# ECMA 31000: Solutions to Problem Set 7

Joe Hardwick

**Question 1 (You must include exogenous regressors in the first stage)** Let $x \in \mathbb{R}^{k+1}$be a random vector with first component $x_0 = 1$. Suppose you observe an iid sample of $\{y_i, x_i, z_i\}_{i=1}^n$. Consider the model

$$y = x'\beta + u; \qquad \mathrm{E}(ux) \neq 0, \mathrm{E}(zu) = 0.$$

Assume $\mathrm{E}(zx') < \infty$ and is invertible. Suppose there is a single endogenous regressor $x_k$.

a) Suppose you estimate this model using Two Stage Least Squares. Describe the two stage procedure. Show that the estimator is equal to the Instrumental Variables estimator. Is this a consistent estimate of $\beta$?

**ANS:** We describe the general case where $\dim(z) = l + 1 \geq k + 1 = \dim(x)$. In the first stage, we regress all the covariates in $x$ on $z$. To do this, stack the observations into matrices $Y, X, Z$ and consider the regression

$$X = Z\Pi + V,$$

where $\Pi$ is an $(l+1) \times (k+1)$ matrix of parameters. The $j$-th column of $\Pi$ represents the coefficients of a regression of $X_j$ on $Z$. Running OLS on each column yields

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

and so we form the fitted values of $X$ as

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

In the second stage, we regress $Y$ on the fitted values of $X$:

$$Y = \hat{X}\beta + \epsilon.$$

We estimate $\beta$ in this regression using OLS to give the two stage least squares estimator:

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'Y \\
&= (X'P_Z'P_Z X)^{-1} X'P_Z'P_Z Y \\
&= (X'P_Z X)^{-1} X'P_Z Y.
\end{aligned}$$

where the third equality holds because $P_Z$ is symmetric and idempotent. Since we have assumed for now that $\mathrm{E}(zx')$ is invertible, it must be square, and so using the formula $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ yields

$$(X'P_ZX)^{-1} = (Z'X)^{-1} Z'Z (X'Z)^{-1},$$

so

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (Z'X)^{-1} Z'Z (X'Z)^{-1} X'P_ZY \\
&= (Z'X)^{-1} Z'Z (X'Z)^{-1} X'Z (Z'Z)^{-1} Z'Y \\
&= (Z'X)^{-1} Z'Y \\
&= \hat{\beta}_{IV}.
\end{aligned}$$

This estimator is consistent because plugging in $Y = X\beta + U$ yields

$$\hat{\beta}_{IV} = \beta + \left(\frac{Z'X}{n}\right)^{-1} \frac{Z'U}{n} \xrightarrow{p} \beta + \mathrm{E}(z_i x_i')^{-1} \mathrm{E}(z_i u_i) = \beta.$$

b) Now suppose you have more instruments than regressors. Show that the 2SLS estimator can be considered as an IV estimator after running the first stage stage regression. What are the instruments in this case? Is the 2SLS estimator consistent given the modelling assumptions?

**ANS:** In the population, a regression of $x$ on $z$ can be written as

$$x' = z'\Pi + v; \quad \mathrm{E}(zv) = 0,$$

where $\Pi$ is an $(l+1) \times (k+1)$ matrix of parameters that has the same interpretation as above. The first stage in the population results in

$$\hat{x}' = z'\mathrm{E}(zz')^{-1} \mathrm{E}(zx').$$

In this particular case, we will obtain

$$\hat{x}' = \left(x_0, x_1, \ldots, x_{k-1}, z'\mathrm{E}(zz')^{-1} \mathrm{E}(zx_k)\right).$$

In sample, the instruments are given by

$$\hat{X} = P_ZX = \begin{pmatrix} | & | & | & | & | \\ X_0 & X_1 & \cdots & X_{k-1} & P_ZX_k \\ | & | & | & | & | \end{pmatrix}.$$

The 2SLS estimator is also consistent, because

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (X'P_ZX)^{-1} X'P_ZY \\
&= \beta + (X'P_ZX)^{-1} X'P_ZU
\end{aligned}$$

$$
\begin{aligned}
&= \beta + \left( \frac{X'Z}{n} \left( \frac{Z'Z}{n} \right)^{-1} \frac{Z'X}{n} \right)^{-1} \frac{X'Z}{n} \left( \frac{Z'Z}{n} \right)^{-1} \frac{Z'U}{n} \\
&\xrightarrow{p} \beta + \left( \mathrm{E}\left(xz'\right) \mathrm{E}\left(zz'\right)^{-1} \mathrm{E}\left(zx'\right) \right) \mathrm{E}\left(xz'\right) \mathrm{E}\left(zz'\right)^{-1} \mathrm{E}\left(zu\right) = \beta.
\end{aligned}
$$

c(i) Return to the exactly identified case. Let $x_{-k} = (x_0, \ldots, x_{k-1})$ and $\gamma_{-k} = (\gamma_0, \ldots, \gamma_{-k})$. Show that the coefficient on $x_k$ in the second stage regression can be found by first running OLS on

$$
y = x'_{-k}\gamma_{-k} + \gamma_k z + \epsilon, \tag{1}
$$

to yield the OLS estimate $\hat{\gamma}_k$, and then dividing this estimate by $\hat{\pi}$, where $\hat{\pi}$ is the OLS coefficient on $z$ in the regression

$$
x_k = x'_{-k}\delta + \pi z + v.
$$

Hint: When considering the second stage regression, substitute $\hat{x}_k = x'_{-k}\hat{\delta} + \hat{\pi}z$. Now compare the OLS minimization problems of the second stage regression and of model (1).

