The Stata Journal (2023)
23, Number 1, pp. 276–280 DOI: 10.1177/1536867X231162021

Stata tip 149: Weighted estimation of fixed-effects and first-differences models

John Gardner
Department of Economics
University of Mississippi
University, MS
jrgardne@olemiss.edu

Applied econometricians frequently use weighted regressions to improve the precision of fitted panel-data models. For example, suppose that the outcome y_{igt} for individual i in group g at time t is

$$y_{igt} = \mathbf{x}'_{gt}\boldsymbol{\beta} + c_g + \varepsilon_{igt}$$

A group-average version of this model is

$$y_{gt} = \mathbf{x}'_{gt}\boldsymbol{\beta} + c_g + \varepsilon_{gt} \tag{1}$$

where $y_{gt} = \sum_i y_{igt}/n_{gt}$, n_{gt} is the number of observations in group g at time t and ε_{gt} is defined similarly. The group-average model might be relevant because individual-level data are not available (for example, because of confidentiality concerns) or for computational reasons.

In such cases, it is common practice to weight the model by n_{gt}.[1] The justification for this practice is that, if the original ε_{igt} are homoskedastic and serially uncorrelated with variance σ^2, then $V(\varepsilon_{gt}) = \sigma^2/n_{gt}$, and the Gauss–Markov theorem applies to the weighted model, which has homoskedastic errors.[2]

This tip clarifies estimation of weighted panel-data models in Stata in two ways. First, it extends the well-known deviation-from-means interpretation of fixed-effects models and the equivalence between fixed-effects and first-differences models with two time periods to the case of weighted estimation. Second, it highlights several ways to fit weighted fixed-effects (WFE) models in Stata. Of course, the tip also applies to models that are weighted for reasons other than heteroskedasticity arising from group averaging.

1. This can be accomplished in Stata using analytic weights, which are "inversely proportional to the variance of an observation" (StataCorp 2021). When you insert the analytic weight into the calculation formula, "you are treating each observation as one or more real observations" (StataCorp 2021). In the regression context, least-squares estimation weighted by n_{gt} is equivalent to least-squares estimation of a transformed model in which each variable for each observation is multiplied by $\sqrt{n_{gt}}$.
2. This is not necessarily a good idea. Solon, Haider, and Wooldridge (2015) show that, if the ε_{igt} are autocorrelated (for example, because of clustering), weighting may increase the estimated standard errors.

To illustrate weighted estimation of models such as (1) in Stata, I begin by generating some heteroskedastic panel data:

```
. set seed 57474
. set obs 100
Number of observations (_N) was 0, now 100.
. generate c = rnormal(1, 2)                          // fixed effects
. generate g = _n                                     // groups
. forvalues t=1/2 {
  2.      generate n`t' = max(1,ceil(uniform()*100)) // group sample sizes
  3.      generate x`t' = rnormal(c) + rnormal()     // x_gt correlated with c_g
  4.      generate e`t' = rnormal(0, 5/sqrt(n`t'))   // heteroskedastic errors
  5.      generate y`t' = 5 + 2*x`t'+ c +e`t'        // y_gt
  6. }
. reshape long x e y n, i(g) j(t)
(j = 1 2)
Data                               Wide   ->   Long
-----------------------------------------------------------------------------
Number of observations              100   ->   200
Number of variables                  10   ->   7
j variable (2 values)                     ->   t
xij variables:
                                 x1 x2   ->   x
                                 e1 e2   ->   e
                                 y1 y2   ->   y
                                 n1 n2   ->   n
-----------------------------------------------------------------------------
```

The simplest route to weighted estimation is via the `regress` command, with group dummy variables and analytic weights equal to the group-time sample sizes:

```
. regress y x i.g [aw=n]
(sum of wgt is 10,289)
      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(100, 99)      =    156.02
       Model |  7952.52137        100  79.5252137   Prob > F        =    0.0000
    Residual |  50.4613647         99  .509710754   R-squared       =    0.9937
-------------+----------------------------------   Adj R-squared   =    0.9873
       Total |  8002.98274        199  40.2159937   Root MSE        =    .71394

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    2.001036   .0557417    35.90   0.000     1.890432     2.11164
             |
           g |
           2 |   -1.184803   .7453728    -1.59   0.115    -2.663785    .2941781
           3 |    1.455865   .5802503     2.51   0.014     .3045227    2.607208
  (output omitted)
------------------------------------------------------------------------------
```

In this case, the weighted estimate compares favorably with the unweighted point estimate of 1.97 with standard error 0.075 (not shown).

