Title

estat sbsingle — Test for a structural break with an unknown break date

Description

`estat sbsingle` performs a test of whether the coefficients in a time-series regression vary over the periods defined by an unknown break date.

`estat sbsingle` requires that the current estimation results be from `regress` or `ivregress 2sls`.

Quick start

Supremum Wald test for a structural break at an unknown break date for current estimation results using default symmetric trimming of 15%

```
estat sbsingle
```

Same as above

```
estat sbsingle, swald
```

As above, but also report average Wald test

```
estat sbsingle, swald awald
```

Supremum Wald test with symmetric trimming of 20%

```
estat sbsingle, trim(20)
```

As above, but use asymmetric trimming with a left trim of 10% and a right trim of 20%

```
estat sbsingle, ltrim(10) rtrim(20)
```

Menu for estat

Statistics > Postestimation

Syntax

```
estat sbsingle [, options]
```

options	Description
breakvars([*varlist*][, constant])	specify variables to be included in the test; by default, all coefficients are tested
trim(#)	specify a trimming percentage; default is trim(15)
ltrim($\#_l$)	specify a left trimming percentage
rtrim($\#_r$)	specify a right trimming percentage
swald	request a supremum Wald test; the default
awald	request an average Wald test
ewald	request an exponential Wald test
all	report all tests
slr	request a supremum likelihood-ratio (LR) test
alr	request an average LR test
elr	request an exponential LR test
generate(*newvarlist*)	create *newvarlist* containing Wald or LR test statistics
nodots	suppress iteration dots

You must tsset your data before using estat sbsingle; see [TS] **tsset**.

Options

breakvars([*varlist*][, constant]) specifies variables to be included in the test. By default, all the coefficients are tested.

constant specifies that a constant be included in the list of variables to be tested. constant may be specified only if the original model was fit with a constant term.

trim(#) specifies an equal left and right trimming percentage as an integer. Specifying trim(#) causes the observation at the #th percentile to be treated as the first possible break date and the observation at the $(100 - \#)$th percentile to be treated as the last possible break date. By default, the trimming percentage is set to 15 but may be set to any value between 1 and 49.

ltrim($\#_l$) specifies a left trimming percentage as an integer. Specifying ltrim($\#_l$) causes the observation at the $\#_l$th percentile to be treated as the first possible break date. This option must be specified with rtrim($\#_r$) and may not be combined with trim(#). $\#_l$ must be between 1 and 99.

rtrim($\#_r$) specifies a right trimming percentage as an integer. Specifying rtrim($\#_r$) causes the observation at the $(100 - \#_r)$th percentile to be treated as the last possible break date. This option must be specified with ltrim($\#_l$) and may not be combined with trim(#). $\#_r$ must be less than $(100 - \#_l)$. Specifying $\#_l = \#_r$ is equivalent to specifying trim(#) with $\# = \#_l = \#_r$.

swald requests that a supremum Wald test be performed. This is the default.

awald requests that an average Wald test be performed.

ewald requests that an exponential Wald test be performed.

all specifies that all tests be displayed in a table.

slr requests that a supremum LR test be performed.

alr requests that an average LR test be performed.

elr requests that an exponential LR test be performed.

generate(*newvarlist*) creates either one or two new variables containing the Wald statistics, LR statistics, or both that are transformed and used to calculate the requested Wald or LR tests. If you request only Wald-type tests (swald, awald, or ewald) or only LR-type tests (slr, alr, or elr), then you may specify only one *varname* in generate(). By default, *newvar* will contain Wald or LR statistics, depending on the type of test specified.

A variable containing Wald statistics and a variable containing LR statistics are created if you specify both Wald-type and LR-type tests and specify two *varnames* in generate(). If you only specify one *varname* in generate() with Wald-type and LR-type tests specified, then Wald statistics are returned.

If no test is specified and generate() is specified, Wald statistics are returned.

nodots suppresses display of the iteration dots. By default, one dot character is displayed for each iteration in the range of possible break dates.

Remarks and examples

stata.com

estat sbsingle constructs a test statistic for a structural break without imposing a known break date by combining the test statistics computed for each possible break date in the sample. estat sbsingle uses the maximum, an average, or the exponential of the average of the tests computed at each possible break date. The test at each possible break date can be either a Wald or an LR test.

