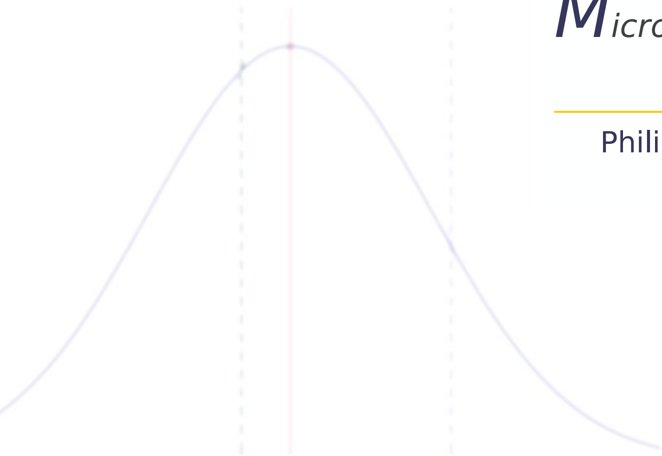
Microeconometrics

Philipp Eisenhauer



Material available on

Visit us!

UNIVERSITÄT BONN

Generalized method of moments

Philipp Eisenhauer

I heavily draw on the material presented in:

- Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology.

General idea

- The generalized method of moments (GMM) is a general estimation principle, where the estimators are derived from so-called moment conditions. It provides a unifying framework for the comparison of alternative estimators.

Structure

- Setup
- Identification
- Asymptotic distribution
- Testing

Setup

Notation

β	$p \times 1$	parameter vector
w_i	$i = 1, \ldots, n$	data points
$g_i(w_i, \beta)$	$m \times 1$	moment

- The GMM estimator is based on a model where, for the true parameter value β_0 the moment conditions $E[g_i(\beta_0)] = 0$ are satisfied.
- The estimator is formed by choosing β so that the sample average of $g_i(\beta)$ is close to its zero population value.

The estimator is formed by choosing β so that the sample average of $g_i(\beta)$ is close to its zero population value. Let

$$\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)$$

- theoretical moments
- empirical moments

Let $\hat{A}$ denote a $m \times m$ positive semi-definite matrix, then the GMM estimator is given by

$$\hat{\beta} = \arg\min_{\beta} \hat{g}(\beta)' \hat{A} \hat{g}(\beta)$$

The GMM estimator chooses $\hat{\beta}$ so the sample average $\hat{g}(\beta)$ is close to zero.

Instrumental variables

Let's work through an example on the blackboard.

Unifying framework

Many other popular estimation strategies can be analyzed in a GMM setup.

Ordinary least squares	$E[x_i(y_i - x_i\beta_0)] = 0$
Instrumental variables	$E[z_i(y_i - x_i\beta_0)] = 0$
Maximum likelihood	$E[\partial \ln f(x_i, \beta_0)/\partial\beta] = 0$

If moments cannot be evaluated analytically then we have an application of the method of simulated moments.

Distance and weighing matrix

Let's look at the role of the weighing matrix for a two dimensional example.

- identity matrix

$$Q(\beta) = \begin{pmatrix} g_1 & g_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = g_1^2 + g_2^2$$

