Web Appendix for "Inference with 'Difference in Differences' with a Small Number of Policy Changes"

Timothy G. Conley[1]
Graduate School of Business
University of Chicago
and
Christopher R. Taber
Department of Economics
University of Wisconsin-Madison

March 19, 2009

[1]We thank Federico Bandi, Alan Bester, Phil Cross, Chris Hansen, Rosa Matzkin, Bruce Meyer, Jeff Russell, and Elie Tamer for helpful comments and Aroop Chaterjee and Nathan Hendren for research assistantship. All errors are our own. Conley gratefully acknowledges financial support from the NSF (SES 9905720) and from the IBM Corporation Faculty Research Fund at the University of Chicago Graduate School of Business. Taber gratefully acknowledges financial support from the NSF (SES 0217032). Stata and Matlab Code to implement the methods here can be found at "http://faculty.chicagogsb.edu/timothy.conley/research/."

1 Appendix

A.1 Proof of Proposition 3

First a standard application of the partitioned inverse theorem and using the fact that $\widetilde{D}'_{jt} = 0$ for $j > N_1$,

$$\begin{aligned}
\widehat{\beta} &= \beta + \left(\frac{1}{N_0+N_1} \sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{X}'_{jt} \right. \qquad \text{(A-1)} \\
&\left. - \frac{1}{N_0+N_1} \left[\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{D}'_{jt} \right] \left[\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{D}_{jt}\widetilde{D}'_{jt} \right]^{-1} \left[\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{D}_{jt}\widetilde{X}'_{jt} \right] \right)^{-1} \\
&\times \left(\frac{1}{N_0+N_1} \sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{\eta}_{jt} \right. \\
&\left. - \frac{1}{N_0+N_1} \left[\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{D}'_{jt} \right] \left[\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{D}_{jt}\widetilde{D}'_{jt} \right]^{-1} \left[\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{D}_{jt}\widetilde{\eta}'_{jt} \right] \right).
\end{aligned}$$

Now consider each piece in turn.

First Assumption 1 states that

$$\frac{1}{N_0+N_1} \sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{X}'_{jt} \xrightarrow{p} \Sigma_x < \infty.$$

The mixing component of Assumption 1 imply that:

$$\frac{1}{N_0+N_1} \sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{\eta}_{jt} \xrightarrow{p} E\left[\sum_{t=1}^{T} \widetilde{X}_{jt}\widetilde{\eta}_{jt} \right] = 0.$$

Consider the ℓ^{th} element of the vector $\widetilde{D}_{jt}(\ell)$,

$$\widetilde{D}_{jt}(\ell) = \begin{cases} d_{jt} - \overline{d_j} - \frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0} & j = \ell \\ -\frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0} & \text{otherwise} \end{cases}.$$

The ℓ^{th} diagonal component of $\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{D}_{jt}\widetilde{D}'_{jt}$ can be written as

$$\begin{aligned}
&\sum_{t=1}^{T} \left(d_{\ell t} - \overline{d_\ell} - \frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0} \right)^2 + \sum_{\substack{j=1 \\ j\neq\ell}}^{N_1+N_0} \sum_{t=1}^{T} \left(-\frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0} \right)^2 \\
&= \sum_{t=1}^{T} \left(d_{\ell t} - \overline{d_\ell} - \frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0} \right)^2 + \frac{(N_1+N_0-1)\sum_{t=1}^{T}\left(\overline{d_\ell} - d_{\ell t}\right)}{(N_1+N_0)^2} \\
&\xrightarrow{p} \sum_{t=1}^{T} \left(d_{\ell t} - \overline{d_\ell} \right)^2.
\end{aligned}$$

A generic (ℓ, k) off diagonal term can be written as

$$
\begin{aligned}
& \left(d_{\ell t} - \overline{d_\ell} - \frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0}\right)\left(-\frac{d_{kt}}{N_1+N_0} + \frac{\overline{d_k}}{N_1+N_0}\right) \\
& + \left(d_{kt} - \overline{d_k} - \frac{d_{kt}}{N_1+N_0} + \frac{\overline{d_k}}{N_1+N_0}\right)\left(-\frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0}\right) \\
& + \sum_{\substack{j=1 \\ j\neq\ell}}^{N_1+N_0}\left(-\frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0}\right)\left(-\frac{d_{kt}}{N_1+N_0} + \frac{\overline{d_k}}{N_1+N_0}\right) \\
& \xrightarrow{p} 0
\end{aligned}
$$

The ℓ^{th} column of $\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{D}'_{jt}$ can be written as

$$
\begin{aligned}
& \sum_{t=1}^{T}\widetilde{X}_{\ell t}\left(d_{\ell t} - \overline{d_\ell}\right) + \sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{X}_{jt}\left(-\frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0}\right) \\
= & \sum_{t=1}^{T}\widetilde{X}_{\ell t}\left(d_{\ell t} - \overline{d_\ell}\right) + \sum_{t=1}^{T}\left(-\frac{d_{\ell t}}{N_1+N_0} + \frac{\overline{d_\ell}}{N_1+N_0}\right)\sum_{j=1}^{N_1+N_0}\widetilde{X}_{jt} \\
= & \sum_{t=1}^{T}\widetilde{X}_{\ell t}\left(d_{\ell t} - \overline{d_\ell}\right),
\end{aligned}
$$

since $\sum_{j=1}^{N_1+N_0}\widetilde{X}_{jt} = 0$ as shown in the proof of Proposition 1.

Similarly the ℓ^{th} column of $\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{\eta}_{jt}$ is equal to $\sum_{t=1}^{T}\left(d_{\ell t} - \overline{d_\ell}\right)\left(\eta_{\ell t} - \overline{\eta_\ell}\right)$.

Thus since all of these terms are $O_p(1)$,

$$
\begin{aligned}
& \frac{1}{N_0+N_1}\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{D}'_{jt}\right]\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{X}'_{jt}\right] \xrightarrow{p} 0 \\
& \frac{1}{N_0+N_1}\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{D}'_{jt}\right]\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{\eta}_{jt}\right] \xrightarrow{p} 0
\end{aligned}
$$

Consistency for $\widehat{\beta}$ follows upon plugging the pieces back into (A-1) and applying Slutsky's theorem.

From the normal equation for $\widehat{\alpha}_j$ it is straightforward to show that

$$\begin{aligned}\widehat{A} &= \left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{D}_{jt}\left(\widetilde{Y}_{jt}-\widetilde{X}'_{jt}\widehat{\beta}\right)\\
&= \left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{D}_{jt}\left(\widetilde{D}'_{jt}A+\widetilde{\eta}_{jt}+\widetilde{X}'_{jt}\left(\beta-\widehat{\beta}\right).\right)\\
&= A+\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{\eta}_{jt}+\\
&+\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{X}'_{jt}\left(\beta-\widehat{\beta}\right).\end{aligned}$$

Since $\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}$ and $\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{X}'_{jt}$ are $O_p(1)$ (as shown above) and $\widehat{\beta}$ is consistent, the last term of this expression converges to zero.

We derived the limits of $\left[\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{D}'_{jt}\right]^{-1}$ and $\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{D}_{jt}\widetilde{\eta}_{jt}$ above, putting this together one sees that for each $j=1,...,N_1$

$$\widehat{A}(\ell)\xrightarrow{p}A(\ell)+\frac{\sum_{t=1}^{T}\left(d_{\ell t}-\overline{d}_{\ell}\right)\left(\eta_{\ell t}-\overline{\eta}_{\ell}\right)}{\sum_{t=1}^{T}\left(d_{\ell t}-\overline{d}_{\ell}\right)^2}$$

where $\widehat{A}(\ell)$ and $A(\ell)$ are the ℓ^{th} component of $\widehat{A}$ and A respectively.

