Econometrics II (Spring 2024)
Suggested Solutions to Problem Set 4

Shuhei Kainuma*

May 8, 2024

Question 1

See the do-file for the Stata implementation.

1.

In this context, we can define that a treatment satisfies SUTVA if

1. *No hidden treatment variation*: each treatment has consistent effect on unit i, as in the standard formulation; and
2. *No interference*: $Y_{ij}(d_1, \ldots, d_J,\ w_{11}, w_{21}, \ldots, w_{12}, \ldots, w_{N_J,J}) = Y_{ij}(d_i, w_{ij})$, i.e., the potential outcome of individual i in village j only depends on d_j and w_{ij}.

The no interference condition would be violated if, for instance, there are

- spillover effects of D_j beyond village j, through, e.g., the across-village general equilibrium effects (the fiscal transfer changes the economy-wide goods price or wage);
- within-village spillover effects from the treated individuals with $W_{ij} = 1$ to the control group with $W_{ij} = 0$, through, e.g., transfers via family/friendship network.

2.

The three ATE can be expressed in the following way:

$$
\begin{aligned}
\tau_w &= \mathbb{E}[Y_{ij}(0,1) - Y_{ij}(0,0)], \\
\tau_d &= \mathbb{E}[Y_{ij}(1,0) - Y_{ij}(0,0)], \\
\tau_{dw} &= \mathbb{E}[Y_{ij}(1,1) - Y_{ij}(0,0)].
\end{aligned}
$$

*`shuhei.kainuma@iies.su.se`. Let me know if you spot any typos/errors!

We will then show that all these ATEs are identified. Because D_j and W_{ij} is randomized and so $\{D_j\}, \{W_{ij}\}$ and $\{Y_{ij}(0), Y_{ij}(1)\}$ are all independent of each other,

$$\begin{aligned}
\tau_w &= \mathbb{E}[Y_{ij}(0,1) - Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}(0,1)] - \mathbb{E}[Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}(0,1)|D_j = 0, W_{ij} = 1] - \mathbb{E}[Y_{ij}(0,0)|D_i = 0, W_{ij} = 0] \\
&= \mathbb{E}[Y_{ij}|D_j = 0, W_{ij} = 1] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0],
\end{aligned}$$

which is the difference in the CEFs of the observable outcome Y_{ij} for two groups in the population, establishing the identification of τ_w. Analogously, we have

$$\begin{aligned}
\tau_d &= \mathbb{E}[Y_{ij}(1,0) - Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}(1,0)] - \mathbb{E}[Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}(1,0)|D_j = 1, W_{ij} = 0] - \mathbb{E}[Y_{ij}(0,0)|D_i = 0, W_{ij} = 0] \\
&= \mathbb{E}[Y_{ij}|D_j = 1, W_{ij} = 0] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0], \\
\tau_{dw} &= \mathbb{E}[Y_{ij}(1,1) - Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}(1,1)] - \mathbb{E}[Y_{ij}(0,0)] \\
&= \mathbb{E}[Y_{ij}(1,1)|D_j = 1, W_{ij} = 1] - \mathbb{E}[Y_{ij}(0,0)|D_i = 0, W_{ij} = 0] \\
&= \mathbb{E}[Y_{ij}|D_j = 1, W_{ij} = 1] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0].
\end{aligned}$$

Therefore, all these ATEs are identified.

This is because both D_j and W_{ij} are randomised, the potential outcomes are balanced across each group defined by D_j and W_{ij}. Hence, the CEF of each group (D_j, W_{ij}) is equal to the population average of the corresponding potential outcome. This enables us to identify each ATE as a difference in the population means of the observed outcome Y_{ij}.

Intuitively, τ_w is identified by the differece in the outcome of the units with a cash transfer ($W_{ij} = 1$) to those without a cash transfer ($W_{ij} = 0$) within the villages without a fiscal transfer ($D_j = 0$). τ_d is identified as the outcome difference between the units in the villages with $D_j = 1$ and those in the villages with $D_j = 0$, all without the cash transfer $W_{ij} = 0$. Similarly, the difference in the consumption between the units with the cash transfer in the village with the fiscal transfer and those without the cash transfer in the villages without the fiscal transfer identifies τ_{dw}.

3.

From the identification results in Question 1.2 above, we have a population estimator for each ATE:

$$\begin{aligned}
\tau_w &= \mathbb{E}[Y_{ij}|D_j = 0, W_{ij} = 1] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0], \\
\tau_d &= \mathbb{E}[Y_{ij}|D_j = 1, W_{ij} = 0] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0], \\
\tau_{dw} &= \mathbb{E}[Y_{ij}|D_j = 1, W_{ij} = 1] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0].
\end{aligned}$$

We can write down a population regression equation to estimate these ATEs:

$$Y_{ij} = \alpha + \beta\ W_{ij}(1 - D_j) + \gamma\ (1 - W_{ij})D_j + \delta\ D_jW_{ij} + \epsilon_{ij}.$$

Then, $\beta = \tau_w,\ \gamma = \beta_d,\ \delta = \tau_{dw}$, because

$$\begin{aligned}
&\mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0] = \alpha, \\
&\mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 1] = \alpha + \beta, \\
&\mathbb{E}[Y_{ij}|D_i = 1, W_{ij} = 0] = \alpha + \gamma, \\
&\mathbb{E}[Y_{ij}|D_i = 1, W_{ij} = 1] = \alpha + \delta, \\
\Rightarrow \quad & \beta = \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 1] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0] = \tau_w, \\
& \gamma = \mathbb{E}[Y_{ij}|D_i = 1, W_{ij} = 0] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0] = \tau_d, \\
& \delta = \mathbb{E}[Y_{ij}|D_i = 1, W_{ij} = 1] - \mathbb{E}[Y_{ij}|D_i = 0, W_{ij} = 0] = \tau_{dw},
\end{aligned}$$

since $\mathbb{E}[\epsilon_{ij}|D_j, W_{ij}] = 0$ by the randomisation of D_j and W_{ij}.

4.

We consider the following population regression equation:

$$Y_{ij} = \mu + \alpha_j + \phi\ W_{ij} + \varepsilon_{ij}.$$

Fixed Effects In a fixed effects perspective, $\{\alpha_j\}_{j\in\{1,\dots,J\}}$ is a set of parameters, and we are interested in estimating them. We can consider that the villages are said to matter for consumption if the consumption levels would be different across villages; and in such a case, there exists some village j, k such that $\alpha_j \neq \alpha_k$, because the level difference would be captured by these parameters $\{\alpha_j\}_j$. Hence, a statistical test is the F-test of the joint null hypothesis $H_i : \alpha_j = 0\ \forall j \in \{1, \dots, J-1\}$, after obtaining $\{\hat{\alpha}_j\}_j$ by estimate the above equation by the OLS, for instance. Note that we exclude one α_j (i.e., normalise to $\alpha_j = 0$) to avoid multicollinearity. If we reject this null, it implies that there are systematic differences in consumption level across villages, at least one village is systematically different from the others.

