Section 8: Endogeneity and Exam Practice

Jeffrey Greenbaum

March 9, 2007

Contents

1 Section Preamble

Today we will complete our discussions on panel data models and endogeneity by discussing fixed effects, random effects, 2SLS estimation, and GMM estimation for endogeneity. For the motivation of these topics and how they fit in with the theme of the course, see last week's section preamble. We will then discuss every exam question time permitting, beginning with endogeneity.

2 Overidentified case: 2SLS

When there are more instruments than regressors, $(Z'X)$ and $E(zx')$ are not invertible because they are not square matrices. We still require that the instrumental variables satisfies rank$(Z'X) = K$. What we do is first regress X on Z, the so called first stage, and then we regress Y on the predicted values of X given Z, the second stage. Intuitively, we are taking out of X all of its correlation with the error term, and then, we are regressing Y on this new term that is still kind of X, but free of its correlation with the error. This description is the intuition, but the idea mathematically is to transform a valid instrument, Z, so that we have ideal dimensionality while keeping all of the relevant information from X and Z.

More formally:

Run the regression:

$$x_i = z_i\gamma + v_i$$

and obtain $\hat{x}_i$, so that $\hat{X} = Z(Z'Z)^{-1}ZX = P_Z X$.

Then, run the regression:

$$y_i = \hat{x}_i\beta + \varepsilon_i$$

and obtain $\hat{\beta}$.

Observe that:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\
&= (X'P_Z P_Z X)^{-1}X'P_Z y \\
&= (X'P_Z X)^{-1}X'P_Z y \\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'y)
\end{aligned}$$

2.1 Asymptotics of 2SLS:

The asymptotic distribution of this estimator is derived very similarly to that of the IV estimator though the projection matrices must be accounted for in the limiting distribution. We are interested in the asymptotic properties to assure that the estimator is consistent and generally for the purposes of inference.
We start by computing $\hat{\beta} - \beta$:

$$\hat{\beta}_{2SLS} = (X'P_Z X)^{-1}X'P_Z(X\beta + \varepsilon) = \beta + (X'P_Z X)^{-1}X'P_Z\varepsilon$$

As a result,

$$\hat{\beta}_{2SLS} - \beta = (X'P_Z X)^{-1}X'P_Z\varepsilon = (\frac{1}{n}\sum_{i=1}^{n} x_i z_i'(z_i z_i')^{-1}z_i x_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'(z_i z_i')^{-1}z_i\varepsilon_i)$$

$\hat{\beta}_{2SLS}$ is consistent if $(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'(z_i z_i')^{-1} z_i x_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} x_i z_i'(z_i z_i')^{-1} z_i \varepsilon_i) \longrightarrow_p 0$.
Indeed this term converges in probability to zero as was the case with $\hat{\beta}_{IV}$ because $\frac{1}{n}\sum_{i=1}^{n} z_i\varepsilon_i \longrightarrow_p$ as already shown for $\hat{\beta}_{IV}$.
As a result of Slutsky's Theorem, $\hat{\beta}_{2SLS}$ is a consistent estimator for β.

Moreover, by the Central Limit Theorem, $\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} z_i\varepsilon_i \longrightarrow_d N(0, E(z^2\varepsilon^2))$.
Again, we repeat the work $\hat{\beta}_{IV}$ to show that:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \longrightarrow_d N\left(0, plim\left(\left(\frac{X'P_Z X}{n}\right)^{-1} \frac{X'P_Z\hat{V}P_Z X}{n}\left(\frac{X'P_Z X}{n}\right)^{-1}\right)\right)$$

We use the same $\hat{V}$ as in the just-identified case.

3 GMM

In the case in which $V_0 \neq \sigma^2 M_{zz}$ then $\hat{\beta}_{2SLS}$ will not have the smallest asymptotic covariance matrix. We define β_{GIV}, the generalized instrumental variable estimator, as $(\hat{\pi}'Z'X)^{-1}(\hat{\pi}'Z'y)$ where $\hat{\pi} \longrightarrow_p \pi$.
In the case of 2SLS, $\hat{\pi}$ is the OLS estimator from the regression of X on Z.

Accordingly, the asymptotic variance of β_{GIV} is $[\pi' M_{zx}]^{-1}\pi' V_0 \pi [M_{xz}\pi]^{-1}$ where $\frac{1}{T}Z'X \longrightarrow_p M_{zx}$.
We obtain an efficient estimator by choosing π that minimizes this asymptotic variance matrix.
This problem is analogous to that of generalized linear regression where we transformed the linear model to find the most efficient estimator.
As such, we multiply the linear model by $\frac{1}{\sqrt{T}}Z'$ so we are analyzing $\frac{1}{\sqrt{T}}Z'y = \frac{1}{\sqrt{T}}Z'X + \frac{1}{\sqrt{T}}Z'\varepsilon$.

We assume that $E[z_i\varepsilon_i] = 0$.
As a result, we can apply the Central Limit Theorem to show that $\frac{1}{\sqrt{T}}Z'\varepsilon \longrightarrow_d N(0, V_0)$.
Moreover, we can show that asymptotically this model satisfies the generalized classical linear assumptions, particularly that asymptotically, the covariance of the new design matrix with the new error matrix is zero.
Thus, we use the GLS framework to define $\hat{\beta}_{GMM}$:

$$\begin{aligned}\hat{\beta}_{GLS} &= ((\frac{1}{\sqrt{T}}Z'X)'V_0^{-1}\frac{1}{\sqrt{T}}Z'X)^{-1}(\frac{1}{\sqrt{T}}Z'X)'V_0^{-1}\frac{1}{\sqrt{T}}Z'y \\ &= [X'ZV_0^{-1}Z'X]^{-1}X'ZV_0^{-1}Z'y\end{aligned}$$

Accordingly, $\hat{\pi} = V_0^{-1}(\frac{1}{T}Z'X)$ where we need a consistent estimator of V_0 either through methods such as Eicker-White or Newey-West.
Moreover, we can use our GLS and 2SLS frameworks to derive $\sqrt{T}(\hat{\beta}_{GMM} - \beta) \longrightarrow_d N(0, [M_{zx}V_0^{-1}M_{xz}]^{-1})$.

