Title

gmmhar—Two-step efficient gmm estimation with HAR standard errors

Syntax Menu Description Options Remarks and example Stored results Methods and formula References Also see

syntax

gmmhar depvar [varlist1] (varlist2=instlist)[if] [in] [, noconstant level(#)]

options	Description
Model	
noconstant	suppress the constant term
Reporting	
level(#)	set the confidence level; default is level(95)

you must tsset your data before using gmmhar; see [TS] tsset
time-series operators are allowed.

Menu

statistics>Time series > Two step gmm estimation with HAR standard errors

Description

Sun (2014) considers the more accurate fixed-smoothing asymptotics in the two-step efficient GMM framework where the weighting matrix and asymptotic variance matrix are based on the orthonormal series long run variance estimator. Hwang and Sun (2017) propose some modifications to the usual test statistics, including the Wald statistic, the quasi-LR type statistic (difference in the GMM criterion functions), and the LM type (score type) statistic and show that the modified test statistics are all asymptotically standard F distributed under the fixed-smoothing asymptotics. The modified statistics are rescaled versions of the original test statistics with the scaling factors depending on the J statistic (Sun and Kim, 2012) for testing over-identifying restrictions.

Options

Model

noconstant; see [R] estimation options.

Reporting

level(#); see [R] estimation options.

Remarks and examples

Hwang and Sun (2017) construct HAR standard errors for two-step GMM estimation and develop asymptotic F approximation for the Wald statistic where the weighting matrix and asymptotic variance matrix are based on the orthonormal series long-run variance estimator. The data-driven choice of K, the number of orthonormal bases, is selected by the AMSE method implemented via the VAR(1) plug-in.

Example 1

To illustrate the use of gmmhar, we estimate a quarterly time-series model relating the change in the U.S. inflation rate (D.inf) to the unemployment rate (UR) for 1959q1–2000q4. As instruments, we use the second lag of quarterly GDP growth, the lagged values of the Treasury bill rate, the trade-weighted exchange rate, and the Treasury medium-term bond rate. We fit our model using the two-step efficient GMM method.

```
. use http://fmwww.bc.edu/ec-p/data/stockwatson/macrodat

.
. generate inf =100 * log( CPI / L4.CPI )
(4 missing values generated)

.
. generate ggdp=100 * log( GDP / L4.GDP )
(10 missing values generated)

.
. gmmhar D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON)

Two-step Efficient GMM Estimation                     Number of obs   =        158
Data-driven optimal K: 46                             F(  1,     43)  =       2.05
                                                      Prob > F        =     0.1597
```

D.inf	Coef.	HAR std.Err.	t	df	P>|t|	[95% Conf.	Interval]
UR	-.0971458	.067901	-1.43	43	0.160	-.2340812	.0397895
_cons	.5631061	.3936908	1.43	43	0.160	-.2308471	1.357059

```
HAR J statistic   =  .92614349
Reference Dist for the J test: F(  3,     44)
P-value of the J test =     0.4361
Instrumented: UR
Instruments: L2.ggdp L.TBILL L.ER L.TBON
```

In this case, the header reports data-driven optimal K by the usual AMSE method.

Example 2

nonparametric orthonormal series approach, noconstant, level case 99%, AMSE automatic bandwidth selection.

```
. gmmhar D.inf (UR=L2.ggdp L.TBILL L.ER L.TBON),nocons l(99)

Two-step Efficient GMM Estimation                    Number of obs   =        158
Data-driven optimal K: 40                            F(  1,     37)  =       0.01
                                                     Prob > F        =     0.9119
```

D.inf	Coef.	HAR std.Err.	t	df	P>|t|	[99% Conf.	Interval]
UR	.0014583	.0130865	0.11	37	0.912	-.0340768	.0369934

```
HAR J statistic  =  .95768181
Reference Dist for the J test: F(  3,    38)
P-value of the J test =    0.4226
Instrumented: UR
Instruments: L2.ggdp L.TBILL L.ER L.TBON
```

Stored results

The gmmhar uses ivregress to get the colname of e(b). So, in addition to the standard stored results from ivregress, gmmhar also stores the following results in e():

Scalars

e(N)	the number of observations
e(sF)	the adjusted F statistic
e(sfdf)	the first degrees of freedom
e(ssdf)	the second degrees of freedom
e(kopt)	the data-driven optimal K for the OS long run variance estimator
e(J)	the J statistic for testing the overidentifying restrictions

Macros

e(cmd)	gmmhar
e(cmdline)	command as typed
e(varline)	variable line as typed
e(carg)	nocons or " " if specified
e(title)	title in estimation output
e(vcetype)	orthonormal series
e(depvar)	name of dependent variable
e(exog)	exogenous variables
e(endog)	endogenous variables
e(inst)	instrument variables

Matrices

e(betahat)	the two-step gmm coefficient vector
e(sstderr)	the adjusted standard error for each individual coefficient
e(sdf)	the degrees of freedom of the t statistic
e(st)	the t statistic for testing a single restriction.

