ECON 4360: Empirical Finance

Overview and Intro to GMM

Sherry Forbes

University of Virginia

Empirics Lecture #01

What are we doing today?

- Beginning the Empirical Half of the Course
- Introduction to GMM

Some Empirical Questions

- Let's look at

$$
\begin{aligned}
1 &= E_t\left(m_{t+1}R_{t+1}\right) \\
m_{t+1} &= \beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}
\end{aligned}
$$

- What values for β and γ best satisfy our central asset pricing equation?
  - We can use GMM as a statistical criteria to pick these parameters.
- Is the CRRA model a good one for asset prices?
  - Can we reject a null hypothesis that the model is correct?

Some Empirical Questions

- Now think about linear factor models...
- Recall our expected return-beta representation

$$E\left(R^{i}\right)=\gamma+\beta_{i,a}\lambda_{a}+\beta_{i,b}\lambda_{b}+\ldots,\ i=1\ldots N$$

- Are the pricing errors in linear factor models "large"?
- Which factors provide more accurate prices?

Overview: Empirical Half of the Course

- The GMM Framework: Formal Statement of Theory and Simple Examples
- Data Issues: Stationarity
- GMM and Robust Standard Errors (HAC)
- Linear Factor Pricing Models
  - Applications to Pricing Stock Market Portfolios, Term Structure of Interest Rates
- Cross-Sectional Factor Pricing Models
  - Linking Stock Returns to Macro Fundamentals
- Estimating and Testing Explicit Factor Pricing Models
  - E.g., CRRA Utility Functions
- H-J Bounds as Tests of Asset Pricing Models
  - 'Exotic' Utility Functions

Overview of Empirical Framework: GMM

- GMM is a 'moment' based estimator (Partial Information)
  - The alternative is Maximum Likelihood (Full Information)
- The advantage of GMM is that it makes few assumptions about the data
  - The data must be covariance stationary
  - But non-normal distributions, persistence, heteroskedasticity, skewness do not pose problems
- Classic regression theory makes many assumptions about regression residuals. (What are they?)
  - GMM allows us to handle important deviations from these assumptions

Basic Idea of GMM

- Our central asset pricing equation predicts that

$$E\left(p_t\right) = E\left[m_{t+1}\left(\text{data}_{t+1}, \text{parameters}\right) x_{t+1}\right]$$

- How should we check this prediction?
  - Looks at sample averages:

$$\frac{1}{T}\Sigma_{t=1}^{T} p_t \text{ and } \frac{1}{T}\Sigma_{t=1}^{T}\left[m_{t+1}\left(\text{data}_{t+1}, \text{parameters}\right) x_{t+1}\right]$$

- We can then evaluate how well our model performs by looking at how close these sample averages are to each other
  - This is equivalent to examining how large "pricing errors" are

Basic Idea of GMM

- Say we want to evaluate the consumption-based model assuming CRRA utility.
- Before evaluating the model, we have to first pick the parameters β and γ.

$$m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}$$

- But which parameters should we choose for our model?
  - If we think we have a good model, we want to pick parameters that "give it the best chance". So how do we do this?
- We can use GMM to give us estimates of the parameters β and γ that make the sample averages

$$\frac{1}{T}\Sigma_{t=1}^{T} p_t \text{ and } \frac{1}{T}\Sigma_{t=1}^{T} \left[m_{t+1} \left(\text{data}_{t+1}, \text{parameters} \right) x_{t+1} \right]$$

as close to each other as possible.
- Then we can use those parameters to test the model.

GMM in Explicit Discount Factor Models

- First, we're going to talk about how to estimate the unknown parameters of the model.
- Let's continue our use of the consumption-based model to provide some content to the theory of GMM we're going to build up...
  - We have $E(p_t) = E\left[m_{t+1}(\text{data}_{t+1}, \text{parameters})\, x_{t+1}\right]$ with $m_{t+1} = \beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}$.
- The discount factor depends on some unknown parameters as well as the data, so we write $m_{t+1}(b)$, where

$$b := \begin{bmatrix} \beta & \gamma \end{bmatrix}'$$

- Now, b is just a vector of parameters to be estimated.
  - And x and p are also typically vectors.