4 Endogeneity Exercises

Endogeneity has appeared on every exam since 2002 in both True/False and free-response questions! In fact recently, it has appeared in both.

4.1 2002 Exam, 2

Question: Suppose the coefficients $\beta = (\beta_1, \beta_2)'$ in the lienar model $y = X\beta + \varepsilon$ are estimated by two-stage least squares, where it is assumed that the errors ε are independent of the matrix Z of instruments with scalar covariance matrix $Var(\varepsilon) = Var(\varepsilon|Z) = \sigma^2 I$. An analysis of $N = 163$ observations yields

$$\hat{\beta}_{2SLS} = \begin{pmatrix} 2 \\ 5 \end{pmatrix} \quad \hat{\sigma}^2_{2SLS} = 4, \quad \hat{X}'\hat{X} = (X'Z)(Z'Z)^{-1}(Z'X) = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}$$

Construct an approximate 95% confidence interval for $\gamma = \beta_1 * \beta_2$, under the (possibly heroic) assumption that the sample size is large enough for the usual limit theorems and linear approximations to be applicable. Is $\gamma_0 = 0$ in this interval?

Answer: The sample size is sufficiently large that we can say that the limiting distribution for $\hat{\beta}_{2SLS}$ provides the correct values for hypothesis testing. Recall that $\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$.

$$\begin{aligned} Var(\hat{\beta}_{2SLS}) &= (\hat{X}'\hat{X})^{-1}\hat{X}'Var(y)\hat{X}(\hat{X}'\hat{X})^{-1} \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'(\sigma^2 I)\hat{X}(\hat{X}'\hat{X})^{-1} \\ &= \sigma^2(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{X}(\hat{X}'\hat{X})^{-1} \\ &= \sigma^2(\hat{X}'\hat{X})^{-1} \end{aligned}$$

As a result, $\sqrt{N}(\hat{\beta}_{2SLS} - \beta) \longrightarrow_d N(0, \frac{1}{N}\sigma^2(\hat{X}'\hat{X})^{-1})$.
However, we are interested in the asymptotic distribution associated with $\gamma = \beta_1 * \beta_2 = g(\beta_1, \beta_2)$.
We thus use the Delta Method to show that $\sqrt{N}(\hat{\gamma} - \gamma) \longrightarrow_d N(0, \frac{1}{N}\sigma^2\hat{G}(\hat{X}'\hat{X})^{-1}\hat{G}')$
where $\hat{G} = \frac{\partial g(\beta_1,\beta_2)}{\partial(\beta_1,\beta_2)} = (\beta_2, \beta_1)$.

We can use the Law of Large Numbers and the Continuous Mapping Theorem to show that

$$\sqrt{\frac{1}{N}\hat{G}(\hat{X}'\hat{X})^{-1}\hat{G}'} = \sqrt{\frac{1}{N}(\beta_2, \beta_1)(\hat{X}'\hat{X})^{-1}(\beta_2, \beta_1)'} \longrightarrow_p \sqrt{\frac{1}{N}(\beta_2, \beta_1)(E(x_i x_i'))^{-1}(\beta_2, \beta_1)'}$$

Moreover, we know that $\hat{\sigma}^2 \longrightarrow_p \sigma^2$ by law of large numbers.
As a result, we can use Slutksy's Theorem twice to show that, $\frac{\sqrt{N}(\hat{\gamma}-\gamma)}{\sqrt{\frac{1}{N}\hat{G}\hat{\sigma}^2(\hat{X}'\hat{X})^{-1}\hat{G}'}} \longrightarrow_d N(0, 1)$.
Equivalently, $\frac{(\hat{\gamma}-\gamma)}{\sqrt{(\beta_2,\beta_1)\hat{\sigma}^2(\hat{X}'\hat{X})^{-1}(\beta_2,\beta_1)'}} \longrightarrow_d N(0, 1)$.

For these data, the estimate of γ is $\hat{\gamma} = \hat{\beta}_1 * \hat{\beta}_2 = 2 * 5 = 10$.
We now calculate the asymptotic standard error, $\sqrt{\hat{G}\hat{V}(\hat{\beta}_{2SLS})\hat{G}'}$:

$$
\begin{aligned}
\hat{V}(\hat{\beta}_{2SLS}) &= \hat{\sigma}^2(\hat{X}'\hat{X})^{-1} \\
& 4 * \left(\frac{1}{4}\right) \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix} \\
&= \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\hat{G}\hat{V}\hat{G}' &= \begin{pmatrix} 5 & 2 \end{pmatrix} * \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix} * \begin{pmatrix} 5 \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} 3 & 5 \end{pmatrix} \begin{pmatrix} 5 \\ 2 \end{pmatrix} \\
&= 25
\end{aligned}
$$

As a result, $SE(\hat{\gamma}) = \sqrt{25} = 5$.

Therefore, an approximate 95% confidence interval for γ is:
$CI = (\hat{\gamma} \pm 1.96 * SE\hat{\gamma}) = (10 \pm 1.96 * 5) = (10 \pm 9.8) = (0.2, 19.8)$.
0 is not in this confidence interval.

4.2 2003 Exam, 1A

Question: True/False/Explain. The Two-Stage Least Squares estimator $\hat{\beta}_{2SLS}$ is unchanged if the original NXL matrix of instrumental variables Z is replaced by a new matrix Z* of instruments if Z* = ZH, where H is an invertible L X L matrix.

Answer: True. The first stage projection matrix, $P_{Z^*} = Z^*((Z^*)'Z^*)^{-1}(Z^*)'$ for the transformed instruments Z^* is identical to the corresponding projection matrix $P_Z = Z(Z'Z)^{-1}Z'$ for the original instruments,

$$
\begin{aligned}
P_{Z^*} &= Z^*((Z^*)'Z^*)^{-1}(Z^*) \\
&= ZH((ZH)'(ZH))^{-1}(ZH)' \\
&= ZH(H'Z'ZH)^{-1}H'Z' \\
&= ZHH^{-1}(Z'Z)^{-1}(H')^{-1}H'Z' \\
&= Z(Z'Z)^{-1}Z' = P_Z
\end{aligned}
$$

Since the two-stage least squares estimator is defined as $\hat{\beta}_{2SLS} = (X'P_ZX)^{-1}X'P_Zy$, it does not change if P_{Z^*} replaces P_Z.
Note that Z^{-1} does not exist because Z is not a square matrix.