Random Effects In the random effects' perspective, we consider that α_j is an i.i.d. random variable and each village α_j in the data is a realisation similar to the idiosyncratic error ε_{ij}, i.e., $(\alpha_j, \varepsilon_{ij}) \sim^{\text{i.i.d.}} F(0, \sigma^2_\alpha) \times G(0, \sigma^2_\varepsilon)$, where $\mathbb{E}[Y_{ij}] = \mathbb{E}[\mu]$, $Var(\alpha_j) = \sigma^2_\alpha$, $Cov(\alpha_j, \alpha_k) = 0\ \forall j, k$ with $j \neq k$. Then, we can consider that the villages matter for consumption if a part of the variation of the consumption across individuals is explained by a systematic variation within each villages. That is, $\sigma_\alpha > 0$. Therefore, a statistical test in this can would be the F-test with the null hypothesis $H_0 : \sigma^2_\alpha = 0$ (and $H_1 : \sigma^2_\alpha > 0$). To do this, we can estimate σ^2_α by the ANOVA while controlling for W_{ij} and treating ϕ as a fixed parameter.

5.

We can interpret the parameter ϕ as the average interaction effect of the treatments W_{ij} and D_j, i.e., $\phi = \tau_{dw} - \tau_d - \tau_w$. To see this, because $\mathbb{E}[Y_{ij}(0,0)] = 0$, we have

$$\begin{aligned} \mathbb{E}[Y_{ij}(1,1)] &= \mathbb{E}\left[Y_{ij}(1,0) + Y_{ij}(0,1) + X_{ij}\right] \\ \Leftrightarrow \underbrace{\mathbb{E}\left[Y_{ij}(1,1) - Y_{ij}(0,0)\right]}_{=\tau_{dw}} &= \underbrace{\mathbb{E}\left[Y_{ij}(1,0) - Y_{ij}(0,0)\right]}_{=\tau_d} + \underbrace{\mathbb{E}[Y_{ij}(0,1) - Y_{ij}(0,0)]}_{=\tau_w} + \underbrace{\mathbb{E}[X_{ij}]}_{=\phi} \\ \Leftrightarrow \phi &= \tau_{dw} - \tau_d - \tau_w. \end{aligned}$$

The parameter μ_j is interpreted as the village specific average treatment effect of W_{ij}, i.e., $\mathbb{E}[Y_{ij}(0,1) - Y_{ij}(0,0)|\mu_j] = \mu_j - \mathbb{E}[Y_{ij}(0,0)] = \mu_j$.

Note that from above, we have

$$\begin{aligned} \tau_d &= \mathbb{E}\left[Y_{ij}(1,0) - Y_{ij}(0,0)\right] = 1, \\ \tau_w &= \mathbb{E}\left[Y_{ij}(0,1) - Y_{ij}(0,0)\right] = \mathbb{E}[\mu_j] = 2, \\ \tau_{dw} &= \tau_d + \tau_w + \phi = 3 + \phi. \end{aligned}$$

6.

See the do-file for the implementation in Stata. Figure 1 presents the estimated coefficients on D_j and W_{ij} for each $\phi = -1, -0.8, \ldots, 0.8, 1$. The estimates are identical across a range of ϕ values and close to the true parameter values in the lower panel with the interaction term. In contrast, in the upper panel without the interaction term included, we can see that the estimates are downward biased with $\phi < 0$ while exhibiting upward bias for $\phi > 0$.[1]

Note that we cluster the standard errors at the village level, since the unit of randomisation of D_j is a village and so the intra-cluster correlation of D_j is 1, for instance. Besides, note that if the number of individuals in each village is relatively small (as in this case of $N_j = 10$), the assignment of W_{ij} is likely to be negatively correlated within a given village.[2] This adds the reason why we want to cluster at the village level.

7.

Table 1 presents the regression results with and without the village fixed effects (FE) for each value of $\phi \in \{0, 1\}$. Comparing Column (1) and (2), we can see that the estimated coefficient on W_{ij} does not change substantially. We obtain the similar result even when $\phi = 1$.

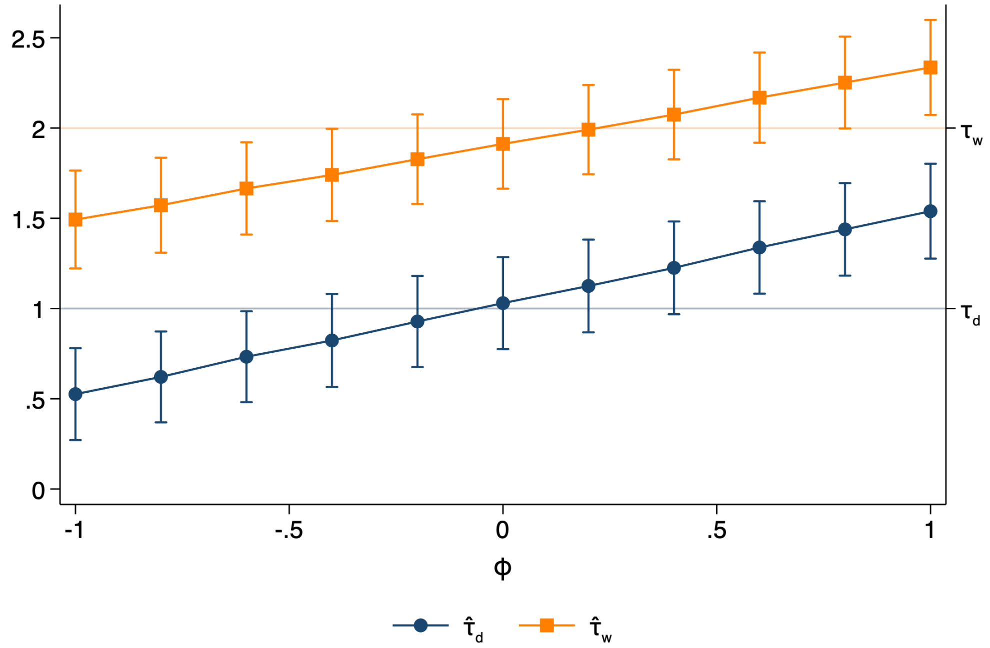
8.

Figure 2 presents the scatter plot of $\hat{\tau}_j$ and μ_j. $\hat{\tau}_j$ and μ_j are highly positively correlated, although there are still some discrepancies between these two in the sense that some village-specific effects are over-estimated while the others are under-estimated.

