Problem Set 5

Artschil Okropiridse*

March 8, 2024

1 Lecture 10 Multivariate and non-linear regression

Notation for multivariate regression is quite intense. To clarify it a bit, let's go trough a simple case. Let's start with the population, say we are interested in a system of two equations ($m = 2$):

$$\begin{aligned} Y_1 &= \beta_0 + X_1\beta_1 + X_2\beta_2 + e_1 &&= X'\beta + e_1 \\ Y_2 &= \delta_0 + Z_1\delta_1 + Z_2\delta_2 + e_2 &&= Z'\delta + e_2 \end{aligned}$$

Now we can rewrite this as a singular expression:

$$\begin{aligned} &\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & X_1 & X_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & Z_1 & Z_2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \delta_0 \\ \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\ &\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X' & 0 \\ 0 & Z' \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\ &\overset{\text{define}}{\Longrightarrow} Y = \bar{X}\beta + e \end{aligned}$$

So in this case Y is 2×1, $\bar{X}$ is 2×6, β is 6×1, and e is 2×1.

If we now move to the sample we can stack this system n times on top of each other where

*many thanks to Jakob Beuschlein

for Y_{ij}, i denotes the individual and j the model, so we write:

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{n1} \\ Y_{n2} \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \equiv \mathbf{Y}, \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \delta_0 \\ \delta_1 \\ \delta_2 \end{pmatrix} = \beta$$

$$\begin{pmatrix} 1 & X_{11} & X_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & Z_{11} & Z_{21} \\ \vdots & \ddots & & & & \\ 1 & X_{1n} & X_{2n} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & Z_{1n} & Z_{2n} \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix} = \bar{\mathbf{X}}, \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \mathbf{e}$$

Which then gives us the sample model:

$$\mathbf{Y} = \bar{\mathbf{X}}\beta + \mathbf{e}$$

Exercise 1

(Ch. 11, ex. 11.1) Show $\Omega = \mathbb{E}\left[\bar{X}_i'\Sigma\bar{X}_i\right]$ (11.10) when errors are conditionally homoskedastic (11.8).

Solution

The vector $\bar{X}_i'e_i$ is iid across i and mean zero. Under existence of second moments, the CLT implies[1]

$$\frac{1}{\sqrt{n}}\sum_i \bar{X}_i'e_i \to_d N(0,\Omega)$$

where $\Omega = \mathbb{E}\left[\bar{X}_i'e_ie_i'\bar{X}_i\right] = \mathbb{E}\left[\bar{X}_i'\Sigma_i\bar{X}_i\right]$. Conditional homoskedasticity implies $\mathbb{E}\left[\Sigma_i|X\right] = \Sigma$ and we can therefore apply the LIE and write

$$\Omega = \mathbb{E}\left[\bar{X}_i'\Sigma_i\bar{X}_i\right] = \mathbb{E}\left[\mathbb{E}\left[\bar{X}_i'\Sigma_i\bar{X}_i|X\right]\right] = \mathbb{E}\left[\bar{X}_i'\mathbb{E}\left[\Sigma_i|X\right]\bar{X}_i\right] = \mathbb{E}\left[\bar{X}_i'\Sigma\bar{X}_i\right].$$

To clarify what we are dealing with here, note that in the example at the top, where $m = 2$ and i denotes an individual, we have:

$$e_ie_i' = \Sigma_i = \begin{pmatrix} e_{i1} \\ e_{i2} \end{pmatrix} \begin{pmatrix} e_{i1} & e_{i2} \end{pmatrix} = \begin{pmatrix} e_{i1}^2 & e_{i1}e_{i2} \\ e_{i1}e_{i2} & e_{i2}^2 \end{pmatrix}$$

[1]Note that we can rewrite $\frac{1}{\sqrt{n}} = \sqrt{n}\frac{1}{n}$. Thus $\frac{1}{\sqrt{n}}\sum_i \bar{X}_i'e_i = \sqrt{n}\left(\frac{1}{n}\sum_i \bar{X}_i'e_i - 0\right)$. Maybe that makes it clearer why the CLT applies.

Exercise 2

(Ch. 11, ex. 11.15) The observations are iid. $(y_{1i}, y_{2i}, X_i : \ i = 1, ..., n)$. The dependent variables y_{1i} and y_{2i} are real-valued. The regressor X_i is a k-vector. The model is the two equation system

$$\begin{aligned} y_1 &= X'\beta_1 + e_1, \quad \mathbb{E}[X_i e_{1i}] = 0 \\ y_2 &= X'\beta_2 + e_2, \quad \mathbb{E}[X_i e_{2i}] = 0. \end{aligned}$$

(a) What are the appropriate estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ for β_1 and β_2?

(b) Find the joint asymptotic distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$.

(c) Describe a test for $H_0 : \beta_1 = \beta_2$.

Solution

(a) We can simply estimate $\hat{\beta}_j$ $j = 1, 2$ via OLS which gives us

$$\hat{\beta}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y_j.$$

This is equivalent to estimating them in a multivariate regression with common regressors and $\mathbf{y} = \begin{pmatrix} y_1 & y_2 \end{pmatrix}'$ which yields the estimator

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1} \bar{\mathbf{X}}'\mathbf{y}$$

(b) We can center our estimators and get

$$\hat{\beta} - \beta = \left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1} \bar{\mathbf{X}}'\mathbf{e}.$$

By applying the WLLN and the CLT we get the asymptotic distribution

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to_d N\left(\mathbf{0}, \mathbf{V}_\beta\right)$$

with

$$\mathbf{V}_\beta = \mathbb{E}\left[\bar{X}_i'\bar{X}_i\right]^{-1} \mathbf{\Omega} \mathbb{E}\left[\bar{X}_i'\bar{X}_i\right]^{-1}.$$

Given that we have only common regressors we can go a little further and write:

$$\mathbb{E}\left[\bar{X}_i'\bar{X}_i\right] = \mathbb{E}\left[\mathbf{I}_k \otimes X_i X_i'\right] = \mathbf{I}_k \otimes \mathbb{E}\left[X_i X_i'\right].$$

