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

Sergio Correia
Duke University

Updated November 2015

Abstract

Maintaining singleton groups in linear regressions where fixed effects are nested within clusters can overstate statistical significance and lead to incorrect inference. Due to this problem, the reghdfe package now automatically drops singletons. However, a broader class of problems related to nested fixed effects and finite-sample adjustments remains.

1 Summary

Singleton groups—groups with only one observation—are common in regressions with multiple levels of fixed effects, such as in the work of Carneiro, Guimarães, and Portugal (2012), who estimate linear regressions that feature fixed effects for each worker, firm, and job title. For instance, in their employer-employee matched dataset, 27% of all fixed effects were singletons (1.5 million singleton groups out of a total of 5.6 million fixed effects). The consequences of having such a large occurrence of singleton group has not been studied in practice, as models with many levels of fixed effects were not feasible until recently due to the lack of practical estimators (Abowd, Creecy, and Kramarz 2002; Guimarães and Portugal 2010; Gaure 2013; Correia 2015).

Keeping singleton groups in such regressions is not only computationally inefficient, but overstates the statistical significance of the regression coefficients and might lead to incorrect

*Thanks to Todd Gormley, Paulo Guimarães and Mark E. Shaffer for their feedback and helpful discussions.

inference. This has been informally recognized by some researchers, who address the problem by dropping singletons from the regression sample. However, an often overlooked fact is that with more than one fixed effect, singletons need to be dropped iteratively. For instance, in a matched CEO-firm regression, dropping a singleton CEO may reduce the observation count of the firm he managed from two observations to one. This turns the firm into a singleton group, which is then dropped, and so on.

In this article, we discuss the effects of keeping singleton groups in the regression sample. We focus on the case where standard errors are cluster-robust (Cameron, Gelbach, and Miller 2011; Petersen 2009) and the fixed effects are nested within clusters[1]. There are several effects of keeping singletons in such a case:

1. Coefficient estimates and conventional variance estimates remain unchanged.
2. Cluster-robust variance estimates will decrease due to the finite-sample adjustment q 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 surprisingly the case in many situations). 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. Estimation will be slower, as there is a larger number of ancillary parameters to estimate.

A possible response to the second argument might be that the finite sample adjustment of the variance q does not matter with very large datasets, as it is will be very close to one. This is false when dealing with many fixed effects. For instance, if a sets of fixed effects is not nested within clusters, the number of estimated parameters would be a significant fraction of the number of observations, so q would never converge to one.

2 Finite-Sample Adjustments

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

[1] A set of fixed effects is nested within clusters if (i) the fixed effect identifier is the same as the cluster identifier, or more generally, (ii) if no fixed effect category spans more than one cluster category. Examples include regressions that cluster at the state level and include county fixed effects, and regressions that cluster by industry, have firm fixed effects, and no firm changes industry.

$$q = \left(\frac{M}{M-1}\right)\left(\frac{N-1}{N-K}\right) \tag{1}$$

If we add M_S singleton groups, the above becomes

$$q^* = \left(\frac{M+M_S}{M+M_S-1}\right)\left(\frac{N+M_S-1}{N+M_S-K}\right) \tag{2}$$

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

2.1 Case When Not All Fixed Effects are Nested Within Clusters

If not all fixed effects are nested within clusters, the effect on q of keeping singleton groups is less clear. For instance, suppose that there are $G + G_S$ fixed effects not nested within clusters, of which G_S are singletons. These need to be considered—together with K—as estimated parameters, so our previous equation becomes:

$$q^* = \left(\frac{M+M_S}{M+M_S-1}\right)\left(\frac{N+M_S-1}{N+M_S-K-G}\right) \tag{3}$$

In this case, dropping singleton groups not only removes M_S from the equation—increasing q—but removes G_S—decreasing q. Therefore, how q and the coefficient standard errors change depends on the specific of each regression.

3 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.

4 Extensions

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

4.1 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:

$$q^* = (M \times Z)/(M \times Z - 1) \times (N \times Z - 1)/(N \times Z - K) \tag{4}$$

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.

5 Solutions

The singleton problem can be easily dealt with by either removing singleton groups, or keeping them while excluding their count from the number of clusters M and observations N.

Solving the more general problem is an open question.

6 Conclusion

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 overstate the statistical significance of fixed effect models.

References

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data*. Longitudinal Employer-Household Dynamics Technical Papers 2002-06. Center for Economic Studies, U.S. Census Bureau. http://ideas.repec.org/p/cen/tpaper/2002-06.html.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics* 29 (2): 238–49. doi:http://doi.org/10.1198/jbes.2010.07136.

Carneiro, Anabela, Paulo Guimarães, and Pedro Portugal. 2012. "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity." *American Economic Journal: Macroeconomics* 4 (2): 133–52. doi:http://doi.org/doi:10.1257/mac.4.2.133.

Correia, Sergio. 2015. *Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator*.

Gaure, Simen. 2013. "OLS with Multiple High Dimensional Category Variables." *Computational Statistics & Data Analysis* 66 (0): 8–18. doi:http://doi.org/http://dx.doi.org/10.1016/j.csda.2013.03.024.

Guimarães, Paulo, and Pedro Portugal. 2010. "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects." *Stata Journal* 10 (4). College Station, TX: Stata Press: 628–649(22). http://www.stata-journal.com/article.html?article=st0212.

Petersen, Mitchell A. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies* 22 (1): 435–80. doi:http://doi.org/10.1093/rfs/hhn053.

Excerpts from Related Discussions

David Matsa's post about dealing with fixed effects link

> "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 link

> IIC eq. 12; "Finite-sample modifications of (11) are typically used, to reduce downwards bias in $Vclu[\beta]$ due to finite G… In general, c ~= 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 β, 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 $u_i g$ 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 link

> 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) link

Contains a 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 link

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 link

> 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).