POLS 607 — Instructor: Casey Crisman-Cox
Panel Data — Spring 2025

Problem Set 1: Answers

This problem set is due February 6th. Upload your write up and any code files to the Github Classroom before midnight. You may work together, but you must turn in separately written (unique) write ups and/or code.

1. Consider the random effects model under the following assumptions:

   RE.1 $y_{it} = \alpha_i + \beta' x_{it} + \gamma' z_i + \varepsilon_{it}$

   RE.2 $\mathrm{E}[\varepsilon_{it}\varepsilon_{js}|\mathbf{X}_i] = 0$, $\forall i \neq j$ and $\forall t \neq s$

   RE.3 $\mathrm{E}[\varepsilon_{it}|\mathbf{X}_i] = \mathrm{E}[\alpha_i|\mathbf{X}_i] = 0$

   RE.4 $\mathrm{E}[\alpha_i \varepsilon_i|\mathbf{X}_i] = 0$

   RE.5 $\mathrm{E}[\alpha_i^2|\mathbf{X}_i] = \sigma_\alpha^2$

   RE.6 $\mathrm{E}[\varepsilon_{it}^2|\mathbf{X}_i] = \sigma_\varepsilon^2$

   where $\mathbf{X}_i = [X_i, Z_i]$ and $X_i = \begin{bmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{bmatrix}$. You may assume you have a balanced panel.

   (a) In your own words, briefly describe what restrictions assumptions RE.2–RE.6 impose on the data.

   Assumption **RE.2** is an independence assumption on the data. This rules out correlations in the errors. Assumption **RE.3** imposes exogeneity on both the error term and the unobserved heterogeneity. This means that anything unobserved is strictly uncorrelated with the observables within each unit. Assumption **RE.4** says that the error term and the unobservered heterogeneity are independent of each other. This is another type of exogeneity assumption, but now it is on the unobserved heterogeneity. This, along with Assumption **RE.2**, tells us that the only unobserved correlation within units comes from the heterogeneity shock. Assumptions **RE.5–6** impose homoscedasticity on the two unobserved terms. This, along with Assumption **RE.2** tells us that the combined shocks are iid both within and across units.

(b) Show that the pooled OLS estimator is consistent for $\theta = (\beta', \gamma')'$ in N for above model and derive its asymptotic distribution. Explain where you use each assumption, but you do not need to list any additional technical moment assumptions. Is it asymptotically efficient? Why or why not?

The pooled OLS estimator for this model is

$$\hat{\theta}_p = \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}\mathbf{x}_{it}' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} y_{it}$$
$$= \theta + \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}\mathbf{x}_{it}' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} [\alpha_i + \varepsilon_{it}] \quad \textbf{RE.1}.$$

Taking this in parts, we find that

$$\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}\mathbf{x}_{it}' \right) \xrightarrow{p} \mathrm{E}\left[T^{-1} \mathbf{X}_i' \mathbf{X}_i \right] \qquad \text{LLN}$$
$$\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}\mathbf{x}_{it}' \right)^{-1} \xrightarrow{p} \mathrm{E}\left[T^{-1} \mathbf{X}_i' \mathbf{X}_i \right]^{-1} \qquad \text{Continuous mapping theorem}$$
$$\left(\frac{1}{N} \sum_{i=1}^{N} a_i \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_{it} \right) \xrightarrow{p} 0 \qquad \textbf{RE.3} \text{ and the LLN}$$
$$\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} \varepsilon_{it} \right) \xrightarrow{p} 0 \qquad \textbf{RE.3} \text{ and the LLN}$$

We can combine these using Slutsky's theorem to get

$$\hat{\theta}_p \xrightarrow{p} \theta + \mathrm{E}\left[T^{-1} \mathbf{X}_i' \mathbf{X}_i \right]^{-1} 0 + \mathrm{E}\left[T^{-1} \mathbf{X}_i' \mathbf{X}_i \right]^{-1} 0 = \theta.$$