This gives the result. ■

A.2 Individual Data

The use of individual-level data has three complications relative to the model in Section 2. We need to worry about a) the fact that $\widehat{\delta}$ is estimated, b) the term $\frac{1}{|M(j,t)|}\sum_{i\in M(j,t)}Z'_i$ which can change with the sample size, and c) the error term involving ε_i. None of these issues are particularly difficult to deal with, but do require verifying that the proofs still hold. We consider each case in turn. However, we first define some notation which is analogous to our earlier notation. Define:

$$\begin{aligned}\widetilde{\overline{\varepsilon}}_{jt} &\equiv \frac{1}{|M(j,t)|}\sum_{i\in M(j,t)}\varepsilon_i-\frac{1}{T\,|M(j,t)|}\sum_{\tau=1}^{T}\sum_{i\in M(j,\tau)}\varepsilon_i\\
&\quad -\frac{1}{N_0}\sum_{\ell=N_1+1}^{N_1+N_0}\frac{1}{|M(\ell,t)|}\sum_{i\in M(\ell,t)}\varepsilon_i+\frac{1}{TN_0}\sum_{\tau=1}^{T}\sum_{\ell=N_1+1}^{N_1+N_0}\sum_{i\in M(\ell,\tau)}\varepsilon_i\\
\widetilde{\overline{Z}}_{jt} &\equiv \frac{1}{|M(j,t)|}\sum_{i\in M(j,t)}Z_i-\frac{1}{T\,|M(j,t)|}\sum_{\tau=1}^{T}\sum_{i\in M(j,\tau)}Z_i\\
&\quad -\frac{1}{N_0}\sum_{\ell=N_1+1}^{N_1+N_0}\frac{1}{|M(\ell,t)|}\sum_{i\in M(\ell,t)}Z_i+\frac{1}{TN_0}\sum_{\tau=1}^{T}\sum_{\ell=N_1+1}^{N_1+N_0}\sum_{i\in M(\ell,\tau)}Z_i.\end{aligned}$$

It is straight forward to verify that

$$\begin{aligned}\widetilde{\widetilde{\lambda}}_{jt} &= \widetilde{\lambda}_{jt} + \widetilde{\widetilde{Z}}'_{jt}\left(\delta - \widehat{\delta}\right) + \widetilde{\widetilde{\varepsilon}}_{jt} \\ &= \alpha \widetilde{d}_{jt} + \widetilde{X}'_{jt}\beta + \widetilde{\widetilde{Z}}'_{jt}\left(\delta - \widehat{\delta}\right) + \widetilde{\eta}_{jt} + \widetilde{\widetilde{\varepsilon}}_{jt}.\end{aligned}$$

We will also use the notation

$$\begin{aligned}\overline{Z}_{jt} &= \frac{1}{|M(j,t)|}\sum_{i \in M(j,t)} Z_i \\ \overline{\varepsilon}_{jt} &= \frac{1}{|M(j,t)|}\sum_{i \in M(j,t)} \varepsilon_i \\ \overline{\varepsilon}_{j} &= \frac{1}{T}\sum_{t=1}^{T} \overline{\varepsilon}_{jt}\end{aligned}$$

A.2.1 Preliminary Lemma

Extending the proof of Proposition 2 to this case takes a bit more work because we can not directly apply the Lemma from Newey and McFadden (1994). Instead to prove the result we directly follow the proof of Lemma 1 of Tauchen (1985).

Lemma A.1 *Let $\theta = (w_j, g, b)$ be contained in a compact parameter space Θ. Under Assumptions 1-3,*

$$F_j(\theta) = \frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0} 1\left(\frac{\sum_{t=1}^{T}\left(d_{jt} - \overline{d}_j\right)\left(\eta_{mt} + \overline{Z}'_{mt}(\delta - g) - X'_{mt}(\beta - b) + \overline{\varepsilon}_{jt}\right)}{\sum_{\ell=1}^{N_1}\sum_{t=1}^{T}\left(d_{\ell t} - \overline{d}_\ell\right)^2} < w_j\right)$$

converges uniformly to

$$\phi_j(\theta) \equiv \Pr\left(\frac{\sum_{t=1}^{T}\left(d_{jt} - \overline{d}_j\right)\left(\eta_{mt} + Z'_{mt}(\delta - g) - X'_{mt}(\beta - b)\right)}{\sum_{\ell=1}^{N_1}\sum_{t=1}^{T}\left(d_{\ell t} - \overline{d}_\ell\right)^2} < w_j\right).$$

Proof.

Define $\theta = (w_j, b, g)$. Generically we will use the notation θ^* to denote (w_j^*, b^*, g^*) and θ_k to denote (w_{jk}, b_k, g_k). We continue to use the notation

$$\rho_{jt} \equiv \frac{\left(d_{jt} - \overline{d}_j\right)}{\sum_{\ell=1}^{N_1}\sum_{t=1}^{T}\left(d_{\ell t} - \overline{d}_\ell\right)^2}$$

Now define

$$\begin{aligned}u(\eta_m, \theta, d) = \sup_{|\theta^* - \theta| \le d}\Bigg| & 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt} + Z'_{mt}(\gamma - g^*) - X'_{mt}(\beta - b^*)\right) \le w_j^*\right) \\ & -1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt} + Z'_{mt}(\gamma - g) - X'_{mt}(\beta - b)\right) \le w_j\right)\Bigg|\end{aligned}$$

where $\eta_m = (\eta_{m1}, ..., \eta_{mT})$. Since the distribution of η_{mt} is continuous, $\lim u(\eta_i, \theta, d) = 0$ as $d \to 0$ with θ fixed almost surely. Applying this means we can define $\overline{d}\,(w_j, \epsilon)$ so that $E\left[u(\eta_i, \theta, 2d(w_j, \epsilon))\right] < \epsilon$ where we take ϵ as given.

Now notice that

$$\begin{aligned}
&\sup_{|\theta-\theta^*|\leq d} \left| 1\left(\sum_{t=1}^{T} \rho_{jt}\left(\eta_{mt} + \overline{Z}'_{mt}(\delta - g) - X'_{mt}(\beta - b) + \overline{\varepsilon}_{jt}\right) < w_j\right)\right. \\
&\qquad \left. -1\left(\sum_{t=1}^{T} \rho_{jt}\left(\eta_{mt} + Z'_{mt}(\delta - g^*) - X'_{mt}(\beta - b^*)\right) \leq w_j^*\right)\right| \\
&\leq \sup_{|\theta-\theta^*|\leq 2d} \left| 1\left(\sum_{t=1}^{T} \rho_{jt}\left(\eta_{mt} + Z'_{mt}(\delta - g) - X'_{mt}(\beta - b)\right) < w_j\right)\right. \\
&\qquad \left. -1\left(\sum_{t=1}^{T} \rho_{jt}\left(\eta_{mt} + Z'_{mt}(\delta - g^*) - X'_{mt}(\beta - b^*)\right) \leq w_j^*\right)\right| \times \\
&\qquad 1\left(\sup_{|g-g^*|\leq d} \left| \left(\overline{Z}_{mt} - Z_{mt}\right)'(\delta - g^*) + \overline{\varepsilon}_{jt}\right| \leq d\right) \\
&\qquad + 1\left(\sup_{|g-g^*|\leq d} \left| \left(\overline{Z}_{mt} - Z_{mt}\right)'(\delta - g^*) + \overline{\varepsilon}_{jt}\right| > d\right) \\
&\leq u(\eta_i, \theta, 2d) + 1\left(\sup_{|g-g^*|\leq d} \left| \left(\overline{Z}_{mt} - Z_{mt}\right)'(\delta - g^*) + \overline{\varepsilon}_{jt}\right| > d\right)
\end{aligned}$$

Let $B(\theta)$ denote an open interval of radius $\overline{d}\,(\theta, \epsilon)$ about θ. By compactness we can form an oven covering $B_k = B(\theta_k, \epsilon)$ for $k = 1, ..., K$. Let $d_k = \overline{d}\,(\theta_k, \epsilon)$ and $\mu_k = E(u(Y_i, \theta_k, 2d_k))$. Now note that if $\theta \in B_k$ then $\mu_k \leq \epsilon$ by definition of $\overline{d}\,(w_{jk}, \epsilon)$. It must also be the case that

$$\left|\phi_j(\theta) - \phi_j(\theta_k)\right| \leq E\left(u\left(\eta_i, w_j, d\right)\right) \leq E\left(u\left(\eta_i, w_j, 2d\right)\right) = \mu_k \leq \epsilon.$$

Now let $\theta \in B_k$ and consider

$$
\begin{aligned}
&\left|\frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0} 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt}+\overline{Z}_{mt}'(\delta-g)-X_{mt}'(\beta-b)+\overline{\varepsilon}_{jt}\right)<w_j\right)-\phi_j(\theta)\right| \\
\le\ & \left|\frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0}\left[1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt}+\overline{Z}_{mt}'(\delta-g)-X_{mt}'(\beta-b)+\overline{\varepsilon}_{jt}\right)<w_j\right)\right.\right. \\
&\quad \left.\left.-1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt}+Z_{mt}'(\delta-g_k)-X_{mt}'(\beta-b_k)\right)\le w_{jk}\right)\right]\right| \\
&+\left|\frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0} 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt}+Z_{mt}'(\delta-g_k)-X_{mt}'(\beta-b_k)\right)\le w_{jk}\right)-\phi_j(\theta_k)\right| \\
&+\left|\phi_j(\theta_k)-\phi_j(\theta)\right| \\
\le\ & \frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0} 1\left(\sup_{|g-g^*|\le d}\left|\left(\overline{Z}_{mt}-Z_{mt}\right)'(\delta-g^*)+\overline{\varepsilon}_{jt}\right|>d\right)+\left(\frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0}u(\eta_m,w_k,2d_k)-\mu_k\right) \\
&+\mu_k+\left|\frac{1}{N_0}\sum_{m=N_1+1}^{N_1+N_0} 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{mt}+Z_{mt}'(\delta-g_k)-X_{mt}'(\beta-b_k)\right)\le w_{jk}\right)-\phi_j(\theta_k)\right|+\epsilon \\
\le\ & 5\varepsilon
\end{aligned}
$$

whenever $N_0 \ge N_k(\epsilon)$ almost surely, by applying twice the strong law of large numbers and taking into account that $1\left(\sup_{|g-g^*|\le d}\left|\left(\overline{Z}_{mt}-Z_{mt}\right)'(\delta-g^*)+\overline{\varepsilon}_{jt}\right|>d\right)$ is iid and converges almost surely to zero so the first term converges to zero.