In the unweighted case, the fixed-effects dummy-variable estimator has a deviation-from-means interpretation: it can be obtained by a *within* regression that replaces y_{gt}

and $\mathbf{x}_{gt}$ with the deviations of those variables from their group-specific means (eliminating the c_g along the way). A natural question is whether WFE estimation has a similar interpretation.

The Frisch–Waugh–Lovell theorem (see, for example, Greene [2018, theorem 3.2]) implies that WFE estimates can be obtained from a weighted regression that replaces y_{gt} and $\mathbf{x}_{gt}$ with the residuals from weighted regressions of those variables on a full set of group dummies. To connect this to a deviation from means interpretation, note that, because the group dummies are mutually orthogonal, the coefficient on the dummy d_{jgt} for group j from a weighted regression of y_{gt} on a full set of group dummies (and no overall constant) can be obtained from a weighted regression of y_{gt} on d_{jgt} alone as

$$\widehat{\lambda}_j = \frac{\sum_{g,t} n_{gt} y_{gt} d_{jgt}}{\sum_{g,t} n_{gt} d_{jgt}^2} = \frac{\sum_{g,t} n_{gt} y_{gt} d_{jgt}}{\sum_{g,t} n_{gt} d_{jgt}} = \frac{\sum_t n_{jt} y_{jt}}{\sum_t n_{jt}}$$

and similarly for $\mathbf{x}_{gt}$. Consequently, weighted dummy-variable estimation of (1) is equivalent to least-squares estimation of the weighted model

$$\begin{aligned}\sqrt{n_{gt}}\left(y_{gt} - \frac{\sum_t n_{gt} y_{gt}}{\sum_t n_{gt}}\right) = \sqrt{n_{gt}}\left(\mathbf{x}_{gt} - \frac{\sum_t n_{gt}\mathbf{x}_{gt}}{\sum_t n_{gt}}\right)'\boldsymbol{\beta} \\ + \sqrt{n_{gt}}\left(\varepsilon_{gt} - \frac{\sum_t n_{gt}\varepsilon_{gt}}{\sum_t n_{gt}}\right)\end{aligned} \tag{2}$$

In other words, the weighted dummy-variable estimator is equivalent to a *weighted* within estimator that replaces y_{gt} and $\mathbf{x}_{gt}$ with deviations from their *weighted* means. This estimator may be preferable when the number of groups is large.

The following illustrates this weighted-deviation-from-weighted-means interpretation:

```
. bysort g: egen sumn=sum(n)
. foreach z in x y {
  2.        generate `z'w=`z'*n
  3.        bysort g: egen `z'wsum = sum(`z'w) // weighted sums
  4.        generate `z'wbar = `z'wsum/sumn    // weighted means
  5.        generate `z'dev = `z'-`z'wbar      // deviations from weighted means
  6. }
. regress ydev xdev [aw=n], nocons
(sum of wgt is 10,289)
```

Source	SS	df	MS	Number of obs	=	200
				F(1, 199)	=	2590.40
Model	656.859929	1	656.859929	Prob > F	=	0.0000
Residual	50.4613654	199	.2535747	R-squared	=	0.9287
				Adj R-squared	=	0.9283
Total	707.321294	200	3.53660647	Root MSE	=	.50356

ydev	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
xdev	2.001036	.0393162	50.90	0.000	1.923506	2.078566

Although the weighted-within point estimates are identical to the dummy-variable estimates, the standard errors are incorrect because they fail to account for the degrees of freedom used in computing the group-level weighted means.[3] Fortunately, the `areg` command does just that:

```
. areg y x [aw=n], absorb(g)
(sum of wgt is 10,289)
Linear regression, absorbing indicators          Number of obs     =       200
Absorbed variable: g                             No. of categories =       100
                                                 F(1, 99)          =   1288.69
                                                 Prob > F          =    0.0000
                                                 R-squared         =    0.9937
                                                 Adj R-squared     =    0.9873
                                                 Root MSE          =    0.7139

------------------------------------------------------------------------------
           y |  Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           x |   2.001036   .0557417    35.90   0.000     1.890432     2.11164
       _cons |   5.644418   .0697073    80.97   0.000     5.506104    5.782733
------------------------------------------------------------------------------
F test of absorbed indicators: F(99, 99) = 6.198              Prob > F = 0.000
```

The `xtreg` command with the `fe` option fits fixed-effects models similarly. However, because `xtreg` does not support time-varying weights, it cannot be used in this application.[4]

Another way to eliminate the group fixed effects in (1) is via the first-differences model

$$\Delta y_{gt} = \Delta \mathbf{x}'_{gt}\boldsymbol{\beta} + \Delta \varepsilon_{gt} \tag{3}$$

Empiricists frequently weight first-differenced models of group averages by $1/(1/n_{gt} + 1/n_{g,t-1})$, the justification being that, if the individual-level errors are homoskedastic and serially uncorrelated, then $V(\Delta\varepsilon_{gt}) = \sigma^2(1/n_{gt} + 1/n_{g,t-1})$.