**ANS:** In the second stage, we run the regression

$$
\begin{aligned}
Y &= \hat{X}\beta + \epsilon \\
&= X_{-k}\beta_{-k} + P_Z X_k \beta_k + \epsilon \\
&= X_{-k}\beta_{-k} + \left[ X_{-k}\hat{\delta} + Z\hat{\pi} \right] \beta_k + \epsilon \\
&= X_{-k}\left[ \beta_{-k} - \hat{\delta} \right] + Z\hat{\pi}\beta_k + \epsilon.
\end{aligned}
$$

The least squares minimization problem is

$$
\min_{b_{-k}, b_k} \sum_{i=1}^{n} \left( y_i - x'_{i,-k}\left( b_{-k} - \hat{\delta} \right) - b_k \hat{\pi} z_i \right)^2 .
$$

The least squares minimization problem when running OLS on (1) is:

$$
\min_{\gamma_{-k}, \gamma_k} \sum_{i=1}^{n} \left( y_i - x'_{i,-k}\gamma_{-k} - \gamma_k z_i \right)^2 .
$$

We can see after some rearranging that both of these regressions are really a regression of $y$ on $x_{-k}$ and $z$ and lead to the same least squares minimization problem up to parameters. Assuming a unique solution to these minimization problems (which will hold with probability approaching 1 by assumption that $\mathrm{E}\left(zx'\right)$ is full rank) yields that

$$
\begin{aligned}
\hat{\beta}_{-k} - \hat{\delta} &= \hat{\gamma}_{-k}; \\
\hat{\beta}_k \hat{\pi} &= \hat{\gamma}_k.
\end{aligned}
$$

The conclusion follows.

c(ii) Is $\hat{\gamma}_k$ a consistent estimate of $\beta_k$?

**ANS:** Generally not, because

$$\hat{\gamma}_k = \hat{\beta}_k \hat{\pi} \xrightarrow{p} \beta_k \pi \neq \beta_k,$$

unless $\pi = 1$.

d) Suppose $k = 2$, the included instruments are $(1, x_1)$ and there is a single endogenous variable $x_2$. There is a single excluded instrument $z_1$, so the vector of instrumental variables is $z = (1, x_1, z_1)$. Explain, first mathematically, and then intuitively, why the following is incorrect and actually yields an inconsistent estimate of $\beta$ when using 2 stage least squares:

"In the first stage regression, it's fine to omit exogenous regressors. Just regress the endogenous variables on the excluded instruments since this produces exogenous variation anyway." (The term "exogenous variation" means $x_k$ is exogenous after projection because it is a linear combination of excluded $z's$).

**ANS:** Mathematically, our answer to part c) can help. First split up the instruments into included and excluded: $Z = (X_{-k}, Z_k)$. If we run OLS on

$$X_k = Z_k \pi + E,$$

and omit all the included instruments, we obtain the estimate $\tilde{\pi} = \left(Z_k' Z_k\right)^{-1} Z_k' X_k$. This will suffer from omitted variables bias and lead to an inconsistent estimate of $\pi$ because

$$\begin{aligned}
\tilde{\pi} &= \left(Z_k' Z_k\right)^{-1} Z_k' \left[X_{-k}\delta + Z_k \pi + V\right] \\
&= \pi + \left(Z_k' Z_k\right)^{-1} Z_k' X_{-k}\delta + \left(Z_k' Z_k\right)^{-1} Z_k' V \\
&\xrightarrow{p} \pi + \frac{\mathrm{E}\left(z_k x_{-k}'\right)\delta}{\mathrm{E}\left(z_k^2\right)} = \pi + \nu,
\end{aligned}$$

where $\nu$ will not generally equal 0 unless $\delta = 0$, or $z_k$ is uncorrelated with all the other included instruments anyway. Our answer to part c) shows that if we run this regression instead of including all instruments the estimate of $\beta$ we obtain in the second stage will be

$$\tilde{\beta} = \frac{\hat{\gamma}_k}{\tilde{\pi}} \xrightarrow{p} \frac{\beta_k \pi}{\pi + \nu} \neq \beta_k.$$

For an intuitive explanation, remember that we are trying to establish the causal effect of a change in $x_k$ on $y$ holding the other variables fixed, and so we measure changes in $x_k$ through changes in $z$ that are <u>not</u> due to the other variables. In the ordinary linear regression model without endogeneity, failing to include relevant variables can lead to omitted variables bias because we falsely attribute the effect of the excluded variable on $y$ to the included variables. The same thing happens here, because in the first stage, we are measuring correlation between $x_k$ and $z_k$ after "controlling for" $x_{-k}$. If we fail to include these variables, the fitted value of $x_k$ will unintentionally capture the variation in $x_k$ associated with $x_{-k}$ and attribute it to $z$ alone. Therefore, when we consider the second stage, we may underestimate the ceteris paribus effect of $x_k$ if an increase in its projected value $\hat{x}_k$ amounts to increasing $z$ mostly through the other $x_{-k}$. This would no longer represent a ceteris paribus effect. For an extreme example, suppose the instrument is in fact irrelevant, so $\pi = 0$ in the regression

$$x_k = x'_{-k}\delta + \pi z + v; \quad \mathrm{E}\left(\begin{pmatrix} x_{-k} \\ z \end{pmatrix} u\right) = 0.$$

In this case, estimating

$$x_k = \pi z + \epsilon; \quad \mathrm{E}\left(z\epsilon\right) = 0$$

by OLS yields an estimate which is only nonzero through omitted variables bias. Changes in $\hat{x}_k$ are associated entirely with changes in $x_{-k}$, not $z$, but our first stage will generally fail to reflect this.

