OLS with two fixed effects

Die Mathematiker sind eine Art Franzosen: redet man zu ihnen, so übersetzen sie es in ihre Sprache, und dann ist es alsobald ganz etwas anderes. – J.W.v. Goethe

Simen Gaure

Frisch Centre

Sep 1, 2010

Our corner of reality

Consider a model

$$y_{ijt} = X_{ijt}\beta + \alpha_i + \gamma_j + \epsilon$$

where y_{ijt} is log-wage of individual i in firm j at time t. X_{ijt} is a set of time-varying covariates for individual i and firm j, α_i is an individual fixed effect, and γ_j is a firm-fixed effect.

- Such models have been used to study correlations between individual effects and firm-effects ("High wage workers and high wage firms"). It picks up arbitrary unobserved heterogeneity in both firms and workers. Correlation between α, γ and X is allowed.
- One might imagine other fixed effects, such as students and schools, or scientists and journals, citizens and their home-location. Or even 3 or more fixed effects.
- In some cases we might not be interested in the α's and γ's, we only need them as controls.
- These models have been difficult to estimate due to the high number of dummy-variables. (One for each firm, one for each individual)

Single fixed effect - familiar within groups estimator

- When only considering one fixed effect, e.g. individual fixed effects, the "within-groups" estimator is frequently used.

$$Y = X\beta + D\alpha + \epsilon$$

- Subtract the group-means from Y and X, and find $\hat{\beta}_{FE}$ from the system

$$\bar{Y} = \bar{X}\beta + \epsilon$$

Single fixed effect - familiar within groups estimator

- When only considering one fixed effect, e.g. individual fixed effects, the "within-groups" estimator is frequently used.

$$Y = X\beta + D\alpha + \epsilon$$

- Subtract the group-means from Y and X, and find $\hat{\beta}_{FE}$ from the system

$$\bar{Y} = \bar{X}\beta + \epsilon$$

- "Time-demeaning", "centering on the means", "creating mean deviations", "sweeping out the fixed effects". Many names for this. I like centering and sweeping.

Single fixed effect - familiar within groups estimator

- When only considering one fixed effect, e.g. individual fixed effects, the "within-groups" estimator is frequently used.

$$Y = X\beta + D\alpha + \epsilon$$

- Subtract the group-means from Y and X, and find $\hat{\beta}_{FE}$ from the system

$$\bar{Y} = \bar{X}\beta + \epsilon$$

- "Time-demeaning", "centering on the means", "creating mean deviations", "sweeping out the fixed effects". Many names for this. I like centering and sweeping.
- If we need $\hat{\alpha}$ too, i.e. not only to control for fixed effects, we may recover $\hat{\alpha}$ by solving $D'D\hat{\alpha} = D'(Y - X\hat{\beta}_{FE})$. This is easy since $D'D$ is diagonal. $\hat{\alpha}$ turns out to be the group means of the residuals.

Two (or more) fixed effects — ineffably tangled mess

- With two fixed effects ("worker" α and "firm" γ)

$$Y = X\beta + D_1\alpha + D_2\gamma + \epsilon \tag{1}$$

it's apparently harder to estimate $\hat{\beta}_{FE}$ by sweeping out *both* fixed effects to get a system like $\bar{Y} = \bar{X}\beta + \epsilon$

Two (or more) fixed effects — ineffably tangled mess

- With two fixed effects ("worker" α and "firm" γ)

$$Y = X\beta + D_1\alpha + D_2\gamma + \epsilon \tag{1}$$

  it's apparently harder to estimate $\hat{\beta}_{FE}$ by sweeping out *both* fixed effects to get a system like $\bar{Y} = \bar{X}\beta + \epsilon$
- *"This fails in our model because creating firm mean deviations destroys the patterning in D_1 and creating worker mean deviations destroys the patterning in D_2"* (Andrews et al, JRSS(2008)). Similarly (but more detailed) in Abowd et al, Econometrica(1999). And elsewhere and in between.

Two (or more) fixed effects — ineffably tangled mess

- With two fixed effects ("worker" α and "firm" γ)

$$Y = X\beta + D_1\alpha + D_2\gamma + \epsilon \tag{1}$$

  it's apparently harder to estimate $\hat{\beta}_{FE}$ by sweeping out *both* fixed effects to get a system like $\bar{Y} = \bar{X}\beta + \epsilon$
- *"This fails in our model because creating firm mean deviations destroys the patterning in D_1 and creating worker mean deviations destroys the patterning in D_2"* (Andrews et al, JRSS(2008)). Similarly (but more detailed) in Abowd et al, Econometrica(1999). And elsewhere and in between.
- Thus, various elaborate estimation schemes have been concocted during the last decade; including sweeping out *one* of the fixed effects, iterated estimations, linear approximations, conjugate gradients with ingenious preconditioners; with all sorts of mathematical formulæ with triple indices, double sums, and even *square* roots.

