Political Science Research and Methods Vol 6, No. 4, 829–835 October 2018

doi:10.1017/psrm.2017.44

Improving the Interpretation of Fixed Effects Regression Results*

JONATHAN MUMMOLO AND ERIK PETERSON

Fixed effects estimators are frequently used to limit selection bias. For example, it is well known that with panel data, fixed effects models eliminate time-invariant confounding, estimating an independent variable's effect using only within-unit variation. When researchers interpret the results of fixed effects models, they should therefore consider hypothetical changes in the independent variable (counterfactuals) that could plausibly occur within units to avoid overstating the substantive importance of the variable's effect. In this article, we replicate several recent studies which used fixed effects estimators to show how descriptions of the substantive significance of results can be improved by precisely characterizing the variation being studied and presenting plausible counterfactuals. We provide a checklist for the interpretation of fixed effects regression results to help avoid these interpretative pitfalls.

The fixed effects regression model is commonly used to reduce selection bias in the estimation of causal effects in observational data by eliminating large portions of variation thought to contain confounding factors. For example, when units in a panel data set are thought to differ systematically from one another in unobserved ways that affect the outcome of interest, unit fixed effects are often used since they eliminate all between-unit variation, producing an estimate of a variable's average effect within units over time (Allison 2009; Wooldridge 2010). Despite this feature, social scientists often fail to acknowledge the large reduction in variation imposed by fixed effects. This article discusses two important implications of this omission: imprecision in descriptions of the variation being studied and the use of implausible counterfactuals—discussions of the effect of shifts in an independent variable that are rarely or never observed within units—to characterize substantive effects.

By using fixed effects, researchers make not only a methodological choice but a substantive one (Bell and Jones 2015), narrowing the analysis to particular dimensions of the data, such as within-country (i.e., over time) variation. By prioritizing within-unit variation, researchers forgo the opportunity to explain between-unit variation since it is rarely the case that between-unit variation will yield plausible estimates of a causal effect. When using fixed effects, researchers should emphasize the variation that is being used in descriptions of their results. Identifying which units actually vary over time (in the case of unit fixed effects) is also crucial for gauging the generalizability of results, since units without variation provide no information during one-way unit-fixed effects estimation (Plümper and Troeger 2007).[1]

In addition, because the within-unit variation is always smaller (or at least, no larger) than the overall variation in the independent variable, researchers should use within-unit variation to motivate counterfactuals when discussing the substantive impact of a treatment. For example, if

* Jonathan Mummolo is an Assistant Professor of Politics and Public Affairs, Princeton University, Robertson Hall, Room 411, Princeton, NJ 08544 (jmummolo@princeton.edu). Erik Peterson is a Post-Doctoral Fellow at Dartmouth College, 026 Silsby Hall, 3 Tuck Mall, Hanover, NH, 03755 (erik.j.peterson@dartmouth.edu). The authors are grateful to Justin Grimmer, Dorothy Kronick, Jens Hainmueller, Brandon Stewart, Jonathan Wand and anonymous reviewers for helpful feedback on this project. To view supplementary material for this article, please visit https://doi.org/10.1017/psrm.2017.44

[1] The same is true for time periods which contain no variation in the treatment when time dummies are used.

only within-country variation in some treatment X is used to estimate a causal effect, researchers should avoid discussing the effect of changes in X that are larger than any changes observed within units in the data. Besides being unlikely to ever occur in the real world, such "extreme counterfactuals," which extrapolate to regions of sparse (or nonexistent) data rest on strong modeling assumptions, and can produce severely biased estimates of a variable's effect if those assumptions fail (King and Zeng 2006).

In what follows, we replicate several recently published articles in top social science journals that used fixed effects. Our intention is not to cast doubt on the general conclusions of these studies, but merely to demonstrate how adequately acknowledging the variance reduction imposed by fixed effects—and formulating plausible counterfactuals after estimation that account for this reduction—can improve descriptions of the substantive significance of results.[2] We conclude with a checklist to help researchers improve discussion of fixed effects results.

