Working Paper No. 782

Feasible Estimation of Linear Models with N-fixed Effects

by

Fernando Rios-Avila*
Levy Economics Institute of Bard College

December 2013

*Acknowledgements: This paper has benefited from the comments of Julie L. Hotchkiss and Thomas Masterson.

Levy Economics Institute
P.O. Box 5000
Annandale-on-Hudson, NY 12504-5000
http://www.levyinstitute.org

ISSN 1547-366X

ABSTRACT

In this paper an alternative approach for the estimation of higher-order linear fixed-effects models is described. The strategy relies on the transformation of the data prior to calculating estimations of the model. While the approach is computationally intensive, the hardware requirements for the estimation process are minimal, allowing for the estimation of models with more than two high-order fixed effects for large datasets. An illustration of the implementation is presented using the US Census Bureau Current Population Survey data with four fixed effects.

Keywords: Fixed-Effects Models; Two-Step Estimation; Stata

JEL Classifications: C13, C23

1. INTRODUCTION

With the availability of large longitudinal datasets in various fields, the interest in and application of models with one or more level fixed effects have increased. The ability to control for unobserved heterogeneity shared across groups using fixed effect models is attractive to researchers in fields like economics, sociology, political science and others, such as firm and worker (Abowd, Kramarz, & Woodcock, 2008), and schools, teachers, and students (Harris & Sass, 2011), as it allows to control for otherwise unobserved heterogeneity within groups. Whenever possible, the implementation of these types of models can be done by adding dummy sets that absorb the specific fixed effects.

For cases where the number of groups within a defined category is large, implementing such models using dummy sets can be difficult using standard statistical software, as it is constrained by the computer's memory capacity to manage large matrixes of estimated parameters. Furthermore, despite the advances in the access to high capacity hardware and software, the process of estimating models with more than one high order fixed effect for large datasets can be a challenge.

While linear models with a single fixed effect can be estimated without the need to include the set of dummies as regressors (within estimator and first-difference estimator, (see Cameron (2005)), there is no simple solution when there is more than one high dimensional fixed effect. Much of the literature dealing with the estimation of these types of models is based on the classical paper by Abowd, Kramarz, and Margolis (1999) in which the authors propose various methods to obtain estimates for a two-fixed effect model.[1]

In recent years, many strategies have been developed and implemented, allowing for the estimation of one and two high order fixed effects models, with different results in terms of systems requirements, computational efficiency and the estimation of standard errors (McCaffrey, Lockwood, Mihaly, & Sass, 2012).[2] Despite the growing literature dealing with high order fixed effect models, the analysis of data with more than two fixed effects is not yet routine. In a recent paper, Guimarães and Portugal (2010) presented an algorithm to estimate linear models with high order fixed effects, using an iterative conditional regression, which is

[1] For details on these methodologies see Andrews, Schank, and Upward (2006).

[2] McCaffrey et al. (2012) presents a review of various commands and strategies created for the statistical software Stata.

later used in Torres, Portugal, Addison, and Guimaraes (2013) to estimate a three way fixed effect model.

The objective of this paper is to provide a feasible methodology that can make the estimation of high order fixed effects models more accessible, and demonstrate the implementation of the methodology using the statistical software Stata. Our methodology is similar to the one suggested in Guimarães and Portugal (2010), but relies on a different theoretical foundation and implements a more intuitive strategy. A simple implementation of this algorithm is presented as an illustration using US Census Current Population Survey (CPS) data.

The rest of the paper is structured as follows. In Section 2, the base model with a one way fixed effect is presented. In section 3, we extend the model to a two-way fixed-effect model, and shows the generalization for three or more fixed effect models. The estimations of standard deviations are discussed in section 4. Section 5 illustrates the implementation of the strategy using CPS data, and we conclude in section 6.

