Two-Way Fixed Effects: Or How I Learned to Stop Worrying and Just Love Unobserved Heterogeneity

Jonathan Kropko
University of Virginia
jkropko@virginia.edu

Robert Kubinec
University of Virginia
rmk7xy@virginia.edu

October 17, 2016

Abstract

Time-series cross-sectional (TSCS) data contain a sample of cases observed at repeated time points. Researchers commonly employ fixed effects (FEs) on the cases to remove cross-sectional unobserved heterogeneity from the model. Recently, a great deal of applied work includes FEs on cases and on time points in the same model with the intention of accounting for omitted variables in both the cross-sectional and time dimensions. The properties of the model that includes FEs on both cases and time are not well understood. We derive the formal two-way FE estimator and show that it does not account for unobserved heterogeneity in either the cross-sectional or the time dimension. We further demonstrate that the two-way FE model is sensitive to whether the panels are balanced while a model that includes FEs only on cases or only on time points is not. Using an analysis of the relationship between a country's wealth and level of democracy, we show that the choice of model has a profound influence on the findings. We recommend that researchers avoid the two-way FE model, and instead use a model with FEs only on cases or only on time points, a choice that depends on the research question.
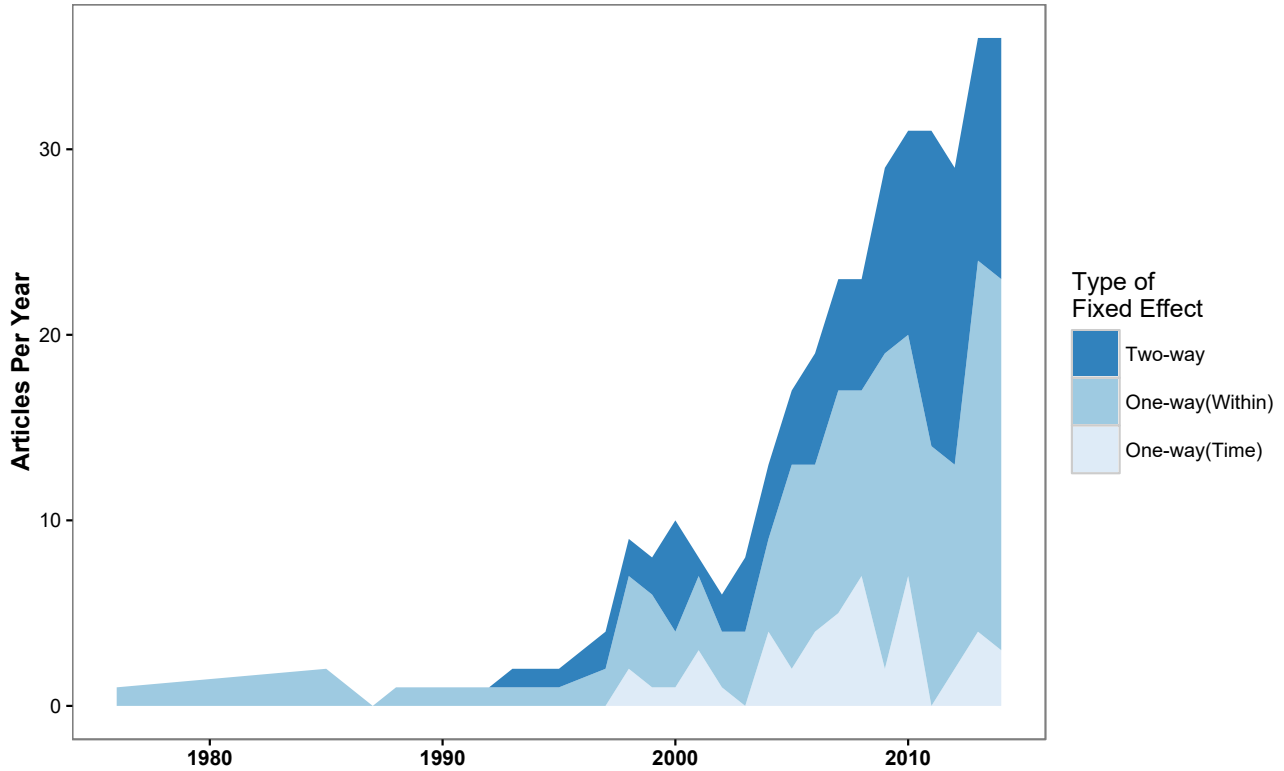
Figure 1: Type of Fixed Effects Employed by Political Science Articles (1976-2014)

1   Meta-Analysis

To examine the prevalence of one and two-way fixed effects models in the political science literature, we collected methodological information on a sample of 363 articles from the *American Political Science Review*, *American Journal of Political Science*, and the *Journal of Politics* for the period 1976 to 2014. The bulk of the articles were collected through JSTOR's Data for Research service, while the most recent years were collected through each journal separately. The search term used to create the dataset was "fixed effect". Papers that were not case, time or case/time fixed effects time-series cross section models, such as hierarchical models, random-effects models, and cross-sectional models, were removed.

What Figure 1 indicates is that the vast majority of fixed effects models employed in the political science literature are either case FX models or case/time FX models. In particular, a noticeable trend over the last ten years has been towards greater usage of case/time FXs, although case FXs remain the dominant empirical estimation strategy for TSCS data.

What is also evident is that the total number of FX models continue to increase year-on-year. This is likely due to the greater prevalence of TSCS datasets, which we regard as a great asset to the discipline. However, despite these models' greater usage, their interpretation does not appear to have grown more consistent or clear over time. Rather, there are several articles which tend to be frequently cited to justify one empirical strategy or another, with decisions often being made around the limitations of the available data.

