# Generalized Method of Moments (GMM)

The setup: Let $\psi_i(\theta)$ be an M-dimensional vector, possibly depending on data specific to each observation $i(i = 1, ..., n)$, such that "theory" specifies:

$$E(\psi_i(\theta)) = 0 \qquad i = 1, ..., n$$

$\theta$ is $k$ dimensional with $m \geq k$.

Notation:

$$\psi(\theta) = \sum_{i=1}^{n} \psi_i(\theta)$$

$$E(\psi(\theta)) = 0$$

$$\bar{\psi}(\theta) = n^{-1}\psi(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi_i(\theta)$$

## GMM Basics

Consider estimating $\theta$ by minimizing:

$$Q(\theta) = \underset{(1\times m)}{\psi'(\theta)} \underset{(m\times m)}{W} \underset{(m\times 1)}{\psi(\theta)} \left(\frac{1}{n^2}\right) = \underset{(1\times m)}{\bar{\psi}'(\theta)} \underset{(m\times m)}{W} \underset{(m\times 1)}{\bar{\psi}(\theta)}$$

where $W$ is pos semidefinite and symmetric. Have a look at the first-order condition:

$$\frac{\partial Q}{\partial \theta} = 2 \underset{(1\times m)}{\bar{\psi}'(\theta)} \underset{(m\times m)}{W} \underset{(m\times k)}{\frac{\partial \bar{\psi}}{\partial \theta}} = 0$$

A natural identification condition gives e.g.

$$\lim_{n\to\infty} \left.\frac{\partial Q}{\partial \theta}\right|_{\theta_0} = 0; \lim_{n\to\infty} \left.\frac{\partial Q}{\partial \theta}\right|_{\theta_0} \neq 0 \text{ if } \theta \neq \theta_0$$

## GMM Basics (2)

The second derivative is:

$$
\begin{aligned}
\frac{\partial^2 Q}{\partial\theta\partial\theta'} &= 2\underbrace{\left(\frac{\partial\bar{\psi}}{\partial\theta}\right)'}_{k\times m}\underset{m\times m}{W}\underbrace{\frac{\partial\bar{\psi}}{\partial\theta}}_{m\times k} \quad +2\bar{\psi}'(\theta)W\frac{\partial}{\partial\theta'}\left(\frac{\partial\bar{\psi}}{\partial\theta}\right) \\
&= \qquad \text{converges} \qquad\qquad \text{a complicated object} \\
& \qquad\qquad\qquad\qquad\qquad \textbf{plim} \text{ is 0 at } \theta_0 \text{ since } \bar{\psi}'(\theta) \text{ has plim 0}
\end{aligned}
$$

Under broad conditions for M-estimation, cf. Davidson and MacKinnon (1993) , Thm 17.2 p.593, “White-style” argument gives:

$$
\sqrt{n}(\widehat{\theta}-\theta_0) \sim N(0, H^{-1}VH^{-1}), \quad
\begin{aligned}
H &= \textbf{plim}\,\frac{\partial^2 Q}{\partial\theta\partial\theta'} \\
V &= AV(\sqrt{n}\frac{\partial Q}{\partial\theta})
\end{aligned}
$$

Write

$$
\begin{aligned}
H &= \mathsf{plim}\,\frac{\partial^2 Q}{\partial\theta\partial\theta} = 2D'WD; \quad D = (\partial\psi/\partial\theta)\frac{1}{n} \\
V &= \mathsf{plim}\,\sqrt{n}\,\underbrace{\frac{\partial Q'}{\partial\theta}\ \frac{\partial Q}{\partial\theta}}_{(k\times 1)(1\times k)}\,\sqrt{n} = \mathsf{plim}\,2D'W\ \{\psi(\theta)\psi'(\theta)\}\ WD\,2
\end{aligned}
$$

So

$$
H^{-1}VH^{-1} = (2D'WD)^{-1}\ 2D'W\{\psi(\theta)\psi'(\theta)\}WD2\ (2D'WD)^{-1}
$$

Suppose

$$
\begin{aligned}
W &= \mathsf{plim}[\psi(\theta)\psi'(\theta)]^{-1} \\
H^{-1}VH^{-1} &= (2D'WD)^{-1}\ 4\,D'W\{W^{-1}\}WD\ (2D'WD)^{-1} \\
&= (2D'WD)^{-1}\,4D'WD\ (2D'WD)^{-1} \\
&= (D'WD)^{-1}
\end{aligned}
$$

This is the efficient choice of $W$ (“GMM efficiency bound”).

GMM when $\psi$ is the score.

## Generalized Instrumental Variables Estimation (GIVE) as a GMM Estimator

What makes it "generalized" is that there are more instruments than coefficients.

