Testing and Dating of Structural Changes in Practice

Achim Zeileis*   Christian Kleiber+   Walter Krämer+   Kurt Hornik†

*Institut für Statistik & Wahrscheinlichkeitstheorie, Technische Universität Wien, Austria

+ Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, Germany

†Institut für Statistik, Wirtschaftsuniversität Wien, Austria

Abstract

The paper presents an approach to the analysis of data that contains (multiple) structural changes in a linear regression setup. We implement various strategies which have been suggested in the literature for testing against structural changes as well as a dynamic programming algorithm for the dating of the breakpoints in the R statistical software package. Using historical data on Nile river discharges, road casualties in Great Britain and oil prices in Germany it is shown that changes in the mean of a time series as well as in the coefficients of a linear regression are easily matched with identifiable historical, political or economic events.

Keywords: structural change, changepoint problem, segmented regressions, R, S, Bellman principle.

1 Introduction

There is a huge literature, both in mathematical statistics and various applications, on testing against changes in the coefficients of a linear regression model. The bibliography by Hackl and Westlund (1989) lists more than 500 references, and at least as many more have since appeared. Many classical tests such as the well-known Chow test assume that there is just a single change under the alternative or that the timing and the type of change are known. More recently, there has been a surge of interest in recovering the date of a shift if one has occurred or in methods which allow for several shifts at once, see Bai (1997); Hawkins (2001); Sullivan (2002); Bai and Perron (2002), among many others. The present paper summarizes

this more recent work and shows how to apply these methods using the R system for statistical computing, thereby a bit rebalancing this literature, which is rather long on theoretical insights, but often rather short on information on how these theories can be applied.

This is what we do below. To illustrate our software, we use three data sets. The first one is the Nile data from Cobb (1978) and later analyzed by Dümbgen (1991) and Balke (1993) which exhibits a level shift associated with the opening of the (first) Aswan dam at the end of the 19th century. The second data set is a time series of British road casualties analyzed by Harvey and Durbin (1986) which exhibits two breaks, one associated with increasing petrol prices in the wake of the first oil crisis, the second with the introduction of compulsory wearing of seatbelts in the early 1980s. Our third example—an index of the oil prices in Germany— was chosen in order to have a series with more than two breaks, in this case associated with events such as the first oil crisis, the Iranian revolution, and the virtual breakup of OPEC in 1985.

2 Model and methods

2.1 The model

We consider the standard linear regression model

$$y_i = x_i^\top \beta_i + u_i \qquad (i = 1, \ldots, n), \tag{1}$$

where at time i, y_i is the observation of the dependent variable, x_i is a $k \times 1$ vector of regressors, with the first component usually equal to unity, and β_i is the $k \times 1$ vector of regression coefficients, which may vary over time. The interpretation of i as time is the most common but not the only one: e.g., in changepoint analysis the observations are often ordered by the size of a variable not included in x_i.

This paper is concerned with testing the hypothesis that the regression coefficients remain constant

$$H_0 : \quad \beta_i \quad = \quad \beta_0 \qquad (i = 1, \ldots, n) \tag{2}$$

against the alternative that they vary over time. In many applications it is reasonable to assume that there are m breakpoints, where the coefficients shift from one stable regression relationship to a different one. Thus, there are $m + 1$ segments in which the regression

2

coefficients are constant, and the model (1) can be rewritten as

$$y_i = x_i^\top \beta_j + u_i \qquad (i = i_{j-1} + 1, \ldots, i_j, \quad j = 1, \ldots, m + 1), \tag{3}$$

where j is the segment index, $\mathcal{I}_{m,n} = \{i_1, \ldots, i_m\}$ denotes the set of the breakpoints ($\mathcal{I}_{m,n}$ is also called m-partition), and by convention $i_0 = 0$ and $i_{m+1} = n$.

In practice, the breakpoints are rarely given exogenously but have to be estimated from the data. This is what we do below. Our methodology is valid under fairly general assumptions on regressors and disturbances, see e.g. Krämer, Ploberger, and Alt (1988) or Bai (1997). Basically, they have to be such that a functional central limit theorem holds. This is for example satisfied if $\{u_i\}$ is a martingale difference with u_i independent of x_i and the regressors $\{x_i\}$ are (almost) stationary, which also allows for lagged dependent variables among the regressors.

