PROBLEM SET # 3

ECONOMICS 240B -SPRING 2005

All answers are the work of Tomas Rau, who was the GSI in 2005. Tomas, I can't thank you enough!

1. Ruud's book questions

20.3 Consider the following,

$$\begin{array}{ll} y_n = x_n'\beta + e_n & \textit{true model} \\ y_n = x_n'\beta + z_n'\alpha + \epsilon_n & \textit{estimated model} \end{array}$$

where $\mathbb{E}(e_n|x_n) = 0$. Now, analyze the following the statement: Including "irrelevant" variables only causes inefficiency.

case 1: $\mathbb{E}(Z'e) = 0$

from the partition regression formula we know that $\hat{\alpha}$ and $\hat{\beta}$ are equal to:

$$\hat{\alpha} = [Z'(I - P_X)Z]^{-1}Z'(I - P_X)y = [Z'(I - P_X)Z]^{-1}Z'(I - P_X)(X\beta + e)$$
$$\hat{\beta} = [X'(I - P_Z)X]^{-1}X'(I - P_Z)y = \beta + [X'(I - P_Z)X]^{-1}X'(I - P_Z)e$$

Note that these estimators are unbiased under the assumption $\mathbb{E}(Z'e) = 0$.

$$\begin{aligned} \mathbb{E}(\hat{\alpha}|X) &= [Z'(I - P_X)Z]^{-1}Z'(I - P_X)X\beta = 0 \\ \mathbb{E}(\hat{\beta}|X) &= \beta \end{aligned}$$

where in the first equality we used that $(I-P_X)X = 0$ and in the second, that $\mathbb{E}(Z'e) = 0$ and $\mathbb{E}_Z[\mathbb{E}(\hat{\beta}|X, Z)] = \mathbb{E}[\hat{\beta}|X]$. Note in addition that in the *true model* $\alpha_0 = 0$. It's easy to see that consistency is not threatened in this case either. However, efficiency is lost.

$$\mathbb{V}(\hat{\beta}|X) = \sigma^2[X'(I - P_Z)X]^{-1} \; \textit{instead of} \; \sigma^2(X'X)^{-1}$$

It's not hard to prove that $\sigma^2[X'(I-P_Z)X]^{-1}$ is higher than $\sigma^2(X'X)^{-1}$ in a p.d. sense. Note that this is equivalent to prove that

$$(X'X) - [X'(I - P_Z)X] \geq 0$$

$$\begin{aligned} (X'X) - [X'(I - P_Z)X] &= X'P_ZX = X'P_Z'P_ZX) \\ &= (P_ZX)'(P_ZX) \geq 0 \end{aligned}$$

Hence, in case 1, it;s true that we loss just efficiency by adding an irrelevant explanatory variable.

Date: March 10 2005.

case 2 Consider in this case that $\mathbb{E}(Z'e) \neq 0$.

It's straightforward to see that $\hat{\beta}$ is biased and inconsistent, so including a "irrelevant" variable mess up our estimation. Watch out with your explanatory variables...

20.7 Consider the following set up,

$$\begin{aligned} y_n &= \beta_0 x_n^* + u_n \\ x_n &= x_n^* + v_n \Rightarrow \\ y_n &= \beta_0 x_n - \beta_0 v_n + u_n \\ &\text{and the reverse is just,} \\ x_n &= \alpha_0 y_n - \alpha_0 u_n + v_n \Rightarrow \\ x_n &= \alpha_0 y_n^* + v_n \\ y_n^* &= y_n - u_n \end{aligned}$$

where $\alpha_0 = \frac{1}{\beta_0}$

a) Show that $\hat{\alpha}$ and $\hat{\beta}_0$ have probability limits that bound β_0

for $\hat{\beta}_0$, Ruud shows that

$$\operatorname{plim} \hat{\beta}_{ols} = \beta_0(1 - \pi_0) < \beta_0$$

where $x_n \pi_0 = \mathbb{E}(v_n | u_n)$. So, we have a lower bound for β_0.