**Question 2** Let $x \in \mathbb{R}^{k+1}$be a random vector with first component $x_0 = 1$. Suppose you observe an iid sample of $\{y_i, x_i\}_{i=1}^n$. Consider the model

$$y = x'\beta + u; \qquad \mathrm{E}\left(u|x\right) = 0, Var\left(u|x\right) = \sigma^2.$$

Suppose you stack the observations and obtain

$$Y = X\beta + U; \mathrm{E}\left(U|X\right) = 0; \quad Var\left(U|X\right) = \sigma^2 I_n.$$

a) Justify each step in the following proof that $\hat{\sigma}^2 = \frac{SSR}{n-k-1}$ is an unbiased estimator of $\sigma^2$ conditional on $X$. Hint: Look up the properties of the trace operator.

First, $SSR = U'M_X U$. Using properties of the trace operator, we have

$$\begin{aligned}
\mathrm{E}\left(U'M_X U|X\right) &= \mathrm{E}\left(tr\left(U'M_X U\right)|X\right) \\
&= \mathrm{E}\left(tr\left(M_X UU'|X\right)\right) \\
&= tr\left(\mathrm{E}\left(M_X UU'|X\right)\right) \\
&= tr\left(M_X \mathrm{E}\left(UU'|X\right)\right) \\
&= \sigma^2 tr\left(M_X\right) \\
&= \sigma^2\left(tr\left(I_n\right) - tr\left(X\left(X'X\right)^{-1}X'\right)\right) \\
&= \sigma^2 n - \sigma^2 tr\left(\left(X'X\right)^{-1}X'X\right) \\
&= \sigma^2\left(n-k-1\right).
\end{aligned}$$

**ANS:** Note that $SSR = \hat{U}'\hat{U}$, where $\hat{U}$ is the vector of residuals of a regression of $Y$ on $X$. Therefore,

$$\hat{U} = M_X Y = M_X\left(X\beta + U\right) = M_X U,$$

so

$$\hat{U}'\hat{U} = U'M'_X M_X U$$

$$= U' M_X U,$$

where the last equality follows because $M_X$ is symmetric and idempotent. In the rest of the proof, note that the trace of a square matrix is the sum of its diagonal elements. The first equality holds because $U' M_X U$ is scalar, the second because whenever both $AB$ and $BA$ are well defined, $tr\,(AB) = tr\,(BA)$. The third equality follows because the trace is linear function, the fourth by taking out what is known from the conditional expectation, the fifth by assumption, since $Var\,(U|X) = \mathrm{E}\,(UU'|X) = \sigma^2 I_n$, the sixth by definition of $M_X$ and linearity of the trace, the seventh by using $tr\,(AB) = tr\,(BA)$ again, and the eighth by noting that $(X'X)^{-1} X'X = I_{k+1}$.

For parts (b)-(c) assume additionally that $y|x \sim \mathcal{N}\,(x'\beta, \sigma^2)$. Use the following facts to help answer the questions.

- Fact 1: Suppose $z \sim \mathcal{N}\,(0,1)$ and $Q \sim \chi^2_{n-k-1}$, and $z$ is independent of $Q$. Then

$$t = \frac{z}{\sqrt{Q/\,(n-k-1)}} \sim t_{n-k-1}.$$

- Fact 2: If $x \sim \mathcal{N}\,(\mu, \Sigma)$ and if $A$ and $B$ are conformable constant matrices, $Ax$ is independent of $Bx$ iff $Cov\,(Ax, Bx) = 0$.

- Fact 3: $\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2}|X \sim \chi^2_{n-k-1}$.

b) Show that the residual vector $\hat{U}$ and $\hat{\beta} - \beta$ are independent conditional on $X$. Argue that $\hat{\sigma}^2$ is independent of $\hat{\beta}$ conditional on $X$. Hint: First show that $\hat{U} = M_X U$, and argue that $U|X \sim \mathcal{N}\,(0, \sigma^2 I_n)$.

