Advanced Econometrics

Professor: Julie Shi

Timekeeper: Rui Zhou

Fall 2023

Contents

4 Endogeneity and Instrumental Variable

4.1 Violation of Exogeneity

Three basic assumptions for linear model:

1. X has full rank
2. Exogeneity: $E[\varepsilon|X] = 0$
3. Homoskedasticity: $Var(\varepsilon|X) = E[\varepsilon\varepsilon'|X] = \sigma^2\mathbf{I}$.

Endogeneity happens when exogeneity is violated, i.e., $E[\varepsilon|X] \neq 0$. Sources of endogeneity usually are:

- Omitted variable bias: endogenous treatment effects, omitted parameter heterogeneity
- Simultaneous/Reverse causality
- Measurement error
- Sample selection bias: non-random sampling, attrition (survivorship bias)

In violation of exogeneity, there must be some correlation between the disturbances and the independent variables. We assume:

$$E[\varepsilon|X] = \eta$$

which means that the regressors now provide information about the expectations of the disturbances. An important implication of this assumption is that, the disturbances and the regressors are now correlated, specifically

$$E[x_i\varepsilon_i] = \gamma$$

for some nonzero γ. If the data are "well-behaved", then we can apply Khinchine's theorem to asset that

$$\text{plim}\ \frac{X'\varepsilon}{n} = \gamma$$

Khinchine's Weak Law of Large Numbers If $x_i (i = 1, 2, \cdots, n)$ is a random (i.i.d.) sample from a distribution with finite mean $E[x_i] = \mu$, then

$$\text{plim}\ \bar{x}_n = \mu$$

Then the estimator of b is biased:

$$\begin{aligned} b &= (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon \\ \Longrightarrow E[b|X] &= \beta + (X'X)^{-1}X'\eta \neq \beta \end{aligned}$$

Original OLS estimator b is also inconsistent:

$$\begin{aligned}\text{plim } b &= \beta + \text{plim } \left(\frac{X'X}{n}\right)^{-1} \cdot \text{plim } \left(\frac{X'\varepsilon}{n}\right) \\ &= \beta + Q_{XX}^{-1}\gamma \neq \beta\end{aligned}$$

The inconsistency of least squares is not confined to the endogenous variable(s). Even though only one of the variables in X is correlated with ε, all of the elements of b are inconsistent, not just the estimator of the coefficient on the endogenous variable. The inconsistency due to the endogeneity of the one variable is smeared across all of the least squares estimators.

4.2 IV Estimation

Luckily, beyond X, there is an additional set of variables, Z, that have two properties:

1. Exogeneity: IV uncorrelated with the disturbance.
2. Relevance: IV correlated with the independent variable X.

In the context of our model, variables that have these two properties are instrumental variables. We assume the following:

1. $[x_i, z_i, \varepsilon_i](i = 1, \cdots, n)$ are i.i.d. sequence of random variables.
2. $E[x_{ik}^2] = Q_{XX}$, a finite constant.
   - plim $\frac{X'X}{n} = Q_{XX}$
3. $E[z_{il}^2] = Q_{ZZ}$, a finite constant.
   - plim $\frac{Z'Z}{n} = Q_{ZZ}$
4. $E[z_{il}x_{ik}] = Q_{ZX}$, a finite constant.
   - plim $\frac{Z'X}{n} = Q_{ZX}$
5. $E[\varepsilon_i|z_i] = 0$
   - plim $\frac{Z'\varepsilon}{n} = 0$

The requirements for IV can also be described as:

1. Exogeneity: plim $\frac{Z'\varepsilon}{n} = 0$.
2. Relevance: plim $\frac{Z'X}{n} = Q_{XX}$, a finite $L \times K$ matrix with rank K.
   - Baseline for coefficient estimation: similar to full column rank of X before.

3. Well-behaved data: plim $\frac{Z'Z}{n} = Q_{ZZ}$, a positive definite matrix.

We discuss two cases, where $L = K$ or $L > K$.

