Generalized Method of Moments

Xiang Ao

June 15, 2009

1 The Method of Moments

The method of moments (MOM) is merely the following proposal:

Proposition 1.1 (MOM)
To estimate a population moment (or a function of population moments) merely use the corresponding sample moment (or a function of sample moments).

A population moment γ can be defined as the expectation of some continuous function g of a random variable x:

$$\gamma = \mathrm{E}[g(x)] \tag{1}$$

On the other hand, a sample moment is the sample version of the population moment in a particular sample:

$$\hat{\gamma} = \frac{1}{n}\sum[g(x)] \tag{2}$$

2 OLS as a moment problem

Consider the simple linear regression

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \quad \mathbf{u} \sim \mathbf{IID}(\mathbf{0}, \sigma^{\mathbf{2}}). \tag{3}$$

If the model is correctly specified, then

$$\mathrm{E}(\mathbf{X}'\mathbf{u}) = \mathbf{0}. \tag{4}$$

The MOM principle suggests that we replace the left-hand side with its sample analog $\frac{1}{n}\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)$.

Since we know that the true β sets the population moment equal to zero in expectation, it seems reasonable to assume that a good choice of $\hat{\beta}$ would be one that sets the sample moment to zero. The MOM procedure suggests an estimate of β that solves

$$\frac{1}{n}\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0}. \tag{5}$$

The MOM estimator is

$$\hat{\beta} = (\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'y}, \tag{6}$$

which is the same as the OLS estimator.

3 IV as a moment problem

Consider the simple linear regression

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \quad \mathbf{u} \sim \mathbf{IID}(\mathbf{0}, \sigma^{\mathbf{2}}). \tag{7}$$

If the model is mis-specified, then

$$\mathrm{E}(\mathbf{X'u}) \neq \mathbf{0}. \tag{8}$$

We have to find an instrumental variable $\mathbf{Z}$ which is

$$\mathrm{E}(\mathbf{Z'u}) = \mathbf{0}. \tag{9}$$

Or,

$$\mathrm{E}(\mathbf{Z'}(\mathbf{y} - \mathbf{X}\beta)) = \mathbf{0}. \tag{10}$$

The sample analogy of this is

$$\frac{1}{n}\mathbf{Z'}(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0}. \tag{11}$$

That gives us the IV estimator

$$\hat{\beta} = (\mathbf{Z'X})^{-\mathbf{1}}\mathbf{Z'y}. \tag{12}$$

4 The Generalized Method of Moments

4.1 Moments

The expectation $\mathrm{E}(Y^r)$ for any $r = 1, 2, \ldots$ is called the r^{th} (raw) moment of Y. The expectation $\mathrm{E}[(Y - \mathrm{E}(Y))^r]$ is called the r^{th} centered moment of Y.

The mean is the first raw moment.

The variance is the second centered moment.

The third centered moment measures the skewness of the distribution.

The fourth centered moment measures the kurtosis of the distribution. Interpreted as a measure of "fatness of tails".

The standardized kurtosis is

$$k = \frac{E[(Y - E(Y))^4]}{E[(Y - E(Y))^2]^2}.$$

For a normal distribution, $k = 3$.

For a t distribution with $v \geq 5$ degrees of freedom, $k = 3 + 6/(v-4) > 4$. i.e., the t distribution has fatter tails than a normal distribution.

The distribution function of a random variable captures all information about the random variable. It can be shown using all moments also captures all information.

This distinction underlies the relative strengths and weaknesses of ML and GMM.

4.2 GMM

The statistical model takes the general form

$$E[m(Y_i; \theta_0)] = 0 \tag{13}$$

where

- $Y_1, \cdots, Y_2$ are random variables from which the sample $y_1, \cdots, y_n$ is drawn,
- $m(Y, \theta)$ is a function specifying the model,
- θ_0 is the "true value" of the parameter.

$E[m(Y_i; \theta_0)] = 0$ are called the population moment conditions.

Two ideas behind GMM:

1. Replace the population mean $E[.]$ with the sample mean calculated from the observed sample $y_1, \cdots, y_n$.
2. Since $E[m(Y_i; \theta_0)] = 0$, choose $\hat{\theta}_{GMM}$ to make $\frac{1}{n}\sum_{i=1}^{n} m(y_i; \hat{\theta}_{GMM})$ as close to zero as possible.

