GMM, Indirect Inference and Bootstrap

General Method of Moments (GMM)

Willi Mutschler

TU Dortmund

Winter 2015/2016

TO IMPROVE

- Example or exercise for weighting matrix, in particular what is optimal what if not.
- Example without gmm package and optimal weighting matrix
- Clarify E_θ:
- Clairfy identification and asymptotic identification: DGP, der zu θ gehrt. Identifikation: Zu θ knnen viele DGPs gehren, aber zu jedem DGP gibt es genau ein θ
- Interpretation of asset pricing models
- Do you really want distinction between elementary zero functions or simply moment conditions? NOTATION!!!
- Slide with motivation and ideas
- List with assumptions needed for consistency and normality, see Pesharan's book!

- Introduced by Lars Peter Hansen (1982, Econometrica)
- Can be applied to time series, cross sectional and panel data
- Many applications: macroeconomics, finance, microeconomics, agricultural economics, environmental economics, labor economics
- Belongs to the class of limited information methods
- Builds upon ideas from method of moments, minimum Chi-Square methods and instrumental variables, however, overcomes (some of) their shortcomings

Hansen, L. (1982), Large Sample Properties of Generalized Method of Moments Estimators, Econometrica 50, 1029-1054:

> *In this paper we study the large sample properties of a class of generalized method of moments (GMM) estimators which subsumes many standard econometric estimators. To motivate this class, consider an econometric model whose parameter vector we wish to estimate. The model implies a family of* **orthogonality conditions that embed any economic theoretical restrictions** *that we wish to impose or test.*

John Cochrane (2005), Asset Pricing, p. 196:

> *Most of the effort involved with GMM is simply mapping a given problem into the very general notation.*

- Describe the model by elementary zero functions

$$E_\theta\left(f_t\left(\theta, y_t\right)\right) = 0$$

where everything can be vector-valued

- Parameter vector θ of length K
- Observation vectors y_t
- Identification condition

$$E_{\theta_0}\left(f_t\left(\theta, y_t\right)\right) \neq 0 \quad \text{for all } \theta \neq \theta_0$$

Example (Linear regression model)

Consider the standard model

$$\begin{aligned} y &= X\beta + u \\ u &\sim N(0, \sigma^2 I), \text{ independent of } X \end{aligned}$$

Parameter vector $\theta =?$
Observations $y_t =?$
Elementary zero functions $f_t(\theta, y_t) =?$

Example (Lognormal distribution)

Suppose there is a random sample $X_1, \dots, X_n$ from

$$X \sim LN(\mu, \sigma^2)$$

Parameter vector $\theta =?$
Observations $y_t =?$
Elementary zero functions $f_t(\theta, y_t) =?$

Example (Asset pricing)

The basic asset pricing formula is

$$p_t = E\left(m_{t+1}x_{t+1}|\Omega_t\right)$$

with asset price p, stochastic discount factor m, payoff x, and information set Ω_t. Assume

$$m_{t+1} = \beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}$$

with c consumption, β the time preference parameter and γ the coefficient of relative risk aversion in the utility function $u(c) = c^{1-\gamma}/(1-\gamma)$.

Parameter vector $\theta =?$
Observations $y_t =?$
Elementary zero functions $f_t(\theta, y_t) =?$

- Stack all elementary zero functions

$$f\left(\theta, y\right) = \begin{bmatrix} f_1\left(\theta, y_1\right) \\ \vdots \\ f_n\left(\theta, y_n\right) \end{bmatrix}$$

- Covariance matrix

$$E\left(f\left(\theta, y\right)\, f\left(\theta, y\right)'\right) = \Omega$$

- Dimension of Ω depends on dimension of $f_t(\theta, y_t)$

Example (Linear regression model)

The covariance matrix Ω is

$$\begin{aligned} E\left(f(\theta, y)\; f(\theta, y)'\right) &= E\left(u\, u'\right) \\ &= \sigma^2 I \end{aligned}$$

If there are autocorrelation and heteroskedasticity

$$E\left(u\, u'\right) = \Omega$$

Example (Lognormal distribution)

The covariance matrix Ω is

$$
\begin{aligned}
E\left(f(\theta,y)\; f\left(\theta,y\right)'\right) &= E\begin{pmatrix}
f_{11}^2 & f_{11}f_{12} & \dots & f_{11}f_{n1} & f_{11}f_{n2}\\
f_{12}f_{11} & f_{12}^2 & \dots & f_{12}f_{n1} & f_{12}f_{n2}\\
\vdots & \vdots & \ddots & \vdots & \vdots\\
f_{n1}f_{11} & f_{n1}f_{12} & \dots & f_{n1}^2 & f_{n1}f_{n2}\\
f_{n2}f_{11} & f_{n2}f_{12} & \dots & f_{n2}f_{n1} & f_{n2}^2
\end{pmatrix}\\
&= \ ?
\end{aligned}
$$

Example (Asset pricing)