4.2.1 $L = K$

First consider the situation when $L = K$. We partition X into x_1, a set of K_1 exogenous variables; and x_2, a set of K_2 endogenous variables, then $Z = [x_1, z_2]$, where z_2 are the instrumental variables for x_2, and x_1 are the instrumental variables for themselves.

Because $E[z_i \varepsilon_i] = 0$ and all terms have finite variances, we have:

$$\begin{aligned}
\text{plim } \left(\frac{Z'\varepsilon}{n}\right) &= \text{plim } \left(\frac{Z'(y - X\beta)}{n}\right) \\
&= \text{plim } \left(\frac{Z'y}{n}\right) - \text{plim } \left(\frac{Z'X\beta}{n}\right) \\
&= 0 \\
\Longrightarrow \text{plim } \left(\frac{Z'y}{n}\right) &= \text{plim } \left(\frac{Z'X\beta}{n}\right) = \beta \cdot \text{plim } \left(\frac{Z'X}{n}\right)
\end{aligned}$$

We have assumed that Z has the same number of variables as X, so $Z'X$ is a square matrix. Moreover, the rank of $Z'X$ is also K, therefore plim $\left(\frac{Z'X}{n}\right)$ is invertible. Then we have:

$$\left[\text{plim } \left(\frac{Z'X}{n}\right)\right]^{-1} \text{plim } \left(\frac{Z'y}{n}\right) = \beta$$

which leads us to the *instrumental variable estimator*:

$$b_{IV} = (Z'X)^{-1} Z'y$$

From the preceding deduction of IV estimator we have already proved that b_{IV} is consistent. Alternatively, in the traditional way we can see:

$$\begin{aligned}
b_{IV} &= (Z'X)^{-1} Z'y \\
&= (Z'X)^{-1} Z'(X\beta + \varepsilon) \\
&= \beta + (Z'X)^{-1} Z'\varepsilon
\end{aligned}$$

It is clear that b_{IV} is consistent since plim $\frac{Z'\varepsilon}{n} = 0$.

We now turn to the asymptotic distribution of b_{IV}. We will use the same method as in previous asymptotic distribution of b_{LS}. First,

$$\sqrt{n}(b_{IV} - \beta) = \left(\frac{Z'X}{n}\right)^{-1} \frac{1}{\sqrt{n}} Z'\varepsilon$$

which has the same limiting distribution as $Q_{ZX}^{-1} \cdot \frac{1}{\sqrt{n}} Z'\varepsilon$. Analysis of distribution of $Z'\varepsilon$ can be the same as that of $X'\varepsilon$, so it follows that:

$$\begin{aligned}
&\left(\frac{1}{\sqrt{n}} Z'\varepsilon\right) \xrightarrow{\text{d}} N\left[0, \sigma^2 Q_{ZZ}\right] \\
\Longrightarrow &\left(\frac{Z'X}{n}\right)^{-1} \left(\frac{1}{\sqrt{n}} Z'\varepsilon\right) \xrightarrow{\text{d}} N\left[0, \sigma^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right]
\end{aligned}$$

Thus, the asymptotic distribution of b_{IV} is

$$b_{IV} \overset{a}{\sim} N\left[\beta, \frac{\sigma^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right]$$

The asymptotic covariance matrix is estimated as:

$$\begin{aligned}
Est.Asy.Var[b_{IV}] &= \hat{\sigma}^2 (Z'X)^{-1} (Z'Z)(X'Z)^{-1} \\
&= \hat{\sigma}^2 \left[X'Z(Z'Z)^{-1} Z'X\right]^{-1}
\end{aligned}$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i' b_{IV}\right)^2$$

$\hat{\sigma}^2$ is a *consistent* estimator of σ^2. A correction for degrees of freedom is superfluous here, as all results here are asymptotic.

