GMM Estimation of Asset Pricing Models

Econ 643: Financial Economics II

Nikolay Gospodinov
Department of Economics, Concordia University

March 16, 2011

Linear Instrumental Variables

- Consider the model

$$y = X\theta + \varepsilon$$

where y is a $T \times 1$ vector, X is a $T \times k$ matrix, θ is a $k \times 1$ parameter vector and ε is a $T \times 1$ vector of errors.

Linear Instrumental Variables

- Consider the model

$$y = X\theta + \varepsilon$$

where y is a $T \times 1$ vector, X is a $T \times k$ matrix, θ is a $k \times 1$ parameter vector and ε is a $T \times 1$ vector of errors.

- For the t^{th} observation, the model is given by

$$y_t = x_t'\theta + \varepsilon_t.$$

Linear Instrumental Variables

- Consider the model
$$y = X\theta + \varepsilon$$
where y is a $T \times 1$ vector, X is a $T \times k$ matrix, θ is a $k \times 1$ parameter vector and ε is a $T \times 1$ vector of errors.
- For the t^{th} observation, the model is given by
$$y_t = x_t'\theta + \varepsilon_t.$$
- Suppose that $E(\varepsilon_t|x_t) \neq 0$ but there is a $T \times m$ matrix of instrumental variables Z such that $E\left(\varepsilon_t|z_t\right) = 0$.

Linear Instrumental Variables

- Consider the model
$$y = X\theta + \varepsilon$$
where y is a $T \times 1$ vector, X is a $T \times k$ matrix, θ is a $k \times 1$ parameter vector and ε is a $T \times 1$ vector of errors.
- For the t^{th} observation, the model is given by
$$y_t = x_t'\theta + \varepsilon_t.$$
- Suppose that $E(\varepsilon_t|x_t) \neq 0$ but there is a $T \times m$ matrix of instrumental variables Z such that $E(\varepsilon_t|z_t) = 0$.
- By the law of iterated expectations (LIE),
$$E(\varepsilon_t z_t) = E[E(\varepsilon_t z_t|z_t)] = E[E(\varepsilon_t|z_t) z_t] = 0$$
since $E(\varepsilon_t|z_t) = 0$ by assumption.

Linear Instrumental Variables

- Consider the model
$$y = X\theta + \varepsilon$$
where y is a $T \times 1$ vector, X is a $T \times k$ matrix, θ is a $k \times 1$ parameter vector and ε is a $T \times 1$ vector of errors.
- For the t^{th} observation, the model is given by
$$y_t = x_t'\theta + \varepsilon_t.$$
- Suppose that $E(\varepsilon_t | x_t) \neq 0$ but there is a $T \times m$ matrix of instrumental variables Z such that $E\left(\varepsilon_t | z_t\right) = 0$.
- By the law of iterated expectations (LIE),
$$E\left(\varepsilon_t z_t\right) = E\left[E\left(\varepsilon_t z_t | z_t\right)\right] = E\left[E\left(\varepsilon_t | z_t\right) z_t\right] = 0$$
since $E\left(\varepsilon_t | z_t\right) = 0$ by assumption.
- This will be used to construct some moment (orthogonality) conditions.

Linear Instrumental Variables

- If we define $g_t(\theta) = \varepsilon_t(\theta) z_t$, then $E[g_t(\theta_0)] = 0$.

Linear Instrumental Variables

- If we define $g_t(\theta) = \varepsilon_t(\theta) z_t$, then $E[g_t(\theta_0)] = 0$.
- Since the population expectation of $g_t(\theta)$ is not observed, we replace it by its sample analog

$$\begin{aligned} g_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \\ &= \frac{1}{T}\sum_{t=1}^{T} z_t \varepsilon_t(\theta) \\ &= \frac{1}{T} Z' \varepsilon(\theta) \end{aligned}$$

where the subscript T indicates dependence on the sample.

Linear Instrumental Variables

- If we define $g_t(\theta) = \varepsilon_t(\theta) z_t$, then $E[g_t(\theta_0)] = 0$.
- Since the population expectation of $g_t(\theta)$ is not observed, we replace it by its sample analog

$$\begin{aligned} g_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \\ &= \frac{1}{T}\sum_{t=1}^{T} z_t \varepsilon_t(\theta) \\ &= \frac{1}{T} Z' \varepsilon(\theta) \end{aligned}$$

  where the subscript T indicates dependence on the sample.
- Note that g is an $m \times 1$ vector ($g(\theta) : \mathbb{R}^k \to \mathbb{R}^m$), where $m \geq k$.

Linear Instrumental Variables: Just-Identified Case

- If $m = k$ (just-identified case), we have k equations (moment condition) with k unknowns

$$\frac{1}{T}\sum_{t=1}^{T} z_t \left(y_t - x_t'\theta\right) = 0$$

or, in matrix form,

$$\frac{1}{T}\left(Z'Y - Z'X\theta\right) = 0.$$

Linear Instrumental Variables: Just-Identified Case

- If $m = k$ (just-identified case), we have k equations (moment condition) with k unknowns

$$\frac{1}{T}\sum_{t=1}^{T} z_t\left(y_t - x_t'\theta\right) = 0$$

or, in matrix form,

$$\frac{1}{T}\left(Z'Y - Z'X\theta\right) = 0.$$

- The solution to this system of equations is given by the simple instrumental variables (IV) estimator

$$\widehat{\theta}_{IV} = \left(Z'X\right)^{-1}\left(Z'Y\right).$$

Linear Instrumental Variables: Just-Identified Case

- If $m = k$ (just-identified case), we have k equations (moment condition) with k unknowns

$$\frac{1}{T}\sum_{t=1}^{T} z_t \left(y_t - x_t'\theta\right) = 0$$

or, in matrix form,

$$\frac{1}{T}\left(Z'Y - Z'X\theta\right) = 0.$$

- The solution to this system of equations is given by the simple instrumental variables (IV) estimator

$$\widehat{\theta}_{IV} = \left(Z'X\right)^{-1}\left(Z'Y\right).$$

  - If $Z = X$ (i.e. if the moment restriction is $E\left(\varepsilon_t | x_t\right) = 0$), we have the OLS estimator $\widehat{\theta}_{OLS} = (X'X)^{-1}(X'Y)$.

Linear Instrumental Variables: Over-Identified Case

- In general, the number of moment conditions may exceed the number of parameters ($m > k$) (over-identified case).

Linear Instrumental Variables: Over-Identified Case

- In general, the number of moment conditions may exceed the number of parameters ($m > k$) (over-identified case).
- In this case, $g_T(\theta) = 0$ cannot be solved exactly. Instead, we want to find an estimator that makes the moment conditions as close to zero as possible by minimizing the quadratic form.

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$$

where W_T is a $m \times m$ symmetric, positive definite weighting matrix.

Linear Instrumental Variables: Over-Identified Case

- In general, the number of moment conditions may exceed the number of parameters ($m > k$) (over-identified case).
- In this case, $g_T(\theta) = 0$ cannot be solved exactly. Instead, we want to find an estimator that makes the moment conditions as close to zero as possible by minimizing the quadratic form.

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$$

  where W_T is a $m \times m$ symmetric, positive definite weighting matrix.
- More specifically,

$$Q_T(\theta) = (Y'Z - \theta' X' Z) W_T (Z'Y - Z'X\theta).$$

Linear Instrumental Variables: Over-Identified Case

- In general, the number of moment conditions may exceed the number of parameters ($m > k$) (over-identified case).
- In this case, $g_T(\theta) = 0$ cannot be solved exactly. Instead, we want to find an estimator that makes the moment conditions as close to zero as possible by minimizing the quadratic form.

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$$

where W_T is a $m \times m$ symmetric, positive definite weighting matrix.

- More specifically,

$$Q_T(\theta) = (Y'Z - \theta' X' Z) W_T (Z'Y - Z'X\theta).$$

- Solving the FOC (k equations in k unknowns)

$$\frac{\partial}{\partial \theta} Q_T(\theta) = -2X'ZW_T(Z'Y - Z'X\theta) = 0$$

gives the generalized IV estimator

$$\widehat{\theta}_{IV} = \left(X'ZW_T Z'X\right)^{-1} X'ZW_T Z'Y.$$

Linear Instrumental Variables: 2SLS Estimator

- Now assume that $E(\varepsilon_t^2 | z_t) = \sigma^2$ (homoskedasticity) and $E\left(\varepsilon_t \varepsilon_{t-s}\right) = 0$ for $s \neq t$ (no serial correlation). Then, $Var\left[T^{-1/2} Z' \varepsilon\left(\theta\right)\right] = \sigma^2 \left(Z'Z\right) \equiv V$.

Linear Instrumental Variables: 2SLS Estimator

- Now assume that $E(\varepsilon_t^2|z_t) = \sigma^2$ (homoskedasticity) and $E\left(\varepsilon_t\varepsilon_{t-s}\right) = 0$ for $s \neq t$ (no serial correlation). Then, $Var\left[T^{-1/2}Z'\varepsilon\left(\theta\right)\right] = \sigma^2\left(Z'Z\right) \equiv V$.
- The optimal weighting matrix turns out to be $W_T = V^{-1}$ or $W_T = \left(Z'Z\right)^{-1}$ which downweighs noisy orthogonality conditions.

