Notes on the estimation of high-dimensional FE with variable returns to ability.

Ekaterina Roshchina

March 20, 2016

1 Proposed procedure

Model to be estimated:

$$log\_earn_{ijart} = \mu_{rt} + \delta_t \alpha_i + \theta_a + \gamma_{jt} + \varepsilon_{iarjt} \tag{1}$$

Algorithm:

1. Initialize the values of $\delta_t^{(0)}$, $\alpha_i^{(0)}$.
2. At iteration k, estimate $\mu_{rt}^{(k)}$, $\theta_a^{(k)}$, $\gamma_{jt}^{(k)}$ by OLS given the values $\delta_t^{(k-1)}\alpha_i^{(k-1)}$. Do it in two steps:
   (a) Estimate $\mu_{rt}^{(k)}$ given $\gamma_{jt}^{(k-1)}$ and $\delta_t^{(k-1)}\alpha_i^{(k-1)}$.
   (b) Estimate $\gamma_{jt}^{(k)}$ and $\theta_a^{(k)}$ given $\mu_{rt}^{(k)}$ and $\delta_t^{(k-1)}\alpha_i^{(k-1)}$.
3. Renew the values for $\delta_t^{(k)}$, $\alpha_i^{(k)}$ using the following expressions:

$$\alpha_i^{(k)} = \sum_t \tilde{\varepsilon}_{it}^{(k)} \delta_t^{(k-1)} \Big/ \sum_t \left(\delta_t^{(k-1)}\right)^2, \tag{2}$$

$$\delta_t^{(k)} = \sum_i \tilde{\varepsilon}_{it}^{(k)} \alpha_i^{(k-1)} \Big/ \sum_i \left(\alpha_i^{(k-1)}\right)^2, \tag{3}$$

   where $\tilde{\varepsilon}_{it}^{(k)} = log\_earn_{ijart} - \mu_{rt}^{(k)} - \theta_a^{(k)} - \gamma_{jt}^{(k)}$.
4. Normalize δ_1 at year 1 to be 1.
5. Iterate steps 2-3 till convergence.
   - Convergence criterion: residual sum of squares of the model is not changing from k to $k-1$.
     Current tolerance: 1e-4

- Convergence check: Coefficients at $\delta_t^k \alpha_i^k$, $\mu_{rt}^{(k)}$ and $\gamma_{jt}^{(k)}$ should equal to 1

This procedure mirrors the suggestion of **?** on pp. 429-430. The basic intuition comes from the fact that the estimates should minimize the NLS objective function Q_n:

$$Q_n = \sum_i \sum_t \left(log\_earn_{ijart} - \mu_{rt} - \delta_t \alpha_i - \theta_a - \gamma_{jt} \right)^2 \tag{4}$$

However, since this is infeasible, at each step only low-dimension coefficients are estimated, and then the person fixed effects are recovered using the first order conditions for estimating δ_t and α_i These FOCs give the expressions for updating. What is required for the procedure to work is that $Q_n^(k)$ decreases at each iteration, i.e. it is a contraction mapping. I do not have a formal proof for this case, though in the empirical implementation $Q_n^(k)$ does decrease at each iteration.

FOCs for the estimates of δ_t and α_i:

$$\sum_t \left(log\_earn_{ijart} - \mu_{rt} - \delta_t \alpha_i - \theta_a - \gamma_{jt} \right) \alpha_i = 0 \tag{α_i}$$

$$\sum_i \left(log\_earn_{ijart} - \mu_{rt} - \delta_t \alpha_i - \theta_a - \gamma_{jt} \right) \delta_t = 0 \tag{δ_t}$$

1.1 Standard errors

Standard errors are implemented using wild bootstrap, as in **?**.

Implementation is based on the guidelines from **?**.

