Generalized **M**ethod of **M**oments (GMM)
Forelæsningsnoter til FR86

Jesper Lund

mail@JesperLund.com
http://www.JesperLund.com

9. oktober 2002

1 Indledning

Formålet med denne forelæsningsnote er at give læseren en introduktion til GMM, specielt med henblik på estimation af parametre i modeller for stochastic discount factors.[1] Der er relativt få konkrete eksempler i denne note, men det skyldes primært at realistiske eksempler må afvente gennemgangen af dynamiske asset-pricing modeller i uge 44-45.

Det vises desuden i forelæsningsnoten, at OLS estimatoren i den lineære regressionsmodel kan opfattes som en GMM estimator. Faktisk, kan de fleste kendte estimationsmetoder (OLS, Maximum Likelihood) opfattes som GMM estimation.

2 Moment baseret estimation

I det følgende er $\boldsymbol{\theta}$ en $d \times 1$ vektor af ukendte parametre, og $\boldsymbol{\theta}_0$ betegner den "sande" men ukendte værdi af parameter vektoren. Vores opgave er at producere en estimator for $\boldsymbol{\theta}$ ud fra et datasæt at T observationer $\mathbf{Z}_{T\times M}$, hvor række t i matricen $\mathbf{Z}$ er $M \times 1$ vektoren $\mathbf{z}_t$ (dimensionerne er ikke væsentlige her). Normalt vil datasættet bestå af tidsserie data, hvorfor der skal tages stilling til issues som autokorrelation.

Vi forestiller os nu at der eksisterer en $n \times 1$ vektor funktion $\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})$ (hvor $n \geq d$) med følgende centrale egenskab:

$$E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\right] = \mathbf{0} \text{ kun hvis } \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{1}$$

Hvis vi nu kendte den forventede værdi af vektor funktionen $\mathbf{f}$, $E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\right]$, kunne vi finde den sande værdi af parameteren ved at løse ligningssystemet (1). Vi har nemlig antaget at $E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\right] = \mathbf{0}$ kun holder hvis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Den type antagelser kaldes formelt *identification*, idet antagelsen sikrer at det er muligt entydigt at estimere den ukendte parameter.

Disse betragtninger er i sagens natur noget fiktive, eftersom vi ikke kender den forventede værdi af funktionen $\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})$ (det skyldes bl.a. at den forventede værdi afhænger af $\boldsymbol{\theta}_0$). Men vi kan erstatte middelværdien med et simpelt gennemsnit over de T observationer:

$$\mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta}) \tag{2}$$

Ved hjælp af et statistisk teorem som hedder "uniform law of large numbers" (ULLN) kan det vises at

$$\mathbf{g}_T(\boldsymbol{\theta}) \to E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\right] \text{ for } T \to \infty. \tag{3}$$

Dette resulat motiverer at bruge $\mathbf{g}_T(\boldsymbol{\theta})$, som vi kan estimere, istedet for $E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\right]$ til at estimere parameteren $\boldsymbol{\theta}$. Den statistiske usikkerhed for estimatoren $\hat{\boldsymbol{\theta}}$ kan direkte henføres til den statistisk usikkerhed ved at estimere middelværdien ved $\mathbf{g}_T(\boldsymbol{\theta})$.

[1] Den oprindelige artikel om GMM er Hansen (1982), som dog er meget teknisk (og svær at læse). Hansen & Singleton (1982) er lettere at læse, og den behandler specifikt brug af GMM til at estimere parametre i en CRRA nyttefunktion ud fra asset returns.

Hvis $n > d$ kan vi ikke bare sætte $\mathbf{g}_T(\boldsymbol{\theta}) = 0$ og forsøge at løse dette ligningssystem. Godt nok har (1) en entydig løsning per vore antagelser (selv om der er flere ligninger end ubekendte), men det samme gælder ikke når vi estimerer $E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\right]$ med $\mathbf{g}_T(\boldsymbol{\theta})$. I stedet vælger man en $n \times n$ symmetrisk (og positiv definit) vægtningsmatrix $\mathbf{W}_T$, og GMM estimatoren $\hat{\boldsymbol{\theta}}$ defineres nu som den værdi der minimerer den kvadratiske form

$$J_T(\boldsymbol{\theta}) = \mathbf{g}_T(\boldsymbol{\theta})' \mathbf{W}_T \mathbf{g}_T(\boldsymbol{\theta}). \tag{4}$$

Generelt er $\mathbf{g}_T(\boldsymbol{\theta})$ en ikke-lineær funktion af $\boldsymbol{\theta}$, hvilket gør det nødvendigt at anvende en numerisk optimeringsmetode for at minimere GMM kriteriefunktionen (4), se afsnit 7.

3 OLS som et eksempel på GMM

På dette tidspunkt sidder de fleste læsere sikkert med em opfattelse af at GMM dels er meget abstrakt, dels er meget forskellig fra estimationsmetoder som OLS, hvor vi har en intuitivt forståelig kriteriefunktion som skal minimeres. Formålet med dette afsnit er at vise at OLS rent faktisk er en GMM estimator.