Linear Instrumental Variables: 2SLS Estimator

- Now assume that $E(\varepsilon_t^2 | z_t) = \sigma^2$ (homoskedasticity) and $E(\varepsilon_t \varepsilon_{t-s}) = 0$ for $s \neq t$ (no serial correlation). Then, $Var\left[T^{-1/2} Z' \varepsilon(\theta)\right] = \sigma^2 (Z'Z) \equiv V$.
- The optimal weighting matrix turns out to be $W_T = V^{-1}$ or $W_T = (Z'Z)^{-1}$ which downweighs noisy orthogonality conditions.
- Plugging $W_T = (Z'Z)^{-1}$ into the IV estimator above gives the two-stage least squares (2SLS) estimator

$$\begin{aligned} \widehat{\theta}_{2SLS} &= \left(X'Z (Z'Z)^{-1} Z'X\right)^{-1} X'Z (Z'Z)^{-1} Z'Y \\ &= (X'P_Z X)^{-1} X'P_Z Y, \end{aligned}$$

where $P_Z = Z (Z'Z)^{-1} Z'$ is the projection matrix that projects orthogonally onto the space spanned by the columns of Z.

- Now assume that $E(\varepsilon_t^2|z_t) = \sigma^2$ (homoskedasticity) and $E(\varepsilon_t \varepsilon_{t-s}) = 0$ for $s \neq t$ (no serial correlation). Then, $Var\left[T^{-1/2} Z'\varepsilon(\theta)\right] = \sigma^2 (Z'Z) \equiv V$.
- The optimal weighting matrix turns out to be $W_T = V^{-1}$ or $W_T = (Z'Z)^{-1}$ which downweighs noisy orthogonality conditions.
- Plugging $W_T = (Z'Z)^{-1}$ into the IV estimator above gives the two-stage least squares (2SLS) estimator

$$\begin{aligned} \widehat{\theta}_{2SLS} &= \left(X'Z (Z'Z)^{-1} Z'X\right)^{-1} X'Z (Z'Z)^{-1} Z'Y \\ &= (X'P_Z X)^{-1} X'P_Z Y, \end{aligned}$$

where $P_Z = Z (Z'Z)^{-1} Z'$ is the projection matrix that projects orthogonally onto the space spanned by the columns of Z.
- This estimation method can be thought of as a two-stage procedure

Linear Instrumental Variables: 2SLS Estimator

- Now assume that $E(\varepsilon_t^2|z_t) = \sigma^2$ (homoskedasticity) and $E(\varepsilon_t \varepsilon_{t-s}) = 0$ for $s \neq t$ (no serial correlation). Then, $Var\left[T^{-1/2} Z'\varepsilon(\theta)\right] = \sigma^2 (Z'Z) \equiv V$.
- The optimal weighting matrix turns out to be $W_T = V^{-1}$ or $W_T = (Z'Z)^{-1}$ which downweighs noisy orthogonality conditions.
- Plugging $W_T = (Z'Z)^{-1}$ into the IV estimator above gives the two-stage least squares (2SLS) estimator

$$\begin{aligned} \widehat{\theta}_{2SLS} &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y \\ &= (X'P_Z X)^{-1} X'P_Z Y, \end{aligned}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix that projects orthogonally onto the space spanned by the columns of Z.

- This estimation method can be thought of as a two-stage procedure
  1. regress X on Z and get the fitted value $\widehat{X}$

Linear Instrumental Variables: 2SLS Estimator

- Now assume that $E(\varepsilon_t^2|z_t) = \sigma^2$ (homoskedasticity) and $E(\varepsilon_t \varepsilon_{t-s}) = 0$ for $s \neq t$ (no serial correlation). Then, $Var\left[T^{-1/2} Z'\varepsilon(\theta)\right] = \sigma^2 (Z'Z) \equiv V$.
- The optimal weighting matrix turns out to be $W_T = V^{-1}$ or $W_T = (Z'Z)^{-1}$ which downweighs noisy orthogonality conditions.
- Plugging $W_T = (Z'Z)^{-1}$ into the IV estimator above gives the two-stage least squares (2SLS) estimator

$$\begin{aligned} \widehat{\theta}_{2SLS} &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y \\ &= (X'P_Z X)^{-1} X'P_Z Y, \end{aligned}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix that projects orthogonally onto the space spanned by the columns of Z.
- This estimation method can be thought of as a two-stage procedure
  1. regress X on Z and get the fitted value $\widehat{X}$
  2. regress Y on the fitted value $\widehat{X}$ to obtain the estimate of θ.

Linear Instrumental Variables: Asymptotic Properties

- Under some regularity conditions,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right)\to^d N\left(0,\Omega\right),$$

where

$$\Omega=\sigma^2\left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1}.$$

Linear Instrumental Variables: Asymptotic Properties

- Under some regularity conditions,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \rightarrow^d N\left(0, \Omega\right),$$

where

$$\Omega = \sigma^2 \left[X' Z \left(Z' Z\right)^{-1} Z' X\right]^{-1}.$$

- If $m = k$, $Z'X$ is square and invertible and the variance-covariance matrix for the simple IV estimator becomes

$$\Omega = \sigma^2 (Z'X)^{-1} \left(Z'Z\right) \left(X'Z\right)^{-1}.$$

Linear Instrumental Variables: Asymptotic Properties

- Under some regularity conditions,

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \to^d N\left(0,\Omega\right),$$

where

$$\Omega=\sigma^2\left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1}.$$

- If $m=k$, $Z'X$ is square and invertible and the variance-covariance matrix for the simple IV estimator becomes

$$\Omega=\sigma^2(Z'X)^{-1}\left(Z'Z\right)\left(X'Z\right)^{-1}.$$

- If $Z=X$, we get the familiar variance-covariance matrix of the OLS estimator

$$\begin{aligned}\Omega &= \sigma^2(X'X)^{-1}\left(X'X\right)\left(X'X\right)^{-1}\\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:
  - $\varepsilon_t(\theta)$ can be a vector rather than a scalar

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:
  - $\varepsilon_t(\theta)$ can be a vector rather than a scalar
  - $\varepsilon_t(\theta)$ can be a nonlinear function of the data and the parameters.

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:
  - $\varepsilon_t(\theta)$ can be a vector rather than a scalar
  - $\varepsilon_t(\theta)$ can be a nonlinear function of the data and the parameters.
  - $\varepsilon_t(\theta)$ can be heteroskedastic and serially correlated

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:
  - $\varepsilon_t(\theta)$ can be a vector rather than a scalar
  - $\varepsilon_t(\theta)$ can be a nonlinear function of the data and the parameters.
  - $\varepsilon_t(\theta)$ can be heteroskedastic and serially correlated
- Define $g_t(\theta) = z_t \otimes \varepsilon_t(\theta)$, where $\otimes$ is the Kronecker product, and let $E[g_t(\theta_0)] = 0$ be a population moment condition implied by economic theory or statistical model.

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:
  - $\varepsilon_t(\theta)$ can be a vector rather than a scalar
  - $\varepsilon_t(\theta)$ can be a nonlinear function of the data and the parameters.
  - $\varepsilon_t(\theta)$ can be heteroskedastic and serially correlated
- Define $g_t(\theta) = z_t \otimes \varepsilon_t(\theta)$, where $\otimes$ is the Kronecker product, and let $E[g_t(\theta_0)] = 0$ be a population moment condition implied by economic theory or statistical model.
- The GMM minimizes a quadratic form in the sample counterparts of these moment conditions

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$$

where $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$ and W_T is a weighting matrix.

Generalized Methods of Moments (GMM)

- The generalized methods of moments (GMM) is an extension of the linear IV estimator in the following directions:
  - $\varepsilon_t(\theta)$ can be a vector rather than a scalar
  - $\varepsilon_t(\theta)$ can be a nonlinear function of the data and the parameters.
  - $\varepsilon_t(\theta)$ can be heteroskedastic and serially correlated
- Define $g_t(\theta) = z_t \otimes \varepsilon_t(\theta)$, where $\otimes$ is the Kronecker product, and let $E[g_t(\theta_0)] = 0$ be a population moment condition implied by economic theory or statistical model.
- The GMM minimizes a quadratic form in the sample counterparts of these moment conditions

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$$

where $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$ and W_T is a weighting matrix.
- The GMM estimator is given by

$$\widehat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} Q_T(\theta).$$

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,
$$\sqrt{T}(\widehat{\theta} - \theta_0) \to^d N(0, \Omega),$$
where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial\theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,
$$\sqrt{T}(\widehat{\theta} - \theta_0) \rightarrow^d N(0, \Omega),$$
where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial \theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.
- The optimal (efficient) GMM sets $W_T = V_T^{-1}$, where
$V_T = E\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \sum_{t=1}^{T} g_t(\theta)'\right] = Var\left[T^{-1/2} g_T(\theta)\right]$.

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,
$$\sqrt{T}(\widehat{\theta} - \theta_0) \to^d N(0, \Omega),$$
where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial\theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.
- The optimal (efficient) GMM sets $W_T = V_T^{-1}$, where
$V_T = E\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\sum_{t=1}^{T} g_t(\theta)'\right] = Var\left[T^{-1/2} g_T(\theta)\right]$.
- If $W_T = V_T^{-1}$, then $\Omega = (M'V^{-1}M)^{-1}$.

