Supplementary Appendix

Myungkou Shin

December 1, 2023

1 Additional empirical results

1.1 Choice of K

To find a stable type assignment over units, we considered $K = 2, \cdots, 10$ for the number of types in the classification steps. Table 1 contains the classification result. The last row is the number of units who are assigned to an 'accidental' type where only a few number of outliers are assigned and a structural break is suspected in the sense that

$$\frac{1}{T_0} \sum_{t=-T_0}^{-1} \left(\hat{\delta}_t(k) - \bar{\delta}(k) \right)^2$$

is high. For $K \geq 3$, the maximum estimated variance of $\hat{\delta}_t(k)$ for a stable type k is 1.062 while the minimum estimated variance of $\hat{\delta}_t(k)$ for an accidental type is 8.587. Also, all of the accidental types are singletons.

K	2	3	4	5	6	7	8	9	10
type 1	28	40	39	38	28	16	7	6	6
type 2	14	-	-	-	10	21	29	29	28
outliers	0	2	3	4	4	5	6	7	8

Table 1: summary of classification result across K

Figure 1 and Figure 2 contain the estimated time fixed-effects $\hat{\delta}_t(k)$ for the two stable types given $K = 5, \ldots, 10$. See that the estimates are stable for $K = 6, \cdots, 9$, give us an anecdotal evidence that the two types capture the heterogeneity in dataset in a stable way. For a more formal discussion, we also considered an information criterion for the classification step. Figure 3 plots the Bayesian information criterion as suggested in Bonhomme and Manresa (2015); Janys and Siflinger (2024):

$$\frac{1}{nT_0}\sum_{i,t}\left(Y_{it} - Y_{it-1} - \hat{\delta}_t(\hat{k}_i) - X_{it}^{\intercal}\hat{\theta}\right)^2 + \hat{\sigma}^2\frac{KT_0 + p}{nT_0}\log nT_0$$

where $\hat{\sigma}^2$ is estimated with the largest $K = 10$. The information criterion is minimized at $K = 9$. Based on this, we presented the estimation results based on the two stable types estimated with $K = 9$.

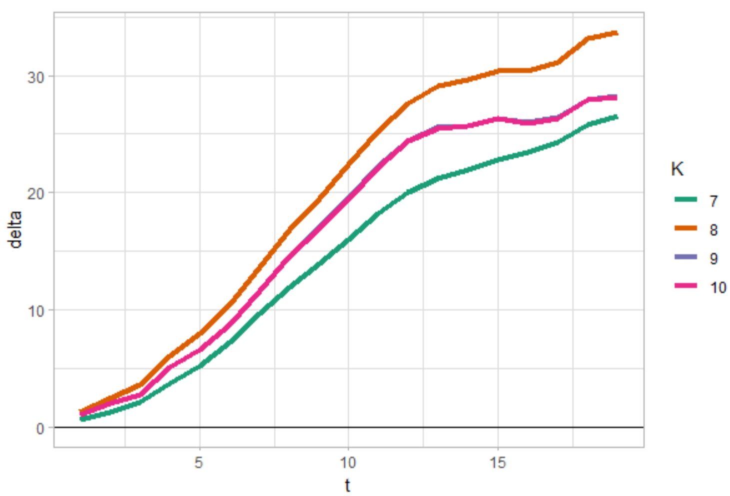


Figure 1: $\sum_{s\leq t}\hat{\delta}_s(1)$ across K

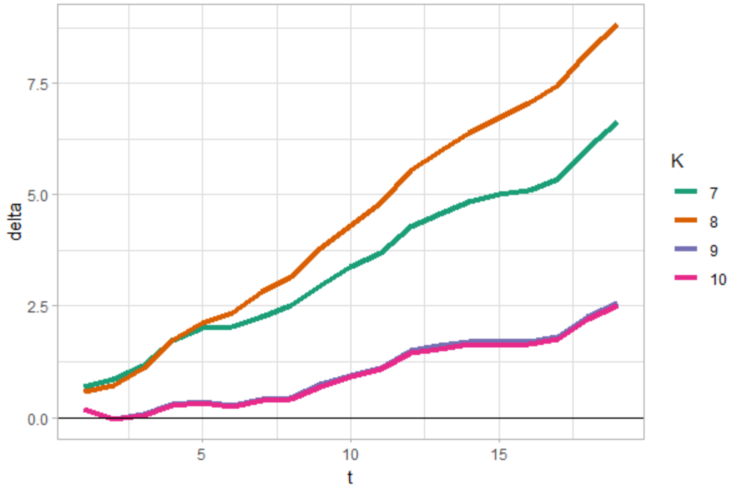


Figure 2: $\sum_{s\leq t}\hat{\delta}_t(2)$ across K

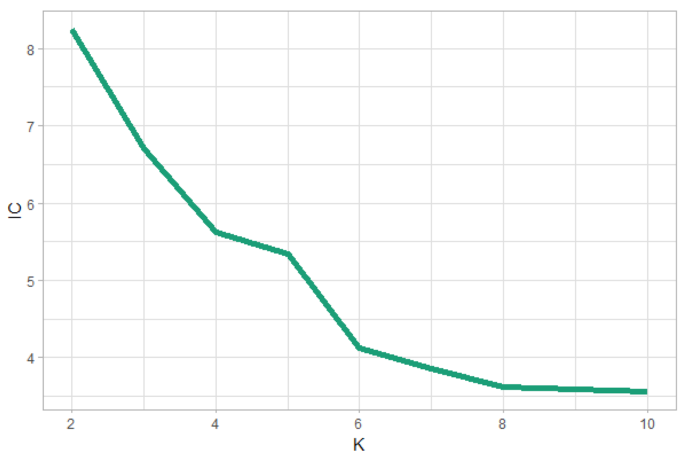


Figure 3: Information criterion for K

1.2 Heterogeneity across types

In the main text, it has been discussed that the two stable types are different in terms of their control covariates X_{it}. Here, we present additional balancedness test for the two types. Figure 4 contains the histogram for the treatment timing for type 1 and type 2. An eyeball test suggests that type 1 is slightly earlier-treated compared to type 2. Table 2 contains formal balancedness test for treatment timing E_i across type 1 and type 2. Any of the mean comparison test that uses indicator for being never-treated, treatment timing, and squared demeaned treatment timing, does not reject the null hypothesis that the two types share the same mean at 0.05 significance level; the joint p-value is 0.537. The two types are similar in terms of their treatment timing distribution, i.e., the treatment timing is well balanced across types.

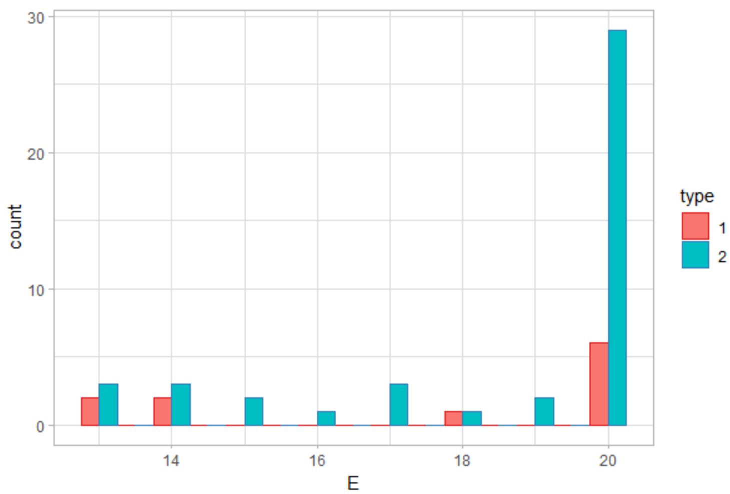


Figure 4: Frequency of E_i across K

	type 1	type 2	Diff
$\mathbf{1}\{E_i = \infty\}$	0.55	0.66	-0.11
	(0.52)	(0.48)	(0.17)
E_i	17.45	18.5	-1.05
	(3.21)	(2.44)	(1.03)
$(E_i - \mathbf{E}[E_i\|k_i])^2$	9.34	10.43	-1.09
	(6.04)	(11.75)	(2.54)
N	11	44	-
joint p-value			0.537

Table 2: Balancedness test on E_i

Lastly, we present the full classification result when $K = 9$. Below the number of school districts in each states for type 1, type 2 and accidental types. The number of treated school districts are denoted with red while the the number of never-treated school districts are denoted with black. Table 3 further summarizes the list and presents the number of school districts for each census region. Type 1 school districts are more concentrated in the South census region compared to Type 2 school districts; 9 out 11 compared to 28 out of 44.

