Generalized Method of Moments

Paolo Zacchia

Econometric Theory

Lecture 12

Limitations of the Method of Moments

- Many econometric estimators can be seen as special cases of the **Method of Moments**.

- The IV estimator, for example, derives from the following **zero moment conditions**:

$$\mathbb{E}\left[\boldsymbol{z}_i\left(Y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right] = \mathbf{0}$$

  where both $\boldsymbol{z}_i$ and $\boldsymbol{x}_i$ have dimension K (while if $\boldsymbol{z}_i = \boldsymbol{x}_i$, this reduces to OLS). Indeed, solving for $\boldsymbol{\beta}_0$ gives:

$$\boldsymbol{\beta}_0 = \mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i^{\mathrm{T}}\right]^{-1}\mathbb{E}\left[\boldsymbol{z}_i Y_i\right]$$

  whose sample analog is precisely the IV estimator.

- The Method of Moments easily handles **semi-parametric** models, but it is limited: for example, it does not allow for overidentified 2SLS (when $\boldsymbol{z}_i$ above has dimension $J > K$).

Generalizing the Method of Moments

- To **generalize** the Method of Moments, posit some $J \geq K$ **zero moment conditions** based on a vector function $\boldsymbol{g}(\cdot)$:

$$\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right] = \mathbf{0}$$

where $\boldsymbol{x}_i = (\boldsymbol{y}_i, \boldsymbol{z}_i)$ while $\boldsymbol{\theta}_0$ has dimension K.

- The **sample analog** of such moment conditions, motivated by the Analogy Principle, is also a J-dimensional vector.

$$\overline{\mathbf{g}}_N\left(\boldsymbol{\theta}\right) \equiv \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}\left(\mathbf{x}_i; \boldsymbol{\theta}\right)$$

- A **Generalized Method of Moments** (GMM) estimator minimizes a quadratic form $\widehat{\mathcal{G}}_N\left(\boldsymbol{\theta}\right)$ as follows:

$$\widehat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \widehat{\mathcal{G}}_N\left(\boldsymbol{\theta}\right) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \overline{\mathbf{g}}_N^{\mathrm{T}}\left(\boldsymbol{\theta}\right)\mathbf{A}_N\overline{\mathbf{g}}_N\left(\boldsymbol{\theta}\right)$$

for *any full rank positive semi-definite* $J \times J$ matrix $\mathbf{A}_N$. It aims to align the empirical with the theoretical moments.

Characteristics of the GMM problem (1/2)

A few remarks about the GMM estimator are in order.

1. The First Order Conditions of the problem are as follows:
$$\frac{\partial}{\partial \boldsymbol{\theta}} \overline{\mathbf{g}}_N^{\mathrm{T}} \left(\widehat{\boldsymbol{\theta}}_{GMM}\right) \cdot \mathbf{A}_N \cdot \overline{\mathbf{g}}_N \left(\widehat{\boldsymbol{\theta}}_{GMM}\right) = \mathbf{0}$$
where the element that precedes $\mathbf{A}_N$ is a $K \times J$ matrix. In general, the GMM estimator must be derived numerically.

2. The GMM estimator is an M-Estimator. Define:
$$\mathcal{G}_0 (\boldsymbol{\theta}) \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left[\boldsymbol{g} (\boldsymbol{x}_i; \boldsymbol{\theta}) \right]^{\mathrm{T}} \mathbf{A}_0 \, \mathbb{E} \left[\boldsymbol{g} (\boldsymbol{x}_i; \boldsymbol{\theta}) \right] \geq 0$$
for a **full rank** positive semi-definite $J \times J$ matrix $\mathbf{A}_0$ such that $\mathbf{A}_N \xrightarrow{p} \mathbf{A}_0$. The objective function $\widehat{\mathcal{G}}_N (\boldsymbol{\theta})$ converges in probability to $\mathcal{G}_0 (\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:
$$\widehat{\mathcal{G}}_N (\boldsymbol{\theta}) \xrightarrow{p} \mathcal{G}_0 (\boldsymbol{\theta})$$
an M-Estimator for $\widehat{\mathcal{G}}_N (\boldsymbol{\theta}) = -\widehat{\mathcal{Q}}_N (\boldsymbol{\theta})$, $\mathcal{G}_0 (\boldsymbol{\theta}) = -\mathcal{Q}_0 (\boldsymbol{\theta})$.

Characteristics of the GMM problem (2/2)

3. The model is **identified** if $\mathcal{G}_0(\boldsymbol{\theta})$ has **one** local **minimum**, which must be equal to the true parameter $\boldsymbol{\theta}_0$ as $\mathcal{G}_0(\boldsymbol{\theta}) \geq 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\mathcal{G}_0(\boldsymbol{\theta}_0) = 0$ by construction. By the First Order Conditions:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{g}^{\mathrm{T}}(\boldsymbol{x}_i; \boldsymbol{\theta}_0)\right] \cdot \mathbf{A}_0 \cdot \mathbb{E}\left[\boldsymbol{g}(\boldsymbol{x}_i; \boldsymbol{\theta}_0)\right] = \mathbf{0}$$

and there is a **unique solution** if the $J \times K$ matrix $\mathbf{G}_0$:

$$\mathbf{G}_0 \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \boldsymbol{g}(\boldsymbol{x}_i; \boldsymbol{\theta}_0)\right]$$

has full column rank K, as otherwise many combination of parameters are equally capable of minimizing $\mathcal{G}_0(\boldsymbol{\theta})$. With *identically distributed* observations, it is as follows.

$$\mathbf{G}_0 = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \boldsymbol{g}(\boldsymbol{x}_i; \boldsymbol{\theta}_0)\right]$$

Asymptotics of the GMM estimator (1/4)

Theorem 1

Asymptotic Properties of GMM. *A GMM estimator that is based on some J zero moment conditions, which is identified and also meets the uniform convergence requirements of M-Estimators from Theorem 2, Lecture 11, is consistent.*

$$\widehat{\boldsymbol{\theta}}_{GMM} = \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \widehat{\mathcal{G}}_N\left(\boldsymbol{\theta}\right) \xrightarrow{p} \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{G}_0\left(\boldsymbol{\theta}\right) = \boldsymbol{\theta}_0$$

If the conditions analogous to those of Theorem 3, Lecture 11 are met, the GMM estimator is also asymptotically normal, with the following sandwiched expression for its variance-covariance:

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1} \mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\boldsymbol{\Omega}_0\mathbf{A}_0\mathbf{G}_0 \left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\right)$$

where $\boldsymbol{\Omega}_0$ is the following $J \times J$ limiting matrix.

$$\boldsymbol{\Omega}_0 \equiv \lim_{N\to\infty} \mathbb{V}\mathrm{ar}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right]$$

Asymptotics of the GMM estimator (2/4)

Theorem 1

Note. Before proceeding with a sketched proof, it is useful to observe that like in similar cases, when the observations are *independent* it is:

$$\boldsymbol{\Omega}_0 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right) \boldsymbol{g}^{\mathrm{T}}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right]$$

and $\boldsymbol{\Omega}_0 = \mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right) \boldsymbol{g}^{\mathrm{T}}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right]$ in the more specialized i.i.d. case.

Proof.

(Sketched.) By some Weak Law of Large Numbers it must be that:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\mathcal{G}}_N\left(\boldsymbol{\theta}_0\right) = \frac{\partial}{\partial \boldsymbol{\theta}} \overline{\mathbf{g}}_N^{\mathrm{T}}\left(\boldsymbol{\theta}_0\right) \cdot \mathbf{A}_N \cdot \overline{\mathbf{g}}_N\left(\boldsymbol{\theta}_0\right) \xrightarrow{p} \mathbf{G}_0 \mathbf{A}_0 \cdot \mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right] = \mathbf{0}$$

since $\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right] = \mathbf{0}$ by assumption; by the FOCs this implies that the GMM estimator is consistent:

$$\widehat{\mathcal{G}}_N\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) \xrightarrow{p} \widehat{\mathcal{G}}_N\left(\boldsymbol{\theta}_0\right) \;\Rightarrow\; \widehat{\boldsymbol{\theta}}_{GMM} \xrightarrow{p} \boldsymbol{\theta}_0$$

if the model is identified ($\mathbf{G}_0$ is of full rank). **(Continues. . .)**

Asymptotics of the GMM estimator (3/4)

Theorem 1

Proof.

