GMM - Generalized Method of Moments

Contents

1 GMM estimation, short introduction

Well know that Maximum Likelihood is generally an optimal method for setting up estimators. If we know we are in a setting where ML applies, we should be using that. However, there are many cases where ML can not be applied.

Important cases

- We are not able to specify the actual probability distribution.
- The calculation of the full probability distribution is not feasible.

In both cases we are not able to use neither OLS nor ML. An alternative way of doing estimation is base on an old idea in statistics, that of "mathcing moments"

I want to spend some time on the analysis of the "Generalized Method of Moments," not only because I like it, but also because it is becoming more and more commonly used in current research. Another reason that we want to spend some time on GMM is that almost all the other methods of analysing data we have seen so far can be viewed as special cases of GMM. We can thus use the general results about GMM problems to show results about all these special cases. This is the value of knowing the most general cases, it is easier to specialize a general case to a special case, than to keep track of 10 different special cases.

Another reason that GMM is so popular is that it is easy to see the mapping from the economic model to the estimation setup.

2 GMM intuition: Matching moments

We will start with the intuition of "matching moments."

Suppose

$$x_t \sim (\mu_0, \sigma^2)$$

Consider the problem of estimating μ_0.

We know that

$$E[x_t] = \mu_0$$

What if this is all we are willing to assume?

Consider

$$\frac{1}{T}\sum_{t=1}^{T} x_t$$

By a law of large numbers, we will have

$$\lim_{T\to\infty}\sum_{t=1}^{T} x_t = E[x_t] = \mu_0$$

Thus, consider

$$E[x_t - \mu_0] = 0$$

If we let T increase,

$$\frac{1}{T}\sum_{t=1}^{T} x_t - \mu_0 \to \ 0 \ \text{ as } \ T \to \infty$$

Given this, a reasonable way to estimate μ_0 is to look at the solution of

$$\frac{1}{T}\sum_{t=1}^{T} x_t - \mu = 0,$$

or take as our estimator

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t,$$

the sample mean, which in this case is the natural estimator.

The name "matching moments" is coming from the fact that e.g. the expectation is the "first moment, " and by constructing the "moment condition"

$$\frac{1}{T}\sum_{t=1}^{T} x_t - \mu = 0,$$

we are in some sense "matching" the first moment, the expectation

$$E[x_t - \mu_0] = 0$$

3 General overview of GMM estimation.

The main ingredient of a GMM estimation is a function

$$h(\theta, w_t),$$

with θ parameters to estimate and w_t data. We often use the term "orthogonality condition" about this expectation.

By assumption

$$E[h(\theta_0, w_t)] = 0$$

under the true parameters $\theta = \theta_0$. Define

$$\mathcal{Y}_T = \{w_1, w_2, \ldots, w_T\}$$

$$g(\theta, \mathcal{Y}_T) = \frac{1}{T}\sum_{t=1}^{T} h(\theta, w_t)$$

If the number of parameters to estimate equals the number of orthogonality conditions, we can find $\hat{\theta}$ directly as the solution to

$$g(\hat{\theta}, \mathcal{Y}_T) = 0$$

Otherwise, if the number of parameters to estimate is less the number of orthogonality conditions, we can find $\hat{\theta}$ as

$$\hat{\theta} = \arg\min_{\theta} J(\theta, \mathcal{Y}_T) = \arg\min_{\theta} g(\theta, \mathcal{Y}_T)' W_T g(\theta, \mathcal{Y}_T)$$

where W_T is some positive definite weighting matrix.

To show consistency of GMM, the main steps will consist of

1. Show

$$g(\theta, \mathcal{Y}_T) \to E[h(\theta, w)] \text{ for all } \theta$$

2. Assume

$$\min E[h(\theta_0, w)' h(\theta_0, w)]$$

is a unique minimum, or alternatively

$$E[h(\theta)] \neq 0 \ \forall \ \theta \neq \theta_0$$

3. Assume continutiy of $h(\cdot)$

4. Given this, we argue that

$$\arg\min_{\theta} g(\theta, \mathcal{Y})'_T W_T g(\theta, \mathcal{Y}_T) \overset{P}{\to} \theta_0$$

We will not go into the details of this argument, it is similar to earlier proofs for other cases, but the laws of large numbers used to show this is somewhat advanced. If you are interested, look at the references given below.

