Econ 2120 - Selected Answers for Final Review

December 3, 2018

Homework 7

Ex 1.

(a) Define the error U_{it} by

$$U_{it} := Y_{it} - \mathbb{E}\left[Y_{it} \mid Y_{i1}, \ldots, Y_{i,t-1}, A_i\right] = Y_{it} - \lambda_t^* - \rho^* Y_{i,t-1} - A_i$$

By definition we have that U_{it} is orthogonal to constant, to A_i and also all previous Y_{it} (that is, Y_{is}, for $s < t$). It also happens to be orthogonal to all previous U_{is}, for $s < t$, because U_{is} can be written as a function of constant, A_i, and previous Y_{it}.

Now just notice that by adding and subtracting A_i we can write

$$V_i = Y_{i3} - \lambda_3^* - \rho^* Y_{i2} - A_i - (Y_{i2} - \lambda_2^* - \rho^* Y_{i1} - A_i) = U_{i3} - U_{i2}$$

Because U_{i3} and U_{i2} are orthogonal to constant and to Y_{i1}, we get our result.

In our orthogonality framework, we can define

$$B_i = \begin{bmatrix} 1 \\ Y_{i1} \end{bmatrix}$$

and

$$R_i = \begin{bmatrix} 1 & Y_{i2} - Y_{i1} \end{bmatrix}$$

and

$$\beta^* = \begin{bmatrix} \lambda_3^* - \lambda_2^* \\ \rho^* \end{bmatrix}$$

And our orthogonality condition is

$$\mathbb{E}\left[B_i V_i\right] = 0$$

In order to β^* to be identified, we need that

$$\mathbb{E}\left[B_i R_i\right] = \mathbb{E}\begin{bmatrix} 1 & Y_{i2} - Y_{i1} \\ Y_{i1} & Y_{i1}(Y_{i2} - Y_{i1}) \end{bmatrix}$$

to be invertible. By playing with it, this will be the case if

$$\mathbb{E}\left[(Y_{i1} - \mathbb{E}\left[Y_{i1}\right])(Y_{i2} - Y_{i1} - \mathbb{E}\left[Y_{i2} - Y_{i1}\right])\right] \neq 0$$

which is the same as $\mathrm{Cov}(Y_{i1}, Y_{i2} - Y_{i1}) \neq 0$.

In words, the change from Y_{i1} to Y_{i2} cannot be completely unrelated to the value of Y_{i1}.

But if we have this condition, then our orthogonality condition identifies ρ^* and we get consistency from the usual argument for estimators based on orthogonality conditions.

(b) The idea here is to first set up the structural equations, that is, what comes from our model and then try to get to the reduced form. To give a simple example before going into the problem, go back to our IV model

$$\begin{aligned} Y_{i,1} &= X_i'\gamma + u_{i,1} \\ Y_{i,2} &= Y_{i,1}\beta + \varepsilon_i \end{aligned}$$

This is our model, and we don't make any assumption about the correlation between $u_{i,1}$ and ε_i, so the way we define the parameters is by saying that these errors will be orthogonal[1] to X_i, the "instruments." This is our structural model,

[1] In the sense that $\mathbb{E}\begin{bmatrix} u_{i,1} \\ \varepsilon \end{bmatrix} \mid X_i\Big] = 0$.

our reduced-form model is giving by substituting $X_i\gamma + u_{i,1}$ for $Y_{i,1}$ in the second equation to get

$$\begin{aligned} Y_{i,1} &= X_i'\gamma + u_{i,1} \\ Y_{i,2} &= X_i'\gamma\beta + u_{i,2} \end{aligned}$$

where $u_{i,2} = \varepsilon_i + u_{i,1}\beta$. Let's do the same transformation in a harder but more generalizable way. First move all the endogenous variables (the Y_i) to the left-hand side.

$$\begin{aligned} Y_{i,1} &= X_i'\gamma + u_{i,1} \\ -Y_{i,1}\beta + Y_{i,2} &= \varepsilon_i \end{aligned}$$