Type 1 Alabama (2/1), Arkansas (1/1), Florida (1/1), North Carolina (1/1), Ohio (1), Wisconsin (1)

Type 2 Alabama (3/1), Arizona (1), Arkansas (1), California (2/1), Connecticut (2), Florida (2/7), Illinois (1), Indiana (2), Kentucky (1), Maryland (1), Michigan (2/1), Mississippi (3), New York (2), North Carolina (2), Pennsylvania (2), Texas (4/3)

Outliers Arkansas (1), Illinois (1), Massachusetts (1), Michigan (1), Mississippi (3/1), North Carolina (1)

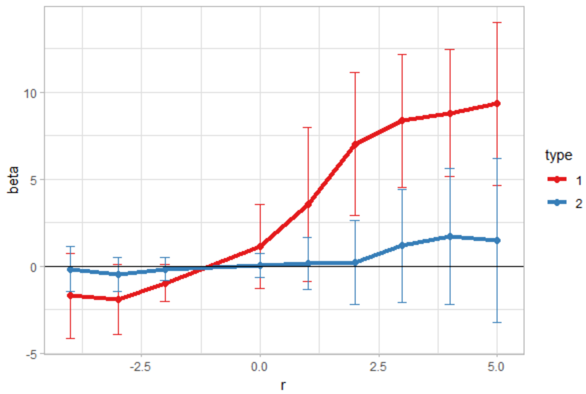
	Northeast	Midwest	South	West
Type 1	-	1/1	5/4	-
Type 2	6	5/1	16/12	2/2
Outliers	1	2	5/1	-

Table 3: Distribution of types across census regions

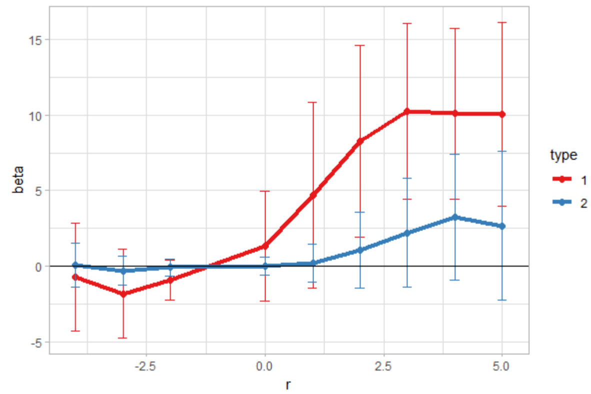
1.3 Robustness check

The qualitative result that we find the treatment effect to be statistically significant for type 1 while not for type 2 stays the same across perturbations in the classification result across K. Dynamic CATT $\beta_r(k)$ for $r \geq 2$ for type 1 are estimated to be significantly away from zero at 0.05 significance level and we find no significant treatment effect for type

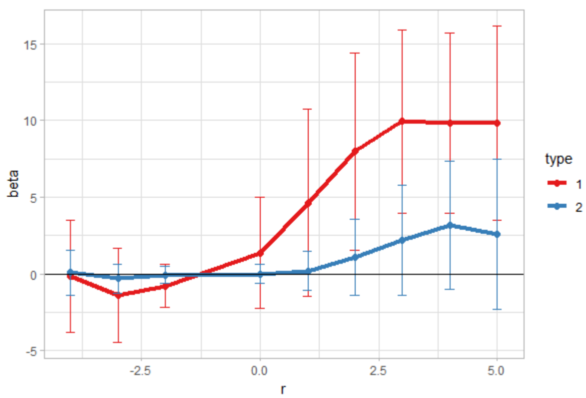
2. Also, we find no pretreatment for both type 1 and type 2 robust to perturbations in the classification, suggesting that the type-specific parallel trend assumption and the no anticipation assumption are not violated for pretreatment outcomes.



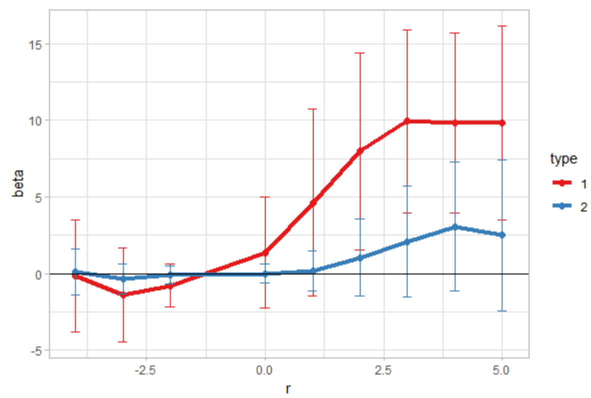
(a) $K = 6$



(b) $K = 7$



(c) $K = 8$



(d) $K = 9$

Figure 5: $\hat{\beta}_r$ across K

We also considered an outcome approach and ran a post-treatment linear regression model: for $t = 1989, \cdots, 2007$

$$Y_{it} - Y_{it-1} = \delta_t(k_i) + X_{it}^\intercal \theta + \sum_{-4 \le r \le 5; r \ne -1} \beta_r(k_i) \mathbf{1}\{t = E_i + r\} + \beta_6(k_i) \mathbf{1}\{t \ge E_i + 6\} + U_{it}.$$

Note that $\beta_r(k)$ is treatment effect on the first-differenced outcome. Thus, to retrieve the treatment effect on level, we need to sum $\beta_r(k)$ over r. Figure 6 contains the type-specific treatment effect estimates and their 95% confidence intervals. We find no pretreatment across type 1 and type 2. The estimates are close to the unweighted type-specific diff-in-diff estimates from Figure 5. Lastly, in Figure 7 contains estimates from the pooled regression specification:

$$Y_{it} - Y_{it-1} = \delta_t(k_i) + X_{it}^\intercal \theta + \sum_{-4 \le r \le 5; r \ne -1} \beta_r \mathbf{1}\{t = E_i + r\} + \beta_6 \mathbf{1}\{t \ge E_i + 6\} + U_{it}.$$

Again, the estimates are close to the averaged estimates of the unweighted type-specific diff-in-diff estimates from Figure 5.

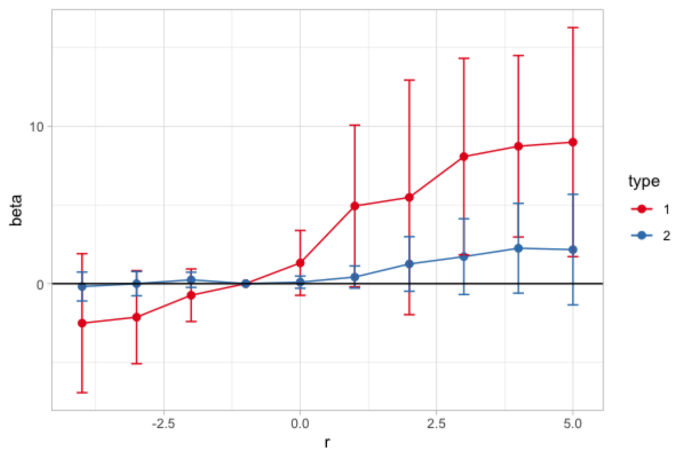


Figure 6: $\sum_{0 \le r' \le r} \hat{\beta}_{r'}(k)$ for $k = -4, \cdots, 5$

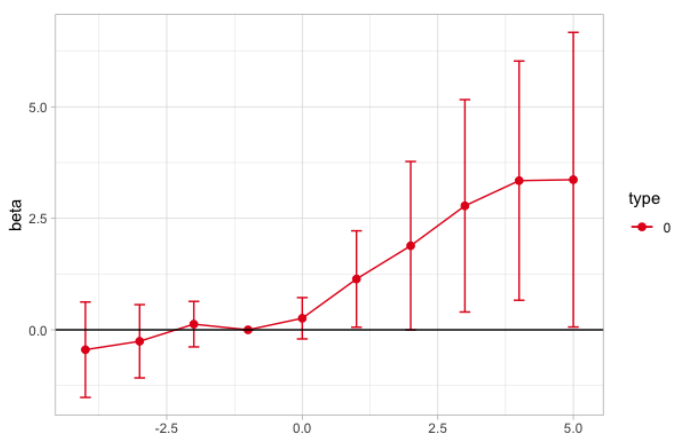


Figure 7: $\sum_{0 \le r' \le r} \hat{\beta}_{r'}$ for $k = -4, \cdots, 5$

2 Proof for Theorem 2

In the proof sections, we will use the dot notation to denote the first difference: $\dot{Y}_{it} = Y_{it} - Y_{it-1}$, $\dot{X}_{it} = X_{it} - X_{it-1}$ and $\dot{U}_{it} = U_{it} - U_{it-1}$. Also, we will use the superscript naught to denote the true values of the parameters and the latent type variable: e.g. k_i^0 is the true type of unit i.

We prove Theorem 2 in the context of a linear model for outcome in level (see *Remark 5* of the main text). This subsumes the case of a linear model for first-differenced outcomes, by replacing $\dot{X}_{it}$ and $\dot{U}_{it}$ with X_{it} and U_{it}. For Theorem 1, replace $\delta_t^0(k)$ and $\dot{U}_{it}$ with $\mathbf{E}[\dot{Y}_{it}(\infty)|k]$ and $\dot{Y}_{it}(E_i) - \mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0]$.

Step 1

The first step is to obtain an approximation of the objective function. Note that

$$\begin{aligned}
\widehat{Q}(\theta, \delta, \gamma) &= \frac{1}{nT_0} \sum_{i=1}^{n} \sum_{t=-T_0}^{-1} \left(\dot{Y}_{it} - \delta_t(k_i) - \dot{X}_{it}^{\intercal} \theta \right)^2 \\
&= \frac{1}{nT_0} \sum_{i,t} \left(\delta_t^0(k_i^0) - \delta_t(k_i) + \dot{X}_{it}^{\intercal}(\theta^0 - \theta) + \dot{U}_{it} \right)^2 \\
&= \frac{1}{nT_0} \sum_{i,t} \left\{ \left(\delta_t^0(k_i^0) - \delta_t(k_i) + \dot{X}_{it}^{\intercal}(\theta^0 - \theta) \right)^2 + \dot{U}_{it}^2 \right\} \\
&\quad + \frac{2}{nT_0} \sum_{i,t} \left(\delta_t^0(k_i^0) - \delta_t(k_i) + \dot{X}_{it}^{\intercal}(\theta^0 - \theta) \right) \dot{U}_{it}.
\end{aligned}$$

Let

$$\tilde{Q}(\theta, \delta, \gamma) = \frac{1}{nT_0} \sum_{i,t} \left\{ \left(\delta_t^0(k_i^0) - \delta_t(k_i) + \dot{X}_{it}^{\intercal}(\theta^0 - \theta) \right)^2 + \dot{U}_{it}^2 \right\}.$$

Then,

$$\begin{aligned}\left|\widehat{Q}(\theta,\delta,\gamma)-\tilde{Q}(\theta,\delta,\gamma)\right| &= \left|\frac{2}{nT_0}\sum_{i,t}\Big(\delta_t^0(k_i^0)-\delta_t(k_i)+\dot{X}_{it}^{\intercal}(\theta^0-\theta)\Big)\dot{U}_{it}\right| \\ &\leq \left|\frac{2}{nT_0}\sum_{i,t}\Big(\delta_t^0(k_i^0)-\delta_t(k_i)\Big)\dot{U}_{it}\right| + \left|\frac{2}{nT_0}\sum_{i,t}\dot{X}_{it}^{\intercal}(\theta^0-\theta)\dot{U}_{it}\right|. \qquad (1)\end{aligned}$$