3.1 Weighting matrix.

What is the matrix W_T in

$$g(\theta, \mathcal{Y}_T)' W_T g(\theta, \mathcal{Y}_T)$$

The easiest way to see what it should be is to argue by the analogy to GLS, where we found that

$$\widehat{\mathbf{b}}^{gls} = \arg\min_b (\mathbf{y} - \mathbf{X}\mathbf{b})' \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b})$$

The matrix Ω is the covariance matrix of the error terms, and the intution was that the higher the variance of a particular observation, the lower weight should it be given in the estimation.

A similar result turns out to be the case for GMM estimation, the optimal W_T to use should be an estimate of the inverse of the covariance matrix of the moment conditions.

$$W_T = \widehat{S}^{-1}$$

$$S = \text{var}\left(\sum_t h(\theta, w_t)' h(\theta, w_t)\right)$$

The weighting matrix W_T is an important part of GMM, it is what makes the method very robust. If we write out

$$\begin{aligned} S &= \text{var}\left(\sum_t h(\theta, w_t)' h(\theta, w_t)\right) \\ &= \sum_i \sum_j E[h(\theta, w_{t+i})' h(\theta, w_{t+j})], \end{aligned}$$

a kind of average of the error terms.

A particular simple version is to assume independence of the error terms, which implies that

$$S = \sum_t E[h(\theta, w_t)' h(\theta, w_t)],$$

which we would estimate by

$$\widehat{S}_T = \frac{1}{T} \sum_t h(\theta, w_t)' h(\theta, w_t)$$

Already we see that to estimate the matrix S you need an estimate of θ, but to estimate θ we need an estimate of S. The way to get around this circularity is to proceed in steps

1. Estimate θ using the identity matrix $\mathbf{I}$ as a weighting matrix.
2. Estimate $\widehat{S}_T$ using this θ.
3. Re–estimate θ using $\widehat{S}_T$ as a weighting matrix.

3.2 Properties of GMM estimators.

The main general results about GMM estimators are that, under the appropriate regularity conditions, which we will not go into,

$$\hat{\theta}_T \xrightarrow{P} \theta_0$$

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, V)$$

where

$$V = \left[D S^{-1} D'\right]^{-1}$$

$$D = E\left[\frac{\partial}{\partial \theta} g(\theta, \mathcal{Y})\right]$$

$$S = \text{ covariance matrix of moment conditions.}$$

3.3 Testing over-identifying restrictions

Specification of a GMM model will consisting in finding a set of "moment conditions" which have expectation zero. It is often the case that the model will supply more moment conditions than we have parameters to estimate. Let

$$E[h(x_t, b)] = 0$$

be the moment conditions. Suppose we have n moment conditions and $r < n$ parameters to estimate. If the model is correctly specified, at the true parameters the sample equivalent of the moment condition will go to zero for all the moment conditions. Suppose we now use only the first r moment conditions to do the estimation of the parameters. This will choose parameters θ to set the sample mean of

$$\sum_{t=1}^{T} \begin{bmatrix} h_1(\theta, w_t) \\ h_2(\theta, w_t) \\ \vdots \\ h_r(\theta, w_t) \end{bmatrix} = 0$$

where $h_i(\theta, w_t)$ signifies the i'th element of $h(\theta, w_t)$. If the parameters are correct, and the model correctly specified, the sample mean of the moment conditions that are not used in the estimation should also be close to zero. If they are not, this is a sign that the model is not correctly specified. The "test of over-identifying restrictions" measure this distance from zero of the "left over" moment conditions, and will reject the model formulation if this statistic is large. You will often see this test termed "Hansen's J-test."

We construct the test statistic as

$$J(\theta) = g_T(\theta, y) S^{-1} g_T(\theta, y)$$

Since $g_T(\theta)$ is asymptotically normal with limiting covariance matrix S, $J(\theta)$ is chi-square distributed with degrees of freedom equal to the number of moment conditions less the number of parameters to estimate.

The J-test is a test of the model formulation, if

$$J(\hat{\theta}) > \text{ critical value},$$

then we may want to think again about the model formulation.

In practice this test is not very powerful, it can be hard to reject a mis-specified model using this particular test.

3.4 Use of conditioning information.

The ability to use *conditioning information* in a meaningful way is one of the major reasons for GMM to be of very wide use.