The limiting distribution of each of these test statistics is known but nonstandard. Not only is each test statistic a function of many sample statistics, but each of these test statistics also depends on the unknown break date, which is not identified under the null hypothesis; see Davies (1987) for a seminal treatment.

Tests that use the maximum of the sample tests are known as supremum tests. Specifically, the supremum Wald test uses the maximum of the sample Wald tests, and the supremum LR test uses the maximum of the sample LR tests. The intuition behind these tests is to compare the maximum sample test with what could be expected under the null hypothesis of no break (Quandt [1960], Kim and Siegmund [1989], and Andrews [1993]).

Supremum tests have much less power than average tests and exponential tests. Average tests use the average of the sample tests, and exponential tests use the natural log of the average of the exponential of the sample tests. An average test is optimal when the alternative hypothesis is a small change in parameter values at the structural break. An exponential test is optimal when the alternative hypothesis is a larger structural break. See Andrews and Ploberger (1994) for details about the properties of average and exponential tests.

All tests implemented in estat sbsingle are a function of the sample statistics computed over a range of possible break dates. However, not all sample observations can be tested as break dates because there are insufficient observations to estimate the parameters for dates too near the beginning or the end of the sample. This identification problem is solved by trimming, which excludes observations too near the beginning or the end of the sample from the set of possible break dates. Andrews (1993) recommends a symmetric trimming of 15% when the researcher has no other information on good trimming values.

Much research went into deriving the properties of the implemented tests, and we have cited only a few of the many papers on the subject. See Perron (2006) for an excellent survey.

▷ Example 1: Test for a structural break with unknown break date

In `usmacro.dta`, we have data for the `fedfunds` series from the third quarter of 1954 to the fourth quarter of 2010 from the Federal Reserve Economic Database (FRED) provided by the Federal Reserve Bank of Saint Louis.

Consider a model for the federal funds rate as a function of its first lag and the inflation rate (`inflation`). We want to test whether coefficients changed at an unknown break date. Below, we fit the model using `regress` and perform the test using `estat sbsingle`.

```
. use http://www.stata-press.com/data/r15/usmacro
(Federal Reserve Economic Data - St. Louis Fed)

. regress fedfunds L.fedfunds inflation
  (output omitted)

. estat sbsingle
────┼──── 1 ────┼──── 2 ────┼──── 3 ────┼──── 4 ────┼──── 5
..................................................    50
..................................................   100
..................................................   150
....
Test for a structural break: Unknown break date

                             Number of obs =       222

Full sample:                 1955q3 - 2010q4
Trimmed sample:              1964q1 - 2002q3
Estimated break date:        1980q4
Ho: No structural break

     Test             Statistic          p-value
───────────────────────────────────────────────

     swald             14.1966            0.0440
───────────────────────────────────────────────

Exogenous variables:           L.fedfunds inflation
Coefficients included in test: L.fedfunds inflation _cons
```

By default, a supremum Wald test is performed. The output indicates that we reject the null hypothesis of no structural break at the 5% level and that the estimated break date is `1980q4`.

Some researchers perform more than one test. Below, we present results for the supremum Wald, average Wald, and average LR tests.

```
. estat sbsingle, swald awald alr
────┼──── 1 ────┼──── 2 ────┼──── 3 ────┼──── 4 ────┼──── 5
..................................................    50
..................................................   100
..................................................   150
....
Test for a structural break: Unknown break date

                             Number of obs =       222

Full sample:                 1955q3 - 2010q4
Trimmed sample:              1964q1 - 2002q3
Ho: No structural break

     Test             Statistic          p-value
───────────────────────────────────────────────

     swald             14.1966            0.0440
     awald              4.5673            0.1474
       alr              4.6319            0.1411
───────────────────────────────────────────────

Exogenous variables:           L.fedfunds inflation
Coefficients included in test: L.fedfunds inflation _cons
```

Only the supremum Wald test rejects the null hypothesis of no break.