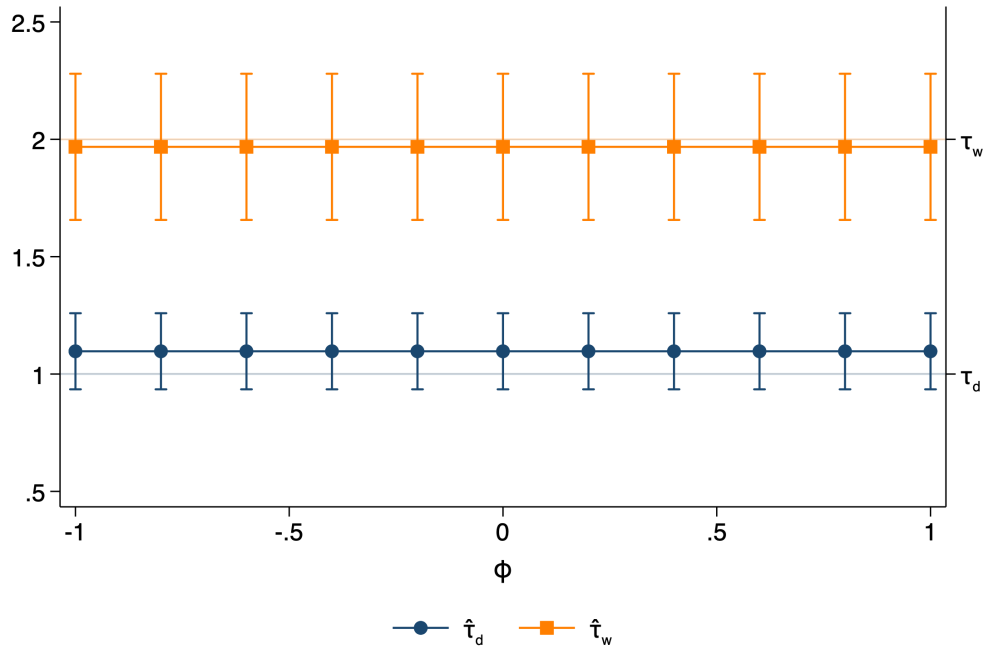
[1] For the related issues, see Muralidharan et al. (2023) "Factorial designs, model selection, and (incorrect) inference in randomized experiments." Review of Economics and Statistics.

[2] This is because, if other individual is to get a treatment and the assignment mechanism is not individualistic, you are less likely to get treated.

Figure 1: Q1.6. ϕ and Estimated Coefficents on D_j and W_{ij} in Two Regressions



(a) Without the Interaction Term $D_j \times W_{ij}$

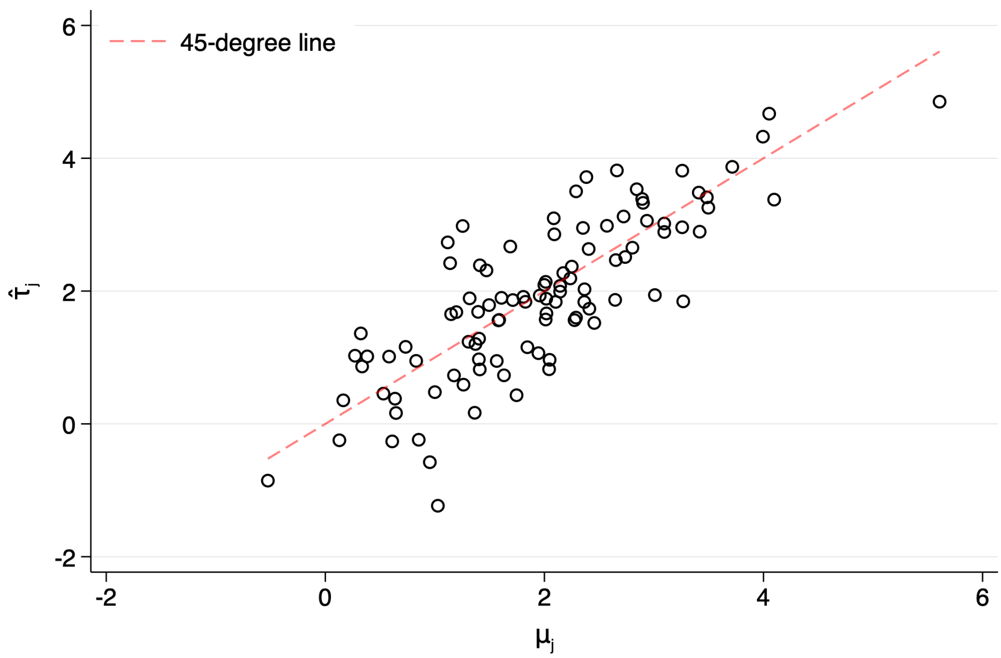


(b) With the Interaction Term $D_j \times W_{ij}$

Table 1: Q1.7. Comparing the Regression Estimates with/without Village FE

	$\phi = 0$		$\phi = 1$	
	(1)	(2)	(3)	(4)
	No Village FE	Village FE	No Village FE	Village FE
D_j	1.0177		1.5286	
	(0.1308)		(0.1317)	
W_{ij}	1.9021	1.8843	2.3268	2.3085
	(0.1267)	(0.1229)	(0.1328)	(0.1278)
Observations	1000	1000	1000	1000

Figure 2: Q1.8. Comparing $\hat{\tau}_j$ to μ_j



9.

For simplicity, let $\sum_{i|j} W_{ij} = N_j/2$ from $\Pr(W_{ij} = 1)$, and formulate the model as follows: for each $d \in \{0, 1\}$,

$$Y_{ij}(d, 1)|\mu_j \sim N(\mu_j, \sigma_Y^2), \;\; Y_{ij}(d, 0)|\mu_j \sim N(0, \sigma_Y^2), \quad \mu_j \sim N(\mu, \sigma_\mu^2).$$

Note that the equal variance assumption of the potential outcomes, meant for simplicity, is easily relaxable. Then, we consider the following within-village difference-in-means estimator for μ_j as an unbiased estimator of μ_j:

$$\hat{\tau}_j = \frac{1}{\sum_i W_{ij}} \sum_i Y_{ij} W_{ij} - \frac{1}{\sum_i (1 - W_{ij})} \sum_i Y_{ij}(1 - W_{ij}).$$

Then, it follows that[3]

$$\hat{\tau}_j|\mu_j \sim N\left(\mu_j,\ \frac{4\sigma_Y^2}{N_j}\right).$$

Then, one can show that in this normal/normal model,

$$\mathbb{E}[\mu_j|\hat{\tau}_j] = \left(\frac{4\sigma_\varepsilon^2/N_j}{\sigma_\mu^2 + 4\sigma_\varepsilon^2/N_j}\right)\mu + \left(\frac{\sigma_\mu^2}{\sigma_\mu^2 + 4\sigma_\varepsilon^2/N_j}\right)\hat{\tau}_j.$$

We take the Empirical Bayes approach to obtain this quantity for each village j:

- Estimate μ with grand mean:

$$\bar{\tau} = \frac{1}{J}\sum_j \hat{\tau}_j,$$

- Estimate σ_Y^2 from the 'within' variance of Y_{ij}:

$$\hat{\sigma}_Y^2 = \frac{1}{J(N_j - 1) - 1}\sum_j\sum_i \left(Y_{ij} - \hat{\alpha}_j - \hat{\tau}_j W_{ij}\right)^2,$$

where $Y_{ij} - \hat{\alpha}_j - \hat{\tau}_j W_{ij}$ are the corresponding OLS residual;