2. ONE FIXED EFFECT MODEL

In the context of an employer-employee linked data, consider the basic model with a single fixed effect:

$$y_{ijk} = a_i + x_{ik}\beta + \varepsilon_{ijk} \qquad (1)$$

Where y_{ijk} represents the outcome of person i, working at firm j at time k, where there is a total of I individuals and J firms across K periods. In this simple model, assume that the outcome of y_{ijt} is a function only of the individual fixed effect a_i, and a set H of observed characteristics x_{ik}, that could vary across individual, firm and/or time. Finally, let ε_{ijt} be a homoscedastic error term with mean zero, and uncorrelated with a_i and x_i.

$$E(\varepsilon_{ijk}|x_{ik}) = 0, E_i(\varepsilon_{ijk}) = 0 \text{ and } corr(\varepsilon_{ijk}, a_i) = corr(\varepsilon_{ijk}, x_{ik}) = 0 \qquad (2)$$

This model can be directly estimated, without estimating the actual individual fixed effects, by subtracting the within person mean from all variables in the model:

$$E(y_{ijk}|i = \mathrm{i}) = i_{\bar{y}} = a_i + i_{\bar{x}}\beta \qquad (3)$$

Where $i_{\bar{y}}$ ($i_{\bar{x}}$) is the within person *i* average across all firms and time periods of variable y (x). Subtracting (3) from (1), we obtain the following transformation of the data (this is referred to as "de-meaning" the data):

$$y_{ijk} - i_{\bar{y}} = (x_{ijk} - i_{\bar{x}})\beta + \varepsilon_{ijk} \tag{4}$$

$$\tilde{y}_{ijk} = \tilde{x}_{ijk}\beta + \varepsilon_{ijk} \tag{5}$$

The latter equations can now be directly estimated using standard ordinary least squares (OLS) procedures. It must be recognized that while the error term ε_{ijk} remains unmodified in equation 5 compared to the original model, the variance-covariance matrix of β ($\sum_{\beta}$) needs to be corrected, to account for the degrees of freedom due to the unestimated fixed effects (to be discussed in section 5).

3. TWO FIXED EFFECT MODEL

Let us now extend the model to allow for two level fixed effects, such that the outcome *y* is a function of the individual fixed effect a_i, and the firm fixed effect b_j where person *i* works at time *k*:

$$y_{ijk} = a_i + b_j + x_{ik}\beta + \varepsilon_{ijk} \tag{6}$$

Similar to the previous case, we assume the error term is well behaved and uncorrelated with the explanatory variables and the firm and individual fixed effects. In this case, if we obtain the within person average and within firm average, we obtain:

$$E(y_{ijk}|i = \mathrm{i}) = i_{\bar{y}} = a_i + i_{\bar{b}_j} + i_{\bar{x}}\beta \tag{7}$$

$$E(y_{ijk}|j = \mathrm{j}) = j_{\bar{y}} = j_{\bar{a}_i} + b_j + j_{\bar{x}}\beta \tag{8}$$

Where $i_{\bar{b}_j}$ is the average firm effect from all the firms where person *i* has ever worked, and $j_{\bar{a}_i}$ is the average individual effect from all individuals who have worked for firm j. In both cases, the averages are weighted by the number of times each worker-employer combination is observed.

Note that while $i_{\bar{b}_j}$ ($j_{\bar{a}_i}$) is fixed within individual i (firms j), it still varies across individuals (firms).[3] From equation 6, we can eliminate part of the impact of the individual and firm fixed effects, by subtracting the means within the group obtained in expression 7 and 8 we obtain:

$$y_{ijk} - i_{\bar{y}} - j_{\bar{y}} = (x_{it} - i_{\bar{x}} - j_{\bar{x}})\beta - j_{\bar{a}_i} - i_{\bar{b}_j} + \varepsilon_{ijk}, \text{ or}$$

$$\tilde{y}_{ijk} = \tilde{x}_{ik}\beta - j_{\bar{a}_i} - i_{\bar{b}_j} + \varepsilon_{ijk} \qquad (9)$$

While the main components of the individual and fixed effects (a_i and b_j) are eliminated in equation 9, some heterogeneity remains in as $j_{\bar{a}_i}$ and $i_{\bar{b}_j}$ vary across firm and persons, respectively. It is possible to eliminate this heterogeneity by continuing to de-mean the variables in the equation 9, obtaining the corresponding averages:

$$E\left(y_{ijk} - i_{\bar{y}} - j_{\bar{y}} | i = \mathrm{i}\right) = i_{\bar{y}} - i_{\bar{y}} - ij_{\bar{y}} = (i_{\bar{x}} - i_{\bar{x}} - ij_{\bar{x}})\beta - ij_{\bar{a}_i} - i_{\bar{b}_j}, \text{ or}$$

$$-ij_{\bar{y}} = (-ij_{\bar{x}})\beta - ij_{\bar{a}_i} - i_{\bar{b}_j} \qquad (10)$$

$$E\left(y_{ijk} - i_{\bar{y}} - j_{\bar{y}} | j = \mathrm{j}\right) = j_{\bar{y}} - ji_{\bar{y}} - j_{\bar{y}} = (j_{\bar{x}} - ji_{\bar{x}} - j_{\bar{x}})\beta - j_{\bar{a}_i} - ji_{\bar{b}_j}, \text{ or}$$

$$-ji_{\bar{y}} = (-ji_{\bar{x}})\beta - j_{\bar{a}_i} - ji_{\bar{b}_j}, \qquad (11)$$