Qualitatively, it is apparent that certain papers have guided the bulk of time series cross section in the discipline. Stimson (1985) first presented the "least-squares with dummy variables" (LSDV) approach, which clarified the nature of fixed effects models although it did not provide specific recommendations as to what kind of case or time FX to use. Beck, Katz, and Tucker (1998) and Beck and Katz (1995) advanced the conversation considerably by providing more specific recommendations for how political scientists should run TSCS models depending on the structure of the data, i.e., different ratios of cases versus time points. More recently, alternative to traditional fixed effects models, such as estimation of "rarely changing variables" (Plümper & Troeger, 2007) and difference-in-difference designs (Abadie, 2005) have diversified fixed effects models to an extent, although the original papers by Beck and Katz and Stimson are still very influential, as can be seen in Figure 2.
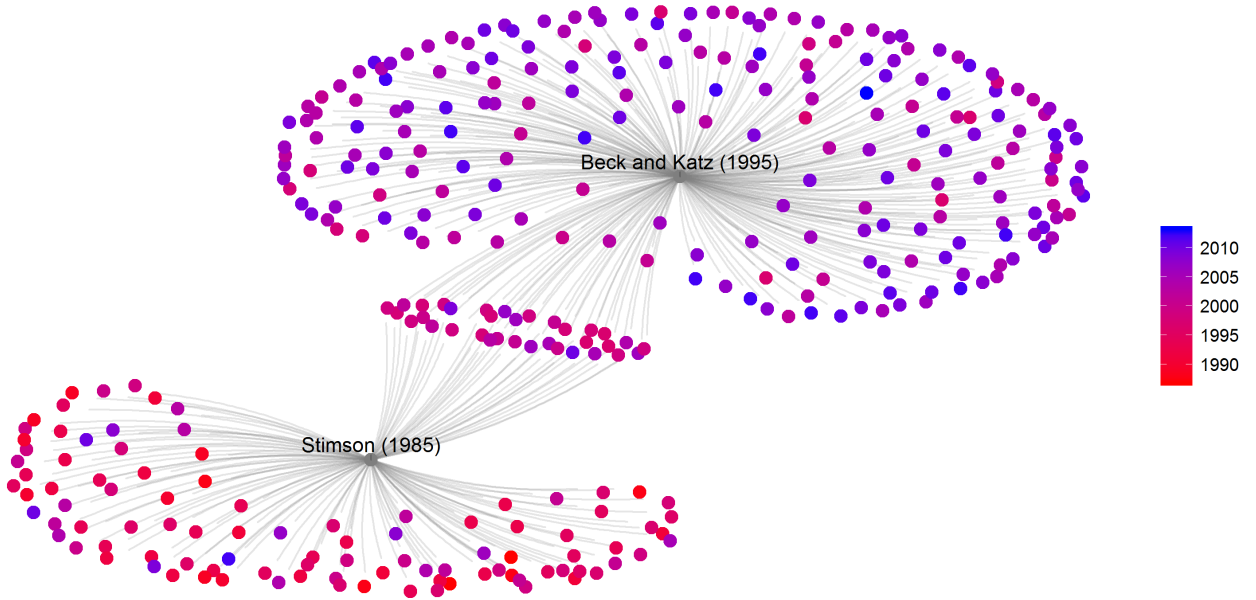
Figure 2: Citations of Beck and Katz (1995) and Stimson (1985) in political science journals in JSTOR

Yet despite the wide uniformity in methods citations in fixed effects papers, there is also diversity in their interpretation and application. Generally speaking, the econometric notion that fixed effects estimators exist to correct for "unobserved heterogeneity" (Arellano, 2003) has meant that authors have come to prefer fixed effects for its ability to remove potential confounders. However, there is evident disagreement over what to do when a certain fixed effects model does not fit the data collected, particularly when most of the variables are constant within cases.

Scholars generally use reasoning along the lines of Donno (2013), who did not use a case FX model because "fixed effects leads to too many observations dropping from the analysis" (p. 709). The problem, however, is not a lack of statistical power, but rather a mis-application of the case FX estimator. In Donno's case, the problem appears to be that she is interested in a variable that does not often change within countries over time, that of the type of electoral system (p. 709). Donno could use a fixed effects model, but as her hypothesis relates more to differences between countries than within countries, a time FX model would probably capture the variation in the dataset more appropriately.

Case/time fixed effects suffer as well from the perception that they are more rigorous than case FX models but also require even more data. For example, Gabel, Carruba, Ansley, and Beaudette (2012) argue that they need both case and time FX because they want to control for "unobserved national factors" (case FX) and "temporal shocks" (time FX) (p. 1132). Similarly, Scheve and Stasavage (2012) use case/time FX within their "difference-in-difference" framework because case FX "control for time-constant unobserved country-level heterogeneity" while time FX will account for "common shocks" (p. 82-83).

The use of case/time effects, particularly in terms of the emergent difference-in-difference framework, introduces a new wrinkle in fixed effects estimation. The standard difference-in-difference framework is based on binary time points (Khandker, Koolwal, & Samad, 2010), but the TSCS application often incorporates many time points (Scheve & Stasavage, 2012). The extension of difference-in-difference to multiple time points is in essence a different model, as the repeated application of treatments over time represents a multiplication of potential outcomes (Imai & Kim, 2016). With a binary treatment and two time points, the case/time effects model yields a valid estimate of the difference-in-difference framework because it removes any variation that is not relevant to the yearly difference between control and treatment groups. It is the fact that this binary model has a clear causal interpretation given certain assumptions that appears to underline its growing use in the literature. However, extending the model to multiple time points and continuous treatment variables produces the same difficulties in interpretation that bedevil case/time FX estimation in general, as we describe in the simulation section.

Given the emphasis on case/time FX and case FX, it is somewhat surprising that time FX have also been used consistently over the past two decades. However, the identification of fixed effects estimation with either case or case/time FX has meant that time FX models are not always recognized as such, and often their usage depends

on traditions within certain streams of scholarship. For example, in the American political science literature on courts, it is common to include time fixed effects for the tenure of a court, administration or Congress (Jenkins & Monroe, 2012; Hall, 2014; Boyd, Epstein, & Martin, 2010; Lazarus, 2009). This usage highlights the fact that case and time FX are merely two sides of the same coin: they splice up a dataset along two different dimensions, but these dimensions can be relabeled without changing the underlying substance.

