Generalized Method of Moments (GMM) Estimation

Heino Bohn Nielsen

Outline of the Lecture

(1) Introduction.

(2) Moment conditions and methods of moments (MM) estimation.
   • Ordinary least squares (OLS) estimation.
   • Instrumental variables (IVE) estimation.

(3) GMM defined in the general case.

(4) Specification test.

(5) Linear GMM.
   • Generalized instrumental variables (GIVE or 2SLS) estimation.
Idea of GMM

Estimation under weak assumptions; based on so-called moment conditions.

Moment conditions are statements involving the data and the parameters. Arise naturally in many contexts. For example:

(A) In a regression model, $y_t = x_t' \beta + \epsilon_t$, we might think that $E[y_t \mid x_t] = x_t' \beta$. This implies the moment condition

$$E[x_t \epsilon_t] = E[x_t (y_t - x_t' \beta)] = 0.$$ 

(B) Consider the economic relation

$$y_t = \beta \cdot E[x_{t+1} \mid I_t] + \epsilon_t$$

$$= \beta \cdot x_{t+1} + (\beta \cdot (E[x_{t+1} \mid I_t] - x_{t+1}) + \epsilon_t)$$

Under rational expectation, the expectation error, $E[x_{t+1} \mid I_t] - x_{t+1}$, should be orthogonal to the information set, $I_t$, and for $z_t \in I_t$ we have the moment condition

$$E[z_t \epsilon_t] = 0.$$ 

Properties of GMM

GMM is a large sample estimator.
Desirable properties as $T \to \infty$.

- Consistent under weak assumptions.
  No distributional assumptions like in maximum likelihood (ML) estimation.

- Asymptotically efficient in the class of models that uses the same amount of information.

- Many estimators are special cases of GMM.
  Unifying framework for comparing estimators.

- GMM is a nonlinear procedure.
  We do not need a regression setup $E[y_t] = h(x_t; \beta)$.
  We can have $E[f(y_t, x_t; \beta)] = 0.$
Moment Conditions and MM Estimation

• Consider a variable $y_t$ with some (possibly unknown) distribution. Assume that the mean $\mu = E[y_t]$ exists. We want to estimate $\mu$.

• We could state the population moment condition:

$$E[y_t - \mu] = 0,$$

or

$$E[f(y_t, \mu)] = 0, \quad \text{where} \quad f(y_t, \mu) = y_t - \mu.$$

• The parameter $\mu$ is identified by the condition if there is a unique solution, in the sense

$$E[f(y_t, \mu)] = 0 \quad \text{only if} \quad \mu = \mu_0.$$

• We cannot calculate $E[f(y_t, \mu)]$ from an observed sample, $y_1, y_2, \ldots, y_t, \ldots, y_T$. Define the sample moment condition as

$$g_T(\mu) = \frac{1}{T} \sum_{t=1}^{T} f(y_t, \mu) = \frac{1}{T} \sum_{t=1}^{T} (y_t - \mu) = 0. \quad (*)$$

• By Law of Large Numbers, sample moments converge to population moments,

$$g_T(\mu) \to E[f(y_t, \mu)] \quad \text{for} \quad T \to \infty. \quad (**$$

The method of moments estimator, $\hat{\mu}_{MM}$, is the solution to $(*)$, i.e.

$$\hat{\mu}_{MM} = \frac{1}{T} \sum_{t=1}^{T} y_t.$$

The sample average can be seen as a MM estimator.

• **MM estimator is consistent.** Under weak regularity conditions $(**)$ implies

$$\hat{\mu}_{MM} \to \mu_0.$$
OLS as a MM Estimator

• Consider the regression model with $K$ explanatory variables

$$y_t = x'_t \beta + \epsilon_t.$$  

Assume no-contemporaneous-correlation (minimum for consistency of OLS):

$$E[x_t \epsilon_t] = E[x_t (y_t - x'_t \beta)] = 0 \quad (K \times 1).$$

$K$ moment conditions for the $K$ parameters in $\beta$.

