

# Empirical Analysis I - Problem Set 5

*Timothy Schwieg*

*Paulo Henrique Ramos*

*Samuel Barker*

*Rafeh Qureshi*

## Question 1

**a**

Consider the lagrangian below:

$$\mathcal{L}(\beta, \lambda) = \frac{1}{2n} \sum_{i=1}^n (Y_i - X_i' \beta)^2 + \lambda' (R\beta - c)$$

Taking gradients and applying the rules of matrix calculus:

$$\nabla_{\beta} \mathcal{L} = -\frac{1}{n} \sum_{i=1}^n X_i (Y_i - X_i' \beta) + R' \lambda = 0$$

$$\nabla_{\lambda} \mathcal{L} = R\beta - c = 0$$

Let  $\tilde{\beta}_n, \tilde{\lambda}_n$  be such that this system is satisfied.

**b**

Expanding the condition on  $\beta$ :

$$\begin{aligned} -\frac{1}{n} \sum_{i=1}^n X_i Y_i + \frac{1}{n} \sum_{i=1}^n X_i X_i' \tilde{\beta}_n + R' \tilde{\lambda}_n &= 0 \\ \frac{1}{n} \sum_{i=1}^n X_i Y_i - R' \tilde{\lambda}_n &= \frac{1}{n} \sum_{i=1}^n X_i X_i' \tilde{\beta}_n \\ \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n X_i Y_i \right) - \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \tilde{\lambda}_n &= \tilde{\beta}_n \\ \hat{\beta}_n - \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \tilde{\lambda}_n &= \tilde{\beta}_n \end{aligned}$$

Note that the inversion is possible because of the same line of reasoning that allowed for it during the calculation of the OLS estimator; that is, because there is no perfect colinearity in  $X$  and  $\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \xrightarrow{p} \mathbb{E}[XX']^{-1}$  due to WLLN, the matrix of samples must have an inverse with probability approaching one.

Substituting the expression above for  $\tilde{\beta}_n$  in the second condition:

$$\begin{aligned} R\tilde{\beta}_n &= c \\ R\hat{\beta}_n - R\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} R'\tilde{\lambda}_n &= c \\ \left(R\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} R'\right)^{-1} (R\hat{\beta}_n - c) &= \tilde{\lambda}_n \end{aligned}$$

**c**

$\tilde{\lambda}_n$  is a continuous function of  $\hat{\beta}_n$  and  $\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1}$ , as matrix multiplication, addition and inversion are all continuous operations. Because  $\hat{\beta}_n \xrightarrow{p} \beta$  and  $\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \xrightarrow{p} \mathbb{E}[XX']^{-1}$  (by the Weak Law of Large Numbers and Continuous Mapping Theorem), and marginal convergence in probability implies joint convergence, we can apply the CMT to conclude:

$$\tilde{\lambda}_n \xrightarrow{p} \left(R\mathbb{E}[XX']^{-1} R'\right)^{-1} (R\beta - c) = 0,$$

when  $(R\beta - c) = 0$ .

**d**

As seen in class, the Limiting distribution of  $\hat{\beta}_n$  is given by:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta) &\xrightarrow{d} \mathcal{N}(0, \Omega) \\ \Omega &= \mathbb{E}[XX']^{-1} \mathbb{V}(Xu) \mathbb{E}[XX']^{-1} \end{aligned}$$

Notice then that we can apply CMT to obtain:

$$\sqrt{n}(R\hat{\beta}_n - c - R\beta + c) = \sqrt{n}(R\hat{\beta}_n - R\beta) \xrightarrow{d} \mathcal{N}(0, R\Omega R')$$

And, when  $(R\beta - c) = 0$ , this implies:  $\sqrt{n}(R\hat{\beta}_n - c) \xrightarrow{d} \mathcal{N}(0, R\Omega R')$ .

Now, because  $\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \xrightarrow{p} \mathbb{E}[XX']^{-1}$ , and using CMT and Slutsky, we can con-

clude that:

$$\begin{aligned}\sqrt{n}\tilde{\lambda}_n &= \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} \sqrt{n} (R\hat{\beta}_n - c) \\ &\xrightarrow{d} \left( R \mathbb{E}[X X']^{-1} R' \right)^{-1} \mathcal{N}(0, R\Omega R') = \mathcal{N}(0, AR\Omega R'A'),\end{aligned}$$

where  $A := (R \mathbb{E}[X X']^{-1} R')^{-1}$ .