2 Empirical Analysis

To demonstrate the difference between case effects, time effects and the simultaneous use of both effects, we undertook an analysis of a research question with a considerable history in time-series cross-section modeling: the regression of gross domestic product (GDP) on democracy. To measure the dependent variable, we used the Electoral Democracy Index from the Varieties of Democratization Project (V-DEM) (Coppedge et al., 2016), which utilizes multiple indicators of electoral democracy in an item-response model to produce a democracy index that takes into account measurement uncertainty. For the independent variable of interest, GDP, and all control variables, we used ancillary data provided by V-DEM. Descriptive statistics are shown in Table 1.

Variable	N	Mean	St. Dev.	Min	Max
Log of GDP Per Capita	10,445	7.812	1.023	5.315	10.667
Electoral Democracy Index	16,224	0.321	0.282	0.008	0.958
# World Democracies	10,154	30.104	7.895	13.043	48.765
Capital as % GDP	5,472	0.520	0.500	0	1
Fuel Income Per Capita	10,211	333.186	2,251.948	0.000	81,161.850
Fertility Rate	10,924	4.433	2.013	0.895	9.223
Internal Conflicts	13,799	0.080	0.271	0	1
Urban Pop (Millions)	15,169	8.629	28.777	0.0001	608.418

Table 1: Descriptive Statistics

Scholars have examined the question of whether economic development causes democratization using a variety of statistical models, including linear models with case effects (Kennedy, 2010) and both case and time fixed effects (Acemoğlu, Robinson, Johnson, & Yared, 2009; Boix, 2011; Haber & Menaldo, 2011). Despite widespread empirical tests, there is as yet no general consensus on the relationship between GDP growth and democratization, with some positing an intermediating effect of inequality Acemoğlu and Robinson, 2006; Houle, 2009; Boix and Stokes, 2003, others the intermediating effect of fuel exports (Haber & Menaldo, 2011; Andersen & Ross, 2014), and still others that the relationship is unconditionally positive, with economic development inevitably leading to democratization (Inglehart & Welzel, 2005). Finally, there are skeptics who argue that there is in fact no effect of GDP growth on democratization, but rather on democratic consolidation (Limongi & Przeworski, 1997)[1].

As we mentioned earlier, the different kinds of fixed effects models naturally relate to different kinds of hypotheses. For the GDP and democracy research question, the hypotheses can be phrased as follows for the different kinds of models:

Case FX — As a country's GDP reaches above-average levels, that country is likely to become more democratic.

Time FX — As a country's GDP rises above the global yearly average, that country is likely to become more democratic.

Case & Time FX — As a country's GDP increases as a ratio of its average level over time to the global average at a particular year, then that country is likely to become more democratic.

We first analyze the case & time effects model, which for illustration purposes we calculate for Ukraine for a ten year period in Table 2. This table shows that the case averages, x_i and y_i, are fixed for Ukraine as each country represents a single case, while the global yearly average for the independent variable, x_t, varies from year to year. The resulting two-way effect, shown in the final column, is a ratio of the case variation to time variation, although

---

[1]For a full review of this literature, see Coppedge (2012)

Year	x_{it}	$\bar{x}_i$	$\bar{x}_t$	y_{it}	$\bar{y}_i$	$(x_{it}-\bar{x}_t)$	$(y_{it}-\bar{y}_t)$	$(x_{it}-\bar{x}_i)$	$\frac{(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t)}{(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_t)}$
2001	8.02	8.31	8.12	0.40	0.51	-0.10	-0.10	-0.30	-0.15
2002	8.08	8.31	8.14	0.41	0.51	-0.06	-0.09	-0.24	-0.15
2003	8.18	8.31	8.16	0.41	0.51	0.01	-0.10	-0.14	-0.15
2004	8.30	8.31	8.20	0.37	0.51	0.09	-0.14	-0.01	-0.15
2005	8.33	8.31	8.24	0.46	0.51	0.09	-0.05	0.02	-0.15
2006	8.41	8.31	8.28	0.57	0.51	0.13	0.06	0.10	-0.15
2007	8.49	8.31	8.33	0.62	0.51	0.17	0.11	0.18	-0.15
2008	8.52	8.31	8.35	0.63	0.51	0.17	0.12	0.21	-0.15
2009	8.37	8.31	8.67	0.64	0.51	-0.30	0.14	0.06	-0.15
2010	8.42	8.31	8.71	0.56	0.51	-0.29	0.05	0.11	-0.15

Table 2: Case/Time Fixed Effect for Ukraine 2001-2010

the dependent variable information is only incorporated in the numerator. As can be seen, it is not straightforward to decipher how the case & time FX estimator relates to the underlying components of variance.

Fixed Effect Type	Coefficient	SE	High 95% CI	Low 95% CI
Between Effect	0.191	0.003	0.198	0.184
Within Effect	0.166	0.003	0.171	0.161
Two-Way Effect	0.059	0.005	0.069	0.048

Table 3: Case, Time and Case/Time Fixed Effects Estimates of GDP on Democracy

The results for the case, time and case/time fixed effects models[2] for the regression of GDP on democracy are shown in Table 11. As can be seen, for this particular model, the case effects and the time effects are very close to each other, with similar standard errors. However, the case/time fixed effect is noticeably different, still positive, but with decreased magnitude and a higher standard error. As we argue, each of these models are statistically identified, but the case and time effects models relate more clearly to research questions of interest, namely whether economic development is associated with democratization over time, whether in terms of within-country economic development or economic development relative to other countries. These two models, case and time effects, can answer complementary aspects of this larger hypothesis. Each dimension of variation–whether within a country over a time or between countries at a single point in time–has different implications for the substantive theories being tested.

To illustrate this point, we also estimated case and time FX models in which we interacted the independent variable with the fixed effect. These results, shown in Figures 4 and 3, reveal that there is substantial heterogeneity in the underlying association, and that this heterogeneity differs whether the model is a time or case FX. For time effects in Figure 4, GDP's positive relationship with democracy was strongest prior to World War II, and has fallen to a third of its prewar level, a similar relationship to that described by Boix, Miller, and Rosato (2013). For case effects shown in Figure 3, the effect of GDP on democracy is in fact negative for a substantial minority of countries, while strongly positive for others. Countries at each end of the spectrum accord well with qualitative knowledge about their regimes, such as Cuba and Belarussia occupying the negative GDP on democracy side, while Benin, Costa Rica and Poland occupy the positive end. This diversity in effects across time and cases suggests that more attention to the relationship between theory and method would benefit this particular research field and help unravel some of these intriguing empirical patterns.