◁

▷ Example 2: Testing for a structural break in a subset of coefficients

Below, we test the null hypothesis that there is a break in the intercept, when we assume that there is no break in either the autoregressive coefficient or the coefficient on inflation.

```
. estat sbsingle, breakvars(, constant)
────┼──── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
..................................................    50
..................................................   100
..................................................   150
....
Test for a structural break: Unknown break date

                             Number of obs =        222

Full sample:                 1955q3 - 2010q4
Trimmed sample:              1964q1 - 2002q3
Estimated break date:        2001q1
Ho: No structural break

     Test            Statistic          p-value
───────────────────────────────────────────────

     swald             6.7794            0.1141
───────────────────────────────────────────────
Exogenous variables:           L.fedfunds inflation
Coefficients included in test:  _cons
```

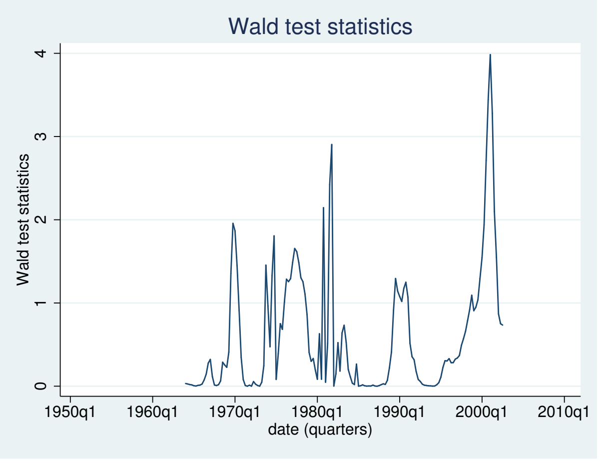
We fail to reject the null hypothesis of no structural break in the intercept when there is no break in any other coefficient.

◁

▷ Example 3: Reviewing sample test statistics

The observation-level Wald or LR test statistics sometimes provide useful diagnostic information. Below, we use the `generate()` option to store the observation-level Wald statistics in the new variable `wald`, which we subsequently plot using `tsline`.

```
. estat sbsingle, breakvars(L.fedfunds) generate(wald)
  (output omitted)
. tsline wald, title("Wald test statistics")
```



◁

We see a spike in the value of the test statistic at the estimated break date of 1980q4. The bump to left of the spike may indicate a second break.

▷ Example 4: Structural break test with an endogenous regressor

We can use `estat sbsingle` to test for a structural break in a regression with endogenous variables. Suppose we want to estimate the New Keynesian hybrid Phillips curve, which defines inflation as a function of the lagged value of inflation (`L.inflation`), the output gap (`ogap`), and the expected value of inflation in $t+1$ $\{E_t(\texttt{F.inflation})\}$, conditional on information available at time t (Gali and Gertler 1999). See **[U] 11.4.4 Time-series varlists**.

Expected future inflation cannot be directly observed, so macroeconomists use instruments to predict the one-step-ahead inflation rate. This prediction is obtained by regressing the one-step ahead inflation rate on a set of instruments.

We can write this mathematically as

$$\texttt{inflation} = \alpha + \texttt{L.inflation}\ \beta + \texttt{ogap}\ \delta + E_t(\texttt{F.inflation})\ \gamma + \epsilon_t$$

and

$$\texttt{F.inflation} = \mathbf{z}_t \boldsymbol{\theta} + \nu_{t+1}$$

where $\mathbf{z}_t$ is a vector of instruments. The forecasted values given by $E_t(\texttt{F.inflation}|\mathbf{z_t}) = \mathbf{z}_t\widehat{\boldsymbol{\theta}}$ are uncorrelated with ν_{t+1} by construction.

In this example, we fit the Phillips curve model for the period 1970q1 to 1997q4. We are interested in testing whether expectation of future inflation changed during this period. We instrument the future value of inflation with the first two lags of inflation, the federal funds rate, and the output gap. We use `ivregress 2sls` to fit the model.

```
. ivregress 2sls inflation L.inflation ogap
> (F.inflation = L(1/2).inflation L(1/2).ogap L(1/2).fedfunds)
> if tin(1970q1,1997q4)
  (output omitted )
. estat sbsingle, breakvars(F.inflation)
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
..................................................    50
............................
Test for a structural break: Unknown break date

                             Number of obs =       112

Full sample:                 1970q1 - 1997q4
Trimmed sample:              1974q2 - 1993q4
Estimated break date:        1981q3
Ho: No structural break

     Test            Statistic           p-value
────────────────────────────────────────────────
     swald             6.7345             0.1164
────────────────────────────────────────────────
Coefficients included in test: F.inflation
```

We fail to reject the null hypothesis of no structural break in the coefficient of expected future inflation.

