

QR.break: An R Package for Structural Breaks in Quantile Regression

Zhongjun Qu*

Tatsushi Oka[†]

Samuel Messer[‡]

Boston University

Keio University

Boston University

April 22, 2025

Abstract

The `QR.break` package provides methods for detecting, estimating, and conducting inference on multiple structural breaks in quantile regression, based on one or multiple quantiles and applicable to both time series and repeated cross-sectional data. The main function, `rq.break()`, returns testing and estimation results based on user specifications of the quantiles of interest, the maximum number of breaks allowed, and the minimum length of a single regime. This note outlines the underlying methods and explains how to use the main function with two datasets: a time series dataset on U.S. real GDP growth rates and a repeated cross-sectional dataset on youth drinking and driving behavior. Both datasets are included in the package available on CRAN.

Keywords: Structural break, quantile regression.

JEL codes: C21, C22, C31, C52, C63.

*Department of Economics, Boston University, qu@bu.edu.

[†]Department of Economics, Keio University, tatsushi.oka@keio.jp.

[‡]Department of Economics, Boston University, snmesser@bu.edu.

1 Introduction

The issue of structural breaks has been extensively studied. Much of the literature has focused on the conditional mean, but in many cases, structural change in the conditional quantile function is more informative. For example, in studying income inequality, it is important to assess whether the wage gap between racial groups, conditional on covariates, has changed over time. Inequality may increase conditional dispersion without affecting the mean. Similarly, when evaluating a policy aimed at improving outcomes for low-performing students, attention should focus on lower quantiles. In both cases, it is desirable to estimate break dates from the data: in the former, the cause of change may be unclear a priori; in the latter, the policy effect may exhibit an unknown time lag.

To address these issues, Qu (2008) and Su and Xiao (2008) developed tests for detecting structural change in conditional quantile functions with unknown break dates, but did not consider estimation and inference for the number of breaks or their locations. Oka and Qu (2011) studied the estimation of multiple structural breaks at unknown dates in conditional quantile functions for two models: a time series model, useful for analyzing macroeconomic data, and a repeated cross-section model, relevant for evaluating social programs and policy effects. Oka and Qu's (2011) framework allows for structural change in single or multiple quantiles. Analyzing multiple quantiles requires stronger assumptions but can improve estimation efficiency.

Key aspects of Oka and Qu's (2011) procedure are as follows: assuming a known number of breaks, the methods construct estimates of break dates and coefficients as global minimizers of the check function over admissible break points. For multiple quantiles, the check function is integrated over the set of quantiles chosen by the user. The assumptions permit dynamic models and impose restrictions only in neighborhoods around quantiles of interest, leaving other quantiles unspecified. This flexibility allows researchers to examine slices of the conditional distribution without imposing global distributional assumptions. The distributions of the break estimators were derived following Picard (1985) and Yao (1987). These distributions involve consistently estimable parameters, which enables confidence interval construction without simulation. Their paper also proposes a test for the number of breaks based on subgradient methods from Qu (2008). These tests do not require variance estimation and have monotonic power even with multiple breaks. Taken together, Qu (2008) and Oka and Qu (2011) offer the most comprehensive treatment to date of multiple structural breaks in quantile regression.

To date, there has been no published software package in any language to test for, estimate, and conduct inference on multiple structural breaks in linear quantile regression models. This led us to develop an R package that performs these tasks in an easy-to-use fashion. The resulting package, `QR.break`, is available on CRAN. Its main function, `rq.break()`, returns all testing and estimation results based on user-specified quantiles of interest, the maximum number of allowed breaks, and the minimum length of a single regime. To use it, the user simply runs:

```
rq.break(y, x, vec.tau, N, trim.e, vec.time, m.max, v.a, v.b, verbose)
```

with the following arguments:

- `y`: the dependent variable, given as a vector or data frame.
- `x`: the regressors, provided as a matrix or data frame. A column of ones should not be included, as it is automatically added during estimation.
- `vec.tau`: the quantiles of interest, e.g., `vec.tau <- seq(0.1, 0.9, by = 0.1)`.
- `N`: the size of the cross-section; set to 1 for time series data.
- `trim.e`: the minimum length of any regime, expressed as a fraction of the total time span (e.g., `trim.e = 0.1`).
- `vec.time`: the time indices of the data, used for reporting the estimated break dates.
- `m.max`: the maximum number of breaks allowed.
- `v.a, v.b`: significance levels over a pre-specified set for testing and estimation, respectively.
- `verbose`: set to 1 to display estimation results in the R console; the default is 0.

We illustrate the application of the function through two empirical examples. The first revisits the “Great Moderation” in U.S. GDP growth using quarterly data. Results suggest the decline in volatility mainly affected the upper tail, with the median and lower quantiles remaining stable. This implies that expansions became less rapid, while recessions remained as severe. The second application analyzes blood alcohol levels of young drivers in California from 1983 to 2007. Two breaks are detected, consistent with the 1984 National Minimum Drinking Age Act and a 1991 beer tax increase. The effects are more pronounced at lower quantiles, indicating a greater impact on lighter drinkers than on heavier ones. These findings are documented in Oka and Qu (2011). Here, the focus is on the detailed steps of application of the methods to obtain these results.