This tells us that the pooled estimator is consistent under these assumptions. It's asymptotic distribution is then based on

$$\sqrt{N} \left(\frac{1}{N} \sum_{i=1}^{N} a_i \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_{it} \right) \xrightarrow{d} N(0, \mathrm{Var}(a_i \bar{\mathbf{x}}_i))$$
$$\sqrt{N} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} \varepsilon_{it} \right) \xrightarrow{d} N(0, \mathrm{Var}(\bar{\mathbf{x}}_i \bar{\varepsilon}_i)),$$

where

$$\begin{aligned}
\operatorname{Var}(a_i\bar{\mathbf{x}}_i) &= \mathrm{E}[\alpha_i^2\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i'] - \mathrm{E}[\alpha_i\bar{\mathbf{x}}_i]\,\mathrm{E}[\alpha_i\bar{\mathbf{x}}_i]' \\
&= \mathrm{E}[\alpha_i^2\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i'] - 0 && \textbf{RE.3} \\
&= \sigma_\alpha^2 E[T^{-2}\mathbf{X}_i'\mathbf{X}_i] - 0 && \textbf{RE.3} \text{ and } \textbf{RE.5}
\end{aligned}$$

and likewise for $\operatorname{Var}(\bar{\mathbf{x}}_i\bar{\varepsilon}_i)$. Combining terms and applying Slutsky's theorem we get

$$\begin{aligned}
\sqrt{N}(\hat{\theta}_p - \theta) &\xrightarrow{d} N\left(0, (\sigma_\varepsilon^2 + \sigma_\alpha^2)\,\mathrm{E}[\mathbf{X}_i'\mathbf{X}_i]^{-1}\right) \\
\hat{\theta}_p &\overset{\text{asym}}{\sim} N\left(0, (\sigma_\varepsilon^2 + \sigma_\alpha^2)(\mathbf{X}'\mathbf{X})^{-1}\right)
\end{aligned}$$

Note that this will not be asymptotically efficient as the pooled estimator is not asymptotically equivalent to the GLS estimator under these conditions unless $\sigma_\alpha^2 = 0$. By the properties of GLS, we know that it will be the best linear unbiased estimator. As the pooled estimator is an unbiased linear estimator that is not the GLS, then the pooled we know it is not the best.

(c) Show that applying the RE-GLS transformation to this model leads to an error term with mean 0 and variance 1. The algebra here can be a little tedious, so feel free to use a computer aid like Mathematica, sympy, Maple, or Matlab's symbolic math toolbox.[1] Include any code you use.

The RE transformation is given as

$$\begin{aligned}
\frac{1}{\sigma_\varepsilon}\,\mathrm{E}[\alpha_i(1-\omega) + \varepsilon_{it} - \omega\bar{\varepsilon}_i] &= \frac{1}{\sigma_\varepsilon}\left(\mathrm{E}[\alpha_i](1-\omega) + \mathrm{E}[\varepsilon_{it}] - \omega\,\mathrm{E}[\bar{\varepsilon}_i]\right) \\
&= 0 \\
\frac{1}{\sigma_\varepsilon^2}\operatorname{Var}(\alpha_i(1-\omega) + \varepsilon_{it} - \omega\bar{\varepsilon}_i) &= \frac{1}{\sigma_\varepsilon^2}\Big(\operatorname{Var}(\alpha_i)(1-\omega)^2 + \operatorname{Var}(\varepsilon_{it}) \\
&\qquad + \omega^2\operatorname{Var}(\bar{\varepsilon}_i) - 2\omega\operatorname{Cov}(\varepsilon_{it}, \bar{\varepsilon}_i)\Big) \\
&= \frac{1}{\sigma_\varepsilon^2}\left(\sigma_\alpha^2(1-\omega)^2 + \sigma_\varepsilon^2 + \omega^2\sigma_\varepsilon^2/T - 2\omega\sigma_\varepsilon^2/T\right) \\
&= 1
\end{aligned}$$

[1]I encourage you to use computer algebra systems for algebra and calculus steps. Sympy is free and open source, see `https://www.sympy.org/en/index.html`. Maple and Matlab freely available to students at TAMU, see `https://software.tamu.edu/`.