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,
$$\sqrt{T}(\widehat{\theta} - \theta_0) \to^d N(0, \Omega),$$
where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial \theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.
- The optimal (efficient) GMM sets $W_T = V_T^{-1}$, where
$V_T = E\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \sum_{t=1}^{T} g_t(\theta)'\right] = Var\left[T^{-1/2} g_T(\theta)\right]$.
- If $W_T = V_T^{-1}$, then $\Omega = (M'V^{-1}M)^{-1}$.
- The two-step GMM is computed as follows:

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,
$$\sqrt{T}(\widehat{\theta} - \theta_0) \to^d N(0, \Omega),$$
where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial \theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.
- The optimal (efficient) GMM sets $W_T = V_T^{-1}$, where
$V_T = E\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \sum_{t=1}^{T} g_t(\theta)'\right] = Var\left[T^{-1/2} g_T(\theta)\right]$.
- If $W_T = V_T^{-1}$, then $\Omega = \left(M'V^{-1}M\right)^{-1}$.
- The two-step GMM is computed as follows:
  1. estimate the model with some arbitrary symmetric positive definite matrix; for example, $W_T = I$. This preliminary estimator $\widetilde{\theta}$ is consistent but not efficient.

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,

$$\sqrt{T}(\widehat{\theta} - \theta_0) \to^d N(0, \Omega),$$

where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial \theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.
- The optimal (efficient) GMM sets $W_T = V_T^{-1}$, where
$V_T = E\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \sum_{t=1}^{T} g_t(\theta)'\right] = Var\left[T^{-1/2} g_T(\theta)\right]$.
- If $W_T = V_T^{-1}$, then $\Omega = (M'V^{-1}M)^{-1}$.
- The two-step GMM is computed as follows:
  1. estimate the model with some arbitrary symmetric positive definite matrix; for example, $W_T = I$. This preliminary estimator $\widetilde{\theta}$ is consistent but not efficient.
  2. set $W_T = V_T(\widetilde{\theta})^{-1}$ and reestimate θ by minimizing $Q_T(\theta) = g_T(\theta)' V_T(\widetilde{\theta})^{-1} g_T(\theta)$.

Generalized Methods of Moments (GMM)

- Under some assumptions and technical conditions,
$$\sqrt{T}(\widehat{\theta} - \theta_0) \rightarrow^d N(0, \Omega),$$
where $\Omega = (M'WM)^{-1} M'WVWM (M'WM)^{-1}$,
$M = E\left(\frac{\partial}{\partial\theta'} g_t(\theta_0)\right)$ and V is the variance of $g_t(\theta_0)$.
- The optimal (efficient) GMM sets $W_T = V_T^{-1}$, where
$V_T = E\left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) \sum_{t=1}^{T} g_t(\theta)'\right] = Var\left[T^{-1/2} g_T(\theta)\right]$.
- If $W_T = V_T^{-1}$, then $\Omega = \left(M'V^{-1}M\right)^{-1}$.
- The two-step GMM is computed as follows:
  1. estimate the model with some arbitrary symmetric positive definite matrix; for example, $W_T = I$. This preliminary estimator $\widetilde{\theta}$ is consistent but not efficient.
  2. set $W_T = V_T(\widetilde{\theta})^{-1}$ and reestimate θ by minimizing $Q_T(\theta) = g_T(\theta)' V_T(\widetilde{\theta})^{-1} g_T(\theta)$.
- If we iterate the procedure until convergence, we obtain the iterated GMM estimator.

GMM Estimation with Possible Heteroskedasticity and Serial Correlation

- In this general case,

$$
\begin{aligned}
V &= \lim_{T\to\infty} E\left[\frac{1}{T}\left(\sum_{t=1}^{T} g_t(\theta)\right)\left(\sum_{t=1}^{T} g_t(\theta)\right)'\right] \\
&= \Gamma_g(j) + \sum_{j=1}^{\infty}\left(\Gamma_g(j) + \Gamma_g(j)'\right)
\end{aligned}
$$

where $\Gamma_g(j) = E\left[g_t(\theta)\, g_{t-j}(\theta)'\right]$ is the j^{th} autocovariance matrix of $g_t(\theta)$.

GMM Estimation with Possible Heteroskedasticity and Serial Correlation

- In this general case,

$$\begin{aligned} V &= \lim_{T\to\infty} E\left[\frac{1}{T}\left(\sum_{t=1}^{T} g_t(\theta)\right)\left(\sum_{t=1}^{T} g_t(\theta)\right)'\right] \\ &= \Gamma_g(j) + \sum_{j=1}^{\infty}\left(\Gamma_g(j) + \Gamma_g(j)'\right) \end{aligned}$$

  where $\Gamma_g(j) = E\left[g_t(\theta)\, g_{t-j}(\theta)'\right]$ is the j^{th} autocovariance matrix of $g_t(\theta)$.

- To estimate V, we replace the true auto-covariances with sample autocovariances

$$\widehat{\Gamma}_g = \frac{1}{T}\sum_{t=j+1}^{T} g_t(\theta)\, g_{t-j}(\theta)'.$$

GMM Estimation with Possible Heteroskedasticity and Serial Correlation

- Then,

$$V_T = \widehat{\Gamma}_g(0) + \sum_{j=1}^{T-1} K\left(\frac{j}{m}\right)\left(\widehat{\Gamma}_g(j) + \widehat{\Gamma}_g(j)'\right)$$

where $K\left(\frac{j}{m}\right)$ is a weight function (kernel) such that $K\left(\frac{j}{m}\right) = 0$ if $j > m$ and m is the bandwidth.

GMM Estimation with Possible Heteroskedasticity and Serial Correlation

- Then,

$$V_T = \widehat{\Gamma}_g(0) + \sum_{j=1}^{T-1} K\left(\frac{j}{m}\right)\left(\widehat{\Gamma}_g(j) + \widehat{\Gamma}_g(j)'\right)$$

  where $K\left(\frac{j}{m}\right)$ is a weight function (kernel) such that $K\left(\frac{j}{m}\right) = 0$ if $j > m$ and m is the bandwidth.

  - m increases with the sample size but not too rapidly ($m^4 / T \to 0$).

GMM Estimation with Possible Heteroskedasticity and Serial Correlation

- Then,

$$V_T = \widehat{\Gamma}_g(0) + \sum_{j=1}^{T-1} K\left(\frac{j}{m}\right)\left(\widehat{\Gamma}_g(j) + \widehat{\Gamma}_g(j)'\right)$$

where $K\left(\frac{j}{m}\right)$ is a weight function (kernel) such that $K\left(\frac{j}{m}\right) = 0$ if $j > m$ and m is the bandwidth.

  - m increases with the sample size but not too rapidly ($m^4/T \to 0$).

- Newey-West (1987, 1994) estimator

$$V_T = \widehat{\Gamma}_g(0) + \sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\left(\widehat{\Gamma}_g(j) + \widehat{\Gamma}_g(j)'\right)$$

with m chosen as $m = 4\left(\frac{T}{100}\right)^{2/9}$.

GMM Estimation of Linear Factor Models: SDF Form

- Recall that the unconditional version of the fundamental asset pricing equation is given by

$$E(mR-1)=0.$$

GMM Estimation of Linear Factor Models: SDF Form

- Recall that the unconditional version of the fundamental asset pricing equation is given by
$$E(mR - 1) = 0.$$
- The stochastic discount factor m is a function of the data x and unknown parameters θ, $m(x, \theta)$.

GMM Estimation of Linear Factor Models: SDF Form

- Recall that the unconditional version of the fundamental asset pricing equation is given by
$$E(mR - 1) = 0.$$
- The stochastic discount factor m is a function of the data x and unknown parameters θ, $m(x, \theta)$.
- For linear factor models,
$$m = a + b'f = \theta' x,$$
where $\theta = (a, b')'$ and $x = (1, f')'$.

GMM Estimation of Linear Factor Models: SDF Form

- Recall that the unconditional version of the fundamental asset pricing equation is given by
$$E(mR-1)=0.$$
- The stochastic discount factor m is a function of the data x and unknown parameters θ, $m(x,\theta)$.
- For linear factor models,
$$m=a+b'f=\theta' x,$$
where $\theta=(a,b')'$ and $x=(1,f')'$.
- Then, the pricing errors for N assets can be expressed as
$$\begin{aligned} g(\theta) &= E(mR)-1_N \\ &= E(Rx'\theta)-1_N \\ &= D\theta-1_N, \end{aligned}$$
where $D=E(Rx')$.

GMM Estimation of Linear Factor Models: SDF Form

- The sample analog is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} R_t x_t'\theta - 1_N.$$

GMM Estimation of Linear Factor Models: SDF Form

- The sample analog is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} R_t x_t' \theta - 1_N.$$

- For a given W_T, the GMM estimator of θ minimizes $g_T(\theta)' W_T g_T(\theta)$ and solves the first-order condition

$$D_T' W_T \left[D_T \widehat{\theta} - 1_N \right] = 0,$$

where $D_T = \frac{\partial g_T(\widehat{\theta})}{\partial \theta'} = T^{-1} \sum_{t=1}^{T} R_t x_t'$.