4.3 2004 Exam, 3

Question: Consider the estimation of two scalar coefficients, β_1 and β_2, in the linear equation

$$y = x_1\beta_1 + x_2\beta_2 + \varepsilon$$

where y, x_1, and x_2 are observable N-dimensional random vectors. In addition, two N-dimensional vectors of instrumental variables, z_1 and z_2, are available. In a sample size of $N = 227$, the following matrix of cross-products of the variables is observed:

$$\begin{bmatrix} y'y & y'x_1 & y'x_2 & y'z_1 & y'z_2 \\ x_1'y & x_1'x_1 & x_1'x_2 & x_1'z_1 & x_1'z_2 \\ x_2'y & x_2'x_1 & x_2'x_2 & x_2'z_1 & x_2'z_2 \\ z_1'y & z_1'x_1 & z_1'x_2 & z_1'z_1 & z_1'z_2 \\ z_2'y & z_2'x_1 & z_2'x_2 & z_2'z_1 & z_2'z_2 \end{bmatrix} = \begin{bmatrix} 22 & -11 & 10 & 8 & 8 \\ -11 & 21 & 10 & -8 & -8 \\ 10 & 10 & 20 & -2 & 0 \\ 8 & -8 & -2 & 6 & 4 \\ 8 & -8 & 0 & 4 & 6 \end{bmatrix}$$

A. For these data, calculate the classical LS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ of the unknown regression coefficients, and compute the instrumental variables estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ using z_1 and z_2 as instruments for x_1 and x_2.

B. Suppose the error terms ε are independent of z_1 and z_2, so that $Var[\varepsilon|z_1, z_2] = \sigma^2 I$, ie., ε has a scalar covariance matrix. If you had to conduct a test of $H_0 : \beta_2 = 1$ versus $H_A : \beta_2 \neq 1$ at an asymptotic 5% level using the IV estimator, and were given a consistent estimator $\tilde{\sigma}^2$ of the unknown variance parameter σ^2, how small would $\tilde{\sigma}^2$ have to be to reject H_0? That is, find the largest value of $\tilde{\sigma}^2$ for which you could (barely) reject the null hypothesis.

Answer: We first calculate the least squares and instrumental variable estimates using the usual formulae.

$$\begin{aligned}
\hat{\beta}_{OLS} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= (X'X)^{-1}X'y = \left(\begin{pmatrix} x_1 & x_2 \end{pmatrix}' \begin{pmatrix} x_1 & x_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} x_1 & x_2 \end{pmatrix}' \begin{pmatrix} y \end{pmatrix} \\
&= \left(\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} \begin{pmatrix} y \end{pmatrix} \\
&= \begin{pmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{pmatrix}^{-1} \begin{pmatrix} x_1'y \\ x_2'y \end{pmatrix} \\
&= \begin{bmatrix} 21 & 10 \\ 10 & 20 \end{bmatrix}^{-1} \begin{pmatrix} -11 \\ 10 \end{pmatrix} \\
&= \frac{1}{320} \begin{bmatrix} 20 & -10 \\ -10 & 21 \end{bmatrix} \begin{pmatrix} -11 \\ 10 \end{pmatrix} \\
&= \frac{1}{320} \begin{pmatrix} -320 \\ 320 \end{pmatrix} \\
&= \begin{pmatrix} -1 \\ 1 \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
\hat{\beta}_{IV} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= (Z'X)^{-1}Z'y = \left(\begin{pmatrix} z_1 & z_2 \end{pmatrix}' \begin{pmatrix} x_1 & x_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} z_1 & z_2 \end{pmatrix}' \begin{pmatrix} y \end{pmatrix} \\
&= \left(\begin{pmatrix} z_1' \\ z_2' \end{pmatrix} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} z_1' \\ z_2' \end{pmatrix} \begin{pmatrix} y \end{pmatrix} \\
&= \begin{pmatrix} z_1'x_1 & z_1'x_2 \\ z_2'x_1 & z_2'x_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1'y \\ z_2'y \end{pmatrix} \\
&= \begin{bmatrix} -8 & -2 \\ -8 & 0 \end{bmatrix}^{-1} \begin{pmatrix} 8 \\ 8 \end{pmatrix} \\
&= \frac{-1}{16} \begin{bmatrix} 0 & 2 \\ 8 & -8 \end{bmatrix} \begin{pmatrix} 8 \\ 8 \end{pmatrix} \\
&= \frac{1}{16} \begin{pmatrix} -16 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -1 \\ 0 \end{pmatrix}
\end{aligned}
$$

Recall that the asymptotic variance matrix of the IV estimator is given by

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

$Var(y) = \sigma^2 I$ and in this model, $Z'X = X'Z$ as evidenced by the data or can be confirmed by matrix multiplication.

$$
\begin{aligned}
AVar(\hat{\beta}_{IV}) &= (Z'X)^{-1}Z'\sigma^2 IZ(X'Z)^{-1} \\
&= \sigma^2(Z'X)^{-1}Z'IZ(Z'X)^{-1} \\
&= \sigma^2(Z'X)^{-1}Z'Z(Z'X)^{-1} \\
&= \sigma^2 \left(\frac{1}{16}\right) \begin{bmatrix} 0 & -2 \\ -8 & 8 \end{bmatrix} \begin{bmatrix} z_1'z_1 & z_1'z_2 \\ z_2'z_1 & z_2'z_2 \end{bmatrix} \left(\frac{1}{16}\right) \begin{bmatrix} 0 & -2 \\ -8 & 8 \end{bmatrix} \\
&= \sigma^2 \left(\frac{1}{16^2}\right) \begin{bmatrix} 0 & -2 \\ -8 & 8 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix} \begin{bmatrix} 0 & -2 \\ -8 & 8 \end{bmatrix} \\
&= \sigma^2 \left(\frac{1}{16^2}\right) \begin{bmatrix} 96 & 80 \\ -120 & 160 \end{bmatrix}
\end{aligned}
$$