Equivalently, we can see the variance matrix of b_{IV} from its expression:

$$\begin{aligned}
b_{IV} &= (Z'X)^{-1} Z'y = (Z'X)^{-1} Z'(X\beta + \varepsilon) = \beta + (Z'X)^{-1} Z'\varepsilon \\
\Longrightarrow Var(b_{IV}|X) &= (Z'X)^{-1} Z' Var(\varepsilon|X) Z (X'Z)^{-1} \\
&= \sigma^2 (Z'X)^{-1} Z'Z (X'Z)^{-1} \\
&= \sigma^2 [X'Z(Z'Z)^{-1} Z'X]^{-1}
\end{aligned}$$

In sum, IV estimation share some technical part in common with OLS. See motivation of IV from linear model:

$$\begin{aligned}
&y = X\beta + \varepsilon \\
\Longrightarrow &X'y = X'X\beta + X'\varepsilon \\
\Longrightarrow &\text{plim } \frac{X'y}{n} = \text{plim } \frac{X'X}{n} \cdot \beta + \text{plim } \frac{X'\varepsilon}{n} \\
\Longrightarrow &Q_{Xy} = Q_{XX}\beta + \gamma
\end{aligned}$$

From $Q_{Xy} = Q_{XX}\beta + \gamma$ we can see that, β and γ cannot be jointly identified without further assumptions. In standard OLS assumption, $\gamma = 0$ is a key assumption of exogeneity, then β can be estimated. However in violation of exogeneity, we introduce IV with assumption of plim $\frac{Z'\varepsilon}{n} = 0$ for estimation:

$$
\begin{aligned}
& y = X\beta + \varepsilon \\
\Longrightarrow & Z'y = Z'X\beta + Z'\varepsilon \\
\Longrightarrow & \text{plim } \frac{Z'y}{n} = \text{plim } \frac{Z'X}{n} \cdot \beta + \text{plim } \frac{Z'\varepsilon}{n} \\
\Longrightarrow & Q_{Zy} = Q_{ZX}\beta
\end{aligned}
$$

4.2.2 $L > K$

The crucial results in all the preceding of IV is based on

$$
\text{plim } \frac{Z'\varepsilon}{n} = \mathbf{0}
$$

That is, every column of Z is asymptotically uncorrelated with ε. That also means that every linear combination of the columns of Z is also uncorrelated with ε, which suggests that one approach would be to choose K linear combinations of the columns of Z. Specially when we consider the situation when $L > K$, i.e., the instrumental variables are more than the endogenous variables. To make use of the L-many IVs, a better choice is the **projection of the columns of X in the column space of Z**:

$$
\hat{X} = Z(Z'Z)^{-1}Z'X
$$

Intuitively, $\hat{X}$ is the predicted value of X when regressing X on Z. After projected into column space of Z, $\hat{X}$ is exogenous and uncorrelated with the disturbance. From our motivation, $\hat{X}$ has K columns of IVs that are combinations of Z. We then have (substitute $\hat{X}$ for "Z" in expression of b_{IV} since $\hat{X}$ is a set of K IVs):

$$
\begin{aligned}
b_{IV} &= (\hat{X}'X)^{-1}\hat{X}'y \\
&= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'y
\end{aligned}
$$

The variance matrix of b_{IV} is the same as before,

$$
\begin{aligned}
b_{IV} &= (\hat{X}'X)^{-1}\hat{X}'y \\
&= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'y \\
&= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'(X\beta + \varepsilon) \\
&= \beta + \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'\varepsilon \\
\Longrightarrow \text{Var}\left[b_{IV}\right] &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'\text{Var}\left[\varepsilon|X\right] Z(Z'Z)^{-1}Z'X \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} \\
&= \sigma^2 \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}
\end{aligned}
$$

Remarks: In computation of the estimated asymptotic covariance matrix, $\hat{\sigma}^2$ should not be based on $\hat{X}$. The estimator $s_{IV}^2 = \dfrac{\left(y - \hat{X}b_{IV}\right)'\left(y - \hat{X}b_{IV}\right)}{n}$ is inconsistent for σ^2, with or without a correction for degrees of freedom.