2. Consider the fixed effects model under the following assumptions:

FE.1 $y_{it} = \alpha_i + \beta' x_{it} + \varepsilon_{it}$

FE.2 Each unit (x_i, ε_i) is drawn from an iid process

FE.3 $\mathrm{E}[X_i' \varepsilon_{it}] = 0$

FE.4 $\mathrm{E}[X_i' \,\mathrm{E}\,[T^{-1} X_i' X_i]]$ exists and has full rank

FE.5 $\mathrm{E}[\varepsilon_{it}^2 | \mathbf{X}_i] = \sigma_\varepsilon^2$

Show the following:

(a) Show that pooled OLS is biased under these assumptions and derive the bias in terms supposing that $\mathrm{E}[\alpha_i | x_{it}] = \gamma_i' z_i$. Explain why a random effects estimator would also be biased. Can you say anything about which would be more biased?

Under these assumptions, the expected value of the pooled estimator is

$$\mathrm{E}[\hat{\theta}_p] = \theta + \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x_{it}' \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \,\mathrm{E}[\alpha_i | x_{it}].$$

Because we have not assumed that $\mathrm{E}[\alpha_i | x_{it}] = 0$, the pooled estimator will be biased to the extent that this expected value differs from 0.

To derive the bias, we have

$$\mathrm{E}[\hat{\theta}_p] - \theta = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x_{it}' \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \,\mathrm{E}[\alpha_i | x_{it}]$$
$$\mathrm{E}[\hat{\theta}_p] - \theta = (X'X)^{-1} X' Z \gamma$$

which is to say it will be equal to the regression coefficients of all the omitted factors contained in α on X times the direct effect of these factors on y.

A random effects estimator will also be biased as it also requires that $\mathrm{E}[\alpha_i | x_{it}] = 0$ to be unbiased. It will be weakly less biased than the pooled estimator as it is a weighting between the biased pooled and unbiased within estimators.

(b) Let M_i be with within unit demeaning matrix such that $M_i = I_T - 1_T(1_T' 1_T)^{-1} 1_T'$ where I is an identity matrix and 1 is a column vector of 1s. Show that

$$X_i' X_i \geq X_i' M_i X_i.$$

Note that to show that one matrix is weakly less than other, you'll want to show that $X_i' X_i - X_i' M_i X_i$ is positive semi-definite.

HINT 1: The trick here is work it into quadratic form, which is to say something that looks like ABA and then show that B is positive semi-definite (it has all weakly positive eigenvalues).

HINT 2: M_i and $I - M_i$ are both idempotent. All the eigenvalues of an idempotent matrix are either 0 or 1.

We can rewrite the inequality to be

$$\begin{aligned} X_i' X_i - X_i' M_i X_i &\geq 0 \\ X_i'(I - M_i) X_i &\geq 0 \end{aligned}$$

Note that this is in quadratic form, and will be positive semi-definite if $(I - M_i)$ is positive semi-definite. Because $(I - M_i)$ is idempotent it has eigenvalues that are all 0 or 1 (i.e., weakly positive) and so $X_i'(I - M_i)X_i$ is positive semi-definite, which is what we had to show.

(c) Use your analysis from the above to argue that under the above assumptions the within estimator of β has a weakly larger variance than the pooled estimator

$$\mathrm{Var}(\hat{\beta}_w | X_i) \geq \mathrm{Var}(\hat{\beta}_p | X_i).$$

What additional assumption to do you need for this inequality to hold? What does this tell us about the choice between the pooled and the within estimators?

Writing out the variances for these estimators under FE assumptions we get

$$\mathrm{Var}(\hat{\beta}_w | X_i) = \sigma_\varepsilon^2 \left[\sum X_i' M_i X_i \right]^{-1}$$

$$\mathrm{Var}(\hat{\beta}_p | X_i) = (\mathrm{Var}(a_i | X_i) + \sigma_\varepsilon^2 + 2\,\mathrm{Cov}(a_i, \varepsilon_{it})) \left[\sum X_i' X_i \right]^{-1}$$

From the above, we know that $X_i' M_i X_i < X_i' X_i$ and so $\sum X_i' M_i X_i < \sum X_i' X_i$ as both of these are the sums of positive semi-definite matrices, and this leads to

$$\left[\sum X_i' M_i X_i\right]^{-1} > \left[\sum X_i' X_i\right]^{-1}.$$

In order for the main inequality to be true, it would be sufficient for

$$\operatorname{Var}(a_i|X_i) + 2\operatorname{Cov}(a_i, \varepsilon_{it}) \leq 0.$$

This will be the case if each α_i is a fixed parameter. If they are draws from a random variable, then this sufficient condition holds if α_i and ε_{it} have a covariance less than $-\operatorname{Var}(a_i|X_i)/2$. Otherwise, the main inequality will only hold if the $\operatorname{Var}(a_i|X_i) + 2\operatorname{Cov}(a_i, \varepsilon_{it})$ is less than the difference between $(\sum X_i' M_i X_i)^{-1}$ and $(\sum X_i' X_i)^{-1}$.