Now rewrite in matrix format

$$\underbrace{\begin{bmatrix} 1 & 0 \\ -\beta & 1 \end{bmatrix}}_{B(\theta)} \begin{bmatrix} Y_{i,1} \\ Y_{i,2} \end{bmatrix} = \begin{bmatrix} X_i'\gamma \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} u_{i,1} \\ \varepsilon_i \end{bmatrix}$$

Now invert the matrix $B(\theta)$ to get

$$\begin{bmatrix} Y_{i,1} \\ Y_{i,2} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix}}_{(B(\theta))^{-1}} \begin{bmatrix} X_i'\gamma \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix} \begin{bmatrix} u_{i,1} \\ \varepsilon_i \end{bmatrix}$$

And if we do the multiplications we get exactly to our reduced-form model. So here we have

$$\mathbb{E}[Y_i \mid X_i] = B(\theta)^{-1} \begin{bmatrix} X_i'\gamma \\ \mathbf{0} \end{bmatrix} = (\mathbf{I}_2 \otimes X_i') \begin{bmatrix} \gamma \\ \gamma\beta \end{bmatrix} =: (\mathbf{I}_2 \otimes X_i')\pi(\theta)$$

and

$$\mathbb{V}[Y_i] = B(\theta)^{-1} \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1,\varepsilon} \\ \sigma_{u_1,\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix} (B(\theta)^{-1})' = \begin{bmatrix} \sigma_{u_1}^2 & \beta\sigma_{u_1}^2 + \sigma_{u_1,\varepsilon} \\ \beta\sigma_{u_1}^2 + \sigma_{u_1,\varepsilon} & \beta^2\sigma_{u_1}^2 + 2\beta\sigma_{u_1,\varepsilon} + \sigma_\varepsilon^2 \end{bmatrix} =: \Sigma(\theta)$$

And our θ here would be

$$\theta = (\gamma, \beta, \sigma_{u_1}^2, \sigma_{u_1,\varepsilon}, \sigma_\varepsilon^2)$$

Ok. Now let's go to the original problem.

We don't observe any exogenous variables, so our X_i here is just a constant. Let's set up the structural equation, if $t \geq 2$,

$$Y_{i,t} = \lambda_t + \rho Y_{i,t-1} + A_i + U_{i,t}$$

if $t = 1$, then the best we can do is to use the best linear predictor

$$Y_{i,1} = \alpha_0 + \alpha_1 A_i + V_i$$

here you could have also used just the best linear predictor on a constant, which is also the conditional expectation on a constant, that is

$$Y_{i,1} = \alpha_0 + \tilde{V}_i = \mathbb{E}[Y_{i,1}] + \tilde{V}_i$$

The advantage of using the first representation is that now V_i is uncorrelated with A_i, that gives a nicer format for the covariance matrix. But you could have used the other one if you kept in your mind that $\tilde{V}_i$ may be correlated with A_i. Remember, we don't observe A_i, so it will go to error anyway.

Now we move all the endogenous to the left-hand side.

If $t = 1$,

$$Y_{i,1} = \alpha_0 + \alpha_1 A_i + V_i$$

If $t \geq 2$

$$Y_{i,t} - \rho Y_{i,t-1} = \lambda_t + A_i + U_{i,t}$$

This can be put in a matrix format

$$
\underbrace{\begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ -\rho & 1 & 0 & \dots & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & -\rho & 1 \end{bmatrix}}_{B(\theta)} \underbrace{\begin{bmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ \vdots \\ Y_{i,T-1} \\ Y_{i,T} \end{bmatrix}}_{Y_i} = \begin{bmatrix} \alpha_0 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_{T-1} \\ \lambda_T \end{bmatrix} + \begin{bmatrix} \alpha_1 A_i + V_i \\ A_i + U_{i,2} \\ \vdots \\ \vdots \\ A_i + U_{i,T-1} \\ A_i + U_{i,T} \end{bmatrix}
$$

