

Lecture 10

GMM

R. Susmel, 2023 (for private use, not to be posted/shared online).

1

Method of Moments (MM): Review

- Idea: Population moment conditions provide information which can be used to estimate population parameters.
- Suppose we want to estimate the population mean μ variance σ^2 of a random variable v_t . These parameters satisfy the population moment conditions:

$$\begin{aligned}E[v_t] - \mu &= 0 \\E[v_t^2] - (\sigma^2 + \mu^2) &= 0\end{aligned}$$

- We move from population conditions to their analogous sample moment conditions:

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^T v_t - \mu^* &= 0 & \Rightarrow \mu^* &= \frac{1}{T} \sum_{t=1}^T v_t \\ \frac{1}{T} \sum_{t=1}^T v_t^2 - (\sigma^{*2} + \mu^{*2}) &= 0 & \Rightarrow \sigma^{*2} &= \frac{1}{T} \sum_{t=1}^T (v_t - \mu^*)^2\end{aligned}$$

Method of Moments: Review

- **Example:** A supply and demand system for wheat:

$$q_t^D = \alpha p_t + u_t^D$$

$$q_t^S = \beta_1 n_t + \beta_2 p_t + u_t^S$$

$$q_t^D = q_t^S = q_t$$

where q_t^D, q_t^S are quantity demanded and supplied; p_t is price, n_t is a weather variable. We want to estimate α .

Problem: Endogeneity, OLS –i.e., regress q_t^D against p_t – will not work

Solution: IV. Find z_t^D such that $\text{cov}(z_t^D, u_t^D) = 0$.

Then,

$$\text{cov}(z_t^D, q_t) - \alpha \text{cov}(z_t^D, p_t) = 0$$

Method of Moments: Review

Example (continuation): $\text{cov}(z_t^D, q_t) - \alpha \text{cov}(z_t^D, p_t) = 0$

Then, if $E[u_t^D] = 0$, we have:

$$E[z_t^D q_t] - \alpha E[z_t^D p_t] = 0 \quad (\text{population condition})$$

The MM leads to

$$\hat{\alpha} = \frac{1}{T} \sum_{i=1}^T z_t^D q_t / \frac{1}{T} \sum_{i=1}^T z_t^D p_t \quad (\text{IV estimator})$$

Method of Moments: Review

- Population moment condition: A vector of observed variables, \mathbf{v}_t , and vector of k parameters $\boldsymbol{\theta}$, satisfy a $k \times 1$ element vector of conditions $E[f(\mathbf{v}_t, \boldsymbol{\theta})] = \mathbf{0}$ for all t .

- The MM estimator $\boldsymbol{\theta}_T^*$ solves the analogous sample moment conditions

$$\mathbf{g}_T(\boldsymbol{\theta}^*) = \frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) = \mathbf{0} \quad (1)$$

where T is the sample size.

- Under the usual regularity conditions, $\boldsymbol{\theta}_T^* \xrightarrow{p} \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is the solution of $E[f(\mathbf{v}_t, \boldsymbol{\theta})] = \mathbf{0}$.

Note: We have k unknowns and k equations \Rightarrow unique solution.

Generalized Method of Moments (GMM)

- Now, suppose f is a $q \times 1$ vector and $q > k$. That is, we have k unknowns and q equations \Rightarrow not a unique solution.

- GMM picks a value for $\boldsymbol{\theta}$ such that it comes closest to satisfy

$$\mathbf{g}_T(\boldsymbol{\theta}^*) = \frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) = \mathbf{0} \quad (q \times 1 \text{ equations})$$

- We define “closeness” by

$$\begin{aligned} Q_T(\boldsymbol{\theta}) &= \left[\frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) \right]' \mathbf{W}_T \left[\frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) \right] \\ &= \mathbf{g}_T(\boldsymbol{\theta}^*)' \mathbf{W}_T \mathbf{g}_T(\boldsymbol{\theta}^*) \end{aligned}$$

where \mathbf{W}_T is $(q \times q)$ psd matrix & $\text{plim}(\mathbf{W}_T) = \mathbf{W}$ pd

GMM: Example 1

- Power utility based asset pricing model –Hansen & Singleton (1982)