Thus

$$\sup_{\theta\in\Theta}\left|F_\eta(\theta)-\phi_j(\theta)\right|\le 5\epsilon$$

whenever $n \ge \max_k N_k(\epsilon)$ almost surely which proves the result.

■

A.2.2 Proof of Proposition 4

The first part of this proof is virtually identical to the proof of Proposition 1 while the second is very similar to that of Proposition 2.

Consistency of $\widehat{\delta}$ follows immediately from the standard OLS argument.

A standard application of the partitioned inverse theorem makes it straight forward to show that

$$
\begin{aligned}
\widehat{\beta} \quad = \quad & \beta+\left(\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{X}_{jt}'}{N_0+N_1}-\frac{\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{X}_{jt}'\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}^2}\right)^{-1} \qquad \text{(A-2)}\\
& \times\left(\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{\eta}_{jt}}{N_0+N_1}-\frac{\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\eta_{jt}\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}^2}\right.\\
& +\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{\overline{Z}}_{jt}'\left(\widehat{\delta}-\delta\right)}{N_0+N_1}-\frac{\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{\overline{Z}}_{jt}\left(\widehat{\delta}-\delta\right)\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}^2}\\
& \left.+\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{\overline{\varepsilon}}_{jt}}{N_0+N_1}-\frac{\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{\overline{\varepsilon}}_{jt}\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}^2}\right).
\end{aligned}
$$

First consider

$$\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{\overline{\varepsilon}}_{jt}}{N_0+N_1}=\sum_{t=1}^{T}\frac{1}{N_0+N_1}\sum_{j=1}^{N_0+N_1}\sum_{i\in M(j,t)}\frac{\widetilde{X}_{jt}\varepsilon_i}{|M(j,t)|}$$

For each t note that

$$E\left(\frac{1}{N_0+N_1}\sum_{j=1}^{N_0+N_1}\sum_{i\in M(j,t)}\frac{\widetilde{X}_{jt}\varepsilon_i}{|M(j,t)|}\right)=0$$

$$
\begin{aligned}
& E\left(\left[\frac{1}{N_0+N_1}\sum_{j=1}^{N_0+N_1}\sum_{i\in M(j,t)}\frac{\widetilde{X}_{jt}\varepsilon_i}{|M(j,t)|}\right]^2\right)\\
= & \left(\frac{1}{N_0+N_1}\right)^2\sum_{j_1=1}^{N_0+N_1}\sum_{j_2=1}^{N_0+N_1}\sum_{i_1\in M(j_1,t)}\sum_{i_2\in M(j_2,t)}E\left(\frac{\widetilde{X}_{j_1t}\varepsilon_{i_1}\widetilde{X}_{j_2t}\varepsilon_{i_2}}{|M(j_1,t)|\,|M(j_2,t)|}\right)\\
= & \left(\frac{1}{N_0+N_1}\right)^2\sum_{j_1=1}^{N_0+N_1}\sum_{i_1\in M(j_1,t)}\frac{E(\widetilde{X}_{j_1t}\widetilde{X}_{j_1t}')E(\varepsilon_{i_1}^2)}{|M(j_1,t)|^2}\\
\xrightarrow{p} & \ 0,
\end{aligned}
$$

Thus the whole term converges to zero in mean squared error and thus in probability.

Now consider the term

$$
\begin{aligned}
\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{\overline{\varepsilon}}_{jt}\right] &= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\widetilde{\overline{\varepsilon}}_{jt}+\sum_{t=1}^{T}\left(\overline{d}-\overline{d}_t\right)\left[\sum_{j=1}^{N_1+N_0}\widetilde{\overline{\varepsilon}}_{jt}\right]\\
&= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\widetilde{\overline{\varepsilon}}_{jt}\\
&\xrightarrow{p} 0,
\end{aligned}
$$

since this last expression involves a finite number of terms and $\widetilde{\widetilde{\varepsilon}}_{jt} \xrightarrow{p} 0$.

Consistency for $\widehat{\beta}$ follows upon plugging the pieces back into (A-2) and applying Slutsky's theorem.

From the normal equation for $\widehat{\alpha}$ it is straightforward to show that

$$\widehat{\alpha} = \alpha + \frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{\widetilde{Z}}'_{jt} \left(\widehat{\delta} - \delta\right)}{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2} + \frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{\widetilde{\varepsilon}}_{jt}}{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2}$$
$$+ \frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{\eta}_{jt}}{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2} + \left[\frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{X}'_{jt}}{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2}\right] (\beta - \widehat{\beta}).$$

The only new term we don't see in the proof of Proposition 1 is $\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{\widetilde{\varepsilon}}_{jt}$ but we have already shown that this converges in probability to zero so the result for W holds.

We now consider consistency of $\widehat{\Gamma}$. Since Γ is defined conditional on d_{jt} for $j = 1, ...N_1$, $t = 1, ..., T$, every probability in the rest of this proof conditions on this set. To simplify the notation, we omit this explicit conditioning. Thus, every probability statement and distribution function in this proof should be interpreted as conditioning on d_{jt} for $j = 1, ...N_1$, $t = 1, ..., T$.

We continue to use the notation

$$\rho_{jt} \equiv \frac{\left(d_{jt} - \overline{d}_j\right)}{\sum_{\ell=1}^{N_1} \sum_{t=1}^{T} \left(d_{\ell t} - \overline{d}_\ell\right)^2}$$

For each $j = 1, ..., N_1$ define the random variable

$$W_j \equiv \sum_{t=1}^{T} \rho_{jt} \eta_{jt}.$$

Note that we used the fact that since $\overline{\eta}_j$ is time invariant, $\sum_{t=1}^{T} \left(d_{jt} - \overline{d}_j\right) \overline{\eta}_j = 0$. Let F_j be the distribution of W_j for $j = 1, .., N_1$.

Then note that

$$\begin{aligned} \Gamma\left(w\right) &= \Pr\left(\sum_{j=1}^{N_1} \sum_{t=1}^{T} \rho_{jt} \eta_{jt} < w\right) \\ &= \int \cdots \int 1\left(\sum_{j=1}^{N_1} W_j < w\right) dF_1(W_1)...dF_{N_1}(W_{N_1}). \end{aligned}$$

We can also write

$$\widehat{\Gamma}\left(w\right) = \int \cdots \int 1\left(\sum_{j=1}^{N_1} W_j < a\right) d\widehat{F}_1(W_1, \widehat{\gamma}, \widehat{\beta})...d\widehat{F}_{N_1}(W_{N_1}, \widehat{\gamma}, \widehat{\beta}),$$

where $\widehat{F}_j(\cdot,\widehat{\delta},\widehat{\beta})$ is the empirical c.d.f. one gets from the residuals using the control states only. That is more generally

$$\widehat{F}_j(w,g,b) \equiv \frac{1}{N_0}\sum_{m=1}^{N_0} 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\widetilde{\eta}_{mt} - \widetilde{\widetilde{Z}}'_{mt}(\delta-g) - \widetilde{X}'_{mt}(\beta-b) + \widetilde{\widetilde{\varepsilon}}_{mt}\right) < w_j\right).$$

It is easy to verify that

$$\begin{aligned}
\widehat{F}_j(w,\widehat{\delta},\widehat{\beta}) &\equiv \frac{1}{N_0}\sum_{m=1}^{N_0} 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\widetilde{\eta}_{mt} - \widetilde{\widetilde{Z}}'_{mt}\left(\delta-\widehat{\delta}\right) - \widetilde{X}'_{mt}\left(\beta-\widehat{\beta}\right) + \widetilde{\widetilde{\varepsilon}}_{mt}\right) < w_j\right) \\
&= \frac{1}{N_0}\sum_{m=1}^{N_0} 1\left(\sum_{t=1}^{T}\rho_{jt}\left(\widetilde{\widetilde{\lambda}}_{mt} - \widetilde{X}'_{mt}\widehat{\beta}\right) < w_j\right).
\end{aligned}$$

To avoid repeating the expression we define

$$\phi_j(w_j,\gamma,b) \equiv \Pr\left(\sum_{t=1}^{T}\rho_{jt}\left(\eta_{jt} - Z'_{jt}(\delta-g) - X'_{jt}(\beta-b)\right) < w_j\right)$$

where $Z_{mt} \equiv E(Z_i \mid j(i)=j, t(i)=t)$.