Building GMM Estimates

- Now, we can write

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

in another form

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

- Equations written in the form $E\left(\cdot\right) = 0$ are easy to work with. Equations like this last equation are the "moment conditions" that we are going to work with.
- This equation should hold in expectation if our model is a good one, so we can think that we might want to minimize the "errors" of this model.
- The errors - from using particular values for b - can be defined as

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

Building GMM Estimates: First-Stage Estimates

- So a first-stage estimate of b solves

$$\widehat{b}_1 = \arg\min_b g_T(b)' W g_T(b)$$

  for some arbitrary matrix W. (E.g., $W = I$).
  - (What is $g_T(b)$?)
- These estimates $\widehat{b}_1$ are:
  - Consistent
  - and Asymptotically Normal.
- (We could just stop here. But we won't...)

Building GMM Estimates: The Weighting Matrix

- Let's think about our weighting matrix, $W = I$...
- What does a weighting matrix do?
  - It directs GMM to emphasize some moments (or linear combinations of moments) at the expense of other moments.
- What does it mean if you start with $W = I$?

Building GMM Estimates: The Weighting Matrix

- Let's think about our weighting matrix, $W = I$...
- What does a weighting matrix do?
  - It directs GMM to emphasize some moments (or linear combinations of moments) at the expense of other moments.
- What does it mean if you start with $W = I$?
  - GMM is trying to price all assets equally well. When would we want to do this? ...

Building GMM Estimates: The Weighting Matrix

- Think about a sample mean $g_T = E_T\left(m_t R_t - 1\right)$. When would you expect it to be an accurate measurement of the population mean $E\left(mR - 1\right)$?

Building GMM Estimates: The Weighting Matrix

- Think about a sample mean $g_T = E_T\left(m_t R_t - 1\right)$. When would you expect it to be an accurate measurement of the population mean $E\left(mR - 1\right)$?
  - Perhaps when the variance of $m_t R_t - 1$ is low.

Building GMM Estimates: The Weighting Matrix

- Think about a sample mean $g_T = E_T\left(m_t R_t - 1\right)$. When would you expect it to be an accurate measurement of the population mean $E\left(mR - 1\right)$?
  - Perhaps when the variance of $m_t R_t - 1$ is low.
- Idea: We want to pay more attention to things we think might be priced more accurately.

Building GMM Estimates: The Weighting Matrix

- Think about a sample mean $g_T = E_T\left(m_t R_t - 1\right)$. When would you expect it to be an accurate measurement of the population mean $E\left(mR - 1\right)$?
  - Perhaps when the variance of $m_t R_t - 1$ is low.
- Idea: We want to pay more attention to things we think might be priced more accurately.
  - This means downweighting things with higher variance.

Building GMM Estimates: The Weighting Matrix

- Think about a weighting matrix that might do this.
  - If we were to replace the 1's in the weighting matrix $W = I$ with $1/\left(var\left[E_T\left(m_t R_t - 1\right)\right]\right)$, that would do it.
  - Think about what would happen if the u_t''s are uncorrelated over time $E_t\left(u_t u_{t-j}'\right) = 0$, then

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

  - This is just a formula for the variance of a sample mean!
- But we actually know more...
  - We know that asset returns are correlated, so if we use a form of the covariance matrix of $\left[E_T\left(m_t R_t - 1\right)\right]$, it will also pay more attention to linear combinations of moments about which the data set has the most information.

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
  - What that means is that the first and second moments of the data have to be finite and independent of time.

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
  - What that means is that the first and second moments of the data have to be finite and independent of time.
- Let Y_t be our time series.

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
  - What that means is that the first and second moments of the data have to be finite and independent of time.
- Let Y_t be our time series.
- Let $\gamma_j = cov\left(Y_t, Y_{t-j}\right)$ and let $\gamma_0 = var\left(Y_t\right)$

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
  - What that means is that the first and second moments of the data have to be finite and independent of time.
- Let Y_t be our time series.
- Let $\gamma_j = cov\left(Y_t, Y_{t-j}\right)$ and let $\gamma_0 = var\left(Y_t\right)$
  - Requirement #1: $E\left(Y_t\right) = \mu$, $\mu < \infty$

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
    - What that means is that the first and second moments of the data have to be finite and independent of time.
- Let Y_t be our time series.
- Let $\gamma_j = cov\left(Y_t, Y_{t-j}\right)$ and let $\gamma_0 = var\left(Y_t\right)$
    - Requirement #1: $E\left(Y_t\right) = \mu,\ \mu < \infty$
    - Requirement #2: $cov\left(Y_t, Y_{t-j}\right) = \gamma_j,\ \gamma_j < \infty\ \forall\ j, t$

