Estimating High-Dimensional Fixed-Effects Models

Paulo Guimarães

guimaraes@moore.sc.edu

University of South Carolina
Motivation

- Data sets are getting larger.
- Estimation of models with many observations and variables poses new challenges.
- A case in point is estimation of models with high-dimensional fixed effects.
- With high-dimensional models explicit introduction of dummy variables to account for fixed effects is not an option.
- With one fixed effect there are other solutions:
  - Condition out the fixed effects (eg: linear regression, poisson, logistic regression)
  - use a modified iterative algorithm for ML maximization (see Greene(2004))
Our problem

In Carneiro, Guimaraes and Portugal (2009) we had a linked employer-employee panel data set with 26 millions observations.

Our objective was:

- To estimate a linear regression model with 26 variables plus two fixed effects (firm and worker).
- To obtain estimates of the fixed effects.

With 541,229 firms and 7,155,898 workers introduction of dummy variables was not an option.

The user written commands `a2reg` (A. Ouazad) and `felsdvreg` (T. Cornelissen) aborted due to memory problems in a Windows machine with 8G RAM running Stata MP.

We developed an alternative estimation strategy.
The Linear Regression

- Consider the linear model $\mathbf{Y} = \mathbf{X}\beta + \epsilon$

- Minimization of the sum of squares (SS) results in a set of equations:

$$\begin{align*}
\frac{\partial SS}{\partial \beta_1} &= \sum_i x_{1i}(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}) = 0 \\
\frac{\partial SS}{\partial \beta_2} &= \sum_i x_{2i}(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}) = 0 \\
&\quad \vdots \\
\frac{\partial SS}{\partial \beta_k} &= \sum_i x_{ki}(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}) = 0
\end{align*}$$

- These equations can easily be solved using

$$\mathbf{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$
The Linear Regression

An alternative approach: the partitioned ("cyclic-ascent" or "zigzag") algorithm:

1. Initialize $\beta_1^{(0)}, \beta_2^{(0)}, \ldots, \beta_k^{(0)}$

2. Solve for $\beta_1^{(1)}$ as the solution to
   \[
   \frac{\partial S}{\partial \beta_1} = \sum_i x_{1i}(y_i - \beta_1 x_{1i} - \beta_2^{(0)} x_{2i} - \ldots - \beta_k^{(0)} x_{ki}) = 0
   \]

3. Solve for $\beta_2^{(1)}$ as the solution to
   \[
   \frac{\partial S}{\partial \beta_2} = \sum_i x_{2i}(y_i - \beta_1^{(1)} x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k^{(0)} x_{ki}) = 0
   \]

3. and so on...

4. Repeat until convergence.
Suppose we have a fixed effect: \( Y = X\beta + D\alpha + \epsilon \)

where \( X \) is \( n \times k \) and \( D \) is a \( n \times G_1 \) matrix of "dummies" and \( G_1 \) is a large number.

The normal equations are:

\[
\begin{bmatrix}
X'X & X'D \\
D'X & D'D
\end{bmatrix}
\begin{bmatrix}
\beta \\
\alpha
\end{bmatrix}
= 
\begin{bmatrix}
X'Y \\
D'Y
\end{bmatrix}
\]

\[
\begin{align*}
X'X\beta + X'D\alpha &= X'Y \\
D'X\beta + D'D\alpha &= D'Y
\end{align*}
\]

\[
\begin{align*}
\beta &= (X'X)^{-1}X'(Y - D\alpha) \\
\alpha &= (D'D)^{-1}D'(Y - X\beta)
\end{align*}
\]
One Fixed Effect

This suggests the following "zigzag" estimation procedure:

\[
\begin{bmatrix}
\beta^{(j+1)} = (X'X)^{-1} X' \left( Y - D\alpha^{(j)} \right) \\
\alpha^{(j)} = (D'D)^{-1} D' \left( Y - X\beta^{(j)} \right)
\end{bmatrix}
\]

The key insight is that \( \eta = D\alpha \) is \( n \times 1 \).

The "zigzag" approach involves running several regressions with \( k \) explanatory variables (1st equation) and repeatedly computing means of residuals (2nd equation).

The variable \( \eta \) contains the estimated fixed effects and if added as a regressor will give the same SS as in a model with the fixed-effects.
One Fixed Effect

Note that

\[ \begin{bmatrix} \beta_{(j+1)} \\ \alpha_{(j)} \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X' \left( Y - D\alpha_{(j)} \right) \\ (D'D)^{-1}D' \left( Y - X\beta_{(j)} \right) \end{bmatrix} \]

The key insight is that \( \eta = D\alpha \) is \( n \times 1 \).