Furthermore

$$\Omega = \mathbb{E}[e_i e_i' \otimes X_i X_i'].$$

It follows that

$$\mathbf{V}_\beta = \left(\mathbf{I}_k \otimes \mathbb{E}\left[X_i X_i'\right]^{-1}\right) \mathbb{E}[e_i e_i' \otimes X_i X_i'] \left(\mathbf{I}_k \otimes \mathbb{E}\left[X_i X_i'\right]^{-1}\right).$$

(c) First, we should decide on a significance level α. Next, we can rephrase $H_0 : \beta_1 - \beta_2 = 0$ and define the restriction matrix

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_k & -\mathbf{I}_k \end{pmatrix}'.$$

Then using an appropriate estimator, e.g. $\hat{\mathbf{V}}_{\hat{\beta}}$ or $\hat{\mathbf{V}}_\beta$, we can perform a Wald test (why can we not use a t-test?) with test statistic

$$W = (\mathbf{R}'\hat{\beta})'(\mathbf{R}'\hat{\mathbf{V}}_{\hat{\beta}}\mathbf{R})^{-1}(\mathbf{R}'\hat{\beta}) = \sqrt{n}(\mathbf{R}'\hat{\beta})'(\mathbf{R}'\hat{\mathbf{V}}_\beta\mathbf{R})^{-1}(\mathbf{R}'\hat{\beta})$$

which asymptotically follows a χ_k^2 distribution. With this information we can calculate the p-value $p = 1 - G(W)$ and decide to reject the hypothesis if $p < \alpha$.

Exercise 3

(Ch. 23, ex. 23.2) Take the model $y(\lambda) = \beta_0 + \beta_1 X + e$ with $\mathbb{E}[e|X] = 0$ where $y(\lambda)$ is the Box-Cox transformation of y. Is this a nonlinear regression model in the parameters $(\lambda, \beta_0, \beta_1)$? (Careful this is tricky.)

Solution

The function is non-linear in the parameters $(\lambda, \beta_0, \beta_1)$. However, we normally require nonlinear regression models to be of the form:

$$y = m(\mathbf{X}, \theta) + e$$

This has an additive error term. Note that for $\lambda \neq 0$ we can write:

$$y = (\lambda(\beta_0 + \beta_1 X + e) + 1)^{\frac{1}{\lambda}}.$$

So while the given model is non-linear in the parameters it is not a **nonlinear regression model**.

Exercise 4

(Ch. 23, ex. 23.10) In Exercise 9.26, you estimated a cost function on a cross-section of electric companies. Consider the nonlinear specification

$$\log \mathrm{TC} = \beta_1 + \beta_2 \log \mathrm{Q} + \beta_3 (\log \mathrm{PL} + \log \mathrm{PK} + \log \mathrm{PF}) + \beta_4 \frac{\log \mathrm{Q}}{1 + \exp(-(\log \mathrm{Q} - \gamma))} + e.$$

This model is called a smooth threshold model. For values of $\log \mathrm{Q}$ much below γ, the variable $\log \mathrm{Q}$ has a regression slope of β_2. For values much above β_7, the regression slope is $\beta_2 + \beta_4$. The model imposes a smooth transition between these regimes.

(a) The model works best when γ is selected so that several values (in this example, at least 10 to 15) of $\log \mathrm{Q}$ are both below and above γ. Examine the data and pick an appropriate range for γ.

(b) Estimate the model by NLLS using a global numerical search over $(\beta_1, \beta_2, \beta_3, \beta_4, \gamma)$.

(c) Estimate the model by NLLS using a concentrated numerical search over γ. Do you obtain the same results?

(d) Calculate the standard errors for all the parameter estimates $(\beta_1, \beta_2, \beta_3, \beta_4, \gamma)$.

Solution

R solution at the end.

Lecture 11 Instrumental Variables

Exercise 5

(Ch 12, ex. 12.1) Consider the single equation model $y = Z\beta + e$ where y and Z are both real valued ((1×1)). Let $\hat{\beta}$ denote the IV estimator of β using as an instrument a dummy variable D (takes only values 0 and 1). Find a simple expression for the IV estimator in this context.

Solution

There is a short answer to this question which is just using the route of the usual IV estimator and writing:

$$
\begin{aligned}
\mathbb{E}\left[De\right] = 0 & \\
&= \mathbb{E}\left[D(Y - Z\beta)\right] \\
&\implies \mathbb{E}\left[DY\right] = \mathbb{E}\left[DZ\right]\beta \\
&\implies \beta = \mathbb{E}\left[DZ\right]^{-1}\mathbb{E}\left[DY\right] \\
&= \frac{\mathbb{E}\left[Y \mid D = 1\right]}{\mathbb{E}\left[Z \mid D = 1\right]}
\end{aligned}
$$

The last step follows from the fact that $\beta \in \mathbb{R}$, in this example.

For the longer answer note that D is an instrumental variable and Hanson's definition of an instrumental variable implies $\mathbb{E}\left[De\right] = 0$ (he says that this means D and e are not correlated). In this example, without an intercept, this is an even stronger assumption. With no intercept we can generally not assume $E[e] = 0$. This means that even independence, which is stronger than uncorrelatedness, does not imply $\mathbb{E}\left[De\right] = 0$. It just implies $\mathbb{E}\left[De\right] = \mathbb{E}\left[D\right]\mathbb{E}\left[e\right]$. So let's just assume both $\mathbb{E}\left[De\right] = 0$ and $\mathbb{E}\left[e\right] = 0$ [2]. Now given that D is binary we can show:

$$
\begin{aligned}
\mathbb{E}\left[De\right] &= P(D = 0)\mathbb{E}\left[De \mid D = 0\right] + P(D = 1)\mathbb{E}\left[De \mid D = 1\right] \\
&= P(D = 1)\mathbb{E}\left[e \mid D = 1\right] \\
&= \mathbb{E}\left[e \mid D = 1\right] = 0, \text{ assuming no degenerate cases}
\end{aligned}
$$

Additionally we can show that:

$$
\begin{aligned}
\mathbb{E}\left[e\right] &= \mathbb{E}\left[\mathbb{E}\left[e \mid D\right]\right] \\
&= P(D = 1)\mathbb{E}\left[e \mid D = 1\right] + P(D = 0)\mathbb{E}\left[e \mid D = 0\right] = 0
\end{aligned}
$$

As we have shown above, the first term in the last equation is zero. This implies that the second term has to be too. So we have shown that in this case $\mathbb{E}\left[De\right] = 0$ and $\mathbb{E}\left[e\right] = 0$ imply $\mathbb{E}\left[e \mid D\right] = 0$.

