## Suggested Answers Problem Set 1 <br> ECOE 60303

Bill Evans
Spring 2014

1. Using partitioned inverses,
$\hat{\beta}_{1}=\left[X_{1}{ }^{\prime}\left(I-X_{2}\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime}\right) X_{1}\right]^{-1} X_{1}{ }^{\prime}\left(I-X_{2}\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime}\right) Y$.
Consider a regression of $X_{2}$ on $X_{1}$ where $X_{1}=X_{2} \gamma+v$ and $\hat{\gamma}=\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime} X_{1}$. If $\hat{\gamma}=0$, the $\mathrm{X}_{2}$ on $\mathrm{X}_{1}$ are uncorrelated. Work with the term in brackets in the portioned inverse, $X_{1}{ }^{\prime}\left(I-X_{2}\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime}\right) X_{1}$ can be re-written to read $X_{1}{ }^{\prime} X_{1}-X_{2}\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime} X_{1}$. The second term is really just $X_{2}\left(X_{2}^{\prime} X_{2}\right)^{-1} X_{2}^{\prime} X_{1}=X_{2} \hat{\gamma}$ which equals zero because $\mathrm{X}_{2}$ on $\mathrm{X}_{1}$ are uncorrelated. This means the first term collapses to $\left(X_{1}{ }^{\prime} X_{1}\right)^{-1}$.

The second term in the portioned inverse $X_{1}{ }^{\prime}\left(I-X_{2}\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime}\right) Y$ can be re-written to read $X_{1}{ }^{\prime} Y-X_{1}{ }^{\prime} X_{2}\left(X_{2}{ }^{\prime} X_{2}\right)^{-1} X_{2}{ }^{\prime} Y$ which can be re-written as $X_{1}{ }^{\prime} Y-\hat{\gamma}^{\prime} X_{2}{ }^{\prime} Y$ which collapses to $X_{1}{ }^{\prime} Y$ because $\hat{\gamma}=0$.
2. This is a dumb idea. The suggested regression of $\hat{\varepsilon}=X \pi+\mathrm{v}$ will produce an estimate of $\hat{\pi}=\left(X^{\prime} X\right)^{-1} X^{\prime} \hat{\varepsilon}$ but from the initial estimate of $\hat{\beta}=\left(X^{\prime} X\right)^{-1} X^{\prime} Y$ we know that by construction, $X^{\prime} \hat{\varepsilon}=0$ in the sample, so for all models $\hat{\pi}=0$. The suggested regression is not informative.
3. Given $\hat{\beta}^{*}=\left(x^{\prime} x\right)^{-1} x^{\prime} y^{*}$ substitute in the true value of $y^{*}$ which equals $y^{*}=y+v=x \beta+\varepsilon+v$ which produces $\hat{\beta}^{*}=\left(x^{\prime} x\right)^{-1} x^{\prime} y^{*}=\left(x^{\prime} x\right)^{-1} x^{\prime}(x \beta+\varepsilon+v)$ which simplifies to read $\hat{\beta}^{*}=\beta+\left(x^{\prime} x\right)^{-1} x^{\prime}(\varepsilon+v)$. Taking expectation, we obtain

$$
E\left[\hat{\beta}^{*} \mid x\right]=\beta+E\left[\left(x^{\prime} x\right)^{-1} x^{\prime}(\varepsilon+v) \mid x\right]=\beta+\left(x^{\prime} x\right)^{-1} x^{\prime}(E[\varepsilon \mid x]+E[v \mid x])=\beta
$$

Because $E[v \mid x]=0 E[\varepsilon \mid x]=0$
Even with measurement error in $y$, the estimate is still unbiased.
We know that $\operatorname{Var}(\hat{\beta})=\sigma_{\varepsilon}^{2}\left(x^{\prime} x\right)^{-1}$
For the estimate with measurement error, note that $V\left[\hat{\beta}^{*} \mid x\right]=E\left[\left(\hat{\beta}^{*}-E\left[\hat{\beta}^{*}\right]\right)\left(\hat{\beta}^{*}-E\left[\hat{\beta}^{*}\right]\right)\right]^{\prime}=E\left[\left(\hat{\beta}^{*}-\beta\right)\left(\hat{\beta}^{*}-\beta\right) '\right]=E\left[\left(x^{\prime} x\right)^{-1} x^{\prime}(\varepsilon+v)\left(\varepsilon^{\prime}+v^{\prime}\right) x\left(x^{\prime} x\right)^{-1}\right]$

Note that
$E\left[(\varepsilon+v)\left(\varepsilon^{\prime}+v^{\prime}\right)\right]=E\left[\varepsilon \varepsilon^{\prime}+v v^{\prime}+2 \varepsilon v^{\prime}\right]$
and by assumption
$E\left[2 \varepsilon v^{\prime}\right]=0$
While
$E\left[\varepsilon \varepsilon^{\prime}+v v^{\prime}\right]=\left(\sigma_{\varepsilon}^{2}+\sigma_{v}^{2}\right)\left(x^{\prime} x\right)^{-1}$
Therefore
$\operatorname{Var}\left(\hat{\beta}^{*}\right)=\left(\sigma_{\varepsilon}^{2}+\sigma_{v}^{2}\right)\left(x^{\prime} x\right)^{-1}>\operatorname{Var}(\hat{\beta})=\sigma_{\varepsilon}^{2}\left(x^{\prime} x\right)^{-1}$
4. The correlation coefficient between $y$ and $\hat{y}$ is $\hat{\rho}(y, \hat{y})=\frac{\left[y^{\prime} M \hat{y} /(n-1)\right]}{\left[y^{\prime} M y /(n-1)\right]^{0.5}\left[\hat{y}^{\prime} M \hat{y} /(n-1)\right]^{0.5}}$