Below, we first outline the methods implemented in the package, followed by a detailed description of their application to the two empirical examples.

2 Methods for estimation and inference

This section explains: (1) the model and the econometric issues of interest; (2) the main steps for estimating the break locations and regression coefficients when the number of breaks is known, based on a single quantile; (3) the main steps for estimating the break locations and regression coefficients when the number of breaks is known, based on multiple quantiles; and (4) a procedure for determining the number of breaks.

2.1 Econometric models and issues of interest

The methods apply to both time series and repeated cross-sectional data. We begin with time series data.

2.1.1 Time series models

Let y_t be a random variable, x_t a p -dimensional vector of covariates, and $Q_{y_t}(\tau|x_t)$ the conditional τ -quantile of y_t given x_t , where t is the time index. Let T denote the sample size. Assume the conditional quantile function is linear and potentially affected by m structural breaks:

$$Q_{y_t}(\tau|x_t) = \begin{cases} x'_t\beta_1^0(\tau), & t = 1, \dots, T_1^0, \\ x'_t\beta_2^0(\tau), & t = T_1^0 + 1, \dots, T_2^0, \\ \vdots & \vdots \\ x'_t\beta_{m+1}^0(\tau), & t = T_m^0 + 1, \dots, T, \end{cases} \quad (1)$$

where $\tau \in (0, 1)$, $\beta_j^0(\tau)$ ($j = 1, \dots, m + 1$) are the unknown parameters, and T_j^0 ($j = 1, \dots, m$) are the unknown break dates. The regressors x_t may include discrete as well as continuous variables. A column of ones is automatically added to the regression when applying the methods. The following examples, taken from Oka and Qu (2011), illustrate the model.

Example 1 *Oka and Qu (2011) studied the following model for U.S. quarterly real GDP*

growth rates for the period 1947:Q2 to 2009:Q2:

$$Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p}) = \mu_j(\tau) + \sum_{i=1}^p \alpha_{i,j}(\tau)y_{t-i} \quad (t = T_{j-1}^0 + 1, \dots, T_j^0).$$

They detect a structural break in 1984 that affects only the upper quantiles of the distribution. The coefficient estimates suggest that growth was slower during expansions, while recessions remained just as severe when they occurred. This dataset is included in the package, and we will use it in Section 3 to illustrate the application of the models in a time-series setting.

Example 2 Cox et al. (1985) considered the following model for a short-term interest rate r_t : $dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{1/2}dW_t$, where W_t is a standard Brownian motion process. The model has the following discrete-time approximation: $r_{t+1} - r_t = \alpha + \beta r_t + (\sigma r_t^{1/2})u_{t+1}$, with $u_{t+1} \sim i.i.d.N(0, 1)$, implying that the τ th conditional quantile of r_{t+1} given r_t is given by $Q_{r_{t+1}}(\tau|r_t) = \alpha + (1 + \beta)r_t + (\sigma r_t^{1/2})F_u^{-1}(\tau)$, where $F_u^{-1}(\tau)$ is the τ th quantile of $N(0, 1)$. The methods in the package can be used to estimate structural changes in α, β and σ assuming that the quantile function is given by the discrete-time approximation.

Example 3 Chernozhukov and Umantsev (2001) studied Value-at-Risk using $Q_{y_t}(\tau|x_t) = x_t'\beta(\tau)$, where y_t is an asset return, and x_t is a vector of information variables that may include returns on other securities, lagged values of y_t , and proxies for volatility. An important question is whether this risk relationship undergoes structural changes over time. The methods in this package can be applied to address this question without specifying the change points a priori.

Example 4 Piehl et al. (2003) applied structural change methods to evaluate the effect of the Boston Gun Project on youth homicides, allowing the break date to be unknown to account for potential policy delays. Their focus was on structural change in the conditional mean; however, the methods in this package could be used to analyze structural changes along the distribution.

2.1.2 Models for repeated cross-sections

For repeated cross-sectional data, let y_{it} be a random variable, x_{it} a p -dimensional vector of covariates, and $Q_{y_{it}}(\tau|x_{it})$ the conditional quantile of y_{it} given x_{it} , where i and t are individual and time indices. Let N be the number of cross-sectional units, assumed constant

over time, and T the number of time periods. The conditional quantile is potentially affected by m breaks:

$$Q_{y_{it}}(\tau|x_{it}) = \begin{cases} x'_{it}\beta_1^0(\tau), & t = 1, \dots, T_1^0, \\ x'_{it}\beta_2^0(\tau), & t = T_1^0 + 1, \dots, T_2^0, \\ \vdots & \vdots \\ x'_{it}\beta_{m+1}^0(\tau), & t = T_m^0 + 1, \dots, T, \end{cases} \quad (2)$$

where $i = 1, \dots, N$; $\tau \in (0, 1)$; $\beta_j^0(\tau)$ ($j = 1, \dots, m + 1$) are the unknown parameters; T_j^0 ($j = 1, \dots, m$) are unknown break dates; and x_{it} can may both discrete and continuous variables. A column of ones is automatically added to the regression. The following application, taken from Oka and Qu (2011), illustrates the model.