The covariance matrix Ω is

$$\begin{aligned} E\left(f(\theta,y)\ f(\theta,y)'\right) &= E\begin{pmatrix} f_{11}^2 & \dots & f_{11}f_{n1} \\ \vdots & \ddots & \vdots \\ f_{n1}f_{11} & \dots & f_{n1}^2 \end{pmatrix} \\ &= ? \end{aligned}$$

- To estimate θ, we need K estimating equations
- In general, they are weighted averages of the f_t
- In most cases, the estimating equations are based on $L \geq K$ instrumental variables W
- If $L > K$, we need to form linear combinations of the instruments
- Let W be the $n \times L$ matrix of instruments and J be an $L \times K$ matrix of full rank (we will discuss how to set J optimally later)
- Define the $n \times K$ matrix $Z = WJ$

- Theoretical **moment conditions** (orthogonality conditions)

$$E\left(Z_t' f_t\left(\theta, y_t\right)\right) = 0$$

- The **estimating equations** are the empirical counterpart

$$\frac{1}{n} Z' f(\theta, y) = 0$$

- Solving this system yields the GMM estimator $\hat{\theta}$

Example (Linear regression model)

The K moment conditions for the linear regression model are

$$E\left(Z_t' f_t\left(\theta, y_t\right)\right) = E\left(X_t'\left(y_t - X_t'\beta\right)\right) = 0$$

and the estimating equations are

$$\frac{1}{n} X'\left(y - X\beta\right) = 0.$$

Example (Lognormal distribution)

The two moment conditions for the lognormal distribution are

$$\begin{aligned}
E\left(Z_t' f_t(\theta, y_t)\right) &= E\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} f_{t1}(\theta, y_t) \\ f_{t2}(\theta, y_t) \end{bmatrix}\right) \\
&= E\left(\begin{bmatrix} X_t - \exp\left(\mu + \frac{1}{2}\sigma^2\right) \\ X_t^2 - \exp\left(2\mu + 2\sigma^2\right) \end{bmatrix}\right) \\
&= \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}$$

Example (contd)

... and the estimating equations are

$$
\begin{aligned}
\frac{1}{n} Z' f(\theta, y) &= \frac{1}{n} \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ \vdots \\ f_{n1} \\ f_{n2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} \left(X_t - \exp\left(\mu + \frac{1}{2}\sigma^2 \right) \right) \\ \frac{1}{n} \sum_{t=1}^{n} \left(X_t^2 - \exp\left(2\mu + 2\sigma^2 \right) \right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\end{aligned}
$$

Consistency

- Assume that a law of large numbers applies to $\frac{1}{n}Z'f(\theta, y)$
- Define the limiting estimation functions

$$\alpha(\theta) = plim \frac{1}{n} Z' f(\theta, y)$$

  and the limiting estimation equations $\alpha(\theta) = 0$
- The GMM estimator $\hat{\theta}$ is consistent if the asymptotic identification condition holds, $\alpha(\theta) \neq \alpha(\theta_0)$ for all $\theta \neq \theta_0$ [P]

Asymptotic normality

- Simplified notation: $f_t(\theta) = f_t(\theta, y_t)$, $f(\theta) = f(\theta, y)$
- Additional assumption: $f_t(\theta)$ is continuously differentiable at θ_0
- First order Taylor series expansion of

$$\frac{1}{n} Z' f(\theta) = 0$$

  in $\hat{\theta}$ around θ_0 [P]

- The asymptotic distribution of $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$ is normal with mean 0 and covariance matrix

$$\left(plim\frac{1}{n}Z'F(\theta_0)\right)^{-1}\left(plim\frac{1}{n}Z'\Omega Z\right)\left(plim\frac{1}{n}F(\theta_0)'Z\right)^{-1}$$

- What is the optimal choice of Z in the estimating equations?
- The optimal choice depends on assumptions about the matrices $F(\theta)$ and Ω

Case I

- If $\Omega = \sigma^2 I$ and $E(F_t(\theta_0) f_t(\theta_0)) = 0$ the optimal choice is

$$Z = F(\theta_0)$$

- Problem: Z depends on the unknown θ_0
- Solution: Solve the estimating equations

$$\frac{1}{n} F'(\theta) f(\theta) = 0$$

Case II

- If $\Omega = \sigma^2 I$ and $E(F_t(\theta_0) f_t(\theta_0)) \neq 0$ but $W_t \in \Omega_t$, the optimal choice is
$$Z = P_W F(\theta_0)$$
- Problem: Z depends on the unknown θ_0
- Solution: Solve the estimating equations
$$\frac{1}{n} F'(\theta) P_W f(\theta) = 0$$

Case III

- Suppose, the covariance matrix Ω is unknown
- Since $Z = WJ$, the covariance matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$ is

$$\left(plim\frac{1}{n}J'W'F_0\right)^{-1}\left(plim\frac{1}{n}J'W'\Omega WJ\right)\left(plim\frac{1}{n}{F_0}'WJ\right)^{-1}$$