Where, $ji_{\bar{y}}$ is the within firm j average of the average outcomes of individuals i, while $ij_{\bar{y}}$ is the within individual i average of the average outcomes in firm j, both weighted by the number of times each combination is observed. One must note that while the expressions $ji_{\bar{y}}$ and $ij_{\bar{y}}$ look similar, they will only be the same in cases of a balanced panel. Subtracting equations 10 and 11 from equation 9, we can further reduce the individual and firm heterogeneity and obtain the following expression:

$$y_{ijk} - i_{\bar{y}} - j_{\bar{y}} + ij_{\bar{y}} + ji_{\bar{y}} = (x_{ik} - i_{\bar{x}} - j_{\bar{x}} + ij_{\bar{x}} + ji_{\bar{x}})\beta + ji_{\bar{a}_i} + ij_{\bar{b}_j} + \varepsilon_{ijk} \text{ or}$$

$$\tilde{y}_{ijk} - i_{\tilde{y}} - j_{\tilde{y}} = (\tilde{x}_{ik} - i_{\tilde{x}} - j_{\tilde{x}})\beta + ji_{\bar{a}_i} + ij_{\bar{b}_j} + \varepsilon_{ijk}, \text{ or just}$$

$$\tilde{\tilde{y}}_{ijk} = \tilde{\tilde{x}}_{ik}\beta + ji_{\bar{a}_i} + ij_{\bar{b}_j} + \varepsilon_{ijk} \qquad (12)$$

[3] The only case when $i_{\bar{b}_j}$ and $j_{\bar{a}_i}$ are constant across both dimensions (*i* and *j*) is when there are the same number of observations for all combinations of *i* and *j*, equivalent to a balanced panel.

Once again, while the heterogeneity observed in equation 9 is no longer present in equation 12 ($j_{\bar{a}_i}$ and $i_{\bar{b}_j}$), some individual and firm heterogeneity in the form of $ji_{\bar{b}_j}$ and $ij_{\bar{a}_i}$ remains. It can be shown, however, that the variation of the heterogeneity that comes from $ji_{\bar{b}_j}$and $ij_{\bar{a}_i}$ is lower than what was observed previously in b_j and a_i (see the proof in appendix A). Furthermore, if we continue to iteratively de-mean the variables, we can achieve a specification where $ji \dots ji_{\bar{b}_j}$ and $ij \dots ij_{\bar{a}_i}$ tend to a constant, with their variance equal to zero:

$$ij \dots ij_{\bar{a}_i} \cong\ jij \dots ij_{\bar{a}_i} \cong a_i \text{ and } ij \dots ij_{\bar{b}_i} \cong\ jij \dots ij_{\bar{b}_i} \cong b_i, \text{ thus}$$

$$Var\left(ij \dots ij_{\bar{a}_i}\right) = Var\left(ji \dots ji_{\bar{b}_j}\right) \cong 0 \qquad (13)$$

This process effectively eliminates the fixed effects components from the specification.

In a similar manner, the expression $ij \dots ij_{\bar{z}}$ and $ji \dots ji_{\bar{z}}$ will also tend to a constant which is equal to the overall mean $\bar{z}$. At this point, the specification can be written as:

$$\tilde{\dddot{\tilde{y}}}_{ijk} = \tilde{\dddot{\tilde{x}}}_{ijk}\beta + \varepsilon_{ijk} \qquad (14)$$

Where:

$$\tilde{\dddot{\tilde{y}}}_{ijk} = y_{ijk} - i_{\bar{y}} - j_{\bar{y}} + \cdots - iji \dots ji_{\bar{y}} - jij \dots ij_{\bar{y}} + 2\bar{y} \qquad (15a)$$

$$\tilde{\dddot{\tilde{x}}}_{ijk} = x_{ijk} - i_{\bar{x}} - j_{\bar{x}} + \cdots - iji \dots ji_{\bar{x}} - jij \dots ij_{\bar{x}} + 2\bar{x} \qquad (15b)$$

As shown above, equation 14, just like equation 5, can be directly estimated using standard OLS procedures to obtain the unbiased β coefficient.