Now for the reverse regression we have, that

$$\hat{\alpha_0} = \frac{\sum x_n y_n}{\sum y_n^2}$$

hence,

$$\operatorname{plim} \hat{\alpha_0} = \frac{\mathbb{E}(x_n y_n)}{\mathbb{E}(y_n^2)}$$

note that, $\mathbb{E}(x_n y_n) = \alpha_0 \mathbb{E}(y_n^2) - \alpha_0 \mathbb{V}(u_n)$, so

$$\operatorname{plim} \hat{\alpha_0} = \alpha_0 - \alpha_0 \frac{\mathbb{V}(u_n)}{\mathbb{E}(y_n^2)} = \alpha_0(1 - \delta_0)$$

where, $0 < \delta_0 = \frac{\mathbb{V}(u_n)}{\mathbb{E}(y_n^2)} < 1$ because $\mathbb{E}(y_n^2) = \mathbb{E}(y_n^{*2}) + \mathbb{V}(u_n)$

Hence,

$$\alpha_0(1 - \delta_0) < \alpha_0 \Rightarrow \beta_0 \left(\frac{1}{1 - \delta_0} \right) > \beta_0$$

so, $\hat{\alpha}_0$ from the reverse regression give us an upper bound for β_0

b) Now, the model includes an intercept, so

$$\begin{aligned} y_n &= \beta_{01} + \beta_{02} x_n + (-\beta_{02} v_n + u_n) \\ x_n &= x_n^* + v_n \end{aligned}$$

using bivariate regression results, we have

$$\hat{\beta}_{02} = \frac{\sum (x_n - \bar{x})(y_n - \bar{y})}{\sum (x_n - \bar{x})^2}$$

$$\text{plim}\,\hat{\beta}_{02} = \frac{\mathbb{C}(x_n, y_n)}{\mathbb{V}(x_n)} = \beta_{02} - \beta_{02}\frac{\mathbb{E}(x_n v_n)}{\mathbb{V}(x_n)} = \beta_{02}\,(1-\delta_0) < \beta_{02}$$

Where $\delta_0 = \frac{\mathbb{E}(x_n v_n)}{\mathbb{V}(x_n)}$. Note that $0 < \gamma < 1$ because $\mathbb{E}(x_n v_n) = \mathbb{V}(v_n)$ and that $\mathbb{V}(x_n) = \mathbb{V}(x_n^*) + \mathbb{V}(v_n)$ hence $\mathbb{V}(v_n) \leq \mathbb{V}(x_n)$

Following the same steps, we get for the reverse regression that

$$\text{plim}\,\hat{\gamma}_{02} = \frac{\mathbb{C}(x_n, y_n)}{\mathbb{V}(y_n)} = \gamma_{02} - \gamma_{02}\frac{\mathbb{E}(y_n u_n)}{\mathbb{V}(y_n)} = \gamma_{02}\,(1-\lambda_0) < \gamma_{02}$$

where $\lambda_{02} = \frac{\mathbb{E}(y_n u_n)}{\mathbb{V}(y_n)}$ and $\gamma_{02} = 1/\beta_{02}$. It is easy to see that $0 < \lambda_{02} < 1$. Hence,

$$\text{plim}\,\hat{\beta}_{02} = \beta_{02}\left(\frac{1}{1-\lambda_0}\right) > \beta_{02}$$

20.12 We know that the MMSE linear predictor of y_n is $x_n'\gamma_0$ and by lemma 20.1, we basically need to prove that $x_n'\gamma_0$ is the MMSE linear predictor of the conditional mean $\mathbb{E}(y_n|x_n)$ given x_n. Here we shouldn't interpret that we want to find the "conditional" MMSE, it is just the MMSE for conditional mean. Hence, need to find

$$\min_{\gamma\in\mathbb{R}^k} \mathbb{E}[(\mathbb{E}(y_n|x_n) - x_n'\gamma)^2]$$

