New York University
Applied Statistics and Econometrics I
Fall 2015

# Problem Set 5

*Cristian Fuenzalida*
*cristian.fuenzalida@nyu.edu*

**(I)** Consider a linear model with measurement error:

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

Here, $x^*$ is not observed directly. Instead, we only have an imperfect measure $x$ (and this is what we can use to run our regression). The relationship between $x$ and $x^*$ is:

$$x_i = x_i^* + u_i$$

We assume that $u$ and $\epsilon$ are uncorrelated.

1. Assume that $\mathbb{E}(x^*) = \mu^*$, $\mathbb{E}(u) = \mathbb{E}(\epsilon) = 0$, and $Var(x^*) = \sigma_*^2, Var(u) = \sigma_u^2, Var(\epsilon) = \sigma_\epsilon^2$. The covariances between these r.v. are zero. Get an expression for the probability limit of $\beta_0$ and $\beta_1$.
2. As an alternative procedure, consider regressing $x$ on a constant and $y$, and then computing the reciprocal of the estimate. Calculate the probability limit of the new estimator.
3. Can you use the estimators in part 1 and 2 to bound the true value of the coefficient?

**(II)** Prove the following properties of the GLS estimator $\hat{\beta}_{GLS}$. In each case, specify if further assumptions are needed to state the result:

1. $\hat{\beta}_{GLS}$ minimizes $(y - X\tilde{\beta})\Omega^{-1}(y - X\tilde{\beta})$.
2. $\hat{\beta}_{GLS}$ is unbiased.

**(III)** Prove that the White estimator of the covariance matrix:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \hat{e}_i^2$$

is consistent.

**(IV)** For the generalized regression model, prove the following: for any unbiased estimator $\tilde{\beta}$, define $q := \tilde{\beta} - \hat{\beta}_{GLS}$. Assume that $\tilde{\beta}$ is such that $V_q := Var(q)$ is nonsingular and positive nondefinite. Prove that:

$$Cov(\hat{\beta}_{GLS}, q) = 0$$

**(V)** We have a model with one endogenous regressor:

$$y_i = \theta z_i + \epsilon_i$$

Suppose we are given an instrument $x_i$ satisfying:

$$z_i = \delta x_i + \nu_i$$

Define:

$$s_{xz} = \frac{1}{n}\sum_{i=1}^{n} x_i z_i, \qquad s_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

1. Show that $\sigma_{xz}^2 := \mathbb{E}(x_i z_i) \neq 0$, and the IV estimator $\theta_{IV}$ is consistent for $\theta$. Express $\theta_{IV}$ as a function of $s_{xz}$ and $s_{xy}$ (actually you will need this to prove consistency anyway).
2. Now suppose that $z_i = \delta_n x_i + \nu_i$ with $\delta_n = 1/\sqrt{n}$. Prove that $s_{xz} \xrightarrow{p} 0$ and that $\sqrt{n} s_{xz} \xrightarrow{d} \sigma_x^2 + \aleph$, where $\aleph \sim N(0, \mathbb{E}(x_i \nu_i)^2)$. Is $\hat{\theta}$ consistent? Interpret this result and relate it to the problem of *weak instruments*.

**(VI)** Write a two-equation system in the form:

$$y_1 = \gamma_1 y_2 + \mathbf{z}_{(1)}\delta_{(1)} + u_1$$

$$y_2 = \gamma_2 y_1 + \mathbf{z}_{(2)}\delta_{(2)} + u_2$$

1. Show that reduced forms exist if and only if $\gamma_1\gamma_2 \neq 1$.
2. State in words the rank condition for identifying each equation.

**(VII)** Consider a SUR model:

$$y_{i1} = \beta_1 x_{i1} + \epsilon_{i1}$$

$$y_{i2} = \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_{i2}$$

Assume that:

$$\Sigma = \begin{pmatrix} \mathbb{E}(\epsilon_{i1}^2) & \mathbb{E}(\epsilon_{i1}\epsilon_{i2}) \\ \mathbb{E}(\epsilon_{i1}\epsilon_{i2}) & \mathbb{E}(\epsilon_{i2}^2) \end{pmatrix}$$

is a known matrix. Now imagine SUR is implemented, but the variable $x_{i3}$ has been (erroneously) omitted. Is the estimator of $\beta_1$ consistent?

