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:

$$\left[\begin{array}{c} \frac{\partial SS}{\partial \beta_1} = \sum_i x_{1i}(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}) = 0 \\ \frac{\partial SS}{\partial \beta_2} = \sum_i x_{2i}(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}) = 0 \\ \ldots \\ \frac{\partial SS}{\partial \beta_k} = \sum_i x_{ki}(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}) = 0 \end{array}\right]$$

- These equations can easily be solved using

$$\widehat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

The Linear Regression

- An alternative approach: the partitioned ("cyclic-ascent" or "zigzag") algorithm:
  - 1. Initialize $\beta_1^{(0)}, \beta_2^{(0)}, ..., \beta_k^{(0)}$
  - 2. Solve for $\beta_1^{(1)}$ as the solution to $\frac{\partial SS}{\partial \beta_1} = \sum_i x_{1i}(y_i - \beta_1 x_{1i} - \beta_2^{(0)} x_{2i} - ... - \beta_k^{(0)} x_{ki}) = 0$
  - 2. Solve for $\beta_2^{(1)}$ as the solution to $\frac{\partial SS}{\partial \beta_2} = \sum_i x_{2i}(y_i - \beta_1^{(1)} x_{1i} - \beta_2 x_{2i} - ... - \beta_k^{(0)} x_{ki}) = 0$
  - 3. and so on...
  - 4. Repeat until convergence.

Linear Regression - One Fixed Effect

- Suppose we have a fixed effect: $\mathbf{Y} = \mathbf{X}\beta + \mathbf{D}\alpha + \epsilon$
- where $\mathbf{X}$ is $n \times k$ and $\mathbf{D}$ is a $n \times G_1$ matrix of "dummies" and G_1 is a large number.
- The normal equations are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{D}'\mathbf{Y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X}\beta + \mathbf{X}'\mathbf{D}\alpha = \mathbf{X}'\mathbf{Y} \\ \mathbf{D}'\mathbf{X}\beta + \mathbf{D}'\mathbf{D}\alpha = \mathbf{D}'\mathbf{Y} \end{bmatrix}$$

$$\begin{bmatrix} \beta = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{Y} - \mathbf{D}\alpha\right) \\ \alpha = \left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}'\left(\mathbf{Y} - \mathbf{X}\beta\right) \end{bmatrix}$$

One Fixed Effect

- This suggests the following "zigzag" estimation procedure:

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

- The key insight is that $\eta = \mathbf{D}\alpha$ 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 η 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

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

- The key insight is that $\eta = \mathbf{D}\alpha$ 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 η 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.

Linear Regressions - Two Fixed Effects

- Suppose we have two fixed effects:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{D}_1\alpha + \mathbf{D}_2\gamma + \epsilon$$

- $\mathbf{D}_1$ is $n \times G_1$ and $\mathbf{D}_2$ 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

$$\left[\begin{array}{c} \beta^{(j+1)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{Y} - \mathbf{D}_1\alpha^{(j)} - \mathbf{D}_2\gamma^{(j)}\right) \\ \alpha^{(j)} = (\mathbf{D}_1'\mathbf{D}_1)^{-1}\mathbf{D}_1'\left(\mathbf{Y} - \mathbf{X}\beta^{(j)} - \mathbf{D}_2\gamma^{(j)}\right) \\ \gamma^{(j)} = (\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'\left(\mathbf{Y} - \mathbf{X}\beta^{(j)} - \mathbf{D}_1\gamma^{(j)}\right) \end{array}\right]$$

Two Fixed Effects

- The final linear regression (with the two fixed effects variables) has the right SS.
- This means that we can estimate σ^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 $V(\widehat{\beta}_j)$ we can use:

$$V(\widehat{\beta}_j) = \frac{\sigma^2}{N s_j^2 (1 - R^2_{j.123...})}$$

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 $\mathbf{D}_1$ and $\mathbf{D}_2$ from $\mathbf{Y}$ and $\mathbf{X}$.
  - Then regress the transformed $\mathbf{Y}$ on the transformed $\mathbf{X}$ to obtain the estimates for β.
  - Then (if needed) recover the estimates of the fixed effects by regressing $\mathbf{u} = \mathbf{Y} - \mathbf{X}\beta$ on $\mathbf{D}_1$ and $\mathbf{D}_2$.
- Regressions on $\mathbf{D}_1$ and $\mathbf{D}_2$ are fast because they only require computation of means.
- We can sweep out one of the fixed effects by demeaning the variables.

Two Fixed Effects - Examples

- 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.
- `gppreg` is available on the SSC server

Non-linear Models: Poisson

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

$$E(y_i) = \lambda_i = \exp(\mathbf{x}_i'\beta + \alpha_1 d_{1i} + \alpha_2 d_{2i} + ... + \alpha_J d_{Ji})$$

- Using the first order conditions:

$$\exp(\alpha_j) = \mathbf{d}_j'\mathbf{y} \times [\mathbf{d}_j' \exp(\mathbf{x}_i'\beta)]^{-1}$$

- Optimization of the maximum-likelihood function requires recursive estimation of a Poisson regression with the $\mathbf{x}$ variables and an offset containing the estimates α 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.