- Estimate σ_μ^2 from "overdispersion" in $\hat{\tau}_j$:

$$\hat{\sigma}_\mu^2 = \frac{1}{J-1}\left[\sum_j \left(\hat{\tau}_j - \bar{\tau}\right)^2 - \frac{4\sigma_Y^2}{N_j}\right],$$

- Empirical Bayes estimate can be:

$$\hat{\mathbb{E}}\left[\mu_j|\hat{\tau}_j\right] = \left(\frac{4\hat{\sigma}_Y^2/N_j}{\hat{\sigma}_\mu^2 + 4\hat{\sigma}_Y^2/N_j}\right)\mu + \left(\frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2 + 4\hat{\sigma}_Y^2/N_j}\right)\hat{\tau}_j.$$

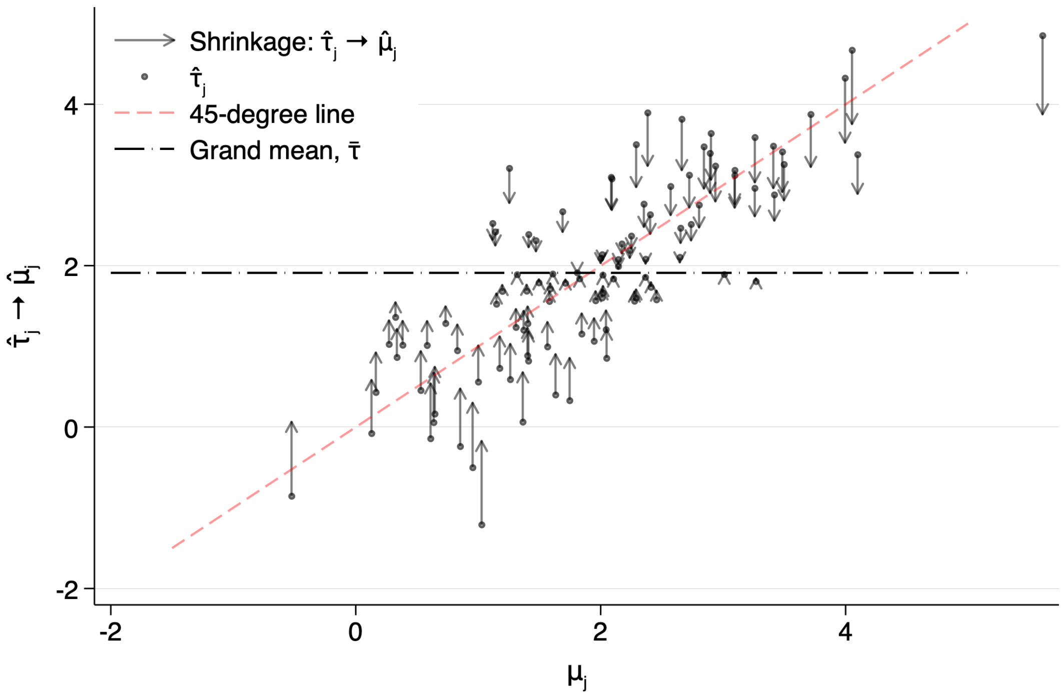
Figure 3 presents the Empirical Bayes estimates and how they move $\hat{\tau}_j$ towards μ_j for each village j. We here denote $\hat{\mu}_j \equiv \hat{\mathbb{E}}[\mu_j|\hat{\tau}_j]$. Are the Empirical Bayes estimates $\{\hat{\mu}_j\}$ "better" than $\{\bar{Y}_j\}$? One measure is the mean-squared errors, i.e, $\frac{1}{J}\sum_j \left(\hat{\mu}_j - \mu_j\right)^2$ and $\frac{1}{J}\sum_j \left(\bar{Y}_j - \mu_j\right)^2$, which indicate the average squared deviation from the true values. In my simulation, the former is 3.69 while the latter is 5.78, showing that the Empirical Bayes method indeed gives the estimates with a smaller MSE than the simple group average estimates do.

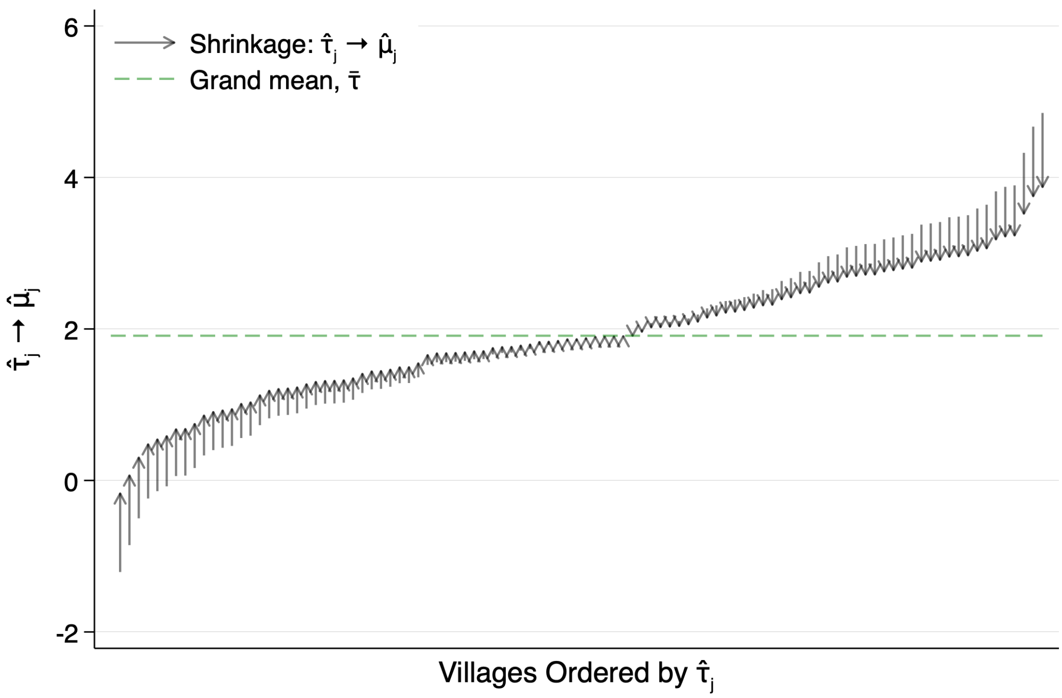
Note that this smaller MSE comes at the cost of increasing bias: as in the lecture note, the MSE is the sum of the bias and variance terms, and the smaller MSE of the Empirical Bayes estimates comes from larger bias but even smaller variance compared to the sample average or the difference-in-means estimator in this case.[4]

[3] Remember the variance of the difference-in-means estimator.

[4] We typically consider the efficiency within the class of unbiased estimator, e.g., the OLS as B(L)UE, since the MSE then consists only of the variance term. This Empirical Bayes approach (and the shrinkage estimator, the James-Stein estimator) does not fall into this class because it is a biased estimator.

Figure 3: Q1.9. Empirical Bayes Estimates $\{\hat{\mu}_j\}$





Question 2

1.

We will show that θ is not identified through the 'rank violations' argument. In this case, because the number of parameters (6) is strictly larger than the number of equations (4), we cannot solve for these parameters. This implies that θ is not identified without any normalisations.