Finally, we also ran models with controls to show how the selection of variables depends on the research question. If the dimension of variation is case effects, then controls should be selected that also vary within cases over time. If the dimension of variation is time effects, then controls should be selected that vary across countries at a certain point in time. We included five control variables in the model results reported in Table 4: the percent of GDP that is comprised by returns to capital, a rough measure of inequality; the number of internal conflicts in a country per year, the size of the urban population, the total amount of GDP income from fuels, and the total number of democracies in the world. Three of these variables, capital as a share of GDP, urbanization and fuel income, vary both within cases over time and vary between cases at particular time points. Internal conflicts is a variable which

---

[2]We followed Bizarro, Coppedge, and Pemstein (2016)'s advice for incorporating measurement uncertainty by bootstrapping the regression coefficients over the dependent variable's MCMC iterations.
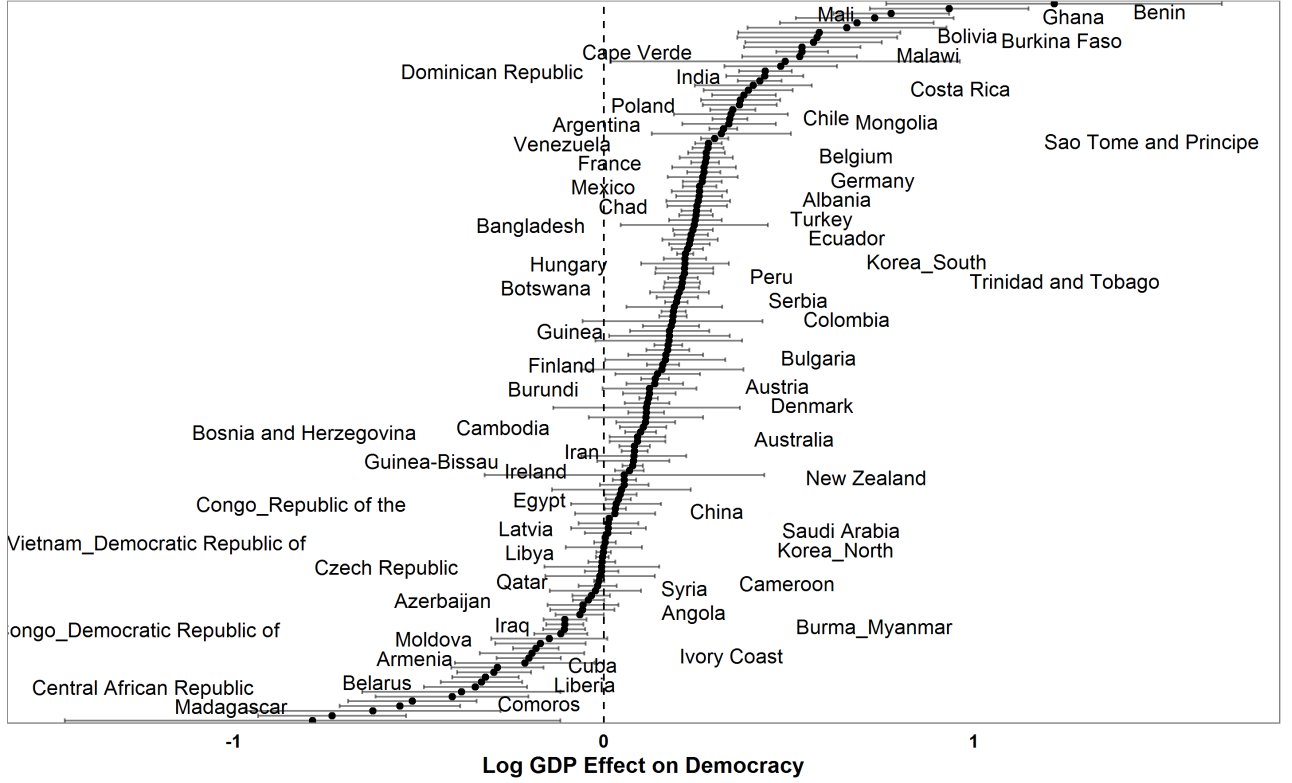
Figure 3: GDP on Democracy with Case Effects and Country Interactions

Note: Dotted line indicates effect of zero.

is more appropriate as a time effect control because it is difficult to conceive of a country having an "average" level of internal violent conflicts. The number of world democracies, on the other hand, is more appropriate in a case effects model because it does not vary between countries in a given year. For that reason, any kind of democratic diffusion model is inherently a case effects model. The coefficients on the included controls in Table 4 are broadly in line with the previous literature, with fuel income having a zero effect for case effects and a negative effect for time effects. We would also note that the GDP per capita coefficient, while it is significantly smaller with case effects and controls, is still positive in both models.

To summarize, GDP appears to have a strong, positive effect on democracy. This effect is present in the case effect, time effect and case/time effects models, but is strongest in the time effect model. In other words, as countries' GDP rises higher than the global average, they are more likely to be democratic. The use of control variables indicates that the common belief that a case effect model has better causal identification does not appear to be true in this case: the coefficient on the case effect model changed significantly with included controls, while the time effect model did not.

