ECOM40006/90013 Econometrics 3

Week 10 Extras: Solutions

Question 1: The Method of Moments

(a) (i.) Population

(ii.) Sample

(iii.) Population

(iv.) Population

(v.) Population

(vi.) Sample

(vii.) Population

For (ii), the population counterpart is (iii). For (vi), the relevant counterpart is (v).

(b) Direct computation yields

$$\begin{aligned}
\mathbb{E}(X'u) &= \mathbb{E}(X'(y - X\beta)) \\
&= \mathbb{E}(X'y) - \mathbb{E}(X'X)\beta = 0 \\
\implies \mathbb{E}(X'X)\beta &= \mathbb{E}(X'y) \\
\implies \beta &= \mathbb{E}(X'X)^{-1}\mathbb{E}(X'y).
\end{aligned}$$

Notice the resemblance that this estimator bears to the classic OLS estimator.

(c) If we replace the population moments with their sample counterparts,

$$\begin{aligned}
\mathbb{E}(X'X)^{-1}\mathbb{E}(X'y) &\overset{\text{match}}{=} \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y \\
&= (X'X)^{-1}N\frac{1}{N}X'y \\
&= (X'X)^{-1}X'y.
\end{aligned}$$

By doing this, we obtain the usual OLS estimator! Hence, the Method of Moments (MM) estimator is, in fact, the OLS estimator. This kind of technique is known as *moment matching*. Appeals are made to the WLLN, Slutsky's Theorem and the Continuous Mapping Theorem in doing this.

Question 2: 2SLS and the Generalized Method of Moments

(a) (i.) To reiterate the definitions: p is the number of instrumental variables and k_2 represents the number of endogenous variables. So, $p \geq k_2$ means that we have at least as many IVs as endogenous variables.

In the case where $p = k_2$, the system is said to be *just-identified.* If $p > k_2$ the system is *over-identified.*

(ii.) Observe that X_1 is $n \times k_1$ and Z is $n \times p$. Since both matrices have the same number of rows, the matrix W has n rows. Now, it remains to determine the number of columns. In this case, add up the columns to get $k_1 + p$. So W is a $n \times (k_1 + p)$ matrix.

In the instrumental variables literature, X_1 represents the exogenous variables in the regression while Z represents the actual IVs that we use in a two-stage least squares (2SLS) process. Sometimes, we refer to the exogenous variables as *included* IVs and the variables in Z as *excluded* IVs. In all the discussion of conditions such as relevance and exogeneity, only the excluded IVs matter.

(b) (i.) The error terms have zero mean when conditioned on the IVs. Specifically: the IVs don't suffer from endgeoeneity. This is one way to implement the *exogeneity* condition in instrumental variables literature. A more convenient way to represent this equation (albeit actually stronger) would be to represent it as $\mathbb{E}(u|W) = 0$.

(ii.) This condition represents *homoskedasticity*: the IVs do not affect the variance of the error term. As before: a more convenient way to represent this – albeit stronger – would be to write it as $\mathbb{E}(uu'|W) = \Omega$.

(c) Using the Law of Iterated Expectations we can write the mean of $W'u$ as

$$\mathbb{E}(W'u) = \mathbb{E}[\mathbb{E}(W'u|W)] = \mathbb{E}(W'\mathbb{E}(u|W)) = \mathbb{E}(W' \times 0) = 0.$$

For the variance, using the fact that $W'u$ has zero mean,

$$\begin{aligned}
\text{Var}(W'u) &= \mathbb{E}[(W'u)(W'u)'] \\
&= \mathbb{E}[W'uu'W] \\
&= \mathbb{E}[\mathbb{E}(W'uu'W|W)] && \text{(LIE)} \\
&= \mathbb{E}[W'\mathbb{E}(uu'|W)W] \\
&= \mathbb{E}[W'\Omega W].
\end{aligned}$$