**ANS:** We have $\hat{U} = M_X U$ and $\hat{\beta} - \beta = (X'X)^{-1} X'U$. Both of these vectors have zero mean conditional on $X$, so

$$\begin{aligned}
Cov\left(M_X U, \hat{\beta} - \beta | X\right) &= \mathrm{E}\left(M_X U U' X\,(X'X)^{-1}\,|X\right) \\
&= M_X \mathrm{E}\,(UU'|X)\,X\,(X'X)^{-1} \\
&= \sigma^2 M_X X\,(X'X)^{-1} = 0.
\end{aligned}$$

Using Fact 2 now implies the vectors are independent conditional on $X$. Since

$$\hat{\sigma}^2 = \frac{\hat{U}'\hat{U}}{n-k-1},$$

it is just a function of $\hat{U}$, and therefore also independent of $\hat{\beta} - \beta$.

c) Suppose you wish to test the null hypothesis $H_0 : \beta_k = \beta_k^0$. Show that under $H_0$:

$$\frac{\hat{\beta}_k - \beta_k^0}{se\left(\hat{\beta}_k\right)} \sim t_{n-k-1}.$$

**ANS:** We show that

$$\frac{\hat{\beta}_k - \beta_k^0}{se\left(\hat{\beta}_k\right)} | X \sim t_{n-k-1},$$

from which the conclusion follows because this implies the distribution of the statistic is independent of $X$. Note that

$$se\left(\hat{\beta}_k\right) = \sqrt{\hat{\sigma}^2 e_k' \left(X'X\right)^{-1} e_k},$$

which is just a function of $\hat{\sigma}^2$ after conditioning on $X$. Now recall from Week 8 that (under $H_0$)

$$b_k = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\sigma^2 e_k' \left(X'X\right)^{-1} e_k}} \sim \mathcal{N}\left(0, 1\right),$$

which is just a function of $\hat{\beta}$ after conditioning on $X$. Therefore

$$\frac{\hat{\beta}_k - \beta_k^0}{se\left(\hat{\beta}_k\right)} = \frac{b_k}{\sqrt{\hat{\sigma}^2/\sigma^2}}.$$

Rewriting Fact 3 shows that $\hat{\sigma}^2/\sigma^2 | X \sim \chi^2_{n-k-1}/\left(n-k-1\right)$. Fact 1 now establishes the conclusion.

Now consider the model with weaker assumptions

$$y = x'\beta + u; \qquad \mathrm{E}\left(xu\right) = 0, Var\left(xu\right) = \mathrm{E}\left(u^2 x x'\right),$$

where $\mathrm{E}\left(u^2 x x'\right)$ is invertible and $\beta = \left(\beta_0, \beta_1, \beta_2\right)' \in \mathbb{R}^3$.

d) Suppose we want to test $\beta_1 = \beta_2 = 1$ vs. the alternative that one of the equalities does not hold. Describe in detail how to conduct this test.

**ANS:** Let $\hat{\beta}$ be the OLS estimator of $\beta$,and let

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The proposed null hypothesis is equivalent to $H_0 : R\beta = c$, and the alternative is $H_1 : R\beta \neq c$. It follows that under $H_0$

$$n \cdot \left(R\hat{\beta}_n - c\right)' \left(R\hat{V}_n R'\right)^{-1} \left(R\hat{\beta}_n - c\right) \xrightarrow{d} \chi^2_p,$$

where

$$\hat{V}_n = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 x_i x_i' \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \xrightarrow{p} \mathrm{E}\left(xx'\right)^{-1} \mathrm{E}\left(u^2 x x'\right) \mathrm{E}\left(xx'\right)^{-1},$$

and $\hat{u}_i = y_i - x_i'\hat{\beta}$. So a test of asymptotic level $\alpha$ would reject iff $T_n > \chi^2_{p,1-\alpha}$.

e) Suppose you want to test the null hypothesis that $\left(\beta_1 - \beta_2\right)^2 = 0$. Can you use the test for non-linear restrictions? If not, how would you conduct a test of this hypothesis?

**ANS:** No. Let $f(\beta_0, \beta_1, \beta_2) = (\beta_1 - \beta_2)^2$. We have

$$D_\beta f(\beta) = [0, 2(\beta_1 - \beta_2), -2(\beta_1 - \beta_2)].$$

Under $H_0$, $f(\beta) = 0$ which implies $\beta_1 - \beta_2 = 0$, so $D_\beta f(\beta) = [0, 0, 0]$ which is not full rank. Let $\hat{\beta}$ denote the OLS estimator of $\beta$. As expected, under $H_0$ :

$$\sqrt{n}\left(\hat{\beta}_1 - \hat{\beta}_2\right)^2 \xrightarrow{d} 0,$$

so we can't use the test for non-linear restrictions. Instead we note that $H_0$ is equivalent to $\beta_1 - \beta_2 = 0$, so we can equivalently characterize it as $H_0 : r'\beta = 0$ where $r = [0, 1, -1]$. Using results in Week 8, it follows that under $H_0$, the test statistic

$$T_n = \frac{\sqrt{n}\left(r'\hat{\beta} - 0\right)}{\sqrt{r'\hat{V}_n r}} \xrightarrow{d} \mathcal{N}(0, 1),$$

The test we use is $\phi_n = \mathbf{1}\left(|T_n| > z_{1-\alpha/2}\right)$, which we proved has asymptotic size $\alpha$.

**Question 3** Suppose $y = \beta x + u$, where $y, x, u$ are scalar random variables, but $\mathrm{E}(xu) \neq 0$. Suppose there exists a scalar random variable $z$ such that $\mathrm{E}(u|z) = 0$. Show by example that the IV estimate of $\beta$ is generally biased.

