

Generalized Method of Moment

CHUNG-MING KUAN

Department of Finance & CRETA

National Taiwan University

May 23, 2011

Lecture Outline

1 Generalized Method of Moments

- Examples of Moment Conditions
- GMM Estimation

2 Asymptotic Properties of the GMM Estimator

- Consistency
- Asymptotic Normality
- Asymptotic Efficiency

3 Different GMM Estimators

- Two-Step Estimator
- Iterative Estimator
- Independently Weighted Estimator
- Continuous Updating Estimator

4 Applications of the GMM

- Regression with Symmetric Error
- Generalized Instrumental Variables Estimator
- Difference of Sharpe Ratios
- Stochastic Volatility Models

Lecture Outline (cont'd)

5 Large Sample Tests

- Over-Identifying Restriction Test
- Hausman Test
- Wald Test

6 Estimation of Conditional Moment Restrictions

- Estimation by GMM
- The Number of Instruments
- Optimal Instruments

7 Applications

- Capital Asset Pricing Model (CAPM)
- Cox, Ingersoll and Ross (CIR) Model

Examples of Moment Conditions

Example 1: Given $y_t = \mathbf{x}_t' \boldsymbol{\beta} + e_t$, consider the moment function:

$$\mathbb{E}(\mathbf{x}_t e_t) = \mathbb{E}[\mathbf{x}_t (y_t - \mathbf{x}_t' \boldsymbol{\beta})].$$

When $\mathbf{x}_t' \boldsymbol{\beta}$ is correctly specified for the linear projection, the following moment condition holds:

$$\mathbb{E}[\mathbf{x}_t (y_t - \mathbf{x}_t' \boldsymbol{\beta}_o)] = \mathbf{0},$$

for some $\boldsymbol{\beta}_o$. The sample counterpart,

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t (y_t - \mathbf{x}_t' \boldsymbol{\beta}) = \mathbf{0},$$

is also the FOC for OLS estimation.

Example 2: given $y_t = f(\mathbf{x}_t; \beta) + e_t$ and the moment function:

$$\mathbb{E}[\nabla f(\mathbf{x}_t; \beta) e_t] = \mathbb{E}\{\nabla f(\mathbf{x}_t; \beta)[y_t - f(\mathbf{x}_t; \beta)]\}.$$

When $f(\mathbf{x}_t; \beta)$ is correctly specified for $\mathbb{E}(y_t|\mathbf{x}_t)$, we have the moment condition:

$$\mathbb{E}\{\mathbf{x}_t[y_t - f(\mathbf{x}_t; \beta_o)]\} = \mathbf{0},$$

for some β_o . Its sample counterpart is the FOC for NLS estimation:

$$\frac{1}{T} \sum_{t=1}^T \nabla f(\mathbf{x}_t; \beta)[y_t - f(\mathbf{x}_t; \beta)] = \mathbf{0}.$$

Example 3: Given the quasi-likelihood function $f(\mathbf{x}_t; \theta)$, the moment condition $\mathbb{E}[\nabla \ln f(\mathbf{x}_t; \theta_o)] = \mathbf{0}$ holds for the minimizer of the Kullback-Leibler information criterion, β_o , and its sample counterpart is the average of the score functions.

Example 4: Given $y_t = \mathbf{x}'_t \beta + e_t$ and the moment function:

$$\mathbb{E}(\mathbf{z}_t e_t) = \mathbb{E}[\mathbf{z}_t (y_t - \mathbf{x}'_t \beta)],$$

for some variables \mathbf{z}_t . When \mathbf{z}_t are proper instrument variables such that

$$\mathbb{E}[\mathbf{z}_t (y_t - \mathbf{x}'_t \beta_o)] = \mathbf{0},$$

its sample counterpart is the FOC for IV estimation:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{z}_t (y_t - \mathbf{x}'_t \beta) = \mathbf{0}.$$

Note: We can **not** solve for unknown parameters if the number of moment conditions (i.e., the dimension of \mathbf{z}_t) is more than the number of parameters.

GMM Estimation

Consider q moment functions $\mathbb{E}[\mathbf{m}(\mathbf{z}_t; \boldsymbol{\theta})]$, where $\boldsymbol{\theta}$ is $k \times 1$, suppose

$$\mathbb{E}[\mathbf{m}(\mathbf{z}_t; \boldsymbol{\theta}_o)] = \mathbf{0}.$$

for some unique parameter vector $\boldsymbol{\theta}_o$.

- These conditions are **exactly identified** if $q = k$ and **over-identified** if $q > k$.
- When the conditions are exactly identified, $\boldsymbol{\theta}_o$ can be estimated by solving their sample counterpart:

$$\bar{\mathbf{m}}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{m}(\mathbf{z}_t; \boldsymbol{\theta}) = \mathbf{0}.$$

This is known as the **method of moment**, which is **not** applicable when the conditions are over-identified.

- When the conditions are over-identified, the following quadratic objective function is minimized at $\theta = \theta_o$:

$$\bar{Q}(\theta; \mathbf{W}_o) := \mathbb{E}[\mathbf{m}(\mathbf{z}_t; \theta)]' \mathbf{W}_o \mathbb{E}[\mathbf{m}(\mathbf{z}_t; \theta)],$$

where \mathbf{W}_o is a $q \times q$ symmetric and p.d. weighting matrix.

- Hansen (1982, *Econometrica*): The **generalized method of moments** (GMM) suggests estimating θ_o by minimizing

$$Q_T(\theta; \mathbf{W}_T) = [\bar{\mathbf{m}}_T(\theta)]' \mathbf{W}_T [\bar{\mathbf{m}}_T(\theta)],$$

where \mathbf{W}_T , possibly dependent on the sample, is a symmetric and p.d. matrix that converges to \mathbf{W}_o in probability. The GMM estimator is:

$$\hat{\theta}_T(\mathbf{W}_T) = \arg \min_{\theta \in \Theta} Q_T(\theta; \mathbf{W}_T),$$

which clearly depends on the choice of \mathbf{W}_T .

- The FOC of GMM estimation contains k equations in k unknowns:

$$\mathbf{G}_T(\boldsymbol{\theta})' \mathbf{W}_T \bar{\mathbf{m}}_T(\boldsymbol{\theta}) = \mathbf{0},$$

where $\mathbf{G}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T \nabla \mathbf{m}(\mathbf{z}_t; \boldsymbol{\theta})$ is $q \times k$. The GMM estimator can be solved using a nonlinear optimization algorithm.

- In the linear regression case,

$$\bar{\mathbf{m}}_T(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t (y_t - \mathbf{x}_t' \boldsymbol{\beta}).$$

When $\mathbf{W}_T = \mathbf{I}_k$, the FOC of GMM estimation is

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

The resulting solution is the OLS estimator.