- Theory condition:

$$E_t[\{\beta(c_{t+1}/c_t)^\gamma (1 + R_{i,t+1}) - 1\}] = 0 \quad \text{with unknown parameters } \beta, \gamma$$

- The q population unconditional moment conditions are

$$E_t[Z_{j,t} \{\beta(c_{t+1}/c_t)^\gamma (1 + R_{i,t+1}) - 1\}] = 0 \quad j=1, \dots, q$$

where $Z_{j,t}$ are instruments in the information set

- The q sample moment conditions are

$$\frac{1}{T} \sum_{t=1}^T Z_{j,t} \{\beta_T^* (c_{t+1}/c_t)^{\gamma_T^*} (1 + R_{i,t+1}) - 1\} = 0.$$

GMM: Example 2

- The CAPM

- Theory condition:

$$E[r_{i,t+1} - \lambda_0(1 - \beta_i) - \beta_i r_{m,t+1}] = 0$$

- The q population moment conditions (Market efficiency):

$$E[(r_{i,t+1} - \lambda_0(1 - \beta_i) - \beta_i r_{m,t+1}) Z_{j,t}] = 0 \quad j=1, \dots, q$$

- The q sample moment conditions:

$$\frac{1}{T} \sum_{t=1}^T \{r_{i,t+1} - \lambda_{0,T}^* (1 - \beta_i) - \beta_{i,T}^* r_{m,t+1}\} Z_{j,t} = 0.$$

GMM: Example 3 - MLE

• Suppose the conditional pdf of the continuous stationary random vector \mathbf{v}_t , given $V_{t-1} = \{\mathbf{v}_{t-1}, \mathbf{v}_{t-2}, \dots\}$ is $p(\mathbf{v}_t; \boldsymbol{\theta}_0, V_{t-1})$

• The MLE of $\boldsymbol{\theta}_0$ based on the conditional log likelihood function is the value of which maximizes $L_T(\boldsymbol{\theta}) = \sum_{t=1}^n \ln\{p(\mathbf{v}_t; \boldsymbol{\theta}, V_{t-1})\}$.

$$\Rightarrow \text{solving } \frac{\delta L_T(\boldsymbol{\theta})}{\delta \boldsymbol{\theta}} = 0$$

• That is, the MLE is just the GMM estimator based on the population moment condition

$$E\left[\frac{\delta \ln\{p(\mathbf{v}_t; \boldsymbol{\theta}, V_{t-1})\}}{\delta \boldsymbol{\theta}}\right] = 0$$

GMM: Summary

• The GMM estimator $\boldsymbol{\theta}_T^* = \arg\min_{\boldsymbol{\theta} \in \Theta} Q_T(\boldsymbol{\theta})$ generates the f.o.c.

$$\left[\frac{1}{T} \sum_{t=1}^T \frac{\delta f(\mathbf{v}_t, \boldsymbol{\theta}_T^*)}{\delta \boldsymbol{\theta}'}\right]' \mathbf{W}_T \left[\frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*)\right] = \mathbf{0} \quad (**)$$

where $\frac{\delta f(\mathbf{v}_t, \boldsymbol{\theta}_T^*)}{\delta \boldsymbol{\theta}'}$ is a $q \times k$ matrix with i, j element $\frac{\delta f_i(\mathbf{v}_t, \boldsymbol{\theta}_T^*)}{\delta \theta_j}$

• There is typically no closed form solution for $\boldsymbol{\theta}_T^*$ so it must be obtained through numerical optimization methods.

Note: From (**), the GMM estimator is the MM estimator based on population moments:

$$\{E\left[\frac{\delta f(\mathbf{v}_t, \boldsymbol{\theta}_0)}{\delta \boldsymbol{\theta}'}\right]\}' \mathbf{W} \{E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)]\} = \mathbf{0} \quad (k \times 1 \text{ conditions})$$

Example: IV estimation of linear model

- Linear IV framework:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}_0 + \boldsymbol{\varepsilon}, \quad \text{with } E[\mathbf{X}'\boldsymbol{\varepsilon}] \neq \mathbf{0}.$$

- Let \mathbf{Z} be a $T \times q$ vector of IV –i.e., $E[\mathbf{Z}'\boldsymbol{\varepsilon}] = \mathbf{0}$ and $E[\mathbf{Z}'\mathbf{X}] \neq \mathbf{0}$.

