θ is a $k \times 1$ parameter vector to be estimated.

w_t is an observed vector $w_t = (y_t, x_t)$.

$h(\theta; w_t)$ is a $r \times 1$ vector function, where $r \geq k$.

Define $g(\theta; W_T)$ as follows:

$$g(\theta; W_T) = \frac{1}{T} \sum_{t=1}^{T} h(\theta; w_t),$$

where $W_T = \{w_T, w_{T-1}, \cdots, w_1\}$.

Compute:

$$\min_{\theta} \; g(\theta; W_T)' S^{-1} g(\theta; W_T)$$

The solution of θ, denoted by $\hat{\theta}_T$, corresponds to the GMM estimator, where S is defined as follows:

$$S = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=-\infty}^{\infty} \mathrm{E}\left(h(\theta; w_t) h(\theta; w_{t-\tau})'\right).$$

In empirical studies, S is replaced by its estimate, i.e., $\hat{S}_T$.

When $h(\theta; w_t)$, $t = 1, 2, \cdots, T$, are not serially correlated, the following $\hat{S}_T$ is consistent, i.e.,

$$\hat{S}_T = \frac{1}{T}\sum_{t=1}^{T} h(\hat{\theta}_T; w_t) h(\hat{\theta}_T; w_t)' \longrightarrow S.$$

When $h(\theta; w_t)$, $t = 1, 2, \cdots, T$, are serially correlated,

$$\hat{S}_T = \hat{\Gamma}(0) + \sum_{\tau=1}^{q} k(\frac{\tau}{q+1})(\hat{\Gamma}(\tau) + \hat{\Gamma}(\tau)'),$$

where $\hat{\Gamma}(\tau) = \dfrac{1}{T}\displaystyle\sum_{t=\tau+1}^{T} h(\hat{\theta}_T; w_t) h(\hat{\theta}_T; w_{t-s})'$.

$k(x) = 1 - x \implies$ Bartlett kernel (Newwey-west estimator),

$k(x) \implies$ Parzen kernel, and etc.

Then, we obtain:

$$\sqrt{T}(\hat{\theta}_T - \theta) \longrightarrow N\left(0, (DS^{-1}D')^{-1}\right),$$

where

$$D = \frac{\partial g(\theta; W_T)}{\partial \theta'}.$$

Note that D is a $r \times k$ matrix.

Let $\hat{D}_T$ be an estimate of D.

The variance estimator of $\hat{\theta}_T$ is given by:

$$\hat{D}_T = \frac{\partial g(\hat{\theta}_T; W_T)}{\partial \theta'}.$$

Asymptotic Normality:

$$\text{Assumption 1}: \quad \hat{\theta}_T \longrightarrow \theta,$$

$$\text{Assumption 2}: \quad \sqrt{T} g(\theta; W_T) \longrightarrow N(0, S).$$

Then, we have the following first-order approximation:

$$\begin{aligned} g(\theta; W_T) &\approx g(\hat{\theta}_T; W_T) + \frac{\partial g(\hat{\theta}_T; W_T)}{\partial \theta'}(\theta - \hat{\theta}_T) \\ &= g(\hat{\theta}_T; W_T) + \hat{D}_T(\theta - \hat{\theta}_T), \end{aligned}$$

where $g(\theta; W_T)$ is linearized around $\theta = \hat{\theta}_T$.

The first-order condition for the minimization problem is:

$$\left(\frac{\partial g(\theta; W_T)}{\partial \theta'}\right)' S^{-1}\Big(g(\theta; W_T)\Big) = 0.$$

Substituting the approximation into the above equation, we obtain the following:

$$D'S^{-1}\Big(g(\theta; W_T)\Big) = D'S^{-1}\Big(g(\hat{\theta}_T; W_T) + \hat{D}_T(\theta - \hat{\theta}_T)\Big)$$
$$= D'S^{-1}g(\hat{\theta}_T; W_T) + D'S^{-1}\hat{D}_T(\theta - \hat{\theta}_T).$$

Therefore,

$$\sqrt{T}(\hat{\theta}_T - \theta) \approx (D'S^{-1}\hat{D}_T)^{-1}D'S^{-1}\sqrt{T}\Big(g(\hat{\theta}_T; W_T) - g(\theta; W_T)\Big).$$

Thus, GMM estimator, $\hat{\theta}_T$, has the following asymptotic distribution:

$$\sqrt{T}(\hat{\theta}_T - \theta) \longrightarrow N\Big(0, (D'S^{-1}D)^{-1}\Big),$$

where $\hat{D}_T \longrightarrow D$ is utilized.