Now, we want to show that indeed, this minimization is equivalent to minimize the MMSE for best linear predictor of y_n.

$$\begin{aligned}\mathbb{E}[(\mathbb{E}(y_n|x_n) - x_n'\gamma)^2] &= \mathbb{E}[(\mathbb{E}(y_n|x_n) - y_n + y_n - x_n'\gamma)^2] \\ &= \mathbb{E}[(\mathbb{E}(y_n|x_n) - y_n)^2 + (y_n - x_n'\gamma)^2 + \\ &-2[\mathbb{E}(y_n|x_n) - y_n][y_n - x_n'\gamma]]\end{aligned}$$

Note that the cross product vanishes because of the *law of iterated expectations*, $\mathbb{E}[\mathbb{E}(y_n|x_n)] = \mathbb{E}(y_n)$. Hence,

$$\min_{\gamma\in\mathbb{R}^k} \mathbb{E}[(\mathbb{E}(y_n|x_n) - x_n'\gamma)^2] = \min_{\gamma\in\mathbb{R}^k} \{\mathbb{E}[(\mathbb{E}(y_n|x_n) - y_n)^2] + \mathbb{E}[(y_n - x_n'\gamma)^2]\}$$

Note that the first element of the curly doesn't depend on γ so this is just the same setup to find a MMSE linear predictor of y_n. Therefore, the MMSE linear predictor is $x_n\gamma_0$.

24.1 Recall that the LSDV estimator is:

$$\hat{\beta}_{LSDV} = [\sum_{i=1}^{N}\sum_{t=1}^{2}(x_{it} - x_{i.})(x_{it} - x_{i.})']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{2}(x_{it} - x_{i.})(y_{it} - y_{i.})$$

and in this case $x_{i.} = \frac{x_{i1}+x_{i2}}{2}$, hence

$$\begin{aligned}
\hat{\beta}_{LSDV} &= [\textstyle\sum_{i=1}^{N}(x_{i1}-x_{i.})(x_{i1}-x_{i.})' + (x_{i2}-x_{i.})(x_{i2}-x_{i.})']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{2}\bar{x}_{it}\bar{y}_{it} \\
&= [\textstyle\sum_{i=1}^{N}(\frac{x_{i1}-x_{i2}}{2})(\frac{x_{i1}-x_{i2}}{2})' + (\frac{x_{i2}-x_{i1}}{2})(\frac{x_{i2}-x_{i1}}{2})']^{-1}\sum_{i=1}^{N}\sum_{t=1}^{2}\bar{x}_{it}\bar{y}_{it} \\
&= [\textstyle\sum_{i=1}^{N}\frac{(x_{i2}-x_{i1})(x_{i2}-x_{i1})'}{2}]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{2}(x_{it}-x_{i.})(y_{it}-y_{i.}) \\
&= [\textstyle\sum_{i=1}^{N}\frac{(x_{i2}-x_{i1})(x_{i2}-x_{i1})'}{2}]^{-1}\sum_{i=1}^{N}\frac{(x_{i2}-x_{i1})(y_{i2}-y_{i1})}{2} \\
&= [\textstyle\sum_{i=1}^{N}(x_{i2}-x_{i1})(x_{i2}-x_{i1})']^{-1}\sum_{i=1}^{N}(x_{i2}-x_{i1})(y_{i2}-y_{i1})
\end{aligned}$$

which is exactly the OLS fitted coefficients from a regression of $y_{i2}-y_{i1}$ on $x_{i2}-x_{i1}$.

