# Problem Set #6

Danny Edgel
Econ 710: Economic Statistics and Econometrics II
Spring 2021

March 1, 2021

*Discussed and/or compared answers with Sarah Bass, Emily Case, Katherine Kwok, Michael Nattinger, and Alex Von Hafften*

## Question 1

(i) The direct representation of the sample average is:

$$\mu_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=1}^{T_i} Y_{it}$$

Since $1_i$ contains $T_i$ elements, $1_i'1_i = T_i$, and $1_i'Y_i = \sum_{t=1}^{T_i} Y_i$. It is clear, then, that

$$\mathbb{E}\left[\hat{\mu}_{OLS}\right] = \frac{\sum_{i=1}^{n} 1_i'Y_i}{\sum_{i=1}^{n} 1_i'1_i}$$

(ii) We can solve for the variance of $\hat{\mu}_{IV}$ as follows:

$$Var(\hat{\mu}_{IV}) = Var\left(\frac{\sum_{i=1}^{n} Z_i'Y_i}{\sum_{i=1}^{n} Z_i'1_i}\right) = \frac{\sum_{i=1}^{n} Z_i'Var(Y_i)Z_i}{\left(\sum_{i=1}^{n} Z_i'1_i\right)^2}$$

Where:

$$\begin{aligned} &Var(Y_i) = Var(\mu_0 + \alpha_i + \varepsilon_{it}) = Var(\alpha_i) + Var(\varepsilon_{it}) + 2Cov(\alpha_i, \varepsilon_{it}) \\ &\quad \Rightarrow \Omega_i = \sigma_\alpha^2 1_i 1_i' + \sigma^2 I_{T_i} \end{aligned}$$

(iii) To determine how we can show that $Var(\hat{\mu}_{IV} \geq \left(\sum_{i=1}^{n} 1_i'\Omega_i^{-1}1_i\right)^{-1}$, we simply need to find that the following inequality holds:

$$\left(\sum_{i=1}^{n} Z_i'1_i\right)^2 \leq \left(\sum_{i=1}^{n} Z_i'\Omega_i Z_i\right)\left(\sum_{i=1}^{n} 1_i'\Omega_i^{-1}1_i\right)$$

Once this inequality is established, it follows that:

$$Var(\hat{\mu}_{IV}) = \frac{\sum_{i=1}^{n} Z_i' Var(Y_i) Z_i}{\left(\sum_{i=1}^{n} Z_i' 1_i\right)^2} \geq \frac{\sum_{i=1}^{n} Z_i' Var(Y_i) Z_i}{\left(\sum_{i=1}^{n} Z_i' \Omega_i Z_i\right)\left(\sum_{i=1}^{n} 1_i' \Omega_i^{-1} 1_i\right)} = \left(\sum_{i=1}^{n} 1_i' \Omega_i^{-1} 1_i\right)^{-1}$$

We can establish the inquality using the Cauchy-Schwarz inequality:

$$\left(\sum_{i=1}^{n} Z_i' 1_i\right)^2 = \left(\sum_{i=1}^{n} Z_i' \Omega^{1/2} \Omega^{-1/2} 1_i\right)^2 \leq \left(\sum_{i=1}^{n} Z_i' \Omega_i Z_i\right)\left(\sum_{i=1}^{n} 1_i' \Omega_i^{-1} 1_i\right)$$

This variance is achieved by $\overline{Z}_i = \Omega_i^{-1} 1_i$

(iv) If $T_i = T$ for all $i$ (i.e., the panel is balanced), then the GLS estimator weights by the entries of $\Sigma^{-1}$, where:

$$\Sigma^{-1} = \frac{1}{\sigma^2}\left(I_T - \frac{\sigma_\alpha^2 T}{\sigma^2 + \sigma_T^2} \frac{1_i 1_i'}{T}\right)$$

Then, the optimal instrument for GLS is:

$$\overline{Z}_i = \Sigma^{-1} 1_i = \frac{1_i}{\sigma^2}\left(1 - \frac{\sigma_\alpha^2 T}{\sigma^2 + \sigma_T^2}\right) = \frac{1_i}{\sigma^2 + \sigma^2 T}$$

This instrument cancels out in the estimator for $\mu_0$, yielding the OLS estimator:

$$\hat{\mu}_{GLS} = \frac{\sum_{i=1}^{n} \overline{Z}_i' Y_i}{\sum_{i=1}^{n} \overline{Z}_i' 1_i} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma^2 + \sigma_T^2} 1_i' Y_i}{\sum_{i=1}^{n} \frac{1}{\sigma^2 + \sigma_T^2} 1_i' 1_i} = \frac{\sum_{i=1}^{n} 1_i' Y_i}{\sum_{i=1}^{n} 1_i' 1_i} = \hat{\mu}_{OLS}$$

Thus, if the panel is balanced, OLS and GLS are identical.

(v) First, note that:

$$\overline{Y} = \frac{1}{T_i} \sum t = 1^{T_i} \mu_0 + \alpha_i + \varepsilon_{it} = \mu_0 + \alpha_i + \frac{1}{T_i} \sum t = 1^{T_i} \varepsilon_{it}$$

And let $\overline{\varepsilon} = \frac{1}{T_i} \sum t = 1^{T_i} \varepsilon_{it}$. Then,

$$
\begin{aligned}
\mathbb{E}\left[\hat{\sigma}_i^2\right] &= \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \mathbb{E}\left[(\mu_0 + \alpha_i + \varepsilon_{it} - \mu_0 - \alpha_i - \overline{\varepsilon})^2\right] \\
&= \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \mathbb{E}\left[(\varepsilon_{it} - \overline{\varepsilon})^2\right] \\
&= \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \mathbb{E}\left[(\varepsilon_{it} - \overline{\varepsilon})\varepsilon_{it}\right] \\
&= \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \mathbb{E}\left[\varepsilon_{it}^2\right] - \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \mathbb{E}\left[\varepsilon_{it}\overline{\varepsilon})\right] \\
&= \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \mathbb{E}\left[\varepsilon_{it}^2\right] - \frac{1}{T_i(T_i - 1)} \sum_{t=1}^{T_i} \mathbb{E}\left[\varepsilon_{it}^2\right] - \frac{1}{T_i(T_i - 1)} \sum_{t=1}^{T_i} \sum_{s \neq t}^{T_i} \mathbb{E}\left[\varepsilon_{it}\varepsilon_{is}\right] \\
&= \frac{1}{T_i - 1} \sum_{t=1}^{T_i} \sigma^2 - \frac{1}{T_i(T_i - 1)} \sum_{t=1}^{T_i} \sigma^2 \\
&= \sigma^2
\end{aligned}
$$

(vi)

(vii)

# Question 2

(i)

(ii)

# Question 3

The table below presents the results of the simulations.[1] As you can see, the pooled OLS estimator of $\beta_0$ is substantially biased upward. An intuitive but less clearly apparent result is that both the fixed effect estimation of $\beta_0$ is generally efficient but much more efficient for larger $n$ and no autocorrelation.

[1] See the attached .do file for the code used to generate this table.