Firstly, find that

$$\begin{aligned}\left|\frac{1}{nT_0}\sum_{i,t}\delta_t^0(k_i^0)\dot{U}_{it}\right| &\leq \sum_{k=1}^{K}\left|\frac{1}{nT_0}\sum_{i,t}\delta_t^0(k)\dot{U}_{it}\mathbf{1}\{k_i^0=k\}\right| \\ &\leq \sum_{k=1}^{K}\left(\frac{1}{T_0}\sum_t \delta_t^0(k)^2\right)^{\frac{1}{2}}\left(\frac{1}{T_0}\sum_t\left(\frac{1}{n}\sum_i \dot{U}_{it}\mathbf{1}\{k_i^0=k\}\right)^2\right)^{\frac{1}{2}} \\ &\leq M\sum_{k=1}^{K}\left(\frac{1}{n^2T_0}\sum_{i,j,t}\dot{U}_{it}\dot{U}_{jt}\mathbf{1}\{k_i^0=k_j^0=k\}\right)^{\frac{1}{2}} \xrightarrow{p} 0.\end{aligned}$$

The first two inequalities are from separating the summation into types and applying Cauchy-Schwartz's inequality to over t. The third is from Assumption 7-b. It remains to prove the convergence in probability; for that we use Assumption 7-a,d. With some constant $C>0$ that only depends on $M>0$ from Assumption 7,

$$\mathbf{E}\left[\dot{U}_{it}\dot{U}_{jt}\mathbf{1}\{k_i^0=k_j^0=k\}\right] = \begin{cases}\mathbf{E}\left[\dot{U}_{it}^2\mathbf{1}\{k_i^0=k\}\right]\leq C & \text{if } i=j\\ \mathbf{E}\left[\dot{U}_{it}\mathbf{1}\{k_i^0=k\}\right]\mathbf{E}\left[\dot{U}_{jt}\mathbf{1}\{k_j^0=k\}\right]=0 & \text{if } i\neq j\end{cases}$$

since $\mathbf{E}[\dot{U}_{it}\mathbf{1}\{k_i^0=k\}]=\mathbf{E}[\dot{U}_{it}|k_i^0=k]\Pr\{k_i^0=k\}=0.$[1] Then,

$$\mathbf{E}\left[\frac{1}{nT_0}\sum_{i,j,t}\dot{U}_{it}\dot{U}_{jt}\mathbf{1}\{k_i^0=k_j^0=k\}\right]\leq C.$$

[1]In the case of Theorem 1,

$$\mathbf{E}[\dot{Y}_{it}(E_i)-\mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0]|k_i^0=k]=\mathbf{E}[\mathbf{E}[\dot{Y}_{it}(E_i)-\dot{Y}_{it}(\infty)|k_i^0,E_i]|k_i^0=k]=0$$

from Assumption 2.

and $\frac{1}{n^2T_0}\sum_{i,j,t}\dot{U}_{it}\dot{U}_{jt}\mathbf{1}\{k_i^0 = k_j^0 = k\} = o_p(1)$. We can repeat this for the other quantity in the first term of (1).

Secondly, again from applying Cauchy-Schwartz's inequality and Jensen's inequality,

$$\left|\frac{1}{nT_0}\sum_{i,t}\dot{X}_{it}^{\intercal}\left(\theta^0-\theta\right)\dot{U}_{it}\right| \leq \frac{1}{T_0}\sum_t\left\|\frac{1}{n}\sum_i\dot{U}_{it}\dot{X}_{it}\right\|_2\cdot\left\|\theta^0-\theta\right\|_2$$
$$\leq \frac{2M}{\sqrt{n}}\cdot\frac{1}{T_0}\sum_t\left(\frac{1}{n}\sum_{i,j}\dot{U}_{it}\dot{U}_{jt}\dot{X}_{it}^{\intercal}\dot{X}_{jt}\right)^{\frac{1}{2}} = \frac{2M}{\sqrt{n}}\cdot O_p(1)\xrightarrow{p}0$$

The convergence in probability is from Assumption 7-a,d. Find that

$$\mathbf{E}\left[\dot{U}_{it}\dot{X}_{it}\right]=\mathbf{0},\qquad\qquad \mathbf{E}\left[\dot{U}_{it}^2\dot{X}_{it}^{\intercal}\dot{X}_{it}\right]\leq C$$

with some constant $C>0$ that only depends on $M>0$ from Assumption 7. Thus,

$$\frac{1}{T_0}\sum_t\mathbf{E}\left[\left(\frac{1}{n}\sum_{i,j}\dot{U}_{it}\dot{U}_{jt}\dot{X}_{it}^{\intercal}\dot{X}_{jt}\right)^{\frac{1}{2}}\right]\leq\frac{1}{T_0}\sum_t\left(\mathbf{E}\left[\frac{1}{n}\sum_{i,j}\dot{U}_{it}\dot{U}_{jt}\dot{X}_{it}^{\intercal}\dot{X}_{jt}\right]\right)^{\frac{1}{2}}\leq\sqrt{C}.$$

Then $\frac{1}{T_0}\sum_t\left(\frac{1}{n}\sum_{i,j}\dot{U}_{it}\dot{U}_{jt}\dot{X}_{it}^{\intercal}\dot{X}_{jt}\right)^{\frac{1}{2}} = O_p(1)$ and $\widehat{Q}(\theta,\delta,\gamma)-\tilde{Q}(\theta,\delta,\gamma)=o_p(1)$.

Step 2

By plugging in the true parameters, $\tilde{Q}(\theta^0,\delta^0,\gamma^0)=\frac{1}{nT_0}\sum_{i,t}\dot{U}_{it}^2$ and

$$\begin{aligned}\tilde{Q}(\theta,\delta,\gamma)-\tilde{Q}(\theta^0,\delta^0,\gamma^0) &= \frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0)-\delta_t(k_i)+\dot{X}_{it}^{\intercal}(\theta^0-\theta)\right)^2\\ &\geq \frac{1}{nT_0}\sum_{i,t}\left(\dot{X}_{it}^{\intercal}(\theta^0-\theta)-\bar{\dot{X}}_{k_i^0\wedge k_i,t}{}^{\intercal}(\theta^0-\theta)\right)^2\\ &= \frac{1}{nT_0}\sum_{i,t}(\theta^0-\theta)^{\intercal}\left(\dot{X}_{it}-\bar{\dot{X}}_{k_i^0\wedge k_i,t}\right)\left(\dot{X}_{it}-\bar{\dot{X}}_{k_i^0\wedge k_i,t}\right)^{\intercal}(\theta^0-\theta)\\ &\geq \min_{\gamma\in\Gamma}\rho_n(\gamma)\cdot\|\theta^0-\theta\|_2^2.\end{aligned}$$

Note that the unknowns in $\tilde{Q}(\theta, \delta, \gamma) - \tilde{Q}(\theta^0, \delta^0, \gamma^0)$ other than $(\theta^0 - \theta)$ are functions of (t, k_i^0, k_i). Thus, subtracting the group mean defined with (t, k_i^0, k_i) from $\dot{X}_{it}^\intercal(\theta^0 - \theta)$ is the lower bound for the sum of squares, giving us the first inequality.

Since the estimator minimizes the objective function,

$$\begin{aligned}
\tilde{Q}(\hat{\theta}, \hat{\delta}, \hat{\gamma}) &= \widehat{Q}(\hat{\theta}, \hat{\delta}, \hat{\gamma}) + o_p(1) \\
&\leq \widehat{Q}(\theta^0, \delta^0, \gamma^0) + o_p(1) \\
&= \tilde{Q}(\theta^0, \delta^0, \gamma^0) + o_p(1).
\end{aligned}$$

Therefore from Assumption 7-h,

$$\begin{aligned}
\min_{\gamma\in\Gamma} \rho_n(\gamma) \cdot \left\|\theta^0 - \hat{\theta}\right\|_2^2 &\leq \tilde{Q}(\hat{\theta}, \hat{\delta}, \hat{\gamma}) - \tilde{Q}(\theta^0, \delta^0, \gamma^0) = o_p(1) \\
\left\|\theta^0 - \hat{\theta}\right\|_2^2 &= \frac{1}{\min_{\gamma\in\Gamma} \rho_n(\gamma)} \cdot \min_{\gamma\in\Gamma} \rho_n(\gamma) \left\|\theta^0 - \hat{\theta}\right\|_2^2 \xrightarrow{p} \frac{1}{\rho} \cdot 0 = 0.
\end{aligned}$$

We have consistency of $\hat{\theta}$.

