Lecture 11: Generalized Method of Moments I

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Nonlinear Least Squares Example in R

Today

- Method of Moments
- Generalized Method of Moments
- Properties of the Generalized Method of Moments Estimator
- Optimal Generalized Method of Moments

Upcoming

- No class on Tuesday! Monday schedule.
- Reading for next time
  - Optional: Crawford and Yurukoglu (2012)
- Problem sets
  - Problem Set 2 is posted, due October 17

MLE and NLS Recap

Maximum likelihood estimation

- Requires a strong assumption about the density of our data
- Has many nice properties (lowest-variance consistent estimator)

Nonlinear least squares

- Relies on weaker distributional assumptions
- Requires a single "regression equation" to be estimated

But what if we do not want to make a strong assumption about our data, and we want to jointly estimate multiple nonlinear equations at once?

Method of Moments

(Generalized) Method of Moments

Method of moments and generalized method of moments

- Common in modern empirical economics
  - Most common in industrial organization, macroeconomics, and finance
- More flexible than MLE or NLS
  - A few weak assumptions required! (Other than moment conditions. . .)
- (Almost?) All estimators you have learned so far are GMM estimators
  - OLS, 2SLS, GLS, MLE, NLS, etc.

Overview of MM and GMM

- MM and GMM use population moment conditions to estimate parameters
  - Moments can come from economic models, model fitting criteria, etc.
- MM and GMM estimators "solve" the sample moment conditions that are analogous to these population moment conditions

Moment Conditions

Moment conditions are functions of parameters and data that equal zero in expectation when evaluated at the true parameter values

$$
\begin{aligned}
\text{Mean: } \mu_y &= E[y] & \Rightarrow E[y-\mu_y] &= 0 \\
\text{Var: } \sigma_y^2 &= E[(y-\mu)^2] & \Rightarrow E[(y-\mu)^2-\sigma_y^2] &= 0 \\
\text{Cov: } \sigma_{xy} &= E[(y-\mu_y)(x-\mu_x)] & \Rightarrow E[(y-\mu_y)(x-\mu_x)-\sigma_{xy}] &= 0
\end{aligned}
$$

How do we normally generate moment conditions?

- Economic models
  - First-order conditions always set something equal to zero
- Econometric or statistical model
  - Example: Instruments must be uncorrelated with errors
- Model fit
  - Example: Predicted market shares must equal realized market shares

Method of Moments Estimator

We have K population moment conditions for K parameters

$$E[m(w_i, \theta)] = 0$$

where $m(\cdot)$ is a vector of K functions, θ is a vector of K parameters, and w_i is observable data (outcome, explanatory variables, instruments, etc.)

The method of moments estimator uses the sample analogs of these moment conditions

$$\frac{1}{n}\sum_{i=1}^{n} m(w_i, \theta)$$

The method of moments estimator is the set of parameters that solves

$$\frac{1}{n}\sum_{i=1}^{n} m(w_i, \hat{\theta}) = 0$$

MM Example: Population Mean

Under the assumption that observations are i.i.d, sample means can be used to estimate population means

If y is distributed i.i.d. with mean μ, then a population moment condition is

$$E[y - \mu] = 0$$

Replacing the expectation with the sample analog gives

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}) = 0$$

Solving this one equation for the one parameter yields

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

MM Example: Linear Regression

Ordinary least squares is a special case of a MM estimator

Consider the most general linear regression model

$$y_i = \beta' x_i + \varepsilon_i$$

where β is a vector of K parameters and $E[\varepsilon_i \mid x_i] = 0$. Then $E[x_i \varepsilon_i] = 0$ (from last semester), which gives K population moment conditions

$$E[x_i(y_i - \beta' x_i)] = 0$$

Replacing the expectation with the sample analog gives

$$\frac{1}{n} \sum_{i=1}^{n} x_i(y_i - \hat{\beta}' x_i) = 0$$

Solving these K equations for the K parameters yields

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} x_i y_i \right)$$

MM Example: Maximum Likelihood

Maximum likelihood is another special case of a MM estimator

We want to maximize the log-likelihood function

$$\ln L(\theta \mid y_i, x_i) = \sum_{i=1}^{n} \ln f(y_i \mid x_i, \theta)$$

which gives K first-order conditions (one for each parameter)

$$\frac{1}{n}\frac{\partial \ln L(\hat{\theta} \mid y_i, x_i)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ln f(y_i \mid x_i, \hat{\theta})}{\partial \theta} = 0$$

These first-order conditions are the sample analogs of K population moment conditions

$$E\left[\frac{\partial \ln f(y_i \mid x_i, \theta)}{\partial \theta}\right] = 0$$

so the MLE is also a MM estimator

Limitation of the Method of Moments

In all of these examples, we have had K moment conditions to estimate K parameters. But what if we have more than K moment conditions?

Examples

- Economic model: Demand parameter appears in demand and supply first-order conditions
- Instruments: Multiple instruments for one endogenous variable
- Model fit: Predicted market shares equal realized market shares
- Statistical assumptions: Errors must be conditional mean-zero and symmetric

The generalized method of moments allows for more moment conditions than parameters

Generalized Method of Moments

Generalized Method of Moments

We have L population moment conditions for K parameters (with $L > K$)

$$E[m(w_i, \theta)] = 0$$

where $m(\cdot)$ is a vector of L functions, θ is a vector of K parameters, and w_i is observable data (outcome, explanatory variables, instruments, etc.)