**e**

To construct a feasible test, we need to estimate the covariance matrix  $\Lambda := AR\Omega R'A'$  in the asymptotic distribution above. First, we may estimate  $\Omega$  with the heteroskedasticity-robust estimators used in class:

$$\hat{\Omega}_n = \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' (\hat{u}_i)^2 \right) \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1}$$

We saw in class that  $\hat{\Omega}_n \xrightarrow{p} \Omega$ . Also, we saw above that  $\left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \xrightarrow{p} \mathbb{E}[X X']^{-1}$ . Thus, by CMT, we have that:

$$\hat{\Lambda}_n := \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} R \hat{\Omega}_n R' \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} \xrightarrow{p} \Lambda$$

So,  $\hat{\Lambda}_n$  is a consistent estimator of the covariance matrix of  $\sqrt{n}\tilde{\lambda}_n$ , under the null hypothesis. Therefore, under  $(R\beta = c)$ , we have, by slusky and the fact that  $\Lambda$  is invertible (it is a multiplication of invertible matrices):

$$\begin{aligned}\hat{\Lambda}_n^{-\frac{1}{2}} \sqrt{n} (\tilde{\lambda}_n) &\xrightarrow{d} \mathcal{N}(0, 1) \\ n \tilde{\lambda}_n' \hat{\Lambda}_n^{-1} \tilde{\lambda}_n &\xrightarrow{d} \chi_p^2\end{aligned}$$

since we have  $p$  restrictions.

This depends completely on known quantities, and can therefore be tested. We define the test statistic as:

$$T_n := n \tilde{\lambda}_n' \hat{\Lambda}_n^{-1} \tilde{\lambda}_n$$

Our test will take the form of:

$$\mathbb{1}_{\{T_n > c_{p,1-\alpha}\}}$$

where  $c_{p,1-\alpha}$  is the critical value for the  $\alpha$  specified in the chi-square distribution with  $p$  degrees of freedom.

And, as shown below, using portmanteau lemma, our test is consistent in level:

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \mathbb{E} [\mathbb{1}_{\{T_n > c_{p,1-\alpha}\}}] &= \limsup_{n \rightarrow \infty} \Pr(T_n > c_{p,1-\alpha}) \\
&\leq \limsup_{n \rightarrow \infty} \Pr\left(n\tilde{\lambda}'_n \hat{\Lambda}_n^{-1} \tilde{\lambda}_n \geq c_{p,1-\alpha}\right) \\
&\leq \Pr(\chi_p^2 \geq c_{p,1-\alpha}) \\
&= \alpha
\end{aligned}$$

**f**

Both tests are actually equivalent. We can see that by substituting the expressions for  $\tilde{\lambda}_n$  and  $\hat{\Lambda}_n^{-1}$  into the test statistic used above:

$$\begin{aligned}
T_n &= n \left\{ \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} (R\hat{\beta}_n - c) \right\}' \\
&\quad \left\{ \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} R\hat{\Omega}_n R' \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} \right\}^{-1} \\
&\quad \left\{ \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} (R\hat{\beta}_n - c) \right\}
\end{aligned}$$

$$\begin{aligned}
T_n &= n (R\hat{\beta}_n - c)' \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} \\
&\quad \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right) (R\hat{\Omega}_n R')^{-1} \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right) \\
&\quad \left( R \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} R' \right)^{-1} (R\hat{\beta}_n - c)
\end{aligned}$$

$$T_n = n (R\hat{\beta}_n - c)' (R\hat{\Omega}_n R')^{-1} (R\hat{\beta}_n - c)$$

Thus, we can see that both test have the same test statistic and the same asymptotic distribution, thus they are equivalent.