We invert $B(\theta)$, to the reduced-form model

$$
\begin{bmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ \vdots \\ Y_{i,T-1} \\ Y_{i,T} \end{bmatrix} = B(\theta)^{-1} \begin{bmatrix} \alpha_0 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_{T-1} \\ \lambda_T \end{bmatrix} + B(\theta)^{-1} \begin{bmatrix} \alpha_1 A_i + V_i \\ A_i + U_{i,2} \\ \vdots \\ \vdots \\ A_i + U_{i,T-1} \\ A_i + U_{i,T} \end{bmatrix}
$$

Because X_i is just a constant, we have that $\mathbb{E}[Y_i \mid X_i] = \mathbb{E}[Y_i]$. Then we have

$$
\mathbb{E}[Y_i] = B(\theta)^{-1} \begin{bmatrix} \alpha_0 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_{T-1} \\ \lambda_T \end{bmatrix} =: \pi(\theta)
$$

For the covariance we have

$$
\mathbb{V}(Y_i) = B(\theta)^{-1} \begin{bmatrix} \alpha_1^2\sigma_A^2 + \sigma_V^2 & \alpha_1\sigma_A^2 & \dots & \dots & \alpha_1\sigma_A^2 \\ \alpha_1\sigma_A^2 & \sigma_A^2 + \sigma_2^2 & \sigma_A^2 & \dots & \sigma_A^2 \\ \alpha_1\sigma_A^2 & \sigma_A^2 & \sigma_A^2 + \sigma_3^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sigma_A^2 \\ \alpha_1\sigma_A^2 & \sigma_A^2 & \dots & \sigma_A^2 & \sigma_A^2 + \sigma_T^2 \end{bmatrix} \left(B(\theta)^{-1}\right)'
$$

So our parameters are

$$\theta = (\rho, \lambda_2, \ldots, \lambda_T, \alpha_0, \alpha_1, \sigma_A^2, \sigma_2^2, \ldots, \sigma_T^2)$$

(c) We can set up the quasi-likelihood to estimate $\hat{\theta}_{\text{QML}}$ as

$$\hat{\theta}_{\text{QML}} = \underset{\theta}{\text{argmax}} \; -\log(\det(\Sigma(\theta))) - \frac{1}{n}\sum_{i=1}^{n} (Y_i - \pi(\theta))' \, (\Sigma(\theta))^{-1} \, (Y_i - \pi(\theta))$$

where $\Sigma(\theta)$ and $\pi(\theta)$ are as defined before

so we ask our research assistant to maximize this by numerical optimization.

(d) Here is just the robustness property of quasi-maximum likelihood with normal linear model likelihood. Even if the model is not actually normal, we get that $\hat{\rho}_{\text{QML}} \to \rho^*$ as long as there is a θ^* such that

$$\mathbb{E}_F[Y_i] = \pi(\theta^*) \qquad \text{and} \qquad \mathbb{V}_F(Y_i) = \Sigma(\theta^*)$$

(e) **How can the limit distribution results in the GMM framework be used with the quasi-likelihood function to obtain a .95 confidence interval for γ?**

We can set up a just-identified GMM by using the first-order conditions from the log likelihood. Call the log likelihood $\ell(Y_i, \theta)$, that is

$$\ell(Y_i, \theta) = -\log(\det(\Sigma(\theta))) - (Y_i - \pi(\theta))' \, (\Sigma(\theta))^{-1} \, (Y_i - \pi(\theta))$$

then our moment function for the GMM will be

$$\Psi(Y_i, \theta) = \frac{\partial \ell}{\partial \theta}(Y_i, \theta)$$

and our moment condition is

$$\mathbb{E}[\Psi(Y_i, \theta^*)] = 0$$

We have that the QML estimator maximizes the log likelihood, so it satisfies the first-order conditions, that is

$$\sum_{i=1}^{n} \Psi(Y_i, \hat{\theta}_{\text{QML}}) = 0$$

that is, $\hat{\theta}_{\text{QML}}$ is the GMM estimator for the moments $\Psi(Y_i, \theta) = \frac{\partial \ell}{\partial \theta}(Y_i, \theta)$.

