Microeconometrics
Supervision 1

Samuel Lee

Question 1

Part I

(a)

The question says there is no data on $\boldsymbol{x}_{it}$, so I think (though I'm not sure) what we really want is an expression for $\hat{\delta}$. Assuming we don't include a constant, we have

$$\hat{\delta} = \arg\min_{\delta} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\log(wage)_{it} - \delta prog_{it}\right]^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \log(wage_{it}) prog_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T} prog_{it}^2}$$

using the first-order condition. This is just the pooled OLS estimator when no constant is included.

(b)

The steps to get the first-difference and fixed effects estimators are the same. To get to the first-difference case, we substitute $\Delta \log(wage_{it}) = \log(wage_{it}) - \log(wage_{it-1})$ for $\log(wage_{it})$, and $\Delta prog_{it} = prog_{it} - prog_{it-1}$ for $prog_{it}$. This yields

$$\hat{\delta}_{FD} = \frac{\sum_{i=1}^{N} \left[\log(wage_{i2}) - \log(wage_{i1})\right] (prog_{i2} - prog_{i1})}{\sum_{i=1}^{N} (prog_{i2} - prog_{i1})^2}$$

In the fixed-effects case, we substitute $\log(wage_{it})^{FE} = \log(wage_{it}) - \frac{1}{T} \sum_{t=1}^{T} \log(wage_{it})$ for $\log(wage_{it})$ and $prog_{it}^{FE} = prog_{it} - \frac{1}{T} \sum_{t=1}^{T} prog_{it}^{FE}$. This yields

$$\hat{\delta}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{2} \left\{\log(wage_{it}) - \frac{1}{2} \left[\log(wage_{i1}) + \log(wage_{i2})\right]\right\} \left[prog_{it} - \frac{1}{2} \left(prog_{i1} + prog_{i2}\right)\right]}{\sum_{i=1}^{N} \sum_{t=1}^{2} \left[prog_{it} - \frac{1}{2} \left(prog_{i1} + prog_{i2}\right)\right]^2}$$

The numerator is equal to

$$\sum_{i=1}^{N}\left\{\frac{1}{2}[\log(wage_{i1})-\log(wage_{i2})]\frac{1}{2}(prog_{i1}-prog_{i2})\right.$$
$$\left.+\frac{1}{2}\left[\log(wage_{i2})-\log(wage_{i1})\right]\frac{1}{2}(prog_{i2}-prog_{i1})\right\}$$
$$=\frac{1}{2}\sum_{i=1}^{N}\left[\log(wage_{i2})-\log(wage_{i1})\right](prog_{i2}-prog_{i1})$$

and the denominator is equal to

$$\sum_{i=1}^{N}\left\{\left[\frac{1}{2}(prog_{i1}-prog_{i2})\right]^2+\left[\frac{1}{2}(prog_{i2}-prog_{i1})\right]^2\right\}=\frac{1}{2}\sum_{i=1}^{N}(prog_{i2}-prog_{i1})^2$$

and substituting them back into the expression gives us the first-difference estimator.

Part II

(a)

Assuming the model is true, for consistency we require $\mathbb{E}[\varepsilon_{is}|prog_{it}]$ for all s and t.

(b)

Strict exogeneity requires that the expectation of the error term in any period, conditional on the regressors also in any period, must be equal to its unconditional expectation (zero in this case). This might be violated in the given context: for example, a negative wage shock yesterday could be correlated with participation in the program tomorrow, which means $\mathbb{E}[\varepsilon_{t-1}|prog_{it}=1]<0$.

(c)

If wages are non-stationary, and if wages are I(1) in particular, a first-difference estimate might be preferred to avoid a spurious regression. If $T=2$ then this will be equal to the fixed effects estimator, but the fixed effects transformation will not generally get rid of nonstationarity.

Part III

In this context I think $\omega=\sigma_c^2/\sigma_c^2+\sigma_\varepsilon^2$, which should really be denoted as $\sigma_c^2/(\sigma_c^2+\sigma_\varepsilon^2)$, is actually supposed to be $\frac{\sigma_c^2}{\sigma_c^2+\sigma_u^2}$. This would be the intraclass correlation coefficient:

$$\frac{\mathrm{Cov}[\varepsilon_t,\varepsilon_{t-\tau}]}{\mathrm{Var}[\varepsilon_t]}=\frac{\sigma_c^2}{\sigma_c^2+\sigma_u^2}$$

assuming there are no time-specific effects. The observed estimate is very close to zero, which suggests the true model might be closer to one with fixed effects than to one with random effects. However, the true model might also be one without any individual-specific effects, in which case OLS is efficient.

