Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

xtbreak: Testing for structural breaks in Stata

2020 Swiss (online) Stata User Group Meeting

Jan Ditzen[1], Yiannis Karavias[2], Joakim Westerlund[3]

[1]Free University of Bozen-Bolzano, Bozen, Italy
www.jan.ditzen.net, jan.ditzen@unibz.it

[2]University of Birmingham, UK
https://sites.google.com/site/yianniskaravias/
i.karavias@bham.ac.uk

[3]Lund University, Lund, Sweden

November 19, 2020

Motivation

- In time series or panel time series structural breaks (or change points) in the relationships between key variables can occur.
- Estimations and forecasts depend on knowledge about structural breaks.
- Structural breaks might influence interpretations and policy recommendations.
- Break can be unknown or known and single and multiple breaks can occur.
- Examples: Financial Crisis, oil price shock, Brexit Referendum, COVID19,...
- Question: Can we estimate when the breaks occur and test them?

Literature

- Time Series:
  - Andrews (1993) test for parameter instability and structure change with unknown change point.
  - Bai and Perron (1998) propose three tests for and estimation of multiple change points.
- Panel (Time) Series:
  - Wachter and Tzavalis (2012) single structural break in dynamic independent panels.
  - Antoch et al. (2019); Hidalgo and Schafgans (2017) single structural break in dependent panel data.
- `xtbreak` introduces tests for multiple structural breaks in time series based on Bai and Perron (1998).

Econometric Model I

- Multiple linear regression model with s breaks:

$$
\begin{aligned}
y_t &= x_t'\beta + z_t'\delta_1 + u_t, && t = 1, ..., T_1 \\
y_t &= x_t'\beta + z_t'\delta_2 + u_t, && t = T_1 + 1, ..., T_2 \\
... \\
y_t &= x_t'\beta + z_t'\delta_{s+1} + u_t, && t = T_s, ..., T
\end{aligned}
$$

- $\tau = (T_1, T_2, ..., T_s)$ are break points of the s breaks.
- x_t is a $(1 \times p)$ vector of variables without structural breaks.
- z_t is a $(1 \times q)$ vector of variables with structural breaks.

Econometric Model II

- The model can be expressed in matrix form:

$$Y = X\beta + \bar{Z}\delta + U \tag{1}$$

- where $Y = (y_1, .., y_T)'$, $X = (x_1, ..., x_T)'$, $\delta = (\delta_1', ..., \delta_{s+1}')'$ and:

$$\bar{Z} = \begin{pmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & z_{s+1} \end{pmatrix}$$

- z_s is $(T_s \times q)$.
- Aim: Test if and when breaks occur.

Hypotheses

- Three hypotheses (Bai and Perron, 1998):
  1. No break vs. s breaks
     $H_0 : \delta_1 = \delta_2 = ... = \delta_{s+1}$ vs $H_1 : \delta_k \neq \delta_j$ for some $j \neq k$.
  2. No break vs $1 \leq s \leq s^*$ breaks
     $H_0 : \delta_1 = \delta_2 = ... = \delta_{s+1}$ vs $H_1 : \delta_k \neq \delta_j$ for some $j \neq k$ and $s = 1, ..., s^*$
  3. s breaks vs $s + 1$ breaks
     $H_0 : \delta_j = \delta_{j+1}$ for one $j = 1, .., s$ vs. $H_1 : \delta_j \neq \delta_{j+1}$ for all $j = 1, ..., s$.
- Next question: know or unknown breakpoints?

Tests

- Main idea: if the model has the true number of breaks, then the SSR should be smaller than for a model with a larger or smaller number of breaks.
- No knowledge of the break points required.

Test Hypothesis 1 I

No break vs. s breaks

$$H_0 : \delta_1 = \delta_2 = ... = \delta_{s+1} \text{ vs } H_1 : \delta_k \neq \delta_j \text{ for some } j \neq k$$

- Wald test with test statistic:

$$F_T(\tau, q) = \frac{T - (s+1)q - p}{sq} \hat{\delta}' R' \left(R \hat{V}(\hat{\delta}) R' \right)^{-1} R\hat{\delta} \tag{2}$$

- R imposes the restrictions such that $R\delta' = (\delta_1' - \delta_2', ..., \delta_s' - \delta_{s+1})'$.
- $\hat{V}(\hat{\delta})$ is an estimate of the variance. For iid errors it is: $\hat{V}(\hat{\delta}) = SSR(\hat{\delta}) \left(\bar{Z}' M_X \bar{Z} \right)^{-1}$.
- For serially correlated errors: $\left(\bar{Z}' M_x \bar{Z} \right)^{-1} \bar{Z}' M_x \Sigma M_x \bar{Z} \left(\bar{Z}' M_x \bar{Z} \right)^{-1}$
- $M_X = I_T - X'(X'X)^{-1}X$ is an annihilator matrix to remove the constant variables in X.