Now we just use that the GMM estimator in the just identified case has the following asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{\text{QML}} - \theta^*) \xrightarrow{\text{d}} \mathcal{N}(0, \Lambda)$$

where

$$\Lambda = \left(\mathbb{E}\left[\frac{\partial \Psi}{\partial \theta}(Y_i, \theta) \right] \right)^{-1} \mathbb{E}\Big[\Psi(Y_i, \theta^*) \Psi(Y_i, \theta^*)' \Big] \left(\mathbb{E}\left[\frac{\partial \Psi}{\partial \theta'}(Y_i, \theta) \right] \right)^{-1}$$

In our setting this gives

$$\Lambda = \left(\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta \partial \theta'}(Y_i, \theta) \right] \right)^{-1} \mathbb{E}\left[\frac{\partial \ell}{\partial \theta}(Y_i, \theta) \frac{\partial \ell}{\partial \theta'}(Y_i, \theta) \right] \left(\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta \partial \theta'}(Y_i, \theta) \right] \right)^{-1}$$

We need to replace this with an estimator, but it is just the sample version, that is

$$\hat{\Lambda} = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ell}{\partial \theta \partial \theta'}(Y_i, \theta) \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ell}{\partial \theta}(Y_i, \theta) \frac{\partial \ell}{\partial \theta'}(Y_i, \theta) \right) \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ell}{\partial \theta \partial \theta'}(Y_i, \theta) \right)^{-1}$$

And to get an confidence interval we just compute

$$\text{SE} = \sqrt{\text{diag}\left(\frac{\hat{\Lambda}}{n} \right)}$$

and construct the confidence interval for ρ as

$$\text{CI}_\rho = [\ \hat{\rho}_{\text{QML}} - 1.96\text{SE}_\rho\ ,\ \hat{\rho}_{\text{QML}} + 1.96\text{SE}_\rho\ ]$$

Ex 2. It is the same as question 1.

Ex 3.

(a) The normal likelihood (for the univariate case) is given by

$$f(Y_i \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2}(Y_i - \mu)^2\right)$$

By pluging

$$\mu = \mathbb{E}\left[Y_i \mid X_i\right] = X_i'\beta$$

and

$$\sigma^2 = \mathbb{V}\left[Y_i \mid X_i\right] = \exp\left(X_i'\gamma\right)$$

we get

$$f(Y_i \mid X_i, \beta, \gamma) = \frac{1}{\sqrt{2\pi \exp\left(X_i'\gamma\right)}} \exp\left(\frac{(Y_i - X_i'\beta)^2}{2\exp\left(X_i'\gamma\right)}\right)$$

Because the sample is iid, the likelihood for the whole sample is just the product

$$f(Y \mid X, \beta, \gamma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \exp\left(X_i'\gamma\right)}} \exp\left(\frac{(Y_i - X_i'\beta)^2}{2\exp\left(X_i'\gamma\right)}\right)$$

Now the RA just need to maximize this function. We would usually first take the log to make it numerically more stable.

(b) Yes! The general idea here is to say that under weak regularity conditions we have that

$$(\hat{\gamma}_{\text{QML}}, \hat{\beta}_{\text{QML}}) \to_{\text{p}} (\gamma_F, \beta_F)$$

And (γ_F, β_F) are defined as

$$(\gamma_F, \beta_F) = \operatorname{argmax} \mathbb{E}\left[-\log(\exp\left(X_i'\gamma\right)) - \frac{(Y_i - X_i'\beta)^2}{\exp\left(X_i'\gamma\right)}\right]$$

the expectation of the log-likelihood for a single observation (ignoring constants that don't affect the argmax).