VARIANCE REDUCTION WITH FIXED EFFECTS

Consider the standard fixed effects dummy variable model:

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}, \tag{1}$$

in which an outcome Y and an independent variable (treatment) X are observed for each unit i (e.g., countries) over multiple time periods t (e.g., years), and a mutually exclusive intercept shift, α, is estimated for each unit i to capture the distinctive, time-invariant features of each unit. This results in an estimate of β that is purged of the influence of between-unit time-invariant confounders.[3]

While the overall variation in the independent variable may be large, the within-unit variation used to estimate β may be much smaller. The same is true when dummies for time are included (i.e., year dummies), in which case β is estimated using only within-year variation in the treatment. When multi-way fixed effects are employed (e.g., country *and* year dummies, as in a generalized difference-in-differences estimator),[4] the variance reduction is even more severe.

PUBLISHED EXAMPLES

To evaluate how often researchers acknowledge the variance reduction imposed by fixed effects when interpreting results, we conducted a literature review of empirical studies published in the *American Political Science Review* or the *American Journal of Political Science*, which used linear fixed effects estimators between 2008 and 2015.[5] Of the 54 studies we identified fitting these

[2] Our critiques also have no bearing on the statistical significance of any published results.

[3] In this scenario, between-unit variation is eliminated by demeaning the data within units (e.g., countries) during estimation (see e.g., Greene 2008, 198; Baltagi 2005; Cameron and Trivedi 2005, 726). That is, we could recover an identical estimate of β by omitting unit dummy variables and estimating the following:

$$\left(Y_{it} - \overline{Y_i}\right) = \beta\left(X_{it} - \overline{X_i}\right) + \left(\varepsilon_{it} - \overline{\varepsilon_i}\right). \tag{2}$$

Equation 2 makes plain that β is estimated solely based on deviations from each unit's average treatment value over time.

[4] Recent scholarship (Kim and Imai, n.d.) has highlighted assumptions that often get overlooked during fixed effects estimation which can bias causal estimates. In the present analysis, we take the research designs in the published work we examine as given to focus solely on the issues associated with variance reduction.

[5] Given our interest is in how researchers formulate counterfactuals, we excluded articles which featured a dichotomous treatment, since there is little discretion for counterfactuals in the case of a 0/1 treatment (i.e., the treatment either occurs or it does not). See Appendix for full details on the parameters of this literature review.

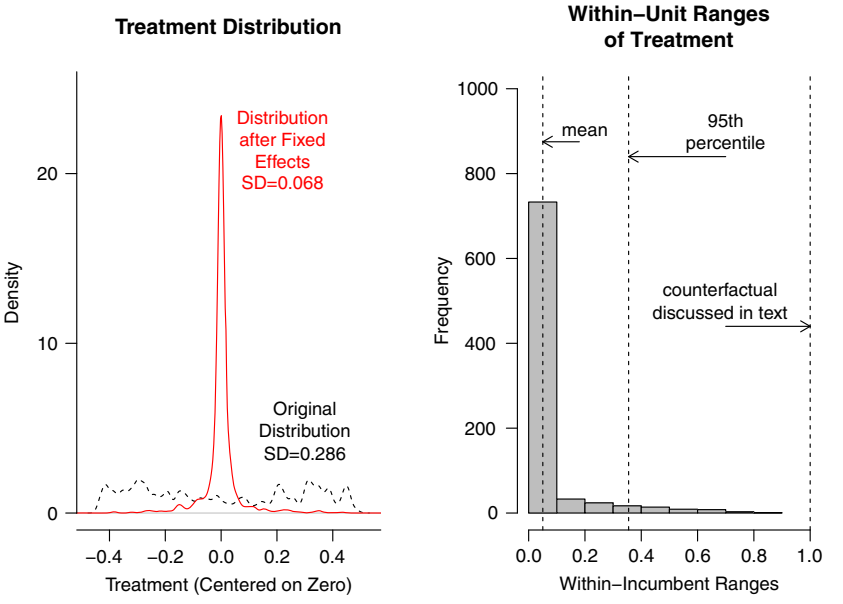

Fig. 1. The left panel displays the distribution of the treatment, media congruence, from Snyder and Strömberg (2010) before and after the incumbent and year fixed effects used in that study are applied
Note: Both distributions were centered on 0 before plotting to ease comparisons of their spread. The right panel shows the within-incumbent ranges of the treatment. While the right panel does not perfectly capture the relevant variation since it does not account for two-way fixed effects, we include it here to help motivate within-unit counterfactuals.