2.2 The tests

Two frameworks for testing for structural change can be distinguished: (i) F statistics (Andrews 1993; Andrews and Ploberger 1994) that are designed for a specific alternative and (ii) generalized fluctuation tests (Kuan and Hornik 1995) that do not assume a particular pattern of deviation from the null hypothesis.

F statistics test against a single-shift alternative of unknown timing, i.e., model (3) with $m = 1$. Tests against this alternative are usually based on a sequence of F statistics for a change at time i: the OLS residuals $\hat{u}(i)$ from a segmented regression, i.e., one regression for each subsample, with breakpoint i, are compared to the residuals $\hat{u}$ from the unsegmented model via

$$F_i = \frac{\hat{u}^\top \hat{u} - \hat{u}(i)^\top \hat{u}(i)}{\hat{u}(i)^\top \hat{u}(i)/(n - 2k)}. \tag{4}$$

These F statistics are then computed for $i = n_h, \ldots, n - n_h$ ($n_h \geq k$) and H_0 is rejected if their supremum—or average or exp functional, see Andrews and Ploberger (1994)—is too large.

In applications, $n_h = \lfloor nh \rfloor$ will be a trimming parameter that can be chosen by the practitioner. In our examples below we use $h = 0.1$ or $h = 0.15$.

Bai and Perron (1998, 2002) extend this approach to F tests for 0 vs. ℓ breaks and ℓ vs. $\ell + 1$ breaks respectively with arbitrary but fixed ℓ.

The generalized fluctuation test framework "includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey .... Essentially, the techniques are designed to bring out departures from constancy in a graphic way instead of parametrizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives." (Brown, Durbin, and Evans 1975, pp. 149–150). More precisely, the model (1) is fitted to the data and an empirical process is derived that captures the fluctuation either in residuals or in parameter estimates. Under the null hypothesis these are governed by functional central limit theorems (see Kuan and Hornik 1995) and therefore boundaries can be found that are crossed with fixed probability α under the null hypothesis. Under the alternative the fluctuation in the process is in general increased. Also, the trajectory of the process often sheds light on the type of deviation from the null hypothesis such as the dating of the structual breaks.

As an example, consider the OLS-based CUSUM test introduced by Ploberger and Krämer (1992) which is based on cumulated sums of standard OLS residuals

$$W_n^0(t) = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \hat{u}_i \qquad (0 \le t \le 1). \tag{5}$$

The limiting process for $W_n^0(t)$ is the standard Brownian bridge $W^0(t) = W(t) - tW(1)$, where $W(\cdot)$ denotes standard Brownian motion; under a single-shift alternative the process should have a peak around the breakpoint. Another test which is used below is the fluctuation or recursive estimates (RE) test of Ploberger, Krämer, and Kontrus (1989), which compares recursive estimates of the regression coefficients with the full sample estimates. If the regression includes only an intercept, i.e., $x_i = 1$, it is equivalent to the OLS-based CUSUM test (5). Additional tests from this framework are the classical CUSUM test of Brown et al. (1975) based on recursive residuals, the MOSUM tests (Chu, Hornik, and Kuan 1995a) and the moving estimates (ME) test (Chu, Hornik, and Kuan 1995b).

2.3 Dating structural changes

Given an m-partition $i_1, \ldots, i_m$ the least squares estimates for the β_j can easily be obtained. The resulting minimal residual sum of squares is given by

$$RSS(i_1, \ldots, i_m) = \sum_{j=1}^{m+1} rss(i_{j-1}+1, i_j), \tag{6}$$

where $rss(i_{j-1} + 1, i_j)$ is the usual minimal residual sum of squares in the jth segment. The problem of dating structural changes is to find the breakpoints $\hat{i}_1, \ldots, \hat{i}_m$ that minimize the objective function

$$(\hat{i}_1, \ldots, \hat{i}_m) \quad = \quad \text{argmin}_{(i_1, \ldots, i_m)} RSS(i_1, \ldots, i_m) \tag{7}$$

over all partitions $(i_1, \ldots, i_m)$ with $i_j - i_{j-1} \geq n_h \geq k$.

