Homework 6

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:

$$\mathbb{E}X_iU_i = \mu \neq 0,$$

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. $Q = \mathbb{E}X_iX_i'$ is finite and positive definite.
A3. $\mathbb{E}\left(U_i - X_i'\delta\right)^2 X_iX_i'$ 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:
$$\mathbb{E}\left(U_i^2|X_i\right) = \sigma^2 = const.$$
Consider the usual estimator of the asymptotic variance of OLS designed for a model with homoskedastic errors and exogenous regressors:
$$n^{-1}\sum_{i=1}^{n}\left(Y_i - X_i'\hat{\beta}_n\right)^2\left(n^{-1}\sum_{i=1}^{n}X_iX_i'\right)^{-1}.$$
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:
$$\left(n^{-1}\sum_{i=1}^{n}X_iX_i'\right)^{-1}\left(n^{-1}\sum_{i=1}^{n}\left(Y_i - X_i'\hat{\beta}_n\right)^2 X_iX_i'\right)\left(n^{-1}\sum_{i=1}^{n}X_iX_i'\right)^{-1}.$$
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 $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$, where $\boldsymbol{X}$ is the $n \times k$ matrix of regressors, $\boldsymbol{Y}$ is the n-vector of observations on the dependent variable, and $\boldsymbol{\beta} \in \mathbb{R}^k$ is the vector of unknown parameters. Let $\boldsymbol{Z}$ be the $n \times k$ matrix of instruments. Assume that:

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

- $\boldsymbol{e}$ is homoskedastic: $\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X},\boldsymbol{Z}\right)=\sigma^2\boldsymbol{I}_n$.
- $\boldsymbol{X}$ and $\boldsymbol{Z}'\boldsymbol{X}$ have rank k.

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

1. Show that $\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}\right)=\boldsymbol{0}$ and $\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}\right)=\sigma^2\boldsymbol{I}_n$.
2. Show that the OLS and IV estimators are unbiased.
3. Find the exact finite sample conditional variances of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$: $\operatorname{Var}\left(\hat{\boldsymbol{\beta}}|\boldsymbol{X},\boldsymbol{Z}\right)$ and $\operatorname{Var}\left(\tilde{\boldsymbol{\beta}}|\boldsymbol{X},\boldsymbol{Z}\right)$. Show that
$$\begin{aligned}&\operatorname{Var}\left(\tilde{\boldsymbol{\beta}}|\boldsymbol{X},\boldsymbol{Z}\right)-\operatorname{Var}\left(\hat{\boldsymbol{\beta}}|\boldsymbol{X},\boldsymbol{Z}\right)\\ =&\sigma^2\left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\left(\boldsymbol{I}_n-\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)\boldsymbol{Z}\left(\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}.\end{aligned}$$
4. When regressors are exogenous, should the econometrician use IV or OLS ? Explain why using the result in in part (iii).

Problem 3. Consider a simple IV regression model:

$$Y_i=\beta X_i+U_i,$$

where X_i is a single regressor, i.e. $\beta\in\mathbb{R}$. Let Z_i be a single exogenous IV:

$$\mathbb{E}Z_iU_i=0,$$

however, assume that Z_i is an irrelevant instrument in the sense that:

$$\mathbb{E}Z_iX_i=0.$$

Assuming that data $\{(Y_i,X_i,Z_i):i=1,\ldots,n\}$ are iid and that the following 2×2 matrix is finite and positive definite

$$\begin{pmatrix}\mathbb{E}(U_i^2Z_i^2) & \mathbb{E}(U_iX_iZ_i^2)\\ \mathbb{E}(U_iX_iZ_i^2) & \mathbb{E}(X_i^2Z_i^2)\end{pmatrix},$$

derive the distribution of the IV estimator by following the steps below.

1. Show that
$$n^{-1/2}\sum_{i=1}^{n}Z_i\begin{pmatrix}U_i\\ X_i\end{pmatrix}\to_d\begin{pmatrix}\Psi_U\\ \Psi_X\end{pmatrix},$$
where Ψ_U and Ψ_X are two random variables following a bivariate normal distribution.
2. Using the result in (i) and the continuous mapping theorem, derive the asymptotic distribution of
$$\hat{\beta}_n^{IV}-\beta,$$
where $\hat{\beta}_n^{IV}$ is the IV estimator of β.
3. Is $\hat{\beta}_n^{IV}$ consistent in this case? Explain why or why not.

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 $l \geq k$,

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

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

$$W_n \to_p W,$$

where W is symmetric and positive definite. Let $\tilde{\beta}_n$ be the restricted GMM estimator: $\tilde{\beta}_n$ minimizes the GMM criterion function $(Y - Xb)'ZW_nZ'(Y - Xb)$ 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
$$\tilde{\beta}_n \to_p \beta^*.$$
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 $\hat{\beta} = (X'P_ZX)^{-1}X'P_ZY$, where $P_Z = Z(Z'Z)^{-1}Z'$. Let $\tilde{\beta}$ 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 $X'MX = X'P_Z X$.

3. Repeat the same steps as in (ii) to show that $X'MY = X'P_Z Y$.