Replacing the unknown value of σ^2 by an estimator $\tilde{\sigma}^2$ would give $V\hat{ar}(\hat{\beta}_2) = \tilde{\sigma}^2 \left(\frac{5}{8}\right)$. Thus, the asymptotic t-test for $H_0 : \beta_2 = 1$ would reject the null hypothesis if

$$
T = \frac{|\beta_2 - 1|}{\sqrt{Var(\beta_2)}} = \frac{|0-1|}{\sqrt{(\tilde{\sigma}^2)\frac{5}{8}}} > 1.96 \Rightarrow \tilde{\sigma}^2 < 1.6 * 1.96^{-2} \approx 0.4.
$$

4.4 2005 Exam, 3

Question: Suppose that, for the sample linear model with no intercept term,

$$y_i = \beta x_i + \varepsilon_i$$

that both $z_i = 1$ and $z_2 = x_i$ are valid instrumental variables for x_i, that is

$$E(z_{i1}\varepsilon_i) = E(\varepsilon_i) = 0$$
$$E(z_{i2}\varepsilon_i) = E(x_i\varepsilon_i) = 0$$

and

$$E(z_{i1}x_i) = E(x_i) = \mu \neq 0$$
$$E(z_{i2}x_i) = E(x_i^2) = \tau^2 \neq 0$$

A. Under the assumption that ε_i and x_i are jointly i.i.d. and ε_i is independent of x_i with $E(\varepsilon_i^2) = \sigma^2$, derive the asymptotic distribution of the IV estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ which use $z_{i1} = 1$ or $z_{i2} = x_i$, respectively, as an instrument for x_i, and compare the asymptotic variances of these two estimators.

B. Under the same assumptions as in part A, explicitly derive the asymptotic variance for the GMM estimator $\hat{\beta}_{GMM}$ which optimally uses both $z_{i1} = 1$ and $z_{i2} = x_i$ as instrumental variables, and show that this variance reduces to the asymptotic variance of one of the estimators in part A. [Hint: the relevant matrices M_{XZ} and V_O can be written in term of the parameters given above.]

Answer: Recall the asymptotic distribution for the IV estimator:

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \longrightarrow_d N(0, E(z_i x_i')^{-1} E(z_i^2 \varepsilon_i^2) E(x_i z_i')^{-1})$$

For $z_{i1} = 1$, $E(z_i^2\varepsilon_i^2) = E(\varepsilon_i^2) = \sigma^2$.
$E(z_i x_i') = E(x_i') = E(x_i) = \mu$
As a result, the asymptotic variance is $E(x_i)^{-1}\sigma^2 E(x_i)^{-1} = \sigma^2 E(x_i)^{-2} = \sigma^2\mu^{-2}$.
Accordingly, $\sqrt{n}(\hat{\beta}_1 - \beta) \longrightarrow_d N(0, \sigma^2\mu^{-2})$

For $z_{i2} = 1$, $E(z_i^2\varepsilon_i^2) = E(x_i^2\varepsilon_i^2) = E(x_i^2)E(\varepsilon_i^2) = \sigma^2 E(x_i^2)$.
$E(z_i x_i') = E(x_i * x_i') = E(x_i^2) = \tau^2$
As a result, the asymptotic variance is $E(x_i^2)^{-1}\sigma^2 E(x_i^2) E(x_i^2)^{-1} = \sigma^2 E(x_i^2)^{-1} = \sigma^2\tau^{-2}$.
Accordingly, $\sqrt{n}(\hat{\beta}_2 - \beta) \longrightarrow_d N(0, \sigma^2\tau^{-2})$.

We know that $0 < Var(x_i) = E(x_i^2) - E(x_i)^2 \Rightarrow E(x_i^2) > E(x_i)^2 \Rightarrow E(x_i^2)^{-1} < E(x_i)^{-2}$.
As a result, $AVar(\hat{\beta}_1) < AVar(\hat{\beta}_2)$ when using instrumental variables estimation.

Recall that the asymptotic variance for the $\hat{\beta}_{GMM}$ is $(\frac{X'Z}{n}(\frac{Z'\hat{V}Z}{n})^{-1}\frac{Z'X}{n})^{-1}$.
In this set-up, $z_i = (1, x_i)'$ and $\hat{V} = \sigma^2 I$.

$\frac{X'Z}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i z_i' = \frac{1}{n}\sum_{i=1}^{n} x_i(1, x_i) = (\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} x_i^2) \longrightarrow_p (\mu, \tau^2)$.

Accordingly, $\frac{XZ'}{n} \longrightarrow_p (\mu, \tau^2)'$.

Looking at the middle term,
$\frac{Z'\hat{V}Z}{n} = \frac{\sigma^2 Z'Z}{n} = \sigma^2(\frac{1}{n}\sum_{i=1}^{n} z_i z_i') \longrightarrow_p \sigma^2 E(z_i z_i')$
By Continuous Mapping Theorem, if $E(z_i z_i') \neq 0$ then $\frac{Z'\hat{V}Z}{n}^{-1} \longrightarrow_p \frac{1}{\sigma^2} E(z_i z_i')^{-1}$.

$$\frac{1}{\sigma^2}E(z_i z_i')^{-1} = \frac{1}{\sigma^2}E[(1, x_i)'(1, x_i)] = \sigma^2 \begin{pmatrix} 1 & \mu \\ \mu & \tau^2 \end{pmatrix}^{-1} = \frac{1}{\sigma^2(\tau^2 - \mu^2)} \begin{pmatrix} \tau^2 & -\mu \\ -\mu & 1 \end{pmatrix}$$

Therefore, Asymptotic Var($\hat{\beta}_{GMM}$) $= (\frac{X'Z}{n}(\frac{Z'\hat{V}Z}{n})^{-1}\frac{Z'X}{n})^{-1} =$

$$\left(\frac{1}{\sigma^2(\tau^2 - \mu^2)} \begin{pmatrix} \mu & \tau^2 \end{pmatrix} \begin{pmatrix} \tau^2 & -\mu \\ -\mu & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau^2 \end{pmatrix}\right)^{-1} = \sigma^2\tau^{-2} = \sigma^2 E(x_i^2)^{-1}.$$

Notice that the GMM estimator is equivalent to the second IV estimator that uses the instrument, x_i.