4. N FIXED EFFECT MODEL

We can now extend the model to allow for N fixed effects. We assume that the outcomes y are a function of a set of H characteristics x, and N fixed effects $n_1, n_2, \dots, n_N$.

$$y = x\beta + \sum_{i=1}^{N} n_i + \varepsilon \qquad (16)$$

As before, we can assume the error ε is well behaved, uncorrelated to all fixed effects and observed characteristics x_i. Following the same strategy used for the two fixed effects model, we first estimate the means with respect to each fixed effect group:[4]

$$E(y|i = \mathrm{i}) = i_{\bar{y}} = \sum_{j \epsilon N} i_{\bar{n}_j} + i_{\bar{x}}\beta, \text{ for all } i = 1,2 \ldots N \tag{17}$$

Subtracting all means in equation 17 from 16, we start to eliminate the variation coming from N fixed effects. Analogous to what was previously seen, however, some heterogeneity will remain from the averaged fixed effects ($i_{\bar{n}_j}$ for $i \neq j$, and $i, j = 1, \ldots, N$).

$$y - \sum_{j=1}^{N} j_{\bar{y}} = \left(x - \sum_{j=1}^{N} j_{\bar{x}}\right)\beta - \sum_{j \neq 1} 1_{\bar{n}_j} - \sum_{j \neq 2} 2_{\bar{n}_j} - \cdots \sum_{j \neq N} N_{\bar{n}_j} + \varepsilon, \text{ or}$$

$$\tilde{y} = \tilde{x}\beta - \sum_{i=1}^{N} \sum_{j \neq i} i_{\bar{n}_j} + \varepsilon \tag{18}$$

Following a strategy similar to the two fixed effect case, we can attempt to eliminate the fixed effects from equation 18, by obtaining the corresponding averages:

$$k_{\tilde{y}} = k_{\tilde{x}}\beta - \sum_{i=1}^{N} \sum_{j \neq i} k i_{\bar{n}_j} + \varepsilon \text{ , for all } k = 1,2 \ldots N \tag{19}$$

Using each of the group averages, we proceed to subtract 19 from 18, in order to eliminate the fixed effects from an iterative de-meaning process. Just as in the two fixed effect case, the transformation will steadily eliminate the influence of the fixed effects from the variables. After multiple iterations, we can obtain a specification similar to equation 14:

$$\tilde{\dddot{\tilde{y}}} = \tilde{\dddot{\tilde{x}}}\beta + \varepsilon \tag{20}$$

Where y and x are defined as:

$$\tilde{\dddot{\tilde{y}}} = y - \sum_{i=1}^{N} i_{\bar{y}} + \sum_{i=1}^{N} \sum_{k \neq i} k i_{\bar{y}} - \cdots$$

$$- \sum_{i=1}^{N} \sum_{k \neq i} \cdots \sum_{g \neq h} gh \ldots k i_{\bar{y}} + N\bar{y} \tag{21}$$

[4] Note that $i_{\bar{n}_i} = n_i$, and that $ii_{\bar{n}_j} = i_{\bar{n}_j}$.

Here, the remaining effect of the fixed effect will be negligible, and equation 20 can be estimated using the transformed data to obtain the unbiased β coefficients.

5. ESTIMATION OF THE STANDARD ERRORS

Up to this point, the discussion in the previous section has focused on obtaining a specification that allows for the estimation of unbiased β coefficients after accounting for all fixed effects. As seen in equations 6, 14 and 20, even after transforming the variables, the error term remains unchanged, and can be used to estimate the variance-covariance matrix.