Now given this information there is two straightforward ways to arrive at an estimator. First, we can take conditional expectations of the structural equation with $D = 0$ and $D = 1$ and then subtract and divide to arrive at the Wald estimator:

[2] Do note however, that imposing $\mathbb{E}\left[e\right] = 0$ is not particularly restrictive, as we can always demean Y which also centers the error term around zero.

$$\beta = \frac{\mathbb{E}\left[Y|\, D = 1\right] - \mathbb{E}\left[Y|\, D = 0\right]}{\mathbb{E}\left[Z|\, D = 1\right] - \mathbb{E}\left[Z|\, D = 0\right]}$$

Note that, in the case of this example, we can already rewrite the structural equations from above as

$$\mathbb{E}[y|D = 1] = \mathbb{E}[Z|D = 1]\beta \implies \beta = \frac{\mathbb{E}\left[Y|\, D = 1\right]}{\mathbb{E}\left[Z|\, D = 1\right]}$$
$$\mathbb{E}[y|D = 0] = \mathbb{E}[Z|D = 0]\beta \implies \beta = \frac{\mathbb{E}\left[Y|\, D = 0\right]}{\mathbb{E}\left[Z|\, D = 0\right]}$$

So we have two expressions that identify β In either case we can estimate this by plugging the sample moments in for population moments, e.g. :

$$\hat{\beta}_{IV} = \frac{\frac{1}{n_1}\sum_i^n d_i y_i}{\frac{1}{n_1}\sum_i^n d_i z_i} = (\mathbf{D'Z})^{\mathbf{-1}}(\mathbf{D'Y})$$

Exercise 6

(Ch. 12, ex. 12.3) Take the linear model $y = X'\beta + e$. Let the estimator for β be $\hat{\beta}$ with OLS residual $\hat{e}_i$. Let the IV estimator for β using some instrument Z be $\tilde{\beta}$ with IV residual $\tilde{e}_i = y_i - X_i\tilde{\beta}$. If X_i is indeed endogenous, will IV fit better than OLS in the sense that $\sum_i \tilde{e}_i^2 \leq \sum_i \hat{e}_i^2$ at least in large samples?

Solution

We have the population model

$$\begin{aligned}
y &= X'\beta + e \\
\beta^* &= \left(\mathbb{E}\left[XX'\right]\right)^{-1}\left(\mathbb{E}\left[XY\right]\right) \\
e^* &= y - X'\beta^* = y - X'\left(\mathbb{E}\left[XX'\right]\right)^{-1}\left(\mathbb{E}\left[XY\right]\right)
\end{aligned}$$

where $*$ indicates that something is defined by projection. This means given endogeneity $\beta^* \neq \beta$.

Using an IV estimator we can write:

$$\begin{aligned}
\beta^{IV} &= (\mathbb{E}[ZX'])^{-1}(\mathbb{E}[ZY]) \\
e^{IV} &= y - X'\beta^{IV} = y - X'(\mathbb{E}[ZX'])^{-1}(\mathbb{E}[ZY])
\end{aligned}$$

The sample analogues of these equations are:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta + \mathbf{e} \\
\widehat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) \\
\widehat{\mathbf{e}} &= \mathbf{Y} - \mathbf{X}\widehat{\beta} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\mathbf{X}'\mathbf{Y} \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\mathbf{X}')\mathbf{Y} \\
&= (\mathbf{I} - \mathbf{P})\mathbf{Y} \\
&= \mathbf{M}\mathbf{Y} = \mathbf{M}(\mathbf{X}\beta + \mathbf{e}) \\
&= \mathbf{M}\mathbf{e}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\beta} &= (\mathbf{Z}'\mathbf{X})^{\mathbf{-1}}\mathbf{Z}'\mathbf{Y} \\
\tilde{\mathbf{e}} &= \mathbf{Y} - \mathbf{X}\tilde{\beta} = \mathbf{Y} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{\mathbf{-1}}\mathbf{Z}'\mathbf{Y} \\
&= \mathbf{X}\beta + \mathbf{e} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{\mathbf{-1}}(\mathbf{X}\beta + \mathbf{e}) \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{\mathbf{-1}}\mathbf{Z}'(\mathbf{X}\beta + \mathbf{e})) \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{\mathbf{-1}}\mathbf{Z}')\mathbf{e}
\end{aligned}$$

We can square these expressions:

$$\begin{aligned}
\widehat{\mathbf{e}}'\widehat{\mathbf{e}} &= (\mathbf{M}\mathbf{e})'\mathbf{M}\mathbf{e} \\
&= \mathbf{e}'\mathbf{M}'\mathbf{M}\mathbf{e} \\
&= \mathbf{e}'\mathbf{M}\mathbf{e} \\
&= \mathbf{e}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\mathbf{X}')\mathbf{e} \\
&= \mathbf{e}'\mathbf{e} - \mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\mathbf{X}'\mathbf{e} \\
\implies \frac{1}{n}\widehat{\mathbf{e}}'\widehat{\mathbf{e}} &= \frac{1}{n}\mathbf{e}'\mathbf{e} - \frac{1}{n}\mathbf{e}'\mathbf{X} * (\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\frac{1}{n}\mathbf{X}'\mathbf{e}
\end{aligned}$$