parameters, roughly 69 percent of studies explicitly described the variation being used during estimation (e.g., "… we include a full set of industry-year fixed effects … so that comparisons are within industry-year cells" (Earle and Gehlbach 2015, 713)). Only one study, Nepal, Bohara and Gawande (2011), explicitly quantified the size of typical shifts in levels of the treatment within units, allowing readers to judge the plausibility of the counterfactuals being discussed.[6]

Attention to this variance reduction has the potential to greatly influence how researchers characterize the substantive significance of results (i.e., whether a treatment imposes an effect large enough to induce a politically meaningful change in the outcome). For example, Figure 1 shows the distribution of the treatment in Snyder and Strömberg (2010) before and after fixed effects are applied.[7] This study estimates the effects of media market congruence (the alignment between print media markets and US congressional incumbents, a continuous variable takes values between 0 and 1), on both the public and political elites. As the left panel of the plot shows, the distribution of congruence is fairly uniform across congressional incumbents and years in the pooled sample. However, once year and incumbent fixed effects are applied, the amount of variation is drastically reduced (the standard deviation of the variable drops from 0.286 to 0.068, a reduction of 76 percent). The reason is that much of the variation in these data is between incumbents, all of which is discarded by the inclusion of incumbent dummies in several models in this study. The right panel of the figure shows the distribution of within-incumbent ranges of the treatment. As the histogram shows, the mean range observed within incumbents is roughly 0.05,

[6] The study supplied within-district standard deviations in table 1.
[7] Specifically, we residualize the treatment with respect to year and incumbent dummies (Lovell 1963).

TABLE 1 *Replication Results from Recent Studies Using Fixed Effects*

Studies	(1) $\hat{\beta}$	(2) $\widehat{\mathrm{SD}}(x)$	(3) $\widehat{\mathrm{SD}}(\tilde{x})$	(4) $\Delta\%$ (3) versus (2)	(5) $\frac{\text{Stated Counterfactual}}{\widehat{\mathrm{SD}}(\tilde{x})}$
Berrebi and Klor (2008)	0.004	1.052	0.839	−20.284	1.192
Gasper and Reeves (2011)	2.65	0.445	0.365	−17.89	2.736
Ichino and Nathan (2013)	0.316	0.219	0.082	−62.512	2.667
Scheve and Stasavage (2012)	23.017	0.159	0.112	−29.563	8.953
Snyder and Strömberg (2010)	0.279	0.286	0.068	−76.33	14.768

Note: Column 1 displays the estimated unstandardized regression coefficient of interest in each study. Columns 2 and 3 display the standard deviations of the key independent variable before and after residualizing with respect to fixed effects, respectively. Column 4 displays the percent difference between these two SDs. Column 5 displays the ratio of the counterfactual change in the treatment discussed in the article versus our preferred counterfactual, the revised standard deviation of X. See Appendix for details on the referenced published counterfactuals.

and is about 0.36 at the 95th percentile of the distribution of ranges. Roughly 44 percent of the time, the treatment does not vary at all within unit. Noting this fact would help to inform the generalizability of the result.

While Snyder and Strömberg (2010) acknowledge the use of a within-unit estimator,[8] the authors go on to consider a 0-to-1 shift in the treatment when discussing its impact on the probability that an individual recalls the name of their member of Congress. The paper states that, "… a change from the lowest to the highest values of congruence is associated with a 28 percent increase in the probability of correctly recalling a candidate's name. This is about as large as the effect of changing a respondent's education from grade school to some college" (Snyder and Strömberg 2010, 372). A 0-to-1 shift is beyond the observed range even in the original data (i.e., before the fixed effects adjustment). When a more realistic counterfactual is used—the standard deviation of the treatment after residualizing with respect to the incumbent and year dummies the authors employ—a more modest effect of a 1.9 percentage point increase in the probability of recalling a representative's name is recovered. A shift the size of the 95th percentile of within-incumbent ranges produces a 9.9 percentage point effect. Apart from rare cases in which very large shifts in congruence occur within units over time, the results suggest that the treatment exerts only a small effect on representative name recall.