Steps:

1. Predict $\widehat{log\_earn}_{ijart}$ and $\hat{\varepsilon}_{iarjt}$.

2. Draw a sample of residuals (with replacement) from $\hat{\varepsilon}_{iarjt}$

3. Construct a bootstrap sample as follows:

$$log\_earn^*_{ijart} = \widehat{log\_earn}_{ijart} + f(\hat{\varepsilon}^*_{ijart}) v^*_{ijart}, \tag{5}$$

where $f(\hat{\varepsilon}^*_{ijart})$ is simply $\hat{\varepsilon}^*_{ijart}$ (resampled residuals), and

$$v^*_{ijart} = \begin{cases} -0.5\left(\sqrt{5}-1\right) & \text{with probability } \left(\sqrt{5}+1\right)/\left(2\sqrt{5}\right) \\ -0.5\left(\sqrt{5}+1\right) & \text{with probability } \left(\sqrt{5}-1\right)/\left(2\sqrt{5}\right) \end{cases} \tag{6}$$

2 Fixed effects: `felsdvreg` vs. `reghdfe`

2.1 Differences between `felsdvreg` vs. `reghdfe`

Normalization used by `reghdfe`:

- Mean of each fixed effect is 0, i.e. $\frac{1}{N}\sum_i \alpha_i = 0$, $\frac{1}{N}\sum_i \mu_{rt} = 0$, $\frac{1}{N}\sum_i \gamma_{jt} = 0$, where N is the total number of observations For industry and region fixed effects that translates into $\frac{1}{N}\sum_{rt} N_{rt}\mu_{rt} = 0$, where N_{rt} is the number of observations in rt cell.

Normalization used by `felsdvreg`:

- Normalization for industry-year effect $gamma_{jt}$ is chose manually, as these effects are not estimated by `felsdvreg`.
- Normalization for region-year effect is based on mobility groups. `felsdvreg` splits all observation on groups connected by worker mobility (0 stands for no movers group, 1 is always the largest group in the data etc.). Region-year effect (or "firm" effect in the usual terminology) is set to 0 for the first region-year observation in each group.

Similarities between the two procedures:

- Estimates for any variables, which are not in the fixed effects (i.e. returns to age here) are *exactly* the same.

Difference between the number of region-year estimation produced by `felsdvreg` and by `reghdfe`:

- Both `felsdvreg` and `reghdfe` use the same sample to estimate the model and yield the same estimates for θ_a. In other words, `e(sample)` marks the same set in Stata for both estimates
- Final estimation sample for `reghdfe` must have at least 2 observations per each group for which fixed effects are estimated. This can be achieved using the following algorithm:
    1. For each fixed effects variable, calculate the number of observations group in the current sample
    2. Mark any observations with less than 2 observations per group as not in the estimation sample
    3. Repeat steps 1–2 till there are no observations to be dropped.

- There are 8 mmc-year cells which have more than 2 observations in the data, but get dropped by `reghdfe`:

	mmc_year	mmc	year	n_obs_mmc_year_data
128.	128	12003	1990	2
129.	129	12003	1991	3
416.	416	13011	2001	2
478.	478	14004	1994	2
1034.	1034	16004	1987	2
1617.	1617	22012	1994	2
4122.	4122	27007	2004	4
11280.	11280	51008	1989	2

- In all cases except mmc_year 1034 (mmc 16004 in 1987), this happens because of the individual observations which get dropped, because these people have fewer than 2 observations. Once all people with 1 observations are dropped, there is only 1 observation per mmc-year cell.

	indiv	subs_ibge_year	mmc_year	n_obs_indiv_data
1.	594277	118	128	10
2.	952024	116	128	1
3.	594277	142	129	10
4.	952023	123	129	1
5.	952165	123	129	1
6.	1247016	374	416	1
7.	2106647	382	416	17
8.	307971	196	478	8
9.	1073372	196	478	1
10.	263675	27	1034	6
11.	779331	30	1034	2
12.	319010	210	1617	20
13.	1080345	195	1617	1

```
14. |  430592        433       4122                    7 |
15. | 2166342        453       4122                    1 |
16. | 2166343        453       4122                    1 |
17. | 2166346        453       4122                    1 |
    |--------------------------------------------------|
18. |  649923         96      11280                    3 |
19. |  777379         73      11280                    1 |
    +--------------------------------------------------+
```

