory of cost minimization for a given level of output, is specified as

$$C = ky^{1/r} p_1^{\alpha_1/r} p_2^{\alpha_2/r} p_3^{\alpha_3/r} u$$

where $C$ denotes COSTS, $k$ denotes a constant (an unknown parameter), $y$ denotes KWH, and $(p_1, p_2, p_3)$ are the three input prices (PL,PF,PK), and $u$ is a multiplicative error term. The parameter $r$ is defined as $r = \alpha_1 + \alpha_2 + \alpha_3$, where $\alpha_1, \alpha_2$, and $\alpha_3$ (and $A$ also) are parameters in the Cobb-Douglas production function

$$y = Ax_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}.$$

(a) Apply the log transformation to all the variables. A variable in lower case with prefix l denotes the log-transformed value of the corresponding variable, e.g. **lpl** means ln(PL) etc. Obtain the correlation matrix of all variables except ORDER. Plot **lcosts** against **lkwh** to form a rough idea of the shape of the cost as a function of output. (5 marks)

(b) Using the production function linearized by log transformation, run two regressions. First regress **lcosts** on **lkwh** and an intercept. Next regress **lcosts** on **lkwh, lpl, lpk, lpf** and an intercept. Compare the coefficient of **lkwh** in the two regressions. Explain why the two estimates are different. (8 marks)

(c) Using the $R^2$ measure of goodness of fit would you say that the first regression in (b) provides a satisfactory fit to the data? Explain. (3 marks)

(d) Generate the fitted values of the dependent variable in the first regression of part (b). Provide a scatter plot of the fitted and observed values of **lcosts.** Comment on the goodness of fit of the model. (10 marks)



(e) Generate a scatter plot of observed values of **lcosts** and the fitted values of the second regression in part (b). Interpret the results. What information does the scatter plot provide regarding the fit of the model? (8 marks)





(f) In the conventional specification of this model it is a standard assumption that the error term $u$ has log-normal distribution, i.e. $\ln(u) \sim \mathcal{N}(0, \sigma^2)$. What advantages does this assumption have? (5 marks)

(g) Suppose we change the functional form of the cost function to

$$C = ky^{1/r} p_1^{\alpha_1/r} p_2^{\alpha_2/r} p_3^{\alpha_3/r} + u,$$

in which the error term enters additively and is assumed to have $N(0, \sigma^2)$ distribution. Is ordinary least squares an appropriate estimator? Justify your answer. (3 marks)

(h) An investigator proposes the following alternative specification:

$$\frac{C}{y} = kp_1^{\alpha_1} p_2^{\alpha_2} p_3^{\alpha_3} u.$$

What advantages if any does the original (with multiplicative error) functional form have relative to this one? (4 marks)

(i) Estimate this alternative model and interpret the regression results. (4 marks)