(d) (i.) The first step is to plug in the linear model $y = X\beta + u$ into the IV estimator:

$$\begin{aligned}
\hat{\beta}_{IV} &= (W'X)^{-1}W'y \\
&= (W'X)^{-1}W'(X\beta + u) \\
&= \beta + (W'X)^{-1}W'u
\end{aligned}$$

Multiplying the term on the right by n/n and shuffling the n terms appropriately we get

$$\hat{\beta}_{IV} = \beta + \left(\frac{1}{n}W'X\right)^{-1}\frac{1}{n}W'u \tag{1}$$

which we'll need later. For now, observe that

- $\left(\frac{1}{n}W'X\right)^{-1} \xrightarrow{p} Q_{W'X}^{-1}$ by the Continuous Mapping Theorem and the assumption made in the question.
- $\frac{1}{n}W'u \xrightarrow{p} 0$.

So using Slutsky's Theorem, taking $n \to \infty$ implies that in equation (1)

$$\hat{\beta}_{IV} \xrightarrow{p} \beta + Q_{W'X}^{-1} \times 0 = \beta,$$

hence implying that $\hat{\beta}_{IV}$ is consistent for β.

(ii.) For the asymptotic distribution, return to equation (1) and bring β over to the left:

$$\hat{\beta}_{IV} - \beta = \left(\frac{1}{n}W'X\right)^{-1}\frac{1}{n}W'u.$$

Now, multiply both sides by $\sqrt{n}$:

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) = \left(\frac{1}{n}W'X\right)^{-1}\frac{1}{\sqrt{n}}W'u.$$

- As before, we have $\left(\frac{1}{n}W'X\right)^{-1} \xrightarrow{p} Q_{W'X}^{-1}$.
- Now we have $\frac{1}{\sqrt{n}}W'u \xrightarrow{d} N(0, Q_{W'\Omega W})$ by the assumption given.

Using Slutsky's Theorem this implies that as we take $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} Q_{W'X}^{-1} \times N(0, Q_{W'\Omega W}).$$

Moving the $Q_{W'X}^{-1}$ into the normal distribution term, we arrive at the asymptotic distribution

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} \times N(0, Q_{W'X}^{-1}Q_{W'\Omega W}Q_{X'W}^{-1}),$$

where $Q_{X'W} = Q'_{W'X}$.[1]

(e) Observe that X is a $n \times k$ matrix and W is a $n \times (k_1 + p)$ matrix with $p > k_2$. Now, observe that X has $k = k_1 + k_2$ columns and W has $k_1 + p$ columns. This implies

$$p > k_2 \implies k_1 + p > k_1 + k_2,$$

[1] If you're having trouble seeing where this line came from, first use the fact that $\text{Var}(AX) = A\text{Var}(X)A'$ and then note that $(A^{-1})' = (A')^{-1}$ i.e. you can swap transposes and inverses.

i.e. W has more columns than X. If we check the dimensions of $W'X$ then we find

$$\underbrace{W'}_{(k_1+p)\times n}\underbrace{X}_{n\times k} = (k_1+p)\times k \text{ matrix.}$$

But since $(k_1 + p) \neq k$ (i.e. it's not square), $W'X$ is not invertible and hence the IV estimator doesn't exist.

(f) A qualitative description of the 2SLS process is:

- *Stage 1.* Regress the dependent variables X on the instrumental variables W and extract the fitted values $\hat{X}$.
- *Stage 2.* Conduct an OLS regression on the original model $y = X\beta + u$, but with X replaced by $\hat{X}$.

What's one way to think about this process intuitively? First, think about the problem that we have to deal with, which is the endogeneity problem $\mathbb{E}(X'u) \neq 0$. When this happens, we can think of the data X as being split into two parts: one that is *exogenous* and one that is *endogenous*.