Note that the $\mathbf{e}$s are the structural errors. Now we can use our tools from asymptotic theory. Specifically, by the WLLN we have the following:

$$\begin{aligned}
\frac{1}{n}\mathbf{e}'\mathbf{e} &\to^P \mathbb{E}[e^2] \\
\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} &\to^P \mathbb{E}[XX']^{-1} \\
\frac{1}{n}\mathbf{X}'\mathbf{e} &\to^P \mathbb{E}[Xe] \\
\frac{1}{n}\mathbf{e}'\mathbf{X} &\to^P \mathbb{E}[Xe]'
\end{aligned}$$

By the CMT we can combine them all and write:

$$\frac{1}{n}\widehat{\mathbf{e}}'\widehat{\mathbf{e}} \to^P \mathbb{E}[e^2] - \mathbb{E}[Xe]'\mathbb{E}[XX']^{-1}\mathbb{E}[Xe]$$

Given potential endogeneity we don't generally have $\mathbb{E}[Xe] = 0$. So this is as much as we can reduce the expression.

Similarly for the IV error we can write:

$$\begin{aligned}
\tilde{\mathbf{e}}'\tilde{\mathbf{e}} &= (\mathbf{e}' - \mathbf{e}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}\mathbf{X}')(\mathbf{e} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{e}) \\
&= \mathbf{e}'\mathbf{e} - \mathbf{e}'\mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{e} - \mathbf{e}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}\mathbf{X}'(\mathbf{e} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{e}) \\
\implies \frac{1}{n}\tilde{\mathbf{e}}'\tilde{\mathbf{e}} &= \frac{1}{n}\mathbf{e}'\mathbf{e} - \frac{1}{n}(e'X(Z'X)^{-1}Z'e - e'Z(X'Z)^{-1}X'(e - X(Z'X)^{-1}Z'e))
\end{aligned}$$

Now given the exogeneity assumption of the instrument we have

$$\begin{aligned}
\mathbf{Z}'\mathbf{e} &\to^P \mathbb{E}[Ze] = 0 \\
\implies \frac{1}{n}\tilde{\mathbf{e}}'\tilde{\mathbf{e}} &\to^P \mathbb{E}[e^2]
\end{aligned}$$

Now we can return to the original question:

$$\begin{aligned}
\sum_i \tilde{e}_i^2 &\overset{?}{=} \sum_i \hat{e}_i^2 \\
\frac{1}{n}\sum_i \tilde{e}_i^2 &\overset{?}{=} \frac{1}{n}\sum_i \hat{e}_i^2 \\
\to^P \mathbb{E}[e^2] &\overset{?}{=} \mathbb{E}[e^2] - \underbrace{\mathbb{E}[Xe]'\mathbb{E}[XX']^{-1}\mathbb{E}[Xe]}_{\geq 0} \\
\implies \mathbb{E}[e^2] &\geq \mathbb{E}[e^2] - \mathbb{E}[Xe]'\mathbb{E}[XX']^{-1}\mathbb{E}[Xe]
\end{aligned}$$

So the IV residuals are asymptotically larger than the residuals of the related OLS estimator. Why is that the case? Because using OLS we use all the variation at hand, while with IV we restrict us to the "exogenous" variation. Also note that if X was not endogenous and we would have $\mathbb{E}[Xe] = 0$, then asymptotically the two terms would be the same.

Exercise 7

(Ch. 12, ex. 12.10) Consider the model

$$\begin{aligned}
y &= X'\beta + e \\
X &= \Gamma Z + u,\ \mathbb{E}[Ze] = 0,\ \mathbb{E}[Zu] = 0
\end{aligned}$$

with y scalar and X and Z each a k vector. You have a random sample $(y_i, X_i, Z_i :\ i = 1, ..., n)$. Take the control function equation $e = u'\gamma + \nu$ with $\mathbf{e}[u\nu] = 0$ and assume for simplicity that u is observed. Inserting into the structural equation we find $y = X'\beta + u'\gamma + \nu$. The control function estimator $(\hat{\beta}, \hat{\gamma})$ is the OLS estimation of this equation.

(a) Show that $\mathbb{E}[X\nu] = 0$ (algebraically).

(b) Derive the asymptotic distribution of $(\hat{\beta}, \hat{\gamma})$.

Solution

(a) There is not all to much to explain, let's just plug one equation into the other and see what we can take out:

$$\begin{aligned}
\mathbb{E}[X\nu] &= \mathbb{E}[(\Gamma' Z + u)\nu] \\
&= \mathbb{E}[(\Gamma' Z + u)(e - u'\gamma)] \\
&= \Gamma' \underbrace{\mathbb{E}[Ze]}_{=0} - \Gamma' \underbrace{\mathbb{E}[Zu']}_{=0} \gamma + \mathbb{E}[ue] - \mathbb{E}[uu']\gamma \\
&= \mathbb{E}[ue] - \mathbb{E}[uu']\mathbb{E}[uu']^{-1}\mathbb{E}[ue] \\
&= \mathbb{E}[ue] - \mathbb{E}[ue] = 0
\end{aligned}$$

(b) Let $\theta' = (\beta' \ \gamma')$ and $\tilde{X}' = (X' \ u')$. u in this example is observed, so we can rewrite the structural equation as

$$y = \tilde{X}'\theta + \nu$$

and estimate θ by OLS since $\mathbb{E}\left[X\nu\right] = 0$ as shown above. Doing exactly the same derivation as last week, we have that

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N(\mathbf{0}, \mathbf{V}_\theta)$$

as $n \to \infty$. The asymptotic variance is given by

$$\mathbf{V}_\theta = (\mathbb{E}[\tilde{X}\tilde{X}'])^{-1}\mathbf{\Omega}(\mathbb{E}[\tilde{X}\tilde{X}'])^{-1},$$

where $\mathbf{\Omega} = \mathbb{E}[\tilde{X}\tilde{X}'e^2]$

Exercise 8

(Ch. 12, ex. 12.11) Consider the structural equation

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e \tag{*}$$

$X \in \mathbb{R}$ treated as exogenous so that $\mathbb{E}[\mathbf{X}e] \neq 0$. We have an instrument $Z \in \mathbb{R}$ which satisfies $\mathbb{E}[e|Z] = 0$ so in particular $\mathbb{E}[e] = 0, \mathbb{E}[Ze] = 0$ and $\mathbb{E}[Z^2 e] = 0$.