Simplest case:

$$y = X\beta + u, \quad Z(n \times m) \text{ instruments}$$
$$u \; i.i.d.$$

"Moment" or "orthogonality" condition:

$$\psi(\theta) = Z'u = \underset{(m\times n)}{Z'} \underset{(n\times 1)}{(y - X\beta)} = 0$$

Let $W$ be an arbitrary weight matrix

$$n^2 Q(\theta) = \psi'(\theta) W \psi(\theta) = (y - X\beta)' Z W Z' (y - X\beta)$$

## GIVE (2)

$$n^2 Q(\theta) = \psi'(\theta) W \psi(\theta) = (y - X\beta)' Z W Z' (y - X\beta)$$

$$\begin{aligned} n^2 \frac{\partial Q}{\partial \theta} &= 2 \underset{(1\times m)}{\psi'(\theta)} \underset{(m\times m)}{W} \underset{(m\times k)}{\frac{\partial \psi}{\partial \theta}} = 0 \qquad \text{is FOC} \\ &= -2(y - X\beta)' Z \; W \; Z' X = 0 \end{aligned}$$

Solving FOC

$$y' Z W Z' X = \beta' X' \, Z W \, Z' X$$

Transpose

$$X' Z W Z' y = X' \underset{k\times k}{Z W Z'} X \beta$$

$$\widehat{\beta} = (X' Z W Z' X)^{-1} (X' Z W Z') y$$

# GIVE (3)

So

$$\widehat{\beta} = (X'ZWZ'X)^{-1}(X'ZWZ')y$$

for arbitrary $W$. By preceding args, efficient choice for $W$ is $[E(\psi\psi')]^{-1}$

$$\begin{aligned} E(\psi\psi') &= Z'uu'Z \\ &= \sigma^2 Z'Z \end{aligned}$$

So $W = (Z'Z)^{-1}$ is the "efficient" choice, ($\sigma^2$ factor doesn't matter).

So we have

$$\widehat{\beta} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z')y$$

Recall in eg 2SLS we regress $X$ on $Z$'s and use "fitted" $X$'s (i.e. $Z\widehat{\gamma}$) as instruments: call these $\widetilde{Z}$.

## GIVE (4): Conclusion

This operation here gives

$$\begin{aligned} \widetilde{Z} &= Z\widehat{\gamma} \\ &= Z(Z'Z)^{-1}Z'X \text{ or} \\ & \widetilde{Z'} = X'Z(Z'Z)^{-1}Z' \end{aligned}$$

Use this expression in $\widehat{\beta}$ above

$$\widehat{\beta} = (\tilde{Z}'X)^{-1}(\tilde{Z}'y)$$

or "optimal" give is "naive" IV. This is very neat because, for the linear case (with homoscedastic and serially independent errors), it shows how to select instruments to reach the GMM efficiency bound. The role of $X$ in the above is that in the more general formulation $y = f(X, \beta) + u$, $X = \frac{\partial f(X,\beta)}{\partial \beta}$ in the linear case. The argument above run in parallel for the nonlinear case presents some problems as it calls for regressing $\partial f / \partial \beta$ on $Z$.

## How a literature about W develops

$W$ is the asymptotic covariance of (standardized) $\psi(\cdot)$. (i.e. $n^{-1/2}\psi$).

That is

$$\psi(\theta) = \sum_{i=1}^{n} \psi_i(\theta)$$

and so the (standardized) covariance is

$$E\frac{1}{n}\left\{\left[\sum_{i=1}^{n} \psi_i(\theta)\right]\left[\sum_{i=1}^{n} \psi_i(\theta)\right]'\right\}$$

Now $\{\cdot\}$ is:

$$\begin{array}{l} \psi_1(\theta) + \psi_2(\theta) + .... \ \psi_n(\theta) \\ \bullet\psi_1'(\theta) + \psi_2'(\theta) + .... \ \psi_n'(\theta) \end{array}$$

$$\psi_1\psi_1' + \psi_1\psi_2' + \dots.\ \psi_1\psi_n'$$
$$\psi_2\psi_1' + \psi_2\psi_2' + \dots\dots\dots.$$

When $\psi_i$ and $\psi_j$ are independent for $i \neq j$. Then $E\psi_i\psi_j' = 0$ and $\{\cdot\}$ becomes

$$E\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_i(\theta)\psi_i'(\theta)\right\}$$

which is pretty straightforward, but this is not true in general and in particular, GMM with conditional moments typically doesn't display the independence structure (cf. Davidson-MacKinnon chap 17.4. 602-614). $E(u_t|X) = 0 \Rightarrow u$ uncorrelated with any function of $X$; contrast with $u_t$ uncorrelated with $X$ and $u_t$ independent of $X$.

## An application: from macro with a micro flavour

Hansen and Singleton (1982); covered in different notation in Hamilton Chap. 14. Estimation of the Euler equation in a consumption based asset-pricing model.