- The case of mmc-year 1034 is similar: Person 779331 in year 1988 gets dropped because this is an only observation for mmc-year 1035 Mmc-year 1034 still has 2 observation per group, but then person 779331 get dropped altogether, because now there is only 1 observation for him. Which also leaves only 1 observation for mmc-year 103, and it's excluded from the estimation sample as well.

```
    +--------------------------------------------------+
    |  indiv   year   mmc_year   n_obs_mmc_year_data |
    |--------------------------------------------------|
 1. | 263675   1987       1034                     2 |
 2. | 263675   1991       1036                     8 |
 3. | 263675   1992       1037                     5 |
 4. | 263675   1993       1038                     4 |
 5. | 263675   1994       1039                     7 |
 6. | 263675   1995       1040                     8 |
    |--------------------------------------------------|
 7. | 779331   1987       1034                     2 |
 8. | 779331   1988       1035                     1 |
    +--------------------------------------------------+
```

2.2 Algorithm to re-normalize results of `reghdfe` to be the same as `felsdvreg`

1. **Industry-year FE.** Denote by γ_{0t} the reference industry within each year (omitted group). Also, denote by $\tilde{\gamma}_{jt}$ the new, re-normalized estimates, and by $\hat{\gamma}_{jt}$ the original `reghdfe` estimates. Then, the modified estimates are:

$$\tilde{\gamma}_{jt} = \hat{\gamma}_{jt} - \hat{\gamma}_{0t} \tag{7}$$

2. **Region-year FE.** Denote by μ_{0t} the reference region-year group within each mobility group (omitted group). Then, the modifies estimates are:

$$\tilde{\mu}_{jt} = \hat{\mu}_{rt} - \hat{\mu}_{0t} + \hat{\gamma}_{0t} - \frac{1}{N}\sum_{t} N_t \hat{\gamma}_{0t} \tag{8}$$

3. **Individual FE.** Denote by $\hat{b}_1$ the constant in the regression of log earnings on age categories and three fixed effects estimates by `reghdfe`, and by $\hat{b}_2$ the constant in the regression of

log earnings on age categories and three fixed effects estimates by `felsdvreg`.

$$\tilde{\alpha}_i = \hat{\alpha}_i + \hat{b}_1 - + \hat{b}_2 + \hat{\mu}_{0t} + \frac{1}{N}\sum_t N_t \hat{\gamma}_{0t} \tag{9}$$

In case of varying return to ability, renormalized individual fixed effects can be obtained as follows:

$$\tilde{\alpha}_i = \hat{\alpha}_i + \frac{1}{\hat{\delta}_t}\left(\hat{b}_1 - + \hat{b}_2 + \hat{\mu}_{0t} + \frac{1}{N}\sum_t N_t \hat{\gamma}_{0t}\right) \tag{10}$$

Caveats:

- `reghdfe` and `felsdvreg` yield estimates for different samples, even tough they use the same samples to estimate the model. `reghdfe` operates on the sample for which there are at least 2 observation for each individual, each industry-year category and each region-year category. It the data data, 3% of individuals for whom `felsdvreg` identifies individual fixed effects having missing individual fixed effect in the `reghdfe` routine. `felsdvreg` used the same sample in the estimation, but then constructs fixed effects for observation for outside of this sample (excluding observations which are in the 0 mobility group), so that residual for these observations is set to 0.

- Because of the fact above, `felsdvreg` occasionally can have region-year reference categories for which the corresponding `reghdfe` estimates are missing. For now, I omitted such observations from the re-normalization.

3 Implementation: iterative procedure with non-linear FE

Do-files description:

- `est_fe_actual_data_cp_as_is_v2.do`
  Estimates standard fixed effects model using `felsdvreg` and `reghdfe`.
- `est_fe_actual_data_cp_renormalize.do`
  Normalizes the estimates from `reghdfe` to be comparable with `felsdvreg` and saves the results.
- `est_nlfe_v14_actual_data.do`
  Estimates the model with variable returns to ability.
- `est_nlfe_v14_bootstrap.do`
  Runs bootstrap for the model with variable returns to ability.
- `est_nlfe_v15_bootstrap_collect.do`
  Calculates bootstrap standard errors and saves all the results in one file per estimate type.

