WEEK 12 NOTES: GMM

NBER
NATIONAL BUREAU OF ECONOMIC RESEARCH

GMM: the general case

L o.c.'s $\quad E\left[\underset{k\times 1}{m_i(\theta)}\right] = 0 \quad \forall$ obs i $\quad$ (population)

Define $\bar{m}_n(\theta) \equiv \frac{1}{n}\sum_i m_i(\theta)$. $\quad$ Then $E\left[\bar{m}_n(\theta)\right] = 0$ $\quad$ (sample o.c.'s)

Ex. instrumental variables $y_i = \underset{(k\times 1)}{x_i'\beta} + \varepsilon_i$ $\quad$ w/ instruments $\underset{(L\times 1)}{z_i}$

$$E\left[\bar{m}_n(\beta)\right] = E\left[\frac{1}{n}\sum_i z_i(y_i - x_i'\beta)\right] = \underline{0}$$

Assump: (GMM1) Empirical moments converge $\quad \bar{m}_n(\theta) \xrightarrow{P} \underline{0}$

sufficient to assume that $m(\theta)$ jointly stationary and ergodic

and derivative of moments converge $\quad \bar{G}_n(\theta) = \frac{\partial \bar{m}_n(\theta)}{\partial \theta'} \xrightarrow{P} \bar{G}(\theta)$

sufficient to assume cts. and cts. differentiable w.r.t. θ

(GMM2) Identification $\quad \theta_1 \neq \theta_2 \Rightarrow \exists$ datasets s.t. $\bar{m}_n(\theta_1) \neq \bar{m}_n(\theta_2)$

idea: need 1 minimum for the criterion function

this definition implies: $\quad L \geq K$

$\underset{L\times K}{\bar{G}_n(\theta)}$ has row rank of K

(GMM3) Empirical moments CLT $\quad \sqrt{n}\,\bar{m}_n(\theta) \xrightarrow{d} N(0, \Phi)$

Either iid or something of our usual lg-sample form

$$E\left[m_i(\theta) \mid m_{i-1}(\theta), m_{i-2}(\theta), \ldots\right] = 0$$

or in IV case: $\quad E\left[z_i\varepsilon_i \mid z_{i-1}\varepsilon_{i-1}, z_{i-2}\varepsilon_{i-2}, \ldots\right] = 0$

Then $\hat{\theta}_{GMM} = \underset{\theta}{\arg\min}\ \bar{m}(\theta)' W \bar{m}(\theta)$ $\quad$ for some fixed W (not depen. on θ)

Thm $\quad \hat{\theta}_{GMM} \xrightarrow{P} \theta \qquad \hat{\theta}_{GMM} \overset{a}{\sim} N\left(\theta, \frac{1}{n}\left[\bar{G}'W\bar{G}\right]^{-1}\bar{G}'W\Phi W\bar{G}\left[\bar{G}'W\bar{G}\right]^{-1}\right)$

where $\sqrt{n}\,\bar{m} \xrightarrow{d} N(0, \Phi) \qquad \Phi = \sum_{j=-\infty}^{\infty} E(m_i m_{i-j}')$

(aka spectral density matrix

NBER
NATIONAL BUREAU OF ECONOMIC RESEARCH

Just as in the GLS/OLS case, this simplifies if we choose the right W, here $W = \bar{\Phi}^{-1}$. (optimal)

If so, then $\hat{\theta}^{opt}_{GMM} \overset{a}{\sim} N\left(\theta, \frac{1}{n}\left[\bar{G}'\bar{\Phi}^{-1}\bar{G}\right]^{-1}\right)$ w/usual GLS vs. OLS tradeoffs

If we're using GMM to do OLS, then $W = I$ for simplicity, and vcv reduces to

$$\hat{\theta}_{GMM} = \hat{\theta}_{OLS} \sim N\left(\theta, \frac{1}{n}(\bar{G}'\bar{G})^{-1}\bar{G}'\bar{\Phi}\bar{G}(\bar{G}'\bar{G})^{-1}\right)$$

which essentially reproduces White or Newey-West, since in this case

$$\bar{m}_n(\beta) = \frac{1}{n}\sum_i x_i(y_i - x_i'\beta)$$

$$\bar{G}(\beta) = \frac{\partial \bar{m}(\beta)}{\partial \beta'} = -\frac{1}{n}\sum_{i=1}^{n} x_i x_i' = -\frac{1}{n}X'X$$