Obtaining the global minimizers in (7) by an extensive grid search would be of order $O(n^m)$ and computationally burdensome for $m > 2$ (and any reasonable sample size n). Therefore, many hierarchical algorithms have been proposed that do recursive partitioning or joining of subsamples, see e.g. Bai (1997) or Sullivan (2002), but these will not necessarily find the global minimizers. These can be found much easier by a dynamic programming approach that is of order $O(n^2)$ for any number of changes m. Hawkins (2001) discusses this for changepoint problems in a maximum likelihood framework; Bai and Perron (2002) present a version of that algorithm for pure and partial structural change models in an OLS regression context, which we adopt here. The basic idea is that of Bellman's principle: the optimal segmentation satisfies the recursion

$$RSS(\mathcal{I}_{m,n}) \quad = \quad \min_{mn_h \leq i \leq n-n_h} [RSS(\mathcal{I}_{m-1,i}) + rss(i+1, n)]. \tag{8}$$

Therefore it suffices to know for each point i the "optimal previous partner" if i was the last breakpoint in an m-partition. This can be derived from a triangular matrix of $rss(i, j)$ with $j - i \geq n_h$, the computation of which is again made easier by the recursive relation $rss(i, j) = rss(i, j - 1) + r(i, j)^2$, where $r(i, j)$ is the recursive residual at time j of a sample starting at i (Brown et al. 1975). For more details on this dynamic programming algorithm see Bai and Perron (2002).

3 Software

We have incorporated all tests described above in the package strucchange in the R system[1] for statistical computing, the GNU implementation of the S language. The package can be downloaded from the Comprehensive R Archive Network (CRAN) at http://cran.R-project.org/ and is described in detail in Zeileis, Leisch, Hornik, and Kleiber

---

[1] http://www.R-project.org/

([2002](#)). Here we show how to use this package to test against and to date structural changes with unknown timing and multiplicity.

For instance, the e̲mpirical f̲luctuation p̲rocesses from the generalized fluctuation test framework can be obtained via the function

```
efp(formula, data, type, ...)
```

where `formula` defines the regression model (1) to be tested, e.g., `y ~ x`. This is a version of the notation introduced by Wilkinson and Rogers (1973), as adapted for S (Chambers and Hastie 1992). The argument `data` is a data frame that might contain the variables `y` and `x` and the argument `type` specifies the type of fluctuation process that should be fitted, e.g., `"OLS-CUSUM"` to fit the OLS-based CUSUM process from (5). The empirical fluctuation process object returned by `efp` can then be plotted together with its boundaries using the function `plot` and the corresponding significance test can be carried out with the function `sctest` (s̲tructural c̲hange te̲st).

Similarly, a sequence of F statistics can be computed with

```
Fstats(formula, data, cov.type, from = 0.15, ...)
```

where `from` specifies the trimming parameter h. The `cov.type` argument allows to calculate the F statistics based on heteroskedasticity robust covariance matrix estimates (the default is spherical errors). The returned object can again be plotted together with its boundaries, and formal significance tests for the sup$F$, ave$F$ and exp$F$ tests can be performed.

If there is evidence for structural changes in the regression relationship these can be dated with the function

```
breakpoints(formula, data, breaks, h = 0.15, ...)
```

which implements the dynamic programming algorithm described above. In particular, it computes the triangular $rss(i, j)$ matrix. The parameter `breaks` is the number of breakpoints m, the default being the largest number allowed by the trimming parameter h. From the object returned by this function any other number of breakpoints m can be extracted (as it contains the triangular $rss(i, j)$ matrix) by another application of the function `breakpoints` to this object. This will be described in more detail in the applications below.

4 Applications

4.1 The Nile data

First, we apply the above methods to a time series of the annual flow of the river Nile at Aswan from 1871 to 1970 (Cobb 1978; Dümbgen 1991; Balke 1993). It measures annual discharge at Aswan in 10^8 m^3 and is depicted in Figure 1.
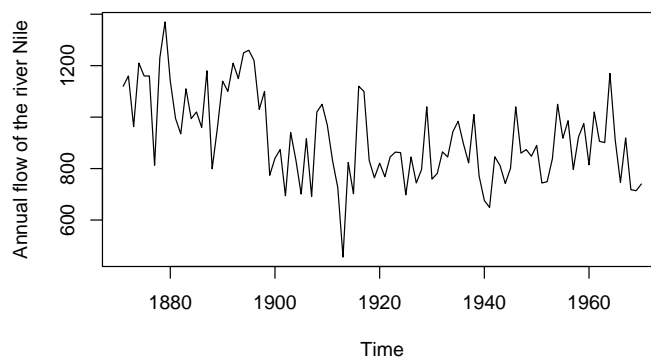

Figure 1: The Nile data

We test whether the mean of the annual flow changes over time, i.e., we fit a constant to the data. In the S language this is written as `Nile ~ 1`, where `Nile` is a time series object containing the data. To check for structural changes in this model we first use the OLS-based CUSUM process from (5). The code below shows how to fit this fluctuation process and to produce the plot shown in Figure 2; it gives the process together with its boundaries at a 5% significance level.