3   Proofs

Lemma 1a. The following equivalence holds in both balanced and unbalanced panels:

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i) = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}) = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x})(y_{it}-\bar{y}_i). \tag{1}$$

Proof. We prove the lemma for unbalanced panels, but the lemma also holds for balanced panels since balanced panels are the special case in which $T_i = T_j, \forall i, j \in \{1, \ldots, N\}$. We demonstrate that all three expressions are
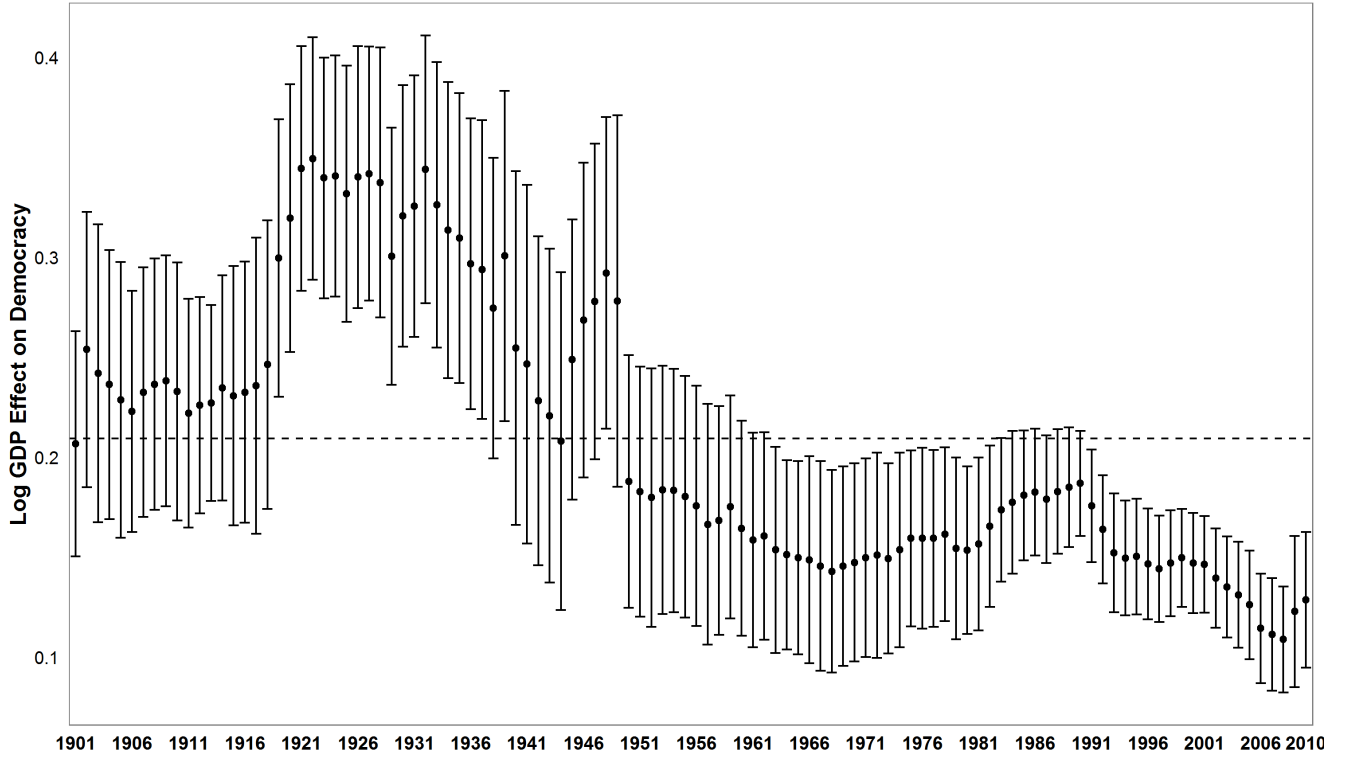
Figure 4: GDP on Democracy with Time Effects and Year Interactions

Note: Dotted line indicates the average effect value over time.

individually equal to $\sum_{i=1}^{N} \sum_{t=1}^{T_i} x_{it} y_{it} - \sum_{i=1}^{N} T_i \bar{x}_i \bar{y}_i$ and are therefore equal to each other:

$$
\begin{aligned}
\sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) &= \sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} y_{it} - x_{it} \bar{y}_i - \bar{x}_i y_{it} + \bar{x}_i \bar{y}_i) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T_i} x_{it} y_{it} - \sum_{i=1}^{N} \sum_{t=1}^{T_i} x_{it} \bar{y}_i - \sum_{i=1}^{N} \sum_{t=1}^{T_i} \bar{x}_i y_{it} + \sum_{i=1}^{N} \sum_{t=1}^{T_i} \bar{x}_i \bar{y}_i \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T_i} x_{it} y_{it} - \sum_{i=1}^{N} \bar{y}_i \sum_{t=1}^{T_i} x_{it} - \sum_{i=1}^{N} \bar{x}_i \sum_{t=1}^{T_i} y_{it} + \sum_{i=1}^{N} T_i \bar{x}_i \bar{y}_i \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T_i} x_{it} y_{it} - \sum_{i=1}^{N} \bar{y}_i (T_i \bar{x}_i) - \sum_{i=1}^{N} \bar{x}_i (T_i \bar{y}_i) + \sum_{i=1}^{N} T_i \bar{x}_i \bar{y}_i \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T_i} x_{it} y_{it} - \sum_{i=1}^{N} T_i \bar{x}_i \bar{y}_i.
\end{aligned}
\tag{2}
$$

Variable	Case Effects	Time Effects
GDP Per Capita	0.036	0.170
	(0.019, 0.054)	(0.163, 0.177)
% Capital of GDP	-0.111	-0.070
	(-0.181,-0.040)	(-0.084,-0.055)
Internal Conflicts		-0.073
		(-0.106,-0.041)
No. Urban Population	0.000	0.000
	(-0.001, 0.000)	(0.000, 0.000)
Total Fuel Income Per Capita	0.000	-0.036
	(-0.001, 0.001)	(-0.043,-0.028)
No. World Democracies	0.012	
	(0.012, 0.013)	

Table 4: GDP on Democracy with Controls

Note: 95% confidence intervals are included below the coefficients.