(Continued.) To show asymptotic normality, apply the Mean Value Theorem *directly* to the empirical moments; with some manipulation:

$$\sqrt{N}\,\bar{\mathbf{g}}_N\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) = \sqrt{N}\,\bar{\mathbf{g}}_N\left(\boldsymbol{\theta}_0\right) + \mathbf{G}_N\left(\widetilde{\boldsymbol{\theta}}_N\right)\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0\right)$$

where as usual $\widetilde{\boldsymbol{\theta}}_N$ is a convex combination of $\widehat{\boldsymbol{\theta}}_{GMM}$ and $\boldsymbol{\theta}_0$, whereas $\mathbf{G}_N\left(\boldsymbol{\theta}\right)$ is defined as:

$$\mathbf{G}_N\left(\boldsymbol{\theta}\right) \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\boldsymbol{g}\left(\mathbf{x}_i;\boldsymbol{\theta}\right)$$

analogously to $\mathbf{G}_0$. Plugging the first expression above into the GMM First Order Conditions delivers the following expression.

$$\mathbf{G}_N^{\mathrm{T}}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right)\mathbf{A}_N\left[\sqrt{N}\,\bar{\mathbf{g}}_N\left(\boldsymbol{\theta}_0\right) + \mathbf{G}_N\left(\widetilde{\boldsymbol{\theta}}_N\right)\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0\right)\right] = \mathbf{0}$$

(Continues. . .)

Asymptotics of the GMM estimator (4/4)

Theorem 1

Proof.
(Continued.) The latter result can be manipulated so as to return the following equation.

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{GMM}-\boldsymbol{\theta}_0\right)=-\left[\mathbf{G}_N^{\mathrm{T}}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right)\mathbf{A}_N\mathbf{G}_N\left(\widetilde{\boldsymbol{\theta}}_N\right)\right]^{-1}\times \\ \times\,\mathbf{G}_N^{\mathrm{T}}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right)\mathbf{A}_N\sqrt{N}\,\overline{\mathbf{g}}_N\left(\boldsymbol{\theta}_0\right)$$

Given that $\mathbf{G}_N\left(\widehat{\boldsymbol{\theta}}_{GMM}\right)\xrightarrow{p}\mathbf{G}_0$ and $\mathbf{G}_N\left(\widetilde{\boldsymbol{\theta}}_N\right)\xrightarrow{p}\mathbf{G}_0$ by consistency of GMM, if

$$\sqrt{N}\,\overline{\mathbf{g}}_N\left(\boldsymbol{\theta}_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\boldsymbol{\Omega}_0\right)$$

that is, a Central Limit Theorem can be applied to the data at hand, all these results can be combined via the Delta Method so as to yield the result sought after. □

Inference on the GMM estimator

The GMM asymptotic variance-covariance is estimated as:

$$\widehat{\mathbb{A}\text{var}}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) = \frac{1}{N}\left(\widehat{\mathbf{G}}_N^{\mathrm{T}}\mathbf{A}_N\widehat{\mathbf{G}}_N\right)^{-1}\widehat{\mathbf{G}}_N^{\mathrm{T}}\mathbf{A}_N\widehat{\boldsymbol{\Omega}}_N\mathbf{A}_N\widehat{\mathbf{G}}_N\left(\widehat{\mathbf{G}}_N^{\mathrm{T}}\mathbf{A}_N\widehat{\mathbf{G}}_N\right)^{-1}$$

...where $\widehat{\mathbf{G}}_N$ is a consistent estimator of $\mathbf{G}_0$:

$$\widehat{\mathbf{G}}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\boldsymbol{g}\left(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_{GMM}\right) \xrightarrow{p} \mathbf{G}_0$$

...whereas $\widehat{\boldsymbol{\Omega}}_N$, the estimator of $\boldsymbol{\Omega}_0$, depends on the statistical assumptions; if the observations are independent:

$$\widehat{\boldsymbol{\Omega}}_N = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}\left(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_{GMM}\right)\boldsymbol{g}^{\mathrm{T}}\left(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_{GMM}\right) \xrightarrow{p} \boldsymbol{\Omega}_0$$

...while the CCE case is as follows (HAC is also feasible).

$$\widehat{\boldsymbol{\Omega}}_{CCE} = \frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N_c}\sum_{j=1}^{N_c}\boldsymbol{g}_{ic}\left(\mathbf{x}_{ic};\widehat{\boldsymbol{\theta}}_{GMM}\right)\boldsymbol{g}_{jc}^{\mathrm{T}}\left(\mathbf{x}_{jc};\widehat{\boldsymbol{\theta}}_{GMM}\right) \xrightarrow{p} \boldsymbol{\Omega}_0$$

Optimal GMM weighting matrix (1/3)

- The asymptotic variance-covariance of $\widehat{\boldsymbol{\theta}}_{GMM}$ depends on the **weighting matrix** $\mathbf{A}_N$ that weighs the J moments.

- An important result, originally by Hansen (1982), showed that the **most efficient** GMM estimator is the one using the following **optimal** weighting matrix.

$$\mathbf{A}_N = \widehat{\boldsymbol{\Omega}}_N^{-1}$$

- Here $\widehat{\boldsymbol{\Omega}}_N^{-1}$ is a matrix that converges in probability to the inverse of $\boldsymbol{\Omega}_0$.

$$\widehat{\boldsymbol{\Omega}}_N^{-1} \xrightarrow{p} \boldsymbol{\Omega}_0^{-1} = \mathbf{A}_0$$

- Therefore, the limiting distribution of the GMM estimator becomes much simpler to express.

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{G}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\mathbf{G}_0\right)^{-1}\right)$$

Optimal GMM weighting matrix (2/3)

- To prove optimality, study the difference between a generic limiting variance-covariance and the optimal one:

$$\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{\Omega}_0\mathbf{A}_0\mathbf{G}_0\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}-\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{\Omega}_0^{-1}\mathbf{G}_0\right)^{-1}=$$
$$=\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{\Omega}_0^{\frac{1}{2}}\cdot\mathbf{M}_{\widetilde{\mathbf{G}}_0}\cdot\mathbf{\Omega}_0^{\frac{1}{2}}\mathbf{A}_0\mathbf{G}_0\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}$$

  where matrix $\mathbf{M}_{\widetilde{\mathbf{G}}_0}$ is symmetric and idempotent:

$$\mathbf{M}_{\widetilde{\mathbf{G}}_0}\equiv\mathbf{I}-\widetilde{\mathbf{G}}_0\left(\widetilde{\mathbf{G}}_0^{\mathrm{T}}\widetilde{\mathbf{G}}_0\right)^{-1}\widetilde{\mathbf{G}}_0^{\mathrm{T}}\text{ where }\widetilde{\mathbf{G}}_0\equiv\mathbf{\Omega}_0^{-\frac{1}{2}}\mathbf{G}_0.$$

- Thus, the above difference is a semi-definite positive matrix whatever $\mathbf{A}_0$ is.

- Intuitively, the **larger** the variance (uncertainty) of a single moment condition, the **smaller** its contribution to $\mathcal{G}_0(\theta)$.

- Simplifying "sandwich" expressions this way typically leads to more efficient variance-covariances, like in the analysis of GLS in the case of the linear model for small samples.

Optimal GMM weighting matrix (3/3)

Theorem 2

Semi-Parametric Efficiency Bound of GMM. *Under its motivating moment conditions, the GMM estimator that obtains through the optimal weighting matrix* $\mathbf{\Omega}_0^{-1}$ *hits the efficiency bound which applies to the class of all semi-parametric estimators of* $\boldsymbol{\theta}_0$*.*

Proof.

(Outline.) The argument proceeds as follows. Let the data $\{\mathbf{x}_i\}_{i=1}^N$ be drawn from a *discrete* support of dimension D; $\mathbb{X}_D = \{\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \dots, \boldsymbol{\chi}_D\}$ – where $\boldsymbol{\chi}_d$ for $d = 1, \dots, D$ is a given point in the support. Under the moment conditions, it follows that:

$$\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right] = \frac{1}{D}\sum_{d=1}^{D} \boldsymbol{g}\left(\boldsymbol{\chi}_d; \boldsymbol{\theta}_0\right) p_d = \mathbf{0}$$

where p_d is the probability attached to the d-th element of $\mathbb{X}_D$. Thus, estimating $\boldsymbol{\theta}_0$ by optimal GMM is equivalent to solving a parametric maximum likelihood problem: the Cramér-Rao bound is binding here. Furthemore, it can be shown that the result approximately holds even when the data have a continuous support. □

Two-step GMM estimation

This result, originally by Chamberlain (1987) is very **powerful**. It motivates a widespread use of GMM and its special cases; yet it is silent on *how* to estimate GMM if $\mathbf{\Omega}_0$ is *ex ante* unknown.

To address this, Hansen (1982) proposed a **two-step** procedure.