Output description:

- `RegionYearFE_all_est_v2_withse_v2.dta`
  Variables:
  - `year`
  - `mmc`
  - `mmc_year`
  - `RegionYearFE1` Estimate of Region-Year FE using `reghdfe`
  - `RegionYearFE2` Estimate of Region-Year FE using `felsdvreg`
  - `RegionYearFE1_renorm` Estimate of Industry-Year FE using `reghdfe`, renormalized to be comparable with `felsdvreg`
  - `RegionYearFE_nlhdfe` Estimate of Region-Year FE using the model with the variable returns to ability
  - `RegionYearFE_nlhdfe_renorm` Estimate of Region-Year FE using the model with the variable returns to ability, renormalized to be comparable with `felsdvreg`
  - `B_effective` Number of bootstrap iterations (500)
  - `RegionYearFE_nl_se` S.e. estimate of Region-Year FE using the model with the variable returns to ability
  - `RegionYearFE_nl_renorm_se` S.e. estimate of Region-Year FE using the model with the variable returns to ability, renormalized to be comparable with `felsdvreg`
- `IndustryYearFE_all_est_v2_withse_v2.dta`
  Variables:

- year
- subs_ibge
- subs_ibge_year
- IndustryYearFE1 Estimate of Industry-Year FE using reghdfe
- IndustryYearFE2 Estimate of Industry-Year FE using felsdvreg
- IndustryYearFE1_renorm Estimate of Industry-Year FE using reghdfe, renormalized to be comparable with felsdvreg
- IndustryYearFE_nlhdfe Estimate of Industry-Year FE using the model with the variable returns to ability
- IndustryYearFE_nlhdfe_renorm Estimate of Industry-Year FE using the model with the variable returns to ability, renormalized to be comparable with felsdvreg
- B_effective Number of bootstrap iterations (500)
- IndustryYearFE_nl_se S.e. estimate of Industry-Year FE using the model with the variable returns to ability
- IndustryYearFE_nl_renorm_se S.e. estimate of Industry-Year FE using the model with the variable returns to ability, renormalized to be comparable with felsdvreg

• delta_nlhdfeall_est_v2_withse_v2.dta
Variables:

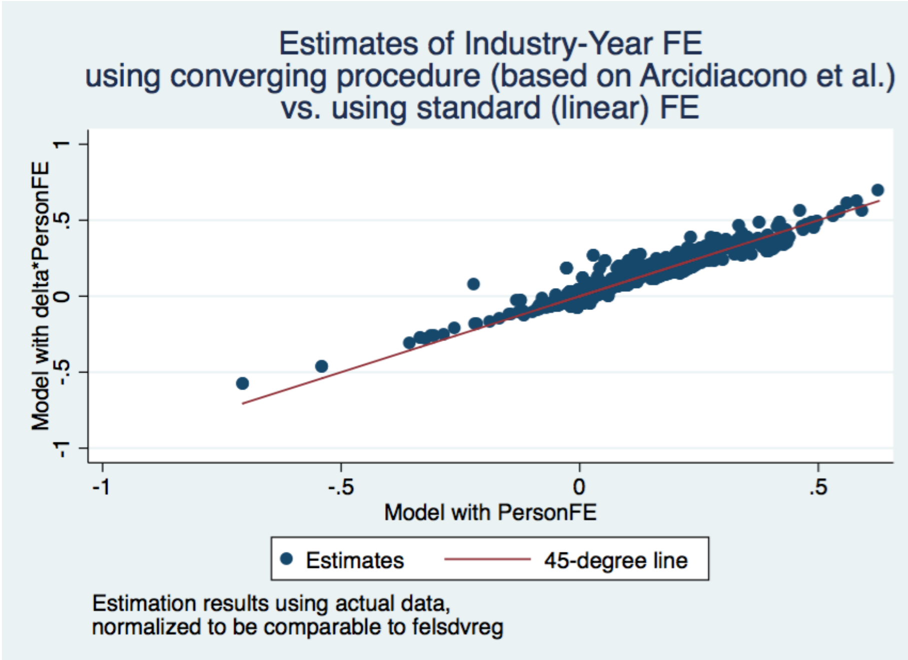
- year
- delta_nlhdfe Estimate of the return to ability using the model with the variable returns to ability
- B_effective Number of bootstrap iterations (500)
- delta_nl_se S.e. for the estimate of the return to ability using the model with the variable returns to ability