**ANS:** Suppose $u \sim \mathcal{N}(0, 1)$ and $x = u$. Then $y = (\beta + 1) u \sim \mathcal{N}\left(0, (\beta + 1)^2\right)$. We have $\mathrm{E}(xu) = \mathrm{E}(u^2) = 1 \neq 0$. Now let $z = 1$. $\mathrm{E}(u|z) = \mathrm{E}(u) = 0$. The IV estimator is given by

$$\begin{aligned}
\hat{\beta}_{IV} &= (Z'X)^{-1} Z'Y \\
&= \beta + (Z'X)^{-1} Z'U \\
&= \beta + \frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n x_i} \\
&= \beta + 1,
\end{aligned}$$

so it is biased for any sample size. In general, the reason we cannot prove unbiasedness is that

$$\mathrm{E}\left(\hat{\beta}_{IV}|Z, X\right) = \beta + (Z'X)^{-1} Z'\mathrm{E}(U|Z, X).$$

The problem is that while $\mathrm{E}(U|Z) = 0$ may hold, we cannot also assume $\mathrm{E}(U|X, Z) = 0$, because $\mathrm{E}(U|X) \neq 0$ to begin with.

**Question 4** Consider the following regression model of the returns to schooling:

$$\ln(w) = \beta_0 + \beta_1 y + \beta_2 x + u,$$

where $w$ denotes annual salary, $y$ denotes years of schooling and $x$ is a measure of previous work experience. Suppose $y$ is endogenous $x$ is exogenous and that $z_1$ is an instrument for $y$. Let $x =$

$(1, y, x)$, $z = (1, z_1, x)$ and $\beta = (\beta_0, \beta_1, \beta_2)$. Suppose $\mathrm{E}(zx'), \mathrm{E}(zz')$ and $Var(zu)$ exist, and that both $\mathrm{E}(zz')$ and $Var(zu)$ are invertible. Suppose $z_1$ is relevant and valid. Suppose we observe an iid sample $\{w_i, y_i, x_i\}_{i=1}^n$.

a) What does it mean for $x$ to be exogenous and $y$ endogenous?

**ANS:** $y$ is endogenous if $\mathrm{E}(yu) \neq 0$, which, given that $\mathrm{E}(u) = 0$, is equivalent to saying that $y$ is correlated with the error term. $x$ is exogenous if $\mathrm{E}(xu) = 0$, or equivalently if it is not correlated with the error term.

b) Can we assume $\mathrm{E}(u) = 0$ without loss of generality?

**ANS:** Yes. Note that $\mathrm{E}(\ln w) = [\beta_0 + \mathrm{E}(u)] + \beta_1 \mathrm{E}(y) + \beta_2 \mathrm{E}(x) + [u - \mathrm{E}(u)]$, so by letting $\beta_0^* = \beta_0 + \mathrm{E}(u)$ and redefining the error as $v = u - \mathrm{E}(u)$, we may write

$$\ln(w) = \beta_0^* + \beta_1 y + \beta_2 x + v; \quad \mathrm{E}(v) = 0.$$

c) What does it mean for $z_1$ to be relevant and valid?

**ANS:** $z_1$ is valid provided $\mathrm{E}(z_1 u) = 0$, otherwise using it as an instrument would also lead to an inconsistent estimate of $\beta$. $z_1$ is relevant if the coefficient $\gamma_2 \neq 0$ in the first stage regression

$$y = \gamma_0 + \gamma_1 x + \gamma_2 z_1 + v; \quad \mathrm{E}(zv) = 0.$$

Informally, $z_1$ must be correlated with $y$ "after controlling for" $x$.

d) Describe how to estimate $\beta$ using the instrumental variables estimator.

**ANS:** Let $Z$ be an $n \times 3$ matrix containing the observations of $z = (1, x, z_1)$, $W$ be an $n \times 1$ vector containing the observations of $\ln(w)$ and $X$ be an $n \times 3$ matrix containing the observations of $x$. Since the model is exactly identified, a consistent estimator of $\beta$ is obtained using the IV estimator:

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

e) Suppose you wish to test the null hypothesis that $z_1$ is irrelevant vs. the alternative that it is relevant at asymptotic level $\alpha$. State the null and alternative hypotheses in terms of features of the joint distribution of $z, x$, and describe how you would perform the test.

**ANS:** Our analysis in Week 9 tells us that $z_1$ is relevant iff the coefficient $\gamma_2 \neq 0$ in the regression

$$y = \gamma_0 + \gamma_1 x + \gamma_2 z_1 + v; \quad \mathrm{E}(zv) = 0.$$

We may use the asymptotic test developed in Week 8. Consider the parameter vector $(\gamma_0, \gamma_1, \gamma_2)'$ and the restriction matrix $r = [0, 0, 1]$. The null hypothesis is $H_0 : r'\gamma = 0$ vs. the alternative $H_1 : r'\gamma \neq 0$. Let $Y$ be an $n \times 1$ vector containing the observations of $y$. The OLS estimate is given by

$$\hat{\gamma} = (Z'Z)^{-1} Z'Y = \gamma + (Z'Z)^{-1} Z'V.$$

By the CLT

$$\sqrt{n}\left(\hat{\gamma}-\gamma\right)=\sqrt{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i v_i \xrightarrow{d} \mathcal{N}\left(0,\mathrm{E}\left(zz'\right)^{-1}\mathrm{E}\left(v^2 zz'\right)\mathrm{E}\left(zz'\right)^{-1}\right).$$