By iterated expectations we have

$$\mathbb{E}\left[-\log(\exp\left(X_i'\gamma\right)) - \frac{\mathbb{E}\left[(Y_i - X_i'\beta)^2 \mid X_i\right]}{\exp\left(X_i'\gamma\right)}\right]$$

So, conditional on γ, the solution for β is to minimize

$$\mathbb{E}\left[(Y_i - X_i'\beta)^2 \mid X_i\right]$$

but this is done by setting $\beta_F = \beta^*$, because[2]

$$\mathbb{E}\left[Y_i \mid X_i\right] = X_i'\beta^*$$

Given that $\beta_F = \beta^*$, and that

$$\mathbb{E}\left[(Y_i - X_i'\beta^*) \mid X_i\right] = \mathbb{V}\left[Y_i \mid X_i\right] = \exp(X_i'\gamma^*)$$

we get that γ_F maximizes

$$\min \mathbb{E}\left[-\log(\exp(X_i'\gamma)) - \frac{\exp(X_i'\gamma^*)}{\exp(X_i'\gamma)}\right]$$

Take first-order conditions to get that we maximize this expression for any X_i by setting

$$\exp(X_i'\gamma_F) = \exp(X_i'\gamma^*)$$

And we conclude that $\gamma_F = \gamma^*$.

[2]This is just the usual orthogonality argument, we have

$$\begin{aligned}\mathbb{E}\left[(Y_i - g(X_i))^2 \mid X_i\right] &= \mathbb{E}\left[(Y_i - \mathbb{E}\left[Y_i \mid X_i\right] + \mathbb{E}\left[Y_i \mid X_i\right] - g(X_i))^2 \mid X_i\right] = \\ &= \mathbb{E}\left[(Y_i - \mathbb{E}\left[Y_i \mid X_i\right])^2 \mid X_i\right] + 2\underbrace{\mathbb{E}\left[Y_i - \mathbb{E}\left[Y_i \mid X_i\right] \mid X_i\right](\mathbb{E}\left[Y_i \mid X_i\right] - g(X_i))}_{=0} + \\ &+ (\mathbb{E}\left[Y_i \mid X_i\right] - g(X_i))^2 = \\ &= \mathbb{E}\left[(Y_i - \mathbb{E}\left[Y_i \mid X_i\right])^2 \mid X_i\right] + (\mathbb{E}\left[Y_i \mid X_i\right] - g(X_i))^2 \geq \mathbb{E}\left[(Y_i - \mathbb{E}\left[Y_i \mid X_i\right])^2 \mid X_i\right]\end{aligned}$$

Ex 4.

(a) As long as we get the conditional moment right, that is, as long as we have a γ such that

$$\mathbb{E}_F[\, Y_{i,t} \mid Z_{i,t}\,] = P(Y_{i,t} = 1 \mid Z_{i,t}) = \Phi(X'_{i,t}\gamma)$$

Then our estimator will be consistent to that. To see that, notice that (under regularity conditions) our estimator is converging to the argmax of the limit function, that is $\hat{\gamma} \xrightarrow{\text{p}} \gamma^*$, where

$$\begin{aligned}\gamma^* &= \operatorname*{argmax}_a \; \mathbb{E}_F\left[\sum_{t=1}^{T} Y_{i,t}\log\left(\Phi(X'_{i,t}a)\right) + (1-Y_{i,t})\log\left(1-\Phi(X'_{i,t}a)\right)\right] = \\ &= \operatorname*{argmax}_a \; \mathbb{E}_F\left[\sum_{t=1}^{T} \Phi(X'_{i,t}\gamma)\log\left(\Phi(X'_{i,t}a)\right) + \left(1-\Phi(X'_{i,t}\gamma)\right)\log\left(1-\Phi(X'_{i,t}a)\right)\right]\end{aligned}$$

Where the second equality follows from iterated expectations and our assumption that $\mathbb{E}_F[\, Y_{i,t} \mid Z_{i,t}\,] = \Phi(X'_{i,t}\gamma)$. Under the regularity conditions required for this to hold, we need that γ^* is the unique maximizer (identification).