The first stage helps separate the two out. From a regression of X we have

$$X = \hat{X} + \text{residuals}.$$

Why is this important? If exogeneity is satisfied, then one way to think of $\hat{X}$ is as some part of X that is *not correlated with the econometric disturbance terms*. The rest are left-overs that we can't say anything about, so we discard them and use just $\hat{X}$. The amount of X we keep in the fitted values will depend on how relevant the IVs are.

In the second stage, we then run our regression using the fitted values of X. As long as we have exogeneity, the main idea is that we will eventually be able to get back to the true value β provided we have enough observations.

(g) In equations, the 2SLS process can be written as the following. For a linear model $y = X\beta + u$,

- *Stage 1.* Regress X on W using the equation

  $$X = W\pi + \eta,$$

  where η is a $n \times 1$ vector of econometric disturbances that follow all the regular assumptions. OLS on this model yields the estimator

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

  with fitted values

  $$\hat{X} = W\hat{\pi} = W(W'W)^{-1}W'X = P_W X.$$

- *Stage 2.* Substitute $\hat{X}$ in place of X in the original regression:

$$y = \hat{X}\beta + u$$

and run OLS on this model:

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y.$$

Expanding this out gives a different version of $\hat{\beta}_{2SLS}$ that we can use in more general derivations and computational procedures:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\
&= ([P_W X]'P_W X)^{-1}(P_W X)'y \\
&= (X'P_W P_W X)^{-1}X'P_W y && (P_W \text{ is symmetric}) \\
&= (X'P_W X)^{-1}X'P_W y, && (P_W \text{ is idempotent})
\end{aligned}$$

which can be written out fully in X, W and y. This is handy for derivations, but is generally slightly less useful when one wishes to commit the formula to memory.

(h) In a similar fashion to the OLS estimator, we can substitute in the linear model $y = X\beta + u$ and do some algebra:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (X'P_W X)^{-1}X'P_W(X\beta + u) \\
&= \beta + (X'P_W X)^{-1}X'P_W u \times \frac{n^3}{n^3} \\
&= \beta + \left[\frac{1}{n}X'W\left(\frac{1}{n}W'W\right)^{-1}\frac{1}{n}W'X\right]^{-1}\frac{1}{n}X'W\left(\frac{1}{n}W'W\right)^{-1}\frac{1}{n}W'u \\
&\xrightarrow{p} \beta + \left[Q_{X'W}Q_{W'W}^{-1}Q_{W'X}\right]^{-1}Q_{X'W}Q_{W'W}^{-1} \times 0 \\
&= \beta.
\end{aligned}$$

Therefore, the 2SLS estimator is consistent for β.

(i) The first step is to take the expression from part (h) above and:

- Subtract β from both sides
- Then, multiply by $\sqrt{n}$ on both sides.

Since we can choose where the $\sqrt{n}$ goes on the RHS, we will send it to $\frac{1}{n}W'u$. Doing this gives us

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = \left[\frac{1}{n}X'W\left(\frac{1}{n}W'W\right)^{-1}\frac{1}{n}W'X\right]^{-1}\frac{1}{n}X'W\left(\frac{1}{n}W'W\right)^{-1}\frac{1}{\sqrt{n}}W'u.$$

The other parts on the RHS continue to converge as normal, but we also know that in this case,

$$\frac{1}{\sqrt{n}}W'u \xrightarrow{d} N(0, \sigma^2 Q_{W'\Omega W}),$$

so that when we take $n \to \infty$ we find

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} \left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1} Q_{X'W} Q_{W'W}^{-1} N(0, Q_{W'\Omega W}) \\
&= N(0, \left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1} Q_{X'W} Q_{W'W}^{-1} Q_{W'\Omega W} Q_{W'W}^{-1} Q_{W'X} \left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1})
\end{aligned}$$

This is the asymptotic distribution of the 2SLS estimator. As you can tell, it takes quite a bit of algebra to reach this point. The steps themselves are not that bad; it's the matrix baggage that we have to carry around that becomes the issue in these kinds of derivation.