Functions

e(sample) marks the estimation sample

*the other results in ereturn list see[R] ivregress

Methods and formulas

Consider the regression model:

$$Y_t = X_t\theta_0 + e_t, t = 1,2,\cdots,T$$

where $\{e_t\}$ is a zero-mean process that may be correlated with the covariate process $\{X_t \in R^{1\times d}\}$. There are instruments $\{Z_t \in R^{1\times m}\}$ such that the moment conditions:

$$EZ_t'(Y_t - X_t\theta_0) = 0$$

hold if and only if $\theta = \theta_0$. We allow the process $\{Z_t'e_t\}$ to have the autocorrelation of unknown forms. The model may be over-identified with the degree of over identification $q = m - d \geq 0$.

Define: $S_{ZX} = \frac{1}{T}\sum_{t=1}^{T} Z_t' X_t$, $S_{ZZ} = \frac{1}{T}\sum_{t=1}^{T} Z_t' Z_t$, $S_{ZY} = \frac{1}{T}\sum_{t=1}^{T} Z_t' Y_t$.

Then the IV estimator of θ_0 is

$$\hat{\theta}_{IV} = [S_{ZX}' W_{0T}^{-1} S_{ZX}]^{-1}[S_{ZX}' W_{0T}^{-1} S_{ZY}],$$

where $W_{0T} = S_{ZZ} \in R^{m\times m}$.

Hwang and Sun (2017) consider two-step efficient GMM estimation and inference where the weighting matrix and asymptotic variance matrix are based on the orthonormal series long run variance estimator. In its general form, the two-step GMM estimator is given by

$$\hat{\theta}_{GMM} = arg \min_{\theta\in\Theta} g_T(\theta)' w_T^{-1}(\hat{\theta}_{IV}) g_T(\theta) = \{S_{ZX}'[w_T(\hat{\theta}_{IV})]^{-1} S_{ZX}\}^{-1}\{S_{ZX}'[w_T(\hat{\theta}_{IV})]^{-1} S_{ZY}\}$$

where

$$w_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_K\left(\frac{t}{T},\frac{s}{T}\right)\left(v_t(\theta) - \bar{v}(\theta)\right)\left(v_t(\theta) - \bar{v}(\theta)\right)', \quad v_t(\theta) = Z_t'(Y_t - X_t\theta),$$

and $\bar{v}(\theta) = \sum_{t=1}^{T} v_t(\theta)/T$. Note that $w_T(\hat{\theta}_{IV})$ is a quadratic estimator of the long run variance of moment process $\{v_t(\theta_0)\}$. $Q_K(\mathrm{r},\mathrm{s})$ is a symmetric weighting function that depends on the smoothing parameter K. Hwang and Sun (2017) focus on the orthonormal series LRV estimation with $Q_K(\mathrm{r},\mathrm{s}) = \frac{1}{K}\sum_{j=1}^{K}\phi_j(\mathrm{r})\,\phi_j(\mathrm{s})$, where $\{\phi_j(\cdot)\}_{j=1}^{K}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \phi_j(r)\,dr = 0$ for $j = 1,2,\cdots,K$.

We are interested in testing the null $H_0\colon R\theta_0 = r$ against the alternative $H_1\colon R\theta_0 \neq r$, where $r \in R^{p\times 1}$ and $R \in R^{p\times d}$ is a matrix of full row rank. Nonlinear restrictions can be converted into linear ones via the delta method. The Wald statistic is given by

$$\mathbb{W}_T(\hat{\theta}_{GMM}) = \sqrt{T}(R\hat{\theta}_{GMM} - r)'\{R(G_T' w_T^{-1}(\hat{\theta}_{GMM})G_T)^{-1}R'\}\sqrt{T}(R\hat{\theta}_{GMM} - r)/p$$

where $G_T = S_{ZX}$. Under the assumptions in Hwang and Sun (2017), the following result holds:

$$\mathbb{W}_T^C(\hat{\theta}_{GMM}) \xrightarrow{d} F_{p,K-p-q+1},$$

where

$$\mathbb{W}_T^C(\hat{\theta}_{GMM}) = \frac{K-p-q+1}{K}\frac{\mathbb{W}_T(\hat{\theta}_{GMM})}{1+\frac{1}{K}J_T(\hat{\theta}_{GMM})}$$

is the modified Wald statistic and

$$J_T(\hat{\theta}_{GMM}) = \frac{K-q+1}{K} \cdot \frac{Tg_T(\hat{\theta}_{GMM})' w_T^{-1}(\hat{\theta}_{GMM}) g_T(\hat{\theta}_{GMM})}{q}$$

is the J statistics for testing overidentifying restrictions. Sun and Kim (2012) show that $J_T(\hat{\theta}_{GMM})$ converges in distribution to $F(q, K-q+1)$. Following Sun and Kim (2012) and Hwang and Sun (2017), we select K based on the AMSE criterion implemented using the VAR(1) plug-in procedure.

References

Huang, J. and Y. Sun, 2017. Asymptotic F and t Tests in an Efficient GMM Setting. Journal of Econometrics, 198(2):277-295.

Sun, Y., 2014. Fixed-smoothing Asymptotics in a Two-step GMM Framework. Econometrica, 82:2327-2370.

Sun, Y. and M.S. Kim, 2012. Simple and Powerful GMM Over-identification Tests with Accurate Size. Journal of Econometrics, 166(2):267-281.

Also see

[TS]tsset—Declare data to be time-series data.