Then we take first-order conditions to get

$$\mathbb{E}_F\left[\sum_{t=1}^{T}\left(\frac{\Phi(X'_{i,t}\gamma)\phi(X'_{i,t}\gamma^*)}{\Phi(X'_{i,t}\gamma^*)} - \frac{\left(1-\Phi(X'_{i,t}\gamma)\right)\phi(X'_{i,t}\gamma^*)}{1-\Phi(X'_{i,t}\gamma^*)}\right)X_{i,t}\right] = 0$$

And if we plug $\gamma^* = \gamma$, we see that the first-order condition is satisfied for γ. The second order condition can be shown to be negative at $\gamma^* = \gamma$ and so, by the identification assumption, γ is the unique maximizer and we get that $\hat{\gamma} \xrightarrow{p} \gamma$.

(b) We use our GMM framework in the just-identified case to get the asymptotic covariance matrix. In this case, our moments will be the derivative of the log likelihood for a single observation with respect to the parameters. That is

$$\begin{aligned}\Psi(Y_{i,t}, X_{i,t}, a) &= \frac{\partial}{\partial a}\left\{Y_{i,t}\log\left(\Phi(X'_{i,t}a)\right) + (1-Y_{i,t})\log\left(1-\Phi(X'_{i,t}a)\right)\right\} = \\ &= \left(\frac{Y_{i,t}\phi(X'_{i,t}a)}{\Phi(X'_{i,t}a)} - \frac{(1-Y_{i,t})\phi(X'_{i,t}a)}{1-\Phi(X'_{i,t}a)}\right)X_{i,t}\end{aligned}$$

Define

$$\Sigma = \mathbb{E}_F \left[\Psi(Y_{i,t}, X_{i,t}, \gamma)\Psi(Y_{i,t}, X_{i,t}, \gamma)' \right] = \mathbb{E}_F \left[\left(\frac{Y_{i,t}\phi(X_{i,t}'\gamma)}{\Phi(X_{i,t}'\gamma)} - \frac{(1-Y_{i,t})\phi(X_{i,t}'\gamma)}{1-\Phi(X_{i,t}'\gamma)} \right)^2 X_{i,t}X_{i,t}' \right]$$

$$H = \mathbb{E}_F \left[\frac{\partial \Psi}{\partial a}(Y_{i,t}, X_{i,t}, \gamma) \right]$$

This will just be the expectation of the second-derivative of the log likelihood.

Our asymptotic covariance matrix is given by

$$\Lambda = H^{-1}\Sigma H^{-1}$$

(c) If we can get samples from the posterior for γ, then we can get samples from the posterior for δ because

$$\delta = P(Y_{i1} = 1 \mid Z_{i1} = d) - P(Y_{i1} = 1 \mid Z_{i1} = c) = \Phi(g(d)'\gamma) - \Phi(g(c)'\gamma)$$

So one possible procedure is to get samples from the posterior for γ by

$$\tilde{\gamma}^{(j)} = \underset{a}{\operatorname{argmax}} \sum_{i=1}^{T} \sum_{t=1}^{T} V_i^{(j)} \left(Y_{it} \log(\Phi(X_{it}'a)) + (1-Y_{it}) \log(1-\Phi(X_{it}'a)) \right)$$

Now get samples from the posterior for δ by

$$\tilde{\delta}^{(j)} = \Phi(g(d)'\tilde{\gamma}^{(j)}) - \Phi(g(c)'\tilde{\gamma}^{(j)})$$

And we can construct a .95 credible interval by using quantiles from these samples.

Ex 5.