Note that $\phi_j(w_j,\delta,\beta) = F_j(w_j)$. The proof strategy is first to demonstrate that $\widehat{F}_j(w_j,\widehat{\delta},\widehat{\beta})$ converges to $\phi_j(w_j,\delta,\beta)$ uniformly over w. The final part of the proof that $\widehat{\Gamma}(w)$ is a consistent estimate of $\Gamma(w)$ is identical to that argument in the proof of Proposition 2.

Following the same line as the proof of Proposition 2, let

$$\widehat{\omega}_j = \sum_{t=1}^{T}\rho_{jt}\left(\overline{\eta}_{mt} - \overline{Z}'_t\left(\delta-\widehat{\delta}\right) - \overline{X}'_t\left(\beta-\widehat{\beta}\right) + \overline{\varepsilon}_t\right).$$

Let Ω be a compact parameter space for w_j, and Θ a compact subset of the parameter space for $\left(\widehat{\omega}_j,\widehat{\gamma},\widehat{\beta}\right)$ in which $(0,\delta,\beta)$ is an interior point.

We use the notation $\sum_m$ rather than $\sum_{m=N_1+1}^{N_1+N_0}$ in order to get the expression to fit on a page.

First, for each $j=1,...,N_1$ consider the difference between $\widehat{F}_j(w_j,\widehat{\gamma},\widehat{\beta})$ and $\phi_j(w_j,\gamma,\beta)$

$$
\begin{aligned}
&\sup_{w_j\in\Omega}|\widehat{F}_j(w_j,\widehat{\delta},\widehat{\beta})-\phi_j(w_j,\delta,\beta)| \\
=&\sup_{w_j\in\Omega}\left|\frac{1}{N_0}\sum_m 1\left(\sum_{t=1}^{T}\alpha_{jt}\left(\widetilde{\eta}_{mt}-\widetilde{\overline{Z}}'_{mt}\left(\delta-\widehat{\delta}\right)-\widetilde{X}'_{mt}\left(\beta-\widehat{\beta}\right)+\widetilde{\overline{\varepsilon}}_{mt}\right)<w_j\right)-\phi_j\left(w_j,\delta,\beta\right)\right| \\
=&\sup_{w_j\in\Omega}\left|\frac{1}{N_0}\sum_m 1\left(\sum_{t=1}^{T}\alpha_{jt}\left(\eta_{mt}-\overline{Z}'_{mt}\left(\delta-\widehat{\delta}\right)-X'_{mt}\left(\beta-\widehat{\beta}\right)+\overline{\varepsilon}_{mt}\right)<w_j+\widehat{\omega}_j\right)-\phi_j\left(w_j,\delta,\beta\right)\right| \\
\leq&\sup_{w_j\in\Omega}\left|\frac{1}{N_0}\sum_m 1\left(\sum_{t=1}^{T}\alpha_{jt}\left(\eta_{mt}-\overline{Z}'_{mt}\left(\delta-\widehat{\delta}\right)-X'_{mt}\left(\beta-\widehat{\beta}\right)+\overline{\varepsilon}_{mt}\right)<w_j+\widehat{\omega}_j\right)-\phi_j\left(w_j+\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)\right| \\
&+\sup_{w_j\in\Omega}\left|\phi_j\left(w_j+\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)-\phi_j\left(w_j,\beta\right)\right| \\
\leq&\sup_{\substack{w_j\in\Omega\\(\omega_j,g,b)\in\Theta}}\left|\frac{1}{N_0}\sum_m^{N_0} 1\left(\sum_{t=1}^{T}\alpha_{jt}\left(\eta_{mt}-\overline{Z}'_{mt}\left(\delta-g\right)-X'_{mt}\left(\beta-b\right)+\overline{\varepsilon}_{mt}\right)<w_j+\omega_j\right)-\phi_j\left(w_j+\omega_j,g,b\right)\right| \\
&+\Pr\left(\left(\widehat{\omega}_j,\widehat{\delta},\hat{\beta},\right)\notin\Theta\right)+\sup_{w_j\in\Omega}\left|\phi_j\left(w_j+\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)-\phi_j\left(w_j,\delta,\beta\right)\right|.
\end{aligned}
$$

First consider $\sup_{w_j\in\Omega}\left|\phi_j\left(w_j+\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)-\phi_j\left(w_j,\delta,\beta\right)\right|$. Using a standard mean-value expansion of ϕ, for some $\left(\widetilde{\omega}_j,\widetilde{\delta},\widetilde{\beta}\right)$

$$
\begin{aligned}
&\sup_{w_j\in\Omega}\left|\phi_j\left(w_j+\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)-\phi_j\left(w_j,\delta,\beta\right)\right| \\
=&\sup_{w_j\in\Omega}\left|\frac{\partial\phi_j\left(w_j+\widetilde{\omega}_j,\widetilde{\delta},\widetilde{\beta}\right)}{\partial w_j}\left(\widehat{\omega}_j\right)+\frac{\partial\phi_j\left(w_j+\widetilde{\omega}_j,\widetilde{\delta},\widetilde{\beta}\right)}{\partial\gamma}\left(\widehat{\gamma}-\gamma\right)+\frac{\partial\phi_j\left(w_j+\widetilde{\omega}_j,\widetilde{\delta},\widetilde{\beta}\right)}{\partial\beta}\left(\widehat{\beta}-\beta\right)\right|.
\end{aligned}
$$

The proof that $\frac{\partial\phi_j\left(w_j+\widetilde{\omega}_j,\widetilde{\delta},\widetilde{\beta}\right)}{\partial\beta}$ is bounded is analogous to that in the proof of Proposition 2. Thus $\sup_{w_j\in\Omega}\left|\phi_j\left(w_j+\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)-\phi_j\left(w_j,\delta,\beta\right)\right|$converges to zero since $\widehat{\beta}$ is consistent. The same argument holds for the other two pieces.

Since $\left(\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)$ converges in probability to $(0,\delta,\beta)$ which is an interior point of Θ, $\Pr\left(\left(\widehat{\omega}_j,\widehat{\delta},\widehat{\beta}\right)\in\Theta\right)$ converges in to zero.

Next consider the term

$$
\sup_{\substack{w_j\in\Omega\\(\omega_j,g,b)\in\Theta}}\left|\frac{1}{N_0}\sum_m^{N_0} 1\left(\sum_{t=1}^{T}\alpha_{jt}\left(\eta_{mt}-\overline{Z}'_{mt}\left(\delta-g\right)-X'_{mt}\left(\beta-b\right)+\overline{\varepsilon}_{mt}\right)<w_j+\omega_j\right)-\phi_j\left(w_j+\omega_j,g,b\right)\right|.
$$

Since w_j and ω_j enter the expression in identical ways we can combine these into one parameter and expand the compact parameter set appropriately and apply Lemma A.1 from section A.4.1 to show that this term converges to zero.

Then putting the three pieces together,

$$\sup_{w_j \in \Omega} |\widehat{F}(w_j, \widehat{\delta}, \widehat{\beta}) - \phi(w_h, \delta, \beta)| \xrightarrow{p} 0.$$

Finally, the final step of the proof that $\widehat{\Gamma}(w)$ converges to $\Gamma(w)$ is identical to that of Proposition 2.

■

A.2.3 Results for $|M(j,t)|$ fixed

We now consider the case in which $|M(j,t)|$ is fixed. This is a straight forward extension of the proofs of Propositions 1 and 2 in which group by time error term becomes $\eta_{jt} + \overline{\varepsilon}_{jt}$ rather than just η_{jt}. We replace Assumption 3 with

Assumption A.1 *ε_i is IID and orthogonal to $[\ Z_i \quad I_i\ ]$, which is full rank. $|M(j,t)|$ is fixed with the sample size and for all j_1, j_2, and t*

$$|M(j_1,t)| = |M(j_2,t)|.$$

Note that without loss of generality we can essentially incorporate $\widetilde{\overline{\varepsilon}}_{jt}$ into $\widetilde{\eta}_{jt}$ and repeat what we have done.

Proposition A.1 *Under Assumptions 1 and A.1,*

$$\begin{aligned}
\widehat{\delta} &\xrightarrow{p} \delta \\
\widehat{\beta} &\xrightarrow{p} \beta \\
\widehat{\alpha} &\xrightarrow{p} \alpha + W
\end{aligned}$$

as $N_0 \to \infty$ *where*

$$W = \frac{\sum_{j=1}^{N_1} \sum_{t=1}^{T} \left(d_{jt} - \overline{d}_j\right)\left(\eta_{jt} + \overline{\varepsilon}_{jt} - \overline{\eta}_j - \overline{\varepsilon}_j\right)}{\sum_{j=1}^{N_1} \sum_{t=1}^{T} \left(d_{jt} - \overline{d}_j\right)^2}.$$

Proof.

This is virtually identical to the proof of Proposition 1.

Consistency of $\widehat{\delta}$ follows directly from the standard argument for consistency of parameters in fixed effects models.