What this tells us is that if the α_i are fixed, then there is a bias-variance trade-off between pooled OLS and the within estimator under the FE assumptions. If the α are random, but correlated with x_{it}, then the within estimator is still unbiased while the pooled estimator is not, but the variance differences are less clear.

3. Consider the following model

$$\begin{aligned}
y_{it} &= \alpha_i + \beta' x_{it} + \gamma' z_i + \varepsilon_{it} & & \\
x_{it} &= [x_{it}^1, x_{it}^2] & z_i &= [z_i^1, z_i^2] \\
\alpha_i &\sim U[-1, 1] & \theta &= (\beta', \gamma')' = (1, 4, -3, 5)' \\
x_{it}^1 &= 15 - 3a_i - 2z_i^2 + 0.25x_{it-1}^1 + x_{it}^{1*} & x_{it}^{1*} &\sim N(0, 1/4) \\
x_{it}^2 &\sim \text{Poisson}(5) & & \\
z_i^1 &\sim N(0, 4) & z_i^2 &\sim N(0, 1/4) \\
\varepsilon_{it} &= 0.8\varepsilon_{it-1} + u_{it} & u_{it} &\sim N(0, 7) \\
x_{i0}^1 &= 15 - 3a_i - 2z_i^2 & \varepsilon_{i0} &= 0 \\
i &= 1, \ldots, N & t &= 1, \ldots, T_i.
\end{aligned}$$

Here α_i and z_2 are assumed to be unobserved to the analyst. Note that the above normal random variables are specified as $N(\mu, \sigma^2)$.

(a) Which of the various A assumptions in the course notes are satisfied in the this model? Which are not?

This model satisfies Assumptions **A1.F**, and **A2-5**. The random effects assumptions **A2.R-A5.R** are all met except for $\mathrm{E}[\alpha_i|x_i, z_i] \neq 0$ (there is correlation between α_i and x^1_{it}) and $\mathrm{E}[\varepsilon_{it}\varepsilon_{js}|x_i, z_i] \neq 0$ for any $t \neq s$ when $i = j$ (there is within-unit correlation in the errors). Assumption **A6.R** is violated as α_i are distributed uniformly, not normally. Assumptions **A1** are **A1.R** partially violated in the sense that z_2 is unobserved.

(b) Based on your answer to part (a), if you were interested in the effect of x^1 on y, what would be an appropriate estimator and specification? What would be an appropriate variance estimator?

Only a fixed effects estimator (e.g., within, LSDV, or Mundlak) would be appropriate here given the correlation of x^1 and α_i. I would include x^2 to improve the variance of the estimates, but it is not necessary to include it from a bias perspective. The clustered variance matrix would be appropriate here.

(c) What if you were only interested in the effect of x_2 on y? How would you specify the model? What estimator would you use for β^2? What would be an appropriate variance estimator?

Because x_2 is fully exogenous, this effect can be found using the pooled, random effects, or within estimator. Under these assumptions, I would use pooled estimator with clustered standard errors as it is unbiased and consistent for x_2. I would control for z_1 and x_1 to improve variance.

(d) Find the "true" values of σ^2_ε and σ^2_α under this data generating process. Note that in this case σ^2_α should include the unobserved z_2. What do they tell you about how similar/different the RE estimates will be from the pooled and within estimates?

HINT 1: you can rewrite ε_{it} to be

$$\varepsilon_{it} = \sum_{s=0}^{t} (\varepsilon_{i0} + u_{it-s})(0.8)^s, \quad u_{i0} = 0$$

which will make it a lot easier to work with expectations and variances. First apply the expectation/variance operator to both sides, and then take $t \to \infty$ (since it's stationary and ergodic we can treat any slice we observe as part of an infinite chain). **I switched the order of these steps, they are backwards, but it doesn't really matter for this exercise.**

HINT 2: Recall that $\sum_{t=0}^{\infty} ar^t = \frac{a}{1-r}$ for $|r| < 1$ and constant a.