Consistency

It can be seen that

$$\begin{aligned} & |Q_T(\theta; \mathbf{W}_T) - \bar{Q}(\theta; \mathbf{W}_o)| \\ & \leq \left| [\bar{\mathbf{m}}_T(\theta) - \mathbb{E} \mathbf{m}(\mathbf{z}_t; \theta)]' \mathbf{W}_T [\bar{\mathbf{m}}_T(\theta) - \mathbb{E} \mathbf{m}(\mathbf{z}_t; \theta)] \right| \\ & \quad + \left| [\bar{\mathbf{m}}_T(\theta) - \mathbb{E} \mathbf{m}(\mathbf{z}_t; \theta)]' \mathbf{W}_T \mathbb{E} [\mathbf{m}(\mathbf{z}_t; \theta)] \right| \\ & \quad + \left| \mathbb{E} [\mathbf{m}(\mathbf{z}_t; \theta)]' \mathbf{W}_T [\bar{\mathbf{m}}_T(\theta) - \mathbb{E} [\mathbf{m}(\mathbf{z}_t; \theta)]] \right| \\ & \quad + \left| \mathbb{E} [\mathbf{m}(\mathbf{z}_t; \theta)]' (\mathbf{W}_T - \mathbf{W}_o) \mathbb{E} [\mathbf{m}(\mathbf{z}_t; \theta)] \right|. \end{aligned}$$

By invoking a suitable ULLN, $Q_T(\theta; \mathbf{W}_T)$ is close to $\bar{Q}(\theta; \mathbf{W}_o)$ uniformly in θ when T is large. Hence, the GMM estimator $\hat{\theta}_T(\mathbf{W}_T)$ ought to be close to θ_o , the minimizer of $\bar{Q}(\theta; \mathbf{W}_o)$, for sufficiently large T . This approach is analogous to that for establishing NLS and QMLE consistency.

Asymptotic Normality

Consider the mean value expansion:

$$\sqrt{T}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T)) = \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) + \mathbf{G}_T(\boldsymbol{\theta}_T^\dagger)\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T) - \boldsymbol{\theta}_o),$$

where $\boldsymbol{\theta}_T^\dagger$ lies between $\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T)$ and $\boldsymbol{\theta}_o$. Using the FOC of GMM estimation,

$$\begin{aligned} \mathbf{0} &= \mathbf{G}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T))' \mathbf{W}_T \sqrt{T}\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T)) \\ &= \mathbf{G}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T))' \mathbf{W}_T \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) \\ &\quad + \mathbf{G}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T))' \mathbf{W}_T \mathbf{G}_T(\boldsymbol{\theta}_T^\dagger) [\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T) - \boldsymbol{\theta}_o)] \\ &= \mathbf{G}_o' \mathbf{W}_o \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) + \mathbf{G}_o' \mathbf{W}_o \mathbf{G}_o [\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T) - \boldsymbol{\theta}_o)] + o_P(1), \end{aligned}$$

where $\mathbf{G}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T))$ converges to $\mathbf{G}_o = \mathbb{E}[\nabla \mathbf{m}(\mathbf{z}_t; \boldsymbol{\theta}_o)]$ uniformly in $\boldsymbol{\theta}$ under a suitable ULLN. It follows that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T) - \boldsymbol{\theta}_o) = -(\mathbf{G}_o' \mathbf{W}_o \mathbf{G}_o)^{-1} \mathbf{G}_o' \mathbf{W}_o \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) + o_P(1).$$

When $\mathbf{m}(\mathbf{z}_t; \theta_o)$ obey a central limit theorem:

$$\sqrt{T} \bar{\mathbf{m}}_T(\theta_o) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{m}(\mathbf{z}_t; \theta_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_o),$$

we have the following result.

Asymptotic Normality

$\sqrt{T}(\hat{\theta}_T(\mathbf{W}_T) - \theta_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Omega_o(\mathbf{W}_o))$, where

$$\Omega_o(\mathbf{W}_o) = (\mathbf{G}'_o \mathbf{W}_o \mathbf{G}_o)^{-1} \mathbf{G}'_o \mathbf{W}_o \boldsymbol{\Sigma}_o \mathbf{W}_o \mathbf{G}_o (\mathbf{G}'_o \mathbf{W}_o \mathbf{G}_o)^{-1}.$$

When $\mathbf{W}_o = \boldsymbol{\Sigma}_o^{-1}$, $\Omega_o(\mathbf{W}_o)$ simplifies to

$$\Omega_o(\boldsymbol{\Sigma}_o^{-1}) = (\mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o)^{-1} \mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o (\mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o)^{-1} = (\mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o)^{-1}.$$

Asymptotic Efficiency

To compare $\Omega_o(\mathbf{W}_o)$ and $\Omega_o(\Sigma_o^{-1})$, note that

$$\begin{aligned}\Omega_o(\Sigma_o^{-1})^{-1} - \Omega_o(\mathbf{W}_o)^{-1} &= \mathbf{G}_o' \Sigma_o^{-1} \mathbf{G}_o - \mathbf{G}_o' \mathbf{W}_o \mathbf{G}_o (\mathbf{G}_o' \mathbf{W}_o \Sigma_o \mathbf{W}_o \mathbf{G}_o)^{-1} \mathbf{G}_o' \mathbf{W}_o \mathbf{G}_o \\ &= \mathbf{G}_o' \Sigma_o^{-1/2} \\ &\quad \left[\mathbf{I} - \Sigma_o^{1/2} \mathbf{W}_o \mathbf{G}_o (\mathbf{G}_o' \mathbf{W}_o \Sigma_o^{1/2} \Sigma_o^{1/2} \mathbf{W}_o \mathbf{G}_o)^{-1} \mathbf{G}_o' \mathbf{W}_o \Sigma_o^{1/2} \right] \\ &\quad \Sigma_o^{-1/2} \mathbf{G}_o,\end{aligned}$$

which is p.s.d., because the matrix in the square bracket is symmetric and idempotent. Thus, $\Omega_o(\mathbf{W}_o) - \Omega_o(\Sigma_o^{-1})$ is p.s.d. Σ_o^{-1} is also known as the **optimal** (limiting) weighting matrix.

Note: In practice, we can obtain an **asymptotically efficient** GMM estimator by computing the GMM estimator $\hat{\theta}(\mathbf{W}_T)$ such that \mathbf{W}_T is consistent for Σ_o^{-1} .

Two-Step Estimator

Hansen and Singleton (1982, *Econometrica*):

- 1 Compute a preliminary, consistent estimator based on the pre-specified weighting matrix $\mathbf{W}_{0,T}$:

$$\hat{\theta}_{1,T}(\mathbf{W}_{0,T}) := \arg \min_{\theta \in \Theta} [\bar{\mathbf{m}}_T(\theta)]' \mathbf{W}_{0,T} [\bar{\mathbf{m}}_T(\theta)].$$

For example, $\mathbf{W}_{0,T}$ may be \mathbf{I}_q .

- 2 Compute a consistent estimator for Σ_o based on $\hat{\theta}_{1,T}$ and use its inverse as the optimal weighting matrix, i.e., $\mathbf{W}_T(\hat{\theta}_{1,T}) = \hat{\Sigma}_T^{-1}$.
- 3 The **two-step** GMM estimator is computed as

$$\hat{\theta}_{2,T}(\hat{\Sigma}_T^{-1}) := \arg \min_{\theta \in \Theta} [\bar{\mathbf{m}}_T(\theta)]' \hat{\Sigma}_T^{-1} [\bar{\mathbf{m}}_T(\theta)].$$

As $\hat{\Sigma}_T^{-1}$ is consistent for Σ_o^{-1} , this is an asymptotically efficient estimator.