Aside: note that if you assume that the errors are distributed $u \sim N(0, \sigma^2 I_n)$, the distribution above simplifies dramatically since $Q_{W'\Omega W}$ is an expression for the expectation

$$\mathbb{E}(W'\Omega W).$$

But if $\Omega = \sigma^2 I_n$ then this becomes

$$\mathbb{E}(W'\sigma^2 I_n W) = \sigma^2 \mathbb{E}(W'W) = \sigma^2 Q_{W'W},$$

and in the asymptotic distribution above this will cancel with one of the $Q_{W'W}^{-1}$ terms, which creates a further cascade: namely, we get

$$\left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1} Q_{X'W} Q_{W'W}^{-1} Q_{W'X} \left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1} = \left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1}.$$

So in that special case the asymptotic distribution of the 2SLS estimator would just be

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \left[Q_{X'W} Q_{W'W}^{-1} Q_{W'X}\right]^{-1}).$$

(j) The implied quadratic form[2] is

$$\begin{aligned}
\hat{m}'\hat{m} &= (y - X\beta)'WW'(y - X\beta) \\
&= (y' - \beta'X')WW'y - (y' - \beta'X')WW'X\beta \\
&= y'WW'y - \beta'X'WW'y - y'WW'X\beta + \beta'X'WW'X\beta \\
&= y'WW'y - 2y'WW'X\beta + \beta'X'WW'X\beta
\end{aligned}$$

since the expressions in the middle are actually scalars (you may verify the dimensions for yourself). Now, the FOC with respect to β gives

$$\begin{aligned}
\frac{\partial \hat{m}'\hat{m}}{\partial \beta} = -2X'WW'y + 2X'WW'X\beta = 0 \\
\implies 2X'WW'X\beta = 2X'WW'y \\
\implies \hat{\beta} = (X'WW'X)^{-1}X'WW'y.
\end{aligned}$$

This is known as a *Generalized Method of Moments* (GMM) estimator. We usually use a slightly different one, which we will explore in the next part.

[2] The $1/n$ terms are ignored: in a maximization problem, these merely scale the value of the quadratic form by a positive constant, which doesn't affect the actual minimum of the objective function.

(k) If we repeat the same steps as what we did above, we end up with

$$\hat{m}'\Lambda\hat{m} = y'W\Lambda W'y - 2y'W\Lambda W'X\beta + \beta'X'W\Lambda W'X\beta$$

from which the FOC gives

$$\frac{\partial \hat{m}'\Lambda\hat{m}}{\partial\beta} = -2X'W\Lambda W'y + 2X'W\Lambda W'X\beta = 0$$
$$\implies \hat{\beta} = (X'W\Lambda W'X)^{-1}X'W\Lambda W'y.$$

If we use $\Lambda = (W'W)^{-1}$, then we get the standard GMM estimator

$$\hat{\beta}_{GMM} = (X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'y.$$

This is quite a mouthful to remember, so if we recognize that the expression contains a projection matrix with respect to W, we can instead write the GMM estimator in a more compact form

$$\hat{\beta}_{GMM} = (X'P_W X)^{-1}X'P_W y.$$

Notice that this is identical to the 2SLS estimator, so the asymptotic distribution can be derived in an identical fashion as well.

Question 3 (bonus): The J-test of overidentifying restrictions

(a) If W is full rank, then for any conformable column vector $x \neq 0$,

$$Wx \neq 0 \qquad \text{and} \qquad x'W' = (Wx)' \neq 0.$$

Therefore, for any such vector $x \neq 0$, one has

$$x'W'Wx = (Wx)'Wx > 0,$$

as this this constitutes a sum of squares where at least one element is non-zero. In this sense, we conclude that $W'W$ is positive definite.

