

# Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix

(or why `reghdfe` will drop singletons from now on...)

Singleton groups (groups with only one observations) are increasingly common in regressions with many fixed effects, such as regressions with worker/firm/job title fixed effects that were previously unfeasible due to computational limitations (see e.g. [Carneiro \*et al\*, 2012](#)). Even though some users may drop them, most are not aware that with more than one fixed effect, singletons need to be dropped iteratively, as e.g. dropping a singleton individual may in turn turn the firm he worked into a singleton, and so on.

Now, what are the effects of keeping singleton groups in regressions where fixed effects are nested within groups/clusters?

1. Coefficients and unadjusted variance estimators don't change.
2. Robust and cluster-robust estimators of the variance decrease due to finite-sample adjustments converging to 1. Note that the asymptotic part of the robust variance estimator (the usual “bread and meat” of the sandwich estimator) remains unaffected, so this is as problem only as long as finite-sample adjustments are relevant (which is the case more often than not). Therefore, **standard errors will be underestimated, and statistical significance will be overstated.**
3. The reported number of clusters will be overstated, potentially misleading users into believing that there are enough clusters to make accurate asymptotic inference (e.g. above 50 clusters).
4. Estimations will run slower, as there is a larger number of ancillary parameters to estimate.

## Finite-Sample Adjustments

Given an estimate of the asymptotic variance of the regression estimates ( $V$ ),  $M$  clusters,  $N$  observations,  $M$  fixed effects (one for each cluster group), and  $K$  regressors of interest, then the finite-sample correction that multiplies  $V$  is:

$$q = M/(M - 1) \times (N - 1)/(N - K)$$

If we add  $M_S$  singleton groups, the above becomes

$$q^* = (M + M_S)/(M + M_S - 1) \times (N + M_S - 1)/(N + M_S - K)$$

Since  $q^*$  converges to 1 as  $M_S$  grows, adding enough singleton observations is enough to deem the standard finite-sample corrections moot.

## Toy Example

As an extreme but illustrative example of the first problem, consider the following regressions using the sample Stata dataset:

```
* Create toy data based on auto.dta
sysuse auto, clear
gen id = _n
replace id = id-1 if _n<8 & mod(id,2)==0
bys id: gen t = _n
xtset id t
bys id: gen is_singleton = (_N==1)
tab is_singleton

* Fixed-effect regression
xtreg price weight length, fe vce(cluster id)
xtreg price weight length, fe vce(cluster id) dfadj
drop if is_singleton
xtreg price weight length, fe vce(cluster id)
xtreg price weight length, fe vce(cluster id) dfadj
```

The first regression reports a P-Value of 0.007 for the *weight* regressor, while the subsequent regressions (those dropping singletons and/or subtracting the fixed effects from the degrees-of-freedom) report much higher P-Values ranging from 0.212 to 0.796.

## Extensions

The inclusion of singletons is part of a larger class of problems. For instance, consider the following scenario:

### Zipcode-level regression of State-level data

Assume all variables are specified at the state level, but we run them at a zipcode level with  $Z$  zipcodes per state. Then, the finite-sample correction becomes:

$$q^* = (M \times Z) / (M \times Z - 1) \times (N \times Z - 1) / (N \times Z - K)$$

Which again converges to 1 and is rendered useless as  $Z$  increases.

A milder but more common version of this extreme scenario occurs whenever there is little variation between zipcodes or counties of the same state, and the regression is clustered by state and contains either state or zipcode fixed effects.

## Solutions

The singleton problem can be easily dealt with by either removing singleton groups, or subtracting them from the number of clusters  $M$  and observations  $N$ .

The broader problem is an open question.

## Summary

Clustered standard errors do not include the number of fixed effects when computing the finite-sample adjustments of the variance estimates, as long as the fixed effects are nested within clusters. This adjustment implies that usually irrelevant specification details, such as adding singleton groups or running regressions on less coarser units, will affect variance estimates and potentially understate the significance of fixed effect models.

## References and Previous Discussions

[David Matsa's post about dealing with fixed effects](#)

"XTREG's approach of not adjusting the degrees of freedom is appropriate when the fixed effects swept away by the within-group transformation are nested within clusters (meaning all the observations for any given group are in the same cluster), as is commonly the case (e.g., firm fixed effects are nested within firm, industry, or state clusters). See Wooldridge (2010, Chapter 20)."

[A. Colin Cameron and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference", Journal of Human Resources, forthcoming, Spring 2015](#)

IIC eq. 12; "Finite-sample modifications of (11) are typically used, to reduce downwards bias in  $V_{clu}[\beta]$  due to finite  $G$ ... In general,  $c \sim G/(G-1)$ ", though see Section IIIB for an important exception when fixed effects are directly estimated"

IIIB: "It is important to note that while LSDV and within estimation lead to identical estimates of  $\beta$ , they can yield different standard errors due to different finite sample degrees-of-freedom correction.

It is well known that if default standard errors are used, i.e. it is assumed that  $u_{ig}$  in (17) is i.i.d., then one can safely use standard errors after LSDV estimation as it correctly views the number of parameters as  $G + K$  rather than  $K$ . If instead the within estimator is used, however, manual OLS estimation

of (18) will mistakenly view the number of parameters to equal  $K$  rather than  $G + K$ . (Built-in panel estimation commands for the within estimator, i.e. a fixed effects command, should remain okay to use, since they should be programmed to use  $G + K$  in calculating the standard errors.)

It is not well known that if cluster-robust standard errors are used, and cluster sizes are small, then inference should be based on the within estimator standard errors... Within estimation sets  $c = G / (G-1) \times (N-1) / (N-K+1)$  since there are only  $(K-1)$  regressors--the within model is estimated without an intercept. LSDV estimation uses  $c = G / (G-1) \times (N-1) / (N-G-K+1)$  since the  $G$  cluster dummies are also included as regressors... Within estimation leads to the correct finite-sample correction"

[Mark Schaffer's Statalist post](#)

"In this panel data context, a singleton is a group in which there is only one observation. Since singletons have zero within-group information, the within (demeaning) transformation will zap them.

Stata's official commands that do linear fixed effects estimation (xtreg, xtivreg, areg) do not adjust the number of observations for the singletons. Explicitly excluding singletons can therefore affect the SEs but will leave the coefficients unchanged

... it is correct to treat singletons as non-observations, no different from observations that are lost because of missing values ..."

[James G. MacKinnon & Halbert White, "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," Journal of Econometrics 29 \(1985\)](#)

(discussion of alternative finite-sample corrections)

[James G. MacKinnon, 2012. "Thirty Years of Heteroskedasticity-Robust Inference," Working Papers 1268, Queen's University, Department of Economics.](#)

(literature review, including an extensive discussion on finite-sample corrections)

[Gormley, Todd A. and Matsa, David A., Common Errors: How to \(and Not to\) Control for Unobserved Heterogeneity \(August 3, 2013\). Review of Financial Studies, 2014, 27\(2\), 617-61](#)

"Typically, the degrees of freedom is adjusted downward (i.e., the estimated standard errors are increased) to account for the number of fixed effects

removed in the within transformation. However, when estimating cluster-robust standard errors (which allows for heteroscedasticity and within-group correlations), this adjustment is not required as long as the fixed effects swept away by the within-group transformation are nested within clusters (meaning all the observations for any given group are in the same cluster), as is commonly the case (e.g., firm fixed effects are nested within firm, industry, or state clusters)."