First a standard application of the partitioned inverse theorem makes it straight forward to show that

$$\begin{aligned}
\widehat{\beta} \;=\; \beta + & \left(\frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt} \widetilde{X}_{jt}'}{N_0+N_1} - \frac{\left[\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{X}_{jt}'\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2} \right)^{-1} \\
& \times \left(\frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt}\left(\widetilde{\eta}_{jt} + \widetilde{\overline{\varepsilon}}_{jt}\right)}{N_0+N_1} - \frac{\left[\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}\left(\eta_{jt} + \overline{\varepsilon}_{jt}\right)\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2} \right. \\
& \left. + \frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{X}_{jt} \widetilde{\overline{Z}}_{jt}'\left(\widehat{\delta} - \delta\right)}{N_0+N_1} - \frac{\left[\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{X}_{jt}\right]\left[\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt} \widetilde{\overline{Z}}_{jt}\left(\widehat{\delta} - \delta\right)\right]}{(N_0+N_1)\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \widetilde{d}_{jt}^2} \right).
\end{aligned}$$

We have two new pieces to consider relative to the proof of Proposition 1. First consider

$$\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{\overline{Z}}'_{jt}\left(\widehat{\delta}-\delta\right)}{N_0+N_1}.$$

Since $\widetilde{Z}_{jt}$ is independent across j, $\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{Z}'_{jt}}{N_0+N_1}$ converges in probability to its mean so by Slutzky,

$$\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{X}_{jt}\widetilde{\overline{Z}}'_{jt}\left(\widehat{\delta}-\delta\right)}{N_0+N_1}\xrightarrow{p}0.$$

Now consider the term

$$\begin{aligned}\left[\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{\overline{Z}}_{jt}\left(\widehat{\delta}-\delta\right)\right] &= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\widetilde{\overline{Z}}'_{jt}\left(\widehat{\delta}-\delta\right)+\sum_{t=1}^{T}\left(\overline{d}-\overline{d}_t\right)\left[\sum_{j=1}^{N_1+N_0}\widetilde{\overline{Z}}'_{jt}\right]\left(\widehat{\delta}-\delta\right)\\ &= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\widetilde{\overline{Z}}'_{jt}\left(\widehat{\delta}-\delta\right)\\ &\xrightarrow{p}0\end{aligned}$$

since this last expression involves a finite number of terms and $\left(\widehat{\delta}-\delta\right)\xrightarrow{p}0$.

Consistency for $\widehat{\beta}$ follows upon plugging the pieces back into the expression for $\widehat{\beta}$ above and applying Slutsky's theorem.

From the normal equation for $\widehat{\alpha}$ it is straightforward to show that

$$\begin{aligned}\widehat{\alpha}=\alpha+&\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}\widetilde{\overline{Z}}'_{jt}\left(\widehat{\delta}-\delta\right)}{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}^2}\\ &+\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}\left(\widetilde{\eta}_{jt}+\widetilde{\overline{\varepsilon}}_{jt}\right)}{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}^2}+\left[\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}\widetilde{X}'_{jt}}{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}^2}\right](\beta-\widehat{\beta}).\end{aligned}$$

We showed in the proof to Proposition 1 that

$$\left[\frac{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}\widetilde{X}'_{jt}}{\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d_{jt}}^2}\right](\beta-\widehat{\beta})\xrightarrow{p}0.$$

and that

$$\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{\eta}_{jt}\xrightarrow{p}\sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}-\overline{\eta}_j\right).$$

By the same logic

$$\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\left(\widetilde{\eta}_{jt}+\widetilde{\overline{\varepsilon}}_{jt}\right)\xrightarrow{p}\sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}+\overline{\varepsilon}_{jt}-\overline{\eta}_j-\overline{\varepsilon}_j\right).$$

Finally, we showed above that

$$\sum_{j=1}^{N_0+N_1}\sum_{t=1}^{T}\widetilde{d}_{jt}\widetilde{\widetilde{Z}}'_{jt}\left(\widehat{\delta}-\delta\right)\xrightarrow{p}0.$$

Putting the terms together the result stands.

■

Proposition A.2 *Under Assumptions 1, 2, and A.1,* $\widehat{\Gamma}(a)$ *converges uniformly to* $\Gamma(a)$.

Proof.

Note that

$$\widehat{\Gamma}(w)\equiv\left(\frac{1}{N_0}\right)^{N_1}\sum_{\ell_1=N_1+1}^{N_1+N_0}\cdots\sum_{\ell_{N_1}=N_1+1}^{N_1+N_0}1\left(\frac{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\left(\widetilde{\widetilde{\lambda}}_{\ell_j t}-\widetilde{X}'_{\ell_j t}\hat{\beta}\right)}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)^2}<w\right)$$

$$=\left(\frac{1}{N_0}\right)^{N_1}\sum_{\ell_1=N_1+1}^{N_1+N_0}\cdots\sum_{\ell_{N_1}=N_1+1}^{N_1+N_0}1\left(\frac{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\left(\widetilde{\eta}_{\ell_j t}-\widetilde{\widetilde{Z}}'_{\ell_j t}\left(\delta-\widehat{\delta}\right)-\widetilde{X}'_{\ell_j t}\left(\beta-\hat{\beta}\right)\right)}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)^2}<w\right).$$

Thus one can see that this result follows directly from the proof of Proposition 2 where we just reinterpret $\left[\begin{array}{cc}\widetilde{\widetilde{Z}}'_{jt} & \widetilde{X}'_{jt}\end{array}\right]$ as "$\widetilde{X}'_{jt}$" ,$\left[\begin{array}{cc}\delta' & \beta'\end{array}\right]$ as "β'" and $\widetilde{\widetilde{\lambda}}_{jt}$ as "$\widetilde{Y}_{jt}$" in that proof.

■

A.2.4 Estimation with Population Weighted Regression

Now we consider estimation of the model directly. That is we imagine that one directly runs the regression

$$Y_i=\alpha d_{jt}+Z_i'\delta+\theta_j+\gamma_t+\eta_{jt}+\varepsilon_i$$

using state dummies and time dummies. Note that the distinction between Z_i and X_{jt} is no longer necessary so we have just incorporated $X'_{j(i)t(i)}\beta$ into $Z_i'\delta$.

We first need to formally define what these objects are. For a generic variable Z_i define

$$\overline{Z}_j=\frac{\sum_{t=1}^{T}\sum_{i\in M(j,t)}Z_i}{\sum_{t=1}^{T}|M(j,t)|}.$$

Since in general, the number of individuals varies across (j,t) cells, derivation of the difference in differences operator requires additional notation. We need to formally define the full set of indicators for groups $\{g_{\ell i}\}_{\ell=1}^{N_1+N_0}$and time periods, $\{p_{\tau i}\}_{\tau=1}^{T-1}$ so that

$$g_{\ell i}\equiv 1(\ell=j(i)) \tag{A-3}$$

$$p_{\tau i}\equiv 1(\tau=t(i)). \tag{A-4}$$

Further define G_i and P_i as the vectors of these dummy variables,

$$G_i \equiv \begin{bmatrix} g_{1i} & g_{2i} & \dots & g_{N_o+N_0,i} \end{bmatrix}' \tag{A-5}$$
$$P_i \equiv \begin{bmatrix} p_{1i} & p_{2i} & \dots & p_{T-1,i} \end{bmatrix}'. \tag{A-6}$$

Then for any individual-specific random variable Z_i, let $\widetilde{Z}_i$ be the residual from a linear regression of Z_i on $\{g_{\ell i}\}_{\ell=1}^{N_1+N_0}$and $\{p_{\tau i}\}_{\tau=1}^{T-1}$. That is

$$\widetilde{Z}_i \equiv Z_i - \begin{bmatrix} G_i \\ P_i \end{bmatrix}' \left(\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \sum_{h \in M(j,t)} \begin{bmatrix} G_h \\ P_h \end{bmatrix} \begin{bmatrix} G_h \\ P_h \end{bmatrix}' \right)^{-1} \left(\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \sum_{h \in M(j,t)} \begin{bmatrix} G_h \\ P_h \end{bmatrix} Z_h \right).$$

We need to strengthen Assumption 1 somewhat. In the two stage approach one does not need to take a stand on the relationship between Z_i and $\eta_{j(i)t(i)}$ because we can obtain consistent estimates of δ via fixed effects. This is no longer the case if we estimate the model in one step. We use the assumption

Assumption A.2 $\left(\left(\eta_{j1}, \{Z_i : i \in M(j,1)\}\right), ..., \left(\eta_{jT}, \{Z_i : i \in M(j,T)\}\right)\right)$ *is stationary and strong mixing across groups;* $\left(\eta_{j1}, ..., \eta_{jT}\right)$ *is expectation zero conditional on* $(d_{j1}, ..., d_{jT})$,$\{Z_i : i \in \cup_{t=1,...,T} M(j,t)\}$ *and all random variables have finite second moments. Further* ε_i *is i.i.d. in uncorrelated with all other random variables in the model.*

We need a regularity condition to guarantee enough degrees of freedom that regressions upon time and group indicators can be run.