The process has a peak around 1900 which exceeds the boundaries and hence indicates a clear structural shift at that time. The obvious reason is the Aswan dam that was built in 1898.

The same conclusion emerges from tests based on F statistics, as shown in Figure 3. The code below computes the F statistics and sets up a plot of the resulting process together with the boundaries corresponding to a sup$F$ test at the 5% significance level.

From this sequence of F statistics the optimal breakpoint for a 2-segment partition can be obtained as it is equivalent to maximize the F statistics (4) or to minimize the residual sum

7

```
R> ocus.nile <- efp(Nile ~ 1, type = "OLS-CUSUM")
R> plot(ocus.nile)
```
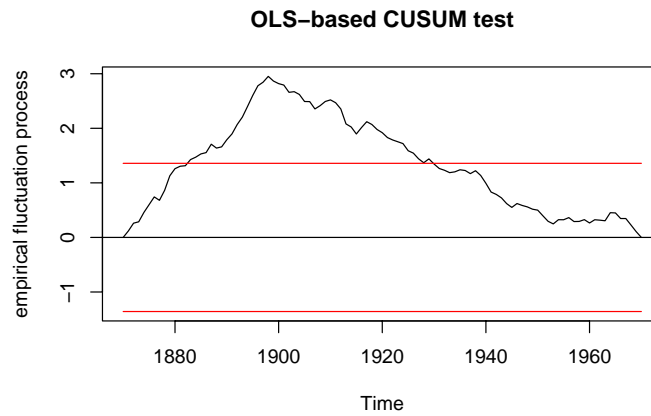
OLS–based CUSUM test

Figure 2: OLS-based CUSUM process for the Nile data

of squares (7). This breakpoint estimate can easily be obtained by `breakpoints(fs.nile)`. Although a 2-segment model seems quite intuitive for these data we also compare it to models with additional breakpoints. The following command computes arbitrary m-segment models
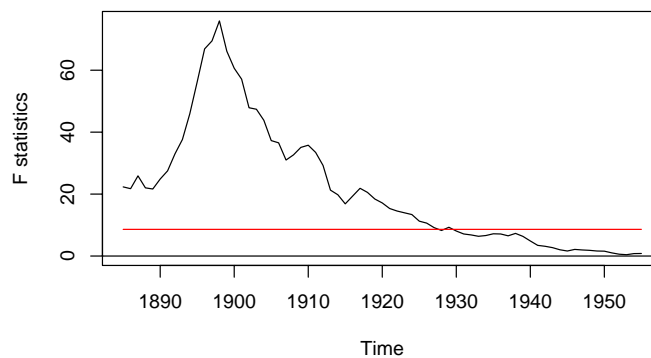
```
R> fs.nile <- Fstats(Nile ~ 1)
R> plot(fs.nile)
```


Figure 3: F statistics for the Nile data

based on the $rss(i, j)$ triangular matrix (with the default trimming of $h = 0.15$):

```
R> bp.nile <- breakpoints(Nile ~ 1)
```
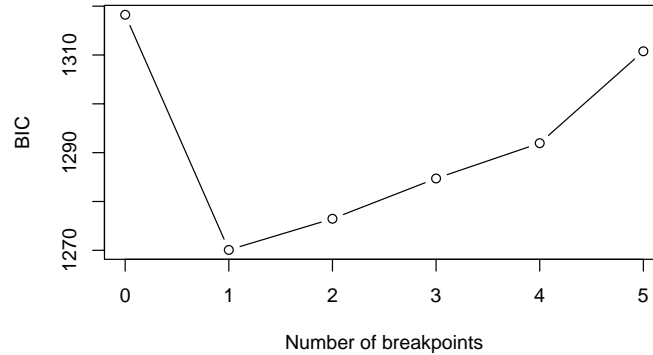