◁

Stored results

`estat sbsingle` stores the following in `r()`:

Scalars

`r(chi2)`	χ^2 test statistic
`r(p)`	p-value for χ^2 test
`r(df)`	degrees of freedom

Macros

`r(ltrim)`	start of trim date
`r(rtrim)`	end of trim date
`r(breakvars)`	list of variables whose coefficients are included in the test
`r(breakdate)`	estimated break date only after supremum tests
`r(test)`	type of test

Methods and formulas

Each supremum test statistic is the maximum value of the test statistic that is obtained from a series of Wald or LR tests over a range of possible break dates in the sample. Let b denote a possible break date in the range $[b_1, b_2]$ for a sample size T.

The supremum test statistic for testing the null hypothesis of no structural change in k coefficients is given by

$$\text{supremum } S_T = \sup_{b_1 \leq b \leq b_2} S_T(b)$$

where $S_T(b)$ is the Wald or LR test statistic evaluated at a potential break date b. The average and the exponential versions of the test statistic are

$$\text{average } S_T = \frac{1}{b_2 - b_1 + 1} \sum_{b=b_1}^{b_2} S_T(b)$$

$$\text{exponential } S_T = \ln\left[\frac{1}{b_2 - b_1 + 1} \sum_{b=b_1}^{b_2} \exp\left\{\frac{1}{2} S_T(b)\right\}\right]$$

respectively.

The limiting distributions of the test statistics are given by

$$\text{supremum } S_T \rightarrow_d \sup_{\lambda \in [\varepsilon_1, \varepsilon_2]} S(\lambda)$$

$$\text{average } S_T \rightarrow_d \frac{1}{\varepsilon_2 - \varepsilon_1} \int_{\varepsilon_1}^{\varepsilon_2} S(\lambda)\ d\lambda$$

$$\text{exponential } S_T \rightarrow_d \ln\left[\frac{1}{\varepsilon_2 - \varepsilon_1} \int_{\varepsilon_1}^{\varepsilon_2} \exp\left\{\frac{1}{2} S(\lambda)\ d\lambda\right\}\right]$$

where

$$S(\lambda) = \frac{\{B_k(\lambda) - \lambda B_k(1)\}'\{B_k(\lambda) - \lambda B_k(1)\}}{\lambda(1-\lambda)}$$

$B_k(\lambda)$ is a vector of k-dimensional independent Brownian motions, $\varepsilon_1 = b_1/T$, $\varepsilon_2 = b_2/T$, and $\lambda = \varepsilon_2(1-\varepsilon_1)/\{\varepsilon_1(1-\varepsilon_2)\}$.

Computing the p-values for the nonstandard limiting distributions is computationally complicated. For each test, the reported p-value is computed using the method in Hansen (1997).

References

Andrews, D. W. K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61: 821–856.

Andrews, D. W. K., and W. Ploberger. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62: 1383–1414.

Davies, R. B. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74: 33–43.

Gali, J., and M. Gertler. 1999. Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics* 44: 195–222.

Hansen, B. E. 1997. Approximate asymptotic p values for structural-change tests. *Journal of Business and Economic Statistics* 15: 60–67.

Kim, H.-J., and D. Siegmund. 1989. The likelihood ratio test for a change-point in simple linear regression. *Biometrika* 76: 409–423.

Perron, P. 2006. Dealing with structural breaks. In *Palgrave Handbook of Econometrics: Econometric Theory, Vol I*, ed. T. C. Mills and K. Patterson, 278–352. Basingstoke, UK: Palgrave.

Quandt, R. E. 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* 55: 324–330.

Also see

[TS] **estat sbcusum** — Cumulative sum test for parameter stability

[TS] **estat sbknown** — Test for a structural break with a known break date

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

[R] **ivregress** — Single-equation instrumental-variables regression

[R] **regress** — Linear regression