In the unweighted case, it is well known that fixed effects and first differences are identical when there are only two time periods. Thus, it may not be surprising that fixed effects weighted by n_{gt} and first differences weighted by $1/(1/n_{gt} + 1/n_{g,t-1})$ are also identical in this case, as the following demonstrates:

```
. xtset g t
Panel variable: g (strongly balanced)
 Time variable: t, 1 to 2
         Delta: 1 unit
. generate wt=1/(1/n+1/l.n)
(100 missing values generated)
```

3. The correct degrees of freedom is $N(T-1)-K$, where N is the number of panel units, T is the number of time periods, and K is the number of regressors (this is also the degrees of freedom for a regression of y_{gt} on $\mathbf{x}_{gt}$ and a full set of N group dummies). If the ε_{gt} are independently distributed, valid standard errors can be obtained by multiplying the default standard errors by $\sqrt{(NT-K)/\{N(T-1)-K\}}$.
4. On the other hand, the **areg** (see [R] **areg**) command is not designed for applications where the number of groups increases with the sample size.

```
. regress d.y d.x [aw=wt], nocons
(sum of wgt is 2,114.55251610279)
      Source |       SS           df       MS      Number of obs   =       100
-------------+----------------------------------   F(1, 99)        =   1288.69
       Model |  1598.07611         1  1598.07611   Prob > F        =    0.0000
    Residual |  122.767572        99  1.24007648   R-squared       =    0.9287
-------------+----------------------------------   Adj R-squared   =    0.9279
       Total |  1720.84368       100  17.2084368   Root MSE        =    1.1136

------------------------------------------------------------------------------
         D.y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           x |
         D1. |   2.001036   .0557417    35.90   0.000     1.890432     2.11164
------------------------------------------------------------------------------
```

To see why this holds, note that, in the two-period case, the left-hand side of (2) is

$$\sqrt{n_{gt}}\frac{n_{gt'}(y_{gt}-y_{gt'})}{n_{gt}+n_{gt'}}$$

and similarly for the right-hand side. Thus, the WFE estimate of $\boldsymbol{\beta}$ is

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{WFE}} &= \left\{\sum_{g,t}\frac{n_{gt}n_{gt'}^2(\mathbf{x}_{gt}-\mathbf{x}_{gt'})(\mathbf{x}_{gt}-\mathbf{x}_{gt'})'}{(n_{g1}+n_{g2})^2}\right\}^{-1} \\
&\quad \left\{\sum_{g,t}\frac{n_{gt}n_{gt'}^2(\mathbf{x}_{gt}-\mathbf{x}_{gt'})(y_{gt}-y_{gt'})}{(n_{g1}+n_{g2})^2}\right\} \\
&= \left\{\sum_{g}\frac{(n_{g1}+n_{g2})n_{g1}n_{g2}\Delta\mathbf{x}_{gt}\Delta\mathbf{x}_{gt}'}{(n_{g1}+n_{g2})^2}\right\}^{-1}\left\{\sum_{g}\frac{(n_{g1}+n_{g2})n_{g1}n_{g2}\Delta\mathbf{x}_{gt}\Delta y_{gt}}{(n_{g1}+n_{g2})^2}\right\} \\
&= \left(\sum_{g}\frac{n_{g1}n_{g2}\Delta\mathbf{x}_{gt}\Delta\mathbf{x}_{gt}'}{n_{g1}+n_{g2}}\right)^{-1}\left(\sum_{g}\frac{n_{g1}n_{g2}\Delta\mathbf{x}_{gt}\Delta y_{gt}}{n_{g1}+n_{g2}}\right)
\end{aligned}$$

The last expression is precisely the vector of coefficients $\widehat{\boldsymbol{\beta}}^{\mathrm{WFE}}$ on $\Delta\mathbf{x}_{gt}$ from a weighted least-squares estimate of (3).

References

Greene, W. H. 2018. *Econometric Analysis*. 8th ed. New York: Pearson.

Solon, G., S. J. Haider, and J. M. Wooldridge. 2015. What are we weighting for? *Journal of Human Resources* 50: 301–316. https://doi.org/10.3368/jhr.50.2.301.

StataCorp. 2021. *Stata 17 User's Guide*. College Station, TX: Stata Press.