Example 5 *Motor vehicle crashes are the leading cause of death among youth aged 15-20, a high proportion of which involve drunk driving. Blood alcohol concentration (BAC) is a key measure of alcohol impairment, and changes in BAC among young drivers provide useful information on how young drivers' drinking behavior has changed over time. Motivated by this, Oka and Qu (2011) studied structural change in BAC among young drivers involved in traffic accidents using the following model:*

$$Q_{y_{it}}(\tau|x_{it}) = \mu_j(\tau) + x'_{it}\gamma_j(\tau) \quad (t = T_{j-1}^0 + 1, \dots, T_j^0),$$

where y_{it} is the BAC level. The regressors are age, gender, and a dummy variable for the fourth quarter. They detect breaks in 1985 and 1992. The changes are negative and meaningful in magnitude. However, the change is smaller for higher quantiles, suggesting that the policies are more effective for “light drinkers” than for “heavy drinkers” in the sample. This is encouraging but falls short of expectations, as heavy drinkers are more likely to cause accidents, suggesting that additional policies are needed to deter heavy drinking. This dataset is included in the package, and later in Section 3, we will use it to illustrate the application of the models in a cross-sectional setting.

2.1.3 Econometric issues of interest

The methods in this package address the following issues:

1. Estimation based on a single quantile when the number of breaks is known. The

method estimates both the break locations and the regression coefficients. If the user specifies more than one quantile level, the analysis is performed independently for each quantile, allowing break locations to differ across quantiles. The program returns the estimated break locations, their confidence intervals, and the corresponding estimates and intervals for the regression coefficients.

2. Estimation based on multiple quantiles when the number of breaks is known. In this case, the break locations are assumed to be common across quantiles and are estimated using information from all specified quantiles. The program returns the estimated break locations, their confidence intervals, and the corresponding estimates and intervals for the regression coefficients.
3. Selection of the number of breaks. The user specifies the maximum number of breaks, and the program determines the number of breaks using a dynamic programming algorithm.

The package is structured so that a main function performs all of these tasks. This will be demonstrated through empirical applications. Next, we describe the methods, starting with estimation of the model when the number of breaks is known, based on a single quantile.

2.2 Estimating break locations based on a single quantile

Consider time series data first. Suppose that the τ th quantile is affected by m structural changes. Define the following function for a set of candidate break dates $T^b = (T_1, \dots, T_m)$:

$$S_T(\tau, \beta(\tau), T^b) = \sum_{j=0}^m \sum_{t=T_j+1}^{T_{j+1}} \rho_\tau(y_t - x_t' \beta_{j+1}(\tau)), \quad (3)$$

where $\rho_\tau(u)$ is the check function given by $\rho_\tau(u) = u(\tau - 1(u < 0))$, see Koenker (2005); $\beta(\tau) = (\beta_1(\tau)', \dots, \beta_{m+1}(\tau)')'$; $T_0 = 0$; and $T_{m+1} = T$. We estimate the break dates and coefficients $\beta(\tau)$ jointly by solving

$$(\hat{\beta}(\tau), \hat{T}^b) = \arg \min_{\beta(\tau), T^b \in \Lambda_\varepsilon} S_T(\tau, \beta(\tau), T^b), \quad (4)$$

where $\hat{\beta}(\tau) = (\hat{\beta}_1(\tau)', \dots, \hat{\beta}_{m+1}(\tau)')'$ and $\hat{T}^b = (\hat{T}_1, \dots, \hat{T}_m)$. In (4), Λ_ε denotes the set of possible partitions to ensure that each estimated regime is a positive fraction of the sample:

$$\Lambda_\varepsilon = \{(T_1, \dots, T_m) : T_j - T_{j-1} \geq \varepsilon T \ (j = 2, \dots, m), T_1 \geq \varepsilon T, T_m \leq (1 - \varepsilon)T\}, \quad (5)$$

where ε is a positive small number specifying the minimum length of a regime. The user will specify ε when using the package. For repeated cross-sections, the estimation procedure is similar, except that the objective function $S_T(\tau, \beta(\tau), T^b)$ is replaced by

$$S_{NT}(\tau, \beta(\tau), T^b) = \sum_{j=0}^m \sum_{t=T_j+1}^{T_{j+1}} \sum_{i=1}^N \rho_\tau(y_{it} - x'_{it}\beta_{j+1}(\tau)),$$

where an additional summation “ $\sum_{i=1}^N$ ” is included to incorporate the cross-sectional observations. For both settings, the computation is carried out using a dynamic programming algorithm, as in Bai and Perron (2003), such that the computation is of order $O(T^2)$ irrespective of the number of breaks allowed in the model.