Two (or more) fixed effects — ineffably tangled mess

- With two fixed effects ("worker" α and "firm" γ)

$$Y = X\beta + D_1\alpha + D_2\gamma + \epsilon \tag{1}$$

  it's apparently harder to estimate $\hat{\beta}_{FE}$ by sweeping out *both* fixed effects to get a system like $\bar{Y} = \bar{X}\beta + \epsilon$
- *"This fails in our model because creating firm mean deviations destroys the patterning in D_1 and creating worker mean deviations destroys the patterning in D_2"* (Andrews et al, JRSS(2008)). Similarly (but more detailed) in Abowd et al, Econometrica(1999). And elsewhere and in between.
- Thus, various elaborate estimation schemes have been concocted during the last decade; including sweeping out *one* of the fixed effects, iterated estimations, linear approximations, conjugate gradients with ingenious preconditioners; with all sorts of mathematical formulæ with triple indices, double sums, and even *square* roots.
- This is the wrong way to think about this problem, we're not supposed to sweep out workers from their firms, but to sweep out workers *and* firms from eq. (1).

- By doing Gaussian elimination on the normal equations, we see that we need to center on both means at once; like this $\bar{Y} = PY; \quad \bar{X} = PX$ where P is the projection onto the orthogonal complement of the column space of $D = [D_1 \quad D_2]$; (i.e. $R(D)^{\perp}$).

Street sweepers — John von Neumann and Israel Halperin

- By doing Gaussian elimination on the normal equations, we see that we need to center on both means at once; like this $\bar{Y} = PY; \quad \bar{X} = PX$ where P is the projection onto the orthogonal complement of the column space of $D = [D_1 \quad D_2]$; (i.e. $R(D)^{\perp}$).
- Centering on individual means only, is also a projection P_1 (onto $R(D_1)^{\perp}$). Similarly with the firm means, P_2.

Street sweepers — John von Neumann and Israel Halperin

- By doing Gaussian elimination on the normal equations, we see that we need to center on both means at once; like this $\bar{Y} = PY; \quad \bar{X} = PX$ where P is the projection onto the orthogonal complement of the column space of $D = [D_1 \quad D_2]$; (i.e. $R(D)^{\perp}$).
- Centering on individual means only, is also a projection P_1 (onto $R(D_1)^{\perp}$). Similarly with the firm means, P_2.
- Moreover, we have $P = P_1 \wedge P_2$, the intersection of P_1 and P_2. (Because $R(D)^{\perp} = R(D_1)^{\perp} \cap R(D_2)^{\perp}$). The problem is to compute Pv given a vector v, using the simple transformations P_1 and P_2.

Street sweepers — John von Neumann and Israel Halperin

- By doing Gaussian elimination on the normal equations, we see that we need to center on both means at once; like this $\bar{Y} = PY; \quad \bar{X} = PX$ where P is the projection onto the orthogonal complement of the column space of $D = [D_1 \quad D_2]$; (i.e. $R(D)^{\perp}$).
- Centering on individual means only, is also a projection P_1 (onto $R(D_1)^{\perp}$). Similarly with the firm means, P_2.
- Moreover, we have $P = P_1 \wedge P_2$, the intersection of P_1 and P_2. (Because $R(D)^{\perp} = R(D_1)^{\perp} \cap R(D_2)^{\perp}$). The problem is to compute Pv given a vector v, using the simple transformations P_1 and P_2.
- If two projections happen to commute, $P_1 P_2 = P_2 P_1$, we have from general theory, $P_1 \wedge P_2 = P_1 P_2$, i.e. we get away by centering on the individual means, then on the firm means. Easy cases.

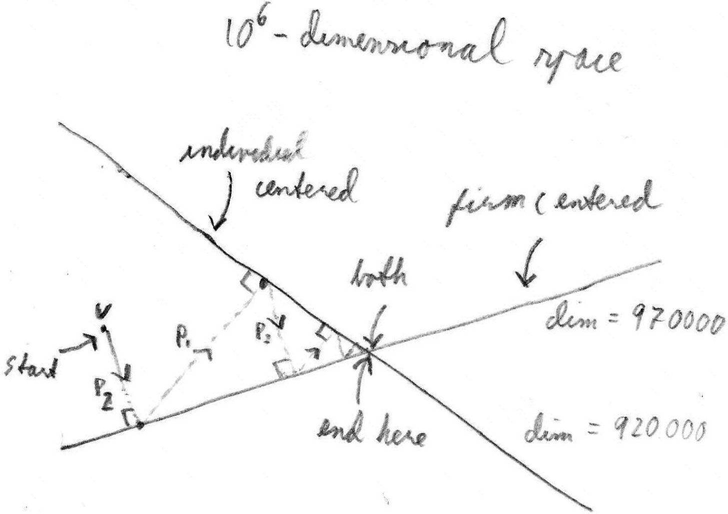
Street sweepers — John von Neumann and Israel Halperin