It follows by the continuous mapping theorem that

$$\sqrt{n}\left(\hat{\gamma}_2-\gamma_2\right)=\sqrt{n}r'\left(\hat{\gamma}-\gamma\right)\xrightarrow{d}\mathcal{N}\left(0,r'\Omega r\right),$$

where $\Omega=\mathrm{E}\left(zz'\right)^{-1}\mathrm{E}\left(v^2zz'\right)\mathrm{E}\left(zz'\right)^{-1}$. A consistent estimate of $\Omega$ is given by

$$\hat{\Omega}=\left(\frac{1}{n}\sum_{i=1}^{n}z_iz_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{v}_i^2z_iz_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_iz_i'\right)^{-1},$$

where $\hat{v}_i=y_i-\hat{\gamma}'z_i$. By Slutsky's theorem, we obtain

$$T_n=\frac{\sqrt{n}\left(\hat{\gamma}_2-\gamma_2\right)}{\sqrt{r'\hat{\Omega}r}}\xrightarrow{d}\mathcal{N}\left(0,1\right).$$

Therefore, we propose the test

$$\phi_n=\mathbf{1}\left(|T_n|>z_{1-\alpha/2}\right),$$

which has asymptotic size $\alpha$, since $|T_n|\xrightarrow{d}|\mathcal{N}\left(0,1\right)|$. That is, we reject the null that $z_1$ is not relevant if $T_n$ is greater than the $1-\alpha/2$ quantile of the standard normal distribution.

f) Suppose $Cov\left(y,x\right)\neq 0$. Can we use $x$ as an instrument for $y$?

**ANS:** No, since $x$ is already an included instrument, if we tried to add it as an instrument for $y$ our vector of instruments would be $z=(1,x,x)$, which would be collinear. The rank condition would therefore fail, because the second row of $\mathrm{E}\left(zx'\right)$ would equal its third row, so $\beta$ would not be identified. On the other hand, if we omitted $x$ in an attempt to use it as an instrument for $y$, relevance would be satisfied because $Cov\left(x,y\right)\neq 0$, but now validity would not, because in the regression

$$\ln\left(w\right)=\beta_0+\beta_1 y+\epsilon, \tag{2}$$

where $\epsilon=\beta_2x+u$, we have $\mathrm{E}\left(x\epsilon\right)=\beta_2\mathrm{E}\left(x^2\right)+\mathrm{E}\left(xu\right)=\beta_2\mathrm{E}\left(x^2\right)$. Thus if $x$ is a relevant variable (meaning $\beta_2\neq 0$) it cannot be used as an instrument for $y$. If $\beta_2=0$ then $\epsilon=u$ in model (2) and $x$ may be used as an instrument for $y$.

**Question 5 (Hausman Test)** Suppose you observe an iid sample of $\{y_i,x_i,z_i\}_{i=1}^n$, where $y$ is a scalar random variable, $x\in\mathbb{R}$ and $z\in\mathbb{R}$ are random variables, and consider the model

$$y=\beta x+u.$$

In this question we will use the instrumental variable $z$ to test the assumption that $\mathrm{E}\left(xu\right)=0$ holds in the linear model.

a) Write down the OLS estimator and the IV estimator of $\beta$ using the variable $z$ as the instrument. Write your answer using sums and also using matrices.

**ANS:** We have

$$\hat{\beta}_{OLS} = \left(X'X\right)^{-1} X'Y = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2};$$
$$\hat{\beta}_{IV} = \left(Z'X\right)^{-1} Z'Y = \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i x_i}.$$

b) Under what conditions is $z$ a relevant and valid instrumental variable? Show that $\hat{\beta}_{IV}$ is a consistent estimator of $\beta$ under these assumptions.

**ANS:** The relevance condition is $\mathrm{E}\left(zx\right) \neq 0$, which is not the same as saying $z$ is correlated with $x$, because there is no constant. The validity condition is $\mathrm{E}\left(zu\right) = 0$. Under these conditions,

$$\hat{\beta}_{IV} = \beta + \frac{\frac{1}{n}\sum_{i=1}^n z_i u_i}{\frac{1}{n}\sum_{i=1}^n z_i x_i} \xrightarrow{a.s.} \beta + \frac{\mathrm{E}\left(zu\right)}{\mathrm{E}\left(zx\right)} = \beta,$$

by the strong law of large numbers, joint convergence almost surely of $\left(\frac{1}{n}\sum_{i=1}^n z_i u_i, \frac{1}{n}\sum_{i=1}^n z_i x_i\right)$ and the continuous mapping theorem.

c) Suppose that $\mathrm{E}\left(xu\right) = 0$ and $\mathrm{E}\left(zu\right) = 0$. Compute the joint asymptotic distribution of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$. Hint: First, show that

$$\begin{aligned}\begin{pmatrix} \hat{\beta}_{OLS} - \beta \\ \hat{\beta}_{IV} - \beta \end{pmatrix} &= \begin{pmatrix} (X'X)^{-1} & 0 \\ 0 & (Z'X)^{-1} \end{pmatrix} \begin{pmatrix} X'U \\ Z'U \end{pmatrix} \\ &= \begin{pmatrix} (X'X)^{-1} & 0 \\ 0 & (Z'X)^{-1} \end{pmatrix} \sum_{i=1}^n \begin{pmatrix} x_i \\ z_i \end{pmatrix} u_i.\end{aligned}$$

Now use the CLT.

