Online Supplementary Information: On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data

Kosuke Imai* In Song Kim†

June 23, 2020

*Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge MA 02138. Phone: 617–384–6778, Email: Imai@Harvard.Edu, URL: https://imai.fas.harvard.edu

†Associate Professor, Department of Political Science, Massachusetts Institute of Technology, Cambridge MA 02142. Phone: 617–253–3138, Email: insong@mit.edu, URL: http://web.mit.edu/insong/www/

1 Proof of Proposition 1

Proof We begin by establishing two algebraic equalities. First, we prove the following equality,

$$
\begin{aligned}
& \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{X_{it}(Y_{it}-\overline{Y}_i-\overline{Y}_t+\overline{Y})-(1-X_{it})(Y_{it}-\overline{Y}_i-\overline{Y}_t+\overline{Y})\right\} \\
= & \sum_{i=1}^{N}\sum_{t=1}^{T}\left[X_{it}\left\{Y_{it}\left(1-\frac{1}{N}-\frac{1}{T}+\frac{1}{NT}\right)-\left(\frac{1}{T}\sum_{t'\neq t}Y_{it'}-\frac{1}{NT}\sum_{t'\neq t}Y_{it'}\right)\right.\right. \\
& \qquad\left.-\left(\frac{1}{N}\sum_{i'\neq i}Y_{i't}-\frac{1}{NT}\sum_{i'\neq i}Y_{i't}\right)+\frac{1}{NT}\sum_{i'\neq i}\sum_{t'\neq t}Y_{i't'}\right\} \\
& -(1-X_{it})\left\{Y_{it}\left(1-\frac{1}{N}-\frac{1}{T}+\frac{1}{NT}\right)-\left(\frac{1}{T}\sum_{t'\neq t}Y_{it'}-\frac{1}{NT}\sum_{t'\neq t}Y_{it'}\right)\right. \\
& \qquad\left.\left.-\left(\frac{1}{N}\sum_{i'\neq i}Y_{i't}-\frac{1}{NT}\sum_{i'\neq i}Y_{i't}\right)+\frac{1}{NT}\sum_{i'\neq i}\sum_{t'\neq t}Y_{i't'}\right\}\right] \\
= & \sum_{i=1}^{N}\sum_{t=1}^{T}\left[X_{it}\left\{\frac{(N-1)(T-1)}{NT}Y_{it}-\frac{N-1}{NT}\sum_{t'\neq t}Y_{it'}-\frac{T-1}{NT}\sum_{i'\neq i}Y_{i't}+\frac{1}{NT}\sum_{i'\neq i}\sum_{t'\neq t}Y_{i't'}\right\}\right. \\
& \qquad\left.-(1-X_{it})\left\{\frac{(N-1)(T-1)}{NT}Y_{it}-\frac{N-1}{NT}\sum_{t'\neq t}Y_{it'}-\frac{T-1}{NT}\sum_{i'\neq i}Y_{i't}+\frac{1}{NT}\sum_{i'\neq i}\sum_{t'\neq t}Y_{i't'}\right\}\right] \\
= & \frac{(T-1)(N-1)}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{X_{it}\left(Y_{it}-\frac{\sum_{t'=1}^{T}Y_{it'}}{T-1}+\frac{\sum_{i'=1}^{N}Y_{i't}}{N-1}-\frac{\sum_{i'\neq i}^{N}\sum_{t'\neq t}^{T}Y_{i't'}}{(T-1)(N-1)}\right)\right. \\
& \qquad\left.-(1-X_{it})\left(\frac{\sum_{t'=1}^{T}Y_{it'}}{T-1}+\frac{\sum_{i'=1}^{N}Y_{i't}}{N-1}-\frac{\sum_{i'\neq i}^{N}\sum_{t'\neq t}^{T}Y_{i't'}-Y_{it}}{(T-1)(N-1)}\right)\right\}. \\
= & \frac{(T-1)(N-1)}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{Y_{it}(1)}-\widehat{Y_{it}(0)}\right) \qquad (1)
\end{aligned}
$$