From Assumption 2, we have the following asymptotic distribution:

$$\Big(\sqrt{T}g(\theta; W_T)\Big)' S^{-1}\Big(\sqrt{T}g(\theta; W_T)\Big) \longrightarrow \chi^2(r).$$

When θ is replaced by GMM estimator $\hat{\theta}_T$, we have the following distribution:

$$\left(\sqrt{T}g(\hat{\theta}_T; W_T)\right)' \hat{S}_T^{-1} \left(\sqrt{T}g(\hat{\theta}_T; W_T)\right) \longrightarrow \chi^2(r-k),$$

which is called a test of the overidentifying restrictions.

$\Longrightarrow$ J test by Hansen (1982)

k linear combinations consisting of a $r \times 1$ vector $g(\hat{\theta}_T; W_T)$ are zeros.

Therefore, the degrees of freedom are $r-k$.

Some Examples:

(a) **OLS:**

Regression Model: $\quad y_t = x_t\beta + \epsilon_t, \quad \mathrm{E}(x_t\epsilon_t) = 0$

$h(\theta; w_t)$ is taken as:

$$h(\theta; w_t) = x_t(y_t - x_t\beta).$$

(b) **IV (Instrumental Variable, 操作変数法):**

Regression Model: $y_t = x_t\beta + \epsilon_t, \quad \mathrm{E}(x_t\epsilon_t) \neq 0, \quad \mathrm{E}(z_t\epsilon_t) = 0$

$h(\theta; w_t)$ is taken as:

$$h(\theta; w_t) = z_t(y_t - x_t\beta),$$

where z_t is a vector of instrumental variables.

(c) **NLS (Nonlinear Least Squares,** 非線形最小二乗法):

Regression Model: $f(y_t, x_t, \beta) = \epsilon_t, \quad \mathrm{E}(x_t\epsilon_t) \neq 0, \quad \mathrm{E}(z_t\epsilon_t) = 0$

$h(\theta; w_t)$ is taken as:

$$h(\theta; w_t) = z_t f(y_t, x_t, \beta)$$

where z_t is a vector of instrumental variables.

Example: Demand function using STATA

二人以上の世帯のうち勤労者世帯（全国）

```
year
y  ＝ 実収入（一月当たり，実質データ）
q1 ＝ 穀類支出額（一年当たり，実質データ）
p1 ＝ 穀類価格　（相対価格＝穀類 CPI ／総合 CPI）
p2 ＝ 魚介類価格（相対価格＝魚介類 CPI ／総合 CPI）
p3 ＝ 肉類価格　（相対価格＝肉類 CPI ／総合 CPI）
```