2.3 Estimating break locations based on multiple quantiles

Suppose that the quantiles in $\mathcal{T}_\omega = [\omega_1, \omega_2]$ with $0 < \omega_1 < \omega_2 < 1$ are affected by structural changes. A natural approach is to consider a partition of this interval and examine a set of quantiles denoted by τ_h , $h = 1, \dots, q$. The estimation can then be carried out in a similar way as in the single quantile case. Specifically, define the following objective function for a set of candidate break dates $T^b = (T_1, \dots, T_m)$ and parameter values $\beta(\mathcal{T}_\omega) = (\beta(\tau_1)', \dots, \beta(\tau_q)')'$:

$$S_T(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b) = \sum_{h=1}^q \sum_{j=0}^m \sum_{t=T_j+1}^{T_{j+1}} \rho_{\tau_h}(y_t - x'_t \beta_{j+1}(\tau_h)),$$

and solve

$$(\hat{\beta}(\mathcal{T}_\omega), \hat{T}^b) = \arg \min_{\beta(\mathcal{T}_\omega), T^b \in \Lambda_\varepsilon} S_T(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b), \quad (6)$$

where Λ_ε has the same definition as in (5). Regarding the partition, we find that a coarse partition, such as quantiles spaced by 0.1, is sufficient to deliver informative results. For repeated cross-sectional data, again the estimation procedure is the same, except that the

objective function should be replaced by

$$S_{NT}(\tau, \beta(\tau), T^b) = \sum_{h=1}^q \sum_{j=0}^m \sum_{t=T_j+1}^{T_{j+1}} \sum_{i=1}^N \rho_\tau(y_{it} - x'_{it} \beta_{j+1}(\tau)),$$

where an additional summation “ $\sum_{i=1}^N$ ” is included to incorporate the cross-sectional observations. For both settings, the computation is carried out using a dynamic programming algorithm, as in Bai and Perron (2003), such that the computation is of order $O(T^2)$ irrespective of the number of breaks allowed in the model.

For all cases discussed above, confidence intervals can be computed based on the limiting distributions of the break point estimates. For example, in the repeated cross-section, multiple-quantile case, the limiting distribution is given by, for $j = 1, \dots, m$:

$$\left(\frac{\bar{\pi}_j^*}{\bar{\sigma}_j^*}\right)^2 v_T^2(\hat{T}_j - T_j^0) \rightarrow^d \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\bar{\sigma}_{j+1}^*/\bar{\sigma}_j^*)W(s) - (\bar{\pi}_{j+1}^*/\bar{\pi}_j^*)|s|/2 & s > 0 \end{cases},$$

where $W(s)$ is the standard two-sided Brownian motion, and $\bar{\pi}_j^*, \bar{\pi}_{j+1}^*, \bar{\sigma}_j^{*2}$, and $\bar{\sigma}_{j+1}^{*2}$ are constants that can be estimated consistently. This distribution has an analytical density function and no simulation is needed to obtain the critical values, therefore reducing the computational cost; see Bai (1995) and Oka and Qu (2011) for details.

2.4 Determining the number of breaks

The package use two test statistics SQ_τ (for a single quantile) and DQ (for multiple quantiles) proposed in Qu (2008) to determine the number of breaks for the single and multiple quantile cases, respectively. We first briefly explain these two tests and then describe a recommended testing procedure based on them.

2.4.1 Testing for a Single Structural Break

The SQ_τ test is designed to detect the presence of a structural break in a given quantile τ :

$$SQ_\tau = \sup_{\lambda \in [0,1]} \left\| (\tau(1-\tau))^{-1/2} \left[H_{\lambda,T}(\hat{\beta}(\tau)) - \lambda H_{1,T}(\hat{\beta}(\tau)) \right] \right\|_\infty,$$

where

$$H_{\lambda,T}(\hat{\beta}(\tau)) = \left(\sum_{t=1}^T x_t x_t' \right)^{-1/2} \sum_{t=1}^{[\lambda T]} x_t \psi_\tau(y_t - x_t' \hat{\beta}(\tau))$$

and $\psi_\tau(u) = \tau - 1(u < 0)$ if we have a single time series, and

$$H_{\lambda,T}(\hat{\beta}(\tau)) = \left(\sum_{t=1}^T \sum_{i=1}^N x_{it} x_{it}' \right)^{-1/2} \sum_{t=1}^{[\lambda T]} \sum_{i=1}^N x_{it} \psi_\tau(y_{it} - x_{it}' \hat{\beta}(\tau))$$

with repeated cross-sectional data, $\hat{\beta}(\tau)$ is the parameter estimate using the full sample assuming no structural change, and $\|\cdot\|_\infty$ is the sup norm to reveal the strongest evidence against the null hypothesis, i.e. for a generic vector $z = (z_1, \dots, z_k)$, $\|z\|_\infty = \max(z_1, \dots, z_k)$.