• Define the sample moments

$$g_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} (x_t (y_t - x'_t \beta)) = \frac{1}{T} X' (Y - X \beta) = 0 \quad (K \times 1).$$

The MM estimator is given by the solution

$$\hat{\beta}_{MM} = \left( \sum_{t=1}^{T} x_t x'_t \right)^{-1} \sum_{t=1}^{T} x_t y_t = (X'X)^{-1}X'Y = \hat{\beta}_{OLS}.$$  

---

Instrumental Variables as a MM Estimator

• Consider the regression model

$$y_t = x'_{1t} \beta_1 + x_{2t} \beta_2 + \epsilon_t,$$

where the $K - 1$ variables in $x_{1t}$ are predetermined and $x_{2t}$ is endogenous:

$$E[x_{1t} \epsilon_t] = 0 \quad (K - 1) \times 1$$

$$E[x_{2t} \epsilon_t] \neq 0. \quad \text{(\#)}$$

OLS is inconsistent!

• Assume there exists a variable, $z_{2t}$, such that

$$\text{corr}(x_{2t}, z_{2t}) \neq 0$$

$$E[z_{2t} \epsilon_t] = 0 \quad (1 \times 1) \quad \text{(#\#)}$$

The new moment condition (#\#) can replace (\#).
Define
\[ \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = z_t \begin{pmatrix} x_{1t} \\ z_{2t} \end{pmatrix}. \]

\( z_t \) are called instruments. \( z_{2t} \) is the new instrument; the predetermined variables, \( x_{1t} \), are instruments for themselves.

- The \( K \) population moment conditions are
  \[ E[z_t \varepsilon_t] = E[z_t (y_t - x_t' \beta)] = 0 \quad (K \times 1). \]

The \( K \) corresponding sample moment conditions are
\[ g_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} (z_t (y_t - x_t' \beta)) = \frac{1}{T} Z'(Y - X \beta) = 0 \quad (K \times 1). \]

The MM estimator is given by the unique solution
\[ \hat{\beta}_{MM} = \left( \sum_{t=1}^{T} z_t x_t' \right)^{-1} \sum_{t=1}^{T} z_t y_t = (Z' X)^{-1} Z' Y = \hat{\beta}_{IV}. \]

(Where do Instruments Come From?)

- Consider the two simple equations
  \[ \begin{align*}
  c_t &= \beta_{10} + \beta_{11} y_t + \beta_{12} w_t + \epsilon_{1t} \\
  y_t &= \beta_{20} + \beta_{21} c_t + \beta_{22} w_t + \beta_{23} r_t + \beta_{24} \tau_t + \epsilon_{2t}
  \end{align*} \]

Say that we are only interested in the first equation.

- Assume that \( w_t \) is predetermined. If \( \beta_{21} \neq 0 \), then \( y_t \) is endogenous and
  \[ E[y_t \epsilon_{1t}] \neq 0. \]

- In this setup \( r_t \) and \( \tau_t \) are possible instruments. We need \( \beta_{23} \) and \( \beta_{24} \) different from zero and
  \[ E[(r_t, \tau_t) \epsilon_{1t}] = 0. \]

- In dynamic models we can often used lagged values as instruments.

- Note, that in this case we have more potential instruments than we have endogenous variables. This is adressed in GMM.
The GMM Problem Defined

- Let \( w_t = (y_t, x'_t)' \) be a vector of model variables and let \( z_t \) be instruments. Consider the \( R \) moment conditions

\[
E[f(w_t, z_t, \theta)] = 0.
\]

Here \( \theta \) is a \( K \times 1 \) vector and \( f(\cdot) \) is a \( R \) dimensional vector function.

- Consider the corresponding sample moment conditions

\[
g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \theta) = 0.
\]

- When can the \( R \) sample moments be used to estimate the \( K \) parameters in \( \theta \)?

Order Condition

\( R < K \)

No unique solution to \( g_T(\theta) = 0 \).
The parameters are not identified.

\( R = K \)

Unique solution to \( g_T(\theta) = 0 \).
Exact identification.
This is the MM estimator (OLS, IV).
Note, that \( g_T(\theta) = 0 \) is potentially a non-linear problem—numerical solution.