GMM Estimation of Linear Factor Models: SDF Form

- The sample analog is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} R_t x_t'\theta - 1_N.$$

- For a given W_T, the GMM estimator of θ minimizes $g_T(\theta)' W_T g_T(\theta)$ and solves the first-order condition

$$D_T' W_T \left[D_T \widehat{\theta} - 1_N \right] = 0,$$

where $D_T = \frac{\partial g_T\left(\widehat{\theta}\right)}{\partial \theta'} = T^{-1}\sum_{t=1}^{T} R_t x_t'$.

- Solving this linear equation for $\widehat{\theta}$ yields

$$\widehat{\theta} = (D_T' W_T D_T)^{-1}(D_T' W_T 1_N).$$

GMM Estimation of Linear Factor Models: Optimal GMM

- The optimal GMM sets $W_T = V_T^{-1}$, where $V_T = Var\left[T^{-1/2} g_T\left(\theta\right)\right]$.

GMM Estimation of Linear Factor Models: Optimal GMM

- The optimal GMM sets $W_T = V_T^{-1}$, where $V_T = Var\left[T^{-1/2} g_T(\theta)\right]$.
- In this case,
$$\widehat{\theta} = (D_T' V_T^{-1} D_T)^{-1} (D_T' V_T^{-1} 1_N),$$
where V_T is evaluated at some preliminary (consistent) estimator $\widetilde{\theta}$ (typically obtained using $W_T = I$).

GMM Estimation of Linear Factor Models: Optimal GMM

- The optimal GMM sets $W_T = V_T^{-1}$, where $V_T = Var\left[T^{-1/2} g_T(\theta)\right]$.
- In this case,

$$\widehat{\theta} = (D_T' V_T^{-1} D_T)^{-1} (D_T' V_T^{-1} 1_N),$$

  where V_T is evaluated at some preliminary (consistent) estimator $\widetilde{\theta}$ (typically obtained using $W_T = I$).
- Model specification test: if the model is correctly specified, the pricing errors $g(\theta) = E(Rx'\theta) - 1_N$ are zero which can be tested using the statistic

$$T g_T(\widehat{\theta})' V_T^{-1} g_T(\widehat{\theta}) \sim \chi^2_{N-k},$$

  where N is the number of test assets and k is the number of factors.

GMM Estimation of Linear Factor Models: Optimal GMM

- The optimal GMM sets $W_T = V_T^{-1}$, where $V_T = Var\left[T^{-1/2} g_T(\theta)\right]$.
- In this case,
$$\widehat{\theta} = (D_T' V_T^{-1} D_T)^{-1}(D_T' V_T^{-1} 1_N),$$
where V_T is evaluated at some preliminary (consistent) estimator $\widetilde{\theta}$ (typically obtained using $W_T = I$).
- Model specification test: if the model is correctly specified, the pricing errors $g(\theta) = E(Rx'\theta) - 1_N$ are zero which can be tested using the statistic
$$Tg_T(\widehat{\theta})' V_T^{-1} g_T(\widehat{\theta}) \sim \chi^2_{N-k},$$
where N is the number of test assets and k is the number of factors.
- If the model is misspecified, the value of the test statistic depends on the choice of W_T.

GMM Estimation of Linear Factor Models: Optimal GMM

- The optimal GMM sets $W_T = V_T^{-1}$, where $V_T = Var\left[T^{-1/2} g_T(\theta)\right]$.
- In this case,
$$\widehat{\theta} = (D_T' V_T^{-1} D_T)^{-1}(D_T' V_T^{-1} 1_N),$$
where V_T is evaluated at some preliminary (consistent) estimator $\widetilde{\theta}$ (typically obtained using $W_T = I$).
- Model specification test: if the model is correctly specified, the pricing errors $g(\theta) = E(Rx'\theta) - 1_N$ are zero which can be tested using the statistic
$$T g_T(\widehat{\theta})' V_T^{-1} g_T(\widehat{\theta}) \sim \chi^2_{N-k},$$
where N is the number of test assets and k is the number of factors.
- If the model is misspecified, the value of the test statistic depends on the choice of W_T.
- Therefore, for model comparison of different asset pricing models, it makes more sense to use the same W_T across models.

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Hansen and Jagannathan (1997) suggest using $W = \Psi^{-1}$, where $\Psi = E(RR')$ is the second moment matrix of returns.

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Hansen and Jagannathan (1997) suggest using $W = \Psi^{-1}$, where $\Psi = E(RR')$ is the second moment matrix of returns.
- Then, the Hansen-Jagannathan (HJ) distance is defined as

$$\delta_{HJ}(\theta) = \sqrt{g_T(\theta)'\Psi_T^{-1}g_T(\theta)}$$

and

$$\begin{aligned} \widetilde{\theta} &= \arg\min_{\theta\in\Theta} \delta_{HJ}^2(\theta) \\ &= (D_T'\Psi_T^{-1}D_T)^{-1}(D_T'\Psi_T^{-1}1_N) \end{aligned}$$

is the resulting GMM estimator.

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Hansen and Jagannathan (1997) suggest using $W = \Psi^{-1}$, where $\Psi = E(RR')$ is the second moment matrix of returns.
- Then, the Hansen-Jagannathan (HJ) distance is defined as

$$\delta_{HJ}(\theta) = \sqrt{g_T(\theta)'\Psi_T^{-1}g_T(\theta)}$$

and

$$\begin{aligned}\widetilde{\theta} &= \arg\min_{\theta\in\Theta} \delta_{HJ}^2(\theta) \\ &= (D_T'\Psi_T^{-1}D_T)^{-1}(D_T'\Psi_T^{-1}1_N)\end{aligned}$$

is the resulting GMM estimator.
- Economic interpretation of HJ distance:

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Hansen and Jagannathan (1997) suggest using $W = \Psi^{-1}$, where $\Psi = E(RR')$ is the second moment matrix of returns.
- Then, the Hansen-Jagannathan (HJ) distance is defined as

$$\delta_{HJ}(\theta) = \sqrt{g_T(\theta)'\Psi_T^{-1} g_T(\theta)}$$

and

$$\begin{aligned} \widetilde{\theta} &= \arg\min_{\theta \in \Theta} \delta_{HJ}^2(\theta) \\ &= (D_T' \Psi_T^{-1} D_T)^{-1} (D_T' \Psi_T^{-1} 1_N) \end{aligned}$$

is the resulting GMM estimator.
- Economic interpretation of HJ distance:
  - measures the minimum distance between the proposed SDF and the set of correct SDFs

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Hansen and Jagannathan (1997) suggest using $W = \Psi^{-1}$, where $\Psi = E(RR')$ is the second moment matrix of returns.
- Then, the Hansen-Jagannathan (HJ) distance is defined as

$$\delta_{HJ}(\theta) = \sqrt{g_T(\theta)' \Psi_T^{-1} g_T(\theta)}$$

and

$$\begin{aligned} \widetilde{\theta} &= \arg\min_{\theta \in \Theta} \delta_{HJ}^2(\theta) \\ &= (D_T' \Psi_T^{-1} D_T)^{-1} (D_T' \Psi_T^{-1} 1_N) \end{aligned}$$

is the resulting GMM estimator.
- Economic interpretation of HJ distance:
  - measures the minimum distance between the proposed SDF and the set of correct SDFs
  - represents the maximum pricing error of a portfolio of R that has a unit second moment.

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Test of model specification is based on

$$T\delta^2_{HJ}(\widetilde{\theta}) \to^d \sum_{j=1}^{N-k} \xi_j v_j,$$

where v_j are χ^2_1 random variables and $\xi_1, ..., \xi_{n-k}$ are non-zero eigenvalues of matrix

$A = V^{1/2}\Psi^{-1/2}[I_N\text{-}(\Psi^{-1/2})'D(D'\Psi^{-1}D)^{-1}D'\Psi^{-1/2}](\Psi^{-1/2})'(V^{1/2})'.$

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Test of model specification is based on

$$T\delta^2_{HJ}(\widetilde{\theta}) \to^d \sum_{j=1}^{N-k} \xi_j v_j,$$

where v_j are χ^2_1 random variables and $\xi_1, ..., \xi_{n-k}$ are non-zero eigenvalues of matrix

$$A{=}V^{1/2}\Psi^{-1/2}[I_N\text{-}(\Psi^{-1/2})'D(D'\Psi^{-1}D)^{-1}D'\Psi^{-1/2}](\Psi^{-1/2})'(V^{1/2})'.$$

  - this is a weighted chi-square distribution.

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Test of model specification is based on

$$T\delta^2_{HJ}(\widetilde{\theta}) \to^d \sum_{j=1}^{N-k} \xi_j v_j,$$

  where v_j are χ^2_1 random variables and $\xi_1, \ldots, \xi_{n-k}$ are non-zero eigenvalues of matrix

  $A{=}V^{1/2}\Psi^{-1/2}[I_N\text{-}(\Psi^{-1/2})'D(D'\Psi^{-1}D)^{-1}D'\Psi^{-1/2}](\Psi^{-1/2})'(V^{1/2})'.$

  - this is a weighted chi-square distribution.
- If $T\delta^2_{HJ}(\widetilde{\theta})$ exceeds the critical value, then the model is misspecified

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Test of model specification is based on

$$T\delta^2_{HJ}(\widetilde{\theta}) \to^d \sum_{j=1}^{N-k} \xi_j v_j,$$

where v_j are χ^2_1 random variables and $\xi_1, ..., \xi_{n-k}$ are non-zero eigenvalues of matrix

$$A{=}V^{1/2}\Psi^{-1/2}[I_N\text{-}(\Psi^{-1/2})'D(D'\Psi^{-1}D)^{-1}D'\Psi^{-1/2}](\Psi^{-1/2})'(V^{1/2})'.$$

  - this is a weighted chi-square distribution.
- If $T\delta^2_{HJ}(\widetilde{\theta})$ exceeds the critical value, then the model is misspecified
  - the standard errors for the estimates $\widetilde{\theta}$ need to be adjusted for model misspecification (see Kan and Robotti, 2009, RFS).