From equation 20, after eliminating the influence of the fixed effects, the corresponding variance-covariance matrix associated with the coefficients β would be:

$$Var(\tilde{\beta}) = \frac{\sum \varepsilon^2}{N-k} * \left(\ddddot{\tilde{\tilde{x}}}' \ddddot{\tilde{\tilde{x}}}\right)^{-1} \tag{22}$$

Because the vector of variables $\ddddot{\tilde{\tilde{x}}}$ is orthogonal to the individual and fixed effects, thus already taking into account the absence of the dummy cross-products in the inverted matrix, the main difference with the estimation of the original specification is the number of degrees of freedom. In the original model estimates, if we were able to estimate it, it would require the estimation of H parameters for each variable in *x*, plus up to $N_1 + N_2 + \cdots N_N$ fixed effects (or I+J in the two fixed effect case). As noted by Abowd et al. (1999), not all fixed effects can be estimated, as there might not be enough observations to fully identify the fixed effects.

In Abowd, Creecy, and Kramarz (2002), an algorithm is presented to identify “mobility groups” for the case of two fixed effects (firms-workers). These mobility groups represent the number of parameters among the fixed effects that cannot be identified, nor estimated in the original model. For the case of three or more fixed effects, there is no exact solution to estimate the total number of unidentifiable parameters in the system. A modification to the algorithm presented in Abowd et al. (2002) is proposed here to find an approximation of the number of unidentifiable parameters in the model.

We will assume the model has N fixed effects (equation 16), and call these fixed effects n_1, n_2 up to n_N.

1. Using groups n_1 and n_2, identify the number of mobility groups using the algorithm in Abowd et al. (2002). Call the number of mobility groups G_1.

2. Create a new index identifying all the interactions of n_1 and n_2, and call it $n_{1,2}$.
3. Using groups $n_{1,2}$ and n_3, identify the number of mobility groups G_2.
4. Repeat 2 and 3 until all fixed effects are used.

The total number of unidentifiable parameters will be $G = \sum_{i=1}^{N-1} G_i$. The algorithm presented above provides a good empirical approximation for the true number of unidentified parameters G. If the results are invariant to the order of fixed effects, G is the exact number of unidentified parameters. On the other hand, if the results vary with respect to the order of fixed effects used, the estimated G becomes an upper bound for the number of unidentifiable parameters.

Once G is estimated, the variance-covariance matrix can be corrected using the correct degrees of freedom as follows:

$$Var(\hat{\beta}) = \frac{\sum \varepsilon^2}{N-\sum N_i+G} * \left(\ddot{\tilde{\tilde{x}}}'\ddot{\tilde{\tilde{x}}}\right)^{-1} \qquad (23)$$

6. ALGORITHM AND ILLUSTRATION

In this section, we present the implementation of the algorithm proposed in the previous section using a sample obtained from the basic CPS monthly survey from 2007 and 2008. The implementation was done using the statistical software Stata v12, using a Xeon CPU 1.8GHz, and 8GB of memory. While the dataset does not provide the richness and complexity typically found in employer-worker linked data or school-teacher data, it still illustrates how the strategy can be used to estimate a linear model with multiple fixed effects.[5]

We assume a simple wage model:

$$\ln w = \alpha + \beta' X + ind + occ + yrm + st + e \qquad (24)$$

Where hourly log wages ($\ln w$) are a function of age, sex, years of education achieved, union status (*X*) , and fixed effects depending on the industry (*ind*) and occupation (*occ*) of the workers, the year*month of the survey (*yrm*), and the state where they are from (*st*). As a

[5] The program used to obtain the results presented in this section can be found in appendix B.

benchmark, the model is estimated using dummy sets to capture the four fixed effects, as is typically done (column 1 in Table 1).

Under the assumption that the previous model is correctly specified, if we were to ignore the fixed effects in the specification, this would generate a bias on all parameters. In this example, the estimates omitting the fixed effects indicate that while the estimation of the union wage gap and age return are similar to the benchmark, returns of education and sex are larger (column 2).

We now proceed to transform the data. The algorithm is an iterative process where each subsequent mean calculation and transformation is applied to the previously transformed data. The algorithm used for the estimation procedure can be described as follows:

1. Identify all variables of interest in the model. (i.e. wages, unions, years of education, sex and age).
2. Identify all fixed effect variables not to be estimated in the model. (i.e. industry, occupation, state and year*month).
3. For each of the variables of interest *i*:
    a. Replace variable *i* with the demeaned variable *i* with respect to the first fixed effect variable.
    b. Repeat the de-meaning process of variable *i*, using each fixed effect variable *k*.
4. Estimate the OLS model using the de-meaned data.
5. Check for changes on the root mean squared error of the model estimated in step 4. If there are no (significant) changes, finish the process and provide estimates; if not, repeat steps 3 and 4.