Note a couple of ways to use conditioning information

- Use of variables in the information set as instruments in the estimation.
- Try to model the conditional expectations directly (latent variables)

4 OLS as a GMM estimation.

As another example of GMM setup, consider an OLS structure

$$y_t = x_t b + \epsilon_t$$

where y_t and x_t are scalars, and the following are satisfied

$$E[\epsilon_t] = 0$$

$$E[x_t \epsilon_t] = 0$$

$$\text{var}(\epsilon_t) = \sigma^2$$

Moment condition

$$E[x_t \epsilon_t] = E[x_t(y_t - x_t b)]$$

Sample equivalent

$$\sum_t x_t(y_t - x_t b) = 0$$

Solve for GMM estimator

$$\hat{b}^{GMM} = \left[\sum_t x_t x_t\right]^{-1} \left[\sum_t x_t y_t\right]$$

If we define the appropriate vectors

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

we can write

$$\hat{b}^{GMM} = (x'x)^{-1} x'y$$

and

$$\text{var}(\epsilon) = \sigma^2 I$$

Now, what is the standard error of the GMM estimate $\hat{b}^{GMM}$.

using the result

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, V)$$

$$V = \left(D S^{-1} D'\right)^{-1}$$

$$D = \frac{\partial}{\partial \theta} g(\theta, \mathcal{Y})$$

$$S = cov(moment\ conditions)$$

Mapping the OLS example into this notation

$$\theta = b$$

$$S = \sigma^2 (X'X)$$

$$D = \frac{\partial}{\partial b} x'(y - xb) = x'x$$

$$DS^{-1}D = (x'x)(\sigma^2(x'x))^{-1}(x'x) = \frac{1}{\sigma^2}(x'x)$$

$$(DS^{-1}D)^{-1} = \sigma^2 (x'x)^{-1}$$

which should be familiar.

5 Running GMM in R

There is a very good implementation of GMM estimation in R, which covers many of the relevant applications for finance.

The user has to do is to write code for calculation of the moment conditions, and then R takes care of the rest. The moment conditions can be specified by writing a function returning the matrix of moment conditions, or, in the case of a linear model, by simply writing the linear model in the same way as a OLS regression.

For documentation of the gmm package, see Chaussé (2010).

We will use some examples to illustrate the use of the package.

Exercise 1.

Consider the model

$$y = a + bx + \varepsilon$$

Simulate the following model letting x be the numbers from 1 to 10, $a = 1$, $b = 1$ and ε is normally distrubuted with mean zero and variance one.

1. Estimate the model using OLS.
2. Estimate the model using GMM.

Solution to Exercise 1.

The following is the R code which does this, and then the output of the two estimations.

First simulating the model

```
x <- 1:10
b <- 1
a <- 1
e <- rnorm(10)
y <- a + b * x + e
```

Running the OLS regression

```
reg <- lm(y~x)
summary(reg)
```

with output

```
lm(formula = y ~ x)
Residuals:
    Min      1Q  Median      3Q     Max
-2.1606 -0.4692 -0.0490  0.7747  1.5867
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.3087     0.7659   1.709    0.126
x             0.9777     0.1234   7.921 4.69e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.121 on 8 degrees of freedom
Multiple R-squared: 0.8869,Adjusted R-squared: 0.8728
F-statistic: 62.74 on 1 and 8 DF,  p-value: 4.689e-05
> reg$coefficients
(Intercept)           x
  1.3087486   0.9776919
```

and then doing the same using GMM. Note the need to load the gmm library.

```
library(gmm)
res <- gmm(y~x,x)
summary(res)
```

results in the output

```
gmm(g = y ~ x, x = x)
Method:  twoStep
Kernel:  Quadratic Spectral
Coefficients:
             Estimate    Std. Error  t value     Pr(>|t|)
(Intercept)  1.3087e+00  9.6023e-01  1.3630e+00  1.7290e-01
x            9.7769e-01  1.2273e-01  7.9663e+00  1.6345e-15
J-Test: degrees of freedom is 0
                J-test                P-value
Test E(g)=0:    4.26970964950873e-29  *******
> res$coefficients
(Intercept)           x
  1.3087486   0.9776919
```

6 Summarizing GMM

Intuition: "Matching Moments"

$$E[] = 0$$

(first moment) Also, potentially

$$E[()^2 - \sigma^2] = 0$$

(second moment)

Basis for constructing estimators.

Allow for estimation in many settings where estimation otherwise impossible.

Generality costs: Less precision when e.g. ML applies.

Important cases:

- Conditional expectations can be the basis for modelling.
- Robust to more general error structure (heteroskedasticity and autocorrelation robust)

References

Pierre Chaussé. Computing generalized method of moments and generalized empirical likelihood with r. *Journal of Statistical Software*, 34(11):1–35, 2010.