We can also see the multicollinearity by expressing the model $Y_{it} = \alpha_i + \gamma_t + \psi_{j(i,t)} + \varepsilon_{it}$ as follows:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}}_{\equiv \mathbf{X}} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \gamma_1 \\ \gamma_2 \\ \psi_1 \\ \psi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{bmatrix}.$$

Denote the k-th column of $\mathbf{X}$ by $\mathbf{x}_k$. Then, we can see that $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_3 + \mathbf{x}_4 = \mathbf{x}_5 + \mathbf{x}_6 = \mathbf{1}$. This brings us to, e.g.,

$$\begin{aligned} \mathbf{x}_3 &= \mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_4, \\ \mathbf{x}_5 &= \mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_6, \end{aligned}$$

indicating that two column vectors in $\mathbf{X}$ are perfectly collinear with the other four column vectors and so at least two parameters cannot be identified.

From this result, we can use the following normalisation and show that θ is identified:

$$\gamma_1 = \psi_1 = 0.$$

Suppose that $\mathbb{E}[\varepsilon_{it}|\theta] = 0 \ \forall i, t$. Then, we can express the conditional mean log wage equations in the following way:

$$\begin{aligned} \mathbb{E}[Y_{11}|\theta] &= \alpha_1, \\ \mathbb{E}[Y_{12}|\theta] &= \alpha_1 + \gamma_2 + \psi_2, \\ \mathbb{E}[Y_{21}|\theta] &= \alpha_2 + \psi_2, \\ \mathbb{E}[Y_{22}|\theta] &= \alpha_2 + \gamma_2. \end{aligned}$$

Rearranging these equations, we can solve for the parameters:

$$\begin{aligned} \Rightarrow \ \alpha_1 &= \mathbb{E}[Y_{11}|\theta], \\ \alpha_2 &= \frac{1}{2}\left[(\mathbb{E}[Y_{22}|\theta] + \mathbb{E}[Y_{21}|\theta]) - (\mathbb{E}[Y_{12}|\theta] - \mathbb{E}[Y_{11}|\theta])\right], \\ \gamma_2 &= \frac{1}{2}\left[(\mathbb{E}[Y_{12}|\theta] - \mathbb{E}[Y_{11}|\theta]) + (\mathbb{E}[Y_{22}|\theta] - \mathbb{E}[Y_{21}|\theta])\right], \\ \psi_2 &= \frac{1}{2}\left[(\mathbb{E}[Y_{12}|\theta] - \mathbb{E}[Y_{11}|\theta]) - (\mathbb{E}[Y_{22}|\theta] - \mathbb{E}[Y_{21}|\theta])\right], \end{aligned}$$

establishing the identification of θ.

2.

In this perfect complement case, θ is not identified even with reasonable normalisations. Let us again consider the system of conditional mean log wage equations:

$$
\begin{aligned}
\mathbb{E}[Y_{11}|\theta] &= \gamma_1 + \min\{\alpha_1, \psi_1\}, \\
\mathbb{E}[Y_{12}|\theta] &= \gamma_2 + \min\{\alpha_1, \psi_2\}, \\
\mathbb{E}[Y_{21}|\theta] &= \gamma_1 + \min\{\alpha_2, \psi_2\}, \\
\mathbb{E}[Y_{22}|\theta] &= \gamma_2 + \min\{\alpha_2, \psi_1\}.
\end{aligned}
$$

If, for instance, $\alpha_1, \alpha_2 < \psi_1, \psi_2$, then one can construct the system with identical $\mathbb{E}[Y_{11}|\theta]$ but $\tilde{\psi}_1 = \psi_1 + c_1,\ \tilde{\psi}_2 = \psi_2 + c_2$ for any $c_1, c_2 > 0$ in place of ψ_1 and ψ_2.

What about, as in part 1, normalising some parameters? It does not work in this case with the minimum operator. For simplicity, let us ignore $\{\gamma_t\}$. Suppose we normalise any one or two of the parameters α_i and ψ_j to 0. Then, we can always find the cases where the normalised parameters are weakly smaller than the rest of the parameters and exactly the same equations hold with any parameter values satisfying this condition and one or more equations with $\mathbb{E}[Y_{it}|\theta] = \min\{\text{non-normalised parameters}\}$. For instance, let $\alpha_1 = \psi_1 = 0$. Then, if $\alpha_1, \psi_1 \leq \alpha_2, \psi_2$, then any $\tilde{\psi}_2 \geq 0$ and $\tilde{\alpha}_2 \geq 0$ with $\min\{\tilde{\alpha}_2, \tilde{\psi}_2\} = \mathbb{E}[Y_{21}|\theta]$ are consistent with this system of equations. θ is therefore not identified in this perfect complement case.

3.

In the presence of the worker $i = 3$, the answer to part 1 and part 2 do not change.

Part 1. Now we have six equations and seven parameters, meaning that at least one parameter needs to be normalised. Expressing the model in a matrix form again, we have

$$
\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix}
=
\underbrace{\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0
\end{bmatrix}}_{\equiv \mathbf{X}}
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \gamma_1 \\ \gamma_2 \\ \psi_1 \\ \psi_2 \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{bmatrix}.
$$

Then, we can see that, for instance, $\mathbf{x}_4 = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_5,\ \mathbf{x}_6 = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_7$, indicating that we need two parameters to be normalised. Indeed, analogous to part 1, normalising $\gamma_1 = \psi_1 = 0$, we can solve the system of conditional mean log wage equations as follows:

$$
\begin{aligned}
&\mathbb{E}[Y_{11}|\theta] = \alpha_1,\ \mathbb{E}[Y_{12}|\theta] = \alpha_1 + \gamma_2 + \psi_2,\\
&\mathbb{E}[Y_{21}|\theta] = \alpha_2 + \psi_2,\ \mathbb{E}[Y_{22}|\theta] = \alpha_2 + \gamma_2.\\
&\mathbb{E}[Y_{31}|\theta] = \alpha_3,\ \mathbb{E}[Y_{32}|\theta] = \alpha_3 + \gamma_2,\\
&\quad \Rightarrow\ \alpha_1 = \mathbb{E}[Y_{11}|\theta],\\
&\qquad\ \alpha_3 = \mathbb{E}[Y_{31}|\theta],\\
&\qquad\ \gamma_2 = \mathbb{E}[Y_{32}|\theta] - \mathbb{E}[Y_{31}|\theta],\\
&\qquad\ \psi_2 = \mathbb{E}[Y_{12}|\theta] - \mathbb{E}[Y_{11}|\theta] - (\mathbb{E}[Y_{32}|\theta] - \mathbb{E}[Y_{31}|\theta])\\
&\qquad\ \alpha_2 = \mathbb{E}[Y_{22}|\theta] - (\mathbb{E}[Y_{32}|\theta] - \mathbb{E}[Y_{31}|\theta])
\end{aligned}
$$

Part 2 The answer to part 2 does not change with the third worker $i = 3$. This is because exactly the same problem also applies to this case. Even if we normalise one or two parameters in $\{\alpha_i\}$ and $\{\psi_j\}$ to 0, there always exist the cases in which non-normalised parameters are larger than the normalised ones and thus take any values within the set defined by the equations (inequalities).

