Gauss code for
"Estimating and Testing for Structural Changes in Multivariate Regressions"
by Zhongjun Qu and Pierre Perron
Econometrica (2007)
This version: November 6, 2007

The package contains three files:

1. mbreak2.prg: This is the main file for estimating and testing models with common breaks, partial structural break models and block partial breaks models.

2. obreak2.prg: This is the main file for estimating and testing a model with locally ordered breaks.

3. mbreak.src: This is the main code. You should not have to modify it.

To use the code, you need to use one of the two .prg files to set up your data and options.

1 Detailed information on using mbreak2.prg

The purpose of mbreak2.prg is to:

1. test for structural changes in the regression coefficients and/or the variance-covariance matrix of the errors;

2. determine the number of changes;

3. estimate the dates of changes and construct confidence intervals.

The output of the program contains two parts:

1. First, the code constructs tests and estimates for an unrestricted model under one of the three specifications:

   (a) pure structural change in regression coefficients, where all the regression coefficients but not the variance-covariance matrix of the errors are allowed to change;

(b) pure structural change in the variance-covariance matrix of the errors, where all the elements of the covariance matrix but not the regression coefficients are allowed to change;

(c) complete pure structural change, where all the parameters are allowed to change.

For the specification considered above, the code provides:

(a) The $\sup LR_T$ (or $\sup F_T$) tests against a fixed number of changes;

(b) The $WDmax$ test for an unknown number of changes up to some pre-specified maximum;

(c) The $Seq(\ell+1|\ell)$ test that allows to select the number of changes;

(d) The estimates of the break dates and confidence intervals when m breaks are allowed;

(e) The estimates of the coefficients.

2. Secondly, the code constructs tests and estimates for a restricted model with restrictions specified by the user. The outputs are:

(a) The $\sup LR_T$ (or $\sup F_T$) tests against a fixed number of changes;

(b) The estimates of the break dates and confidence intervals when m breaks are allowed;

(c) The estimated of the coefficients.

- Note that the current version of the code only allows restrictions of the following form:

$$\beta = R\delta$$

where β is the vector of regression coefficients in the unrestricted model, R is a matrix of constants (refer to the code for an example), δ is the set of basic parameters of the model (including parameters from all equations and all regimes). Hence, in particular, a subset of coefficients, as well as a subset of equations can be restricted not to change. The code can be changed accordingly to allow other forms of restrictions.

To use the code, you need to:

- Set up you data in four parts: a) the dependent variables y as a matrix of dimension T by n, where T is the sample size and n is the number of equations; b) the regressors z, as a matrix of dimension T by q, including regressors from all equations (without duplication); c) the matrix $\_S$, which imposes restrictions within regimes or selects regressors appearing in each regression; d) the restrictions specified by matrix R.
- Specify m, the maximal number of breaks allowed.
- Specify trm, which is the minimal length of a segment as a fraction of the sample.
- Specify other global control variables related to the distributions of the regressors and the errors, see the code.

2 The detailed information on using obreak2.prg

The purpose of obreak2.prg is to:

1. test for structural changes in the regression coefficients in a two equations system when the breaks are local and the ordering is known a priori; the specification used is such that the break in the first equation occurs no later than that in the second.
2. estimate the dates of changes and construct confidence intervals;
3. estimate the other coefficients of the model.

The output contains related information for the above procedures.

To use the code, you need to:

- Set up you data in three parts: a) the dependent variables as a matrix of dimension T by 2, where T is the sample size and 2 is the number of equations; b) the regressors, including regressors from all equations (without duplication); c) the matrix $\_S$, which specifies the restrictions within regimes or selects regressors appearing in each regression.
- Specify trm, which is the minimal length of a segment as a fraction of the sample.
- Specify other global control variables related to the distributions of the regressors, see the code.

- Set getcon=1 if performing simulation to obtain the confidence interval for the breaks, which is computationally intensive.

Note that the current version of the code does not allow for additional restrictions on the coefficients.

3 Two examples

We use two examples to illustrate how to estimate a linear multivariate regressions model with multiple breaks using the Gauss code provided. It involves two steps. In the first step, we need to re-write the model in the form

$$y_t = (I_n \otimes z_t')S\beta_j + u_t, \tag{1}$$

and in the second step, we need to specify the restrictions imposed on the regression coefficients. As a matter of notation, let m denote the number of breaks, n the number of equations, and T the sample size. Notice that β_j is a column vector that includes all regression coefficients from regime j, i.e., $\beta_j = (\beta_{1,j}, ..., \beta_{k,j})'$; define β as a column vector obtained by stacking coefficients from different regimes, i.e., $\beta = (\beta_1', ..., \beta_{m+1}')'$.