This procedure is a modification of the base algorithm described in section 4 that facilitates the programming. This alternative algorithm centers all variables at zero, so the models should be estimated with no constant. For this example, the process is repeated until the model shows no significant improvements using double precision.[6]

As we can see, the parameter estimates of the transformed data are the correct ones, as they match those of the benchmark model (column 3). As discussed in section 5, however, the

[6] While the algorithm estimates the regression on every stage, it can be modified to check convergence on individual parameters, or every certain number of “steps”. To obtain exact values, we use double precision criteria, however, with parameters as large as 0.001 the estimates are close to the benchmark ones.

standard errors of the model need to be corrected to match those of the benchmark. For this purpose, we estimate the number of unidentifiable parameters by applying the algorithm described in section 5. The results are shown in column 4 of Table 1.

Table 1. Alternative specifications of wage equation model

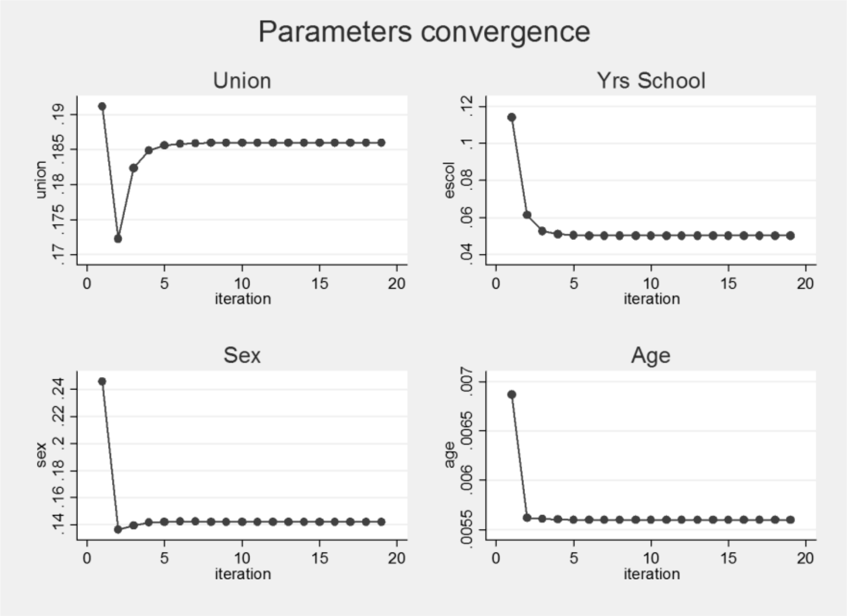
	OLS with FE	OLS no FE	OLS - Trans data	OLS - Trans data, correct
union	0.185947	0.191169	0.185947	0.185947
	(0.004203)	(0.004867)	(0.004193)	(0.004203)
escol	0.050235	0.114177	0.050235	0.050235
	(0.000516)	(0.000485)	(0.000515)	(0.000516)
sex	0.142023	0.245948	0.142023	0.142023
	(0.002561)	(0.002523)	(0.002555)	(0.002561)
age	0.005598	0.006870	0.005598	0.005598
	(0.000082)	(0.000100)	(0.000082)	(0.000082)
N	165439	165439	165439	165439

For this particular example, the algorithm takes 62 iterations before it achieves convergence using double precision, taking a total of 1:10 minutes. This is faster than the algorithm used in *gpreg*.[7] It is unclear if the same increase in processing speed would be observed for larger datasets. The comparability of the algorithms is partial, as there are no alternative commands that deal with more than 2 fixed effects without introducing the additional fixed effects as explicit dummy sets in the models.[8] As it can be appreciated in figure 1, the parameters are very close to the true ones after less than 10 iterations. This should be taken into consideration, bearing in mind that the processing time could increase geometrically for larger datasets and more complex specifications.

[7] The command *gpreg,* which is also described to be computationally intensive, takes 2:24 min to estimate the model.

[8] As shown in McCaffrey et al. (2012), commands like *areg*, *xtreg*, and *felsdvreg* perform better for models with fewer levels on the second fixed effect. In this case, *gpreg* is slower, as it needs to estimate more parameters for the additional fixed effects.

