Chapter 10

GMM in Explicit Discount Factor Models

Main Content

- The Basic Idea Of The GMM
- The Recipe
- Interpreting The GMM Procedure
- Applying GMM

The Basic Idea of GMM

- The asset pricing model predicts

$$E(p_t) = E(m(data_{t+1}, parameters)x_{t+1})$$

- The most natural way to check this prediction is to examine sample averages, i.e. to calculate

$$\frac{1}{T}\sum_{t=1}^{T} p_t \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}[m(data_{t+1}, paramenters)x_{t+1}]$$

- GMM *estimates* the parameters by making the sample averages as close to each other as possible.

- It suggests that we *evaluate* the model by looking at how close the sample averages of price and discounted payoff *are* to each other.

- For Example: use GMM method to estimate the degrees of freedom of t distribution.

$$f_{Y_t}(y_t;v)=\frac{\tau[(v+1)/2]}{(\pi v)^{1/2}\tau(v/2)}[1+(y_t^2/v)]^{-(v+1)/2};$$

- the second moment of t distribution is:

$$u_2 = E(Y_t^2) = v/(v-2)$$

the forth moment of t distribution is:

$$u_4 = E(Y_t^4) = \frac{3v^2}{(v-2)(v-4)}$$

- If we want to choice the degrees freedom v that make the second moment and the forth moment close to the second moment and the forth moment of sample,we need minimize:

$$Q(v; y_T; \cdots ; y_1) = g'Wg \qquad g = \begin{pmatrix} \hat{u}_{2,T} - \dfrac{v}{v-2} \\ u_{4,T} - \dfrac{3v^2}{(v-2)(v-4)} \end{pmatrix}$$

where w is one of (2*2) matrix, $\hat{u}_{2,T} = \sum_{t=1}^{T} y_t^2 \quad \hat{u}_{4,T} = \sum_{t=1}^{T} y_t^4$

- When you want to go on GMM estimation,you need find one map g satisfy E(g)=0

10.1 The Recipe

- Discount factor models involve some unknown parameters I write $m_{t+1}(b)$,so any asset pricing model implies:

$$E(p_t) = E[m_{t+1}(b)x_{t+1}]$$

- It's easiest to write this equation in the form

$$E[m_{t+1}(b)x_{t+1} - p_t] = 0$$

There x and p are vector.we typically check whether a model for m can price a number of assets simultaneously.

- It's convenient to define the *errors* $u_t(b)$ as the object whose mean should be zero

$$u_{t+1}(b) = m_{t+1}(b)x_{t+1} - p_t$$

- Define $g_T(b)$ as the sample mean of the errors u_{t+1}

$$g_{T(b)} = \frac{1}{T}\sum_{t=1}^{T} u_t(b) = E_T[u_t(b)] = E_T[m_{t+1}(b)x_{t+1} - p_t]$$

- The second equality introduces the handy notation E_t for sample means

$$E_t = \frac{1}{T}\sum_{t=t+1}^{T}(\bullet)$$

- The *first stage estimate* of b minimizes a quadratic form of the sample mean of the errors:

$$\hat{b}_1 = \arg\min_{\{\hat{b}\}} g_T(\hat{b})' W g_T(\hat{b})$$

for some arbitrary matrix W ▹This estimate is consistent and asymptotically normal.

- Using $\hat{b}_1$,form an estimate $\hat{S}$ of

$$S = \sum_{j=-\infty}^{\infty} E[u_t(b)u_{t-j}(b)']$$

- Form a *second stage* estimate $\hat{b}_2$ using the matrix $\hat{S}$ in the quadratic form:

$$\hat{b}_2 = \arg\min_b g_T(b)' \hat{S}^{-1} g_T(b)$$

- $\hat{b}_2$ is a consistent, asymptotically normal, and asymptotically efficient estimate of the parameter vector b.
- “Efficient” means that it has the smallest variance-covariance matrix among all estimators that set different linear combinations of $g_T(b)$ to zero.
- The variance-covariance matrix of $\hat{b}_2$ is

$$\mathrm{var}(\hat{b}_2) = \frac{1}{T}(d'S^{-1}d)^{-1}$$

where

$$d = \frac{\partial g_T(b)}{\partial b}$$

or more explicitly

$$d = E_T(\frac{\partial}{\partial b}[(m_{t+1}(b)x_{t+1} - p_t)])|_{b=\hat{b}}$$

- This variance-covariance matrix can be used to test whether a parameter or group of parameters are equal to zero, via $\frac{\hat{b}_i}{\sqrt{\mathrm{var}(\hat{b})_{ii}}} \sim N(0,1)$

  and $\hat{b}_j[\mathrm{var}(\hat{b}_{jj})]^{-1}\hat{b}_j \sim \chi^2$ (#included b)

  where $b_j = subvector$, $\mathrm{var}(b)_{jj} = submatrix$.
- Finally, the *test of overidentifying restrictions* is a test of the overall fit of the model. the minimized value of the second-stage objective is distributed χ^2 with degrees of freedom equal to the number of moments less the number of estimated parameters.