Drawbacks of Two-Step Estimators

- The finite-sample performance of the two-step estimator clearly depends on the initial weighting matrix $\mathbf{W}_{0,T}$ and the resulting, preliminary GMM estimator $\hat{\theta}_{1,T}$.
- The second step hinges on a consistent estimator of Σ_o .
 - When $\mathbf{m}(\mathbf{z}_t; \theta)$ are not serially correlated,

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{m}(\mathbf{z}_t; \hat{\theta}_{1,T}) \mathbf{m}(\mathbf{z}_t; \hat{\theta}_{1,T})'.$$

- When $\mathbf{m}(\mathbf{z}_t; \theta)$ are serially correlated, a Newey-West type estimator is needed.
- It has been shown that, as $\hat{\Sigma}_T$ is determined by $\mathbf{m}(\mathbf{z}_t; \theta)$, the correlation between $\bar{\mathbf{m}}_T$ and $\hat{\Sigma}_T$ may induce finite-sample bias in $\hat{\theta}_{2,T}$.

Iterative Estimator

- 1 At the j^{th} iteration, compute the j^{th} **iterative** GMM estimator using the weighting matrix $\mathbf{W}_T(\hat{\theta}_{j-1,T})$:

$$\hat{\theta}_{j,T} := \arg \min_{\theta \in \Theta} [\bar{\mathbf{m}}_T(\theta)]' \mathbf{W}_T(\hat{\theta}_{j-1,T}) [\bar{\mathbf{m}}_T(\theta)].$$

The initial weighting matrix, $\mathbf{W}_{0,T}$, may be \mathbf{I}_q .

- 2 Use $\hat{\theta}_{j,T}$ to construct the optimal weighting matrix $\mathbf{W}_T(\hat{\theta}_{j,T})$, which is consistent for Σ_o^{-1} , and set $j = j + 1$.
- 3 The convergence criteria for iterations: For some pre-specified ε ,

$$\|Q_T(\hat{\theta}_{j,T}) - Q_T(\hat{\theta}_{j-1,T})\| \leq \varepsilon, \quad \text{or} \quad \|\hat{\theta}_{j,T} - \hat{\theta}_{j-1,T}\| \leq \varepsilon.$$

Note: More iterations may (or may not) improve finite-sample performance but do **not** affect asymptotic efficiency.

Independently Weighted Estimator

Altonji and Segal (1996, *JBES*): The **independently weighted** estimator avoids possible correlation between $\bar{\mathbf{m}}_T$ and $\hat{\Sigma}_T$ by splitting the sample into sub-samples and computing $\bar{\mathbf{m}}_T$ and $\hat{\Sigma}_T$ based on different sub-samples.

- Split the sample into ℓ groups, and let $\bar{\mathbf{m}}_{T_j}(\theta)$ be the sample average of $\mathbf{m}(\mathbf{z}_t, \theta)$ for t in the j th group with T_j observations.
- Also let $\hat{\Sigma}_{T_j}^{-1}$ be the optimal weighting matrix based on the observations not in the j th group.
- The resulting GMM estimator is computed as:

$$\hat{\theta}_T := \arg \min_{\theta \in \Theta} \sum_{j=1}^{\ell} [\bar{\mathbf{m}}_{T_j}(\theta)]' \hat{\Sigma}_{T_j}^{-1} [\bar{\mathbf{m}}_{T_j}(\theta)].$$

A common choice of ℓ is 2.

Continuous Updating Estimator

Hansen, Heaton and Yaron (1996, *JBES*): Instead of estimating in 2 (or more) steps, the **continuous updating** (CU) estimator is based on one-time optimization:

$$\hat{\theta}_T := \arg \min_{\theta \in \Theta} [\bar{\mathbf{m}}_T(\theta)]' \mathbf{W}_T(\theta) [\bar{\mathbf{m}}_T(\theta)].$$

Computing this estimator may be computationally cumbersome.

- The limiting distribution of the resulting estimator is the same as that of the two-step estimator; see Pakes and Pollard (1989, *Econometrica*).
- The CU estimator is usually invariant when the moment conditions are re-scaled, even when the scale factor is parameter dependent; the two-stage or iterative GMM estimator is sensitive to such transformation, however.

Regression with Symmetric Error

Given the specification $y_t = \mathbf{x}_t' \boldsymbol{\beta} + e_t$, let

$$\mathbf{m}(y_t, \mathbf{x}_t; \boldsymbol{\beta}) = \begin{bmatrix} \mathbf{x}_t(y_t - \mathbf{x}_t' \boldsymbol{\beta}) \\ \mathbf{x}_t(y_t - \mathbf{x}_t' \boldsymbol{\beta})^3 \end{bmatrix}.$$

The moment condition $\mathbb{E}[\mathbf{m}(y_t, \mathbf{x}_t; \boldsymbol{\beta}_o)] = \mathbf{0}$ suggests estimating $\boldsymbol{\beta}_o$ while taking into account symmetry of the error term. The gradient vector of \mathbf{m} is:

$$\nabla \mathbf{m}(y_t, \mathbf{x}_t; \boldsymbol{\beta}) = \begin{bmatrix} -\mathbf{x}_t \mathbf{x}_t' \\ -3\mathbf{x}_t \mathbf{x}_t' (y_t - \mathbf{x}_t' \boldsymbol{\beta})^2 \end{bmatrix}.$$

If the data are independent over t ,

$$\boldsymbol{\Sigma}_o = \mathbb{E} \begin{bmatrix} \epsilon_t^2 \mathbf{x}_t \mathbf{x}_t' & \epsilon_t^4 \mathbf{x}_t \mathbf{x}_t' \\ \epsilon_t^4 \mathbf{x}_t \mathbf{x}_t' & \epsilon_t^6 \mathbf{x}_t \mathbf{x}_t' \end{bmatrix}.$$

Let $\hat{e}_t = y_t - \mathbf{x}_t' \hat{\beta}_{1,T}$, where $\hat{\beta}_{1,t}$ is a first-step GMM estimator based on a preliminary weighting matrix. Then, Σ_o may be estimated by the sample counterpart:

$$\hat{\Sigma}_T(\hat{\beta}_{1,T}) = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} \hat{e}_t^2 \mathbf{x}_t \mathbf{x}_t' & \hat{e}_t^4 \mathbf{x}_t \mathbf{x}_t' \\ \hat{e}_t^4 \mathbf{x}_t \mathbf{x}_t' & \hat{e}_t^6 \mathbf{x}_t \mathbf{x}_t' \end{bmatrix}.$$

Note that $\hat{\beta}_{1,T}$ here may be the OLS estimator; a consistent estimator for β_o suffices.

The two-step GMM estimator is computed with $[\hat{\Sigma}_T(\hat{\beta}_{1,T})]^{-1}$ as the weighting matrix. That is,

$$\hat{\beta}_{2,T} := \arg \min_{\theta \in \Theta} \bar{\mathbf{m}}_T(\beta) [\hat{\Sigma}_T(\hat{\beta}_{1,T})]^{-1} \bar{\mathbf{m}}_T(\beta).$$

Generalized Instrumental Variables Estimator

For the specification $y_t = \mathbf{x}_t' \beta + e_t$, consider the moment condition:

$$\mathbb{E}[\mathbf{m}_t(\beta_o)] = \mathbb{E}[\mathbf{z}_t(y_t - \mathbf{x}_t' \beta_o)] = \mathbf{0},$$

where \mathbf{z}_t contains $q > k$ instrumental variables. The GMM estimator minimizes

$$\left(\frac{1}{T} \sum_{t=1}^T \mathbf{z}_t (y_t - \mathbf{x}_t' \beta) \right)' \mathbf{W}_T \left(\frac{1}{T} \sum_{t=1}^T \mathbf{z}_t (y_t - \mathbf{x}_t' \beta) \right),$$

and solves

$$\left(\sum_{t=1}^T \mathbf{x}_t \mathbf{z}_t' \right) \mathbf{W}_T \left(\sum_{t=1}^T \mathbf{z}_t (y_t - \mathbf{x}_t' \beta) \right) = (\mathbf{X}' \mathbf{Z}) \mathbf{W}_T [\mathbf{Z}' (\mathbf{y} - \mathbf{X} \beta)] = \mathbf{0}.$$

where \mathbf{Z} ($T \times q$) is the matrix of instrumental variables and \mathbf{X} ($T \times k$) is the matrix of regressors.

