Introduction to reghdfe and Conjugate Gradient Descent

Building a Robot Judge - TA Session 3

References form Last TA

Multicollinearity:

- *Introductory Econometrics - A Modern Approach*, by Jeffrey Wooldridge
  - Comments on multicollinearity: Chapter 3.4, 4.5
  - Extreme case of perfect collinearity: assumption MLR.3 in Chapter 3.3
- *Econometric Analysis*, by William Greene
  - Chapter 4.9.1

Advantages of reghdfe

- Allows to run regressions with multiple levels of fixed effects (e.g. in a model of health insurance include both insurance companies and buyers fixed effects)
- Supports different slopes per individuals
- Uses Conjugate Gradient Descent (CGD) $\rightarrow$ allows to produce results faster than other convergence methods
- Supports Instrumental Variable estimation
- Allows for two-way clustering
- Automatically drops singletons observations

Multi-Way Fixed Effects

Consider the following model and suppose that we want to compute $\hat{\beta}$

$$y = X\beta + D\alpha + \varepsilon$$

- y: vector of outcome observations
- X: matrix of explanatory variables
- β: vector of coefficients
- $D = [D_1 ... D_F]$ vector of F indicator matrices, each corresponding to a different level of fixed effects

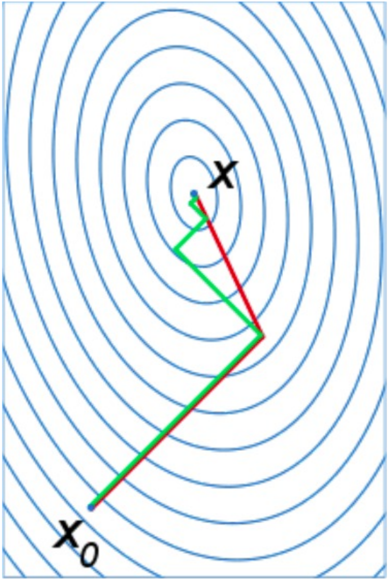
Problem: Cannot estimate including dummies because if D is too large ($F > 1 \implies$ more variables than observations)

Solution Algorithm

1. Apply the Frisch-Waugh-Lovell theorem
   a. Compute the residual ($\tilde{y} = y - \hat{y}$) of regressions of y and each X with respect to each of the fixed effects ($\tilde{y}$ and $\tilde{X}$)
   b. Get the least square estimate $\hat{\beta}$ by regressing $\tilde{y}$ on $\tilde{X}$
2. Apply method of alternating projection (MAP) to compute $\tilde{y}$ and $\tilde{X}$. Consider the dependent variable y
   a. Regress y, if $i = 0$, or $\tilde{y}_1$, if $i > 0$, on one of the fixed effects dimension and get the residuals
   b. Regress the obtained residuals on the next fixed effect and get residuals, and so on for all fixed effects to obtain $\tilde{y}_i$
   c. Check distance between $\tilde{y}_{i-1}$ and $\tilde{y}_i$
   d. Repeat points a. to c. until the estimated residual converge to the partialled-out estimate
3. Accelerate the algorithm with conjugate gradient

CGD - Conjugate Direction Method

- Algorithm for numerical solution of linear equations with symmetric matrices
- Algorithm for seeking minima of nonlinear equations
- Iterative method for systems too large to be solved in closed forms
- Conjugate vectors converge much faster than gradient descent (red line is the conjugate vector, green line is the gradient descent with optimal step size)



CGD - Conjugate Direction

Definition

Let Q be a symmetric matrix. $\{d_1, d_2, ...d_k\}$ vectors ($d_i \in \mathbb{R}^n$, $d_i \neq 0$) are Q-conjugate w.r.t. Q if

$$d_i' Q d_j = 0, \quad \forall i \neq j$$

Particular definition of orthogonality (recall the definition of orthogonality entails $d_i' d_j = 0$)

Lemma

Let Q be positive definite. If $\{d_1, d_2, ...d_k\}$ vectors are Q-conjugate, then they are linearly independent.

CDG - Example I

Suppose we have the following problem to solve (let $Q = X'X \in \mathbb{R}^{n \times n}$ be positive definite and $S = YX'$)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\beta' Q \beta - S'\beta$$

The solution to the problem $\hat{\beta}$ satisfies the following

$$Q\beta = S$$

Consider the Q-conjugate vectors $\{d_0, d_1, ...d_{n-1}\}$, which are linearly independent. Thus, by linear independence

$$\hat{\beta} = \alpha_0 d_0 + ... + \alpha_{n-1} d_{n-1}$$

CDG - Example II

$$
\begin{aligned}
d_i' Q\hat{\beta} &= d_i' Q(\alpha_0 d_0 + ... + \alpha_{n-1} d_{n-1}) = \alpha_i d' Q d_i \\
\implies \alpha_i &= \frac{d_i' Q\hat{\beta}}{d_i' Q d_i} = \frac{d_i' S}{d_i' Q d_i} \\
\implies \hat{\beta} &= \sum_{i=0}^{n-1} \alpha_i d_i = \sum_{i=0}^{n-1} \frac{d_i' S}{d_i' Q d_i} d_i
\end{aligned}
$$

Note that

- We do not need $\hat{\beta}$ to calculate α_i
- We can find the solution without inverting the matrix!

Conjugate Gradient Algorithm I

Algorithm to get convergence to the solution. Let $\{d_0, d_1, ...d_{n-1}\}$ vectors be Q-conjugate and $\beta_0 \in \mathbb{R}^n$ an arbitrary starting point. Let

$$\begin{aligned} \beta_{k+1} &= \beta_k + \alpha_k d_k \qquad \text{update rule} \\ g_k &= Q\beta_k - S \qquad \text{gradient of the function} \\ \alpha_k &= -\frac{g_k' d_k}{d_k' Q d_k} = -\frac{(Q\beta_k - S)' d_k}{d_k' Q d_k} \end{aligned}$$

then after n steps $\beta_n \to \hat{\beta}$.

Conjugate Gradient Algorithm II

How do we choose $\{d_0, d_1, ...d_{n-1}\}$?

- Conjugate direction vectors are determined sequentially as a conjugate version of the successive gradients obtained
- At each step d_{k+1} is chosen in the following way

$$
\begin{aligned}
g_{k+1} &= Q\beta_{k+1} - S \\
d_{k+1} &= -g_{k+1} + \delta_k d_k \\
\delta_k &= \frac{g'_{k+1} Q d_k}{d'_k Q d_k}
\end{aligned}
$$

Advantages:

- The selection based on gradients allows to progress toward solution
- Slightly more complicated to implement than steepest descent but converges in a finite number of steps
- It does not require matrix inversion!!

References

- Description of reghdfe
- Paper on multi-way fixed effects (Correia (2016))
- Technical note on singletons (Correia (2015))
- Frisch-Waugh-Lowell theorem: Chapter 3.3 of *Econometric Analysis*, by William Greene
- Papers on points 1. and 2. of the solution algorithm (Gaure (2013a), Guimaraes and Portugal (2010))
- Notes on conjugate methods (slides, notes)