Generalized Method of Moments (GMM)

Jakub Mućk
SGH Warsaw School of Economics

Introduction

- The standard classical methods, e.g., the Maximum Likelihood (ML) method, requires a complete specification of the model that is considered to be estimated. This includes also the probability of distribution of the variable of interest.
- Contrary to the ML method, the **Generalized Method of Moments (GMM)** requires only a set moment conditions that are implied by assumption of the underlying econometric model. The GMM method is attractive when:
  - there is a variety of moment or orthogonality conditions that are deduced from the assumption of the theoretical model;
  - the economic model is complex, i.e., it is difficult to write down a tractable and applicable likelihood function,
  - to overcome the computational complexities associated with the ML estimator.

Generalized Method of Moments (GMM)

- Let's assume that a sample of T observations is drawn from the joint probability distribution:

$$f\left(w_1, w_2, \ldots, w_T, \theta_0\right) \tag{1}$$

where θ_0 is the $(q \times 1)$ vector of true parameters and w_t contains one or more endogenous and/or exogenous variables.

- Population moments condition:

$$\mathbb{E}\left[m\left(w_T, \theta_0\right)\right] = 0, \quad \text{for all } t. \tag{2}$$

- where $m(\cdot)$ is the r-dimensional vector of functions.
- Three cases:
  1. $q > r \implies$ the parameters in θ are **not identified**;
  2. $q = r \implies$ the parameters in θ are **exactly identified**;
  3. $q < r \implies$ the parameters in θ are **overidentified** and the moments conditions have to be restricted in order to deliver a unique θ in estimation. This can be done by the means of a weighting matrix (A_T).
- Estimation bases on the empirical counterpart of $\mathbb{E}\left[m\left(w_T, \theta_0\right)\right]$:

$$M_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} m\left(w_T, \theta_0\right), \tag{3}$$

where $M_T(\theta)$ is the r-dimensional vector of sample moments.

- **Linear regression:**
  - Consider the standard linear regression:
  $$y_t = x_t'\beta + \varepsilon_t \tag{4}$$
  Under the standard (classical) assumption, the population conditions is following:
  $$\mathbb{E}\left(x_t, \varepsilon_t\right) = \mathbb{E}\left[x_t, \left(y_t - x_t'\beta\right)\right] = 0 \quad \text{for } t \in 1, \ldots, T. \tag{5}$$
- **Linear regression with endogenous variables:**
  - Consider the standard linear regression with endogenous variables:
  $$y_t = x_t'\beta + \varepsilon_t \tag{6}$$
  where $\mathbb{E}(x_t, \varepsilon_t) \neq 0$.
  - The population conditions:
  $$\mathbb{E}\left(z_t, \varepsilon_t\right) = \mathbb{E}\left[z_t, \left(y_t - x_t'\beta\right)\right] = 0 \quad \text{for } t \in 1, \ldots, T \tag{7}$$
  where z_T is the set of the instrumental variables that satisfies the above orthogonality conditions.

The GMM and GIVE estimators

- **The GMM estimator** of θ bases on:

$$\hat{\theta}_T = \text{argmin}_{\theta \in \Theta} \left\{ M_T'(\theta) W_T M_T(\theta) \right\}, \quad (8)$$

where W_T is a $r \times r$ positive semi-define, possibly random weighting matrix.

- We wish to choose the weighting matrix that minimizes the covariance matrix of $\hat{\theta}$.
  - This provides the efficient estimator. Other weighting matrices would lead to less efficient estimators of θ.
- **The general instrumental variable estimator (GIVE)** combines all available instruments to estimates the unknown parameters. In this case, **the number of instruments can be larger than number of parameters to estimate ($r > k$).**
- The starting point: the r population conditions:

$$\mathbb{E}\left(z_t, \varepsilon_t\right) = \mathbb{E}\left[z_t, \left(y_t - x_t'\beta\right)\right] = 0 \quad \text{for } t \in 1, \ldots, T \quad (9)$$

where z_t is the set of (r) instruments, x_t is the k-dimensional vector of regressors. The regressors are endogenous, i.e., $\mathbb{E}\left(x_t, \varepsilon_t\right) \neq 0$ while the error term is idiosyncratic, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The instruments are correlated with x_t but not correlated with the error term.

- This implies the following sample moments:

$$M_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} z_t \left(y_t - \beta' x_t\right). \tag{10}$$

- It can be shown that the GIVE estimator is given as:

$$\hat{\beta}^{GIVE} = \left(X' W_z X\right)^{-1} X' W_z Y, \tag{11}$$

where $W_z = Z\left(Z'Z\right)^{-1} Z'$. The matrix Z collects all instruments, the matrix X stands for the regressors while Y denotes the observations of the dependent variable.

- The estimator of the variance matrix of $\hat{\beta}^{GIVE}$ is as follows:

$$Var\left(\hat{\beta}^{GIVE}\right) = \hat{\sigma}^2_{GIVE}\left(X' W_z X\right)^{-1}. \tag{12}$$

where the estimated variance of the error term bases on the variance of the residuals from the considered regression:

$$\hat{\sigma}^2_{GIVE} = \frac{1}{T-K}\hat{\varepsilon}'_{GIVE}\hat{\varepsilon}_{GIVE}. \tag{13}$$

- In analogous fashion to the basic linear models, the robust standard error (e.g. hetereoskedasticity-consistent or Newey-West) can be computed.

- In the GIVE estimation we use r instruments. Are these instruments valid?
- Consider following test statistics:

$$\chi^2_{SM} = \frac{Q(\hat{\beta}^{GIVE})}{\hat{\sigma}^2_{GIVE}}, \tag{14}$$

where

$$Q(\hat{\beta}^{GIVE}) = \left(y - X\hat{\beta}^{GIVE}\right)' W_z \left(y - X\hat{\beta}^{GIVE}\right) \tag{15}$$

- Under the null the regression is correctly specified and the r instruments Z are valid instruments.
- The Sargan misspecification statistics is χ^2 distributed with $r - k$ degrees of freedom.

General remarks

- GMM approach bases on strong economic assumptions.
- Partial/weak identification.
- Efficiency is related to the moment condition.
- The GMM estimates are less robust $\Longrightarrow$ robustness can be achieved through weights and variance.