\( R > K \)

More equations than parameters.
Over-identified case. No solution in general (\( Z'X \) is a \( R \times K \) matrix).
- Not optimal to drop moments!
- Instead, choose \( \theta \) to make \( g_T(\theta) \) as close as possible to zero.
GMM Estimation \( (R > K) \)

- We want to make the \( R \) moments \( g_T(\theta) \) as close to zero as possible...how?

- Assume we have a \( R \times R \) symmetric and positive definite weight matrix \( W_T \). Then we could define the quadratic form

\[
Q_T(\theta) = g_T(\theta)' W_T g_T(\theta) \quad (1 \times 1).
\]

The GMM estimator is defined as the vector that minimizes \( Q_T(\theta) \), i.e.

\[
\hat{\theta}_{GMM}(W_T) = \arg \min \theta \left\{ g_T(\theta)' W_T g_T(\theta) \right\}.
\]

- The matrix \( W_T \) tells how much weight to put on each moment condition. Different \( W_T \) give different estimators,

\[
\hat{\theta}_{GMM}(W_T).
\]

GMM is consistent for any weight matrix, \( W_T \).

What is the optimal choice of \( W_T \)?

---

Optimal GMM Estimation

- The \( R \) sample moments \( g_T(\theta) \) are estimators of \( E[f(\cdot)] \); and random variables. The law of large numbers implies:

\[
g_T(\theta) \to E[f(\cdot)] \quad \text{for} \quad T \to \infty.
\]

A central limit theorem implies:

\[
\sqrt{T} \cdot g_T(\theta) \to N(0, S),
\]

where \( S \) is the asymptotic variance of the moments, \( \sqrt{T} \cdot g_T(\theta) \).

- Intuitively, moments with little variance should have large weights. The optimal weight matrix for GMM is a matrix \( W_{opt}^T \) such that

\[
\lim_{T \to \infty} W_{opt}^T = W_{opt} = S^{-1}.
\]
• Without autocorrelation, a natural estimator \( \hat{S} \) of \( S \) is

\[
\hat{S} = V \left[ \sqrt{T} \cdot g_T(\theta) \right] \\
= T \cdot V [g_T(\theta)] \\
= T \cdot V \left[ \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \theta) \right] \\
= \frac{1}{T} \cdot \sum_{t=1}^{T} f(w_t, z_t, \theta) f(w_t, z_t, \theta)'.
\]

This implies that

\[
W_T^{opt} = \hat{S}^{-1} = \left( \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \theta) f(w_t, z_t, \theta)' \right)^{-1}.
\]

• Note, that \( W_T^{opt} \) depends on \( \theta \) in general.

---

**Estimation in Practice**

• **Two-step (efficient) GMM.**

  1. Choose some initial weight matrix \( W_{[1]} \). E.g. \( W_{[1]} = I \) or \( W_{[1]} = (Z'Z)^{-1} \).

     Find a (consistent) estimate

     \[
     \hat{\theta}_{[1]} = \arg\min_{\theta} g_T(\theta)'W_{[1]}g_T(\theta).
     \]

     Estimate the optimal weights, \( W_T^{opt} \).

  2. Find the optimal GMM estimate

     \[
     \hat{\theta}_{GMM} = \arg\min_{\theta} g_T(\theta)'W_T^{opt}g_T(\theta).
     \]

• **Iterated GMM.**

  Start with some initial weight matrix \( W_{[1]} \).

  1. Find an estimate \( \hat{\theta}_{[1]} \).

  2. Find a new weight matrix, \( W_{[2]}^{opt} \).

     Iterate between \( \hat{\theta}_{[1]} \) and \( W_{[1]}^{opt} \) until convergence.
Properties of Optimal GMM

- The GMM estimator, \( \hat{\theta}_{GMM}(W_{opt}^T) \), is asymptotically efficient.
  Lowest variance in a class of models that uses same information.

- The GMM estimator is asymptotically normal, i.e.
  \[
  \sqrt{T} \cdot (\hat{\theta}_{GMM} - \theta) \rightarrow N(0, V),
  \]
  where
  \[
  V = (D'W_{opt}^T D)^{-1} = (D'S^{-1} D)^{-1},
  \]
  \[
  D = \text{plim} \frac{\partial g_T(\theta)}{\partial \theta'} (R \times K).
  \]
  - \( S \) measures the variance of the moments. The larger \( S \) the larger \( V \).
  - \( D \) measures the sensitivity of the moments wrt. changes in \( \theta \).
    If this is large the parameter can be estimated precisely.