Figure 1: Parameters Convergence by Iteration



7. CONCLUSIONS

In this paper an alternative methodology for estimating linear models with N-fixed effects model has been presented, which uses an intuitive strategy already used for cases with fewer fixed effects. While other alternatives have been suggested in the literature, none of them present feasible alternatives for estimating linear models with more than two fixed effects. The closest proposition to solve this problem has been Guimarães and Portugal (2010) and its application in Torres et al. (2013), which does not elaborate on the details of the expansion to N-fixed effects.

While this strategy is computationally demanding, the ability to freely exchange fixed effects from the explanatory variables to the fixed effects sets allows to better control the memory requirements for the estimation of the models. For instance in the paper by Hotchkiss, Quispe and Rios-Avila (2013), this methodology is used to estimate a fixed effect model with 4 fixed effects 3,376,102 workers, 93,021 firms, 40 quarters, and 159 counties, a model that could not be estimated using other available strategies.

REFERENCES

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

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). "High Wage Workers and High Wage Firms." *Econometrica, 67*(2), 251-333. doi: 10.2307/2999586

Abowd, J. M., Kramarz, F., & Woodcock, S. (2008). "Econometric Analyses of Linked Employer–Employee Data." In L. Mátyás & P. Sevestre (Eds.), *The Econometrics of Panel Data* (Vol. 46, pp. 727-760): Springer Berlin Heidelberg.

Andrews, M., Schank, T., & Upward, R. (2006). "Practical Fixed-effects Estimation Methods for the Three-way Error-components Model." *Stata Journal, 6*(4), 461-481.

Cameron, A. C. (2005). *Microeconometrics: Methods and Applications*: Cambridge University Press.

Guimarães, P., & Portugal, P. (2010). "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects." *Stata Journal, 10*(4), 628-649.

Harris, D. N., & Sass, T. R. (2011). "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics, 95*(7–8), 798-812.

Hotchkiss, J., Quispe-Agnoli, M., & Rios-Avila, F. (Forthcoming). *The Wage Impact of Undocumented Workers: Evidence from Administrative Data*.

McCaffrey, D. F., Lockwood, J. R., Mihaly, K., & Sass, T. R. (2012). "A Review of Stata Commands for Fixed-effects Estimation in Normal Linear Models." *Stata Journal, 12*(3), 406-432.

Torres, S., Portugal, P., Addison, J. T., & Guimaraes, P. (2013). "The Sources of Wage Variation: A Three-Way High-Dimensional Fixed Effects Regression Model." *IZA Discussion Papers*(7276).

Appendix A

Let y_i be a variable with an overall mean $\bar{y}$ and variance σ_y^2. Without loss of generality, assume that the i*th* means of y_i ($i_{\bar{y}}$) are all different from each other, i.e. $\sigma_{i_{\bar{y}}}^2 \neq 0$.

The variance of variable *y* can then be written as:

$$\sigma_y^2 = var(y_i) = E[(y_i - \bar{y})^2] \qquad (1A)$$

Maintaining the equality on the expression, we can add and subtract the i*th* mean to expression in parenthesis, obtaining the alternative variance expression:

$$\sigma_y^2 = E\left[\left(y_i - i_{\bar{y}} + i_{\bar{y}} - \bar{y}\right)^2\right] \qquad (2A)$$

Expanding this expression, we obtain:

$$\sigma_y^2 = E\left[\left(y_i - i_{\bar{y}}\right)^2 + \left(i_{\bar{y}} - \bar{y}\right)^2 + 2\left(y_i - i_{\bar{y}}\right)\left(i_{\bar{y}} - \bar{y}\right)\right]$$
$$\sigma_y^2 = E\left[\left(y_i - i_{\bar{y}}\right)^2\right] + E\left[\left(i_{\bar{y}} - \bar{y}\right)^2\right] + 2E\left[\left(y_i - i_{\bar{y}}\right)\left(i_{\bar{y}} - \bar{y}\right)\right]$$
$$\sigma_y^2 = E\left[\left(y_i - i_{\bar{y}}\right)^2\right] + E\left[\left(i_{\bar{y}} - \bar{y}\right)^2\right] + 2\left[E\left(y_i i_{\bar{y}} - i_{\bar{y}}^2\right)\right] \qquad (3A)$$

Using iterative expectations, the third term of the expression is equal to zero.

$$E\left(y_i i_{\bar{y}} - i_{\bar{y}}^2\right) = E\left[E\left(y_i i_{\bar{y}} - i_{\bar{y}}^2 | i = i\right)\right] = 0$$