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Test of model specification is based on

$$T\delta^2_{HJ}(\widetilde{\theta}) \to^d \sum_{j=1}^{N-k} \xi_j v_j,$$

where v_j are χ^2_1 random variables and $\xi_1, ..., \xi_{n-k}$ are non-zero eigenvalues of matrix

$$A = V^{1/2}\Psi^{-1/2}[I_N\text{-}(\Psi^{-1/2})'D(D'\Psi^{-1}D)^{-1}D'\Psi^{-1/2}](\Psi^{-1/2})'(V^{1/2})'.$$

  - this is a weighted chi-square distribution.
- If $T\delta^2_{HJ}(\widetilde{\theta})$ exceeds the critical value, then the model is misspecified
  - the standard errors for the estimates $\widetilde{\theta}$ need to be adjusted for model misspecification (see Kan and Robotti, 2009, RFS).
- Statistical comparison of HJ distances of two competing asset pricing models

GMM Estimation of Linear Factor Models: Hansen-Jagannathan Distance

- Test of model specification is based on

$$T\delta^2_{HJ}(\widetilde{\theta}) \rightarrow^d \sum_{j=1}^{N-k} \xi_j v_j,$$

where v_j are χ^2_1 random variables and $\xi_1, \ldots, \xi_{n-k}$ are non-zero eigenvalues of matrix

$$A = V^{1/2}\Psi^{-1/2}[I_N\text{-}(\Psi^{-1/2})'D(D'\Psi^{-1}D)^{-1}D'\Psi^{-1/2}](\Psi^{-1/2})'(V^{1/2})'.$$

  - this is a weighted chi-square distribution.
- If $T\delta^2_{HJ}(\widetilde{\theta})$ exceeds the critical value, then the model is misspecified
  - the standard errors for the estimates $\widetilde{\theta}$ need to be adjusted for model misspecification (see Kan and Robotti, 2009, RFS).
- Statistical comparison of HJ distances of two competing asset pricing models
  - depends on whether the models are correctly specified or misspecified, nested or non-nested (see Kan and Robotti, 2009, RFS).

GMM Estimation of Conditional Linear Factor Models

- The GMM estimators discussed so far are based on the unconditional moment restriction $E[m_{t+1}(\theta)R_{t+1} - 1] = 0$.

GMM Estimation of Conditional Linear Factor Models

- The GMM estimators discussed so far are based on the unconditional moment restriction $E[m_{t+1}(\theta)R_{t+1} - 1] = 0$.
- If $E[m_{t+1}(\theta)R_{t+1} - 1 | I_t] = 0$ and $z_t \in I_t$ is a vector of observable conditioning variables (instruments), then (by LIE) the estimation can be based on the moment conditions

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

where $\otimes$ denotes the Kronecker product.

GMM Estimation of Conditional Linear Factor Models

- The GMM estimators discussed so far are based on the unconditional moment restriction $E[m_{t+1}(\theta)R_{t+1} - 1] = 0$.
- If $E[m_{t+1}(\theta)R_{t+1} - 1|I_t] = 0$ and $z_t \in I_t$ is a vector of observable conditioning variables (instruments), then (by LIE) the estimation can be based on the moment conditions

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

  where $\otimes$ denotes the Kronecker product.
- Then estimation of the unknown parameters θ is performed by substituting the sample analog of these moment conditions into the GMM objective function.

GMM Estimation of Linear Factor Models Using Excess Returns

- When excess returns are used to estimate and test asset pricing models, the moment condition is given by

$$E[mR^e] = 0_N.$$

GMM Estimation of Linear Factor Models Using Excess Returns

- When excess returns are used to estimate and test asset pricing models, the moment condition is given by

$$E[mR^e] = 0_N.$$

- In this case, the mean of the SDF cannot be identified or, equivalently, the parameters a and b in $m = a - b'f$ cannot be identified separately.

GMM Estimation of Linear Factor Models Using Excess Returns

- When excess returns are used to estimate and test asset pricing models, the moment condition is given by

$$E[mR^e] = 0_N.$$

- In this case, the mean of the SDF cannot be identified or, equivalently, the parameters a and b in $m = a - b'f$ cannot be identified separately.
- This requires a particular choice of normalization

GMM Estimation of Linear Factor Models Using Excess Returns

- When excess returns are used to estimate and test asset pricing models, the moment condition is given by

$$E[mR^e] = 0_N.$$

- In this case, the mean of the SDF cannot be identified or, equivalently, the parameters a and b in $m = a - b'f$ cannot be identified separately.
- This requires a particular choice of normalization
  - one popular normalization is to set $a = 1$ in which case $m = 1 - b'f$

GMM Estimation of Linear Factor Models Using Excess Returns

- When excess returns are used to estimate and test asset pricing models, the moment condition is given by

$$E[mR^e] = 0_N.$$

- In this case, the mean of the SDF cannot be identified or, equivalently, the parameters a and b in $m = a - b'f$ cannot be identified separately.
- This requires a particular choice of normalization
  - one popular normalization is to set $a = 1$ in which case $m = 1 - b'f$
  - an alternative (preferred) normalization is to set $a = 1 + b'E(f)$ in which case $m = 1 - b'[f - E(f)]$ with $E(m) = 1$.

GMM Estimation of Linear Factor Models Using Excess Returns

- When excess returns are used to estimate and test asset pricing models, the moment condition is given by

$$E[mR^e] = 0_N.$$

- In this case, the mean of the SDF cannot be identified or, equivalently, the parameters a and b in $m = a - b'f$ cannot be identified separately.
- This requires a particular choice of normalization
  - one popular normalization is to set $a = 1$ in which case $m = 1 - b'f$
  - an alternative (preferred) normalization is to set $a = 1 + b'E(f)$ in which case $m = 1 - b'[f - E(f)]$ with $E(m) = 1$.
- It turns out that these two normalizations in specifying the SDF can give rise to very different results (see Kan and Robotti, 2008, JEF; Burnside, 2007).

GMM Estimation of Beta Representation

- Recall that the asset pricing model can also be rewritten as

$$E(R^e) = \beta'\lambda.$$

GMM Estimation of Beta Representation

- Recall that the asset pricing model can also be rewritten as

$$E(R^e) = \beta' \lambda.$$

- Then, the moment conditions $g(\theta)$ are given by

$$g(\theta) = \begin{bmatrix} E(R^e - a - \beta' f) \\ E(R^e - a - \beta' f) f \\ E(R^e - \beta' \lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

GMM Estimation of Beta Representation

- Recall that the asset pricing model can also be rewritten as

$$E(R^e) = \beta' \lambda.$$

- Then, the moment conditions $g(\theta)$ are given by

$$g(\theta) = \begin{bmatrix} E(R^e - a - \beta' f) \\ E(R^e - a - \beta' f) f \\ E(R^e - \beta' \lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

- Substituting the sample analogs of these moment conditions into the GMM objective function and minimizing, we obtain the GMM estimator of a, β and λ.

GMM Estimation of Beta Representation

- Recall that the asset pricing model can also be rewritten as

$$E(R^e) = \beta'\lambda.$$

- Then, the moment conditions $g(\theta)$ are given by

$$g(\theta) = \begin{bmatrix} E(R^e - a - \beta' f) \\ E(R^e - a - \beta' f) f \\ E(R^e - \beta'\lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

- Substituting the sample analogs of these moment conditions into the GMM objective function and minimizing, we obtain the GMM estimator of a, β and λ.
- Advantages:

GMM Estimation of Beta Representation

- Recall that the asset pricing model can also be rewritten as

$$E(R^e) = \beta'\lambda.$$

- Then, the moment conditions $g(\theta)$ are given by

$$g(\theta) = \begin{bmatrix} E(R^e - a - \beta' f) \\ E(R^e - a - \beta' f) f \\ E(R^e - \beta'\lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

- Substituting the sample analogs of these moment conditions into the GMM objective function and minimizing, we obtain the GMM estimator of a, β and λ.
- Advantages:
  - no need to correct standard errors for generated regressors

GMM Estimation of Beta Representation

- Recall that the asset pricing model can also be rewritten as

$$E(R^e) = \beta'\lambda.$$

- Then, the moment conditions $g(\theta)$ are given by

$$g(\theta) = \begin{bmatrix} E(R^e - a - \beta' f) \\ E(R^e - a - \beta' f)f \\ E(R^e - \beta'\lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

- Substituting the sample analogs of these moment conditions into the GMM objective function and minimizing, we obtain the GMM estimator of a, β and λ.
- Advantages:
    - no need to correct standard errors for generated regressors
    - valid for serially correlated and heteroskedastic errors.