The true value of σ_ε^2 is given by

$$\begin{aligned}
\text{Var}(\varepsilon_{it}) &= \text{Var}\left(\sum_{s=0}^{t}(u_{it-s})(0.8)^s\right) \\
&= \sum_{s=0}^{\infty} \text{Var}\left((u_{is})(0.8)^s\right) \\
&= \sum_{s=0}^{\infty} \text{Var}\left(u_{is}\right)(0.8)^{2s} \\
&= \sum_{s=0}^{\infty} 7(0.8)^{2s} \\
&= \sum_{s=0}^{\infty} 7(0.8^2)^{s} \\
&= \frac{7}{1-0.64} \approx 19.4.
\end{aligned}$$

Likewise for σ_α^2 we get

$$\begin{aligned}
\text{Var}(\alpha_i + 5z_i^2) &= \text{Var}(\alpha_i) + \text{Var}(5z_i^2) \\
&= \frac{2^2}{12} + 25(1/4) \\
&\approx 6.58
\end{aligned}$$

Based on these values, we would except an average ω_i of about 0.67 (i.e., when $T_i = 25$). In this case, I would except notable bias in the RE estimator, and as such, more in the pooled estimator.

(e) Find the correlation between x_{it}^1 and the unobserved heterogeneity $\alpha_i + 5z_i^2$.

Recall that

$$\text{cor}(x_{it}^1, \alpha_i + 5z_i^2) = \frac{\text{Cov}(x_{it}^1, \alpha_i + 5z_i^2)}{\text{sd}(x_{it}^1)\text{sd}(\alpha_i + 5z_i^2)}.$$

We will consider these in parts. For the covariance we get

$$\begin{aligned}
\operatorname{Cov}(x_{it}^1, \alpha_i + 5z_i^2) &= \operatorname{E}[x_{it}^1\alpha_i + x_{it}^1 5z_i^2] - \operatorname{E}[x_{it}^1]\operatorname{E}[\alpha_i + 5z_i^2] \\
&= \operatorname{E}[x_{it}^1\alpha_i] + \operatorname{E}[x_{it}^1 5z_i^2] \\
&= \operatorname{E}\left[\alpha_i \sum_{s=0}^{t} (x_{i0}^1 + x_{it-s}^*)(0.25)^s\right] + \operatorname{E}\left[5z_i^2 \sum_{s=0}^{t} (x_{i0}^1 + x_{it-s}^*)(0.25)^s\right] \\
&= \sum_{s=0}^{\infty} \operatorname{E}\left[\alpha_i x_{i0}^1\right](0.25)^s + \sum_{s=0}^{\infty} \operatorname{E}[\alpha_i x_{is}^*](0.25)^s \\
&\quad + 5\sum_{s=0}^{\infty} \operatorname{E}[z_i^2 x_{i0}^1](0.25)^s + 5\sum_{s=0}^{\infty} \operatorname{E}[z_i^2 x_{is}^*](0.25)^s \\
&= \frac{\operatorname{E}[\alpha_i x_{i0}^1] + 5\operatorname{E}[z_i^2 x_{i0}^1]}{1-0.25} \\
&= \frac{\operatorname{E}[\alpha_i(15 - 3a_i - 2z_i^2)] + 5\operatorname{E}[z_i^2(15 - 3a_i - 2z_i^2)]}{1-0.25} \\
&= \frac{-3\sigma_\alpha^2 - 10\sigma_{z2}^2}{1-0.25} = \frac{-3.5}{3/4} \approx -4.67.
\end{aligned}$$