- Little is known in finite samples.

Specification Test

- If \( R > K \), we have more moments than parameters.
  All moments have expectation zero.
  In a sense \( K \) moments are zero by estimating the parameters. Test if the additional \( R - K \) moments are close to zero.
  If not, some orthogonality condition is violated.

- Remember, that
  \[
  \sqrt{T} \cdot g_T(\theta) \rightarrow N(0, S).
  \]
  This implies that if the weights are optimal, \( W_{opt}^T \rightarrow S^{-1} \), then
  \[
  \xi = g_T(\hat{\theta}_{GMM})' \left( \frac{1}{T} S \right)^{-1} g_T(\hat{\theta}_{GMM}) \]
  \[
  = T \cdot g_T(\hat{\theta}_{GMM})' W_{opt}^T g_T(\hat{\theta}_{GMM}) \rightarrow \chi^2(R - K).
  \]
  Hansen test for overidentifying restrictions. (J-test, Sargan test).
  A test for \( R - K \) overidentifying conditions.
Famous Example: Hansen and Singleton (1982)

• Consider an optimizing agent with a power utility function on consumption, 
  \[ U(C_t) = \frac{C_{t}^{1-\gamma}}{1-\gamma} \]. The first order condition for maximizing the discounted utility of future consumption is given by 
  \[ E \left[ \delta \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} (1 + r_{t+1}) - 1 \mid I_t \right] = 0, \]
  where \( I_t \) is the conditioning information set at time \( t \).

• Assume rational expectations. Now if \( z_t \in I_t \), then it must be orthogonal to the expectation error, i.e.
  \[ f(C_{t+1}, C_t, r_{t+1}; z_t; \delta; \gamma) = E \left[ \left( \delta \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} (1 + r_{t+1}) - 1 \right) z_t \right] = 0. \]
  This is a moment condition. We need at least \( R = 2 \) instruments in \( z_t \).

• Note: Specification is theory driven, nonlinear, and not in regression format.

Linear GMM and GIVE

• Consider the linear regression model with \( K \) explanatory variables
  \[ y_t = x'_{1t}\beta_1 + x'_{2t}\beta_2 + \epsilon_t = x'_{t}\beta + \epsilon_t, \]
  where \( E[x_{1t}\epsilon_t] = 0 \), but the variables in \( x_{2t} \) are endogenous \( E[x_{2t}\epsilon_t] \neq 0 \).

• Assume there exist \( R > K \) instruments \( z_t = (x'_1, z'_2) \), such that
  \[ E[z_t\epsilon_t] = E[z_t(y_t - x'_t\beta)] = 0 \] \((R \times 1)\).
  Identification requires a non-zero correlation of \( z_{2t} \) and \( x_{2t} \). Rank condition.

• The sample moments are
  \[ g_R(\beta) = \frac{1}{T} \sum_{t=1}^{T} (z_t(y_t - x'_t\beta)) = \frac{1}{T} Z' (Y - X\beta). \]
  Note, that we cannot solve \( g_R(\beta) = 0 \) directly.
  \( Z'X \) is a \( R \times K \) matrix (of rank \( K \)) and cannot be inverted.
• Instead we want to minimize the quadratic form

\[ Q_T(\beta) = g_T(\beta)' W_T g_T(\beta) \]
\[ = \left( \frac{1}{T} Z'(Y - X\beta) \right)' W_T \left( \frac{1}{T} Z'(Y - X\beta) \right) \]
\[ = \frac{1}{T^2} (Y'Z - \beta'X'Z) W_T (Z'Y - Z'X\beta) \]
\[ = \frac{1}{T^2} (Y'ZW_TZ'Y - 2\beta'X'ZW_TZ'Y + \beta'X'ZW_TZ'X\beta). \]

• To minimize \( Q_T(\beta) \) we take the derivative and solve the \( K \) equations

\[ \frac{\partial Q_T(\beta)}{\partial \beta} = 0 \quad (K \times 1). \]