**ANS:** The tricky part of deriving the joint asymptotic distribution is separating out a sum which the CLT can be applied to, but the trick above works well. Note that by the multivariate CLT and Slutsky's Theorem,

$$\begin{aligned}\sqrt{n}\begin{pmatrix} \hat{\beta}_{OLS} - \beta \\ \hat{\beta}_{IV} - \beta \end{pmatrix} &= \sqrt{n}\begin{pmatrix} \left(X'X\right)^{-1} X'U \\ \left(Z'X\right)^{-1} Z'U \end{pmatrix} \\ &= \sqrt{n}\begin{pmatrix} (X'X)^{-1} & 0 \\ 0 & (Z'X)^{-1} \end{pmatrix} \begin{pmatrix} X'U \\ Z'U \end{pmatrix} \\ &= \begin{pmatrix} \left(\frac{X'X}{n}\right)^{-1} & 0 \\ 0 & \left(\frac{Z'X}{n}\right)^{-1} \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} x_i u_i \\ z_i u_i \end{pmatrix} \\ &\xrightarrow{d} \begin{pmatrix} \frac{1}{\mathrm{E}(x^2)} & 0 \\ 0 & \frac{1}{\mathrm{E}(zx)} \end{pmatrix} \times \mathcal{N}\left(0, \begin{pmatrix} \mathrm{E}\left(u^2x^2\right) & \mathrm{E}\left(u^2xz\right) \\ \mathrm{E}\left(u^2xz\right) & \mathrm{E}\left(u^2z^2\right) \end{pmatrix}\right)\end{aligned}$$

$$\stackrel{d}{=} \mathcal{N}\left(0, \begin{pmatrix} \frac{\mathrm{E}\left(u^2x^2\right)}{\mathrm{E}(x^2)^2} & \frac{\mathrm{E}\left(u^2xz\right)}{\mathrm{E}(x^2)\mathrm{E}(zx)} \\ \frac{\mathrm{E}\left(u^2xz\right)}{\mathrm{E}(zx)\mathrm{E}(x^2)} & \frac{\mathrm{E}\left(u^2z^2\right)}{\mathrm{E}(zx)^2} \end{pmatrix}\right). \tag{3}$$

d) Explain how to use your result to conduct a test of asymptotic size $\alpha$ of $H_0 : \mathrm{E}\left(xu\right) = 0$ vs. $H_1 : \mathrm{E}\left(xu\right) \neq 0$. Assume that all relevant matrix inverses exist.

**ANS:** Under the alternative hypothesis, $\hat{\beta}_{OLS} \stackrel{p}{\nrightarrow} \beta$, so $\hat{\beta}_{OLS} - \hat{\beta}_{IV} \stackrel{p}{\nrightarrow} 0$. Under the null, $\hat{\beta}_{OLS} - \hat{\beta}_{IV} \xrightarrow{p} 0$, which suggests using a test statistic based on (3) and then using the continuous mapping theorem. We obtain under $H_0$ :

$$\sqrt{n}\left(\hat{\beta}_{OLS} - \hat{\beta}_{IV}\right) \xrightarrow{d} \mathcal{N}\left(0, V\right).$$

where $V = \frac{\mathrm{E}\left(u^2x^2\right)}{\mathrm{E}(x^2)^2} - 2\frac{\mathrm{E}\left(u^2xz\right)}{\mathrm{E}(x^2)\mathrm{E}(zx)} + \frac{\mathrm{E}\left(u^2z^2\right)}{\mathrm{E}(zx)^2}$. We further assume that the asymptotic variance of $\hat{\beta}_{OLS} - \hat{\beta}_{IV}$ is nonzero, which is guaranteed provided $\mathrm{P}\left(xu \neq zu\right) > 0$, since in this case the Cauchy-Schwarz inequality provides that

$$\begin{pmatrix} \mathrm{E}\left(u^2x^2\right) & \mathrm{E}\left(u^2xz\right) \\ \mathrm{E}\left(u^2xz\right) & \mathrm{E}\left(u^2z^2\right) \end{pmatrix}$$

is a full rank matrix because its determinant is non-zero. It only remains to estimate the unknown expectations consistently. This can be done by noting that under $H_0$, $\hat{\beta}_{OLS}$ converges at rate $\sqrt{n}$ to $\beta$, and so does $\hat{\beta}_{IV}$. Therefore, set $\hat{u}_i = y_i - \hat{\beta}_{OLS}x_i$ or $\hat{u}_i = y_i - \hat{\beta}_{IV}x_i$. Mimicking the proof in Week 8 again yields that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i^2 \xrightarrow{p} \mathrm{E}\left(u^2x^2\right);$$
$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 z_i^2 \xrightarrow{p} \mathrm{E}\left(u^2z^2\right);$$
$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i z_i \xrightarrow{p} \mathrm{E}\left(u^2xz\right).$$

It follows by Slutsky's Theorem that under $H_0$:

$$T_n = \frac{\sqrt{n}\left(\hat{\beta}_{OLS} - \hat{\beta}_{IV}\right)}{\sqrt{\hat{V}}} \xrightarrow{d} \mathcal{N}\left(0, 1\right), \tag{4}$$

where

$$\hat{V} = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right]^2} - 2\frac{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i z_i}{\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right]\left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i\right]} + \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 z_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n} z_i x_i\right]^2}.$$

The test rejects if $|T_n| > z_{1-\alpha/2}$, and has asymptotic size $\alpha$ by result (4).