Test Hypothesis 1 II

No break vs. s breaks

- If the break dates are known, then (Andrews, 1993)

$$F_T(\tau) \sim \chi^2(sq).$$

- If the break dates are unknown, then *supF* test statistic is used:

$$\sup F_T(s,q) = \sup_{\tau \in \tau_\eta} F_T(\tau,q)$$

- τ_ϵ is a subset of $[0, T]^s$ and represent all possible combination of break points with a minimal length of each set of η.
- Asymptotic critical values depending on the number of breaks s and regressors q are given in Bai and Perron (1998, Table 1).

Test Hypothesis 2 I

No break vs. $1 \leq s \leq s^*$ breaks

- Test if a maximum of s^* breaks occurs.
- "Double Maximum" test, where the maximum of the test using hypothesis 1 for the number of breaks between 1 and s^* is taken.

$$\text{WDmax}F_T(s, q) = \max_{1 \leq s \leq s^*} \left\{ \frac{c_{\alpha,1,q}}{c_{\alpha,s,q}} \sup_{\tau \in \tau_\eta} F_T(\tau, q) \right\}$$

- $c_{\alpha,s,q}$ is the critical value at a level of α for s breaks and q regressors.
- Asymptotic critical values depending on the number of breaks s and regressors q are given in Bai and Perron (1998, Table 1).

Test Hypothesis 3 I

s breaks vs. $s+1$ breaks

- Idea: test each s segments for an additional break within the segment.

$$F(s+1|s) = \frac{SSR(\hat{T}_1, ..., \hat{T}_s) - \min_{1 \leq j \leq s+1} \left\{ \inf_{\tau \in \Lambda_{j,\eta}} SSR(\hat{T}_1, ..., \hat{T}_{j-1}, \tau, \hat{T}_j, ..., \hat{T}_s) \right\}}{\hat{\sigma}_s^2}$$

$$\Lambda_{j,\eta} = \left\{ \tau; \hat{T}_{j-1} + \left(\hat{T}_j - \hat{T}_{j-1} \right) \eta \leq \tau \leq \hat{T}_j - \left(\hat{T}_j - \hat{T}_{j-1} \right) \eta \right\}$$

$$\hat{\sigma}_s^2 = \frac{SSR(\hat{T}_1, ..., \hat{T}_s)}{N(T-1) - sq - p}$$

$$SSR(\hat{T}_1, ..., \hat{T}_{s+1}) = \min_{\tau \in \tau_\eta} SSR(\tau)$$

Test Hypothesis 3 II

s breaks vs. $s+1$ breaks

- Looks complicated.... but it is essentially the difference of the minimum of combinations of the SSR with s and $s+1$ breaks.
- Asymptotic critical values depending on the number of breaks s and regressors q are given in Bai and Perron (1998, Table 2).

xtbreak[1]

`xtbreak test` *depvar* [*indepvars*] [*if*] [, hypothesis(1|2|3) *break_point_options* nobreakvariables(varlist ts) noconstant breakconstant vce(ssr|hac|nw)]

If the breakpoint is known then `break_point_options` are:

`breakpoints(numlist [ ,index ])`

If the breakpoint is unknown then `break_point_options` are:

`breaks(real) minlength(real) level(real)`

- `breaks(real)` sets the number of breaks.
- `breakpoints(numlist)` sets the breakpoints.
- `vce` is the variance/covariance estimator.

[1] This command is work in progress. Options, functions and results might change.

Excess deaths in the UK I

- Question: can we identify structural breaks in the excess deaths in the UK in 2020 due to COVID19?
- Data from Office of National Statistics (ONS) for weekly deaths in the UK for 2020.
- $d_{y,w}$ are the deaths in year y and week w.
- Excess death is defined as: $ed_{y,w} = d_{y,w} - \frac{1}{5}\sum_{j=1}^{5} d_{y-j,w}$, i.e. the difference between the actual deaths and the average of the past 5 years.
- Assume the excess deaths vary around a long run mean (β_0):

$$ed_{y,w} = \beta_0 + \epsilon_{y,w}, \epsilon_{y,w} \sim IID(0, \sigma^2)$$

- To find out if excess deaths varied due to COVID, we need to test if there are breaks in the long run mean β_0.

