Notes on "Using and Interpreting Fixed Effects Models"

Paper: Breuer, Matthias, and Ed Dehaan, "Using and Interpreting Fixed Effects Models," *Journal of Accounting Research*, 62 (2024), 1183–1226. https://doi.org/10.1111/1475-679X.12559

What FE do

- The constant is a special version of FE, one with only one group.
- In essence, although the single constant recenters the entire sample around the origin, FE separately recenter each group's observations around the origin. Therefore, variation resulting from differences in means across group is implicitly eliminated. This is why researchers often say that FE "restrict analyses to within-group variation in X."
- The demeaning results in more narrow distributions on variables, which is especially important when using the distribution properties (e.g., standard deviation) of X to interpret economic magnitudes.
- FE groups with multiple observations but no within-group variation in the X of interest (so-called "no-variation" groups) affect standard errors but may or may not affect $\hat{\beta}$.
    - No-variation groups have no effect on $\hat{\beta}$ in the simple case of one X and one set of FE.
    - For the more common case of multiple X's or FE groupings, no-variation groups can contribute to estimating $\hat{\beta}$ as long as they have variation in at least one regressor. They do so by helping to estimate the coefficients on control variables and other sets of FE, which in turn influence the estimate of $\hat{\beta}$. In this regard, no-variation groups act much like control groups in matching designs or simple DD models.
- The most influential observations in FE regressions tend to be those that have the largest within-group variation in Y or X, as opposed to those that are outliers in the raw data.
- FE can help alleviate omitted-variable bias as they control for all factors that vary only at the group level. This is true for both observed and unobserved factors.
- However, FE can also affect standard errors and test power, and often for the worse. The impact of FE on test power depends on the correlation structure between Y ($y_{i,t}$), the X ($x_{i,t}$) of interest, and other regressors ($z_{i,t}$), which is given by

  $$\hat{\sigma}_{\hat{\beta}} = \sqrt{\left(\frac{1}{N-p-1}\right)\left(\frac{\sum \hat{\epsilon}_{i,t}^2}{\sum (x_{i,t}-\bar{x})^2(1-\hat{R}_x^2)}\right)},$$

  where $\hat{\epsilon}_{i,t}^2$ is the variance of the regression residual and $\hat{R}_x^2$ is the fraction of variation in $x_{i,t}$ that is explained by all other controls and FE (this can be proved by FWL theorem). Consider the following DGP:

  $$y_{i,t} = \alpha + \beta x_{i,t} + \delta z_g + \varepsilon_{i,t},$$

  where z_g is unobservable, so we instead control for it using FE for group g:

  $$y_{i,t} = \alpha_g + \beta x_{i,t} + \varepsilon_{i,t}.$$

    - *Case 1:* z_g is not a determinant of $y_{i,t}$ (i.e., $\delta = 0$) and is not correlated with $x_{i,t}$. Including FE will decrease power by increasing p and $\hat{R}_x^2$.
    - *Case 2:* z_g is not a determinant of $y_{i,t}$ but it is correlated with $x_{i,t}$. Again, including FE will decrease power by increasing p and $\hat{R}_x^2$. Additionally, as now z_g does explain $x_{i,t}$, the increase in

$\hat{R}_x^2$ would be larger.

  - *Case 3:* z_g is a determinant of $y_{i,t}$ but is not correlated with $x_{i,t}$. Now, including FE could decrease power by increasing p and $\hat{R}_x^2$ but also could increase power by reducing $\hat{\epsilon}_{i,t}^2$. Typically, FE can be used to increase test power, especially if the number of observations is large relative to the number of group.
  - *Case 4:* z_g is both a determinant of $y_{i,t}$ and correlated with $x_{i,t}$. The test power could be increased or decreased, depending on all the power-increasing and -decreasing effects noted in the above prior three cases. However, in this case, including FE effectively eliminate omitted-variable bias, and, typically, Type 2 errors are less concerning than Type 1 errors. Therefore, the potential decrease in test power should not be used as an excuse to neglect FE.
- Spillover effects arise if a treatment affects both the treated and control groups. Spillovers bias the treatment-effect estimate upwards (downwards) if they affect control group in the opposite (same) direction as the treatment effect on the treated group. It should be noted that FE can aggravate spillover bias by restricting analyses to more comparable treatment and control firms. Although we generally prefer more comparable control firms, in the case of local spillovers, restricting to local control group reduces concerns about bias due to confounding differences but potentially adds bias due to an indirect impact of the treatment on the control group.
- Interacted FE play a key role in cases when the X of interest is a differential slope. When employing interacted FE, we also face trade-offs: interacting fine FE with X can consume many degree of freedom and remove too many variation in X for the estimation, while interacting coarse FE with X may not control confounding heterogeneity. An alternative remedy for impracticable high-frequency FE interactions is to attempt to impose more structure on the coefficient distribution. Bayesian hierarchical modeling, for example, can account for rich heterogeneity while reducing concerns about limited within-group data.
- A more subtle but likely pervasive problem is that estimates of parameters of interest in linear models that include subject FE can be strongly biased in the presence of dynamics or feedback effects. The bias arises because in-sample means combine and reflect the (recent) dynamics in firms' variables, which are driven by observable but also unobservable factors.
- In some cases, FE can even induce dependence. For example, including FE deducts the group-level mean of the error term from the already independent error. In the case of small group sizes, deducting the group's mean error links all group-level errors with each other and thus introduces correlation between within-group errors. Therefore, FE groups should ideally be "nested" within the clusters to help eliminate any effects of FE-induced dependence.
- Maximum-likelihood estimator (MLE) is typically biased and inconsistent when using firm FE. This "incidental parameters problem" is particularly severe if the time dimension is small and fixed. The use of FE in nonlinear models is an active area of research among econometricians, and recent advances are starting to make the use of FE in nonlinear models more practicable. For example, Correia, Guimarães, and Zylkin (2020) develop an implementable Poisson model with high-frequency FE (i.e., `ppmlhdfe` Stata package).