4.

Let $\gamma_1 = \psi_1 = 0$. Suppose $\varepsilon_{it} = 0$ for all (i, t). Since this is the same normalisations as in part 1, we can use the sample analogue of the population estimators to find the values of θ:

$$
\begin{aligned}
\alpha_1 &= Y_{11} = 2,\\
\alpha_2 &= \frac{1}{2}\left[(Y_{22} + Y_{21}) - (Y_{12} - Y_{11})\right] = 2,\\
\gamma_2 &= \frac{1}{2}\left[(Y_{12} - Y_{11}) + (Y_{22} - Y_{21})\right] = 1,\\
\psi_2 &= \frac{1}{2}\left[(Y_{12} - Y_{11}) - (Y_{22} - Y_{21})\right] = 1,
\end{aligned}
$$

5.

See the do-file for the Stata implementation.

6.

In matrix notation, we can express the model in which workers and machines are perfect subsitutes as

$$\mathbf{Y} = \mathbf{D}\alpha + \mathbf{X}\gamma + \mathbf{F}\psi + \varepsilon,$$

where $\mathbf{D}$ is $NT \times N$ matrix with each $((k, s), i)$ element being $\mathbb{1}\{k = i\}$ (a matrix of worker dummies), $\mathbf{X}$ is $NT \times T$ matrix with each $((k, s), t)$ element being $\mathbb{1}\{s = t\}$ (a matrix of period dummies), and $\mathbf{F}$ is $NT \times J$ matrix with each $((k, s), j)$ element being $\mathbb{1}\{j(k, s) = j\}$ (a matrix of machine dummies).

The identification of the worker and machine effects requires the following conditions:

$$\mathbb{E}\left[\mathbf{D}'\varepsilon\right] = 0,\ \mathbb{E}\left[\mathbf{F}'\varepsilon\right] = 0;$$
$$\text{In each connected set, one } \psi_j = 0.$$

The first set of conditions can be expressed as $\mathbb{E}\left[\sum_t \varepsilon_{it}\right] = 0$ for each i, and $\mathbb{E}\left[\sum_{(i,t)} \mathbb{1}\{j(i,t) = j\}\ \varepsilon_{it}\right] = 0$ for each j. According to the DGP in this simulation, the noise $\varepsilon_{it} \sim N(0, 0.2)$ is i.i.d., and thus $\mathbb{E}\left[\sum_t \varepsilon_{it}\right] = 0$ holds for all i, indicating that $\mathbb{E}\left[\mathbf{D}'\varepsilon\right] = 0$ is satisfied. In addition, the random assignment of workers to machines implies that $j(i,t)$ and ε_{it} are independent (and so uncorrelated), ensuring $\mathbb{E}[\mathbf{F}'\varepsilon] = 0$.

The second, rank condition requires us to normalise one ψ_j to 0 in each connected set of machines via worker assignment, akin to our discussion in part 1 and 3.

7.

The upper panel of Figure 4 presents the scatter plot of the true α_i and ψ_i against their estimates. The dotted line is the 45-degree line. We can see that $\hat{\alpha}_i$ is largely on the 45-degree line, showing that the estimates are close to the true values. While $\hat{\psi}_i$ all above the 45-degree line, the estimates and the true values are perfectly collinear for ψ as well, and it looks like the parameter estimates are shifted downwards from the true values by about one. This is due to, as we have seen above, normalisation—missing the "intercept" of $\psi \sim \mathcal{N}(1, 1)$.

8.

The lower panel of Figure 4 shows the same scatter plot as in part 7 except that the dataset is generated with $T = 5$, which is substantially smaller than $T = 200$ in part 7. Because of this 'short panel' structure, the estimates are slightly more noisy and exhibits more variability.

Furthermore, this shorter periods mean that the number of worker movements and that of machines connected with other machines can be smaller, reducing the number of estimable worker and machine fixed effects. In my simulation, there are two individuals and two machines whose effects are not estimated.

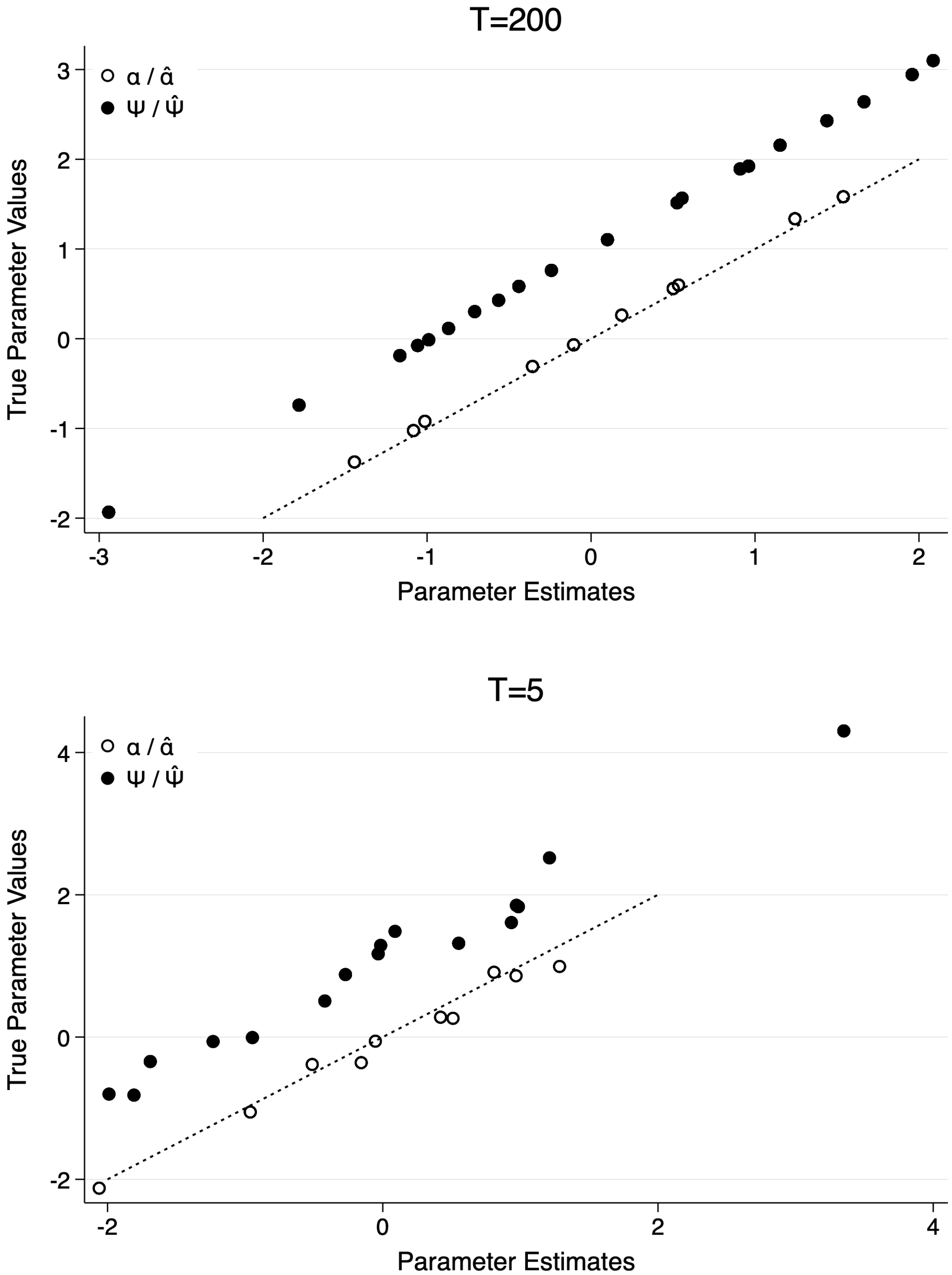
9.