- We want to estimate $\boldsymbol{\theta}_0$ using GMM. Then, the GMM estimator

$$\boldsymbol{\theta}_T^* = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \{Q_T(\boldsymbol{\theta}) = [\boldsymbol{\varepsilon}(\boldsymbol{\theta})\mathbf{Z}'/T] \mathbf{W}_T [\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta})/T]\}$$

($k \times 1$) f.o.c.:

$$(\mathbf{X}'\mathbf{Z}/T) \mathbf{W}_T (\mathbf{Z}' \boldsymbol{\varepsilon}(\boldsymbol{\theta}_T^*)/T) = 0$$

$$(\mathbf{X}'\mathbf{Z}/T) \mathbf{W}_T (\mathbf{Z}' (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_T^*)/T) = 0$$

or

$$\Rightarrow (\mathbf{X}'\mathbf{Z}/T) \mathbf{W}_T (\mathbf{Z}'\mathbf{y}/T) = (\mathbf{X}'\mathbf{Z}/T) \mathbf{W}_T (\mathbf{Z}'\mathbf{X}/T) \boldsymbol{\theta}_T^*$$

Example: IV estimation of linear model

- From f.o.c.: $(\mathbf{X}'\mathbf{Z}/T) \mathbf{W}_T (\mathbf{Z}'\mathbf{y}/T) = (\mathbf{X}'\mathbf{Z}/T) \mathbf{W}_T (\mathbf{Z}'\mathbf{X}/T) \boldsymbol{\theta}_T^*$

CASE 1: $q = k$ –i.e., *just-identified*– and $(\mathbf{Z}'\mathbf{X}/T)$ is nonsingular then

$$\boldsymbol{\theta}_T^* = (\mathbf{Z}'\mathbf{X}/T)^{-1} (\mathbf{Z}'\mathbf{y}/T)$$

independently of the weighting matrix \mathbf{W}_T .

CASE 2: $q > k$ –i.e., *over-identified*.

$$\boldsymbol{\theta}_T^* = \left\{ \left(\frac{\mathbf{X}'\mathbf{Z}}{T} \right) \mathbf{W}_T \left(\frac{\mathbf{Z}'\mathbf{X}}{T} \right) \right\}^{-1} \left(\frac{\mathbf{X}'\mathbf{Z}}{T} \right) \mathbf{W}_T \left(\frac{\mathbf{Z}'\mathbf{y}}{T} \right)$$

Note: GMM = MM based on ($k \times 1$) population moment conditions

$$E[\mathbf{X}'\mathbf{Z}] \mathbf{W} E[\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$$

(1) When $q = k$ GMM = MM based on $E[\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$.

(2) When $q > k$ GMM sets k linear combinations of $E[\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$.

But, in order to estimate $\boldsymbol{\theta}_0$, we only need k conditions!

Identifying and overidentifying restrictions

- Recall (**). The GMM estimator is the MM estimator based on $(k \times 1)$ population moments

$$\{E[\frac{\delta f(\mathbf{v}_t, \boldsymbol{\theta}_0)}{\delta \boldsymbol{\theta}}]\}' \mathbf{W} \{E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)]\} = 0 \quad (***)$$

- Let $\mathbf{F}(\boldsymbol{\theta}_0) = \mathbf{W}^{1/2} E[\frac{\delta f(\mathbf{v}_t, \boldsymbol{\theta}_0)}{\delta \boldsymbol{\theta}}]$ (a $q \times k$ matrix), with $\text{rank}(\mathbf{F}(\boldsymbol{\theta}_0)) = k$. (The *rank condition* is necessary for identification of $\boldsymbol{\theta}_0$).

- Rewrite (***) as

$$\mathbf{F}(\boldsymbol{\theta}_0) \mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)] = \mathbf{0} \quad \text{or} \quad (k \text{ equations})$$

$$\mathbf{F}(\boldsymbol{\theta}_0)[\mathbf{F}(\boldsymbol{\theta}_0)' \mathbf{F}(\boldsymbol{\theta}_0)]^{-1} \mathbf{F}(\boldsymbol{\theta}_0)' \mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)] = \mathbf{P}_F' \mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$$

- The LS projection of $\mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)]$ on to the column space of $\mathbf{F}(\boldsymbol{\theta}_0)$ is $\mathbf{0}$.