Thus:

$$\sigma_y^2 = E\left[\left(y_i - i_{\bar{y}}\right)^2\right] + E\left[\left(i_{\bar{y}} - \bar{y}\right)^2\right] \rightarrow \ \sigma_y^2 = \sigma_{y-i_{\bar{y}}}^2 + \sigma_{i_{\bar{y}}}^2 \qquad (4A)$$

Finally, the overall variance of y can be decomposed into two components, one corresponding to the within variation $\sigma_{y-i_{\bar{y}}}^2$, and one corresponding to the across variation $\sigma_{i_{\bar{y}}}^2$. Given that all variances must be positive by construction, this implies that the variance of de-meaned data, and the within group means are smaller than that of the original data.

In a similar matter, we can further decompose the across individual variance $\sigma_{i_{\bar{y}}}^2$, respect to an alternative subgroup. Say we decompose respect to groups *j*, it follows that:

$$\sigma^2_{i_{\bar{y}}} = \sigma^2_{i_{\bar{y}}-ji_{\bar{y}}} + \sigma^2_{ji_{\bar{y}}}$$

Using the same strategy, with iterative decompositions, it follows that variance of subsequent transformations and tends to zero:

$$\sigma^2_y > \sigma^2_{i_{\bar{y}}} > \sigma^2_{ji_{\bar{y}}} > \cdots > \sigma^2_{ij\ldots ji_{\bar{y}}} \cong 0$$

Appendix B

The following the code was used for the estimation of the results in this paper. It provides one possible implementation of the algorithm described in this paper.

```
set more off
clear all
use C:\Users\user\Downloads\cps_sample , clear

* benchmark
codebook ind occ state yrm
recast double lnwageh union escol sex age ind occ state yrm
set matsize 1000
reg lnwageh union escol sex age i.ind i.occ i.state i.yrm
est sto eq1
reg lnwageh union escol sex age
est sto eq2

global varXY union escol sex lnwageh age
global ffee ind occ state yrm
* Data is itself modified
local a0=0
local a1=10
display "$S_TIME"
global t0="$S_TIME"
while abs(`a0'-`a1')>epsfloat(){
local a0=`a1'
qui: reg lnwageh union escol sex age,
matrix b=nullmat(b) \ e(b)
local a1=e(rmse)
qui: foreach h of global ffee {
  foreach i of global varXY  {
   capture drop i_
   bysort `h':egen double i_=mean(`i')
   replace `i'=`i'-i_
  }
 }
}

reg lnwageh union escol sex age, nocons
est sto eq3

*** estimating unidentified parameters
*Nr of Fixed effects:4
#delimit;
gen id0=_n;gen id1=_n;gen id2=_n;gen id3=_n;gen flag=1;
#delimit cr
egen fe1= group(ind)
egen fe2= group(occ)
egen fe3= group(state)
egen fe4= group(yrm)
egen fe1_2=group(ind occ)
egen fe1_2_3=group(ind occ state)
```

```
while flag {
 capture drop id0a id1a id2a id0b id1b id2b
 bysort fe1:egen  id0a=min(id0)
 bysort fe1_2:egen  id1a=min(id1)
 bysort fe1_2_3:egen  id2a=min(id2)
 bysort fe2:egen  id0b=min(id0a)
 bysort fe3:egen  id1b=min(id1a)
 bysort fe4:egen  id2b=min(id2a)
 count if ![(id0==id0b)&(id1==id1b)&(id2==id2b)]
 if r(N)==0 {
   replace flag=0
 }
 replace id0=id0b
 replace id1=id1b
 replace id2=id2b
}
scalar G=0
foreach h in id0 id1 id2 {
capture drop _k
egen _k=group(`h')
sum _k
scalar G=G+r(max)
}
** TOtal parameters among FE
scalar DF=0
foreach h of global ffee {
 capture drop _k
 egen _k=group(`h')
 sum _k
 scalar DF=DF+r(max)
}
* Correcting for DF
ssc install erepost
reg lnwageh union escol sex age, nocons
matrix V=e(V)*e(df_r)/(e(df_r)-DF+G)
erepost V=V
display "$S_TIME"
global t1="$S_TIME"
ereturn display
est sto eq4
est tab eq1 eq2 eq3 eq4, se keep(union escol sex age ) se(%6.5f) b(%6.5f)
```