- By doing Gaussian elimination on the normal equations, we see that we need to center on both means at once; like this $\bar{Y} = PY; \quad \bar{X} = PX$ where P is the projection onto the orthogonal complement of the column space of $D = [D_1 \quad D_2]$; (i.e. $R(D)^{\perp}$).
- Centering on individual means only, is also a projection P_1 (onto $R(D_1)^{\perp}$). Similarly with the firm means, P_2.
- Moreover, we have $P = P_1 \wedge P_2$, the intersection of P_1 and P_2. (Because $R(D)^{\perp} = R(D_1)^{\perp} \cap R(D_2)^{\perp}$). The problem is to compute Pv given a vector v, using the simple transformations P_1 and P_2.
- If two projections happen to commute, $P_1 P_2 = P_2 P_1$, we have from general theory, $P_1 \wedge P_2 = P_1 P_2$, i.e. we get away by centering on the individual means, then on the firm means. Easy cases.
- In general, they don't commute, but there's a theorem by von Neumann (1933) stating that $P_1 \wedge P_2 = \lim_{n\to\infty} (P_1 P_2)^n$. Israel Halperin extended it in 1962 to any finite number of projections.

Street sweepers — John von Neumann and Israel Halperin

- By doing Gaussian elimination on the normal equations, we see that we need to center on both means at once; like this $\bar{Y} = PY; \quad \bar{X} = PX$ where P is the projection onto the orthogonal complement of the column space of $D = [D_1 \quad D_2]$; (i.e. $R(D)^{\perp}$).
- Centering on individual means only, is also a projection P_1 (onto $R(D_1)^{\perp}$). Similarly with the firm means, P_2.
- Moreover, we have $P = P_1 \wedge P_2$, the intersection of P_1 and P_2. (Because $R(D)^{\perp} = R(D_1)^{\perp} \cap R(D_2)^{\perp}$). The problem is to compute Pv given a vector v, using the simple transformations P_1 and P_2.
- If two projections happen to commute, $P_1 P_2 = P_2 P_1$, we have from general theory, $P_1 \wedge P_2 = P_1 P_2$, i.e. we get away by centering on the individual means, then on the firm means. Easy cases.
- In general, they don't commute, but there's a theorem by von Neumann (1933) stating that $P_1 \wedge P_2 = \lim_{n\to\infty} (P_1 P_2)^n$. Israel Halperin extended it in 1962 to any finite number of projections.
- Thus, we may sweep out both (or more) fixed effects by centering Y and X on the individual means, then on the firm means, then on the individual means again, then the firm means, individual means, firm means, i.m., f.m., ..., until ... they're gone.

A graphical rendition of the process



So ... What about the fixed effects?

- Once we have acquired $\hat{\beta}$, we compute the residuals $R = Y - X\hat{\beta}$, go back to the normal equations and note that $D'D \begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = D'R$ where $D = \begin{bmatrix} D_1 & D_2 \end{bmatrix}$. (Just like the within-groups estimator).

So ... What about the fixed effects?

- Once we have acquired $\hat{\beta}$, we compute the residuals $R = Y - X\hat{\beta}$, go back to the normal equations and note that $D'D \begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = D'R$ where $D = \begin{bmatrix} D_1 & D_2 \end{bmatrix}$. (Just like the within-groups estimator).
- Unlike the case with a single fixed effect, $D'D$ is not diagonal. Luckily, identification is a complicated issue (though, solved by Abowd et al) which lowers the dimension of this system.

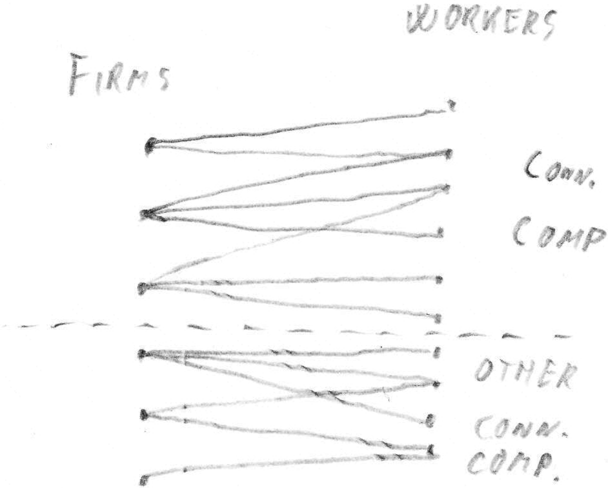
So ... What about the fixed effects?