1. Derive a first step estimate $\widehat{\boldsymbol{\theta}}_1$ with some arbitrary $\mathbf{A}_N$ like the identity matrix. The resulting estimate $\widehat{\boldsymbol{\theta}}_1$ is consistent but inefficient; yet it affords a consistent estimator for $\mathbf{\Omega}_0$.

$$\widehat{\mathbf{\Omega}}_N = \frac{1}{N} \sum_{i=1}^{N} \left[\boldsymbol{g}\left(\mathbf{x}_i; \widehat{\boldsymbol{\theta}}_1\right) \boldsymbol{g}^{\mathrm{T}}\left(\mathbf{x}_i; \widehat{\boldsymbol{\theta}}_1\right) \right] \xrightarrow{p} \mathbf{\Omega}_0$$

2. Obtain a second step, final GMM estimate $\widehat{\boldsymbol{\theta}}_{GMM} = \widehat{\boldsymbol{\theta}}_2$ by minimizing function $\widehat{\mathcal{G}}_2(\boldsymbol{\theta}) = \overline{\mathbf{g}}_N^{\mathrm{T}}(\boldsymbol{\theta}) \widehat{\mathbf{\Omega}}_N^{-1} \overline{\mathbf{g}}_N(\boldsymbol{\theta})$.

Ultimately, the variance-covariance is estimated as follows.

$$\widehat{\mathbb{A}\mathrm{var}}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) = \frac{1}{N} \left(\widehat{\mathbf{G}}_N^{\mathrm{T}} \widehat{\mathbf{\Omega}}_N^{-1} \widehat{\mathbf{G}}_N \right)^{-1}$$

Alternative GMM estimation procedures

Hansen's two-step procedure is popular, but is **biased in small samples**. Consequently, sometimes the second step is repeated: $\boldsymbol{\Omega}_0$ is re-estimated using $\widehat{\boldsymbol{\theta}}_2$. Often, two **alternative** approaches are used, although they are typically computationally intensive.

- **Iterated GMM estimation**: with this method, Hansen's procedure is repeated many times, each time re-estimating $\boldsymbol{\Omega}_0$, until yet another iteration makes little difference.

- **Continuously updating GMM estimation** (CUGMM): here the idea is to estimate the weighting matrix as well as the parameters *jointly*. The objective function is as follows.

$$\widehat{\mathcal{G}}_N(\boldsymbol{\theta}) = \left[\sum_{i=1}^{N} \boldsymbol{g}(\mathbf{x}_i; \boldsymbol{\theta})\right]^{\mathrm{T}} \left[\sum_{i=1}^{N} \boldsymbol{g}(\mathbf{x}_i; \boldsymbol{\theta})\, \boldsymbol{g}^{\mathrm{T}}(\mathbf{x}_i; \boldsymbol{\theta})\right]^{-1} \left[\sum_{i=1}^{N} \boldsymbol{g}(\mathbf{x}_i; \boldsymbol{\theta})\right]$$

  In numerical optimization, $\widehat{\boldsymbol{\Omega}}_N$ is updated at every step.

GMM and Instrumental Variables: overview

In many applications, GMM estimators are based upon so-called **conditional moment conditions**, which take the form:

$$\mathbb{E}\left[\left.\boldsymbol{h}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta_0\right)\right| \boldsymbol{z}_i\right] = \mathbf{0}$$

where $\boldsymbol{h}(\cdot)$ is here a P-valued function, whereas $\boldsymbol{z}_i$ is a vector of J **instrumental variables**.

As per the Law of Iterated Expectations, the above conditional moments deliver PJ **zero moment conditions** that are usable for estimation.

$$\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i; \theta_0\right)\right] = \mathbb{E}\left[\boldsymbol{z}_i \otimes \boldsymbol{h}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta_0\right)\right] = \mathbf{0}$$

The ensuing discussion shows that this type of GMM estimators encompass and generalize many typical econometric estimators; among the others: 1. 2SLS; 2. 3SLS; 3. extensions of NLLS, like Instrumental Variables Non-Linear Least Squares (IV-NLLS).

2SLS as a GMM estimator (1/4)

Recall the IV estimator framed as a MM estimator, and let $\boldsymbol{z}_i$ have dimension $J \geq K$ (the earlier motivation for GMM). The moment conditions' sample analogs are as follows.

$$\overline{\mathbf{g}}_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \left(y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} \right) = 0$$

The associated GMM estimator is:

$$\widehat{\boldsymbol{\beta}}_{GMM} =$$

$$= \underset{\boldsymbol{\beta} \in \mathbb{R}^K}{\arg\min} \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \left(y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} \right) \right]^{\mathrm{T}} \mathbf{A}_N \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \left(y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} \right) \right]$$

which has an **analytic solution**. The First Order Conditions of the problem above are as follows.

$$-2 \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i^{\mathrm{T}} \right] \mathbf{A}_N \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \left(y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{GMM} \right) \right] = 0$$

2SLS as a GMM estimator (2/4)

The solution is expressed as:

$$\widehat{\boldsymbol{\beta}}_{GMM} =$$
$$= \left[\left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i^{\mathrm{T}}\right) \mathbf{A}_N \left(\sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i^{\mathrm{T}}\right)\right]^{-1} \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i^{\mathrm{T}}\right) \mathbf{A}_N \sum_{i=1}^{N} \mathbf{z}_i y_i$$

or, in compact matrix notation, as follows.

$$\widehat{\boldsymbol{\beta}}_{GMM} = \left(\mathbf{X}^{\mathrm{T}} \mathbf{Z} \mathbf{A}_N \mathbf{Z}^{\mathrm{T}} \mathbf{X}\right)^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{Z} \mathbf{A}_N \mathbf{Z}^{\mathrm{T}} \mathbf{y}$$

This *looks like* the 2SLS estimator. If the weighting matrix $\mathbf{A}_N$ were chosen as:

$$\widetilde{\mathbf{A}}_N = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}}\right)^{-1} = \left(\frac{1}{N} \mathbf{Z}^{\mathrm{T}} \mathbf{Z}\right)^{-1}$$

the solution would be *exactly* the 2SLS estimator.

2SLS as a GMM estimator (3/4)

The theory of GMM allows additional efficiency gains. If, under the assumption that *the observations are independent* $\mathbf{A}_N$ were:

$$\begin{aligned}\mathbf{A}_N = \widehat{\mathbf{\Omega}}_N^{-1} &= \left\{\widehat{\mathbb{A}\text{var}}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{z}_i\left(y_i - \mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{GMM}\right)\right]\right\}^{-1} \\ &= \left(\frac{1}{N}\sum_{i=1}^{N} e_i^2 \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}}\right)^{-1} \\ &= \left(\frac{1}{N}\mathbf{Z}^{\mathrm{T}}\widehat{\mathbf{E}}_N\mathbf{Z}\right)^{-1}\end{aligned}$$

where $e_i \equiv y_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{GMM}$ for $i = 1, \dots, N$ and $\widehat{\mathbf{E}}_N$ is as defined in Lecture 10. The GMM estimator would thus become:

$$\widehat{\boldsymbol{\beta}}_{GMM} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\widehat{\mathbf{E}}_N\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\widehat{\mathbf{E}}_N\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{y}$$

which differs slightly from standard 2SLS, as this a generalized version (in the GLS sense) of the overidentified 2SLS estimator.

2SLS as a GMM estimator (4/4)

The estimated asymptotic variance is no longer "sandwiched."

$$\widehat{\mathbb{A}\text{var}}\left[\widehat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{N}\left[\mathbf{X}^{\mathrm{T}}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\widehat{\mathbf{E}}_N\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}\right]^{-1}$$

How does optimal GMM for this problem fare against standard 2SLS? In the extreme i.i.d. case, it is:

$$\mathbf{A}_N^{-1} = \widehat{\boldsymbol{\Omega}}_N = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}} \xrightarrow{p} \sigma^2\, \mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}\right]$$

where $\sigma^2 \equiv \mathbb{V}\text{ar}\left[Y_i - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right]$ does not depend on $\boldsymbol{z}_i$; while:

$$\widetilde{\mathbf{A}}_N^{-1} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}} \xrightarrow{p} \mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^{\mathrm{T}}\right]$$

therefore the two estimators would asymptotically **coincide**, as σ^2 would vanish from the analytical expression for $\widehat{\boldsymbol{\beta}}_{GMM}$. The equivalence **collapses** beyond the i.i.d. case, but the differences are typically minimal in practical settings.

3SLS as a GMM estimator (1/6)

Recall the SEM model estimated by 3SLS from Lecture 10:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$$

that is more extensively written as follows.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_P \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_{20} \\ \vdots \\ \boldsymbol{\beta}_{P0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_P \end{bmatrix}$$

with dimension $N \times P$ (N observations for P equations).