Guidance on diagnostics and reporting

1. Papers should clearly articulate what variation is needed (or "wanted") to test their research questions and how their research design choices, including their FE structure, help isolate this variation.
2. Singletons have just one observation within a FE group, do not affect coefficient estimates, and can bias standard errors, so should be dropped before running a FE regression. Singletons must be identified iteratively in regressions with multiple FE groupings, meaning that one first drops singletons for the first

FE grouping, and then drops singletons for the second FE grouping, and then repeats until no singletons remain. FE software such as `reghdfe` Stata package can iteratively drop singletons.

3. So-called "no-variation FE groups" have multiple observations but no variation in the X of interest, so only affect $\hat{\beta}$ by acting as control groups to help estimate relations between Y and other determinants. Much like control groups in matching and standard DD designs, no-variation FE groups are only helpful if they share a similar data generating process as the groups that do have variation in X. If they do, their inclusion in the sample improves test power and specificity. To test this, one can regress Y on the controls and FE excluding the X of interest, separately for the no-variation groups and other groups. Significant differences in coefficient estimates can indicate that the two subsamples have different data generating processes. One can also run regressions with and without no-variation groups and compare their $\hat{\beta}$'s.
4. FE groups with no variation in Y do directly contribute to estimating $\hat{\beta}$ in linear models. Whether to drop groups without variation in Y often depends on whether it is interesting to examine only groups with variation. For example, in a study of the determinants of R&D expense, it plausibly makes sense to drop banks (which rarely have material R&D), but it likely would not make sense to drop consumer product firms (for which immaterial R&D is likely a strategic choice).
5. Because FE implicitly transform variables into within-FE variation, standard descriptive statistics should be supplemented with descriptive statistics for each variable residualized to the FE. It is recommended reporting both the raw standard deviation of each variable and the within-FE standard deviation of each variable.
6. One way to be transparent about a regression's explanatory power coming from FE versus other regressors is to supplement the reported regression R^2 with the "within-R^2," which is the portion of the R^2 generated by the explanatory variables other than FE.
7. Because FE regression estimates are based on the within-FE variation in X, the raw variation in X is no longer a meaningful unit of change for characterizing magnitudes. The economic magnitude of $\hat{\beta}$ should instead be characterized using the within-FE variation in X, which is never larger than the pooled sample variation, and is often much smaller. For example, one can instead use the standard deviation of the residualized X. Whether or not to use the within-FE variation of Y depends on whether the within-FE variation in Y is an economically interesting benchmark (e.g., whether the authors are more interested in the policy effects on the whole economy or on a single firm/industry).
8. Because FE use only within-group variation to estimate effects, empirical results cannot directly speak to whether similar cause-and-effect relations exist across FE groups, which could undermine the generalizability of the research. As always, transparently describing the context of the study's empirical results and potential generalizability to other settings will help readers understand the need for further research.
9. Finer FE (e.g., year-month instead of year FE) can increase credibility by further reducing concerns about potentially unwanted variation, but can also exacerbate all the issues about power, contributing observations, and generalizability. Accordingly, finer FE are not always preferred for main specifications, but can nevertheless be informative as a robustness test.
10. Heterogeneous effects can bias DD models estimated via two-way FE specifications, so should be carefully evaluated when using such models.
11. Spillovers from treated to control firms can introduce bias that can be aggravated through FE. It is recommended gauging potential spillover bias in FE models, especially if theory predicts that spillovers likely occur. One way to do so is to examine whether a treatment effect estimate is sensitive to varying levels of FE. If a treatment effect changes systematically when using finer FE, the effect may be driven by positive spillovers to local control firms. Alternatively, one can try to explicitly control for spillovers to alleviate concerns about their confounding influence.

12. Subjects' characteristics are unlikely to be completely fixed over time. When subjects' characteristics evolve over time, subject FE become less effective as panel periods lengthen. In such cases, finer FE (e.g., subject-decade FEs), first-difference estimators, or rolling estimators may be more effective.
13. When the strict exogeneity assumption is violated, for example, when assignment of the treatment of interest depends on past or future outcomes, Imai and Kim (2019) recommend that researchers try to measure and control for time-varying confounders rather than use subject FE to adjust for unobserved time-invariant confounders. In a similar vein, Plümper and Troeger (2019) recommend that, in case of dynamics and feedback effects, researchers should try to explicitly model and account for the dynamic relationships (e.g., via dynamic panel estimators) instead of rely on subject FE as a default solution.
14. The most influential observations in FE regressions are likely to be those with the most extreme within-FE variation. Truncating or winsorizing raw variables will potentially fail to mitigate observations with high leverage in FE regressions, in which case, it is useful to truncate or winsorize the residualized variables instead.