Excess deaths in the UK II

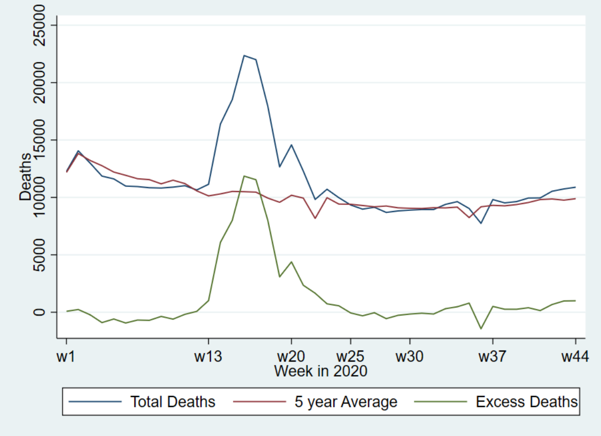

Figure: Excess Deaths in the UK. Data from ONS.

Excess deaths in the UK III

- Until week 13 excess deaths were normally moving around 0.
- From around week 19 excess deaths slowly declined and returned from around week 25 to the long run mean.
- First wave is clearly visible.
- Question: can we test how many breaks happened and when?

Unknown Breakdates

Test for no vs up to 4 breaks

- We can test if the number of breaks is up to or smaller than a given number.
- Assumptions that we have at most 4 breaks. That is we test: H_0 : no breaks vs $H_1 : 1 \leq s \leq 4$ breaks.
- There are 33 different break combinations for 1 break, 378 for 2 breaks, 1771 for 3 and 3060 for 4 break points.
- `xtbreak` loops through all of them and selects the one with the largest value of $W(\tau)$.
- `xtbreak` displays the 1%, 5% and 10% critical values from Bai and Perron (1998)
- We reject the hypothesis of no breaks against the alternative that there are at most 4 breaks.
- We also find that there are two breaks at period 13 and 20.

```
. xtbreak test  ExcessDeaths , breakconstant breaks(1 4) hypothesis(2)

Testing combinations for 1 break(s) (33)
----+--- 10 ---+--- 20 ---+--- 30 ---+--- 40 ---+--- 50   %
..................................................    50
..................................................   100

Testing combinations for 2 break(s) (378)
----+--- 10 ---+--- 20 ---+--- 30 ---+--- 40 ---+--- 50   %
..................................................    50
..................................................   100

Testing combinations for 3 break(s) (1771)
----+--- 10 ---+--- 20 ---+--- 30 ---+--- 40 ---+--- 50   %
..................................................    50
..................................................   100

Testing combinations for 4 break(s) (3060)
----+--- 10 ---+--- 20 ---+--- 30 ---+--- 40 ---+--- 50   %
..................................................    50
..................................................   100

Test for multiple breaks at unknown breakdates
(Bai & Perron. 1998. Econometrica)
H0: no break(s) vs. H1: 1 <= s <= 4 break(s)
```

		Bai & Perron Critical Values		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
max supW(tau)*	88.85	15.02	10.91	9.14

```
Estimated break points: 13 20
* evaluated at a level of 0.95.
```

Known Breakdates

Test for no vs 3 breaks

- We can test if there is a break in weeks 13 and 20 against the hypothesis of no break.
- That would be 3 breaks at known break dates:

```
. xtbreak test ExcessDeaths , breakconstant hypothesis(1) ///
>                                 breakpoints(13 20, index)
Test for multiple breaks at known breakdates
(Bai & Perron. 1998. Econometrica)
H0: no breaks vs. H1: 2 break(s)
 W(tau)    =     81.01
 p-value   =      0.00
```

- The p-value of the $\chi(2)^2$ distribution is almost 0, thus we can reject the hypothesis of no breaks.

Known Breakdates

Test for no vs 2 breaks

- We can use a HAC consistent estimator rather than the SSR.
- We use $\Sigma = \hat{\sigma}^2 I$ and $\hat{V}(\hat{\delta}) = \left(\bar{Z}' M_x \bar{Z}\right)^{-1} \bar{Z}' M_x \Sigma M_x \bar{Z} \left(\bar{Z}' M_x \bar{Z}\right)^{-1}$

```
. xtbreak test  ExcessDeaths , breakconstant hypothesis(1) ///
>                                breakpoints(13 20, index) vce(hac)

Test for multiple breaks at known breakdates
(Bai & Perron. 1998. Econometrica)
H0: no breaks vs. H1: 2 break(s)

 W(tau)    =      9.36
 p-value   =      0.01
```

- Hypothesis of no breaks against the alternative of 2 breaks can be rejected.

