1. Given \( Y_i = \alpha + X_i \beta + u_i + v_i + \epsilon_i \), but only two observations per block, generate the difference

\[
Y_{i2} = \alpha + X_{i2} \beta + u_i + v_2 + \epsilon_{i2} \\
-Y_{i1} = \alpha + X_{i1} \beta + u_i + v_1 + \epsilon_{i1} \\
\Delta Y_i = \Delta X_i \beta + (v_2 - v_1) + \Delta \epsilon_i
\]

Where \( \theta = v_2 - v_1 \) in the estimating equation \( \Delta Y_i = \theta + \Delta X_i \beta + \Delta \epsilon_i \), and this represents the growth in outcomes between years 1 and 2.

To figure out what \( \Delta Y_i \) is estimating, write the original model as

\[
Y_{it} = \alpha + X_{it} \beta + u_i + v_{it} + \lambda_i \text{time}_i + \epsilon_{it}
\]

Where \( \text{time}_i \) is a time trend that equals 1 in period 1, 2 in period 2, etc. The equation allows for a unique time trend within a particular state. The new model can be written as

\[
Y_{it} = \alpha + X_{it} \beta + u_i + v_{it} + \lambda_{it} + \epsilon_{it}
\]

And the first difference between any two adjacent years is

\[
Y_{it} = \alpha + X_{it} \beta + u_i + v_{it} + \lambda_{it} + \epsilon_{it}
\]

\[
Y_{it-1} = \alpha + X_{it-1} \beta + u_i + v_{it-1} + \lambda_{i(t-1)} + \epsilon_{it-1}
\]

\[
\Delta Y_{it} = \Delta X_{it} \beta + (v_{it} - v_{it-1}) + \lambda_{it} + \Delta \epsilon_{it}
\]

And the estimating equation is \( \Delta Y_{it} = \theta + \Delta X_{it} \beta + \lambda_{it} + \Delta \epsilon_{it} \). In this equation, the year effects pick up growth rates between adjacent years and \( \lambda_{it} \) picks up the state-specific time trend.

2. Given the equation \( y_{it} = \alpha_{it} + x_{it} \beta + u_i + \lambda_i + \epsilon_{it} \)

- take the mean over all \( t \) for person \( i \): \( \bar{y}_i = \alpha_i + \bar{x}_i \beta + \bar{u}_i + \bar{\lambda}_i + \bar{\epsilon}_i \) where \( \bar{\lambda}_i = \frac{1}{t_i} \sum_{t=1}^{t_i} \lambda_{it} \). Note that the means of the year fixed effects differ across \( y \) because each panel has different numbers of observations.

- take the mean over all \( n \) for year \( t \): \( \bar{y}_t = \alpha + \bar{x}_t \beta + \bar{\mu}_t + \bar{\lambda}_t + \bar{\epsilon}_t \) where \( \bar{\mu}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} u_i \) where \( n_t \) is the number of panels that are observed in time period \( t \). Note that the means of the individual effects will vary by \( t \) because each year has different numbers of observations.

- take the mean over observations: \( \bar{y} = \alpha + \bar{x} \beta + \bar{\mu} + \bar{\lambda} + \bar{\epsilon} \). Let \( M \) be the number of observations in the sample and hence \( \bar{\lambda} = \frac{1}{m} \sum_{i=1}^{m} \sum_{t=1}^{t_i} \lambda_{it} \) and \( \bar{\mu} = \frac{1}{m} \sum_{i=1}^{m} \sum_{t=1}^{t_i} u_i \).
Construct the difference \( \tilde{y}_{it} = y_{it} - \bar{y}_{i} - \bar{y}_{t} + \bar{y} \). Because the observations vary by panel and the number of panels within a year vary as well, the differencing does not eliminate the means of the group and year effects and as a result, 
\[
\tilde{y}_{it} = (x_{it} - \bar{x}_{i} - \bar{x} + \bar{x})\beta - (\bar{y}_{i} + \bar{y}) - (\bar{u}_{i} + \bar{u}) + (\varepsilon_{it} - \bar{\varepsilon}_{i} - \bar{\varepsilon} + \bar{\varepsilon})
\]

3. a. When data is sorted by year then group. Let \( Y_i = X_i\beta + V_i \) be the \( n \) observations for year \( t \). In this case, \( E[V'_V] \) is \((n x n)\) and equals \( \Omega_n = \sigma^2_n I_n + \sigma^2_i i'_i \) and then \( E[V'V] \) which is \((nt x nt)\) is \( I_t \otimes \Omega_n \).

