GMM and SMM

LIAO Chengwang

Structural Estimation Workshop, CUHK Econ

October 16, 2023

Method of Moments Estimation

- Set population moment conditions (equations), then replace population moments with sample moments
  - e.g., $X_1, \ldots, X_n \sim$ i.i.d. $\mathcal{N}(\mu, 1)$, population moment condition $\mathbb{E}(X_i) = \mu$, sample moment condition $\frac{1}{n}\sum_i X_i = \hat{\mu}$
- In general, suppose we have k parameters $\beta_1, \ldots, \beta_k$ to estimate, and have k moment conditions

$$\mathbb{E}[\boldsymbol{g}_i(\boldsymbol{X}_i, \boldsymbol{\beta})] = \mathbf{0}$$

  Then the sample moment conditions

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_i(\boldsymbol{X}_i, \boldsymbol{\beta}) = \mathbf{0}$$

  Solve for $\boldsymbol{\beta}$ and denote it $\hat{\boldsymbol{\beta}}_{\mathrm{MM}}$
- $\boldsymbol{\beta}$ is **just identified**: number of equations = number of parameters

Method of Moments Estimation

Many estimators be seen as MM estimators.

- Mean: Set $g_i(\mu) = X_i - \mu$, then $\hat{\mu}_{\text{MM}} = \frac{1}{n}\sum_{i=1}^{n} \mu$
- OLS: Set $\boldsymbol{g}_i(\boldsymbol{\beta}) = \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$, then $\hat{\boldsymbol{\beta}}_{\text{MM}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
- IV: Set $\boldsymbol{g}_i(\boldsymbol{\beta}) = \boldsymbol{Z}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$, then $\hat{\boldsymbol{\beta}}_{\text{MM}} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$
- NLS: Set $\boldsymbol{g}_i(\boldsymbol{\beta}) = \boldsymbol{X}_i[Y_i - f(\boldsymbol{X}_i, \boldsymbol{\beta})]$, then $\hat{\boldsymbol{\beta}}_{\text{MM}}$ is the solution to $\frac{1}{n}\sum_i \boldsymbol{g}_i(\boldsymbol{\beta}) = 0$

Generalized Method of Moments

- What if # of moment equations > # of parameters, i.e., $\boldsymbol{\beta}$ is **over-identified**? The system of equations could possibly have no solution
- Instead of solving the system of equations, we can try to find a criterion making $\bar{\boldsymbol{g}}_i(\boldsymbol{\beta}) := \frac{1}{n}\sum_i \boldsymbol{g}_i(\boldsymbol{\beta})$ as close to zero as possible
- GMM estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{GMM}} = \arg\min_{\boldsymbol{\beta}} \bar{\boldsymbol{g}}_i(\boldsymbol{\beta})' \boldsymbol{W} \bar{\boldsymbol{g}}_i(\boldsymbol{\beta})$$

  where $\boldsymbol{W}$ is a (p.s.d. and symmetric) weight matrix
- Under fairly general assumptions, $\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}$ asymptotically follows a normal distribution and is consistent (converges to the real $\boldsymbol{\beta}$)

Two-step GMM

- Though asymptotic normality does not depend on the choice of weight matrix, asymptotic variance does
- The optimal weight matrix that minimizes avar($\hat{\boldsymbol{\beta}}_{\text{GMM}}$) is given by $\boldsymbol{W} = \left(\mathbb{E}[\boldsymbol{g}_i(\boldsymbol{\beta})\boldsymbol{g}_i(\boldsymbol{\beta})']\right)^{-1}$
- Two steps to achieve feasible efficient GMM
  1. Choose a weight matrix as you like (e.g., $\boldsymbol{W} = \boldsymbol{I}$) and obtain the first step GMM estimator, $\hat{\boldsymbol{\beta}}^{(1)}$
  2. Let $\boldsymbol{W} = \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_i(\hat{\boldsymbol{\beta}}^{(1)})\boldsymbol{g}_i(\hat{\boldsymbol{\beta}}^{(1)})'\right]^{-1}$ and obtain the feasible efficient GMM estimator, $\hat{\boldsymbol{\beta}}_{\text{EGMM}}$