The DQ test is designed to detect structural changes in quantiles in an interval \mathcal{T}_ω :

$$DQ = \sup_{\tau \in \mathcal{T}_\omega} \sup_{\lambda \in [0,1]} \left\| H_{\lambda,T}(\hat{\beta}(\tau)) - \lambda H_{1,T}(\hat{\beta}(\tau)) \right\|_\infty.$$

The expressions in the sup norm as defined as before.

2.4.2 Testing l against $l + 1$ breaks

We also need the following tests for the purpose of testing l against $l + 1$ breaks, labelled as $SQ_\tau(l + 1|l)$ test and $DQ(l + 1|l)$ test. Suppose a model with l breaks has been estimated with the break estimates denoted by $\hat{T}_1, \dots, \hat{T}_l$. These values partition the sample into $(l + 1)$ segments, with the j th segment being $[\hat{T}_{j-1} + 1, \hat{T}_j]$. The strategy proceeds by testing each of the $(l + 1)$ segments for the presence of an additional break. We let $SQ_{\tau,j}$ and DQ_j denote the SQ_τ and DQ test applied to the j th segment, i.e.,

$$\begin{aligned} SQ_{\tau,j} &= \sup_{\lambda \in [0,1]} \left\| (\tau(1 - \tau))^{-1/2} \left[H_{\lambda, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) - \lambda H_{1, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) \right] \right\|_\infty, \\ DQ_j &= \sup_{\tau \in \mathcal{T}_\omega} \sup_{\lambda \in [0,1]} \left\| H_{\lambda, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) - \lambda H_{1, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) \right\|_\infty, \end{aligned}$$

where

$$H_{\lambda, T_{j-1}, T_j}(\hat{\beta}_j(\tau)) = \left(\sum_{t=T_{j-1}+1}^{T_j} x_t x_t' \right)^{-1/2} \sum_{t=T_{j-1}+1}^{[\lambda(T_j - T_{j-1})]} x_t \psi_\tau(y_t - x_t' \hat{\beta}_j(\tau))$$

if we have a single time series, and

$$H_{\lambda, T_{j-1}, T_j}(\hat{\beta}_j(\tau)) = \left(\sum_{t=T_{j-1}+1}^{T_j} \sum_{i=1}^N x_{it} x'_{it} \right)^{-1/2} \sum_{t=T_{j-1}+1}^{\lfloor \lambda(T_j - T_{j-1}) \rfloor} \sum_{i=1}^N x_{it} \psi_{\tau}(y_{it} - x'_{it} \hat{\beta}_j(\tau))$$

for repeated cross-sectional data, where $\hat{\beta}_j(\tau)$ is the parameter estimate using the j th regime. $SQ_{\tau}(l+1|l)$ and $DQ(l+1|l)$ represent the maximum of the $SQ_{\tau,j}$ and DQ_j over the $l+1$ segments:

$$\begin{aligned} SQ_{\tau}(l+1|l) &= \max_{1 \leq j \leq l+1} SQ_{\tau,j}, \\ DQ(l+1|l) &= \max_{1 \leq j \leq l+1} DQ_j. \end{aligned}$$

We reject the null hypothesis in favor of a model with $l+1$ breaks if the resulting value exceed the corresponding critical value.

2.4.3 Critical values

These tests are asymptotically nuisance parameter free, and tables for critical values are provided in Qu (2008). They do not require the estimation of any variance parameter, thus having monotonic power even when multiple breaks are present. Qu (2008) provided a simulation study. The results suggest that these two tests perform favorably compared to Wald-based tests (see Figure 1 in Qu, 2008).

2.4.4 The recommended procedure

We can use this procedure to determine the number of breaks (we consider the interval \mathcal{T}_{ω} and focus on the quantile grid $\tau_1, \dots, \tau_q \in \mathcal{T}_{\omega}$).

- **Step 1.** Apply the DQ test. If the test does not reject, conclude that there is no break and terminate the procedure. If it rejects, then estimate the model allowing one break. Save the estimated break date and proceed to Step 2.
- **Step 2.** Apply the $DQ(l+1|l)$ tests starting with $l = 1$. Increase the value of l if the test rejects the null hypothesis. In each stage, the model is re-estimated and the break dates are the global minimizers of the objective function allowing l breaks. Continue the process until the test fails to reject the null.

- **Step 3.** Let \hat{l} denote the first value for which the test fails to reject. Estimate the model allowing \hat{l} breaks. Save the estimated break dates and confidence intervals.
- **Step 4.** This step treats the q quantiles separately. Specifically, for every quantile τ_h ($h = 1, \dots, q$), apply the SQ_τ and $SQ_\tau(l+1|l)$ tests. Carry out the same operations as in Steps 1 to 3. Examine whether the estimated breaks are in agreement with those from Step 3.

2.4.5 Size control

Since this is a sequential procedure, it is important to consider its overall rejection error under the null hypothesis. Suppose there is no break, and a 5% significance level is used. Then, there is a 95% chance that the procedure will be terminated in Step 1, implying the probability of finding one or more breaks is 5% in large samples. If there are m breaks with $m > 0$, then, similarly, the probability of finding more than m breaks will be at most 5%. The probability of finding less than m breaks in finite samples will vary from case to case depending on the magnitude of the breaks. This means that the overall significance level of this procedure is bounded from above by 5% under the null hypothesis. It does not over-reject the null hypothesis.