Alternatively, we can rewrite b_{IV} as

$$\begin{aligned} b_{IV} &= (\hat{X}'X)^{-1}\hat{X}'y \\ &= (X'P_Z'X)^{-1}\hat{X}'y \\ &= (X'P_Z'P_ZX)^{-1}\hat{X}'y \\ &= \left[(P_ZX)'(P_ZX)\right]^{-1}\hat{X}'y \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \end{aligned}$$

$b_{IV} = (\hat{X}'X)^{-1}\hat{X}'y$ has practical meanings, which is the theoretical foundation for Two Stage Least Squares (2SLS). In practice, b_{IV} can be estimated in two steps:

1. X is regressed on Z, and predict $\hat{X}$.
2. y is regressed on $\hat{X}$ to get $b_{IV} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$.

2SLS estimator is not only intuitive and easy to compute, but also efficient in the IV class of linear combinations of Z.

Efficiency of 2SLS Estimator Of all the different linear combinations of Z, $\hat{X}$ is the most efficient in the sense that the asymptotic covariance matrix of an IV estimator based on a linear combination ZF is smaller when $F = (Z'Z)^{-1}Z'X$ than with any other F that uses all L columns of Z.

Proof Denote a combination of columns of Z as $\tilde{Z} = ZF$, where F is a $L \times K$ matrix. IV estimator with $\tilde{Z}$ is

$$\begin{aligned} b_{\tilde{Z}} &= (\tilde{Z}'X)^{-1}\tilde{Z}'y \\ &= (\tilde{Z}'X)^{-1}\tilde{Z}'(X\beta + \varepsilon) \\ &= \beta + (\tilde{Z}'X)^{-1}\tilde{Z}'\varepsilon \end{aligned}$$

Therefore, the $(\tilde{Z}'X)^{-1}\tilde{Z}'\varepsilon$ part in $b_{\tilde{Z}}$ is its random part. We have

$$\begin{aligned} Var[b_{\tilde{Z}}|X] &= (\tilde{Z}'X)^{-1}\tilde{Z}'Var[\varepsilon|X]\tilde{Z}(X'\tilde{Z})^{-1} \\ &= \sigma^2(\tilde{Z}'X)^{-1}\tilde{Z}'\tilde{Z}(X'\tilde{Z})^{-1} \end{aligned}$$

Since $\tilde{Z}'X, \tilde{Z}'Z$, and $X'\tilde{Z}$ are square and full rank matrices, thus invertible. We rewrite the covariance matrix for $b_{\tilde{Z}}$ as:

$$\begin{aligned} Var[b_{\tilde{Z}}|X] &= \sigma^2(\tilde{Z}'X)^{-1}\tilde{Z}'\tilde{Z}(X'\tilde{Z})^{-1} \\ & \sigma^2\{X'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X\}^{-1} \end{aligned}$$

The 2SLS IV is $\hat{Z} = P_Z X = Z(Z'Z)^{-1}Z'X$ (the same as preceding $\hat{X}$; write as $\hat{Z}$ for uniformity in notations), and its coefficients estimator is:

$$b_{\hat{Z}} = (\hat{Z}'X)^{-1}\hat{Z}'y = \beta + (\hat{Z}'X)^{-1}\hat{Z}'\varepsilon$$

Same as before, we focus on the random part in $b_{\hat{Z}}$ to obtain its covariance matrix:

$$\begin{aligned}
Var[b_{\hat{Z}}|X] =& \sigma^2(\hat{Z}'X)^{-1}\hat{Z}'\hat{Z}(X'\hat{Z})^{-1} \\
=& \sigma^2\{X'Z(Z'Z)^{-1}Z'X\}^{-1}\{X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\}\{X'Z(Z'Z)^{-1}Z'X\}^{-1} \\
=& \sigma^2\{X'Z(Z'Z)^{-1}Z'X\}^{-1}
\end{aligned}$$

Denote $P_Z = Z(Z'Z)^{-1}Z', P_{\tilde{Z}} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'$. So the covariance matrices for $b_{\tilde{Z}}$ and $b_{\hat{Z}}$ can be simplified to

$$\begin{aligned}
Var[b_{\tilde{Z}}|X] = \sigma^2\{X'P_{\tilde{Z}}X\}^{-1} \\
Var[b_{\hat{Z}}|X] = \sigma^2\{X'P_Z X\}^{-1}
\end{aligned}$$