2SLS

- 2SLS is a special case of GMM in linear models, with the weight matrix $\boldsymbol{W} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$
- Under homoskedasticity, 2SLS is equivalent to efficient GMM (the optimal weight matrix is $[\mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i' u_i^2)]^{-1}$). But in general, they are not equivalent

MLE and GMM

- MLE can also be thought of as a special case of GMM
- MLE assumes $\beta_0 = \max_{\boldsymbol{\beta}} \mathbb{E}[\ln f(\boldsymbol{X}_i, \boldsymbol{\beta})]$, which yields FOC

  $$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}} \ln f(\boldsymbol{X}_i, \beta_0)\right] = \mathbf{0}$$

  This gives us population moment conditions. The sample analog

  $$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} \ln f(\boldsymbol{X}_i, \hat{\beta}) = 0$$

  corresponds to the FOC of maximizing log-likelihood
- *Remark.* The way we form moment conditions is important. Under MLE assumptions, the estimator is the most efficient, while in general, the moments may fail to fully utilize sample information

Simulated Method of Moments

- SMM is essentially GMM, but simulation-based
- SMM was originally developed by McFadden (1989) and Pakes and Pollard (1989) to estimate discrete-choice models in i.i.d. environments, and then extended by Lee and Ingram (1991) and Duffie and Singleton (1993) to time-series models with serially correlated shocks
- The motivation is to circumvent *computational difficulties* such as evaluating intractable aggregation formulae (numerical integration) and evaluating discrete response probabilities
- GMM implicity assumes that $g_i(\beta)$ is a known function having closed-form expression, which may not always be the case
    - For example, $g_i(\beta) = h_i(\beta) - \mathbb{E}[h_i(\beta)]$, the expectation $\mathbb{E}[h_i(\beta)]$ may turn out to be very hard to evaluate

Simulated Method of Moments

- Suppose $g_i(X_i, \beta)$ can be expressed as a conditional expectation of a tractable function G

  $$g_i(X_i, \beta) = \int G_i(X_i, \zeta_i, \beta)\, dF(\zeta|X_i) = \mathbb{E}_{\zeta|X_i}[G_i(X_i, \zeta_i, \beta)]$$

  Then given each X_i, one could generate observations $\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{im}$ from distribution $F(\zeta|X_i)$ and form the average

  $$\hat{g}_i(X_i, \beta) = \frac{1}{m} \sum_{j=1}^{m} G_i(X_i, \zeta_{ij}, \beta)$$

  The SMM estimator is obtained by making $\frac{1}{n} \sum_i \hat{g}_i(X_i, \beta)$ as close to zero as possible (as GMM does)
- Under regularity conditions, the SMM estimator preserves asymptotic normality and consistency (but less efficient than GMM due to simulation errors)

Example 1: A Discrete-Choice Model

- Model setup
  - A consumer has ℓ alternatives to choose between
  - His choice is determined by a set of ℓ vector covariates $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_\ell$ and a *random* vector $\boldsymbol{\alpha}$ of weights: alternative i is chosen if $\boldsymbol{\alpha}'\boldsymbol{X}_i$ is larger than all the other $\boldsymbol{\alpha}'\boldsymbol{X}_j$ (we can interpret $u_i = \boldsymbol{\alpha}'\boldsymbol{X}_i$ as utility from choosing i)
  - The weight vector α is generated from multivariate normal: $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0)$ where $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0)$ are unknown parameters
- The probability of choosing k (conditional on $\boldsymbol{X}$)

$$P(k|\boldsymbol{X}, \boldsymbol{\theta}) = \int 1\{\boldsymbol{\alpha}'\boldsymbol{X}_k \geq \boldsymbol{\alpha}'\boldsymbol{X}_{-k}\} f(\boldsymbol{\alpha}|\boldsymbol{\theta})\, d\boldsymbol{\alpha}$$

When ℓ and dimension of α are large, numerical integration is hard to obtain

Example 1: A Discrete-Choice Model

- There are n consumers so that we have a sample of size n: decisions $(\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(n)})$ and covariates $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)})$
- Moment conditions are generated from MLE FOCs. Log likelihood