The second algebraic equality we prove is the following,

$$
\begin{aligned}
& \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{X_{it}(X_{it}-\overline{X}_i-\overline{X}_t+\overline{X})-(1-X_{it})(X_{it}-\overline{X}_i-\overline{X}_t+\overline{X})\right\} \\
= & \sum_{i=1}^{N}\sum_{t=1}^{T}\left[X_{it}\left\{X_{it}\left(1-\frac{1}{N}-\frac{1}{T}+\frac{1}{NT}\right)-\left(\frac{1}{T}\sum_{t'\neq t}X_{it'}-\frac{1}{NT}\sum_{t'\neq t}X_{it'}\right)\right.\right. \\
& \qquad\left.-\left(\frac{1}{N}\sum_{i'\neq i}X_{i't}-\frac{1}{NT}\sum_{i'\neq i}X_{i't}\right)+\frac{1}{NT}\sum_{i'\neq i}\sum_{t'\neq t}X_{i't'}\right\} \\
& -(1-X_{it})\left\{X_{it}\left(1-\frac{1}{N}-\frac{1}{T}+\frac{1}{NT}\right)-\left(\frac{1}{T}\sum_{t'\neq t}X_{it'}-\frac{1}{NT}\sum_{t'\neq t}X_{it'}\right)\right.
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad - \left(\frac{1}{N} \sum_{i' \neq i} X_{i't} - \frac{1}{NT} \sum_{i' \neq i} X_{i't} \right) + \frac{1}{NT} \sum_{i' \neq i} \sum_{t' \neq t} X_{i't'} \Bigg\} \Bigg] \\
= & \sum_{i=1}^{N} \sum_{t=1}^{T} \left[X_{it} \left\{ \frac{(N-1)(T-1)}{NT} X_{it} - \frac{N-1}{NT} \sum_{t' \neq t} X_{it'} - \frac{T-1}{NT} \sum_{i' \neq i} X_{i't} + \frac{1}{NT} \sum_{i' \neq i} \sum_{t' \neq t} X_{i't'} \right\} \right. \\
& \qquad \left. -(1 - X_{it}) \left\{ \frac{(N-1)(T-1)}{NT} X_{it} - \frac{N-1}{NT} \sum_{t' \neq t} X_{it'} - \frac{T-1}{NT} \sum_{i' \neq i} X_{i't} + \frac{1}{NT} \sum_{i' \neq i} \sum_{t' \neq t} X_{i't'} \right\} \right] \\
= & \frac{(T-1)(N-1)}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\{ X_{it} \left(\frac{\sum_{t'=1}^{T} (1 - X_{it'})}{T-1} + \frac{\sum_{i'=1}^{N} (1 - X_{i't})}{N-1} - \frac{\sum_{i' \neq i}^{N} \sum_{t' \neq t}^{T} (1 - X_{i't'})}{(T-1)(N-1)} \right) \right. \right. \\
& \qquad \left. \left. + (1 - X_{it}) \left(\frac{\sum_{t'=1}^{T} X_{it'}}{T-1} + \frac{\sum_{i'=1}^{N} X_{i't}}{N-1} - \frac{\sum_{i' \neq i}^{N} \sum_{t' \neq t}^{T} X_{i't'}}{(T-1)(N-1)} \right) \right\} \right] \\
= & K(T-1)(N-1) \qquad (2)
\end{aligned}
$$

Finally, using the above algebraic equalities, we can derive the desired result as follows,

$$
\begin{aligned}
\hat{\beta}_{\mathsf{FE2}} &= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \overline{X}_i - \overline{X}_t + \overline{X})(Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \overline{X}_i - \overline{X}_t + \overline{X})^2} \\
&= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} Y_{it} - T \sum_{i=1}^{N} \overline{X}_i \overline{Y}_i - N \sum_{t=1}^{T} \overline{X}_t \overline{Y}_t + NT \overline{XY}}{NT\overline{X} - T \sum_{i=1}^{N} \overline{X}_i^2 - N \sum_{t=1} T \overline{X}_t^2 + NT \overline{X}^2} \\
&= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (2X_{it} - 1)(Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (2X_{it} - 1)(X_{it} - \overline{X}_i - \overline{X}_t + \overline{X})} \\
&= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ X_{it}(Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}) - (1 - X_{it})(Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}) \right\}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ X_{it}(X_{it} - \overline{X}_i - \overline{X}_t + \overline{X}) - (1 - X_{it})(X_{it} - \overline{X}_i - \overline{X}_t + \overline{X}) \right\}} \\
&= \frac{1}{K} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)} \right) \right\}
\end{aligned}
$$

where the last equality follows from equation (1) and (2). □

2 Proof of Proposition 2

Proof We first establish the following equality.