**(VIII)** Consider a linear model with measurement error:

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

Here, $x^*$ is not observed directly. Instead, we only have an imperfect measure $x$ (and this is what we can use to run our regression). The relationship between $x$ and $x^*$ is:

$$x_i = x_i^* + u_i$$

We assume that $u$ and $\epsilon$ are uncorrelated.

1. Assume that $\mathbb{E}(x^*) = \mu^*$, $\mathbb{E}(u) = \mathbb{E}(\epsilon) = 0$, and $Var(x^*) = \sigma_*^2, Var(u) = \sigma_u^2, Var(\epsilon) = \sigma_\epsilon^2$. The covariances between these r.v. are zero. Get an expression for the probability limit of $\beta_0$ and $\beta_1$.
2. As an alternative procedure, consider regressing $x$ on a constant and $y$, and then computing the reciprocal of the estimate. Calculate the probability limit of the new estimator.
3. Can you use the estimators in part 1 and 2 to bound the true value of the coefficient?

**(IX)** We have a two equation model:

$$y_1 = \beta_1 x + \epsilon_1$$

$$y_2 = \beta_2 x_2 + \beta_3 x_3 + \epsilon_2$$

Assume that the error covariances are known. Suppose that we applies GLS to this model but we omit the variable $x_3$ from the second equation. What effect does this specification error have on the consistency of the estimator of $\beta_1$?

**(X)** We are interested in estimating a system of equations:

$$y_i = X_i \beta + u_i$$

Here, $i$ is an index for the cross sectional observation, $y_i$ and $u_i$ are of dimension $G \times 1$, $X_i$ is $G \times K$, $Z_i$ is the $G \times L$ matrix of instruments, and $\beta$ is $K \times 1$. Denote by $\Omega$ the covariance matrix of the errors $u_i$. Assume that $\mathbb{E}(Z_i' u_i) = 0$, the rank of $\mathbb{E}(Z_i' X_i)$ is K and both $\mathbb{E}(Z_i' Z_i)$ and $\mathbb{E}(Z_i' \Omega Z_i)$ are nonsingular.

1. What are the main properties of the 3SLS estimator?
2. Find the asymptotic variance matrix of $\sqrt{N}(\hat{\beta}_{3SLS} - \beta)$.
3. How would you estimate Avar($\hat{\beta}_{3SLS}$)?

**(XI)** We are trying to estimate the model:

$$y_i = \beta z_i^2 + \epsilon_i$$

But the researcher suspects there could be an endogeneity problem with this regression.

1. Propose a test to test if there is endogeneity in the regression. Tell exactly what is the statistic to compute the null hypothesis and its distribution. Explain how you would carry out the test.
2. Suppose the researcher has an instrumental $x$ such that:

$$z_i = \gamma x_i + \eta_i$$

with $\mathbb{E}((\epsilon, \eta)|x) = 0$ and $Var(\epsilon, \eta)|x) = \Sigma$. Here $\Sigma$ is unknown. A research assistant suggests to estimate the following way: first, regress $z$ on $x$ using OLS. Then, use the predicted values $\hat{z}_i$ in the first equation (that is, regress $y$ on $\hat{z}_i$) to obtain $\hat{\beta}$ by OLS. Is this estimator consistent?

3. Indicate how you would estimate this equation by 2SLS. How does affect this estimation the fact that $\Sigma$ is unknown?

**(XII)** An i.i.d. sample is given $\{x_i\}_{i=1}^{n}$. Define:

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(x_i \leq x)$$

as the empirical distribution function of $x_i$. Now consider two density estimators:

$$\hat{f}_1(x) = \frac{\hat{F}(x+h) - \hat{F}(x)}{h}$$

$$\hat{f}_2(x) = \frac{\hat{F}(x+h/2) - \hat{F}(x-h/2)}{h}$$

1. Prove that $\hat{F}(x)$ is an unbiased estimator of $F(x)$.
2. The bias of $\hat{f}_1(x)$ is $O(h^a)$. Find the value of $a$.
3. The bias of $\hat{f}_2(x)$ is $O(h^b)$. Find the value of $b$.