Assumption A.3

$$\frac{\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \sum_{i \in M(j,t)} P_i P_i'}{\sum_{j=1}^{N_1} \sum_{t=1}^{T} |M(j,t)|} - \frac{\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \sum_{i \in M(j,t)} P_i G_i' \left(\sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \sum_{i \in M(j,t)} G_i G_i' \right)^{-1} \sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \sum_{i \in M(j,t)} G_i P_i'}{\sum_{j=1}^{N_1} \sum_{t=1}^{T} |M(j,t)|} \xrightarrow{p} \Omega$$

where Ω *is of full rank.*

Under this condition, we can rewrite the model as:

$$\widetilde{Y}_i = \alpha \widetilde{d}_{j(i)t(i)} + \widetilde{Z}_i' \delta + \widetilde{\eta}_{j(i)t(i)} + \widetilde{\varepsilon}_i. \tag{A-7}$$

We estimate α and β in equation (A-7) by OLS, letting $\widehat{\alpha}$ and $\widehat{\beta}$ denote the corresponding estimators. This requires the usual OLS rank condition stated as

Assumption A.4

$$\frac{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \sum_{i \in M(j,t)} \widetilde{Z}_i \widetilde{Z}_i'}{\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} |M(j,t)|} \xrightarrow{p} \Sigma_z$$

where Σ_z *is finite and of full rank.*

Finally since the asymptotic weights for the treatment groups will enter the expression we put more structure on the way in which groups grow

Assumption A.5 *For each $j = 1, ..., N_0 + N_1$, $|M(j,t)|$ grows at the same rate as N_1. For all j and t, defining*

$$\phi_{jt} \equiv \lim_{N_1 \to \infty} \frac{|M(j,t)|}{\sum_{j=1}^{N_0} \sum_{t=1}^{T} |M(j,t)|},$$

we assume that where $\phi_{jt} > 0$ and bounded from above. For all t, defining

$$\phi_t \equiv \lim_{N_1 \to \infty} \frac{1}{N_0 + N_1} \sum_{j=1}^{N_0+N_1} \frac{|M(j,t)|}{\sum_{j=1}^{N_0} \sum_{t=1}^{T} |M(j,t)|},$$

we assume that $0 < \phi_t < \infty$.

We will make use of the following Lemma

Lemma A.2 *Consider a regression of $d_{j(i)t(i)}$ on group dummies (G_i) and time dummy variables (P_i) as defined in equations (A-3)-(A-6). Let $\widehat{a}_t$ be coefficient on the time variable for time period $t = 1, .., T-1$ and $\widehat{a}_T \equiv 0$. Under Assumptions A.2-A.5,*

$$\widetilde{d}_{j(i)t(i)} = d_{j(i)t(i)} - \overline{d}_{j(i)} - \left(\widehat{a}_{t(i)} - \frac{\sum_{\tau=1}^{T-1} |M(j(i),\tau| \widehat{a}_\tau}{\sum_{\tau=1}^{T} |M(j(i),\tau|} \right)$$

and $\widehat{a}_\tau = O_p(\frac{1}{N_0}), \tau = 1, ..., T-1$.

Proof. To streamline the notation, let $\sum_i$ denote $\sum_{j=1}^{N_0+N_1} \sum_{t=1}^{T} \sum_{i \in M(j,t)}$ and let

$$\begin{aligned} m_0 &\equiv \sum_{j=1}^{N_1} \sum_{t=1}^{T} |M(j,t)| \\ m_1 &\equiv \sum_{j=N_1+1}^{N_0+N_1} \sum_{t=1}^{T} |M(j,t)| \\ m &\equiv m_0 + m_1 \end{aligned}$$

Note that m_0 is fixed but m_1 and m get large as $N_0 \to \infty$. We will use this notation in the proof of Proposition A.3.

Now consider a regression of $d_{j(i)t(i)}$ on group dummies and time dummies. We will write this regression equation as

$$d_{j(i)t(i)} = P_i' \widehat{a} + G_i' \widehat{b} + \widetilde{d}_{j(i)t(i)}$$

where P_i and G_i are as defined equations (A-3)-(A-6).

The first part of our lemma is a standard regression result with dummy variables. Note that we can rewrite this regression equation as

$$d_{j(i)t(i)} - \widehat{a}_{t(i)} = G_i' \widehat{b} + \widetilde{d}_{j(i)t(i)}.$$

Since $\widetilde{d}_{j(i)t(i)}$ is orthogonal to G_i we could construct residuals by regressing $d_{j(i)t(i)} - \widehat{a}_{t(i)}$ on a full set of group dummies and taking residuals. However, it is well known that this will lead to taking deviations of the left hand side variable from group means so that

$$\begin{aligned}
\widetilde{d}_{j(i)t(i)} &= \left(d_{j(i)t(i)} - \widehat{a}_{t(i)}\right) - \frac{\sum_{\tau=1}^{T} \sum_{\ell \in M(j(i),\tau)} \left(d_{j(\ell)\tau} - \widehat{a}_{t(\ell)}\right)}{\sum_{\tau=1}^{T} |M(j(i),\tau)|} \\
&= \left(d_{j(i)t(i)} - \overline{d}_{j(i)}\right) - \left(\widehat{a}_{t(i)} - \frac{\sum_{\tau=1}^{T-1} |M(j(i),\tau)| \, \widehat{a}_{\tau}}{\sum_{\tau=1}^{T} |M(j(i),\tau)|}\right).
\end{aligned}$$

Next consider the derivation of $\widehat{a}$. Using the partitioned inverse theorem,

$$\begin{aligned}
\widehat{a} = \; & \frac{1}{m}\left(\frac{1}{m}\sum_i P_i P_i' - \frac{1}{m}\sum_i P_i G_i' \left(\sum_i G_i G_i'\right)^{-1} \sum_i G_i P_i'\right)^{-1} \times \\
& \left[\sum_i P_i d_{j(i)t(i)} - \sum_i P_i G_i' \left(\sum_i G_i G_i'\right)^{-1} \sum_i G_i d_{j(i)t(i)}\right].
\end{aligned}$$

Assumption A.3 implies that we can rewrite this as

$$\widehat{a} = \frac{1}{m}\left(\Omega + o_p(1)\right)^{-1} \left[\sum_i P_i d_{j(i)t(i)} - \sum_i P_i G_i' \left(\sum_i G_i G_i'\right)^{-1} \sum_i G_i d_{j(i)t(i)}\right].$$

Now consider the last term, $\sum P_i G_i' \left(\sum G_i G_i'\right)^{-1} \sum G_i d_{j(i)t(i)}$. It is straightforward to show that this is a $(T-1) \times 1$ vector with generic element t

$$\sum_{j=1}^{N_1+N_0} \frac{|M(j,t)| \sum_{\tau=1}^{T} |M(j,\tau)| \, d_{j\tau}}{\sum_{\tau=1}^{T} |M(j,\tau)|} = \sum_{j=1}^{N_1+N_0} |M(j,t)| \, \overline{d}_j.$$

Thus the $(T-1) \times 1$ vector $\left[\sum P_i d_{j(i)t(i)} - \sum P_i G_i' \left(\sum G_i G_i'\right)^{-1} \sum G_i d_{j(i)t(i)}\right]$ has generic t element

$$\begin{aligned}
\sum_{j=1}^{N_1+N_0} |M(j,t)| \, d_{jt} - \sum_{j=1}^{N_1+N_0} |M(j,t)| \, \overline{d}_j &= \sum_{j=1}^{N_1+N_0} |M(j,t)| \left(d_{jt} - \overline{d}_j\right) \\
&= \sum_{j=1}^{N_1} |M(j,t)| \left(d_{jt} - \overline{d}_j\right).
\end{aligned}$$

We can write

$$\begin{aligned}
\hat{a} = & \frac{1}{N_1 + N_0}\left(\Omega + o_p(1)\right)^{-1} \times \\
& \left[\frac{N_1 + N_0}{m} \sum_i P_i d_{j(i)t(i)} - \frac{N_1 + N_0}{m} \sum_i P_i G_i' \left(\sum_i G_i G_i'\right)^{-1} \sum_i G_i d_{j(i)t(i)}\right].
\end{aligned}$$

As above the last term in brackets is a $(T-1)\times 1$ vector with a generic element t that can be written as

$$
\begin{aligned}
\frac{N_1+N_0}{m}\sum_{j=1}^{N_1}|M(j,t)|\left(d_{jt}-\overline{d}_j\right) &= \frac{\sum_{j=1}^{N_1}|M(j,t)|\left(d_{jt}-\overline{d}_j\right)}{\frac{1}{N_1+N_0}\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}|M(j,t)|} \\
&= \frac{\sum_{j=1}^{N_1}\frac{|M(j,t)|}{\sum_{\ell=1}^{N_1}\sum_{\tau=1}^{T}|M(\ell,\tau)|}\left(d_{jt}-\overline{d}_j\right)}{\frac{1}{N_1+N_0}\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\left(\frac{|M(j,t)|}{\sum_{\ell=1}^{N_1}\sum_{\tau=1}^{T}|M(\ell,\tau)|}\right)} \\
&\xrightarrow{p} \frac{\sum_{j=1}^{N_1}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)}{\sum_{t=1}^{T}\phi_t}
\end{aligned}
$$

which is $O_p(1)$. ■

Proposition A.3 *Under Assumptions A.2-A.5,*

$$
\begin{aligned}
&\widehat{\delta}\xrightarrow{p}\delta \\
&\widehat{\alpha}\xrightarrow{p}\alpha+\frac{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}-\overline{\eta}_j\right)}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j^2\right)}
\end{aligned}
$$

as $N_0\to\infty$.

