Note 3: Replication of Shrinkage estimation of wage by David and Minchul

Ignacio Sarmiento

October 7, 2020

Abstract

Replicate what is in the note, document

To estimate the average wage by cities and by occupations.

Data description

- 276 cmsa
- 64 occupation classifications
- number of observations: 4,284,808
- year: 2015

1 Empirical Model

1.1 A model with 64 occupations

$$y_{ij} = D_{ij}'\beta_j + (X_{ij}^f)'\beta^f + e_{ij}, \quad e_{ij} \sim_{i.i.d.} N(0, \sigma^2)$$

where

- cities are indexed by j, $(j = 1, 2, ..., 276)$.
- D_{ij}: 64×1, occupation indicator.
- β_j: 64×1. Average wage of occupations in city j after controlling for other demographic info.
- X_{ij}^f: $n_f \times 1$, fixed regressors, demographic characteristics such as age, gender, US born, etc.
- β^f: $n_f \times 1$

and we put prior on β_j

$$\beta_j = m_0 + m_1 \log p_j + \varepsilon_j, \quad \varepsilon_j \sim N(0, \Sigma)$$

where the average wage in city j depends on log population. m_0 and m_1 are 64×1 vector, and Σ is 64×64 matrix.

Few remarks.

- If $m_{k,1} > 0$, then the model implies that average wage of the kth occupation is higher for the larger city.
- Suppose there are two cities with the same size (population), then the model implies that average wage are similar to each other. (so that we can borrow information from other cities).
- We also borrow wage information from other occupations. We use more information from similar occupation. "similarity" among occupations is measured by the implied correlation from Σ. Note that we estimate Σ from the data,

Occupations, D_{ij}. $D_{ij} = (1, 0, 0, ..., 0)$ if the occupation of the person i in city j is the first occupation. If kth element in D_{ij} is 1 and other elements take 0, then the occupation of the person i in city j is kth occupation.

$||D_{ij}|| = 1$. That is, the person can have only one occupation. 64 occupations are described in "Occupation_names.xlsx".

***No observations in occupation occ=35 in the data set. Actually, we have 63 occupations...

Fixed regressors, X_{ij}^f These regressors are "fixed" in the sense that their coefficients do not vary by cities. In our application, we have the following as fixed regressors.

- $X_{1,ij}$: 1 if home owner (p), 0 if renter (r)
- $X_{2,ij}$: 1 if male (m), 0 if female (f)
- $X_{3,ij}$: 1 if education = g
- $X_{4,ij}$: 1 if education = c
- $X_{5,ij}$: 1 if education = s
- $X_{6,ij}$: 1 if education = h
- $X_{7,ij}$: 1 if race = w
- $X_{8,ij}$: 1 if race = b
- $X_{9,ij}$: 1 if race = l
- $X_{10,ij}$: 1 if race = a
- $X_{11,ij}$: 1 if age 1
- $X_{12,ij}$: 1 if age 2
- $X_{13,ij}$: 1 if age 3
- $X_{14,ij}$: 1 if US born, 0 if immigrants

Note that if all education related variables take 0, then i is education=d (drop-out). Note also that if all race related variables take 0, then i is "Other race." If all age related variables are 0, then age = 4 (oldest group).

Then, β_j is the average wage of renter, female, dropout, other race, oldest age group, immigrants.