Consumption CAPM: CRRA Utility

- Recall that the SDF is given by

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

Consumption CAPM: CRRA Utility

- Recall that the SDF is given by

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

- Consider the constant relative risk aversion (CRRA) or power utility function

$$u(c_t) = \frac{c_t^{1-\rho} - 1}{1-\rho},$$

where $\rho > 0$ is the coefficient of relative risk aversion.

Consumption CAPM: CRRA Utility

- Recall that the SDF is given by

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

- Consider the constant relative risk aversion (CRRA) or power utility function

$$u(c_t) = \frac{c_t^{1-\rho} - 1}{1-\rho},$$

where $\rho > 0$ is the coefficient of relative risk aversion.

  - as $\rho \to 1$, we obtain the log utility function $u(c_t) = \log(c_t)$.

Consumption CAPM: CRRA Utility

- Recall that the SDF is given by

$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

- Consider the constant relative risk aversion (CRRA) or power utility function

$$u(c_t) = \frac{c_t^{1-\rho} - 1}{1-\rho},$$

where $\rho > 0$ is the coefficient of relative risk aversion.

  - as $\rho \to 1$, we obtain the log utility function $u(c_t) = \log(c_t)$.
  - Arrow-Pratt coefficient of relative risk aversion is

$$-\frac{c_t u''(c_t)}{u'(c_t)} = -\frac{c_t \left(-\rho c_t^{-\rho-1}\right)}{c_t^{-\rho}} = \rho,$$

  i.e. the relative risk aversion is constant.

Consumption CAPM: CRRA Utility

- Recall that the SDF is given by
$$m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$
- Consider the constant relative risk aversion (CRRA) or power utility function
$$u(c_t) = \frac{c_t^{1-\rho} - 1}{1-\rho},$$
where $\rho > 0$ is the coefficient of relative risk aversion.
  - as $\rho \to 1$, we obtain the log utility function $u(c_t) = \log(c_t)$.
  - Arrow-Pratt coefficient of relative risk aversion is
$$-\frac{c_t u''(c_t)}{u'(c_t)} = -\frac{c_t\left(-\rho c_t^{-\rho-1}\right)}{c_t^{-\rho}} = \rho,$$
i.e. the relative risk aversion is constant.
- Substituting into the fundamental pricing equation, we obtain
$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho} R_{t+1} - 1 \middle| I_t\right] = 0.$$

Consumption CAPM: CRRA Utility

- Drawbacks of CRRA utility

Consumption CAPM: CRRA Utility

- Drawbacks of CRRA utility
    1. Equity premium puzzle (Mehra and Prescott, 1985): requires very large values of risk aversion (ρ as high as 30 to 50) in order to fit US data

Consumption CAPM: CRRA Utility

- Drawbacks of CRRA utility
  1. Equity premium puzzle (Mehra and Prescott, 1985): requires very large values of risk aversion (ρ as high as 30 to 50) in order to fit US data
  2. In the CRRA framework, ρ is inversely related to the elasticity of intertemporal substitution (EIS)

Consumption CAPM: CRRA Utility

- Drawbacks of CRRA utility
  1. Equity premium puzzle (Mehra and Prescott, 1985): requires very large values of risk aversion (ρ as high as 30 to 50) in order to fit US data
  2. In the CRRA framework, ρ is inversely related to the elasticity of intertemporal substitution (EIS)
     - this is inappropriate because EIS is about the willingness of an investor to transfer consumption between time periods whereas ρ is about transferring consumption between states of the world.

Consumption CAPM: CRRA Utility

- Drawbacks of CRRA utility
  1. Equity premium puzzle (Mehra and Prescott, 1985): requires very large values of risk aversion (ρ as high as 30 to 50) in order to fit US data
  2. In the CRRA framework, ρ is inversely related to the elasticity of intertemporal substitution (EIS)
     - this is inappropriate because EIS is about the willingness of an investor to transfer consumption between time periods whereas ρ is about transferring consumption between states of the world.
- Non-expected and time non-separable (habit persistence) utility functions separate risk aversion and intertemporal substitution.

Consumption CAPM: Non-Expected Utility

- Non-expected (Epstein-Zin-Weil) utility

$$u_t = \left[(1-\beta) c_t^{1-\gamma} + \beta \left(E_t u_{t+1}^{1-\rho} \right)^{(1-\gamma)/(1-\rho)} \right]^{1/(1-\gamma)}.$$

Consumption CAPM: Non-Expected Utility

- Non-expected (Epstein-Zin-Weil) utility

$$u_t = \left[(1-\beta) c_t^{1-\gamma} + \beta \left(E_t u_{t+1}^{1-\rho} \right)^{(1-\gamma)/(1-\rho)} \right]^{1/(1-\gamma)}.$$

- The fundamental pricing equation is given by

$$E \left[\beta^{\lambda} \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma\lambda} (R_{t+1}^m)^{\lambda - 1} R_{t+1}^i - 1 \middle| I_t \right] = 0,$$

where $\lambda = (1-\rho)/(1-\gamma)$.

Consumption CAPM: Non-Expected Utility

- Non-expected (Epstein-Zin-Weil) utility

$$u_t = \left[(1-\beta) c_t^{1-\gamma} + \beta \left(E_t u_{t+1}^{1-\rho} \right)^{(1-\gamma)/(1-\rho)} \right]^{1/(1-\gamma)}.$$

- The fundamental pricing equation is given by

$$E \left[\beta^{\lambda} \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma\lambda} (R_{t+1}^m)^{\lambda-1} R_{t+1}^i - 1 \middle| I_t \right] = 0,$$

where $\lambda = (1-\rho)/(1-\gamma)$.

- It can be reduced to time-separable (CRRA) utility for $\lambda = 1$ (or, equivalently, $\gamma = \rho$).

Consumption CAPM: Time Non-Separable Utility

- Utility function with habit persistence and durability

$$u_t = \frac{s_t^{1-\rho} - 1}{1-\rho},$$

where $s_t = c_t + \delta c_{t-1}$.

- Utility function with habit persistence and durability

$$u_t = \frac{s_t^{1-\rho} - 1}{1-\rho},$$

where $s_t = c_t + \delta c_{t-1}$.

- The fundamental pricing equation is given by

$$E\left[\beta\left(s_{t+1}^{-\rho} + \beta\delta s_{t+2}^{-\rho}\right) R_{t+1}^{i} - \left(s_t^{-\rho} + \beta\delta s_{t+1}^{-\rho}\right) \middle| I_t\right] = 0.$$

Consumption CAPM: Time Non-Separable Utility

- Utility function with habit persistence and durability

$$u_t = \frac{s_t^{1-\rho} - 1}{1-\rho},$$

where $s_t = c_t + \delta c_{t-1}$.

- The fundamental pricing equation is given by

$$E\left[\beta\left(s_{t+1}^{-\rho} + \beta\delta s_{t+2}^{-\rho}\right) R_{t+1}^{i} - \left(s_t^{-\rho} + \beta\delta s_{t+1}^{-\rho}\right) \Big| I_t\right] = 0.$$

- It can be reduced to time-separable (CRRA) utility for $\delta = 0$.

Hansen-Jagannathan Volatility Bound

- For excess returns $R^e = R - R^f$, we have

$$\begin{aligned} 0 &= E(mR^e) \\ &= E(m)E(R^e) + Corr(m, R^e)\sigma(m)\sigma(R^e), \end{aligned}$$

by the covariance decomposition and the definition of correlation.

Hansen-Jagannathan Volatility Bound

- For excess returns $R^e = R - R^f$, we have

$$\begin{aligned} 0 &= E(mR^e) \\ &= E(m)E(R^e) + Corr(m, R^e)\sigma(m)\sigma(R^e), \end{aligned}$$

  by the covariance decomposition and the definition of correlation.

- Then,

$$\begin{aligned} \frac{\sigma(m)}{E(m)} &= -\frac{1}{Corr(m, R^e)} \frac{E(R^e)}{\sigma(R^e)} \\ &\geq \frac{|E(R^e)|}{\sigma(R^e)} \end{aligned}$$

  since $|Corr(m, R^e)| \leq 1$.

Hansen-Jagannathan Volatility Bound

- For excess returns $R^e = R - R^f$, we have

$$\begin{aligned} 0 &= E(mR^e) \\ &= E(m)E(R^e) + Corr(m, R^e)\sigma(m)\sigma(R^e), \end{aligned}$$

  by the covariance decomposition and the definition of correlation.
- Then,

$$\begin{aligned} \frac{\sigma(m)}{E(m)} &= -\frac{1}{Corr(m, R^e)}\frac{E(R^e)}{\sigma(R^e)} \\ &\geq \frac{|E(R^e)|}{\sigma(R^e)} \end{aligned}$$

  since $|Corr(m, R^e)| \leq 1$.
  - this inequality or volatility bound (Hansen and Jagannathan, 1991) uses data on returns to define combinations of mean and standard deviation of m that are consistent with the fundamental asset pricing equation.