The GMM estimator is

$$\hat{\beta}(\mathbf{W}_T) = (\mathbf{X}'\mathbf{Z}\mathbf{W}_T\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_T\mathbf{Z}'\mathbf{y}.$$

which is known as the **generalized instrumental variables estimator** (GIVE).

- When the data are independent and there is **no** condition heteroskedasticity,

$$\boldsymbol{\Sigma}_o = \text{var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{z}_t (y_t - \mathbf{x}'_t \beta_o) \right) = \frac{\sigma_o^2}{T} \sum_{t=1}^T \mathbb{E}(\mathbf{z}_t \mathbf{z}'_t).$$

Ignoring σ_o^2 in $\boldsymbol{\Sigma}_o$, we can estimate $T^{-1} \sum_{t=1}^T \mathbb{E}(\mathbf{z}_t \mathbf{z}'_t)$ by $\mathbf{Z}'\mathbf{Z}/T$.

- This leads to the following two-step estimator:

$$\hat{\beta}_{2,T}(\mathbf{Z}'\mathbf{Z}) = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = [\tilde{\mathbf{X}}'\tilde{\mathbf{X}}]^{-1}\tilde{\mathbf{X}}'\mathbf{y},$$

where $\tilde{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ is the matrix of fitted values from the OLS regression of \mathbf{X} on \mathbf{Z} . As such, this is also known as the **two-stage least squares** (2SLS) estimator.

- When the data are independent and there is conditional heteroskedasticity, the 2SLS estimator remains consistent (why?), but it is **not** asymptotically efficient.
- A two-step GMM estimator with a properly estimated weighting matrix $[\hat{\Sigma}_T(\hat{\beta}_{1,T})]^{-1}$ would be more efficient asymptotically. For example, the weighting matrix may be

$$\hat{\Sigma}_T(\hat{\beta}_{1,T}) = \frac{1}{T} \sum_{t=1}^T \hat{e}_t^2 \mathbf{z}_t \mathbf{z}_t',$$

where $\hat{e}_t = y_t - \mathbf{x}_t' \hat{\beta}_{1,T}$ are the residuals from the first-step GMM estimation.

- When the data are dependent over time, a Newey-West type estimator of the weighting matrix would be needed.

Difference of Sharpe Ratios

Lo (2002, FAJ)

- Let $r_{t,j}$, $j = 1, 2$, be the monthly return of the j -th mutual fund with mean μ_j and r_f the risk free rate. The Difference of Sharpe ratios (DSR) is a criterion to evaluate their relative risk-adjusted performance:

$$\text{DSR}(\theta) := \frac{\mu_1 - r_f}{\sqrt{\gamma_1 - \mu_1^2}} - \frac{\mu_2 - r_f}{\sqrt{\gamma_2 - \mu_2^2}},$$

where $\gamma_j := \mathbb{E}(r_{t,j}^2)$ so that $\text{var}(r_{t,j}) = \gamma_j - \mu_j^2$.

- Let $\mathbf{z}_t = (r_{t,1}, r_{t,2}, r_{t,1}^2, r_{t,2}^2)'$ and $\theta_o = (\mu_1, \mu_2, \gamma_1, \gamma_2)'$. We may estimate θ_o based on the moment conditions: $\mathbb{E}[\mathbf{z}_t - \theta_o] = 0$ so that the (G)MM estimator is $\hat{\theta}_T = \sum_{t=1}^T \mathbf{z}_t / T$.

- The (G)MM estimator is such that $\sqrt{T}(\hat{\theta}_T - \theta_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma_o)$, where

$$\Sigma_o = \lim_{T \rightarrow \infty} \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T (\mathbf{z}_t - \theta_o)(\mathbf{z}_s - \theta_o)' \right],$$

because \mathbf{z}_t may be serially dependent due to, e.g., volatility clustering. Σ_o may be estimated using a Newey-West type estimator $\hat{\Sigma}_T$.

- Using the delta method

$$\sqrt{T}[\text{DSR}(\hat{\theta}_T) - \text{DSR}(\theta_o)] \xrightarrow{D} \mathcal{N}(0, \nabla_{\theta} \text{DSR}(\theta_o) \Sigma_o \nabla_{\theta} \text{DSR}(\theta_o)'),$$

where $\nabla_{\theta} \text{DSR}(\theta_o)$ is

$$\left(\frac{\partial \text{DSR}(\theta)}{\partial \mu_1} \quad \frac{\partial \text{DSR}(\theta)}{\partial \mu_2} \quad \frac{\partial \text{DSR}(\theta)}{\partial \gamma_1} \quad \frac{\partial \text{DSR}(\theta)}{\partial \gamma_2} \right) \bigg|_{\theta=\theta_o}.$$

The covariance matrix estimator of DSR is therefore

$$\nabla_{\theta} \text{DSR}(\hat{\theta}_T) \hat{\Sigma}_T \nabla_{\theta} \text{DSR}(\hat{\theta}_T)'$$

Stochastic Volatility Models

- Stochastic volatility (SV) models: See Shephard (1996) and Taylor (2005) for some surveys. Taylor (1986) model daily return as

$$r_t = \mu + \sigma_t u_t,$$
$$\log(\sigma_t) - \alpha = \phi(\log(\sigma_{t-1}) - \alpha) + \eta_t, \quad |\phi| < 1,$$

where u_t are i.i.d. $\mathcal{N}(0, 1)$, η_t are i.i.d. $\mathcal{N}(0, \beta^2(1 - \phi^2))$, and u_t and η_t are mutually independent. Also assume σ_t is positive stationary and $\mathbb{E}(\sigma_t^4)$ is finite. The parameter vector is $\theta = (\mu, \alpha, \beta, \phi)'$.

- The moment conditions:

$$\mathbb{E}|r_t - \mu|^i = \mathbb{E}(\sigma_t^i) \mathbb{E}(|u_t|^i),$$
$$\mathbb{E}[|r_t - \mu|^k |r_{t+\tau} - \mu|^k] = \mathbb{E}(\sigma_t^k \sigma_{t+\tau}^k) \mathbb{E}(|u_t|^k)^2.$$

- Consider the sample moment functions:

$$\frac{1}{T} \sum_{t=1}^T |r_t - \bar{r}|^i - \mathbb{E}(\sigma_t^i) \mathbb{E}(|u_t^i|), \quad i = 1, 2, 3, 4,$$

where $\bar{r} = \sum_{t=1}^T r_t / T$, and for $\tau > 0$,

$$\frac{1}{T} \sum_{t=1}^T (|r_t - \bar{r}|)^k (|r_{t+\tau} - \bar{r}|)^k - \mathbb{E}(\sigma_t^k \sigma_{t+\tau}^k) \mathbb{E}(|u_t^k|)^2, \quad k = 1, 2.$$

Under certain distribution conditions, it can be shown that $\mathbb{E}(\sigma_t^i)$, $\mathbb{E}(\sigma_t^k \sigma_{t+\tau}^k)$, and $\mathbb{E}(|u_t^i|)$ have analytic forms.