Figure 4: BIC for models with m breakpoints

A `summary` of this object reports the breakpoints for $m + 1$-segment models with $m = 0, \ldots, 5$ (the maximum possible with $h = 0.15$) as well as the associated RSS and BIC. Such information criteria are often used for model selection, which in this case means selection of the number m of breakpoints. Bai and Perron (2002) argue that the AIC usually overestimates the number of breaks but that the BIC is a suitable selection procedure in many situations. For the Nile data, Figure 4 shows that the BIC selects a model with $m = 1$ breakpoint, which confirms the results of the previous tests. The breakpoint for this model is observation 28 – or equivalently the year 1898 – and can be extracted by

```
R> bp1 <- breakpoints(bp.nile, breaks = 1)
```

To summarize our results we fit two linear models to the data. The first model `fm0` is the model under the null hypothesis without any breaks and `fm1` is the estimated 2-segment model. The factor `nile.fac` is a suitable coding of the partition implied by the estimated breakpoint.

```
R> fm0.nile <- lm(Nile ~ 1)
R> coef(fm0.nile)
```

9

```
(Intercept)

     919.35


R> nile.fac <- breakfactor(bp1)

R> fm1.nile <- lm(Nile ~ nile.fac - 1)

R> coef(fm1.nile)


nile.facsegment1 nile.facsegment2

       1097.7500          849.9722
```

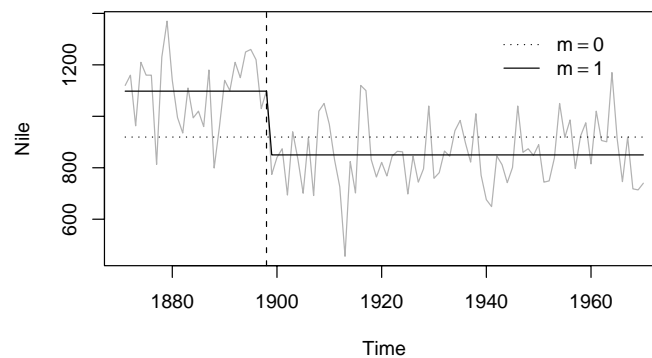The results can also be visualized as in Figure 5, which shows the fitted model for $m = 0, 1$.

Figure 5: Fitted models for the Nile data

4.2 The seatbelt data

This section analyzes a monthly time series (from 1969(1) to 1984(12)) of the number of car drivers in Great Britain killed or seriously injured in traffic accidents (Harvey and Durbin 1986). An appropriate model for analyzing this time series in a least squares framework is to take the logarithm of the data and regress it on its lagged values at lag 1 and 12. This corresponds to a multiplicative SARIMA$(1, 0, 0)(1, 0, 0)_{12}$ model fitted by OLS. Another possibility would be to take first differences instead of logs which leads to very similar results, with slightly inferior fits.
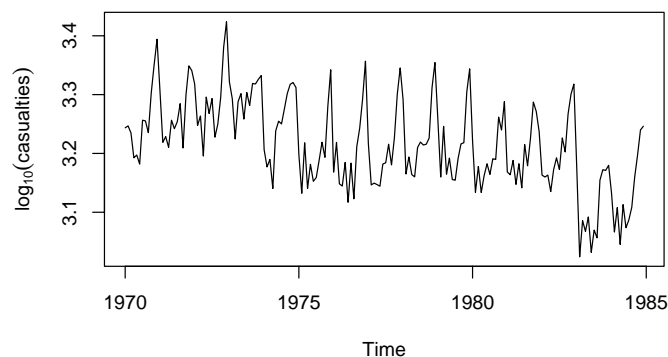
10

Figure 6: The seatbelt data

The log-transformed series is depicted in Figure 6.

To test for structural changes we use the fluctuation test (RE test) and F statistics with a trimming parameter of $h = 0.1$. Both processes are fitted and plotted in the code given below and the results are shown in Figure 7.

Both processes show significant departures from the null at a 5% level, indicating at least one break in the data. But both processes have two clear peaks, the first in 1973(10) and the second in 1982(12) and 1983(1) respectively, and favour a model with $m = 2$ breaks. To date the structural changes in this regression model we again estimate the breakpoints, with a trimming parameter of $h = 0.1$ and a maximum of $m = 5$ breaks.

```
R> bp.seat <- breakpoints(y ~ ylag1 + ylag12, data = seatbelt, h = 0.1,
+       breaks = 5)
```

The BIC as shown in Figure 8 for the models with $m = 0, \ldots, 5$ breakpoints would choose a model without any breaks, even though all the structural change tests above indicate at least one break. This might be caused by the fact that lagged regressors are included in which case the BIC might perform badly as pointed out by Bai and Perron (2002).