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{k=1}^{\ell} d_k^{(i)} \ln P(k|\boldsymbol{X}^{(i)}, \boldsymbol{\theta})$$

The FOCs

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \sum_{k=1}^{\ell} \frac{\partial \ln P(k|\boldsymbol{X}^{(i)}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [d_k^{(i)} - P(k|\boldsymbol{X}^{(i)}, \boldsymbol{\theta})] = \mathbf{0}$$

where $\boldsymbol{Z}_k^{(i)} := \partial \ln P(k|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ can be seen as instruments

Example 1: A Discrete-Choice Model

- For a fixed $\boldsymbol{\theta}$, for each individual i, independently generate m random vectors $\boldsymbol{\alpha}_1^{(i)}, \ldots, \boldsymbol{\alpha}_m^{(i)}$ from $\mathcal{N}(\boldsymbol{\theta})$, then evaluate for each k

$$\hat{P}(k|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}) = \frac{1}{m} \sum_{j=1}^{m} 1\{\boldsymbol{\alpha}_j^{(i)\prime} \boldsymbol{X}_k^{(i)} \geq \boldsymbol{\alpha}_j^{(i)\prime} \boldsymbol{X}_{-k}^{(i)}\}$$

- Finally, the SMM estimator for $\boldsymbol{\theta}_0$ is obtained by

$$\min_{\boldsymbol{\theta}} \left\| \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{\ell} \boldsymbol{Z}_k^{(i)} [d_k^{(i)} - \hat{P}(k|\boldsymbol{X}^{(i)}, \boldsymbol{\theta})] \right\|$$

Example 2: A DSGE Model

- We are interested in estimating a version of the Brock-Mirman model
- Model setup
  - Representative agent maximizes lifetime utility $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}-1}{1-\gamma} \xi_t$ subject to the resource constraint $c_t + k_{t+1} = z_t k_t^{\alpha} + (1-\delta) k_t$
  - ξ_t: taste shock; z_t: productivity shock. They follow Markov processes $\ln z_t = \rho_z \ln z_{t-1} + \epsilon_t^z$ and $\ln \xi_t = \rho_\xi \ln \xi_{t-1} + \epsilon_t^\xi$
- Data: $(\{c_t\}_{t=0}^T, \{k_t\}_{t=0}^T)$. Parameters: $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta, \rho_z, \rho_\xi)'$
- Solving the model (analytically or numerically), we would have a policy function $\boldsymbol{\pi}$ generating the state process $\boldsymbol{Y}_t = (z_t, \xi_t, k_t)'$:

$$\boldsymbol{Y}_{t+1} = \boldsymbol{\pi}(\boldsymbol{Y}_t, \epsilon_t, \boldsymbol{\theta})$$

- The conventional GMM that constructs moment conditions by Euler equations does not work because of the taste shock

Example 2: A DSGE Model

- We can actually simulate moments of any form we are interested in
- Suppose we observe some function $\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{\theta})$ of the data of ℓ-history $\boldsymbol{x}_t = (c_t, \ldots, c_{t-\ell+1}, k_t, \ldots, k_{t-\ell+1})$
  - e.g., let $\boldsymbol{h}(\boldsymbol{x}_t) = \ln(c_t / c_{t-1})$
- However, $\mathbb{E}[\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{\theta})]$ has no analytic form so that the sample moment conditions $\frac{1}{T}\sum_{i=1}^{T} \boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{\theta}) - \mathbb{E}[\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{\theta})] = \mathbf{0}$ is not available
  - Remark. This problem becomes salient if we solve the model by, e.g., second-order polynomial approximation, since then the policy function must be nonlinear
- We can instead estimate $\mathbb{E}[\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{\theta})]$ by simulating state process $\{\tilde{\boldsymbol{Y}}_\tau\}_{\tau=1}^{\mathcal{T}}$ according to the policy function:

$$\tilde{\boldsymbol{Y}}_\tau = \boldsymbol{\pi}(\tilde{\boldsymbol{Y}}_\tau, \tilde{\boldsymbol{\epsilon}}_\tau, \boldsymbol{\theta})$$

$\tilde{\boldsymbol{\epsilon}}_\tau$ are drawn from the same distribution as $\boldsymbol{\epsilon}_t$