- For a normally distributed u_t ,

$$\mathbb{E}(|u_t|^p) = 2^{p/2} \pi^{-1/2} \Gamma((p+1)/2).$$

- As $\log(\sigma_t) \sim \mathcal{N}(\alpha, \beta^2)$, we have for any positive number p ,

$$\log(\sigma_t^p) = p \log(\sigma_t) \sim \mathcal{N}(p\alpha, p^2\beta^2).$$

By the moment generating function of normal r.v.,

$$\mathbb{E}(\sigma_t^p) = \exp\left(p\alpha + \frac{1}{2}p^2\beta^2\right).$$

- The AR(1) structure of $\log(\sigma_t)$ implies

$$\text{cov}(\log \sigma_t^p, \log \sigma_{t+\tau}^p) = p^2 \text{cov}(\log \sigma_t, \log \sigma_{t+\tau}) = p^2 \beta^2 \phi^{|\tau|}.$$

Therefore,

$$\log(\sigma_t^p) + \log(\sigma_{t+\tau}^p) \sim \mathcal{N}(2p\alpha, 2p^2\beta^2(1 + \phi^{|\tau|})).$$

Using moment generating function and $\log A + \log B = \log AB$ we have

$$\mathbb{E}(\sigma_t^p \sigma_{t+\tau}^p) = \exp(2p\alpha + p^2\beta^2(1 + \phi^{|\tau|})).$$

Over-Identifying Restrictions Test

To test whether the model for $\mathbb{E}[\mathbf{m}(\mathbf{z}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$ is correctly specified, it is natural to check if $\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T)$ is sufficiently close to zero. For example, when the moment conditions are the Euler equations in different capital asset pricing models, this amounts to checking if the “pricing errors” are zero.

The **over-identifying restrictions** (OIR) test of Hansen (1982), also known as the \mathcal{J} test, is based on the value of the GMM objective function:

$$\mathcal{J}_T(\mathbf{W}_T) := T [\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T))]'\mathbf{W}_T [\bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T))].$$

This test is quite unusual, because the test statistic involves the GMM estimator obtained from the same objective function.

To derive its limiting distribution, note that

$$\begin{aligned}\mathbf{W}_T^{1/2} \sqrt{T} \bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T)) \\ = \mathbf{W}_T^{1/2} \sqrt{T} \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) + \mathbf{W}_T^{1/2} \mathbf{G}_T(\boldsymbol{\theta}_T^\dagger) \sqrt{T} (\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T) - \boldsymbol{\theta}_o).\end{aligned}$$

As $\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T) - \boldsymbol{\theta}_o) = -(\mathbf{G}_o' \mathbf{W}_o \mathbf{G}_o)^{-1} \mathbf{G}_o' \mathbf{W}_o \sqrt{T} \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) + o_P(1)$,

$$\mathbf{W}_T^{1/2} \sqrt{T} \bar{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T(\mathbf{W}_T)) = \mathbf{P}_o \mathbf{W}_o^{1/2} \sqrt{T} \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) + o_P(1)$$

where $\mathbf{P}_o = \mathbf{I} - \mathbf{W}_o^{1/2} \mathbf{G}_o (\mathbf{G}_o' \mathbf{W}_o \mathbf{G}_o)^{-1} \mathbf{G}_o' \mathbf{W}_o^{1/2}$ which is symmetric and idempotent with rank $q - k$ (why?), and

$$\mathbf{W}_o^{1/2} \sqrt{T} \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{W}_o^{1/2} \boldsymbol{\Sigma}_o \mathbf{W}_o^{1/2}).$$

When $\mathbf{W}_o = \boldsymbol{\Sigma}_o^{-1}$, $\mathbf{W}_o^{1/2} \sqrt{T} \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$. Consequently,

$$\mathcal{J}_T(\mathbf{W}_T) = T \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o)' \mathbf{W}_o^{1/2} \mathbf{P}_o \mathbf{P}_o \mathbf{W}_o^{1/2} \bar{\mathbf{m}}_T(\boldsymbol{\theta}_o) \xrightarrow{D} \chi^2(q - k).$$

The Limiting Distribution of the OIR Test

Let $\hat{\Sigma}_T$ be a consistent estimator of Σ_o . Then,

$$\mathcal{J}_T(\hat{\Sigma}_T^{-1}) := T [\bar{\mathbf{m}}_T(\hat{\theta}_T(\hat{\Sigma}_T^{-1}))]' \hat{\Sigma}_T^{-1} [\bar{\mathbf{m}}_T(\hat{\theta}_T(\hat{\Sigma}_T^{-1}))] \xrightarrow{D} \chi^2(q - k),$$

where $\hat{\theta}_T(\hat{\Sigma}_T^{-1})$ is the optimal two-step GMM estimator.

Remarks:

- The weighting matrix in the $\mathcal{J}_T(\hat{\Sigma}_T^{-1})$ statistic and the weighting matrix for the GMM estimator in $\bar{\mathbf{m}}_T$ must be the **same**. That is, the OIR test requires the optimal two-step GMM estimator.
- Lee and Kuan (2010) propose an OIR test that does **not** requires the weighting matrix to converge to Σ_o and hence avoids the optimal GMM estimation.

Hausman Test

Given two estimators $\hat{\theta}_T$ and $\check{\theta}_T$ of the parameter θ_o , suppose that both are consistent under the null hypothesis of correct model specification, but only one, say $\check{\theta}_T$, is also consistent under the alternative. The **Hausman** test suggests testing the null hypothesis by comparing these two estimators.

The Hausman test is particularly useful for testing the model specification that can **not** be expressed as parameter restrictions. For example, consider the null hypothesis of exogenous regressors and the alternative of endogenous regressors. Under “classical” conditions, the OLS estimator $\hat{\theta}_T$ and the 2SLS estimator $\check{\theta}_T$ are consistent under the null, but only the 2SLS estimator is consistent under the alternative. Thus, we can test for **endogeneity** by checking if $\hat{\theta}_T$ and $\check{\theta}_T$ are sufficiently close to each other.