Vi tager udgangspunkt i den klassiske regressionsmodel $y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$, og vi lader $\mathbf{z}_t = (y_t, \mathbf{x}_t')'$ og definerer funktionen $\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})$ som

$$\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta}) = \mathbf{x}_t \left(y_t - \mathbf{x}_t'\boldsymbol{\theta}\right). \tag{5}$$

Da der her er ligeså mange momentbetingelser som parametere, kan vi beregne GMM estimatoren ved at løse ligningssystemet

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \left(y_t - \mathbf{x}_t'\boldsymbol{\theta}\right) = \mathbf{0}. \tag{6}$$

Da dette er et lineært ligningssystem, kan løsningen bestemmes som

$$\begin{aligned} &\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right) - \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)\boldsymbol{\theta} = \mathbf{0} \;\Leftrightarrow \\ &\hat{\boldsymbol{\theta}} = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right), \end{aligned} \tag{7}$$

hvilket er den sædvanlige formel for OLS estimatoren $\hat{\boldsymbol{\theta}}$. Den opmærksomme læser vil måske have bemærket, at momentbetingelsen (5) er første-ordens betingelsen for OLS kriteriefunktionen (summen af de kvadrerede afvigelser). Dette er ingen tilfældighed, idet minimering af en funktion er ækvivalent med at beregne gradienten, sætte den lig med nul-vektoren, og løse det ligningssystem som fremkommer derved. På samme måde kan det vises at maximum likelihood er en GMM estimator, hvor gradienten til $\log L(\boldsymbol{\theta})$ udgør momentbetingelsen.

4 Vægtningsmatricen for GMM

Hvis $d = n$, altså hvis der er lige mange momentbetingelser som parametre, er det ligegyldigt hvad vi vælger som vægtningsmatrix. Der vil generelt være en $\hat{\boldsymbol{\theta}}$, som sikrer at $\mathbf{g}_T(\hat{\boldsymbol{\theta}}) = 0$. I de andre tilfælde, hvor $n > d$, kan det vises at der eksisterer en optimal vægtningsmatrix (hermed menes: den vægtningsvarians som giver GMM estimatoren den lavest mulige varians), nemlig

$$\mathbf{W}_0 = \left(\mathrm{Cov}\left[\sqrt{T}\mathbf{g}_T(\boldsymbol{\theta})\right]\right)^{-1} \tag{8}$$

Hvis der ikke er autokorrelation i den stokastiske process $\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})$, kan W_0 umiddelbart estimeres som

$$\mathbf{W}_T = \left(\sum_{t=1}^{T} \frac{1}{T}\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})'\right)^{-1}. \tag{9}$$

Eftersom det vides at middelværdien for $\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})$ er 0, er dette den inverse kovariansmatricen for den stokastiske vektor $\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta})$.

Bestemmelse af den optimale vægtningsmatrix er mere kompliceret, hvis vi ikke kan antage fravær af autokorrelation. Generelt er det nødvendigt at bruge en HAC (heteroskedasticitet og autokorrelation konsistent) estimator som Newey-West (1987) metoden, men da generelt ikke vil støde på situationer hvor dette er nødvendigt (i forbindelse med GMM), holder vi os til den basale formel (9).

4.1 Iterativ GMM

Et mere praktisk problem er at den optimale vægtningsmatrix afhænger af $\boldsymbol{\theta}$, og meningen med at bestemme $\mathbf{W}_T$ er netop at estimere den ukendte parameter $\boldsymbol{\theta}$. Det problem løser man normalt ved at bruge den iterative GMM estimator.

I første trin estimeres $\boldsymbol{\theta}$ med den ikke-optimale vægtningsmatrix $\mathbf{W}_T = \mathbf{I}$. Det kan vises at denne estimator er konsistent (dvs. $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ for $T \to \infty$), men naturligvis ikke efficient (da vi ikke bruger den optimale vægtningsmatrix). Men vi kan nu bruge dette GMM estimat til at beregne $\mathbf{W}_T$ ud fra formel (9), og baseret på denne vægtningsmatrix udfører vi en ny GMM estimation (trin to i den iterative udgave af GMM). Normalt vil man nu beregne vægtningsmatricen ud fra det seneste GMM estimat (dvs. trin to), og bruge denne vægtningsmatrix til en ny GMM estimation. Denne iterative procedure fortsætter indtil der ikke sker nogen ændring i den estimerede værdi af $\boldsymbol{\theta}$ i forhold til den forudgående iteration (stop kriteriet). Under normale omstændigher vil denne procedure konvergere ret hurtigt.

5 Asymptotiske fordeling for GMM estimatoren

Generelt har vi kun asymptotiske resultater for GMM estimatorens fordeling, og vi bruger disse som approksimation for fordelingen i en endelig stikprøve. De centrale resultater er

- GMM estimatoren er konsistent, dvs. $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ for $T \to \infty$. Husk at dette gælder uanset hvad vi vælger som vægtningsmatrix.