In another example, Ichino and Nathan (2013) discuss the effect of changing the level of an ethnic group surrounding a polling station, stating that, "… a one standard deviation increase in the spatially weighted population share of Akans beyond a polling station (about 0.21) results in a predicted 6.9 percentage point greater NPP presidential vote share" (Ichino and Nathan 2013, 351–2). This counterfactual is based on the standard deviation of the overall distribution of the independent variable. However, the fixed effects estimator employed in this analysis only uses the within-parliamentary constituency distribution, and a 1-SD shift in this revised distribution produces a 2.6 percentage point effect, a decrease of 63 percent compared with the reported effect.

Table 1 displays results from replications of five published studies.[9] In each case, we generated the revised version of the independent variable after residualizing with respect to the included fixed effects (we label this residualized variable $\tilde{x}$), in order to assess the consequences

[8] Snyder and Strömberg (2010) states that the estimation strategy, "… focuses on variation within congressional districts," (357).

[9] These studies were chosen based on the availability of functional replication files and also to represent different substantive fields in political science.

of variance reduction. As the results show, considering a 1 SD shift in $\tilde{x}$, as opposed to a 1-SD shift in the original distribution, reduces estimated treatment effects by as much as 76 percent (column 4). Column 5 compares the counterfactual shifts in X discussed in each article to $\widehat{\text{SD}}(\tilde{x})$. In the case of Berrebi and Klor (2008), the discussed counterfactual is quite plausible (i.e., the ratio is close to 1), while the counterfactuals discussed in Gasper and Reeves (2011) and Ichino and Nathan (2013) are close to three times the size of the revised standard deviation. The changes in X discussed in Scheve and Stasavage (2012) and Snyder and Strömberg (2010) —both shifts from the minimum to the maximum of the independent variable's range—are roughly 9 and 15 times the size of a typical shift in the revised distribution in X, respectively, indicating that changes this large are rarely observed in the data.[10]

Since most political science studies using fixed effects seek to characterize on-average effects rather than effects within any particular unit, we consider the within-unit standard deviation a sensible counterfactual to consider since it represents the average amount the independent variable deviates from the mean after fixed effects are employed. Still, some might wish to convey treatment effects in the common scenario of a one-unit shift. This may be especially useful if the treatment being studied is dichotomous or a count variable that takes only integer values, in which case unit shifts arguably make more substantive sense to consider (this is the case with Berrebi and Klor (2008) and Gasper and Reeves (2011), for example). Here too, variance reduction should be noted, since in some cases even a one-unit shift can represent a change that is several times the size of a typical shift within units. Researchers are of course free to discuss such counterfactuals, but we recommend discussing the magnitude of the hypothetical change in X relative to the variation being used during estimation (i.e., the ratio displayed in column 5) so readers have a sense of whether the hypothetical change in X is rare.

CONCLUSION

Because it reduces concerns that omitted variables drive any associations between dependent and independent variables, many researchers use linear fixed effects regression. This estimator reduces the variance in the independent variable and narrows the scope of a study to a subset of the overall variation in the data set. While some researchers readily acknowledge these features, many studies can benefit from more specificity in the description of the variation being studied and the consideration of counterfactuals that are plausible given the variation being used for estimation (e.g., a typical within-unit shift in X).

CHECKLIST FOR INTERPRETING FIXED EFFECTS RESULTS

When researchers employ linear fixed effects models, we recommend[11] the following method for evaluating results after estimation:

1. Isolate relevant variation in the treatment: Residualize the key independent variable with respect to the fixed effects being employed (Lovell 1963). That is, regress the treatment on the dummy variables which comprise the fixed effects and store the resulting vector of residuals. This vector represents the variation in X that is used to estimate the coefficient of interest in the fixed effects model.[12]

[10] Note that several articles discussed more than one counterfactual scenario. We highlight one from each article here for clarity of discussion.