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}) &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}y_{it}-x_{it}\bar{y}-\bar{x}_i y_{it}+\bar{x}_i\bar{y}) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}\bar{y}-\sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}_i y_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}_i\bar{y} \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\bar{y}\sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}-\sum_{i=1}^{N}\bar{x}_i\sum_{t=1}^{T_i}y_{it}+\bar{y}\sum_{i=1}^{N}T_i\bar{x}_i \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\bar{y}\sum_{i=1}^{N}T_i\bar{x}_i-\sum_{i=1}^{N}\bar{x}_i(T_i\bar{y}_i)+\bar{y}\sum_{i=1}^{N}T_i\bar{x}_i \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\sum_{i=1}^{N}T_i\bar{x}_i\bar{y}_i.
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x})(y_{it}-\bar{y}_i) &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}y_{it}-x_{it}\bar{y}_i-\bar{x}y_{it}+\bar{x}\bar{y}_i) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}\bar{y}_i-\sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}y_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}\bar{y}_i \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\sum_{i=1}^{N}\bar{y}_i\sum_{t=1}^{T_i}x_{it}-\bar{x}\sum_{i=1}^{N}\sum_{t=1}^{T_i}y_{it}+\bar{x}\sum_{i=1}^{N}T_i\bar{y}_i \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\sum_{i=1}^{N}\bar{y}_i(T_i\bar{x}_i)-\bar{x}\sum_{i=1}^{N}T_i\bar{y}_i+\bar{x}\sum_{i=1}^{N}T_i\bar{y}_i \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}y_{it}-\sum_{i=1}^{N}T_i\bar{x}_i\bar{y}_i. \quad\blacksquare
\end{aligned}
\tag{4}
$$

Lemma 1b. The following equivalence holds in both balanced and unbalanced panels:

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_t)(y_{it}-\bar{y}_t) = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_t)(y_{it}-\bar{y}) = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x})(y_{it}-\bar{y}_t). \tag{5}$$

Proof. The proof follows the proof of lemma 1a, substituting $\bar{x}_t$ for $\bar{x}_i$ and $\bar{y}_t$ for $\bar{y}_i$ and rewriting the summation as $\sum_{t=1}^{T}\sum_{i=1}^{N_t}$. ∎

Lemma 2a. The following equivalence holds in both balanced and unbalanced panels:

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)^2 = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}). \tag{6}$$

Proof. We prove the lemma for unbalanced panels, but the lemma also holds for balanced panels since balanced panels are the special case in which $T_i = T_j, \forall i,j \in \{1,\ldots,N\}$. We demonstrate that both expressions are individually equal to $\sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - \sum_{i=1}^{N}T_i\bar{x}_i^2$ and are therefore equal to each other:

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)^2 &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}^2 - 2x_{it}\bar{x}_i + \bar{x}_i^2)\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - 2\sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}\bar{x}_i + \sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}_i^2\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - 2\sum_{i=1}^{N}\bar{x}_i\sum_{t=1}^{T_i}x_{it} + \sum_{i=1}^{N}T_i\bar{x}_i^2\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - 2\sum_{i=1}^{N}T_i\bar{x}_i^2 + \sum_{i=1}^{N}T_i\bar{x}_i^2\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - \sum_{i=1}^{N}T_i\bar{x}_i^2. \tag{7}
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}) &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}^2 - \bar{x}x_{it} - \bar{x}_i x_{it} + \bar{x}_i\bar{x})\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - \sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}x_{it} - \sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}_i x_{it} + \sum_{i=1}^{N}\sum_{t=1}^{T_i}\bar{x}_i\bar{x}\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - \bar{x}\sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it} - \sum_{i=1}^{N}\bar{x}_i\sum_{t=1}^{T_i}x_{it} + \bar{x}\sum_{i=1}^{N}T_i\bar{x}_i\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - \bar{x}\sum_{i=1}^{N}T_i\bar{x}_i - \sum_{i=1}^{N}T_i\bar{x}_i^2 + \bar{x}\sum_{i=1}^{N}T_i\bar{x}_i\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}x_{it}^2 - \sum_{i=1}^{N}T_i\bar{x}_i^2. \quad ∎ \tag{8}
\end{aligned}$$

Lemma 2b. The following equivalence holds in both balanced and unbalanced panels:

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_t)^2 = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\bar{x}_t)(x_{it}-\bar{x}). \tag{9}$$

Proof. The proof follows the proof of lemma 2a, substituting $\bar{x}_t$ for $\bar{x}_i$ and rewriting the summation as $\sum_{t=1}^{T}\sum_{i=1}^{N_t}$. ∎

Theorem 1. The two-way fixed effects estimator is a weighted average of five coefficient estimates: (1) the pooled OLS estimator (β_{pool}), (2) the case-level fixed effects estimator (β_{caseFE}), (3) the time-level fixed effects estimator (β_{timeFE}), (4) the OLS estimator applied to the model that removes the case-level means from the outcome and the time-level means from the predictor (β_{casetime}), and (5) the OLS estimator applied to the model that removes the time-level means from the outcome and the case-level means from the predictor (β_{timecase}). Specifically, the two-way fixed effects estimator is

$$\beta_{TW} = \frac{\omega_1 \beta_{\text{pool}} - \omega_2 \beta_{\text{caseFE}} - \omega_3 \beta_{\text{timeFE}} + \omega_4 \beta_{\text{casetime}} + \omega_5 \beta_{\text{timecase}}}{\omega_1 - \omega_2 - \omega_3 + \omega_4 + \omega_5}, \tag{10}$$

where

$$\omega_1 = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x})^2, \quad \omega_2 = -\sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)^2, \quad \omega_3 = -\sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x}_t)^2, \quad \omega_4 = \omega_5 = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t).$$