Infinitely-lived agent maxes expected utility given by:

$$E_o\left[\sum_{t=0}^{\infty}\delta^t u(c_t)\right] \tag{1}$$

$E_t(\cdot) =$ expectation given information available at $t$. In an period t: the agent can either buy a consumption good or an asset, a unit of which pays $R_{t+1}$ in the next period. (Divisible amounts). Let

$$\begin{aligned} P_t &= \text{Price of asset (in consumption terms)} \\ Q_t &= \text{Quantity of asset} \\ W_t &= \text{Wage} \end{aligned}$$

The budget constraint is:

$$C_t + P_t Q_t \leq Q_{t-1} + W_t \tag{2}$$

Max (1) subject to (2):

$$E_o \left\{ \sum_{t=0}^{\infty} \{ \delta^t u(C_t) + \lambda_t (C_t + P_t Q_t - R_t Q_{t-1} - W_t) \right\}$$

$$\begin{aligned} \frac{\partial}{\partial C_t} &= E_0 \{ \delta^t u'(C_t) + \lambda_t \} = 0 \text{ or } \lambda_t = \delta^t u'(C_t) \\ \frac{\partial}{\partial Q_t} &= E_0 \{ \lambda_t P_t - \lambda_{t+1} R_{t+1} \} = 0 \end{aligned}$$

# GMM Estimation of an Euler Equation

$$
\begin{aligned}
\lambda_t &= \delta^t u'(C_t) \\
\lambda_t P_t &- \lambda_{t+1} R_{t+1} = 0
\end{aligned}
$$

$$E_0\{\delta^t u'(C_t)P_t - \delta^{t+1}u'(C_{t+1})R_{t+1}\} = 0$$

In particular above holds at $t$, when the decision has to be mine

$$u'(C_t)P_t = \delta E_t(u'(C_{t+1})R_{t+1})$$

So

$$E_t\left[\delta \frac{R_{t+1}}{P_t}\frac{u'(C_{t+1})}{u'(C_t)} - 1\right] = 0$$

Let $u(C_t) = C_t^\gamma/\gamma \quad \gamma < 1$.

Then

$$E_t \left[ \delta \frac{R_{t+1}}{P_t} \left[ \frac{C_{t+1}}{C_t} \right]^{\gamma - 1} - 1 \right] = 0$$

This would serve as one moment condition (remove the $E_t$ via iterated expectation)

Let $Z_t$ be a variable in the information set at $t$. Then

$$E \left( \left\{ \delta \frac{R_{t+1}}{P_t} \left[ \frac{C_{t+1}}{C_t} \right]^{\gamma - 1} - 1 \right\} Z_t \right) = 0$$

is another condition.

## Inference in GMM

Texts: Davidson and MacKinnon, section 17.6, pp. 614-620; (also Hamilton, *Time Series Analysis*, Chapter 14, especially p. 415 and pp. 427-430.)

Hansen's test of overidentifying restrictions: Since $m > k$, $m - k$ is the number of overidentifying restrictions, i.e. the number of moment conditions in excess of the strict minimum necessary to estimate the $k$ parameters.

With the efficient choice of $W$, $Q(\theta)$ from p.1 converges to 0 and $nQ(\theta)$ is $\chi^2$. (Compare eqn. 17.68 p. 615 DM). Heuristic: what degrees of freedom characterizes nQ($\theta$)? Well if $m = k$, $nQ(\theta)$ is identically zero (since we pick the $k$ parameters to satisfy the $m = k$ moment conditions) so we might expect $nQ(\theta)$ is $\chi^2(m - k)$. This is indeed correct, cf. proof in DM, pp. 615-616.

Hansen's test of overidentifying restrictions compares $nQ(\theta)$ with $\chi^2(m - k)$, using an efficient estimate of $W$.

## Testing hypotheses on $\theta$

Now suppose we want to test whether one component of $\theta$ equals a particular value, i.e. $H_0$: $\theta_j = \theta_j^0$. Then estimating with GMM with this imposed gives an $nQ(\theta)$ with one more d.f. than without the restriction imposed. The difference between the two objective functions is 'thus' $\chi^2(1)$. And so on for hypotheses of greater dimension. This resembles the LR test (and is the statistic $D$ in Newey-West.) (Note that the *same* choice of $W$ must be used in both optimization problems; the natural choice is $W$ estimated without imposing the $\theta_j = \theta_j^0$ hypothesis).

Additionally, Wald and score-style tests can be constructed. For a Wald test of $r(\theta) = 0$, let $R(\theta) = \partial r / \partial \theta$. Then the normality of $\theta$ means that

$$S_w = nr(\hat{\theta})'[\hat{R}(D'WD)^{-1}\hat{R}']^{-1}r(\hat{\theta})$$

is $\chi^2(\text{dim } r)$.

If one estimates under the null hypothesis, calling the resulting estimate $\tilde{\theta}$, then the quasi-score in the full model evaluated at $\tilde{\theta}$ is normally distributed and so can be made the basis of an LM-style test. The quasi-score is

$$g(\tilde{\theta}) = n^{-1/2} D(\tilde{\theta})' W \psi(\tilde{\theta})$$

and the test statistic is

$$S_{LM} = g(\tilde{\theta})' [D(\tilde{\theta})' W D(\tilde{\theta})]^{-1} g(\tilde{\theta})$$