Also 3SLS is a GMM estimator. The moment conditions are:

$$\mathbb{E}\left[\varepsilon_{pi} \middle| \boldsymbol{z}_i\right] = 0 \Rightarrow \mathbb{E}\left[\boldsymbol{z}_i \varepsilon_{pi}\right] = \mathbf{0}$$

for all equations $p = 1, \ldots, P$, and where $\boldsymbol{z}_i$ are the exogenous variables.

3SLS as a GMM estimator (2/6)

The sample analog of the moment conditions is:

$$\overline{\mathbf{g}}_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\begin{pmatrix} y_{1i} - \mathbf{x}_{1i}^{\mathrm{T}}\boldsymbol{\beta}_1 \\ y_{2i} - \mathbf{x}_{2i}^{\mathrm{T}}\boldsymbol{\beta}_2 \\ \vdots \\ y_{Pi} - \mathbf{x}_{Pi}^{\mathrm{T}}\boldsymbol{\beta}_P \end{pmatrix} = \mathbf{0}$$

a PQ-dimensional vector (P equations for Q instruments). It is better to express this in compact matrix notation:

$$\frac{1}{N}\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

where the Kronecker product diagonally stacks the transpose of matrix $\mathbf{Z}$ just P times (it has dimension $PQ \times PN$).

$$\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}} = \begin{bmatrix} \mathbf{Z}^{\mathrm{T}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}^{\mathrm{T}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}^{\mathrm{T}} \end{bmatrix}$$

3SLS as a GMM estimator (3/6)

The GMM problem is written as:

$$\widehat{\boldsymbol{\beta}}_{GMM} = \underset{\boldsymbol{\beta}\in\mathbb{R}^B}{\arg\min} \left[\frac{1}{N}\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]^{\mathrm{T}} \cdot \\ \cdot \mathbf{A}_N \cdot \left[\frac{1}{N}\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]$$

where $B = |\boldsymbol{\beta}|$. The First Order Conditions are:

$$-2\left[\frac{1}{N}\mathbf{X}^{\mathrm{T}}(\mathbf{I}\otimes\mathbf{Z})\right]\mathbf{A}_N\left[\frac{1}{N}\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)\left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{GMM}\right)\right] = \mathbf{0}$$

with solution:

$$\widehat{\boldsymbol{\beta}}_{GMM} = \left[\mathbf{X}^{\mathrm{T}}(\mathbf{I}\otimes\mathbf{Z})\mathbf{A}_N\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)\mathbf{X}\right]^{-1} \cdot \\ \cdot \mathbf{X}^{\mathrm{T}}(\mathbf{I}\otimes\mathbf{Z})\mathbf{A}_N\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)\mathbf{y}$$

which again varies with the weighting matrix $\mathbf{A}_N$.

3SLS as a GMM estimator (4/6)

This estimator is equivalent to 3SLS if the weighting matrix is:

$$\begin{aligned}\widetilde{\mathbf{A}}_N &= \left[\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right)\left(\widehat{\boldsymbol{\Sigma}}_N\otimes\mathbf{I}\right)(\mathbf{I}\otimes\mathbf{Z})\right]^{-1}\\ &= \left[\widehat{\boldsymbol{\Sigma}}_N\otimes\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right]^{-1} = \widehat{\boldsymbol{\Sigma}}_N^{-1}\otimes\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\end{aligned}$$

with $\widehat{\boldsymbol{\Sigma}}_N$ the estimate of the conditional error covariance matrix $\boldsymbol{\Sigma}$, as in Lecture 10. Because projection matrices are symmetric and idempotent, it is:

$$\begin{aligned}(\mathbf{I}\otimes\mathbf{Z})\left[\widehat{\boldsymbol{\Sigma}}_N^{-1}\otimes\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\right]\left(\mathbf{I}\otimes\mathbf{Z}^{\mathrm{T}}\right) &= \widehat{\boldsymbol{\Sigma}}_N^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\\ &= \left(\widehat{\boldsymbol{\Sigma}}_N^{-1}\otimes\mathbf{I}\right)(\mathbf{I}\otimes\mathbf{P}_{\mathbf{Z}})\\ &= (\mathbf{I}\otimes\mathbf{P}_{\mathbf{Z}})\left(\widehat{\boldsymbol{\Sigma}}_N^{-1}\otimes\mathbf{I}\right)(\mathbf{I}\otimes\mathbf{P}_{\mathbf{Z}})\end{aligned}$$

where $\mathbf{P}_{\mathbf{Z}} = \mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}$ – showing equivalence with 3SLS if the latter is obtained under both hypotheses of within-equation homoscedasticity and cross-equation dependence.

3SLS as a GMM estimator (5/6)

Further efficiency gains are obtained with an optimal weighting matrix. Under the hypothesis of *independent observations*, it is:

$$\mathbf{A}_N = \widehat{\mathbf{\Omega}}_N^{-1} = \left\{ \widehat{\mathbb{A}\text{var}} \left[\frac{1}{\sqrt{N}} \left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \left(\mathbf{y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{GMM} \right) \right] \right\}^{-1}$$

$$= \left(\frac{1}{N} \left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \widehat{\mathbf{S}}_N \left(\mathbf{I} \otimes \mathbf{Z} \right) \right)^{-1}$$

where $\widehat{\mathbf{S}}_N$ is analogous to $\widehat{\mathbf{\Sigma}}_N \otimes \mathbf{I}$, yet slightly more complex:

$$\widehat{\mathbf{S}}_N = \begin{bmatrix} \widehat{\mathbf{S}}_{N,11} & \widehat{\mathbf{S}}_{N,12} & \cdots & \widehat{\mathbf{S}}_{N,1P} \\ \widehat{\mathbf{S}}_{N,21} & \widehat{\mathbf{S}}_{N,22} & \cdots & \widehat{\mathbf{S}}_{N,2P} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathbf{S}}_{N,P1} & \widehat{\mathbf{S}}_{N,P2} & \cdots & \widehat{\mathbf{S}}_{N,PP} \end{bmatrix}$$

To express it properly, define the following for $i = 1, \ldots, N$.

$$e_{pqi} = \left(y_{pi} - \mathbf{x}_{pi}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{pGMM} \right) \left(y_{qi} - \mathbf{x}_{qi}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{qGMM} \right)$$

(Continues. . .)

3SLS as a GMM estimator (6/6)

(Continued.) Thus, for any $p, q = 1, \dots, P$:

$$\widehat{\mathbf{S}}_{N,pq} = \begin{bmatrix} e^2_{pp1} & 0 & \cdots & 0 \\ 0 & e^2_{pq2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^2_{pqN} \end{bmatrix}$$

The resulting GMM estimator (which is seldom used) is:

$$\widehat{\boldsymbol{\beta}}_{GMM} = \left\{ \mathbf{X}^{\mathrm{T}} \left(\mathbf{I} \otimes \mathbf{Z} \right) \left[\left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \widehat{\mathbf{S}}_N \left(\mathbf{I} \otimes \mathbf{Z} \right) \right]^{-1} \left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \mathbf{X} \right\}^{-1} \times \\ \times \mathbf{X}^{\mathrm{T}} \left(\mathbf{I} \otimes \mathbf{Z} \right) \left[\left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \widehat{\mathbf{S}}_N \left(\mathbf{I} \otimes \mathbf{Z} \right) \right]^{-1} \left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \mathbf{y}$$

and its asymptotic variance-covariance is estimated as follows.

$$\widehat{\mathbb{A}\mathrm{var}} \left[\widehat{\boldsymbol{\beta}}_{GMM} \right] = \\ = \frac{1}{N} \left\{ \mathbf{X}^{\mathrm{T}} \left(\mathbf{I} \otimes \mathbf{Z} \right) \left[\left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \widehat{\mathbf{S}}_N \left(\mathbf{I} \otimes \mathbf{Z} \right) \right]^{-1} \left(\mathbf{I} \otimes \mathbf{Z}^{\mathrm{T}} \right) \mathbf{X} \right\}^{-1}$$

IVs in non-linear models (1/4)

The GMM framework is well suited to extend the Instrumental Variables approach to **non-linear** estimators like (for example) generalizations of the MLE score that allow for *exogenous* IVs.