Proof. The two-way fixed effects estimator is the OLS estimator applied to the following model,

$$y_{it}^* = \alpha_{TW} + \beta_{TW} x_{it}^* + \varepsilon_{it}, \tag{11}$$

where $*$ denotes the transformation

$$x_{it}^* = x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}. \tag{12}$$

By OLS, the coefficient in equation 11 is given by

$$\begin{aligned}
\beta_{TW} &= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left(x_{it} - \bar{x}_i - \bar{x}_t + \bar{x} \right) \left(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y} \right)}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left(x_{it} - \bar{x}_i - \bar{x}_t + \bar{x} \right)^2} \\[2mm]
&= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left(x_{it} + x_{it} - x_{it} - \bar{x}_i - \bar{x}_t + \bar{x} \right) \left(y_{it} + y_{it} - y_{it} - \bar{y}_i - \bar{y}_t + \bar{y} \right)}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left(x_{it} + x_{it} - x_{it} - \bar{x}_i - \bar{x}_t + \bar{x} \right)^2} \\[2mm]
&= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left[(x_{it} - \bar{x}_i) + (x_{it} - \bar{x}_t) - (x_{it} - \bar{x}) \right] \left[(y_{it} - \bar{y}_i) + (y_{it} - \bar{y}_t) - (y_{it} - \bar{y}) \right]}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left[(x_{it} - \bar{x}_i) + (x_{it} - \bar{x}_t) - (x_{it} - \bar{x}) \right]^2}, \\[2mm]
&= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} A_{it}}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} B_{it}}, \tag{13}
\end{aligned}$$

where

$$\begin{aligned}
A_{it} ={}& (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t) - (x_{it} - \bar{x}_i)(y_{it} - \bar{y}) \\
&+ (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i) + (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_t) - (x_{it} - \bar{x}_t)(y_{it} - \bar{y}) \\
&- (x_{it} - \bar{x})(y_{it} - \bar{y}_i) - (x_{it} - \bar{x})(y_{it} - \bar{y}_t) + (x_{it} - \bar{x})(y_{it} - \bar{y}), \tag{14}
\end{aligned}$$

and

$$\begin{aligned}
B_{it} ={}& (x_{it} - \bar{x}_i)^2 + (x_{it} - \bar{x}_t)^2 + (x_{it} - \bar{x})^2 \\
&+ 2(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t) - 2(x_{it} - \bar{x}_i)(x_{it} - \bar{x}) - 2(x_{it} - \bar{x}_t)(x_{it} - \bar{x}). \tag{15}
\end{aligned}$$

From lemmas 1a, 1b, 2a, and 2b, these expressions reduce to

$$\begin{aligned}
A_{it} ={}& (x_{it} - \bar{x})(y_{it} - \bar{y}) - (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_t) \\
&+ (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t) + (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i), \tag{16}
\end{aligned}$$

and

$$B_{it} = (x_{it} - \bar{x})^2 - (x_{it} - \bar{x}_i)^2 - (x_{it} - \bar{x}_t)^2 + 2(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t). \tag{17}$$

Note that the pooled OLS estimator for the coefficient is

$$\beta_{\text{pool}} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x})(y_{it} - \bar{y})}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x})^2}, \tag{18}$$

the case fixed effects estimator for the coefficient is

$$\beta_{\text{caseFE}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)^2}, \tag{19}$$

and the time fixed effects estimator for the coefficient is

$$\beta_{\text{timeFE}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_t)(y_{it} - \bar{y}_t)}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_t)^2}. \tag{20}$$

In addition, define β_{casetime} to be the OLS coefficient obtained by removing the case-level means from the outcome and the time-level means from the predictor, given by

$$\beta_{\text{casetime}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t)}. \tag{21}$$

Likewise, define β_{timecase} to be the OLS coefficient obtained by removing the time-level means from the outcome and the case-level means from the predictor, given by

$$\beta_{\text{timecase}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t)}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t)}. \tag{22}$$

The two-way fixed effects estimator is therefore a weighted average of the preceding five estimates, given by

$$\beta_{TW} = \frac{\omega_1\beta_{\text{pool}} - \omega_2\beta_{\text{caseFE}} - \omega_3\beta_{\text{timeFE}} + \omega_4\beta_{\text{casetime}} + \omega_5\beta_{\text{timecase}}}{\omega_1 - \omega_2 - \omega_3 + \omega_4 + \omega_5}, \tag{23}$$

where

$$\omega_1 = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x})^2, \quad \omega_2 = -\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)^2, \quad \omega_3 = -\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_t)^2, \quad \omega_4 = \omega_5 = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t).$$

Lemma 3. If the panels in the data are balanced, then

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t) = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i). \tag{24}$$

Proof. This proof depends on the fact that

$$\bar{x} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}}{NT} = \frac{\sum_{i=1}^{N}\bar{x}_i}{N} = \frac{\sum_{t=1}^{T}\bar{x}_t}{T} \tag{25}$$

if and only if the panels are balanced:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t) = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}y_{it}-x_{it}\bar{y}_t-\bar{x}_i y_{it}+\bar{x}_i\bar{y}_t)$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it} - \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{y}_t - \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i y_{it} + \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i\bar{y}_t$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it} - \sum_{t=1}^{T}\bar{y}_t\sum_{i=1}^{N}x_{it} - \sum_{i=1}^{N}\bar{x}_i\sum_{t=1}^{T}y_{it} + \sum_{i=1}^{N}\bar{x}_i\sum_{t=1}^{T}\bar{y}_t$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it} - N\sum_{t=1}^{T}\bar{x}_t\bar{y}_t - T\sum_{i=1}^{N}\bar{x}_i\bar{y}_i + \sum_{i=1}^{N}\bar{x}_i\sum_{t=1}^{T}\bar{y}_t$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it} - N\sum_{t=1}^{T}\bar{x}_t\bar{y}_t - T\sum_{i=1}^{N}\bar{x}_i\bar{y}_i + NT\bar{x}\bar{y}$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it} - \sum_{t=1}^{T}\bar{x}_t\sum_{i=1}^{N}y_{it} - \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{y}_i + \sum_{i=1}^{N}\bar{x}_t\sum_{t=1}^{T}\bar{y}_i$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}y_{it}-\bar{x}_t y_{it}-x_{it}\bar{y}_i+\bar{x}_t\bar{y}_i)$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_t)(y_{it}-\bar{y}_i). \quad \blacksquare \tag{26}$$

Lemma 4. If the panels in the data are balanced, then

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)+(x_{it}-\bar{x}_t)(y_{it}-\bar{y}_t)-(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t)\right] = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})(y_{it}-\bar{y}). \tag{27}$$

Proof.