For the standard deviation of x_{it}^1,

$$\begin{aligned}
\operatorname{sd}(x_{it}^1) &= \left(\operatorname{Var}\left(\sum_{s=0}^{t} (x_{i0}^1 + x_{it-s}^*)(0.25)^s\right)\right)^{1/2} \\
&= \left(\operatorname{Var}\left(x_{i0}^1 \sum_{s=0}^{\infty} (0.25)^s\right) + \operatorname{Var}\left(\sum_{s=0}^{\infty} x_{is}^*(0.25)^s\right)\right)^{1/2} \\
&= \left(\operatorname{Var}\left(x_{i0}^1(1-0.25)^{-1}\right) + \sum_{s=0}^{\infty} (0.25)^{2s}\operatorname{Var}\left(x_{is}^*\right)\right)^{1/2} \\
&= \left((1-0.25)^{-2}\operatorname{Var}\left(x_{i0}^1\right) + \sum_{s=0}^{\infty} (0.25^2)^s(1/4)\right)^{1/2} \\
&= \left((1-0.25)^{-2}\operatorname{Var}\left(x_{i0}^1\right) + (1/4)(1-(0.25^2))^{-1}\right)^{1/2} \\
&= \left((1-0.25)^{-2}(9\sigma_\alpha^2 + 4\sigma_{z2}^2) + (1/4)(1-(0.25^2))^{-1}\right)^{1/2} \\
&= \left((1-0.25)^{-2}(4) + (1/4)(1-(0.25^2))^{-1}\right)^{1/2} \approx 2.71,
\end{aligned}$$

and finally for the standard deviation of the unobserved heterogeneity

$$\mathrm{sd}(\alpha_i + 5z_i^2) = \sqrt{\sigma_\alpha^2 + 25\sigma_{z2}^2} \approx 2.6.$$

As such the correlation is about

$$\begin{aligned}\mathrm{cor}(x_{it}^1, \alpha_i + 5z_i^2) &\approx \frac{-4.67}{2.71(2.6)} \\ &\approx -0.67\end{aligned}$$

(f) Conduct a Monte Carlo experiment with this data generating process and

i. $N \in \{50, 200\}$ (the number of U.S. states and roughly the number of countries in the world)
ii. $T_i - 1 \sim \mathrm{Poisson}(24)$

Code and implement your own versions of the RE-GLS, RE-MLE (with gradients), the within, and the Mundlak estimators. By your own version, I mean that `lm` and `optim` are fine, but I don't want to see `lmer` or `feols` here. The only packages I'd like to see would be for data manipulation, graphics, or tabling (e..g, `dplyr`, `data.table`, `ggplot`, `xtable`, `matrixStats`), but you can complete this assignment without loading any packages. If you decide to parallelize your simulations, you may use any packages you like for that.

The gradients for the RE-MLE are given by

$$\begin{aligned}
D_\theta L(\theta|y) &= \sum_{i=1}^{N} \frac{(\varepsilon_i - \bar{\varepsilon}_i\omega_i)'(\mathbf{x}_i - \bar{\mathbf{x}}_i\omega_i)}{\sigma_\varepsilon^2} \\
D_{\sigma_\alpha^2} L(\theta|y) &= \frac{1}{2}\sum_{i=1}^{N} \frac{-T_i\bar{\varepsilon}_i \sum_t (\varepsilon_{it} - \bar{\varepsilon}_i\omega_i)}{\sqrt{\sigma_\varepsilon^2}\,(\sigma_i^2)^{\frac{3}{2}}} - \frac{T_i}{\sigma_i^2} \\
D_{\sigma_\varepsilon^2} L(\theta|y) &= \frac{1}{2}\sum_{i=1}^{N} \frac{-\bar{\varepsilon}_i \sum_t (\varepsilon_{it} - \bar{\varepsilon}_i\omega_i)\left(\sqrt{\sigma_\varepsilon^2}\,(\sigma_i^2)^{-3/2} - (\sigma_\varepsilon^2\sigma_i^2)^{-1/2}\right)}{\sigma_\varepsilon^2} \\
&\quad + \frac{(\varepsilon_i - \bar{\varepsilon}_i\omega_i)'(\varepsilon_i - \bar{\varepsilon}_i\omega_i)}{(\sigma_\varepsilon^2)^2} - \frac{T_i - 1}{\sigma_\varepsilon^2} - \frac{1}{\sigma_i^2},
\end{aligned}$$

where

$$\sigma_i^2 = T_i \sigma_\alpha^2 + \sigma_\varepsilon^2$$
$$\varepsilon_{it} = y_{it} - \mathbf{x}_{it}'\theta.$$