b. When data is sorted by group then year. \( Y_i = X_i\beta + V_i \) be the \( t \) observations for group \( i \). In this case, \( E[V'_V] \) is \((t x t)\) and equals \( \Omega_t = (\sigma^2_t + \sigma^2_i) I_t \). However, the 1st observation in panel 1 will be correlated with the 1st observation in panel 2, panel 3, etc. Therefore, \( E[V'V] \) can be generalized to read

\[
E[V'V] = \begin{bmatrix}
\sigma_i + \sigma^2_i & 0 & \ldots & 0 \\
0 & \sigma_i + \sigma^2_i & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_i + \sigma^2_i
\end{bmatrix}
\begin{bmatrix}
\sigma_i & 0 & \ldots & 0 \\
0 & \sigma_i & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_i
\end{bmatrix}
\begin{bmatrix}
\sigma_i & 0 & \ldots & 0 \\
0 & \sigma_i & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_i
\end{bmatrix}
\]

Which can be re-written as

\[
E[\tilde{V}' \tilde{V}] = \sigma^2_n I_n + (i_n i'_n \otimes \sigma^2_i I_i).
\]

5. a. Putting the results from class and the previous problem together, it is clear that

Therefore \( E[V'V] = \sigma^2_n I_n + (i_n i'_n \otimes \sigma^2_i I_i) + (I_n \otimes \sigma^2_u i'_u) = \Omega \)

b. \( \hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) \)

6. Note that \( \bar{y}^{00} = \alpha + \bar{e}^{00} \) and \( \bar{y}^{01} = \alpha + \theta + \bar{e}^{01} \). Because \( y_i = \alpha + G_i\gamma + A_i\theta + D_i\beta + \varepsilon_i \) then \( y_i \) or only the observations for the treatment group become \( y_i = \alpha + \gamma + A_i\theta + A_i\beta + \varepsilon_i \). because \( A_i = D_i \) in this sample. Let \( y_{i1}^{10} \) be the observations for the treatment group before the intervention and \( y_{i1}^{11} \) be the observations after the intervention. Then
\[
\hat{\gamma}^1 = [(1 - A_1)\bar{y}^{00} + A_1\bar{y}^{01}] \text{ can be written as } (1 - A_1)(y_i^{10} - \bar{y}^{00}) + A_1(y_i^{11} - \bar{y}^{01}). \] Let's take each of these terms separately.

\[
(y_i^{10} - \bar{y}^{00}) = \alpha + \gamma + \epsilon_i - \alpha - \bar{\epsilon}^{00} = \gamma + (\epsilon_i - \bar{\epsilon}^{00})
\]

\[
(y_i^{11} - \bar{y}^{01}) = \alpha + \gamma + \theta + \epsilon_i - \alpha - \bar{\epsilon}^{01} = \gamma + \beta + \epsilon_i - \bar{\epsilon}^{01}
\]

Therefore

\[
y_i^1 - [(1 - A_1)\bar{y}^{00} + A_1\bar{y}^{01}] = (1 - A_1)(y_i^{10} - \bar{y}^{00}) + A_1(y_i^{11} - \bar{y}^{01}) = (1 - A_1)[\gamma + (\epsilon_i - \bar{\epsilon}^{00})] + A_1[\gamma + \beta + \epsilon_i - \bar{\epsilon}^{01}]
\]

Which can be reduced to read

\[
y_i^1 = \gamma + A_1\beta + \epsilon_i - (1 - A_1)\bar{\epsilon}^{00} - A_1\bar{\epsilon}^{01} = \gamma + A_1\beta + u_i
\]

7. There are two things wrong with the numbers. We know that random effects are a weighted average of between and within group estimates so they random effect estimates should lie between the two values. In this case it does not. Second, random effect is efficient estimation and the variance should be smaller than the fixed-effect estimate. Again, it is not.

The Hausman test is \( \hat{q} = (\hat{\beta}_{fe} - \hat{\beta}_{re})^T\left[\text{Var}(\hat{\beta}_{fe}) - \text{var}(\hat{\beta}_{re})\right]^{-1}(\hat{\beta}_{fe} - \hat{\beta}_{re}) \) but when the parameter is a scaler, the test statistic reduces to

\[
\hat{q} = (\hat{\beta}_{fe} - \hat{\beta}_{re})^2 / \left[\text{Var}(\hat{\beta}_{fe}) - \text{var}(\hat{\beta}_{re})\right] = (0.04813 - 0.0635)^2 / [(0.02429)^2 - (0.02137)^2] = 1.77
\]

The p-value for a chi-square distribution with 1 degree of freedom evaluated at 1.77 is 0.18, so we cannot reject the null that the random and fixed-effect estimates are the same.