(a) This is almost by definition.

$$Y_i = \sum_{j=0}^{2} \mathbf{I}(T_i(S_i) = t_j) Y_i(t_j) = \sum_{j=0}^{2} T_{i,j}(S_i) Y_i(t_j)$$

(b) This one is easy too

$$\begin{aligned}
\mathbb{E}[Y_i \mid S_i = s] &= \mathbb{E}\left[\sum_{j=0}^{2} T_{i,j}(S_i) Y_i(t_j) \mid S_i = s\right] = \\
&= \mathbb{E}\left[\sum_{j=0}^{2} T_{i,j}(s) Y_i(t_j) \mid S_i = s\right] = \\
&= \sum_{j=0}^{2} \mathbb{E}\left[T_{i,j}(s) Y_i(t_j) \mid S_i = s\right] = \\
&\overset{(1)}{=} \sum_{j=0}^{2} \mathbb{E}\left[T_{i,j}(s) Y_i(t_j)\right]
\end{aligned}$$

In (1) we used that $(Y_i(2), Y_i(1), Y_i(0), T(s))$ is indenpendent of S_i.

(c) Here the trick is to use that $T_{i,0}(s) + T_{i,1}(s) + T_{i,2} = 1$.

$$\begin{aligned}
\mathbb{E}[Y_i \mid S_i = b] - \mathbb{E}[Y_i \mid S_i = a] &= \sum_{j=0}^{2} \mathbb{E}\left[T_{i,j}(b) Y_i(t_j)\right] - \mathbb{E}\left[T_{i,j}(a) Y_i(t_j)\right] = \\
&= \sum_{j=0}^{2} \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a)) Y_i(t_j)\right] = \\
&= \mathbb{E}[(T_{i,0}(b) - T_{i,0}(a)) Y_i(t_0)] + \sum_{j=1}^{2} \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a)) Y_i(t_j)\right]
\end{aligned}$$

Now, because $T_{i,0}(s) + T_{i,1}(s) + T_{i,2} = 1$, we have that

$$\begin{aligned}
T_{i,0}(b) - T_{i,0}(a) &= 1 - T_{i,1}(b) - T_{i,2}(b) - 1 + T_{i,1}(a) - T_{i,2}(a) = \\
&= -(T_{i,1}(b) - T_{i,1}(a)) - (T_{i,2}(b) - T_{i,2}(a))
\end{aligned}$$

this implies that

$$\mathbb{E}[(T_{i,0}(b) - T_{i,0}(a)) Y_i(t_0)] = -\sum_{j=1}^{2} \mathbb{E}[(T_{i,j}(b) - T_{i,j}(a)) Y_i(t_0)]$$

And we finally get that

$$
\begin{aligned}
\mathbb{E}[Y_i \mid S_i = b] - \mathbb{E}[Y_i \mid S_i = a] &= \left(\sum_{j=1}^{2} \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))Y_i(t_j)\right]\right) + \mathbb{E}[(T_{i,0}(b) - T_{i,0}(a))Y_i(t_0)] = \\
&= \sum_{j=1}^{2} \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))Y_i(t_j)\right] - \sum_{j=1}^{2} \mathbb{E}[(T_{i,j}(b) - T_{i,j}(a))Y_i(t_0)] = \\
&= \sum_{j=1}^{2} \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0))\right]
\end{aligned}
$$

(d) Now we are almost there, notice that $T_{i,j}(b) - T_{i,j}(a)$ can, in principle, have only 3 values $-1, 0, 1$. We can split $\mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0))\right]$ into 3 pieces

$$
\begin{aligned}
&\mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 1\right] P(T_{i,j}(b) - T_{i,j}(a) = 1) + \\
&+ \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 0\right] P(T_{i,j}(b) - T_{i,j}(a) = 0) + \\
&+ \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) - 1\right] P(T_{i,j}(b) - T_{i,j}(a) = -1)
\end{aligned}
$$