(b) Observe that

$$\begin{aligned}\mathbb{E}(z) &= \mathbb{E}[(W'W)^{-1/2}W'u] \\ &= (W'W)^{-1/2}W'\mathbb{E}(u) \\ &= 0\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(z) &= \mathrm{Var}[(W'W)^{-1/2}W'u] \\ &= (W'W)^{-1/2}W'\mathrm{Var}(u)W(W'W)^{-1/2} \\ &= \sigma^2(W'W)^{-1/2}W'W(W'W)^{-1/2} \\ &= \sigma^2(W'W)^{-1/2}(W'W)^{1/2}(W'W)^{1/2}(W'W)^{-1/2} \\ &= \sigma^2 I_n.\end{aligned}$$

Therefore, the regular rules of variance mean that if we normalize z by σ, we have

$$\frac{z}{\sigma} \sim N(0, I_n).$$

(c) Direct substitution of the GMM estimator into the objective function yields

$$\begin{aligned}
f(\hat{\beta}_{GMM}) &= (y - X\hat{\beta})'P_W(y - X\hat{\beta}) \\
&= (y - X[X'P_WX]^{-1}X'P_Wy)'P_W(y - X[X'P_WX]^{-1}X'P_Wy).
\end{aligned}$$

Let's simplify the terms in brackets first before proceeding:

$$\begin{aligned}
y - X[X'P_WX]^{-1}X'P_Wy &= y - X[X'P_WX]^{-1}X'P_W(X\beta + u) \\
&= \underbrace{y - X\beta}_{=u} - X[X'P_WX]^{-1}X'P_Wu \\
&= u - X[X'P_WX]^{-1}X'P_Wu \\
&= [I - X[X'P_WX]^{-1}X'P_W]u.
\end{aligned}$$

So if we put the whole thing together, we have a lot of painful algebra, shown below. Note that most of it is just factorization and expansion of brackets, while in the second last equality we take advantage of the fact that the expressions that we are dealing with are, in fact, scalars (so they are equal to their transpose):

$$\begin{aligned}
f(\hat{\beta}_{GMM}) &= \left([I - X[X'P_WX]^{-1}X'P_W]u\right)' P_W \left([I - X[X'P_WX]^{-1}X'P_W]u\right) \\
&= u'[I - P_WX[X'P_WX]^{-1}X']P_W[I - X[X'P_WX]^{-1}X'P_W]u \\
&= u'[P_W - P_WX[X'P_WX]^{-1}X'P_W][I - X[X'P_WX]^{-1}X'P_W]u \\
&= [u'P_W - u'P_WX[X'P_WX]^{-1}X'P_W][u - X[X'P_WX]^{-1}X'P_Wu] \\
&= u'P_Wu - u'P_WX[X'P_WX]^{-1}X'P_Wu - u'P_WX[X'P_WX]^{-1}X'P_Wu \\
&\quad + u'P_WX[X'P_WX]^{-1}X'P_WX[X'P_WX]^{-1}X'P_Wu \\
&= u'P_Wu - 2e'P_WX[X'P_WX]^{-1}X'P_Wu + u'P_WX[X'P_WX]^{-1}X'P_Wu \\
&= \underbrace{u'P_Wu}_{(1)} - \underbrace{u'P_WX[X'P_WX]^{-1}X'P_Wu}_{(2)}.
\end{aligned}$$

Now we can start to disassemble this expression. As in the question, we may write

$$(W'W)^{-1} = (W'W)^{-1/2}(W'W)^{-1/2}$$

so that

$$\begin{aligned}
(1) = u'P_Wu &= u'W(W'W)^{-1}W'u \\
&= u'W(W'W)^{-1/2}(W'W)^{-1/2}W'u \\
&= z'z
\end{aligned}$$

as per the definition of z that we were given earlier. Also, we can write

$$u'P_W = u'W(W'W)^{-1}W' = u'W(W'W)^{-1/2}(W'W)^{-1/2}W' = z'(W'W)^{-1/2}W'$$

which means that we can manipulate (2) to give

$$\begin{aligned}
(2) &= u'P_WX[X'P_WX]^{-1}X'P_Wu \\
&= z'\underbrace{(W'W)^{-1/2}W'X}_{A}[X'P_WX]^{-1}\underbrace{X'W(W'W)^{-1/2}}_{A}z \\
&= z'A[X'P_WX]^{-1}A'z.
\end{aligned}$$