Identifying and overidentifying restrictions

- That is, the GMM estimator is based on $\text{rank}\{\mathbf{P}_F\} = k$ restrictions on the $(q \times 1)$ (*transformed*) population moment condition

$$\mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)].$$

- These are the *identifying restrictions*; GMM picks $\boldsymbol{\theta}_T^*$ to satisfy them.

- The restrictions that are left over are

$$\{\mathbf{I}_q - \mathbf{P}_F\}' \mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$$

- That is, the projection of $\mathbf{W}^{1/2} E[f(\mathbf{v}_t, \boldsymbol{\theta}_0)]$ on to the orthogonal complement of $\mathbf{F}(\boldsymbol{\theta}_0)$ is zero, generating $q - k$ restrictions on the transformed population moment condition.

- These *over-identifying restrictions* are ignored by the GMM estimator!

Identifying and overidentifying restrictions

- The **over-identifying restrictions** are ignored by the GMM estimator, so they need not be satisfied in the sample.

- From (**),

$$\mathbf{W}_T^{1/2} \left[\frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) \right] = \{\mathbf{I}_q - \mathbf{P}_F\}' \mathbf{W}_T^{1/2} \left[\frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) \right]$$

Thus, $Q_T(\boldsymbol{\theta}_T^*)$ is like a sum of squared residuals, and can be interpreted as a measure of how far the sample is from satisfying the over-identifying restrictions.

Asymptotic properties of GMM IVE

- Under the usual regularity conditions, it can be shown that:

$$(1) \boldsymbol{\theta}_T^* \xrightarrow{p} \boldsymbol{\theta}_0$$

$$(2) (\boldsymbol{\theta}_{T,i}^* - \boldsymbol{\theta}_{0,i}) / [\sqrt{\mathbf{V}_{T,ii}^*}/T] \xrightarrow{d} N(0, 1)$$

where $\mathbf{V}_T^* = (\mathbf{X}'\mathbf{Z}\mathbf{W}_T\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_T\mathbf{S}_T^*\mathbf{W}_T\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{Z}\mathbf{W}_T\mathbf{Z}'\mathbf{X})^{-1}$

$$\mathbf{S}_T^* = \lim_{T \rightarrow \infty} \text{Var}[\mathbf{Z}'\boldsymbol{\varepsilon}/T]$$

- Assuming $\boldsymbol{\varepsilon}$ is serially uncorrelated

$$\begin{aligned} \text{Var}[\mathbf{Z}'\boldsymbol{\varepsilon}/T] &= E\left[\left\{\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{z}_t \varepsilon_t\right\} \left\{\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{z}_t \varepsilon_t\right\}'\right] \\ &= \frac{1}{T} \sum_{t=1}^T E[\varepsilon_t^2 \mathbf{z}_t \mathbf{z}_t'] \end{aligned}$$

- Thus,

$$\mathbf{S}_T^* = \frac{1}{T} \left[\sum_{t=1}^T \boldsymbol{\varepsilon}(\boldsymbol{\theta}_T^*)_t^2 \mathbf{z}_t \mathbf{z}_t' \right]$$

GMM: Two-step estimator

- The asymptotic variance depends on the weighting matrix \mathbf{W}_T .

The optimal choice is $\mathbf{W}_T = \mathbf{S}_T^{*-1}$ to give $\mathbf{V}_T^* = (\mathbf{X}'\mathbf{Z} \mathbf{S}_T^{*-1} \mathbf{Z}'\mathbf{X})^{-1}$

- But we need $\boldsymbol{\theta}_T^*$ to construct \mathbf{S}_T^* . This suggests a two-step (iterative) GMM procedure:

- (1) Start with sub-optimal $\mathbf{W}_T(0)$, say \mathbf{I}
- (2) Using $\mathbf{W}_T(0)$ estimate $\boldsymbol{\theta}_T^*(1)$ & $\mathbf{S}_T^*(1)$
- (3) Estimate with $\mathbf{W}_T(1) = \mathbf{S}_T^*(1)^{-1}$
- (4) Using $\mathbf{W}_T(j)$ repeat steps (2)-(3) to get $\boldsymbol{\theta}_T^*(j+1)$ & $\mathbf{S}_T^*(j+1)$ until convergence.

