Lecture 10
GMM

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:

$$E v_t] - \mu = 0$$
$$E[v_t^2] - (\sigma^2+\mu^2) = 0$$

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

$$\frac{1}{T}\sum_{t=1}^{T} v_t - \mu* = 0 \qquad \Rightarrow \mu* = \frac{1}{T}\sum_{t=1}^{T} v_t$$
$$\frac{1}{T}\sum_{i=1}^{T} v_t^2 - (\sigma*^2 + \mu*^2) = 0 \quad \Rightarrow \sigma*^2 = \frac{1}{T}\sum_{t=1}^{T}(v_t - \mu*)^2$$

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 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:

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

The MM leads to

$$a^* = \frac{1}{T}\sum_{i=1}^{T} z_t^D q_t \,/ \frac{1}{T}\sum_{i=1}^{T} z_t^D p_t \qquad \text{(IV estimator)}$$

Method of Moments: Review

- Population moment condition: A vector of observed variables, $\boldsymbol{v}_t$, and vector of k parameters $\boldsymbol{\theta}$, satisfy a kx1 element vector of conditions E[$f(\boldsymbol{v}_t, \boldsymbol{\theta})$] = $\mathbf{0}$ for all t

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

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

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 (1).

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

Generalized Method of Moments (GMM)

- Now, suppose $\mathbf{f}$ is a qx1 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

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

- We define "closeness" by

$$Q_T(\boldsymbol{\theta}) = [\frac{1}{T}\sum_{t=1}^{T} f(\boldsymbol{v}_t, \boldsymbol{\theta}^*_T)]' \, \boldsymbol{W}_T \, [\frac{1}{T}\sum_{t=1}^{T} f(\boldsymbol{v}_t, \boldsymbol{\theta}^*_T)]$$
$$= \boldsymbol{g}_T(\boldsymbol{\theta}*)' \, \boldsymbol{W}_T \, \boldsymbol{g}_T(\boldsymbol{\theta}*)$$

where $\boldsymbol{W}_T$ is psd & plim($\boldsymbol{W}_T$) = $\boldsymbol{W}$ pd

GMM: Example 1

- Power utility based asset pricing model –Hansen and Singleton (1982)
  - Theory condition:
    $E_t[\{\beta(c_{t+1}/c_t)^{-\gamma}(1 + R_{i,t+1}) - 1\}] = 0$ with unknown parameters β, γ

  - The q population unconditional moment conditions are
    $E_t[Z_{j,t}\ \{\beta(c_{t+1}/c_t)^{-\gamma}(1 + R_{i,t+1}) - 1\}] = 0 \qquad j = 1, \ldots, q$
    where $Z_{j,t}$ are instruments in the information set

- The q sample moment conditions are
  $\frac{1}{T}\sum_{t=1}^{T} Z_j\ \{\beta_T^* (c_{t+1}/c_t)^{\gamma_T^*}\ (1 + R_{i,t+1}) - 1\} = 0.$

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

  - 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 probability density function of the continuous stationary random vector v_t, given $V_{t-1}=\{v_{t-1},v_{t-2},\ldots\}$ is $p(v_t;\boldsymbol{\theta}_0,V_{t-1})$

• The MLE of $\boldsymbol{\theta}_0$ based on the conditional log likelihood function is the value of which maximizes $L_T(\boldsymbol{\theta}) = \Sigma\ln\{p(v_t;\boldsymbol{\theta},V_{t-1})\}$

$$\Rightarrow \text{solving } \partial L_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$$

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

$$E[\,\partial\ln\{p(v_t;\boldsymbol{\theta},V_{t-1})\}/\partial\boldsymbol{\theta}\}\,] = 0$$

GMM: Summary

• The GMM estimator $\boldsymbol{\theta}_T^* = \text{argmin}_{\theta\,\varepsilon\,\Theta} Q_T(\boldsymbol{\theta})$ generates the f.o.c.

$$[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(\boldsymbol{v}_t,\boldsymbol{\theta}_T^*)}{\partial\boldsymbol{\theta}'}]'\,\boldsymbol{W}_T\,[\frac{1}{T}\sum_{t=1}^{T} f(\boldsymbol{v}_t,\boldsymbol{\theta}_T^*)] = \boldsymbol{0} \qquad (**)$$

where $\frac{\partial f(\boldsymbol{v}_t,\boldsymbol{\theta}_T^*)}{\partial\boldsymbol{\theta}'}$ is a qxk matrix with *i,j* element $\frac{\partial f_i(\boldsymbol{v}_t,\boldsymbol{\theta}_T^*)}{\partial\theta_j}$

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

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

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

Example: IV estimation of linear model

• Linear IV framework:

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

• Let $\mathbf{Z}$ be a Txq vector of IV –i.e., $\text{E}[\mathbf{Z}'\boldsymbol{\varepsilon}] = \mathbf{0}$. and $\text{E}[\mathbf{Z}'\mathbf{X}] \neq \mathbf{0}$.