3 Applications

The section illustrates how to use the main function using: a time series dataset on U.S. real GDP growth rates and a repeated cross-sectional dataset on youth drinking and driving behavior. Both datasets are included in the package.

3.1 Time series example

After loading the package using `library(QR.break)`, the US GDP data can be loaded using the command

```
> data(gdp)
```

Moreover, the dataset can be viewed by running:

```
> gdp
      yq      gdp      lag1      lag2
1947 Q4  6.06552 -0.32868 -0.61348
```

```

1948 Q1    6.35664    6.06552   -0.32868
...
2009 Q1   -6.58992   -5.48400   -2.70336
2009 Q2   -0.73980   -6.58992   -5.48400

```

The `yq` column contains the dates, the `gdp` column contains the dependent variable, and the remaining two columns contain the first and second lagged values of `gdp` to be used as regressors.

3.1.1 Using the function

Below is the main function to estimate this model:

```
> result = rq.break(y, x, vec.tau, N, trim.e, vec.time, m.max, v.a, v.b, verbose)
```

We now take a closer look at the inputs to this function. The following commands define the `y` and `x` variables in the quantile regression:

```

> y = gdp[, "gdp"]
> x = gdp[, c("lag1", "lag2")]

```

A column of ones are always added to the regressors when estimating the model. Therefore, in this case, the model has three parameters that are allowed to be affected by structural breaks: the intercept, the coefficient for the first lag, and that for the second lag.

The next command specifies the quantiles of interest:

```
> vec.tau = seq(0.20, 0.80, by = 0.15)
```

Using these inputs, the function will perform two sets of calculations. First, it conducts analysis using the quantiles in `vec.tau` independently, which means that the number of breaks and their locations can be different across quantiles. Then, it carries out the analysis using all quantiles simultaneously. In this case, the breaks are assumed to be common across quantiles, and information across quantiles is pooled together to estimate the break dates.

Since this is a time series regression, the number of cross-sectional units should be set as:

```
> N=1
```

The program requires the specification of the minimum length of a regime. This parameter is important because if the regime length is too small, then the model fit might be non-unique and the estimation might pick up spurious breaks. In this example, we can set:

```
> trim.e = 0.15
```

This implies a regime is at least 15% of the sample size. We recommend setting this value between 10% and 20% in applications.

An issue related to the minimum length of a regime is the maximum number of breaks allowed. In this example, we can set:

```
> m.max = 3
```

This means that we allow at most 3 breaks and therefore 4 regimes for the entire sample period.

The next two parameters are for the significance levels of the analysis. The input `v.a` controls the significance level used for determining the number of breaks; 1, 2 or 3 for 10%, 5% or 1%, respectively. The input `v.b` controls the coverage level for the confidence intervals of break dates; 1 or 2 for 90% and 95%, respectively. In this example, we can set:

```
> v.a = 2
```

```
> v.b = 2
```

There are two additional inputs that control the printing and displaying functions while running the code. If we set:

```
> vec.time = gdp["yq"]
```

then the program will use `vec.time` as the time index to display the break dates. Here a break date is the final date of the existing regime, not the starting date of a new regime. Alternatively, one can set:

```
> vec.time = NULL
```

so that the break dates are expressed in terms of integers, e.g., 50 if the break is estimated to be the 50-th observation of the sample. The input `verbose` controls whether to automatically display the results in the R console. The default is `FALSE`, while setting it to `TRUE` will display the results:

```
> verbose = TRUE
```

Either way, the results from estimation and testing are all saved when running the code, and they can be displayed using:

```
> print(result)
```

3.1.2 Interpreting the results

The results consist of two parts: one based on individual quantiles (`result$s.out`) and one based on all quantiles (`result$m.out`).

Output based on separate quantiles The entries in `result$s.out` are ordered according to `vec.tau`: first for $\tau = 0.2$, then for $\tau = 0.35$, and so on. For each quantile, the break testing results are shown first; if at least one significant break is detected, the break locations and parameter estimation results follow.

For example, for $\tau = 0.2$, the results are:

```
> $s.out
$s.out$test_0.2          #testing results at the chosen significance level
      1 Breaks 2 Breaks 3 Breaks
SQ test      1.423269 1.373012      0
Critical values 1.529859 1.637547      0

$s.out$nbreak_0.2        #number of breaks detected
[1] 0
```