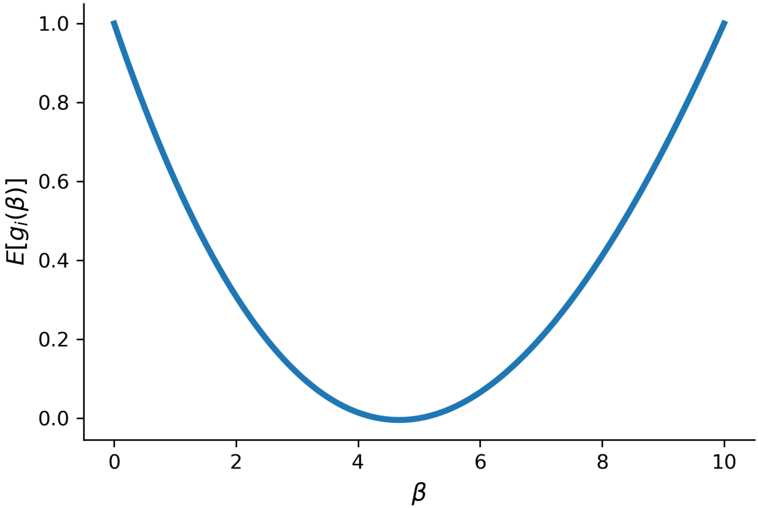
- alternative

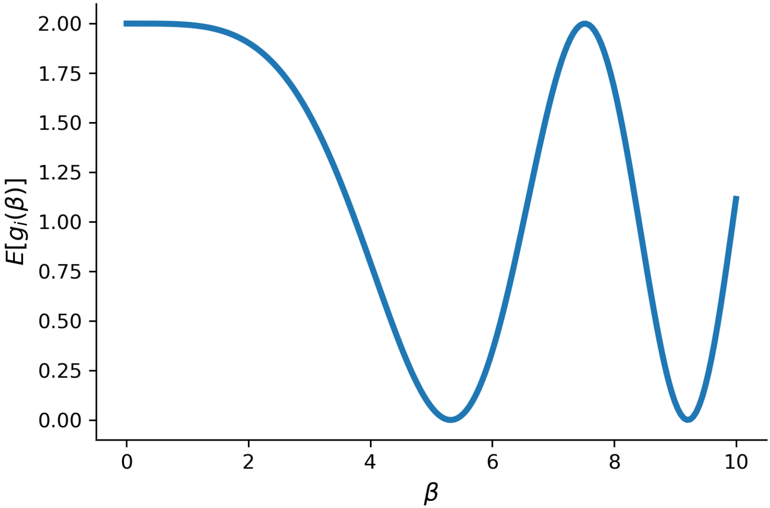
$$Q(\beta) = \begin{pmatrix} g_1 & g_2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 2\dot{g}_1^2 + g_2^2$$

Our alternative attaches more weight to the first coordinate in the distance.

Identification

The parameters β_0 are identified if β_0 is the only solution to $E[g_i(\beta)] = 0$.


1.0
0.8
0.6
0.4
0.2
0.0
$E[g_i(\beta)]$
0
2
4
6
8
10
β



2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
$E[g_i(\beta)]$
0
2
4
6
8
10
β

- Necessary condition for identification is that $m \geq p$. When $m \leq p$, i.e. there are fewer equations to solve than parameters, there will typically be multiple solutions to the moment conditions.

- Let $G = E[\partial g_i(\beta_0)/\partial\beta]$. Rank condition is $rank(G) = p$. Necessary and sufficient for identification when $g_i(\beta)$ is linear in β.
- In the general nonlinear case it is difficult to specify conditions for uniqueness of the solution to $E[g_i(\beta)] = 0$.

- m = p, exact identification, $\hat{g}(\hat{\beta}) = 0$ asymptotically
- m > p, overidentification, $\hat{g}(\hat{\beta}) > 0$ asymptotically

In the case of overidentification, the choice of A matters and affects the estimator's asymptotic distribution.

Asymptotic distribution

Asymptotic distribution

Under some regularity conditions, the GMM estimator has the following asymptotic distribution.

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathbb{N}(0, V),$$

where $V = (G'AG)^{-1}G'A\Omega AG(G'AG)^{-1}$ with $G = E[\partial g_i(\beta_0)/\partial\beta]$ and $\Omega = E[g_i(\beta_0)g_i(\beta_0)']$.

$\Rightarrow$ asymptotic variance depends on the choice of the weighing matrix A

The optimal weighing matrix $A = \Omega^{-1}$ the asymptotic variance simplifies to

$$V = (G'\Omega^{-1}G)^{-1}$$

What makes a good moment?