Step 3

In this step, we show that $\{\hat{\delta}_t(\hat{k}_i)\}_{i,t}$ is close to $\{\delta_t^0(k_i^0)\}_{i,t}$ in terms of the l_2 norm.

$$\begin{aligned}
&\left|\tilde{Q}(\hat{\theta}, \hat{\delta}, \hat{\gamma}) - \tilde{Q}(\theta^0, \hat{\delta}, \hat{\gamma})\right| \\
&= \left|\frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(\hat{k}_i) + \dot{X}_{it}^\intercal(\theta^0 - \hat{\theta})\right)^2 - \frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(\hat{k}_i)\right)^2\right| \\
&\leq \left|\frac{2}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(\hat{k}_i)\right)\dot{X}_{it}^\intercal(\theta^0 - \hat{\theta}) + \frac{1}{nT_0}\sum_{i,t}\left(\dot{X}_{it}^\intercal(\theta^0 - \hat{\theta})\right)^2\right| \\
&\leq \frac{4M}{nT_0}\sum_{i,t}\|\dot{X}_{it}\|_2 \cdot \left\|\theta^0 - \hat{\theta}\right\|_2 + \frac{1}{nT_0}\sum_{i,t}\|\dot{X}_{it}\|_2^2 \cdot \left\|\theta^0 - \hat{\theta}\right\|_2^2 = o_p(1).
\end{aligned}$$

The second inequality is from Assumption 7-b and Cauchy-Schwartz inequality on $|\dot{X}_{it}^{\intercal}(\theta^0 - \hat{\theta})|$. Note that for any n, $\frac{1}{nT_0}\sum_{i,t}\|\dot{X}_{it}\|_2^2$ is bounded in expectation by $4M$ from Assumption 7.d and thus $O_p(1)$. Likewise, $\frac{1}{nT_0}\sum_{i,t}\|\dot{X}_{it}\|_2$ is bounded in expectation by $2\sqrt{M}$. Since we have shown $\hat{\theta} \xrightarrow{p} \theta^0$, we have the last equality. Then,

$$
\begin{aligned}
&\frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(\hat{k}_i)\right)^2 + \frac{1}{nT_0}\sum_{i,t}\dot{U}_{it}^2 \\
&= \tilde{Q}(\theta^0, \hat{\delta}, \hat{\gamma}) = \tilde{Q}(\hat{\theta}, \hat{\delta}, \hat{\gamma}) + o_p(1) = \widehat{Q}(\hat{\theta}, \hat{\delta}, \hat{\gamma}) + o_p(1) \\
&\leq \widehat{Q}(\theta^0, \delta^0, \gamma^0) + o_p(1) = \frac{1}{nT_0}\sum_{i,t}\dot{U}_{it}^2 + o_p(1).
\end{aligned}
$$

$\frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(\hat{k}_i)\right)^2 = o_p(1)$. For Theorem 1, the result holds directly from Step 1.

Step 4

In this step, we find some permutation on $\left\{\hat{\delta}_t(k)\right\}_{t,k}$ so that $\frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(k_i^0)\right)^2$ is close to zero. Note that $\widehat{Q}(\theta, \delta, \gamma)$ does not vary for any $(\theta, \tilde{\delta}, \tilde{\gamma})$ defined with a permutation on $(1, \cdots, K)$: with σ, a permutation on $\{1, \cdots, K\}$, letting $\tilde{k}_i = \sigma(k_i)$ and $\tilde{\delta}_t(\sigma(k)) = \delta_t(k)$ gives us $\widehat{Q}(\theta, \delta, \gamma) = \widehat{Q}(\theta, \tilde{\delta}, \tilde{\gamma})$. Thus, we want to define a bijection on $\{1, \cdots, K\}$ to match $\hat{k}$ with true k^0, to have classification result. Define a function σ by letting

$$
\sigma(k) = \arg\min_{\tilde{k}} \frac{1}{T_0}\sum_{t=-T_0}^{-1}\left(\delta_t^0(k) - \hat{\delta}_t(\sigma(k))\right)^2
$$

for each k. First, let us show that σ actually lets the objective go to zero for each k: fix k,

$$
\begin{aligned}
&\min_{\tilde{k}} \frac{1}{T_0}\sum_{t=-T_0}^{-1}\left(\delta_t^0(k) - \hat{\delta}_t(\sigma(k))\right)^2 \\
&\leq \frac{n}{\sum_i \mathbf{1}\{k_i^0 = k\}} \cdot \min_{\tilde{k}} \frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k) - \hat{\delta}_t(\sigma(k))\right)^2 \mathbf{1}\{k_i^0 = k\} \\
&\leq \frac{n}{\sum_i \mathbf{1}\{k_i^0 = k\}} \cdot \frac{1}{nT_0}\sum_{i,t}\left(\delta_t^0(k_i^0) - \hat{\delta}_t(\hat{k}_i)\right)^2 \xrightarrow{p} 0
\end{aligned}
$$

as $n \to \infty$. From Assumption 7-f, we have the convergence.

For some $k, \tilde{k}$ such that $k \neq \tilde{k}$,

$$
\begin{aligned}
&\left(\frac{1}{T_0} \sum_t \left(\hat{\delta}_t(\sigma(k)) - \hat{\delta}_t(\sigma(\tilde{k})) \right)^2 \right)^{\frac{1}{2}} \\
&\geq \left(\frac{1}{T_0} \sum_t \left(\delta_t^0(k) - \delta_t^0(\tilde{k}) \right)^2 \right)^{\frac{1}{2}} \\
&\qquad\qquad - \left(\frac{1}{T_0} \sum_t \left(\delta_t^0(k) - \hat{\delta}_t(\sigma(k)) \right)^2 \right)^{\frac{1}{2}} - \left(\frac{1}{T_0} \sum_t \left(\delta_t^0(\tilde{k}) - \hat{\delta}_t(\sigma(\tilde{k})) \right)^2 \right)^{\frac{1}{2}} \\
&\xrightarrow{p} c(k, \tilde{k}) > 0
\end{aligned}
$$

from Assumption 7.c. Thus, $\Pr\{\sigma \text{ is not bijective}\} \leq \sum_{k \neq \tilde{k}} \Pr\left\{\sigma(k) = \sigma(\tilde{k})\right\} \to 0$ as $n \to \infty$. Note that σ depends on the dataset.

Before proceeding to the next step, let us drop the σ notation. Based on σ, we can construct a bijection $\tilde{\sigma} : \{1, \cdots, K\} \to \{1, \cdots, K\}$ such that

$$
\frac{1}{T} \sum_t \left(\delta_t^0(k) - \hat{\delta}_t(\tilde{\sigma}(k)) \right)^2 \xrightarrow{p} 0 \tag{2}
$$

as $n \to \infty$ for all k, by letting $\tilde{\sigma} = \sigma$ whenever σ is bijective. From now on, I will drop $\tilde{\sigma}$ by always rearranging $(\hat{\theta}, \hat{\delta}, \hat{\gamma})$ so that $\tilde{\sigma}(k) = k$.

Step 5

Here, we study the probability of the K-means algorithm assigning a wrong type to an arbitrary unit i.

$$
\begin{aligned}
\Pr\left\{\hat{k}_i \neq k_i^0\right\} &\leq \sum_{\tilde{k} \neq k_i^0} \Pr\left\{ \frac{1}{T_0} \sum_t \left(\dot{Y}_{it} - \hat{\delta}_t(\tilde{k}) - \dot{X}_{it}^\intercal \hat{\theta} \right)^2 \leq \frac{1}{T_0} \sum_t \left(\dot{Y}_{it} - \hat{\delta}_t(k_i^0) - \dot{X}_{it}^\intercal \hat{\theta} \right)^2 \right\} \\
&= \sum_{\tilde{k} \neq k_i^0} \Pr\left\{ \frac{2}{T_0} \sum_t \left(\hat{\delta}_t(k_i^0) - \hat{\delta}_t(\tilde{k}) \right) \cdot \left(\dot{Y}_{it} - \frac{\hat{\delta}_t(k_i^0) + \hat{\delta}_t(\tilde{k})}{2} - \dot{X}_{it}^\intercal \hat{\theta} \right) \leq 0 \right\}.
\end{aligned}
$$

The inequality is from the second stage of the K-means algorithm. Then,

$$
\begin{aligned}
&\Pr\left\{\hat{k}_i \neq k_i^0\right\} \\
&= \sum_{\tilde{k}\neq k_i^0} \Pr\left\{\frac{2}{T_0}\sum_t \left(\hat{\delta}_t(k_i^0)-\hat{\delta}_t(\tilde{k})\right)\cdot\left(\delta_t^0(k_i^0)-\frac{\hat{\delta}_t(k_i^0)+\hat{\delta}_t(\tilde{k})}{2}+\dot{X}_{it}^{\intercal}(\theta^0-\hat{\theta})+\dot{U}_{it}\right)\leq 0\right\} \\
&\leq \sum_k \sum_{\tilde{k}\neq k} \Pr\left\{\frac{2}{T}\sum_t \left(\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\cdot\left(\delta_t^0(k)-\frac{\hat{\delta}_t(k)+\hat{\delta}_t(\tilde{k})}{2}+\dot{X}_{it}^{\intercal}(\theta^0-\hat{\theta})+\dot{U}_{it}\right)\leq 0\right\}.
\end{aligned}
$$

Let

$$
\begin{aligned}
A_{ik\tilde{k}} &= \frac{1}{T_0}\sum_t \left(\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\dot{U}_{it} + \frac{1}{T_0}\sum_t \left(\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\dot{X}_{it}^{\intercal}(\theta^0-\hat{\theta}) \\
&\quad + \frac{1}{T_0}\sum_t \left(\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\cdot\left(\dot{\delta}_t^0(k)-\frac{\hat{\delta}_t(k)+\hat{\delta}_t(\tilde{k})}{2}\right) \\
B_{ik\tilde{k}} &= \frac{1}{T_0}\sum_t \left(\delta_t^0(k)-\delta_t^0(\tilde{k})\right)\dot{U}_{it} + \frac{1}{2T_0}\sum_t \left(\delta_t^0(k)-\delta_t^0(\tilde{k})\right)^2.
\end{aligned}
$$

Note that $A_{ik\tilde{k}}$ depends on the estimator $(\hat{\theta},\hat{\delta},\hat{\gamma})$ while $B_{ik\tilde{k}}$ does not. Then,

$$
\Pr\left\{\hat{k}_i \neq k_i^0\right\} \leq \sum_k\sum_{\tilde{k}\neq k}\Pr\left\{A_{ik\tilde{k}}\leq 0\right\} \leq \sum_k\sum_{\tilde{k}\neq k}\Pr\left\{B_{ik\tilde{k}}\leq \left|B_{ik\tilde{k}}-A_{ik\tilde{k}}\right|\right\} \tag{3}
$$

We will show that $A_{ik\tilde{k}}$ and $B_{ik\tilde{k}}$ are sufficiently close to each other and that $\Pr\left\{B_{ik\tilde{k}}\leq 0\right\}$ converges to zero sufficiently fast.