(all we need is a p.def. estimate of $\bar{\Phi}$). — getting $\hat{\bar{\Phi}}$:
① calc $\hat{\theta}$ using $W = I$ or something pre-spec'd
② calc vcv of moments (might need some structure)
③ use $W = \hat{\bar{\Phi}}^{-1}$ to get $\hat{\theta}$

Testing the validity of the o.c.'s.

if $L = K$, it's exactly identified, so we can't test anything

if $L \geq K$ and the model isn't right, some o.c's will be violated

Consider a Wald test that all the moments are zero, specifically:

$$(\sqrt{n}\,\bar{m}(\hat{\theta})')\underbrace{\left[avar(\sqrt{n}\,\bar{m}(\hat{\theta}))\right]^{-1}}_{W^{-1}\,!}(\sqrt{n}\,\bar{m}(\hat{\theta}))$$

So $n\,\bar{m}(\hat{\theta})'W\bar{m}(\hat{\theta}) \overset{d}{\to} \chi^2_{L-K}$ sometimes called a J test

$J \equiv \bar{m}(\hat{\theta})'W\bar{m}(\hat{\theta})$

You can also do an LR equivalent of this:

$$TJ_{RESTRICTED} - TJ_{UNRESTRICTED} \sim \chi^2_{\#\text{ of restrictions}}$$

Week 12 An intro to GMM in asset pricing

Recall the basic SDF result. If $\nexists$ arbitrage, the $\exists$ positive r.v m_t s.t

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

$\forall$ assets w/non-zero price, or more generally

$$E_t(m_{t+1} x_{t+1}) = p_t$$, where x_{t+1} is the payoff next period

Basic empirical strategy:

postulate a model for m: $m(\text{parameters, realizations}_{t+1})$

check sample moments $\frac{1}{T}\sum_{i=1}^{T} m_t(\text{parms, data}_{t+1}) R_{t+1}$

and see whether they're close to 1.

All of GMM is designed to help us do statistical tests of H_0: sample moments

let's do this in the context of a specific model, and why not start with the original? Hansen & Singleton (1982)

Basic consumption asset-pricing model result

From Wk. 1

investor consumes over 2 dates

can trade in a risky asset w/ price p_t and payoff x_{t+1}

has expected utility of consumption that's separable

Then the investor's FOC implies $p_t = E_t\left(\underbrace{\beta}_{\text{discount rate in utility fn.}} \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1}\right)$

and if all investors have similar utility fns, $\exists$ a representative investor, and prices will satisfy their FOC for that investor

Assume power utility: $u(c_t) = \dfrac{c_t^{1-\gamma}}{1-\gamma} \Rightarrow u'(c_t) = c_t^{-\gamma}$

Then the FOC becomes

$$p_t = E_t\Big(\underbrace{\beta \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}}_{m_{t+1}} x_{t+1}\Big) \quad \text{or} \quad 1 = E_t\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1}\right)$$

2 parameters + consumption growth

Define the pricing errors $u_{t+1} = m_{t+1} R_{t+1} - 1$

or as $u_{t+1} = m_{t+1}(\beta, \gamma)\, x_{t+1} - p_t$

Sample mean of these pricing errors: (over time)

$$g_T(\beta,\gamma) = \frac{1}{T}\sum_{t=1}^{T} u_{t+1}(\beta,\gamma) \equiv E_T\left[u_{t+1}(\beta,\gamma)\right]$$

Hansen's notation
(terrible, but we're stuck with it)

Goal: choose β, γ to min. pricing errors
what metric?