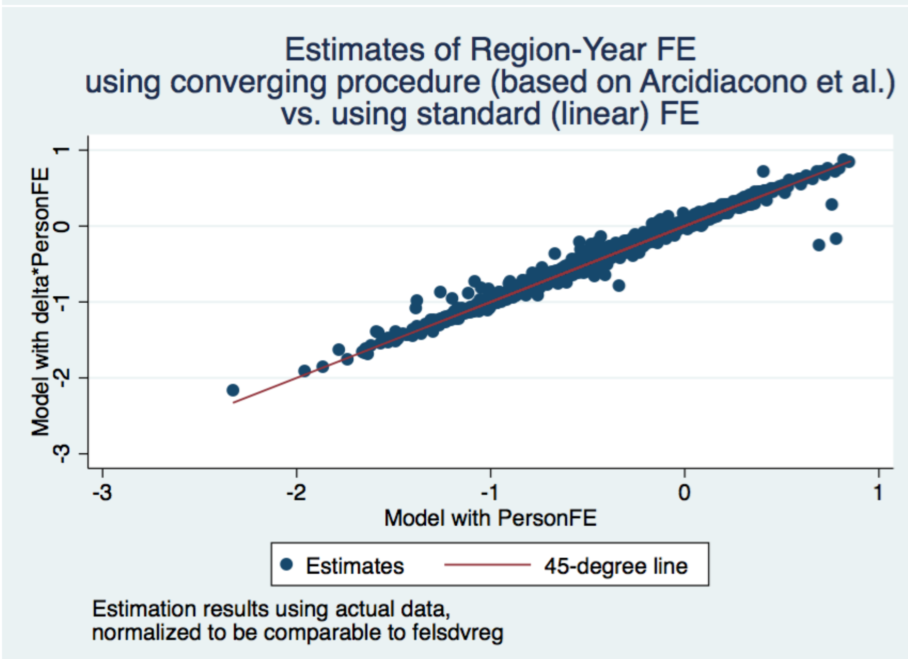
• IndFE_all_est_v2.dta
Variables:

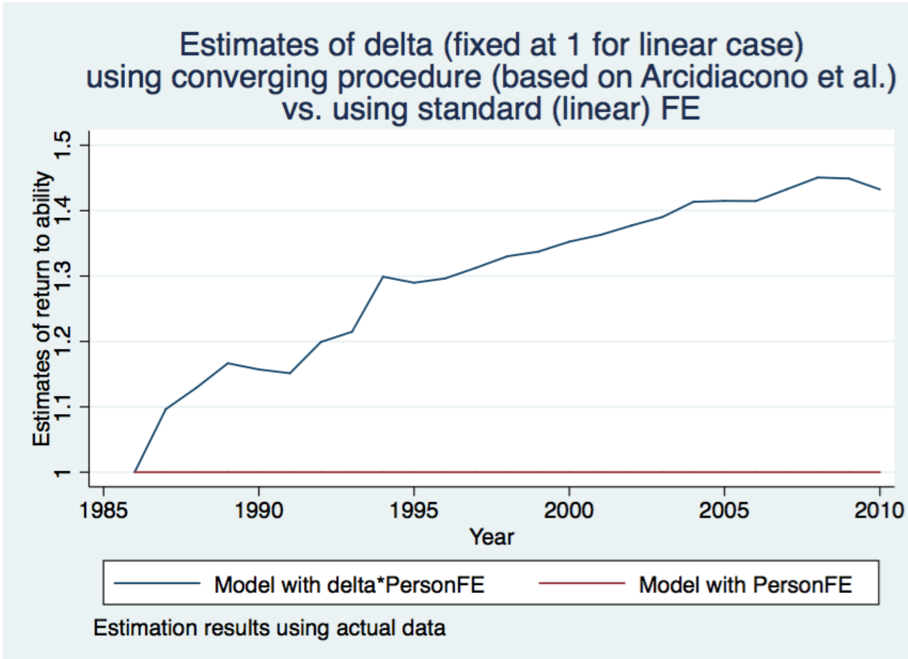
- indiv
- PersonFE1 Estimate of individual FE using reghdfe
- PersonFE2 Estimate of individual FE using felsdvreg
- PersonFE_nlhdfe Estimate of individual FE using the model with the variable returns to ability
- PersonFE1_renorm Estimate of individual FE using reghdfe renormalized to be comparable with felsdvreg
- PersonFE_nlhdfe_renorm Estimate of individual FE using the model with the variable returns to ability renormalized to be comparable with felsdvreg

