Code documentation

César Garro-Marín*

July 5, 2023

1 Notation

Our GMM model is based on the moment equations:

$$\mathbb{E}\left(\psi(w_i, \mu)\right) = 0$$

where:

- w_i is the vector of data for observation i.
- ψ is a $P \times 1$ vector of functions.
- μ is an $R \times 1$ vector of parameters.

Our estimate is the solution to the problem:

$$\hat{\mu} = \arg\min_c \left(\frac{1}{N}\sum_{i=1}^{N}\psi(w_i, c)\right)' A \left(\frac{1}{N}\sum_{i=1}^{N}\psi(w_i, c)\right)$$

which we can write as:

$$\hat{\mu} = \arg\min_c \left(\frac{1}{N}\varepsilon(w_i, c)'Z\right) A \left(\frac{1}{N}Z'\varepsilon(w_i, c)\right)$$

where $\varepsilon(w_i, c)$ is the $N \times 1$ vector of errors of the model, and Z is the $N \times P$ matrix of instruments. Here $N = N_1 + N_2 + N_3$ is the total number of observations and N_i denotes the number of observations that belong to equation i. The matrix Z is structured so that the first N_1 observations correspond to the first equation, the next N_2 to the second, and so on and so forth.

*Boston University, email: cesarlgm@bu.edu

2 GMM standard errors

I minimize the quadratic form:

$$Q = \left(\frac{1}{N} \sum_{i=1}^{N} \psi(w_i, c) \right)' A \left(\frac{1}{N} \sum_{i=1}^{N} \psi(w_i, c) \right) \tag{1}$$

The GMM estimates have a distribution of:

$$\sqrt{N}(\hat{\mu} - \mu) \to N(0, \tilde{V})$$

where $\bar{V} = (D'AD)^{-1} D'AVAD(D'AD)^{-1}$. Here:

- D is the model gradient.
- V is defined as:

$$V = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left(\psi(w_i, \mu) \psi(w_i, \mu)' \right)$$

- A is the weighting matrix $(Z'Z)^{-1}$

2.1 Estimating the variance matrix

I will now describe how do I estimate each component of the variance covariance matrix A:

- **Estimating V:** our estimate is:

$$\hat{V} = \frac{1}{N} \sum_{i=1}^{N} \psi(w_i, \hat{\mu}) \psi(w_i, \hat{\mu})'$$

  we compute this estimate this component in the function `estimate_v`

- **Estimating D:** this is the gradient of the model's errors.

2.2 The gradient:

Our estimate of the gradient D is:

$$\hat{D} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \psi(w_l, \hat{\mu})}{\partial c'}$$

now, let us describe the form of this gradient in more detail. Each element of $\psi(w_l, c)$ is the function $z_{lp} \varepsilon_i(c)$, where $\varepsilon_l(c)$ is the error term for the l-th observation.

To avoid making the notation a mess, define the functions $J(l)$, $E(l)$, $T(l)$, and $I(l)$ which return the job, education level, year, and skill that correspond to observation l. In addition, for the third equation observations I define as $EET(l)$ as the function that returns the education pair-year cell of the observation.

$$\varepsilon_l(\mu) = \begin{cases} \Delta\overline{\ln S_{ijet+1}} - \sum_k \theta_{ke}\overline{S_{kejt}}\pi_{kjt} + \pi_{ijt} & l \leq N_1 \\ 1 - \sum_k \theta_{ke}\overline{S_{kejt}} & N_1 < l \leq N_1 + N_2 \\ \Delta\left[\ln \frac{q_{ejt+1}}{q_{e'jt+1}}\right] - \beta_j\left[\sum_k \left(\theta_{ke}\overline{S_{kjet}} - \theta_{ke'}\overline{S_{kje't}}\right)\pi_{kjt}\right] - \gamma_{e,e't} & N_1 + N_2 < l \leq N \end{cases}$$

Note that the parameter vector μ has four types of parameters: θ_{ke}, π_{kjt}, and β_j. Let us go case by case:

$$\frac{\partial\varepsilon_l(\mu)}{\partial\beta_j} = \begin{cases} -\beta_j\left[\sum_k \left(\theta_{ke}\overline{S_{kjet}} - \theta_{ke'}\overline{S_{kje't}}\right)\pi_{kjt}\right] & N_1 + N_2 < l \leq N, j = J(l) \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial\varepsilon_l(\mu)}{\partial\gamma_{ee't}} = \begin{cases} -1 & N_1 + N_2 < l \leq N, (e, e', t) = EET(l) \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial\varepsilon_l(\mu)}{\partial\pi_{ijt}} = \begin{cases} -\theta_{ie}\overline{S_{iejt}} & l \leq N_1, i \neq I(l), j = J(l), t = T(l) \\ -\theta_{ie}\overline{S_{iejt}} + 1 & l \leq N_1, i = I(l), j = J(l), t = T(l) \\ -\beta_j\left(\theta_{ie}\overline{S_{ijet}} - \theta_{ie'}\overline{S_{ije't}}\right) & N_1 + N_2 < l \leq N \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial\varepsilon_l(\mu)}{\partial\theta_{ie}} = \begin{cases} -\overline{S_{kejt}}\pi_{ijt} & l \leq N_1, e = E(l) \\ -\overline{S_{iejt}} & N_1 < l \leq N_1 + N_2, e = E(l) \\ -\beta_j\theta_{ie}\overline{S_{ijet}}\pi_{ijt} & N_1 + N_2 < l \leq N, (e, \cdot, \cdot) = EET(l) \\ \beta_j\theta_{ie}\overline{S_{ijet}}\pi_{ijt} & N_1 + N_2 < l \leq N, (\cdot, e, \cdot) = EET(l) \\ 0 & \text{otherwise} \end{cases}$$

Let Ξ be the $N \times R$ matrix with general term $\Xi_{ir} = \frac{\partial\varepsilon(w_i, c)}{\partial c_r}$. Then our estimate of the gradient is $P \times R$ matrix:

$$\hat{D} = \frac{1}{N}Z'\Xi$$