The GMM problem associated with $\mathbb{E}\left[\boldsymbol{z}_i \otimes \boldsymbol{h}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\theta}_0\right)\right] = \mathbf{0}$ is:

$$\widehat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i \cdot \boldsymbol{h}\left(\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}\right)\right]^{\mathrm{T}} \cdot \mathbf{A}_N \cdot \left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i \cdot \boldsymbol{h}\left(\mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}\right)\right]$$

which generally lacks an explicit solution. Yet an expression for both the limiting and asymptotic variances can be obtained via the theory of GMM; one notices that, in this case:

$$\mathbf{G}_0 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\boldsymbol{z}_i \cdot \frac{\partial}{\partial\boldsymbol{\theta}} \boldsymbol{h}^{\mathrm{T}}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\theta}_0\right)\right]$$

and similarly for $\mathbf{G}_N\left(\boldsymbol{\theta}\right)$.

IVs in non-linear models (2/4)

A typical application is in single-equation non-linear models if:

$$\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}h\left(\mathbf{x}_i;\boldsymbol{\theta}_0\right)\cdot\left(y_i-h\left(\mathbf{x}_i;\boldsymbol{\theta}_0\right)\right)\right]\neq\mathbf{0}$$

the error term $\varepsilon_i = y_i - h\left(\mathbf{x}_i;\boldsymbol{\theta}_0\right)$ is **not mean-independent** of the implicit set of instruments that is defined by the Non-Linear Least Squares estimator, because of some case of endogeneity.

The solution is to use a J-dimensional vector of IVs $\boldsymbol{z}_i$ with:

$$\mathbb{E}\left[\mathbf{z}_i\cdot\boldsymbol{h}\left(\boldsymbol{y}_i,\boldsymbol{z}_i;\boldsymbol{\theta}\right)\right]=\mathbb{E}\left[\boldsymbol{z}_i\left(Y_i-h\left(\boldsymbol{x}_i;\boldsymbol{\theta}_0\right)\right)\right]=\mathbf{0}$$

hence the **Non-Linear Two-Stages Least Squares** (NL2SLS) estimator follows from the solution of this GMM problem. Note:

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{NL2SLS}-\boldsymbol{\theta}_0\right)\stackrel{d}{\rightarrow}$$
$$\mathcal{N}\left(\mathbf{0},\left(\mathbf{J}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{J}_0\right)^{-1}\mathbf{J}_0^{\mathrm{T}}\mathbf{A}_0\boldsymbol{\Omega}_0\mathbf{A}_0\mathbf{J}_0\left(\mathbf{J}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{J}_0\right)^{-1}\right)$$

(Continues. . .)

IVs in non-linear models (3/4)

(Continued.) …where $\mathbf{J}_0$, the analogue of $\mathbf{G}_0$, is defined as:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{h}_{0i}^{\mathrm{T}} \xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{h}_{0i}^{\mathrm{T}}\right] \equiv \mathbf{J}_0$$

$\boldsymbol{h}_{0i}$ is as in Lecture 11, while $\mathbf{A}_0$ and $\mathbf{\Omega}_0$ are as before.

The estimated asymptotic variance of the NL2SLS estimator is:

$$\widehat{\mathbb{A}\mathrm{var}}\left[\widehat{\boldsymbol{\theta}}_{NL2SLS}\right] =$$
$$= \frac{1}{N}\left(\widehat{\mathbf{J}}_N^{\mathrm{T}}\mathbf{A}_N\widehat{\mathbf{J}}_N\right)^{-1}\widehat{\mathbf{J}}_N^{\mathrm{T}}\mathbf{A}_N\widehat{\mathbf{\Omega}}_N\mathbf{A}_N\widehat{\mathbf{J}}_N\left(\widehat{\mathbf{J}}_N^{\mathrm{T}}\mathbf{A}_N\widehat{\mathbf{J}}_N\right)^{-1}$$

where:

$$\widehat{\mathbf{J}}_N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\widehat{\mathbf{h}}_i^{\mathrm{T}}$$

and $\widehat{\mathbf{h}}_i$ is as in Lecture 11.

IVs in non-linear models (4/4)

The particular choice of the weighting matrix entails the same considerations as in the case of "linear" GMM-2SLS:

1. $\mathbf{A}_N = N \left(\sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}} \right)^{-1}$ delivers the traditional (standard) version of the NL2SLS estimator, akin to standard 2SLS;

2. with **independent** observations, $\widehat{\mathbf{\Omega}}_N = N \left(\sum_{i=1}^{N} e_i^2 \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}} \right)^{-1}$ where e_i is the usual residual for each i-th observation;

3. in such a case, however, the **optimal** NL2SLS estimator is obtained by setting $\mathbf{A}_N = \widehat{\mathbf{\Omega}}_N^{-1}$, and like in the 2SLS-3SLS cases it is asymptotically equivalent to "standard" NL2SLS only under homoscedasticity.

Note that if one specifies $\boldsymbol{z}_i = \boldsymbol{h}_{0i}$, so that the instruments enter as $\mathbf{z}_i = \widehat{\mathbf{h}}_i$ in the estimation problem, GMM yields the standard NLLS estimator from Lecture 11.

Optimal Instruments (1/3)

- In this framework, **every** function of the instruments $\boldsymbol{l}(\boldsymbol{z}_i)$ which takes values upon a J'-dimensional set yields **valid** moment conditions of the kind:

$$\mathbb{E}\left[\boldsymbol{l}(\boldsymbol{z}_i) \otimes \boldsymbol{h}(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta_0)\right] = \mathbf{0}$$

  so long as $PJ' \geq K$, where K is the number of parameters.

- A relevant question is to what extent it is possible to build appropriate **optimal instruments**, so that GMM delivers the **most efficient** estimate available with the information enclosed in the conditional moment conditions.

- It can be shown that this objective is achieved via matrix:

$$\begin{aligned}\mathbf{L}(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta_0) &= \\ &= \mathbb{E}\left[\left.\frac{\partial}{\partial \theta}\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta_0)\right| \boldsymbol{z}_i\right]\left\{\mathbb{V}\mathrm{ar}\left[\boldsymbol{h}(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta_0)\right| \boldsymbol{z}_i]\right\}^{-1}\end{aligned}$$

  where the first term is $K \times P$, while the second is $P \times P$.

Optimal Instruments (2/3)

- The efficient estimate of $\boldsymbol{\theta}_0$ is then obtained through the following K "optimal" moment conditions:

$$\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\theta}_0\right)\right] = \mathbb{E}\left[\mathbf{L}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\theta}_0\right) \cdot \boldsymbol{h}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\theta}_0\right)\right] = \mathbf{0}$$

- ...and the corresponding **estimate** $\widehat{\boldsymbol{\theta}}_{MM}$ solves a simple Method of Moments sample analog system of equations.

$$\frac{1}{N}\sum_{i=1}^{N}\left[\mathbf{L}\left(\mathbf{y}_i, \mathbf{z}_i; \widehat{\boldsymbol{\theta}}_{MM}\right) \cdot \boldsymbol{h}\left(\mathbf{y}_i, \mathbf{z}_i; \widehat{\boldsymbol{\theta}}_{MM}\right)\right] = \mathbf{0}$$

- Since the model is just-identified, $\mathbf{A}_N$ is redundant.

- While this result holds in theory, any practical application requires *a priori* knowledge of the elements that enter into $\mathbf{L}\left(\boldsymbol{y}_i, \boldsymbol{z}_i; \boldsymbol{\theta}_0\right)$, which is generally not the case.

- Thus, practice favors simple, untransformed IVs $\boldsymbol{z}_i$.

Optimal Instruments (3/3)

- This is best illustrated with the linear model. In fact, this problem is analogous to the specification of $\mathbf{\Sigma}$ in "feasible" GLS for small samples.

- This is not just an analogy: if $\boldsymbol{z}_i = \boldsymbol{x}_i$, $P = 1$ and:

$$\boldsymbol{h}\left(Y_i, \boldsymbol{x}_i; \boldsymbol{\beta}_0\right) = Y_i - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0$$

  the resulting Method of Moments estimator is just GLS!

$$\mathbf{L}_{GLS}\left(Y_i, \boldsymbol{x}_i; \boldsymbol{\beta}_0\right) = -\boldsymbol{x}_i \cdot \frac{1}{\sigma_L^2\left(\boldsymbol{x}_i\right)}$$

- Similarly, if $\boldsymbol{z}_i = \boldsymbol{x}_i$, $P = 1$ but the model is non-linear:

$$\boldsymbol{h}\left(Y_i, \boldsymbol{x}_i; \theta_0\right) = Y_i - h\left(\boldsymbol{x}_i; \theta_0\right)$$

  the model yields the "Generalized" version of NLLS.

$$\mathbf{L}_{GNLLS}\left(Y_i, \boldsymbol{x}_i; \theta_0\right) = -\boldsymbol{h}_{0i} \cdot \frac{1}{\sigma_{NL}^2\left(\boldsymbol{x}_i\right)}$$

Testing overidentification (1/3)

- Is it possible to **test** that the zero moment condition that motivate GMM are valid in the data (real world)?