The third term is 0 because we assumed that $P(T_{i,j}(b) - T_{i,j}(a) = -1) = 0$. And the second is 0 because

$$
\begin{aligned}
&\mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 0\right] P(T_{i,j}(b) - T_{i,j}(a) = 0) = \\
&\qquad = \mathbb{E}\left[\, 0 \times (Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 0\right] P(T_{i,j}(b) - T_{i,j}(a) = 0) = 0
\end{aligned}
$$

So, in the end, we have $\mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0))\right]$ is just

$$
\begin{aligned}
&\mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 1\right] P(T_{i,j}(b) - T_{i,j}(a) = 1) = \\
&\qquad = \mathbb{E}\left[(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 1\right] P(T_{i,j}(b) - T_{i,j}(a) = 1)
\end{aligned}
$$

So we plug it back to get

$$
\begin{aligned}
&\mathbb{E}[Y_i \mid S_i = b] - \mathbb{E}[Y_i \mid S_i = a] = \sum_{j=1}^{2} \mathbb{E}\left[(T_{i,j}(b) - T_{i,j}(a))(Y_i(t_j) - Y_i(t_0))\right] = \\
&\qquad = \sum_{j=1}^{2} \mathbb{E}\left[(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 1\right] P(T_{i,j}(b) - T_{i,j}(a) = 1)
\end{aligned}
$$

We could do the same thing for $\mathbb{E}[T_i \mid S_i = b] - \mathbb{E}[T_i \mid S_i = a]$ to get that

$$\mathbb{E}[T_i \mid S_i = b] - \mathbb{E}[T_i \mid S_i = a] = \sum_{j=1}^{2}(t_j - t_0)P(T_{i,j}(b) - T_{i,j}(a) = 1)$$

So what we get in the end is that

$$\frac{\mathbb{E}[Y_i \mid S_i = b] - \mathbb{E}[Y_i \mid S_i = a]}{\mathbb{E}[T_i \mid S_i = b] - \mathbb{E}[T_i \mid S_i = a]} = \sum_{j=1}^{2} \mathbb{E}\left[(Y_i(t_j) - Y_i(t_0)) \mid T_{i,j}(b) - T_{i,j}(a) = 1\right] \omega_j$$

where

$$\omega_j = \frac{P(T_{i,j}(b) - T_{i,j}(a) = 1)}{(t_1 - t_0)P(T_{i,1}(b) - T_{i,1}(a) = 1) + P((t_2 - t_0)T_{i,2}(b) - T_{i,2}(a) = 1)}$$

So what got is that LATE is identifying a weighted average of average treatment effects for the two groups of "compliers", the ones that when moved from a to b changed from treatment t_0 to treatment t_1, and the ones that when moved from a to b changed from treatment t_0 to treatment t_2.

Not perfect, but still something.

(e) Now we could have some people that would move from treatment t_1 to treatment t_2 and also the other way around when we changed the subsidy from a to b. For these people we would have either

$$T_{i,1}(b) - T_{i,1}(a) < 0 \qquad \text{or} \qquad T_{i,2}(b) - T_{i,2}(a) < 0$$

And this violates our assumption. Now, intuitively, we would have some "defiers" that would appear in our LATE with the wrong sign.

One example, imagine that drug 1 has severe side effects, but has a higher rate of sucess in treating people. drug 2 is milder, but less effective. Some people when in the control group would be taking the drug 1 even if they are not in that bad shape just because they don't know that drug 2 exists. We send some encouragement that make then aware that drug 2 also exists, so we would expect that some of the people would take drug 1 without encouragement, but would take drug 2 if they were encouraged, especially if they are having the side effects.

The group of people that moved from 1 to 2 may be the healthier among the ones taking drug 1 without encouragement. So we look at the final outcome and it seems like drug 1 is worse than drug 2, but it is coming from just healthier people moving from drug 1 to drug 2 when we give the encouragement.

In any case, the point is that it gets much harder to think about monotoniciy and what exactly we get from LATE when we have more than one treatment.