It seems like we're almost there! We just need to check the term in square brackets:

$$\begin{aligned}
X'P_W X &= X'W(W'W)^{-1}W'X \\
&= X'W(W'W)^{-1/2}(W'W)^{-1/2}X \\
&= A'A
\end{aligned}$$

which is in line with our definitions. Hence, we can now write

$$\begin{aligned}
f(\hat{\beta}_{GMM}) &= (1) - (2) \\
&= z'z - z'A(A'A)^{-1}A'z \\
&= z'z - z'P_A z \\
&= z[I - P_A]z \\
&= z'M_A z,
\end{aligned}$$

as required.

(d) Before starting, it would be good to check the dimensions of A. Since W is $n \times (k_1 + p)$ and X is $n \times k$,

$$A = \underbrace{(W'W)^{-1/2}}_{(k_1+p)\times(k_1+p)} \underbrace{W'X}_{(k_1+p)\times k} = (k_1 + p) \times k \text{ matrix}$$

so $A'A$ is a $k \times k$ matrix. The reason why this is useful comes from its use in determining the rank of M_A. In particular, since M_A is a residual maker, it is symmetric and idempotent. For such matrices, we know that the rank of the matrix is equal to their trace so

$$\begin{aligned}
\text{rank}(M_A) &= \text{tr}(I_{k_1+p} - A(A'A)^{-1}A') \\
&= \text{tr}(I_{k_1+p}) - \text{tr}(A(A'A)^{-1}A') \\
&= k_1 + p - \text{tr}(A'A(A'A)^{-1}) && \because \text{tr}(AB) = \text{tr}(BA) \\
&= k_1 + p - \text{tr}(I_k) \\
&= k_1 + p - k \\
&= k_1 + p - (k_1 + k_2) && \because k = k_1 + k_2 \\
&= p - k_2 \\
&= \#\text{ IVs} - \#\text{ endogenous}.
\end{aligned}$$

In this case, we obtain the $\text{tr}(I_k)$ term since we know that $A'A$ is a $k \times k$ matrix.

(e) Here is one way to show it: since our previous results depend on the true value σ, we must manipulate the GMM objective function so that we can get σ back into it:

$$\frac{f(\hat{\beta})}{\hat{\sigma}^2} = \frac{z'M_A z}{\hat{\sigma}^2} \times \frac{\sigma^2}{\sigma^2} = \frac{z'M_A z}{\sigma^2} \div \frac{\hat{\sigma}^2}{\sigma^2}.$$

In this case we can write

$$\frac{z'M_A z}{\sigma^2} = \left(\frac{z}{\sigma}\right)' M_A \left(\frac{z}{\sigma}\right) \xrightarrow{d} \chi^2_{p-k_2}$$

using our previous results.[3] Since $\hat{\sigma}^2$ is a consistent estimator for σ^2, one also has

$$\frac{\hat{\sigma}^2}{\sigma^2} \xrightarrow{p} 1.$$

Therefore, the use of Slutsky's Theorem allows us to say that

$$\frac{f(\hat{\beta})}{\hat{\sigma}^2} \xrightarrow{d} \chi^2_{p-k_2} \times 1 \xrightarrow{d} \chi^2_{p-k_2},$$

as required.

[3]Specifically, we need the following theorem: if $u \sim N(0, I_n)$ and A is a $n \times n$ symmetric idempotent matrix with rank q, then $u'Au \sim \chi^2_q$.