Hansen-Jagannathan Volatility Bound

- For excess returns $R^e = R - R^f$, we have

$$\begin{aligned} 0 &= E(mR^e) \\ &= E(m)E(R^e) + Corr(m, R^e)\sigma(m)\sigma(R^e), \end{aligned}$$

  by the covariance decomposition and the definition of correlation.
- Then,

$$\begin{aligned} \frac{\sigma(m)}{E(m)} &= -\frac{1}{Corr(m, R^e)} \frac{E(R^e)}{\sigma(R^e)} \\ &\geq \frac{|E(R^e)|}{\sigma(R^e)} \end{aligned}$$

  since $|Corr(m, R^e)| \leq 1$.
  - this inequality or volatility bound (Hansen and Jagannathan, 1991) uses data on returns to define combinations of mean and standard deviation of m that are consistent with the fundamental asset pricing equation.
  - this approach is useful when we are uncertain of the form of the utility function.

Equity Premium Puzzle

- In US data, $\frac{|E(R^e)|}{\sigma(R^e)} \approx \frac{0.08}{0.16} = 0.5$.

Equity Premium Puzzle

- In US data, $\frac{|E(R^e)|}{\sigma(R^e)} \approx \frac{0.08}{0.16} = 0.5$.
- Since $E(m) = 1/E(R^f) \approx 0.99$ (where the risk-free rate is proxied by the T-bill rate), then $\sigma(m) \geq 0.5$.

Equity Premium Puzzle

- In US data, $\frac{|E(R^e)|}{\sigma(R^e)} \approx \frac{0.08}{0.16} = 0.5$.
- Since $E(m) = 1/E(R^f) \approx 0.99$ (where the risk-free rate is proxied by the T-bill rate), then $\sigma(m) \geq 0.5$.
- In C-CAPM with CRRA preferences, $m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\rho}$.

Equity Premium Puzzle

- In US data, $\frac{|E(R^e)|}{\sigma(R^e)} \approx \frac{0.08}{0.16} = 0.5$.
- Since $E(m) = 1/E(R^f) \approx 0.99$ (where the risk-free rate is proxied by the T-bill rate), then $\sigma(m) \geq 0.5$.
- In C-CAPM with CRRA preferences, $m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\rho}$.
- Since $\sigma\left(\frac{c_{t+1}}{c_t}\right) \approx 0.01$ in US data, log utility ($\rho \to 1$) implies that $\sigma(m) \approx \rho\sigma(\triangle c) = 0.01$ which is 50 times smaller than the volatility bound requires.

Equity Premium Puzzle

- In US data, $\frac{|E(R^e)|}{\sigma(R^e)} \approx \frac{0.08}{0.16} = 0.5$.
- Since $E(m) = 1/E(R^f) \approx 0.99$ (where the risk-free rate is proxied by the T-bill rate), then $\sigma(m) \geq 0.5$.
- In C-CAPM with CRRA preferences, $m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\rho}$.
- Since $\sigma\left(\frac{c_{t+1}}{c_t}\right) \approx 0.01$ in US data, log utility ($\rho \to 1$) implies that $\sigma(m) \approx \rho\sigma(\triangle c) = 0.01$ which is 50 times smaller than the volatility bound requires.
  - alternatively, it requires an implausibly high value of the risk aversion parameter.

Hansen-Jagannathan Volatility Bound

- In the general case,

$$p = E(mx) = E(m)E(x) + \Sigma\beta, \qquad (1)$$

where $\Sigma = E[(x - E(x))(x - E(x))']$ and β is the vector of slope coefficients in the regression of m on the asset payoffs

$$m = E(m) + (x - E(x))'\beta + \varepsilon \qquad (2)$$

with $E(\varepsilon) = 0$ and $E(\varepsilon x) = 0$.

Hansen-Jagannathan Volatility Bound

- In the general case,

$$p = E(mx) = E(m)E(x) + \Sigma\beta, \tag{1}$$

where $\Sigma = E[(x - E(x))(x - E(x))']$ and β is the vector of slope coefficients in the regression of m on the asset payoffs

$$m = E(m) + (x - E(x))'\beta + \varepsilon \tag{2}$$

with $E(\varepsilon) = 0$ and $E(\varepsilon x) = 0$.

- From (1),

$$\beta = \Sigma^{-1}[p - E(m)E(x)]. \tag{3}$$

Hansen-Jagannathan Volatility Bound

- In the general case,

$$p = E(mx) = E(m)E(x) + \Sigma\beta, \tag{1}$$

where $\Sigma = E[(x - E(x))(x - E(x))']$ and β is the vector of slope coefficients in the regression of m on the asset payoffs

$$m = E(m) + (x - E(x))'\beta + \varepsilon \tag{2}$$

with $E(\varepsilon) = 0$ and $E(\varepsilon x) = 0$.
- From (1),

$$\beta = \Sigma^{-1}[p - E(m)E(x)]. \tag{3}$$

- By definition, $\sigma^2(m) = E[(m - E(m))(m - E(m))']$. Then, from (2) and (3), it follows that the HJ volatility bound is given by

$$\sigma^2(m) \geq [p - E(m)E(x)]'\Sigma^{-1}[p - E(m)E(x)].$$

Hansen-Jagannathan Volatility Bound

- In the general case,

$$p = E(mx) = E(m)E(x) + \Sigma\beta, \tag{1}$$

where $\Sigma = E[(x - E(x))(x - E(x))']$ and β is the vector of slope coefficients in the regression of m on the asset payoffs

$$m = E(m) + (x - E(x))'\beta + \varepsilon \tag{2}$$

with $E(\varepsilon) = 0$ and $E(\varepsilon x) = 0$.

- From (1),

$$\beta = \Sigma^{-1}[p - E(m)E(x)]. \tag{3}$$

- By definition, $\sigma^2(m) = E[(m - E(m))(m - E(m))']$. Then, from (2) and (3), it follows that the HJ volatility bound is given by

$$\sigma^2(m) \geq [p - E(m)E(x)]'\Sigma^{-1}[p - E(m)E(x)].$$

  - this is a parabola in $\{E(m), \sigma^2(m)\}$ or hyperbola in $\{E(m), \sigma(m)\}$.

Hansen-Jagannathan Volatility Bound

- In the general case,

$$p = E(mx) = E(m)E(x) + \Sigma\beta, \tag{1}$$

where $\Sigma = E[(x - E(x))(x - E(x))']$ and β is the vector of slope coefficients in the regression of m on the asset payoffs

$$m = E(m) + (x - E(x))'\beta + \varepsilon \tag{2}$$

with $E(\varepsilon) = 0$ and $E(\varepsilon x) = 0$.

- From (1),

$$\beta = \Sigma^{-1}[p - E(m)E(x)]. \tag{3}$$

- By definition, $\sigma^2(m) = E[(m - E(m))(m - E(m))']$. Then, from (2) and (3), it follows that the HJ volatility bound is given by

$$\sigma^2(m) \geq [p - E(m)E(x)]'\Sigma^{-1}[p - E(m)E(x)].$$

  - this is a parabola in $\{E(m), \sigma^2(m)\}$ or hyperbola in $\{E(m), \sigma(m)\}$.
  - for given data, we can verify if the SDFs of different asset pricing models satisfy the HJ volatility bound.

GMM Estimation of Nonlinear Asset Pricing Models

- Consider first C-CAPM with time separable (CRRA) utility.

GMM Estimation of Nonlinear Asset Pricing Models

- Consider first C-CAPM with time separable (CRRA) utility.
- The Euler equation is given by

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}R_{t+1}-1\middle| I_t\right]=0.$$

GMM Estimation of Nonlinear Asset Pricing Models

- Consider first C-CAPM with time separable (CRRA) utility.
- The Euler equation is given by

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}R_{t+1}-1\middle| I_t\right]=0.$$

- Let $z_t \in I_t$ with $z_t = (1, \frac{c_t}{c_{t-1}}, \frac{c_{t-1}}{c_{t-2}}, ..., R_t, R_{t-1}, ...)$ for example.

GMM Estimation of Nonlinear Asset Pricing Models

- Consider first C-CAPM with time separable (CRRA) utility.
- The Euler equation is given by

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}R_{t+1}-1\middle| I_t\right]=0.$$

- Let $z_t \in I_t$ with $z_t = (1, \frac{c_t}{c_{t-1}}, \frac{c_{t-1}}{c_{t-2}}, ..., R_t, R_{t-1}, ...)$ for example.
- Then, by the law of iterated expectations, we obtain the unconditional moment restriction

$$E\left\{\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}R_{t+1}-1\right]z_t\right\}=0$$

or

$$E\left\{\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}\mathbf{R}_{t+1}-\mathbf{1}\right]\otimes z_t\right\}=\mathbf{0}$$

where $\mathbf{R}_{t+1}$ is a vector of asset returns, $\mathbf{1}$ is a vector of ones, $\mathbf{0}$ is a vector of zeros and $\otimes$ is the Kronecker product.

GMM Estimation of Nonlinear Asset Pricing Models

- The sample analog of the moment condition

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}\mathbf{R}_{t+1} - \mathbf{1}\right]\otimes z_t$$

is an $m \times 1$ vector $(m = \dim(\mathbf{R}_{t+1})\dim(z_t))$ and $\theta = (\beta, \rho)$.

- The sample analog of the moment condition

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}\mathbf{R}_{t+1} - \mathbf{1}\right]\otimes z_t$$

is an $m \times 1$ vector $(m = \dim(\mathbf{R}_{t+1})\dim(z_t))$ and $\theta = (\beta, \rho)$.