## Question 2

**a**

$$\begin{aligned}\mathbb{V}(\hat{X}) &= \mathbb{V}(X) + \mathbb{V}(V) + 2\text{Cov}(X, V) \\ &= \mathbb{V}(X) + \mathbb{V}(V) + 2(\mathbb{E}[XV] - \mathbb{E}[X]\mathbb{E}[V]) \\ &= \mathbb{V}(X) + \mathbb{V}(V)\end{aligned}$$

As  $\mathbb{E}[V] = \mathbb{E}[XV] = 0$ .

$$\begin{aligned}\mathbb{V}(Y) &= \mathbb{V}(\beta_0) + \mathbb{V}(\beta_1 X) + \mathbb{V}(U) + 2(\text{Cov}(\beta_0, U) + \text{Cov}(\beta_1 X, U) + \text{Cov}(\beta_0, \beta_1 X)) \\ &= \beta_1^2 \mathbb{V}(X) + \mathbb{V}(U) + 2(\beta_1 \mathbb{E}[XU] - \beta_1 \mathbb{E}[X]\mathbb{E}[U]) \\ &= \beta_1^2 \mathbb{V}(X) + \mathbb{V}(U)\end{aligned}$$

As  $\beta_0$  is a constant, and  $\mathbb{E}[U] = \mathbb{E}[XU] = 0$ .

$$\begin{aligned}\text{Cov}(\hat{X}, Y) &= \text{Cov}(X, Y) + \text{Cov}(V, Y) \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[VY] - \mathbb{E}[V]\mathbb{E}[Y] \\ &= \mathbb{E}[\beta_0 X + \beta_1 X^2 + UX] - \mathbb{E}[X](\beta_0 + \beta_1 \mathbb{E}[X]) + \mathbb{E}[\beta_0 V + \beta_1 XV + UV] - 0 \\ &= \beta_0 \mathbb{E}[X] + \beta_1 \mathbb{E}[X^2] - \beta_0 \mathbb{E}[X] - \beta_1 \mathbb{E}[X]^2 \\ &= \beta_1 \mathbb{V}(X)\end{aligned}$$

As  $\mathbb{E}[V] = \mathbb{E}[XV] = \mathbb{E}[U] = \mathbb{E}[XU] = 0$ .

**b**

First let's show that  $\frac{\text{Cov}(\hat{X}, Y)}{\mathbb{V}(\hat{X})} \leq \beta_1$ . Let  $\beta_1 \geq 0$ . If  $\beta_1 = 0$ , then  $\text{Cov}(\hat{X}, Y) = 0$  and this result is true trivially. So let  $\beta_1 > 0$ . Then we get that:

$$\frac{\text{Cov}(\hat{X}, Y)}{\mathbb{V}(\hat{X})} = \frac{\beta_1 \mathbb{V}(X)}{\mathbb{V}(X) + \mathbb{V}(V)} \leq \beta_1,$$

since  $\mathbb{V}(X) \leq \mathbb{V}(X) + \mathbb{V}(V)$  and  $\beta_1 > 0$ .

Now, we show that  $\beta_1 \leq \frac{\mathbb{V}(Y)}{\text{Cov}(\hat{X}, Y)}$ . If  $\beta_1 = 0$ , then  $\text{Cov}(\hat{X}, Y) = 0$  and the division is not well-defined (although we may define  $\frac{c}{0} := \infty$ , for  $c > 0$ , and then we have the inequality). Consider  $\beta_1 > 0$ . Then we have:

$$\frac{\mathbb{V}(Y)}{\text{Cov}(\hat{X}, Y)} = \frac{\beta_1^2 \mathbb{V}(X) + \mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} = \beta_1 + \frac{\mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} \geq \beta_1$$

Because  $\beta_1 > 0$  and the variances are positive.

We know that  $\frac{Cov(\hat{X}, Y)}{\mathbb{V}(\hat{X})}$  is the coefficient from the regression  $Y = \beta_0^* + \beta_1^* \hat{X} + U^*$ . Thus we can interpret the lower bound as a proof that the coefficient on the “wrong” regression, in this simple case, will be always weakly smaller than the coefficient on the “true” regression. Similarly, the upper bound is the inverse of the coefficient of the reverse regression  $\hat{X} = \alpha_0 + \alpha_1 Y + z$ , and thus we can interpret it as stating the the true coefficient will be always smaller than the inverse of the coefficient of the reverse regression involving the  $\hat{X}$ .

### c

We already saw that the result holds when  $\beta_1 = 0$  (letter *b* above), so let us examine the case when  $\beta_1 < 0$ . We can do this similarly to letter *b* above:

First the upper bound:

$$\frac{Cov(\hat{X}, Y)}{\mathbb{V}(\hat{X})} = \frac{\beta_1 \mathbb{V}(X)}{\mathbb{V}(X) + \mathbb{V}(V)} \geq \beta_1,$$

because now  $\beta_1 < 0$ , and multiplying it by a number smaller than one decreases its absolute value and makes it bigger, since it is negative.