$$TJ_T = T\min_{\{b\}}[g_T(b)'S^{-1}g_T(b)] \sim \chi^2(\#moments - \#parameters)$$

10.2 Interpreting the GMM Procedure

Pricing errors

- The moment conditions

$$g_T(b) = E_T[m_{t+1}(b)x_{t+1}] - E_T[p_t]$$

- we pick parameters so that the model's predicted prices are as close as possible to the actual prices.
- We can evaluate the model by how large these pricing error are.
- Do we have other method to evaluate the model?

- In the language of expected returns

$$E(R^e) = -\frac{\mathrm{cov}(m, R^e)}{E(m)}$$

we can write the pricing error as

$$g(b) = E(mR^e) = E(m)(E(R^e) - (-\frac{\mathrm{cov}(m, R^e)}{E(m)}))$$

- g(b)=1/R^f(#actual mean return – #predicted mean return)
- If we express the model in expected return-beta language

$$E(R^{ei}) = \alpha_i + \beta_i' \lambda$$

we can write the pricing error as

$$g(b) = \frac{1}{R^f} \alpha_i$$

First Stage Estimates

- We like to pick b to make every element of $g_T(b) = 0$
- But there are usually more moment conditions (returns times instruments) than there are parameters.
- we choose b to make the pricing error $g_T(b)$ as small as possible, by minimizing a quadratic form

$$\min_{\{b\}} g_T(b)' W g_T(b)$$

- W is a *weighting matrix* that tells us how much attention to pay to each moment.
- When $g_T(b)$ is nonlinear function of b,you can use a numerical search to find the value of b.

Second-Stage Estimates:Why S^{-1}

- The weighting matrix directs GMM to emphasize some moments or linear combinations of moments at the expense of others.
- The second –stage estimate picks a weighting matrix based on statistical considerations.
- We should pay less attention to pricing errors from asset with high variance of pricing errors.
- One could implement this idea by using a matrix composed of inverse variances of $g_T(b)$ on the diagonal.

- Exploiting the assumption that $E(u_t)=0$,and that u_t is stationary,we have

$$\mathrm{var}(g_T)=\mathrm{var}\left(\frac{1}{T}\sum_{t=1}^{T}u_{t+1}\right)$$

$$=\frac{1}{T^2}[TE(u_t u_t{}')+(T-1)(E(u_t u'_{t-1})+E(u_t u'_{t+1}))+\cdots].$$

As $T\rightarrow+\infty$ $(T-j)/T\rightarrow 1$, so

$$\mathrm{var}(g_T)\rightarrow\frac{1}{T}\sum_{j=-\infty}^{\infty}E(u_t u'_{t-j})=\frac{1}{T}S.$$

- The last equality denotes S,known as the *spectral density matrix at frequency zero of* u_t .
- When the u_t are uncorrelated over time ,the previous equation reduces to

$$\mathrm{var}\left(\frac{1}{T}\sum_{t=1}^{T}u_{t+1}\right)=\frac{1}{T}E(uu')=\frac{\mathrm{var}(u)}{T}$$

- Hansen(1982) shows formally that the choice

  $$W = S^{-1} \qquad S = \sum_{j=-\infty}^{\infty} E(u_t u'_{t-j}),$$

  is the statistically optimal weighing matrix.
- It mean that it produces estimates with lowest asymptotic variance.
- The first- and second-stage estimates are all consistent. But the second-stage is more efficient, meaning that sampling variation in the estimated parameters is lower.
- If we consider the mean of economic,we usually use the first-stage estimates.