• The GMM estimator solves the \( K \) equations

\[ \frac{\partial Q_T(\beta)}{\partial \beta} = \frac{\partial \left( T^{-2} \left( Y'ZW_TZ'Y - 2\beta'X'ZW_TZ'Y + \beta'X'ZW_TZ'X\beta \right) \right)}{\partial \beta} \]
\[ = -2T^{-2}X'ZW_TZ'Y + 2T^{-2}X'ZW_TZ'X\beta \]
\[ = 0 \]

i.e.

\[ \hat{\beta}_{GMM}(W_T) = (X'ZW_TZ'X)^{-1} X'ZW_TZ'Y. \]

• The optimal weight matrix is the inverse variance of the moments, i.e.

\[ W_T^{opt} = S^{-1}, \]

where

\[ S = V \left[ \sqrt{T} \cdot g_T(\theta) \right] = \frac{1}{T} V [Z'\epsilon] = \frac{1}{T} E [Z'\epsilon'Z] = \frac{1}{T} Z'\Omega Z, \]

where we define \( E[\epsilon\epsilon'] = \Omega. \)
Case 1: Homoscedastic Errors

- If $E[\varepsilon \varepsilon'] = \Omega = \sigma^2 I$, the natural estimator of $S$ is
  \[
  \hat{S} = \frac{1}{T} Z' \hat{\Omega} Z = \frac{1}{T} \hat{\sigma}^2 Z' Z,
  \]
  where $\hat{\sigma}^2$ is a consistent estimator for $\sigma^2$.

- Then the GMM estimator becomes
  \[
  \hat{\beta}_{GMM} = \left( X' Z \hat{S}^{-1} Z' X \right)^{-1} X' Z \hat{S}^{-1} Z' Y
  \]
  \[
  = \left( X' Z \left( \frac{1}{T} \hat{\sigma}^2 Z' Z \right)^{-1} Z' X \right)^{-1} X' Z \left( \frac{1}{T} \hat{\sigma}^2 Z' Z \right)^{-1} Z' Y
  \]
  \[
  = \left( X' Z (Z' Z)^{-1} Z' X \right)^{-1} X' Z (Z' Z)^{-1} Z' Y
  \]
  \[
  = \hat{\beta}_{GIVE} = \hat{\beta}_{2SLS}
  \]
  Under homoscedasticity the optimal GMM estimator is GIVE (and 2SLS).

- Recall, that
  \[
  \hat{\beta}_{GMM} \rightarrow N \left( \beta, \frac{1}{T} \left( D' W_{opt} D \right)^{-1} \right).
  \]
  The derivative is given by
  \[
  D_T \left( R \times K \right) = \frac{\partial g_T(\beta)}{\partial \beta'} = \frac{\partial \left( \frac{1}{T} Z' (Y - X \beta) \right)}{\partial \beta'} = -\frac{1}{T} Z' X = -\frac{1}{T} \sum_{t=1}^{T} z_t x_t'.
  \]

- The variance can be estimated by
  \[
  V \left[ \hat{\beta}_{GMM} \right] = \frac{1}{T} \left( D_T' W_{opt} D_T \right)^{-1}
  \]
  \[
  = \frac{1}{T} \left( \left( -\frac{1}{T} Z' X \right)' \left( \frac{1}{T} \hat{\sigma}^2 Z' Z \right)^{-1} \left( -\frac{1}{T} Z' X \right) \right)^{-1}
  \]
  \[
  = \hat{\sigma}^2 (X' Z (Z' Z)^{-1} Z' X)^{-1},
  \]
  known from 2SLS.
• The specification test is given by

\[ \xi = T \cdot g_T(\hat{\theta}_{GMM})' \hat{S}^{-1} g_T(\hat{\theta}_{GMM}) \]

\[ = T \cdot \left( \sum_{t=1}^{T} \tilde{e}_t z_t \right)' \left( \frac{\sigma^2}{T} \sum_{t=1}^{T} z_t' z_t \right)^{-1} \left( \sum_{t=1}^{T} \tilde{e}_t z_t \right) \]

\[ = \left( \sum_{t=1}^{T} \tilde{e}_t z_t \right)' \left( \frac{\sigma^2}{T} \sum_{t=1}^{T} z_t' z_t \right)^{-1} \left( \sum_{t=1}^{T} \tilde{e}_t z_t \right) \]

\[ \rightarrow \chi^2(R - K). \]

In the linear case it is denoted the Sargan test.