Unknown Breakdates

Test for no vs 2 breaks

- Test for 2 breaks at unknown dates.

```
. xtbreak test  ExcessDeaths , breakconstant breaks(2) hypothesis(1)
Testing combinations for 2 break(s) (378)
────┼──── 10 ──┼─── 20 ──┼─── 30 ──┼─── 40 ──┼─── 50    %
..................................................   50
..................................................  100

Test for multiple breaks at unknown breakdates
(Bai & Perron. 1998. Econometrica)
H0: no break(s) vs. H1: 2 break(s)
```

	Test Statistic	Bai & Perron Critical Values: 1% Critical Value	5% Critical Value	10% Critical Value
supW(tau)	81.01	10.95	8.78	7.87

```
Estimated break points: 13 20
```

- Output is similar to the one for testing up to 4 breaks.
- We can reject the hypothesis that there are no breaks against the alternative of 2 breaks.
- Estimated break points are as expected.

Unknown Breakdates

Test for no vs 2 breaks

- We can use the HAC consistent estimator instead.

```
. xtbreak test  ExcessDeaths , breakconstant breaks(2) hypothesis(1) vce(hac)
Testing combinations for 2 break(s) (378)
----+--- 10 --+--- 20 --+--- 30 --+--- 40 --+--- 50    %
..................................................    50
..................................................   100

Test for multiple breaks at unknown breakdates
(Bai & Perron. 1998. Econometrica)
H0: no break(s) vs. H1: 2 break(s)
```

		Bai & Perron Critical Values		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
supW(tau)	10.05	10.95	8.78	7.87

```
Estimated break points: 13 19
```

- We can still reject the hypothesis, but at a lower level.
- Note: Estimated break points changed from 20 to 19!

Unknown Breakdates

Test for 2 vs 3 breaks

```
. xtbreak test  ExcessDeaths , breakconstant breaks(2) hypothesis(3)
Testing combinations for 2 break(s) (378)
────┼──── 10 ──┼─── 20 ──┼─── 30 ──┼─── 40 ──┼─── 50    %
..................................................    50
..................................................   100

Testing combinations for 3 break(s) (1771)
────┼──── 10 ──┼─── 20 ──┼─── 30 ──┼─── 40 ──┼─── 50    %
..................................................    50
..................................................   100

Test for multiple breaks at unknown breakpoints
(Bai & Perron. 1998. Econometrica)
H0: 2 vs. H1: 3 break(s)
```

		Bai & Perron Critical Values		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
F(s+1|s)*	2.74	15.62	12.16	10.45

```
* s = 2
```

- We cannot reject the hypothesis of 2 breaks.

Unknown Breakdates

Test for 1 vs 2 breaks

- Finally, let's test for 1 vs. 2 breaks.

```
. xtbreak test  ExcessDeaths , breakconstant breaks(1) hypothesis(3)
Testing combinations for 1 break(s) (33)
----+--- 10 --+--- 20 --+--- 30 --+--- 40 --+--- 50   %
..................................................    50
..................................................   100

Testing combinations for 2 break(s) (378)
----+--- 10 --+--- 20 --+--- 30 --+--- 40 --+--- 50   %
..................................................    50
..................................................   100

Test for multiple breaks at unknown breakpoints
(Bai & Perron. 1998. Econometrica)
H0: 1 vs. H1: 2 break(s)
```

		Bai & Perron Critical Values		
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
F(s+1|s)*	31.45	15.03	11.14	9.56

* s = 1

- We can reject the hypothesis of 1 breaks, implying the we found 2 breaks.
- For estimation of break dates we would need confidence intervals though....

Conclusion

- Introduced new community contributed package called `xtbreak`
- Test for breaks at known and unknown points in time.
- Three tests for time series included, following Bai and Perron (1998).
- What's next:
  - Extensions for panel data models.
  - Confidence intervals for estimated break dates.
  - Improve speed.
  - Monte Carlo Simulations.

References I

Andrews, D. W. K. 1993. Tests for Parameter Instability and Structural Change With Unknown Change Point. Econometrica 61(4): 821–856.

Antoch, J., J. Hanousek, L. Horvath, M. Huskova, and S. Wang. 2019. Structural breaks in panel data: Large number of panels and short length time series. Econometric Reviews 38(7).

Bai, B. Y. J., and P. Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica, 66(1): 47–78.

Hidalgo, J., and M. Schafgans. 2017. Inference and testing breaks in large dynamic panels with strong cross sectional dependence. Journal of Econometrics 96(2).

Wachter, S. D., and E. Tzavalis. 2012. Detection of structural breaks in linear dynamic panel data models. Computational Statistics & Data Analysis .