Then the lower bound:

$$\frac{\mathbb{V}(Y)}{Cov(\hat{X}, Y)} = \frac{\beta_1^2 \mathbb{V}(X) + \mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} = \beta_1 + \frac{\mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} \leq \beta_1$$

Because  $\beta_1 < 0$  and the variances are positive, so adding something negative to a negative number decreases the result.

## Question 3

$X, Z$  are both  $k + 1$  dimensional vectors (thus  $\mathbb{E}[ZX']$  is a square matrix with dimension  $k + 1$ ), and  $\text{rank } \mathbb{E}[ZX'] = k + 1$ . Approach by contrapositive. Let us assume that there is perfect colinearity in  $Z$  and show that this implies  $\text{rank } \mathbb{E}[ZX'] < k + 1$  (it cannot have rank higher than its dimension). That is, assume  $\exists c \neq 0$  such that  $1 = \Pr(c'Z = 0)$ . Thus using this  $c$ , we have:

$$c' \mathbb{E}[ZX'] = \mathbb{E}[(c'Z)X'] = 0$$

But this implies that there exists  $c \neq 0$  such that  $c' \mathbb{E}[ZX'] = 0$ . This means that the rows of  $\mathbb{E}[ZX']$  are not linearly independent. As  $\mathbb{E}[ZX']$  is a square matrix, a row being linearly dependent implies that the determinant of the matrix is zero. Therefore the transpose of this matrix,  $\mathbb{E}[ZX']'$  has a determinant of zero, and linearly dependent rows. Linear dependence in the rows of this matrix is equivalent to linear dependence in the columns of  $\mathbb{E}[ZX']$ . This means the matrix does not have full rank.

Therefore having rank  $k + 1$  in  $\mathbb{E}[ZX']$  implies no perfect colinearity in  $Z$ .

## Question 4

We are given  $V = X - BLP(X|Z)$ . Let us define  $BLP(X|Z) = \Pi'Z$  where  $\Pi$  is a  $\ell + 1$  by  $k + 1$  matrix.

Consider  $BLP(U|V) = V'\gamma$ , note that this has an implied first-order condition of orthogonality.

$$\mathbf{0} = \mathbb{E}[V(U - V'\gamma)] = \mathbb{E}[V\tilde{U}]$$

This is exactly exogeneity for  $V$ .

From the first-order conditions on the  $BLP(X|Z)$  we know that for all  $j$ ,

$$\begin{aligned}\mathbb{E}[Z(X_j - \Pi'Z)] &= 0 \\ \mathbb{E}[ZV_j] &= 0 \\ \mathbb{E}[ZV'] &= 0\end{aligned}$$

Where the last result accumulates all of the  $j$  vector conditions into one matrix condition. We may also note that:

$$\begin{aligned}\mathbb{E}[X\tilde{U}] &= \mathbb{E}[V\tilde{U} + BLP(X|Z)\tilde{U}] = \mathbb{E}[BLP(X|Z)\tilde{U}] \\ &= \mathbb{E}[\Pi'Z(U - V'\gamma)] \\ &= \Pi'\mathbb{E}[ZU] - \Pi'\mathbb{E}[ZV']\gamma \\ &= \Pi'\mathbf{0} - \Pi'\mathbf{0}\gamma = \mathbf{0}\end{aligned}$$

## b

Using the sub-vector results from linear regression before, we may solve for  $\beta$  as follows. Define  $\tilde{Y} = Y - BLP(Y|V)$  and  $\tilde{X} = X - BLP(X|V)$ .

Notice that  $BLP(X|Z)$  is simply  $X - \Pi'Z$ , since  $V = X - \Pi'Z$ , and thus predicting  $X$  on  $V$  should return  $X$  plus some difference, which would be the best linear predictor of  $X$  given the term in  $V$  different from  $X$ . This would give  $\tilde{X} = X - X + \Pi'Z = \Pi'Z$ , since  $\Pi'Z$  is already  $BLP(X|Z)$ .