Standard Errors

- The formula for the standard error of the estimate

$$\mathrm{var}(\overset{\frown}{b}_2) = \frac{1}{T}(d'S^{-1}d)^{-1}$$

- Suppose there is only one parameter and one moment. S/T is the variance matrix of the moment $g_T(b)$. d^{-1} is $[\frac{\partial g_T}{\partial b}]^{-1} = \frac{\partial b}{\partial g_T}$ .Then the delta method formula gives:

$$\mathrm{var}(\overset{\frown}{b}_2) = \frac{1}{T}\frac{\partial b}{\partial g_T}\mathrm{var}(g_T)\frac{\partial b}{\partial g_T}$$

- The delta method mean that the asymptotic variance of $f(x)$ is $f'(x)^2\,\mathrm{var}(x)$

J_T Test

- The J_T test asks whether the pricing error are "big" by statistical standards – if the model is true.

$$TJ_T = T[g_T(\overset{\frown}{b})' S^{-1} g_T(\overset{\frown}{b})] \sim \chi^2(\#moments - \#parameters)$$

- The reduction in degrees of freedom corrects for the fact that S is really the covariance matrix of g_T for fixed b.

10.3 Applying GMM

Notation; Instruments and Returns

- Most of the effort involved with GMM is simply mapping a given problem into the very general notation. In the asset pricing model,we can use this equation:

$$E(m_{t+1}x_{t+1} - p_t) = 0$$

- We often test asset pricing models using returns, in which case the moment conditions are

$$E(m_{t+1}R_{t+1} - 1) = 0$$

- It is common to add *instruments* as well. Mechanically, you can multiply both sides of

$$1 = E_t(m_{t+1}(b)R_{t+1})$$

by any variable z_t observed at time t before taking unconditional expectations, resulting in

$$E(z_t) = E(m_{t+1}(b)R_{t+1}z_t)$$

- Expressing the result in $E(\leq) = 0$ form

$$0 = E\{[m_{t+1}R_{t+1} - 1]z_t\}$$

- We can do this for a whole vector of returns and instruments, multiplying each return by each instrument.
- For example, if we start with two returns $R = [R^a, R^b]'$ and one instrument z, the equation looks like

$$E\left\{\begin{bmatrix} m_{t+1}(b)R^a_{t+1} \\ m_{t+1}(b)R^b_{t+1} \\ m_{t+1}(b)R^a_{t+1} \\ m_{t+1}(b)R^b_{t+1} \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ z_t \\ z_t \end{bmatrix}\right\} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

- Using the Kronecker product $\otimes$ “meaning multiply every element by every other element” we can denote the same relation compactly by

$$E\{[m_{t+1}R_{t+1}-1]\otimes z_t\}=0$$

Forecast Error and Instruments

- The error $u_{t+1} = m_{t+1}R_{t+1} - 1$ is the ex-post discounted return.

  $u_{t+1} = m_{t+1}R_{t+1} - 1$ represents a forecast error.
- Like any forecast error, u_{t+1} should be conditionally and unconditionally mean zero.
- In an econometric context, z is an *instrument* .It should be uncorrelated with the error u_{t+1}
- Adding instruments basically checks that the ex-post discounted return is unforecastable by linear regressions.
- If an asset's return is higher than predicted when z_t is unusually high, but not on average,scaling by z_t will pick up this feature of the data.

Stationarity and distributions

- The GMM distribution theory does require m‘ p‘ and x must be *stationary* random variables.
- Statistical definition of stationarity is that the joint distribution of x_t, x_{t-j} depends only on j and not on t.
- Sample averages must converge to population means as the sample size grows, and stationarity implies this result.
- Assuring stationarity usually amounts to a choice of sensible units.

- For example, we could express the pricing of a stock as $p_t = E_t[m_{t+1}(d_{t+1} + p_{t+1})]$
- It would not be wise to do so. For stocks, p and d rise over time and so are typically not stationary; their unconditional means are not defined.
- It is better to divide by p_t and express the model as

$$1 = E_t[m_{t+1}\frac{d_{t+1} + p_{t+1}}{p_t}] = E_t(m_{t+1}R_{t+1})$$

The stock *return* is plausibly stationary.

- Dividing by dividends is an alternative and I think underutilized way to achieve stationarity.

$$\frac{p_t}{d_t} = E_t[m_{t+1}(1 + \frac{p_{t+1}}{d_{t+1}})\frac{d_{t+1}}{d_t}]$$

- Now we map $(1+\frac{p_{t+1}}{d_{t+1}})\frac{d_{t+1}}{d_t}$ into x_{t+1} and $\frac{p_t}{d_t}$ into p_t .
- This formulation allows us to focus on *prices* rather than one-period returns.
- bond prices and yields do not grow over time.so it might be all right to examine

$$p_t^b = E(m_{t+1}1)$$

- Stationarity is not always a clear-cut question in practice. As variables become less stationary, as they experience longer swings in a sample, the asymptotic distribution can becomes a less reliable guide to a finite-sample distribution.
- Many econometric techniques require assumptions about distributions. GMM *do not* include the usual assumptions

The End

Thank you!