4.5 2006 Exam, 1A

Question: True/False/Explain. The Two-Stage Least Squares estimator $\hat{\beta}_{2SLS}$ is unchanged if the original NXL matrix of instrumental variables Z is replaced by a new matrix Z^* of instruments if $Z^* = HZ$, where H is an invertible NXN matrix.

Answer: False. $\hat{\beta}_{2SLS}$ is unchanged by using Z^* in place of Z if $P_Z = P_{Z^*}$ because the projection matrix is the only place in the estimator in which Z comes up.

$$\begin{aligned} P_{Z^*} &= Z^*((Z^*)'Z^*)^{-1}(Z^*)' \\ &= HZ((HZ)'(HZ))^{-1}(HZ)' \\ &= HZ(Z'H'HZ)^{-1}Z'H' \end{aligned}$$

Because we assume that $N > L$, neither Z nor ZH is invertible. As a result, this expression cannot be simplied further.

$P_Z = Z(Z'Z)^{-1}Z'$.

P_{Z^*} need not equal P_Z for any invertible NXN matrix H. We know the two are equal if H is the identity matrix, and for a specific Z, there are many possible matrices for H that can satisfy $P_Z = P_{Z^*}$. However, there are many possible matrices for H for which they would not be equal. As a result the answer is false.

However, if $Z^* = ZH$ and H is LXL so that the dimensions are valid for multiplicaiton then the statement would be true:

$$\begin{aligned}
P_{Z^*} &= Z^*((Z^*)'Z^*)^{-1}(Z^*) \\
&= ZH((ZH)'(ZH))^{-1}(ZH)' \\
&= ZH(H'Z'ZH)^{-1}H'Z' \\
&= ZHH^{-1}(Z'Z)^{-1}(H')^{-1}H'Z' \\
&= Z(Z'Z)^{-1}Z' = P_Z
\end{aligned}$$

4.6 2006 Exam, 3

Question: Consider the two equation system

$$\begin{aligned}
y &= x\beta + \varepsilon \\
x &= Z\pi + \eta
\end{aligned}$$

where y and x are observable T-dimensional vectors, ε and η are T-vectors of errors assumed jointly independent across rows, β is a scalar unknown parameter, Z is an observable TXL matrix of instrumental variables, and π is a L-dimensional vector of unknown coefficients. The error terms ε and η are jointly independent of the instruments Z, and assumed to have $E[\varepsilon] = E[\eta] = 0, E[\varepsilon\varepsilon'] = \sigma^2 I, E[\eta\eta'] = \tau^2 I$, and $E[\varepsilon\eta'] = \gamma I$, where γ may be nonzero (so x is endogenous in the equation for y).

Suppose you are given an estimator $\hat{\pi}$ from a separate sample (so it is statistically independent of ε and η) that satisfies

$$\sqrt{T}(\hat{\pi} - \pi) \longrightarrow_d N(0, V)$$

Defining $\hat{x} = Z\hat{\pi}$, consider the following two estimators of the scalar parameter β: the "instrumental variables" estimator

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

and the "two-stage plug-in" estimator

$$\hat{\beta}_{2S} = (\hat{x}'\hat{x})^{-1}(\hat{x}'y)$$

Assuming $plimT^{-1}Z'Z = M_{ZZ} = E[z_t z_t']$ has $\pi' M_Z Z\pi \neq 0$, derive the limiting distributions of $\sqrt{T}(\hat{\beta}_{IV} - \beta)$ and $\sqrt{T}(\hat{\beta}_{2S} - \beta)$, assuming the relevant Law of Large Numbers and Central Limit Theorems apply. Are these asymptotic distributions the same? If not, is one more efficient than the other in general?

Answer: The asymptotic distributions are not the same. We first analyze $\hat{\beta}_{IV}$ and use many of the same calculations to analyze $\hat{\beta}_{2S}$.

$$
\begin{aligned}
\hat{\beta}_{IV} &= (\hat{x}'x)^{-1}\hat{x}'y \\
&= (\hat{x}'x)^{-1}\hat{x}'(x\beta + \varepsilon) \\
&= (\hat{x}'x)^{-1}\hat{x}'x\beta + (\hat{x}'x)^{-1}\hat{x}'\varepsilon \\
&= \beta + (\hat{x}'x)^{-1}\hat{x}'\varepsilon \\
\Rightarrow \hat{\beta}_{IV} - \beta &= (\hat{x}'x)^{-1}\hat{x}'\varepsilon \\
&= ((Z\hat{\pi})'(Z\pi + \eta))^{-1}((Z\hat{\pi})'\varepsilon) \\
&= (\hat{\pi}'Z'(Z\pi + \eta))^{-1}(\hat{\pi}'Z'\varepsilon) \\
&= (\hat{\pi}'Z'Z\pi + \pi'Z'\eta)^{-1}(\hat{\pi}'Z'\varepsilon) \\
&= (T^{-1}\hat{\pi}'Z'Z\pi + T^{-1}\pi'Z'\eta)^{-1}(T^{-1}\hat{\pi}'Z'\varepsilon) \\
\Rightarrow \sqrt{T}(\hat{\beta}_{IV} - \beta) &= (T^{-1}\hat{\pi}'Z'Z\pi + T^{-1}\hat{\pi}'Z'\eta)^{-1}\sqrt{T}(T^{-1}\hat{\pi}'Z'\varepsilon)
\end{aligned}
$$

We proceed by analyzing each of the three terms separately. We will then apply Slutsky's Theorem to the sum and then again to the product to determine the overall distribution. However, we must also apply Slutsky's Theorem to each term along the way. We will work from left to right.