Note that the ( $\mathrm{n}-1$ )'s in the numerator and denominator cancel so the squared correlation coefficient is then $\hat{\rho}(y, \hat{y})^{2}=\frac{\left[y^{\prime} M \hat{y}\right]\left[y^{\prime} M \hat{y}\right]}{\left[y^{\prime} M y\right]\left[\hat{y}^{\prime} M \hat{y}\right]}$. Note further that because $y=x \hat{\beta}+\hat{\varepsilon}$
$y^{\prime} M \hat{y}=(\hat{y}+\hat{\varepsilon})^{\prime} M \hat{y}=\hat{y}^{\prime} M \hat{y}+\hat{\varepsilon}^{\prime} M \hat{y}$ and since $\hat{y}=x \hat{\beta}$ then $\hat{\varepsilon}^{\prime} M \hat{y}=\hat{\varepsilon}^{\prime} M x \hat{\beta}=\hat{\varepsilon}^{\prime} x \hat{\beta}=0$ because $X^{\prime} \hat{\varepsilon}=0$. Therefore $\hat{\rho}(y, \hat{y})^{2}=\frac{\left[y^{\prime} M \hat{y}\right]\left[y^{\prime} M \hat{y}\right]}{\left[y^{\prime} M y\right]\left[\hat{y}^{\prime} M \hat{y}\right]}=\frac{\left[\hat{y}^{\prime} M \hat{y}\right]\left[\hat{y}^{\prime} M \hat{y}\right]}{\left[y^{\prime} M y\right]\left[\hat{y}^{\prime} M \hat{y}\right]}=\frac{\left[\hat{y}^{\prime} M \hat{y}\right]}{\left[y^{\prime} M y\right]}=R^{2}$
5. This is trivial. Reduce this to a bivariate regression model. $y_{i t}=\alpha+x_{i t} \beta+u_{i}+\varepsilon_{i t}$. Note that in first difference $y_{i 2}-y_{i 1}=\Delta y_{i}=\left(x_{i 2}-x_{i 1}\right) \beta+\left(\varepsilon_{i 2}-\varepsilon_{i 1}\right)=\Delta x_{i} \beta+\Delta \varepsilon_{i}$. The OLS estimate is then $\hat{\beta}=\frac{\sum_{i=1}^{n} \Delta y_{i} \Delta x_{i}}{\sum_{i=1}^{n} \Delta x_{i}^{2}}$. If there is no within-panel variation in x , then $\Delta x_{i}=0$ and by including the observation in the model, neither the numerator or the denominator is altered at all.
6. If the data is sorted by year then group, the first N observations in D will be an identify matrx $\mathrm{I}_{\mathrm{n}}$. This will be repeated T times, do as a result, the matrix D will look like $D=i_{t} \otimes I_{n}$
$M=I_{n t}-D\left(D^{\prime} D\right)^{-1} D^{\prime}$
$\left(D^{\prime} D\right)^{-1}=\left(\left(i_{t}{ }^{\prime} \otimes I_{n}\right)\left(i_{t} \otimes I_{n}\right)\right)^{-1}=\left(i_{t}{ }^{\prime} i_{t} \otimes I_{n}\right)^{-1}=\left(t \otimes I_{n}\right)^{-1}=\frac{1}{t} I_{n}$
$D\left(D^{\prime} D\right)^{-1} D^{\prime}=\left(i_{t} \otimes I_{n}\right)\left[\frac{1}{t} \otimes I_{n}\right]\left(i_{t}{ }^{\prime} \otimes I_{n}\right)=\left(\frac{i_{t}}{t} \otimes I_{n}\right)\left(i_{t}{ }^{\prime} \otimes I_{n}\right)=\left(\frac{i_{t} i_{t}{ }^{\prime}}{t} \otimes I_{n}\right)$
$M=I_{n t}-\left(\frac{i_{t} i_{t}{ }^{\prime}}{t} \otimes I_{n}\right)$

Which can be re-written to read

$$
M=\left(I_{t}-\frac{i_{t} i_{t}{ }^{\prime}}{t}\right) \otimes I_{n}
$$

7. A program is available on the web page - ps1_q7.do
b. Note that the sample means of within panel deviations in means are all zero out to about 6 decimal placed. Since all the x's and y's have zero mean, you should not need to include an intercept in these regressions. The standard errors in this part are so much smaller because they od not take into consideration that we SHOULD be estimating dummy variables for all groups. The DOF are NT-K but they should be NT-N-K.
c. The coefficient is 0.048 when means that workers make $4.8 \%$ more when then move from a non-union to a union job.
d. Note that the coefficient stays the same - the fixed effect estimate does not utilize information from people who do not change union status - which is the majority of people.
e. Although we delete people without within panel variation in union status the estimates from this question and part b differ because union status is correlated with other variables in the model
