Alastair Hall

ECON61001:
Econometric Methods

Semester 1, 2020-21

Problem Set for Tutorial 6

In some cases, our data are group averages. In this first question, you explore the use of OLS and GLS to estimate a simple linear regression model when the data take this form. As will emerge this is a setting in which heteroscedasticity naturally arises even if the errors in the individual level data are homoscedastic.

1. Suppose that

$$y_i \;=\; \beta_{0,1} \,+\, \beta_{0,2} h_i \,+\, u_i \;=\; x_i' \beta_0 \,+\, u_i, \qquad i = 1, 2, \ldots N \tag{1}$$

where h_i is scalar, and $x_i = (1, h_i)'$, u_i satisfy Assumptions CS1-CS5 in Lecture 4 (or equivalently in Section 3.2 of the Lecture Notes).

Suppose that the observations are collected into $\mathcal{G}$ groups as follows. Group 1 consists of observations $i = 1, 2, \ldots, N_1$, group 2 consists of observations $i = N_1 + 1, N_1 + 2, \ldots, N_2$, and so on with group $\mathcal{G}$ consisting of observations $i = N_{\mathcal{G}-1} + 1, N_{\mathcal{G}-1} + 2, \ldots N$. This structure can be presented in generic notation as follows. Group g consists of observations $i = N_{g-1} + 1, N_{g-1} + 2, \ldots, N_g$ for $g = 1, 2, \ldots, \mathcal{G}$ where we set $N_0 = 0$ and $N_{\mathcal{G}} = N$.

Now consider the case where the researcher only observes the the group average data,

$$\bar{y}_g = n_g^{-1} \sum_{i=N_{g-1}+1}^{N_g} y_i, \qquad \bar{h}_g = n_g^{-1} \sum_{i=N_{g-1}+1}^{N_g} h_i, \quad \text{for } n_g = N_g - N_{g-1} \text{ and } g = 1, 2, \ldots, \mathcal{G},$$

and estimates the regression model

$$\bar{y}_g \;=\; \beta_{0,1} \,+\, \beta_{0,2} \bar{h}_g \,+\, v_g \qquad g = 1, 2 \ldots \mathcal{G}, \tag{2}$$

where v_g denotes the error term.

(a) Derive $E[v_g | \bar{h}_g]$ and $Var[v_g | \bar{h}_g]$.
Hint: (i) if (1) and (2) hold then v_g is a function of $\{u_i\}$ - what function? (ii) use the version of the LIE-II in Lemma 3.9 of the Lecture Notes with "G"= $\bar{h}_g$, "H"= $\{h_i;\ i = N_{g-1}+1,\ N_{g-1}+2,\ \ldots, N_g\}$, and note that expectations conditional on "G" and "H" is the same as expectations conditional on "H" in this case.

(b) Derive $E[v | \bar{X}]$ and $Var[v | \bar{X}]$ where $v = (v_1, v_2, \ldots v_{\mathcal{G}})'$ and $\bar{X}$ is the $\mathcal{G} \times 2$ matrix with g^{th} row $(1, \bar{h}_g)$.

(c) What are the properties of the OLS estimators of $\beta_0 = (\beta_{0,1}, \beta_{0,2})'$ based on (2)?

(d) What is the GLS estimator of β_0 in (2)? Is it a feasible estimator?

In this question, you explore the connection between linear models with parameter variation and linear regression models with heteroscedasticity.

2. Consider the model
$$y_i \;=\; x_i^{'}\beta_i \tag{3}$$
in which $\beta_i|x_i \sim N(\beta_0, \sigma_0^2 I_K)$. Rewrite (3) in the standard linear regression model framework: $y_i = x_i'\beta_0 + u_i$. What are the mean and variance of the error term of u_i conditional on x_i? *Hint: Substitute for β_i in (3).*

In this question, you explore the finite sample properties of the WLS estimator.

3. Consider the linear regression model
$$y \;=\; X\beta_0 \,+\, u \tag{4}$$
in which Assumptions CA1-CA4, CA5-H and CA6 hold. Let $\hat{\beta}_W = (X'W_2X)^{-1}X'W_2y$ be the Weighted Least Squares estimator of β_0 based on (4) and with $W_2 = diag(w_1^2, w_2^2, \ldots, w_N^2)$ for positive constants $\{w_i\}_{i=1}^N$.

   (a) Show that
   $$\hat{\beta}_W = \beta_0 \,+\, (X'W_2X)^{-1}X'W_2u$$
   (b) Show that $E[\hat{\beta}_W] = \beta_0$.

   (c) Show that $Var[\hat{\beta}_W] = (X'W_2X)^{-1}X'W_2\Sigma W_2X(X'W_2X)^{-1}$, where $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$.

   (d) Show that $\hat{\beta}_W \sim N\left(\beta_0, Var[\hat{\beta}_W]\right)$.

   (e) Suppose now that the regressors are stochastic and conditions *SR1-SR4* hold. Assuming $E[\hat{\beta}_W]$ exists, is $\hat{\beta}_W$ an unbiased estimator for β_0? Explain briefly.

In ths question, you consider the large sample properties of the WLS estimator and an associated test statistic.

4. Consider the model
$$y_i \;=\; x_i'\beta_0 \,+\, u_i$$
where Assumptions *CS1-CS4* and *CS5-H* hold. Let $\hat{\beta}_W$ be the WLS estimator defined in Question 3.

(a) Show that

$$\hat{\beta}_W = \beta_0 + \left(\sum_{t=1}^{N} \breve{x}_i \breve{x}_i' \right)^{-1} \sum_{t=1}^{N} \breve{x}_i \breve{u}_i$$

where $\breve{x}_i = w_i x_i$ and $\breve{u}_i = w_i u_i$.

(b) Show that $\{ (\breve{x}_i', \breve{u}_i) \}_{t=1}^{N}$ form an independently but not identically distributed sequence.

(c) Assuming that

$$\begin{aligned} N^{-1} X' W_2 X &\xrightarrow{p} Q_w, \text{ a positive definite matrix of finite constants,} \\ N^{-1/2} X' W_2 u &\xrightarrow{d} N(0, \Omega_w), \end{aligned}$$

where $\Omega_w = plim_{N \to \infty} N^{-1} X' W_2 \Sigma W_2 X$, $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ and $\sigma_i^2 = h(x_i)$.

(i) Show that $\hat{\beta}_W$ is consistent for β_0.

(ii) Show that $N^{1/2}(\hat{\beta}_W - \beta_0) \xrightarrow{d} N(0, Q_w^{-1} \Omega_w Q_w^{-1})$.

(iii) Suppose it is desired to test $H_0 : R\beta_0 = r$ versus $H_A : R\beta_0 \neq r$ where R is a $n_r \times k$ matrix of constants with $rank\{R\} = n_r$ and r is a $n_r \times 1$ vector of constants. Propose a test statistic based on the WLS estimator and state its distribution under the null hypothesis.