Because $\sqrt{T}(\hat{\pi} - \pi) \longrightarrow_d N(0, V), \hat{\pi} \longrightarrow_p \pi \Rightarrow \hat{\pi}' \longrightarrow_p \pi'$
It is given that $plimT^{-1}Z'Z = M_{ZZ} = E[z_t z_t']$.
As a result of Slutstky's Theorem, $T^{-1}\hat{\pi}'Z'Z\pi \longrightarrow_p \pi' M_{ZZ}\pi$

$T^{-1}(\frac{1}{T}\sum_{t=1}^{T} z_t\eta_t) \longrightarrow_p E(z_t\eta_t) = E(z_t)E(\eta_t) = 0$ by law of large numbers because $z_t\eta_t$ is iid and $Var(z_t\eta_t) = E(z_t\eta_t\eta_t' z_t) - E(z_t\eta_t)^2 = \tau^2 E(z_t z_t')$ is finite.
As a result of Slutsky's Theorem, $T^{-1}\hat{\pi}'Z'\eta \longrightarrow_p \pi' * 0 = 0$.

As a result of Slutsky's Theorem, $(T^{-1}\hat{\pi}'Z'Z\pi + T^{-1}\pi'Z'\eta) \longrightarrow_p \pi'M_{ZZ}\pi + 0 = \pi'M_{ZZ}\pi$.
As a result of the Continous Mapping Theorem, because $\pi'M_{ZZ}\pi \neq 0$, it is invertible, and so $(T^{-1}\hat{\pi}'Z'Z\pi + T^{-1}\pi'Z'\eta)^{-1} \longrightarrow_p (\pi'M_{ZZ}\pi)^{-1}$.

$E(z_t\varepsilon_t) = E(z_t)E(\varepsilon_t) = 0$ and $z_t\varepsilon_t$ is iid.
$Var(z_t\varepsilon_t) = E(z_t\varepsilon_t\varepsilon_t' z_t') - 0 = \sigma^2 E(z_t z_t') = \sigma^2 M_{zz}$.
As a result of the Central Limit Theorem, $\sqrt{T}(T^{-1}Z'\varepsilon) \longrightarrow_d N(0, Var(z_t\varepsilon_t)) = N(0, \sigma^2 M_{zz})$.

By Slutsky's Theorem, $\sqrt{T}(T^{-1}\hat{\pi}'Z'\varepsilon) \longrightarrow_d \pi' N(0, \sigma^2 M_{zz}) = N(0, \pi'\sigma^2 M_{zz}\pi) = N(0, \sigma^2\pi' M_{zz}\pi)$.

Finally, apply Slutsky's Theorem to the entire expression:
$\sqrt{T}(\hat{\beta}_{IV} - \beta) \longrightarrow_d (\pi'M_{zz}\pi)^{-1}N(0, \sigma^2\pi'M_{zz}\pi) = N(0, (\pi'M_{zz}\pi)^{-1}\sigma^2\pi'M_{zz}\pi(\pi'M_{zz}\pi)^{-1}) = N(0, \sigma^2(\pi'M_{zz}\pi)^{-1})$.

Now we consider the asymptotic distribution for $\hat{\beta}_{2S}$.

$$
\begin{aligned}
\hat{\beta}_{2S} &= (\hat{x}'\hat{x})^{-1}\hat{x}'y \\
&= (\hat{x}'\hat{x})^{-1}\hat{x}'(x\beta + \varepsilon) \\
&= (\hat{x}'\hat{x})^{-1}\hat{x}'((\hat{x} + Z(\pi - \hat{\pi}) + \eta)\beta + \varepsilon) \\
&= (\hat{x}'\hat{x})^{-1}\hat{x}'\hat{x}\beta + (\hat{x}'\hat{x})^{-1}\hat{x}'((Z(\pi - \hat{\pi}) + \eta)\beta + \varepsilon) \\
\Rightarrow \hat{\beta}_{2S} - \beta &= (\hat{x}'\hat{x})^{-1}\hat{x}'((Z(\pi - \hat{\pi}) + \eta)\beta + \varepsilon) \\
&= (T^{-1}\hat{x}'\hat{x})^{-1}\hat{x}'(T^{-1}(Z(\pi - \hat{\pi}) + \eta)\beta + T^{-1}\varepsilon) \\
\Rightarrow \sqrt{T}(\hat{\beta}_{2S} - \beta) &= (T^{-1}\hat{x}'\hat{x})^{-1}\hat{x}'\sqrt{T}(T^{-1}Z(\pi - \hat{\pi}) + T^{-1}\eta)\beta + T^{-1}\varepsilon) \\
&= (T^{-1}(z\hat{\pi})'(Z\hat{\pi}))^{-1}(Z\hat{\pi})'\sqrt{T}(T^{-1}Z(\pi - \hat{\pi}) + T^{-1}\eta)\beta + T^{-1}\varepsilon) \\
&= (T^{-1}\hat{\pi}'Z'Z\hat{\pi})^{-1}\hat{\pi}'Z\sqrt{T}(T^{-1}Z(\pi - \hat{\pi}) + T^{-1}\eta)\beta + T^{-1}\varepsilon) \\
&= (T^{-1}\hat{\pi}'Z'Z\hat{\pi})^{-1}\hat{\pi}'\sqrt{T}(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon)
\end{aligned}
$$

Using the previous calculations, we can use Slutsky's Theorem twice to show that $(T^{-1}\hat{\pi}'Z'Z\hat{\pi}) \longrightarrow_p \pi' M_{zz}\pi$.

As a result of the Continuous Mapping Theorem, $(T^{-1}\hat{\pi}'Z'Z\hat{\pi})^{-1} \longrightarrow_p (\pi' M_{ZZ}\pi)^{-1}$.

We will analyze $\sqrt{T}(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon)$ by using the Central Limit Theorem, and exploiting that $\hat{\pi}$ comes from a sample that is independent of the observable data and unobservable terms.