- For the optimal $J = \left(W'\Omega W\right)^{-1}W'F_0$ this becomes

$$\left(plim\frac{1}{n}F_0'W\left(W'\Omega W\right)^{-1}W'F_0\right)^{-1}$$

- Although Ω **cannot** be estimated consistently, the term $\frac{1}{n}W'\Omega W$ **can** be estimated consistently (we will do that later)
- If $\hat{\Sigma}$ is an estimator of $\frac{1}{n}W'\Omega W$,the optimal estimating equations are

$$\frac{1}{n}J'W'f(\theta) = \frac{1}{n}F(\theta)'W\hat{\Sigma}^{-1}W'f(\theta) = 0$$

and the estimated covariance matrix of $\hat{\theta}$ is

$$\widehat{Cov}(\hat{\theta}) = n\left(\hat{F}'W\hat{\Sigma}^{-1}W'\hat{F}\right)^{-1}$$

Attention

- Many textbooks use a different notation (and so does the `gmm` package in R)
- The two approaches are equivalent
- The moment conditions are notated as

$$E\left(g\left(\theta, y_t\right)\right) = E\left(W_t' f_t\left(\theta, y\right)\right) = 0$$

- The number of moment conditions L can be larger than the number of parameters K

- The L estimating equations cannot be solved exactly

$$\bar{g}_n(\theta, y) = \frac{1}{n}\sum_{t=1}^{n} g(\theta, y_t) = 0$$

- The GMM estimator is defined by

$$\hat{\theta} = \arg\min \bar{g}_n(\theta, y)' A_n \, \bar{g}_n(\theta, y)$$

where A_n is a sequence of $L \times L$ weighting matrices (which can be chosen by the user) with limit A

- The GMM estimator based on $\bar{g}_n$ is consistent, $\hat{\theta} \xrightarrow{p} \theta$
- Asymptotic normality: Define the $L \times K$ matrix

$$G(\theta) = \frac{\partial \bar{g}_n(\theta, y_t)}{\partial \theta'} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial g(x_t, \theta)}{\partial \theta'}$$

- Assume that $\sqrt{n}\bar{g}_n(\theta, y) \xrightarrow{d} N(0, V)$, then [P]

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \left(G'AG\right)^{-1} G'AVAG \left(G'A'G\right)^{-1}\right)$$

- Asymptotically optimal weighting matrix A [P]

- The two GMM approaches (based on f_t and g) are equivalent
- The first order condition of $\bar{g}(\theta)' A \bar{g}(\theta)$ is

$$\underset{K\times L}{G'} \underset{L\times L}{A} \underset{L\times 1}{g} = \underset{K\times 1}{0}$$

which is the same as

$$\underset{K\times L}{J'} \underset{L\times n}{W'} \underset{n\times 1}{f} = \underset{K\times 1}{0}$$

- List of equivalences [P]

- The covariance matrix of the elementary zero functions

$$E\left(f\left(\theta, y\right) f\left(\theta, y\right)'\right) = \Omega$$

  is often unknown
- There may be heteroskedasticity and autocorrelation in Ω
- Although Ω cannot be estimated consistently, the term $\frac{1}{n}W'\Omega W$ can be estimated consistently

- Write
$$\Sigma = plim_{n\to\infty}\frac{1}{n}W'\Omega W$$
- Assume that a suitable law of large numbers holds,
$$\Sigma = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}E\left(f_t f_s W_t' W_s\right)$$

where $f_t = f_t\left(\theta, y_t\right)$

- Define the autocovariance matrices

$$\Gamma(j) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} E\left(f_t f_{t-j} W_t' W_{t-j}\right) & \text{for } j \geq 0 \\ \frac{1}{n}\sum_{t=-j+1}^{n} E\left(f_{t+j} f_t W_{t+j}' W_t\right) & \text{for } j < 0 \end{cases}$$

- Then

$$\Sigma = \lim_{n\to\infty} \sum_{j=-n+1}^{n-1} \Gamma(j) = \lim_{n\to\infty} \left(\Gamma(0) + \sum_{j=1}^{n-1} \left(\Gamma(j) + \Gamma'(j)\right) \right)$$

- The autocovariance matrix $\Gamma(j)$, $j \geq 0$, can be estimated by

$$\hat{\Gamma}(j) = \frac{1}{n} \sum_{t=j+1}^{n} \hat{f}_t \hat{f}_{t-j} W_t' W_{t-j}$$

- Newey-West estimator of Σ

$$\hat{\Sigma} = \hat{\Gamma}(0) + \sum_{j=1}^{p} \underbrace{\left(1 - \frac{j}{p+1}\right)}_{\text{Adjustment term}} \left(\hat{\Gamma}(j) + \hat{\Gamma}'(j)\right)$$

- Adjustment term ensures positive (semi)definiteness in finite samples
- Estimator tends to underestimate the autocovariance matrices, unless p increase at an appropriate rate of $n^{1/3}$, see Newey and West (1994) for a procedure to select p automatically

- The GMM estimators minimize the criterion function

$$\frac{1}{n} f'(\theta) W \hat{\Sigma}^{-1} W' f(\theta)$$

- Asymptotically, the minimized value (Hansen's J statistics, Hansen's overidentification statistic, Hansen-Sargan statistic) is distributed as χ^2_{L-K} if the overidentifying restrictions hold
- If the null hypothesis is rejected, then something went wrong, e.g. the model is misspecified