1^{st} stage: $\min \; g_T(\beta,\gamma)' \, W \, g_T(\beta,\gamma)$ for any W
(usually $W = I$, so SSE)
if c-s test, treat all assets the same

The resulting parameter estimates are consistent
but not necessarily efficient (just like OLS).

what are these pricing errors?
essentially actual − expected returns over whole sample
which is just Jensen's alpha

we'll spend most of our time thinking about C-S tests, where
there are lots of different assets, and the pricing errors
should be close to zero for all of them if the model is right

but Hansen & Singleton go a different direction. They consider
instruments z_t that are in the info set at time t,
and they only consider 1 asset at a time in most of the p

1 asset, no instruments: is the H-S model identified?

start with $E_t\left(m_{t+1}(\gamma,\beta) R_{t+1}\right) = 1$

mult by z_t, take uncond expectation:

$$E\left(m_{t+1}(\gamma,\beta) R_{t+1} z_t\right) = E(z_t)$$

or $E\left[\left(m_{t+1}(\gamma,\beta) R_{t+1} - 1\right) z_t\right] = 0$

So here's an orthogonality condition, and we can test whether the sample moment $\frac{1}{T}\sum_{t=1}^{T}(m_{t+1}R_{t+1}-1)z_t = 0$

that's why GMM is sometimes called gen'l'd IV estimation

Adding an instrument: basically checks that ~~returns~~ ex-post discounted returns can't be forecasted using the instrument in a linear regression

Hansen-Singleton: instruments are lagged returns & lagged $\frac{C_{t+1}}{C_t}$

Suppose you use 2 lags of each as instruments (they consider 1, 2, 4, 6 monthly lags)

How many O.C.'s?
How many ~~[illegible]~~ parms?
How many d.f.'s?

$\frac{1}{T}\sum e_t$

"

So here we have a vector of pricing errors $g_T(\beta,\gamma) = \frac{1}{T}\sum \begin{bmatrix} u_{t+1} \\ u_{t+1}z_{1t} \\ \vdots \\ u_{t+1}z_{nt} \end{bmatrix}$

To test whether these are jointly zero, we'll do a χ^2 test:

$$g_T(\beta,\gamma)' \underbrace{\hat{var}^{-1}(g_T(\beta,\gamma))}_{} g_T(\beta,\gamma) \sim \chi^2_{d.f.} \text{ under } H_0$$

This is the challenge, it turns out

Let's go back to our minimization problem to get the parms.

$$\min g_T(\beta,\gamma)' W g_T(\beta,\gamma)$$

Is there an optimal W? Yes. Hansen shows that

$W = \hat{S}^{-1}$, $\hat{S}$ an estimate of $S = \underbrace{\sum_{j=-\infty}^{\infty} E\, u_t u_{t-j}'}_{\text{spectral density at frequency zero}}$, is optimal

This is just like GLS.

If one O.C. is noisy, we'll downweight it... and so on.

2^{nd}-stage estimate $\hat{b}_2 = \begin{pmatrix}\beta \\ \gamma\end{pmatrix}$ mins. $g_T(\beta,\gamma)' \hat{S}^{-1} g_T(\beta,\gamma)$

$\hat{b}_2$ is consistent, asymp. normal, and efficient, just like GLS.

and $var(\hat{b}_2) = \frac{1}{T}(d' S^{-1} d)^{-1}$, where $d = \frac{1}{T}\Sigma\left(\frac{\partial}{\partial b} g(\beta,\gamma)\right)\Big|_{b=\hat{b}} = \frac{\partial g_T(\beta,\gamma)}{\partial b}$

we can use this to test hypotheses about the params

we can also test the model as we ~~did~~ started to above

$$T\, g_T(\beta,\gamma)' S^{-1} g_T(\beta,\gamma) \sim \chi^2(\#\text{ moments} - \#\text{ params})$$

(Using the 2^{nd} stage estimates of β & γ)

Note: it's easy to get the 2^{nd} stage estimate var-cov matrix
it's much hardest to get it for the 1^{st} stage estimate

H-S usually reject the model when they use 1 lag, 1 stock ret
accept more lags, ..
reject 2 stock ret or more (say, EW & VW)

2 stock ret, 1 lag of instruments

how many instruments?
how many O.C.'s?
how many params?
how many d.f.'s?

whole idea: overfit the model & see how it does

one last thing: why is $var(\hat{b}_2) = \frac{1}{T}(d' S^{-1} d)^{-1}$?

Delta method: $avar(f(x)) = f'(x)^2\, avar(x)$

or for vectors $= \frac{\partial f}{\partial x}'\, avar(x)\, \frac{\partial f}{\partial x}$

What have we left undone so far?