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
  - What that means is that the first and second moments of the data have to be finite and independent of time.
- Let Y_t be our time series.
- Let $\gamma_j = cov\left(Y_t, Y_{t-j}\right)$ and let $\gamma_0 = var\left(Y_t\right)$
  - Requirement #1: $E\left(Y_t\right) = \mu$, $\mu < \infty$
  - Requirement #2: $cov\left(Y_t, Y_{t-j}\right) = \gamma_j$, $\gamma_j < \infty \; \forall \; j, t$
    - What that means is that the covariances can depend on *the interval* j, but not on *where you are at* t, e.g., $E\left(u_1 u_2'\right) = E\left(u_t u_{t+1}'\right)$

As a Primer...

- Recall, I mentioned that for GMM, we need the data to be 'weakly stationary' or 'covariance stationary.'
  - What that means is that the first and second moments of the data have to be finite and independent of time.
- Let Y_t be our time series.
- Let $\gamma_j = cov\left(Y_t, Y_{t-j}\right)$ and let $\gamma_0 = var\left(Y_t\right)$
  - Requirement #1: $E\left(Y_t\right) = \mu$, $\mu < \infty$
  - Requirement #2: $cov\left(Y_t, Y_{t-j}\right) = \gamma_j$, $\gamma_j < \infty \; \forall \; j, t$
    - What that means is that the covariances can depend on *the interval* j, but not on *where you are at* t, e.g., $E\left(u_1 u_2'\right) = E\left(u_t u_{t+1}'\right)$
- Note that since $cov\left(Y_t, Y_{t-j}\right) = E\left[\left(Y_t - EY_t\right)\left(Y_{t-j} - EY_{t-j}\right)\right]$, if the time-series has mean zero, this simplifies to $cov\left(Y_t, Y_{t-j}\right) = E\left(Y_t, Y_{t-j}\right)$

Building GMM Estimates: S Matrix

- Now, we can exploit $E\left(u_t\right) = 0$ and covariance stationarity to build a weighting matrix S:

Building GMM Estimates: S Matrix

- Now, we can exploit $E(u_t) = 0$ and covariance stationarity to build a weighting matrix S:
- Now, we first look at

$$
\begin{aligned}
var(g_T) &= var\left(\frac{1}{T}\Sigma_{t=1}^{T} u_{t+1}\right) \\
&= \frac{1}{T^2}\left[TE\left(u_t u_t'\right) + (T-1)\left(E\left(u_t u_{t-1}'\right) + E\left(u_t u_{t+1}'\right)\right) + \ldots\right.
\end{aligned}
$$

Building GMM Estimates: S Matrix

- Now, we can exploit $E\left(u_t\right) = 0$ and covariance stationarity to build a weighting matrix S:
- Now, we first look at

$$
\begin{aligned}
var\left(g_T\right) &= var\left(\frac{1}{T}\Sigma_{t=1}^{T} u_{t+1}\right) \\
&= \frac{1}{T^2}\left[TE\left(u_t u_t'\right) + \left(T-1\right)\left(E\left(u_t u_{t-1}'\right) + E\left(u_t u_{t+1}'\right)\right) + \ldots\right.
\end{aligned}
$$

- Now, at $T \rightarrow \infty$, $\left(T - j\right)/T \rightarrow 1$, so

$$
var\left(g_T\right) \rightarrow \frac{1}{T}\Sigma_{j=-\infty}^{\infty} E\left(u_t u_{t-j}'\right) = \frac{1}{T} S
$$

where

$$
S = \Sigma_{j=-\infty}^{\infty} E\left(u_t u_{t-j}'\right)
$$

Building GMM Estimates: S Matrix

- Now, we can exploit $E\left(u_t\right)=0$ and covariance stationarity to build a weighting matrix S:
- Now, we first look at

$$\begin{aligned} var\left(g_T\right) &= var\left(\frac{1}{T}\Sigma_{t=1}^{T} u_{t+1}\right) \\ &= \frac{1}{T^2}\left[TE\left(u_t u_t'\right)+\left(T-1\right)\left(E\left(u_t u_{t-1}'\right)+E\left(u_t u_{t+1}'\right)\right)+\ldots\right. \end{aligned}$$

- Now, at $T\rightarrow\infty$, $\left(T-j\right)/T\rightarrow 1$, so

$$var\left(g_T\right)\rightarrow\frac{1}{T}\Sigma_{j=-\infty}^{\infty}E\left(u_t u_{t-j}'\right)=\frac{1}{T}S$$

  where

$$S=\Sigma_{j=-\infty}^{\infty}E\left(u_t u_{t-j}'\right)$$

- Hansen (1982) shows that $W=S^{-1}$ is the statistically optimal weighting matrix. What does that mean?