The Hausman test reads:

$$\mathcal{H}_T = T(\hat{\theta}_T - \check{\theta}_T)' \widehat{\mathbf{V}}_T^{-1} (\hat{\theta}_T - \check{\theta}_T) \xrightarrow{D} \chi^2(k),$$

where $\widehat{\mathbf{V}}_T$ is a consistent estimator for the asymptotic covariance matrix of $\sqrt{T}(\hat{\theta}_T - \check{\theta}_T)$:

$$\mathbf{V}(\hat{\theta}_T - \check{\theta}_T) = \mathbf{V}(\hat{\theta}_T) + \mathbf{V}(\check{\theta}_T) - 2\text{cov}(\hat{\theta}_T, \check{\theta}_T).$$

This asymptotic covariance matrix is simplified when $\hat{\theta}_T$ is also **asymptotically efficient** under the null. In this case,

$$\mathbf{V}_{1,2} := \text{cov}(\hat{\theta}_T, \check{\theta}_T) = \mathbf{V}(\hat{\theta}_T),$$

so that $\mathbf{V}(\hat{\theta}_T - \check{\theta}_T)$ depends only on the respective asymptotic covariance matrices of these two estimators:

$$\mathbf{V}(\hat{\theta}_T - \check{\theta}_T) = \mathbf{V}(\check{\theta}_T) - \mathbf{V}(\hat{\theta}_T).$$

Note that

$$\text{cov}(\check{\theta}_T - \hat{\theta}_T, \hat{\theta}_T) = \mathbf{V}_{1,2} - \mathbf{V}(\hat{\theta}_T).$$

If this covariance is not zero, we may combine $\check{\theta}_T - \hat{\theta}_T$ and $\hat{\theta}_T$ to form a new estimator that is more efficient than $\hat{\theta}_T$. Consider the new estimator

$$\hat{\theta}_T^\dagger = \hat{\theta}_T + [\mathbf{V}(\hat{\theta}_T) - \mathbf{V}_{1,2}] \mathbf{V}(\hat{\theta}_T - \check{\theta}_T)^{-1} (\check{\theta}_T - \hat{\theta}_T),$$

with the variance:

$$\mathbf{V}(\hat{\theta}_T^\dagger) = \mathbf{V}(\hat{\theta}_T) - [\mathbf{V}(\hat{\theta}_T) - \mathbf{V}_{1,2}] \mathbf{V}(\hat{\theta}_T - \check{\theta}_T)^{-1} [\mathbf{V}(\hat{\theta}_T) - \mathbf{V}_{1,2}]'.$$

When $\mathbf{V}(\hat{\theta}_T)$ is not the same as $\mathbf{V}_{1,2}$, the second term on the RHS is p.d., so that $\hat{\theta}_T^\dagger$ is asymptotically more efficient than $\hat{\theta}_T$. This contradicts the assumption that $\hat{\theta}_T$ is asymptotically efficient. Therefore, we must have $\mathbf{V}(\hat{\theta}_T) = \mathbf{V}_{1,2}$.

Wald Test

Hypothesis: $\mathbf{R}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{R} : \mathbb{R}^K \mapsto \mathbb{R}^r$. A mean-value expansion yields

$$\mathbf{R}(\hat{\boldsymbol{\theta}}_T) = \mathbf{R}(\boldsymbol{\theta}_0) + \nabla \mathbf{R}(\boldsymbol{\theta}^\dagger)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

where $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_0)$. Therefore,

$$\sqrt{T}[\mathbf{R}(\hat{\boldsymbol{\theta}}_T) - \mathbf{R}(\boldsymbol{\theta}_0)] \xrightarrow{D} \mathcal{N}(\mathbf{0}, [\nabla \mathbf{R}(\boldsymbol{\theta}_0)]\boldsymbol{\Omega}_0[\nabla \mathbf{R}(\boldsymbol{\theta}_0)]').$$

It follows that, under the null hypothesis,

$$\mathcal{W}_T := T\mathbf{R}(\hat{\boldsymbol{\theta}}_T)' \left([\nabla \mathbf{R}(\hat{\boldsymbol{\theta}}_T)]\hat{\boldsymbol{\Omega}}_T[\nabla \mathbf{R}(\hat{\boldsymbol{\theta}}_T)]' \right)^{-1} \mathbf{R}(\hat{\boldsymbol{\theta}}_T) \xrightarrow{D} \chi^2(r).$$

where $\hat{\boldsymbol{\Omega}}_T$ is a consistent estimator of $\boldsymbol{\Omega}_0$. For the linear hypothesis $\mathbf{R}\boldsymbol{\theta}_0 = \mathbf{r}$ with \mathbf{R} a $r \times k$ matrix,

$$\mathcal{W}_T = T(\mathbf{R}\hat{\boldsymbol{\theta}}_T - \mathbf{r})'(\mathbf{R}\hat{\boldsymbol{\Omega}}_T\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\theta}}_T - \mathbf{r}) \xrightarrow{D} \chi^2(r).$$

Estimation by GMM

Suppose that an economic model yields some conditional moment restrictions. That is, there exists unique θ_o such that

$$\mathbb{E}[\mathbf{h}(\eta_t; \theta_o) | \mathcal{F}^t] = \mathbf{0},$$

where \mathcal{F}^t is the information set up to time t , η_t is not \mathcal{F}^t -measurable, and \mathbf{h} is $r \times 1$. To estimate θ_o , consider the **implied**, unconditional moment conditions:

$$\mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{h}(\eta_t; \theta_o)] = \mathbf{0},$$

where \mathbf{w}_t is a set of variables from \mathcal{F}^t and $\mathbf{D}(\mathbf{w}_t)$ is a $(r \times n)$ matrix of **instruments** based on \mathbf{w}_t and has full rank. Then, we can estimate θ_o by applying the GMM to the sample moment functions:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{m}_t(\theta) = \frac{1}{T} \sum_{t=1}^T \mathbf{D}(\mathbf{w}_t)' \mathbf{h}(\eta_t; \theta).$$

Q: How do we choose \mathbf{w}_t and $\mathbf{D}(\mathbf{w}_t)$?

The Number of Instruments

The optimal GMM estimator based on the instruments $\mathbf{D}(\mathbf{w}_t)$ has the asymptotic covariance matrix: $(\mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o)^{-1}$, with

$$\boldsymbol{\Sigma}_o = \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{h}(\eta_t; \theta_o) \mathbf{h}(\eta_t; \theta_o)' \mathbf{D}(\mathbf{w}_t)], \quad (n \times n)$$

$$\mathbf{G}_o = \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \nabla \mathbf{h}(\eta_t; \theta_o)]. \quad (n \times k)$$

Consider another optimal GMM estimator based on a smaller set of instruments $\mathbf{D}(\mathbf{w}_t)\mathbf{C}$, with \mathbf{C} an $n \times p$ ($p < n$) matrix. For example, \mathbf{C} may be a selection matrix $[\mathbf{I}_p \ \mathbf{0}]$. Such estimator is obtained from the sample moment conditions:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{C}' \mathbf{D}(\mathbf{w}_t)' \mathbf{h}(\eta_t; \theta),$$

and has the asymptotic covariance matrix:

$$[\mathbf{G}'_o \mathbf{C} (\mathbf{C}' \boldsymbol{\Sigma}_o \mathbf{C})^{-1} \mathbf{C}' \mathbf{G}_o]^{-1}.$$

The difference between these two asymptotic covariance matrices is

$$\begin{aligned} \mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o - \mathbf{G}'_o \mathbf{C} (\mathbf{C}' \boldsymbol{\Sigma}_o \mathbf{C})^{-1} \mathbf{C}' \mathbf{G}_o \\ = \mathbf{G}'_o \boldsymbol{\Sigma}_o^{-1/2} [\mathbf{I}_n - \boldsymbol{\Sigma}_o^{1/2} \mathbf{C} (\mathbf{C}' \boldsymbol{\Sigma}_o^{1/2} \boldsymbol{\Sigma}_o^{1/2} \mathbf{C})^{-1} \mathbf{C}' \boldsymbol{\Sigma}_o^{1/2}] \boldsymbol{\Sigma}_o^{-1/2} \mathbf{G}_o, \end{aligned}$$

which is p.s.d., because the term in the square bracket is symmetric and idempotent. Thus, the optimal GMM estimator based on a larger set of moment conditions is asymptotically more efficient than that based on a subset of moment conditions. This suggests that, as far as GMM efficiency is concerned, **more** instruments would be preferred. It has been found, however, that the estimator based on more instruments tends to have a **larger bias** in finite samples.