For this quantile, the test for the null hypothesis of no break against a single break is equal to 1.423269, while the critical value is 1.529859. The value is insignificant at the chosen level, so no break is detected. The final column is equal to zero because the test is not computed when the previous tests are insignificant at the 10% level. No break estimation results are produced. The results for other quantiles also show insignificance until the quantile level 0.65. At that point, one significant break is detected, followed by its confidence intervals reported in two ways: first in terms of index values, then in terms of dates.

```
$s.out$test_0.65          #testing results at the chosen significance level
      1 Breaks 2 Breaks 3 Breaks
SQ test      1.817933 1.023126      0
Critical values 1.529859 1.637547      0

$s.out$nbreak_0.65        #number of breaks detected
[1] 1

$s.out$br_est_0.65        #point estimate and confidence interval for break date
      Estimate CI_Lower_Bound CI_Upper_Bound
Break 1      147           83           161

$s.out$br_est_time_0.65
      Estimate CI_Lower_Bound CI_Upper_Bound
Break 1 "1984 Q2" "1968 Q2"      "1987 Q4"
```

Therefore, one break is detected, with a point estimate of $t = 147$, corresponding to the second quarter of 1984, with a confidence interval of $[83, 161]$, or equivalently $[1968Q2:1987Q4]$. For this quantile, the output also includes the estimated coefficients for the two regimes:

```

$s.out$coef_0.65
$s.out$coef_0.65$Regime_1
              Value Std. Error  t value      Pr(>|t|)
Intercept  4.5169837 0.63003316  7.169438 3.625833e-11
x1          0.4199469 0.09643037  4.354924 2.512308e-05
x2         -0.1051057 0.09458095 -1.111277 2.683010e-01

```

```

$s.out$coef_0.65$Regime_2
              Value Std. Error  t value      Pr(>|t|)
Intercept  2.2855100 0.4866286  4.696620 8.711282e-06
x1          0.1724133 0.1218581  1.414869 1.603091e-01
x2          0.2386392 0.1320517  1.807164 7.383669e-02

```

In this case, the sum of the autoregressive coefficients changes little, but the intercept decreases significantly, indicating a notably lower 0.65 quantile after the break:

```

$s.out$`bsize_0.65_Regime_2_minus_Regime_1`
              Value Std. Error  t value      Pr(>|t|)
Intercept -2.2314737 0.7957371 -2.804285 0.005453863
x1         -0.2475336 0.1375265 -1.799898 0.073127343
x2          0.3437448 0.1418386  2.423492 0.016109478

```

Similarly, the method detects a break in the 0.80 quantile; we omit the details.

Output based on multiple quantiles The results are structured in a similar way as the single quantile case. Here the break testing and estimation results are based on all chosen quantiles in `vec.tau`:

```

> $m.out
$m.out$test_joint      #testing results at the chosen significance level
              1 Breaks  2 Breaks 3 Breaks
DQ test          1.0275870 0.5892746      0
Critical values 0.9098714 0.9584567      0

$m.out$nbreak_joint    #number of breaks detected
[1] 1

$m.out$br_est_joint    #point estimate and confidence interval for break date
              Estimate CI_Lower_Bound CI_Upper_Bound
Break 1          146          120          147

$m.out$br_est_joint_time
              Estimate CI_Lower_Bound CI_Upper_Bound
Break 1 "1984 Q1" "1977 Q3"          "1984 Q2"

```


A single break is detected, as in the analysis based on individual upper quantiles. The rest of the output contains the coefficient estimates and their confidence intervals as in the single quantile case; we omit the details.

In summary, the findings here shed light on the “Great Moderation” debate on U.S. GDP growth. Results suggest the decline in volatility mainly affected the upper tail, with the median and lower quantiles remaining stable. This implies that expansions became less rapid, while recessions remained as severe.

3.1.3 Computational time

The package uses a dynamic programming algorithm, as in Bai and Perron (2003), to determine the globally optimal break partitions. The resulting computational cost grows with the square of the sample size, regardless of the number of breaks allowed. In this example, the program is expected to finish within a few minutes on a typical desktop computer with a single processor.

The package has built-in critical values for common configurations: for the SQ test when the number of regressors is less than 100, and for the DQ test when the number of regressors is less than 20 and the quantile trimming is symmetric, as in the example. With more than 20 regressors, or when the quantile trimming is asymmetric—as in the cross-sectional example below—the critical values for the DQ test are computed via simulation, which can add a few minutes or more of computational time when running the function.

3.1.4 Potential error messages

The function is structured to display error messages if the input configurations are not set up properly. Here we give some examples. Suppose we set:

```
> trim.e = 0.2  
> m.max = 6
```

The product of these two values exceeds 1 because it is not possible to allow 6 breaks when each regime is at least 20% of the sample size. Running the code will produce:

```
Error: m.max*trim.e exceeds 1. This occurs because too many regimes are allowed  
or the minimum length of a regime is too large. Consider decreasing m.max,  
trim.e, or both.
```

As another example, if we set `trim.e` too small, we will get an error message with a suggestion to increase it:

```
> trim.e = 0.01
Error: trim.e * nrow(x) must be at least the number of regressors; otherwise,
the estimation results are not unique. Consider increasing trim.e.
```

When an error message is produced, the program will exit with no saved results and the user can modify the input parameters and re-start the program.