Proof:

In this proof we make use of the notation defined in the proof of the Lemma A.2.

First a standard application of the partitioned inverse theorem makes it straightforward to show that

$$
\begin{aligned}
\widehat{\delta} &= \delta+\left(\frac{1}{m}\sum_i\widetilde{Z}_i\widetilde{Z}_i'-\frac{m_0}{m}\frac{\left[\frac{1}{m_0}\sum_i\widetilde{d}_{j(i)t(i)}\widetilde{Z}_i\right]\left[\frac{1}{m_0}\sum_i\widetilde{d}_{j(i)t(i)}\widetilde{Z}_i'\right]}{\frac{1}{m_0}\sum_i\widetilde{d}_{j(i)t(i)}^2}\right)^{-1} \\
&\quad\times\left(\frac{1}{m}\sum_i\widetilde{Z}_i\left(\widetilde{\eta}_{j(i)t(i)}+\widetilde{\varepsilon}_i\right)-\frac{m_0}{m}\frac{\left[\frac{1}{m_0}\sum_i\widetilde{d}_{j(i)t(i)}\widetilde{Z}_i\right]\left[\frac{1}{m_0}\sum_i\widetilde{d}_{j(i)t(i)}\left(\widetilde{\eta}_{j(i)t(i)}+\widetilde{\varepsilon}_i\right)\right]}{\frac{1}{m_0}\sum_i\widetilde{d}_{j(i)t(i)}^2}\right).
\end{aligned}
$$

Now consider each piece in turn.

Assumption A.4 states that

$$\frac{1}{m}\sum_i\widetilde{Z}_i\widetilde{Z}_i'\xrightarrow{p}\Sigma_z.$$

Using Assumptions A.2-A.3 and invoking the law of large numbers,

$$\frac{1}{m}\sum_i\widetilde{Z}_i\left(\widetilde{\eta}_{j(i)t(i)}+\tilde{\varepsilon}_i\right)\xrightarrow{p}0.$$

Define $\widehat{a}_t$ as in the statement of Lemma $A.2$ and then define

$$\widehat{\widehat{a}}_{jt}\equiv\left(\widehat{a}_t-\frac{\sum_{\tau=1}^{T-1}|M(j,\tau)|\,\widehat{a}_\tau}{\sum_{\tau=1}^{T}|M(j,\tau)|}\right).$$

Lemma A.2 states that $\widetilde{d}_{j(i)t(i)} = d_{j(i)t(i)} - \overline{d}_{j(i)} - \widehat{\widehat{a}}_{j(i)t(i)}$. Note also that for $j > N_1$, $d_{jt} - \overline{d}_j = 0$. Thus

$$\begin{aligned}
\frac{1}{m_0}\sum_i \widetilde{d}^2_{j(i)t(i)} =& \frac{1}{m_0}\sum_{j=1}^{N_1}\sum_{t=1}^{T} |M(j,t)|\, \widetilde{d}^2_{jt} + \frac{1}{m_0}\sum_{j=N_1+1}^{N_1+N_0}\sum_{t=1}^{T} |M(j,t)|\, \widetilde{d}^2_{jt} \\
=& \frac{1}{m_0}\sum_{j=1}^{N_1}\sum_{t=1}^{T} |M(j,t)| \left[\left(d_{jt} - \overline{d}_j\right)^2 - 2\widehat{\widehat{a}}_{jt}\left(d_{jt} - \overline{d}_j\right) + \widehat{\widehat{a}}^2_{jt}\right] + \frac{1}{m_0}\sum_{j=N_1+1}^{N_1+N_0}\sum_{t=1}^{T} \widehat{\widehat{a}}^2_{jt}\, |M(j,t)| \\
\xrightarrow{p}& \sum_{j=1}^{N_1}\sum_{t=1}^{T} \phi_{jt}\left(d_{jt} - \overline{d}_j\right)^2.
\end{aligned}$$

This result follows because

$$\begin{aligned}
&\frac{1}{m_0}\sum_{j=N_1+1}^{N_1+N_0}\sum_{t=1}^{T} \widehat{\widehat{a}}^2_{jt}\, |M(j,t)| \\
&= \frac{1}{m_0}\sum_{j=N_1+1}^{N_1+N_0}\left[\sum_{t=1}^{T}\left(\widehat{a}_t - \frac{\sum_{\tau=1}^{T-1} |M(j,\tau|\,\widehat{a}_\tau}{\sum_{\tau=1}^{T}|M(j,\tau|}\right)^2 |M(j,t)|\right] \\
&= \sum_{j=N_1+1}^{N_1+N_0}\sum_{t=1}^{T-1} \widehat{a}^2_t \frac{|M(j,t)|}{m_o} - 2\sum_{j=N_1+1}^{N_1+N_0}\sum_{t=1}^{T-1} \widehat{a}_t \frac{\sum_{\tau=1}^{T-1}|M(j,\tau)|\,\widehat{a}_\tau}{m_0\sum_{\tau=1}^{T}|M(j,\tau)|} \\
&+ \sum_{j=N_1+1}^{N_1+N_0}\sum_{t=1}^{T}\left(\frac{\sum_{\tau=1}^{T-1}|M(j,\tau)|\,\widehat{a}_\tau}{\sum_{\tau=1}^{T}|M(j,\tau)|}\right)^2 \frac{|M(j,t)|}{m_0} \\
&= \sum_{t=1}^{T-1}\widehat{a}^2_t \sum_{j=N_1+1}^{N_1+N_0}\frac{|M(j,t)|}{m_0} - 2\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\widehat{a}_t\widehat{a}_\tau \sum_{j=N_1+1}^{N_1+N_0}\frac{|M(j,\tau)|}{m_0\sum_{s=1}^{T}|M(j,s)|} \\
&+ \sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\widehat{a}_\tau\widehat{a}_t \sum_{j=N_1+1}^{N_1+N_0}\frac{|M(j,\tau)|\,|M(j,t)|}{m_0\sum_{s=1}^{T}|M(j,s)|} \\
&= \sum_{t=1}^{T-1} O_p\left(\frac{1}{N_0}\right)\sum_{j=N_1+1}^{N_1+N_0}\frac{|M(j,t)|}{m_0} - \frac{2}{m_0}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1} O_p\left(\frac{1}{N_0}\right)\sum_{j=N_1+1}^{N_1+N_0}\frac{|M(j,\tau)|}{\sum_{s=1}^{T}|M(j,s)|} \\
&+ \sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1} O_p\left(\frac{1}{N_0}\right)\sum_{j=N_1+1}^{N_1+N_0}\frac{|M(j,\tau)|\,|M(j,t)|}{m_0\sum_{s=1}^{T}|M(j,s)|} \\
&\xrightarrow{p} 0.
\end{aligned}$$