**Computer problem**

A sample program (psid_measurement_error.do) that generates these results is on the web and the some key results from that are reported below. In the OLS model, the attenuation bias is estimated to be \( \hat{\theta} = \frac{\hat{\sigma}_\tau^2}{\hat{\sigma}_\tau^2 + \sigma^2} \), which is exactly what you find. In the fixed-effects case (first difference with two observations in the panel only), with two observations, we showed in class that the attenuation bias is \( 1 - \frac{\sigma^2}{\sigma_\tau^2(1 - \rho) + \sigma^2} \) where \( \rho \) is the correlation in tenure over time. To estimate \( \hat{\rho} \) estimated a random effects model and calculated the fraction of variance that is due to within-panel correlation which is \( \hat{\rho} = 0.89 \)

\[
\hat{\sigma}_\tau^2 = (6.91)^2 = 47.75
\]

\[
\hat{\beta}_{w/out}^{ols} = 0.0238 \text{ and } \hat{\beta}_{w/out}^{fe} = 0.0081
\]

\[
\bar{\sigma}_\tau^2 = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\sigma}_\tau^2(i) = 1.999^2 = 4
\]

\[
\bar{\beta}_{w/\text{with}}^{OLS} = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_{w/\text{with}}^{OLS}(i) = 0.0219
\]

\[
\bar{\beta}_{w/\text{with}}^{fe} = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_{w/\text{with}}^{fe}(i) = 0.00455
\]
\[\hat{\theta} = \frac{\hat{\sigma}_i^2}{\bar{\hat{\sigma}}_i^2 + \hat{\sigma}_i^2} = \frac{47.75}{4 + 47.75} = 0.92\]

\[\frac{\bar{\beta}_{\text{OLS}}^{\text{with}}}{\bar{\beta}_{\text{OLS}}^{\text{w/out}}} = \frac{0.0219}{0.0238} = 0.92\]

\[\frac{\bar{\beta}_{\text{FE}}^{\text{with}}}{\bar{\beta}_{\text{FE}}^{\text{w/out}}} = \frac{0.00455}{0.00809} = 0.563\]

\[1 - \frac{\sigma_v^2}{\sigma_i^2(1 - \rho) + \sigma_v^2} = 1 - \frac{4}{47.75(1 - 0.89) + 4} = 0.568\]

. **************************************;
. * part a --- get the variance of tenure;
. **************************************;
. sum wagel tenure;

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>wagel</td>
<td>1578</td>
<td>2.330113</td>
<td>.5435628</td>
<td>.7558519</td>
<td>4.299593</td>
</tr>
<tr>
<td>tenure</td>
<td>1578</td>
<td>7.6109</td>
<td>6.913697</td>
<td>0</td>
<td>33</td>
</tr>
</tbody>
</table>

. **************************************;
. * part h --- estimate random effects model;
. * get within panel correlation in tenure;
. * that is sigma u squared;
. **************************************;
. iis id;
.
. sort id year;
.
. xtreg tenure;

Random-effects GLS regression

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of obs = 1578</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of groups = 789</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-sq: within = 0.0000
between = 0.0000
overall = 0.0000

Observations: min = 2
avg = 2.0
max = 2

Random effects u_i ~ Gaussian
Wald chi2(0) = 0.00
Prob > chi2 = .

| tenure | Coef.    | Std. Err. |     z   |   P>|z|   | [95% Conf. Interval] |
|--------|----------|-----------|---------|--------|----------------------|
| _cons  | 7.6109   | .2394447  | 31.79   | 0.000  | 7.141597 - 8.080203  |

| sigma_u | 6.5304204 |               |         |       |                      |
| sigma_e | 2.2759307 |               |         |       |                      |
| rho     | .8916942  | (fraction of variance due to u_i) | | | |

. **************************************;
. * part b --- get OLS and FE estimates;
. **************************************;
. * regression of ln(wagel) on tenure;
. reg wagel tenure;

    Source |        SS      df       MS
----------|----------|----------|----------
 Model    | 42.663768  1 42.663768
 Residual | 423.277514 1576 .268577103
----------|----------|----------|----------
 Total    | 465.941283 1577 .295460547
----------|----------|----------|----------
F(  1,  1576) = 158.85
Prob > F = 0.0000
R-squared = 0.0916
Adj R-squared = 0.0910
Root MSE = .51824

. * regression of ln(wagel) on tenure;
. areg wagel tenure, absor

Linear regression, absorbing indicators Number of obs = 1578
F(  1,   788) =   4.91
Prob > F = 0.0270
R-squared = 0.9077
Adj R-squared = 0.8153
Root MSE = .2336

. qui save `main', replace ;
. use psid_measurementerror_results;

. sum;

    Variable |    Obs   Mean     Std. Dev.     Min    Max
-------------|--------|--------------|--------------|--------|--------
     v2_m    | 1000   -.0000245  .0506558  -.164754  .1637709
     v2_sd    | 1000   1.999933   .0356311   1.892301  2.10275
    beta_ols  | 1000   0.0219413  .005363   0.0203601  0.0238582
    beta_fe   | 1000   0.0045495  .0018241  -.0007662  .0108822