First, $E(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon) = 0$ because the expectation of each individual term equals zero. We have already shown that $E(T^{-1}\hat{\pi}'Z'\varepsilon) = 0$ and because $E[\eta] = 0$, it also follows that $E(T^{-1}\hat{\pi}'Z'\eta)\beta) = 0$. Moreover, $E(\hat{\pi}) = \pi, soE(T^{-1}\hat{\pi}'Z'Z(\pi - \hat{\pi})\beta) = 0$.
We also know that we could rewrite the expression so that it is iid because all of the terms come from iid variables.

We now compute the corresponding variance in the limiting distribution:

$$
\begin{aligned}
&Var(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon) = \\
&Var(T^{-1}Z'Z(\pi - \hat{\pi})\beta) + Var(T^{-1}Z'\eta)\beta) + Var(T^{-1}Z'\varepsilon) + \\
&2Cov(T^{-1}Z'Z(\pi - \hat{\pi})\beta), (T^{-1}Z'\eta)\beta)) + 2Cov(T^{-1}Z'Z(\pi - \hat{\pi})\beta), (T^{-1}Z'\varepsilon)) + \\
&2Cov((T^{-1}Z'\eta)\beta), (T^{-1}Z'\varepsilon))
\end{aligned}
$$

Because π is fixed and $\hat{\pi}$ is statistically independent of η and ε,
$2Cov(T^{-1}Z'Z(\pi - \hat{\pi})\beta), (T^{-1}Z'\varepsilon)) = 0$ and
$2Cov(T^{-1}Z'Z(\pi - \hat{\pi})\beta), (T^{-1}Z'\eta\beta)) = 0$.
We know from before that $Var(T^{-1}Z'\varepsilon) = \sigma^2 M_{ZZ}$.
Similarly, $Var(T^{-1}Z'\eta\beta) = \beta^2 E(T^{-1}Z'Z)E(\eta\eta') = \beta^2\tau^2 M_{ZZ}$ because of statistical independence.

$2Cov((T^{-1}Z'\eta)\beta),(T^{-1}Z'\varepsilon)) = 2\beta\gamma M_{ZZ}$.
$Var(T^{-1}Z'Z(\pi - \hat{\pi})\beta) = \beta^2 M_{ZZ} V M_{ZZ}$.
Therefore, the asymptotic variance is is $(\sigma^2 + \tau^2\beta^2 M_{ZZ} + \beta^2 V M_{ZZ} + 2\beta\gamma)M_{ZZ}$

By the Central Limit Theorem, $\sqrt{T}(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon) \longrightarrow_d N(0, (\sigma^2 + \tau^2\beta^2 M_{ZZ} + \beta^2 V M_{ZZ} + 2\beta\gamma)M_{ZZ})$.

By Slutsky's Theorem, $\hat{\pi}'\sqrt{T}(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon) \longrightarrow_d N(0, (\sigma^2 + \tau^2\beta^2 M_{ZZ} + \beta^2 V M_{ZZ} + 2\beta\gamma)\pi' M_{ZZ}\pi)$.

Applying Slutsky's Theorem once more, $(T^{-1}\hat{\pi}'Z'Z\hat{\pi})^{-1}\hat{\pi}'\sqrt{T}(T^{-1}Z'Z(\pi - \hat{\pi}) + T^{-1}Z'\eta)\beta + T^{-1}Z'\varepsilon) \longrightarrow_d N(0, (\pi' M_{ZZ}\pi)^{-1}(\sigma^2 + \tau^2\beta^2 M_{ZZ} + \beta^2 V M_{ZZ} + 2\beta\gamma)\pi' M_{ZZ}\pi)$

Therefore, $\sqrt{T}(\hat{\beta}_{2S} - \beta) \longrightarrow_d N(0, \sigma^2 + \tau^2\beta^2 M_{ZZ} + \beta^2 V M_{ZZ} + 2\beta\gamma)$
All of these terms in the variance are positive because they correspond with variances, which must be positive.

Therefore, the asymptotic variance associated with $\sqrt{T}(\hat{\beta}_{2S} - \beta)$ is larger than $\sqrt{T}(\hat{\beta}_{IV} - \beta)$.

5 Additional Exercises

This section includes solutions to previous 240B exam questions that have not already been distributed. Excluded questions are those about maximum likelihood estimation from the 2002 and 2003 exam as well as recycled questions as seen in the 2005 exam and question 1D in 2006.

5.1 2003 Exam, 1D

Question: By the Continuous Mapping theorem, if $\hat{\theta}$ is root-n consistent and asymptotically normal for the scalar parameter θ_0, then its squared value, when multiplied by an appropriate function of n, should have a limiting chi-square distribution.

Answer: False. As we will show though the statement is true only for $\theta_0 = 0$.
If $\theta_0 \neq 0$ then the delta method implies that the asymptotic distribution of θ^2 is nomral, not chi-squared. Letting $g(\theta) = \theta^2$ with derivative $g'(\theta) = 2\theta$, and assuming

$$\sqrt{n}(\hat{\theta} - \theta_0) \longrightarrow_d N(0, V_0)$$

the delta method implies that

$$\sqrt{n}(\hat{\theta}^2 - \theta_0^2) \longrightarrow_d N(0, [g'(\theta)]^2 V_0) = N(0, 4\theta_0^2 V_0)$$

Thus, the squared value $\hat{\theta}^2$ is equal to θ_0^2 plus an asymptotically normal variable, and cannot be rescaled to be asymptotically chi-squared.

However, if $\theta_0 = 0$, then this statement is true if we scale $\hat{\theta}^2$ by n.

Recall that if $Z_N = \sqrt{N}(\theta_N - \theta_0) \longrightarrow_d N(0, I_p)$ then $T = Z_N' Z_N = N(\hat{\theta} - \theta_0)'(\hat{\theta} - \theta_0) \longrightarrow_d \chi_p^2$ where $p = dim(\hat{\theta_N})$

*Note that this statement comes directly from Professor Powell's lecture notes though he has a very important typo. Instead he has Σ listed as the variance-covariance matrix rather than I_p.

Therefore, if $\theta_0 = 0$, $\left(\frac{n}{V_0}\right)\hat{\theta}^2 \longrightarrow_d \chi_1^2$ because $\left(\frac{\sqrt{n}}{\sqrt{V_O}}\right)\hat{\theta} \longrightarrow_d N(0,1)$.