[11] See Appendix for R code to apply these recommendations.

[12] This approach is often preferable to manually demeaning the data since it easily accommodates multi-way fixed effects.

2. Identify a plausible counterfactual shift in X given the data: Generate a histogram (as in Figure 1) of the within-unit ranges of the treatment to get a sense of the relevant shifts in X that occur in the data. Compute the standard deviation of the transformed (residualized) independent variable, which can be thought of as a typical shift in the portion of the independent variable that is being used during the fixed effects estimation. Multiply the estimated coefficient of interest by the revised standard deviation of the independent variable to assess substantive importance. Note for readers what share of observations do not exhibit any variation within units to help characterize the generalizability of the result. Alternatively, if describing the effect of a one-unit shift, or any other quantity, note the ratio of this shift in X to the within-unit standard deviation, as well as its location on the recommended histogram, to gauge how typically a shift of this size occurs within units.
3. Clarify the variation being studied: In describing the scope of the research and in discussing results after fixed effects estimation, researchers should clarify which variation is being used to estimate the coefficient of interest. For example, if only within-unit variation is being used, then phrases like "as X changes within countries over time, Y changes …" should be used when describing treatment effects.
4. Consider the outcome scale: Consider characterizing this new effect in terms of both the outcome's units and in terms of standard deviations of the original and transformed outcome (i.e., the outcome residualized with respect to fixed effects). The substance of particular studies should be used to guide which outcome scale is most relevant.

REFERENCES

Allison, Paul D. 2009. *Fixed Effects Regression Models*. London: Sage.

Baltagi, Badi. 2005. *Econometric Analysis of Panel Data*. New York: Wiley & Sons.

Bell, Andrew, and Kelvyn Jones. 2015. 'Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data'. *Political Science Research and Methods* 3(1): 133–53.

Berrebi, Claude, and Esteban F. Klor. 2008. 'Are Voters Sensitive to Terrorism? Direct Evidence from the Israeli Electorate'. *American Political Science Review* 102(3):279–301.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics*. Cambridge: Cambridge University Press.

Earle, John S., and Scott Gehlbach. 2015. 'The Productivity Consequences of Political Turnover: Firm? Level Evidence from Ukraine's Orange Revolution'. *American Journal of Political Science* 59(3): 708–23.

Gasper, John T., and Andrew Reeves. 2011. 'Make it Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters'. *American Journal of Political Science* 55(92):340–55.

Greene, William H. 2008. *Econometric Analysis* 6th ed. Upper Saddle River, NJ: Pearson Education, Inc.

Ichino, Nahomi, and Noah L. Nathan. 2013. 'Crossing the Line: Local Ethnic Geography and Voting in Ghana'. *American Political Science Review* 107(2):344–61.

Kim, In Song, and Kosuke Imai. n.d. 'When Should We Use Linear Fixed Effects Regression Models for Causal Inference With Longitudinal Data?'. Working Paper, ifundefinedselectfont https://imai.princeton.edu/research/FEmatch.html, accessed 9 January 2017.

King, Gary, and Langche Zeng. 2006. 'The Dangers of Extreme Counterfactuals'. *Political Analysis* 14:131–59.

Plümper, Thomas, and Vera E. Troeger. 2007. 'Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses With Unit Fixed Effects'. *Political Analysis.* 15(2):124–39.

Lovell, Michael C. 1963. 'Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis'. *Journal of the American Statistical Association* 58(204):993–1010.

Nepal, Mani, Alok K. Bohara, and Kishore Gawande. 2011. 'More Inequality, More Killings: The Maoist Insurgency in Nepal'. *American Journal of Political Science* 55(4):886–906.

Scheve, Kenneth, and David Stasavage. 2012. 'Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation'. *American Political Science Review* 106(1):81–102.

Snyder, James M., and David Strömberg. 2010. 'Press Coverage and Political Accountability'. *Journal of Political Economy* 118(2):355–408.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, London: MIT Press.