Then

$$\begin{aligned}\beta &= \mathbb{E}[\tilde{X}\tilde{X}']^{-1}\mathbb{E}[\tilde{X}\tilde{Y}] \\ &= \mathbb{E}[\tilde{X}\tilde{X}']^{-1}\mathbb{E}[\tilde{X}Y] \\ &= \mathbb{E}[\Pi'ZZ'\Pi']^{-1}\mathbb{E}[\Pi'ZY] \\ &= \Pi'\mathbb{E}[ZZ']\Pi^{-1}\Pi'\mathbb{E}[ZY]\end{aligned}$$

This is exactly our instrumental variables estimator that we derived in class.

More formally, we can show that, using the conditions that  $\mathbb{E}[VX'] = \mathbb{E}[VV']\omega$  and  $\mathbb{E}[VY'] = \mathbb{E}[VV']\phi$ , and defining  $BLP(X|V) = \omega'V$  and  $BLP(Y|V) = V'\phi$ :

$$\begin{aligned}\beta &= \mathbb{E}[\tilde{X}\tilde{X}']^{-1}\mathbb{E}[\tilde{X}\tilde{Y}] \\ &= (\mathbb{E}[XX'] - \mathbb{E}[XV']\omega)^{-1}(\mathbb{E}[XY] - \mathbb{E}[XV']\phi)\end{aligned}$$

Also from defining  $BLP(X|V) = \omega'V$ , we can show:

$$\begin{aligned}\omega &= \mathbb{E}[VV']^{-1}\mathbb{E}[VX] \\ &= (\mathbb{E}[XX'] - \mathbb{E}[XZ']\Pi - \Pi'\mathbb{E}[ZX'] + \Pi'\mathbb{E}[ZZ']\Pi)^{-1}(\mathbb{E}[XX'] - \Pi'\mathbb{E}[ZX']) \\ &= (\mathbb{E}[XX'] - \mathbb{E}[XZ']\Pi)^{-1}(\mathbb{E}[XX'] - \Pi'\mathbb{E}[ZX'])\end{aligned}$$

where we expanded  $V$  and used that:  $\mathbb{E}[ZX'] = \mathbb{E}[ZZ']\Pi$  by the definition of best linear predictor.

Replacing the expression for  $\omega$  in the first part of the expression for  $\beta$  above we can show:

$$\beta = (\Pi'\mathbb{E}[ZX'])^{-1}(\mathbb{E}[XY] - \mathbb{E}[XV']\phi)$$

where we also used that  $V = X - \Pi'Z$  to simplify the first part of the expression for  $\beta$ .

Similarly, simplifying the second part of  $\beta$  and substituting the expression for  $\phi$  (which is similar to what we did for  $\omega$  above) we get:

$$\begin{aligned}\beta &= (\Pi'\mathbb{E}[ZX'])^{-1}(\mathbb{E}[XY] - \mathbb{E}[XV']\phi) \\ &= (\Pi'\mathbb{E}[ZX'])^{-1}(\mathbb{E}[\Pi'ZY])\end{aligned}$$

And this is exactly the expression seen in class.

## Question 5

**a**

Since  $Z'\lambda = BLP(Y|Z)$ , we can use the expression for the best linear predictor and: (i) that  $\mathbb{E}[ZZ']$  exists and is invertible, since there is no perfect colinearity in  $Z$ , and (ii)  $\mathbb{E}[ZY] = \mathbb{E}[ZX']\beta + \mathbb{E}[ZU]$  exists (since both elements in the summation exist by assumption), to get:

$$\begin{aligned}\lambda &= \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZY] \\ &= \mathbb{E}[ZZ']^{-1}\mathbb{E}[Z(X'\beta + U)] \\ (\text{Because } \mathbb{E}[ZU] &= 0) = \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX']\beta\end{aligned}$$

Now we also have that  $\Gamma'Z = BLP(X|Z)$ , and, using the same facts stated in the first paragraph above, we can also get the expression for  $\Gamma = \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX']$ . Therefore, using the result above, we get:  $\lambda = \Gamma\beta$  as we wanted.



Now we can substitute this into the reduced form equation for  $Y$  to get:

$$\begin{aligned}\epsilon &= Y - Z'\lambda \\ &= X'\beta + U - Z'\Gamma\beta \\ &= (X' - Z'\Gamma)\beta + U \\ (\text{Because } \eta &= X - \Gamma'Z) = \eta'\beta + U,\end{aligned}$$

which is the result we wanted.