Building GMM Estimates: S Matrix

- Now, we can exploit $E\left(u_t\right)=0$ and covariance stationarity to build a weighting matrix S:
- Now, we first look at

$$
\begin{aligned}
var\left(g_T\right) &= var\left(\frac{1}{T}\Sigma_{t=1}^{T}u_{t+1}\right)\\
&= \frac{1}{T^2}\left[TE\left(u_t u_t'\right)+\left(T-1\right)\left(E\left(u_t u_{t-1}'\right)+E\left(u_t u_{t+1}'\right)\right)+\right.
\end{aligned}
$$

- Now, at $T\rightarrow\infty$, $\left(T-j\right)/T\rightarrow 1$, so

$$
var\left(g_T\right)\rightarrow\frac{1}{T}\Sigma_{j=-\infty}^{\infty}E\left(u_t u_{t-j}'\right)=\frac{1}{T}S
$$

  where

$$
S=\Sigma_{j=-\infty}^{\infty}E\left(u_t u_{t-j}'\right)
$$

- Hansen (1982) shows that $W=S^{-1}$ is the statistically optimal weighting matrix. What does that mean?
  - It produces estimates with the lowest asymptotic variance.

Building GMM Estimates: Second-Stage Estimates

- Now, from our first-stage estimate of b from

$$\widehat{b}_1 = \arg\min_b g_T(b)' W g_T(b)$$

- We can use $\widehat{b}_1$ to form an estimate of $\widehat{S}$

$$\widehat{S} = \Sigma_{j=-\infty}^{\infty} E\left[u_t\left(\widehat{b}_1\right) u_{t-j}\left(\widehat{b}_1\right)'\right]$$

- Next, we can form second-stage estimates according to

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

Building GMM Estimates

- Estimates for $\widehat{b}_2$ are
  - Consistent
  - Asymptotically Normal
  - and, now, also Asymptotically Efficient
    - Where efficient means that it has the smallest variance-covariance matrix among all estimators that set linear combinations of $g_T(b)$ equal to zero or all choices of weighting matrices W.

First- and Second-Stage Estimates

- The estimates we have done should remind you of standard linear regression models.
- The first-stage estimates are *like* OLS.
  - For OLS, if the errors are not i.i.d., OLS estimates are consistent, but not efficient.
- To get efficient estimates, we can use the OLS estimates to construct a series of residuals to estimate a variance-covariance matrix of the residuals to then use for GLS.
  - GLS is also consistent, but more efficient (meaning the sampling variation in the estimated parameters is lower).

Does the Weighting Matrix Matter?

- It Depends.

Two Cases of GMM

- Case 1: We have the same number of moment conditions as parameters.
  - The parameters are exactly identified.
  - We can set all of the moment conditions equal to zero (exactly).
  - The weighting matrix:
    - is irrelevant for solving the minimization problem.
    - is needed for inference
    - can be constructed after solving for the parameters.

Two Cases of GMM

- Case 2: We have more moment conditions than parameters.
  - The parameters are over-identified.
  - We cannot set all of the moment conditions equal to zero.
  - The weighting matrix is key to solving the minimization problem.
  - The weighting matrix is needed for estimating the parameters, yet it depends on the parameters
  - So we solve the minimization problem numerically on a computer:
    - Start with any weighting matrix, e.g., $W = I$ to find $\widehat{b}_1$
    - Use $\widehat{b}_1$ to construct $\widehat{S}_1$
    - Find $\widehat{b}_2$ using $\widehat{S}_1$
    - Continue until $\widehat{b}_{i+1} \approx \widehat{b}_i$

GMM Estimator: Formal Statement

- Let $\overline{Y}_T$ be a matrix of data with T time-series observations, and let b be a vector of parameters to be estimated.
- Let $f\left(Y_t, b\right)$ denote the moment condition that relates the data and parameters.
  - And let $g_T\left(\overline{Y}_T, b\right) = (1/T)\Sigma_{t=1}^{\infty} f\left(Y_t, b\right)$ denote the sample average of $f\left(Y_t, b\right)$
- The GMM estimate for b solves the following minimization problem

$$\widehat{b}_{GMM} = \arg\min_b g_T\left(\overline{Y}_{T,} b\right)' \widehat{S}^{-1} g_T\left(\overline{Y}_{T,} b\right)$$