1. how to estimate S in practice
2. how to conduct inference if we just use the 1^{st} stage estimator

Remember that $S = \sum_{j=-\infty}^{\infty} E(e_t e_{t-j}')$

1

A

obviously we have to cut this off at some finite j
even if j_{max} is small, sometimes $\hat{S}$ isn't positive definite.

So it's nice to use something guaranteed to be p.d.
There are lots, but perhaps the most popular is Bartlett aka Newey-West

$$\hat{S} = \sum_{j=-k}^{k} \left(\frac{k-|j|}{k}\right) \frac{1}{T} \sum_{t=1}^{T} e_t e_{t-j}'$$

Keep k as ~~too~~ small as possible for best finite-sample behavior

(note the similarity to the variance ratio formula from Wk. 2)

what are we doing?: picking up c-s and autocorr all at once!

lots of APM's imply $k=0$

B

Hansen-Singleton also recommend removing means to calc $\hat{S}$

so $(e_t - \bar{e})(e_t - \bar{e})'$ rather than $e_t e_t'$

same under the null, since $Ee_t = 0$, but demeaned pricing errors are ~~tes~~ usually further away from being singular

C

Use $W = I$ to start, but make sure the O.C.'s have similar variances

For example, if you're going to test whether 10 pfls of stocks fit a low-dim model, $W = I$ is probably fine, since pfls. have similar moments

but if one instrument is, say, divd yld ($\sim 2\%$), and another is, say, turnover ($\sim 100\%$), it might make sense to scale 'em up so they're of similar magnitude

D

Keep the number of assets / O.C.'s small — you're estimating $O(n^2)$ correlations

OLS is GMM

OLS: $\min_{\beta} \; E_T\left[(y_t - \beta' x_t)^2\right]$ ← sample moment notation

F.O.C. $E_T\left[x_t(y_t - \beta' x_t)\right] = \underset{\sim}{0} = g_T(\hat\beta)$ these are the GMM O.C.'s

This is exactly ID'd: # parms = # O.C.
so all sample moments can be set to zero exactly
and the weighting matrix is just $W = I$

Now $var(\hat\beta) = \frac{1}{T}(d' S^{-1} d)^{-1}$

and plugging into our formulas, we get

$$d = \frac{\partial g_T(\hat\beta)}{\partial \beta} = E_T(x_t x_t') \qquad S = \sum_{j=-\infty}^{\infty} E(\varepsilon_t x_t x_{t-j}' \varepsilon_{t-j})$$

~~and $\hat S = \sum_{j=-\infty}^{\infty} E_T(g_T(\hat\beta) g_T(\hat\beta)') = E_T(x_t \varepsilon_t$~~ since $g_T(\hat\beta) = E_T(x_t\varepsilon_t)$

So $$var(\hat\beta) = \frac{1}{T}\left[E_T(x_t x_t')^{-1}\right]\left[\sum_{j=-\infty}^{\infty} E_T(\varepsilon_t x_t x_{t-j}' \varepsilon_{t-j})\right]\left[E_T(x_t x_t')^{-1}\right]$$

If ε_t uncorr'd w/everything else, then only $j=0$ in middle term survives

if conditional $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, then that $j=0$ term reduces to $\sigma_\varepsilon^2 E_T(x_t x_t')$
(homoskedasticity)

and this whole thing reduces to $var(\hat\beta) = \sigma_\varepsilon^2 \left[E_T(x_t x_t')\right]^{-1}$
the usual formula

under general heteroskedasticity
but no autocorrelation... $var(\hat\beta) = \frac{1}{T} E_T(x_t x_t')^{-1} E_T(\varepsilon_t^2 x_t x_t') E_T(x_t x_t')^{-1}$
these are White std. errs.

and for autocorrelation up to lag k, we keep everything, and it turns out that Newey-West std. errors involve a consistent estimate for $\sum_{j=-k}^{k} E(\varepsilon_t x_t' x_{t-j} \varepsilon_{t-j})$

Let's do an example to make this clear
rederive GRS statistic to test if a pfl is M-V efficient

