Economics 671
Problem Set #5:

(**1**) Consider a regression model with just an intercept parameter:

$$y_i = \mu + \epsilon_i, \quad E(\epsilon|X) = 0, \quad E(\epsilon\epsilon'|X) = \Omega,$$

where

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

(1a) Given a random sample of n observations from this model, calculate the OLS estimator, $\hat{\mu}$ and derive its variance.

(1b) Calculate the GLS estimator, $\hat{\mu}_{GLS}$ and derive its variance, $\text{Var}(\hat{\mu}_{GLS})$.

(1c) Formally compare the variances in (1a) and (1b). When doing so, you might make note of the following inequality that orders arithmetic and harmonic means:

If $a_1, \cdots a_n$ are positive real numbers, then

$$\overline{a}_n \geq \frac{n}{\sum_{i=1}^n a_i^{-1}}.$$

(**2**) Go back to the cars2 data set used in the previous problem set. Now, consider the regression:

$$\begin{aligned} Price_i &= \beta_0 + \beta_1 NotReliable_i + \beta_2 RoadScore_i + \beta_3 MPG_i + & (1) \\ & \quad \beta_4 Zeroto60_i + \beta_5 German_i + \epsilon_i & (2) \end{aligned}$$

where

$$E(\epsilon|X) = 0, \quad E(\epsilon_i^2|X) = \exp(\alpha_0 + \alpha_1 German_i + \alpha_2 RoadScore_i + \alpha_3 Zeroto60_i).$$

Using methods discussed in class, estimate this model using FGLS. In particular, do the following:

(a) Estimate (1) using OLS and obtain the variance-covariance matrix of the OLS estimator.

(b) Estimate $\hat{\alpha}$ and comment on the results. Note: Here you can ignore an issues with the intercept estimator $\hat{\alpha}_0$ and not worry about the fact that the errors ϵ_i have been replaced with estimated values $\hat{\epsilon}_i$.

(c) Calculate the (asymptotic) variance-covariance matrix of the FGLS estimator and compare those results to what you obtained in part (a).

(**3**) Consider the simple regression model:

$$y_i = \beta_0 + \beta_1 x_i^* + \epsilon_i$$

where the "true" scalar x_i^* is never observed, but instead, we observe and use an error-ridden measure x_i where

$$x_i = \alpha_0 + \alpha_1 x_i^* + u_i.$$

Suppose we assume that the error of measurement u is independent of x^* and ϵ and ϵ is also independent of x^*. In this case, derive

$$\text{plim}(\hat{\theta}_1) \quad \text{and} \quad \text{plim}(\hat{\theta}_0)$$

obtained from the model

$$y_i = \theta_0 + \theta_1 x_i + \eta_i.$$

Comment on the role of α_0 on these results, and whether (and perhaps when) the standard attenuation result will hold.

Questions to think about (not to be turned in, nor will solutions be provided):

(4) Suppose you have a regression model where a single (i.e., a scalar) right-hand side variable, say x_1, is believed to be endogenous. Specifically, you seek to estimate the model

$$y = \beta_0 + \beta_1 x_1 + u, \quad E(u|x_1) \neq 0.$$

In addition, suppose you know the following:

- A scalar variable z_1 is a valid instrument and is, in fact, independent of the regression error u
- A scalar variable w is known to be an invalid instrument (i.e., it is known to be correlated with u).
- A scalar variable z_2, which is potentially an instrument, is known to be independent of z_1 (the valid instrument) and is also known to be correlated with w (the invalid instrument).

Do the above conditions imply that z_2 is an invalid instrument? Prove or explain in detail why not.

(5:) Consider the *simple* regression model (i.e., x_i is a scalar):

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where x_i is regarded as potentially endogenous, meaning $E(\epsilon|x) \neq 0$.

Furthermore, there is a *BINARY* variable (dummy variable) z_i, $z_i \in \{0, 1\}$ that is corelated with x and satisfies $E(\epsilon|z) = 0$.

(5a:) Show that

$$\beta_1 = \frac{E(y|z=1) - E(y|z=0)}{E(x|z=1) - E(x|z=0)}.$$

(5b:) Based on the result in (5a), suggest a simple, consistent estimator of β_1.

(5c:) Now, consider the estimator (known as the instrumental variables estimator) of $\beta = [\beta_0\ \beta_1]'$:

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y,$$

where

$$Z \equiv \begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_n \end{bmatrix}, \quad X \equiv \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Show that the IV estimator of the *slope coefficient* (β_1), from the above, reduces to a natural estimator of the expression in (5a).