(a) Should X^2 be treated as endogenous or exogenous?

(b) Suppose we have a scalar instrument Z which satisfies

$$X = \gamma_0 + \gamma_1 Z + u \qquad (+)$$

with u independent of Z and mean zero.

Consider using $(1, Z, Z^2)$ as instruments. Is this a sufficient number of instruments? Is the model (*) just-identified, over-identified or under-identified?

(c) Write out the reduced form equation for X^2. Under what condition on the reduced form parameters in (+) are the parameters in (*) identified?

Solution

(a) First note that by the LIE $\mathbb{E}[Xe] = \mathbb{E}[X\mathbb{E}[e|X]]$. Since $\mathbb{E}[Xe] \neq 0$ $\mathbb{E}[e|X] \neq 0$ as well. We can use the conditioning theorem and write

$$\mathbb{E}[X^2 e] = \mathbb{E}[f(X)e] = \mathbb{E}[f(X)\mathbb{E}[e|X]] \neq 0$$

generally. Hence, X^2 is generally endogenous.

(b) We have two endogenous regressors X and X^2 and two external instruments Z and Z^2. Also, the intercept is an exogenous regressor. Therefore, the model is just-identified.

(c) Let's write the reduced form for all regressors:

$$\begin{aligned}
1 &= 1 \\
X &= \gamma_0 + \gamma_1 Z + \gamma_2 Z^2 + u \\
X^2 &= \delta_0 + \delta_1 Z + \delta_2 Z^2 + \epsilon \\
\overset{\text{mv reg.}}{\Longrightarrow} \bar{X} &= \Gamma' \bar{Z} + \nu \\
&= \begin{pmatrix} 1 & 0 & 0 \\ \gamma_0 & \gamma_1 & \gamma_2 \\ \delta_0 & \delta_1 & \delta_2 \end{pmatrix} \begin{pmatrix} 1 \\ Z \\ Z^2 \end{pmatrix} + \begin{pmatrix} 0 \\ u \\ \epsilon \end{pmatrix} \\
\Longrightarrow \Gamma &= \mathbb{E}\left[\bar{Z}\bar{Z}'\right]^{-1} \mathbb{E}\left[\bar{Z}\bar{X}'\right]
\end{aligned}$$

Under the invertibility assumption (i.e $\mathbb{E}\left[\bar{Z}\bar{Z}'\right] > 0$), we have that Γ is identified. But we are asked about identification of β. The reduced form of the LHS variable Y

is:

$$
\begin{aligned}
Y &= \eta_0 + \eta_1 Z + \eta_2 Z^2 + \varepsilon = \bar{Z}'\eta + \varepsilon \\
\implies \eta &= \mathbb{E}\left[\bar{Z}\bar{Z}'\right]^{-1} \mathbb{E}\left[\bar{Z}Y\right]
\end{aligned}
$$

Note that we can also arrive at this equation if we start from the structural equation and plug the reduced form of endogenous regressors in:

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X + \beta_2 X^2 + e \\
Y &= \beta_0 + \beta_1(\gamma_0 + \gamma_1 Z + \gamma_2 Z^2 + u) + \beta_2(\delta_0 + \delta_1 Z + \delta_2 Z^2 + \epsilon) + e \\
&= \underbrace{\beta_0 + \beta_1\gamma_0 + \beta_2\delta_0}_{\eta_0} + \underbrace{(\beta_1\gamma_1 + \beta_2\delta_1)}_{=\eta_1} Z + \underbrace{(\beta_1\gamma_2 + \beta_2\delta_2)}_{=\eta_2} Z^2 + \underbrace{\beta_1 u + \beta_2\epsilon + e}_{=\varepsilon}
\end{aligned}
$$

It turns out that $\eta = \Gamma\beta$ (mulitply it out to see). So

$$
\begin{aligned}
\mathbb{E}\left[\bar{Z}\bar{Z}'\right]^{-1} \mathbb{E}\left[\bar{Z}Y\right] &= \mathbb{E}\left[\bar{Z}\bar{Z}'\right]^{-1} \mathbb{E}\left[\bar{Z}\bar{X}'\right]\beta \\
\mathbb{E}\left[\bar{Z}Y\right] &= \mathbb{E}\left[\bar{Z}\bar{X}'\right]\beta \\
\beta &= \mathbb{E}\left[\bar{Z}\bar{X}'\right]^{-1} \mathbb{E}\left[\bar{Z}Y\right]
\end{aligned}
$$

So β only has a unique solution if $\text{rank}(\mathbb{E}\left[\bar{Z}\bar{X}'\right]) = k$, given that in our example this is a square matrix full rank implies invertibility. This is basically the relevance condition. I.e. we need Z not independent of X and Z^2 not independent of X^2. Which follows from inspecting the matrix:

$$
\begin{aligned}
\mathbb{E}\left[\bar{Z}\bar{X}'\right] &= \mathbb{E}\left[\begin{pmatrix} 1 \\ Z \\ Z^2 \end{pmatrix} \begin{pmatrix} 1 & X & X^2 \end{pmatrix}\right] \\
&= \begin{pmatrix} 1 & \mathbb{E}[X] & \mathbb{E}[X^2] \\ \mathbb{E}[Z] & \mathbb{E}[ZX] & \mathbb{E}[ZX^2] \\ \mathbb{E}[Z^2] & \mathbb{E}[Z^2X] & \mathbb{E}[Z^2X^2] \end{pmatrix}
\end{aligned}
$$

If the variables were independent we could just multiply column 1 with $\mathbb{E}[X]$ (a constant) to get column 2. This would imply rank deficiency.