24.3 Using Greene's notation, let

$$\begin{aligned}
S_{xx}^{w} &= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{i.})(x_{it}-x_{i.})' \\
S_{xy}^{w} &= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{i.})(y_{it}-y_{i.}) \\
S_{xx}^{B} &= \textstyle\sum_{i=1}^{N}(x_{i.}-x_{..})(x_{i.}-x_{..})' \\
S_{xy}^{B} &= \textstyle\sum_{i=1}^{N}(x_{i.}-x_{..})(y_{i.}-y_{..}) \\
S_{xx}^{t} &= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{..})(x_{it}-x_{..})' \\
S_{xy}^{t} &= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{..})(y_{it}-y_{..})
\end{aligned}$$

hence,

$$\begin{aligned}
\hat{\beta}_{LS} &= [S_{xx}^{t}]^{-1}S_{xy}^{t} \\
\hat{\beta}_{W} &= [S_{xx}^{w}]^{-1}S_{xy}^{w} \\
\hat{\beta}_{B} &= [S_{xx}^{B}]^{-1}S_{xy}^{B}
\end{aligned}$$

now, consider S_{xy}^{t},

$$\begin{aligned}
S_{xy}^{t} &= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{..})(y_{it}-y_{..}) \\
&= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-x_{..}+x_{i.}-x_{i.})(y_{it}-y_{..}+y_{i.}-y_{i.}) \\
&= \textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}[(x_{it}-x_{..})(y_{it}-y_{..})+(x_{it}-x_{..})(y_{i.}-y_{..})+(x_{i.}-x_{..})(y_{it}-y_{..})+ \\
&\quad +(x_{i.}-x_{..})(y_{i.}-y_{..})] \\
&= S_{xy}^{w}+S_{xy}^{B}
\end{aligned}$$

the last equality is because the the two middle products vanish when adding over t. Analog proof can be done for $S_{xx}^{t}=S_{xx}^{w}+S_{xx}^{B}$

Hence,

$$\begin{aligned}
\hat{\beta}_{LS} &= [S_{xx}^{w}+S_{xx}^{B}]^{-1}(S_{xy}^{w}+S_{xy}^{B}) \\
&= [S_{xx}^{w}+S_{xx}^{B}]^{-1}S_{xy}^{w}+[S_{xx}^{w}+S_{xx}^{B}]^{-1}S_{xy}^{B} \\
&= [S_{xx}^{w}+S_{xx}^{B}]^{-1}S_{xx}^{w}\hat{\beta}_{w}+[S_{xx}^{w}+S_{xx}^{B}]^{-1}S_{xx}^{B}\hat{\beta}_{B} \\
&= F_{w}\hat{\beta}_{w}+(I-F_{w})\hat{\beta}_{B}
\end{aligned}$$

where $F_{w}=[S_{xx}^{w}+S_{xx}^{B}]^{-1}S_{xx}^{w}$

24.6

a) This part is trivial, just need to solve for σ_α^2

$$\hat{\sigma}_\alpha^2 = \sum_{n=1}^{N} \frac{[\bar{y}_n - \bar{\bar{y}} - (\bar{x}_n - \bar{\bar{x}})\hat{\beta}_B]^2}{N-1-K} - \frac{1}{T}\left(\sum_{n=1}^{N}\sum_{t=1}^{T} \frac{[y_{nt} - \bar{y}_n - (x_{nt} - \bar{x}_n)\hat{\beta}_{DV}]^2}{NT-N-K}\right)$$

Note the typo in Ruud's book, $\hat{\sigma}_\epsilon^2$ should be divided by $NT-N-K$. Note that this estimator is unbiased but can be negative...

b) This is negative obviously when,

$$\sum_{n=1}^{N} \frac{[\bar{y}_n - \bar{\bar{y}} - (\bar{x}_n - \bar{\bar{x}})\hat{\beta}_B]^2}{N-1-K} < \frac{1}{T}\left(\sum_{n=1}^{N}\sum_{t=1}^{T} \frac{[y_{nt} - \bar{y}_n - (x_{nt} - \bar{x}_n)\hat{\beta}_{DV}]^2}{NT-N-K}\right)$$

so,

$$\sum_{n=1}^{N}[\bar{y}_n - \bar{\bar{y}} - (\bar{x}_n - \bar{\bar{x}})\hat{\beta}_B]^2 < \frac{N-1-K}{T(NT-N-K)}\left(\sum_{n=1}^{N}\sum_{t=1}^{T}[y_{nt} - \bar{y}_n - (x_{nt} - \bar{x}_n)\hat{\beta}_{DV}]^2\right)$$

since, for $T > 1$, $\frac{N-1-K}{T(NT-N-K)}$ is less than one, then one alternative of having this is for no significant values of $\hat{\beta}_{DV}$ in comparison to $\hat{\beta}_B$

c) It might cast some doubt on the appropriateness of the random effect model.