## b

To get an expression for  $\beta$ , we can substitute for  $\lambda$  in the reduced form for  $Y$  to get:

$$Y = (\Gamma'Z)'\beta + \epsilon$$

Now we can check that:  $\mathbb{E}[\Gamma'Z\epsilon] = 0$

$$\begin{aligned}\mathbb{E}[\Gamma'Z\epsilon] &= \mathbb{E}[\Gamma'Z(\eta'\beta + U)] \\ &= \Gamma'\mathbb{E}[Z\eta']\beta + \Gamma'\mathbb{E}[ZU] = 0,\end{aligned}$$

because  $\eta = X - BLP(X|Z)$ , so the first order condition for the best linear predictor gives us that  $\mathbb{E}[Z\eta'] = 0$ , and also we have by  $\mathbb{E}[ZU]$  assumption.

Thus, since  $\mathbb{E}[\Gamma'Z\epsilon] = 0$ , we can obtain  $\beta$  as  $BLP(Y|\Gamma'Z)$ , which results in:  $\beta = \mathbb{E}[\Gamma'ZZ'\Gamma]^{-1}\mathbb{E}[\Gamma'ZY] = \Gamma'\mathbb{E}[ZZ']\Gamma^{-1}\Gamma'\mathbb{E}[ZY]$ . Notice that this expression is exactly the one derived in class, except that in class we called  $\Gamma'Z = BLP(X|Z)$  as  $\Pi'Z$ .

## Question 6

### a

Suppose we have that  $\mathbb{E}(U) = \alpha \neq 0$ . Then, we have

$$E(U) = E(\log(Y) - \beta_0 - \beta_1X_1 - \beta_2X_2) = \alpha$$

But, then we can write

$$\log(Y) = \tilde{\beta}_0 + \beta_1X_1 + \beta_2X_2 + \tilde{U}$$

, wherein  $\tilde{\beta}_0 = \beta_0 + \alpha$  and  $\tilde{U} = U - \alpha$ . Then, we see that

$$\begin{aligned}E(\tilde{U}) &= E(\log(Y) - \tilde{\beta}_0 - \beta_1X_1 - \beta_2X_2) \\ &= E(\log(Y) - \beta_0 - \alpha - \beta_1X_1 - \beta_2X_2) \\ &= E(\log(Y) - \beta_0 - \beta_1X_1 - \beta_2X_2) - E(\alpha) \\ &= \alpha - \alpha = 0\end{aligned}$$

Note, as this states, we cannot separate  $E(U)$  from  $E(\beta_0 + U)$ , because  $U$  is not a variable included in the regression. Namely, the mean of  $U$  will be included in the expectation of the calculated constant in the regression.

**b**

We can say  $X_k$  is exogenous if  $E(X_k U) = 0$ . This means  $X_k$  is orthogonal to the error term. With our examples from class, we can say that there is no measurement error in  $X_k$ , there are no variables omitted in the regression that are correlated with  $X_k$  and  $X_k$  is not determined simultaneously with  $Y$ .

Similarly, we can say  $X_k$  is endogenous if  $E(X_k U) \neq 0$ .

In our particular example, we would guess that  $E(UX_1) \neq 0$  as there are variables not included in the regression that are correlated with years of education and are also correlated with  $Y$ , but are not included in our regression. The classic example of this is *ability*. An individual's ability, such as her intelligence and sedulousness is likely to influence her decision to enroll in more schooling and are also likely to cause her to have a higher hourly wage. Without controlling for ability however, we are likely to attribute to years of education what is partially caused by ability.

**c**

We say the instrument is exogenous if it is orthogonal to the error term;  $E(Zu) = 0$ . Intuitively, it says that the instrument does not itself affects directly the outcome variable of interest (but affects it through the instrumented variables, as seen in the relevance condition below).

We say the instrument is relevant if the  $\text{rank}(\mathbb{E}(ZX')) = k + 1$ , where  $k$  is the number of covariates in  $X$  other than the constant (in this case,  $k = 2$ ).