Next consider the object

$$
\begin{aligned}
\frac{1}{m_0}\sum_i \widetilde{d}_{j(i)t(i)}\widetilde{Z}_i &= \frac{1}{m_0}\sum_{j=1}^{N_1}\sum_{t=1}^{T}\sum_{i\in m(j,t)}\left(d_{jt}-\overline{d}_j\right)\widetilde{Z}_i + \frac{1}{m_0}\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T}\sum_{i\in m(j,t)}\widehat{\widetilde{a}}_{jt}\widetilde{Z}_i \\
&= \frac{1}{m_0}\sum_{j=1}^{N_1}\sum_{t=1}^{T}\sum_{i\in m(j,t)}\left(d_{jt}-\overline{d}_j\right)\widetilde{Z}_i - \frac{1}{m_0}\sum_{j=1}^{N_1+N_0}\sum_{t=1}^{T-1}\sum_{i\in m(j,t)}\left(\widehat{a}_t - \frac{\sum_{\tau=1}^{T-1}|M(j,\tau)|\,\widehat{a}_\tau}{\sum_{\tau=1}^{T}|M(j,\tau)|}\right)\widetilde{Z}_i \\
&+ \frac{1}{m_0}\sum_{j=1}^{N_1+N_0}\sum_{i\in m(j,t)}\frac{\sum_{\tau=1}^{T-1}|M(j,T)|\,\widehat{a}_\tau}{\sum_{\tau=1}^{T}|M(j,T)|}\widetilde{Z}_i \\
&= \frac{1}{m_0}\sum_{j=1}^{N_1}\sum_{t=1}^{T}\sum_{i\in m(j,t)}\left(d_{jt}-\overline{d}_j\right)\widetilde{Z}_i - \frac{1}{m_0}\sum_{t=1}^{T-1}\widehat{a}_t\sum_{j=1}^{N_1+N_0}\sum_{i\in m(j,t)}\widetilde{Z}_i \\
&+ \frac{1}{m_0}\sum_{j=1}^{N_1+N_0}\frac{\sum_{\tau=1}^{T-1}|M(j,\tau)|\,\widehat{a}_\tau}{\sum_{\tau=1}^{T}|M(j,\tau)|}\sum_{t=1}^{T}\sum_{i\in m(j,t)}\widetilde{Z}_i \\
&= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\left[\frac{|M(j,t)|}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}|M(j,t)|}\right]\left[\frac{1}{|M(j,t)|}\sum_{i\in m(j,t)}\widetilde{Z}_i\right] \\
&\xrightarrow{p} \sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)E(\widetilde{Z}_i \mid i\in M(j,t)) \\
&= O_p(1).
\end{aligned}
$$

We used the fact that $\widetilde{Z}_i$ is the residual from a regression on time and state dummies so $\sum_{j=1}^{N_1+N_0}\sum_{i\in m(j,t)}\widetilde{Z}_i = 0$ and $\sum_{t=1}^{T}\sum_{i\in m(j,t)}\widetilde{Z}_i = 0$.

An analogous argument gives

$$
\begin{aligned}
\frac{1}{m_0}\sum_i \widetilde{d}_{j(i)t(i)}\left(\widetilde{\eta}_{j(i)t(i)}+\widetilde{\varepsilon}_i\right) &= \frac{1}{m_0}\sum_{j=1}^{N_1}\sum_{t=1}^{T}\sum_{i\in m(j,t)}\left(d_{jt}-\overline{d}_j\right)\left(\widetilde{\eta}_{jt}+\widetilde{\varepsilon}_i\right) \\
&= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\left(d_{jt}-\overline{d}_j\right)\left[\frac{|M(j,t)|}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}|M(j,t)|}\right]\left[\frac{1}{|M(j,t)|}\sum_{i\in m(j,t)}\left(\widetilde{\eta}_{jt}+\widetilde{\varepsilon}_i\right)\right] \\
&\xrightarrow{p} \sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)E\left(\widetilde{\eta}_{jt}+\widetilde{\varepsilon}_i \mid i\in M(j,t)\right) \\
&= \sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}-\overline{\eta}_j\right) \\
&= O_p(1).
\end{aligned}
$$

The last term follows because for any $\tau = 1,..,T$ $E\left(\eta_{j\tau}-\overline{\eta}_j\right) = E(\varepsilon_i - \overline{\varepsilon}_j \mid t(i)=\tau) = 0$. So for a regression of either $\left(\eta_{j\tau}-\overline{\eta}_j\right)$ or $(\varepsilon_i-\overline{\varepsilon}_j)$ on time dummies, the coefficient on the dummy variables will converge to zero so $\widetilde{\eta}_{jt} \xrightarrow{p} \left(\eta_{j\tau}-\overline{\eta}_j\right)$ and $\widetilde{\varepsilon}_i \xrightarrow{p} (\varepsilon_i-\overline{\varepsilon}_j)$.

Putting all the objects into the expression for $\widehat{\delta}$, one can see that $\widehat{\delta}$ is consistent.

Now consider $\widehat{\alpha}$. It is straight forward to show that

$$(\widehat{\alpha}-\alpha)=\frac{\frac{1}{m_0}\sum_i \widetilde{d}_{j(i)t(i)}\left(\widetilde{\eta}_{j(i)t(i)}+\widetilde{\varepsilon}_i\right)}{\frac{1}{m_0}\sum_i \widetilde{d}^2_{j(i)t(i)}}+\frac{\frac{1}{m_0}\sum_i \widetilde{d}_{j(i)t(i)}\widetilde{Z}_i'\left(\delta-\delta\right)}{\frac{1}{m_0}\sum_i \widetilde{d}^2_{j(i)t(i)}}.$$

We have shown that

$$\frac{1}{m_0}\sum_i \widetilde{d}^2_{jt} \xrightarrow{p} \sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)^2$$

$$\frac{1}{m_0}\sum_i \widetilde{d}_{j(i)t(i)}\widetilde{Z}'_{j(i)t(i)} \xrightarrow{p} \sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)E(\widetilde{Z}_i \mid i\in M(j,t))$$

$$(\delta-\widehat{\delta}) \xrightarrow{p} 0$$

$$\frac{1}{m_0}\sum_i \widetilde{d}_{j(i)t(i)}\left(\widetilde{\eta}_{j(i)t(i)}+\widetilde{\varepsilon}_i\right) \xrightarrow{p} \sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}-\overline{\eta}_j\right).$$

Thus we are left with:

$$\begin{aligned}(\widehat{\alpha}-\alpha) &= \frac{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}-\overline{\eta}_j\right)+o_p(1)}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)^2+o_p(1)}+o_p(1)\\ &\xrightarrow{p} \frac{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)\left(\eta_{jt}-\overline{\eta}_j\right)}{\sum_{j=1}^{N_1}\sum_{t=1}^{T}\phi_{jt}\left(d_{jt}-\overline{d}_j\right)^2}.\end{aligned}$$

This gives the result.

■

A.3 Cross Sectional Dependence and Heteroskedasticity

This section provides a specific example model replacing the IID assumption for η of Assumption 2.

Consider a model that builds up η_{jt} from two Gaussian processes μ_{jt} and v_{jt} that are independent of each other with the first capturing dependence and the second heteroskedasticity:

$$\eta_{jt}=\mu_{jt}+v_{jt}.$$

Suppose the μ process is expectation zero and stationary across both space and time, potentially spatially and temporally correlated. The covariance between μ_{it} and μ_{ks} depends on a time-invariant observed 'economic' distance, denoted $dist_{ik}$, and the time lag $t-s$. Thus a covariance function for μ can be defined as $Cov(\mu_{it},\mu_{ks})=f(dist_{ik},t-s;\theta_\mu)$, for some finite-dimensional parameter θ_μ. The function f can be chosen to be a valid covariance function for any θ_μ parameter value. See e.g. Cressie and Huang (1999) or Ma (2003) for classes of valid families for f. One simple example parameterization with $\theta_\mu = (\theta_1, \theta_2, \theta_3) > 0$ is:

$$Cov(\mu_{it}, \mu_{ks}) = \theta_1 \exp\{-\theta_2 dist_{ik} - \theta_3 |t-s|\}.$$

Further suppose that the second component v_{jt} of this process is independent across groups and time with a variance that is a function $g(pop_{jt}; \theta_v)$ of observed group population, pop_{jt}, so that:

$$v_{jt} \sim N[0, g(pop_{jt}; \theta_v)].$$

These specifications for the components of η_{jt} imply that the probability limit of residuals in equation (5), $(\eta_{jt} - \bar{\eta}_j)$, is Gaussian process with a known parametric covariance structure that depends on the data, $\{dist_{ik}\}_{all\ ik}$ and $\{pop_{jt}\}_{all\ jt}$, and the parameters θ_μ and θ_v. It is straightforward to consistently estimate θ_μ and θ_v by spatial GMM (Conley, 1999) using covariances and variances of residuals in equation (5) as moment conditions. The 'plug-in' estimator of the joint distribution of $(\eta_{jt} - \bar{\eta}_j)$ using GMM estimators of θ_μ and θ_v will be consistent and can directly be used to estimate the distribution of W.[1]

References

Conley, Timothy G. "GMM Estimation with Cross Sectional Dependence" *Journal of Econometrics* 92:1-45, 1999.

Cressie, N. and C-H Huang "Classes of Nonseparable Spatiotemporal Stationary Covariance Functions" *JASA* 94:1330-1340, 1999.

Ma, C. "Spatio-temporal Stationary Covariance Models" *Journal of Multivariate Analysis.* 86:97-107, 2003.

Newey, Whitney and Daniel McFadden, "Large Sample Estimation and Hypothesis Testing." *Handbook of Econometrics, Vol. 4*, Engle and McFadden eds. Elsevier, 1994,2113-2245.

Tauchen, G., "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30: 415-443, 1985.

[1] Parametric tests for spatial and/or temporal independence are directly available from such a model via tests of restrictions upon θ_μ. Nonparametric tests for spatial/temporal independence are also available and easy to implement, see e.g. Conley and Dupor (2003).