year	y	q1	p1	p2	p3
2000	567865	7087.0	1.043390	0.884965	0.818365
2001	561722	6993.1	1.032520	0.886179	0.822154
2002	553768	6934.4	1.031800	0.891282	0.834872
2003	539928	6816.8	1.050410	0.876543	0.843621
2004	547006	6651.6	1.089510	0.865226	0.868313
2005	541367	6615.8	1.020640	0.862745	0.887513
2006	540863	6523.7	1.000000	0.878601	0.891975
2007	543994	6680.5	0.994856	0.886831	0.908436
2008	541821	6494.7	1.043610	0.894523	0.932049
2009	533154	6477.3	1.066870	0.898148	0.934156
2010	539577	6458.2	1.040420	0.889119	0.924352
2011	529750	6448.4	1.025960	0.894081	0.925234
2012	538988	6377.6	1.057170	0.904366	0.917879
2013	542018	6360.7	1.047620	0.909938	0.916149
2014	523953	6174.6	1.016130	0.971774	0.960685
2015	525669	6268.0	1.000000	1.000000	1.000000
2016	527501	6244.8	1.018020	1.019020	1.017020

```
2017 531693 6106.6 1.027890 1.066730 1.025900


. tsset year
        time variable:  year, 2000 to 2017
                delta:  1 unit

. reg q1 y p1 p2 p3 if year>2000.5

      Source |       SS           df       MS      Number of obs   =        17
-------------+----------------------------------   F(4, 12)        =     25.83
       Model |  913640.443         4  228410.111   Prob > F        =    0.0000
    Residual |  106100.077        12  8841.67308   R-squared       =    0.8960
-------------+----------------------------------   Adj R-squared   =    0.8613
       Total |  1019740.52        16  63733.7825   Root MSE        =     94.03

------------------------------------------------------------------------------
          q1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |   .0067843   .0045443     1.49   0.161     -.003117    .0166856
          p1 |  -1128.834   998.7698    -1.13   0.280    -3304.966    1047.299
          p2 |   356.8095   806.2301     0.44   0.666    -1399.815    2113.434
          p3 |  -3442.221   1130.078    -3.05   0.010    -5904.448   -979.9931
       _cons |   6850.563   3179.316     2.15   0.052    -76.57278     13777.7
------------------------------------------------------------------------------

. gmm (q1-{b0}-{b1}*y-{b2}*p1-{b3}*p2-{b4}*p3) if year>2000.5, instruments(y p1
> p2 p3)
```

```
Step 1
Iteration 0:   GMM criterion Q(b) =   42400764
Iteration 1:   GMM criterion Q(b) =  6.781e-12
Iteration 2:   GMM criterion Q(b) =  6.781e-12  (backed up)

Step 2
Iteration 0:   GMM criterion Q(b) =  1.966e-15
Iteration 1:   GMM criterion Q(b) =  1.963e-15  (backed up)
convergence not achieved
    The Gauss-Newton stopping criterion has been met but missing standard errors
    indicate some of the parameters are not identified.

GMM estimation

Number of parameters =   5
Number of moments    =   5
Initial weight matrix: Unadjusted                 Number of obs   =         17
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |   6850.563   17645.71     0.39   0.698    -27734.4    41435.53
         /b1 |   .0067843   .0282325     0.24   0.810   -.0485504     .062119
         /b2 |  -1128.834   1057.915    -1.07   0.286   -3202.309    944.6415
         /b3 |   356.8095    1565.86     0.23   0.820   -2712.219    3425.838
         /b4 |  -3442.221   5085.561    -0.68   0.498   -13409.74    6525.296
------------------------------------------------------------------------------
```

```
Instruments for equation 1: y p1 p2 p3 _cons
Warning: convergence not achieved

. gmm (q1-{b0}-{b1}*y-{b2}*p1-{b3}*p2-{b4}*p3) if year>2000.5, instruments(p1 p2
> p3 l.p1 l.p2 l.p3)

Step 1
Iteration 0:   GMM criterion Q(b) =   42404066
Iteration 1:   GMM criterion Q(b) =  2790.3146
Iteration 2:   GMM criterion Q(b) =  2790.3146

Step 2
Iteration 0:   GMM criterion Q(b) =   .3201826
Iteration 1:   GMM criterion Q(b) =   .2469289
Iteration 2:   GMM criterion Q(b) =   .2469289

GMM estimation

Number of parameters =   5
Number of moments    =   7
Initial weight matrix: Unadjusted                 Number of obs   =         17
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |  -1192.466   4669.012    -0.26   0.798    -10343.56     7958.63
         /b1 |   .0186312   .0067682     2.75   0.006     .0053657    .0318967
```

```
         /b2 |  -1016.864    780.979    -1.30   0.193    -2547.554    513.8271
         /b3 |  -905.5585   598.0885    -1.51   0.130     -2077.79    266.6734
         /b4 |  -499.8064   1147.985    -0.44   0.663    -2749.815    1750.202
------------------------------------------------------------------------------
Instruments for equation 1: p1 p2 p3 L.p1 L.p2 L.p3 _cons

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(2) = 4.19779 (p = 0.1226)
```