To compare the two covariance matrices, we only need to compare $X'P_{\tilde{Z}}X$ and $X'P_Z X$. To compare $X'P_{\tilde{Z}}X$ and $X'P_Z X$, we only need to compare $P_{\tilde{Z}}$ and P_Z. Hence, we first construct the "difference matrix":

$$D = P_Z - P_{\tilde{Z}}$$

Apparently, since both P_Z and $P_{\tilde{Z}}$ are projection matrices, they are symmetric and idempotent. So D is immediately symmetric. We are then interested to see if D is also idempotent.

$$\begin{aligned}
D^2 = DD &= (P_Z - P_{\tilde{Z}})(P_Z - P_{\tilde{Z}}) \\
&= P_Z - P_Z p_{\tilde{Z}} - P_{\tilde{Z}}P_Z + P_{\tilde{Z}}
\end{aligned}$$

Thus, our interest flows to $P_Z P_{\tilde{Z}}$ and $P_{\tilde{Z}}P_Z$:

$$\begin{aligned}
P_Z P_{\tilde{Z}} &= Z(Z'Z)^{-1}Z'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}' \\
&= Z(Z'Z)^{-1}Z'ZF(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}' \\
&= ZF(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}' \\
&= \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}' \\
&= P_{\tilde{Z}} \\
P_{\tilde{Z}}P_Z &= \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'Z(Z'Z)^{-1}Z' \\
&= \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}F'Z'Z(Z'Z)^{-1}Z' \\
&= \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}F'Z' \\
&= \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}' \\
&= P_{\tilde{Z}}
\end{aligned}$$

Plug in those two results to D^2, we can see:

$$\begin{aligned} D^2 = DD &= (P_Z - P_{\tilde{Z}})(P_Z - P_{\tilde{Z}}) \\ &= P_Z - P_Z p_{\tilde{Z}} - P_{\tilde{Z}} P_Z + P_{\tilde{Z}} \\ &= P_Z - P_{\tilde{Z}} \\ &= D \end{aligned}$$

Therefore, our conjecture of D being idempotent has been proved!

Since D is both symmetric and idempotent, then take any vector x,

$$x'Dx = X'DDx = X'D'Dx = (Dx)'(DX) = ||DX|| \geq 0$$

So D is non-negative definite. From this result, we can see that $X'DX$ is also non-negative definite:

$$y'X'DXy = (Xy)'D(Xy) \geq 0$$

Or alternatively, $y'X'DXy = y'X'D'DXy = (DXy)'(DXy) = ||DXy|| \geq 0$.

Since $X'DX$ is non-negative, that is, semi-positive definite, we have $Var[b_{\hat{Z}}|X] \leq Var[b_{\tilde{Z}}|X]$. In conclusion, the 2SLS estimator is (asymptotically) efficient in the class of all IV estimators using linear combination of Z. ■

4.3 Specification Tests

4.3.1 Hausman Test

Before introducing IV, we should conduct tests to inspect endogeneity or exogeneity. Under classical Guass-Markov assumptions, if X is uncorrelated with ε, then b_{LS} is the most efficient estimator. So if we do not have endogeneity, going with b_{LS} is apparently a better choice with more estimation precision. Consider a comparison of the two covariance matrices under the hypothesis that both estimators are consistent, that is, assuming plim $\frac{X'\varepsilon}{n} = 0$. The difference between the asymptotic covariance matrices of the two estimators is

$$\begin{aligned} \text{Asy.Var}\left[b_{IV}\right] - \text{Asy.Var}\left[b_{LS}\right] &= \frac{\sigma^2}{n} \cdot \text{plim} \left(\frac{X'Z(Z'Z)^{-1}Z'X}{n}\right) - \frac{\sigma^2}{n} \cdot \text{plim} \left(\frac{X'X}{n}\right)^{-1} \\ &= \frac{\sigma^2}{n} \cdot \text{plim} \left(n\left[\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} - (X'X)^{-1}\right]\right) \end{aligned}$$