Recall the exercise. For each asset i, excess returns r_{it}

$$r_{it} = \alpha_i + \beta_i f_t + \varepsilon_{it} \qquad \Sigma = E(\varepsilon_t \varepsilon_t')$$

If the model is right $E r_{it} = \beta_i E f_t \Rightarrow \alpha_i = 0 \quad \forall i$

This is a joint test $\forall i$, so we form the GRS test statistic

$$\underbrace{T\left[1+\left(\frac{E_T f}{\hat\sigma(f)}\right)^2\right]^{-1}}_{\text{scalars}} \underset{1\times N}{\hat\alpha'} \underset{N\times N}{\hat\Sigma^{-1}} \underset{N\times 1}{\hat\alpha} \sim \chi^2_N \qquad (N \text{ assets})$$

You might already see how this looks like a GMM test, but here goes more formally

Rewrite in vector form: $\underset{N\times 1}{r_t} = \alpha + \beta f_t + \varepsilon_t$

OLS moments to get estimates $g_T(b) = E_T\begin{bmatrix} \varepsilon_t \\ f_t \varepsilon_t \end{bmatrix} = 0$ 2N o.c.'s, 2N parms

(no overfitting, so no weighting matrix)

$$\text{var}\begin{pmatrix} \hat\alpha \\ \hat\beta \end{pmatrix} = \frac{1}{T}(d' S^{-1} d)^{-1}, \quad \text{so} \quad \hat\alpha' \text{var}(\hat\alpha)^{-1}\hat\alpha \sim \chi^2_N$$

$$d \equiv \frac{\partial g_T(\hat b)}{\partial b'} = \begin{bmatrix} -I_N & -I_N E_T(f_t) \\ -I_N E_T(f_t) & -I_N E_T(f_t^2) \end{bmatrix}$$

$$S = \sum_{j=-\infty}^{\infty} \begin{bmatrix} E\, \varepsilon_t \varepsilon_{t-j}' & E\, \varepsilon_t \varepsilon_{t-j}' f_{t-j} \\ E\, f_t \varepsilon_t \varepsilon_{t-j}' & E\, f_t \varepsilon_t \varepsilon_{t-j}' f_{t-j} \end{bmatrix}$$

If we assume the errors are uncorrelated over time and homoskedastic

$$S = \begin{bmatrix} E\,\varepsilon_t\varepsilon_t' & E\,\varepsilon_t\varepsilon_t'\, E f_t \\ E f_t\, E\,\varepsilon_t\varepsilon_t' & E\,\varepsilon_t\varepsilon_t'\, E f_t^2 \end{bmatrix}$$ and a bunch of algebra $\Rightarrow$

[and note it would be easy to do this if we didn't make the simplifying assumps.

$$\boxed{\text{var}(\hat\alpha) = \frac{1}{T}\left(1 + \frac{(Ef)^2}{\text{var}(f)}\right)\Sigma}$$

(2) How to use 1^{st}-stage estimate (prespecified W)

GLS is more efficient than OLS, but less robust
slight misspecs of Ω may ~~render~~ have big effects on $\hat{\beta}$
estimation of Ω in small samples may do weird things

same with GMM, so there are lots of reasons to use a weighting matrix that's exogenously spec'd

Cochrane 11.5 derives asymptotics from a very general GMM setup in 11.1
We'll skip the details but just give the results:

$$var(\hat{b}) = \frac{1}{T}(d'Wd)^{-1} d'WSWd(d'Wd)^{-1}$$

$$var(g_T) = \frac{1}{T}\left(I - d(d'Wd)^{-1}d'W\right) S \left(I - Wd(d'Wd)^{-1}d'\right)$$

Note that this reduces to our earlier result when $W = S^{-1}$

The only problem is that the matrices we have to invert are singular. So instead of a standard inverse we need to calculate a generalized inverse (which Matlab is good at)

But we can test the model in the expected way:

$$g_T' \, var(g_T)^{-1} g_T \sim \chi^2_{d.f. = \# \cancel{moments}\ oc - \# parms}$$

The most common prespecified $W = I$, i.e. where we treat all the assets identically in a c-s. framework

Hansen and Jagannathan (1997) (not 1991) use a different prespecified W in measuring distances btw $\hat{m}$ and m