Additional questions:

1.

a) First we need to calculate Ω.

$$\Omega = E(\varepsilon\varepsilon'/X) = \begin{bmatrix} x_{11}'\Gamma_{11}x_{11} & & 0 & x_{11}'\Gamma_{12}x_{12} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & x_{N1}'\Gamma_{11}x_{N1} & 0 & & x_{N1}'\Gamma_{11}x_{N2} \\ x_{12}'\Gamma_{21}x_{11} & & 0 & x_{12}'\Gamma_{22}x_{12} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & x_{N2}'\Gamma_{21}x_{N1} & 0 & & x_{N2}'\Gamma_{22}x_{N2} \end{bmatrix}$$

$= X(\Sigma \otimes I_N)X'$, where $X = \begin{bmatrix} X_1 & 0 \\ 0 & X2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$

We need a consistent estimator of Σ. Observe that

$\beta_{ij} = (x_{ij}x_{ij}')x_{ij}y_{ij}$, and the β_{ij} are iid in i, with finite variances, so, assuming that larger moments are still finite, we have:

$$\frac{1}{N}\sum_{i=1}^{N}(\beta_{ij} - \beta_j)(\beta_{ik} - \beta_k)' \longrightarrow_p \Gamma_{jk}$$

Now, we know that $\widehat{\beta}_j \longrightarrow_p \beta_j$, since OLS is still consistent, so

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(\beta_{ij}-\widehat{\beta}_j)(\beta_{ik}-\widehat{\beta}_k)' &= \\
& \frac{1}{N}\sum_{i=1}^{N}(\beta_{ij}-\beta_j)(\beta_{ik}-\beta_k)' + (\beta_j-\widehat{\beta}_j)\frac{1}{N}\sum_{i=1}^{N}(\beta_{ik}-\beta_k)' + \\
& + \left(\frac{1}{N}\sum_{i=1}^{N}(\beta_{ij}-\beta_j)\right)(\beta_k-\widehat{\beta}_k)' + (\beta_j-\widehat{\beta}_j)(\beta_k-\widehat{\beta}_k)' \\
&\longrightarrow \; {}_p\Gamma_{jk}
\end{aligned}
$$

because all 3 last terms converge in prob. to zero.

Let $\widehat{\Gamma}_{jk} = \frac{1}{N}\sum\limits_{i=1}^{N}(\beta_{ij}-\widehat{\beta}_j)(\beta_{ik}-\widehat{\beta}_k)'$, then substitute it to obtain $\widehat{\Sigma}$, and the FGLS estimator will be:

$$\widehat{\beta} = (X'(X(\widehat{\Sigma}\otimes I_N)X')^{-1}X)^{-1}X'(X(\widehat{\Sigma}\otimes I_N)X')^{-1}y$$

b) In this context, we are not assuming the random coefficient structure, so:

$$
\Omega = \begin{bmatrix}
\sigma_{1,11} & & 0 & \sigma_{1,12} & & 0 \\
& \ddots & & & \ddots & \\
0 & & \sigma_{N,11} & 0 & & \sigma_{N,12} \\
\sigma_{1,21} & & 0 & \sigma_{1,22} & & 0 \\
& \ddots & & & \ddots & \\
0 & & \sigma_{N,21} & 0 & & \sigma_{N,22}
\end{bmatrix}
$$

and, since the matrix is simmetric, we know that $\sigma_{i,12} = \sigma_{i,21}$, and, since $\sigma_{i,jk}$ varies with the person, we can't consistently estimate it. The solution using OLS would be to estimate each equation separately, which is the same as estimating OLS for the whole model (Since it is somehow simmilar to the SUR model, only the variances vary per individual), and then treat each equation as a simple heteroskedasticity case.