Exercise 9

(Ch. 12, ex. 12.22) You will replicate and extend the work reported in Acemoglu, Johnson, and Robinson (2001). The authors provided an expanded set of controls when they published their 2012 extension and posted the data on the AER website.This dataset is AJR2001on the textbook website.

(a) Estimate the OLS regression (12.86), the reduced form regression(12.87), and the 2SLS regression (12.88). (Which point estimate is different by 0.01 from the reported values? This is a common phenomenon in empirical replication).

(b) For the above estimates calculate both homoskedastic and heteroskedastic-robust standard errors. Which were used by the authors (as re-ported in (12.86)-(12.87)-(12.88)?)

(c) Calculate the 2SLS estimates by the Indirect Least Squares formula. Are they the same?

(d) Calculate the 2SLS estimates by the two-stage approach. Are they the same?

(e) Calculate the 2SLS estimates by the control variable approach. Are they the same?

(f) Acemoglu, Johnson, and Robinson (2001) reported many specifications including alternative regressor controls, for example *latitude* and *africa*. Estimate by least squares the equation for *log(GDP)* adding *latitude* and *africa* as regressors. Does this regression suggest that *latitude* and *africa* are predictive of the level of GDP?

(g) Now estimate the same equation as in (f) but by 2SLS using *log(mortality)* as an instrument for risk. How does the interpretation of the effect on *latitude* and *africa* change?

(h) Return to our baseline model (without including *latitude* and *africa*). the authors' reduced from equation uses *log(mortality)* as the instrument, rather than, say, the level of mortality. Estimate the reduced from for risk with mortality as the instrument. (This variable is not provided in the daatset so you need to take the exponential of *log(mortality)*.) Can you explain why the authors preferred the equation with *log(mortality)*?

(i) Try an alternative reduced form including both *log(mortality)* and the square of *log(mortality)*. Interpret the results. Re-estimate the structural equation by 2SLS using both *log(mortality)* and its square as instruments. How do the results change?

(j) Calculate and interpret a test for exogeneity of the instruments.

(k) Estimate the equation by LIML using the instruments *log(mortality)* and the square of *log(mortality)*.

Solution

R solution at the end.

Metrics1_PS6

Artschil Okropiridse

Housekeeping

```r
rm(list=ls())

library(readr)
library(tidyverse)
library(knitr)
library(estimatr)
library(sandwich)
library(matrixStats)
library(ivreg)
library(ivmodel)
```

Question 4

```r
# Load data
data <- read.table("Data/Nerlove1963.txt",header=TRUE)

# Prepare data
data <- data |>
  mutate(lC = log(Cost)) |>
  mutate(lQ = log(output)) |>
  mutate(lPL = log(Plabor))  |>
  mutate(lPK = log(Pcapital)) |>
  mutate(lPF = log(Pfuel)) |>
  mutate(l_sum = lPF + lPK + lPL) |>
  tibble()
```

```r
y <- matrix(data$lC,ncol=1)
x <- as.matrix(cbind(matrix(1,nrow(matrix(data$lQ,ncol=1)),1),data$lQ,data$l_sum))
```

(a)

We are going to pick bounds for γ such that exactly 15 observations are below the lower bound and 15 observations are above the upper bound.

```r
# Derive bounds for gamma
gamma_lb <- sort(data$lQ, decreasing = F)[16]
gamma_ub <- sort(data$lQ, decreasing = T)[16]

gamma_lb
```

```
[1] 4.143135
```

```r
gamma_ub
```

```
[1] 8.644883
```

(b)

We want to minimize the objective function $S(\theta) = \frac{1}{n}\sum_i (Y_i - m(X_i, \theta))^2$.

```r
# Use median lQ as starting point for gamma
gamma_med <- median(data$lQ)

# Use NLS function
nls <- nls(lC ~ a + b*lQ + c*l_sum + d*(lQ/(1+exp(-(lQ-gamma)))), # Model
    data = data,                                                  # Data
    start = list(a = 1, b= 1, c= 1, d = 1, gamma = gamma_med),    # Starting values
    lower = c(-Inf, -Inf, -Inf, -Inf, gamma_lb),                  # Restrict grid for gamma
    upper = c(Inf, Inf, Inf, Inf, gamma_ub),
    algorithm = 'port')

summary(nls)
```

```
Formula: lC ~ a + b * lQ + c * l_sum + d * (lQ/(1 + exp(-(lQ - gamma))))

Parameters:
      Estimate Std. Error t value Pr(>|t|)
a     -5.32109    0.59774  -8.902 2.56e-15 ***
b      0.43779    0.04788   9.143 6.35e-16 ***
c      0.37074    0.06244   5.938 2.17e-08 ***
d      0.22401    0.02944   7.609 3.68e-12 ***
gamma  6.87514    0.32087  21.427  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3193 on 140 degrees of freedom

Algorithm "port", convergence message: relative convergence (4)
```

(c)

The model is linear in the beta coefficients. This implies that we can compute an estimate for beta conditional on γ using the standard OLS formula $\hat{\beta}(\gamma) = (X(\gamma)'X(\gamma))^{-1} X(\gamma))'y$.