Question 2

Part I

We have

$$
\begin{aligned}
\mathrm{Var}[\boldsymbol{w}_i] &= \mathbb{E}[\boldsymbol{w}_i \boldsymbol{w}_i'] \\
&= \mathbb{E}\left[(\alpha_i \boldsymbol{i}_T + \varepsilon_i)(\alpha_i \boldsymbol{i}_T + \varepsilon_i)'\right] \\
&= \mathbb{E}[\alpha_i^2 \boldsymbol{i}_T \boldsymbol{i}_T' + \alpha_i \boldsymbol{i}_T \varepsilon_i' + \alpha \varepsilon_i \boldsymbol{i}_T' + \varepsilon_i \varepsilon_i'] \\
&= \sigma_\alpha^2 \boldsymbol{i}_T \boldsymbol{i}_T' + \mathbb{E}[\varepsilon_i \varepsilon_i'] \\
&= \sigma_\alpha^2 \boldsymbol{i}_T \boldsymbol{i}_T' + \sigma_\varepsilon^2 \boldsymbol{I}_T \\
&= \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \cdots & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}
\end{aligned}
$$

Part II

We can represent the random effects model in matrix form:

$$
\boldsymbol{y}_i = \lambda \boldsymbol{i}_T + \boldsymbol{X}_i \boldsymbol{\beta} + \alpha_i \boldsymbol{i}_T + \varepsilon_i = \lambda \boldsymbol{i}_T + \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{w}_i
$$

where $\boldsymbol{X}_i$ is a $T \times k$ matrix with row t given by $\boldsymbol{x}_{it}'$, and $\boldsymbol{y}_i$ is a $T \times 1$ vector. Denoting the $T \times T$ covariance matrix for $\boldsymbol{w}_i$ derived above by Ω, which is real and symmetric and therefore has an inverse, we have

$$
\Omega^{-\frac{1}{2}} \boldsymbol{y}_i = \lambda \Omega^{-\frac{1}{2}} \boldsymbol{i}_T + \Omega^{-\frac{1}{2}} \boldsymbol{X}_i \boldsymbol{\beta} + \Omega^{-\frac{1}{2}} \boldsymbol{w}_i
$$

And since symmetry of Ω implies its inverse is symmetric, we have

$$
\mathrm{Var}\left[\Omega^{-\frac{1}{2}} \boldsymbol{w}_i\right] = \Omega^{-\frac{1}{2}} \mathbb{E}[\boldsymbol{w}_i \boldsymbol{w}_i'] \Omega^{-\frac{1}{2}} = \boldsymbol{I}_T
$$

Therefore, the random effects estimator is akin to a generalised least squares estimator with weighting matrix $\Omega^{-\frac{1}{2}}$.

Part III

As shown above, the random effects estimator is akin to running a regression of $\Sigma^{-\frac{1}{2}} \boldsymbol{y}$ on $\Omega^{-\frac{1}{2}} \boldsymbol{X}$, where $\boldsymbol{y}$ is a $NT \times 1$ stacked vector, $\boldsymbol{X}$ is a $NT \times k$ stacked matrix, and we now define $\Sigma = \boldsymbol{I}_N \otimes \Omega$ to be a $NT \times NT$ block diagonal matrix with the Ω on the diagonals. Therefore,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{RE} &= \left[\left(\Sigma^{-\frac{1}{2}} \boldsymbol{X}\right)' \left(\Sigma^{-\frac{1}{2}} \boldsymbol{X}\right)\right]^{-1} \left(\Sigma^{-\frac{1}{2}} \boldsymbol{X}\right)' \Sigma^{-\frac{1}{2}} \boldsymbol{y} \\
&= \left(\boldsymbol{X}' \Sigma^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}' \Sigma^{-1} \boldsymbol{y}
\end{aligned}
$$

We found before that

$$\Omega = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \cdots & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

As $\sigma_\alpha^2 \to 0$, $\Omega \to \sigma_\varepsilon^2 \boldsymbol{I}_T$ and $\Sigma \to \sigma_\varepsilon^2 \boldsymbol{I}_{NT}$

$$\lim_{\sigma_\alpha^2 \to 0} \hat{\boldsymbol{\beta}}_{RE} = \left(\frac{1}{\sigma_\varepsilon^2} \boldsymbol{X}'\boldsymbol{X} \right)^{-1} \frac{1}{\sigma_\varepsilon^2} \boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y} = \hat{\boldsymbol{\beta}}_{OLS}$$

On the other hand, as $\sigma_\alpha^2 \to \infty$, $\hat{\boldsymbol{\beta}}_{RE} \to \hat{\boldsymbol{\beta}}_{FE}$, though this is a mess to show (Mundlak (1978) derived this). But if one takes this on faith then the fixed effects estimator represents another limiting case of the random effects estimator.

Question 3

The number of parameters is increasing in N, since every individual carries an individual-specific effect η_i. Therefore, the large N will not help us to estimate the η_i consistently if that's what we are concerned with.