We use Eicker-White like estimator for the variance estimator:

$$\sqrt{N}(\widehat{\beta}-\beta) \longrightarrow_d N\left(0, p\lim\left(\frac{X'X}{N}\right)^{-1}\left(\frac{X'\Omega X}{N}\right)\left(\frac{X'X}{N}\right)^{-1}\right)$$

and the middle term is:

$$
\begin{bmatrix}
X_1'Diag(\sigma_{i,11})X_1 & X_1'Diag(\sigma_{i,12})X_2 \\
X_2'Diag(\sigma_{i,12})X_1 & X_2'Diag(\sigma_{i,22})X_2
\end{bmatrix}
= \begin{bmatrix}
\Gamma_{11} & \Gamma_{12} \\
\Gamma_{12}' & \Gamma_{22}
\end{bmatrix}
$$

So, simmilarly to the Eicker White case, we estimate them by:

$$\widehat{\Omega} = \begin{bmatrix} \widehat{\Gamma}_{11} & \widehat{\Gamma}_{12} \\ \widehat{\Gamma}'_{12} & \widehat{\Gamma}_{22} \end{bmatrix} = \begin{bmatrix} X_1' Diag(e_{i1}^2) X_1 & X_1' Diag(e_{i1} e_{i2}) X_2 \\ X_2' Diag(e_{i1} e_{i2}) X_1 & X_2' Diag(e_{i2}^2) X_2 \end{bmatrix}$$

where $e_{ij} = y_{ij} - x'_{ij} \widehat{\beta}_{LS}$.

The logic is the same as for Eicker-White: the difference between the LS residuals and the true error terms gonverges to zero, so the sample covariance of residuas times regressors goes to the same plim as the sample covariance of true errors times regressors, which converges to the true covariances. Observe that the blocks in the main diagonal are just Eicker-White applied to each equation separatelly.

So we can rewrite the result as:

$$\sqrt{N}(\widehat{\beta} - \beta) \longrightarrow_d N \left(0, p\lim \left(\frac{X'X}{N} \right)^{-1} \left(\frac{X'\widehat{\Omega}X}{N} \right) \left(\frac{X'X}{N} \right)^{-1} \right)$$

Now, to test the hypothesis $H_0 : g(\beta_1, \beta_2) = 0$, we use the Delta Method, and we know that:

$$\sqrt{N}(g(\widehat{\beta}_1, \widehat{\beta}_2) - g(\beta_1, \beta_2)) \longrightarrow_d N \left(0, G_0 \left(p\lim \left(\frac{X'X}{N} \right)^{-1} \left(\frac{X'\widehat{\Omega}X}{N} \right) \left(\frac{X'X}{N} \right)^{-1} \right) G_0' \right)$$

with

$$G_0 = \left[\frac{\partial g}{\partial \beta_1}(\beta_1, \beta_2), \frac{\partial g}{\partial \beta_2}(\beta_1, \beta_2) \right]$$

2. Consider the usual setting, $y_{N\text{x}1}$, $X_{N\text{x}K}$, $Z_{N\text{x}L}$ with $L \geq K$ and $\hat{X}_{N\text{x}K} \equiv Z\hat{\Pi}$, $\hat{\Pi}_{L\text{x}K} = (Z'Z)^{-1}Z'X$

i)

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

ii) This is Theil's approach:

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

iii) This is Basmann's approach. Consider $\hat{\pi} = (Z'Z)^{-1}Z'y$ and $\hat{y} = Z\hat{\pi}$