Due to the two peaks in the recursive estimates process and the sequence of F statistics respectively we decide in favour of a model with two breakpoints:

```
R> bp2 <- breakpoints(bp.seat, breaks = 2)
```

```
R> re.seat <- efp(y ~ ylag1 + ylag12, data = seatbelt, type = "fluctuation")
R> plot(re.seat)
R> fs.seat <- Fstats(y ~ ylag1 + ylag12, data = seatbelt, from = 0.1)
R> plot(fs.seat, main = "supF test")
```
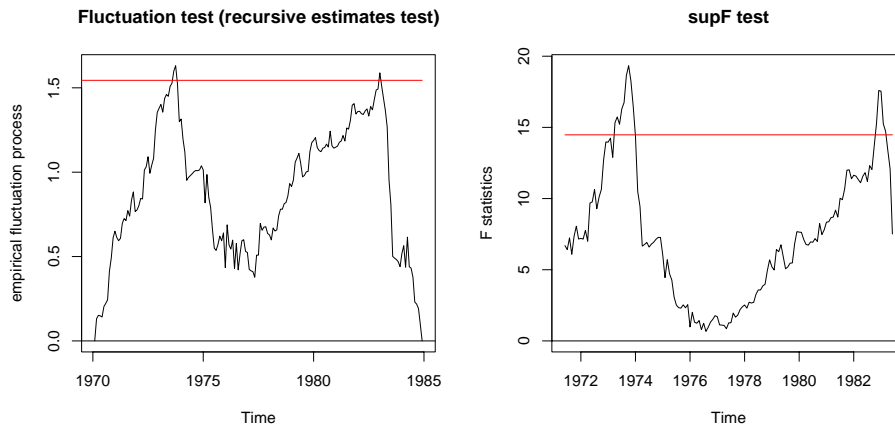

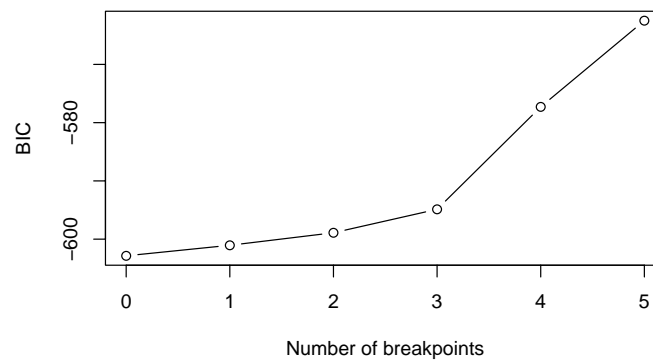Figure 7: Structural change tests for the seatbelt data

Figure 8: BIC for models with m breakpoints

The optimal breakpoints are then 1973(10)—associated with petrol rationing and the introduction of lower speed limits during the first oil crisis, see also the time series of oil prices in the following subsection—and 1983(1)— a level shift associated with the seat belt law introduced in the UK on 1983-01-31 (see Harvey and Durbin 1986, for further details).

The fitted dependent variable with the two breaks can be seen in Figure 9.

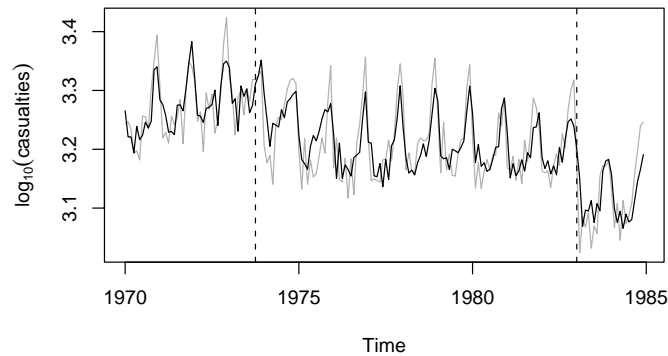Figure 9: Fitted model for the seatbelt data

4.3 The oil price data

This section considers a quarterly index of import prices of petroleum products – hereafter referred to as the oil price data – from 1960(1) to 1994(4) (base year: 1991). The data was obtained from the Statistisches Bundesamt Deutschland (Federal Statistical Office, Germany) and is given in Figure 10 (in logs).