GMM: Two-step estimator

- Note that if \mathbf{e} is homoskedastic:

$$\text{Var}[\boldsymbol{\varepsilon} | \mathbf{X}] = \sigma^2 \mathbf{I}_T \quad \Rightarrow \quad \mathbf{S}_T^* = s^{*2} \mathbf{Z}\mathbf{Z}'$$

where s^{*2} is a consistent estimator of σ^2 .

- Choosing this \mathbf{S}_T^* to construct the weighting matrix $\mathbf{W}_T = \mathbf{S}_T^{*-1}$. Then,

$$\begin{aligned} \boldsymbol{\theta}_T^* &= (\mathbf{X}'\mathbf{Z} (s^{*2} \mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} (s^{*2} \mathbf{Z}\mathbf{Z}')^{-1} \mathbf{Z}'\mathbf{y} \\ &= \{\widehat{\mathbf{X}}'\widehat{\mathbf{X}}\}^{-1} \widehat{\mathbf{X}}'\mathbf{y} \end{aligned}$$

where $\widehat{\mathbf{X}} = \mathbf{P}_Z \mathbf{X}$ is the predicted value of \mathbf{X} from a regression of \mathbf{X} on \mathbf{Z} . This is the 2SLS estimator.

Model specification test

- Identifying restrictions are satisfied in the sample regardless of whether the model is correct.
- Over-identifying restrictions are not imposed in the sample. We can use them to test the model.

Recall the $q \times 1$ population moment conditions $E[\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$. We can construct a Wald-type test to check if these q conditions are met in sample. The **overidentifying restrictions test**:

$$J_T = T Q_T(\boldsymbol{\theta}_T^*) = T^{-1/2} \boldsymbol{\varepsilon}(\boldsymbol{\theta}_T^*)' \mathbf{Z} \mathbf{S}_T^{*-1} T^{-1/2} \mathbf{Z}' \boldsymbol{\varepsilon}(\boldsymbol{\theta}_T^*)$$

Under $H_0: E[\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$, $J_T \xrightarrow{d} \chi_{q-k}^2$

Testing C-CAPM: GMM

- GMM can naturally be applied in the C-CAPM. The Euler's equation, gives us a starting point for a moment condition:

$$0 = E_t[m(\mathbf{x}_{t+1}, \boldsymbol{\theta}_0) (1 + R_{i,t+1}) - 1]$$

- Let \mathbf{Z}_t be a set of l ($l \geq k$) instruments, available at time t . Then, for each asset i :

$$E_t[\mathbf{Z}_{j,t} \{\beta(c_{t+1}/c_t)^\gamma (1 + R_{i,t+1}) - 1\}] = 0 \quad i=1, \dots, N; j=1, \dots, l.$$

Note: Now we have a lot of moments: $l \times N$!

- GMM works with sample analogues of the population moments:

$$g(\mathbf{v}_t, \boldsymbol{\theta}_T^*) = \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{j,t} \{\beta_T^* (c_{t+1}/c_t)^{\gamma_T^*} (1 + R_{i,t+1}) - 1\} = 0.$$

Testing C-CAPM: GMM - Remarks

- Q: How do we choose \mathbf{Z}_t the l instruments? Not a trivial question. In general, predetermined regressors used to be viewed with a positive light. Today, they are viewed as “acceptable,” in some settings. In many settings, especially, with macro series/indexes they are not.
- Note: Recall that weak instruments are a problem. In theory, we only need small correlation between \mathbf{Z}_t and the model’s variables. However, the bigger the correlation, the better:
 \Rightarrow 50 weak instruments are no substitute for a good IV!
- Advantages of GMM approach:
 - All we need is a moment condition.
 - No need to log-linearize anything.
 - Non-linearities are not a problem.
 - Robust to heteroscedasticity and distributional assumptions.

Testing C-CAPM: GMM - Remarks

- Practical Considerations:
 - We need at least as many moment conditions as parameters (just-identified case).