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

iv) Let $\hat{V} = X - \hat{X}$, consider

$$y = X\beta + \hat{V}\gamma + \epsilon$$

partition regression implies,

$$\hat{\beta} = (X'(I - P_{\hat{V}})X)^{-1}X'(I - P_{\hat{V}})y$$

Note that,

$$\begin{aligned}
\hat{V} &= (I - P_Z)X \\
\hat{V}'X &= X'\hat{V} = X'(I - P_Z)X \\
\hat{V}'\hat{V} &= X'(I - P_Z)(I - P_Z)X \\
&= X'(I - P_Z)X
\end{aligned}$$

hence,

$$\begin{aligned}
\hat{\beta} &= (X'(I - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}')X)^{-1}X'(I - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}')y \\
\hat{\beta} &= (X'X - X'\hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'X)^{-1}(X'y - X'\hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'y) \\
\hat{\beta} &= (X'X - \hat{V}'X)^{-1}(X'y - \hat{V}'y) \\
\hat{\beta} &= (X'X - X'(I - P_Z)X)^{-1}(X'y - X'(I - P_Z)y) \\
\hat{\beta} &= (X'P_ZX)^{-1}(X'P_Zy)
\end{aligned}$$

v)

$$\begin{aligned}
\hat{\beta} &= (X'(I - P_{\hat{V}})X)^{-1}X'(I - P_{\hat{V}})P_Zy \\
\hat{\beta} &= (X'P_ZX)^{-1}X'(I - P_{\hat{V}})P_Zy \\
\hat{\beta} &= (X'P_ZX)^{-1}(X'P_Zy - X'P_{\hat{V}}P_Z)y \\
\hat{\beta} &= (X'P_ZX)^{-1}(X'P_Zy)
\end{aligned}$$

where the last equality comes from,

$$X'P_{\hat{V}}P_Z = X'\hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'P_Z = 0$$

since $\hat{V}$ are the residuals of the regression of X on Z, hence $\hat{V} \perp Z$

Empirical Question

OLS regression of Months Worked on

Variable	Bols	std errors	t-test
cons	11.7477	0.3781	31.0703
Monthly earn	0.2711	0.2596	1.0441
Fam size	-0.0528	0.0733	-0.7207

R2 0.01618

IV estimation of Months Worked on

Variable	Biv	std errors	t-test
cons	11.4204	0.4636	24.6346
Monthly earn	0.7668	0.4720	1.6247
Fam size	-0.0542	0.0747	-0.7257

Pseudo R2 0.03290

R2s original variables on all the instruments:

Monthly earnings	0.3140
Family size	1.0000

Comments

First, both regressions show very low R^2's, so the amount of variance explained for the models is very small. The R^2 for IV estimation is higher than the OLS regression one. In the IV estimation, the standard errors increases and the t-test decrease except for monthly earnings.

According to the plausibility of the exclusion restrictions there are several considerations. First, when doing 2SLS estimations, note that the matrix $\hat{X}$ is the projections of the columns of X onto the columns space of Z. This is important because this avoid us to have conformability problems, and more important identification problems. Now, if there is some column of Z uncorrelated with the matrix X, the inclusion of this instrument is as to include an irrelevant variable in the first stage. Therefore, depending on the correlation of the instrument (irrelevant) with the error term we will have inefficient estimations. Recall a previous problem set question related to this.

Second, in this particular exercise, we have that the relevant variable that we are "instrumentalizing" is monthly earnings. This is because in the set of instruments we are incorporating a constant and Family size that are included also in the X matrix. Therefore, the projection of these columns (constant and family size) of X onto the column space of Z will give us the same two columns in $\hat{X}$. Now, according to the *Becker-Mincer* human capital view, age, education and race would explain monthly earnings. Actually, Mincer includes age indirectly when he propose the "potential labor market experience" defined as $(Age - Education - 6)$. Now, economically speaking the inclusion of these variables as instrument would be correct. In addition, the relation between monthly earnings and Month worked is apparently from monthly earnings to month worked. The participation rate varies over time and depends on income among other variables. This is the typical income effect in the labor supply. So, if an individual earns big amounts of money per month, she can adjust her labor supply by cutting the number of month worked, assuming a very flexible labor market. Then I do not think that monthly earnings are endogenous.