- Well, not always...but in **overidentified** models, one can test **some** of them if the others are maintained.

- The ensuing discussion assumes overidentification: $J > K$.

- The starting point is the **Hansen *J*** statistic:

$$\mathrm{J}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) = N\,\overline{\mathbf{g}}_N^{\mathrm{T}}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right)\widehat{\boldsymbol{\Omega}}_N^{-1}\overline{\mathbf{g}}_N\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) \xrightarrow{d} \chi^2_{J-K}$$

  the GMM quadratic objective evaluated at $\widehat{\boldsymbol{\theta}}_{GMM}$ times N.

- Under the **null** hypothesis that the moment conditions are "true," the statistic should be close to zero.

- Too high values may lead to rejecting the null hypothesis.

Testing overidentification (2/3)

- The Hansen J statistic is not very helpful if there are many moments relative to the number of parameters: $J >> K$.

- In this case, one may want to test **blocks** of moments. The so-called **"incremental" Sargan test** performs this task.

- Intuitively, if some moments are **redundant** the others can be used to estimate θ_0; such an estimate is used to test the redundant moments.

- Suppose that the moment conditions are split in two blocks:

$$\mathbb{E}\left[\boldsymbol{g}\left(\boldsymbol{x}_i;\theta_0\right)\right] = \mathbb{E}\begin{bmatrix}\boldsymbol{g}_1\left(\boldsymbol{x}_i;\theta_0\right)\\ \boldsymbol{g}_2\left(\boldsymbol{x}_i;\theta_0\right)\end{bmatrix} = 0$$

  with $\left|\boldsymbol{g}_1\left(\boldsymbol{x}_i;\theta_0\right)\right| = J_1 > K$, $\left|\boldsymbol{g}_2\left(\boldsymbol{x}_i;\theta_0\right)\right| = J_2$, $J_1 + J_2 = J$.

- Assume that $\mathbb{E}\left[\boldsymbol{g}_1\left(\boldsymbol{x}_i;\theta_0\right)\right] = 0$ is **true**; the null hypothesis is $\mathrm{H}_0: \ \mathbb{E}\left[\boldsymbol{g}_2\left(\boldsymbol{x}_i;\theta_0\right)\right] = 0$.

Testing overidentification (3/3)

- The "incremental" **Sargan statistic** is:

$$\mathrm{J}_S\left(\widehat{\boldsymbol{\theta}}_{GMM}, \widetilde{\boldsymbol{\theta}}\right) = \mathrm{J}\left(\widehat{\boldsymbol{\theta}}_{GMM}\right) - \widetilde{\mathrm{J}}\left(\widetilde{\boldsymbol{\theta}}\right) \xrightarrow{d} \chi^2_{J_2}$$

  for:

$$\widetilde{\mathrm{J}}\left(\widetilde{\boldsymbol{\theta}}\right) = N\overline{\mathbf{g}}^{\mathrm{T}}_{N1}\left(\widetilde{\boldsymbol{\theta}}\right)\widehat{\boldsymbol{\Omega}}^{-1}_{N_1}\overline{\mathbf{g}}_{N1}\left(\widetilde{\boldsymbol{\theta}}\right) \xrightarrow{d} \chi^2_{J_1-K}$$

  and:

$$\widetilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \overline{\mathbf{g}}^{\mathrm{T}}_{N1}(\boldsymbol{\theta})\,\widehat{\boldsymbol{\Omega}}^{-1}_{N_1}\overline{\mathbf{g}}_{N1}(\boldsymbol{\theta})$$

  where here $\overline{\mathbf{g}}_{N1}(\boldsymbol{\theta})$ is the sample mean of $\boldsymbol{g}_1(\mathbf{x}_i;\boldsymbol{\theta})$, $\widehat{\boldsymbol{\Omega}}_{N_1}$ is some consistent **estimate** of its variance-covariance matrix and $\mathrm{J}(\cdot)$ is Hansen's J-statistic.

- Intuitively, the Sargan test results from the original Hansen J-statistics, *minus* **another** Hansen J-statistic that obtains from a "reduced" GMM model only based on the first block of J_1 moment conditions.

Example: overidentified Mincer equation (1/2)

- Return to the Mincer equation from Lectures 7, 9 and 10.

$$\log W_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 S_i + \alpha_i + \epsilon_i$$

- Suppose that there are **three** IVs: "distance from college" Z_i; "fellowship grant" G_i and "friends in college" F_i.

- Some of these instruments may not be exogenous!

- The moment conditions are as follows.

$$\mathbb{E}\left[\begin{pmatrix} 1 \\ X_i \\ X_i^2 \\ Z_i \\ G_i \\ F_i \end{pmatrix} \underbrace{\left(\log W_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 S_i\right)}_{=\alpha_i + \epsilon_i}\right] = 0$$

Example: overidentified Mincer equation (2/2)

- This model is **estimable** via 2SLS-GMM for $Y_i = \log W_i$, $\boldsymbol{x}_i^{\mathrm{T}} = (1, X_i, X_i^2, S_i)$ and $\boldsymbol{z}_i^{\mathrm{T}} = (1, X_i, X_i^2, Z_i, G_i, F_i)$.

- Moreover, one can test if either moment condition specific to one of the three IVs is invalid, *one at the time*.

- The Hansen J-statistic here is:

$$\mathrm{J}\left(\widehat{\boldsymbol{\beta}}_{2SLS}\right) = N \left[\sum_{i=1}^{N} \mathbf{z}_i e_i\right]^{\mathrm{T}} \left[\sum_{i=1}^{N} e_i^2 \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}}\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{z}_i e_i\right] \xrightarrow{d} \chi_2^2$$

  where, as usual, $e_i = y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{2SLS}$.

- To proceed, one would sequentially remove each IV: Z_i, G_i, F_i from the model, calculate the Hansen J-statistic for such a reduced model, and then the associated Sargan statistic.

- Rejecting a Sargan test is suggestive that an IV is invalid.

Unsolvable Moment Conditions

- Occasionally, the functions $\boldsymbol{g}(\boldsymbol{x}_i; \boldsymbol{\theta})$ that express the GMM moment conditions are difficult to calculate, like when they are defined by integrals without closed form solution.

- Like in the analogous problem with Maximum (Likelihood) Estimation, this one calls for **simulation-based** solutions.

- Let the moment conditions be $\mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_i; \boldsymbol{\theta}_0)] = \mathbf{0}$, where:

$$\boldsymbol{g}(\mathbf{x}_i; \boldsymbol{\theta}) = \int_{\mathbb{U}} \boldsymbol{g}_{\boldsymbol{u}}(\mathbf{x}_i, \mathbf{u}_i; \boldsymbol{\theta})\, dH_{\boldsymbol{u}}(\mathbf{u}_i)$$

  and where $\boldsymbol{u}_i$ is a random vector with c.d.f. $H_{\boldsymbol{u}}(\mathbf{u}_i)$ that is integrated out over its support $\mathbb{U}$. This integral may lack a closed form solution given $\boldsymbol{\theta}$ and $\mathbf{x}_i$.

- Simulators would be based upon a set $\{\mathbf{u}_s\}_{s=1}^{S}$ of S random draws of $\boldsymbol{u}_i$ from $H_{\boldsymbol{u}}(\mathbf{u}_i)$: Direct **Monte Carlo** Sampling.

Method of Simulated Moments (1/2)

- A consistent moment **simulator** is, for $i = 1, \dots, N$:

$$\widehat{\boldsymbol{g}}_S\left(\mathbf{x}_i; \boldsymbol{\theta}\right) = \frac{1}{S}\sum_{s=1}^{S} \widetilde{\boldsymbol{g}}_{\boldsymbol{u}}\left(\mathbf{x}_i, \mathbf{u}_s; \boldsymbol{\theta}\right)$$

where $\widetilde{\boldsymbol{g}}_{\boldsymbol{u}}\left(\mathbf{x}_i, \mathbf{u}_s; \boldsymbol{\theta}\right)$ is a **subsimulator**.

- The subsimulator is ideally **unbiased**, so that it guarantees $\widehat{\boldsymbol{g}}_S\left(\mathbf{x}_i; \boldsymbol{\theta}\right) \xrightarrow{p} \boldsymbol{g}\left(\mathbf{x}_i; \boldsymbol{\theta}\right)$ by standard asymptotic arguments.