$$
\begin{aligned}
\left|B_{ik\tilde{k}}-A_{ik\tilde{k}}\right| &\leq \left|\frac{1}{T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)\dot{U}_{it}\right| + \left|\frac{1}{T_0}\sum_t \left(\delta_t^0(\tilde{k})-\hat{\delta}_t(\tilde{k})\right)\dot{U}_{it}\right| \\
&\quad + \left|\frac{1}{T_0}\sum_t \left(\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\dot{X}_{it}^{\intercal}(\theta^0-\hat{\theta})\right| \\
&\quad + \left|\frac{1}{2T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)\cdot\left(-\delta_t^0(k)+\delta_t^0(\tilde{k})+\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\right| \\
&\quad + \left|\frac{1}{2T_0}\sum_t \left(\delta_t^0(\tilde{k})-\hat{\delta}_t(\tilde{k})\right)\cdot\left(\delta_t^0(k)-\delta_t^0(\tilde{k})++\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\right|.
\end{aligned}
$$

We apply Cauchy-Schwartz's inequality to each of the five terms so that we can use the consistency result in (2). For the first term,

$$\left|\frac{1}{T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)\dot{U}_{it}\right| \le \left(\frac{1}{T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)^2\right)^{\frac{1}{2}} \left(\frac{1}{T_0}\sum_t \dot{U}_{it}^2\right)^{\frac{1}{2}}$$

and similarly for the second term. As for the third term, from Assumption 7-b,

$$\begin{aligned}\left|\frac{1}{T_0}\sum_t \left(\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\dot{X}_{it}^{\intercal}(\theta^0-\hat{\theta})\right| &\le \frac{1}{T_0}\sum_t \left|\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right|\cdot\|\dot{X}_{it}\|_2\cdot\left\|\theta^0-\hat{\theta}\right\|_2 \\ &\le 2M\left(\frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2\right)\cdot\left\|\theta^0-\hat{\theta}\right\|_2\end{aligned}$$

Last, for the fourth term, from Assumption 7-b,

$$\begin{aligned}&\left|\frac{1}{2T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)\cdot\left(-\delta_t^0(k)+\delta_t^0(\tilde{k})+\hat{\delta}_t(k)-\hat{\delta}_t(\tilde{k})\right)\right| \\ &\le M\left(\frac{1}{T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)^2\right)^{\frac{1}{2}}\end{aligned}$$

and similarly for the fifth term. From Assumption 7-d, both $\frac{1}{T_0}\sum_t \dot{U}_{it}^2$ and $\frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2$ are bounded in expectation by the same bound for every n and thus $O_p(1)$. To use (2), choose an arbitrary $\eta > 0$ and focus only on the event of

$$\left\|\theta^0-\hat{\theta}\right\|_2, \left(\frac{1}{T_0}\sum_t \left(\delta_t^0(k)-\hat{\delta}_t(k)\right)\right)^{\frac{1}{2}} < \eta \tag{4}$$

for all k. When (4) is true, with some constant $C > 0$,

$$|B_{ik\tilde{k}}-A_{ik\tilde{k}}| \le \eta C\left(\left(\frac{1}{T_0}\sum_t \dot{U}_{it}^2\right)^{\frac{1}{2}}+\frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2+1\right).$$

Note that C only depend on M from Assumption 7 and does not depend on η. Let $D(\eta)$ be a binary random variable which equals one if (4) holds true for all k. Then,

$$
\begin{aligned}
&\Pr\left\{B_{ik\tilde{k}} \leq \left|B_{ik\tilde{k}} - A_{ik\tilde{k}}\right|, D(\eta) = 1\right\} \\
&\leq \Pr\left\{B_{ik\tilde{k}} \leq \eta C\left(\left(\frac{1}{T_0}\sum_t \dot{U}_{it}^2\right)^{\frac{1}{2}} + \frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2 + 1\right)\right\} \\
&\leq \Pr\left\{\frac{1}{T_0}\sum_t \dot{U}_{it}^2 \geq M^*\right\} + \Pr\left\{\frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2 \geq M^*\right\} \\
&\qquad + \Pr\left\{B_{ik\tilde{k}} \leq \eta C(M^* + \sqrt{M^*} + 1)\right\} \qquad (5)
\end{aligned}
$$

for any arbitrary $M^* > 0$. Let $M^* = \max\{4\sqrt{M} + 1, 4\tilde{M}\}$ since $\mathbf{E}[\dot{U}_{it}^2]$ is uniformly bounded by $4\sqrt{M}$ from Assumption 7-d.[2]

Now, we show that all three probabilities in (5) go to zero. For that, we use Lemma B5 of Bonhomme and Manresa (2015). For the first quantity, find that

$$
\begin{aligned}
\Pr\left\{\frac{1}{T_0}\sum_t \dot{U}_{it}^2 \geq M^*\right\} &\leq \Pr\left\{\frac{1}{T_0}\sum_t \dot{U}_{it}^2 \geq 4\sqrt{M} + 1\right\} \\
&\leq \Pr\left\{\frac{1}{T_0}\sum_t \left(\dot{U}_{it}^2 - \mathbf{E}[\dot{U}_{it}^2]\right) \geq 1\right\}.
\end{aligned}
$$

Let $Z_t = \dot{U}_{it}^2 - \mathbf{E}[\dot{U}_{it}^2]$. WTS $\{Z_t\}_{t=1}^{T_0}$ satisfies the condition given in Assumption 7-g.

$$
\begin{aligned}
\Pr\{|Z_t| \geq z\} &= \Pr\left\{|U_{it} - U_{it-1}| \geq \sqrt{\mathbf{E}[\dot{U}_{it}^2] + z}\right\} + \Pr\left\{|U_{it} - U_{it-1}| \leq \sqrt{\mathbf{E}[\dot{U}_{it}^2] - z}\right\} \\
&\leq \Pr\left\{|U_{it}| \geq \frac{\sqrt{\mathbf{E}[\dot{U}_{it}^2] + z}}{2}\right\} + \Pr\left\{|U_{it-1}| \geq \frac{\sqrt{\mathbf{E}[\dot{U}_{it}^2] + z}}{2}\right\} + \mathbf{1}\{z \leq \mathbf{E}[\dot{U}_{it}^2]\} \\
&\leq 2\exp\left(1 - \left(\frac{\sqrt{\mathbf{E}[\dot{U}_{it}^2] + z}}{2b}\right)^{d_2}\right) + \mathbf{1}\{z \leq \mathbf{E}[\dot{U}_{it}^2]\} \\
&\leq 2\exp\left(1 - \left(\frac{z}{2b}\right)^{d_2}\right) + \mathbf{1}\{z \leq \mathbf{E}[\dot{U}_{it}^2]\}.
\end{aligned}
$$

[2] In cases of the linear model for first-differenced outcomes and Theorem 1, a similar uniform bound on $\mathbf{E}[U_{it}^2]$ and $\mathbf{E}[(\dot{Y}_{it}(E_i) - \mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0])^2]$ can be found.

We want to find some $\tilde{b}$ and $\tilde{d}_2$ such that

$$\Pr\left\{|Z_t| \geq z\right\} \leq \exp\left(1 - \left(\frac{z}{\tilde{b}}\right)^{\tilde{d}_2}\right).$$

Note that the RHS crosses one when $z = \tilde{b}$. It suffices to show

$$2\exp\left(1 - \left(\frac{z}{2b}\right)^{d_2}\right) + \mathbf{1}\{z \leq \mathbf{E}[\dot{U}_{it}^2]\} \leq \exp\left(1 - \left(\frac{z}{\tilde{b}}\right)^{\tilde{d}_2}\right) \tag{6}$$

for $z \geq \tilde{b}$. Fix some $\tilde{d}_2 \in (0, d_2)$ and let

$$\tilde{b} = \max\left\{4\sqrt{M} + 1, 2b\left(1 + \log 2\right)^{\frac{1}{d_2}}, 2b\left(\frac{\tilde{d}_2}{d_2}\right)^{\frac{1}{d_2}}\right\}.$$

Since $\tilde{b} > \sqrt{M} \geq \mathbf{E}[\dot{U}_{it}^2]$, (6) for $z \geq \tilde{b}$ is equivalent with

$$\exp\left(\left(\frac{z}{2b}\right)^{d_2} - \left(\frac{z}{\tilde{b}}\right)^{\tilde{d}_2}\right) \geq 2 \qquad \Leftrightarrow \qquad \left(\frac{z}{2b}\right)^{d_2} - \left(\frac{z}{\tilde{b}}\right)^{\tilde{d}_2} \geq \log 2.$$

The inequality holds at $z = \tilde{b}$ and the LHS in the last inequality strictly increases in z since

$$\frac{d_2 z^{d_2 - 1}}{(2b)^{d_2}} - \frac{\tilde{d}_2 z^{\tilde{d}_2 - 1}}{\tilde{b}^{\tilde{d}_2}} = z^{\tilde{d}_2 - 1}\left(\frac{d_2}{(2b)^{d_2}} z^{d_2 - \tilde{d}_2} - \frac{\tilde{d}_2}{\tilde{b}^{\tilde{d}_2}}\right) \geq 0$$

for all $z \geq \tilde{b}$. Z_t is strongly mixing since $\dot{U}_{it}^2$ is a measurable function of (U_{it}, U_{it-1}). By adjusting a and d_1, we can satisfy Assumption 7-g for Z_t. Thus, from Lemma B5 of Bonhomme and Manresa (2015), for any $\nu > 0$,

$${T_0}^{\nu} \Pr\left\{\frac{1}{T_0}\sum_t \dot{U}_{it}^2 \geq M^*\right\} = o(1).$$

For Theorem 1, find that a similar result holds with $\Pr\left\{\frac{1}{T_0}\sum_t(\dot{Y}_{it}(e) - \mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0])^2 \geq M^*\right\}$. Since E_i has finite support, ${T_0}^{\nu} \Pr\left\{\frac{1}{T_0}\sum_t(\dot{Y}_{it}(E_i) - \mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0])^2 \geq M^*\right\} = o(1)$.