- If there are more moments –the usual case–, we have “over-identifying restrictions.” Use them to test the model (J -test):

$$J = T Q_T(\boldsymbol{\theta}_T^*) = T^{-1/2} g(\mathbf{v}_t, \boldsymbol{\theta}_T^*)' \mathbf{S}_T^{*-1} g(\mathbf{v}_t, \boldsymbol{\theta}_T^*) \sim \chi_{L \times N - k}^2$$

where $\mathbf{S}_T^* = \text{Var}[g(\mathbf{v}_t, \boldsymbol{\theta}_T^*)]$

- Too many moments are not desirable in practice.
- The instruments (conditioning information) matter.

Testing C-CAPM: GMM - Remarks

- Practical Considerations:
 - Estimating \mathbf{S} is tricky. In general, the moments will be serially dependent. Newey-West (1987) does not work well when the dimensions of the system is large. Small changes to \mathbf{S} produces big swings in estimated $\boldsymbol{\theta}$. (Sometimes is better to work with $\mathbf{W}=\mathbf{I}$!)
 - Some questions regarding the small sample properties of GMM.
 - The over-identifying restrictions are subject to a “which moments to choose?” critique.
 - The J test also depends crucially on \mathbf{S} ; difficult to estimate accurately
 - Not surprisingly, the J test rejects a lot of models. We should be aware of its problems.

Testing C-CAPM: GMM - Example

- Taken from Hansen and Singleton (1982).

For each asset i , H&S have:

$$E_t[\mathbf{Z}_t \{ \beta (c_{t+1}/c_t)^\gamma (1 + R_{i,t+1}) - 1 \}] = 0 \quad i=1, \dots, N.$$

$R_{i,t}$ = NYSE stock returns (VW and EW).

c_t = Consumption (Non-durables (ND) & ND plus services (NDS).)

\mathbf{Z}_t = lagged $R_{i,t+1}$ and c_{t+1}/c_t . (H&S use 1, 2, 4 and 6 lags.)

Findings: $\beta \approx 1$ (around .99) and γ small (between .32 to .03.) J-tests reject C-CAPM.

- General problem with IVE of the C-CAPM: weak instruments. It's difficult to find IVs highly correlated with consumption growth.
- According to Hall's (1978) consumption follows a random walk: lagged $R_{i,t+1}$ and c_{t+1}/c_t should have low correlation with c_{t+1}/c_t !

GMM estimation – General Case

- Go back to GMM estimation but let \mathbf{f} be a vector of continuous nonlinear functions of the data and unknown parameters.

- In our case, we have N assets and the moment condition is:

$$E[\{m(\mathbf{x}_t, \theta_0) (1 + R_{i,t}) - 1\} z_{j,t-1}] = E[\varepsilon_t z_{j,t-1}] = 0,$$

using (lagged) instruments $z_{j,t-1}$ for each asset $i = 1, \dots, N$ and each instrument $j = 1, \dots, q$.

- Collect these as $\mathbf{f}(\mathbf{v}_t, \theta) = \mathbf{z}_{t-1}' \otimes \varepsilon_t(\mathbf{x}_t, \theta)$, where \mathbf{z}_t is a $1 \times q$ vector of instruments and ε_t is a $N \times 1$ vector. \mathbf{f} is a column vector with qN elements – it contains the cross-product of each instrument with each element of ε .

- Population moment condition: $E[\mathbf{f}(\mathbf{v}_t, \theta_0)] = 0$

GMM estimation – General Case

- The population moment condition is

$$E[\mathbf{f}(\mathbf{v}_t, \theta_0)] = 0$$

- As before, let $\mathbf{g}_T(\theta) = [\frac{1}{T} \sum_{t=1}^T \mathbf{f}(\mathbf{v}_t, \theta)]$. The GMM estimator is

$$\theta_T^* = \operatorname{argmin}_{\theta \in \Theta} Q_T(\theta)$$

- The f.o.c.'s are

$$\mathbf{G}_T(\theta_T^*)' \mathbf{W}_T \mathbf{g}_T(\theta_T^*) = 0$$

where $\mathbf{G}_T(\theta_T^*)$: Matrix of partial derivatives with i, j element $\delta \mathbf{g}_{T,i} / \delta \theta_j$