Since there are obvious shifts in the mean, all structural change tests are highly significant. We therefore omit all details and directly proceed to estimating the breakpoints (with the default trimming of $h = 0.15$).

```
R> bp.oil <- breakpoints(log(OilPrice) ~ 1)
```

Again, a summary of this object would give information about the estimated breakpoints and the associated RSS and BIC of partitions with $m = 0, \ldots, 5$ breakpoints. For illustration,
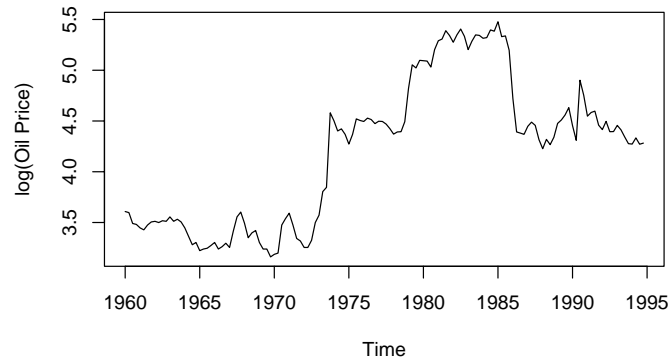
Figure 10: The oil price data

Figure 11 depicts the BIC, which is almost identical for 3 and 4 breaks. Hence, we first extract the segmentation with 3 breaks

```
R> bp3 <- breakpoints(bp.oil, breaks = 3)
```
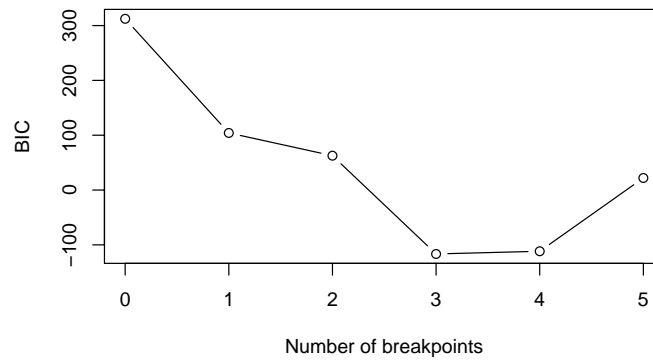

Figure 11: BIC for models with m breakpoints

and then we use the OLS-based CUSUM for checking for additional breaks in the mean. The CUSUM process in Figure 12 exhibits various peaks, the highest of which is in 1964, but none of these seems to be too extreme. So we stick to the model with 3 breakpoints which are 1973(3), 1979(1) and 1985(4). The fitted model is shown in Figure 13.

14

```
R> ocus.oil <- efp(log(OilPrice) ~ breakfactor(bp3), type = "OLS-CUSUM")
R> plot(ocus.oil)
```
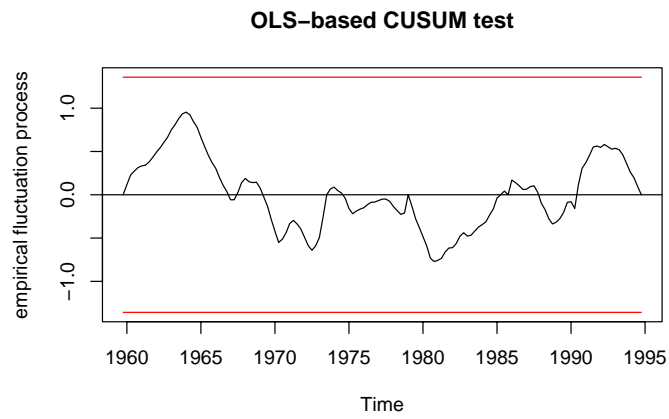
OLS−based CUSUM test