To compare the two matrices in the brackets, we can compare their inverses. The inverse of the first is $X'Z(Z'Z)^{-1}Z'X = X'\left(1 - M_Z\right)X = X'X - X'M_ZX$. Because M_Z is a non-negative definite matrix, it follows that $X'M_ZX$ is also. So $X'Z(Z'Z)^{-1}Z'X$ equals $X'X$ minus a non-negative definite matrix. Because $X'Z(Z'Z)^{-1}Z'X$ is smaller, in the matrix sense, than $X'X$, so its inverse is larger. Under the hypothesis, the asymptotic covariance matrix of the LS estimator is never larger than that of the IV estimator, and it will actually be smaller unless all the columns of X are perfectly predicted by regressions on Z. Thus, we have established that if plim $\frac{X'\varepsilon}{n} = \mathbf{0}$—that is, if LS is consistent—then it is a preferred estimator.

Our interest in the difference between these two estimators goes beyond the question of efficiency. The null hypothesis of interest will usually be specifically whether plim $\frac{X'\varepsilon}{n} = \mathbf{0}$. Seeking the covariance between X and ε through $\frac{X'e}{n}$ is fruit- less, of course, because the normal equations produce $X'e = \mathbf{0}$. The logic of Hausman's test is as follows.

- Null hypothesis: X is uncorrelated with ε;
- Alternative hypothesis: X is correlated with ε.

Under the null hypothesis, both b_{LS} and b_{IV} are consistent estimators of β. While under the alternative hypothesis. Only b_{IV} is consistent. We construct $d = b_{IV} - b_{LS}$. Under the null hypothesis, plim $d = 0$, whereas under the alternative, plim $d \neq 0$. Hence, we have a Wald statistic,

$$H = d'\{Est.Asy.Var[d]\}^{-1}d$$

The asymptotic covariance matrix we need for the test is

$$\begin{aligned}\text{Asy.Var}\left[b_{IV} - b_{LS}\right] =& \text{Asy.Var}\left[b_{IV}\right] + \text{Asy.Var}\left[b_{LS}\right] \\ & - \text{Asy.Cov}\left[b_{IV}, b_{LS}\right] - \text{Asy.Cov}\left[b_{LS}, b_{IV}\right]\end{aligned}$$

Hausman gives a fundamental result that allows us to proceed.

Hausman Result The covariance between an efficient estimator b_E of a parameter vector β and its difference from an inefficient estimator b_I of the same parameter vector $b_E - b_I$ is zero, which means

$$\begin{aligned} & Cov(b_E, b_E - b_I) = 0 \\ \Longleftrightarrow\ & Cov(b_E, b_I) = Var(b_E)\end{aligned}$$

Hence, plug the result back to asymptotic variance matrix of $d = b_{IV} - b_{LS}$, we have

$$\begin{aligned}\text{Asy.Var}\left[b_{IV} - b_{LS}\right] =& \text{Asy.Var}\left[b_{IV}\right] + \text{Asy.Var}\left[b_{LS}\right] \\ & - \text{Asy.Cov}\left[b_{IV}, b_{LS}\right] - \text{Asy.Cov}\left[b_{LS}, b_{IV}\right] \\ =& \text{Asy.Var}\left[b_{IV}\right] + \text{Asy.Var}\left[b_{LS}\right] \\ & - \text{Asy.Var}\left[b_{LS}\right] - \text{Asy.Var}\left[b_{LS}\right] \\ =& \text{Asy.Var}\left[b_{IV}\right] - \text{Asy.Var}\left[b_{LS}\right]\end{aligned}$$

Inserting this useful and critical result into our Wald statistic and reverting to our empirical estimates of these quantities, we have

$$H = (b_{IV} - b_{LS})' \left\{ \text{Asy.Var}\left[b_{IV}\right] - \text{Asy.Var}\left[b_{LS}\right] \right\}^{-1} (b_{IV} - b_{LS}) = (b_{IV} - b_{LS})' \left\{ \sigma^2_{IV} (\hat{X}'\hat{X})^{-1} - \sigma^2_{LS} (X'X$$

Remarks:

1. Under large samples, $s^2_{IV} = s^2_{LS} = s^2$ and both are consistent estimators to σ^2_{IV} and σ^2_{LS}, respectively. Here we do not make adjustments for substituting s^2 for the unknown σ^2 in the statistic. Because rigorously speaking, if we do make such adjustments, the H is then reverted to a F statistic, and the corresponding F distribution is asymptotically a χ^2 distribution as if we had not done adjustments.