5.2 2004 Exam, 2

Question: A feasible GLS fit of the generalized regression model with $K = 3$ regressors yields the estimates $\hat{\beta} = (2, -1, 2)$ where the GLS covariance matrix $V = \sigma^2[X'\Omega^{-1}X]^{-1}$ is estimated as

$$\hat{V} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

using consistent estimators of σ^2 and Ω. The sample size $N = 403$ is large enough so that it is reasonable to assume a normal approximation holds for the GLS estimator.

Use these results to test the null hypothesis $H_O : \theta = 1$ against a two-sided alternative asymptotic 5% level, where

$$\theta = g(\beta) = ||\beta|| = (\beta_1^2 + \beta_2^2 + \beta_3^2)^{\frac{1}{2}}$$

Answer: We reject the null hypothesis by constructing an asymptotic approximation by using the delta method

Because $\hat{\beta} \overset{A}{\sim} N(\beta, \hat{V})$, $\theta = g(\hat{\beta}) \overset{A}{\sim} N(\theta, \hat{G}\hat{V}\hat{G}')$ where

$$\hat{G} = \frac{\partial g(\beta)}{\partial \beta'} = \frac{1}{(\beta_1^2 + \beta_2^2 + \beta_3^2)^{\frac{1}{2}}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$$

For these data, $\hat{\theta} = (2^2 + (-1)^2 + (-2)^2)^{\frac{1}{2}} = 9^{\frac{1}{2}} = 3$ and $\hat{G} = \frac{1}{3}(2, -1, -2)$.

$$\begin{aligned}
\hat{G}\hat{V}\hat{G}' &= \frac{1}{3}\begin{pmatrix} 2, & -1, & -2 \end{pmatrix} * \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} * \frac{1}{3}\begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} \\
&= \frac{1}{9}\begin{pmatrix} 3, & 1, & -2 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix} \\
&= 1
\end{aligned}$$

Accordingly, to test $H_0 : \theta = 1$, the absolute value of the t-statistic is

$$\frac{|\hat{\theta} - \theta_0|}{\sqrt{\hat{G}\hat{V}\hat{G}'}} = \frac{|3-1|}{1} = 2$$

which exceeds 1.96, the upper 97.5% critical value of a standard normal. We thus (barely) reject H_0 at an asymptotic 5% level. As is often the case, the sample size $N = 403$ does not directly figure into the solution, though it is implicit in the estimate $\hat{V}$ of the approximate covariance matrix of $\hat{\beta}$.

5.3 2006 Exam, 1B

Question: Suppose that, for the population of firms in the U.S., the relationship over time between dividends and some observable regressors (which include firm size) follows the assumptions of the Classical Normal Linear Regression model, conditional on the realized values of the regressors. Rather than a random sample of firms over time, though, suppose only that a sample of T average values of dividends and the regressors are available for the Fortune 500 largest firms. Using this sample, the Classical Least Squares estimators of the regression coefficients will be efficient, and F-tests using the usual normal-theory for the linear regression model will have correct size.

Answer: True. Although this scenario appears to be a classic case of a grouped-data regression model in which WLS is more appropriate than OLS, the weighting matrix would be a diagonal matrix based on $\sqrt{500}$ for each observation. Accordingly, transforming the linear regression model by a diagonal matrix in which all elements along the diagonal are equal does not change the ordinary least squares estimation. Such is the case here because the inverse of the diagonal matrix remains a diagonal matrix with all identical elements along the diagonal. In other words, this matrix can be factored into the product of a scalar and the identity. We know that GLS, or WLS in this case, reduces to OLS if the variance matrix is the identity.

5.4 2006, Exam 2

Question: Suppose $\hat{\theta}$ is an asymptotically normal estimator of a 3-dimensional parameter $\theta = (\theta_1, \theta_2, \theta_3)'$, which has the asymptotic distribution

$$\sqrt{N}(\hat{\theta} - \theta) \longrightarrow_d N(0, V)$$

Suppose that $\hat{\theta} = (1, -1, -2)'$ is the realized value of this estimator, and that a consistent estimator $\hat{V}$ of V has the realized value

$$\begin{pmatrix} 50 & 0 & 0 \\ 0 & 50 & 0 \\ 0 & 0 & 100 \end{pmatrix}$$

where it is assumed that the sample $N = 597$ is large enough so that the normal approximation is accurate for this problem.
Use these results to test the joint null hypothesis $H_O : \theta_1^2 + \theta_2^2 = 1$ and $\theta_3^2 = 1$, against the alternative that one or both of these restrictions fail, at an asymptotic 5% level.

Answer: We seek to conduct an asymptotic F-test, which converges to a chi-squared distribution. We use the Delta Method to derive a test statistic that tests this null hypothesis and converges to a chi-squared distribution.
We effectively want to derive a limiting distribution for

$$\sqrt{N}\left(\begin{pmatrix} \hat{\theta_1}^2 + \hat{\theta_2}^2 \\ \hat{\theta_3}^2 \end{pmatrix} - \begin{pmatrix} (\theta_1)^2 + (\theta_2)^2 \\ (\theta_3)^2 \end{pmatrix}\right)$$

$$\hat{G} = \frac{\partial g}{\partial \theta_1, \theta_2, \theta_3)} = \begin{pmatrix} 2\hat{\theta}_1, & 2\hat{\theta}_2, & 0 \\ 0, & 0, & 2\hat{\theta}_3 \end{pmatrix} = \begin{pmatrix} 2, & -2, & 0 \\ 0, & 0, & -4 \end{pmatrix}$$

Use the usual probability theorems and delta method to derive a distribution that asymptotically resembles the F-test in a way that is very similar to the asymptotic t-test.
That is, subtract the matrix of the hypothesized values from the estimated matrix. Multiply the transpose of this matrix by the inverse of the asymptotic variance matrix by this subtracted matrix. This expression converges to a distribution that is χ^2_2.
Reject the null hypothesis if the result of this matrix multiplication is larger than the corresponding value of χ^2_2 at the asymptotic 5% level.