Optimal Instruments

Consider the matrix of k instruments:

$$\mathbf{D}^*(\mathbf{w}_t; \boldsymbol{\theta}_o) = \underbrace{\left[\text{var}(\mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) | \mathcal{F}^t) \right]^{-1}}_{\mathbf{V}_t} \underbrace{\mathbb{E}[\nabla \mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) | \mathcal{F}^t]}_{\mathbf{J}_t}, \quad (r \times k)$$

In this case,

$$\begin{aligned}\boldsymbol{\Sigma}_o &= \mathbb{E}(\mathbf{J}_t' \mathbf{V}_t^{-1} \mathbf{V}_t \mathbf{V}_t^{-1} \mathbf{J}_t) = \mathbb{E}(\mathbf{J}_t' \mathbf{V}_t^{-1} \mathbf{J}_t), \\ \mathbf{G}_o &= \mathbb{E}[\mathbf{J}_t' \mathbf{V}_t^{-1} \nabla \mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)] = \mathbb{E}(\mathbf{J}_t' \mathbf{V}_t^{-1} \mathbf{J}_t).\end{aligned}$$

The asymptotic covariance matrix of the optimal GMM estimator based on the instruments $\mathbf{D}^*(\mathbf{w}_t; \boldsymbol{\theta}_o)$ is

$$(\mathbf{G}_o' \boldsymbol{\Sigma}_o^{-1} \mathbf{G}_o)^{-1} = \mathbb{E}(\mathbf{J}_t' \mathbf{V}_t^{-1} \mathbf{J}_t)^{-1}.$$

For the GMM estimator based on the instruments $\mathbf{D}(\mathbf{w}_t)$, the asymptotic covariance matrix is

$$\begin{aligned} & \left\{ \mathbb{E}[\nabla \mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)' \mathbf{D}(\mathbf{w}_t)] \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o) \mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)' \mathbf{D}(\mathbf{w}_t)]^{-1} \right. \\ & \quad \left. \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \nabla \mathbf{h}(\boldsymbol{\eta}_t; \boldsymbol{\theta}_o)] \right\}^{-1} \\ &= \left\{ \mathbb{E}[\mathbf{J}_t' \mathbf{D}(\mathbf{w}_t)] \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{V}_t \mathbf{D}(\mathbf{w}_t)]^{-1} \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{J}_t] \right\}^{-1}. \end{aligned}$$

Given the difference below:

$$\mathbb{E}(\mathbf{J}_t' \mathbf{V}_t^{-1} \mathbf{J}_t) - \mathbb{E}[\mathbf{J}_t' \mathbf{D}(\mathbf{w}_t)] \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{V}_t \mathbf{D}(\mathbf{w}_t)]^{-1} \mathbb{E}[\mathbf{D}(\mathbf{w}_t)' \mathbf{J}_t],$$

it is readily verified that its sample counterpart is p.s.d. Passing to the limit, we can conclude that the optimal GMM estimator based on the instruments $\mathbf{D}^*(\mathbf{w}_t; \boldsymbol{\theta}_o)$ has the smallest asymptotic covariance matrix $\mathbb{E}(\mathbf{J}_t' \mathbf{V}_t^{-1} \mathbf{J}_t)^{-1}$. $\mathbf{D}^*(\mathbf{w}_t; \boldsymbol{\theta}_o)$ is thus the matrix of **optimal instruments**. As $\mathbf{D}^*(\mathbf{w}_t; \boldsymbol{\theta}_o)$ contains k instruments, the implied moment conditions are exactly identified.

- Suppose the conditional moment restriction is $\mathbb{E}(y_t - \mathbf{x}'_t \beta_o | \mathcal{F}^t) = 0$. Letting $\varepsilon_t = y_t - \mathbf{x}'_t \beta_o$ and $\sigma_{\varepsilon_t}^2 = \text{var}(\varepsilon_t | \mathcal{F}^t)^{-1}$, the optimal instruments are $\sigma_{\varepsilon_t}^{-2} \mathbf{x}'_t$. The implied unconditional moments are

$$\mathbb{E}[\sigma_{\varepsilon_t}^{-2} \mathbf{x}_t (y_t - \mathbf{x}'_t \beta_o)] = \mathbf{0}.$$

This is precisely the first order condition of a weighted least squares objective function (also the GLS objective function), as it ought to be.

- In practice, the optimal instruments are unknown and need to be estimated. Estimating the optimal instruments is cumbersome because \mathbf{J}_t and \mathbf{V}_t are conditional expectations of unknown form.
- A common choice of the instruments $\mathbf{D}(\mathbf{w}_t)$ is a low-order polynomial in the elements of \mathbf{w}_t . With arbitrary instruments, there is **no** guarantee that the parameter θ_o is still identified in the implied, unconditional moment restrictions; see Domínguez and Lobato (2004, *Econometrica*) and Hsu and Kuan (2011, *J. of Econometrics*) for details.

Capital Asset Pricing Model

Cochrane (2005)

- Consider a utility function of the current and future consumptions:

$$U(c_t, c_{t+1}) = u(c_t) + \mathbb{E}[\beta u(c_{t+1}) | \mathcal{F}^t],$$

where β is the subjective discount factor. A common choice of u is

$$u(c_t) := \frac{c_t^{1-\gamma} - 1}{1-\gamma}.$$

Note that $u(c) \approx c^{1-\gamma} \ln c \rightarrow \ln c$ when $\gamma \rightarrow 1$.

- Writing $\mathbb{E}(\cdot | \mathcal{F}^t)$ as $\mathbb{E}_t(\cdot)$, an investor's choice problem is

$$\max_{q_t} u(c_t) + \mathbb{E}_t[\beta u(c_{t+1})],$$

$$\text{s.t. } c_t = e_t - p_t q_t, \quad c_{t+1} = e_{t+1} + x_{t+1} q_t,$$

where e is the endowment, and q, p, x are, respectively, the quantity, price, and payoff of the asset.

- The Euler equation (first-order condition) yields

$$p_t u'(c_t) = \mathbb{E}_t[\beta u'(c_{t+1})x_{t+1}],$$

where $p_t u'(c_t)$ is the marginal loss in utility from buying one more unit of asset at time t , whereas $\mathbb{E}_t[\beta u'(c_{t+1})x_{t+1}]$ is the increase in the discounted utility of getting extra payoff at time $t + 1$. Thus, the asset choice, and hence the optimal consumption, is determined by the equality of marginal gain and marginal loss.

- We can also write

$$p_t = \mathbb{E}_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right].$$

This is the basic asset pricing formula which indicates, given the consumption choices of c_t and c_{t+1} , what the market price p_t should be.

- Equivalently, $p_t = \mathbb{E}_t(\psi_{t+1}x_{t+1})$, where

$$\psi_{t+1} := \beta \frac{u'(c_{t+1})}{u'(c_t)},$$

is also known as the **stochastic discount factor** (or **pricing kernel**, **marginal rate of substitution**).