Define the notation

$$\bar{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(y_i; \theta). \tag{14}$$

$\hat{\theta}_{GMM}$ is chosen to make $\bar{m}(\theta)'\bar{m}(\theta)$ as close to zero as possible.

More generally, $\hat{\theta}_{GMM}$ is chosen to minimize $\bar{m}(\theta)'W\bar{m}(\theta)$ for some weighting matrix W.

4.3 A GMM example

Suppose we have people's income data, which are non-negative, highly skewed and contain large outliers.

Consider the gamma distribution with pdf

$$f(y; \alpha, \beta) = \frac{y^{\alpha-1}\exp(-y/\beta)}{\Gamma(\alpha)\beta^{\alpha}}, \; y, \alpha, \beta > 0. \tag{15}$$

If Y_i has a $\gamma(\alpha_0, \beta_0)$ distribution then

$$E(Y_i) = \frac{\alpha_0}{\beta_0},\ E(Y_i^2) = \frac{\alpha_0 + \alpha^2}{\beta_0^2}. \tag{16}$$

So the two moment conditions are

$$E[m(Y_i; \theta_0)] = 0 \tag{17}$$

where $\theta = (\alpha, \beta)$ and

$$m(Y_i; \theta) = \begin{bmatrix} Y_i - \alpha/\beta \\ Y_i^2 - (\alpha + \alpha^2)/\beta^2 \end{bmatrix} \tag{18}$$

The sample moment conditions are

$$\frac{1}{n} m(y_i; \hat{\theta}_{GMM}) = 0, \tag{19}$$

i.e.

$$\begin{bmatrix} \frac{1}{n}\sum_{i=1}^n y_i - \hat{\alpha}_{GMM}/\hat{\beta}_{GMM} \\ \frac{1}{n}\sum_{i=1}^n y_i^2 - (\hat{\alpha}_{GMM} + \hat{\alpha}_{GMM}^2)/\hat{\beta}_{GMM}^2 \end{bmatrix} = 0. \tag{20}$$

The solution is

$$\begin{bmatrix} \hat{\alpha}_{GMM} \\ \hat{\beta}_{GMM} \end{bmatrix} = \begin{bmatrix} \frac{\bar{y}^2}{s^2} \\ \frac{\bar{y}}{s^2} \end{bmatrix}. \tag{21}$$

4.4 Conditioning

Independence

If X and Y are independent then

$$f(x, y) = f(x)f(y) \tag{22}$$

and hence

$$f(y|x) = f(y). \tag{23}$$

If X and Y are independent then

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)] \tag{24}$$

and hence

$$Cov[g(X), h(Y)] = 0. \tag{25}$$

i.e. all functions of X and Y are uncorrelated.

Law of Iterated Expectations

$$E[Y] = E[E(Y|X)]. \tag{26}$$

Dependence Concepts

- X, Y independent:
$$Cov[g(X), h(Y)] = 0 \tag{27}$$

- X, Y uncorrelated:
$$Cov[X, Y] = 0 \tag{28}$$

- $E[Y|X] = 0$:
$$Cov[g(X), Y] = 0 \tag{29}$$

4.5 Regression

A regression model is a model of $E[Y_i|X_i]$. For example,

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{30}$$

where $E[u_i|X_i] = 0$.

GMM regression

The regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad E[u_i|X_i] = 0 \tag{31}$$

implies the moment condition

$$E[u_i] = 0 \quad \text{and} \quad E[X_i u_i] = 0 \tag{32}$$

That is,

$$E[Y_i - \beta_0 - \beta_1 X_i] = 0 \tag{33}$$

$$E[X_i(Y_i - \beta_0 - \beta_1 X_i)] = 0 \tag{34}$$

The sample moment conditions are

$$\frac{1}{n}\sum_{i=1} n(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{35}$$

$$\frac{1}{n}\sum_{i=1} n x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{36}$$

These are just normal equations for OLS.

A characteristic of GMM: the specification of the model generates the estimator. i.e. only $E[Y_i|X_i] = \beta_0 + \beta_1 X_i$ is assumed.

Note there are no assumptions that u_i is homoscedastic, not autocorrelated or normally distributed. These properties affect the statistical properties of the GMM estimator, not its definition.