- Hvis vi bruger den optimale vægtningsmatrix er

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\left(\mathbf{G}_T(\hat{\boldsymbol{\theta}})'\mathbf{W}_T(\hat{\boldsymbol{\theta}})\mathbf{G}_T(\hat{\boldsymbol{\theta}})\right)^{-1}, \tag{10}$$

hvor

$$\boldsymbol{G}_T(\boldsymbol{\theta})_{n\times d} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \mathbf{f}(\mathbf{z}_t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}. \tag{11}$$

- Hvis modellen er korrekt specificeret og $n > d$ (samt at vi bruger den optimale vægtningsmatrix), er

$$TJ_T(\hat{\boldsymbol{\theta}}) \sim \chi^2(n-d), \tag{12}$$

hvor $J_T(\hat{\boldsymbol{\theta}})$ er den minimerede værdi af GMM kriteriefunktionen. Dette test er standard inden for GMM, og det bruges som et generelt (omnibus) specifikationstest. Man kan opfatte $n-d$ som modellens frihedsgrader, nemlig forskellen mellem antal af momentbetingelser og antal parametre.

6 Estimation af SDF modeller vha. GMM

I forbindelse med intertemporale asset-pricing modeller, støder vi jævnligt på restriktioner (kaldet Euler ligninger) af typen

$$E_t\left[m_{t+1}(\boldsymbol{\theta})(1+R_{t+1})\right] = 1, \tag{13}$$

hvor $R_{t+1} = (P_{t+1} + D_{t+1})/P_t - 1$ er afkastet med diskret tilskrivning, $E_t[\cdot]$ er den betingede forventning givet information på tidspunkt t, og $m_{t+1}(\boldsymbol{\theta})$ er en *stochastic discount factor* (SDF) funktion.[2] Denne SDF afhænger af nogle parametre $\boldsymbol{\theta}$, som vi skal estimere. Mest mest hyppige SDF specifikation er baseret på CRRA nyttefunktionen (constant relative risk aversion), hvor

$$m_{t+1}(\rho,\gamma) = \rho\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{14}$$

Hvis vi lader $\mathbf{x}_t$ være en $n \times 1$ vektor af variable som er indeholdt i informationssættes på tidspunkt t (læs: er kendt på tidspunkt t), kan vi bruge følgende moment betingelse til en GMM estimation:

$$\mathbf{f}(\mathbf{z}_t,\boldsymbol{\theta}) = \mathbf{x}_t\left[m_{t+1}(\boldsymbol{\theta})(1+R_{t+1}) - 1\right]. \tag{15}$$

[2] I det konkrete tilfælde er R_{t+1} en skalar, idet der kun er et aktiv. Estimationsmetoden, der beskrives i det følgende, kan imidlertid nemt generaliseres til situationer hvor vi inddrager afkastet på flere aktiver, dvs. hvor $\mathbf{R}_{t+1}$ er en vektor.

Dette er en valid momentbetingelse idet den ubetingede forventning er $\mathbf{0}$ hvis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Beviset herfor er som følger:

$$\begin{aligned} E\left[\mathbf{f}(\mathbf{z}_t, \boldsymbol{\theta}_0)\right] &= E\left\{E_t\left(\mathbf{x}_t\left[m_{t+1}(\boldsymbol{\theta}_0)(1+R_{t+1}) - 1\right]\right)\right\} \\ &= E\left\{\mathbf{x}_t E_t\left[m_{t+1}(\boldsymbol{\theta}_0)(1+R_{t+1}) - 1\right]\right\} \\ &= E\left\{\mathbf{x}_t \mathbf{0}\right\} = \mathbf{0}. \end{aligned} \tag{16}$$

Bemærk at vi bruger loven om itererede forventninger i den første linie af ligning (16), mens vi i den anden ligning udnytter at $\mathbf{x}_t$ er en konstant når vi betinger på informationen på tidspunkt t, og vi kan derfor flytte $\mathbf{x}_t$ uden for forventningsoperatoren. Endelig bruger vi Euler-ligning restriktionen (13) til at indse at udtrykket $E_t[\cdots]$ i den anden linie af (16) er lig nul.

7 GMM estimation "i praksis"

Medmindre vi ser på et specialtilfælde af GMM såsom OLS, vil det kræve en numerisk optimeringsmetode for at minimere kriteriefunktionen (4). Dette aspekt af GMM estimation ligger dog uden for rammerne af denne note, men i en snæver vending kan man måske være heldig at Excel's `solver` kan klare opgaven. De fleste moderne statistiske programpakker vil indeholde en eller anden form for funktionalitet til GMM estimation (som bør være beskrevet i manualen eller on-line dokumentationen).

Literatur

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. & K. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1288.

Newey, W. and K.D. West (1987), "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.