$$
\begin{aligned}
& \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} \\
= & \sum_{i'=1}^{N} \sum_{t'=1}^{T} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} w_{it}^{i't'} \right) \\
= & \sum_{i'=1}^{N} \sum_{t'=1}^{T} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{i't'} w_{it}^{i't'} + (1 - X_{i't'}) w_{it}^{i't'} \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'}\left\{X_{i't'}\left(\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{M}^*_{i't'}\cdot\#\mathcal{N}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{N}^*_{i't'}\cdot\#\mathcal{M}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\left(a_{i't'}-\#\mathcal{A}^*_{i't'}+a_{i't'}\right)}{\#\mathcal{A}^*_{i't'}+a_{i't'}}\right)\right.\\
&\quad \left.+(1-X_{i't'})\left(\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{M}^*_{i't'}\cdot\#\mathcal{N}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{N}^*_{i't'}\cdot\#\mathcal{M}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}-\frac{\left(\#\mathcal{A}^*_{i't'}-a_{i't'}-a_{i't'}\right)}{\#\mathcal{A}^*_{i't'}+a_{i't'}}\right)\right\}\\
&= \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'}\left\{X_{i't'}\left(\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\left(a_{i't'}-\#\mathcal{A}^*_{i't'}+a_{i't'}\right)}{\#\mathcal{A}^*_{i't'}+a_{i't'}}\right)\right.\\
&\quad \left.+(1-X_{i't'})\left(\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}-\frac{\left(\#\mathcal{A}^*_{i't'}-a_{i't'}-a_{i't'}\right)}{\#\mathcal{A}^*_{i't'}+a_{i't'}}\right)\right\}\\
&= \sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'}\left(2X_{i't'}+2(1-X_{i't'})\right) = 2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}. \qquad (3)
\end{aligned}
$$

The third equality follows from the fact that for a given unit (i',t') there are $\#\mathcal{M}^*_{i't'}$ matched observations $(i,t)\in\mathcal{M}^*_{i't'}$ with weights equal to $\frac{D_{i't'}K_{i't'}}{\#\mathcal{M}^*_{i't'}} = \frac{D_{i't'}\#\mathcal{A}^*_{i't'}}{\#\mathcal{M}^*_{i't'}(\#\mathcal{A}^*_{i't'}+a_{i't'})} = \frac{D_{i't'}\#\mathcal{N}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}$. Similarly, there are $\#\mathcal{N}^*_{i't'}$ observations $(i,t)\in\mathcal{N}^*_{i't'}$ with weights $\frac{D_{i't'}\#\mathcal{M}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}$. The final matched set $\mathcal{A}^*_{i't'}$ is composed of $a_{i't'}$ observations with the same treatment status with (i',t') and $\mathcal{A}^*_{i't'}-a_{i't'}$ observations with the opposite treatment status. When $X_{i't'}$, the former type gets weight equal to $\frac{D_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}$ while the latter type is weighted by $-\frac{D_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}$. The unit itself gets weight equal to $\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}$. All the other observations will get zero weight.

Following the same logic from above, it is straightforward to show that $\overline{X}^*_i = \overline{X}^*_t = \overline{X}^* = \frac{1}{2}$, and thus

$$
X_{it} - \overline{X}^*_i - \overline{X}^*_t + \overline{X}^* = \begin{cases} \frac{1}{2} & \text{if } X_{it}=1 \\ -\frac{1}{2} & \text{if } X_{it}=0 \end{cases} \qquad (4)
$$

For instance,

$$
\begin{aligned}
\overline{X}^* &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}X_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}}\\
&= \frac{\sum_{i'=1}^{N}\sum_{t'=1}^{T}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}w_{it}^{i't'}\right)}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}}\\
&= \frac{\sum_{i'=1}^{N}\sum_{t'=1}^{T}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{i't'}X_{it}w_{it}^{i't'}+(1-X_{i't'})X_{it}w_{it}^{i't'}\right)}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}}\\
&= \frac{\sum_{i'=1}^{N}\sum_{t'=1}^{T} D_{i't'}X_{i't'}\left(\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}+\frac{a_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}\right)+D_{i't'}(1-X_{i't'})\left(\frac{\#\mathcal{A}^*_{i't'}}{\#\mathcal{A}^*_{i't'}+a_{i't'}}-\frac{-a_{i't'}}{\#\mathcal{A}^*_{i't'}a_{i't'}}\right)}{2\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}}
\end{aligned}
$$

$$= \frac{\sum_{i=1}^N \sum_{t=1}^T D_{it}}{2\sum_{i=1}^N \sum_{t=1}^T D_{it}} = \frac{1}{2}$$