For the second quantity, find that

$$\Pr\left\{\frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2 \geq M^*\right\} \leq \Pr\left\{\frac{2}{T_0}\sum_{t=-T_0-1}^{-1} \|X_{it}\|_2 \geq 4\tilde{M}\right\}$$
$$\leq \Pr\left\{\frac{1}{T_0+1}\sum_{t=-T_0-1}^{-1} \|X_{it}\|_2 \geq \tilde{M}\right\}$$

From Assumption 7-d, for any $\nu > 0$,

$$T_0{}^{\nu}\Pr\left\{\frac{1}{T_0}\sum_t \|\dot{X}_{it}\|_2 \geq M^*\right\} = o(1).$$

For the last quantity, let $\eta^* = \frac{c^*}{4C(M^*+\sqrt{M^*}+1)}$ with $c^* = \frac{\min_{k,k'} c(k,k')}{2} > 0$. Then,

$$\begin{aligned}
&\Pr\left\{B_{ik\tilde{k}} \leq \eta^* C(M^* + \sqrt{M^*} + 1)\right\} \\
&\leq \Pr\left\{\frac{1}{T_0}\sum_t \left(\delta_t^0(k) - \delta_t^0(\tilde{k})\right)\dot{U}_{it} \leq \eta^* C(M^* + \sqrt{M^*} + 1) - \frac{c^*}{2}\right\} \\
&\qquad + \mathbf{1}\left\{\frac{1}{T_0}\sum_t \left(\delta_t^0(k) - \delta_t^0(\tilde{k})\right)^2 \leq c^*\right\} \\
&\leq \Pr\left\{\frac{1}{T_0}\sum_t \left(\delta_t^0(k) - \delta_t^0(\tilde{k})\right)\dot{U}_{it} \leq -\frac{c^*}{4}\right\} + \mathbf{1}\left\{\frac{1}{T_0}\sum_t \left(\delta_t^0(k) - \delta_t^0(\tilde{k})\right)^2 \leq c^*\right\}.
\end{aligned}$$

For the first term, use Lemma B5 of Bonhomme and Manresa (2015) again. From Assumption 7-b, we have

$$\Pr\left\{\left|\left(\delta_t^0(k) - \delta_t^0(\tilde{k})\right)\dot{U}_{it}\right| \geq z\right\} \leq \Pr\left\{|\dot{U}_{it}| \geq \frac{z}{2M}\right\}.$$

By applying similar argument from before, we can prove the tail property given in Assumption 7-g for $\left(\delta_t^0(k) - \delta_t^0(\tilde{k})\right)\dot{U}_{it}$ with any k and $\tilde{k}$. Also, the first part of Assumption 7-g is satisfied since $\left(\delta_t^0(k)-\delta_t^0(\tilde{k})\right)\dot{U}_{it}$ is a measurable function of $\left(U_{it},U_{it-1}\right)$.[3] For any $\nu>0$,

$$T_0{}^{\nu}\Pr\left\{\frac{1}{T_0}\sum_t\left(\delta_t^0(k)-\delta_t^0(\tilde{k})\right)\dot{U}_{it}\le-\frac{c^*}{4}\right\}=o(1).$$

Again, in the case of Theorem 1, note that E_i has finite support and repeat

$$T_0{}^{\nu}\Pr\left\{\frac{1}{T_0}\sum_t\left(\mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0=k]-\mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0=\tilde{k}]\right)\left(\dot{Y}_{it}(e)-\mathbf{E}[\dot{Y}_{it}(\infty)|k_i^0]\right)\right\}=o(1)$$

for every e. For the second term, Assumption 7-c assumes that $\mathbf{1}\{\frac{1}{T_0}\sum_t\left(\delta_t^0(k)-\delta_t^0(\tilde{k})\right)^2\le c^*\}=0$ when n is large and therefore $o(T^{-\nu})$ for any $\nu>0$.

Finally, going back to (3) and (5), thanks to K being fixed,

$$\Pr\left\{\hat{k}_i\neq k_i^0,D(\eta^*)=1\right\}=o(T^{-\nu}).\tag{7}$$

Step 6

In this step let us discuss the probability of assigning a wrong type at least to one unit. As $n\to\infty$, for any $\nu>0$

$$\begin{aligned}
&\Pr\left\{\sup_i\mathbf{1}_{\{\hat{k}_i\neq k_i^0\}}>0\right\}\\
&\le\Pr\left\{\sum_i\mathbf{1}_{\{\hat{k}_i\neq k_i^0\}}>0,D(\eta^*)=1\right\}+\Pr\{D(\eta^*)=0\}\\
&\le n\cdot\Pr\left\{\hat{k}_i\neq k_i^0,D(\eta^*)=1\right\}+\Pr\{D(\eta^*)=0\}\\
&=o(nT_0{}^{-\nu})+o(1).
\end{aligned}$$

The last equality holds from (7).

□

[3] Here, I am treating $\left\{\delta_t^0(k)\right\}_{t,k}$ as if uniformly fixed across n. This can be relaxed by assuming $\left\{\delta_t^0(k)\right\}_{t,k}$ is also a strongly mixing random process as in Bonhomme and Manresa (2015).

3 Proof for Corollary 3

The first part of the proof is the same with Corollary 2. The second part follows the proof of Theorem 2 of Callaway and Sant'Anna (2021). Fix some t, k and e such that $0 \le e \le t \le T_1 - 1$ and $\mu(k,e) > 0$. Then, it satisfies that $t - e \le \bar{r}_k$ from Assumption 6.

Step 1

Firstly, let us show that $\widehat{CATT}_t(k,e)$ is close to the infeasible estimator using the true types $\{k_i^0\}_{i=1}^n$:

$$\begin{aligned}\widetilde{CATT}_t(k,e) = & \frac{\sum_{i=1}^n \left(Y_{it} - Y_{i,e-1}\right) \mathbf{1}\{k_i^0 = k, E_i = e\}}{\sum_{i=1}^n \mathbf{1}\{k_i^0 = k, E_i = e\}} \\ & - \frac{\sum_{i=1}^n \left(Y_{it} - Y_{i,e-1}\right) \mathbf{1}\{k_i^0 = k, E_i = \infty\}\pi_e(X_i, k, \hat{\xi})/\pi_\infty(X_i, k, \hat{\xi})}{\sum_{i=1}^n \mathbf{1}\{k_i^0 = k, E_i = \infty\}\pi_e(X_i, k, \hat{\xi})/\pi_\infty(X_i, k, \hat{\xi})}.\end{aligned}$$

Find that

$$\begin{aligned}&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(Y_{it} - Y_{i,e-1}\right)\left(\mathbf{1}\{\hat{k}_i = k, E_i = e\} - \mathbf{1}\{k_i^0 = k, E_i = e\}\right)\frac{\pi_e(X_i, k, \hat{\xi})}{\pi_\infty(X_i, k, \hat{\xi})}\right| \\ &\le \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \left(Y_{i,e+r} - Y_{i,e-1}\right)^2\right)^{\frac{1}{2}} \cdot \left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\hat{k}_i \neq k_i^0\}\right)^{\frac{1}{2}} \sup_i \left|\frac{\pi_e(X_i, k, \hat{\xi})}{\pi_\infty(X_i, k, \hat{\xi})}\right|.\end{aligned}$$

$\sup_i \pi_e/\pi_\infty$ is bounded by $1/\varepsilon^\pi$ from Assumption 9-c. $\frac{1}{n}\sum_{i=1}^n \left(Y_{i,e+r} - Y_{i,e-1}\right)^2$ is bounded in expectation uniformly over e and r from Assumption 9-a and therefore $O_p(1)$. From Theorem 2,

$$\Pr\left\{\sum_{i=1}^n \mathbf{1}\{\hat{k}_i \neq k_i\} > \varepsilon^2\right\} \le \Pr\left\{\sup_i \mathbf{1}\{\hat{k}_i \neq k_i\} > 0\right\} = o(n{T_0}^{-\nu}) + o(1)$$

for any $\nu, \epsilon > 0$. Since $n{T_0}^{-\nu^*} \to 0$ as $n \to \infty$, $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\hat{k}_i \neq k_i^0\}\right)^{\frac{1}{2}} = o_p(1)$.

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_{it} - Y_{i,e-1})\mathbf{1}\{\hat{k}_i = k, E_i = e\}\frac{\pi_e(X_i, k, \hat{\xi})}{\pi_\infty(X_i, k, \hat{\xi})} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_{it} - Y_{i,e-1})\mathbf{1}\{k_i^0 = k, E_i = e\}\frac{\pi_e(X_i, k, \hat{\xi})}{\pi_\infty(X_i, k, \hat{\xi})} + o_p(1)
\end{aligned}$$

By the same argument,

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\hat{k}_i = k, E_i = e\}\frac{\pi_e(X_i, k, \hat{\xi})}{\pi_\infty(X_i, k, \hat{\xi})} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{k_i^0 = k, E_i = e\}\frac{\pi_e(X_i, k, \hat{\xi})}{\pi_\infty(X_i, k, \hat{\xi})} + o_p(1).$$

The same applies to the other term without π_e/π_∞. Note that $\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{k_i^0 = k, E_i = e\}$ and $\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{k_i^0 = k, E_i = \infty\}\frac{\pi_e}{\pi_\infty}$ both have nonzero probabilistic limits; for the latter, apply Assumption 9-c and find that it is bounded from below by $\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{k_i^0 = k, E_i = \infty\}\varepsilon^\pi$. Thus,

$$\sqrt{n}\left(\widehat{CATT}_t(k, e) - \widetilde{CATT}_t(k, e)\right) = o_p(1).$$

Step 2

In this step, we rewrite $CATT_t(k, e)$ in a way that it connects to $\widetilde{CATT}_t(k, e)$:

$$CATT_t(k, e) = \mathbf{E}\left[Y_{it}(e) - Y_{i,e-1}(e) | k_i^0 = k, E_i = e\right] - \mathbf{E}\left[Y_{it}(\infty) - Y_{i,e-1}(\infty) | k_i^0 = k, E_i = e\right].$$