3.2 Repeated cross-section example

The main steps and the output structure are both similar to the time series case. In particular, the data can be loaded and the variables defined by running the following commands:

```
> data(driver)
> Driving_data<-driver
> y <- Driving_data[, "bac"]
> x <- Driving_data[, c("age", "gender", "winter")]
> N <- 108
```

Here the variable y is the dependent variable, x contains the three regressors, N is the number of entities in each time period. The data are organized such that the first N rows are for the first time period, the next N rows for the next time period, and so forth.

Other input parameters can be specified as follows, with the same interpretations as in the time series case:

```
> vec.tau = seq(0.70, 0.85, 0.05)
> trim.e <- 0.05
> vec.time <- unique(Driving_data[, "yq"])
> m.max <- 3
> v.a <-2
> v.b <-2
```

The function to run is:

```
> result <- rq.break(y, x, vec.tau, N, trim.e, vec.time, m.max, v.a, v.b)
```

We now highlight some outputs from the function. As in the time series case, the first part of the output is for analysis based on separate quantiles. The first quantile is $\tau = 0.7$, for which two breaks are detected, along with their confidence intervals:

```
$s.out$test_0.7
          1 Breaks 2 Breaks 3 Breaks
SQ test      5.179373 2.205150 1.339123
Critical values 1.574681 1.679331 1.737955
```

```
$s.out$nbreak_0.7
```

```
[1] 2
```

```
$s.out$br_est_0.7
```

	Estimate	CI_Lower_Bound	CI_Upper_Bound
Break 1	10	6	15
Break 2	38	35	39

```
$s.out$br_est_time_0.7
```

	Estimate	CI_Lower_Bound	CI_Upper_Bound
Break 1	"1985 Q2"	"1984 Q2"	"1986 Q3"
Break 2	"1992 Q2"	"1991 Q3"	"1992 Q3"

The breaks are consistent with the 1984 National Minimum Drinking Age Act and a 1991 beer tax increase. The output continues with the parameter estimates for the three regimes; we omit the details. After that, it continues with the next quantile, $\tau = 0.75$, and reports the same types of results.

The results for the multiple-quantile case are saved in

```
$m.out
```

```
$m.out$test_joint
```

	1 Breaks	2 Breaks	3 Breaks
DQ test	2.3734870	1.0105268	0.5995825
Critical values	0.7761262	0.8223711	0.8486304

```
$m.out$nbreak_joint
```

```
[1] 2
```

```
$m.out$br_est_joint
```

	Estimate	CI_Lower_Bound	CI_Upper_Bound
Break 1	9	5	13
Break 2	38	34	39

```
$m.out$br_est_joint_time
```

	Estimate	CI_Lower_Bound	CI_Upper_Bound
Break 1	"1985 Q1"	"1984 Q1"	"1986 Q1"
Break 2	"1992 Q2"	"1991 Q2"	"1992 Q3"

The rest of the output contains the coefficient estimates for each quantile followed by break size estimates at each quantile. We omit the details.

4 Conclusion

This paper has introduced the `QR.break` package, which implements a set of tools for detecting and estimating structural breaks in quantile regression models. The methods accommodate both time series and repeated cross-sectional data and allow researchers to work with a single quantile or a range of quantiles. Estimation is based on a dynamic programming algorithm, and confidence intervals for break dates are derived from analytical limiting distributions. The package also includes sequential testing procedures to determine the number of breaks.

There are several directions in which the current framework can be extended. Allowing break points to depend on covariates, adapting the methods to high-dimensional settings, incorporating partial structural breaks where only a subset of coefficients is allowed to change, or developing versions for panel data with fixed effects are all natural next steps. We hope this package provides a useful foundation for researchers working with quantile models in environments where structural stability cannot be taken for granted.

References

- Bai, J. (1995), “Least Absolute Deviation Estimation of a Shift,” *Econometric Theory*, 11, 403-436.
- Bai, J., and Perron, P.(2003), “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics*, 18, 1-22.
- Chernozhukov, V., and Umantsev, L. (2001), “Conditional Value-at-Risk: Aspects of Modeling and Estimation,” *Empirical Economics*, 26, 271-292.
- Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), “A Theory of the Term Structure of Interest Rates,” *Econometrica*, 53, 385-407.
- Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.
- Picard, D. (1985), “Testing and Estimating Change-Points in Time Series,” *Advances in Applied Probability*, 17, 841-867.
- Piehl, A. M., Cooper, S. J., Braga, A. A., and Kennedy, D. M. (2003), “Testing for Structural Breaks in the Evaluation of Programs,” *The Review of Economics and Statistics*, 85, 550-558.
- Oka, T., and Qu, Z. (2011), “Estimating Structural Changes in Regression Quantiles,” *Journal of Econometrics*, 162, 248–267.
- Qu, Z. (2008), “Testing for Structural Change in Regression Quantiles,” *Journal of Econometrics*, 146, 170-184.
- Su, L., and Xiao, Z. (2008), “Testing for Parameter Stability in Quantile Regression Models,” *Statistics and Probability Letters*, 78, 2768-2775.
- Yao, Y. C. (1987), “Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables,” *The Annals of Statistics*, 15, 1321-1328.