- The **Method of Simulated Moments** (MSM) estimator is defined as:

$$\widehat{\boldsymbol{\theta}}_{MSM} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\min} \left[\frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{g}}_S\left(\mathbf{x}_i; \boldsymbol{\theta}\right)\right]^{\mathrm{T}} \mathbf{A}_N \left[\frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{g}}_S\left(\mathbf{x}_i; \boldsymbol{\theta}\right)\right]$$

where $\mathbf{A}_N$ is a $J \times J$ weighting matrix that has probability limit $\mathbf{A}_0$, as in standard GMM.

Method of Simulated Moments (2/2)

- The MSM estimator accommodates overidentification, like regular GMM itself.

- However, the issue of unsolvable moments could very easily occur in just-identified models ($J = K$). In such cases, it is $\mathbf{A}_N = \mathbf{A}_0 = \mathbf{I}$ (again like in regular GMM).

- If the moment function is $\boldsymbol{g}(\boldsymbol{x}_i; \theta) = \boldsymbol{z}_i \otimes \boldsymbol{h}(\boldsymbol{y}_i, \boldsymbol{z}_i; \theta)$: that is, it is based on instrumental variables, the simulator is:

$$\widehat{\boldsymbol{g}}_S(\mathbf{x}_i; \theta) = \frac{1}{S} \sum_{s=1}^{S} \mathbf{z}_i \widetilde{\boldsymbol{h}}_{\boldsymbol{u}}(\mathbf{y}_i, \mathbf{z}_i, \mathbf{u}_s; \theta)$$

  where $\widetilde{\boldsymbol{h}}_{\boldsymbol{u}}(\mathbf{y}_i, \mathbf{z}_i, \mathbf{u}_s; \theta)$ is again a suitable subsimulator that here applies for $\boldsymbol{h}(\mathbf{y}_i, \mathbf{z}_i; \theta)$.

- This case is illustrated through the next example.

Example: IV random coefficients logit (1/2)

- Recall the random coefficients logit model from Lecture 11. That model leads to the moment condition:

$$\mathbb{E}\left[Y_i - \Lambda\left(\beta_0 + \beta_{1i}X_i\right)\middle|\, X_i\right] = 0$$

where $\Lambda(\cdot)$ is the **logistic** c.d.f. of the **binary outcome** Y_i conditional on X_i.

- This moment can be interpreted like in linear models, more easily so if one defines the "error" $Y_i - \Lambda\left(\beta_0 + \beta_{1i}X_i\right)$.

- This moment might be difficult to calculate. Let once again be $\beta_{1i} \sim \mathcal{N}\left(\beta_1, \sigma^2\right)$ and $u_i = \left(\beta_{1i} - \beta_1\right)/\sigma$, then:

$$\Lambda\left(\beta_0 + \beta_{1i}X_i\right) = \int_{\mathbb{R}} \Lambda\left(\beta_0 + \left(\beta_1 + \sigma u_i\right)X_i\right)\phi\left(u_i\right)du_i$$

and the integral clearly lacks a closed form solution.

Example: IV random coefficients logit (2/2)

- A **just-identified** MSM estimator for $\theta = (\beta_0, \beta_1, \sigma^2)$ here would be based on the simulator:

$$\widehat{g}_S(y_i, x_i; \theta) = \frac{1}{N}\sum_{i=1}^{N} x_i \left[y_i - \frac{1}{S}\sum_{s=1}^{S} \widetilde{\Lambda}(\beta_0 + (\beta_1 + \sigma u_s) x_i)\right]$$

given the subsimulator $\widetilde{\Lambda}(\beta_0 + (\beta_1 + \sigma u_s) x_i)$.

- Now, suppose that because of endogeneity, the regressor X_i is correlated with the model error $Y_i - \Lambda(\beta_0 + \beta_{1i} X_i)$. Yet, $J \geq K$ valid IVs $\boldsymbol{z}_i$ are available so that the following holds.

$$\mathbb{E}[Y_i - \Lambda(\beta_0 + \beta_{1i} X_i) | \boldsymbol{z}_i] = 0$$

- Hence, in this setting the simulator

$$\widehat{g}_S(y_i, x_i, \mathbf{z}_i; \theta) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i \left[y_i - \frac{1}{S}\sum_{s=1}^{S} \widetilde{\Lambda}(\beta_0 + (\beta_1 + \sigma u_s) x_i)\right]$$

affords a suitable **overidentified** MSM estimator.

Method of Simulated Moments: Observations

- To properly discuss the asymptotic properties of the MSM estimator, it is useful to make some observations.

- Let the $J \times J$ limiting variance-covariance of the simulators be:

$$\widetilde{\boldsymbol{\Omega}}_0 \equiv \lim_{N \to \infty} \mathbb{V}\text{ar}\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{\boldsymbol{g}}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right]$$

this matrix also features noise due to the simulation!

- One can show that, by the Law of Total Variance:

$$\widetilde{\boldsymbol{\Omega}}_0 = \boldsymbol{\Omega}_0 + \lim_{N \to \infty} \mathbb{E}_{\boldsymbol{x}}\left[\mathbb{V}\text{ar}_{\boldsymbol{u}}\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{\boldsymbol{g}}\left(\boldsymbol{x}_i; \boldsymbol{\theta}_0\right)\right]\right]$$

with $\boldsymbol{\Omega}_0$ as in regular GMM. Regarding the second element on the right-hand side, the **outer** expectation is taken with respect to $\boldsymbol{x}_i$ – while the **inner** variance-covariance is taken with respect to $\boldsymbol{u}_i$. As $S \to \infty$, this element **vanishes**.

Method of Simulated Moments: Asymptotics

The following result was originally given by McFadden (1989).

Theorem 3

Asymptotic Efficiency of the Method of Simulated Moments estimators. *If an MSM estimator is based upon an ubiased simulator* $\widehat{\boldsymbol{g}}_S(\boldsymbol{x}_i;\boldsymbol{\theta})$*, and all conditions implicit in the statement of Theorem 1 are met, even with fixed* S *the estimator is consistent and its limiting distribution is as follows.*

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{MSM}-\boldsymbol{\theta}_0\right)\stackrel{d}{\to}$$
$$\mathcal{N}\left(\mathbf{0},\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\widetilde{\boldsymbol{\Omega}}_0\mathbf{A}_0\mathbf{G}_0\left(\mathbf{G}_0^{\mathrm{T}}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\right)$$

Proof.

(Outline.) The proof proceeds through the manipulation of the First Order Conditions of the MSM problem as in the proof of Theorem 1, relying on the unbiasedness of the simulator for the sake of simplifying some key expressions. □

Method of Simulated Moments: Remarks

- This asymptotic result is better than the analogous one for SML: here, consistency does **not** require $S \to \infty$!

- This is quite an advantage for MSM over SML – besides all usual considerations about MM versus MLE.

- However, having a large simulation size is still helpful, as it makes $\widetilde{\mathbf{\Omega}}_0$ asymptotically coincident with $\mathbf{\Omega}_0$.

- When performing **inference** on the MSM estimator, $\mathbf{A}_0$, $\mathbf{G}_0$ and $\mathbf{\Omega}_0$ must be estimated. $\mathbf{A}_0$ is "estimated" by $\mathbf{A}_N$.

- Regarding $\mathbf{G}_0$ and $\mathbf{\Omega}_0$, they are estimated using the formulae from regular GMM, but **substituting** $\boldsymbol{g}(\mathbf{x}_i; \boldsymbol{\theta})$ evaluated at the GMM estimator with the simulator $\widehat{\boldsymbol{g}}_S(\mathbf{x}_i; \boldsymbol{\theta})$ evaluated at the MSM estimator.

- With small S, the second element of $\widetilde{\mathbf{\Omega}}_0$ must be estimated.

Applications of GMM: overview

The importance of the Generalized Method of Moments is largely theoretical, as it encompasses many common estimators.

Consequently, it is hard to identify non-generic examples of GMM applications: GMM is a first of all a *framework*. What follows is an attempt to illustrate the usefulness of this framework.

Three concrete "domains" of GMM are overviewed next.

1. Dynamic linear models for panel data.

2. The estimation of production functions.

3. The estimation of rational expectations models.

Dynamic linear models for panel data (1/3)

- Consider the following model for panel data.

$$y_{it} = \boldsymbol{\alpha} + \mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\beta} + \gamma y_{i(t-1)} + \varepsilon_{it}$$

  In many macroeconomic models the past affects the present: this is captured here by parameter γ.

- Time feedback may also affect the error term, which in such a model is typically **autoregressive**:

$$\varepsilon_{it} = \rho\varepsilon_{i(t-1)} + \xi_{it}$$

  for some $\rho \in [-1, 1]$, and where ξ_{it} is i.i.d. ("white noise").