The "zigzag" approach involves running several regressions with \( k \) explanatory variables (1st equation) and repeatedly computing means of residuals (2nd equation).

The variable \( \eta \) contains the estimated fixed effects and if added as a regressor will give the same SS as in a model with the fixed-effects.
One Fixed Effect - Example

- Estimation of a linear regression with one fixed effect.
- See EXAMPLE1.
Suppose we have two fixed effects:

\[ Y = X\beta + D_1\alpha + D_2\gamma + \epsilon \]

\(D_1\) is \(n \times G_1\) and \(D_2\) is \(n \times G_2\) and both \(G_1\) and \(G_2\) are large numbers.

Estimation of this model is complicated. See Abowd, Kramarz and Margolis (Ectrca 1999).

A "zigzag" approach is simple to implement.

\[
\begin{align*}
\beta^{(j+1)} &= (X'X)^{-1}X' \left( Y - D_1\alpha^{(j)} - D_2\gamma^{(j)} \right) \\
\alpha^{(j)} &= (D_1'D_1)^{-1}D_1' \left( Y - X\beta^{(j)} - D_2\gamma^{(j)} \right) \\
\gamma^{(j)} &= (D_2'D_2)^{-1}D_2' \left( Y - X\beta^{(j)} - D_1\gamma^{(j)} \right)
\end{align*}
\]
Two Fixed Effects

- The final linear regression (with the two fixed effects variables) has the right SS.
- This means that we can estimate $\sigma^2$ if we can figure out the degrees of freedom.
- Because some coefficients of the fixed effects are not identifiable we need to use $N - k - G_1 - G_2 + M$ where $M$ is the number of mobility groups (see Abowd et al 2002).

- To estimate $V(\hat{\beta}_j)$ we can use:

$$V(\hat{\beta}_j) = \frac{\sigma^2}{Ns_j^2(1 - R_{j.123}^2)}$$
Two Fixed Effects

In practical applications it may make more sense to estimate in steps using the Frisch-Waugh-Lovell theorem.

First remove the effects of $D_1$ and $D_2$ from $Y$ and $X$.
Then regress the transformed $Y$ on the transformed $X$ to obtain the estimates for $\beta$.
Then (if needed) recover the estimates of the fixed effects by regressing $u = Y - X\beta$ on $D_1$ and $D_2$.

Regressions on $D_1$ and $D_2$ are fast because they only require computation of means.
We can sweep out one of the fixed effects by demeaning the variables.
Estimates a linear regression with two fixed effects
Check EXAMPLE2

A faster approach to the same problem
Check EXAMPLE3
Two command gpreg

The command `gpreg` programmed by Johannes F. Schmieder implements the two-step approach for estimation of linear regression models with two high dimensional fixed effects.

Command Syntax:
```
gpreg depvar indepvars [if] [in] ,
ivar(varname) jvar(varname) [ ife(new varname) jfe(new varname) 
maxiter(integer) tolerance(float) nodots 
Algorithm(integer) ]
```

There are 4 options for choice of algorithm 2 of them implemented in Mata.

`gpreg` is available on the SSC server
Non-linear Models: Poisson

- This approach can be extended to non-linear models.
- An example with Poisson regression:

\[ E(y_i) = \lambda_i = \exp(x_i' \beta + \alpha_1 d_{1i} + \alpha_2 d_{2i} + \ldots + \alpha_J d_{Ji}) \]

- Using the first order conditions:

\[ \exp(\alpha_j) = d_j'y \times [d_j' \exp(x_i' \beta)]^{-1} \]

- Optimization of the maximum-likelihood function requires recursive estimation of a Poisson regression with the \(x\) variables and an offset containing the estimates \(\alpha\) obtained from the expression above.
Non-linear Models: Examples

- A Poisson regression with one fixed effect
  see EXAMPLE4

- A Poisson regression with two fixed effects
  see EXAMPLE5

- A Negative Binomial regression with one fixed effect
  see EXAMPLE6
Final Remarks

- The main advantage of this approach is that it does not require much memory.
- The approach can be extended to non-linear models.
- The approach can be extended to 3 or more high-dimensional fixed effects.
- This approach tends to be slow but there is room for improvement.
- This presentation is based in: Guimaraes and Portugal (2009), "A simple feasible alternative procedure to estimate models with high-dimensional fixed-effects" IZA Discussion Papers 3935.