- Since the moment conditions are possibly serially correlated, we need to use Newey-West estimator of the long-run variance of $g(\theta)$ and use its inverse as optimal weighting matrix

$$W_T = \left[\widetilde{\Gamma}_g(0) + \sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\left(\widetilde{\Gamma}_g(j) + \widetilde{\Gamma}_g(j)'\right)\right]^{-1}$$

where $\widehat{\Gamma}_g(j)$ is the estimated jth autocovariance matrix of $g\left(\widetilde{\theta}\right)$ and $\widetilde{\theta}$ is a preliminary estimator obtained by setting $W_T = I$.

GMM Estimation of Nonlinear Asset Pricing Models

- Plugging the expressions for $g_T(\theta)$ and W_T into the GMM objective function $Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$, the GMM estimator of θ can be obtained by minimizing $Q_T(\theta)$.

GMM Estimation of Nonlinear Asset Pricing Models

- Plugging the expressions for $g_T(\theta)$ and W_T into the GMM objective function $Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$, the GMM estimator of θ can be obtained by minimizing $Q_T(\theta)$.
- The standard errors are computed as the square root of the diagonal elements of matrix $\left(\widehat{M}'\widehat{V}^{-1}\widehat{M}\right)^{-1}/T$, where $\widehat{M} = \sum_{t=1}^{T} \frac{\partial g_t(\widehat{\theta})}{\partial \theta'}$ and $\widehat{V}$ is the Newey-West variance estimator of $g_t(\widehat{\theta})$.

GMM Estimation of Nonlinear Asset Pricing Models

- Plugging the expressions for $g_T(\theta)$ and W_T into the GMM objective function $Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$, the GMM estimator of θ can be obtained by minimizing $Q_T(\theta)$.
- The standard errors are computed as the square root of the diagonal elements of matrix $\left(\widehat{M}'\widehat{V}^{-1}\widehat{M}\right)^{-1}/T$, where $\widehat{M} = \sum_{t=1}^{T} \frac{\partial g_t(\widehat{\theta})}{\partial \theta'}$ and $\widehat{V}$ is the Newey-West variance estimator of $g_t(\widehat{\theta})$.
- The model specification is tested using the statistic

$$TQ_T(\widehat{\theta}) \to^d \chi^2_{m-k},$$

where k is the dimension of the parameter vector θ.

GMM Estimation of Nonlinear Asset Pricing Models

- Plugging the expressions for $g_T(\theta)$ and W_T into the GMM objective function $Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$, the GMM estimator of θ can be obtained by minimizing $Q_T(\theta)$.
- The standard errors are computed as the square root of the diagonal elements of matrix $\left(\widehat{M}' \widehat{V}^{-1} \widehat{M}\right)^{-1} / T$, where $\widehat{M} = \sum_{t=1}^{T} \frac{\partial g_t(\widehat{\theta})}{\partial \theta'}$ and $\widehat{V}$ is the Newey-West variance estimator of $g_t(\widehat{\theta})$.
- The model specification is tested using the statistic

$$TQ_T(\widehat{\theta}) \rightarrow^d \chi^2_{m-k},$$

where k is the dimension of the parameter vector θ.

  - a value of the test statistic that exceeds the critical value from the chi-square distribution results in a rejection of the model specification (model assumptions, choice of utility function, choice of instruments).

GMM Estimation of Nonlinear Asset Pricing Models

- Tesing parameter restrictions of the form $h(\theta) = 0$, where $h(\cdot)$ is continuous and twice differentiable function.

GMM Estimation of Nonlinear Asset Pricing Models

- Tesing parameter restrictions of the form $h(\theta) = 0$, where $h(\cdot)$ is continuous and twice differentiable function.
- Wald statistic
$$F_T = Th(\widehat{\theta})' \left(H\widehat{\Omega}H' \right)^{-1} h(\widehat{\theta}),$$
where $H == \frac{\partial}{\partial\theta'} h\left(\widehat{\theta}\right)$ and $\widehat{\Omega} = \left(\widehat{M}'\widehat{V}^{-1}\widehat{M}\right)^{-1}$.

GMM Estimation of Nonlinear Asset Pricing Models

- Tesing parameter restrictions of the form $h(\theta) = 0$, where $h(\cdot)$ is continuous and twice differentiable function.
- Wald statistic
$$F_T = Th(\widehat{\theta})' \left(H\widehat{\Omega}H' \right)^{-1} h(\widehat{\theta}),$$
where $H == \frac{\partial}{\partial \theta'} h\left(\widehat{\theta}\right)$ and $\widehat{\Omega} = \left(\widehat{M}'\widehat{V}^{-1}\widehat{M}\right)^{-1}$.
- Under the null hypothesis $h(\theta) = 0$,
$$F_T \xrightarrow{d} \chi^2_q,$$
where q is the number of restrictions.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with habit persistence (time non-separable preferences), the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta\frac{s_{t+1}^{-\rho} + \beta\delta s_{t+2}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{R}_{t+1} - \frac{s_t^{-\rho} + \beta\delta s_{t+1}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{1}\right] \otimes z_t = \mathbf{0},$$

  where we divide the original moment condition by $s_t^{-\rho}(1+\beta\delta)$ to induce stationarity and rule out trivial solutions (note that if $\rho = 0$ and $\beta\delta = -1$, the moment conditions are trivially satisfied).

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with habit persistence (time non-separable preferences), the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta \frac{s_{t+1}^{-\rho} + \beta\delta s_{t+2}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{R}_{t+1} - \frac{s_t^{-\rho} + \beta\delta s_{t+1}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{1}\right] \otimes z_t = \mathbf{0},$$

  where we divide the original moment condition by $s_t^{-\rho}(1+\beta\delta)$ to induce stationarity and rule out trivial solutions (note that if $\rho = 0$ and $\beta\delta = -1$, the moment conditions are trivially satisfied).
- The parameter vector θ is $\theta = (\beta, \rho, \delta)$.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with habit persistence (time non-separable preferences), the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta\frac{s_{t+1}^{-\rho}+\beta\delta s_{t+2}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{R}_{t+1} - \frac{s_t^{-\rho}+\beta\delta s_{t+1}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{1}\right]\otimes z_t = \mathbf{0},$$

  where we divide the original moment condition by $s_t^{-\rho}(1+\beta\delta)$ to induce stationarity and rule out trivial solutions (note that if $\rho = 0$ and $\beta\delta = -1$, the moment conditions are trivially satisfied).
- The parameter vector θ is $\theta = (\beta, \rho, \delta)$.
- The hypothesis $\delta = 0$ can be tested statistically to see if the model can be reduced to time additive preferences.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with habit persistence (time non-separable preferences), the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta\frac{s_{t+1}^{-\rho} + \beta\delta s_{t+2}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{R}_{t+1} - \frac{s_t^{-\rho} + \beta\delta s_{t+1}^{-\rho}}{s_t^{-\rho}(1+\beta\delta)}\mathbf{1}\right]\otimes z_t = \mathbf{0},$$

  where we divide the original moment condition by $s_t^{-\rho}(1+\beta\delta)$ to induce stationarity and rule out trivial solutions (note that if $\rho = 0$ and $\beta\delta = -1$, the moment conditions are trivially satisfied).
- The parameter vector θ is $\theta = (\beta, \rho, \delta)$.
- The hypothesis $\delta = 0$ can be tested statistically to see if the model can be reduced to time additive preferences.
  - In our notation, this hypothesis can be written as $H\theta = 0$, where $H = [0, 0, 1]$.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with non-expected (Epstein-Zin-Weil) utility, the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta^{\lambda}\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma\lambda}(R_{t+1}^{m})^{\lambda-1}\mathbf{R}_{t+1} - \mathbf{1}\right]\otimes z_t = \mathbf{0},$$

where the parameter vector θ is $\theta = (\beta, \gamma, \lambda)$.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with non-expected (Epstein-Zin-Weil) utility, the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta^{\lambda}\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma\lambda}(R^m_{t+1})^{\lambda-1}\mathbf{R}_{t+1} - \mathbf{1}\right]\otimes z_t = \mathbf{0},$$

  where the parameter vector θ is $\theta = (\beta, \gamma, \lambda)$.
- The time separable utility again is a testable hypothesis by testing if $\lambda = 1$.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with non-expected (Epstein-Zin-Weil) utility, the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta^{\lambda}\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma\lambda}(R^m_{t+1})^{\lambda-1}\mathbf{R}_{t+1} - \mathbf{1}\right]\otimes z_t = \mathbf{0},$$

  where the parameter vector θ is $\theta = (\beta, \gamma, \lambda)$.
- The time separable utility again is a testable hypothesis by testing if $\lambda = 1$.
  - In our notation, this hypothesis can be written as $H\theta - 1 = 0$, where $H = [0, 0, 1]$.

GMM Estimation of Nonlinear Asset Pricing Models

- For the model with non-expected (Epstein-Zin-Weil) utility, the sample moment condition is

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\beta^{\lambda}\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma\lambda}(R^m_{t+1})^{\lambda-1}\mathbf{R}_{t+1} - \mathbf{1}\right]\otimes z_t = \mathbf{0},$$

  where the parameter vector θ is $\theta = (\beta, \gamma, \lambda)$.

- The time separable utility again is a testable hypothesis by testing if $\lambda = 1$.
  - In our notation, this hypothesis can be written as $H\theta - 1 = 0$, where $H = [0, 0, 1]$.

- The standard errors and the test for model specification are constructed as described above.