Suppose that the assignment process of workers to machines is switched to the process so that the output is maximised. Then, since $\psi_j \sim \mathcal{N}(1, 1)$ for all i and t, the assignment function can be formulated as $\forall i, t$,

$$j(i,t) = \arg\max_j Y_{it} = \arg\max_j \ \left\{\alpha_i + \gamma_t + \psi_{j(i,t)} + \varepsilon_{it}\right\} = \arg\max_j \psi_j,$$

which is $j \in \{1, \ldots, J\}$ such that $\max \psi_j$. In this setting, therefore, every worker is assigned to the single machine whose ψ_j is the largest. This then implies that, because no workers move across machines, the connected set is singleton and no machine effects are identified.

Figure 4: Q2.7/8. Worker and Machine Effects



Question 3

1.

Denote each group in the grouped treatment structure by $g \in \{2, 3\}$, where $g = 2$ is the group treated in $t = 2$ and $g = 3$ corresponds to the one treated in $t = 3$. A two-way fixed effects regression can be written as

$$Y_{it} = \alpha_{g(i)} + \gamma_t + \tau_{\mathrm{TWFE}}\ D_{it} + \varepsilon_{it},$$

where τ_{TWFE} is the (population) OLS estimator. We typically impose the following parallel trends assumption in order τ_{TWFE} to correspond to *some* causal effect: $\forall t \geq 2$,

$$\mathbb{E}[Y_{it}(0) - Y_{it-1}(0)|G_i = 2] = \mathbb{E}[Y_{it}(0) - Y_{it-1}(0)|G_i = 3].$$

Now, we consider the average treatment effect (ATT), as the causal effect of interest:

$$\tau_{\mathrm{ATT}} = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \mid D_{it} = 1\right].$$

For $\hat{\tau}_{\mathrm{TWFE}}$ to correspond to τ_{ATT}, the average treatment effect for each treated cell is homogeneous. More precisely, if τ_{gt}, defined in Question 3.2., is homogeneous across (g, t), i.e., $\tau_{gt} = \tau$, then τ_{TWFE} corresponds to the ATT.

2.

We compute the weights associated with four τ_{gt} in the TWFE regression, i.e., $\{\tau_{22}, \tau_{23}, \tau_{33}, \tau_{34}\}$. Following Lecture 9, the weight for each treated cell is given by

$$\frac{N_{gt}}{N_1} w_{gt} = \frac{N_{gt}}{N_1} \frac{\tilde{D}_{gt}}{\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1} \tilde{D}_{gt}},$$

where $\{\tilde{D}_{gt}\}$ are the residuals from the regression $D_{it} = \alpha^D_{g(i)} + \gamma^D_t + \epsilon_{it}$, $N_1 = \sum_{(g,t):D_{it}=1}$. From the FWL Theorem, we have $\tilde{D}_{gt} = D_{it} - \bar{D}_g - \bar{D}_t + \bar{D}$. Using this, we have

$$\begin{aligned}
\tilde{D}_{22} &= 1 - \frac{1}{2} - \frac{1}{2} + \frac{1}{2} = \frac{1}{2}, \\
\tilde{D}_{23} &= 1 - \frac{1}{2} - 1 + \frac{1}{2} = 0, \\
\tilde{D}_{33} &= 1 - \frac{1}{2} - 1 + \frac{1}{2} = 0, \\
\tilde{D}_{34} &= 1 - \frac{1}{2} - \frac{1}{2} + \frac{1}{2} = \frac{1}{2}.
\end{aligned}$$

Combining these results gives

$$\begin{aligned}
\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1}\tilde{D}_{gt} &= \frac{N_{22}}{N_1}\cdot\frac{1}{2} + \frac{N_{23}}{N_1}\cdot 0 + \frac{N_{33}}{N_1}\cdot 0 + \frac{N_{34}}{N_1}\cdot\frac{1}{2} = \frac{N_{22}+N_{34}}{2N_1}, \\
\Rightarrow \quad \frac{N_{22}}{N_1}w_{22} &= \frac{N_{22}}{N_1}\cdot\frac{1}{2}\cdot\left(\frac{N_{22}+N_{34}}{2N_1}\right)^{-1} = \frac{N_{22}}{N_{22}+N_{34}} \\
\frac{N_{23}}{N_1}w_{23} &= \frac{N_{23}}{N_1}\cdot 0\cdot\left(\frac{N_{22}+N_{34}}{2N_1}\right)^{-1} = 0, \\
\frac{N_{33}}{N_1}w_{33} &= \frac{N_{33}}{N_1}\cdot 0\cdot\left(\frac{N_{22}+N_{34}}{2N_1}\right)^{-1} = 0, \\
\frac{N_{34}}{N_1}w_{22} &= \frac{N_{34}}{N_1}\cdot\frac{1}{2}\cdot\left(\frac{N_{22}+N_{34}}{2N_1}\right)^{-1} = \frac{N_{34}}{N_{22}+N_{34}}.
\end{aligned}$$

3.

Leveraged Comparisons As in Lecture 9, the switching estimator leverages the comparisons of [0 1] against [0 0] and [1 0] against [1 1]. More precisely, using the grouped treatment structure, the comparisons we leverage are

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

The former measures "joiner-effect" while the latter "leaver-effect".