GMM Asymptotics: General Case

It can be shown that:

$$(1) \boldsymbol{\theta}_T^* \xrightarrow{p} \boldsymbol{\theta}_0$$

$$(2) (\boldsymbol{\theta}_{T,i}^* - \boldsymbol{\theta}_{0,i}) / [\sqrt{\mathbf{V}_{T,ii}^*} / T] \xrightarrow{d} N(0, 1)$$

where $\text{Asy Var}(\boldsymbol{\theta}_{T,i}^*) = \mathbf{V} = \mathbf{M} \mathbf{S} \mathbf{M}'$ where

$$- \mathbf{M} = (\mathbf{G}_0' \mathbf{W} \mathbf{G}_0)^{-1} \mathbf{G}_0' \mathbf{W}$$

$$- \mathbf{G}_0 = E[\partial f(\mathbf{v}_t, \boldsymbol{\theta}_0) / \partial \boldsymbol{\theta}']$$

$$- \mathbf{S} = \lim_{T \rightarrow \infty} \text{Var}[T^{1/2} \mathbf{g}_T(\boldsymbol{\theta}_0)]$$

(3) A test of the model's over-identifying restrictions is given by

$$J_T = T Q_T(\boldsymbol{\theta}_T^*) \xrightarrow{d} \chi_{qN-k}^2$$

Covariance matrix estimation for GMM

• In practice $\mathbf{V}_T^* = \mathbf{M}_T^* \mathbf{S}_T^* \mathbf{M}_T^{*'} is a consistent estimator of \mathbf{V} , where$

$$- \mathbf{M}_T^* = [\mathbf{G}_T(\boldsymbol{\theta}_T^*)' \mathbf{W}_T \mathbf{G}_T(\boldsymbol{\theta}_T^*)]^{-1} \mathbf{G}_T(\boldsymbol{\theta}_T^*)' \mathbf{W}_T$$

- \mathbf{S}_T^* is a consistent estimator of \mathbf{S}

- Estimator of \mathbf{S} depends on time series properties of $f(\mathbf{v}_t, \boldsymbol{\theta}_0)$. In general, it is

$$\mathbf{S} = \boldsymbol{\Gamma}_0 + \Sigma(\boldsymbol{\Gamma}_i + \boldsymbol{\Gamma}_i')$$

where $\boldsymbol{\Gamma}_i = E\{f_t - E(f_t)\} \{f_{t-i} - E(f_{t-i})\}' = E[f_t f_{t-i}']$ is the i -th autocovariance matrix of $f_t = f(\mathbf{v}_t, \boldsymbol{\theta}_0)$.

• We can consistently estimate \mathbf{S} with

$$\mathbf{S}_T^* = \boldsymbol{\Gamma}_0^*(\boldsymbol{\theta}_T^*) + \Sigma \{ \boldsymbol{\Gamma}_i^*(\boldsymbol{\theta}_T^*) + \boldsymbol{\Gamma}_i^{*'}(\boldsymbol{\theta}_T^*) \}$$

where

$$\boldsymbol{\Gamma}_i^*(\boldsymbol{\theta}_T^*) = \frac{1}{T} \sum_{t=1}^T f(\mathbf{v}_t, \boldsymbol{\theta}_T^*) * f(\mathbf{v}_{t-i}, \boldsymbol{\theta}_T^*)'$$

Covariance matrix estimation for GMM

- If theory implies that the autocovariances of $f(\mathbf{v}_t, \boldsymbol{\theta}_0) = \mathbf{0}$ for some lag i , then we can exclude these from \mathbf{S}_T^* –e.g., $\boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}(\mathbf{v}_t, \boldsymbol{\theta}_T^*)$ are serially uncorrelated implies

$$\mathbf{S}_T^* = \frac{1}{T} \sum_{t=1}^T \{ \boldsymbol{\varepsilon}(\mathbf{v}_t, \boldsymbol{\theta}_T^*) \boldsymbol{\varepsilon}(\mathbf{v}_t, \boldsymbol{\theta}_T^*)' \otimes (\mathbf{z}_t' \mathbf{z}_t) \}$$

GMM Adjustments

- Iterated GMM is recommended in small samples
- More powerful tests by subtracting sample means of $f(\mathbf{v}_t, \boldsymbol{\theta}_T^*)$ in calculating $\boldsymbol{\Gamma}_i^*(\boldsymbol{\theta}_T^*)$
- Asymptotic standard errors may be understated in small samples: multiply asymptotic variances by “degrees of freedom adjustment” $T/(T - k)$ or $\left\{ \frac{(N+q)T}{(N+q)T - k} \right\}$.