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\Big[(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)+(x_{it}-\bar{x}_t)(y_{it}-\bar{y}_t)-(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t)\Big]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\Big[(x_{it}y_{it}-x_{it}\bar{y}_i-\bar{x}_iy_{it}+\bar{x}_i\bar{y}_i)+(x_{it}y_{it}-x_{it}\bar{y}_t-\bar{x}_ty_{it}+\bar{x}_t\bar{y}_t)-(x_{it}y_{it}-x_{it}\bar{y}_t-\bar{x}_iy_{it}+\bar{x}_i\bar{y}_t)\Big]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\Big[(x_{it}y_{it}+x_{it}y_{it}-x_{it}y_{it})-x_{it}(\bar{y}_i+\bar{y}_t-\bar{y}_t)-(\bar{x}_i+\bar{x}_t-\bar{x}_i)y_{it}+\bar{x}_i\bar{y}_i+\bar{x}_t\bar{y}_t-\bar{x}_i\bar{y}_t\Big]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\Big[x_{it}y_{it}-x_{it}\bar{y}_i-\bar{x}_ty_{it}+\bar{x}_i\bar{y}_i+\bar{x}_t\bar{y}_t-\bar{x}_i\bar{y}_t\Big]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it}-\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{y}_i-\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_ty_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i\bar{y}_i+\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_t\bar{y}_t-\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i\bar{y}_t$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it}-T\sum_{i=1}^{N}\bar{x}_i\bar{y}_i-N\sum_{t=1}^{T}\bar{x}_t\bar{y}_t+T\sum_{i=1}^{N}\bar{x}_i\bar{y}_i+N\sum_{t=1}^{T}\bar{x}_t\bar{y}_t-NT\bar{x}\bar{y}$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it}-NT\bar{x}\bar{y}$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it}-NT\bar{x}\bar{y}-NT\bar{x}\bar{y}+NT\bar{x}\bar{y}$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}y_{it}-\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{y}-\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}y_{it}+\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}\bar{y}$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}y_{it}-x_{it}\bar{y}-\bar{x}y_{it}+\bar{x}\bar{y})$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})(y_{it}-\bar{y}).\quad\blacksquare \tag{28}$$

Lemma 5. If the panels in the data are balanced, then

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\Big[(x_{it}-\bar{x}_i)^2+(x_{it}-\bar{x}_t)^2-(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_t)\Big]=\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2. \tag{29}$$

Proof.

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(x_{it}-\bar{x}_i)^2 + (x_{it}-\bar{x}_t)^2 - (x_{it}-\bar{x}_i)(x_{it}-\bar{x}_t)\right]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(x_{it}^2 - 2x_{it}\bar{x}_i + \bar{x}_i^2) + (x_{it}^2 - 2x_{it}\bar{x}_t + \bar{x}_t^2) - (x_{it}^2 - x_{it}\bar{x}_t - x_{it}\bar{x}_i + \bar{x}_i\bar{x}_t)\right]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(x_{it}^2 + x_{it}^2 - x_{it}^2) - x_{it}(2\bar{x}_i + 2\bar{x}_t - \bar{x}_i - \bar{x}_t) + \bar{x}_i^2 + \bar{x}_t^2 - \bar{x}_i\bar{x}_t\right]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\left[x_{it}^2 - x_{it}\bar{x}_i + x_{it}\bar{x}_t + \bar{x}_i^2 + \bar{x}_t^2 - \bar{x}_i\bar{x}_t\right]$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{x}_i + \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{x}_t + \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i^2 + \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_t^2 - \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i\bar{x}_t$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - T\sum_{i=1}^{N}\bar{x}_i^2 + N\sum_{t=1}^{T}\bar{x}_t^2 + T\sum_{i=1}^{N}\bar{x}_i^2 + N\sum_{t=1}^{T}\bar{x}_t^2 - NT\bar{x}^2$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - NT\bar{x}^2$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - NT\bar{x}^2 + NT\bar{x}^2 - NT\bar{x}^2$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - 2\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}\bar{x} + \sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}^2$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}^2 - 2x_{it}\bar{x} + \bar{x}^2)$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2. \quad \blacksquare \tag{30}$$

Theorem 2. If the panels in the data are balanced, then the two-way fixed effect estimator is given by

$$\beta_{TW} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_t)}, \tag{31}$$

or equivalently as

$$\beta_{TW} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_t)(y_{it}-\bar{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_t)}. \tag{32}$$

Proof. From the proof of theorem 1, the two-way fixed effects estimator can be written as

$$\beta_{TW} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}A_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}B_{it}}, \tag{33}$$

where

$$A_{it} = (x_{it}-\bar{x})(y_{it}-\bar{y}) - (x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i) - (x_{it}-\bar{x}_t)(y_{it}-\bar{y}_t)$$
$$+ (x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t) + (x_{it}-\bar{x}_t)(y_{it}-\bar{y}_i), \tag{34}$$

and

$$B_{it} = (x_{it}-\bar{x})^2 - (x_{it}-\bar{x}_i)^2 - (x_{it}-\bar{x}_t)^2 + 2(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_t). \tag{35}$$

By lemma 3, A_{it} simplifies to

$$A_{it} = (x_{it}-\bar{x})(y_{it}-\bar{y}) - (x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i) - (x_{it}-\bar{x}_t)(y_{it}-\bar{y}_t) + 2(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_t), \tag{36}$$

or equivalently to

$$A_{it} = (x_{it} - \bar{x})(y_{it} - \bar{y}) - (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_t) + 2(x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i). \tag{37}$$

Next, by lemma 4, these expressions further simplify to

$$A_{it} = \left[(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_t) - (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t) \right]$$
$$- (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_t) + 2(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t)$$

$$= (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t), \tag{38}$$

or equivalently to $(x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i)$ by lemma 3. Applying lemma 5, B_{it} reduces as follows:

$$B_{it} = \left[(x_{it} - \bar{x}_i)^2 + (x_{it} - \bar{y}_t)^2 - (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t) \right]$$
$$- (x_{it} - \bar{x}_i)^2 - (x_{it} - \bar{x}_t)^2 + 2(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t)$$

$$= (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t). \tag{39}$$

Therefore, the two-way fixed effect estimator is given by

$$\beta_{TW} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_t)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t)}, \tag{40}$$

or equivalently as

$$\beta_{TW} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_t)(y_{it} - \bar{y}_i)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_t)}. \quad \blacksquare \tag{41}$$