We can derive the desired result.

$$\begin{aligned}
\hat{\beta}_{\mathsf{WFE2}} &= \frac{\sum_{i=1}^N \sum_{t=1}^T W_{it}(X_{it} - \overline{X}_i^* - \overline{X}_t^* + \overline{X}^*)(Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*)}{\sum_{i=1}^N \sum_{t=1}^T W_{it}(X_{it} - \overline{X}_i^* - \overline{X}_t^* + \overline{X}^*)^2} \\
&= \frac{\frac{1}{2}\sum_{i=1}^N \sum_{t=1}^T W_{it}(2X_{it}-1)(Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*)}{\frac{1}{4}\sum_{i=1}^N \sum_{t=1}^T W_{it}} \\
&= \frac{1}{\sum_{i=1}^N \sum_{t=1}^T D_{it}} \sum_{i=1}^N \sum_{t=1}^T W_{it}(2X_{it}-1)(Y_{it} - \overline{Y}_i^* - \overline{Y}_t^* + \overline{Y}^*) \\
&= \frac{1}{\sum_{i=1}^N \sum_{t=1}^T D_{it}} \sum_{i=1}^N \sum_{t=1}^T W_{it}(2X_{it}-1)Y_{it} \\
&= \frac{1}{\sum_{i=1}^N \sum_{t=1}^T D_{it}} \sum_{i=1}^N \sum_{t=1}^T \left\{ \left(\sum_{i'=1}^N \sum_{t'=1}^T w_{it}^{i't'} \right) (2X_{it}-1)Y_{it} \right\} \\
&= \frac{1}{\sum_{i=1}^N \sum_{t=1}^T D_{it}} \sum_{i'=1}^N \sum_{t'=1}^T \left\{ X_{i't'} \left(\sum_{i=1}^N \sum_{t=1}^T w_{it}^{i't'}(2X_{it}-1)Y_{it} \right) + (1 - X_{i't'}) \left(\sum_{i=1}^N \sum_{t=1}^T w_{it}^{i't'}(2X_{it}-1)Y_{it} \right) \right\} \\
&= \frac{1}{\sum_{i=1}^N \sum_{t=1}^T D_{it}} \sum_{i'=1}^N \sum_{t'=1}^T \frac{D_{i't'}}{K_{i't'}} \left\{ X_{i't'} \left(Y_{it} - \frac{\sum_{(i,t)\in\mathcal{M}_{i't'}^*} Y_{it}}{\#\mathcal{M}_{i't'}^*} - \frac{\sum_{(i,t)\in\mathcal{N}_{i't'}^*} Y_{it}}{\#\mathcal{N}_{i't'}^*} + \frac{\sum_{(i,t)\in\mathcal{A}_{i't'}^*} Y_{it}}{\#\mathcal{A}_{i't'}^*} \right) \right. \\
&\qquad \left. + (1 - X_{i't'}) \left(\frac{\sum_{(i,t)\in\mathcal{M}_{i't'}^*} Y_{it}}{\#\mathcal{M}_{i't'}^*} + \frac{\sum_{(i,t)\in\mathcal{N}_{i't'}^*} Y_{it}}{\#\mathcal{N}_{i't'}^*} - \frac{\sum_{(i,t)\in\mathcal{A}_{i't'}^*} Y_{it}}{\#\mathcal{A}_{i't'}^*} - Y_{it} \right) \right\} \\
&= \frac{1}{\sum_{i=1}^N \sum_{t=1}^T D_{it}} \sum_{i=1}^N \sum_{t=1}^T \frac{D_{it}}{K_{it}} (\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)}) = \hat{\tau}_{\mathsf{match2}}
\end{aligned}$$

where the second and third equality follows from equation (3) and (4). The last two equalities follow from applying the definition of K_{it}, W_{it} $\widehat{Y_{it}(1)}$ and $\widehat{Y_{it}(0)}$ given in Proposition 2. □