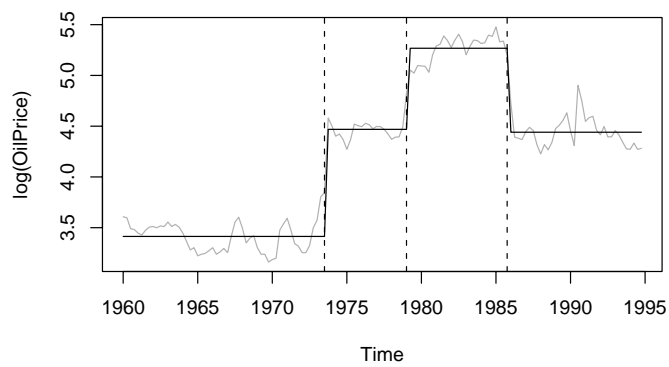Figure 12: OLS-based CUSUM process for the oil price data

Figure 13: Fitted model for the oil price data

The period 1973(3), for quarterly data, is essentially the same breakpoint as 1973(10) for the monthly seatbelt data, which corresponds to the Arab oil embargo after the Yom Kippur war. 1979(1) marks the start of the Iranian revolution followed by the war between Iran and Iraq. The break in 1985(1) is, in hindsight, a joint product of various minor events such as a worldwide slowdown of demand, the entering of Great Britain, Norway and Mexiko as major suppliers in international oil markets, and internal quarrels in the OPEC cartel, which led Saudi Arabia to increase its production and to abandon its role as the "residual supplier" which had until then softened all shocks in demand.

5 Outlook and summary

We have shown how recent methodological advances in testing against and dating multiple structural changes can be applied to "real" data. These methods are implemented in the R package `strucchange`: it allows for visualization and graphical analysis of empirical fluctuation processes and sequences of F statistics which often convey information about the presence and location of breakpoints in the data. In addition, it provides formal significance tests and a dynamic programming algorithm for computing breakpoint estimates that are global minimizers of the residual sum of squares. Although our empirical results are rather encouraging, they should not conceal that the difficulties with BIC-based model selection for one of our data sets suggest that the problem of determining the number of breakpoints in changepoint analysis deserves further study. This is currently under investigation.

Acknowledgements

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 Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics*, 79, 551–563.

Bai, J. and Perron, P. (1998), "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica*, 66, 47–78.

— (2002), "Computation and Analysis of Multiple Structural Change Models," Forthcoming.

Balke, N. S. (1993), "Detecting Level Shifts in Time Series," *Journal of Business and Economic Statistics*, 11, 81–92.

Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society*, B 37, 149–163.

Chambers, J. M. and Hastie, T. J. (1992), *Statistical Models in S*, London: Chapman & Hall.

Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995a), "MOSUM Tests for Parameter Constancy," *Biometrika*, 82, 603–617.

— (1995b), "The Moving-Estimates Test for Parameter Stability," *Econometric Theory*, 11, 669–720.

Cobb, G. W. (1978), "The Problem of the Nile: Conditional Solution to a Change-Point Problem," *Biometrika*, 65, 243–251.

Dümbgen, L. (1991), "The Asymptotic Behavior of Some Nonparametric Change-Point Estimators," *Annals of Statistics*, 19, 1471–1495.

Hackl, P. and Westlund, A. H. (1989), "Statistical Analysis of "Structural Change": An Annotated Bibliography," in *Econometrics of Structural Change*, ed. Krämer, W., Heidelberg: Physica-Verlag, pp. 103–128.

Harvey, A. C. and Durbin, J. (1986), "The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling (with Discussion)," *Journal of the Royal Statistical Society*, A 149, 187–227.

Hawkins, D. M. (2001), "Fitting Multiple Change-Point Models to Data," *Computational Statistics & Data Analysis*, 37, 323–341.

Krämer, W., Ploberger, W., and Alt, R. (1988), "Testing for Structural Change in Dynamic Models," *Econometrica*, 56, 1355–1369.

Kuan, C.-M. and Hornik, K. (1995), "The Generalized Fluctuation Test: A Unifying View," *Econometric Reviews*, 14, 135–161.

Ploberger, W. and Krämer, W. (1992), "The CUSUM Test With OLS Residuals," *Econometrica*, 60, 271–285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989), "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

Sullivan, J. H. (2002), "Estimating the Locations of Multiple Change Points in the Mean," *Computational Statistics*, 17, 289–296.

Wilkinson, G. N. and Rogers, C. E. (1973), "Symbolic Description of Factorial Models for Analysis of Variance," *Applied Statistics*, 22, 392–399.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002), "`strucchange`: An R Package for Testing for Structural Change in Linear Regression Models," *Journal of Statistical Software*, 7, 1–38.