In class, we saw that, given the lack of perfect colinearity of  $Z$ , the rank of  $\mathbb{E}[ZX']$  is the same as the rank of  $\Pi$ , such that  $\Pi'Z = BLP(X|Z)$ . Thus, the condition above translates into rank of  $\Pi$  being  $k + 1$ . In this particular case, where the dimension of  $Z$  is also  $k + 1$ , we can easily interpret this condition by looking at the matrix  $\Pi$ :

$$\Pi' = \begin{pmatrix} 1 & 0 & 0 \\ \pi_0 & \pi_1 & \pi_2 \\ 0 & 0 & 1 \end{pmatrix}$$

The matrix looks like that because predicting a constant and  $X_2$  given a constant and  $X_2$  results in the first and last lines, and the coefficients on the second line are used to predict  $X_1$  given  $Z_1$ . For this matrix to have rank  $k + 1$ , where  $k = 2$ , we must have  $\pi_1 \neq 0$ . That is, after controlling for the constant and  $X_2$ ,  $Z_1$  must still have some predictive power over  $X_1$ .

**d**

We can use the IV estimator, since we are in the exactly identified case; namely

$$\hat{\beta}_{IV} = \left( \frac{1}{N} \sum_{i=1}^N Z_i X_i' \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^N Z_i Y_i' \right)$$

In the exactly identified case, this would be equivalent to the 2SLS estimator. However, for the overidentified case (where we have more than one instrumental variable for  $X_1$ ), we go over the 2SLS procedure. We would go about

1. Regressing  $X_1$  on the entire matrix  $Z$ , and we can call the estimated coefficients  $\pi$ .
2. Getting the estimated  $X_1$  from the regression in (1). Namely,  $\hat{X}_1 = \hat{\pi}'Z$ .
3. Regressing  $Y$  on  $\hat{X}_1$  and  $X_2$ . This will give us

$$\hat{\beta}_{TSLS} = \left( \frac{1}{N} \sum_{i=1}^N \hat{\pi}' Z_i Z_i' \hat{\pi} \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^N \hat{\pi}' Z_i Y_i \right)$$

**e**

We can look at the “first stage” regression of

$$X_1 = \gamma_0 + \gamma_1 Z_1 + \gamma_3 X_2 + \epsilon$$

As seen above, for  $Z$  to be a relevant instrument, we must have  $\gamma_1 \neq 0$ . Thus, we would like to test the null hypothesis  $\gamma_1 = 0$  (not relevant) against the alternative hypothesis  $\gamma_1 \neq 0$ . Interpreting this regression as best linear predictor and using OLS to estimate  $\gamma$ , we recall that for the vector of coefficients  $\gamma$ ,

$$\sqrt{N}(\hat{\gamma} - \gamma) \rightarrow N(0, \Omega)$$

Where is  $\Omega = \mathbb{E}[ZZ']^{-1} \mathbb{V}(Z\epsilon) \mathbb{E}[ZZ']^{-1}$ , as seen in class. Also, using CMT and  $r = (0, 1, 0)'$ , we get that:  $\sqrt{N}(r'\hat{\gamma} - r'\gamma) \rightarrow N(0, r'\Omega r)$ , or equivalently:  $\sqrt{N}(\hat{\gamma}_1 - \gamma_1) \rightarrow N(0, \Omega_{1,1})$ , where  $\Omega_{1,1}$  is the central element of the matrix  $\Omega$ .

We can now build a consistent estimator  $\hat{\Omega}$  (and thus of  $\Omega_{1,1}$ , by CMT), and reject the null if the test statistic below is above  $z_{1-\frac{\alpha}{2}}$ :

$$T_n := \left| \frac{\sqrt{N}(\hat{\gamma}_1)}{\sqrt{\Omega_{2,2}}} \right| > z_{1-\frac{\alpha}{2}}$$

This test has level alpha, since, under the null,  $\left| \frac{\sqrt{N}(\hat{\gamma}_1)}{\sqrt{\Omega_{2,2}}} \right| \xrightarrow{d} |N(0, 1)|$  (by Slutsky and CMT), and  $P(|z| \geq z_{1-\frac{\alpha}{2}}) = \alpha$ , where  $z \sim N(0, 1)$ .

**f**

Note, as  $Z = (1, Z_1, X_2)'$ , we see that the rank condition would be violated if we tried using  $Z_1 = X_2$  as the instrument (we are adding fewer instrumental variables to the model here than the number of endogenous variables we have).

Now, suppose, alternatively, that we remove  $X_2$  from the regression and instead try to estimate

$$\log(Y) = \beta_0 + \beta_1 X_1 + \tilde{U}$$

Then, we may no longer have that  $\text{Cov}(Z, \tilde{U}) = 0$ . Indeed, if we are right about the model and  $\beta_2 \neq 0$ , this would have to be the case.