- This leads to endogeneity, since by construction:

$$\mathbb{E}[Y_{i(t-1)}\varepsilon_{it}] \neq 0$$

  and OLS estimation is therefore inconsistent.

Dynamic linear models for panel data (2/3)

- Moreover, typical solutions to **unobserved heterogeneity** are not available in this model. Suppose that $\alpha = 0$ but the error term features a **fixed effect**.

$$\varepsilon_{it} = \alpha_i + \epsilon_{it}$$

- By construction, such a fixed effect is correlated at the very least with the lagged outcome variable.

$$\mathbb{E}[Y_{i(t-1)}\alpha_{it}] \neq 0$$

- The within transformation here reads:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^{\mathrm{T}} \boldsymbol{\beta} + \gamma \left(y_{i(t-1)} - \bar{y}_i \right) + \varepsilon_{it} - \bar{\varepsilon}_i$$

  yet **endogeneity** persists, as the demeaned lagged outcome is mechanically correlated to the demeaned error term.

$$\mathbb{E}\left[\left(Y_{i(t-1)} - \bar{Y}_i \right) (\varepsilon_{it} - \bar{\varepsilon}_i) \right] \neq 0$$

Dynamic linear models for panel data (3/3)

The GMM approach for this class of models is based on two types of moment conditions, which can be possibly combined within a larger overidentified system of moments.

- **Moment in levels**, featuring a product between the first difference of the error term and the higher lags (one period and beyond) of the dependent variable (Arellano and Bond, 1991). For $s \geq 2$, this is:

$$\mathbb{E}\left[Y_{i(t-s)}\Delta\varepsilon_{it}\right] = 0.$$

- **Moment in differences**, featuring a product between the error term and the higher lags (one period and beyond) of the first difference of the dependent variable (Blundell and Bond, 1998). For $s \geq 2$, this is:

$$\mathbb{E}\left[\Delta Y_{i(t-s)}\varepsilon_{it}\right] = 0.$$

Estimation of production functions (1/3)

- Consider a Cobb-Douglas production function model like the one from Lecture 7, but adapted to panel data:

$$y_{it} = \alpha_i + \beta_K k_{it} + \beta_L \ell_{it} + \omega_{it} + \varepsilon_{it}$$

where $y_{it} \equiv \log Y_{it}$, $k_{it} \equiv \log K_{it}$, $\ell_{it} \equiv \log L_{it}$ and:

$$\log A_{it} = \alpha_i + \omega_{it} + \varepsilon_{it}$$

the log of "total" productivity A_{it} is split between a constant factor α_i and two time-varying factors ω_{it} and ε_{it}.

- Why this distinction? While part of the error can be treated as exogenous:

$$\mathbb{E}\left[\varepsilon_{it} \middle| k_{it}, \ell_{it}\right] = 0$$

(think about lucky events), the other part may not:

$$\mathbb{E}\left[\alpha_i, \omega_{it} \middle| k_{it}, \ell_{it}\right] \neq \mathbf{0}$$

as firm adapt their inputs k_{it}, ℓ_{it} to their own circumstances.

Estimation of production functions (2/3)

- Complications arise if the "endogenous" error component ω_{it} is autoregressive, like ε_{it} in dynamic linear models.

$$\omega_{it} = \rho\omega_{i(t-1)} + \xi_{it}$$

- A suitable GMM approach is akin to the Blundell-Bond one for dynamic linear models. Subtract $\rho y_{i(t-1)}$ from both sides:

$$\begin{aligned} y_{it} - \rho y_{i(t-1)} = \alpha_i (1-\rho) + \beta_K \left(k_{it} - \rho k_{i(t-1)}\right) \\ + \beta_L \left(\ell_{it} - \rho\ell_{i(t-1)}\right) + \upsilon_{it} \end{aligned}$$

where $\upsilon_{it} \equiv \xi_{it} + \varepsilon_{it} - \rho\varepsilon_{i(t-1)}$. Note that ω_{it} vanishes.

- Estimation is based on the following moments, for $s \geq 2$.

$$\mathbb{E}\left[\begin{pmatrix} \Delta k_{i(t-s)} \\ \Delta\ell_{i(t-s)} \end{pmatrix} \left(\alpha_i (1-\rho) + \xi_{it} + \varepsilon_{it} - \rho\varepsilon_{i(t-1)}\right)\right] = 0$$

Yet, estimates from actual applications are often imprecise.

Estimation of production functions (3/3)

- Modern **control function** methods – see Ackerberg, Caves and Frazer (2015) – are based on moment conditions like:

$$\mathbb{E}\left[\begin{pmatrix} k_{i(t-s)} \\ \ell_{i(t-s)} \end{pmatrix}\left(y_{it} - \beta_K k_{it} - \beta_L \ell_{it} - g\left(\widehat{\varphi}_{i(t-1)} - \beta_K k_{i(t-1)} - \beta_L \ell_{i(t-1)}\right)\right)\right] = 0$$

  for $s = 2, \dots, t$; where $\widehat{\varphi}_{i(t-1)} = \widehat{\varphi}\left(k_{i(t-s)}, k_{i(t-s)}, m_{i(t-s)}\right)$ is a non-parametric prediction function of $\alpha_i + \omega_{i(t-1)}$.

- Here m_{it} is a "shifter" variable (for example, variable input materials) and $g(\cdot)$ is another non-parametric function.

- The **implicit** "timing" assumption is: $\mathbb{E}\left[\varepsilon_{it}, \xi_{it} \middle| k_{it}\right] = \mathbf{0}$. It is motivated if firms cannot see ξ_{it} before "adjusting" k_{it}.

Estimation of rational expectation models (1/3)

- The motivation of GMM in econometrics was **structural** in origin, because the zero moment conditions may be derived **directly** from economic theory.

- The **rational expectations models** from macroeconomics provide a good example of this. The GMM treatment of the **permanent income** model by Hall (1978) is reviewed next.

- Consider representative consumers who maximize their own **lifetime intertemporal utility function**:

$$\mathcal{U}_t\left(C_t, C_{t+1}, \ldots, C_T\right) = \mathbb{E}_t\left[\sum_{\tau=0}^{T-t}\left(\frac{1}{1+\delta}\right)^{\tau} U\left(C_{t+\tau}\right)\bigg| \mathbb{I}_t\right]$$

  subject to an **intertemporal budget constraint**.

$$\sum_{\tau=0}^{T-t}\left(\frac{1}{1+\delta}\right)^{\tau}\left(C_{t+\tau} - W_{t+\tau}\right) = A_t$$

Estimation of rational expectation models (2/3)

- Here C_s is **consumption**; $U(C_s)$ is the per-period **utility**; W_s are the **earnings**; A_s are the **assets**; r is the **interest rate**; δ is the **discount factor**; and $\mathbb{I}_t$ is the **information set**. Most variables are referred to a generic time period s.

- Individuals form **expectations** $\mathbb{E}_t\left[W_{t+\tau}\middle|\,\mathbb{I}_t\right]$ regarding their future earnings. The **Euler equation** is thus as follows.

$$\mathbb{E}_t\left[U'(C_{t+1})\middle|\,\mathbb{I}_t\right] = \frac{1+\delta}{1+r}U'(C_t)$$

- With a "Constant Relative Risk Aversion" (CRRA) utility:

$$U(C_t) = \frac{1}{1-\alpha}C_t^{1-\alpha}$$

  the Euler equation specializes as:

$$\mathbb{E}_t\left[\beta(1+r)R_{t+1}^{\lambda} - 1\middle|\,\mathbb{I}_t\right] = 0$$

  where $\beta \equiv (1+\delta)^{-1}$, $\lambda \equiv -\alpha$ and $R_{t+1} = C_{t+1}/C_t$.

Estimation of rational expectation models (3/3)

- A researcher may want to estimate parameters β and λ.

- For this sake, GMM may be applied directly to the moment conditions motivated by this theory of optimal choice.

$$\mathbb{E}_t\left[\begin{pmatrix}1\\R_t\end{pmatrix}\left(\beta\,(1+r)\,R_{t+1}^{\lambda}-1\right)\right]=0$$

- Suppose that the researchers observes **several** variables $\boldsymbol{z}_i$ that are known to affect the information set $\mathbb{I}_t$. In this case, the moment conditions become:

$$\mathbb{E}_t\left[\boldsymbol{z}_t\left(\beta\,(1+r)\,R_{t+1}^{\lambda}-1\right)\right]=0$$

  leading to overidentification whenever $|\boldsymbol{z}_i|>2$.

- Observe that this model requires no statistical assumption beyond those that are derived from economic theory.