• We want to estimate θ_0 using GMM. Then, the GMM estimator

$$\boldsymbol{\theta}_T^* = \text{argmin}_{\theta\ \varepsilon\ \Theta} \{Q_T(\boldsymbol{\theta}) = [T^{-1}\ \boldsymbol{\varepsilon}(\boldsymbol{\theta})'\mathbf{Z}]\ \boldsymbol{W}_T\ [T^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\theta})]\}$$

(kx1) f.o.c.:

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

or

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

Example: IV estimation of linear model

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

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

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

independently of the weighting matrix $\boldsymbol{W}_T$.

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

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

Note: GMM = MM based on (kx1) population moment conditions

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

(1) When $q = k$ GMM = MM based on $\text{E}[\mathbf{Z}'\ \boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$.
(2) When $q > k$ GMM sets k linear combinations of $\text{E}[\mathbf{Z}'\ \boldsymbol{\varepsilon}(\boldsymbol{\theta}_0)] = \mathbf{0}$.
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 (kx1) population moments

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

• Let $\mathbf{F}(\boldsymbol{\theta}_0) = \boldsymbol{W}^{1/2} E[\frac{\partial f(\boldsymbol{v}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}]$ (a qxk matrix), with rank($\mathbf{F}(\theta_0)$)= k. (The *rank condition* is necessary for identification of $\boldsymbol{\theta}_0$).

• Rewrite (***) as

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

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

• The LS projection of $\boldsymbol{W}^{1/2} E[f(\boldsymbol{v}_t, \boldsymbol{\theta}_0)]$ 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 rank$\{\mathbf{P_F}\} = k$ restrictions on the (q x1) (*transformed*) population moment condition
$\boldsymbol{W}^{1/2} E[f(\boldsymbol{v}_t, \boldsymbol{\theta}_0)]$.

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

• The restrictions that are left over are

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

• That is, the projection of $\boldsymbol{W}^{1/2} E[f(\boldsymbol{v}_t, \boldsymbol{\theta}_0)]$ 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 (**),

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

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

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

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

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

$$\text{Var}[\mathbf{Z}'\boldsymbol{\varepsilon}/T] = E[\{\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t\varepsilon_t\}\{\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t\varepsilon_t\}']$$
$$= \tfrac{1}{T}\sum_{t=1}^{T} E[\varepsilon_t^2 z_t z_t']$$

• Thus,

$$\boldsymbol{S}_T^* = \tfrac{1}{T}\,[\sum_{t=1}^{T} \varepsilon(\boldsymbol{\theta}_T^*)_t^2 z_t z_t']$$

GMM: Two-step estimator

• The asymptotic variance depends on the weighting matrix $\boldsymbol{W}_T$.

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

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

(1) Start with sub-optimal $\boldsymbol{W}_T(0)$, say $\mathbf{I}$

(2) Using $\boldsymbol{W}_T(0)$ estimate $\boldsymbol{\theta}_T^*(1)$ & $\boldsymbol{S}_T^*(1)$

(3) Estimate with $\boldsymbol{W}_T(1) = \boldsymbol{S}_T^*(1)^{-1}$

(4) Using $\boldsymbol{W}_{T\text{T}}(j)$ repeat steps (2)-(3) to get $\boldsymbol{\theta}_T^*(j+1)$ & $\boldsymbol{S}_T^*(j+1)$ until convergence.

GMM: Two-step estimator

• Note that if $\boldsymbol{\varepsilon}$ is homoskedastic: $\text{var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}_T \quad \Rightarrow \boldsymbol{S}_T^* = s^{*2}\mathbf{ZZ}'$ where s^{*2} is a consistent estimator of σ^2.

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

$$\boldsymbol{\theta}_T^* = \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1}\, \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$
$$= \{\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\}^{-1}\, \hat{\boldsymbol{X}}'\boldsymbol{y}$$

where $\hat{\boldsymbol{X}} = \mathbf{P}_z\mathbf{X}$ 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 qx1 population moment conditions $E[\mathbf{Z}'\varepsilon(\boldsymbol{\theta}_0)] = \mathbf{0}$. We can construct a Wald-type test to check if these q conditions are met in sample. The *overidentifying restrictions test*:

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

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

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:

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

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

$$E_t[Z_j\ \{\beta(c_{t+1}/c_t)^{-\gamma}\,(1 + R_{i,t+1}) - 1\}] = 0 \qquad i = 1, \ldots, N;\ j = 1, \ldots, l.$$

<u>Note</u>: Now we have a lot of moments: lxN!

