

Homework 5

Problem 1. In this question, you will derive the asymptotic distribution of the OLS estimator under endogeneity. Consider the usual linear regression model $Y_i = X_i' \beta + U_i$, where β is a $k \times 1$ vector. Assume, however, that X_i 's are endogenous:

$$\mathbb{E} X_i U_i = \mu \neq 0,$$

where μ is an unknown $k \times 1$ vector. Let $\hat{\beta}_n$ denote the OLS estimator of β . Make the following additional assumptions:

A1. Data are iid.

A2. $Q = \mathbb{E} X_i X_i'$ is finite and positive definite.

A3. $\mathbb{E} (U_i - X_i' \delta)^2 X_i X_i'$ is finite and positive definite, where $\delta = Q^{-1} \mu$.

1. Find the probability limit of $\hat{\beta}_n$.
2. Re-write the model as $Y_i = X_i'(\beta + \delta) + (U_i - X_i' \delta)$ and find $\mathbb{E} X_i (U_i - X_i' \delta)$.
3. Using the result in (ii), derive the asymptotic distribution of $\hat{\beta}_n$ and find its asymptotic variance. Explain how this result differs from the asymptotic normality of OLS with exogenous regressors. Hint: To establish asymptotic normality, $\hat{\beta}_n$ must be properly re-centered based on the result in (i).
4. Can $\hat{\beta}_n$ and its asymptotic distribution be used for inference about β ? Explain why or why not.
5. Suppose that the errors U_i 's are homoskedastic:

$$\mathbb{E} (U_i^2 | X_i) = \sigma^2 = \text{const.}$$

Consider the usual estimator of the asymptotic variance of OLS designed for a model with homoskedastic errors and exogenous regressors:

$$n^{-1} \sum_{i=1}^n (Y_i - X_i' \hat{\beta}_n)^2 \left(n^{-1} \sum_{i=1}^n X_i X_i' \right)^{-1}.$$

Is it consistent for the asymptotic variance of the OLS estimator if X_i 's are in fact endogenous? Explain why or why not.

6. Continue to assume that U_i 's are homoskedastic as in (v). Consider the usual heteroskedasticity-robust asymptotic variance estimator designed for a model with exogenous regressors:

$$\left(n^{-1} \sum_{i=1}^n X_i X_i' \right)^{-1} \left(n^{-1} \sum_{i=1}^n (Y_i - X_i' \hat{\beta}_n)^2 X_i X_i' \right) \left(n^{-1} \sum_{i=1}^n X_i X_i' \right)^{-1}.$$

Is it consistent for the asymptotic variance of the OLS estimator if X_i 's are in fact endogenous? Explain why or why not.

Problem 2. Consider the linear regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where \mathbf{X} is the $n \times k$ matrix of regressors, \mathbf{Y} is the n -vector of observations on the dependent variable, and $\beta \in \mathbb{R}^k$ is the vector of unknown parameters. Let \mathbf{Z} be the $n \times k$ matrix of instruments. Assume that:

- \mathbf{X} and \mathbf{Z} are strongly exogenous: $\mathbb{E}(\mathbf{e} | \mathbf{X}, \mathbf{Z}) = \mathbf{0}$.

- e is homoskedastic: $\mathbb{E}(ee'|X, Z) = \sigma^2 I_n$.
- X and $Z'X$ have rank k .

Let $\hat{\beta} = (X'X)^{-1} X'Y$ and $\tilde{\beta} = (Z'X)^{-1} Z'Y$ be the OLS and IV estimators of β respectively.

1. Show that $\mathbb{E}(e|X) = 0$ and $\mathbb{E}(ee'|X) = \sigma^2 I_n$.
2. Show that the OLS and IV estimators are unbiased.
3. Find the exact finite sample conditional variances of $\hat{\beta}$ and $\tilde{\beta}$: $\text{Var}(\hat{\beta}|X, Z)$ and $\text{Var}(\tilde{\beta}|X, Z)$. Show that

$$\begin{aligned} & \text{Var}(\tilde{\beta}|X, Z) - \text{Var}(\hat{\beta}|X, Z) \\ &= \sigma^2 (Z'X)^{-1} Z' (I_n - X(X'X)^{-1} X') Z (X'Z)^{-1}. \end{aligned}$$

4. When regressors are exogenous, should the econometrician use IV or OLS ? Explain why using the result in in part (iii).