If we only wanted to estimate α, β_0, and β_1, this would not be a problem since we can remove the individual-specific effects using either a fixed effects or first-difference transformation. But the presence of the lagged dependent variable creates problems: it implies that strict exogeneity must necessarily fail. This is because y_{it-1} contains v_{it-1}, and is therefore correlated with past values of v_{it}. To show that this creates problems for both the fixed effects and first-difference estimates, we show the transformations that are applied to the original equation:

$$\begin{aligned} \tilde{y}_{it} &= \alpha \tilde{y}_{it-1} + \beta_0 \tilde{x}_{it} + \beta_1 \tilde{x}_{it-1} + \tilde{v}_{it} \\ \Delta y_{it} &= \alpha \Delta y_{it-1} + \beta_0 \Delta x_{it} + \beta_1 \Delta x_{it-1} + \Delta v_{it} \end{aligned}$$

where $\tilde{z}_{it} = z_{it} - \frac{1}{T} \sum_{\tau=1}^{T} z_{i\tau}$. Then, for the fixed effects specification, we have

$$\mathrm{Cov}[\tilde{y}_{it-1}, \tilde{v}_{it}] = \mathrm{Cov}[\tilde{v}_{it-1}, \tilde{v}_{it}] = \mathrm{Cov}\left[v_{it-1} - \frac{1}{T} \sum_{\tau=1}^{T} v_{i\tau}, v_{it} - \frac{1}{T} \sum_{\tau=1}^{T} v_{i\tau} \right] = -\frac{1}{T}\sigma_v^2 + \frac{1}{T} \times \text{Constant}$$

where the unknown constant is there since it is said there is serial correlation in v_{it}. For the first-difference specification, we have

$$\mathrm{Cov}\left[\Delta y_{it-1}, \Delta v_{it}\right] = -\sigma_v^2 + 2\mathrm{Cov}[v_{it}, v_{it-1}] - \mathrm{Cov}[v_{it}, v_{it-2}]$$

We can see that in the two cases, the errors are correlated with a regressor, so the estimates of the parameters will be biased. In the case of the first-difference specification, the estimates are also inconsistent, given that the covariance of Δy_{it-1} and Δv_{it} is independent of T. In general, the bias shrinks at rate $\frac{1}{T}$ for the fixed effects estimator, though this may be less relevant when T is small.

The form of serial correlation in v_{it} may help us to choose between estimators. As shown above, without serial correlation in v_{it}, the covariance between $\tilde{y}_{it-1}$ and $\tilde{v}_{it}$ is smaller in magnitude than that between Δy_{it-1} and Δv_{it}, and furthermore the former shrinks with T while the latter does not.

If v_{it} follows a random walk (which means there is no cointegrating vector between the dependent variable and the regressors), then a first-difference specification helps to make the series stationary as mentioned before. Furthermore, if v_{it} follows a random walk then Δv_{it} is just white noise and we no longer have correlation between Δy_{it-1} and Δv_{it}. Therefore the first-difference estimator will be more efficient.

In reality, the true dynamics of v_{it} could be somewhere in between, and there is no easy way to determine which model is superior without further assumptions. If we know that the autocorrelation in the errors does not persist after a certain number of lags, we might be able to use the first-difference specification, except that we instrument for Δy_{it-1} using $y_{it-\tau}$ for some τ chosen such that $y_{it-\tau}$ is not correlated with the error. We can take this further and employ a GMM-style estimator using all values of y_{it} after τ lags to instrument for the first-differenced regressor.

Question 4

(a)

The partial effect of z_2 on the response probability is

$$\frac{\partial}{\partial z_2}\Pr(y=1|\boldsymbol{z}_1, z_2) = \frac{\partial}{\partial z_2}\Phi(\boldsymbol{z}_1\boldsymbol{\delta}_1 + \gamma_1 z_2 + \gamma_2 z_2^2) = \phi(\boldsymbol{z}_1\boldsymbol{\delta}_1 + \gamma_1 z_2 + \gamma_2 z_2^2)(\gamma_1 + 2\gamma_2 z_2)$$

We could estimate the model above by maximum likelihood estimation, which would allow us to find the estimated (heterogeneous) partial effect given values of $\boldsymbol{z}_1$ and z_2 by plugging our estimates of $\boldsymbol{\delta}_1$, γ_1, and γ_2 into the formula above.

(b)

Using the same steps as before, the partial effect is

$$\phi(\boldsymbol{z}_1\boldsymbol{\delta}_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1)(\gamma_1 + \gamma_3 d_1)$$

We cannot use calculus to find the effect of d_1 on the response probability since d_1 is discrete. But finding the effect is a simple matter of taking differences:

$$\Pr(y=1|\boldsymbol{z}_1, z_2, d_1 = 1) - \Pr(y=1|\boldsymbol{z}_1, z_2, d_1 = 0) = \Phi(\boldsymbol{z}_1\boldsymbol{\delta}_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1) - \Phi(\boldsymbol{z}_1\boldsymbol{\delta}_1 + \gamma_1 z_2)$$

Again, we can find estimates for the coefficients using maximum likelihood estimation, and just plug the estimates into the formula above.

If at any time we wanted to find the partial effect at the average variable values, we would just plug the average values of $\boldsymbol{z}_1$ and z_2 into the formulas above. If we wanted to find the average partial effect over all variable values, we would have to take a weighted sum of the partial effect formula evaluated at different values of $\boldsymbol{z}_1$ and z_2, with the weights being the empirical densities of $(\boldsymbol{z}_1, z_2)$.