We include the non-linear part in the initial model into our regressor matrix. Since the residual $\hat{e} = y - X(\gamma)'\hat{\beta}(\gamma)$ is a non-linear function of γ we can apply the NLS estimator to the objective function

$$S(\gamma) = \frac{1}{n} \sum_i \left(Y_i - X_i(\gamma)'\hat{\beta}(\gamma) \right)^2$$

to obtain an estimate $\hat{\gamma}$ for γ. Finally, we can then use this estimate to get an estimate $\bar{\beta} = \hat{\beta}(\hat{\gamma})$ of β.

```
# Regressor matrix conditional on gamma
x.gam <- function(gamma){
 non.lin <- (data$lQ/(1+exp(gamma - data$lQ)))
 X <- cbind(x,non.lin)
 return(X)
}

# OLS estimates of beta1 - beta4 conditional on gamma
beta.gamma <- function(gamma){
  beta.hat <- solve(t(x.gam(gamma))%*%x.gam(gamma))%*%t(x.gam(gamma))%*%y
```

```r
  return(beta.hat)
}

# Objective function conditional on gamma
S.gamma <- function(gamma){
  1/nrow(x.gam(gamma)) * sum((data$lC - x.gam(gamma)%*%beta.gamma(gamma))^2)
}
gamma.hat <- optimize(f=S.gamma,
                lower = gamma_lb,
                upper = gamma_ub)$minimum

coefs <- rbind(beta.gamma(gamma.hat), gamma.hat)
coefs
```

```
                [,1]
          -5.3210907
           0.4377924
           0.3707373
non.lin    0.2240123
gamma.hat  6.8751290
```

d)

See results in (a) for the respective standard errors. We can derive the standard errors of the model in (c) by estimating

$$\hat{\mathbf{V}} = \hat{\mathbf{Q}}^{-1}\hat{}\hat{\mathbf{Q}}^{-1} = \left(\frac{1}{n}\sum_i \hat{m}_i\hat{m}_i'\right)^{-1}\left(\frac{1}{n}\sum_i \hat{m}_i\hat{m}_i'\hat{e}_i\right)\left(\frac{1}{n}\sum_i \hat{m}_i\hat{m}_i'\right)^{-1}$$

where

$$\hat{m}_i = \frac{\partial}{\partial\theta}m(X_i,\hat{\theta}) = \begin{pmatrix} 1 \\ \log \mathrm{Q}_i \\ \log \mathrm{PL}_i + \log \mathrm{PK}_i + \log \mathrm{PF}_i \\ \frac{\log \mathrm{Q}_i}{1+\exp(\hat{\gamma}-\log \mathrm{Q}_i)} \\ \frac{\hat{\beta}_4 \log \mathrm{Q}_i \exp(\hat{\gamma}-\log \mathrm{Q}_i)}{(1+\exp(\hat{\gamma}-\log \mathrm{Q}_i))^2} \end{pmatrix}$$

```r
n <- nrow(matrix(data$lQ,ncol=1))

e.hat <- data$lC - x.gam(gamma.hat)%*%beta.gamma(gamma.hat)
```

```r
mhat <- cbind(matrix(1,n,1),
              data$lQ,
              data$l_sum,
              data$lQ/(1+exp(gamma.hat-data$lQ)),
              (data$lQ*coefs[4]*exp(gamma.hat-data$lQ))/(1+exp(gamma.hat-data$lQ))^2
              )

Q.hat.inv <- solve((t(mhat)%*%mhat)/n)
Omega.hat <- (t(mhat)%*%diag(as.vector(e.hat^2), nrow = n)%*%mhat)/n
V.hat <- Q.hat.inv%*%Omega.hat%*%Q.hat.inv

se <- sqrt(diag(V.hat)/n)
se
```

```
[1] 0.49378280 0.09957152 0.04414803 0.05455272 0.35864877
```

Question 9

```r
data <- read.delim("Data/AJR2001.txt")
```

(a)

```r
ols_a <- lm(loggdp ~ risk, data = data)
fs_a <- lm(risk ~ logmort0, data)
iv_a <- ivreg(loggdp ~ risk | logmort0, data=data)

summary(ols_a)
```

```
Call:
lm(formula = loggdp ~ risk, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8649 -0.4604  0.1741  0.4682  1.1608

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.68741    0.41744  11.229  < 2e-16 ***
risk         0.51619    0.06252   8.257 1.42e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7289 on 62 degrees of freedom
Multiple R-squared:  0.5237,    Adjusted R-squared:  0.516
F-statistic: 68.17 on 1 and 62 DF,  p-value: 1.421e-11
```

```r
summary(fs_a)
```

```
Call:
lm(formula = risk ~ logmort0, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6507 -0.9889  0.0324  0.8498  3.3768

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.3659     0.6106  15.339  < 2e-16 ***
logmort0     -0.6133     0.1269  -4.831 9.27e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.262 on 62 degrees of freedom
Multiple R-squared:  0.2735,    Adjusted R-squared:  0.2618
F-statistic: 23.34 on 1 and 62 DF,  p-value: 9.273e-06
```

```r
summary(iv_a)
```

```
Call:
ivreg(formula = loggdp ~ risk | logmort0, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.41118 -0.54771  0.08268  0.69741  1.68871
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.9943     1.0240   1.947    0.056 .
risk          0.9295     0.1561   5.955 1.33e-07 ***

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   1  62     23.34 9.27e-06 ***
Wu-Hausman         1  61     22.04 1.56e-05 ***
Sargan             0  NA        NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9517 on 62 degrees of freedom
Multiple R-Squared: 0.188,  Adjusted R-squared: 0.1749
Wald test: 35.46 on 1 and 62 DF,  p-value: 1.327e-07
```

We can see that our 2sls estimate is slightly different.

(b)

```
se.ols.hm <- sqrt(diag(vcovHC(ols_a, type='const')))
se.ols.HC1 <- sqrt(diag(vcovHC(ols_a, type='HC1')))

se.fs.hm <- sqrt(diag(vcovHC(fs_a, type='const')))
se.fs.HC1 <- sqrt(diag(vcovHC(fs_a, type='HC1')))

se.iv.hm <- sqrt(diag(vcovHC(iv_a, type='const')))
se.iv.HC1 <- sqrt(diag(vcovHC(iv_a, type='HC1')))

se.ols.hm
```

```
(Intercept)        risk
 0.41744102  0.06251859
```

```
se.ols.HC1
```

```
(Intercept)        risk
 0.32447413  0.05110098
```

```r
se.fs.hm
```

```
(Intercept)    logmort0
  0.6105941   0.1269412
```

```r
se.fs.HC1
```

```
(Intercept)    logmort0
  0.7084137   0.1517849
```

```r
se.iv.hm
```

```
(Intercept)        risk
  1.0240313   0.1560901
```

```r
se.iv.HC1
```

```
(Intercept)        risk
  1.1531461   0.1728088
```

It seems like the authors used homoskedastic SEs.