Third, econometrically speaking it seems that the set of instruments is not too bad. We can observe that the t-test for monthly earnings improves a lot in the 2SLS equation but it is not enough to determine significance. Actually, at a 12% level the monthly earning parameter is significant in the 2SLS. Regarding to the explanatory power of the instruments on monthly earnings it is not bad: $R^2 = 31.4\%$ (The usual R^2 in Mincer equations). Now, even though the bias of IV estimators is directly correlated to the fit of the explanatory variables on those instruments, this can be reduced in 2SLS by dropping irrelevant instruments.

Matlab Code

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Tomas Rau                                        %%
%% Econ 240b                                        %%
%% Problem Set 3, Empirical question                %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% load Earnings.txt;
x2=Earnings(:,2); x3=Earnings(:,3);
x4=Earnings(:,4); x7=Earnings(:,7); x6=Earnings(:,6);
x9=Earnings(:,9); x96=x9./x6; on=ones(size(x6,1),1);
 x=[on x96 x2]; [n,k]=size(x);
b=x x6; % a shortcut for the normal equations, returns the ols e=x6-x*b;
%predicted value of error
s2=(e'*e)/(n-k); cov=s2*inv(x'*x); r2=1-(e'*e)/((n-1)*var(x6));
clc; fprintf(' tOLS regression of Months Worked on  n n')
fprintf(' tVariable            Bols    std errors   t-test n')
fprintf(' tcons              %.4f   %.4f       %.4f  n',b(1),
sqrt(cov(1,1)),b(1)/sqrt(cov(1,1)))
fprintf(' tMonthly earn      %.4f    %.4f       %.4f  n',b(2),
sqrt(cov(2,2)),b(2)/sqrt(cov(2,2)))
fprintf(' tFam size         %.4f    %.4f      %.4f  n n',b(3),
sqrt(cov(3,3)),b(3)/sqrt(cov(3,3)))
fprintf(' tR2       %.5f n n',r2);
%instruments
z=[on x2 x3 x4 x7]; pi_hat=inv(z'*z)*z'*x; x_hat=z*pi_hat;
biv=x_hat x6;
ei=x6-x*biv; %predicted value of error
s2i=(ei'*ei)/(size(x,1)-size(x,2));
covi=inv(pi_hat'*z'*x)*pi_hat'*(z'*z)*pi_hat*inv(x'*z*pi_hat)*s2i;
%Pseudo R2
ef=x6-x_hat*biv; r2_i=1-(ef'*ef)/((n-1)*var(x6));
% R2's of each original variables on all the instruments
ei1=x(:,2)-x_hat(:,2); r2_i1=1-ei1'*ei1/((n-1)*var(x(:,2)));
ei2=x(:,3)-x_hat(:,3); r2_i2=1-ei2'*ei2/((n-1)*var(x(:,3)));
fprintf(' tIV estimation of Months Worked on  n n')
fprintf(' tVariable            Biv     std errors   t-test n')
fprintf(' tcons              %.4f   %.4f       %.4f  n',biv(1),
sqrt(covi(1,1)),biv(1)/sqrt(covi(1,1)))
fprintf(' tMonthly earn      %.4f    %.4f       %.4f  n',biv(2),
sqrt(covi(2,2)),biv(2)/sqrt(covi(2,2)))
fprintf(' tFam size         %.4f    %.4f      %.4f  n n',biv(3),
sqrt(covi(3,3)),biv(3)/sqrt(covi(3,3)))
fprintf(' tPseudo R2   %.5f n n',r2_i);
fprintf(' tR2s original variables on all the instruments:  n');
fprintf(' t    Monthly earnings %.4f n',r2_i1);
fprintf(' t    Family size      %.4f n',r2_i2);
```

E-mail address: `carolina@econ.berkeley.edu`

ANSWER KEY