Example 1: A two equations SUR model. Consider the model

$$\begin{aligned} y_{1,t} &= x_{1,t}\beta_{1,j} + x_{2,t}\beta_{2,j} + u_{1,t} \\ y_{2,t} &= x_{3,t}\beta_{3,j} + x_{4,t}\beta_{4,j} + u_{2,t}, \end{aligned}$$

where all variables involved are scalar random variables. The model can be re-written as

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \left(I_2 \otimes \begin{bmatrix} x_{1,t} & x_{2,t} & x_{3,t} & x_{4,t} \end{bmatrix}\right) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \beta_{1,j} \\ \beta_{2,j} \\ \beta_{3,j} \\ \beta_{4,j} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}.$$

We have, corresponding to (1),

$$z_t' = \begin{bmatrix} x_{1,t} & x_{2,t} & x_{3,t} & x_{4,t} \end{bmatrix}, \beta_j = \begin{pmatrix} \beta_{1,j} \\ \beta_{2,j} \\ \beta_{3,j} \\ \beta_{4,j} \end{pmatrix}, \ S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The relevant gauss commands needed are (suppose $data[.,.]$ is a T by 6 matrix, with the first two columns containing y$_t'$ $(t = 1, ..., T)$ and the last four columns containing z_t' $(t = 1, ..., T)$):

$$\begin{aligned} y &= data[.,1:2]; \\ z &= data[.,3:6]; \\ \_S &= \{1\ 0\ 0\ 0,\ 0\ 1\ 0\ 0,\ 0\ 0\ 0\ 0,\ 0\ 0\ 0\ 0,\ 0\ 0\ 0\ 0,\ 0\ 0\ 0\ 0,\ 0\ 0\ 1\ 0,\ 0\ 0\ 0\ 1\}; \end{aligned}$$

If there is no restriction, i.e., all regression coefficients are allowed to change, then R is given by:

$$R = eye(cols(\_S) * (m+1)).$$

Otherwise, R is chosen to solve:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_{m+1} \end{pmatrix} = R\delta,$$

where δ is the set of basic parameters of the model (including parameters from all equations and all regimes without repetition). Consider the case with $m = 2$ and assume that only the coefficients on $x_{1,t}$ and $x_{3,t}$ can change, then

$$\delta = \begin{pmatrix} \beta_{1,1} & \beta_{2,1} & \beta_{3,1} & \beta_{4,1} & \beta_{1,2} & \beta_{3,2} & \beta_{1,3} & \beta_{3,3} \end{pmatrix}',$$

and

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Example 2: VAR models. For VARs things are extremely simple, because z_t contains simply the lagged dependent variables and the deterministic terms, and S is an identity matrix. For illustration, consider a bivariate VAR with an intercept and one lag:

$$\begin{aligned} y_{1,t} &= \beta_{1,j} + y_{1,t-1}\beta_{2,j} + y_{2,t-1}\beta_{3,j} + u_{1,t} \\ y_{2,t} &= \beta_{4,j} + y_{1,t-1}\beta_{5,j} + y_{2,t-1}\beta_{6,j} + u_{2,t}. \end{aligned}$$

Then,

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \Big(I_2 \otimes \begin{pmatrix} 1 & y_{1,t-1} & y_{2,t-1} \end{pmatrix}\Big) \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \beta_{1,j} \\ \beta_{2,j} \\ \beta_{3,j} \\ \beta_{4,j} \\ \beta_{5,j} \\ \beta_{6,j} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

The relevant gauss commands for this specific example are (suppose $data[.,.]$ is a T by 2 matrix containing the time series)

$$\begin{aligned} y &= data[2:T,.]; \\ z &= ones(T-1,1)\tilde{}data[1:T-1,.]; \\ \_S &= eye(6); \end{aligned}$$

For R, if there is no restriction, then,

$$R = eye(cols(\_S) * (m+1)).$$

If there are restrictions, then again R solves

$$\beta = R\delta.$$

Consider the restriction that only the coefficients of the first equation change, and $m = 1$, then R solves

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = R \begin{pmatrix} \beta_{1,1} & \beta_{2,1} & \beta_{3,1} & \beta_{4,1} & \beta_{5,1} & \beta_{6,1} & \beta_{1,2} & \beta_{2,2} & \beta_{3,2} \end{pmatrix}'$$

which gives

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

This program is distributed freely for non-profit academic purposes only. For other uses, please contact Zhongjun Qu at qu@bu.edu. A lot of effort has been put to construct this program and we would appreciate that you acknowledge using this code in your research and cite the relevant paper on which it is based:

- Qu, Z. and P. Perron (2007): "Estimating and Testing for Structural Changes in Multivariate Regressions", Econometrica.

Although a lot of efforts has been put in constructing the program, we cannot be held responsible for any consequences that could result from remaining errors. Comments about errors, possible improvements and so on are most welcomed. Thank you for your interest and good luck with the program.

Zhongjun Qu
(qu@bu.edu)
Department of Economics
Boston University
Boston, MA, 02215