Problem 3. Consider the model

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + e_i \\ \mathbb{E}(e_i) &= 0 \\ \mathbb{E}(X_i e_i) &= 0 \end{aligned}$$

with both Y_i and X_i scalar. Assume $\beta_0 > 0$ and $\beta_1 < 0$. Suppose the parameter of interest is the area under the regression curve (e.g., consumer surplus), which is $A = -\frac{\beta_0^2}{2\beta_1}$. Let $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1)'$ be the LS estimator of $\theta = (\beta_0, \beta_1)'$ so that $\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N(0, V_\theta)$. Let \hat{V}_θ be a standard consistent estimator for V_θ . You do not need to write out these estimators.

1. Given the above, describe an estimator of A .
2. Construct an asymptotic $1 - \alpha$ coverage probability confidence interval for A .
3. Construct an asymptotic $1 - \alpha$ coverage probability bootstrap percentile confidence interval for A .

Problem 4. Consider a regression model with potentially endogenous regressors:

$$Y_i = X_i' \beta + U_i, \quad \beta \in \mathbb{R}^k.$$

Let Z_i be the l -vector of instruments such that $l \geq k$,

$$\begin{aligned} \text{rank}(\mathbb{E}Z_i X_i') &= k, \\ \mathbb{E}Z_i U_i &= 0. \end{aligned}$$

Let R be a $q \times k$ matrix of rank q , and let r be a $q \times 1$ vector; both R and r are known. Let W_n be an $l \times l$ matrix such that

$$W_n \rightarrow_p W,$$

where W is symmetric and positive definite. Let $\tilde{\beta}_n$ be the restricted GMM estimator: $\tilde{\beta}_n$ minimizes the GMM criterion function $(Y - Xb)' Z W_n Z' (Y - Xb)$ subject to the restriction $Rb - r = 0$, where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix}, \quad Z = \begin{pmatrix} Z_1' \\ \vdots \\ Z_n' \end{pmatrix}.$$

1. Show that $\tilde{\beta}_n$ solves

$$-X'ZW_nZ'(Y - X\tilde{\beta}_n) + R'\tilde{\lambda}_n = 0,$$

where $\tilde{\lambda}_n$ is the q -vector of Lagrange multipliers.

2. Show that

$$\tilde{\beta}_n = \hat{\beta}_n - (X'ZW_nZ'X)^{-1}R'\tilde{\lambda}_n,$$

where $\hat{\beta}_n$ is the unconstrained GMM estimator, i.e.

$$\hat{\beta}_n = \arg \min_{b \in \mathbb{R}^k} (Y - Xb)'ZW_nZ'(Y - Xb).$$

3. Using the the result from (ii) and the fact that $\tilde{\beta}_n$ satisfies the constraint, show that

$$\tilde{\beta}_n = \hat{\beta}_n - (X'ZW_nZ'X)^{-1}R'(R(X'ZW_nZ'X)^{-1}R')^{-1}(R\hat{\beta}_n - r).$$

4. Suppose that data are iid, and the instruments and regressors have finite second moments. Find the probability limit of the restricted GMM estimator, i.e. find the expression for β^* in

$$\tilde{\beta}_n \rightarrow_p \beta^*.$$

Under what condition the restricted GMM estimator $\tilde{\beta}_n$ is consistent?

5. Suppose that $R\beta = r$. Find the probability limit of $\tilde{\lambda}_n/n^2$. Explain how the result can be used for testing $H_0 : R\beta = r$ against $H_1 : R\beta \neq r$. You do not have to figure out the details of such a test, only to explain why the probability limit of $\tilde{\lambda}_n/n^2$ is useful for construction of the test.

Problem 5. Consider the following regression model:

$$Y = X\beta + U,$$

where Y is an $n \times 1$ vector of observations on the dependent variable and X is an $n \times k$ matrix of observations on the regressors. Let Z be an $n \times l$ matrix of observations on the instruments, $l \geq k$. The 2SLS estimator of β can be written as $\tilde{\beta} = (X'P_ZX)^{-1}X'P_ZY$, where $P_Z = Z(Z'Z)^{-1}Z'$. Let $\hat{\beta}$ be the OLS estimator of the coefficients on X in the regression of Y against X and \hat{V} :

$$Y = X\beta + \hat{V}\gamma + U,$$

where \hat{V} is the matrix of the fitted residuals from the regression of X against Z ,

$$X = Z\hat{\Pi} + \hat{V},$$

and $\hat{\Pi} = (Z'Z)^{-1}Z'X$ is the OLS estimator from the regression of X against Z . Show that $\tilde{\beta} = \hat{\beta}$ by following the steps below:

1. Use the partitioned regression result to write $\tilde{\beta} = (X'MX)^{-1}X'MY$, and define the matrix M in terms of \hat{V} .
2. Using the definition of M from part (i) and the definition of \hat{V} , show that $X'MX = X'P_ZX$.
3. Repeat the same steps as in (ii) to show that $X'MY = X'P_ZY$.