2. It is tempting to invoke our results for the full rank quadratic form in a normal vector and conclude the degrees of freedom for this chi-squared statistic is K. But that method will usually be incorrect, and worse yet, unless X and Z have no variables in common, the rank of the matrix in this statistic is less than K, and the ordinary inverse will not even exist. In most cases, at least some of the variables in X will also appear in Z. (In almost any application, X and Z will both contain the constant term.) That is, some of the variables in X are known to be uncorrelated with the disturbances. In this case, our hypothesis, plim $\frac{X'X}{n} = \mathbf{0}$, does not really involve all K variables, because a subset of the elements in this vector are known to be zero, say K_0-many. As such, the quadratic form in the Wald test is being used to test only $K^* = K - K_0$ hypothesis. In the meantime, it is easy and useful to show that, H is in fact a rank K^* quadratic form. Since $Z(Z'Z)^{-1}Z'$ is an idempotent matrix, $\left(\hat{X}'\hat{X}\right) = \hat{X}'X$. Using this result and expanding d, we find

$$\begin{aligned} d &= b_{IV} - b_{LS} \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y - (X'X)^{-1}X'y \\ &= (\hat{X}'\hat{X})^{-1}\left[\hat{X}'y - (\hat{X}'\hat{X})(X'X)^{-1}X'y\right] \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'\left[y - X(X'X)^{-1}X'y\right] \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'e \end{aligned}$$

where e is the vector of least squares residuals. Recall that for exogenous variables in X, they are IV for themselves, and satisfy $X'e = 0$. Denote the endogenous part of X as $\hat{X}^*$, then d can be expressed in the form:

$$d = (\hat{X}'\hat{X})^{-1} \begin{pmatrix} 0 \\ \hat{X^*}'e \end{pmatrix} = (\hat{X}'\hat{X})^{-1} \begin{pmatrix} 0 \\ q \end{pmatrix}$$

Finally, denote the entire matrix in H by W. (Because that the ordinary inverse may not exist, this matrix will have to be a generalized inverse.) Then, denoting the whole matrix product by P, we obtain

$$H = \begin{bmatrix} \mathbf{0}' & q' \end{bmatrix} (\hat{X}'\hat{X})^{-1} W (\hat{X}'\hat{X})^{-1} \begin{bmatrix} \mathbf{0} \\ q \end{bmatrix} = \begin{bmatrix} \mathbf{0}' & q' \end{bmatrix} P \begin{bmatrix} \mathbf{0} \\ q \end{bmatrix} = q' P_{**} q$$

where P_{**} is the lower right $K^* \times K^*$ submatrix of P. We now have the end result, that algebraically H is actually a quadratic form in a K^* vector, so K^* is the degrees of freedom for the test.

1. Hausman test can be applied to extensive situations. Suppose we have a pair of estimators $\widehat{\theta}_E$ and $\widehat{\theta}_I$, such that under the null hypothesis that $\widehat{\theta}_E$ and $\widehat{\theta}_I$ are both consistent, while under the alternative hypothesis only $\widehat{\theta}_I$ remains consistent and $\widehat{\theta}_E$ is inconsistent, then we can form a test of the hypothesis by referring the Hausman statistic:

$$H = \left(\hat{\theta}_I - \hat{\theta}_E\right)' \left\{ \text{Est.Asy.Var}\left[\hat{\theta}_I\right] - \text{Est.Asy.Var}\left[\hat{\theta}_E\right] \right\} \left(\hat{\theta}_I - \hat{\theta}_E\right) \xrightarrow{\text{d}} \chi^2 \left[J\right]$$

where the appropriate degrees of freedom for the test, J, will depend on the context.