**Question 6** Suppose you observe an iid sample of $\{y_i, x_i\}_{i=1}^{n}$, where $y$ and $x$ are scalar random variables, and you assume

$$y = \beta x + u; \quad \mathrm{E}\left(u|x\right) = 0.$$

a) Is the assumption $\mathrm{E}\left(u|x\right) = 0$ without loss of generality?

**ANS:** No, $\mathrm{E}\left(y|x\right)$ need not be linear in $x$, e.g. $y = x^2$ would imply $u = x^2 - \beta x$ and $\mathrm{E}\left(u|x\right) = x^2 - \beta x$, which is almost surely nonzero for any $\beta$ if $x$ is continuously distributed.

b) Show that $\mathrm{E}\left(xu\right) = 0$ and $\mathrm{E}\left(x^2 u\right) = 0$. Explain why assuming that $\mathrm{E}\left(u|x\right) = 0$ yields many different representations of $\beta$.

**ANS:** By the law of iterated expectation,

$$\begin{aligned} \mathrm{E}\left(xu\right) &= \mathrm{E}\left(x\mathrm{E}\left(u|x\right)\right) = 0 \\ \mathrm{E}\left(xu\right) &= \mathrm{E}\left(x^2\mathrm{E}\left(u|x\right)\right) = 0. \end{aligned}$$

The conditional mean zero assumption is much stronger than the unconditional mean zero assumption because it implies that for any function $f \in L^2\left(x\right)$,

$$\mathrm{E}\left(uf\left(x\right)\right) = 0.$$

It follows that

$$\mathrm{E}\left(\left[y - \beta x\right] f\left(x\right)\right) = 0 \implies \beta = \frac{\mathrm{E}\left(yf\left(x\right)\right)}{\mathrm{E}\left(xf\left(x\right)\right)},$$

provided $\mathrm{E}\left(xf\left(x\right)\right) \neq 0$. There are therefore many representations of $\beta$ (all of which are equal in the population, but the choice of $f$ will affect the properties of resulting estimator).

c) Suppose now that you wish to estimate $\beta$ using a combination of the moments $\mathrm{E}\left(xu\right) = 0$ and $\mathrm{E}\left(x^2 u\right) = 0$ using GMM. Explain in detail how to calculate an asymptotically efficient GMM estimator of $\beta$ based on these moments. Consider 2 cases − conditional homoskedasticity and heteroskedasticity of $u$.

**ANS:** The instruments are $z = \left(x, x^2\right)$, where $x$ is included and $x^2$ is excluded. Consider first conditional homoskedasticity: $\mathrm{E}\left(u^2|x\right) = \mathrm{E}\left(u^2\right)$. In Week 9, we showed that an optimal GMM estimator is in fact the 2SLS estimator. In the first stage we regress all the covariates on all the instruments:

$$x = \gamma_0 x + \gamma_1 x^2 + v, \quad \mathrm{E}\left(\begin{pmatrix} x \\ x^2 \end{pmatrix} v\right) = 0,$$

but this simply produces $\gamma_0 = 1, \gamma_1 = 0$. In sample, a perfect fit is obtained because $x$ is already an included instrument, so under conditional homoskedasticity the OLS estimator is asymptotically optimal. Under heteroskedasticity, we let $Z$ be an $n \times 2$ matrix of the observations of $z = \left(x, x^2\right)$ and $X$ be an $n \times 1$ column vector containing the observations of $x$. Our derivations in class yield that

$$\hat{\beta}_{OGMM} = \left(X'Z\hat{\Omega}^{-1}Z'X\right)^{-1} X'Z\hat{\Omega}^{-1}Z'Y,$$

is an optimal GMM estimator, where $\hat{\Omega}$ is a consistent estimator of $\Omega = \mathrm{E}\left(u^2 z z'\right)$. We assume that $\Omega$ exists and is invertible. This estimator will typically not equal the OLS estimator, which can be considered an IV estimator that simply 'throws out' the second instrument. Instead, the optimal GMM estimator finds an optimal linear combination of $x$ and $x^2$ in the sense that

$$\sqrt{n}\left(\hat{\beta}_{OGMM} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \left[\mathrm{E}\left(x_i z_i'\right)\Omega^{-1}\mathrm{E}\left(z_i x_i'\right)\right]^{-1}\right),$$

which has the smallest variance among GMM estimators. A consistent estimate of $\Omega$ is obtained by noting that the OLS estimator of $\beta$ converges at rate $\sqrt{n}$ and mimicking the proof given in Week 8, and so

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 z_i z_i' \xrightarrow{p} \Omega,$$

where $\hat{u}_i = y_i - x_i'\hat{\beta}_{OLS}$.