We cannot ensure all L moment conditions are solved with only K parameters. Instead, we seek to get as close as possible to solving all L moment conditions by minimizing

$$Q_N(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} m(w_i, \theta)\right]' W_N \left[\frac{1}{n}\sum_{i=1}^{n} m(w_i, \theta)\right]$$
$$\Rightarrow \hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q_N(\theta)$$

where W_N is a $L \times L$ positive definite weighting matrix

GMM Example: Instrumental Variables

Two-stage least squares is a special case of a GMM estimator

Consider the linear regression model

$$y_i = \beta' x_i + \varepsilon_i$$

where β is a vector of K parameters. We have a vector of L instruments, z_i, that are correlated with x_i and must satisfy $E[\varepsilon_i \mid z_i] = 0$, which (with some math) gives L population moment conditions

$$E[z_i(y_i - \beta' x_i)] = 0$$

Replacing the expectation with the sample analog gives

$$\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - \beta' x_i) = 0$$

The GMM estimator, $\hat{\beta}$, is the set of parameters that minimizes

$$Q_N(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - \beta' x_i)\right]' W_N \left[\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - \beta' x_i)\right]$$

Weighting Matrix

From the previous example, the GMM estimator, $\hat{\beta}$, is the set of parameters that minimizes

$$Q_N(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - \beta' x_i)\right]' W_N \left[\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - \beta' x_i)\right]$$

which depends on the choice of weighting matrix, W_N

- W_N must be a $L \times L$ positive definite matrix

The GMM estimator will be equivalent to the 2SLS estimator if we use

$$W_N = \sum_{i=1}^{n} z_i z_i'$$

But another weighting matrix might yield a “better” estimator

Properties of the Generalized Method of Moments Estimator

GMM Assumptions

1. The empirical moments obey the law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} m_i(\theta_0) \xrightarrow{p} 0$$

2. The empirical moments obey the central limit theorem

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n} m_i(\theta_0) \xrightarrow{d} N(0, S_0)$$

   where

$$S_0 = \operatorname{plim} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} m_i(\theta_0) m_j(\theta_0)'$$

3. The derivatives of the empirical moments converge

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial m(\theta)}{\partial \theta'}\right)_{\theta_0} \xrightarrow{p} G_0$$

GMM Assumptions

4. The parameters are identified

$$\theta_1 \neq \theta_2 \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} m_i(\theta_1) \neq \frac{1}{n}\sum_{i=1}^{n} m_i(\theta_2)$$

5. The weighting matrix converges to a finite symmetric positive definite matrix

$$W_N \xrightarrow{p} W_0$$

Consistency of the GMM Estimator

The GMM estimator, $\hat{\theta}$, is a consistent estimator of the true parameter values, θ_0

$$\hat{\theta} \xrightarrow{p} \theta_0$$

Asymptotic Normality of the GMM Estimator

The GMM estimator, $\hat{\theta}$, is asymptotically normal with mean θ_0 and known variance

$$\hat{\theta} \overset{a}{\sim} N\left(\theta_0, \frac{1}{n}(G_0' W_0 G_0)^{-1}(G_0' W_0 S_0 W_0 G_0)(G_0' W_0 G_0)^{-1}\right)$$

where

$$G_0 = \text{plim}\, \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial m(\theta)}{\partial \theta'}\right)_{\theta_0}$$

$$W_0 = \text{plim}\, W_N$$

$$S_0 = \text{plim}\, \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} m_i(\theta_0) m_j(\theta_0)'$$

Optimal Generalized Method of Moments

Optimal Weighting Matrix

If every weighting matrix yields a GMM estimator that is consistent, then we would ideally like to use the weighting matrix that minimizes the variance of the estimator

$$Var(\hat{\theta}) = \frac{1}{n}(G_0' W_0 G_0)^{-1}(G_0' W_0 S_0 W_0 G_0)(G_0' W_0 G_0)^{-1}$$

This variance is minimized when

$$W_0 = S_0^{-1} = \left[\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} m_i(\theta_0)m_j(\theta_0)'\right]^{-1}$$

$$\Rightarrow \quad Var(\hat{\theta}) = \frac{1}{n}(G_0' S_0^{-1} G_0)^{-1}$$

S_0 is the variance-covariance matrix of the empirical moments, so weighting by its inverse minimizes the variance of the GMM estimator

- This is a generalization of FGLS

Optimal (or Two-Step) GMM Estimator

To estimate the optimal (or two-step) GMM estimator, perform GMM twice

1. Perform GMM with any arbitrary weighting matrix
   - The identity matrix is a simple matrix to use
   - Calculate $\widehat{S}$ evaluated at the first-step GMM estimator
2. Perform GMM with $W_N = \widehat{S}^{-1}$ from the first step
   - Calculate $\widehat{S}$ and $\widehat{G}$ evaluated at the second-step GMM estimator
   - Use $\widehat{S}$ and $\widehat{G}$ to estimate the variance covariance matrix

$$\widehat{Var}(\hat{\theta}) = \frac{1}{n}\left(\widehat{G}'\widehat{S}^{-1}\widehat{G}\right)^{-1}$$

where (assuming independence between observations)

$$\widehat{G} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial m_i(\theta)}{\partial \theta'}\right)_{\hat{\theta}}$$

$$\widehat{S} = \frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta}) m_i(\hat{\theta})'$$

Announcements

No class on Tuesday! Monday schedule.

Reading for next time

- Optional: Crawford and Yurukoglu (2012)

Upcoming

- Problem Set 2 is posted, due October 17