• A simple way to calculate \( \xi \) is to consider the regression

\[ \tilde{e}_t = z_t' \gamma + \text{residual}, \]

and calculate the test statistic as \( \xi = T \cdot R^2. \)

Case 2: Heteroscedasticity, No Autocorrelation

• In the case of heteroscedasticity but no autocorrelation,

\[ E[\epsilon\epsilon'] = \Omega = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \sigma_T^2 \end{pmatrix}, \]

we can use the estimator

\[ \hat{S} = \frac{1}{T} Z' \hat{\Omega} Z = \frac{1}{T} \sum_{t=1}^{T} \tilde{e}_t^2 z_t z_t'. \]

We only need an estimate of the \( K \times K \) matrix \( Z' \Omega Z \) and not the \( T \times T \) matrix \( \Omega \). We get

\[ \beta_{GMM}(\hat{S}^{-1}) = \left( X' Z' \hat{S}^{-1} Z' X \right)^{-1} X' Z' \hat{S}^{-1} Z' Y. \]

Note, that a constant in the weight is not important for the estimation.
• The variance of the estimator becomes

\[
V \left[ \hat{\beta}_{GMM} \right] = \frac{1}{T} (D_T W_{opt} D_T)^{-1}
\]

\[
= \frac{1}{T} \left( \left( -\frac{1}{T} \sum_{t=1}^{T} x_t' z_t' \right) \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{\epsilon}_t^2 z_t' z_t' \right)^{-1} \left( -\frac{1}{T} \sum_{t=1}^{T} z_t x_t' \right) \right)^{-1}
\]

\[
= \left( \sum_{t=1}^{T} x_t' z_t' \right)^{-1} \sum_{t=1}^{T} \tilde{\epsilon}_t^2 z_t^{-1} \left( \sum_{t=1}^{T} z_t x_t' \right)^{-1},
\]

which is the heteroscedasticity consistent variance estimator of White.

---

Case 3: Autocorrelation

• The weight-matrix is \( W_{opt} = \hat{S}^{-1} \), where

\[
\hat{S} = V \left[ \sqrt{T} \cdot g_T(\theta) \right] = T^{-1} V \left[ \sum_{t=1}^{T} (z_t \epsilon_t) \right].
\]

With autocorrelation, we need to take into account the covariances.

• This is done by the heteroscedasticity and autocorrelation consistent (HAC) estimator. Let

\[
\Gamma_j = \text{cov}(z_t \epsilon_t, z_{t-j} \epsilon_{t-j}) = E \left[ (z_t \epsilon_t)(z_{t-j} \epsilon_{t-j})' \right]
\]

be a covariance matrix for lag \( j \). Then

\[
T \cdot S = T^{-1} \left\{ V(z_t \epsilon_t) + \text{Cov}(z_t \epsilon_t, z_{t-1} \epsilon_{t-1}) + \text{Cov}(z_t \epsilon_t, z_{t-2} \epsilon_{t-2}) + \ldots \right.
\]

\[
+ \text{Cov}(z_t \epsilon_t, z_{t+1} \epsilon_{t+1}) + \text{Cov}(z_t \epsilon_t, z_{t+2} \epsilon_{t+2}) + \ldots \left\}
\]

\[
= T^{-1} \sum_{j=-\infty}^{\infty} \Gamma_j.
\]
• If we can argue that $\Gamma_j = 0$ for $j$ larger than some lag, $q$, we can use the estimator

$$\hat{S} = T^{-1} \sum_{j=-q}^{q} \hat{\Gamma}_j,$$

where we estimate the covariances by

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} (z_t \hat{e}_t)(z_{t-j} \hat{e}_{t-j})'.$$

• The obtained $\hat{S}$ is not necessarily positive definite. Instead the covariances can be given decreasing weights, Newey-West estimator. Finite sample properties are unknown.

• The HAC covariance estimator can also be used for OLS.