GMM - Generalized method of moments

GMM Intuition: Matching moments

You want to estimate properties of a data set $\{x_t\}_{t=1}^{T}$. You assume that x_t has a constant mean and variance.

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

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?
How can this be used as a basis of estimation?
Why is this called "Matching moments?"

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

OLS as GMM

Consider the classical linear model,

$$y_t = x_t' b + u_t$$

Under the standard assumption of $E[x_t u_t] = 0$, we have $E[x_t(y_t - x_t' b)] = 0$. In GMM applications, we term these the *moment conditions*. These moment conditions will always follow from our model. Estimation in this case will be to set the sample equivalent of the moment conditions,

$$\frac{1}{T}\sum_{t=1}^{T} x_t \left(y_t - x_t' b\right)$$

equal to zero.

1. Show that doing this results in the GMM estimator

$$\hat{b}_T^{gmm} = \left[\sum_{t=1}^{T} x_t x_t'\right]^{-1} \left[\sum_{t=1}^{T} x_t y_t\right]$$

   and show that this is the same as the OLS estimator in this case.

We can use this to find the estimated parameters b_T^{gmm}:

$$\frac{1}{T}\sum_{t=1}^{T} x_t\left(y_t - x_t' b\right) = 0$$

$$\sum_{t=1}^{T} x_t y_t - \sum_{t=1}^{T} x_t x_t' b = 0$$

$$\left[\sum_{t=1}^{T} x_t y_t\right] = \left[\sum_{t=1}^{T} x_t x_t'\right] b$$

$$\hat{b}_T^{gmm} = \left[\sum_{t=1}^{T} x_t x_t'\right]^{-1} \left[\sum_{t=1}^{T} x_t y_t\right]$$

Note that the GMM estimator coincides with the OLS estimator in this case.

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

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

We will not go into the details of this argument

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

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

The matrix Ω is the covariance matrix of the error terms.
GMM estimation similar, the optimal W_T should be an estimate of the inverse of the covariance matrix of the moment conditions.

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

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

Weighting matrix.

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

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

$$= \sum_i \sum_j E[h(\theta, w_{t+i})' h(\theta, w_{t+j})],$$

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.

Properties of GMM estimators

The main general results about GMM estimators are that, under the appropriate regularity conditions,

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

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

where

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

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

$S =$ covariance matrix of moment conditions.

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.

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)

Running GMM in R

There is a very good implementation of GMM estimation in R, which covers many of the relevant applications for finance.
Essentially, all 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.
We will use some examples to illustrate the use of the package.

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.

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
```

Summarizing GMM

Intuition: "Matching Moments"

$$E[] = 0$$

(first moment) Also

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

(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)

Asset pricing applications

Examples

- Consumption based pricing: $m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}$
- CAPM: $m = f(r_m)$
- Fama French: $m = f(r_m, SMB, HML)$.