- The basic pricing formula is

$$p_t = \mathbb{E}_t(\psi_{t+1}x_{t+1}) \quad \text{or} \quad 1 = \mathbb{E}_t(\psi_{t+1}r_{t+1}),$$

where $r_{t+1} = x_{t+1}/p_t$ is the gross return.

- The pricing formula is a conditional moment restriction, in which the SDF, ψ_{t+1} , contains unknown parameters. For example, ψ may depend on the risk adverse parameter γ when $u(c_t) := \frac{c_t^{1-\gamma}-1}{1-\gamma}$, and ψ may depend on γ and the habit persistent parameter δ when

$$u(c_t) = \frac{(c_t + \delta c_{t-1})^{1-\gamma} - 1}{1-\gamma}.$$

In our notations,

$$\mathbb{E}_t[\psi_{t+1}(\theta_o)x_{t+1} - p_t] = \mathbb{E}_t[h_{t+1}(\eta_t; \theta_o)] = 0,$$

where $h_{t+1}(\theta_o)$ is the **pricing error** or **test function**.

Examples of Test Asset

	price (p_t)	payoff (x_{t+1})
Stock	p_t	$p_{t+1} + d_{t+1}$
Return	1	r_{t+1}
Price-dividend ratio	$\frac{p_t}{d_t}$	$(\frac{p_{t+1}}{d_{t+1}} + 1)\frac{d_{t+1}}{d_t}$
Excess return	0	$r_{t+1}^e = r_{t+1}^a - r_{t+1}^b$
One-period bond	p_t	1
Risk-free rate	1	r_f
Option	C	$\max(S_T - K, 0)$

- The pricing formula for stock is

$$p_t = \mathbb{E}_t[\psi_{t+1}(d_{t+1} + p_{t+1})].$$

As GMM requires stationarity of data, it would be better to consider

$$1 = \mathbb{E}_t[\psi_{t+1}(d_{t+1} + p_{t+1})/p_t],$$

where the stock return $(d_{t+1} + p_{t+1})/p_t$ is more likely to be stationary.

- The pricing error for stock is then

$$\mathbb{E}_t\left[\psi_{t+1}(\theta_o)\frac{d_{t+1} + p_{t+1}}{p_t} - 1\right] = \mathbb{E}_t[h(\eta_{t+1}; \theta_o)] = 0.$$

- The implied unconditional moment restrictions are

$\mathbb{E}[\mathbf{D}(\mathbf{w}_t)'h(\eta_{t+1}; \theta_o)] = \mathbf{0}$, where $\mathbf{D}(\mathbf{w}_t)$ is in \mathcal{F}^t , and the GMM estimation of θ_o is based on the sample moments:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{D}(\mathbf{w}_t)'h(\eta_{t+1}; \theta).$$

Some Examples

- Specifying a utility function: $u(c_t) := \frac{c_t^{1-\gamma} - 1}{1-\gamma}$, $0 < \gamma \neq 1$. Then,

$$h(\eta_{t+1}; \theta) = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} r_{t+1} - 1.$$

- Specifying a pricing kernel: A nonlinear kernel of Dittmar (2002, JF):

$$\psi_{t+1}(\eta_{t+1}; \theta) := d_0 + d_1 r_{W,t+1} + d_2 r_{W,t+1}^2 + d_3 r_{W,t+1}^3,$$

where $r_{W,t+1}$ represents the return on end-of-period aggregate wealth. We may write the pricing error $\mathbb{E}_t(\psi_{t+1} r_{t+1} - 1)$ as

$$\mathbb{E}_t(r_{t+1}) = \frac{1}{\mathbb{E}_t(\psi_{t+1})} - \text{cov}_t(r_{t+1}, \psi_{t+1}) \frac{1}{\mathbb{E}_t(\psi_{t+1})}.$$

Note that with the nonlinear pricing kernel, the terms $\text{cov}_t(r_{t+1}, r_{W,t+1}^j)$, $j = 1, 2, 3$, are consistent with higher-order moment CAPM.

Cox, Ingersoll and Ross (CIR) Model

- Cox et al. (1985, *Econometrica*) assume the instantaneous short rate follows the diffusion:

$$dr_t = \kappa(\bar{r} - r_t) dt + \sigma\sqrt{r_t} dW_t.$$

Thus, r_t is **mean reverting** to the long-run level \bar{r} and has conditional volatility $\sigma\sqrt{r_t}$.

- The conditional density of r_{t+1} given r_t is

$$f(r_{t+1}|r_t) = c \exp(-u_t - v_{t+1}) \left(\frac{n_{t+1}}{u_t} \right)^{q/2} I_q(2(y_t v_{t+1})^{1/2}),$$

where $c = 2\kappa/[\sigma^2(1 - e^{-\kappa})]$, $u_t = ce^{-\kappa}r_t$, $v_{t+1} = cr_{t+1}$, $q = 2\kappa\bar{r}/\sigma^2 - 1$, and I_q is the modified Bessel function of the first kind of order q . This is a non-central χ^2 distribution with $2q + 2$ degrees of freedom and the non-centrality parameter $2u_t$.

- From the density above, we can integrate to obtain the conditional mean:

$$\mathbb{E}(r_{t+\Delta}|r_t) = r_t e^{-\Delta\kappa} + \bar{r}(1 - e^{-\Delta\kappa}).$$

For $\Delta = 1$, we have the conditional moment restriction:

$$\mathbb{E}_t[r_{t+1} - r_t e^{-\kappa} - \bar{r}(1 - e^{-\kappa})] = 0.$$

- With the instruments $\mathbf{D}(\mathbf{w}_t)$, the implied, unconditional moment conditions are

$$\mathbb{E}[(r_{t+1} - r_t e^{-\kappa} - \bar{r}(1 - e^{-\kappa}))\mathbf{D}(\mathbf{w}_t)] = \mathbf{0}.$$

Then, $\theta = (\bar{r}, \kappa)$ can be estimated based on the sample moment functions:

$$\frac{1}{T} \sum_{t=1}^T h_{t+1}(\theta) = \frac{1}{T} \sum_{t=1}^T [r_{t+1} - r_t e^{-\kappa} + \bar{r}(1 - e^{-\kappa})]\mathbf{D}(\mathbf{w}_t).$$

- Similarly, we can compute the conditional variance as

$$\text{var}_t(r_{t+1}) = r_t \frac{\sigma}{\kappa} \left(e^{-\kappa} - e^{-2\kappa} \right) + \bar{r} \frac{\sigma^2}{2\kappa} (1 - e^{-\kappa})^2.$$

Hence, $\mathbb{E}_t(r_{t+1}^2) = \text{var}_t(r_{t+1}) + (\mathbb{E}_t(r_{t+1}))^2$.

- With the instruments $\mathbf{D}(\mathbf{w}_t)$, the additional sample moment functions are

$$\frac{1}{T} \sum_{t=1}^T h_{t+1}(\theta) = \frac{1}{T} \sum_{t=1}^T [r_{t+1}^2 - \text{var}_t(r_{t+1}) + (\mathbb{E}_t(r_{t+1}))^2] \mathbf{D}(\mathbf{w}_t),$$

where we need to replace $\text{var}_t(r_{t+1})$ and $\mathbb{E}_t(r_{t+1})$ by their functional forms given above.