Find that

$$\begin{aligned}
&\mathbf{E}\left[Y_{it}(\infty) - Y_{i,e-1}(\infty) | k_i^0 = k, E_i = e\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[Y_{it}(\infty) - Y_{i,e-1}(\infty) | X_i, k_i^0 = k\right] | k_i^0 = k, E_i = e\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[Y_{it} - Y_{i,e-1} | X_i, k_i^0 = k, E_i = \infty\right] | k_i^0 = k, E_i = e\right] \\
&= \frac{\mathbf{E}\left[\mathbf{E}\left[Y_{it} - Y_{i,e-1} | X_i, k_i^0 = k, E_i = \infty\right]\mathbf{1}\{k_i^0 = k, E_i = e\}\right]}{\Pr\{k_i^0 = k, E_i = e\}}
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbf{E}\left[\mathbf{E}\left[Y_{it}-Y_{i,e-1}|X_i,k_i^0=k,E_i=\infty\right]\mathbf{1}\{k_i^0=k,E_i=e\}\right]\\
&=\mathbf{E}\left[\frac{\mathbf{E}\left[(Y_{it}-Y_{i,e-1})\,\mathbf{1}\{k_i^0=k,E_i=\infty\}|X_i\right]\Pr\left\{k_i^0=k,E_i=e|X_i\right\}}{\Pr\left\{k_i^0=k,E_i=\infty|X_i\right\}}\right]\\
&=\mathbf{E}\left[(Y_{it}-Y_{i,e-1})\,\mathbf{1}\{k_i^0=k,E_i=\infty\}\cdot\frac{\Pr\left\{k_i^0=k,E_i=e|X_i\right\}}{\Pr\left\{k_i^0=k,E_i=\infty|X_i\right\}}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
\Pr\left\{k_i^0=k,E_i=e\right\}&=\mathbf{E}\left[\mathbf{1}\{k_i^0=k,E_i=e\}\cdot\frac{\Pr\left\{k_i^0=k,E_i=\infty|X_i\right\}}{\Pr\left\{k_i^0=k,E_i=\infty|X_i\right\}}\right]\\
&=\mathbf{E}\left[\mathbf{1}\{k_i^0=k,E_i=\infty\}\cdot\frac{\Pr\left\{k_i^0=k,E_i=e|X_i\right\}}{\Pr\left\{k_i^0=k,E_i=\infty|X_i\right\}}\right].
\end{aligned}
$$

The second to the last equality holds since $\Pr\left\{E_i=\infty|k_i^0,X_i\right\}\geq\varepsilon^\pi>0$ from Assumption 9-c and $\mu(k,\infty)>0$ for every k from Assumption 6.

For notational brevity, let

$$
\begin{aligned}
W_i&=\mathbf{1}\{k_i^0=k,E_i=\infty\}\pi_e(X_i,k,\xi^0)/\pi_\infty(X_i,k,\xi^0),\\
\widehat{W}_i&=\mathbf{1}\{k_i^0=k,E_i=\infty\}\pi_e(X_i,k,\hat{\xi})/\pi_\infty(X_i,k,\hat{\xi}).
\end{aligned}
$$

Then,

$$
\begin{aligned}
CATT_t(k,e)&=\frac{\mathbf{E}\left[(Y_{it}-Y_{i,e-1})\mathbf{1}\{k_i^0=k,E_i=e\}\right]}{\mathbf{E}\left[\mathbf{1}\{k_i^0=k,E_i=e\}\right]}-\frac{\mathbf{E}\left[(Y_{it}-Y_{i,e-1})W_i\right]}{\mathbf{E}\left[W_i\right]}\\
\widetilde{CATT}_t(k,e)&=\frac{\frac{1}{n}\sum_i(Y_{it}-Y_{i,e-1})\mathbf{1}\{k_i^0=k,E_i=e\}}{\frac{1}{n}\sum_i\mathbf{1}\{k_i^0=k,E_i=e\}}-\frac{\frac{1}{n}\sum_i(Y_{it}-Y_{i,e-1})\widehat{W}_i}{\frac{1}{n}\sum_i\widehat{W}_i}
\end{aligned}
$$

Step 3

Now, let us derive an asymptotic linear approximation of $\widetilde{CATT}_t(k,e)$. Find that

$$\sqrt{n}\left(\widetilde{CATT}_t(k,e) - CATT_t(k,e)\right) = A_n - B_n$$

where

$$A_n = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_{it} - Y_{i,e-1}\right)\mathbf{1}\{k_i^0 = k, E_i = e\}}{\hat{\mu}(k,e)} - \sqrt{n}\frac{\mathbf{E}\left[(Y_{it} - Y_{i,e-1})\mathbf{1}\{k_i^0 = k, E_i = e\}\right]}{\mu(k,e)}$$
$$B_n = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_{it} - Y_{i,e-1}\right)\widehat{W}_i}{\overline{\widehat{W}}_n} - \sqrt{n}\frac{\mathbf{E}\left[(Y_{it} - Y_{i,e-1})W_i\right]}{\mathbf{E}\left[W_i\right]}$$

where $\overline{\widehat{W}}_n = \frac{1}{n}\sum_{i=1}^{n}\widehat{W}_i$.

Before deriving the asymptotic approximation, let us provide some useful expansions and probabilistic convergences. Firstly, apply the first-order Taylor's expansion to $\widehat{W}_i$ with regard to $\hat{\xi}$ around ξ^0:

$$\widehat{W}_i = W_i + \mathbf{1}\{k_i^0 = k, E_i = \infty\}\frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)^{\intercal}\Bigg|_{\xi\in\left(\xi^0,\hat{\xi}\right)}\left(\hat{\xi} - \xi^0\right). \tag{8}$$

The first-order remainder term is $O_p(1/\sqrt{n})$ since $\|\hat{\xi} - \xi^0\|_2 = O_p(1/\sqrt{n})$ from asymptotic normality of $\hat{\xi}$ and $\frac{\partial}{\partial \xi}\frac{\pi_e}{\pi_\infty} = O_p(1)$ from Assumption 9-d and the convergence of $\hat{\xi}$ to ξ^0:

$$\begin{aligned}\left|\frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)^{\intercal}\Bigg|_{\xi\in\left(\xi^0,\hat{\xi}\right)}\left(\hat{\xi} - \xi^0\right)\right| &\leq \left\|\frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)\Bigg|_{\xi\in\left(\xi^0,\hat{\xi}\right)}\right\|_2\left\|\hat{\xi} - \xi^0\right\|_2 \\ &= O_p(1)O_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Now, apply the second-order Taylor's expansion to $\widehat{W}_i$:

$$\widehat{W}_i = W_i + \mathbf{1}\{k_i^0 = k, E_i = \infty\} \frac{\partial}{\partial \xi} \left(\frac{\pi_e(X_i, k, \xi)}{\pi_\infty(X_i, k, \xi)} \right)^{\intercal} \Bigg|_{\xi = \xi^0} \left(\hat{\xi} - \xi^0 \right)$$
$$+ \mathbf{1}\{k_i^0 = k, E_i = \infty\} \left(\hat{\xi} - \xi^0 \right)^{\intercal} \frac{\partial^2}{\partial \xi \partial \xi^{\intercal}} \left(\frac{\pi_e(X_i, k, \xi)}{\pi_\infty(X_i, k, \xi)} \right) \Bigg|_{\xi \in \left(\xi^0, \hat{\xi}\right)} \left(\hat{\xi} - \xi^0 \right). \tag{9}$$

Note that the second-order remainder term is $o_p(1/\sqrt{n})$ from Assumption 9-d and the asymptotic normality of $\hat{\xi}$. An abuse of notation is used when we write $\xi \in (\xi^0, \hat{\xi})$ to say ξ lies between ξ^0 and $\hat{\xi}$. Lastly, find that from (8) and $\frac{1}{n}\sum_i (Y_{it} - Y_{i,e-1})^2$ being bounded in expectation,

$$\left| \frac{1}{n} \sum_{i=1}^{n} (Y_{it} - Y_{i,e-1}) \left(\widehat{W}_i - W_i \right) \right| = O_p\left(\frac{1}{\sqrt{n}} \right),$$
$$\frac{1}{n} \sum_{i=1}^{n} (Y_{it} - Y_{i,e-1}) \widehat{W}_i = \mathbf{E}\left[(Y_{it} - Y_{i,e-1}) W_i \right] + O_p\left(\frac{1}{\sqrt{n}} \right). \tag{10}$$

The $O_p(1/\sqrt{n})$ term in the second equality comes from applying the CLT to $(Y_{it} - Y_{i,e-1}) W_i$ and the $O_p(1/\sqrt{n})$ term from the first equality. Likewise, we have

$$\overline{\widehat{W}}_n = \mathbf{E}[W_i] + O_p(1/\sqrt{n}). \tag{11}$$

As argued in the Step 1, $\mathbf{E}[W_i] > 0$ from Assumption 9-c.