(c)-(d)

```r
# Indirect IV
# We divide the reduced form by the first stage
fs <- lm(risk ~ logmort0, data=data)$coefficients[2]
rf <- lm(loggdp ~ logmort0, data=data)$coefficients[2]
ils <- rf/fs
ils
```

```
 logmort0
0.9294897
```

```r
# Two-stage approach

# Predict risk from first stage
data$risk.hat <- fs_a$fitted.values
tsa <- lm(loggdp ~ risk.hat, data=data)$coefficients[2]
tsa
```

```
 risk.hat
0.9294897
```

```r
# Control variable approach

# Use residuals from first stage as additional regressor
data$u.hat <- fs_a$residuals
cfa <- lm(loggdp ~ risk + u.hat, data=data)$coefficients[2]
cfa
```

```
     risk
0.9294897
```

As we can see all three approaches yield the same result.

f)

```r
# OLS
ols.f <- lm(loggdp ~ risk + africa + latitude, data=data)
summary(ols.f)
```

```

Call:
lm(formula = loggdp ~ risk + africa + latitude, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.74326 -0.25591  0.04807  0.42696  1.01128 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  5.65223    0.41529  13.610  < 2e-16 ***
risk          0.37652    0.06084   6.189 5.94e-08 ***
africa       -0.72327    0.17130  -4.222 8.33e-05 ***
latitude      1.38246    0.64404   2.147   0.0359 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6226 on 60 degrees of freedom
Multiple R-squared:  0.6637,    Adjusted R-squared:  0.6469
F-statistic: 39.47 on 3 and 60 DF,  p-value: 3.254e-14
```

The p-values indicate that both latitude and africa are predictive of log GDP.

g)

```r
# 2SLS
iv.f <- ivreg(loggdp ~ risk + africa + latitude| logmort0 + africa + latitude, data=data)
summary(iv.f)
```

```
Call:
ivreg(formula = loggdp ~ risk + africa + latitude | logmort0 +
    africa + latitude, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.26852 -0.46397 -0.02031  0.64671  1.47220

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.99507    1.63339   1.834  0.07167 .
risk          0.79997    0.25793   3.101  0.00293 **
africa       -0.34793    0.31630  -1.100  0.27573
latitude     -0.05531    1.19980  -0.046  0.96338

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   1  60     6.708  0.0120 *
Wu-Hausman         1  59     5.854  0.0186 *
Sargan             0  NA        NA      NA
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.837 on 60 degrees of freedom
Multiple R-Squared: 0.3922, Adjusted R-squared: 0.3618
Wald test: 17.98 on 3 and 60 DF,  p-value: 1.926e-08
```

Unlike in the OLS case, africa and latitude are not predictive of log GDP anymore.

h)

```r
# Mortality
data$mortality <- exp(data$logmort0)
# First stage
fs.h <- lm(risk ~ mortality, data=data)
summary(fs.h)
```

```
Call:
lm(formula = risk ~ mortality, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0207 -0.9024  0.0026  0.8095  3.3024 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.7094189  0.2021894  33.184   <2e-16 ***
mortality   -0.0007862  0.0003819  -2.059   0.0437 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.432 on 62 degrees of freedom
Multiple R-squared:  0.06399,	Adjusted R-squared:  0.04889 
F-statistic: 4.239 on 1 and 62 DF,  p-value: 0.04372
```

The effect of the level of mortality on risk is not significant. However, we need a strong first stage to get reliable estimates since the 2SLS estimator is biased if the first stage is small relative to the sample size.

i-j)

```r
# Mortality
data$logm2 <- data$logmort0^2

# First stage
fs.i <- lm(risk ~ logmort0 + logm2, data=data)
summary(fs.i)
```

```
Call:
lm(formula = risk ~ logmort0 + logm2, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.31884 -0.99818  0.08128  0.85518  2.44086 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.94859    1.55170   8.989 8.92e-13 ***
logmort0    -2.64568    0.65090  -4.065  0.00014 ***
logm2        0.21011    0.06617   3.176  0.00235 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.179 on 61 degrees of freedom
Multiple R-squared:  0.3766,    Adjusted R-squared:  0.3561 
F-statistic: 18.42 on 2 and 61 DF,  p-value: 5.509e-07
```

```r
# 2SLS
iv.i <- ivreg(loggdp ~ risk| logmort0 + logm2, data=data)
summary(iv.i)
```

```
Call:
ivreg(formula = loggdp ~ risk | logmort0 + logm2, data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.13540 -0.49188  0.08412  0.55741  1.30176 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0188     0.7553   3.997 0.000173 ***
risk          0.7723     0.1148   6.725 6.43e-09 ***

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   2  61    18.423 5.51e-07 ***
Wu-Hausman         1  61    11.918  0.00102 **
Sargan             1  NA     5.135  0.02344 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8216 on 62 degrees of freedom
Multiple R-Squared: 0.3948, Adjusted R-squared: 0.3851
Wald test: 45.22 on 1 and 62 DF,  p-value: 6.427e-09
```

Including the additional instrument increases the fit of the first stage. It furthermore, reduces the main effect from ca. 0.93 to 0.77. From the output we can also see the statistic of the Sargan test which is 5.135. The corresponding p-value is 0.023 which is relatively high.

k)

```r
# LIML using the ivmodel package
ivmodel.k <- ivmodel(Y = as.numeric(data$loggdp),
                     D = as.numeric(data$risk),
                     Z = cbind(as.numeric(data$logmort0), as.numeric(data$logm2)))
liml.k <- LIML(ivmodel.k)
liml.k$point.est
```

```
      Estimate
[1,] 0.8371439
```