- small Ω, small sample variation of the moment
- large G, moment informative on true value

Figure: Weak identification

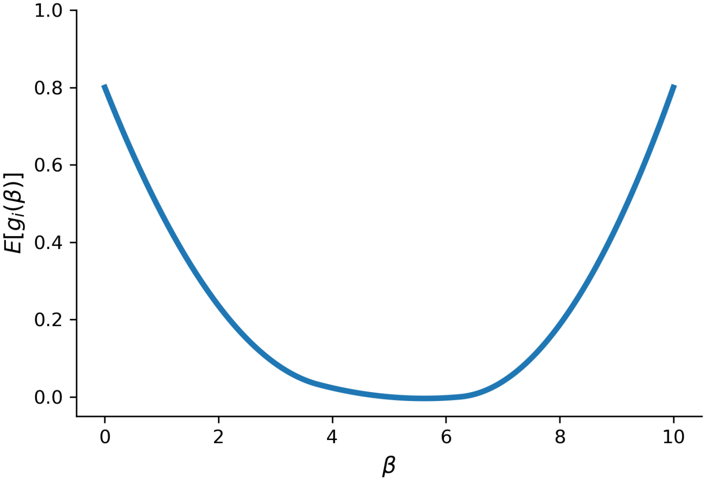
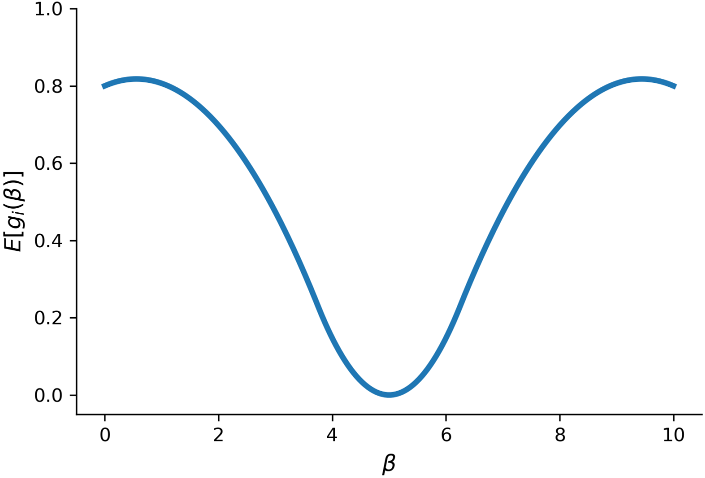


Figure: Sharp identification



Instrumental variables

Let's continue our example on the blackboard.

Testing

An important statistic for GMM is the test of overidentifying restrictions that is given by

$$T = n\,\hat{g}(\hat{\beta})'\,\hat{\Sigma}^{-1}\,\hat{g}(\hat{\beta})$$

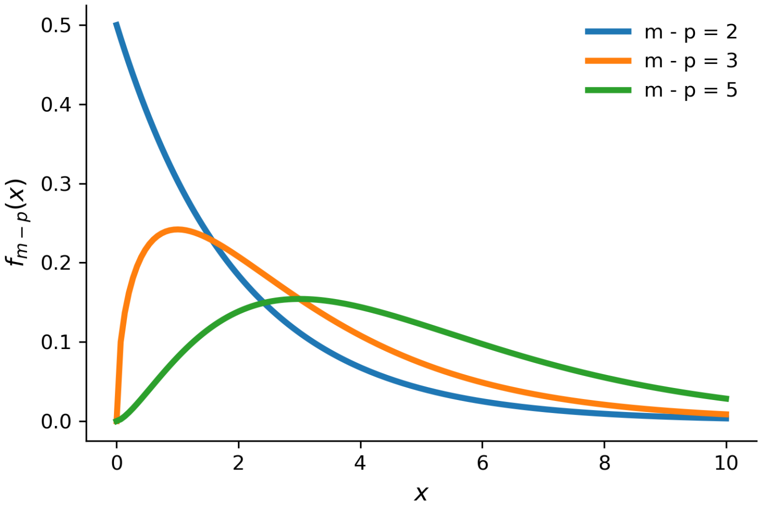
which converges in distribution to

$$T \xrightarrow{d} \chi^2(m-p)$$

under H_0 that the model is correctly specified.

Figure: Density of $\chi^2(2)$

Figure: Density of $\chi^2(m-p)$



Wrapping up

Feasible efficient GMM

The optimal weighing matrix depends on moment evaluations at β_0 which is unknown.

- iterated feasible GMM
- continuously updating GMM

Let’s turn to our Jupyter notebooks ...

Appendix

References

Hahn, J., Todd, P. E., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317.