To drive the asymptotic approximation of B_n, apply the second-order Taylor's expansion to B_n with regard to $\overline{\widehat{W}}_n$ around $\mathbf{E}[W_i]$:

$$\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{it} - Y_{i,e-1}) \widehat{W}_i}{\overline{\widehat{W}}_n}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{it} - Y_{i,e-1}) \widehat{W}_i \left(\frac{1}{\mathbf{E}[W_i]} - \frac{1}{\mathbf{E}[W_i]^2} \left(\overline{\widehat{W}}_n - \mathbf{E}[W_i] \right) + \frac{2}{\widetilde{W}_n^3} \left(\overline{\widehat{W}}_n - \mathbf{E}[W_i] \right)^2 \right)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(Y_{it} - Y_{i,e-1}) \widehat{W}_i}{\mathbf{E}[W_i]} - \frac{\mathbf{E}[(Y_{it} - Y_{i,e-1}) W_i]}{\mathbf{E}[W_i]^2} \sqrt{n} \left(\overline{\widehat{W}}_n - \mathbf{E}[W_i] \right) + o_p(1).$$

with some $\widetilde{W}_n$ between $\widehat{\widehat{W}}_n$ and $\mathbf{E}[W_i]$. The second equality holds from $\mathbf{E}[W_i] > 0$, (11) and (10). Then, from (9) and $\frac{1}{n}\sum_i (Y_{it} - Y_{i,e-1})^2$ being bounded in expectation,

$$
\begin{aligned}
&\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_{it} - Y_{i,e-1})\widehat{W}_i}{\widehat{\widehat{W}}_n} \\
&= \frac{1}{\mathbf{E}[W_i]} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{it} - Y_{i,e-1}) W_i + o_p(1) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \frac{(Y_{it} - Y_{i,e-1})\mathbf{1}\{k_i^0 = k, E_i = \infty\}}{\mathbf{E}[W_i]} \frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)^\intercal \Bigg|_{\xi=\xi^0} \cdot \sqrt{n}\left(\hat{\xi} - \xi^0\right) \\
&\quad - \frac{\mathbf{E}[(Y_{it} - Y_{i,e-1})W_i]}{\mathbf{E}[W_i]^2} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i - \mathbf{E}[W_i]) \\
&\quad - \frac{\mathbf{E}[(Y_{it} - Y_{i,e-1})W_i]}{\mathbf{E}[W_i]} \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{1}\{k_i^0 = k, E_i = \infty\}}{\mathbf{E}[W_i]} \frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)^\intercal \Bigg|_{\xi=\xi^0} \cdot \sqrt{n}\left(\hat{\xi} - \xi^0\right).
\end{aligned}
$$

Let

$$
\begin{aligned}
\bar{B}_1 &= \frac{1}{\mathbf{E}[W_i]} \cdot \mathbf{E}\left[(Y_{it} - Y_{i,e-1})\,\mathbf{1}\{k_i^0 = k, E_i = \infty\}\frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)\Bigg|_{\xi=\xi^0}\right] \\
\bar{B}_2 &= \frac{1}{\mathbf{E}[W_i]} \cdot \mathbf{E}\left[\mathbf{1}\{k_i^0 = k, E_i = \infty\}\frac{\partial}{\partial \xi}\left(\frac{\pi_e(X_i,k,\xi)}{\pi_\infty(X_i,k,\xi)}\right)\Bigg|_{\xi=\xi^0}\right].
\end{aligned}
$$

Note that the sample analogues for $\bar{B}_1$ and $\bar{B}_2$ with ξ^0 replaced with $\hat{\xi}$ are consistent for $\bar{B}_1$ and $\bar{B}_2$ from Assumption 9-d. Consequently,

$$
\begin{aligned}
B_n = &\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_i}{\mathbf{E}[W_i]}\left(Y_{it} - Y_{i,e-1} - \frac{\mathbf{E}[(Y_{it} - Y_{i,e-1})W_i]}{\mathbf{E}[W_i]}\right) \\
&+ \left(\bar{B}_1 - \frac{\mathbf{E}[(Y_{it} - Y_{i,e-1})W_i]}{\mathbf{E}[W_i]}\bar{B}_2\right)^\intercal \cdot \sqrt{n}\left(\hat{\xi} - \xi^0\right) + o_p(1).
\end{aligned}
$$

By repeating the same argument for A_n,

$$
A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\mathbf{1}\{k_i^0 = k, E_i = e\}}{\mu(k,e)}\left(Y_{it} - Y_{i,e-1} - \frac{\mathbf{E}[(Y_{it} - Y_{i,e-1})\mathbf{1}\{k_i^0 = k, E_i = e\}]}{\mu(k,e)}\right) + o_p(1).
$$

Note the asymptotic linear approximation given in Corollary 3 holds for $\hat{\xi}$ as well from the proof for Corollary 2. We can construct score functions l^1 and l^0 as follows:

$$
\begin{aligned}
A_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^1_{tke} \left(\{Y_{it}\}_{t \geq -1}, k^0_i, E_i \right) + o_p(1), \\
B_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^0_{tke} \left(\{Y_{it}\}_{t \geq -1}, X_i, k^0_i, E_i \right) + o_p(1).
\end{aligned}
$$

Note that l^π appears in l^0. Now we have

$$
\begin{aligned}
&\sqrt{n} \left(\widehat{CATT}_t(k,e) - CATT_t(k,e) \right) \\
&= \begin{pmatrix} 1, -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^1_{tke} \left(\{Y_{it}\}_{t \geq -1}, k^0_i, E_i \right) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^0_{tke} \left(\{Y_{it}\}_{t \geq -1}, X_i, k^0_i, E_i \right) \end{pmatrix} + o_p(1).
\end{aligned}
$$

The asymptotic linear approximation is derived for $\widehat{CATT}_t(k,e)$.

Step 4

To derive asymptotic distribution of $\hat{\beta}_r(k)$, consider

$$
\begin{aligned}
&\frac{\hat{\mu}(k,e)}{\sum_{e' \leq T_1-1-r} \hat{\mu}(k,e')} \cdot \sqrt{n}\widehat{CATT}_t(k,e) - \frac{\mu(k,e)}{\sum_{e' \leq T_1-1-r} \mu(k,e')} \cdot \sqrt{n} CATT_t(k,e) \\
&= \frac{\hat{\mu}(k,e)}{\sum_{e' \leq T_1-1-r} \hat{\mu}(k,e')} \cdot \sqrt{n} \left(\widehat{CATT}_t(k,e) - CATT_t(k,e) \right) \\
&\quad + \sqrt{n} \left(\frac{\hat{\mu}(k,e)}{\sum_{e' \leq T_1-1-r} \hat{\mu}(k,e')} - \frac{\mu(k,e)}{\sum_{e' \leq T_1-1-r} \mu(k,e')} \right) \cdot CATT_t(k,e).
\end{aligned}
$$

By taking the second-order Taylor's expansion of $\sum_{e'} \hat{\mu}(k,e')$ around $\sum_{e'} \mu(k,e')$,

$$
\begin{aligned}
\sqrt{n}\left(\frac{\hat{\mu}(k,e)}{\sum_{e'}\hat{\mu}(k,e')} - \frac{\mu(k,e)}{\sum_{e'}\mu(k,e')}\right) = \sqrt{n}&\left(\frac{\hat{\mu}(k,e)}{\sum_{e'}\mu(k,e')} - \frac{\mu(k,e)}{\sum_{e'}\mu(k,e')}\right)\\
&- \frac{\hat{\mu}(k,e)}{(\sum_{e'}\mu(k,e'))^2}\sqrt{n}\left(\sum_{e'}\Big(\hat{\mu}(k,e') - \mu(k,e')\Big)\right)\\
&+ \frac{2\hat{\mu}(k,e)}{\tilde{\mu}^3}\sqrt{n}\left(\sum_{e'}\Big(\hat{\mu}(k,e') - \mu(k,e')\Big)\right)^2
\end{aligned}
$$

with some $\tilde{\mu}$ between $\sum_{e'} \mu(k,e')$ and $\sum_{e'} \hat{\mu}(k,e')$. The second-order remainder term is $o_p(1)$ since $\sqrt{n}\,(\sum_{e'}(\hat{\mu}(k,e') - \mu(k,e'))) = O_p(1)$ and $\sum_{e'} \mu(k,e')$ is nonzero by taking $r \leq \bar{r}_k$ from Assumption 6. Thus,

$$
\begin{aligned}
&\sqrt{n}\left(\frac{\hat{\mu}(k,e)}{\sum_{e'}\hat{\mu}(k,e')} - \frac{\mu(k,e)}{\sum_{e'}\mu(k,e')}\right)\\
&= \sqrt{n}\left(\frac{\hat{\mu}(k,e) - \mu(k,e)}{\sum_{e'}\mu(k,e')}\right) - \frac{\mu(k,e)}{(\sum_{e'}\mu(k,e'))^2}\sqrt{n}\left(\sum_{e'}\Big(\hat{\mu}(k,e') - \mu(k,e')\Big)\right) + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{1}\{k_i^0 = k, E_i = e\} - \mu(k,e)}{\sum_{e'}\mu(k,e')}\\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mu(k,e)\big(\mathbf{1}\{k_i^0 = k, E_i \leq T_1 - 1 - r\} - \sum_{e'}\mu(k,e')\big)}{\big(\sum_{e'}\mu(k,e')\big)^2} + o_p(1).
\end{aligned}
$$

Let l^μ denote the score function in the asymptotic linear approximation:

$$
\sqrt{n}\left(\frac{\hat{\mu}(k,e)}{\sum_{e'}\hat{\mu}(k,e')} - \frac{\mu(k,e)}{\sum_{e'}\mu(k,e')}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_{ke}^{\mu}(k_i^0, E_i) + o_p(1).
$$

Combining all of the results so far, we get

$$
\begin{aligned}
&\sqrt{n}\left(\hat{\beta}_r(k) - \beta_r(k)\right) \\
&= \sum_{e \leq T_1-1-r} \left(\frac{\hat{\mu}(k,e)}{\sum_{e'} \hat{\mu}(k,e')} \cdot \sqrt{n}\widehat{CATT}_t(k,e) - \frac{\mu(k,e)}{\sum_{e'} \mu(k,e')} \cdot \sqrt{n} CATT_t(k,e) \right) \\
&= \sum_{e \leq T_1-1-r} \frac{\mu(k,e)}{\sum_{e'} \mu(k,e')} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(l^1_{e+r,k,e}\left(\{Y_{it}\}_{t\geq 0}, k_i^0, E_i\right) - l^0_{e+r,k,e}\left(\{Y_{it}\}_{t\geq 0}, X_i, k_i^0, E_i\right) \right) \\
&\qquad + \sum_{e \leq T_1-1-r} CATT_t(k,e) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^{\mu}_{ke}(k_i^0, E_i) + o_p(1).
\end{aligned}
$$

References

Bonhomme, Stéphane and Elena Manresa, “Grouped patterns of heterogeneity in panel data,” *Econometrica*, 2015, *83* (3), 1147–1184.

Callaway, Brantly and Pedro HC Sant’Anna, “Difference-in-differences with multiple time periods,” *Journal of Econometrics*, 2021, *225* (2), 200–230.

Janys, Lena and Bettina Siflinger, “Mental health and abortions among young women: Time-varying unobserved heterogeneity, health behaviors, and risky decisions,” *Journal of Econometrics*, 2024, *238* (1), 105580.