- Once we have acquired $\hat{\beta}$, we compute the residuals $R = Y - X\hat{\beta}$, go back to the normal equations and note that $D'D \begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = D'R$ where $D = \begin{bmatrix} D_1 & D_2 \end{bmatrix}$. (Just like the within-groups estimator).
- Unlike the case with a single fixed effect, $D'D$ is not diagonal. Luckily, identification is a complicated issue (though, solved by Abowd et al) which lowers the dimension of this system.
- Consider a case where we have 15000 firms with 300000 employees moving between them. Then we have another set of 9000 firms, with a set of 120000 employees. But no employee in one group ever moves into the other group. In each group we may add a constant c to every individual effect, and subtract the same constant from every firm effect, and we will not be able to tell:

  $$y_{ijt} = x_{ijt}\beta + (\alpha_i + c) + (\gamma_j - c) = x_{ijt}\beta + \alpha_i + \gamma_j$$

  We may have a different constant c in each group, thus estimates are not comparable across groups.

Connection components



That's all there is to it

- Divide the dataset into (graph-theoretic) connection components (of firms and individuals, with no relation to the other components).
- In each component we must pick a reference (either a firm or an employee). (This is due to a somewhat lengthy argument by Abowd et al, but is also a direct consequence of a theorem about the signless Laplacian spectrum of bipartite graphs). Then we may solve for the fixed effects separately in each component, like $B'B \begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = B'R$ (these are "sparse" systems, i.e. mostly zeroes). All the systems are conditional on the jointly estimated $\hat{\beta}$ (via the residuals $R = Y - X\hat{\beta}$).
- The fixed effects are identified up to adding a constant c to the individual effects and subtracting the same c from the firm effects. Thus, differences (and variances) between individual effects within a component are identified, and so are differences between firm effects.

Practical estimation

We have installed software (under the name "LFE" - linear fixed effects) doing the above on "leif", our compute cluster. Both centering and solving the component systems are *embarrassingly parallel* tasks, thus we do this in parallel over 8 cpus. A typical specification file looks like this

```
file middata.csv                          # name of data file
vars x x2 year id firm y ife ffe yfe      # layout of data-file
model y ~ x + x2 + year                   # R-style model-spec
dummy year                                # tell'em it's categories
complim 10                                # ditch small components
fe firm id                                # fixed effects
```

A package "lfe" will shortly be uploaded to "CRAN", the R package repository for public download. (Our installed "LFE" is just a wrapper around this package.)

Elsewhere

This way of finding the intersection of projections is known as *MAP - Method of Alternating Projections* and has been in use in image processing, not in Photoshop (I think), but in computed tomography (CT) for a while, where it is known as *ART - Algebraic Reconstruction Technique*. In numerical linear algebra it is known as *The Kaczmarz Method*.

J.M. Abowd, R.H. Creecy, and F. Kramarz, *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data*, Tech. Report TP-2002-06, U.S. Census Bureau, 2002.

M. Andrews, L. Gill, T. Schank, and R. Upward, *Hige wage workers and low wage firms: negative assortative matching or limited mobility bias?*, J.R. Stat. Soc.(A) **171(3)** (2008), 673–697.

J.M. Abowd, F. Kramarz, and D.N. Margolis, *High Wage Workers and High Wage Firms*, Econometrica **67** (1999), no. 2, 251–333.

L. Aronszajn, *Theory of Reproducing Kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.

F. Deutsch and H. Hundal, *The Rate of Convergence for the Method of Alternating Projections, II*, J. Math. Anal. App. **205** (1997), no. 2, 381–405.

William B. Gearhart and Mathew Koshy, *Acceleration schemes for the method of alternating projections*, Journal of Computational and Applied Mathematics **26** (1989), no. 3, 235–249.

I. Halperin, *The Product of Projection Operators*, Acta Sci. Math. (Szeged) **23** (1962), 96–99.

A. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bulletin International de l'Academie Polonaise des Sciences et des Lettres **35** (1937), 355–357.

B.G. Salomon and H. Ur, *Accelerating the Convergence of the von Neumann-Halperin Method of Alternating Projections*, Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop, 4th, 2006, pp. 328 –331.

J. von Neumann, *On Rings of Operators. Reduction Theory*, Ann. Math. **50** (1949), 401–485.