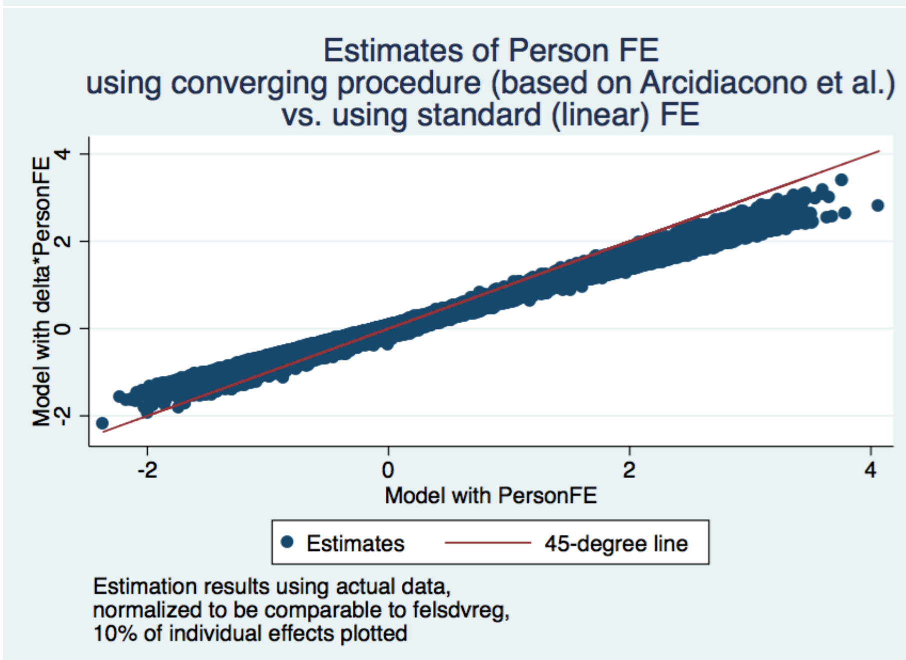
3.1 Results comparison: constant vs. varying returns to individual effects

Comparison of results using convergence procedure with results using (standard) linear FE.






Estimates of delta (fixed at 1 for linear case)
using converging procedure (based on Arcidiacono et al.)
vs. using standard (linear) FE
Estimates of return to ability
1
1.1
1.2
1.3
1.4
1.5
1985
1990
1995
2000
2005
2010
Year
Model with delta*PersonFE
Model with PersonFE
Estimation results using actual data


Estimates of Person FE
using converging procedure (based on Arcidiacono et al.)
vs. using standard (linear) FE
Model with delta*PersonFE
-2
0
2
4
-2
0
2
4
Model with PersonFE
Estimates
45-degree line
Estimation results using actual data,
normalized to be comparable to felsdvreg,
10% of individual effects plotted