  - where $\widehat{S}$ is a weighting matrix defined as

$$\widehat{S} = \Sigma_{j=-\infty}^{\infty} E\left[f\left(Y_{t,b}\right) f\left(Y_{t-j}, b\right)'\right]$$

- And $\widehat{b}_{GMM} \overset{a}{\sim} N\left[b, \frac{1}{T}\left(d\widehat{S}^{-1}d'\right)^{-1}\right]$, where d is just the derivative of the moment condition w.r.t. b, $d = \frac{\partial g\left(\overline{Y}_{T,} b\right)}{\partial b}$

The Standard Errors

- What is d?
  - Recall that we're trying to find estimates of b, and we know that the GMM estimates are distributed asymptotically normal.
- Recall, the Delta Method:
  - It's easy to see in the univariate case. Basically, if we have

$$\sqrt{n}\left[X_n - \theta\right] \rightarrow N\left(0, \sigma^2\right)$$

then

$$\sqrt{n}\left[h\left(X_n\right) - h\left(\theta\right)\right] \rightarrow N\left(0, \left[h'\left(\theta\right)\right]^2 \sigma^2\right)$$

- So think of $var\left(\widehat{b}_2\right) = \frac{1}{T}\left(d\widehat{S}^{-1}d'\right)^{-1}$ as just an application of the delta method, where

$$\begin{aligned} d &= \frac{\partial g\left(\overline{Y}_{T}, b\right)}{\partial b} \\ &= E_T\left[\frac{\partial}{\partial b}\left(m_{t+1}\left(b\right)x_{t+1} - p_t\right)\right]\Big|_{b=\widehat{b}} \end{aligned}$$

Testing

- We now have all the pieces we need to test if a parameter or group of parameters is equal to zero.
- Since we have the asymptotic distribution,

$$\widehat{b}_{GMM} \overset{a}{\sim} N\left[b, \frac{1}{T}\left(d\widehat{S}^{-1}d'\right)^{-1}\right]$$

- We just use, for an individual parameter,

$$\frac{\widehat{b}_i}{\sqrt{var\left(\widehat{b}\right)_{ii}}} \sim N\left(0,1\right)$$

or, for a group,

$$\widehat{b}_j\left[var\left(\widehat{b}\right)_{jj}\right]^{-1}\widehat{b}_j \sim \chi^2\left(\dim\left(\widehat{b}_j\right)\right)$$

where b_j is a subvector of b, and $var\left(b\right)_{jj}$ is a submatrix of the variance matrix $\frac{1}{T}\left(d\widehat{S}^{-1}d'\right)^{-1}$

What Else Can We Do? The J Test

- Now, we've used GMM to estimate parameters to make the model fit the best it possibly can. But how well does the *model* fit?
  - We're going to now look at the pricing errors and see if they are "large"
- J_T Test: If the model is true, how often should we see a weighted sum of squared pricing errors as big as what we got?
  - If the answer is "not too often", then the model is rejected.
- The J_T test is also called a *test of overidentifying restrictions*

$$TJ_T = T\left[g_T\left(\widehat{b}_{GMM}\right)' S^{-1} g_T\left(\widehat{b}_{GMM}\right)\right] \sim \chi^2\,(\#\text{ mom's - }\#\text{ par's})$$

and recall that S is the variance-covariance matrix for g_T, where this statistic is the minimized pricing errors divided by their variance-covariance matrix.

Interpreting GMM

- So what have we done?
- We've constructed $g_T(b)$ and interpreted it as a pricing error.
- We've used GMM to pick parameters that minimize a weighted sum of squared pricing errors.
  - First and second stage estimates of the parameters are like OLS and GLS regressions - the second stage estimates pick the linear combinations of pricing errors that are 'best measured', interpreted as having the smallest variation in the sample.
- We've constructed the asymptotic distribution of the parameters through an application of the delta method for use in testing parameters.
- We've developed the J_T test as a test of the overall model.

End of Today's Lecture.

- That's all for today. Today's material corresponds roughly to 10 in Cochrane (2005).