The first comparison identifies $\tau_{22} = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0) \mid G_i = 2]$, under the following parallel trends assumption, because it is a standard 2×2 DID:

$$\mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 3] = \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 2].$$

The second comparison, on the other hand, concerns $\tau_{24} = \mathbb{E}[Y_{i4}(1) - Y_{i4}(0) \mid G_i = 2]$. This is identified under the parallel trends assumption for $\{Y(1)\}$:

$$\mathbb{E}[Y_{i4}(1) - Y_{i3}(1) \mid G_i = 3] = \mathbb{E}[Y_{i4}(1) - Y_{i3}(1) \mid G_i = 2],$$

because now $\mathbb{E}[Y_{i4}(1) \mid G_i = 2]$ is unobservable, but this assumption provides

$$\begin{aligned}
\mathbb{E}[Y_{i4}(1) \mid G_i = 2] &= \mathbb{E}[Y_{i4}(1) - Y_{i3}(1) \mid G_i = 3] + \mathbb{E}[Y_{i3}(1) \mid G_i = 2] \\
&= \mathbb{E}[Y_{i4} \mid G_i = 3] - \mathbb{E}[Y_{i3} \mid G_i = 3] + \mathbb{E}[Y_{i3} \mid G_i = 2].
\end{aligned}$$

Avoided Comparison From the structure, in turn, we avoid the following comparison in this estimator:

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

For this comparison to be valid, we need, e.g., the following assumptions to hold:

$$\begin{aligned}
\mathbb{E}[Y_{i3}(0) - Y_{i2}(0) \mid G_i = 3] &= \mathbb{E}[Y_{i3}(0) - Y_{i2}(0) \mid G_i = 2], \\
\mathbb{E}[Y_{i3}(0) - Y_{i2}(0) \mid G_i = 2] &= \mathbb{E}[Y_{i3}(1) - Y_{i2}(1) \mid G_i = 2].
\end{aligned}$$

While the first is the standard parallel trends assumption for $\{Y(0)\}$, the second additional one states that the trends of $\{Y(0)\}$ and $\{Y(1)\}$ are identical on average for the 'control group' ($G_i = 2$ in this case). If the treatment effect is constant, i.e., $\mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid G_i = 3] = \tau$, then $\mathbb{E}[Y_{i3}(1) - Y_{i2}(1) \mid G_i = 3]$ differences out such constant effect, leaving the trends of $\mathbb{E}[Y_{i3}(0) - Y_{i2}(0) \mid G_i = 3]$. This is, thus, violated if the treatment effects are heterogeneous across periods, for instance.

4.

Switching Estimator Suppose that the two types of parallel trends assumptions in part 3 hold, the population switching estimators can be then expressed as

$$\begin{aligned}
\tau_{22} &= (\mathbb{E}[Y_{i2} \mid G_i = 2] - \mathbb{E}[Y_{i1} \mid G_i = 2]) - (\mathbb{E}[Y_{i2} \mid G_i = 3] - \mathbb{E}[Y_{i1} \mid G_i = 3]), \\
\tau_{24} &= (\mathbb{E}[Y_{i4} \mid G_i = 3] - \mathbb{E}[Y_{i3} \mid G_i = 3]) - (\mathbb{E}[Y_{i4} \mid G_i = 2] - \mathbb{E}[Y_{i3} \mid G_i = 2]).
\end{aligned}$$

Using their sample analogue, we obtain the estimates

$$\hat{\tau}_{22} = (4 - 1) - (3 - 2) = 2, \ \hat{\tau}_{24} = (3 - 4) - (3 - 5) = 1.$$

If the weights for these comparisons are equal, meaning that each weight $= 1/2$, then the estimate is $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 1 = \frac{3}{2}$.

Another Identified τ_{gt} In addition to these two $\{\tau_{gt}\}$, we can show that τ_{34} is identified under the parallel trends assumption for $\{Y(0)\}$ for all $t \geq 2$:

$$\mathbb{E}[Y_{it}(0) - Y_{it-1}(0) \mid G_i = 3] = \mathbb{E}[Y_{it}(0) - Y_{it-1}(0) \mid G_i = 2].$$

To show the identification of τ_{34}, note that

$$\begin{aligned}
\tau_{34} &= \mathbb{E}[Y_{i4}(1) - Y_{i4}(0) \mid G_i = 3] \\
&= \mathbb{E}[Y_{i4} \mid G_i = 3] - \underbrace{\mathbb{E}[Y_{i4}(0) \mid G_i = 3]}_{\text{Unobservable}}.
\end{aligned}$$

To impute the second unobservable term with the known, observable quantities, we can recursively apply the set of the parallel trends assumptions:

$$\begin{aligned}
\mathbb{E}[Y_{i4}(0) - Y_{i3}(0) \mid G_i = 3] &= \mathbb{E}[Y_{i4}(0) - Y_{i3}(0) \mid G_i = 2] \\
\Leftrightarrow \ \mathbb{E}[Y_{i4}(0) \mid G_i = 3] &= \underbrace{\mathbb{E}[Y_{i3}(0) \mid G_i = 3]}_{\text{Unobservable}} + \underbrace{\mathbb{E}[Y_{i4}(0) \mid G_i = 2]}_{=\mathbb{E}[Y_{i4} \mid G_i=2]} - \underbrace{\mathbb{E}[Y_{i3}(0) \mid G_i = 2]}_{\text{Unobservable}}, \\
\mathbb{E}[Y_{i3}(0) - Y_{i2}(0) \mid G_i = 3] &= \mathbb{E}[Y_{i3}(0) - Y_{i2}(0) \mid G_i = 2] \\
\Leftrightarrow \ \mathbb{E}[Y_{i3}(0) \mid G_i = 3] - \mathbb{E}[Y_{i3}(0) \mid G_i = 2] &= \underbrace{\mathbb{E}[Y_{i2}(0) \mid G_i = 3]}_{=\mathbb{E}[Y_{i2} \mid G_i=3]} - \underbrace{\mathbb{E}[Y_{i2}(0) \mid G_i = 2]}_{\text{Unobservable}} \\
\mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 3] &= \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 2] \\
\Leftrightarrow \ \mathbb{E}[Y_{i2}(0) \mid G_i = 2] &= \mathbb{E}[Y_{i1}(0) \mid G_i = 2] + \mathbb{E}[Y_{i2}(0) - Y_{i1}(0) \mid G_i = 3] \\
&= \mathbb{E}[Y_{i1} \mid G_i = 2] + \mathbb{E}[Y_{i2} \mid G_i = 3] - \mathbb{E}[Y_{i1} \mid G_i = 3],
\end{aligned}$$

Combining these results, we have

$$\begin{aligned}
\mathbb{E}[Y_{i4}(0) \mid G_i = 3] &= \mathbb{E}[Y_{i4} \mid G_i = 2] + \mathbb{E}[Y_{i2} \mid G_i = 3] \\
&\qquad\qquad - (\mathbb{E}[Y_{i1} \mid G_i = 2] + \mathbb{E}[Y_{i2} \mid G_i = 3] - \mathbb{E}[Y_{i1} \mid G_i = 3]) \\
&= \mathbb{E}[Y_{i4} \mid G_i = 2] + (\mathbb{E}[Y_{i1} \mid G_i = 3] - \mathbb{E}[Y_{i1} \mid G_i = 2]) , \\
\Rightarrow \ \tau_{34} &= \mathbb{E}[Y_{i4} \mid G_i = 3] - \mathbb{E}[Y_{i4} \mid G_i = 2] - (\mathbb{E}[Y_{i1} \mid G_i = 3] - \mathbb{E}[Y_{i1} \mid G_i = 2]) ,
\end{aligned}$$

which is the comparison of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ in period 4 to $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ in period 1. Assuming that these parallel trends assumptions all hold, we obtain the estimate of τ_{34} as

$$\hat{\tau}_{34} = (3 - 3) - (2 - 1) = -1.$$