4.3.2 Wu Test

The preceding Wald test requires a generalized inverse, so it is going to be a bit cumbersome. In fact, one need not actually approach the test in this form, and it can be carried out with any regression program. The alternative variable addition test approach devised by Wu (1973) is simpler. An F statistic with K^* and $n - K - K^*$ degrees of freedom can be used to test the joint significance of the elements of γ in the augmented regression:

$$y = X\beta + \hat{X}^* \gamma + \varepsilon^*$$

where $\hat{X}^*$ are the fitted values in the regressions of the variables in X^* on Z. This result is equivalent to the Hausman test for this model.

4.4 Measurement Error

The general assessment of measurement error problem is not particularly optimistic. The bias introduced by measurement error can be rather severe. There are almost no known finite-sample results for the models of measurement error; nearly all the results that have been developed are asymptotic.

Consider a regression model with a single regressor and no constant term:

$$y^* = \beta x^* + \varepsilon$$

where y^* and x^* are not available. Instead we can only observe y and x:

$$\begin{aligned} y &= y^* + v, \text{ with } v \sim N[0, \sigma_v^2] \\ x &= x^* + v, \text{ with } u \sim N[0, \sigma_u^2] \end{aligned}$$

First assume for the moment that only y^* is measured with error:

$$y = \beta x^* + \varepsilon + v = \beta x^* + \varepsilon'$$

This result conforms to the assumptions of the classical regression model. As long as the regressor is measured properly, measurement error on y^* can be absorbed in the disturbance of the regression and hence ignored.

Then suppose x has measurement error (y without), we have

$$\begin{aligned} y &= \beta x^* + \varepsilon \\ &= \beta(x - u) + \varepsilon \\ &= \beta x + (\varepsilon - \beta u) \end{aligned}$$

To check exogeneity, we inspect covariance of x and the disturbance term $(\varepsilon - \beta u)$ as a whole:

$$\text{Cov}\,[x, \varepsilon - \beta u] = \text{Cov}\,[x^* + u, \varepsilon - \beta u] = -\beta\sigma_u^2 \neq 0$$

The nonzero correlation violates assumption of exogeneity. As $b = \dfrac{\frac{1}{n}\sum_{i=1}^n x_i y_i}{\frac{1}{n}\sum_{i=1}^n x_i^2}$, we have

$$\text{plim } b = \frac{\text{plim } \frac{1}{n}\sum_{i=1}^n (x_i^* + u_i)(\beta x_i^* + \varepsilon_i)}{\text{plim } \frac{1}{n}\sum_{i=1}^n (x_i^* + u_i)^2}$$

Because x^*, ε and u are mutually independent, this equation reduces to

$$\text{plim } b = \frac{\beta Q^*}{Q^* + \sigma_u^2} = \frac{\beta}{1 + \sigma_u^2/Q*} < \beta$$

where $Q^* = \text{plim } \dfrac{\sum_{i=1}^n x_i^2}{n}$.

As long as σ_u^2 is positive, b is then inconsistent, with a persistent bias towards 0. Clearly, the greater the variability in the measurement error, the worse the bias. The effect of biasing the coefficient towards zero is called **attenuation**.

In a multiple regression model, matters only get worse. When only a single variable is measured with error (assume the first variable), we have

$$\text{plim } b_1 = \frac{\beta_1}{1+\sigma_u^2 q^{*11}}$$
$$\text{plim } b_k = \beta_k - \beta_1 \left[\frac{\sigma_u^2 q^{*k1}}{1+\sigma_u^2 q^{*11}}\right], \text{ for } k \neq 1$$

where q^{*k1} is the $(k,1)$th element in $(Q^*)^{-1}$.

The result in multiple regression case is not required. Notice that the coefficient on the badly measured variable is still biased toward zero. The other coefficients are all biased as well, although in *unknown directions*. A badly measured variable contaminates all the least squares estimates. If more than one variable is measured with error, there is very little that can be said.