When conducting your simulation, use a 25 period burn-in in generating x^1 and ε to ensure they are not too tied to the initial conditions. Use the specification(s) you described in part (b). Build your simulations to answer the following questions:

i. Is the pooled estimator biased for β_1? Does it roughly match what you would expect for omitted variable bias based on your answer to question 1?
The pooled estimator is biased. Table 1 lists the average pooled estimate of β_1 as about 0.4 regardless of N. The expected omitted variable bias based on the simulated data is about -0.6 in both cases or an expected value of about 0.4, which is nearly exactly what we observe.

ii. Is the RE-GLS estimator biased for estimating β_1? Does the RE-MLE do any better?[2] Do you notice any differences in the RE estimates of β_1? If so, what might explain it?
Both RE estimators are also biased, these results are listed in Table 1. Both return an average estimate of around 0.6 regardless of N. I do not see any major differences between the two approaches here.

iii. Is the within estimator biased for estimating β_1?
The within estimator is unbiased under these assumptions and this results is borne out as shown in Table 1. On average, the within estimates are right around 1 which is the true value.

iv. For the RE-GLS, pooled, and within estimators, calculate the standard error on $\hat{\beta}_1$ using both classical and clustered variance estimator. For the clustered variance you can either estimate the asymptotic variance or conduct an appropriate bootstrap. Which one better reflects the observed uncertainty in your estimates across simulations? Why?

[2]Use the MLE estimator we described in class that assumes that the unobserved heterogeneity is normal.

For the pooled and within estimators we see that the clustered variance matrix provides a value much closer the observed spread across simulations. This is most notable with the pooled estimator. Interestingly, see little change with in the RE estimators. This is interesting, because there is definitely still serial correlation in the residuals, but it seems that the RE transformation, which includes the autocorrelation due to the unobserved heterogeneity, accommodates enough of that to limit the observed differences here. To the extent that there are differences, however, the clustered variance matrix provides a more accurate description of the uncertainty.

v. Conduct the Hausman test for comparing RE-GLS with the within estimator. Does a rejection reflect statistical size or power in this simulation? Also analyze the performance of the Wald test we described on the Mundlak-specified model fit with OLS and clustered standard errors.
   In these simulations, the null hypothesis for the Hausman and this Wald test are false. Rejections of these nulls are correct and the observed rejection rates are estimates of the power of these tests.

vi. Conduct a second simulation that changes the DGP to satisfy the random effects assumptions. In this simulation record only the Hausman test results and the Wald test on the Mundlak model results. Does a rejection here reflect size or power?
   Under the RE assumptions, the Hausman and Mundlak nulls are true and so a rejection is a false positive. The observed rejection rates are now an estimate of each test size.

vii. Using the above results on size and power, compare the Hausman test to a Wald test. Which has better properties in these simulations? Is it surprising?
   In both cases, we find that the Hausman test has better power, but not by much. This surprised me as it uses variance matrices that are not correct when the null is false and we are looking at cases where the null is false. Making the Wald test robust, may reduce its power as accounting for the increased uncertainty makes it more difficult to distinguish the estimates. The size of the tests are interesting, the Hausman

test is not particularly close to the expected value of 0.05, although the Wald test when $N = 50$. As N increases, neither produces any false rejections in these simulations. Although the power of the Hausman test in this situation is appealing.

Table 1: Different approaches for estimating $\beta_1 = 1$

		Pooled	Within	RE-MLE	RE-GLS
N=50	Average estimate	0.40	0.99	0.61	0.61
	Observed st. dev.	0.13	0.25	0.14	0.14
	Classical st. err.	0.05	0.20	0.12	0.12
	Clustered st. err.	0.14	0.25	0.14	0.14
N=200	Average estimate	0.37	1.00	0.61	0.61
	Observed st. dev.	0.06	0.13	0.07	0.07
	Classical st. err.	0.03	0.10	0.07	0.07
	Clustered st. err.	0.08	0.13	0.07	0.07

Table 2: Comparing panel specification tests

		Hausman	Mundlak-Wald
N=50	Power	0.579	0.505
	Size	0.001	0.043
N=200	Power	0.993	0.991
	Size	0	0