• GMM works with sample analogues of the population moments:

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

Testing C-CAPM: GMM - Remarks

• Q: How do we choose $\mathbf{Z}_t$ the l instruments? Not a trivial question. In general, predetermined regressors are viewed as OK.

• Note: Weak instruments are a problem. In theory, we only need small correlation between $\mathbf{Z}$ and the model's variables. However, the bigger the correlation, the better:

⇒ 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 heteroscedasticiy 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):

$$J = T\, Q_T(\boldsymbol{\theta}_T^*) = T^{-1/2}\, g(\boldsymbol{v}_t, \boldsymbol{\theta}_T^*)'\, \boldsymbol{S}_T^{*\,-1}\, g(\boldsymbol{v}_t, \boldsymbol{\theta}_T^*) \sim \chi^2_{LxN-k}$$

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

- Too many moments are not desirable in practice.

- The instruments (conditioning information) matter.

Testing C-CAPM: GMM - Remarks

• Practical Considerations:
- Estimating **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 **S** produces big swings in estimated **θ**. (Sometimes is better to work with **W**=**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 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:

$$E_t[\, Z_t \; \{\beta(c_{t+1}/c_t)^{-\gamma}\, (1 + R_{i,t+1}) - 1\}] = 0, \qquad i = 1, \ldots, N.$$

$R_{i,t}$ = NYSE stock returns (VW and EW).
c_t= Consumption (Non-durables (ND) & ND plus services (NDS).)
Z_t= lagged R_{t+1} and c_{t+1}/c_t. (H&S use 1, 2, 4 and 6 lags.)

<u>Findings</u>: $\beta \approx 1$ (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_{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 **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:

$$E[\{m(\boldsymbol{x}_t, \boldsymbol{\theta}_0)\ (1 + R_{i,t}) - 1\}\ z_{j,t-1}] = E[\varepsilon_t\ z_{j,t-1}] = 0,$$

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

• Collect these as $\boldsymbol{f}(\boldsymbol{v}_t, \boldsymbol{\theta}) = z_{t-1}{}' \otimes \varepsilon_t(\boldsymbol{x}_t, \boldsymbol{\theta})$, where $\boldsymbol{z}_t$ is a 1xq vector of instruments and ε_t is a Nx1 vector. $\boldsymbol{f}$ is a column vector with qN elements – it contains the cross-product of each instrument with each element of $\boldsymbol{\varepsilon}$.

• Population moment condition: $E[\boldsymbol{f}(\boldsymbol{v}_t, \boldsymbol{\theta}_0)] = 0$

GMM estimation – General Case

• The population moment condition is

$$E[\boldsymbol{f}(\boldsymbol{v}_t, \boldsymbol{\theta}_0)] = 0$$

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

$$\boldsymbol{\theta}_T^* = \text{argmin}_{\theta \varepsilon \Theta}\ Q_T(\boldsymbol{\theta})$$

• The FOCs are

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

where $\mathbf{G}_T(\boldsymbol{\theta})$ is a matrix of partial derivatives with i, j element

$$\delta \boldsymbol{g}_{T,i} / \delta \boldsymbol{\theta}_j$$

GMM Asymptotics: General Case

It can be shown that:

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

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

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

- $\mathbf{M} = (\mathbf{G}_0'\, \boldsymbol{W}\, \mathbf{G}_0)^{-1}\, \mathbf{G}_0' \boldsymbol{W}$
- $\mathbf{G}_0 = E[\partial f(v_t,\theta_0)/\partial\theta']$
- $\mathbf{S} = \lim_{T\to\infty}$ Var$[T^{-1/2}\mathbf{g}_T(\boldsymbol{\theta}_0)]$

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

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

Covariance matrix estimation for GMM

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

- $\boldsymbol{M}_T^* = [\mathbf{G}_T(\boldsymbol{\theta}_T^*)'\, \boldsymbol{W}_T\, \mathbf{G}_T(\boldsymbol{\theta}_T^*)]^{-1}\, \mathbf{G}_T(\boldsymbol{\theta}^*_T)' \boldsymbol{W}_T$
- $\boldsymbol{S}_T^*$ is a consistent estimator of S
- Estimator of $\mathbf{S}$ depends on time series properties of $f(\boldsymbol{v}_t, \boldsymbol{\theta}_0)$. 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 $f_t = f(\boldsymbol{v}_t, \boldsymbol{\theta}_0)$.

• We can consistently estimate S with

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

where

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

Covariance matrix estimation for GMM

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

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

GMM Adjustments

• Iterated GMM is recommended in small samples

• More powerful tests by subtracting sample means of $f(\boldsymbol{v}_t, \boldsymbol{\theta}_T^*)$ 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*" $T/(T\text{-}k)$ or $(N+q)T/\{(N+q)T - k\}$ where d = k + $((Nq)^2+Nq)/2$